Kenneth A. **Ross**, Charles R. B. **Wright**

Matematyka dyskretna

Wydawnictwo Naukowe PWN

Oddajemy do rąk Czytelnika podręcznik matematyki dyskretnej — bardzo ważnego działu matematyki, który ma obecnie szerokie zastosowanie w informatyce. Podstawowy materiał wykładu jest zawarty w czterech pierwszych rozdziałach. Tematy przedstawione dalej to: zliczanie, grafy, algorytmy rekurencyjne, rachunek prawdopodobieństwa, struktury algebraiczne, rachunek predykatów i zbiory nieskończone. Zamierzeniem Autorów książki jest kształtowanie matematycznej dojrzałości Czytelnika, a stosowana przez nich metoda dydaktyczna już się sprawdziła w praktyce. W podręczniku Czytelnik znajdzie szereg przykładów i ćwiczeń do samodzielnego wykonania. Autorzy twierdzą, że materiał każdego paragrafu książki można opanować w ciągu jednego dnia.

Publikacja przeznaczona jest dla studentów pierwszych lat matematyki, informatyki i innych kierunków ścisłych na uniwersytetach, wyższych uczelniach pedagogicznych oraz wyższych uczelniach technicznych.

ISBN 83-01-12129-7

9 788301 121297

Wykaz oznaczeń

Wyróżnione zbiory

$\mathbb{B} = \{0, 1\}$, $\mathbb{B}^n$ **588**
$\mathfrak{M}_{m,n}$ (macierze wymiaru $m \times n$) **155**
$\mathbb{N}$ (liczby naturalne, czyli liczby całkowite nieujemne) **16**
$\mathbb{P}$ (liczby całkowite dodatnie) **16**
$\mathbb{Q}$ (liczby wymierne) **16, 183**
$\mathbb{R}$ (liczby rzeczywiste) **16**
Σ^*, Σ **20**
$\mathbb{Z}$ (wszystkie liczby całkowite) **16**
$\mathbb{Z}_p$ **189**
$[a, b]$, (a, b) itp. **18**

Oznaczenia dotyczące zbiorów

$a \in A$, $a \notin A$ **15**
A^c (dopełnienie, uzupełnienie zbioru) **27**
$A \backslash B$ **25**
$A \cup B$ (suma zbiorów) **25**
$A \cap B$ (przecięcie, iloczyn zbiorów) **25**
$A \oplus B$ **25**
$\mathcal{P}(S)$ (zbiór potęgowy) **19**
(s, t), $(s_1, \ldots, s_n)$ **33, 34**
$S \times T$, $S^2 = S \times S$ **32**
$S_1 \times S_2 \times \ldots \times S_n$, S^n **34**
$\bigcup_{k \in I} A_k$, $\bigcap_{k \in I} A_k$, $\bigcup_{k=1}^{m} A_k$ itp. **62**
$T \subseteq S$ **18**
$T \subset S$ **18**
$\{\,\} = \emptyset$ (zbiór pusty) **19**

Funkcje

χ_A (funkcja charakterystyczna) **42**
$\mathrm{Dom}(f)$ (dziedzina funkcji f) **37**
$f + g$, $f \cdot g$ **763**
$f \circ g$ (złożenie funkcji) **43**
$f(A)$ **52**
$f : S \to T$ **38**
f^{-1} (funkcja odwrotna) **49**
$f^{\leftarrow}(B)$, $f^{\leftarrow}(y)$ **53**
$\mathrm{FUN}(S, T)$ **276**
wykres(f) **38**
1_S (funkcja identycznościowa) **42**
$\mathrm{Im}(f)$ (przeciwdziedzina, zbiór wartości funkcji f) **37**
$\log x$, $\ln x$ **50**
(s_n) (ciąg) **59**

Inne oznaczenia

$a := b$ **202**
$\lfloor x \rfloor$ (część całkowita) **188**
$n!$ (silnia) **58**
$|x|$ (wartość bezwzględna) **38**
$m \equiv n \pmod{p}$ **137**
$a +_p b$, $a *_p b$ **192**
$a * b$ (iloczyn) **449–450**
$a^{\wedge} b$ (a^b) **450**
$n \,\mathrm{DIV}\, p$, $n \,\mathrm{MOD}\, p$ **189**
$\mathrm{NWD}(m, n)$ **257**
$O(n^2)$, $O(n \log n)$ itp. **68–69**
$\prod$ (iloczyn) **58**
$\sum$ (suma) **58**
$\approx$ (wartość przybliżona) **60**
∞, $-\infty$ **19**
■ **18**

Logika

$\neg p$ (negacja) **89**
$p|q$ (kreska Sheffera) **121**
$p \wedge q$, $p \vee q$ (i, lub) **89–90**
$p \to q$ (implikacja) **90**
$P \Rightarrow Q$ **96**
$p \leftrightarrow q$ (równoważność) **92**
$P \Leftrightarrow Q$ **94**
$p \oplus q$ (alternatywa wykluczająca) **99**
t, c **95**
$\forall$, $\exists$ **83**
$\exists!$ **798**

Σ^*

λ (słowo puste) **21**
długość(w) **22**
Σ (alfabet) **20**
Σ^* (zbiór słów) **20**
Σ^k (zbiór słów długości k) **61**
$\overleftarrow{w}$ (odwrócenie słowa) **411**
$w_1 w_2$ (konkatenacja, iloczyn słów) **761**

Macierze

$\mathbf{A} = [a_{jk}]$ **154**
$\mathbf{A}[j, k] = a_{jk}$ **154**
$\mathbf{A}^{-1}$ (macierz odwrotna do $\mathbf{A}$) **172**
$\mathbf{A}^T$ (macierz transponowana) **155**

$\mathbf{A} + \mathbf{B}$ (suma macierzy) **156**
$\mathbf{AB}$ (iloczyn macierzy) **167**
$c\mathbf{A}$ (iloczyn macierzy przez skalar) **158**
$-\mathbf{A}$ (macierz przeciwna do $\mathbf{A}$) **157**
$\mathbf{I}$, $\mathbf{I}_n$ (macierz jednostkowa) **170**
$\mathfrak{M}_{m,n}$ (macierze wymiaru $m \times n$) **155**
$\mathbf{0}$ (macierz zerowa) **157**

Grafy i drzewa

$V(G)$, $E(G)$ **143, 147**
$\deg(v)$, $\mathrm{indeg}(v)$, $\mathrm{outdeg}(v)$ **333, 483**
$D_k(G)$ **333**
F (rezerwa czasowa) **499**
$G\backslash\{e\}$ **331**
$G \simeq H$ (grafy izomorficzne) **332**
K_n (graf pełny) **334**
$K_{m,n}$ **379**
M (waga maksymalna) **496**
$\mathbf{M}_R$ **159**
$R(v)$ **483**
$\mathrm{NAST}(v)$, $\mathrm{DOST}(v)$ **437, 480**
T_r, T_v (drzewa z wyróżnionym korzeniem) **363, 364**
W (waga) **492**
W^* (waga minimalna) **492**
$W(G)$ (waga grafu) **385**
$W(T)$ (waga drzewa) **459**

Zliczanie i prawdopodobieństwo

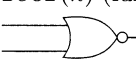
$\binom{n}{r}$ **277–278**
$P(n,r)$ **278**
$|S|$ **33**
Ω (przestrzeń zdarzeń elementarnych) **283**
$E(X) = \mu$ (wartość oczekiwana) **556**
f_X (rozkład prawdopodobieństwa) **547**
F_X (dystrybuanta) **548**
$P(E)$ (prawdopodobieństwo zdarzenia E) **284**
$P(E|S)$ (prawdopodobieństwo warunkowe) **529**
$P(X = 2)$ itp. **543**
σ (odchylenie standardowe) **565**
$V(X) = \sigma^2$ **565**
$\tilde{X}$, $\tilde{F}$ (zmienne unormowane) **578**
Φ (dystrybuanta rozkładu normalnego) **579**

Relacje

R_f (dla funkcji f) **139**
$R^{\leftarrow}$ (relacja odwrotna) **138**
$f^{\leftarrow}$ (jako relacja) **138**
$\sim$ (równoważność) **176**
$[s]$ (klasa równoważności, klasa abstrakcji) **179**
$[S]$ **180**
$\preceq$, $(S, \preceq)$ **636**
$\max(S)$, $\min(S)$ **643**
$\sup(S)$, $\inf(S)$ **643**
$x \vee y$, $x \wedge y$ **644**
$\mathrm{FUN}(S,T)$ **276**
$\preceq$ (na zbiorze $\mathrm{FUN}(S,T)$) **654**
$\preceq^k$ (porządek leksykograficzny) **656**
$\preceq^*$ (porządek standardowy) **657**
$\preceq_L$ (porządek leksykograficzny) **658**
E (relacja równości) **669**
$R_1 R_2 = R_2 \circ R_1$ **664**
$\mathbf{A}_1 * \mathbf{A}_2$ (iloczyn booleowski) **667–668**
R^0, R^n **669**
$\mathbf{A}_1 \leq \mathbf{A}_2$ (macierze booleowskie) **671**
$\mathbf{A}_1 \vee \mathbf{A}_2$, $\mathbf{A}_1 \wedge \mathbf{A}_2$ **672**
$z(R)$, $s(R)$, $p(R)$ **676**
$\mathbf{z}(\mathbf{A})$, $\mathbf{s}(\mathbf{A})$, $\mathbf{p}(\mathbf{A})$ **679–680**
$psz(R)$ itd. **681**

Systemy algebraiczne

$x \vee y$, $x \wedge y$ **589**
x' (dopełnienie) **589**
$x \leq y$ **592**
0, 1 **588**
$\mathbb{B}$, $\mathbb{B}^n$ **588**
$\mathrm{BOOL}(n)$ (funkcje booleowskie) **598**


itp. **613**

$\mathrm{PERM}(X)$ (permutacje) **687**
S_n (grupa symetryczna zbioru n-elementowego) **687**
(1 2 3) itd. **688**
$\langle g \rangle$ (grupa generowana przez g) **696**
$\langle A \rangle$ (podgrupa generowana przez A) **738**
$Gx = \{gx\colon g \in G\}$ **700**
$\mathrm{AUT}(D)$ (automorfizmy) **702**
$\mathrm{FIX}_G(x) = \{g \in G\colon g(x) = x\}$ **704**
$\mathrm{FIX}_X(g) = \{x \in X\colon g(x) = x\}$ **709**
$C(k)$ (kolorowania) **722**
g^* **714**
g^{-1} **734**
gH, Hg (warstwy) **740**
G/H (zbiór warstw lewostronnych) **742**
A^+ (półgrupa generowana przez A) **764**
R/I (pierścień ilorazowy) **778**

Matematyka dyskretna

Naszym żonom **Ruth** i **Leslie** za ich miłość i dodawanie otuchy

Kenneth A. **Ross**, Charles R. B. **Wright**

Matematyka dyskretna

Z języka angielskiego przełożyli
E. Sepko-Guzicka, W. Guzicki, P. Zakrzewski

Wydawnictwo Naukowe PWN Warszawa 1996

Dane oryginału
Discrete Mathematics
Kenneth A. Ross, Charles R. B. Wright
Third Edition

Copyright © 1992 by Prentice Hall Inc.
All rights reserved

Okładkę i strony tytułowe projektowali
Bokiewicz & Freudenreich

Redaktor: *Agnieszka Grabarczyk*

Redaktor techniczny: *Mariola Grzywacka*

Korekta: *Małgorzata Kopczyńska*

Tytuł dotowany przez Ministra Edukacji Narodowej

Copyright © for the Polish edition by
Wydawnictwo Naukowe PWN Sp. z o.o.
Warszawa 1996

ISBN 83-01-12129-7

Wydawnictwo Naukowe PWN Sp. z o.o.
Wydanie pierwsze
Ark. drukarskich 56,25
Druk ukończono w październiku 1996 r.
Skład i łamanie: ArtGraph, Warszawa
Druk i oprawa: Drukarnia Wydawnictw Naukowych S.A.
Łódź, ul. Żwirki 2

SPIS TREŚCI

Z przedmowy do trzeciego wydania 9

1. Zbiory, ciągi i funkcje 15

1.1. Niektóre szczególne zbiory 15
1.2. Działania na zbiorach 25
1.3. Funkcje 37
1.4. Funkcje odwrotne 48
1.5. Ciągi 57
1.6. Notacja O 64

2. Elementy logiki 77

2.1. Nieformalne wprowadzenie 78
2.2. Rachunek zdań 89
2.3. Metody dowodzenia 101
2.4. Rachunek zdań – ciąg dalszy 109
2.5. Analiza rozumowań 122

3. Relacje 135

3.1. Relacje 135
3.2. Grafy i grafy skierowane 142
3.3. Macierze 154
3.4. Mnożenie macierzy 165
3.5. Relacje równoważności i podziały zbioru 175
3.6. Algorytm dzielenia i zbiory $\mathbb{Z}_p$ 187

4. Indukcja i rekurencja 200

4.1. Niezmienniki pętli 201
4.2. Indukcja matematyczna 217
4.3. Definicje rekurencyjne 228
4.4. Zależności rekurencyjne 239
4.5. Więcej o indukcji 249
4.6. Algorytm Euklidesa 257

5. Zliczanie 272

5.1. Podstawowe techniki zliczania 272
5.2. Elementarny rachunek prawdopodobieństwa 282
5.3. Zasada włączeń i wyłączeń, metody dwumianowe 294
5.4. Zliczanie i podziały 304
5.5. Zasada szufladkowa Dirichleta 314

6. Wprowadzenie do grafów i drzew 327

6.1. Grafy 327
6.2. Zagadnienia związane z poruszaniem się po krawędziach . . 340
6.3. Drzewa 352
6.4. Drzewa z wyróżnionym korzeniem 360
6.5. Zagadnienia związane z przechodzeniem przez wierzchołki . 372
6.6. Minimalne drzewa spinające 382

7. Rekurencja, drzewa i algorytmy 398

7.1. Ogólna postać definicji rekurencyjnych i dowodów indukcyjnych 398
7.2. Algorytmy rekurencyjne 413
7.3. Algorytmy przeszukiwania w głąb 427
7.4. Notacja polska 449
7.5. Drzewa z wagami 459

8. Grafy skierowane 477

8.1. Grafy skierowane 477
8.2. Grafy skierowane z wagami 490
8.3. Algorytmy na grafach skierowanych 503
8.4. Modyfikacje i zastosowania algorytmów na grafach skierowanych 517

9. Rachunek prawdopodobieństwa 527

9.1. Niezależność 527
9.2. Zmienne losowe 542
9.3. Wartość oczekiwana i odchylenie standardowe 556
9.4. Rozkład dwumianowy i inne rozkłady z nim związane . . . 570

10. Algebry Boole'a 587

10.1. Algebry Boole'a 587
10.2. Wyrażenia booleowskie 601
10.3. Sieci logiczne 612
10.4. Tablice Karnaugha 624

11. Więcej o relacjach 634

11.1. Zbiory częściowo uporządkowane 634
11.2. Szczególne porządki 650

11.3. Ogólne własności relacji 663
11.4. Domknięcia relacji 675

12. Struktury algebraiczne 687

12.1. Permutacje 687
12.2. Działania grup na zbiorach 699
12.3. Działania grup na zbiorach, część 2 708
12.4. Zastosowania działań grup na zbiorach do problemów kolorowania 720
12.5. Grupy 733
12.6. Twierdzenie o izomorfizmie 747
12.7. Półgrupy 760
12.8. Inne systemy algebraiczne 771

13. Rachunek predykatów i zbiory nieskończone 790

13.1. Kwantyfikatory 790
13.2. Elementarny rachunek predykatów 799
13.3. Zbiory nieskończone 809

Odpowiedzi i wskazówki 822

Skorowidz 893

Z PRZEDMOWY DO TRZECIEGO WYDANIA

Pierwotnie planowaliśmy dodanie w trzecim wydaniu książki *Matematyka dyskretna* jedynie rozdziału poświęconego rachunkowi prawdopodobieństwa, aby uczynić niektóre trudniejsze paragrafy mniej abstrakcyjnymi i poprawić sposób przedstawienia indukcji i rekurencji. Kiedy zaczęliśmy realizować ten zamiar, stwierdziliśmy, że znaczniejsze zmiany w układzie treści mogłyby istotnie zwiększyć możliwości wykorzystania książki jako podręcznika, a także zebrać razem blisko ze sobą związane zagadnienia. W rezultacie wprowadziliśmy znaczne zmiany, do których należy zarówno dodanie istotnych partii nowego materiału, jak i jaśniejsze, jak sądzimy, przedstawienie i ułożenie na nowo zagadnień z drugiego wydania tej książki w taki sposób, by pasowały do większej niż przedtem liczby rozmaitych cykli wykładów.

Jednym z naszych głównych celów jest kształtowanie matematycznej dojrzałości Czytelnika. Materiał ten był przez wiele lat pomyślnie wykorzystywany przez nas i naszych kolegów w pracy z przeciętnymi początkującymi studentami matematyki i stwierdzaliśmy, że pod koniec drugiego semestru byli oni dobrze przygotowani do dalszej nauki na wyższych latach studiów. Na początku książki prezentowane zagadnienia przedstawiamy w sposób intuicyjny, zwiększając ścisłość wykładu w miarę jak studenci uczą się doceniać wartość dowodów i zdobywają umiejętność ich przeprowadzania. Nasze wyjaśnienia są staranne, lecz nieformalne. W miarę posuwania się naprzód pokazujemy, w jaki sposób matematycy dążą do rozwiązywania problemów i ukazujemy wartość podejścia abstrakcyjnego. Staraliśmy się przedstawić materiał na tyle prosto, by student był w stanie się go nauczyć, i na tyle kompletnie, by nie musiał później uczyć się go ponownie.

W pierwszych dwóch wydaniach stosowana była metoda dydaktyczna, polegąca na stopniowym przedstawianiu omawianych zagadnień na coraz głębszym poziomie, z licznymi odnośnikami do wcześniejszego materiału tak, by studenci byli w stanie dostrzec związki między różnymi pojęciami. Wadą tej, pod innymi względami rozsądnej, strategii było to, że wykład podstawowych zagadnień czyniła niepotrzebnie trudnym. W nowym układzie to, co uważamy za podstawowy materiał jakiegokolwiek poważnego wykładu matematyki dyskretnej, znalazło się w czterech pierwszych rozdziałach. Tematy te można przerobić z łatwością w ciągu połowy semestru. Wykład semestralny można by uzupełnić kombinatoryką i elementami probabilistyki lub też wybrać grafy, drzewa i algorytmy rekurencyjne. Wykładowcy, wykorzystujący tę książkę do prowadzenia rocznego wykładu, zechcą prawdopodobnie wykładać materiał w takiej kolejności, jaką my przyjęliśmy i przedstawiać rozdziały w takim porządku, w jakim pojawiają się one w książce.

Aby dołączyć nowy materiał, a zarazem utrzymać rozsądną objętość książki, zmuszeni byliśmy usunąć z niej pewne tematy, w tym takie, które lubiliśmy.

W rozdziale 1 znajduje się teraz wprowadzenie do funkcji i ciągów. Zawiera on też cały paragraf z w miarę gruntownym wstępem do notacji O. Mamy nadzieję, że paragraf ten będzie dla Czytelników przydatnym źródłem odwołań podczas studiowania tej książki.

W rozdziale 2 koncentrujemy się na logice elementarnej jako na przydatnym narzędziu. W stopniu większym nawet niż we wcześniejszych wydaniach książki pomniejszamy tu rolę dowodów formalnych, kładąc nacisk na metody dowodzenia i analizę rozumowań. Indukcja jest przeniesiona do rozdziału 4 z dwóch powodów. Po pierwsze, aby podać sensowne jej zastosowania musimy przyjąć, że Czytelnik zna pewne podstawowe fakty, podane w rozdziale 3. Po drugie, jednoczesne nauczanie indukcji i rekurencji w logicznie powiązany ze sobą sposób ma naprawdę sens. Kwantyfikatory w nieformalny sposób pojawiają sie w rozdziale 2; byłoby dobrze, żeby studenci do nich przywykli, gdyż dają wygodną możliwość skracania zapisu. Rozdział 13 zawiera więcej informacji na temat kwantyfikatorów oraz krótki wstęp do rachunku predykatów. Jak sugeruje jego umiejscowienie, tego bardziej formalnego ujęcia nie wykorzystuje się w pozostałej części książki.

W rozdziale 3, który zatytułowaliśmy po prostu „Relacje", zebrane są podstawowe narzędzia matematyki dyskretnej: relacje, grafy i grafy skierowane oraz macierze. Umożliwiają one spojrze-

nie na ten sam krąg pojęć z wielu różnych punktów widzenia. Starannie omówione są tu relacje równoważności, występujące w całej abstrakcyjnej matematyce; pokazane jest ich zastosowanie do arytmetyki modulo n w zbiorze $\mathbb{Z}$.

Większość tekstu poświęconego indukcji matematycznej wchodzi teraz w skład jednego rozdziału, przy czym nastąpiła zmiana w sposobie jej wprowadzenia. Rozdział 4 rozpoczyna się omówieniem pętli i niezmienników pętli, które nasi studenci uważają za naturalne i stosunkowo łatwe do zrozumienia. Materiał ten będzie później potrzebny do sprawdzania poprawności algorytmów. Stwierdziliśmy, że na wykładach z informatyki widać korzyści, jakie studenci odnoszą z wcześniejszego zapoznania się z czysto matematycznym przedstawieniem niezmienników. Indukcję matematyczną można przedstawić w taki sposób, by główne idee były zbliżone do tych związanych z niezmiennikami pętli; nasze uzasadnienie indukcji opieramy więc na pojęciu pętli. Takie podejście zdało u nas egzamin. Użytkowników tej książki prosimy o komentarze na temat ich doświadczeń w innych uczelniach. Tak czy inaczej łatwo jest na wykładzie przedstawić indukcję w jakiś bardziej tradycyjny sposób.

Rozdział 4 zawiera też wprowadzenie do rekurencyjnego definiowania ciągów i do relacji rekurencyjnych. Dodaliśmy krótką informację na temat równań rekurencyjnych typu „dziel i rządź". Nowy jest też ostatni paragraf tego rozdziału, poświęcony wariantom algorytmu Euklidesa. Algorytmy pojawiają się tu jako interesujące przykłady, jednakże same rezultaty tego paragrafu są wykorzystywane dopiero pod koniec rozdziału 12.

W rozdziale 5 jeden z tematów usunęliśmy, a jeden dodaliśmy. Paragraf dotyczący zbiorów nieskończonych zawiera fascynujący materiał, który jednakże nie jest potrzebny w dalszej części podręcznika. Został on przeniesiony do rozdziału 13. Dodaliśmy paragraf dotyczący rachunku prawdopodobieństwa, w którym wprowadzamy terminologię, a który ma stanowić przygotowanie do rozdziału 9.

Materiał dotyczący grafów i drzew w nowym układzie obejmuje trzy rozdziały. Rozsądny wstęp do tych zagadnień zawiera rozdział 6, po którym studiować można rozdział 7 bądź rozdział 8. Nowym materiałem w rozdziale 6 są pewne szybkie teoriografowe algorytmy. W rozdziale 7 nacisk położony jest teraz na rekurencję. Dwa pierwsze paragrafy zawierają definicje rekurencyjne i ogólne algorytmy rekurencyjne. Reszta tego rozdziału poświęcona jest algorytmom rekurencyjnym związanym z drzewami i ich ciekawym zastosowaniom. W rozdziale 8 omawiamy teraz wyłącznie

grafy skierowane i algorytmy z nimi związane.

W pierwszych dwóch wydaniach rozwijaliśmy teorię algebr Boole'a, opierając się na pewnych szczególnych systemach częściowo uporządkowanych. W odpowiedzi na uwagi Czytelników przedstawiamy teraz algebry Boole'a w rozdziale 10 w sposób czysto algebraiczny, kładąc nacisk na funkcje booleowskie. Zachowaliśmy ich zastosowania do układów logicznych i tablic Karnaugha.

Omówienie porządków częściowych w rozdziale 11 nie uległo większym zmianom w stosunku do wydania drugiego. Ponieważ szczegółowa dyskusja ogólnych własności relacji znajduje się teraz w rozdziale 3, sposób ich potraktowania w rozdziale 11 został uproszczony. Omówienie operatorów domknięcia jest teraz mniej abstrakcyjne niż w poprzednich wydaniach.

Rozdział 12, dotyczący struktur algebraicznych, został całkowicie przerobiony. Pierwsze cztery paragrafy są teraz poświęcone permutacjom i stosunkowo konkretnej sytuacji, w której grupa działa na zbiorze. Podajemy pewne nietrywialne zastosowania działań grup na zbiorach do problemów kolorowania. Reszta rozdziału dotyczy tradycyjnych tematów związanych z grupami i półgrupami i kończy się bardzo krótkim wprowadzeniem do teorii pierścieni, w tym pierścieni wielomianów oraz ciał.

Przestaliśmy kłaść nacisk na niektóre tematy bądź też je w ogóle pominęliśmy. Gdziekolwiek pojawia się prawo łączności, tam mamy do czynienia z półgrupą, ale aż do rozdziału 12 unikaliśmy terminologii dotyczącej półgrup. Półgrupa wolna Σ^*występuje w całej książce, choć nie jest tak nazywana, usunęliśmy natomiast paragraf dotyczący półgrupy $\mathcal{P}(\Sigma^*)$.

Utrzymaliśmy te charakterystyczne cechy książki, które, jak się wydaje, były udane w poprzednich jej wydaniach. Nadal jest bardzo dużo przykładów, z którymi studenci mogą zapoznać się samodzielnie, co daje prowadzącemu zajęcia możliwość poświęcenia czasu wybranym tematom i pozostawienia reszty materiału do przerobienia w domu. Poza paroma wyjątkami, każdy z paragrafów można nadal przerobić w ciągu jednego dnia.

Otrzymaliśmy cenne uwagi i wskazówki od wielu kolegów, w tym od następujących recenzentów:

Richard Makohon, University of Portland
John Chollet, Towson State University
Bo Green, Abilene Christian University
Richard H. Austing, University of Maryland
Charles Searcy, New Mexico Highlands University
Hyeong-Ah Choi, George Washington University

Pragniemy podziękować tym, którzy w wydawnictwie Prentice Hall przyczynili się do ukazania się tego wydania książki. Oto ich nazwiska: Priscilla McGeehon, Steve Conmy, Nick Romanelli.

Wierzymy, że obecne trzecie wydanie jest znacznie lepsze od pierwszych dwóch. Prosimy Czytelników o przesyłanie uwag na adres University of Oregon Department of Mathematics albo pocztą elektroniczną na adres ross@math.uoregon.edu lub wright@math.uoregon.edu.

Specjalnie do studentów

Wiemy, że słowa takie jak „oczywiście" i „jest jasne, że" mogą być bardzo denerwujące; nas też one czasem męczyły. Kiedy od czasu do czasu natkniesz się na nie w tej książce, przyjmij je jako wskazówki dla Ciebie. Jeśli dany fragment tekstu nie jest dla Ciebie oczywisty ani jasny, to prawdopodobnie komplikujesz zanadto daną sytuację bądź też robisz jakieś założenia, których w tekście nie ma. Zrób sobie przerwę; potem przeczytaj dany materiał raz jeszcze. Podobnie, przykłady są po to, by Ci pomóc. Jeśli jesteś całkiem pewien, że znasz niezbędne pojęcia i fakty, ale przykład wydaje Ci się zbyt trudny, to opuść go przy pierwszym czytaniu i wróć do niego później. Jeśli natomiast nie jesteś całkiem pewny, czy rozumiesz teorię, na której jest on oparty, to przeanalizuj go dokładniej.

Ważną częścią tej książki są ćwiczenia. Dają Ci one możliwość sprawdzenia, czy rozumiesz dany materiał oraz nabrania wprawy w jasnym matematycznym myśleniu i pisaniu. Posuwając się w głąb książki napotkasz coraz więcej ćwiczeń, polegających na dowodzeniu czegoś. Używamy zamiennie słów „wykaż" i „udowodnij", chociaż słowo „wykaż" częściej występuje wtedy, gdy dla uzyskania wyniku wystarczy dokonać pewnych obliczeń, natomiast słowo „udowodnij" sugeruje, że należy przeprowadzić jakieś rozumowanie. „Udowodnij" znaczy „podaj przekonywający argument bądź przeprowadź rozważania pokazujące, dlaczego dane twierdzenie jest prawdziwe". To, co piszesz, powinno być przekonywające dla osoby prowadzącej zajęcia, dla innego studenta, a także następnego dnia dla Ciebie samego. Dowody powinny zawierać słowa i zdania, a nie jedynie obliczenia — muszą umożliwić czytelnikowi śledzenie Twojego toku myśli. Używaj jako wzorców, zwłaszcza na początku, dowodów zamieszczonych w tej książce. Pomoże Ci też znajdujące się w rozdziale 2 omówienie dowodów logicznych. Doskonalenie umiejętności pisania „dobrych" dowodów jest podobne do doskonalenia umiejęt-

ności pisania „dobrych" esejów bądź przedstawiania „dobrych" referatów. Wszystkie one wymagają praktyki. Nie zniechęcaj się, jeśli jakiś Twój dowód nie przekona eksperta (powiedzmy, wykładowcy lub prowadzącego zajęcia). Postaraj się raczej stwierdzić, co nie było przekonywające.

Artykuł Sandry Z. Keith w numerze z maja 1991 roku pisma *Ume Trends*, poświęconego nauczaniu matematyki w szkole wyższej, daje następujące ogólne rady dotyczące studiowania:

„Załóż zeszyt do prac domowych, ponawiaj próby rozwiązania zadania, staraj się nie zaglądać do odpowiedzi, opracowuj notatki zrobione na zajęciach, przygotowuj streszczenia przerobionego materiału, wymieniaj się nimi z przyjaciółmi i studiuj je przed egzaminami nawet, jeśli nigdy nie robiłeś tego w szkole średniej!"

Do tych doskonałych rad chcielibyśmy dodać jeszcze jedną: czytaj z wyprzedzeniem. Przeglądaj dany materiał przed zajęciami, aby nabrać o nim jakiegoś pojęcia i zlokalizować trudne miejsca. Gdy na zajęciach dojdzie do omawiania tych trudnych punktów, śmiało będziesz mógł prosić o dodatkowe wyjaśnienia będąc pewnym, że nie marnujesz czasu zajęć na kwestie, które po lekturze książki stałyby się oczywiste. Jeśli jesteś przygotowany do zajęć, to możesz skorzystać z pomocy prowadzącego i zaoszczędzić sobie wiele wysiłku.

Każdy rozdział kończy się listą tego, co w nim było najważniejsze, wraz z sugestiami, jak tę listę wykorzystać do powtórzenia materiału. Jednym z najlepszych sposobów uczenia się rzeczy, które planujesz później wykorzystywać, jest powiązanie każdego nowego pojęcia i faktu z tak dużą liczbą znanych pojęć i sytuacji, jak to możliwe, i wyobrażenie sobie okoliczności, w jakich dany fakt może się przydać. Aby ułatwić ten proces, umieściliśmy w tekście bardzo dużo przykładów. Z list, służących do powtórzenia materiału, można korzystać w ten sam sposób zarówno pracując samodzielnie, jak i ucząc się z innymi studentami.

Na końcu książki znajdują się odpowiedzi i wskazówki do większości ćwiczeń o numerach nieparzystych. Rozsądni studenci będą zaglądać do odpowiedzi dopiero po podjęciu poważnej próby zrobienia ćwiczenia. W przypadku polecenia przeprowadzenia dowodu podajemy zwykle wskazówkę bądź szkic dowodu, który należy najpierw zrozumieć, a następnie rozwinąć.

Na wewnętrznych stronach okładki znajdują się: wykaz oznaczeń, alfabet grecki oraz nazwy algorytmów. Warto się z nimi zapoznać przed rozpoczęciem czytania rozdziału 1.

K.A. Ross, C.R.B. Wright

1. ZBIORY, CIĄGI I FUNKCJE

Jest to rozdział wstępny zawierający dość dużą ilość podstawowych definicji i oznaczeń. Większość materiału zawartego w tym rozdziale jest pewnie Czytelnikowi znana, chociaż prawdopodobnie zetknął się on z innymi oznaczeniami lub innym poziomem ścisłości. Ponadto, wprowadzając w tym rozdziale pojęcia i metody, ukazujemy styl wykładu, który będzie zachowany w dalszym ciągu książki.

§ 1.1. Niektóre szczególne zbiory

W ostatnich kilkudziesięciu latach stało się tradycją używanie teorii mnogości jako podstaw matematyki. To znaczy, że pojęcia „zbiór" i „należenie" są podstawowymi, niezdefiniowanymi terminami, a reszta matematyki jest zdefiniowana lub opisana za pomocą tych pojęć. Zbiór jest „kolekcją" obiektów; definicja zbioru musi być niedwuznaczna w tym sensie, że musi być jasne, czy dany konkretny obiekt należy do tego zbioru. Będziemy zazwyczaj oznaczać zbiory wielkimi literami, takimi jak A, B, S czy X. Obiekty są zazwyczaj oznaczane małymi literami, takimi jak a, b, s czy x. Obiekt a, który należy do zbioru S, jest nazywany **elementem** zbioru S. Jeśli a jest obiektem a A jest zbiorem, piszemy $a \in A$ mając na myśli, że a jest elementem zbioru A oraz piszemy $a \notin A$ dla zaznaczenia, że a nie jest elementem zbioru A. Znak $\in$ może być czytany jako „jest elementem zbioru" lub „należy do" lub „jest w" lub jako spójnik „w", w zależności od kontekstu.

Poszczególne zbiory mogą być zapisywane na wiele sposobów. Niektóre szczególnie często występujące i ważne zbiory będą miały swoje własne nazwy, tzn. swoje własne symbole. Rezerwujemy symbol $\mathbb{N}$ dla oznaczania zbioru **liczb naturalnych**:

$$\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, \ldots\}.$$

Zauważmy, że liczbę 0 zaliczyliśmy do liczb naturalnych.

Przez $\mathbb{P}$ oznaczamy zbiór **liczb całkowitych dodatnich**:

$$\mathbb{P} = \{1, 2, 3, 4, 5, 6, 7, \ldots\}.$$

W wielu tekstach matematycznych ten zbiór jest oznaczany symbolem $\mathbb{N}$. Zbiór wszystkich **liczb całkowitych**, dotatnich, zero czy ujemnych, będzie oznaczany przez $\mathbb{Z}$ (od niemieckiego słowa „Zahl"). Liczby postaci m/n, gdzie $m \in \mathbb{Z}$, $n \in \mathbb{Z}$ i $n \neq 0$, są nazywane **liczbami wymiernymi**. Zbiór wszystkich liczb wymiernych jest oznaczany przez $\mathbb{Q}$. Zbiór wszystkich **liczb rzeczywistych**, wymiernych czy niewymiernych, jest oznaczany przez $\mathbb{R}$. Zatem zbiór $\mathbb{R}$ zawiera wszystkie liczby należące do $\mathbb{Q}$, a także $\sqrt{2}, \sqrt{3}, \sqrt[3]{2}, -\pi, e$ i wiele, wiele innych liczb. Elementy małych zbiorów skończonych mogą być wypisane w nawiasach klamrowych, oddzielone przecinkami. Na przykład $\{2, 4, 6, 8, 10\}$ jest zbiorem złożonym z pięciu dodatnich liczb całkowitych parzystych, mniejszych od 12, a zbiór $\{2, 3, 5, 7, 11, 13, 17, 19\}$ składa się z ośmiu liczb pierwszych mniejszych od 20. Dwa zbiory są równe, jeśli mają te same elementy. Zatem

$$\{2, 4, 6, 8, 10\} = \{10, 8, 6, 4, 2\} = \{2, 8, 2, 6, 2, 10, 4, 2\};$$

kolejność wypisywania nie ma znaczenia i nic nie daje (ani nie szkodzi) wypisanie tego samego elementu więcej niż jeden raz.

Elementy dużych zbiorów skończonych, a nawet zbiorów nieskończonych, mogą być wypisane z użyciem matematycznego „i tak dalej", a mianowicie wielokropka, ..., przy założeniu, że znaczenie wielokropka jest jasne. Zatem $\{1, 2, 3, \ldots, 1000\}$ oznacza zbiór liczb całkowitych dodatnich mniejszych lub równych 1000 i zakładamy, że $\{3, 6, 9, 12, \ldots\}$ oznacza nieskończony zbiór liczb całkowitych dodatnich podzielnych przez 3. Z drugiej strony, znaczenie $\{1, 2, 3, 5, 8, \ldots\}$ może być nie do końca jasne. Takie nieprecyzyjne użycie wielokropka nie zawsze spełnia swoją rolę, szczególnie w informatyce; więc podamy metody jednoznacznego opisu takich zbiorów bez użycia wielokropka.

Zbiory są często opisywane przez opis własności swych elementów, za pomocą notacji

$$\{ : \quad \}.$$

Zmienna (na przykład n czy x) jest wypisana przed dwukropkiem, a własności podane są po dwukropku. Na przykład

$$\{n\colon n \in \mathbb{N} \text{ i } n \text{ jest parzyste}\}$$

oznacza zbiór nieujemnych liczb całkowitych parzystych, tzn. zbiór $\{0, 2, 4, 6, 8, 10, \ldots\}$. Dwukropek jest zazwyczaj czytany jako „takich, że", a więc powyższy napis czytamy „zbiór wszystkich n takich, że n należy do $\mathbb{N}$ i n jest parzyste". Podobnie

$$\{x\colon x \in \mathbb{R} \text{ i } 1 \leq x < 3\}$$

oznacza zbiór wszystkich liczb rzeczywistych, które są większe lub równe 1 i mniejsze od 3. Liczba 1 należy do tego zbioru, ale 3 nie należy. Upraszczając nieco notację, ostatnie dwa zbiory można zapisać jako

$$\{n \in \mathbb{N}\colon n \text{ jest parzyste}\} \text{ i } \{x \in \mathbb{R}\colon\ 1 \leq x < 3\}.$$

Pierwszy zbiór czytamy „zbiór wszystkich n należących do $\mathbb{N}$, takich że n jest parzyste".

Innym sposobem wyszczególnienia elementów zbioru jest podanie reguły pozwalającej otrzymać te elementy z elementów innego zbioru. Na przykład $\{n^2\colon n \in \mathbb{N}\}$ oznacza zbiór wszystkich liczb całkowitych, które są kwadratami liczb całkowitych ze zbioru $\mathbb{N}$, tzn.

$$\begin{aligned}\{n^2\colon n \in \mathbb{N}\} &= \{m \in \mathbb{N}\colon m = n^2 \text{ dla pewnego } n \in \mathbb{N}\}\\ &= \{0, 1, 4, 9, 16, 25, 36, \ldots\}.\end{aligned}$$

Zauważmy, że ten zbiór jest równy zbiorowi $\{n^2\colon\ n \in \mathbb{Z}\}$. Podobnie, $\{(-1)^n\colon n \in \mathbb{N}\}$ oznacza zbiór, który otrzymujemy obliczając wartości $(-1)^n$ dla wszystkich $n \in \mathbb{N}$, a więc

$$\{(-1)^n\colon n \in \mathbb{N}\} = \{-1, 1\}.$$

Zbiór ten ma tylko dwa elementy.

Rozważmy teraz dwa zbiory S i T. Mówimy, że S jest **podzbiorem** zbioru T, jeśli każdy element zbioru S należy do zbioru T. Jeśli zbiór S jest podzbiorem zbioru T, piszemy $S \subseteq T$. Symbol $\subseteq$ czytamy jako „jest podzbiorem". Często też mówimy, że „S jest zawarty w T", w przypadku, gdy $S \subseteq T$, ale trzeba zwrócić uwagę na możliwość nieporozumienia, jeśli powiemy także „x jest zawarty w T", mając na myśli „$x \in T$". Zawieranie się zbiorów i zawieranie elementów oznaczają zupełnie różne rzeczy.

Dwa zbiory S i T są **równe**, jeśli mają dokładnie te same elementy. Zatem $S = T$ wtedy i tylko wtedy, gdy $S \subseteq T$ i $T \subseteq S$.

PRZYKŁAD 1 (a) Mamy $\mathbb{P} \subseteq \mathbb{N}$, $\mathbb{N} \subseteq \mathbb{Z}$, $\mathbb{Z} \subseteq \mathbb{Q}$, $\mathbb{Q} \subseteq \mathbb{R}$. Tak jak w przypadku znanej relacji nierówności $\leq$, możemy zapisać te stwierdzenia razem

$$\mathbb{P} \subseteq \mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}.$$

(b) Ponieważ 2 jest jedyną parzystą liczbą pierwszą, mamy

$$\{n \in \mathbb{P}\colon n \text{ jest liczbą pierwszą i } n \geq 3\} \subseteq \{n \in \mathbb{P}\colon n \text{ jest liczbą nieparzystą}\}.$$

(c) Weźmy znów dowolny zbiór S. Oczywiście, jeśli $x \in S$, to $x \in S$, a więc $S \subseteq S$. To oznacza, że dowolny zbiór jest swoim własnym podzbiorem. To właśnie dlatego używaliśmy oznaczenia $\subseteq$, a nie $\subset$. To oznaczenie jest podobne do oznaczenia $\leq$ dla liczb rzeczywistych. Nierówność $x \leq 5$ jest prawdziwa dla wielu liczb, takich jak na przykład $3,1$ oraz -73. Jest ona też prawdziwa dla $x = 5$, tzn. $5 \leq 5$. Ta ostatnia nierówność wygląda trochę dziwnie, ponieważ tak naprawdę wiemy więcej, mianowicie, że $5 = 5$. Ale „$5 \leq 5$" oznacza, że „5 jest mniejsze od 5 lub 5 jest równe 5" i jest to stwierdzenie prawdziwe. Podobnie, stwierdzenie, że $S \subseteq S$ jest prawdziwe, chociaż wiemy więcej, mianowicie, że $S = S$. Stwierdzenia takie, jak „$5 = 5$", „$5 \leq 5$", „$S = S$" czy „$S \subseteq S$" nie przeszkadzają, a często przydają się do zwrócenia uwagi na to, że zachodzi szczególny przypadek stwierdzenia ogólniejszego.[1]

■

Będziemy czasami pisać $T \subset S$ mając na myśli, że $T \subseteq S$ i $T \neq S$, tzn. zbiór T jest podzbiorem zbioru S, różnym od S. Znaku $\subset$ używamy podobnie jak znaku $<$ dla liczb rzeczywistych. Jeśli $T \subset S$, to mówimy, że T jest **właściwym podzbiorem** S.

Wprowadzimy teraz oznaczenia dla pewnych szczególnych podzbiorów zbioru R, nazywanych **przedziałami**. Dla $a, b \in \mathbb{R}$, gdzie $a < b$, określamy

$$[a,b] = \{x \in \mathbb{R}\colon a \leq x \leq b\}; \quad (a,b) = \{x \in \mathbb{R}\colon a < x < b\};$$
$$[a,b) = \{x \in \mathbb{R}\colon a \leq x < b\}; \quad (a,b] = \{x \in \mathbb{R}\colon a < x \leq b\}.$$

Ogólna zasada jest taka: nawiasy [,] oznaczają, że końce przedziału należą do niego, a nawiasy (,) oznaczają, że przedział ich nie zawiera. Przedziały postaci $[a,b]$ nazywamy przedziałami **domkniętymi**, przedziały postaci (a,b) nazywamy przedziałami **otwartymi**. Wygodnie jest też używać nazwy „przedział" dla pewnych nieograniczonych zbiorów, które zapisujemy za po-

[1] Będziemy używać znaku ■ do zaznaczenia końca przykładu lub dowodu.

mocą symboli $+\infty$ i $-\infty$, nie oznaczających liczb rzeczywistych, ale stanowiących po prostu fragmenty oznaczeń tych zbiorów. Zatem

$$[a,\infty) = \{x \in \mathbb{R}\colon a \leq x\}; \quad (a,\infty) = \{x \in \mathbb{R}\colon a < x\};$$
$$(-\infty,b] = \{x \in \mathbb{R}\colon x \leq b\}; \quad (-\infty,b) = \{x \in \mathbb{R}\colon x < b\}.$$

Oznaczeń zbiorów i przedziałów należy używać starannie. Na przykład $[0,1]$, $(0,1)$ i $\{0,1\}$ oznaczają różne zbiory. W rzeczywistości, przedziały $[0,1]$ i $(0,1)$ są zbiorami nieskończonymi, podczas gdy zbiór $\{0,1\}$ ma tylko dwa elementy.

Rozważmy następujące zbiory:

$$\{n \in \mathbb{N}\colon 2 < n < 3\}, \quad \{x \in \mathbb{R}\colon x^2 < 0\},$$
$$\{r \in \mathbb{Q}\colon r^2 = 2\}, \quad \{x \in \mathbb{R}\colon x^2 + 1 = 0\}.$$

Zbiory te mają jedną wspólną własność: nie mają żadnych elementów. Z czysto logicznego punktu widzenia, wszystkie one mają te same elementy, a więc są równe, mimo że są zapisane w różny sposób. Ten jedyny zbiór nie mający w ogóle elementów, nazywamy **zbiorem pustym**. Będziemy używać dla niego dwóch oznaczeń, przemawiającego do wyobraźni oznaczenia $\{\ \}$ i standardowego oznaczenia $\emptyset$. Symbol $\emptyset$ nie jest grecką literą ϕ (fi); jest on literą zapożyczoną z alfabetu norweskiego i należy go czytać jako „zbiór pusty". Przyjmujemy, że $\emptyset$ jest podzbiorem każdego zbioru S, gdyż zdanie „jeśli $x \in \emptyset$, to $x \in S$" uważamy za logicznie prawdziwe. Musisz przyjąć to wyjaśnienie „na wiarę", dopóki nie przeczytasz paragrafu 2.3.

Zbiory są obiektami, a więc mogą być elementami innych zbiorów. Zbiór $\{\{1,2\},\{1,3\},\{2\},\{3\}\}$ ma cztery elementy, mianowicie zbiory $\{1,2\}$, $\{1,3\}$, $\{2\}$ i $\{3\}$. Gdybyśmy mieli pudełko zawierające dwa woreczki z kulkami, to traktowalibyśmy je jako pudełko z woreczkami, a nie pudełko z kulkami, więc miałoby ono dwa elementy. Podobnie, jeśli A jest zbiorem, to $\{A\}$ jest zbiorem mającym jeden element, mianowicie A, niezależnie od tego, ile elementów ma sam zbiór A. Pudełko zawierające pusty woreczek zawiera coś, mianowicie woreczek, a więc nie jest pustym pudełkiem. W ten sam sposób, $\{\emptyset\}$ jest zbiorem mającym jeden element, podczas gdy $\emptyset$ jest zbiorem nie mającym elementów, zatem $\{\emptyset\}$ i $\emptyset$ są różnymi zbiorami. Mamy $\emptyset \in \{\emptyset\}$, a nawet $\emptyset \subseteq \{\emptyset\}$, ale $\emptyset \notin \emptyset$.

Zbiór wszystkich podzbiorów zbioru S nazywamy **zbiorem potęgowym** zbioru S i oznaczamy symbolem $\mathcal{P}(S)$. Oczywiście zbiór pusty $\emptyset$ i sam zbiór S są elementami $\mathcal{P}(S)$, tzn. $\emptyset \in \mathcal{P}(S)$ i $S \in \mathcal{P}(S)$.

PRZYKŁAD 2 (a) Mamy $\mathcal{P}(\emptyset) = \{\emptyset\}$, gdyż $\emptyset$ jest jedynym podzbiorem $\emptyset$.

(b) Rozważmy typowy zbiór jednoelementowy, na przykład $S = \{a\}$. Wtedy zbiór $\mathcal{P}(S) = \{\emptyset, \{a\}\}$ ma dwa elementy.

(c) Jeśli $S = \{a, b\}$ i $a \neq b$, to zbiór

$$\mathcal{P}(S) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$$

ma cztery elementy.

(d) Jeśli zbiór $S = \{a, b, c\}$ ma trzy elementy, to zbiór

$$\mathcal{P}(S) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$$

ma osiem elementów.

(e) Niech S będzie zbiorem skończonym. Zauważmy, że jeśli zbiór S ma n elementów i $n \leq 3$, to zbiór $\mathcal{P}(S)$ ma 2^n elementów, jak to wykazaliśmy w powyższych punktach od (a) do (d). Nie jest to przypadkowe, jak pokażemy w przykładzie 4(b) w § 4.2.

(f) Jeśli zbiór S jest nieskończony, to oczywiście zbiór $\mathcal{P}(S)$ jest też nieskończony. ∎

Wprowadzimy jeszcze jeden szczególny rodzaj zbioru, oznaczany przez Σ^*, który będzie często pojawiał się w tej książce. Naszym celem jest w miarę ogólne, ale precyzyjne z matematycznego punktu widzenia badanie języków. Po pierwsze, powiemy, że **alfabet** jest to skończony zbiór niepusty Σ (wielka grecka litera sigma), którego elementami są symbole, często nazywane **literami** alfabetu Σ, i na który nałożymy pewne drobne ograniczenia. Omówimy je pod koniec tego paragrafu. **Słowem** danego alfabetu Σ nazywamy dowolny skończony ciąg liter tego zbioru Σ. Wreszcie, zbiór wszystkich słów zbudowanych z liter alfabetu Σ oznaczymy przez Σ^* (sigma z gwiazdką). Dowolny podzbiór zbioru Σ^* nazywamy **językiem** nad alfabetem Σ.

PRZYKŁAD 3 (a) Niech $\Sigma = \{a, b, c, d, \ldots, z\}$ składa się z 26 liter alfabetu angielskiego. Dowolny ciąg liter z Σ należy do Σ^*. Zatem Σ^* zawiera *math*, *is*, *fun*, *aint*, *lieblich*, *amour*, *zzyzzoomph*, *etcetera* etc. Ponieważ Σ^* zawiera *a*, *aa*, *aaa*, *aaaa*, *aaaaa* itd., więc oczywiście Σ^* jest zbiorem nieskończonym. Moglibyśmy dla ścisłości zdefiniować język amerykański L jako podzbiór Σ^* składający się ze słów najnowszego wydania *Webster's New World Dictionary of the American Language*. Zatem

$$L = \{a, aachen, aardvark, aardwolf, \ldots, zymurgy\}$$

jest to duży, ale jednak skończony zbiór.

(b) Aby podać proste przykłady dobrze ilustrujące omawiane pojęcia, często jako Σ bierzemy dwuelementowy zbiór $\{a, b\}$.

W tym przypadku zbiór Σ^* zawiera a, b, ab, ba, bab, $babbabb$ itp. Znów zbiór Σ^* jest nieskończony.

(c) Jeśli $\Sigma = \{0, 1\}$, to zbiór B tych słów ze zbioru Σ^*, które zaczynają się od 1, jest dokładnie zbiorem rozwinięć dwójkowych liczb całkowitych dodatnich. To znaczy

$$B = \{1, 10, 11, 100, 101, 110, 111, 1000, 1001, \ldots\}. \quad \blacksquare$$

Istnieje w Σ^* pewne szczególne słowo, w pewnym sensie analogiczne do zbioru pustego, nazywane **słowem pustym**; jest to ciąg nie zawierający wcale liter i oznaczany przez λ (mała grecka litera lambda).

PRZYKŁAD 4 (a) Jeśli $\Sigma = \{a, b\}$, to

$$\Sigma^* = \{\lambda, a, b, aa, ab, ba, bb, aaa, aab, aba, abb, baa, bab, bba, \ldots\}.$$

(b) Jeśli $\Sigma = \{0, 1, 2\}$, to

$$\Sigma^* = \{\lambda, 0, 1, 2, 00, 01, 02, 10, 11, 12, 20, 21, 22, 000, 001, 002, \ldots\}.$$

(c) Jeśli $\Sigma = \{a\}$, to

$$\Sigma^* = \{\lambda, a, aa, aaa, aaaa, aaaaa, aaaaaa, \ldots\}.$$

W tym przykładzie nie znajdziemy żadnych interesujących języków, ale będzie on dobrze ilustrował różne wprowadzane pojęcia.

(d) Różne języki programowania spełniają naszą definicję języka. Na przykład, alfabet Σ pewnej wersji języka ALGOL ma 113 elementów; zbiór Σ składa się z liter, cyfr 0, 1, 2, ..., 9 i wielu operatorów, w tym również operatorów zbudowanych z ciągów liter, takich jak „go to" i „if". Jak zwykle, Σ^* zawiera wszystkie możliwe skończone ciągi liter ze zbioru Σ, niezależnie od ich znaczenia. Podzbiór zbioru Σ^*, składający się z tych ciągów, które zostają zaakceptowane w czasie działania kompilatora języka ALGOL na danym komputerze, jest dobrze zdefiniowanym i użytecznym podzbiorem zbioru Σ^*; możemy go nazwać językiem ALGOL określonym przez ten kompilator. $\blacksquare$

Omówimy teraz niezbędne ograniczenia nakładane na zbiór Σ. Problem pojawia się wtedy, gdy litery alfabetu Σ same są zbudowane z innych liter z tego samego zbioru Σ lub z jakiegoś innego alfabetu. Na przykład, jeśli zbiór Σ zawiera jako litery symbole a, b i ab, to ciąg aab może być rozumiany jako ciąg trzech liter a, a i b z Σ lub jako ciąg dwóch liter a i ab. Nie można stwierdzić, którym z tych ciągów ma on być, i maszyna

czytająca litery a, a, b po jednej nie byłaby w stanie zinterpretować tych danych wejściowych jednoznacznie. Weźmy inny przykład. Jeśli Σ zawiera ab, aba i bab, to ciąg wejściowy $ababab$ może być zinterpretowany albo jako $(ab)(ab)(ab)$, albo jako $(aba)(bab)$. Aby uniknąć takich problemów, nie pozwolimy na to, by zbiór Σ zawierał jakiekolwiek litery, które same są ciągami liter rozpoczynającymi się od litery należącej do Σ. Zatem dopuścimy zbiory $\Sigma = \{a, b, c\}$, $\Sigma = \{a, b, ca\}$, $\Sigma = \{a, b, Ab\}$, ale nie dopuścimy ani $\Sigma = \{a, b, c, ac\}$, ani nawet $\Sigma = \{a, b, ac\}$.

Po przyjęciu tej umowy możemy jednoznacznie określić **długość słowa** w z Σ^* (długość(w)) jako liczbę liter alfabetu Σ w słowie w, zliczając każde wystąpienie litery. Na przykład, jeśli $\Sigma = \{a, b\}$, to długość(aab) = długość(bab) = 3. Jeśli $\Sigma = \{a, b, Ab\}$, to długość$(abbAb)$ = 4. Określmy także długość$(\lambda) = 0$. Bardziej precyzyjna definicja długości jest podana w § 7.1.

Jeszcze jedna uwaga na zakończenie: będziemy używać symboli w, w_1 itd. jako nazw zmiennych dla słów. Nie powinno prowadzić to do nieporozumień, pomimo że litera w jest też literą alfabetu polskiego.

PRZYKŁAD 5 Jeśli $\Sigma = \{a, b\}$ oraz $A = \{w \in \Sigma^*\text{: długość}(w) = 2\}$, to $A = \{aa, ab, ba, bb\}$. Jeśli

$$B = \{w \in \Sigma^*\text{: długość}(w) \text{ jest liczbą parzystą}\},$$

to B jest zbiorem nieskończonym $\{\lambda, aa, ab, ba, bb, aaaa, aaab, aaba, aabb, \ldots\}$. Zauważmy, że A jest podzbiorem B. ∎

ĆWICZENIA DO § 1.1

1. Wypisz po pięć elementów każdego z następujących zbiorów:
(a) $\{n \in \mathbb{N}\text{: liczba } n \text{ jest podzielna przez } 5\}$,
(b) $\{2n + 1\text{: } n \in \mathbb{P}\}$,
(c) $\mathcal{P}(\{1, 2, 3, 4, 5\})$,
(d) $\{2^n\text{: } n \in \mathbb{N}\}$,
(e) $\{1/n\text{: } n \in \mathbb{P}\}$,
(f) $\{r \in \mathbb{Q}\text{: } 0 < r < 1\}$,
(g) $\{n \in \mathbb{N}\text{: liczba } n + 1 \text{ jest pierwsza}\}$.

2. Wypisz elementy następujących zbiorów:
(a) $\{1/n\text{: } n = 1, 2, 3, 4\}$,
(b) $\{n^2 - n\text{: } n = 0, 1, 2, 3, 4\}$,
(c) $\{1/n^2\text{: } n \in \mathbb{P},\ \text{liczba } n \text{ jest parzysta i } n < 11\}$,
(d) $\{2 + (-1)^n\text{: } n \in \mathbb{N}\}$.

3. Wypisz po pięć elementów każdego z następujących zbiorów:
(a) Σ^*, gdzie $\Sigma = \{a, b, c\}$,
(b) $\{w \in \Sigma^*$: długość$(w) \leq 2\}$, gdzie $\Sigma = \{a, b\}$,
(c) $\{w \in \Sigma^*$: długość$(w) = 4\}$, gdzie $\Sigma = \{a, b\}$.
Które z tych zbiorów zawierają słowo puste λ?

4. Wyznacz poniższe zbiory, tzn. wypisz ich elementy, jeśli zbiory te są niepuste i napisz $\emptyset$, jeśli są puste.
(a) $\{n \in \mathbb{N}: n^2 = 9\}$
(b) $\{n \in \mathbb{Z}: n^2 = 9\}$
(c) $\{x \in \mathbb{R}: x^2 = 9\}$
(d) $\{n \in \mathbb{N}: 3 < n < 7\}$
(e) $\{n \in \mathbb{Z}: 3 < |n| < 7\}$
(f) $\{x \in \mathbb{R}: x^2 < 0\}$
(g) $\{n \in \mathbb{N}: n^2 = 3\}$
(h) $\{x \in \mathbb{Q}: x^2 = 3\}$
(i) $\{x \in \mathbb{R}: x < 1$ i $x \geq 2\}$
(j) $\{3n + 1: n \in \mathbb{N}$ i $n \leq 6\}$
(k) $\{n \in \mathbb{P}$: liczba n jest pierwsza i $n \leq 15\}$ (pamiętaj o tym, że 1 nie jest liczbą pierwszą.)

5. Ile elementów mają poniższe zbiory? Napisz ∞, jeśli zbiór jest nieskończony.
(a) $\{n \in \mathbb{N}: n^2 = 2\}$
(b) $\{n \in \mathbb{Z}: 0 \leq n \leq 73\}$
(c) $\{n \in \mathbb{Z}: 5 \leq |n| \leq 73\}$
(d) $\{n \in \mathbb{Z}: 5 < n < 73\}$
(e) $\{n \in \mathbb{Z}$: liczba n jest parzysta i $|n| \leq 73\}$
(f) $\{x \in \mathbb{Q}: 0 \leq x \leq 73\}$
(g) $\{x \in \mathbb{Q}: x^2 = 2\}$
(h) $\{x \in \mathbb{R}: x^2 = 2\}$
(i) $\{x \in \mathbb{R}: 0{,}99 < x < 1{,}00\}$
(j) $\mathcal{P}(\{0, 1, 2, 3\})$
(k) $\mathcal{P}(\mathbb{N})$
(l) $\{n \in \mathbb{N}$: liczba n jest parzysta$\}$
(m) $\{n \in \mathbb{N}$: liczba n jest pierwsza$\}$
(n) $\{n \in \mathbb{N}$: liczba n jest parzysta i pierwsza$\}$
(o) $\{n \in \mathbb{N}$: liczba n jest parzysta lub pierwsza$\}$

6. Ile elementów mają poniższe zbiory? Napisz ∞, jeśli zbiór jest nieskończony.
(a) $\{-1, 1\}$
(b) $[-1, 1]$
(c) $(-1, 1)$
(d) $\{n \in \mathbb{Z}: -1 \leq n \leq 1\}$
(e) Σ^*, gdzie $\Sigma = \{a, b, c\}$
(f) $\{w \in \Sigma^*$: długość$(w) \leq 4\}$, gdzie $\Sigma = \{a, b, c\}$

7. Rozważ zbiory

$$A = \{n \in \mathbb{P}\text{: liczba } n \text{ jest nieparzysta}\},$$
$$B = \{n \in \mathbb{P}\text{: liczba } n \text{ jest pierwsza}\},$$
$$C = \{4n + 3\text{: } n \in \mathbb{P}\}, \qquad D = \{x \in \mathbb{R}\text{: } x^2 - 8x + 15 = 0\}.$$

Które z tych zbiorów są podzbiorami innych z tych zbiorów? Rozpatrz wszystkie szesnaście możliwości.

8. Rozważ zbiory $\{0,1\}$, $(0,1)$ i $[0,1]$. Czy następujące stwierdzenia są prawdziwe?
(a) $\{0,1\} \subseteq (0,1)$
(b) $\{0,1\} \subseteq [0,1]$
(c) $(0,1) \subseteq [0,1]$
(d) $\{0,1\} \subseteq \mathbb{Z}$
(e) $[0,1] \subseteq \mathbb{Z}$
(f) $[0,1] \subseteq \mathbb{Q}$
(g) $1/2$ i $\pi/4$ są elementami $\{0,1\}$
(h) $1/2$ i $\pi/4$ są elementami $(0,1)$
(i) $1/2$ i $\pi/4$ są elementami $[0,1]$

9. Rozważ następujące trzy alfabety: $\Sigma_1 = \{a, b, c\}$, $\Sigma_2 = \{a, b, ca\}$ oraz $\Sigma_3 = \{a, b, Ab\}$. Sprawdź, do którego ze zbiorów Σ_1^*, Σ_2^* i Σ_3^* należy każde poniższe słowo i określ długość tego słowa jako elementu każdego zbioru, do którego ono należy.
(a) *aba*
(b) *bAb*
(c) *cba*
(d) *cab*
(e) *caab*
(f) *baAb*

10. Zastanów się nad następującym problemem. Załóżmy, że $\Sigma = \{a, b\}$ i wyobraźmy sobie, jeśli to możliwe, słownik zawierający wszystkie niepuste słowa z Σ^*, ułożone w zwykłym porządku alfabetycznym. Wszystkie słowa *a*, *aa*, *aaa*, *aaaa* itd. muszą znajdować się przed słowem *ba*. Jak daleko musimy szukać w słowniku słowa *ba*? Jak zmieniłaby się odpowiedź, gdyby słownik zawierał tylko te słowa z Σ^*, których długość jest nie większa niż 5?

11. Przypuśćmy, że w jest niepustym słowem w Σ^*.
(a) Jeśli usuniemy pierwszą (od lewej strony) literę słowa w, to czy otrzymane słowo należy do Σ^*?
(b) A jeśli będziemy usuwać litery z obu końców słowa w? Czy otrzymane słowa będą nadal należeć do Σ^*?
(c) Jeśli mamy do dyspozycji maszynę, która potrafi rozpoznawać litery alfabetu Σ oraz potrafi usuwać litery ze słów, to w jaki sposób możemy wykorzystać ją do rozstrzygania, czy dany ciąg symboli należy do Σ^*?

§ 1.2. Działania na zbiorach

W tym paragrafie wprowadzimy działania na zbiorach, pozwalające na tworzenie nowych zbiorów ze starych. Definiujemy **sumę** $A \cup B$ i **przecięcie** $A \cap B$ zbiorów A i B w następujący sposób:

$$A \cup B = \{x\colon x \in A \text{ lub } x \in B \text{ lub } x \text{ należy do obu zbiorów}\};$$

$$A \cap B = \{x\colon x \in A \text{ i } x \in B\}.$$

Dodaliśmy „lub x należy do obu zbiorów" w definicji $A \cup B$ dla podkreślenia, że istnieje taka możliwość. W potocznym języku polskim słowo „lub" może być rozumiane na dwa sposoby. Czasami oznacza ono alternatywę nie wykluczającą, czyli stwierdza, że prawdziwe jest pierwsze zdanie, drugie zdanie lub oba. W ten sposób można na przykład interpretować program nauczania stwierdzający, że student musi wybrać lektorat języka angielskiego lub niemieckiego. Czasami jednak „lub" oznacza **alternatywę wykluczającą**, co znaczy, że prawdziwe ma być albo jedno zdanie, albo drugie, ale nie oba jednocześnie. Jest to interpretacja taka jak w menu oferującym zupę lub przystawkę. W matematyce będziemy zawsze interpretować „lub" jako alternatywę nie wykluczającą, chyba że wyraźnie zaznaczymy, iż mamy na myśli znaczenie przeciwne. Mówimy, że zbiory A i B są rozłączne, jeśli nie mają wspólnych elementów, tzn. gdy $A \cap B = \emptyset$.

Dla danych zbiorów A i B definiujemy ich **różnicę** $A \backslash B$ jako zbiór obiektów należących do A i nie należących do B:

$$A \backslash B = \{x\colon x \in A \text{ i } x \notin B\} = \{x \in A\colon x \notin B\}.$$

Jest to zbiór powstały przez usunięcie ze zbioru A tych wszystkich elementów zbioru B, które należały też do A.

Różnicą symetryczną $A \oplus B$[1] zbiorów A i B nazywamy zbiór

$$A \oplus B = \{x\colon x \in A \text{ lub } x \in B, \text{ ale } x \text{ nie należy do obu zbiorów jednocześnie}\}.$$

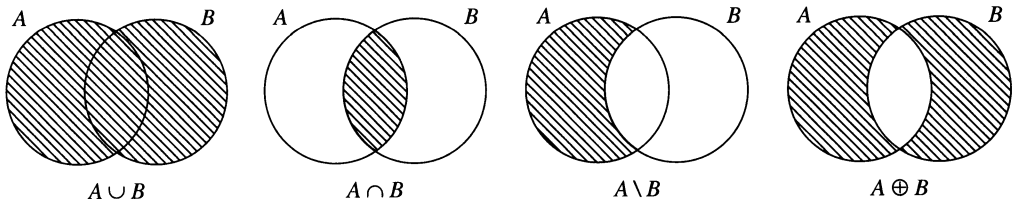
Zauważmy, że w tej definicji użyliśmy alternatywy wykluczającej. Bezpośrednio z definicji wynika, że

$$A \oplus B = (A \cup B) \backslash (A \cap B) = (A \backslash B) \cup (B \backslash A).$$

Czasami wygodnie jest ilustrować związki pomiędzy zbiorami za pomocą rysunków, zwanych **diagramami Venna**, na których

[1] W Polsce tradycyjnie oznacza się różnicę symetryczną symbolem $A \div B$ (obecnie używa się też symbolu $A \triangle B$).

zbiory odpowiadają podzbiorom płaszczyzny. Popatrzmy na rysunek 1.1, na którym wskazane zbiory zostały zakreskowane.



Rysunek 1.1

PRZYKŁAD 1 (a) Niech $A = \{n \in \mathbb{N}\colon n \leq 11\}$, $B = \{n \in \mathbb{N}\colon$ liczba n jest parzysta i $n \leq 20\}$ i $E = \{n \in \mathbb{N}\colon$ liczba n jest parzysta$\}$. Wtedy mamy

$$A \cup B = \{0,1,2,3,4,5,6,7,8,9,10,11,12,14,16,18,20\},$$
$$A \cap B = \{0,2,4,6,8,10\},$$
$$A \setminus B = \{1,3,5,7,9,11\},$$
$$B \setminus A = \{12,14,16,18,20\},$$
$$A \oplus B = \{1,3,5,7,9,11,12,14,16,18,20\}.$$

Ponadto $E \cap B = B$, $B \setminus E = \{\ \}$,

$$E \setminus B = \{n \in \mathbb{N}\colon \text{liczba } n \text{ jest parzysta i } n \geq 22\}$$
$$= \{22,24,26,28,\ldots\},$$
$$\mathbb{N} \setminus E = \{n \in \mathbb{N}\colon \text{liczba } n \text{ jest nieparzysta}\}$$
$$= \{1,3,5,7,9,11,\ldots\},$$
$$A \oplus E = \{1,3,5,7,9,11\}$$
$$\cup \{n \in \mathbb{N}\colon \text{liczba } n \text{ jest parzysta i } n \geq 12\}$$
$$= \{1,3,5,7,9,11,12,14,16,18,20,22,\ldots\}.$$

(b) Rozważmy przedziały $[0,2]$ i $(0,1]$. Wtedy $(0,1] \subseteq [0,2]$, a więc

$$(0,1] \cup [0,2] = [0,2] \quad \text{oraz} \quad (0,1] \cap [0,2] = (0,1].$$

Ponadto

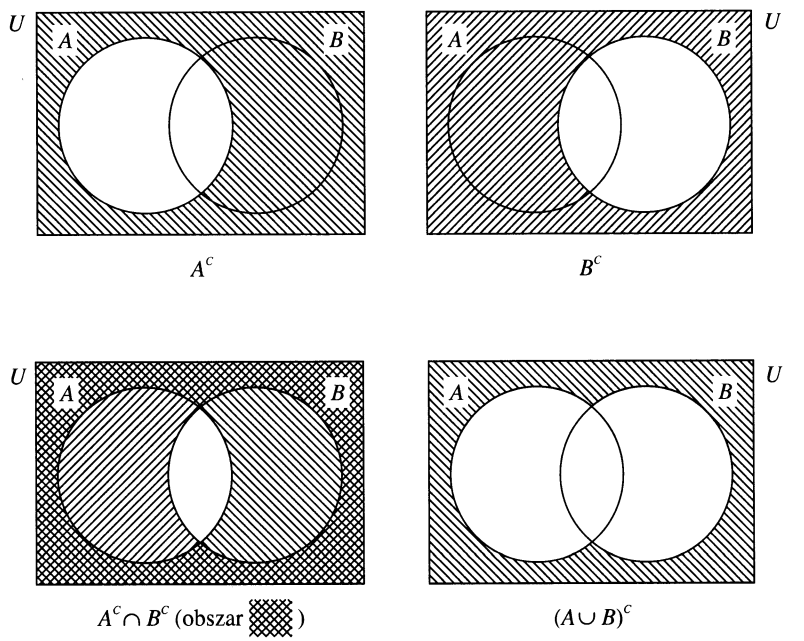
$$(0,1] \setminus [0,2] = \{\ \},$$
$$[0,2] \setminus (0,1] = \{0\} \cup (1,2] \quad \text{oraz} \quad [0,2] \setminus (0,2) = \{0,2\}.$$

(c) Niech $\Sigma = \{a,b\}$, $A = \{\lambda, a, aa, aaa\}$, $B = \{\lambda, b, bb, bbb\}$ i $C = \{w \in \Sigma^*\colon$ długość$(w) \leq 2\}$. Wtedy mamy

$$
\begin{aligned}
A \cup B &= \{\lambda, a, b, aa, bb, aaa, bbb\},\\
A \cap B &= \{\lambda\},\\
A \backslash B &= \{a, aa, aaa\},\\
B \backslash A &= \{b, bb, bbb\},\\
A \cap C &= \{\lambda, a, aa\},\\
B \backslash C &= \{bbb\},\\
C \backslash A &= \{b, ab, ba, bb\},\\
A \backslash \Sigma &= \{\lambda, aa, aaa\}.
\end{aligned}
$$

■

Często wygodnie jest użyć pewnego ustalonego zbioru U, takiego jak $\mathbb{N}$, $\mathbb{R}$ lub Σ^*, nazywanego **uniwersum**, **zbiorem uniwersalnym** lub **przestrzenią**, i rozważać tylko elementy i podzbiory tego zbioru U. Dla zbioru $A \subseteq U$ różnicę zbiorów $U \backslash A$ nazywamy **dopełnieniem** lub **uzupełnieniem** zbioru A i oznaczamy symbolem A^c (czasami stosuje się też oznaczenia $-A$, $\bar{A}$, A' - przyp. tłum.). Zauważmy, że różnica zbiorów może być zapisana za pomocą dopełnienia: $A \backslash B = A \cap B^c$. Na diagramach Venna na rysunku 1.2 uniwersum U zostało narysowane w postaci prostokąta i wskazane zbiory zostały zakreskowane.



Rysunek 1.2

PRZYKŁAD 2 (a) Jeśli przestrzenią jest zbiór $\mathbb{N}$ i zbiory A i E są takie jak w przykładzie 1(a), to

$$A^c = \{n \in \mathbb{N}\colon n \geq 12\} \text{ oraz}$$

$$E^c = \{n \in \mathbb{N}\colon \text{liczba } n \text{ jest nieparzysta}\}.$$

(b) Jeśli przestrzenią jest $\mathbb{R}$, to $[0,1]^c = (-\infty, 0) \cup (1, \infty)$, $(0,1)^c = (-\infty, 0] \cup [1, \infty)$ i $\{0,1\}^c = (-\infty, 0) \cup (0,1) \cup (1, \infty)$. Dla dowolnej liczby $a \in \mathbb{R}$ mamy $[a, \infty)^c = (-\infty, a)$ oraz $(a, \infty)^c = (-\infty, a]$. ■

Zauważmy, że dwa ostatnie diagramy Venna na rysunku 1.2 pokazują, że $A^c \cap B^c = (A \cup B)^c$. Ta i wiele innych równości między zbiorami są równościami prawdziwymi dla dowolnych zbiorów. Tablica 1.1 podaje pewne podstawowe tożsamości dotyczące działań na zbiorach. Ich liczba nie powinna nas przytłaczać, przyjrzyjmy się każdej z nich z osobna. Jak wskazują niektóre nazwy praw, wiele z nich odpowiada prawom znanym z algebry. Nowe są prawa idempotentności działań (prawo $a + a = a$ nie jest prawdziwe dla większości liczb), jak również jest tylko jedno prawo rozdzielności dotyczące liczb. Oczywiście nowe są też prawa dotyczące dopełnienia. Zakładamy, że wszystkie zbiory, których

Tablica 1.1. Prawa algebry zbiorów

1a. $A \cup B = B \cup A$ **b.** $A \cap B = B \cap A$	prawa przemienności
2a. $(A \cup B) \cup C = A \cup (B \cup C)$ **b.** $(A \cap B) \cap C = A \cap (B \cap C)$	prawa łączności
3a. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ **b.** $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	prawa rozdzielności
4a. $A \cup A = A$ **b.** $A \cap A = A$	prawa idempotentności
5a. $A \cup \emptyset = A$ **b.** $A \cup U = U$ **c.** $A \cap \emptyset = \emptyset$ **d.** $A \cap U = A$	prawa identyczności
6. $(A^c)^c = A$	prawo podwójnego dopełnienia
7a. $A \cup A^c = U$ **b.** $A \cap A^c = \emptyset$	
8a. $U^c = \emptyset$ **b.** $\emptyset^c = U$	
9a. $(A \cup B)^c = A^c \cap B^c$ **b.** $(A \cap B)^c = A^c \cup B^c$	prawa De Morgana

dotyczy tablica 1.1, są podzbiorami pewnej ustalonej przestrzeni U. Ze względu na prawa łączności działań, możemy zapisywać zbiory takie, jak $A \cup B \cup C$ i $A \cap B \cap C$ bez użycia nawiasów i nie będzie to prowadziło do nieporozumień.

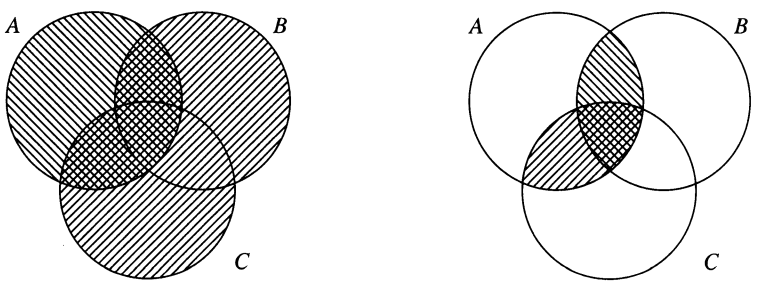
Poprawność równości podanych w tablicy 1.1 można sprawdzić dwoma sposobami. Można narysować odpowiednie diagramy Venna, zakreskować właściwe zbiory i zobaczyć, że są one równe. Można też wykazać, że zbiory S i T są równe, dowodząc, że $S \subseteq T$ i $T \subseteq S$; prawdziwości tych inkluzji dowodzimy pokazując, że $x \in S$ implikuje $x \in T$ oraz $x \in T$ implikuje $x \in S$. Pokażemy przykłady obu tych sposobów argumentacji, pozostawiając większość dowodów zainteresowanemu Czytelnikowi.

PRZYKŁAD 3 Diagramy Venna na rysunku 1.2 ilustrują prawo De Morgana 9a. Podamy teraz dowód, w którym pokażemy najpierw, że $(A \cup B)^c \subseteq A^c \cap B^c$, a następnie, że $A^c \cap B^c \subseteq (A \cup B)^c$.

Aby pokazać, że $(A \cup B)^c \subseteq A^c \cap B^c$, bierzemy element $x \in (A \cup B)^c$. Wtedy $x \notin A \cup B$. W szczególności $x \notin A$, a więc musimy mieć $x \in A^c$. Podobnie, $x \notin B$, a więc $x \in B^c$. Zatem $x \in A^c \cap B^c$. Pokazaliśmy więc, że $x \in (A \cup B)^c$ implikuje $x \in A^c \cap B^c$; zatem $(A \cup B)^c \subseteq A^c \cap B^c$.

Aby dowieść odwrotnej inkluzji, $A^c \cap B^c \subseteq (A \cup B)^c$, bierzemy $x \in A^c \cap B^c$. Wtedy $x \in A^c$, a więc $x \notin A$. Również $x \in B^c$, więc $x \notin B$. Ponieważ $x \notin A$ i $x \notin B$, więc $x \notin A \cup B$, czyli $x \in (A \cup B)^c$. Zatem $A^c \cap B^c \subseteq (A \cup B)^c$. ∎

PRZYKŁAD 4 Diagramy Venna na rysunku 1.3 wykazują prawdziwość prawa rozdzielności 3b z tablicy 1.1. Ilustracją zbioru $A \cap (B \cup C)$ jest zakreskowany w obie strony obszar na diagramie po lewej stronie; po prawej stronie zbiór $(A \cap B) \cup (A \cap C)$ jest reprezentowany przez obszar zakreskowany w jedną stronę lub w obie strony.

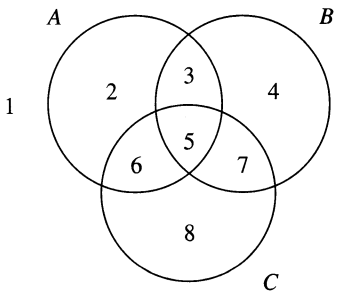


$A \cap (B \cup C)$ $(A \cap B) \cup (A \cap C)$

Rysunek 1.3

A oto dowód, w którym pokazujemy, że każdy z tych zbiorów jest podzbiorem drugiego. Weźmy najpierw $x \in A \cap (B \cup C)$. Wtedy na pewno x jest elementem zbioru A. Element x należy również do $B \cup C$. Zatem albo $x \in B$ i w tym przypadku $x \in A \cap B$, albo $x \in C$ i wtedy $x \in A \cap C$. W obu przypadkach mamy $x \in (A \cap B) \cup (A \cap C)$. Widać, że $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$.

Teraz weźmy $y \in (A \cap B) \cup (A \cap C)$. Wtedy albo $y \in A \cap B$, albo $y \in A \cap C$; te dwa przypadki rozpatrzymy oddzielnie. Jeśli $y \in A \cap B$, to $y \in A$ i $y \in B$, a więc $y \in B \cup C$, skąd wynika, że $y \in A \cap (B \cup C)$. Podobnie, jeśli $y \in A \cap C$, to $y \in A$ i $y \in C$, zatem $y \in B \cup C$ i znów $y \in A \cap (B \cup C)$. Ponieważ w obu przypadkach mieliśmy $y \in A \cap (B \cup C)$, więc pokazaliśmy, że $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. Udowodniliśmy już inkluzję odwrotną, więc te dwa zbiory są równe. ∎



Rysunek 1.4

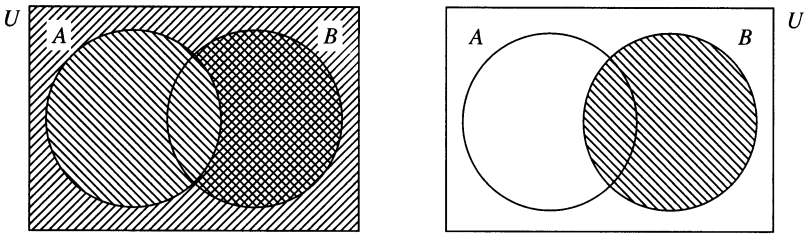
Dowody za pomocą diagramów Venna wydają się znacznie prostsze niż dowody, w których dowodzimy inkluzji analizując należenie elementów do zbiorów. Dowody rysunkowe budzą niepokój wielu ludzi; jednak diagram Venna dla zbiorów A, B, C składa się z ośmiu obszarów (rys. 1.4), co wyczerpuje wszystkie logiczne możliwości. Zatem dowody za pomocą diagramów Venna są całkowicie poprawne. Znacznie poważniejszym zarzutem wobec diagramów Venna jest to, że ukrywają one przebieg rozumowania; nie ukazują one całej logiki związanej z zakreskowywaniem diagramów. Gdybyśmy zapisali całe rozumowanie prowadzące do rysunku 1.3, dowód byłby równie długi, jak oparty wyłącznie na logice dowód analizujący należenie elementów do poszczególnych zbiorów, podany w przykładzie 4. Innym powodem unikania diagramów Venna jest to, że trudno je narysować, gdy mamy do czynienia z więcej niż trzema zbiorami. Jednakże prawie każdy, kto zajmuje się matematyką, korzysta z rysunków, w tym również z diagramów Venna, by lepiej rozumieć konkretne sytuacje matematyczne.

Tablica 1.1 podaje niektóre podstawowe zależności teorii mnogości. Istnieje jeszcze wiele innych zależności. Można je udowodnić jedną z trzech metod: (1) metodą diagramów Venna, (2) metodą rozumowań wykorzystujących elementy zbiorów, takich jak w przykładach 3 i 4 lub (3) metodą korzystania z praw podanych w tablicy 1.1. Czasami dowody będą łączyły metody (2) i (3).

PRZYKŁAD 5 Podamy trzy dowody zależności

$$(A \cup B) \cap A^c \subseteq B.$$

Dowód 1. Popatrzmy na rysunek 1.5. Obszar odpowiadający zbiorowi $(A \cup B) \cap A^c$ jest zakreskowany w obie strony i oczywiście jest on zawarty w obszarze oznaczającym zbiór B.



$(A \cup B) \cap A^c$ B

Rysunek 1.5

Dowód 2. Pokażemy, że $x \in (A \cup B) \cap A^c$ implikuje $x \in B$. Weźmy element x zbioru $(A \cup B) \cap A^c$. Wtedy $x \in A^c$, a więc $x \notin A$. Ponieważ x należy także do zbioru $A \cup B$, zatem należy do A lub do B, a więc stąd wynika, że musi należeć do B.

Dowód 3. Korzystając z praw algebry zbiorów podanych w tablicy 1.1, otrzymujemy

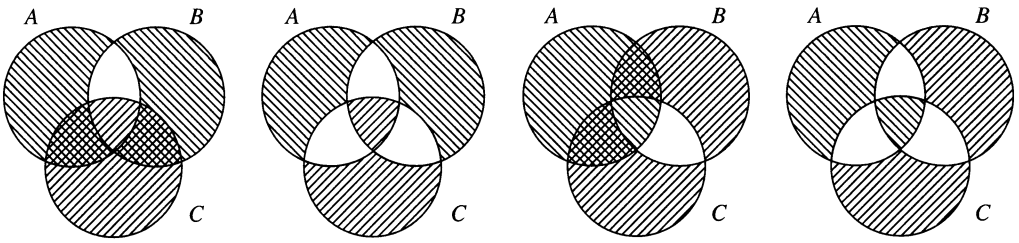
$$\begin{aligned}
(A \cup B) \cap A^c &= A^c \cap (A \cup B) && \text{przemienność 1b} \\
&= (A^c \cap A) \cup (A^c \cap B) && \text{rozdzielność 3b} \\
&= (A \cap A^c) \cup (A^c \cap B) && \text{przemienność 1b} \\
&= \emptyset \cup (A^c \cap B) && \text{7b} \\
&= (A^c \cap B) \cup \emptyset && \text{przemienność 1a} \\
&= A^c \cap B. && \text{prawo identyczności 5a}
\end{aligned}$$

Ta równość zgadza się oczywiście z diagramem po lewej stronie rysunku 1.5. Teraz jest jasne, że $A^c \cap B \subseteq B$, gdyż jeśli $x \in A^c \cap B$, to x musi być elementem B. ∎

Również różnica symetryczna $\oplus$ jest działaniem łącznym:

$$(A \oplus B) \oplus C = A \oplus (B \oplus C).$$

Możemy się o tym przekonać patrząc na diagramy Venna na rysunku 1.6. Po lewej stronie mamy zakreskowany w jedną stronę zbiór $A \oplus B$ i zakreskowany w drugą stronę zbiór C. Wtedy zbiór $(A \oplus B) \oplus C$ jest obszarem zakreskowanym w jedną stronę, wszystko jedno którą, ale nie w obie strony. Jeśli zrobimy to samo ze zbiorem $A \oplus (B \oplus C)$, otrzymamy ten sam obszar, a więc zbiory $(A \oplus B) \oplus C$ i $A \oplus (B \oplus C)$ są równe.



Rysunek 1.6

Oczywiście można też dowieść tej równości bez odwoływania się do rysunku. Być może będziecie chcieli sami przeprowadzić taki dowód. Zwracamy jednak uwagę na to, że szczegółowe rozumowanie będzie dość skomplikowane.

Ponieważ działanie $\oplus$ jest łączne, zapis $A \oplus B \oplus C$ nie prowadzi do nieporozumień. Zauważmy, że dowolny obiekt jest elementem tego zbioru wtedy i tylko wtedy, gdy należy do dokładnie jednego ze zbiorów A, B i C lub gdy należy do wszystkich trzech.

Rozważmy dwa zbiory S i T. Dla każdego elementu s zbioru S i każdego elementu t zbioru T tworzymy **parę uporządkowaną** (s,t). Element s jest tu pierwszym elementem pary uporządkowanej (poprzednikiem), t jest drugim elementem (następnikiem) i kolejność tych elementów jest istotna. Zatem $(s_1,t_1) = (s_2,t_2)$ wtedy i tylko wtedy, gdy $s_1 = s_2$ i $t_1 = t_2$. Zbiór wszystkich par uporządkowanych (s,t) nazywamy **iloczynem kartezjańskim** (produktem) zbiorów S i T i oznaczamy przez $S \times T$:

$$S \times T = \{(s,t)\colon s \in S \text{ i } t \in T\}.$$

Jeśli $S = T$, to czasami piszemy S^2 zamiast $S \times S$.

PRZYKŁAD 6 (a) Niech $S = \{1, 2, 3, 4\}$ i $T = \{a, b, c\}$. Wtedy $S \times T$ składa się z dwunastu par uporządkowanych wypisanych po lewej stronie tablicy 1.2. Moglibyśmy także przedstawić te pary jako odpowiednie punkty przecięcia numerowanych wierszy i kolumn, jak pokazuje to prawa część tego rysunku. Czytelnik powinien wypisać lub narysować elementy zbioru $T \times S$ i zauważyć, że $T \times S \neq S \times T$.

Tablica 1.2

$(1,c)$	$(2,c)$	$(3,c)$	$(4,c)$	c	o	o	o	o
$(1,b)$	$(2,b)$	$(3,b)$	$(4,b)$	b	o	o	o	o
$(1,a)$	$(2,a)$	$(3,a)$	$(4,a)$	a	o	o	o	o
					1	2	3	4
Elementy zbioru $[1,2,3,4] \times \{a,b,c\}$				**Rysunek zbioru** $[1,2,3,4] \times \{a,b,c\}$				

(b) Jeśli $S = \{1, 2, 3, 4\}$, to $S^2 = S \times S$ ma szesnaście par uporządkowanych (zob. tabl. 1.3). Zauważmy, że $(2, 4) \neq (4, 2)$; w tych parach uporządkowanych występują te same dwie liczby, ale w innej kolejności. Dla porównania, zbiory $\{2, 4\}$ i $\{4, 2\}$ są tym samym zbiorem. Zauważmy też, że $(2, 2)$ jest całkiem dobrą parą uporządkowaną, w której tak się składa, że pierwszy element jest równy drugiemu. Z drugiej strony, zbiór $\{2, 2\}$ jest po prostu zbiorem $\{2\}$, w którym liczba 2 została napisana dwa razy.

Tablica 1.3

$(1,4)$	$(2,4)$	$(3,4)$	$(4,4)$	4	o	o	o	o
$(1,3)$	$(2,3)$	$(3,3)$	$(4,3)$	3	o	o	o	o
$(1,2)$	$(2,2)$	$(3,2)$	$(4,2)$	2	o	o	o	o
$(1,1)$	$(2,1)$	$(3,1)$	$(4,1)$	1	o	o	o	o
					1	2	3	4
Elementy zbioru $\{1,2,3,4\}^2$				**Rysunek zbioru** $\{1,2,3,4\}^2$				

Nasze oznaczenie par uporządkowanych, na przykład $(2, 4)$, stoi w jawnej sprzeczności z oznaczeniem przedziałów z § 1.1, gdzie napis $(2, 4)$ oznaczał zbiór $\{x \in \mathbb{R}\colon 2 < x < 4\}$. Jednakże w obu przypadkach jest to standardowe oznaczenie. Na szczęście, zamierzone znaczenie prawie zawsze jasno wynika z kontekstu. ■

Dla każdego zbioru skończonego S przez $|S|$ oznaczamy liczbę elementów tego zbioru. Zatem $|S| = |T|$ dokładnie wtedy, gdy

zbiory S i T są tej samej wielkości. Zauważmy, że

$$|\emptyset| = 0 \text{ oraz } |\{1, 2, \ldots, n\}| = n \text{ dla } n \in \mathbb{P}.$$

Ponadto, $|S \times T| = |S| \cdot |T|$. Widzimy, skąd pochodzi oznaczenie $\times$ dla iloczynu kartezjańskiego dwóch zbiorów. Okazuje się, że $|\mathcal{P}(S)| = 2^{|S|}$, więc $\mathcal{P}(S)$ oznacza się też przez 2^S.

Możemy zdefiniować produkt dowolnej skończonej rodziny zbiorów $S_1, S_2, \ldots, S_n$. **Produkt zbiorów** $S_1 \times S_2 \times \ldots \times S_n$ składa się ze wszystkich uporządkowanych ciągów n-elementowych $(s_1, s_2, \ldots, s_n)$, gdzie $s_1 \in S_1$, $s_2 \in S_2$ itd. To znaczy, że

$$S_1 \times S_2 \times \ldots \times S_n = \{(s_1, s_2, \ldots, s_n)\colon s_k \in S_k \text{ dla } k = 1, 2, \ldots, n\}.$$

Tak jak w przypadku par uporządkowanych, dwa ciągi n-elementowe $(s_1, s_2, \ldots, s_n)$ i $(t_1, t_2, \ldots, t_n)$ uważamy za równe, jeśli ich odpowiednie wyrazy są równe: $s_k = t_k$ dla $k = 1, 2, \ldots, n$. Jeśli zbiory $S_1, S_2, \ldots, S_n$ są wszystkie równe zbiorowi S, to możemy napisać S^n zamiast $S_1 \times S_2 \times \ldots \times S_n$.

ĆWICZENIA DO § 1.2

1. Niech $U = \{1, 2, 3, 4, 5, \ldots, 12\}$, $A = \{1, 3, 5, 7, 11\}$, $B = \{2, 3, 5, 7, 11\}$, $C = \{2, 3, 6, 12\}$ i $D = \{2, 4, 8\}$. Wyznacz następujące zbiory:
(a) $A \cup B$,
(b) $A \cap C$,
(c) $(A \cup B) \cap C^c$,
(d) $A \backslash B$,
(e) $C \backslash D$,
(f) $B \oplus D$.
(g) Ile podzbiorów ma zbiór C?

2. Niech $A = \{1, 2, 3\}$, $B = \{n \in \mathbb{P}\colon$ liczba n jest parzysta$\}$ oraz $C = \{n \in \mathbb{P}\colon$ liczba n jest nieparzysta$\}$.
(a) Wyznacz zbiory $A \cap B$, $B \cap C$, $B \cup C$ i $B \oplus C$.
(b) Wypisz wszystkie podzbiory zbioru A.
(c) Które z następujących zbiorów: $A \oplus B$, $A \oplus C$, $A \backslash C$, $C \backslash A$ są nieskończone?.

3. W tym ćwiczeniu zbiorem uniwersalnym jest $\mathbb{R}$. Wyznacz następujące zbiory:
(a) $[0, 3] \cap [2, 6]$,
(b) $[0, 3] \cup [2, 6]$,
(c) $[0, 3] \backslash [2, 6]$,

(d) $[0,3] \oplus [2,6]$,
(e) $[0,3]^c$,
(f) $[0,3] \cap \emptyset$.

4. Niech $\Sigma = \{a, b\}$, $A = \{a, b, aa, bb, aaa, bbb\}$, $B = \{w \in \Sigma^*$: długość$(w) \geq 2\}$ i $C = \{w \in \Sigma^*$: długość$(w) \leq 2\}$.
(a) Wyznacz zbiory $A \cap C$, $A \backslash C$, $C \backslash A$ i $A \oplus C$.
(b) Wyznacz zbiory $A \cap B$, $B \cap C$, $B \cup C$ i $B \backslash A$.
(c) Wyznacz zbiory $\Sigma^* \backslash B$, $\Sigma \backslash B$ i $\Sigma \backslash C$.
(d) Wypisz wszystkie podzbiory Σ.
(e) Ile zbiorów należy do zbioru $\mathcal{P}(\Sigma)$?

5. W tym ćwiczeniu zbiorem uniwersalnym jest Σ^*, gdzie $\Sigma = \{a, b\}$. Niech zbiory A, B i C będą takie jak w ćwiczeniu 4. Wyznacz następujące zbiory:
(a) $B^c \cap C^c$,
(b) $(B \cap C)^c$,
(c) $(B \cup C)^c$,
(d) $B^c \cup C^c$,
(e) $A^c \cap C$,
(f) $A^c \cap B^c$.
(g) Które z tych zbiorów są równe? Dlaczego?

6. Następujące zdania dotyczą podzbiorów pewnego ustalonego niepustego zbioru uniwersalnego U. Wskaż, które z tych zdań jest prawdziwe, a które fałszywe. Dla każdego fałszywego zdania podaj przykład pokazujący, że jest ono fałszywe.
(a) $A \cap (B \cup C) = (A \cap B) \cup C$ dla wszystkich zbiorów A, B, C.
(b) $A \cup B \subseteq A \cap B$ implikuje $A = B$.
(c) $(A \cap \emptyset) \cup B = B$ dla wszystkich zbiorów A, B.
(d) $A \cap (\emptyset \cup B) = A$, jeśli tylko $A \subseteq B$.
(e) $A \cap B = A^c \cup B^c$ dla wszystkich zbiorów A, B.

7. Jakim zbiorem jest $A \oplus A$ dla dowolngo zbioru A? A jakim $A \oplus \emptyset$?

8. Udowodnij następujące równości za pomocą diagramów Venna:
(a) $A \cap (B \oplus C) = (A \cap B) \oplus (A \cap C)$,
(b) $A \oplus B \subseteq (A \oplus C) \cup (B \oplus C)$.

9. Udowodnij uogólnione prawo De Morgana $(A \cap B \cap C)^c = A^c \cup B^c \cup C^c$. *Wskazówka*: Najpierw zastosuj prawo De Morgana 9b do zbiorów A i $B \cap C$. Można uniknąć stosowania metody elementów zbiorów.

10. Udowodnij prawdziwość następujących zdań nie używając diagramów Venna:
(a) $A \cap B \subseteq A$ i $A \subseteq A \cup B$ dla dowolnych zbiorów A i B.
(b) Jeśli $A \subseteq B$ i $A \subseteq C$, to $A \subseteq B \cap C$.
(c) Jeśli $A \subseteq C$ i $B \subseteq C$, to $A \cup B \subseteq C$.
(d) $A \subseteq B$ wtedy i tylko wtedy, gdy $B^c \subseteq A^c$.

11. Niech $A = \{a, b, c\}$ i $B = \{a, b, d\}$.

(a) Wypisz lub narysuj wszystkie pary uporządkowane ze zbioru $A \times A$.
(b) Wypisz lub narysuj wszystkie pary uporządkowane ze zbioru $A \times B$.
(c) Wypisz lub narysuj elementy zbioru $\{(x, y) \in A \times B\colon x = y\}$.

12. Niech $S = \{0, 1, 2, 3, 4\}$ i niech $T = \{0, 2, 4\}$.
(a) Ile par uporządkowanych należy do zbioru $S \times T$, a ile do zbioru $T \times S$?
(b) Wypisz lub narysuj elementy zbioru $\{(m, n) \in S \times T\colon m < n\}$.
(c) Wypisz lub narysuj elementy zbioru $\{(m, n) \in T \times S\colon m < n\}$.
(d) Wypisz lub narysuj elementy zbioru $\{(m, n) \in S \times T\colon m + n \geq 3\}$.
(e) Wypisz lub narysuj elementy zbioru $\{(m, n) \in T \times S\colon mn \geq 4\}$.
(f) Wypisz lub narysuj elementy zbioru $\{(m, n) \in S \times S\colon m + n = 10\}$.

13. Wypisz elementy tych spośród następujących zbiorów, które mają mniej niż siedem elementów. Dla większych zbiorów wypisz dokładnie siedem elementów danego zbioru.
(a) $\{(m, n) \in \mathbb{N}^2\colon m = n\}$,
(b) $\{(m, n) \in \mathbb{N}^2\colon$ liczba $m + n$ jest pierwsza$\}$,
(c) $\{(m, n) \in \mathbb{P}^2\colon m = 6\}$,
(d) $\{(m, n) \in \mathbb{P}^2\colon \min\{m, n\} = 3\}$,
(e) $\{(m, n) \in \mathbb{P}^2\colon \max\{m, n\} = 3\}$,
(f) $\{(m, n) \in \mathbb{N}^2\colon m^2 = n\}$.

14. Narysuj diagram Venna dla czterech zbiorów A, B, C i D. Zadbaj o to, by był narysowany każdy z szesnastu obszarów, takich jak na przykład $A \cap B^c \cap C^c \cap D$.
Uwaga. W pozostałych ćwiczeniach możesz obrać dowolną metodę dowodu.

15. Dla każdego z następujących zdań udowodnij je lub wykaż, że jest ono fałszywe. (Dowód wymaga rozumowania ogólnego, ale jeden kontrprzykład wystarczy do wykazania, że zdanie jest fałszywe.)
(a) $A \cap B = A \cap C$ implikuje $B = C$.
(b) $A \cup B = A \cup C$ implikuje $B = C$.
(c) $A \cap B = A \cap C$ i $A \cup B = A \cup C$ implikują $B = C$.
(d) $A \cup B \subseteq A \cap B$ implikuje $A = B$.
(e) $A \oplus B = A \oplus C$ implikuje $B = C$.

16. (a) Udowodnij, że $A \subseteq B$ wtedy i tylko wtedy, gdy $A \cup B = B$. Oznacza to, że musisz udowodnić dwa zdania: „$A \subseteq B$ implikuje $A \cup B = B$" i „$A \cup B = B$ implikuje $A \subseteq B$".
(b) Udowodnij, że $A \subseteq B$ wtedy i tylko wtedy, gdy $A \cap B = A$.

17. (a) Udowodnij, że odejmowanie zbiorów nie jest działaniem przemiennym, tzn. może nie być prawdziwa równość $A \setminus B = B \setminus A$.
(b) Udowodnij, że odejmowanie zbiorów nie jest działaniem łącznym, tzn. równość $(A \setminus B) \setminus C = A \setminus (B \setminus C)$ może nie być prawdziwa.
(c) Wykaż jednakże, że $(A \setminus B) \setminus C \subseteq A \setminus (B \setminus C)$ dla dowolnych zbiorów A, B i C.

§ 1.3. Funkcje

Na początku podamy roboczą, opisową definicję funkcji. **Funkcja** f przyporządkowuje każdemu elementowi x z pewnego zbioru S dokładnie jeden element pewnego zbioru T. Mówimy wtedy, że taka funkcja f **jest określona na zbiorze** S **i ma wartości w zbiorze** T. Zbiór S nazywamy **dziedziną funkcji** f i czasami oznaczamy przez $\mathbf{Dom}(f)$[1]. Element przyporządkowany elementowi x jest zazwyczaj oznaczany przez $f(x)$. Należy szczególnie uważać, by nie pomylić samej funkcji f z jej wartościami $f(x)$, zwłaszcza wtedy, gdy tak jak często się to robi, będziemy pisać „funkcja $f(x)$". Funkcja f jest wyznaczona jednoznacznie przez:

(a) zbiór, na którym jest określona, mianowicie $\mathrm{Dom}(f)$;

(b) przyporządkowanie, regułę lub wzór podające wartość $f(x)$ dla każdego $x \in \mathrm{Dom}(f)$.

Dla x należących do $\mathrm{Dom}(f)$, $f(x)$ nazywamy też **wartością elementu** x **przy funkcji** f. Zbiór wszystkich wartości $f(x)$ jest podzbiorem zbioru T, nazywamy go **przeciwdziedziną funkcji** f lub zbiorem wartości funkcji f i oznaczamy przez $\mathrm{Im}(f)$[2]. Mamy zatem

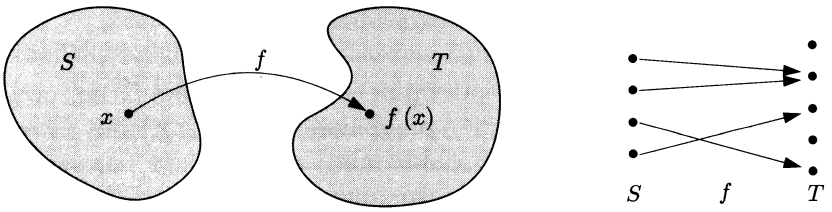
$$\mathrm{Im}(f) = \{f(x)\colon x \in \mathrm{Dom}(f)\}.$$

Często wygodnie jest wyróżnić zbiór T możliwych wartości funkcji f, tzn. zbiór zawierający $\mathrm{Im}(f)$. Mówimy wtedy, że f jest **funkcją o wartościach w zbiorze** T. Funkcja f ma dokładnie jedną dziedzinę $\mathrm{Dom}(f)$ i dokładnie jedną przeciwdziedzinę $\mathrm{Im}(f)$, natomiast dowolny zbiór zawierający $\mathrm{Im}(f)$ może być podany jako zbiór, w którym funkcja f ma wartości. Oczywiście, jeśli podajemy zbiór, w którym funkcja f ma wartości, to staramy się wybrać go w sposób użyteczny lub dający informacje w danym kontekście. Oznaczenie $f\colon S \to T$ jest skrótem stwierdzenia: „f jest funkcją o dziedzinie S i wartościach w zbiorze T". Czasami funkcję f nazywamy **przekształceniem** lub **odwzorowaniem** i mówimy, że przekształca (odwzorowuje) ona zbiór S w zbiór T. Kiedy czujemy potrzebę graficznego przedstawienia funkcji, to robimy rysunki, takie jak rysunek 1.7.

PRZYKŁAD 1 (a) Rozważmy funkcję $f\colon \mathbb{R} \to \mathbb{R}$. Oznacza to, że $\mathrm{Dom}(f) = \mathbb{R}$ oraz dla każdej liczby $x \in \mathbb{R}$, $f(x)$ oznacza jedną liczbę

[1] Stosowane są również inne oznaczenia dziedziny funkcji, np. $\mathrm{D}(f)$, $\mathrm{dom}(f)$, $\mathrm{dm}(f)$.

[2] Przeciwdziedzinę funkcji f oznaczamy też inaczej, np. $\mathrm{R}(f)$, $\mathrm{Rg}(f)$, $\mathrm{rg}(f)$.



funkcja f przekształcająca zbiór S w zbiór T

Rysunek 1.7

ze zbioru $\mathbb{R}$. Zatem funkcja f przyjmuje wartości w zbiorze $\mathbb{R}$, ale jej przeciwdziedzina może być znacznie mniejszym zbiorem. Na przykład, jeśli $f_1(x) = x^2$ dla wszystkich $x \in \mathbb{R}$, to $\mathrm{Im}(f_1) = [0, \infty)$ i możemy napisać, że $f_1\colon \mathbb{R} \to [0, \infty)$. Jeśli funkcja f_2 jest zdefiniowana w następujący sposób:

$$f_2(x) = \begin{cases} 1 & \text{dla} \quad x \geq 0, \\ 0 & \text{dla} \quad x < 0, \end{cases}$$

to $\mathrm{Im}(f_2) = \{0, 1\}$ i możemy napisać, że $f_2\colon \mathbb{R} \to [0, \infty)$, $f_2\colon \mathbb{R} \to \mathbb{N}$ lub $f_2\colon \mathbb{R} \to \{0, 1\}$ itp.

(b) Przypomnijmy, że **wartość bezwzględna** $|x|$ liczby rzeczywistej x jest określona wzorem:

$$|x| = \begin{cases} x & \text{dla} \quad x \geq 0, \\ -x & \text{dla} \quad x < 0. \end{cases}$$

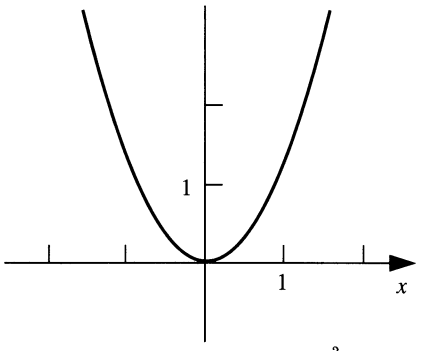
Funkcja f zdefiniowana za pomocą wzoru $f(x) = |x|$ jest funkcją o dziedzinie $\mathbb{R}$ i przeciwdziedzinie $[0, \infty)$; zauważmy, że $|x| \geq 0$ dla wszystkich liczb $x \in \mathbb{R}$. Wartość bezwzględna ma dwie ważne własności, z których skorzystamy w § 1.6 i do których powrócimy w następnym rozdziale: $|x \cdot y| = |x| \cdot |y|$ i $|x + y| \leq |x| + |y|$ dla wszystkich $x, y \in \mathbb{R}$.

(c) Rozważmy funkcję $g\colon \mathbb{N} \to \mathbb{N}$ zdefiniowaną wzorem $g(n) = n^2 - n$. W tym przypadku może być korzystne podanie zbioru $\mathbb{N}$ jako zbioru, w którym funkcja g przyjmuje swoje wartości, gdyż prawdopodobnie nie będzie nas interesować sam zbiór $\mathrm{Im}(g)$. ■

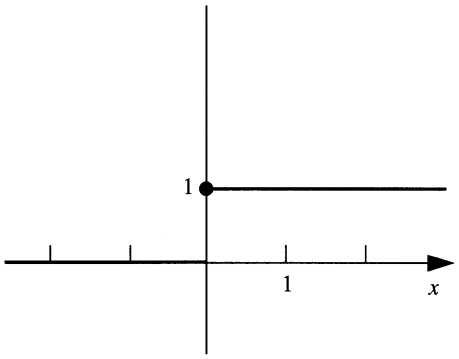
Rozważmy funkcję $f\colon S \to T$. **Wykresem funkcji** f nazywamy następujący podzbiór zbioru $S \times T$:

$$\mathrm{Wykres}(f) = \{(x, y) \in S \times T\colon y = f(x)\}.$$

Ta definicja jest zgodna z definicją wykresu funkcji w algebrze i analizie matematycznej. Wykresy funkcji z przykładu 1(a) są naszkicowane na rysunku 1.8.



wykres funkcji $f_1(x) = x^2$



wykres funkcji f_2

Rysunek 1.8

Nasza robocza definicja „funkcji" jest niepełna; termin „przyporządkowuje" nie został zdefiniowany. Można podać bardzo dokładną, teoriomnogościową definicję funkcji. Opiera się ona na następującej podstawowej obserwacji: nie tylko funkcja określa swój wykres, ale jest też ona jednoznacznie wyznaczona przez wykres. W rzeczywistości wykres funkcji $f\colon S \longrightarrow T$ jest podzbiorem G zbioru $S \times T$, mającym następującą własność:

dla każdego $x \in S$ istnieje dokładnie jeden element $y \in T$ taki, że $(x, y) \in G$.

Dla danego zbioru G mamy $\mathrm{Dom}(f) = S$ i dla każdego $x \in S$, $f(x)$ jest jedynym elementem zbioru T takim, że $(x, f(x)) \in G$. Należy zauważyć, że niczego nie tracimy przyjmując, że funkcje i wykresy są tym samym, a zyskujemy większą ścisłość definicji. Funkcja o dziedzinie S i wartościach w zbiorze T jest podzbiorem

G zbioru $S \times T$ spełniającym warunek:

dla każdego $x \in S$ istnieje dokładnie jeden element $y \in T$ taki, że $(x, y) \in G$.

Jeśli S i T są podzbiorami zbioru $\mathbb{R}$ oraz jeśli zbiór $S \times T$ jest narysowany tak, że S jest częścią osi poziomej, a T jest częścią osi pionowej, to podzbiór G zbioru $S \times T$ jest funkcją (lub wykresem funkcji), jeśli każda pionowa linia prosta przechodząca przez punkt zbioru S przecina zbiór G w dokładnie jednym punkcie.

Funkcja $f\colon S \to T$ jest **funkcją różnowartościową** (własność tę oznaczamy często symbolem „1 – 1"), jeśli różnym elementom zbioru S funkcja f przyporządkowuje różne wartości w zbiorze T:

$$\text{jeśli } x_1, x_2 \in S \text{ i } x_1 \neq x_2, \text{ to } f(x_1) \neq f(x_2).$$

Warunek ten jest logicznie równoważny z często używanym warunkiem:

$$\text{jeśli } x_1, x_2 \in S \text{ i } f(x_1) = f(x_2), \text{ to } x_1 = x_2.$$

Korzystając z definicji funkcji f jako wykresu G, mówimy, że funkcja f jest różnowartościowa wtedy i tylko wtedy, gdy:

dla każdego $y \in T$ istnieje co najwyżej jeden element $x \in S$ taki, że $(x, y) \in G$.

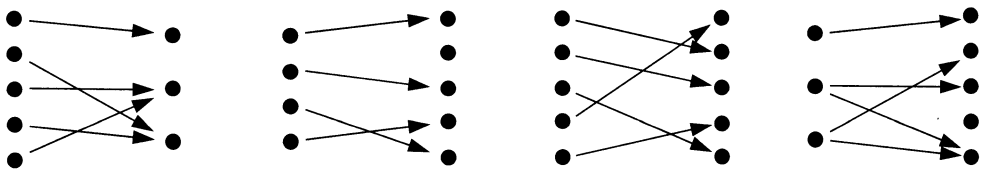
Jeśli S i T są podzbiorami zbioru $\mathbb{R}$ i funkcję f identyfikujemy z jej wykresem G, to warunek ten oznacza, że poziome linie proste przecinają G w co najwyżej jednym punkcie.

Dla danej funkcji $f\colon S \to T$ mówimy, że **funkcja** f przekształca zbiór S **na** podzbiór B zbioru T, jeśli $B = \mathrm{Im}(f)$. W szczególności mówimy, że funkcja f przekształca zbiór S na zbiór T, jeśli $\mathrm{Im}(f) = T$. Jeśli utożsamimy funkcję f z jej wykresem G, to powiemy, że f przekształca zbiór S na zbiór T wtedy i tylko wtedy, gdy:

dla każdego $y \in T$ istnieje co najmniej jeden $x \in S$ taki, że $(x, y) \in G$.

Funkcję $f\colon S \to T$, która jest różnowartościowa i przekształca zbiór S na zbiór T, nazywamy **przekształceniem wzajemnie jednoznacznym** zbioru S na zbiór T. Zatem f jest przekształceniem wzajemnie jednoznacznym wtedy i tylko wtedy, gdy:

dla każdego $y \in T$ istnieje dokładnie jedno $x \in S$ takie, że $(x, y) \in G$.



Rysunek 1.9

Te trzy specjalne rodzaje funkcji są zilustrowane na rysunku 1.9.

Zanim zajmiemy się przykładami matematycznymi, przedstawimy te pojęcia w przypadku niematematycznym.

PRZYKŁAD 2 Przypuśćmy, że każdy student w grupie S ma przypisany numer miejsca ze zbioru $T = \{1, 2, \ldots, 75\}$. To przypisanie określa funkcję $f\colon S \to T$; zatem dla każdego studenta s, $f(s)$ określa jego (lub jej) numer miejsca. Funkcja będzie różnowartościowa, jeśli dwaj różni studenci nie będą przypisani do tego samego miejsca. W tym przypadku grupa nie może liczyć więcej niż 75 studentów. Funkcja będzie przekształcać zbiór S na zbiór T, jeśli każdy numer w zbiorze T będzie przypisany co najmniej jednemu studentowi. Zauważmy, że w tym przypadku grupa musi liczyć co namniej 75 studentów. Jedyną możliwością, by f było przekształceniem wzajemnie jednoznacznym zbioru S na zbiór T jest, by grupa liczyła dokładnie 75 studentów.

Jeśli potraktujemy tę funkcję f jako zbiór par uporządkowanych, to będzie się ona składać z par ze zbioru $S \times T$, takich jak na przykład (Anna Kowalska, 73). ■

PRZYKŁAD 3 (a) Definiujemy funkcję $f\colon \mathbb{N} \to \mathbb{N}$ korzystając ze wzoru $f(n) = 2n$. Wtedy funkcja f jest różnowartościowa, ponieważ

$$f(n_1) = f(n_2) \text{ implikuje } 2n_1 = 2n_2,$$
$$\text{co z kolei implikuje } n_1 = n_2.$$

Jednakże funkcja f nie przekształca zbioru $\mathbb{N}$ na zbiór $\mathbb{N}$, ponieważ $\mathrm{Im}(f)$ składa się tylko z liczb naturalnych parzystych.

(b) Niech Σ będzie alfabetem. Wtedy długość$(w) \in \mathbb{N}$ dla każdego słowa w ze zbioru Σ^*; zob. § 1.1. Zatem „długość” jest funkcją ze zbioru Σ^* na zbiór $\mathbb{N}$. (Zauważ, że funkcje mogą mieć bardziej wyszukane nazwy niż „f”). Aby się o tym przekonać, zauważmy, że zbiór Σ jest niepusty, a więc Σ zawiera jakąś literę, na przykład a. Wtedy 0 = długość(λ), 1 = długość(a),

2 = długość(aa) itd. Funkcja długość nie jest funkcją różnowartościową, chyba, że zbiór Σ ma tylko jeden element. ∎

PRZYKŁAD 4 Udowodnimy, że funkcja $f\colon \mathbb{R} \to \mathbb{R}$ określona wzorem $f(x) = 3x - 5$ jest przekształceniem wzajemnie jednoznacznym zbioru $\mathbb{R}$ na $\mathbb{R}$. Aby sprawdzić, że f jest różnowartościowa, musimy pokazać, że

$$\text{jeśli } f(x) = f(x'), \text{ to } x = x',$$

to znaczy

$$\text{jeśli } 3x - 5 = 3x' - 5, \text{ to } x = x'.$$

Ale jeżeli $3x - 5 = 3x' - 5$, to $3x = 3x'$ (dodaj 5 do obu stron), a to implikuje, że $x = x'$ (podziel obie strony przez 3).

Aby pokazać, że f przekształca zbiór $\mathbb{R}$ na $\mathbb{R}$, weźmy element y ze zbioru $\mathbb{R}$. Musimy znaleźć w zbiorze $\mathbb{R}$ taki element x, że $f(x) = y$, tzn. $3x - 5 = y$. Rozwiązujemy równanie z niewiadomą x i otrzymujemy $x = (y + 5)/3$. Zatem dla danego y z $\mathbb{R}$ liczba $(y + 5)/3$ należy do $\mathbb{R}$ i $f((y + 5)/3) = 3((y + 5)/3) - 5 = y$. To pokazuje, że każdy element y ze zbioru $\mathbb{R}$ należy do $\mathrm{Im}(f)$, a zatem f przekształca $\mathbb{R}$ na $\mathbb{R}$. ∎

Pewne szczególne funkcje pojawiają się tak często, że mają one specjalne nazwy. Niech S będzie zbiorem niepustym. **Funkcją identycznościową** 1_S na zbiorze S nazywamy funkcję, która przekształca każdy element zbioru S na siebie samego:

$$1_S(x) = x \text{ dla wszystkich } x \in S.^3$$

Zatem funkcja identycznościowa jest przekształceniem wzajemnie jednoznacznym zbioru S na S.

Funkcję $f\colon S \to T$ nazywamy **funkcją stałą**, jeśli istnieje element $y_0 \in T$ taki, że $f(x) = y_0$ dla wszystkich $x \in S$. Wartość, jaką przyjmuje funkcja stała, nie zmienia się, gdy x przebiega zbiór S.

Weźmy zbiór S i jego podzbiór A. Funkcję określoną na zbiorze S, która przyjmuje wartość 1 dla elementów zbioru A i wartość 0 dla innych elementów zbioru S, nazywamy **funkcją charakterystyczną** zbioru A i oznaczamy przez χ_A (mała grecka litera chi z indeksem A). Zatem

$$\chi_A(x) = \begin{cases} 1 & \text{dla } x \in A, \\ 0 & \text{dla } x \in S \backslash A. \end{cases}$$

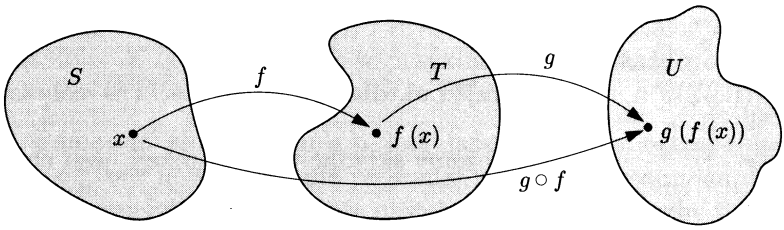
[3] Funkcję identycznościową często oznaczamy przez id_S, Id_S lub I_S.

Zauważmy, że funkcja χ_A: $S \to \{0,1\}$ rzadko jest funkcją różnowartościową i często jest przekształceniem „na". W rzeczywistości funkcja χ_A przekształca zbiór S na zbiór $\{0,1\}$, chyba, że $A = S$ lub $A = \emptyset$. Jeśli A lub $S \backslash A$ ma co najmniej dwa elementy, to funkcja χ_A nie jest funkcją różnowartościową.

Weźmy teraz funkcje f: $S \to T$ i g: $T \to U$; zob. rysunek 1.10. Definiujemy **złożenie funkcji** $g \circ f$: $S \to U$ za pomocą wzoru

$$g \circ f(x) = g(f(x)) \quad \text{dla wszystkich} \quad x \in S.$$

Lewą stronę można czytać jako „g złożona z f od x" lub „g od f od x". Bardziej złożone operacje wykonywane w analizie matematycznej lub skomplikowane działania wykonywane na kalkulatorze mogą być przedstawione jako złożenie prostszych funkcji.



składanie funkcji

Rysunek 1.10

PRZYKŁAD 5 (a) Weźmy funkcję h: $\mathbb{R} \to \mathbb{R}$ daną wzorem

$$h(x) = (x^3 + 2x)^7.$$

Wartość $h(x)$ otrzymujemy obliczając najpierw $x^3 + 2x$, a następnie podnosząc wynik do siódmej potęgi. Przez f oznaczymy pierwszą (lub wewnętrzną) funkcję: $f(x) = x^3 + 2x$. Przez g oznaczymy drugą (lub zewnętrzną) funkcję: $g(x) = x^7$. Nazwa zmiennej x jest nieistotna; możemy równie dobrze napisać $g(y) = y^7$ dla $y \in \mathbb{R}$. Tak czy inaczej, widzimy, że

$$g(f(x)) = g(x^3 + 2x) = (x^3 + 2x)^7 = h(x) \quad \text{dla} \quad x \in \mathbb{R}.$$

Zatem $h = g \circ f$. Umiejętność przedstawienia skomplikowanych funkcji jako złożenia prostszych funkcji jest bardzo ważną w analizie matematycznej. Zauważmy, że kolejność funkcji f i g ma znaczenie. Istotnie

$$f \circ g(x) = f(x^7) = (x^7)^3 + 2(x^7) = x^{21} + 2x^7 \quad \text{dla} \quad x \in \mathbb{R}.$$

(b) Przypuśćmy, że chcemy obliczyć za pomocą kalkulatora $h(x) = \sqrt{\log x}$ dla pewnych dodatnich wartości x. Kalkulator ma funkcje $\sqrt{x}$ i $\log x$ (co oznacza $\log_{10} x$). Zaczynamy obliczenia „od wewnątrz". Na przykład, jeśli $x = 73$, wprowadzamy tę wartość, obliczamy $\log x$, otrzymując $1,8633$, a następnie obliczamy $\sqrt{x}$, otrzymując $1,3650$. Zauważmy, że $h = g \circ f$, gdzie $f(x) = \log x$ i $g(x) = \sqrt{x}$. Podobnie jak w przykładzie (a) kolejność jest istotna: $h \neq f \circ g$, tzn. $\sqrt{\log x}$ na ogół nie jest równy $\log \sqrt{x}$. Na przykład, jeśli $x = 73$, to $\sqrt{x}$ wynosi w przybliżeniu $8,5440$, a $\log \sqrt{x}$ jest równy w przybliżeniu $0,9317$.

(c) Oczywiście złożenie niektórych funkcji f i g jest przemienne, tzn. $f \circ g = g \circ f$. Na przykład, jeśli $f(x) = \sqrt{x}$ i $g(x) = 1/x$ dla $x \in (0, \infty)$, to $f \circ g = g \circ f$, ponieważ

$$\sqrt{\frac{1}{x}} = \frac{1}{\sqrt{x}} \quad \text{dla} \quad x \in (0, \infty).$$

Na przykład, dla $x = 9$ mamy $\sqrt{1/9} = 1/3 = 1/\sqrt{9}$. ∎

Możemy składać więcej funkcji, jeśli chcemy.

PRZYKŁAD 6 Definiujemy funkcje f, g i h, które przekształcają zbiór $\mathbb{R}$ w $\mathbb{R}$ za pomocą wzorów

$$f(x) = x^4, \quad g(y) = \sqrt{y^2 + 1}, \quad h(z) = z^2 + 72.$$

Użyliśmy różnych nazw zmiennych x, y i z, aby poniższe obliczenia były bardziej czytelne. Obliczmy $h \circ (g \circ f)$ oraz $(h \circ g) \circ f$ i porównajmy wyniki. Po pierwsze, dla $x \in \mathbb{R}$ mamy

$$\begin{aligned}
(h \circ (g \circ f))(x) &= h(g \circ f(x)) && \text{z definicji } h \circ (g \circ f) \\
&= h(g(f(x))) && \text{z definicji } g \circ f \\
&= h(g(x^4)) && \text{ponieważ } f(x) = x^4 \\
&= h(\sqrt{x^8 + 1}) && y = x^4 \\
& && \text{w definicji funkcji } g \\
&= (\sqrt{x^8 + 1})^2 + 72 && z = \sqrt{x^8 + 1} \\
& && \text{w definicji funkcji } h \\
&= x^8 + 73 && \text{przekształcenia} \\
& && \text{algebraiczne.}
\end{aligned}$$

Z drugiej strony,

$$\begin{aligned}((h\circ g)\circ f)(x) &= (h\circ g)(f(x)) && \text{z definicji } (h\circ g)\circ f\\ &= h(g(f(x))) && \text{z definicji } h\circ g\\ &= x^8+73 && \text{dokładnie tak jak wyżej.}\end{aligned}$$

Stąd wynika, że

$(h\circ(g\circ f))(x) = ((h\circ g)\circ f)(x) = x^8+73$ dla wszystkich $x\in\mathbb{R}$,

zatem funkcje $h\circ(g\circ f)$ i $(h\circ g)\circ f$ są równe. To nie jest przypadek, jak pokazuje następne ogólne twierdzenie. ∎

Łączność złożenia funkcji

Weźmy funkcje $f\colon S\to T$, $g\colon T\to U$ i $h\colon U\to V$. Wtedy $h\circ(g\circ f) = (h\circ g)\circ f$.

Dowód tego podstawowego faktu sprowadza się do sprawdzenia, że obie funkcje $h\circ(g\circ f)$ i $(h\circ g)\circ f$ przekształcają zbiór S w zbiór V oraz że tak jak w przykładzie 6, dla każdego $x\in S$ wartości $(h\circ(g\circ f))(x)$ i $((h\circ g)\circ f)(x)$ są równe $h(g(f(x)))$.

Ponieważ złożenie funkcji jest łączne, więc możemy pisać $h\circ g\circ f$ bez nawiasów i nie prowadzi to do nieporozumień. Możemy również zapisać złożenie dowolnej skończonej liczby funkcji nie używając nawiasów.

PRZYKŁAD 7 (a) Jeśli $f(x) = x^4$ dla $x\in[0,\infty)$, $g(x) = \sqrt{x+2}$ dla $x\in[0,\infty)$ oraz $h(x) = x^2+1$ dla $x\in\mathbb{R}$, to

$$\begin{aligned}h\circ g\circ f(x) &= h(g(x^4)) = h(\sqrt{x^4+2}) = (x^4+2)+1\\ &= x^4+3 \quad \text{dla} \quad x\in[0,\infty),\\ f\circ g\circ h(x) &= f(g(x^2+1)) = f(\sqrt{x^2+1+2})\\ &= (x^2+3)^2 \quad \text{dla} \quad x\in\mathbb{R},\\ f\circ h\circ g(x) &= f(h(\sqrt{x+2})) = f(x+2+1)\\ &= (x+3)^4 \quad \text{dla} \quad x\in[0,\infty).\end{aligned}$$

(b) Funkcja F dana wzorem

$$F(x) = (\sqrt{x^2+1}+3)^5 \quad \text{dla} \quad x\in\mathbb{R}$$

może być zapisana jako funkcja $k\circ h\circ g\circ f$, gdzie

$$\begin{aligned}f(x) &= x^2+1 && \text{dla} \quad x\in\mathbb{R},\\ g(x) &= \sqrt{x} && \text{dla} \quad x\in[0,\infty),\\ h(x) &= x+3 && \text{dla} \quad x\in\mathbb{R},\\ k(x) &= x^5 && \text{dla} \quad x\in\mathbb{R}.\end{aligned}$$

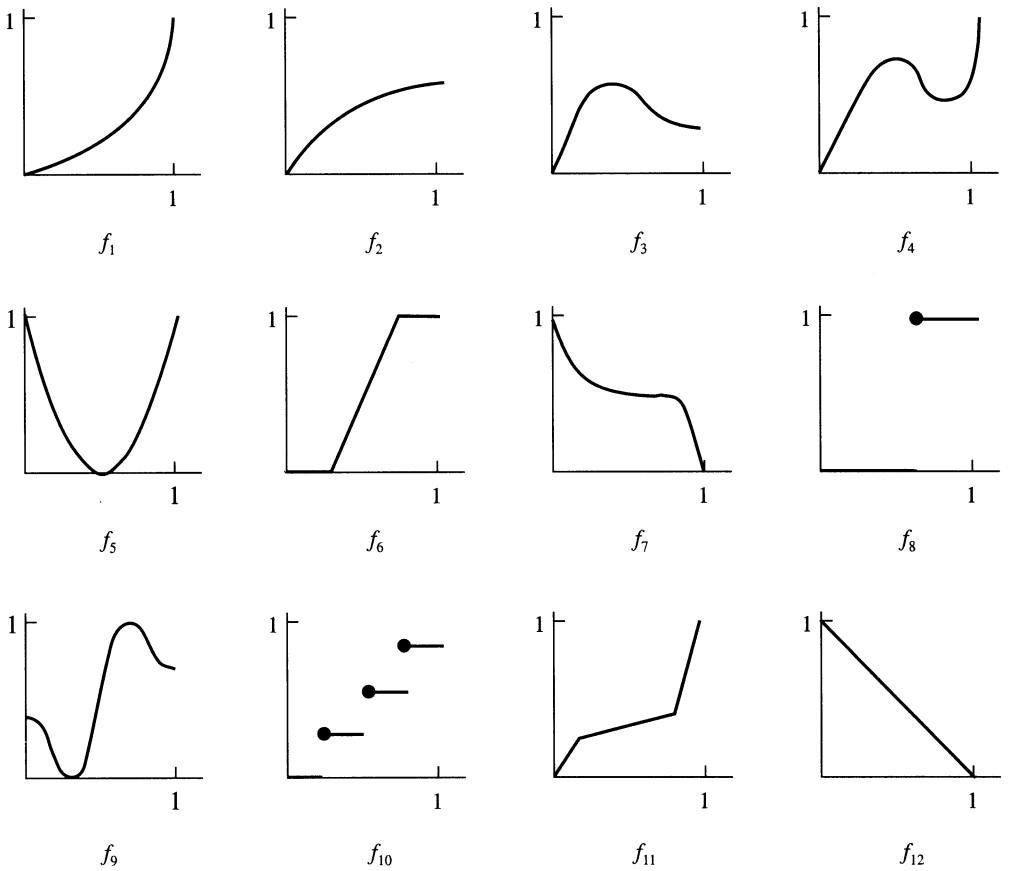
∎

ĆWICZENIA DO § 1.3

1. Definiujemy funkcję $f\colon \mathbb{R} \to \mathbb{R}$ w następujący sposób:

$$f(x) = \begin{cases} x^3, & \text{jeśli } x \geq 1, \\ x, & \text{jeśli } 0 \geq x < 1, \\ -x^3, & \text{jeśli } x < 0. \end{cases}$$

(a) Oblicz $f(3)$, $f(1/3)$, $f(-1/3)$ oraz $f(-3)$.
(b) Naszkicuj wykres funkcji f.
(c) Znajdź $\mathrm{Im}(f)$.



Rysunek 1.11

2. Funkcje przedstawione na rysunku 1.11 mają dziedzinę $[0, 1]$ i przyjmują wartości z tego samego przedziału.
(a) Które z tych funkcji są różnowartościowe?

(b) Które z tych funkcji przekształcają przedział $[0,1]$ na przedział $[0,1]$?
(c) Które z tych funkcji są przekształceniami wzajemnie jednoznacznymi?

3. Niech $S = \{1,2,3,4,5\}$, a $T = \{a,b,c,d\}$. Dla każdego z poniższych pytań: podaj przykład, jeśli odpowiedź brzmi TAK; podaj krótkie wyjaśnienia, jeśli odpowiedź brzmi NIE.
(a) Czy istnieją funkcje różnowartościowe z S w T?
(b) Czy istnieją funkcje różnowartościowe z T w S?
(c) Czy istnieją funkcje przekształcające S na T?
(d) Czy istnieją funkcje przekształcające T na S?
(e) Czy istnieją przekształcenia wzajemnie jednoznaczne z S na T?

4. Niech $S = \{1,2,3,4,5\}$ i weźmy następujące funkcje ze zbioru S w S: $1_S(n) = n$, $f(n) = 6 - n$, $g(n) = \max\{3,n\}$, $h(n) = \max\{1, n-1\}$.
(a) Zapisz każdą z tych funkcji jako zbiór par uporządkowanych, tzn. wypisz elementy ich wykresów.
(b) Naszkicuj wykres każdej z tych funkcji.
(c) Które z tych funkcji są jednocześnie różnowartościowe i „na"?

5. Wzór $f((m,n)) = 2^m 3^n$ określa funkcję różnowartościową ze zbioru $\mathbb{N} \times \mathbb{N}$ w zbiór $\mathbb{N}$. *Uwaga.* Jeśli funkcje są definiowane dla par uporządkowanych, to opuszcza się jedną parę nawiasów. Zatem będziemy pisać $f(m,n) = 2^m 3^n$.
(a) Oblicz $f(m,n)$ dla pięciu różnych par (m,n) ze zbioru $\mathbb{N} \times \mathbb{N}$.
(b) Wyjaśnij, dlaczego funkcja f jest różnowartościowa.
(c) Czy funkcja f przekształca zbiór $\mathbb{N} \times \mathbb{N}$ na zbiór $\mathbb{N}$? Uzasadnij odpowiedź.
(d) Pokaż, że wzór $g(m,n) = 2^m 4^n$ określa funkcję na zbiorze $\mathbb{N} \times \mathbb{N}$, która nie jest różnowartościowa.

6. Weźmy następujące funkcje ze zbioru $\mathbb{N}$ w $\mathbb{N}$: $1_{\mathbb{N}}(n) = n$, $f(n) = 3n$, $g(n) = n + (-1)^n$, $h(n) = \min\{n, 100\}$, $k(n) = \max\{0, n-5\}$.
(a) Które z tych funkcji są różnowartościowe?
(b) Które z tych funkcji przekształcają zbiór $\mathbb{N}$ na zbiór $\mathbb{N}$?

7. Niech A i B będą zbiorami niepustymi. Przekształcenie rzutowania RZUT polega na braniu pierwszego elementu z każdej pary ze zbioru $A \times B$, tzn. funkcja RZUT: $A \times B \longrightarrow A$ jest określona wzorem $\text{RZUT}(a,b) = a$. (Przypomnijmy umowę dotyczącą funkcji określonych na parach uporządkowanych, którą przyjęliśmy w ćwiczeniu 5).
(a) Czy taka funkcja przekształca $A \times B$ na zbiór A? Uzasadnij odpowiedź.
(b) Czy funkcja RZUT jest różnowartościowa? Co będzie, gdy zbiór B ma tylko jeden element?

8. Niech $\Sigma = \{a,b,c\}$ i niech Σ^*będzie zbiorem wszystkich słów w utworzonych za pomocą liter ze zbioru Σ; zob. przykład 3(b). Określamy $L(w) =$ długość(w) dla wszystkich $w \in \Sigma^*$.

(a) Oblicz $L(w)$ dla słów $w_1 = cab$, $w_2 = ababac$ oraz $w_3 = \lambda$.
(b) Czy L jest funkcją różnowartościową? Odpowiedź uzasadnij.
(c) Funkcja L przekształca Σ^* w zbiór $\mathbb{N}$. Czy L przekształca Σ^* na zbiór $\mathbb{N}$? Odpowiedź uzasadnij.
(d) Znajdź wszystkie słowa w takie, że $L(w) = 2$.

9. Dla $n \in \mathbb{Z}$ niech $f(n) = \frac{1}{2}[(-1)^n + 1]$. Funkcja f jest funkcją charakterystyczną pewnego podzbioru zbioru $\mathbb{Z}$. Jaki to podzbiór?

10. W przykładzie 5(b) porównywaliśmy funkcje $\sqrt{\log x}$ i $\log \sqrt{x}$. Pokaż, że funkcje te przyjmują tę samą wartość dla $x = 10000$.

11. Określamy trzy funkcje przekształcające zbiór $\mathbb{R}$ w zbiór $\mathbb{R}$ w następujący sposób: $f(x) = x^3 - 4x$, $g(x) = 1/(x^2 + 1)$, $h(x) = x^4$. Znajdź
(a) $f \circ g \circ h$,
(b) $f \circ h \circ g$,
(c) $h \circ g \circ f$,
(d) $f \circ f$,
(e) $g \circ g$,
(f) $h \circ g$,
(g) $g \circ h$.

12. Pokaż, że jeśli $f\colon S \to T$ i $g\colon T \to U$ są funkcjami różnowartościowymi, to funkcja $g \circ f$ jest też różnowartościowa.

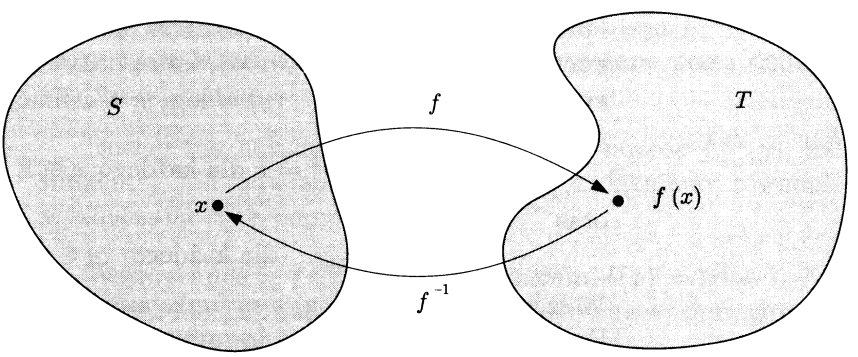
13. Udowodnij, że złożenie funkcji jest łączne.

14. Wartości kilku ważnych funkcji można obliczać za pomocą kalkulatora. Dlaczego nie ma wśród nich funkcji identycznościowej, tzn. funkcji $1_{\mathbb{R}}$, gdzie $1_{\mathbb{R}}(x) = x$ dla wszystkich $x \in \mathbb{R}$?

15. Weźmy funkcje f i g przekształcające zbiór $\mathbb{Z}$ w zbiór $\mathbb{Z}$, gdzie $f(n) = n - 1$ dla $n \in \mathbb{Z}$, a g jest funkcją charakterystyczną χ_E zbioru $E = \{n \in \mathbb{Z}\colon n \text{ jest parzysta}\}$.
(a) Oblicz $(g \circ f)(5)$, $(g \circ f)(4)$, $(f \circ g)(7)$ i $(f \circ g)(8)$.
(b) Oblicz $(f \circ f)(11)$, $(f \circ f)(12)$, $(g \circ g)(11)$ i $(g \circ g)(12)$.
(c) Wyznacz funkcje $g \circ f$ oraz $f \circ f$.
(d) Pokaż, że $g \circ g = g \circ f$ oraz że funkcja $f \circ g$ przyjmuje wartości przeciwne do $g \circ f$.

§ 1.4. Funkcje odwrotne

Z grubsza mówiąc, funkcją odwrotną do funkcji f jest taka funkcja, która cofa działanie f. Wyznaczając najpierw wartość funkcji f, a następnie funkcji do niej odwrotnej, otrzymujemy z powrotem ten element dziedziny funkcji f, dla którego obliczaliśmy wartość funkcji (zob. rysunek 1.12).



funkcja i funkcja do niej odwrotna

Rysunek 1.12

PRZYKŁAD 1 (a) Funkcje x^2 i $\sqrt{x}$ o dziedzinach $[0, \infty)$ są funkcjami wzajemnie odwrotnymi. Jeśli wykonasz na jakiejś liczbie te dwie operacje w dowolnej kolejności, otrzymasz tę początkową liczbę. Wypróbuj to na kalkulatorze! Możemy to zapisać za pomocą wzorów

$$\sqrt{x^2} = x \quad \text{i} \quad (\sqrt{x})^2 = x \text{ dla } x \in [0, \infty).$$

(b) Funkcja $1/x$ jest funkcją odwrotną do siebie samej. Jeśli wykonasz tę operację dwukrotnie na jakiejś liczbie, otrzymasz z powrotem tę liczbę. To znaczy

$$\frac{1}{1/x} = x \text{ dla wszystkich różnych od zera } x \text{ ze zbioru } \mathbb{R}. \quad \blacksquare$$

Oto dokładna definicja. **Funkcją odwrotną** do funkcji $f\colon S \to T$ jest funkcja $f^{-1}\colon T \to S$ taka, że $f^{-1} \circ f = 1_S$ oraz $f \circ f^{-1} = 1_T$, tzn. taka, że

$$f^{-1}(f(x)) = x \quad \text{dla wszystkich} \quad x \in S$$

oraz

$$f(f^{-1}(y)) = y \quad \text{dla wszystkich} \quad y \in T.$$

Nie wszystkie funkcje mają funkcje odwrotne; te, które mają, nazywamy **funkcjami odwracalnymi**. Zobaczymy w dowodzie twierdzenia, że definicja funkcji odwrotnej f^{-1} określa ją jednoznacznie, jeśli taka funkcja istnieje, zatem funkcja odwracalna nie może mieć dwóch różnych funkcji odwrotnych.

PRZYKŁAD 2 Weźmy dodatnią liczbę rzeczywistą $b \neq 1$. W tym przykładzie szczególnie ważne będą trzy wartości liczby b: 2, 10 oraz często występująca w analizie liczba e, w przybliżeniu równa 2,718.

Funkcja f_b dana wzorem $f_b(x) = b^x$ dla $x \in \mathbb{R}$ ma funkcję odwrotną f_b^{-1} o dziedzinie $(0, \infty)$, którą nazywamy **funkcją logarytmiczną**. Piszemy $f_b^{-1}(y) = \log_b y$; z definicji funkcji odwrotnej mamy

$$\log_b b^x = x \text{ dla każdego } x \in \mathbb{R}$$

oraz

$$b^{\log_b y} = y \quad \text{dla każdego} \quad y \in (0, \infty).$$

W szczególności, e^x i $\log_e x$ są funkcjami wzajemnie odwrotnymi. Funkcję $\log_e x$ nazywamy **logarytmem naturalnym** i często oznaczamy $\ln x$. Funkcje 10^x i $\log_{10} x$ są wzajemnie odwrotne, podobnie 2^x i $\log_2 x$. Wartości funkcji $\log_{10} x = \log x$ i $\log_e x = \ln x$ można obliczać na wielu kalkulatorach; takie kalkulatory pozwalają też obliczać wartości funkcji odwrotnych 10^x i e^x. Aby obliczyć $\log_2 x$ na kalkulatorze, korzystamy z jednego z następujących wzorów

$$\log_2 x = \frac{\log_{10} x}{\log_{10} 2} \approx 3,321928 \cdot \log x$$

lub

$$\log_2 x = \frac{\log_e x}{\log_e 2} = \frac{\ln x}{\ln 2} \approx 1,442695 \cdot \ln x. \qquad \blacksquare$$

Następne twierdzenie pokaże nam, jakie funkcje są odwracalne.

Twierdzenie

Funkcja $f\colon S \to T$ jest odwracalna wtedy i tylko wtedy, gdy f jest różnowartościowa i przekształca zbiór S na zbiór T.

Dowód. Przypuśćmy, że f ma funkcję odwrotną f^{-1}. Jeśli $x_1, x_2 \in S$ oraz $f(x_1) = f(x_2)$, to

$$x_1 = f^{-1}(f(x_1)) = f^{-1}(f(x_2)) = x_2.$$

Zatem funkcja f jest różnowartościowa. Ponadto, jeśli $y \in T$, to $f^{-1}(y)$ należy do S i $f(f^{-1}(y)) = y$; zatem $y \in \mathrm{Im}(f)$. Stąd $T = \mathrm{Im}(f)$, a więc f przekształca zbiór S na zbiór T.

Na odwrót, jeśli f przekształca S na T, to dla każdego y ze zbioru T istnieje jakiś $x \in S$ taki, że $f(x) = y$. Jeśli f jest także różnowartościowa, to istnieje dokładnie jeden taki x i funkcję f^{-1} możemy określić wzorem:

$(*)$ $\qquad f^{-1}(y) =$ ten jedyny $x \in S$ taki, że $f(x) = y$.

Z definicji tej mamy bezpośrednio $f(f^{-1}(y)) = y$, a $f^{-1}(f(x))$ jest jedynym elementem zbioru S, który funkcja f przekształca

w $f(x)$, czyli samym elementem x. Zatem funkcja f^{-1} określona za pomocą wzoru $(*)$ spełnia warunki nałożone przez definicję funkcji odwrotnej. ∎

Dowód ten pokazuje też, jak otrzymać wartość $f^{-1}(y)$, jeśli funkcja f jest odwracalna. Po prostu, rozwiązujemy równanie, wyznaczając x za pomocą y.

PRZYKŁAD 3 Weźmy funkcję $f\colon \mathbb{R} \to \mathbb{R}$ daną wzorem $f(x) = x^3 + 1$. Aby przekonać się, że funkcja f jest różnowartościowa, zauważmy, że

$$f(x_1) = f(x_2) \text{ implikuje } x_1^3 + 1 = x_2^3 + 1$$
$$\text{implikuje } x_1^3 = x_2^3 \text{ implikuje } x_1 = x_2;$$

ta ostatnia implikacja zachodzi, ponieważ każda liczba rzeczywista ma dokładnie jeden pierwiastek trzeciego stopnia.

Aby sprawdzić, że funkcja f przekształca $\mathbb{R}$ na $\mathbb{R}$, weźmy y ze zbioru $\mathbb{R}$. Musimy znaleźć liczbę $x \in \mathbb{R}$ taką, że $f(x) = y$, tzn. musimy rozwiązać równanie $x^3 + 1 = y$ z niewiadomą x. Kiedy to zrobimy, otrzymamy liczbę $x = \sqrt[3]{y-1}$, która należy do $\mathbb{R}$. To oznacza, że f przekształca $\mathbb{R}$ na $\mathbb{R}$.

Ponieważ funkcja f jest różnowartościowa i przekształca $\mathbb{R}$ na $\mathbb{R}$, więc na podstawie twierdzenia jest odwracalna. Funkcję f^{-1} wyznaczyliśmy w ostatnim akapicie, kiedy rozwiązaliśmy równanie. Zatem $f^{-1}(y) = \sqrt[3]{y-1}$. Ten wzór ma sens dla każdego y ze zbioru $\mathbb{R}$, a więc funkcja f^{-1} jest dobrze określona. ∎

PRZYKŁAD 4 Weźmy funkcję $g\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ określoną wzorem $g(m,n) = (-n,-m)$. Sprawdzimy, że funkcja g jest różnowartościowa i „na", a następnie znajdziemy funkcję odwrotną do niej. Aby pokazać, że funkcja g jest różnowartościowa, musimy pokazać, że

$$g(m,n) = g(m',n') \text{ implikuje } (m,n) = (m',n').$$

Jeśli $g(m,n) = g(m',n')$, to $(-n,-m) = (-n',-m')$. Ponieważ te pary uporządkowane są równe, otrzymujemy: $-n = -n'$ oraz $-m = -m'$. Stąd $m = m'$ i $n = n'$, a więc $(m,n) = (m',n')$, czego mieliśmy dowieść.

Aby pokazać, że funkcja g przekształca $\mathbb{Z}\times\mathbb{Z}$ na $\mathbb{Z}\times\mathbb{Z}$, weźmy parę (p,q) ze zbioru $\mathbb{Z}\times\mathbb{Z}$. Musimy znaleźć parę (m,n) ze zbioru $\mathbb{Z}\times\mathbb{Z}$ taką, że $g(m,n) = (p,q)$. Zatem musi być spełniona równość $(-n,-m) = (p,q)$, a to oznacza, że n powinno być równe $-p$, a m powinno być równe $-q$. Innymi słowy, dla danej pary (p,q) ze zbioru $\mathbb{Z}\times\mathbb{Z}$, para $(-q,-p)$ jest elementem zbioru $\mathbb{Z}\times\mathbb{Z}$ takim, że $g(-q,-p) = (p,q)$. Zatem g przekształca zbiór $\mathbb{Z}\times\mathbb{Z}$ na $\mathbb{Z}\times\mathbb{Z}$.

Aby znaleźć funkcję odwrotną do funkcji g, musimy wziąć parę (p, q) ze zbioru $\mathbb{Z} \times \mathbb{Z}$ i znaleźć $g^{-1}(p, q)$. Zrobiliśmy to właśnie w ostatnim akapicie; funkcja g przekształca parę $(-q, -p)$ na (p, q), a więc $g^{-1}(p, q) = (-q, -p)$ dla wszystkich par (p, q) ze zbioru $\mathbb{Z} \times \mathbb{Z}$.

Warto zauważyć, że w tym przypadku $g = g^{-1}$. ∎

Funkcje odwrotne są tak użyteczne, że czasami ograniczamy dziedziny funkcji nieróżnowartościowych do zbiorów, na których są one różnowartościowe. Jeśli następnie przyjmiemy, że zbiorem, w którym znajdują się wartości funkcji, jest jej przeciwdziedzina, to otrzymamy funkcję odwracalną.

PRZYKŁAD 5 (a) Weźmy funkcję $f\colon \mathbb{R} \to \mathbb{R}$ określoną wzorem $f(x) = x^2$. Wtedy funkcja f nie jest różnowartościowa, ale będzie różnowartościowa, jeśli ograniczymy jej dziedzinę do zbioru $[0, \infty)$. Zatem definiujemy nową funkcję F tym samym wzorem, $F(x) = x^2$, ale przyjmując $\mathrm{Dom}(F) = [0, \infty)$. Wtedy funkcja F jest różnowartościowa. W rzeczywistości funkcja $F\colon [0, \infty) \to [0, \infty)$ jest różnowartościowa i „na". Jest to ta funkcja, której funkcją odwrotną jest $F^{-1}(x) = \sqrt{x}$; zob. przykład 1(a).

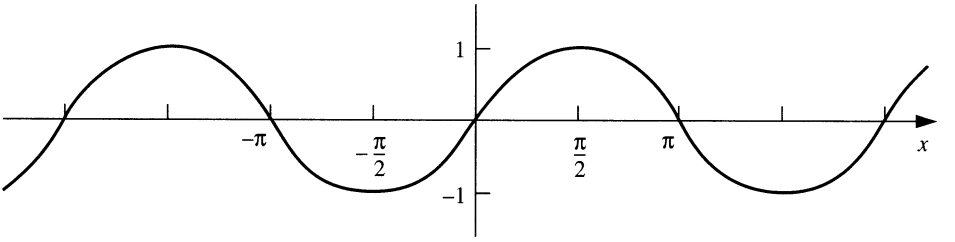
Funkcję F nazywamy **obcięciem funkcji** f do zbioru $[0, \infty)$. Takie obcięcia funkcji są możliwe i potrzebne w wielu różnych sytuacjach.

(b) Będziemy mogli prześledzić następujący przykład, nawet jeśli nie znamy trygonometrii. Okazuje się, że żadna z funkcji trygonometrycznych nie jest różnowartościowa. Na przykład, popatrzmy na wykres funkcji $\sin x$ przedstawiony na rysunku 1.13. Jednakże $\sin x$ jest funkcją różnowartościową, jeśli obetniemy jej dziedzinę na przykład do zbioru $[-\pi/2, \pi/2]$. Na rysunku 1.14(a) obcięcie funkcji $\sin x$ oznaczyliśmy przez $\mathrm{Sin}\, x$. Jest to funkcja odwracalna, której zbiorem wartości jest przedział $[-1, 1]$; funkcja odwrotna jest przedstawiona na rysunku 1.14(b). Jest to funkcja arc sin, spotykana często w trygonometrii, analizie matematycznej i na wielu kalkulatorach. ∎

Weźmy funkcję $f\colon S \to T$. Dla dowolnego podzbioru A zbioru S określamy

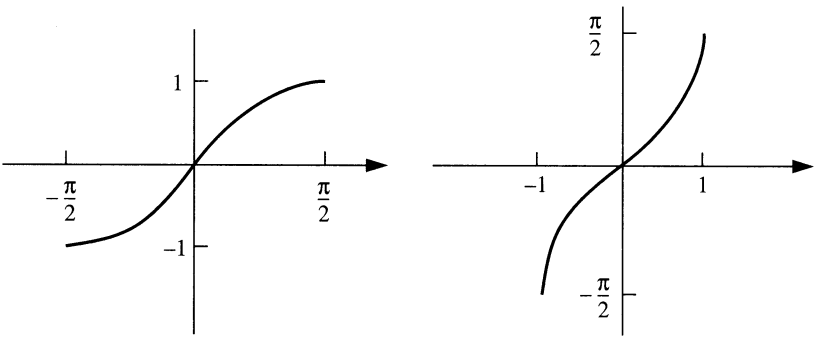
$$f(A) = \{f(x)\colon\ x \in A\}.$$

Zatem $f(A)$ jest zbiorem wszystkich wartości $f(x)$, gdy x przebiega zbiór A. Zbiór $f(A)$ nazywamy **obrazem zbioru** A względem funkcji f. Interesuje nas też operacja odwrotna do obrazu



funkcja sin x

Rysunek 1.13



(a)
funkcja sin x w przedziale $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$

(b)
funkcja odwrotna do (a): arc sin x

Rysunek 1.14

zbioru, dla zbioru B zawartego w zbiorze T:

$$f^{\leftarrow}(B) = \{x \in S\colon\ f(x) \in B\}.$$

Zbiór $f^{\leftarrow}(B)$ nazywamy **przeciwobrazem zbioru B względem funkcji f.**

Jeśli f jest funkcją odwracalną, to (ćwiczenie 16) przeciwobraz podzbioru B zbioru T względem funkcji f jest równy obrazowi zbioru B względem funkcji f^{-1}, tzn. w tym przypadku

$$f^{\leftarrow}(B) = \{f^{-1}(y)\colon\ y \in B\} = f^{-1}(B).$$

Jeśli funkcja f jest nieodwracalna, to oczywiście nie ma sensu pisanie $f^{-1}(y)$ czy $f^{-1}(B)$. Ponieważ $f^{-1}(B)$ nie może mieć żadnego innego znaczenia poza $f^{\leftarrow}(B)$, niektórzy autorzy rozszerzają znaczenie f^{-1} i przez $f^{-1}(B)$ oznaczają to, co my oznaczamy przez $f^{\leftarrow}(B)$, nawet jeśli funkcja f jest nieodwracalna. Bądź ostrożny!

Dla $y \in T$ przez $f^{\leftarrow}(y)$ oznaczamy zbiór $f^{\leftarrow}(\{y\})$. To znaczy

$$f^{\leftarrow}(y) = \{x \in S\colon\ f(x) = y\}.$$

Zbiór ten nazywamy **przeciwobrazem elementu y względem funkcji f**. Zauważmy, że rozwiązanie równania $f(x) = y$ z niewiadomą x jest równoważne znalezieniu zbioru $f^{\leftarrow}(y)$. To znaczy, że $f^{\leftarrow}(y)$ jest zbiorem rozwiązań równania $f(x) = y$. Tak jak w przypadku równań algebraicznych zbiór $f^{\leftarrow}(y)$ może mieć jeden element, wiele elementów lub nie mieć ich wcale.

PRZYKŁAD 6 (a) Weźmy funkcję $f\colon \mathbb{R} \to \mathbb{R}$ daną wzorem $f(x) = x^2$. Wtedy zbiór

$$f^{\leftarrow}(4) = \{x \in \mathbb{R}\colon\ x^2 = 4\} = \{-2, 2\}$$

jest zbiorem rozwiązań równania $x^2 = 4$. Przeciwobraz zbioru $[1, 9]$ jest równy

$$\begin{aligned} f^{\leftarrow}([1,9]) &= \{x \in \mathbb{R}\colon\ x^2 \in [1,9]\} = \{x \in \mathbb{R}\colon\ 1 \leq x^2 \leq 9\} \\ &= [-3,-1] \cup [1,3]. \end{aligned}$$

Mamy również $f^{\leftarrow}([-1,0]) = \{0\}$ i $f^{\leftarrow}([-1,1]) = [-1,1]$.

(b) Weźmy funkcję $g\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ określoną wzorem $g(m,n) = m^2 + n^2$. Wtedy $g^{\leftarrow}(0) = \{(0,0)\}$, $g^{\leftarrow}(1) = \{(0,1),(1,0)\}$, $g^{\leftarrow}(2) = \{(1,1)\}$, $g^{\leftarrow}(3) = \emptyset$, $g^{\leftarrow}(4) = \{(0,2),(2,0)\}$ itd. Na przykład mamy $g^{\leftarrow}(25) = \{(0,5),(3,4),(4,3),(5,0)\}$. ∎

PRZYKŁAD 7 (a) Niech Σ będzie alfabetem i niech L będzie funkcją długości określoną na zbiorze Σ^*: $L(w) =$ długość(w) dla $w \in \Sigma^*$. Jak już zauważyliśmy w przykładzie 3(b) w § 1.3, funkcja L przekształca zbiór Σ^* na $\mathbb{N}$. Dla każdej liczby $k \in \mathbb{N}$

$$L^{\leftarrow}(k) = \{w \in \Sigma^*\colon\ L(w) = k\} = \{w \in \Sigma^*\colon\ \text{długość}(w) = k\}.$$

Zauważmy, że zbiory $L^{\leftarrow}(k)$ dla różnych k są rozłączne i ich sumą jest Σ^*:

$$L^{\leftarrow}(0) \cup L^{\leftarrow}(1) \cup L^{\leftarrow}(2) \cup \ldots = \Sigma^*.$$

(b) Weźmy funkcję $h\colon \mathbb{Z} \to \{-1,1\}$ taką, że $h(n) = (-1)^n$. Wtedy

$$h^{\leftarrow}(1) = \{n \in \mathbb{Z}\colon\ \text{liczba } n \text{ jest parzysta}\}$$

oraz

$$h^{\leftarrow}(-1) = \{n \in \mathbb{Z}\colon\ \text{liczba } n \text{ jest nieparzysta}\}.$$

Te dwa zbiory są rozłączne i ich sumą jest cały zbiór $\mathbb{Z}$:

$$h^{\leftarrow}(1) \cup h^{\leftarrow}(-1) = \mathbb{Z}.$$ ∎

Nie jest to przypadkowe, że przeciwobrazy elementów dzielą dziedzinę funkcji na zbiory rozłączne, jak w ostatnim przykładzie. W paragrafie 3.5 przekonamy się, że tak dzieje się zawsze

i nauczymy się wykorzystywać te podziały dziedziny na zbiory rozłączne.

ĆWICZENIA DO § 1.4

1. Znajdź funkcje odwrotne do następujących funkcji przekształcających $\mathbb{R}$ w $\mathbb{R}$:
(a) $f(x) = 2x + 3$,
(b) $g(x) = x^3 - 2$,
(c) $h(x) = (x-2)^3$,
(d) $k(x) = \sqrt[3]{x} + 7$.

2. W wielu kalkulatorach dane są funkcje $\log x$, x^2, $\sqrt{x}$ i $1/x$.
(a) Określ dziedziny tych funkcji.
(b) Które z tych funkcji są funkcjami wzajemnie odwrotnymi?
(c) Dla których par tych funkcji operacja złożenia jest przemienna?
(d) W niektórych kalkulatorach dane są też funkcje $\sin x$, $\cos x$ i $\operatorname{tg} x$. Jeśli choć trochę znasz trygonometrię, wykonaj polecenia (a), (b), (c) dla tych funkcji.

3. Oto kilka funkcji ze zbioru $\mathbb{N} \times \mathbb{N}$ w zbiór $\mathbb{N}$:
$\text{SUMA}(m,n) = m + n$, $\quad \text{ILOCZYN}(m,n) = m \cdot n$,
$\text{MAX}(m,n) = \max\{m,n\}$, $\quad \text{MIN}(m,n) = \min\{m,n\}$.
(a) Które z tych funkcji przekształcają zbiór $\mathbb{N} \times \mathbb{N}$ na $\mathbb{N}$?
(b) Pokaż, że żadna z tych funkcji nie jest funkcją różnowartościową.
(c) Jak duży jest zbiór $F^{\leftarrow}(4)$ dla każdej z tych funkcji F?

4. Oto kilka funkcji ze zbioru $\mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N})$ w zbiór $\mathcal{P}(\mathbb{N})$:
$\text{SUMA}(A,B) = A \cup B$, $\text{ILOCZYN}(A,B) = A \cap B$ i $\text{SYM}(A,B) = A \oplus B$.
(a) Pokaż, że każda z tych funkcji przekształca $\mathcal{P}(\mathbb{N}) \times \mathcal{P}(\mathbb{N})$ na $\mathcal{P}(\mathbb{N})$.
(b) Pokaż, że żadna z tych funkcji nie jest funkcją różnowartościową.
(c) Jak duży jest zbiór $F^{\leftarrow}(\emptyset)$ oraz zbiór $F^{\leftarrow}(\{0\})$ dla każdej z tych funkcji F?

5. Oto dwie funkcje przekształcające zbiór $\mathbb{N}$ w $\mathbb{N}$: $f(n) = n + 1$ oraz $g(n) = \max\{0, n-1\}$ dla $n \in \mathbb{N}$.
(a) Oblicz $f(n)$ dla $n = 0, 1, 2, 3, 4, 73$.
(b) Oblicz $g(n)$ dla $n = 0, 1, 2, 3, 4, 73$.
(c) Pokaż, że funkcja f jest różnowartościowa, ale nie przekształca zbioru $\mathbb{N}$ na $\mathbb{N}$.
(d) Pokaż, że funkcja g przekształca zbiór $\mathbb{N}$ na $\mathbb{N}$, ale nie jest różnowartościowa.
(e) Pokaż, że $g \circ f = 1_{\mathbb{N}}$, ale $f \circ g \neq 1_{\mathbb{N}}$.

6. Definiujemy $f\colon \mathbb{N} \to \mathbb{N}$ oraz $g\colon \mathbb{N} \to \mathbb{N}$ w następujący sposób: $f(n) = 2n$ dla wszystkich $n \in \mathbb{N}$, $g(n) = n/2$, jeśli liczba n jest parzysta oraz $g(n) = (n-1)/2$, jeśli liczba n jest nieparzysta.

(a) Oblicz $g(n)$ dla $n = 0, 1, 2, 3, 4, 73$.
(b) Pokaż, że $g \circ f = 1_{\mathbb{N}}$, ale $f \circ g \neq 1_{\mathbb{N}}$.

7. Jeśli $f\colon S \to S$ oraz $f \circ f = 1_S$, to funkcja f jest funkcją odwrotną do siebie samej. Pokaż, że następujące funkcje są odwrotne do siebie samych.
(a) Funkcja $f\colon (0,\infty) \to (0,\infty)$ dana wzorem $f(x) = 1/x$.
(b) Funkcja $\Phi\colon \mathcal{P}(S) \to \mathcal{P}(S)$ określona wzorem $\Phi(A) = A^c$.
(c) Funkcja $g\colon \mathbb{R} \to \mathbb{R}$ dana wzorem $g(x) = 1 - x$.
(d) Funkcja ODWR: $C \times C \to C \times C$ (gdzie C jest dowolnym zbiorem) określona wzorem $\text{ODWR}(x, y) = (y, x)$.

8. Niech A będzie podzbiorem pewnego zbioru S i niech funkcja χ_A będzie funkcją charakterystyczną zbioru A. Znajdź $\chi_A^{\leftarrow}(1)$ oraz $\chi_A^{\leftarrow}(0)$.

9. Niech $f\colon S \to T$ i $g\colon T \to U$ będą funkcjami odwracalnymi. Pokaż, że $g \circ f$ jest funkcją odwracalną oraz że $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$.

10. Niech $f\colon S \to T$ będzie funkcją odwracalną. Pokaż, że f^{-1} jest funkcją odwracalną oraz że $(f^{-1})^{-1} = f$.

11. Weźmy funkcje $f\colon S \to T$ i $g\colon T \to S$ takie, że $g \circ f = 1_S$. Nietrywialne przykłady takich par funkcji podaliśmy w ćwiczeniach 5 i 6.
(a) Udowodnij, że funkcja f jest różnowartościowa.
(b) Udowodnij, że funkcja g przekształca zbiór T na zbiór S.

12. Weźmy funkcję $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ określoną wzorem

$$f(x, y) = (x + y, x - y).$$

(a) Udowodnij, że f jest funkcją różnowartościową w zbiorze $\mathbb{R} \times \mathbb{R}$.
(b) Udowodnij, że funkcja f przekształca zbiór $\mathbb{R} \times \mathbb{R}$ na $\mathbb{R} \times \mathbb{R}$.
(c) Znajdź funkcję odwrotną f^{-1}.
(d) Znajdź funkcje złożone $f \circ f^{-1}$ i $f \circ f$.

13. Niech $f\colon S \to T$.
(a) Pokaż, że $f(f^{\leftarrow}(B)) \subseteq B$ dla dowolnego podzbioru B zbioru T.
(b) Pokaż, że $A \subseteq f^{\leftarrow}(f(A))$ dla dowolnego podzbioru A zbioru S.
(c) Pokaż, że $f^{\leftarrow}(B_1 \cap B_2) = f^{\leftarrow}(B_1) \cap f^{\leftarrow}(B_2)$ dla dowolnych podzbiorów B_1 i B_2 zbioru T.
(d) Jakie warunki musi spełniać zbiór B, by w punkcie (a) zachodziła równość?

14. Niech $f\colon S \to T$. Udowodnij, że następujące zdania są prawdziwe lub wykaż, że są fałszywe. Jeśli są fałszywe, to wystarczy podać jeden kontrprzykład.
(a) $f(A_1 \cap A_2) = f(A_1) \cap f(A_2)$, jeśli A_1 i A_2 są podzbiorami zbioru S.
(b) $f(A_1 \setminus A_2) = f(A_1) \setminus f(A_2)$, jeśli A_1 i A_2 są podzbiorami zbioru S.
(c) $f(A_1) = f(A_2)$ implikuje $A_1 = A_2$.

15. (a) Można pokazać, że jeśli funkcja $f\colon T \to U$ jest różnowartościowa oraz funkcje $g\colon S \to T$ i $h\colon S \to T$ spełniają warunek $f \circ g = f \circ h$,

to $g = h$. Podaj przykłady funkcji f, g i h spełniających warunek $f \circ g = f \circ h$ i takich, że $g \neq h$.
(b) Podaj przykłady funkcji f, g i h takich, że $g \circ f = h \circ f$ oraz $g \neq h$.
(c) Sformułuj warunek, jaki musi spełniać funkcja f, by z równości $g \circ f = h \circ f$ wynikała równość $g = h$.

16. Załóżmy, że funkcja $f\colon S \to T$ jest odwracalna i niech $B \subseteq T$.
(a) Pokaż, że jeśli $y \in B$, to $f^{-1}(y) \in f^{\leftarrow}(B)$.
(b) Pokaż, że jeśli $x \in f^{\leftarrow}(B)$, to $x \in f^{-1}(B)$.
(c) Pokaż, że $f^{\leftarrow}(B) = f^{-1}(B)$.

§ 1.5. Ciągi

Ten paragraf dotyczy ciągów obiektów. W tym kontekście właściwe jest użycie indeksów, podobnie jak zawsze wtedy, gdy mamy do czynienia z wieloma obiektami; „wiele" znaczy często „więcej niż 3 lub 4". Na przykład, wybór liter x, y i z jest właściwy, gdy mamy do czynienia z równaniami o trzech (lub mniejszej liczbie) niewiadomych. Jednakże, jeśli mamy dziesięć niewiadomych lub chcemy omawiać sytuację ogólną dla n niewiadomych, gdzie n jest niesprecyzowaną liczbą ze zbioru $\mathbb{P}$, to właściwszy byłby wybór $x_1, x_2, \ldots, x_n$ jako nazw dla tych niewiadomych. Niewiadome rozróżniamy tu za pomocą małych numerków $1, 2, \ldots, n$ napisanych u dołu litery x, nazywanych **indeksami**. Innym przykładem jest ogólna postać wielomianu niezerowego

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0,$$

gdzie $a_n \neq 0$. Liczba n jest tu stopniem wielomianu, a $n+1$ możliwych współczynników oznaczyliśmy używając indeksów $a_0, a_1, \ldots, a_n$. Na przykład wielomian $x^3 + 4x^2 - 73$ jest zapisany według tego schematu, jeśli $n = 3$, $a_3 = 1$, $a_2 = 4$, $a_1 = 0$ i $a_0 = -73$.

Litery Σ używaliśmy jako nazwy alfabetu. W matematyce wielka grecka litera $\sum$ standardowo oznacza znak sumowania. Składniki napisane po niej mają być dodane do siebie, zgodnie z zakresem zapisanym u dołu i u góry litery $\sum$. Na przykład, weźmy wyrażenie

$$\sum_{k=1}^{10} k^2.$$

Napisy $k = 1$ i 10 mówią nam, że należy dodać do siebie liczby k^2 otrzymane przez podstawianie kolejno $k = 1$, następnie $k = 2$,

następnie $k = 3$ itd. aż do $k = 10$. To znaczy

$$\sum_{k=1}^{10} k^2 = 1 + 4 + 9 + 16 + 25 + 36 + 49 + 64 + 81 + 100 = 385.$$

Litera k jest zmienną (zmienia się ona od 1 do 10), którą można zastąpić dowolną inną zmienną. Zatem

$$\sum_{k=1}^{10} k^2 = \sum_{j=1}^{10} j^2 = \sum_{r=1}^{10} r^2.$$

Możemy również rozważać bardziej ogólne sumy, takie jak

$$\sum_{k=1}^{n} k^2,$$

w których górna granica sumowania może przyjmować różne wartości. Każda wartość n daje pewną szczególną wartość sumy; dla każdej wybranej liczby n zmienna k przebiega wartości od 1 do n. Oto kilka przypadków sumy $\sum_{k=1}^{n} k^2$ dla różnych n:

Wartość n	**Suma**
$n = 1$	$1^2 = 1$
$n = 2$	$1^2 + 2^2 = 1 + 4 = 5$
$n = 3$	$1^2 + 2^2 + 3^2 = 14$
$n = 4$	$1^2 + 2^2 + 3^2 + 4^2 = 30$
$n = 10$	$1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2 = 385$
$n = 73$	$1^2 + 2^2 + 3^2 + 4^2 + \ldots + 73^2 = 132349$

Możemy rozpatrywać nawet jeszcze bardziej ogólne sumy, takie jak

$$\sum_{k=1}^{n} x_k \quad \text{oraz} \quad \sum_{j=m}^{n} a_j.$$

Należy rozumieć, że $\{x_k\colon\ 1 \leq k \leq n\}$ i $\{a_j\colon\ m \leq j \leq n\}$ oznaczają tu zbiory liczb. Milcząco zakładamy, że $m \leq n$, ponieważ w przeciwnym przypadku nie byłoby nic do sumowania.

Podobnie, jak grecka litera $\sum$, wielka grecka litera $\prod$ jest znakiem mnożenia. Dla $n \in \mathbb{P}$ iloczyn pierwszych n liczb całkowitych nazywamy n **silnia** i oznaczamy przez $n!$. Zatem

$$n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n = \prod_{k=1}^{n} k.$$

Wyrażenie $1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$ jest nieco mylące dla małych wartości n, takich jak 1 czy 2; tak naprawdę oznacza ono „pomnóż przez

siebie kolejne liczby naturalne aż do n". Wyrażenie $\prod_{k=1}^{n} k$ jest bardziej jednoznaczne. Oto kilka pierwszych wartości $n!$: $1! = 1$, $2! = 1 \cdot 2 = 2$, $3! = 1 \cdot 2 \cdot 3 = 6$, $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$. Więcej wartości $n!$ podajemy w tablicach 1.4 i 1.5. Z pewnych powodów $n!$ jest również zdefiniowana dla $n = 0$; $0!$ z definicji jest równe 1. Do definicji $n!$ powrócimy w § 4.1.

Nieskończony ciąg obiektów może być zapisany za pomocą indeksów ze zbioru $\mathbb{N} = \{0, 1, 2, \ldots\}$ liczb naturalnych (lub $\{m, m+1, m+2, \ldots\}$ dla pewnej liczby całkowitej m). Zatem **ciąg** określony na zbiorze $\mathbb{N}$ jest listą $s_0, s_1, \ldots, s_n, \ldots$, w której dla każdej liczby $n \in \mathbb{N}$ obiekt s_n ma określoną wartość. Często s_n nazywamy **n-tym wyrazem ciągu**. Czasami wygodniej jest oznaczać sam ciąg symbolem (s_n) lub $(s_n)_{n\in\mathbb{N}}$, lub $(s_0, s_1, s_2, \ldots)$. Niekiedy będziemy pisać $s(n)$ zamiast s_n. Informatycy używają często oznaczenia $s[n]$, częściowo dlatego, że łatwo się to zapisuje na komputerze.

Oznaczenie $s(n)$ wygląda tak jak nasze oznaczenie funkcji. W rzeczywistości ciąg jest funkcją, której dziedziną jest zbiór $\mathbb{N} = \{0, 1, 2, \ldots\}$ liczb naturalnych lub zbiór $\{m, m+1, m+2, \ldots\}$ dla pewnej liczby całkowitej m. Każda liczba całkowita n z dziedziny ciągu wyznacza wartość $s(n)$ n-tego wyrazu ciągu.

W pierwszych przykładach będziemy rozważać ciągi liczb rzeczywistych.

PRZYKŁAD 1 (a) Ciąg $(s_n)_{n\in\mathbb{N}}$, gdzie $s_n = n!$, jest ciągiem silnii $(1, 1, 2, 6, 24, \ldots)$. Zbiorem wartości tego ciągu jest $\{1, 2, 6, 24, \ldots\} = \{n!\colon\ n \in \mathbb{P}\}$.

(b) Ciąg $(a_n)_{n\in\mathbb{N}}$ zdefiniowany za pomocą wzoru $a_n = (-1)^n$ dla $n \in \mathbb{N}$ jest ciągiem $(1, -1, 1, -1, 1, -1, 1, \ldots)$, którego zbiorem wartości jest $\{-1, 1\}$. ∎

Jak pokazują to dwa ostatnie przykłady, ważne jest, by rozróżniać ciąg i zbiór jego wartości. Zawsze używamy nawiasów { }, kiedy wypisujemy elementy zbioru lub opisujemy zbiór i nigdy nie używamy tych nawiasów do opisu ciągu. Ciąg $(a_n)_{n\in\mathbb{N}}$ dany wzorem $a_n = (-1)^n$ z przykładu 1(b) ma nieskończenie wiele wyrazów, chociaż jego wartości ciągle się powtarzają. Z drugiej strony, zbiór wartości $\{(-1)^n\colon\ n \in \mathbb{N}\}$ jest dokładnie zbiorem $\{-1, 1\}$, składającym się z dwóch liczb.

Ciągom nadaje się często przemawiające do wybraźni skrótowe nazwy, takie jak SILNIA, SUMA itp.

PRZYKŁAD 2 (a) Niech $\text{SILNIA}(n) = n!$ dla $n \in \mathbb{N}$. Jest to dokładnie ten sam ciąg, co w przykładzie 1(a); zmieniła się tylko jego nazwa (SILNIA

zamiast s). Zauważmy, że $\text{SILNIA}(n+1) = (n+1) \cdot \text{SILNIA}(n)$ dla $n \in \mathbb{N}$.

(b) Niech $\text{DWA}(n) = 2^n$ dla $n \in \mathbb{N}$. Wtedy DWA jest ciągiem. Zauważmy, że $\text{DWA}(n+1) = 2 \cdot \text{DWA}(n)$ dla $n \in \mathbb{N}$. ∎

Nasza definicja ciągu dopuszcza, by dziedziną był dowolny zbiór postaci $\{m, m+1, m+2, \ldots\}$, gdzie m jest liczbą całkowitą.

PRZYKŁAD 3

(a) Dziedzina ciągu (b_n) danego wzorem $b_n = 1/n^2$ dla $n \geq 1$ musi być zbiorem liczb różnych od zera. Kilka pierwszych wyrazów tego ciągu to: $1, \frac{1}{4}, \frac{1}{9}, \frac{1}{16}, \frac{1}{25}$.

(b) Weźmy ciąg, którego n-tym wyrazem jest $\log_2 n$. Zauważmy, że $\log_2 0$ nie ma sensu, tak więc ten ciąg musi zaczynać się od $n = 1$. Mamy $\log_2 1 = 0$, ponieważ $2^0 = 1$, $\log_2 2 = 1$, ponieważ $2^1 = 2$, $\log_2 4 = 2$, ponieważ $2^2 = 4$, $\log_2 8 = 3$, ponieważ $2^3 = 8$ itd. Pośrednie wartości $\log_2 n$ mogą być obliczone tylko w przybliżeniu (zob. tabl. 1.4). Na przykład $\log_2 5 \approx 2,3219$ jest tylko przybliżoną wartością, ponieważ $2^{2,3219} \approx 4,9999026$. ∎

Tablica 1.4

$\log_2 n$	$\sqrt{n}$	n	n^2	2^n	$n!$	n^n
0	1,0000	**1**	1	2	1	1
1,0000	1,4142	**2**	4	4	2	4
1,5850	1,7321	**3**	9	8	6	27
2,0000	2,0000	**4**	16	16	24	256
2,3219	2,2361	**5**	25	32	120	3125
2,5850	2,4495	**6**	36	64	720	46 656
2,8074	2,6458	**7**	49	128	5040	823 543
3,0000	2,8284	**8**	64	256	40 320	$1,67 \cdot 16^7$
3,1699	3,0000	**9**	81	512	362 880	$3,87 \cdot 10^8$
3,3219	3,1623	**10**	100	1 024	3 628 800	10^{10}

PRZYKŁAD 4

(a) Będziemy się zajmować porównywaniem szybkości wzrostu znanych ciągów, takich jak $\log_2 n$, $\sqrt{n}$, n^2, 2^n, $n!$ oraz n^n. Nawet dla względnie małych wartości n wydaje się oczywiste na podstawie tabl. 1.4, że ciąg n^n rośnie dużo szybciej niż ciąg $n!$, który z kolei rośnie dużo szybciej niż ciąg 2^n itd., chociaż wydaje się, że ciągi $\log_2 n$ i $\sqrt{n}$ rosną mniej więcej jednakowo szybko. W paragrafie 1.6 omówimy to zagadnienie dokładniej i podamy argumenty, które nie zależą od obserwacji dokonanych na podstawie kilku obliczeń.

(b) Naprawdę interesuje nas porównanie ciągów z przykładu (a) dla dużych wartości n (zob. tablicę 1.5[1]).

Tablica 1.5

$\log_2 n$	$\sqrt{n}$	n	n^2	2^n	$n!$	n^n
3,32	3,16	**10**	100	1,024	$3,63 \cdot 10^6$	10^{10}
6,64	10	**100**	10 000	$1,27 \cdot 10^{30}$	$9,33 \cdot 10^{157}$	10^{200}
9,97	31,62	**1 000**	10^6	$1,07 \cdot 10^{301}$	$4,02 \cdot 10^{2567}$	10^{3000}
13,29	100	**10 000**	10^8	$2,00 \cdot 10^{3010}$	$2,85 \cdot 10^{35659}$	10^{40000}
16,61	316,2	**100 000**	10^{10}	$1,00 \cdot 10^{30103}$	$2,82 \cdot 10^{456573}$	10^{500000}
19,93	1000	$\mathbf{10^6}$	10^{12}	$9,90 \cdot 10^{301029}$	$8,26 \cdot 10^{5565708}$	$10^{6000000}$
39,86	10^6	$\mathbf{10^{12}}$	10^{24}	duża liczba	większa	jeszcze większa

Okazuje się teraz, że $\log_2 n$ rośnie (dużo) wolniej niż $\sqrt{n}$; przekonamy się o tym w następnym paragrafie. Ten wzrost jest wolniejszy, ponieważ 2^n rośnie (dużo) szybciej niż n^2, a $\log_2 x$ i $\sqrt{x}$ są, odpowiednio, funkcjami odwrotnymi do funkcji 2^x i x^2. ∎

Jak dotąd, wszystkie nasze ciągi miały wartości będące liczbami rzeczywistymi. Jednakże, nie ma takiego ograniczenia w definicji i tak naprawdę będziemy zajmować się ciągami o wartościach innych typów.

PRZYKŁAD 5 Następujące ciągi mają wartości, które są zbiorami.

(a) Ciąg $(D_n)_{n\in\mathbb{N}}$ podzbiorów zbioru $\mathbb{Z}$ jest określony wzorem

$$\begin{aligned} D_n &= \{m \in \mathbb{Z}\colon\ m \text{ jest wielokrotnością } n\} \\ &= \{0, \pm n, \pm 2n, \pm 3n, \ldots\}. \end{aligned}$$

(b) Niech Σ będzie alfabetem. Dla każdej liczby $k \in \mathbb{N}$ zbiór Σ^k jest określony jako zbiór wszystkich słów ze zbioru Σ^*, mających długość k. Zapisujemy to:

$$\Sigma^k = \{w \in \Sigma^*\colon\ \text{długość}(w) = k\}.$$

Ciąg $(\Sigma^k)_{k\in\mathbb{N}}$ jest ciągiem podzbiorów zbioru Σ^*, którego sumą $\bigcup_{k\in\mathbb{N}} \Sigma^k$ jest zbiór Σ^*. Zauważmy, że zbiory Σ^k są rozłączne oraz że $\Sigma^0 = \{\lambda\}$, a $\Sigma^1 = \Sigma$. W przypadku, gdy $\Sigma = \{a, b\}$, mamy $\Sigma^0 = \{\lambda\}$, $\Sigma^1 = \Sigma = \{a, b\}$, $\Sigma^2 = \{aa, ab, ba, bb\}$ itd.

Zbiory Σ^k pojawiły się już w przykładzie 7 w § 1.4 jako zbiory $L^{\leftarrow}(k)$, gdzie L była funkcją długości w zbiorze Σ^*. Odtąd będziemy używać oznaczenia Σ^k dla tych zbiorów. ∎

[1] Dziękujemy naszemu koledze Richardowi M. Kochowi za dostarczenie do tej tablicy danych dla większych wartości n. Używał on programu *Mathematica*.

W poprzednim przykładzie symbolem $\bigcup_{k\in\mathbb{N}} \Sigma^k$ oznaczyliśmy sumę nieskończonego ciągu zbiorów. Słowo jest elementem tego zbioru, jeśli należy do jednego ze zbiorów Σ^k. Zbiory Σ^k są rozłączne, ale ogólnie możemy rozważać sumy zbiorów przecinających się. Jeśli $(A_k)_{k\in\mathbb{N}}$ jest ciągiem zbiorów, definiujemy

$$\bigcup_{k\in\mathbb{N}} A_k = \{x\colon\ x \in A_k \text{ dla co najmniej jednego } k \text{ ze zbioru } \mathbb{N}\}.$$

Ta definicja oczywiście ma również sens, jeśli te zbiory są określone dla k ze zbioru $\mathbb{P}$ lub z jakiegoś innego zbioru. Podobnie definiujemy

$$\bigcap_{k\in\mathbb{N}} A_k = \{x\colon\ x \in A_k \text{ dla wszystkich } k \in \mathbb{N}\}.$$

Oznaczenie $\bigcup_{k=0}^{\infty} A_k$ ma podobną interpretację, z tym tylko, że ∞ odgrywa tu szczególną rolę. Napis $\bigcup_{k=0}^{\infty}$ oznacza, że k przyjmuje wartości $0, 1, 2, \ldots$, ale k nie przyjmuje wartości ∞. Zatem

$$\bigcup_{k=0}^{\infty} A_k = \{x\colon\ x \in A_k \text{ dla co najmniej jednej liczby całkowitej } k \geq 0\},$$

podczas gdy

$$\bigcap_{k=1}^{\infty} A_k = \{x\colon\ x \in A_k \text{ dla wszystkich liczb całkowitych } k \geq 1\}.$$

W przykładzie 5 mogliśmy równie dobrze napisać $\Sigma^* = \bigcup_{k=0}^{\infty} \Sigma^k$.

Niektóre ciągi nie są nieskończone. **Ciągiem skończonym** jest ciąg obiektów, które można wypisać używając jako indeksów elementów skończonego podzbioru zbioru $\mathbb{Z}$ postaci $\{m, m+1, \ldots, n\}$. Często m będzie równe 0 lub 1. Taki ciąg $(a_m, a_{m+1}, \ldots, a_n)$ jest funkcją o dziedzinie $\{m, m+1, \ldots, n\}$, tak jak nieskończony ciąg $(a_m, a_{m+1}, \ldots)$ ma dziedzinę $\{m, m+1, \ldots\}$.

PRZYKŁAD 6

(a) Na początku tego paragrafu wspomnieliśmy o sumach uogólnionych, takich jak $\sum_{j=m}^{n} a_j$. Wartości, które mają być dodane do siebie, są wyrazami skończonego ciągu $(a_m, a_{m+1}, \ldots, a_n)$.

(b) Cyfry w rozwinięciu dziesiętnym liczby całkowitej tworzą ciąg skończony. Ciągiem cyfr liczby 8832 jest $(8, 8, 3, 2)$, jeśli wypisujemy cyfry od najbardziej znaczącej lub $(2, 3, 8, 8)$, jeśli wypisujemy cyfry od najmniej znaczącej.

ĆWICZENIA DO § 1.5

1. Oblicz:
(a) $\frac{7!}{5!}$, (b) $\frac{10!}{6!4!}$, (c) $\frac{9!}{0!9!}$,
(d) $\frac{8!}{4!}$, (e) $\sum_{k=0}^{5} k!$, (f) $\prod_{j=3}^{6} j$.

2. Uprość ułamki:
(a) $\frac{n!}{(n-1)!}$, (b) $\frac{(n!)^2}{(n+1)!(n-1)!}$.

3. Oblicz:
(a) $\sum_{k=1}^{n} 3^k$ dla $n = 1, 2, 3, 4$; (b) $\sum_{k=3}^{n} k^3$ dla $n = 3, 4, 5$;
(c) $\sum_{j=n}^{2n} j$ dla $n = 1, 2, 5$.

4. Oblicz:
(a) $\sum_{i=1}^{10} (-1)^i$, (b) $\sum_{k=0}^{3} (k^2 + 1)$, (c) $\left(\sum_{k=0}^{3} k^2\right) + 1$,
(d) $\prod_{n=1}^{5} (2n + 1)$, (e) $\prod_{j=4}^{8} (j - 1)$.

5. (a) Oblicz $\prod_{r=1}^{n} (r - 3)$ dla $n = 1, 2, 3, 4, 73$.
(b) Oblicz $\prod_{k=1}^{m} \frac{k+1}{k}$ dla $m = 1, 2, 3$. Podaj wzór na ten iloczyn dla wszystkich $m \in \mathbb{P}$.

6. (a) Oblicz $\sum_{k=0}^{n} 2^k$ dla $n = 1, 2, 3, 4, 5$.
(b) Wykorzystaj odpowiedzi w ćwiczeniu (a), by odgadnąć wzór ogólny dla tej sumy.

7. Weźmy ciąg dany wzorem $a_n = \frac{n-1}{n+1}$ dla $n \in \mathbb{P}$.
(a) Wypisz sześć pierwszych wyrazów tego ciągu.
(b) Oblicz $a_{n+1} - a_n$ dla $n = 1, 2, 3$.
(c) Pokaż, że $a_{n+1} - a_n = \frac{2}{(n+1)(n+2)}$ dla $n \in \mathbb{P}$.

8. Weźmy ciąg dany wzorem $b_n = \frac{1}{2}(1 + (-1)^n)$ dla $n \in \mathbb{N}$.
(a) Wypisz siedem pierwszych wyrazów tego ciągu.
(b) Jaki jest zbiór wartości tego ciągu?

9. Niech $\text{CIĄG}(n) = n^2 - n$ dla $n \in \mathbb{N}$.
(a) Oblicz $\text{CIĄG}(n)$ dla $n \leq 6$.
(b) Pokaż, że $\text{CIĄG}(n + 1) = \text{CIĄG}(n) + 2n$ dla wszystkich $n \in \mathbb{N}$.
(c) Pokaż, że $\text{CIĄG}(n + 1) = \frac{n+1}{n-1} \cdot \text{CIĄG}(n)$ dla $n \geq 2$.

10. Niech $\text{SUMAKW}(n) = \sum_{i=1}^{n} i^2$ dla $n = 1, 2, 3 \ldots$.
(a) Oblicz $\text{SUMAKW}(n)$ dla $n = 1, 2, 3, 5$.
(b) Zauważ, że $\text{SUMAKW}(n + 1) = \text{SUMAKW}(n) + (n + 1)^2$ dla $n \geq 1$.
(c) Okazuje się, że $\text{SUMAKW}(73) = 132349$. Wykorzystaj tę równość do obliczenia $\text{SUMAKW}(74)$ i $\text{SUMAKW}(72)$.

11. Wypisz kilka początkowych wyrazów następujących ciągów; wypisuj je dotąd, aż stanie się jasna reguła, według której jest zbudowany ciąg.
(a) $a_n = (2n - 1 + (-1)^n)/4$ dla $n \in \mathbb{N}$.
(b) (b_n), gdzie $b_n = a_{n+1}$ dla $n \in \mathbb{N}$ i ciągu a_n z ćwiczenia (a).

(c) WEK$(n) = (a_n, b_n)$ dla $n \in \mathbb{N}$.

12. Znajdź wyrazy ciągów $\log_2 n$ oraz $\sqrt{n}$ dla $n = 16, 64, 256, 4096$ i porównaj je.

13. (a) Używając kalkulatora lub innego urządzenia, uzupełnij tablicę 1.6. (Napisz E, jeśli wynik wykracza poza możliwości kalkulatora.)
(b) Omów zauważone względne szybkości wzrostu ciągów n^4, 4^n, n^{20}, 20^n i $n!$.

Tablica 1.6

n	n^4	4^n	n^{20}	20^n	$n!$
5			$9,54 \cdot 10^{13}$	$3,2 \cdot 10^6$	
10				$1,02 \cdot 10^{13}$	$3,63 \cdot 10^6$
25	$3,91 \cdot 10^5$				
50		$1,27 \cdot 10^{30}$			

14. Powtórz ćwiczenie 13 dla tablicy 1.7.

Tablica 1.7

n	$\log_{10} n$	$\sqrt{n}$	$20 \cdot \sqrt[4]{n}$	$\sqrt[4]{n} \cdot \log_{10} n$
50	1,70	7,07	53,18	4,52
100				
10^4				
10^6				

15. Zwróć uwagę na wykładniki potęg występujące w kolumnie 2^n w tablicy 1.5. Zauważ też, że $\log_{10} 2 \approx 0,30103$. Jak to wytłumaczyć?

§ 1.6. Notacja O

W informatyce ciągi pojawiają się w naturalny sposób jako listy kolejno wyliczanych wartości; ciągi SILNIA i DWA z przykładu 2 w § 1.5 są ciągami tego rodzaju. Inne ważne zastosowanie, szczególnie gdy analizujemy algorytmy, znajdują ciągi w zagadnieniu szacowania czasu wykonywania obliczeń dla określonych danych wejściowych.

Zastanówmy się, na przykład, nad posortowaniem w kolejności rosnącej listy n danych liczb całkowitych. Jest wiele algorytmów, które wykonują to zadanie; możemy prawdopodobnie samodzielnie wymyślić kilka różnych sposobów. Niektóre algorytmy są szybsze niż inne i wszystkie z nich wymagają więcej czasu, gdy n staje się większe. Jeśli n jest małe, prawdopodobnie nie ma różnicy, którą metodę wybierzemy, ale dla dużych n właściwy wybór metody może prowadzić do istotnego zaoszczędzenia

czasu. Potrzebny jest nam sposób opisu czasu działania naszych algorytmów.

W naszym przykładzie dotyczącym sortowania, niech ciąg t mierzy czas działania danego algorytmu, a więc niech $t(n)$ oznacza czas sortowania listy długości n. Na szybszych komputerach możemy skrócić ten czas $t(n)$ dwa, 100 lub nawet 1000 razy. Ale wtedy wszystkie nasze algorytmy będą również działały szybciej. To, co jest naprawdę istotne przy wyborze metody, to nie jest absolutna wielkość $t(n)$, ale szybkość, z jaką $t(n)$ rośnie, gdy zwiększa się n. Czy ta wielkość rośnie tak szybko jak 2^n, n^2, n lub $\log_2 n$, czy tak jak jakaś inna funkcja zmiennej n, omawiana w § 1.5?

Głównym zadaniem tego paragrafu jest omówienie notacji służącej do opisu szybkości wzrostu. Zanim to zrobimy, przyjrzymy się bliżej zależnościom między znanymi nam ciągami, takimi jak $\log_2 n$, $\sqrt{n}$, n, n^2 i 2^n.

PRZYKŁAD 1

(a) Dla wszystkich liczb całkowitych dodatnich n mamy

$$\ldots \leq \sqrt[4]{n} \leq \sqrt[3]{n} \leq \sqrt{n} \leq n \leq n^2 \leq n^3 \leq n^4 \leq \ldots.$$

Oczywiście, do tego ciągu nierówności można wstawić inne potęgi n. Na przykład

$$n \leq n\sqrt{n} \leq n^2 \quad \text{dla wszystkich} \quad n;$$

przypomnijmy, że $n\sqrt{n} = n^{3/2}$.

(b) Dla wszystkich $n \in \mathbb{N}$ mamy $n \leq 2^n$. Tak naprawdę, dla wszystkich n mamy $n \leq 2^{n-1}$. Jest to oczywiste dla małych wartości n, takich jak 1, 2 czy 3. Ogólnie,

$$n = 2 \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \frac{5}{4} \cdot \ldots \cdot \frac{n-1}{n-2} \cdot \frac{n}{n-1}.$$

Po prawej stronie mamy $n-1$ czynników i każdy z nich jest co najwyżej równy 2, a więc $n \leq 2^{n-1}$.

(c) $n^2 \leq \frac{9}{8} \cdot 2^n$ dla $n \geq 1$. Łatwo to sprawdzić dla $n = 1, 2, 3, 4$. Zauważmy, że dla $n = 3$ mamy równość. Dla $n > 4$ otrzymujemy

$$n^2 = 4^2 \cdot \left(\frac{5}{4}\right)^2 \cdot \ldots \cdot \left(\frac{n-1}{n-2}\right)^2 \cdot \left(\frac{n}{n-1}\right)^2.$$

Każdy czynnik po prawej stronie, oprócz pierwszego, jest równy co najwyżej $\left(\frac{5}{4}\right)^2$ i jest $n-4$ takich czynników. Ponieważ $\left(\frac{5}{4}\right)^2 = 1,5625 < 2$, więc

$$n^2 < 4^2 \cdot 2^{n-4} = 2^4 \cdot 2^{n-4} = 2^n, \quad \text{jeśli} \quad n > 4. \qquad \blacksquare$$

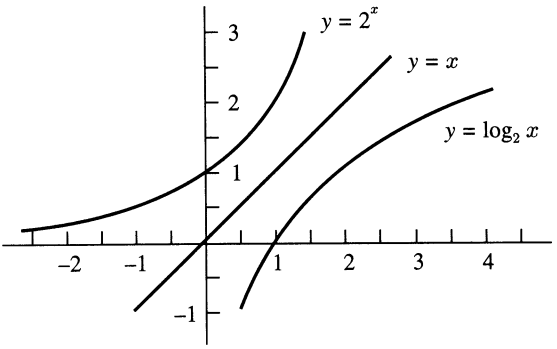
PRZYKŁAD 2 (a) Ponieważ $n \leq 2^{n-1}$ dla wszystkich n ze zbioru $\mathbb{N}$, więc mamy

$$\log_2 n \leq \log_2 2^{n-1} = n - 1 \quad \text{dla} \quad n \geq 1.$$

Możemy wykorzystać ten fakt do pokazania, że

$$\log_2 x < x \text{ dla wszystkich liczb rzeczywistych } x > 0;$$

zob. rys. 1.15. Rzeczywiście, jeśli n jest najmniejszą liczbą całkowitą większą od x, to mamy $n - 1 \leq x < n$, a więc $\log_2 x < \log_2 n \leq n - 1 \leq x$.



Rysunek 1.15

(b) W przykładzie 1(c) pokazaliśmy, że $n^2 < 2^n$ dla $n > 4$. W rzeczywistości, dla danej liczby dodatniej m mamy

$$n^m < 2^n \quad \text{dla odpowiednio dużych} \quad n.$$

Aby to pokazać, po pierwsze zauważmy, że $\frac{1}{2}\log_2 n = \log_2 n^{1/2} = \log_2 \sqrt{n} < \sqrt{n}$ na podstawie (a), a więc

$$\log_2 n^m = m \cdot \log_2 n < 2m \cdot \sqrt{n} = \frac{2m}{\sqrt{n}} \cdot n.$$

Następnie, $2m/\sqrt{n} \leq 1$ dla $n \geq 4m^2$, a więc

$$\log_2 n^m < n \quad \text{dla} \quad n \geq 4m^2.$$

Stąd otrzymujemy

$$n^m < 2^m \quad \text{dla} \quad n \geq 4m^2.$$

(c) Niech m będzie liczbą całkowitą dodatnią. Na podstawie przykładu (b) mamy

$$\log_2 n^m < n \quad \text{dla} \quad n \geq 4m^2.$$

Nierówność ta zachodzi nawet wtedy, gdy n nie jest liczbą całkowitą, a więc możemy zastąpić liczbę n liczbą $\sqrt[m]{n}$, aby otrzymać $\log_2(\sqrt[m]{n})^m < \sqrt[m]{n}$ dla $\sqrt[m]{n} \geq 4m^2$, tzn.

$$\log_2 n < \sqrt[m]{n} \quad \text{dla} \quad n \geq (4m^2)^m.$$

Zatem $\log_2 n < \sqrt[m]{n}$ dla dostatecznie dużych n. ∎

Z grubsza mówiąc, przykład 2(b) pokazuje, że 2^n rośnie szybciej niż jakakolwiek potęga n, a przykład 2(c) mówi nam, że $\log_2 n$ rośnie dużo wolniej niż jakikolwiek pierwiastek z n. Zanim wyrazimy te spostrzeżenia w sposób bardziej precyzyjny, popatrzmy na kilka następnych nierówności.

PRZYKŁAD 3 (a) Mamy

$$2^n < n! < n^n \quad \text{dla} \quad n \geq 4.$$

Dla $n = 4$ nierówności te są oczywiste: $16 < 24 < 256$. Dla $n > 4$ mamy $n! = (4!) \cdot 5 \cdot 6 \cdot \ldots \cdot (n-1) \cdot n$. Pierwszy czynnik $4!$ jest większy od 2^4, a każdy z pozostałych $n-4$ czynników jest większy od 2. Zatem $n! > 2^4 \cdot 2^{n-4} = 2^n$.

Nierówność $n! < n^n$ jest oczywista, ponieważ $n!$ jest iloczynem liczb naturalnych, z których wszystkie poza jedną są mniejsze od n.

(b) Nie ograniczajmy się do liczby 2 i pokażmy, że $40^n < n!$ dla dostatecznie dużych n. Dowód będzie bardziej pomysłowy niż dowód nierówności $2^n < n!$. Zauważmy, że dla $n > 80$ możemy napisać

$$\begin{aligned} n! &> n(n-1) \cdot \ldots \cdot 81 \quad (n - 80 \text{ czynników}) \\ &> 80 \cdot 80 \cdot \ldots \cdot 80 \quad (n - 80 \text{ czynników}) \\ &= 80^{n-80} = 40^n \cdot \left(2^n \cdot \tfrac{1}{80^{80}}\right) > 40^n \end{aligned}$$

przy założeniu, że $2^n > 80^{80}$ lub $n > \log_2(80^{80}) = 80 \cdot \log_2 80 \approx 505,8$. Było to „grube" szacowanie, w tym sensie, że pominęliśmy duży czynnik (tak naprawdę $80!$), pisząc $n! > n(n-1) \cdot \ldots \cdot 81$. Gdybyśmy potrzebowali bliższej informacji o tym, gdzie $n!$ staje się większe od 40^n, moglibyśmy dokonać staranniejszego oszacowania. To, co zrobiliśmy, było jednak wystarczające do pokazania, że $40^n < n!$ dla dostatecznie dużych n. ∎

Aby wyrazić dokładniej, co mamy na myśli mówiąc „... rośnie jak ... dla dużych n", musimy wprowadzić nowe oznaczenie, nazywane **notacją** O. Głównym zastosowaniem tej notacji będzie opisanie czasu działania algorytmów.

Mając dane dwa ciągi, na przykład s i t, o wartościach rzeczywistych nieujemnych, stwierdzenie „$s(n) = O(t(n))$" (czytamy „$s(n)$ jest O od $t(n)$") ma oznaczać, że dla dużych n wartości s są nie większe niż wartości t, pomnożone przez pewną stałą.

PRZYKŁAD 4 (a) Przykład 1(a) pokazuje, że $\sqrt{n} = O(n)$, $n = O(n^2)$ itd. Przykład 1(b) pokazuje, że $n = O(2^n)$, a przykład 1(c) pokazuje, że $n^2 = O(2^n)$. W rzeczywistości, na podstawie przykładu 2(b), $n^m = O(2^n)$ dla każdego m. Przykład 3(a) pokazuje, że $2^n = O(n!)$ oraz że $n! = O(n^n)$.

(b) Najbardziej znaczącym składnikiem wyrażenia, które ma postać: $6n^4 + 20n^2 + 2000$ jest $6n^4$, ponieważ dla dużych n wartość n^4 jest dużo większa niż n^2 czy stała 2000. Zapiszemy tę obserwację w następujący sposób

$$6n^4 + 20n^2 + 2000 = O(n^4),$$

aby zaznaczyć, że wyrażenie po lewej stronie „ rośnie nie szybciej niż pewna wielokrotność n^4". Tak naprawdę, rośnie odrobinę szybciej niż $6n^4$. To, co jednak ma znaczenie, to fakt, że dla wystarczająco dużych n (w tym przypadku $n \geq 8$ jest wystarczająco duże) $20n^2 + 2000 < n^4$, zatem $6n^4 + 20n^2 + 2000 < 7n^4$, tzn. wartość całego wyrażenia jest nie większa niż pewna ustalona wielokrotność n^4. Możemy też oczywiście powiedzieć, że to wyrażenie rośnie nie szybciej niż pewna wielokrotność n^5, ale nie jest to już tak przydatna informacja jak ta, którą otrzymaliśmy. ∎

A oto dokładna definicja. Niech f i g będą ciągami liczb rzeczywistych. Piszemy

$$f(n) = O(g(n))$$

wtedy, gdy istnieje stała dodatnia C taka, że

$$|f(n)| \leq C \cdot |g(n)| \text{ dla dostatecznie dużych wartości } n.$$

Dopuszczamy, by ta nierówność nie była spełniona dla pewnej liczby małych wartości n, między innymi dlatego, że ciągi $f(n)$ lub $g(n)$ mogą nie być zdefiniowane dla tych n. Zależy nam na dużych wartościach n. W praktyce $f(n)$ oznacza ciąg, którym w danym momencie się zajmujemy (na przykład górne ograniczenie czasu działania jakiegoś algorytmu), podczas gdy $g(n)$ będzie pewnym prostym ciągiem, takim jak n, $\log_2 n$, n^3 itd., którego szybkość wzrostu jest nam znana. Następujące twierdzenie przedstawia własności, które poznaliśmy w przykładach 1-3.

Twierdzenie 1

Oto hierarchia pewnych znanych ciągów uporządkowanych w ten sposób, że każdy z nich jest O od wszystkich ciągów na prawo od niego:

$$1, \log_2 n, \ldots, \sqrt[4]{n}, \sqrt[3]{n}, \sqrt{n}, n, n \cdot \log_2 n, n\sqrt{n}, n^2, n^3, n^4, \ldots, 2^n, n!, n^n.$$

Ciąg stały 1 w powyższym twierdzeniu jest określony wzorem $1(n) = 1$ dla wszystkich n. On w ogóle nie rośnie.

PRZYKŁAD 5

(a) Przypuśćmy, że $g(n) = n$ dla wszystkich $n \in \mathbb{N}$. Stwierdzenie $f(n) = O(n)$ oznacza, że $|f(n)|$ jest ograniczony z góry przez pewną stałą wielokrotność n, tzn. że istnieje taka liczba $C > 0$, że $|f(n)| \leq Cn$ dla odpowiednio dużych $n \in \mathbb{N}$.

(b) Przypuśćmy, że $g(n) = 1$ dla wszystkich $n \in \mathbb{N}$. Mówimy, że $f(n)$ jest $O(1)$, jeśli istnieje taka stała C, że $|f(n)| \leq C$ dla wszystkich dużych n, tzn. że wtedy wartości $|f(n)|$ są ograniczone z góry przez pewną stałą.

(c) Ciąg s określony wzorem $s_n = 3n^2 + 15n$ spełnia warunek $s_n = O(n^2)$, ponieważ $n \leq n^2$ dla $n \geq 1$, a zatem mamy $|s_n| \leq 3n^2 + 15n^2 = 18n^2$ dla wszystkich dostatecznie dużych n.

(d) Ciąg t dany wzorem $t_n = 3n^2 + (-1)^n \cdot 15n$ również spełnia warunek $t_n = O(n^2)$. A więc, tak jak w przykładzie (c), mamy $|t_n| \leq 3n^2 + 15n^2 \leq 18n^2$ dla $n \geq 1$.

(e) Możemy uogólnić przykłady (c) oraz (d). Jeśli $f(n) = a_m n^m + a_{m-1} n^{m-1} + \ldots + a_0$, gdzie $a_m \neq 0$, jest wielomianem stopnia m zmiennej n, to $|a_k n^k| \leq |a_k| \cdot n^m$ dla $k = 0, 1, \ldots, m-1$, więc

$$\begin{aligned} |f(n)| &\leq |a_m n^m| + |a_{m-1} n^{m-1}| + \ldots + |a_0| \\ &\leq (|a_m| + |a_{m-1}| + \ldots + |a_0|) \cdot n^m, \end{aligned}$$

a zatem $f(n) = O(n^m)$. Pierwsza nierówność zachodzi dlatego, że

$$|x_1 + x_2 + \ldots + x_i| \leq |x_1| + |x_2| + \ldots + |x_i|$$

dla dowolnego skończonego ciągu $x_1, x_2, \ldots, x_i$ ze zbioru $\mathbb{R}$. ∎

Ze względu na przykład 5(e), mówimy, że ciąg $f(n)$ **rośnie wielomianowo**, jeśli $f(n) = O(n^m)$ dla jakiejś dodatniej liczby całkowitej m. Z teoretycznego punktu widzenia, algorytmy, których czas wykonywania rośnie wielomianowo, są uważane za praktyczne. Oczywiście, że są one praktyczne, jeśli się porówna je na przykład z algorytmami, których czas działania jest rzędu $O(2^n)$.

W praktyce najbardziej pożądany jest efektywny czas działania rzędu $O(n)$ lub $O(n \cdot \log_2 n)$. W następnym przykładzie rozpatrujemy dwa ciągi, które często spotykamy przy szacowaniu czasu działania algorytmów.

PRZYKŁAD 6 (a) Niech $s_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}$ dla $n \geq 1$. Twierdzimy, że

$$s_n = O(\log_2 n).$$

Zauważmy, że

$$s_2 = 1 + \frac{1}{2} < 2,$$

$$s_4 = s_2 + \left(\frac{1}{3} + \frac{1}{4}\right) < 2 + \left(\frac{1}{2} + \frac{1}{2}\right) = 3,$$

$$s_8 = s_4 + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) < 3 + \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4}\right) = 4$$

itd. Ogólnie $s_{2^k} < k + 1$. (Rozumowanie indukcyjne z § 4.2 przekona nas o tym całkowicie). Weźmy jakąś liczbę całkowitą $n > 2$. Ograniczmy liczbę n kolejnymi potęgami dwójki, powiedzmy $2^k < n \leq 2^{k+1}$. Ponieważ $k < \log_2 n$, więc mamy

$$s_n \leq s_{2^{k+1}} < (k+1) + 1 < \log_2 n + 2.$$

Jeśli $n \geq 4$, to $\log_2 n \geq 2$, a więc

$$s_n < 2 \log_2 n \text{ dla } n \geq 4.$$

Stąd $s_n = O(\log_2 n)$.

(b) Jeśli $t_n = n + \frac{n}{2} + \frac{n}{3} + \ldots + \frac{n}{n}$ dla $n \geq 1$, to $t_n = n \cdot s_n$, a więc na podstawie przykładu (a) $t_n < 2n \cdot \log_2 n$ dla $n \geq 4$. Zatem mamy $t_n = O(n \cdot \log_2 n)$. ∎

Następne twierdzenie podaje kilka ogólnych faktów dotyczących notacji O. Pamiętajmy, że napis $f(n) = O(g(n))$ po prostu oznacza, że $f(n)$ jest pewnym ciągiem, który jest $O(g(n))$.

Twierdzenie 2

(a) Jeśli $f(n) = O(g(n))$ i c jest stałą, to $c \cdot f(n) = O(g(n))$.
(b) Jeśli $f(n) = O(g(n))$ i $h(n) = O(g(n))$, to $f(n) + h(n) = O(g(n))$.
(c) Jeśli $f(n) = O(a(n))$ i $g(n) = O(b(n))$, to $f(n) \cdot g(n) = O(a(n) \cdot b(n))$.
(d) Jeśli $a(n) = O(b(n))$ i $b(n) = O(c(n))$, to $a(n) = O(c(n))$.

Dowód. Punkty (a) i (c) zostawiamy jako ćwiczenie 13.

(b) Jeśli $f(n) = O(g(n))$ i $h(n) = O(g(n))$, to istnieją liczby dodatnie C i D takie, że

$$|f(n)| \leq C \cdot |g(n)| \quad \text{dla dostatecznie dużych} \quad n$$

oraz

$$|h(n)| \leq D \cdot |g(n)| \quad \text{dla dostatecznie dużych} \quad n.$$

Ponieważ $|x + y| \leq |x| + |y|$ dla $x, y \in \mathbb{R}$, więc wynika stąd, że

$$|f(n) + h(n)| \leq |f(n)| + |h(n)| \leq (C + D) \cdot |g(n)|$$

dla dostatecznie dużych n. Zatem mamy $f(n) + h(n) = O(g(n))$.

(d) Jeśli $a(n) = O(b(n))$ i $b(n) = O(c(n))$, to istnieją liczby dodatnie C i D takie, że

$$|a(n)| \leq C \cdot |b(n)| \text{ oraz } |b(n)| \leq D \cdot |c(n)|$$

dla dostatecznie dużych n. Zatem

$$|a(n)| \leq C \cdot |b(n)| \leq C \cdot D \cdot |c(n)| \quad \text{dla dostatecznie dużych} \quad n,$$

a więc $a(n) = O(c(n))$. ∎

Zasady ogólne podane w twierdzeniu 2 pozwalają skrócić rozumowania w przykładach 5 i 6. I tak na przykład, wiemy, że $n^k = O(n^m)$, jeśli $k \leq m$, a więc z twierdzenia 2 wynika

$$\begin{aligned} a_m n^m + a_{m-1} n^{m-1} + \ldots + a_1 n + a_0 & \\ &= O(n^m) + O(n^{m-1}) + \ldots + O(n) + O(1) \\ &\qquad \text{(na podstawie (a))} \\ &= O(n^m) + O(n^m) + \ldots + O(n^m) + O(n^m) \\ &\qquad \text{(na podstawie (d))} \\ &= O(n^m) \qquad \text{(na podstawie (b)).} \end{aligned}$$

Możemy skorzystać z własności (b), gdyż liczba składników, $m + 1$, nie zależy od n.

W przykładzie 6(a) tak naprawdę zrobiliśmy już wszystko, gdy pokazaliśmy, że $s_n < \log_2 n + 2$, ponieważ

$$\log_2 n + 2 = O(\log_2 n) + O(1) = O(\log_2 n).$$

Przykład 6(b) wynika natychmiast z twierdzenia 2(c), gdyż

$$t_n = n \cdot s_n = O(n \cdot \log_2 n).$$

Patrząc na przykłady 4(b) i 5 widzimy, że byłoby wygodnie opisywać ciągi, podając ich najbardziej znaczące składniki

i składniki niższego rzędu dla dużych n. Zatem moglibyśmy napisać

$$6n^4 + 20n^2 + 1000 = 6n^4 + O(n^2),$$

gdzie składnik zawierający O dotyczy wyrażenia $20n^2 + 1000$, o którym wiemy, że spełnia warunek $20n^2 + 1000 = O(n^2)$. Takie użycie notacji O w wyrażeniach typu $f(n) + O(g(n))$ różni się nieco od sposobu użycia jej w równościach $f(n) = O(g(n))$. Oba sposoby jednak są zgodne z interpretacją, że $O(g(n))$ oznacza „ciąg, którego wyrazy są ograniczone przez pewną wielokrotność $g(n)$ dla dostatecznie dużych n". Odtąd będziemy używać tej notacji w takim znaczeniu.

Podobnie jak nadaliśmy sens wyrażeniom $a(n) + O(b(n))$, możemy napisać $a(n) \cdot O(b(n))$, mając na myśli „$a(n) \cdot f(n)$, gdzie $f(n) = O(b(n))$". Interpretując to w taki właśnie sposób, możemy pisać „równości" takie jak w następnym twierdzeniu.

Twierdzenie 3

Dla dowolnych ciągów $a(n)$ i $b(n)$ mamy
(a) $O(a(n)) + O(b(n)) = O(\max\{|a(n)|, |b(n)|\})$.
(b) $O(a(n)) \cdot O(b(n)) = O(a(n) \cdot b(n))$.

Te zdania po prostu oznaczają: jeśli $f(n) = O(a(n))$ oraz $g(n) = O(b(n))$, to $f(n) + g(n) = O(\max\{|a(n)|, |b(n)|\})$ oraz $f(n) \cdot g(n) = O(a(n) \cdot b(n))$. Na przykład $O(n^3) + O(n^4) = O(n^4)$, ponieważ $\max\{n^3, n^4\} = n^4$, a $O(n^3) \cdot O(n^4) = O(n^7)$. W przeciwieństwie do prawdziwych równości, stwierdzenia te niewiele znaczą, jeśli czytamy je od prawej do lewej.

Dowód twierdzenia 3. (a) Niech $f(n) = O(a(n))$ oraz $g(n) = O(b(n))$. Wtedy istnieją liczby dodatnie C i D takie, że

$$|f(n)| \leq C \cdot |a(n)| \text{ oraz } |g(n)| \leq D \cdot |b(n)|$$

dla dostatecznie dużych n.

Zatem mamy

$$\begin{aligned} |f(n) + g(n)| &\leq |f(n)| + |g(n)| \leq C \cdot |a(n)| + D \cdot |b(n)| \\ &\leq C \cdot \max\{|a(n)|, |b(n)|\} + D \cdot \max\{|a(n)|, |b(n)|\} \\ &= (C + D) \cdot \max\{|a(n)|, |b(n)|\} \end{aligned}$$

dla dostatecznie dużych n. Tak więc $f(n) + g(n) = O(\max\{|a(n)|, |b(n)|\})$.

Punkt (b) wynika natychmiast z twierdzenia 2(c). ∎

PRZYKŁAD 7 (a) Ponieważ $n^2 + 13n = O(n^2)$ oraz $(n+1)^3 = O(n^3)$ (przemyśl to), więc na podstawie twierdzenia 3(a) mamy $(n^2 + 13n) + (n+1)^3 = O(n^3)$ oraz na podstawie twierdzenia 3(b) mamy $(n^2 + 13n)(n+1)^3 = O(n^5)$.

(b) Jeśli $a(n) = O(n^4)$ i $b(n) = O(\log_2 n)$, to $a(n) \cdot b(n) = O(n^4 \cdot \log_2 n)$, $a(n)^2 = O(n^8)$ oraz $b(n)^2 = O(\log_2^2 n)$. Zauważmy, że $\log_2^2 n$ oznacza $(\log_2 n)^2$.

(c) Wnioskiem z przykładu 5(e) jest też równość

$$\sum_{k=0}^{m} a_k n^k = a_m n^m + O(n^{m-1}), \quad \text{jeśli} \quad a_m \neq 0. \qquad \blacksquare$$

ĆWICZENIA DO § 1.6

1. Dla każdego z poniższych ciągów znajdź najmniejszą liczbę k taką, że $f(n) = O(n^k)$.
(a) $f(n) = 13n^2 + 4n - 73$ (b) $f(n) = (n^2+1)(2n^4 + 3n - 8)$
(c) $f(n) = (n^3 + 3n - 1)^4$ (d) $\sqrt{n+1}$

2. Powtórz ćwiczenie 1 dla ciągów:
(a) $f(n) = (n^2 - 1)^7$, (b) $f(n) = \sqrt{n^2 - 1}$,
(c) $f(n) = \sqrt{n^2 + n}$, (d) $f(n) = (n^2 + n + 1)^2 \cdot (n^3 + 5)$.

3. Dla każdego z poniższych ciągów podaj ciąg $a(n)$ z hierarchii w twierdzeniu 1 taki, że $f(n) = O(a(n))$ oraz $a(n)$ znajduje się możliwie najbardziej na lewo w tej hierarchii.
(a) $f(n) = 3^n$ (b) $f(n) = n^3 \cdot \log_2 n$
(c) $f(n) = \sqrt{\log_2 n}$

4. Powtórz ćwiczenie 3 dla ciągów:
(a) $f(n) = n + 3 \cdot \log_2 n$, (b) $f(n) = (n \cdot \log_2 n + 1)^2$,
(c) $f(n) = (n+1)!$.

5. Sprawdź, czy każda z poniższych równości jest prawdziwa czy fałszywa. W każdym z przypadków uzasadnij swoją odpowiedź.
(a) $2^{n+1} = O(2^n)$ (b) $(n+1)^2 = O(n^2)$
(c) $2^{2n} = O(2^n)$ (d) $(200n)^2 = O(n^2)$

6. Powtórz ćwiczenie 5 dla ciągów:
(a) $\log_2^{73} n = O(\sqrt{n})$, (b) $\log_2 n^{73} = O(\log_2 n)$,
(c) $\log_2 n^n = O(\log_2 n)$, (d) $(\sqrt{n}+1)^4 = O(n^2)$.

7. Czy poniższe zdania są prawdziwe czy fałszywe? Uzasadnij odpowiedź.
(a) $40^n = O(2^n)$ (b) $(40n)^2 = O(n^2)$
(c) $(2n)! = O(n!)$ (d) $(n+1)^{40} = O(n^{40})$

8. Niech A będzie pewną stałą dodatnią. Pokaż, że $A^n < n!$ dla dostatecznie dużych n. *Wskazówka*: przeanalizuj przykład 3(b).

9. (a) Dla $n \in \mathbb{P}$ niech $s_n = \sum_{k=1}^{n} \frac{1}{k^2}$, więc $s_n = 1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2}$ dla $n \geq 4$. Pokaż, że $s_n = O(1)$. *Wskazówka*: pokaż, że $\frac{1}{k^2} \leq \left(\frac{1}{k-1} - \frac{1}{k}\right)$ dla $k \geq 2$, a więc, że $s_n \leq 1 + \sum_{k=2}^{n} \left(\frac{1}{k-1} - \frac{1}{k}\right) = 2 - \frac{1}{n}$.

(b) Pokaż, że $t_n = O(n^2)$, jeśli $t_n = \sum_{k=1}^{n} k = 1 + 2 + \ldots + n$.

10. (a) Pokaż, że jeśli $s_n = \sum_{k=1}^{n} k^2$, to $s_n = O(n^3)$.

(b) Niech m będzie ustaloną liczbą ze zbioru $\mathbb{P}$. Definiujemy $t_n = \sum_{k=1}^{n} k^m$. Pokaż, że $t_n = O(n^{m+1})$.

11. Pokaż, że jeśli $f(n) = 3n^4 + O(n)$ i $g(n) = 2n^3 + O(n)$, to:
(a) $f(n) + g(n) = 3n^4 + O(n^3)$, (b) $f(n) \cdot g(n) = 6n^7 + O(n^5)$.

12. Pokaż, że
(a) $(5n^3 + O(n^2)) \cdot (3n^4 + O(n^3)) = 15n^7 + O(n^6)$,
(b) $(5n^3 + O(n)) \cdot (3n^4 + O(n^2)) = 15n^7 + O(n^5)$.

13. Wyjaśnij, dlaczego prawdziwe są własności (a) i (c) w twierdzeniu 2.

14. (a) Udowodnij bezpośrednio z twierdzenia 2, że jeśli $a(n) = O(c(n))$ i $b(n) = O(c(n))$, to $O(a(n)) + O(b(n)) = O(c(n))$.

(b) Zauważ, że to daje inny dowód twierdzenia 3(a).

15. To ćwiczenie pokazuje, że trzeba być ostrożnym, używając dzielenia w obliczeniach wykorzystujących notację O.

(a) Niech $a(n) = n^5$ i $b(n) = n$. Zauważ, że $a(n) = O(n^5)$, natomiast $b(n) = O(n^2)$, ale $a(n)/b(n)$ nie jest $O(n^3)$.

(b) Podaj przykłady ciągów $a(n)$ i $b(n)$ takich, że $a(n) = O(n^6)$ i $b(n) = O(n^2)$, ale $a(n)/b(n)$ nie jest $O(n^4)$.

16. Pokaż, że $\log_{10} n = O(\log_2 n)$.

17. Dla każdej liczby naturalnej $n \in \mathbb{P}$ niech $\text{CYFR}(n)$ oznacza liczbę cyfr w rozwinięciu dziesiętnym liczby n.

(a) Pokaż, że $10^{\text{CYFR}(n)-1} \leq n < 10^{\text{CYFR}(n)}$.

(b) Pokaż, że $\log_{10} n$ jest $O(\text{CYFR}(n))$.

(c) Pokaż, że $\text{CYFR}(n)$ jest $O(\log_{10} n)$.

(d) Niech $\text{CYFR2}(n)$ będzie liczbą cyfr w rozwinięciu dwójkowym liczby n. Jaki jest związek między $O(\text{CYFR}(n))$ i $O(\text{CYFR2}(n))$?

To, co jest najważniejsze w tym rozdziale

Aby sprawdzić, czy dobrze rozumiesz treść tego rozdziału:

1. Przekonaj się, że potrafisz zdefiniować każde pojęcie i oznaczenie oraz możesz opisać każdą metodę.

2. Podaj przynajmniej jeden powód, dla którego dany temat został omówiony w tym rozdziale.

3. Zastanów się nad co najmniej jednym przykładem ilustrującym dane pojęcie oraz co najmniej jedną sytuacją, w której dany fakt czy metoda są przydatne.

Ten rozdział jest rozdziałem wstępnym i zawiera stosunkowo dużo podstawowych definicji i oznaczeń. Ważne jest, aby dobrze przyswoić sobie jego treść teraz, ponieważ reszta tej książki na nim się opiera.

Pojęcia

zbiór (pojęcie niedefiniowane)
- element, podzbiór
- równe, rozłączne
- działania na zbiorach
- zbiór uniwersalny (przestrzeń), dopełnienie (uzupełnienie)
- diagramy Venna

para uporządkowana, iloczyn kartezjański zbiorów
alfabet, język, słowo, długość słowa
funkcja = przekształcenie = odwzorowanie
- dziedzina
- obraz elementu x, przeciwdziedzina funkcji $f = \mathrm{Im}(f)$
- wykres funkcji
- funkcja różnowartościowa, funkcja „na", przekształcenie wzajemnie jednoznaczne
- złożenie funkcji
- funkcja odwrotna
- obcięcie funkcji, $f(A)$
- przeciwobraz zbioru, $f^{\leftarrow}(B)$

ciąg, ciąg skończony
notacja O

Przykłady i oznaczenia

Zbiory $\mathbb{N}$, $\mathbb{P}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$
$\in$, $\notin$, $\{ : \}$, $\subseteq$, $\subset$
$\emptyset = \{\ \} =$ zbiór pusty
$[a,b]$, (a,b), $[a,b)$, $(a,b]$ – oznaczenia przedziałów
$\mathcal{P}(S)$, $\cup$, $\cap$, $A \backslash B$, $A \oplus B$, A^c
(s,t) – oznaczenie pary uporządkowanej, $S \times T, S^n$
$|S|$ = liczba elementów zbioru S
$\sum$ – oznaczenie sumy, $\prod$ – oznaczenie iloczynu
$n!$ – n silnia

Zbiór Σ^*, λ = słowo puste
Funkcje specjalne $1_S, \chi_A, \log_b$
$f(n) = O(g(n))$, $f(n) = g(n) + O(h(n))$

Fakty

Podstawowe prawa algebry zbiorów (tablica 1.1 w § 1.2).
Złożenie funkcji jest łączne.
Funkcja jest odwracalna wtedy i tylko wtedy, gdy jest przekształceniem wzajemnie jednoznacznym.
Porównanie szybkości wzrostu znanych ciągów (tablica 1.5 w § 1.5, twierdzenie 1 w § 1.6).

Metody

Stosowanie diagramów Venna.
Wnioskowanie z definicji i wcześniej udowodnionych faktów.

2. ELEMENTY LOGIKI

Rozdział ten stanowi nieformalne wprowadzenie do logiki, będącej zarówno zbiorem narzędzi do prowadzenia rozumowań, jak i samodzielnym obiektem badań. Każdy, kto w swojej pracy zajmuje się wnioskowaniem, musi umieć rozpoznawać rozumowania poprawne i niepoprawne. Matematycy kładą wielki nacisk na tworzenie takich dowodów, które są z logicznego punktu widzenia bez zarzutu. Informatycy oczywiście muszą umieć rozumować logicznie i do tego muszą znać formalne reguły logiki, według których działają ich maszyny. My położymy nacisk na logikę jako narzędzie pracy. Będziemy starali się robić to w sposób nieformalny, ale będziemy też zaznaczać, jak dany temat powinien być rozwijany bardziej formalnie. Omówimy też kilka metod logiki symbolicznej potrzebnych w informatyce. Związek między sprzętem komputerowym i układami logicznymi będzie dokładniej omówiony w rozdziale 10.

W paragrafie 2.1 wprowadzamy terminologię i stosowane oznaczenia, w tym również kwantyfikatory $\forall$ i $\exists$. W paragrafie 2.2 zajmujemy się wprowadzeniem do rachunku zdań. Pojęcia ogólne są istotne i powinny być dobrze przyswojone. Nie należy się przerażać znajdującymi się tam tablicami, których nie trzeba uczyć się na pamięć. Zostały one umieszczone po to, by można było do nich się odwołać. Ostatecznie celem dowodu jest podanie przekonujących argumentów. W paragrafie 2.3 omawiamy dowody spotykane „w rzeczywistości", formalizujemy te idee w § 2.4, a następnie powracamy do analizy nieformalnych rozumowań w § 2.5.

§ 2.1. Nieformalne wprowadzenie

Rachunek zdań polega na badaniu związków logicznych między zdaniami, które zazwyczaj można interpretować jako sensowne stwierdzenia odnoszące się do rzeczywistości. Dla nas **zdaniem** będzie dowolne stwierdzenie, które jest albo prawdziwe, albo fałszywe, i które nie może być jednocześnie prawdziwe i fałszywe. Zatem jest to stwierdzenie, któremu można przypisać wartość logiczną **prawdy** lub wartość logiczną **fałszu**, ale nie obie wartości jednocześnie. Nie musimy wiedzieć, jaka jest wartość logiczna zdania, aby się nim zajmować.

PRZYKŁAD 1 Następujące stwierdzenia są zdaniami:

(a) Juliusz Cezar był prezydentem Stanów Zjednoczonych.
(b) Związek Radziecki był krajem o największej powierzchni na świecie.
(c) $2 + 2 = 4$.
(d) $2 + 3 = 7$.
(e) Liczba 4 jest dodatnia, a liczba 3 jest ujemna.
(f) Jeśli zbiór ma n elementów, to ma on 2^n podzbiorów.
(g) $2^n + n$ jest liczbą pierwszą dla nieskończenie wielu n.
(h) Każda liczba całkowita parzysta większa od 4 jest sumą dwóch liczb pierwszych.

Zauważmy, że zdania (c) i (d) są zdaniami matematycznymi, w których „=" oznacza czasownik „równa się" lub „jest równe". Zdanie (e) jest fałszywe, ponieważ 3 nie jest liczbą ujemną. Jeśli nie jest to oczywiste teraz, to wkrótce stanie się jasne, zdanie (e) jest bowiem zdaniem złożonym: „4 jest liczbą dodatnią i 3 jest liczbą ujemną". Nie mamy pojęcia, czy zdanie (g) jest prawdziwe czy fałszywe, chociaż jacyś matematycy mogą znać odpowiedź. Z drugiej strony, w czasie pisania tej książki nikt nie wiedział, czy zdanie (h) jest prawdziwe; znane jest jako „hipoteza Goldbacha". ∎

PRZYKŁAD 2 Oto kilka następnych zdań:

(a) $x + y = y + x$ dla wszystkich liczb $x, y \in \mathbb{R}$.
(b) $2^n = n^2$ dla pewnych liczb $n \in \mathbb{N}$.
(c) Nie jest prawdą, że 3 jest liczbą parzystą lub 7 jest liczbą pierwszą.
(d) Jeśli świat jest płaski, to $2 + 2 = 4$.

Zdanie (a) jest w rzeczywistości nieskończonym zbiorem zdań utworzonym za pomocą sformułowania „dla wszystkich" lub „dla

każdych". Zdanie (b) jest szczególnym typem zdania, ze względu na sformułowanie „dla pewnych". Zdania tego typu omówimy dalej w tym paragrafie, a dokładniej zbadamy w rozdziale 13. Zdanie (c) jest nieco mylącym zdaniem złożonym, którego wartość logiczną będzie łatwiej przeanalizować po przeczytaniu tego rozdziału. Nasz rachunek zdań pozwoli budować zdania takie jak w przykładzie (d), nawet jeśli wydają się one być głupie czy paradoksalne. ∎

PRZYKŁAD 3 Następujące sformułowania nie są zdaniami:

(a) To twoje czy moje miejsce?
(b) Dlaczego indukcja jest ważna?
(c) Idź prosto do więzienia.
(d) Albo ten pies wyjdzie, albo ja.
(e) $x - y = y - x$.

Problem z przykładem (d) polega na tym, że tylko osoba wypowiadająca tę kwestię, może wiedzieć, czy to jest prawda czy nie. Gdyby ta wartość logiczna była znana, byłoby to zupełnie dobre zdanie. Powód, dla którego wyrażenie (e) nie jest zdaniem, jest taki, że symbole nie zostały dobrze określone. Jeśli zamierzeniem było

$$(e')\ x - y = y - x \text{ dla wszystkich } x, y \in \mathbb{R},$$

to jest to zdanie fałszywe. Jeśli zaś intencją było

$$(e'')\ x - y = y - x \text{ dla pewnych } x, y \in \mathbb{R}$$

lub

$$(e''')\ x - y = y - x \text{ dla wszystkich } x, y \text{ ze zbioru } \{0\},$$

to jest to zdanie prawdziwe. Problemem nieokreślonych symboli zajmiemy się pod koniec tego paragrafu. ∎

PRZYKŁAD 4 Oczywiście w życiu codziennym występują zdania niejednoznaczne:

(a) Nauczyciele są przepłacani.
(b) Lekarze są bogaci.
(c) Było zimno w Minneapolis w styczniu 1924 r.
(d) Matematyka jest zabawna.
(e) $A^2 = 0$ implikuje $A = 0$ dla wszystkich A.

Problem z wyrażeniem (e) polega na tym, że nie został określony zbiór dopuszczalnych A. Wyrażenie w (e) jest prawdziwe dla wszystkich $A \in \mathbb{R}$. Okazuje się, że (e) nie jest pozbawione sensu,

choć jest fałszywe w zbiorze wszystkich macierzy A wymiaru 2×2. Dwuznaczne zdania powinno się przekształcić w niedwuznaczne lub należy je pominąć. My nie będziemy się tym zajmować, ale założymy, że nasze zdania są jednoznaczne. (Zob. jednakże ćwiczenia 17 i 18 do § 2.1, gdzie znajdziesz następne przykłady zdań niejednoznacznych). ■

W rachunku zdań będziemy zazwyczaj używać małych liter p, q, r, ... do oznaczenia zdań i będziemy łączyć zdania, by otrzymać zdania złożone, używając standardowych symboli dla spójników:

$\neg$ jako „nie", czyli negacja;
$\wedge$ jako „i";
$\vee$ jako „lub" (alternatywa niewykluczająca);
$\rightarrow$ jako „implikuje", czyli zdanie warunkowe;
$\leftrightarrow$ jako „wtedy i tylko wtedy, gdy", czyli równoważność.

Inne spójniki, takie jak $\oplus$, pojawią się w ćwiczeniach do § 2.2 i § 2.4. Czasami będziemy budować zdanie złożone ze zdań prostszych, a czasami będziemy analizować skomplikowane zdanie na podstawie jego części składowych.

W następnym paragrafie bardzo starannie omówimy każdy ze spójników i wyjaśnimy, jak wpływają one na wartości logiczne zdań złożonych. Na razie jednak będziemy traktować te symbole nieformalnie, jako pewnego rodzaju skróty słów lub sformułowań w języku polskim i spróbujemy nabrać doświadczenia w używaniu ich do modyfikowania lub łączenia zdań. Następny przykład pokazuje, jak niektóre ze zdań z przykładów 1 i 2 można potraktować jako zdania złożone.

PRZYKŁAD 5 (a) Przypomnijmy zdanie (e) z przykładu 1: „liczba 4 jest dodatnia, a liczba 3 jest ujemna". To zdanie można potraktować jako zdanie złożone $p \wedge q$ (czytamy „p i q"), gdzie p = „liczba 4 jest dodatnia", a q = „liczba 3 jest ujemna".

(b) Zdanie (d) z przykładu 2, „jeśli świat jest płaski, to $2+2=4$", może być potraktowane jako zdanie złożone $r \rightarrow s$ (czytamy „r implikuje s"), gdzie r = „świat jest płaski", a s = „$2+2=4$".

(c) Zdanie (c) z przykładu 2 mówi: „nie jest prawdą, że 3 jest liczbą parzystą lub 7 jest liczbą pierwszą". Jest to zdanie $\neg(p \vee q)$, gdzie p = „3 jest liczbą parzystą", a q = „7 jest liczbą pierwszą". Tak naprawdę zdanie to jest źle napisane i może ono być również interpretowane jako $(\neg p) \vee q$. Kiedy przeczytamy głośno zdanie (c) jako „nie p lub q", musimy wyjaśnić, gdzie są nawiasy. Czy to jest negacja p, czy negacja obu zdań p i q? ■

Zdanie złożone $p \to q$ czytamy „p implikuje q", ale jest kilka innych równoważnych zwrotów w języku polskim, takich jak na przykład „jeśli p, to q". Tak naprawdę w przykładzie 5(b) zdanie złożone $r \to s$ miało postać „jeśli r, to s". To zdanie mogłoby też być napisane: „świat jest płaski implikuje $2 + 2 = 4$". Inne równoważne formy w języku polskim dla zdania $p \to q$ to: „q, jeśli p", „q, o ile p" itp. Na ogół będziemy ich unikać, zob. jednak ćwiczenia 15-18 do § 2.1.

Wydaje się, że istnieje związek między zdaniami złożonymi postaci $p \to q$, $q \to p$, $\neg p \to \neg q$ itp. i często te zdania są mylone. Ważne jest, by je rozróżniać. Zdanie $q \to p$ jest **zdaniem odwrotnym** do zdania $p \to q$. Jak się przekonamy, jego znaczenie jest inne niż zdania $p \to q$. Okazuje się, że zdanie $p \to q$ jest równoważne zdaniu $\neg q \to \neg p$, które nazywamy **zdaniem przeciwstawnym** lub **kontrapozycją** zdania $p \to q$.

PRZYKŁAD 6 Weźmy zdanie: „jeśli pada deszcz, to na niebie są chmury". Jest to zdanie złożone $p \to q$, gdzie $p =$ „pada deszcz" i $q =$ „na niebie są chmury". Jest to zdanie prawdziwe. Zdanie odwrotne $q \to p$ ma postać: „jeśli na niebie są chmury, to pada deszcz". Niestety jest to zdanie fałszywe. Kontrapozycja $\neg q \to \neg p$ ma postać: „jeśli na niebie nie ma chmur, to nie pada deszcz". Nie tylko jest to zdanie prawdziwe, ale większość ludzi zgodziłaby się, że wynika ono „logicznie" ze zdania $p \to q$, bez konieczności odwoływania się do fizycznego związku między chmurami i deszczem. Ono rzeczywiście wynika i ten logiczny związek będzie omówiony dokładniej w § 2.2 (tabl. 2.1, poz. 9). ∎

W logice zajmujemy się określaniem wartości logicznych zdań na podstawie wartości logicznych innych zdań. Przykład 6 pokazuje, że z prawdziwości zdania $p \to q$ nie wynika, że zdanie $q \to p$ jest prawdziwe, ale sugeruje, że prawdziwość kontrapozycji $\neg q \to \neg p$ wynika z prawdziwości zdania $p \to q$. A oto inny przykład pokazujący, dlaczego musimy być ostrożni, gdy zajmujemy się wyrażeniami logicznymi.

PRZYKŁAD 7 Popatrzmy na rozumowanie: „Jeśli Tomek nie pójdzie jutro do pracy, nie będzie miał pracy. On pójdzie jutro do pracy, a więc będzie ją miał." Nie interesuje nas to, czy Tomek rzeczywiście utraci pracę (jest to jego sprawa), ale to czy Tomek będzie miał pracę, wynika logicznie z poprzednich dwóch stwierdzeń: „Jeśli Tomek nie pójdzie jutro do pracy, nie będzie miał pracy" oraz „On pójdzie jutro do pracy". Okazuje się, że to rozumowanie nie jest poprawne. Pierwsze zdanie mówi nam tylko tyle, że Tomek

znajdzie się w kłopocie, jeśli nie pójdzie jutro do pracy; nie mówi nam nic, co się stanie w przeciwnym przypadku. Być może Tomek utraci pracę ze względu na brak kwalifikacji. Gdyby powyższe rozumowanie było poprawne, poprawne byłoby też: „Jeśli Karolina nie kupi losu na loterię, nie wygra 1000000$. Karolina kupi los na loterię, więc wygra 1000000$".

Oba te niepoprawne rozumowania można zapisać symbolicznie w postaci: jeśli zdania $\neg p \to \neg q$ i p są prawdziwe, to zdanie q jest prawdziwe. Rachunek zdań, który przedstawimy w § 2.2 i § 2.4, dostarczy formalnych podstaw do sprawdzania poprawności rozumowań takich jak powyższe. ∎

Prawdziwość zdania $p \to q$ wyrażamy czasami, mówiąc, że p jest **warunkiem wystarczającym** dla q lub, że q jest **warunkiem koniecznym** dla p. Powiedzenie, że p jest warunkiem koniecznym i wystarczającym dla q, jest innym sposobem stwierdzenia, że zdanie $p \leftrightarrow q$ jest prawdziwe.

PRZYKŁAD 8 (a) Aby zdać ten egzamin, trzeba ciężko pracować. To znaczy, że implikacja zdać $\to$ ciężko pracować jest prawdziwa. Ciężka praca jednakże nie wystarczy. Znamy przykłady pokazujące, że implikacja ciężko pracować $\to$ zdać nie zawsze jest prawdziwa.

(b) Aby zabić muchę, wystarczy trafić ją kulą armatnią, ale nie jest to konieczne. Zatem implikacja trafić kulą armatnią $\to$ zabić muchę jest prawdziwa, ale implikacja zabić muchę $\to$ trafić kulą armatnią nie jest.

(c) Warunkiem koniecznym i wystarczającym na to, by liczba pierwsza p była parzysta, jest $p = 2$. ∎

Zajmiemy się teraz zdaniami złożonymi zawierającymi sformułowania „dla każdego" lub „dla wszystkich" i odnoszącymi się do wielu zdań, być może nieskończenie wielu.

PRZYKŁAD 9 Weźmy hipotezę Goldbacha z przykładu 1: „każda liczba całkowita parzysta większa od 4 jest sumą dwóch liczb pierwszych". Zdanie to daje się rozłożyć w następujący sposób

$$„p(6) \wedge p(8) \wedge p(10) \wedge \ldots"$$

lub

$$„p(n) \text{ dla każdego } n \text{ ze zbioru } \mathbb{N}, \text{ większego od } 4",$$

gdzie $p(n)$ jest zdaniem prostym „n jest sumą dwóch liczb pierwszych". Jednakże zasady łączenia zdań za pomocą spójników w rachunku zdań nie dopuszczają takich konstrukcji, w których

występuje więcej niż skończona liczba zdań, czy zawierających sformułowania takie jak „dla każdego" lub „dla pewnego". ∎

W tak zwanym rachunku predykatów występują dwa wygodne sformułowania, które pozwalają na przedstawienie w sposób symboliczny takich zdań jak w przykładzie 9. Omówimy znaczenie logiczne tych sformułowań, nazywanych **kwantyfikatorami**, bardziej szczegółowo w rozdziale 13; na razie potraktujmy je jako nieformalne skróty.

Przypuśćmy, że $\{p(x)\colon x \in U\}$ jest rodziną zdań, gdzie U jest zbiorem, być może nieskończonym. Innymi słowy, p jest funkcją zdaniową określoną na zbiorze U. **Kwantyfikator ogólny** $\forall$ (odwrócona litera A) jest używany do tworzenia zdań złożonych postaci

$$\forall x\, p(x),$$

które czytamy „dla każdego x, $p(x)$". Można je czytać również „dla wszystkich", „dla dowolnego". Jednakże należy uważać na sformułowanie „dla dowolnego", ponieważ może ono być mylnie interpretowane jako „dla pewnego". Wyrażenie „statek przybija do dowolnego portu we Francji" nie oznacza, że „statek przybija do każdego portu we Francji". Zdanie złożone $\forall x\, p(x)$ ma przypisaną wartość logiczną w następujący sposób:

zdanie $\forall x\, p(x)$ jest prawdziwe, jeśli zdanie $p(x)$ jest prawdziwe dla każdego x ze zbioru U;

w przeciwnym przypadku zdanie $\forall x\, p(x)$ jest fałszywe.

Kwantyfikator szczegółowy (egzystencjalny) $\exists$ (odwrócona litera E) jest używany do formułowania zdań postaci

$$\exists x\, p(x),$$

które czytamy „istnieje x taki, że $p(x)$" lub „dla pewnego x, $p(x)$". Zdanie złożone $\exists x\, p(x)$ ma następujące wartości logiczne:

zdanie $\exists x\, p(x)$ jest prawdziwe, jeśli zdanie $p(x)$ jest prawdziwe dla co najmniej jednego x ze zbioru U;

zdanie $\exists x\, p(x)$ jest fałszywe, jeśli zdanie $p(x)$ jest fałszywe dla każdego x ze zbioru U.

PRZYKŁAD 10 (a) Dla każdego n ze zbioru $\mathbb{N}$, niech $p(n)$ oznacza zdanie „$n^2 = n$". Wtedy $\forall n\, p(n)$ jest zdaniem fałszywym, ponieważ na przykład zdanie $p(3)$, tzn. $3^2 = 3$ jest fałszywe. Z drugiej strony, zdanie $\exists n\, p(n)$ jest prawdziwe, ponieważ co najmniej jedno zdanie $p(n)$ jest prawdziwe; tak naprawdę, dokładnie dwa z tych zdań są prawdziwe, mianowicie $p(0)$ i $p(1)$.

Dla $n \in \mathbb{N}$ niech $q(n)$ będzie zdaniem „$(n+1)^2 = n^2+2n+1$". Możemy używać $p(n)$, $q(n)$ i spójników logicznych z rachunku zdań, aby otrzymać inne zdania z kwantyfikatorami. Na przykład $\exists n(\neg q(n))$ jest zdaniem fałszywym; każde zdanie $q(n)$ jest prawdziwe, a więc każde zdanie $\neg q(n)$ jest fałszywe. Zdanie $\forall n(p(n) \vee q(n))$ jest prawdziwe, ponieważ każde zdanie $p(n) \vee q(n)$ jest prawdziwe. (Pamiętajmy, że znak $\vee$ oznacza „lub"). Oczywiście zdanie słabsze $\exists n(p(n) \vee q(n))$ też jest prawdziwe.

(b) Niech $p(x)$ oznacza „$x \leq 2x$", a $q(x)$ oznacza „$x^2 \geq 0$" dla $x \in \mathbb{R}$. Ponieważ elementów zbioru $\mathbb{R}$ nie można wypisać w postaci ciągu, byłoby zupełnie niemożliwe zapisanie symbolicznie zdań $\exists x\, p(x)$ czy $\forall x\, q(x)$ w rachunku zdań. Jasne jest, że zdanie $\exists x\, p(x)$ jest prawdziwe; zdanie $\forall x\, p(x)$ jest fałszywe, ponieważ zdanie $p(x)$ jest fałszywe dla x ujemnych. Oba zdania $\forall x\, q(x)$ i $\exists x\, q(x)$ są prawdziwe, a zdanie $\exists x(\neg q(x))$ jest fałszywe.

(c) Kwantyfikatory są również przydatne wtedy, gdy mamy do czynienia ze skończonymi, ale dużymi zbiorami zdań. Przypuśćmy np., że mamy zdania $p(n)$ dla n ze zbioru $\{n \in \mathbb{N}\colon 0 \leq n \leq 65535\}$. Wolimy oczywiście zapis $\forall n\, p(n)$ zamiast

$$p(0) \wedge p(1) \wedge p(2) \wedge p(3) \wedge \ldots \wedge p(65535),$$

chociaż moglibyśmy wprowadzić teoretycznie dopuszczalny zapis

$$\bigwedge_{n=0}^{65535} p(n).$$

(d) Hipoteza Goldbacha może być teraz zapisana jako $\forall n\, p(n)$, gdzie $p(n)$ = „jeśli liczba n ze zbioru $\mathbb{N}$ jest parzysta i większa od 4, to n jest sumą dwóch liczb pierwszych". ∎

PRZYKŁAD 11 (a) Kwantyfikatory $\forall$ i $\exists$ są używane często nieformalnie jako skróty. Pierwsze dwa zdania z przykładu 2 mogłyby być zapisane w postaci „$x + y = y + x$ $\forall x, y \in \mathbb{R}$" oraz „$\exists n \in \mathbb{N}$ takie, że $2^n = n^2$".

(b) W praktyce często opuszczamy oczywiste kwantyfikatory. Prawa łączności i skracania dla działań w zbiorze $\mathbb{R}$ często zapisujemy w postaci

(Ł) $\qquad (x + y) + z = x + (y + z)$,

(S) $\qquad xz = yz \quad$ oraz $\quad z \neq 0$ implikuje $x = y$.

Mamy na myśli oczywiście

(Ł) $\qquad \forall x\, \forall y\, \forall z[(x + y) + z = x + (y + z)]$,

(S) $\qquad \forall x\, \forall y\, \forall z\, [(xz = yz \wedge z \neq 0) \rightarrow x = y]$,

gdzie x, y i z należą do $\mathbb{R}$. W codziennej praktyce równość (Ł) mogłaby również być zapisana jako

$$(x+y)+z=x+(y+z) \quad \forall x\,\forall y\,\forall z,$$

lub

$$(x+y)+z=x+(y+z) \quad \forall x,y,z\in\mathbb{R},$$

lub

$$(x+y)+z=x+(y+z) \quad \text{dla wszystkich} \quad x,y,z\in\mathbb{R}. \qquad \blacksquare$$

Często będziemy mogli dowodzić zdań postaci $\forall n\,p(n)$, gdzie $n\in\mathbb{N}$ za pomocą bardzo ważnej metody, mianowicie indukcji matematycznej, którą opiszemy w rozdziale 4.

Zdanie złożone postaci $\forall x\,p(x)$ będzie **fałszywe**, jeśli jedno (lub więcej) zdań $p(x)$ będzie fałszywe. Zatem, aby wykazać, że takie zdanie złożone jest fałszywe, wystarczy pokazać, że jedno z jego zdań składowych jest fałszywe. Innymi słowy, wystarczy pokazać jeden przykład zaprzeczający zdaniu ogólnemu, tzw. **kontrprzykład**.

Hipoteza Goldbacha nie została dotychczas rozstrzygnięta, gdyż nikomu nie udało się pokazać, że każda liczba parzysta większa od 4 jest sumą dwóch liczb pierwszych, ani też nikomu nie udało się znaleźć kontrprzykładu. Hipoteza ta została sprawdzona dla bardzo wielu liczb parzystych.

PRZYKŁAD 12 (a) Liczba 2 jest kontrprzykładem na stwierdzenie mówiące, że „wszystkie liczby pierwsze są nieparzyste".

(b) Liczba 7 jest kontrprzykładem na stwierdzenie „każda dodatnia liczba całkowita jest sumą trzech kwadratów liczb całkowitych". Można jednak dowieść, że każda dodatnia liczba całkowita jest sumą czterech kwadratów liczb całkowitych, na przykład $1=1^2+0^2+0^2+0^2$, $7=2^2+1^2+1^2+1^2$, $73=8^2+3^2+0^2+0^2$.

(c) Wartość liczby $n=3$ jest kontrprzykładem na stwierdzenie „$n^2\leq 2^n$ dla wszystkich $n\in\mathbb{N}$", które możemy zapisać jako „$n^2\leq 2^n\ \forall n\in\mathbb{N}$". Nie ma innych kontrprzykładów, co wykazaliśmy już w ćwiczeniu 1(c) w § 1.6.

(d) Gerald Ford jest kontrprzykładem na stwierdzenie, że „wszyscy prezydenci Stanów Zjednoczonych byli praworęczni". Istnieją trzy inne kontrprzykłady. $\blacksquare$

Jeżeli mamy dane zdanie ogólne, którego wartości logicznej nie znamy, to często jedyną strategią jest odgadnięcie tej wartości. Jeśli zgadniemy, że to zdanie jest prawdziwe, to należy przeanalizować tę sytuację, by zobaczyć, dlaczego zawsze wydaje się ono

być prawdziwe. Taka analiza może prowadzić do dowodu. Jeśli nie umiemy znaleźć dowodu i widzimy, dlaczego nie możemy go znaleźć, to może znajdziemy kontrprzykład. Jeśli teraz nie udaje nam się znaleźć kontrprzykładu, powinniśmy zacząć podejrzewać ponownie, że teza jest prawdziwa i próbować podać powody, dlaczego tak ma być. Jest to bardzo typowe, szczególnie przy bardzo trudnych zagadnieniach, że czynimy duże wysiłki próbując osiągnąć każdą z dwóch możliwości, do czasu, aż któraś okaże się być właściwa. Jeden z autorów tej książki poświęcił dużo energii na poszukiwanie kontrprzykładu do tezy, która wydawała mu się być fałszywa, a nieco później młody matematyk angielski udowodnił, że teza ta jest prawdziwa.

ĆWICZENIA DO § 2.1

1. Niech p, q, r będą następującymi zdaniami:

$$p = \text{„pada deszcz”},$$
$$q = \text{„słońce świeci”},$$
$$r = \text{„na niebie są chmury”}.$$

Zapisz następujące zdania za pomocą symboliki logicznej, używając p, q, r i spójników logicznych:

(a) Pada deszcz i świeci słońce.
(b) Jeśli pada deszcz, to na niebie są chmury.
(c) Jeśli nie pada deszcz, to nie świeci słońce i na niebie są chmury.
(d) Słońce świeci wtedy i tylko wtedy, gdy nie pada deszcz.
(e) Jeśli nie ma chmur na niebie, to świeci słońce.

2. Niech p, q i r będą takie jak w ćwiczeniu 1. Przetłumacz następujące zdania na język polski:.

(a) $(p \wedge q) \to r$. (b) $(p \to r) \to q$.
(c) $\neg p \leftrightarrow (q \vee r)$. (d) $\neg(p \leftrightarrow (q \vee r))$.
(e) $\neg(p \vee q) \wedge r$.

3. (a) Podaj wartości logiczne zdań z przykładów 1 (a)-(e).
(b) Zrób to samo dla przykładów 2 (a) i (b).

4. Które z następujących wyrażeń są zdaniami? Podaj wartości logiczne tych zdań.

(a) $x^2 = x$ $\forall x \in \mathbb{R}$.
(b) $x^2 = x$ dla pewnego $x \in \mathbb{R}$.
(c) $x^2 = x$.
(d) $x^2 = x$ dla dokładnie jednego $x \in \mathbb{R}$.
(e) $xy = xz$ implikuje $y = z$.
(f) $xy = xz$ implikuje $y = z$ $\forall x, y, x \in \mathbb{R}$.

(g) $w_1w_2 = w_1w_3$ implikuje $w_2 = w_3$ dla wszystkich słów $w_1, w_2, w_3 \in \Sigma^*$.

5. Weźmy wyrażenie niejednoznaczne „$x^2 = y^2$ implikuje $x = y$ $\forall x, y$".
(a) Zrób z tego zdanie jednoznaczne, którego wartością logiczną jest prawda.
(b) Zrób z tego zdanie jednoznaczne, którego wartością logiczną jest fałsz.

6. Podaj zdania odwrotne do następujących zdań:
(a) $q \rightarrow r$.
(b) Jeśli jestem bystry, to jestem bogaty.
(c) Jeśli $x^2 = x$, to $x = 0$ lub $x = 1$.
(d) Jeśli $2 + 2 = 4$, to $2 + 4 = 8$.

7. Podaj kontrapozycje zdań z ćwiczenia 6.

8. (a) Sprawdź, że hipoteza Goldbacha jest prawdziwa dla małych liczb, takich jak 6, 8 czy 10.
(b) Sprawdź to dla liczby 98.

9. (a) Pokaż, że wartość $n = 3$ jest kontrprzykładem na stwierdzenie $n^3 < 3^n$ $\forall n \in \mathbb{N}$.
(b) Czy umiesz znaleźć inne kontrprzykłady?

10. (a) Pokaż, że $(m, n) = (4, -4)$ jest kontrprzykładem na stwierdzenie: „jeśli m, n są niezerowymi liczbami całkowitymi, które są nawzajem podzielne przez siebie, to $m = n$".
(b) Podaj inny kontrprzykład.

11. (a) Pokaż, że $x = -1$ jest kontrprzykładem na „$(x+1)^2 \geq x^2$ $\forall x \in \mathbb{R}$".
(b) Znajdź inny kontrprzykład.
(c) Czy liczba nieujemna może być kontrprzykładem? Wyjaśnij to.

12. Znajdź kontrprzykłady na następujące stwierdzenia.
(a) $2^n - 1$ jest liczbą pierwszą dla każdego $n \geq 2$.
(b) $2^n + 3^n$ jest liczbą pierwszą $\forall n \in \mathbb{N}$.
(c) $2^n + n$ jest liczbą pierwszą dla każdej nieparzystej liczby dodatniej n.

13. (a) Podaj kontrprzykład na „$x > y$ implikuje $x^2 > y^2$ $\forall x, y \in \mathbb{R}$". Rozwiązaniem powinna być para uporządkowana (x, y).
(b) Jak mógłbyś ograniczyć x i y, by zdanie z ćwiczenia (a) było prawdziwe?

14. Niech S będzie zbiorem niepustym. Określ, które z następujących stwierdzeń są prawdziwe. Dla tych, które są prawdziwe, podaj powody. Dla tych, które są fałszywe, podaj kontrprzykłady.
(a) $A \cup B = B \cup A$ $\forall A, B \in \mathcal{P}(S)$.
(b) $(A \backslash B) \cup B = A$ $\forall A, B \in \mathcal{P}(S)$.
(c) $(A \cup B) \backslash A = B$ $\forall A, B \in \mathcal{P}(S)$.
(d) $(A \cap B) \cap C = A \cap (B \cap C)$ $\forall A, B, C \in \mathcal{P}(S)$.

15. Nawet jeśli zazwyczaj używamy sformułowań „implikuje" czy „jeśli ..., to", aby opisać implikację, to często w praktyce pojawiają się inne słowa czy sformułowania, tak jak w poniższym przykładzie. Niech p, q i r będą zdaniami:

p = „znacznik jest ustawiony",
q = „$I = 0$",
r = „podprogram S zakończył działanie".

Zapisz każde z poniższych zdań za pomocą symboliki logicznej, używając liter p, q, r i spójników logicznych.
(a) Jeśli znacznik jest ustawiony, to $I = 0$.
(b) Podprogram S zakończył działanie, jeśli znacznik jest ustawiony.
(c) Znacznik jest ustawiony, jeśli podprogram S nie zakończył działania.
(d) Kiedykolwiek $I = 0$, znacznik jest ustawiony.
(e) Podprogram S zakończył działanie tylko wtedy, gdy $I = 0$.
(f) Podprogram S zakończył działanie tylko wtedy, gdy $I = 0$ lub znacznik jest ustawiony. Zwróć uwagę na dwuznaczność: są dwa różne rozwiązania, każde mające swoją wartość logiczną. Czy pomogłoby użycie znaków przestankowych?

16. Weźmy następujące zdania:

r = „NIEPARZYSTA$(N) = T$",
m = „wyniki są wyświetlane na ekranie",
p = „wyniki są drukowane na drukarce".

Zapisz następujące zdania tak jak w ćwiczeniu 15.
(a) Wyniki są wyświetlane na ekranie, jeśli NIEPARZYSTA$(N) = T$.
(b) Wyniki są drukowane na drukarce, gdy tylko NIEPARZYSTA$(N) = T$ nie jest prawdą.
(c) NIEPARZYSTA$(N) = T$ tylko wtedy, gdy wyniki są wyświetlane na ekranie.
(d) Wyniki są wyświetlane na ekranie, jeśli wyniki są drukowane na drukarce.
(e) NIEPARZYSTA$(N) = T$ lub wyniki są wyświetlane na ekranie, jeśli wyniki są drukowane na drukarce.

17. Każde z następujących zdań wyraża implikację. Przepisz każde z nich na nowo w postaci „jeśli p, to q".
(a) Dotknij tych ciastek, jeśli chcesz dostać lanie.
(b) Dotknij tych ciastek, a będziesz żałował.
(c) Odejdź stąd albo poszczuję cię psem.
(d) Zrobię to, jeśli ty to zrobisz.
(e) Pójdę sobie, chyba że przestaniesz.

18. Zapisz kontrapozycję każdego ze zdań z ćwiczenia 17 w postaci „jeśli p, to q".

§ 2.2. Rachunek zdań

W tym rozdziale chcemy osiągnąć dwa cele. Chcemy wprowadzić formalne metody analizowania zdań i operowania zdaniami, coś na kształt algebry zdań, w pewnym sensie podobnej do algebry liczb. Potrzebna jest nam również mechaniczna metoda obliczania wartości logicznych skomplikowanych zdań. Właśnie ten obliczeniowy aspekt nadał rachunkowi zdań jego nazwę.

Jeśli zdanie jest zbudowane z innych zdań za pomocą spójników logicznych, to jego prawdziwość lub fałszywość jest całkowicie wyznaczona przez wartości logiczne tych zdań prostszych i przez sposób, w jaki to zdanie złożone jest z nich zbudowane. Dla danych zdań p i q wartości logiczne zdań złożonych $\neg p$, $p \wedge q$, $p \vee q$, $p \rightarrow q$ i $p \leftrightarrow q$ będą całkowicie wyznaczone za pomocą wartości logicznych zdań p i q. Ponieważ istnieją tylko cztery różne kombinacje wartości logicznych zdań p i q, możemy po prostu podać tablice opisujące wartości logiczne tych zdań złożonych we wszystkich możliwych kombinacjach.

Jednym ze sposobów zaznaczania wartości logicznych w tablicy jest użycie liter T (ang. Truth — prawda) i F (ang. False — fałsz). Jednak wybraliśmy inną metodę, zgodną z użyciem zmiennych booleowskich w większości języków programowania i będziemy używać liczby 1 dla oznaczenia prawdy i liczby 0 dla oznaczenia fałszu.

Zdanie $\neg p$ powinno być fałszywe, gdy zdanie p jest prawdziwe i prawdziwe, gdy p jest fałszywe. Zatem nasza tablica dla spójnika $\neg$ wygląda następująco:

p	$\neg p$
0	1
1	0

W kolumnie na lewo od linii pionowej są wypisane wszystkie możliwe wartości logiczne zdania p. Po prawej stronie znajdują się odpowiadające im wartości logiczne zdania $\neg p$.

Tablica wartości logicznych (matryca logiczna) dla spójnika $\wedge$ wygląda następująco:

p	q	$p \wedge q$
0	0	0
0	1	0
1	0	0
1	1	1

Cztery możliwe kombinacje wartości logicznych zdań p i q są wypisane po lewej stronie pionowej linii i odpowiadające im wartości

logiczne zdania $p \wedge q$ są podane po prawej stronie. Zauważmy, że wartością logiczną zdania $p \wedge q$ jest prawda dokładnie wtedy, gdy oba zdania p i q są prawdziwe.

Jak już wyjaśniliśmy na początku § 1.2, użycie spójnika „lub" w języku polskim jest nieco mylące, jednak nasze użycie $\vee$ nie będzie niejednoznaczne. Definiujemy spójnik $\vee$ w następujący sposób:

p	q	$p \vee q$
0	0	0
0	1	1
1	0	1
1	1	1

Większość ludzi zgodziłaby się z przypisaniem wartości logicznych w pierwszych trzech wierszach. Wiersz czwarty oznacza, że uważamy zdanie $p \vee q$ za prawdziwe, jeśli oba zdania p i q są prawdziwe. Jest to „alternatywa niewykluczająca", czasami zapisywana jako „i/lub". Zatem zdanie $p \vee q$ jest prawdziwe, jeśli p jest prawdziwe lub q jest prawdziwe lub oba są prawdziwe. Alternatywa wykluczająca, zapisywana za pomocą symbolu $\oplus$, oznacza, że albo jedno zdanie, albo drugie jest prawdziwe, ale nie oba jednocześnie, por. ćwiczenie 13.

Zdanie warunkowe $p \to q$ oznacza, że prawdziwość p implikuje prawdziwość q. Innymi słowy, jeśli zdanie p jest prawdziwe, to zdanie q jest prawdziwe. Jedynym przypadkiem, w którym ta implikacja może nie być prawdziwa, jest ten, gdy zdanie p jest prawdziwe, a q fałszywe.

p	q	$p \to q$
0	0	1
0	1	1
1	0	0
1	1	1

Pierwsze dwa wiersze tej matrycy logicznej dla implikacji $p \to q$ mogą zaniepokoić niektórych Czytelników, gdyż wygląda na to, że ze zdania fałszywego wynika wszystko. W rzeczywistości, po prostu definiujemy, że zdanie złożone $p \to q$ jest prawdziwe, jeśli zdanie p jest fałszywe. Takie rozumienie implikacji występuje w języku polskim. Przypuśćmy, że jakiś polityk obiecuje „jeśli zostanę wybrany, to w przyszłym roku podatki będą obniżone". Jeśli ten polityk nie zostanie wybrany, to na pewno nie uznamy go za kłamcę niezależnie od tego, jak zmienią się podatki.

Równoważność $p \leftrightarrow q$ omówimy po wprowadzeniu ogólnych matryc logicznych.

Matrycą logiczną (tablicą wartości logicznych) zdania złożonego, zbudowanego ze zdań prostych p, q, r, ..., nazywamy tablicę podającą wartości logiczne tego zdania złożonego w zależności od wartości logicznych zdań p, q, r, ... Wtedy p, q, r, ... nazywamy **zmiennymi zdaniowymi** tej matrycy i tego zdania złożonego. Wartość logiczną zdania złożonego możemy obliczyć wyznaczając po kolei wartości logiczne zdań prostszych, z których jest ono zbudowane. Pokażemy to teraz na przykładach.

PRZYKŁAD 1 Oto matryca logiczna zdania złożonego $(p \wedge q) \vee \neg(p \to q)$. Zauważmy, że występują w niej nadal tylko cztery wiersze, gdyż istnieją tylko cztery różne kombinacje wartości logicznych p i q.

kolumna	1	2	3	4	5	6
	p	q	$p \wedge q$	$p \to q$	$\neg(p \to q)$	$(p \wedge q) \vee \neg(p \to q)$
	0	0	0	1	0	0
	0	1	0	1	0	0
	1	0	0	0	1	1
	1	1	1	1	0	1

Wartości w kolumnach 3 i 4 są wyznaczone przez wartości w kolumnach 1 i 2. Wartości w kolumnie 5 są wyznaczone przez wartości w kolumnie 4. Wartości w kolumnie 6 są wyznaczone przez wartości w kolumnach 3 i 5. Szósta kolumna zawiera wartości logiczne całego zdania złożonego.

Można również użyć prostszej tablicy wartości, zbudowanej w ten sam sposób, w której jednak wartości logiczne są napisane pod spójnikami:

	p	q	$(p \wedge q)$	$\vee$	$\neg(p \to q)$	
	0	0	0	0	0	1
	0	1	0	0	0	1
	1	0	1	1	0	1
	1	1	1	1	0	1
krok	1	1	2	4	3	2

W każdym kroku wartości logiczne są wyznaczane przez wartości znalezione we wcześniejszych krokach. Na przykład wartości w kroku trzecim są wyznaczone przez wartości w ostatniej kolumnie. Wartości w kroku czwartym są wyznaczone przez wartości w trzeciej i piątej kolumnie. Kolumna utworzona w ostatnim kroku zawiera wartości logiczne całego zdania złożonego. ∎

Te prostsze matryce logiczne są bardziej poręczne, gdy mamy do czynienia z bardziej skomplikowanymi zdaniami złożonymi.

PRZYKŁAD 2 Oto matryca logiczna zdania

$$(p \to q) \wedge [(q \wedge \neg r) \to (p \vee r)].$$

p	q	r	$(p \to q)$	$\wedge$	$[(q \wedge \neg r) \to (p \vee r)]$
0	0	0	1	1	0 1 1 0
0	0	1	1	1	0 0 1 1
0	1	0	1	0	1 1 0 0
0	1	1	1	1	0 0 1 1
1	0	0	0	0	0 1 1 1
1	0	1	0	0	0 0 1 1
1	1	0	1	1	1 1 1 1
1	1	1	1	1	0 0 1 1
krok	1	1	1	2	5	3 2 4 2

Zauważmy, że wiersze matrycy logicznej mogą być wypisane w dowolnej kolejności. Wybraliśmy systematyczną metodę wypisywania kombinacji wartości logicznych zdań p, q, r częściowo z tego powodu, by mieć pewność, że wypisaliśmy je wszystkie. ∎

Równoważność $p \leftrightarrow q$ jest zdefiniowana za pomocą matrycy logicznej zdania

$$(p \to q) \wedge (q \to p).$$

p	q	$(p \to q)$	$\wedge$	$(q \to p)$
0	0	1	1	1
0	1	1	0	0
1	0	0	0	1
1	1	1	1	1
krok	1	1	2	3	2

Zatem

p	q	$(p \leftrightarrow q)$
0	0	1
0	1	0
1	0	0
1	1	1

Zdanie $p \leftrightarrow q$ jest więc prawdziwe, jeśli oba zdania p i q są prawdziwe lub oba są fałszywe. Zdanie $p \leftrightarrow q$ można wyrazić w języku polskim w następujący sposób: „p wtedy i tylko wtedy, gdy q", „p jest warunkiem koniecznym i wystarczającym dla q", „p dokładnie wtedy, gdy q".

Warto podkreślić, że zdanie złożone $p \to q$ i zdanie odwrotne do niego $q \to p$ są różne; mają one różne matryce logiczne.

Ważną klasę zdań złożonych tworzą zdania, które są zawsze prawdziwe, niezależnie od wartości logicznych zmiennych zdaniowych p, q itd. Takie zdanie złożone nazywamy **tautologią**. Dlaczego w ogóle interesujemy się zdaniami zawsze prawdziwymi, a przez to dość nudnymi? Odpowiedź jest taka, że będziemy zajmować się dość skomplikowanymi zdaniami i będziemy chcieli pokazać, że są one prawdziwe. Metoda, którą do tego celu wybierzemy, będzie polegała na wykorzystywaniu innych zdań, o których wiemy, że są zawsze prawdziwe. Poczekajmy. Najpierw obejrzyjmy bardzo prostą tautologię.

PRZYKŁAD 3

(a) Klasyczną tautologią jest zdanie złożone $p \to p$:

p	$p \to p$
0	1
1	1

(b) Zdanie złożone $[p \wedge (p \to q)] \to q$ jest tautologią:

	p	q	$[p \wedge (p \to q)]$		$\to$	q
	0	0	0	1	1	
	0	1	0	1	1	
	1	0	0	0	1	
	1	1	1	1	1	
krok	1	1	3	2	4	

(c) Zdanie $\neg(p \vee q) \leftrightarrow (\neg p \wedge \neg q)$ jest tautologią:

p	q	$\neg(p \vee q)$		$\leftrightarrow$	$(\neg p \wedge \neg q)$		
0	0	1	0	1	1	1	1
0	1	0	1	1	1	0	0
1	0	0	1	1	0	0	1
1	1	0	1	1	0	0	0
krok 1	1	3	2	4	2	3	2

■

Zdaniem sprzecznym nazywamy zdanie złożone, które jest zawsze fałszywe. Oczywiście zdanie złożone P jest zdaniem sprzecznym wtedy i tylko wtedy, gdy zdanie $\neg P$ jest tautologią.

PRZYKŁAD 4

Klasycznym zdaniem sprzecznym jest zdanie złożone $p \wedge \neg p$:

p	p	$\wedge$	$\neg p$
0	0	0	1
1	1	0	0

■

Dwa zdania złożone P i Q są **zdaniami logicznie równoważnymi**, jeśli mają takie same wartości logiczne dla wszystkich kombinacji wartości logicznych ich zmiennych zdaniowych p, q itd. Innymi słowy, kolumny ostatecznych wartości logicznych w ich matrycach logicznych są takie same. W takim przypadku piszemy $P \Leftrightarrow Q$. Ponieważ w matrycy logicznej dla zdania $P \leftrightarrow Q$ wartość prawdy występuje dokładnie tam, gdzie wartości logiczne zdań P i Q zgadzają się, więc widzimy, że

$P \Leftrightarrow Q$ wtedy i tylko wtedy, gdy zdanie $P \leftrightarrow Q$ jest tautologią.

Zauważenie, że $P \Leftrightarrow Q$, jest szczególnie przydatne wtedy, gdy zdania P i Q znacznie się różnią (por. na przykład zdania w tabl. 2.1).

PRZYKŁAD 5

(a) W przykładzie 3(c) zobaczyliśmy, że zdania złożone $\neg(p \vee q)$ i $\neg p \wedge \neg q$ są logicznie równoważne. Zatem $\neg(p \vee q) \Leftrightarrow (\neg p \wedge \neg q)$. Stwierdzenie, że nieprawdą jest, iż pójdę na spacer lub będę oglądał telewizję, jest tym samym, co stwierdzenie, że nie pójdę na spacer i nie będę oglądał telewizji.

(b) Sama natura spójników $\vee$ i $\wedge$ sugeruje, że $p \vee q \Leftrightarrow q \vee p$ i $p \wedge q \Leftrightarrow q \wedge p$. Można sprawdzić, że tak jest naprawdę, pokazując, iż zdania $p \vee q \leftrightarrow q \vee p$ i $p \wedge q \leftrightarrow q \wedge p$ są tautologiami. ∎

Warto podkreślić różnicę między $\leftrightarrow$ i $\Leftrightarrow$. Wyrażenie $P \leftrightarrow Q$ po prostu oznacza pewne zdanie złożone, które może być tautologią, ale które też może nie być tautologią. Wyrażenie $P \Leftrightarrow Q$ jest stwierdzeniem o zdaniach, mówiącym, że zdania P i Q są logicznie równoważne, tzn. że zdanie $P \leftrightarrow Q$ jest tautologią.

W tablicy 2.1 wymienionych jest wiele par zdań logicznie równoważnych, wybranych ze względu na ich przydatność. Aby otrzymać z nich tautologie, trzeba po prostu zastąpić każdy znak $\Leftrightarrow$ znakiem $\leftrightarrow$. Za pomocą matryc logicznych można sprawdzić, że te wszystkie zdania są tautologiami. Większość z nich powinna być przy tym intuicyjnie oczywista.

Wiele równoważności w tablicy 2.1 ma swoje nazwy, które podaliśmy, jednak nie ma potrzeby pamiętania większości z nich. Równoważności 2, 3 i 4 mają znane nazwy. Równoważności 8 — prawa De Morgana — i 9 — **prawo kontrapozycji** — pojawiają się wystarczająco często, by warto było zapamiętać ich nazwy.

Zauważmy, że w prawach De Morgana zdanie po jednej stronie znaku $\Leftrightarrow$ zawiera spójnik $\wedge$, podczas gdy zdanie po drugiej stronie zawiera spójnik $\vee$. W paragrafie 2.4 zobaczymy, że prawa De Morgana pozwalają zastąpić dane zdanie zdaniem logicznie

Tablica 2.1. Zdania logicznie równoważne[1]

1. $(\neg\neg p) \Leftrightarrow p$	prawo podwójnego przeczenia
2a. $(p \vee q) \Leftrightarrow (q \vee p)$ **b.** $(p \wedge q) \Leftrightarrow (q \wedge p)$ **c.** $(p \leftrightarrow q) \Leftrightarrow (q \leftrightarrow p)$	prawa przemienności
3a. $[(p \vee q) \vee r] \Leftrightarrow [p \vee (q \vee r)]$ **b.** $[(p \wedge q) \wedge r] \Leftrightarrow [p \wedge (q \wedge r)]$	prawa łączności
4a. $[p \vee (q \wedge r)] \Leftrightarrow [(p \vee q) \wedge (p \vee r)]$ **b.** $[p \wedge (q \vee r)] \Leftrightarrow [(p \wedge q) \vee (p \wedge r)]$	prawa rozdzielności
5a. $(p \vee p) \Leftrightarrow p$ **b.** $(p \wedge p) \Leftrightarrow p$	prawa idempotentności
6a. $(p \vee c) \Leftrightarrow p$ **b.** $(p \vee t) \Leftrightarrow t$ **c.** $(p \wedge c) \Leftrightarrow c$ **d.** $(p \wedge t) \Leftrightarrow p$	prawa identyczności
7a. $(p \vee \neg p) \Leftrightarrow t$ **b.** $(p \wedge \neg p) \Leftrightarrow c$	
8a. $\neg(p \vee q) \Leftrightarrow (\neg p \wedge \neg q)$ **b.** $\neg(p \wedge q) \Leftrightarrow (\neg p \vee \neg q)$ **c.** $(p \vee q) \Leftrightarrow \neg(\neg p \wedge \neg q)$ **d.** $(p \wedge q) \Leftrightarrow \neg(\neg p \vee \neg q)$	prawa De Morgana
9. $(p \to q) \Leftrightarrow (\neg q \to \neg p)$	prawo kontrapozycji
10a. $(p \to q) \Leftrightarrow (\neg p \vee q)$ **b.** $(p \to q) \Leftrightarrow \neg(p \wedge \neg q)$	określenie implikacji za pomocą alternatywy lub koniunkcji
11a. $(p \vee q) \Leftrightarrow (\neg p \to q)$ **b.** $(p \wedge q) \Leftrightarrow \neg(p \to \neg q)$	
12a. $[(p \to r) \wedge (q \to r)] \Leftrightarrow [(p \vee q) \to r]$ **b.** $[(p \to q) \wedge (p \to r)] \Leftrightarrow [p \to (q \wedge r)]$	
13. $(p \leftrightarrow q) \Leftrightarrow [(p \to q) \wedge (q \to p)]$	określenie równoważności
14. $[(p \wedge q) \to r] \Leftrightarrow [p \to (q \to r)]$	prawo eksportacji
15. $(p \to q) \Leftrightarrow [(p \wedge \neg q) \to c]$	reductio ad absurdum

[1] W tej tablicy t jest dowolną tautologią, a c dowolnym zdaniem sprzecznym.

równoważnym, w którym pewne lub wszystkie spójniki $\wedge$ zostały zamienione na $\vee$ lub na odwrót.

Dla danych dwóch zdań złożonych P i Q mówimy, że zdanie P implikuje logicznie zdanie Q, jeśli zdanie Q ma wartość logiczną prawdy zawsze wtedy, gdy zdanie P ma wartość logiczną prawdy. Piszemy wtedy $P \Rightarrow Q$. Zauważmy, że

$P \Rightarrow Q$ wtedy i tylko wtedy, gdy zdanie złożone
$P \to Q$ jest tautologią.

Równoważnie, $P \Rightarrow Q$ oznacza, że zdania P i Q nie mogą nigdy jednocześnie przyjąć odpowiednio wartości 1 i 0; gdy P jest prawdziwe, to Q też jest prawdziwe i jeśli Q jest fałszywe, to P też jest fałszywe.

PRZYKŁAD 6 (a) Mamy $[p \wedge (p \to q)] \Rightarrow q$, gdyż $[p \wedge (p \to q)] \to q$ jest tautologią; por. przykład 3(b).

(b) Zdanie $(A \wedge B) \Rightarrow C$ oznacza, że $(A \wedge B) \to C$ jest tautologią. Ponieważ $(A \wedge B) \to C \Leftrightarrow A \to (B \to C)$ na podstawie prawa 14 (eksportacji), więc zdanie $(A \wedge B) \to C$ jest tautologią wtedy i tylko wtedy, gdy zdanie $A \to (B \to C)$ jest tautologią, tzn. wtedy i tylko wtedy, gdy $A \Rightarrow (B \to C)$. Zatem stwierdzenia $(A \wedge B) \Rightarrow C$ i $A \Rightarrow (B \to C)$ oznaczają to samo. ∎

Tablica 2.2. Implikacje logiczne

16.	$p \Rightarrow (p \vee q)$	wprowadzanie alternatywy
17.	$(p \wedge q) \Rightarrow p$	opuszczanie koniunkcji
18.	$(p \to c) \Rightarrow \neg p$ (c — dowolne zdanie sprzeczne)	sprowadzenie do sprzeczności
19.	$[p \wedge (p \to q)] \Rightarrow q$	modus ponendo ponens
20.	$[(p \to q) \wedge \neg q] \Rightarrow \neg p$	modus tollendo tollens
21.	$[(p \vee q) \wedge \neg p] \Rightarrow q$	modus ponendo tollens
22.	$p \Rightarrow [q \to (p \wedge q)]$	
23.	$[(p \leftrightarrow q) \wedge (q \leftrightarrow r)] \Rightarrow (p \leftrightarrow r)$	przechodniość $\leftrightarrow$
24.	$[(p \to q) \wedge (q \to r)] \Rightarrow (p \to r)$	przechodniość $\to$
25a.	$(p \to q) \Rightarrow [(p \vee r) \to (q \vee r)]$	
b.	$(p \to q) \Rightarrow [(p \wedge r) \to (q \wedge r)]$	
c.	$(p \to q) \Rightarrow [(q \to r) \to (p \to r)]$	
26a.	$[(p \to q) \wedge (r \to s)] \Rightarrow [(p \vee r) \to (q \vee s)]$	prawa dylematu konstrukcyjnego
b.	$[(p \to q) \wedge (r \to s)] \Rightarrow [(p \wedge r) \to (q \wedge s)]$	
27a.	$[(p \to q) \wedge (r \to s)] \Rightarrow [(\neg q \vee \neg s) \to (\neg p \vee \neg r)]$	prawa dylematu destrukcyjnego
b.	$[(p \to q) \wedge (r \to s)] \Rightarrow [(\neg q \wedge \neg s) \to (\neg p \wedge \neg r)]$	

W tablicy 2.2 podajemy wiele przydatnych implikacji logicznych. Każde wynikanie stanie się tautologią, gdy znak $\Rightarrow$ zastą-

pimy znakiem $\rightarrow$. Podobnie jak w tablicy 2.1, wiele implikacji ma swoje nazwy, nie trzeba jednak ich zapamiętywać.

Aby wykazać logiczną implikację $P \Rightarrow Q$, wystarczy tylko wyszukać te wiersze matrycy logicznej, w których zdanie P jest prawdziwe, a zdanie Q fałszywe. Jeśli są takie wiersze, to implikacja $P \Rightarrow Q$ nie zachodzi. W przeciwnym przypadku implikacja $P \Rightarrow Q$ jest prawdziwa. Zatem możemy zaniedbać te wiersze, w których zdanie P jest fałszywe oraz te, w których zdanie Q jest prawdziwe.

PRZYKŁAD 7

(a) Sprawdzimy prawdziwość implikacji $(p \wedge q) \Rightarrow p$. Musimy rozważać jedynie przypadek, w którym zdanie $p \wedge q$ jest prawdziwe, tzn. gdy oba zdania p i q są prawdziwe. Zatem rozważamy matrycę ograniczoną

p	q	$(p \wedge q)$	$\rightarrow$	p
1	1	1	1	1

(b) Aby sprawdzić prawdziwość implikacji $\neg p \Rightarrow (p \rightarrow q)$ wystarczy rozpatrzeć tylko te przypadki, w których zdanie $\neg p$ jest prawdziwe, tzn. gdy zdanie p jest fałszywe. Matryca ograniczona ma wtedy postać:

p	q	$\neg p$	$\rightarrow$	$(p \rightarrow q)$
0	0	1	1	1
0	1	1	1	1

Można to zrobić szybciej, mianowicie rozważyć przypadek, w którym zdanie $p \rightarrow q$ jest fałszywe, tzn. p jest prawdziwe i q fałszywe. Wtedy matryca ma jeden wiersz:

p	q	$\neg p$	$\rightarrow$	$(p \rightarrow q)$
1	0	0	1	0

(c) Sprawdzimy prawdziwość implikacji 26a. Pełna matryca logiczna miałaby 16 wierszy. Jednakże wystarczy sprawdzić tylko te przypadki, w których implikacja $(p \vee r) \rightarrow (q \vee s)$ może być fałszywa. Zatem wystarczy sprawdzić tylko te przypadki, w których zdanie $q \vee s$ jest fałszywe, tzn. oba zdania q i s są fałszywe.

p	q	r	s	$[(p \rightarrow q) \wedge (r \rightarrow s)]$			$\rightarrow$	$[(p \vee r) \rightarrow (q \vee s)]$		
0	0	0	0	1	1	1	1	0	1	0
0	0	1	0	1	0	0	1	1	0	0
1	0	0	0	0	0	1	1	1	0	0
1	0	1	0	0	0	0	1	1	0	0
krok	1	1	1	2	3	2	4	2	3	2

■

Bezpośrednio z definicji równoważności logicznej $\Leftrightarrow$ wynika, że jeśli $P \Leftrightarrow Q$ i $Q \Leftrightarrow R$, to $P \Leftrightarrow R$ i wszystkie trzy zdania są sobie równoważne. Jeśli mamy ciąg równoważności $P_1 \Leftrightarrow P_2 \Leftrightarrow \ldots \Leftrightarrow P_n$, to wszystkie zdania P_i są równoważne. Podobnie (por. ćwiczenie 23(a)), jeśli $P \Rightarrow Q$ i $Q \Rightarrow R$, to $P \Rightarrow R$. Symbole $\Leftrightarrow$ i $\Rightarrow$ zachowują się trochę podobnie do symboli $=$ oraz $\geq$ w algebrze. Mówimy czasem, że zdanie P jest **silniejsze niż** Q lub Q jest **słabsze niż** P, gdy $P \Rightarrow Q$.

ĆWICZENIA DO § 2.2

1. Podaj zdanie odwrotne i przeciwstawne (kontrapozycję) dla każdego z następujących zdań:
(a) $p \to (q \wedge r)$.
(b) Jeśli $x + y = 1$, to $x^2 + y^2 \geq 1$.
(c) Jeśli $2 + 2 = 4$, to $3 + 3 = 8$.

2. Weźmy zdanie „jeśli $x > 0$, to $x^2 > 0$". Zakładamy, że $x \in \mathbb{R}$.
(a) Podaj zdanie odwrotne i przeciwstawne do tego zdania.
(b) Które z następujących zdań są prawdziwe: zdanie pierwotne, odwrotne do niego i przeciwstawne?

3. Weźmy następujące zdania:

$$p \to q, \quad \neg p \to \neg q, \quad q \to p, \quad \neg q \to \neg p,$$
$$q \wedge \neg p, \quad \neg p \vee q, \quad \neg q \vee p, \quad p \wedge \neg q.$$

(a) Które zdanie jest zdaniem odwrotnym do zdania $p \to q$?
(b) Które zdanie jest zdaniem przeciwstawnym do zdania $p \to q$?
(c) Które zdania są logicznie równoważne ze zdaniem $p \to q$?

4. Określ wartości logiczne następujących zdań złożonych:
(a) Jeśli $2 + 2 = 4$, to $2 + 4 = 8$.
(b) Jeśli $2 + 2 = 5$, to $2 + 4 = 8$.
(c) Jeśli $2 + 2 = 4$, to $2 + 4 = 6$.
(d) Jeśli $2 + 2 = 5$, to $2 + 4 = 6$.
(e) Jeśli świat jest płaski, to Juliusz Cezar był pierwszym prezydentem Stanów Zjednoczonych.
(f) Jeśli świat jest płaski, to George Washington był pierwszym prezydentem Stanów Zjednoczonych.
(g) Jeśli George Washington był pierwszym prezydentem Stanów Zjednoczonych, to świat jest płaski.
(h) Jeśli George Washington był pierwszym prezydentem Stanów Zjednoczonych, to $2 + 2 = 4$.

5. Załóżmy, że wiadomo, iż zdanie $p \to q$ jest fałszywe. Podaj wartości logiczne zdań:

(a) $p \wedge q$. (b) $p \vee q$. (c) $q \to p$.

6. Zbuduj matrycę logiczną dla zdań:
(a) $p \wedge \neg p$. (b) $p \vee \neg p$. (c) $p \leftrightarrow \neg p$. (d) $\neg\neg p$.

7. Zbuduj matrycę logiczną dla zdań:
(a) $\neg(p \wedge q)$. (b) $\neg(p \vee q)$. (c) $\neg p \wedge \neg q$. (d) $\neg p \vee \neg q$.

8. Zbuduj matrycę logiczną dla zdania: $(p \to q) \to [(p \vee \neg q) \to (p \wedge q)]$.

9. Zbuduj matrycę logiczną dla zdania: $[(p \vee q) \wedge r] \to (p \wedge \neg q)$.

10. Zbuduj matrycę logiczną dla zdania: $[(p \leftrightarrow q) \vee (p \to r)] \to (\neg q \wedge p)$.

11. Zbuduj matrycę logiczną dla zdań:
(a) $\neg(p \vee q) \to r$. (b) $\neg((p \vee q) \to r)$.
Ćwiczenie to pokazuje, że trzeba ostrożnie używać nawiasów. Zajmiemy się tym dokładniej w § 7.2, szczególnie w ćwiczeniu 15.

12. W których z poniższych zdań spójnik lub oznacza alternatywę niewykluczającą?
(a) Do wyboru zupa lub sałatka.
(b) Aby wstąpić na uniwersytet, kandydat musi zaliczyć roczny kurs fizyki lub chemii w szkole średniej.
(c) Publikuj lub giń. (Publish or perish).
(d) Znajomość Fortranu lub Pascala pożądana.
(e) Praca zostanie zakończona w czwartek lub piątek.
(f) Zniżka dla osób poniżej 20 lat lub powyżej 60 lat.
(g) Łowienie ryb lub polowanie jest zabronione.
(h) Szkoła nie będzie otwarta w lipcu lub sierpniu.

13. **Alternatywa wykluczająca** lub spójnik $\oplus$ (informatycy używają oznaczenia **XOR**) jest zdefiniowana za pomocą matrycy:

p	q	$p \oplus q$
0	0	0
0	1	1
1	0	1
1	1	0

(a) Pokaż, że $p \oplus q$ ma te same wartości logiczne co zdanie $\neg(p \leftrightarrow q)$.
(b) Zbuduj matrycę logiczną dla zdania $p \oplus p$.
(c) Zbuduj matrycę logiczną dla zdania $(p \oplus q) \oplus r$.
(d) Zbuduj matrycę logiczną dla zdania $(p \oplus p) \oplus p$.

14. (a) Napisz zdanie złożone, które jest prawdziwe wtedy, gdy dokładnie jedno z trzech zdań p, q i r jest prawdziwe.
(b) Napisz zdanie złożone, które jest prawdziwe wtedy, gdy dokładnie dwa z trzech zdań p, q i r są prawdziwe.

15. (a) Zapisz zdania (g) i (h) z ćwiczenia 12 używając praw De Morgana.
(b) Czy prawa De Morgana zachodzą dla alternatywy wykluczającej $\oplus$? Uzasadnij odpowiedź.

16. Udowodnij następujące równoważności lub wykaż, że są one nieprawdziwe:
(a) $[p \to (q \to r)] \Leftrightarrow [(p \to q) \to (p \to r)]$,
(b) $[p \oplus (q \to r)] \Leftrightarrow [(p \oplus q) \to (p \oplus r)]$,
(c) $[(p \to q) \to r] \Leftrightarrow [p \to (q \to r)]$,
(d) $[(p \leftrightarrow q) \leftrightarrow r] \Leftrightarrow [p \leftrightarrow (q \leftrightarrow r)]$,
(e) $(p \oplus q) \oplus r \Leftrightarrow p \oplus (q \oplus r)$.

17. Sprawdź za pomocą matryc logicznych prawdziwość:
(a) prawa 12a,
(b) prawa 14,
(c) prawa 15.

18. Sprawdź za pomocą matryc logicznych prawdziwość następujących implikacji logicznych:
(a) modus tollendo tollens (prawo 20),
(b) modus ponendo tollens (prawo 21).

19. Sprawdź za pomocą matryc logicznych, stosując skrótową metodę opisaną w przykładzie 7, prawdziwość:
(a) prawa 25b,
(b) prawa 25c,
(c) prawa 26b.

20. Udowodnij następujące równoważności i implikacje lub wykaż, że są one nieprawdziwe. Zauważ, że wystarczy tylko jeden wiersz w matrycy logicznej, by pokazać, że zdanie nie jest tautologią.
(a) $(q \to p) \Leftrightarrow (p \wedge q)$.
(b) $(p \wedge \neg q) \Rightarrow (p \to q)$.
(c) $(p \wedge q) \Rightarrow (p \vee q)$.

21. Matka, będąca z zawodu logikiem, powiedziała swojemu synowi: „jeśli nie dokończysz kolacji, nie będziesz mógł oglądać dłużej telewizji dziś wieczorem". Syn zjadł kolację, ale wtedy został natychmiast wysłany do łóżka. Przedyskutuj tę sytuację.

22. Rozważ zdanie: „beton nie zwiąże się, jeśli go nie polejesz wodą".
(a) Zbuduj zdanie przeciwstawne.
(b) Zbuduj zdanie odwrotne.
(c) Zbuduj zdanie odwrotne do zdania przeciwstawnego.
(d) Które ze zdań: zdanie oryginalne oraz zdania z punktów (a), (b) i (c) jest prawdziwe?

23. (a) Pokaż, że jeśli $A \Rightarrow B$ i $B \Rightarrow C$, to $A \Rightarrow C$.
(b) Pokaż, że jeśli $P \Leftrightarrow Q$, $Q \Rightarrow R$ i $R \Rightarrow S$, to $P \Rightarrow S$.
(c) Pokaż, że jeśli $P \Rightarrow Q$, $Q \Rightarrow R$ i $R \Rightarrow P$, to $P \Leftrightarrow Q$.

§ 2.3. Metody dowodzenia

Duży nacisk kładziony na logikę i dowodzenie odróżnia matematykę od innych dziedzin. Dowody stosowane w codziennej praktyce matematycznej opierają się na podstawach logicznych, które wprowadziliśmy wraz z rachunkiem zdań. W paragrafie 2.4 wprowadzimy pojęcie dowodu formalnego, używając do tego symboliki rachunku zdań, a w § 2.5 zastosujemy ten formalizm do przeanalizowania typowych rozumowań z życia codziennego.

Paragraf ten zawiera wstępne, nieformalne omówienie powszechnie stosowanych metod dowodzenia i związanej z nimi standardowej terminologii. Zazwyczaj mamy do czynienia ze zbiorem założeń $H_1, \ldots, H_n$, z których chcemy wyprowadzić wniosek C. Jedną z najbardziej naturalnych metod dowodzenia jest **dowód wprost**, w którym pokazujemy

$$H_1 \wedge H_2 \wedge \ldots \wedge H_n \Rightarrow C. \tag{1}$$

Dowody podane w rozdziale 1 były głównie dowodami tego typu.

Jednym z rodzajów **dowodu nie wprost** jest dowód **kontrapozycji**

$$\neg C \Rightarrow \neg(H_1 \wedge H_2 \wedge \ldots \wedge H_n). \tag{2}$$

Zgodnie z prawem 9 w tablicy 2.1 w § 2.2, implikacja (2) jest prawdziwa wtedy i tylko wtedy, gdy prawdziwa jest implikacja (1), a więc dowód implikacji (2) pozwoli nam wyprowadzić wniosek C z założeń $H_1, \ldots, H_n$.

PRZYKŁAD 1 Niech $m, n \in \mathbb{N}$. Chcemy dowieść, że jeśli $m + n \geq 73$, to $m \geq 37$ lub $n \geq 37$. W tym celu dowodzimy kontrapozycji: jeśli nieprawdą jest, że „$m \geq 37$ lub $n \geq 37$", to nieprawdą jest, że „$m + n \geq 73$". Na podstawie prawa De Morgana zaprzeczeniem zdania „$m \geq 37$ lub $n \geq 37$" jest „nieprawda, że $m \geq 37$ i nieprawda, że $n \geq 37$", tzn. „$m \leq 36$ i $n \leq 36$". Zatem zdaniem przeciwstawnym jest zdanie: „jeśli $m \leq 36$ i $n \leq 36$, to $m + n \leq 72$". To zdanie wynika natychmiast z ogólnej własności nierówności: jeśli $a \leq c$ i $b \leq d$, to $a + b \leq c + d$ dla dowolnych liczb rzeczywistych a, b, c, d. ∎

Innym rodzajem **dowodu nie wprost** jest dowód przez **sprowadzenie do sprzeczności**:

$$H_1 \wedge H_2 \wedge \ldots \wedge H_n \wedge \neg C \Rightarrow \text{ sprzeczność}. \tag{3}$$

Prawo 15 w tablicy 2.1 w § 2.2 mówi nam, że implikacja (3) jest prawdziwa wtedy i tylko wtedy, gdy prawdziwa jest implikacja (1).

PRZYKŁAD 2 Chcemy dowieść, że $\sqrt{2}$ jest liczbą niewymierną, tzn. jeśli $x \in \mathbb{R}$ i $x^2 = 2$, to x nie jest liczbą wymierną. Własność bycia liczbą niewymierną jest własnością o charakterze negatywnym i nie jest łatwo sprawdzić ją bezpośrednio. Możemy jednakże pokazać, że założenia $x^2 = 2$ i x jest liczbą wymierną (tzn. nie niewymierną) razem prowadzą do sprzeczności.

Przyjmijmy, że $x \in \mathbb{R}$, że $x^2 = 2$ i że x jest liczbą wymierną. Wtedy z definicji liczby wymiernej mamy $x = p/q$, gdzie $p, q \in \mathbb{Z}$ oraz $q \neq 0$. Po ewentualnym skróceniu ułamka możemy przyjąć, że p i q nie mają wspólnych dzielników. W szczególności p i q nie mogą być jednocześnie parzyste. Ponieważ $2 = x^2 = p^2/q^2$, mamy $p^2 = 2q^2$, a zatem p^2 jest liczbą parzystą. Stąd wynika, że p jest liczbą parzystą, jak pokażemy to w przykładzie 4. Stąd $p = 2k$ dla jakiegoś $k \in \mathbb{Z}$. Wtedy $(2k)^2 = 2q^2$, a zatem $2k^2 = q^2$. Tak więc q^2 i q są również parzyste. Ale wtedy obie liczby p i q byłyby parzyste, co jest sprzeczne z naszym wcześniejszym stwierdzeniem. Zatem $\sqrt{2}$ jest liczbą niewymierną. ∎

PRZYKŁAD 3 Udowodnimy przez sprowadzenie do sprzeczności, że istnieje nieskończenie wiele liczb pierwszych. Załóżmy zatem, że istnieje skończenie wiele liczb pierwszych, powiedzmy, że jest ich k. Oznaczmy je przez $p_1, p_2, \ldots, p_k$, tak więc $p_1 = 2$, $p_2 = 3$ itd. Niech $n = 1 + p_1 p_2 \cdot \ldots \cdot p_k$. Ponieważ $n > p_j$ dla wszystkich $j = 1, 2, \ldots, k$, więc liczba n nie jest pierwsza. Jednakże jest ona iloczynem liczb pierwszych (dowód tego łatwego do uwierzenia faktu nie jest trywialny, podamy go później, w przykładzie 1 w § 4.5). Zatem co najmniej jedna liczba p_j musi dzielić liczbę n. Ponieważ każda liczba p_j dzieli $n - 1$, więc co najmniej jedna liczba p_j musi dzielić obie liczby n i $n - 1$, co jednak jest niemożliwe. Rzeczywiście, jeśli p_j dzieli n i $n - 1$, to dzieli też ich różnicę, czyli 1, co jest niemożliwe. ∎

Powinno się jednak unikać sztucznych dowodów przez sprowadzenie do sprzeczności, takich jak dowód w następnym przykładzie.

PRZYKŁAD 4 Udowodnimy przez sprowadzenie do sprzeczności, że iloczyn dwóch liczb nieparzystych jest liczbą nieparzystą. Załóżmy, że liczby $m, n \in \mathbb{N}$ są nieparzyste, ale ich iloczyn mn jest parzysty. Istnieją wtedy liczby $k, l \in \mathbb{N}$ takie, że $m = 2k + 1$ i $n = 2l + 1$. Wtedy

$$mn = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1,$$

czyli mn jest liczbą nieparzystą, co przeczy założeniu, że mn jest liczbą parzystą.

Ten dowód przez sprowadzenie do sprzeczności jest sztuczny, gdyż nie skorzystaliśmy z naszego założenia, że „liczba mn jest parzysta" dotąd, aż stwierdziliśmy, że „liczba mn jest nieparzysta". Znacznie lepiej jest przeprowadzić następujący dowód wprost.

Weźmy liczby nieparzyste $m, n \in \mathbb{N}$. Istnieją liczby $k, l \in \mathbb{N}$ takie, że $m = 2k + 1$ i $n = 2l + 1$. Wtedy liczba

$$mn = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1$$

jest nieparzysta. ∎

Implikacja postaci

$$H_1 \vee H_2 \vee \ldots \vee H_n \Rightarrow C$$

jest równoważna

$$(H_1 \Rightarrow C) \text{ i } (H_2 \Rightarrow C) \text{ i } \ldots \text{ i } (H_n \Rightarrow C)$$

(dla $n = 2$ zob. prawo 12a w tablicy 2.1 w § 2.2), a więc można jej dowieść rozpatrując przypadki, tzn. dowodząc każdej implikacji $H_1 \Rightarrow C, \ldots, H_n \Rightarrow C$ oddzielnie. Następny przykład pokazuje, jak nudny i zawierający powtórzenia może być dowód polegający na rozpatrywaniu przypadków.

PRZYKŁAD 5 Przypomnijmy, że **wartość bezwzględna** $|x|$ liczby $x \in \mathbb{R}$ jest określona wzorem:

$$|x| = \begin{cases} x, & \text{jeśli } x \geq 0 \\ -x, & \text{jeśli } x < 0. \end{cases}$$

Zakładając znane własności relacji porządku $\leq$ w zbiorze $\mathbb{R}$, dowodzimy, że

$$|x + y| \leq |x| + |y| \quad \text{dla} \quad x, y \in \mathbb{R}.$$

Rozważamy cztery przypadki: **1.** $x \geq 0$ i $y \geq 0$; **2.** $x \geq 0$ i $y < 0$; **3.** $x < 0$ i $y \geq 0$; **4.** $x < 0$ i $y < 0$.

Przypadek 1. Jeśli $x \geq 0$ i $y \geq 0$, to $x + y \geq 0$, a więc $|x + y| = x + y = |x| + |y|$.

Przypadek 2. Jeśli $x \geq 0$ i $y < 0$, to

$$x + y < x + 0 = |x| \leq |x| + |y|$$

oraz

$$-(x + y) = -x + (-y) \leq 0 + (-y) = |y| \leq |x| + |y|.$$

Mamy albo $|x+y| = x+y$, albo $|x+y| = -(x+y)$; w każdym z tych przypadków otrzymujemy, że $|x+y| \leq |x|+|y|$ na podstawie powyższych nierówności.

Przypadek 3. Przypadek $x < 0$ i $y \geq 0$ jest podobny do przypadku 2.

Przypadek 4. Jeśli $x < 0$ i $y < 0$, to $x+y < 0$ i $|x+y| = -(x+y) = -x+(-y) = |x|+|y|$.

Zatem we wszystkich czterech przypadkach $|x+y| \leq |x|+|y|$.■

PRZYKŁAD 6 Dla każdego $n \in \mathbb{N}$ liczba n^3+n jest parzysta. Możemy tego dowieść rozpatrując przypadki.

Przypadek 1. Przypuśćmy, że n jest liczbą parzystą. Wtedy $n = 2k$ dla jakiejś liczby $k \in \mathbb{N}$, a więc liczba

$$n^3+n = 8k^3+2k = 2(4k^3+k)$$

jest parzysta.

Przypadek 2. Przypuśćmy, że liczba n jest nieparzysta; wtedy $n = 2k+1$ dla jakiejś liczby $k \in \mathbb{N}$, a zatem liczba

$$n^3+n = (8k^3+12k^2+6k+1)+(2k+1) = 2(4k^3+6k^2+4k+1)$$

jest parzysta.

Oto bardziej elegancki dowód też polegający na rozpatrywaniu przypadków. Dla danej liczby $n \in \mathbb{N}$ mamy $n^3+n = n(n^2+1)$. Jeśli n jest liczbą parzystą, to $n(n^2+1)$ też jest liczbą parzystą. Jeśli n jest liczbą nieparzystą, to n^2 jest liczbą nieparzystą, czyli n^2+1 jest liczbą parzystą, a więc $n(n^2+1)$ jest liczbą parzystą. ■

Czasami mówi się, że implikacja $P \Rightarrow Q$ jest spełniona **„w próżni"**, jeśli zdanie P jest fałszywe. Przykład 7(b) w § 2.2 pokazał, że $\neg p \Rightarrow (p \to q)$, a więc jeśli zdanie $\neg P$ jest prawdziwe, to zdanie $P \to Q$ też jest prawdziwe. **Dowód „w próżni"** jest dowodem implikacji $P \Rightarrow Q$, w którym pokazuje się, że zdanie P jest fałszywe. Implikacje takie rzadko mają istotne znaczenie. Zazwyczaj pojawiają się one w dowodach ogólnych zdań, polegających na rozpatrywaniu przypadków. Przypadek, w którym występuje dowód „w próżni", jest zazwyczaj tym przypadkiem, w którym założenie P zostaje wykluczone; w pewnym sensie w tym przypadku nie ma nic do sprawdzenia. Chociaż implikacja $P \Rightarrow Q$ jest prawdziwa w tym przypadku, nie dowiadujemy się niczego o Q.

PRZYKŁAD 7 (a) Weźmy skończone zbiory A i B. Zdanie „Jeśli zbiór A ma mniej elementów niż zbiór B, to istnieje przekształcenie wzajemnie jednoznaczne A na pewien podzbiór właściwy zbioru B"

jest prawdziwe „w próżni", jeśli B jest zbiorem pustym, ponieważ w takim przypadku założenie musi być fałszywe. Dowód „w próżni" polega tutaj na prostej obserwacji, że to założenie jest niemożliwe.

(b) Zdanie „$n \geq 4m^2$ implikuje $n^m < 2^n$", które rozpatrywaliśmy w przykładzie 2(b) w § 1.6, jest prawdziwe „w próżni" dla $n = 0, 1, 2, 3, \ldots, 4m^2 - 1$. Dla tych wartości n jego wartość logiczna nie zależy od tego, czy zdanie $n^m < 2^n$ jest prawdziwe oraz nie mówi nam nic o tym, czy $n^m < 2^n$. ∎

Czasami mówimy, że implikacja $P \Rightarrow Q$ jest **trywialna**, jeśli zdanie Q jest prawdziwe. W takim przypadku wartość logiczna zdania P jest nieistotna. **Trywialnym dowodem** implikacji $P \Rightarrow Q$ jest dowód, w którym wykazujemy prawdziwość Q bez odwoływania się do założenia P.

PRZYKŁAD 8 Jeśli x i y są takimi liczbami rzeczywistymi, że $xy = 0$, to $(x + y)^n = x^n + y^n$ dla $n \geq 1$. To stwierdzenie jest trywialne dla $n = 1$; równość $(x + y)^1 = x^1 + y^1$ jest oczywiście prawdziwa i założenie $xy = 0$ nie było potrzebne. Dla $n \geq 2$ to założenie jest potrzebne. ∎

Czasami mówi się o **konstruktywnych** i **niekonstruktywnych dowodach** istnienia obiektów matematycznych, mających określone własności. Dowód konstruktywny albo wskazuje ten obiekt (na przykład liczbę, macierz itp.), albo pokazuje, w jaki sposób ten obiekt może być znaleziony za pomocą określonej procedury czy algorytmu. Dowód niekonstruktywny stwierdza tylko istnienie obiektów za pomocą pewnych metod nie wprost, takich jak dowód przez sprowadzenie do sprzeczności, nie dając jednocześnie wskazówek, w jaki sposób można te obiekty znaleźć.

PRZYKŁAD 9 W przykładzie 3 udowodniliśmy przez sprowadzenie do sprzeczności, że istnieje nieskończenie wiele liczb pierwszych. Nie skonstruowaliśmy nieskończonego ciągu liczb pierwszych. Nasz dowód może jednak być tak przerobiony, by opisywał konstruktywną metodę tworzenia dowolnie długich ciągów różnych liczb pierwszych, przy założeniu, że znamy jakąś metodę rozkładania liczb całkowitych na czynniki. (Jest to ćwiczenie 14). ∎

PRZYKŁAD 10 Jeśli wszystkie wyrazy nieskończonego ciągu $a_1, a_2, a_3, \ldots$ należą do zbioru skończonego S, to pewne dwa wyrazy tego ciągu muszą być równe. W rzeczywistości, musi istnieć element s zbioru S taki, że $a_i = s$ dla nieskończenie wielu indeksów i. Możemy tylko powiedzieć, że taki element s istnieje; nie możemy natomiast powiedzieć, który to element, ani które wyrazy a_i są jemu równe, dopóki nie wiemy czegoś więcej o tym ciągu. Jednak nawet ta informacja egzystencjalna (tzn. o istnieniu) może być użyteczna.

Na przykład, twierdzimy, że w ciągu $1, 2, 2^2, 2^3, 2^4, \ldots$ istnieją dwa wyrazy różniące się o wielokrotność liczby pierwszej 7. To bardzo proste: łatwo zauważyć, że $2^3 - 1 = 7$ oraz również $2^4 - 2 = 16 - 2 = 2 \cdot 7$, $2^5 - 2^2 = 4 \cdot 7$ i $2^6 - 1 = 9 \cdot 7$. A jak będzie dla dużych liczb pierwszych p, na przykład $p = 8191$? Czy jakieś dwa wyrazy tego ciągu muszą różnić się o wielokrotność liczby p? Okazuje się, że tak. Wyobraźmy sobie, że dzielimy każdą potęgę 2^k przez p i wkładamy do odpowiedniego pudełka, w zależności od tego, jaka jest otrzymana reszta. Potęgi, które dają resztę 1 zostaną włożone do pudełka o numerze 1, potęgi, które dają resztę 2 pójdą do pudełka o numerze 2 itd. Mamy tylko p pudełek, odpowiadających możliwym resztom $0, 1, 2, \ldots, p-1$. Co najmniej jedno pudełko zawiera dwie liczby. (To właśnie jest naszym stwierdzeniem egzystencjalnym). Przypuśćmy, że liczby 2^k i 2^l znajdują się w m-tym pudełku, czyli $2^k = s \cdot p + m$ oraz $2^l = t \cdot p + m$ dla pewnych liczb całkowitych s i t. Wtedy liczba $2^k - 2^l = s \cdot p + m - t \cdot p - m = (s - t) \cdot p$ jest wielokrotnością p.

To niekonstruktywne rozumowanie nie mówi nam, jak znaleźć k i l, ale przekonuje nas, że muszą one istnieć. Mając już informację o ich istnieniu, możemy powiedzieć coś więcej. Jeśli na przykład $2^l < 2^k$, to $2^k - 2^l = 2^l \cdot (2^{k-l} - 1)$. Ponieważ p jest nieparzystą liczbą pierwszą i jest dzielnikiem tej różnicy, więc $2^{k-l} - 1$ musi być wielokrotnością p, a zatem muszą istnieć co najmniej dwie liczby w pudełku z numerem 1, mianowicie 1 i 2^{k-l}. ∎

PRZYKŁAD 11 Każda dodatnia liczba całkowita n ma postać $2^k m$, gdzie $k \in \mathbb{N}$ i m jest liczbą nieparzystą. Można tego dowieść na kilka sposobów, sugerujących następujące postępowanie konstruktywne. Jeśli liczba n jest nieparzysta, to niech $k = 0$ i $m = n$. W przeciwnym przypadku dzielimy n przez 2 i stosujemy to postępowanie do liczby $n/2$. Kontynuujemy to aż do otrzymania liczby nieparzystej. Wtedy liczba k będzie równa liczbie potrzebnych dzieleń przez 2. W ćwiczeniu 15 poprosimy o sprawdzenie tego rozumowania. ∎

Rozpoczęliśmy ten rozdział od omówienia bardzo ograniczonego systemu logicznego, mianowicie rachunku zdań. W tym paragrafie zrezygnowaliśmy z formalizmu na rzecz omówienia kilku metod dowodzenia spotykanych w tej książce i poza nią. Mamy nadzieję, że macie teraz lepsze pojęcie o tym, czym jest dowód matematyczny. W paragrafach 2.4 i 2.5 wrócimy do rachunku zdań i jego zastosowań. W rozdziale 13 zajmiemy się bardziej wymyślnymi aspektami logiki.

Poza logiką jako dziedziną zainteresowania i w szczególności w tej książce dowody służą przekonaniu Czytelnika o prawdziwości wypowiadanych stwierdzeń. Logika stoi u podstaw tego procesu, ale będzie pozostawać w tle. To znaczy, nie zawsze będzie konieczne świadome odwoływanie się do metod logiki przedstawionych w tym rozdziale, ale zawsze jeśli jakiś dowód przedstawiony w tej książce lub gdziekolwiek indziej, będzie intrygujący, możecie przeanalizować go dokładniej. Jakie są dokładnie założenia? Czy występują ukryte założenia? Czy autor stosuje metodę dowodzenia nie wprost?

Wreszcie możliwe jest, że autor popełnił błąd lub napisał coś innego niż zamierzał. Możecie wykazać, że twierdzenie jest fałszywe lub przynajmniej pokazać, że rozumowanie jest błędne. Nawet bardzo dobrzy matematycy popełniają błąd polegający na tym, że próbują dowieść równoważności $P \Leftrightarrow Q$ wykazując, że $P \Rightarrow Q$ i $\neg Q \Rightarrow \neg P$, choć ten błąd jest na ogół dobrze ukryty. Powrócimy do tych zagadnień po następnym paragrafie.

ĆWICZENIA DO § 2.3

We wszystkich ćwiczeniach, w których prosimy o dowód, wskaż, jaka metoda dowodu została użyta.

1. Udowodnij, że iloczyn dwóch liczb parzystych jest wielokrotnością 4.

2. Udowodnij, że iloczyn liczby parzystej i nieparzystej jest liczbą parzystą.

3. Udowodnij, że $|xy| = |x| \cdot |y|$ dla $x, y \in \mathbb{R}$.

4. Udowodnij, że liczba $n^4 - n^2$ jest podzielna przez 3 dla wszystkich $n \in \mathbb{N}$.

5. Udowodnij, że liczba $n^2 - 2$ nigdy nie jest podzielna przez 3 dla $n \in \mathbb{N}$.

6. (a) Udowodnij, że liczba $\sqrt{3}$ jest niewymierna.
(b) Udowodnij, że liczba $\sqrt[3]{2}$ jest niewymierna.

7. Udowodnij każde z następujących stwierdzeń lub wykaż, że jest ono fałszywe:
(a) Suma dwóch liczb parzystych jest liczbą parzystą.

(b) Suma dwóch liczb nieparzystych jest liczbą nieparzystą.
(c) Suma dwóch liczb pierwszych nigdy nie jest liczbą pierwszą.
(d) Suma trzech kolejnych liczb całkowitych jest podzielna przez 3.
(e) Suma czterech kolejnych liczb całkowitych jest podzielna przez 4.
(f) Suma pięciu kolejnych liczb całkowitych jest podzielna przez 5.

8. (a) Nie wiadomo, czy istnieje nieskończenie wiele par liczb pierwszych bliźniaczych, tzn. nieparzystych liczb pierwszych, których różnica wynosi 2. Przykładami liczb pierwszych bliźniaczych są: $(3,5)$, $(5,7)$, $(11,13)$ i $(71,73)$. Podaj trzy inne przykłady par liczb bliźniaczych.
(b) Udowodnij, że trójka $(3,5,7)$ jest jedynym przykładem „trojaczków". *Wskazówka*: mając dane liczby $2k+1$, $2k+3$, $2k+5$, gdzie $k \in \mathbb{N}$, pokaż, że jedna z nich musi być podzielna przez 3.

9. Udowodnij następujące stwierdzenia dotyczące liczby rzeczywistej x i $n = 1$:
(a) Jeśli $x \geq 0$, to $(1+x)^n \geq 1+nx$.
(b) Jeśli $x^n = 0$, to $x = 0$.
(c) Jeśli n jest liczbą parzystą, to $x^n \geq 0$.

10. Udowodnij twierdzenie z przykładu 8. Wykorzystaj fakt, że jeśli $xy = 0$, to $x = 0$ lub $y = 0$.

11. Udowodnij, że istnieją dwie różne liczby pierwsze p i q, które mają takich samych sześć ostatnich cyfr rozwinięcia dziesiętnego. Czy Twój dowód jest konstruktywny? Jeśli tak, znajdź takie liczby pierwsze.

12. Pokaż, że ciąg $a_1, a_2, \ldots$ określony wzorem $a_n = 40^n - n!$ ma wyraz największy. Czy Twój dowód jest konstruktywny? *Wskazówka*: przykład 3(b) z § 1.6 pokazuje, że jeśli $n \geq 80 \cdot \log_2 80$, to $n! > 40^n$.

13. (a) Udowodnij, że dla danego $n \in \mathbb{N}$ istnieje n kolejnych liczb całkowitych, które nie są pierwsze; tzn. zbiór liczb pierwszych ma dowolnie duże luki. *Wskazówka*: zacznij od $(n+1)! + 2$.
(b) Czy dowód jest konstruktywny? Jeśli tak, użyj go do znalezienia sześciu kolejnych liczb, które nie są pierwsze.
(c) Podaj siedem kolejnych liczb, które nie są pierwsze.

14. Przyjmijmy, że $p_1, p_2, \ldots, p_k$ jest daną listą różnych liczb pierwszych. Wyjaśnij, jak można użyć algorytmu rozkładania liczb na czynniki pierwsze do znalezienia liczby, której nie ma na tej liście. *Sugestia*: rozłóż $1 + p_1 p_2 \cdot \ldots \cdot p_k$.

15. Wykorzystaj procedurę podaną w przykładzie 11 do zapisania następujących liczb całkowitych dodatnich w postaci $2^k m$, gdzie $k \in \mathbb{N}$ i m jest liczbą nieparzystą:
(a) 14, (b) 73, (c) 96, (d) 1168,

16. (a) Rozumowanie w przykładzie 10 stosuje się do $p = 5$. Znajdź dwie potęgi liczby 2, które różnią się o wielokrotność liczby 5.
(b) Zrób to samo dla $p = 11$.

§ 2.4. Rachunek zdań — ciąg dalszy

W tym paragrafie zastanowimy się najpierw nad tym, w jaki sposób uzyskać nowe równoważności logiczne i nowe implikacje ze starych, bez użycia matryc logicznych. Następnie sformalizujemy pojęcia dowodu i poprawnego rozumowania.

Zaczniemy od dwóch użytecznych reguł podstawiania; są one naturalne, ale wymagają ostrożnego użycia.

Reguła podstawiania (a)

Jeśli zdanie złożone P jest tautologią i wszystkie wystąpienia pewnej zmiennej, na przykład p, występującej w zdaniu P zastąpimy tym samym zdaniem E, to otrzymane zdanie złożone P^* będzie też tautologią.

PRZYKŁAD 1

(a) Zgodnie z prawem 19 w tablicy 2.2 w § 2.2 (modus ponendo ponens, w skrócie modus ponens) zdanie

$$P = \text{„}[p \wedge (p \to q)] \to q\text{”}$$

jest tautologią. Zatem wszystkie wartości logiczne zdania P w matrycy logicznej dla P są równe 1, niezależnie od wartości p i q (por. matrycę logiczną w przykładzie 3(b) w § 2.2). Przypuśćmy, że zastąpimy każde wystąpienie zmiennej p zdaniem $E = \text{„}q \to r\text{”}$, by otrzymać nowe zdanie

$$P^* = \text{„}[(q \to r) \wedge ((q \to r) \to q)] \to q\text{”}.$$

Zgodnie z regułą podstawiania (a) zdanie P^* jest tautologią. Zobaczmy, dlaczego. Matrycą logiczną zdania P^* jest:

q	r	$[(q \to r) \wedge ((q \to r) \to q)]$				$\to$	q
0	0	1	0	1	0	1	
0	1	1	0	1	0	1	
1	0	0	0	0	1	1	
1	1	1	1	1	1	1	

W pierwszych dwóch wierszach zdanie E ma wartość logiczną 1 (cyfry wyróżnione), a q ma wartość logiczną 0. Ponieważ zastąpiliśmy każde wystąpienie zmiennej p w zdaniu P zdaniem E, pozostałe wartości logiczne w tych wierszach będą takie same, jak w wierszu odpowiadającym wartościom $p = 1$, $q = 0$ matrycy logicznej wyjściowego zdania P. Ponieważ zdanie P jest tautologią, więc wartości logiczne dla P, jak również dla P^* muszą być równe 1. W trzecim wierszu zdanie E ma wartość logiczną 0, a q ma wartość logiczną 1, a więc pozostałe wartości logiczne

będą takie same jak w wierszu dla $p = 0$ i $q = 1$ matrycy logicznej zdania P. Znów końcowa wartość logiczna będzie równa 1. Podobnie, czwarty wiersz odpowiada wierszowi $p = 1$, $q = 1$ matrycy logicznej zdania P, a więc końcowa wartość logiczna będzie równa 1.

Zauważmy, że gdy tylko określiliśmy wartość logiczną zdania E, wiedzieliśmy, na który wiersz matrycy logicznej zdania P należy patrzeć. Ale to nie miało znaczenia. Wszystkie wiersze dla tautologii P i tak dawały końcową wartość logiczną 1, a zatem zdanie P^* również miało wartość logiczną 1. Tak więc zdanie P^* też było tautologią. To rozumowanie może być uogólnione tak, by można było pokazać, że reguła podstawiania (a) jest zawsze słuszna.

(b) Jeżeli natomiast zastąpimy każde wystąpienie q zdaniem $E =$„$q \to r$”, to otrzymamy tautologię

$$[p \wedge (p \to (q \to r))] \to (q \to r).$$

Zauważmy, że matryca logiczna tej tautologii będzie miała 8 wierszy. ∎

Reguła podstawiania (a) może być stosowana w celu utworzenia nowych równoważności logicznych ze starych. Przypuśćmy, że znana jest równoważność $A \Leftrightarrow B$, na przykład z tablicy 2.1 w § 2.2. Wtedy zdanie $A \leftrightarrow B$ jest tautologią. Jeśli każde wystąpienie zmiennej p, zarówno w A, jak i w B, zastąpimy zdaniem E i otrzymamy odpowiednio A^* i B^*, to reguła podstawiania (a) mówi, że zdanie $A^* \leftrightarrow B^*$ jest również tautologią. Stąd równoważność $A^* \Leftrightarrow B^*$ jest prawdziwa. To znaczy, jeśli A i B są logicznie równoważne, to również A^* i B^* są logicznie równoważne. Podobnie, jeśli implikacja $A \Rightarrow B$ jest prawdziwa, to również prawdziwa jest implikacja $A^* \Rightarrow B^*$.

PRZYKŁAD 2 (a) Oto kilka prostych ilustracji. Prawo łączności 3a z tablicy 2.1 w § 2.2 ma postać

$$[(p \vee q) \vee r] \Leftrightarrow [p \vee (q \vee r)].$$

Możemy zastąpić wszędzie zmienną p zdaniem $p \wedge q$, aby otrzymać

$$[((p \wedge q) \vee q) \vee r] \Leftrightarrow [(p \wedge q) \vee (q \vee r)],$$

a następnie zastąpić wszystkie q zmienną p, aby otrzymać

$$[((p \wedge p) \vee p) \vee r] \Leftrightarrow [(p \wedge p) \vee (p \vee r)].$$

Dokonując podstawień w odwrotnej kolejności otrzymalibyśmy najpierw

$$[(p \vee p) \vee r] \Leftrightarrow [p \vee (p \vee r)]$$

i ostatecznie

$$[((p \wedge q) \vee (p \wedge q)) \vee r] \Leftrightarrow [(p \wedge q) \vee ((p \wedge q) \vee r)].$$

Możemy też jednocześnie zastąpić każde p zdaniem $p \wedge q$, a każde q zmienną p, aby otrzymać równoważność

$$[((p \wedge q) \vee p) \vee r] \Leftrightarrow [(p \wedge q) \vee (p \vee r)].$$

Ta równoważność różni się od każdej z równoważności otrzymanych w poprzednim akapicie. Jednoczesne podstawianie może wydawać się niedopuszczalne, ponieważ nie mówi się o nim wyraźnie w regule podstawiania (a), ale może być ono uzasadnione tym, że można je zastąpić ciągiem pojedynczych podstawień, używając tymczasowych nazw. W naszym przykładzie możemy zastąpić każde q tymczasową literą s, następnie każde p zdaniem $p \wedge q$ i wreszcie każde s zmienną p. Zatem

$$[(p \vee q) \vee r] \Leftrightarrow [p \vee (q \vee r)]$$

stanie się

$$[(p \vee s) \vee r] \Leftrightarrow [p \vee (s \vee r)],$$

które stanie się

$$[((p \wedge q) \vee s) \vee r] \Leftrightarrow [(p \wedge q) \vee (s \vee r)]$$

i wreszcie

$$[((p \wedge q) \vee p) \vee r] \Leftrightarrow [(p \wedge q) \vee (p \vee r)].$$

(b) Prawo definiowania implikacji 10a z tablicy 2.1 w § 2.2 ma postać

$$(p \to q) \Leftrightarrow (\neg p \vee q)$$

i odpowiada tautologii $(p \to q) \leftrightarrow (\neg p \vee q)$. Zastępując każde wystąpienie p zdaniem $\neg p$ i każde wystąpienie q zdaniem $p \to q$ w tej tautologii, otrzymamy zdanie

$$[\neg p \to (p \to q)] \leftrightarrow [\neg\neg p \vee (p \to q)],$$

które zgodnie z regułą podstawiania (a) jest także tautologią. Stąd zdanie $\neg p \to (p \to q)$ jest logicznie równoważne zdaniu $\neg\neg p \vee (p \to q)$.

(c) Prawo 18 z tablicy 2.2, sprowadzenia do sprzeczności, ma postać $(p \to c) \Rightarrow \neg p$. Odpowiada ono tautologii $(p \to c) \to \neg p$. Zastępując p zdaniem $q \vee r$, zgodnie z regułą podstawiania (a) otrzymamy tautologię $((q \vee r) \to c) \to \neg(q \vee r)$, która odpowiada implikacji logicznej $((q \vee r) \to c) \Rightarrow \neg(q \vee r)$.

Zastępując c zmienną p w implikacji $(p \to c) \to \neg p$ otrzymamy zdanie $(p \to p) \to \neg p$, które nie jest tautologią. Problem polega na tym, że c nie jest zmienną, a więc reguła podstawiania (a) nie ma tu zastosowania. ∎

Nasza druga reguła podstawiania jest podobna do reguły algebraicznej mówiącej, że możemy zawsze zastąpić wielkości innymi wielkościami równymi im. W tym przypadku wystarczy, by nasze zdanie, które podstawiamy, było tylko równoważne ze zdaniem, które mamy nim zastąpić, a wynik będzie równoważny z tym, od czego zaczęliśmy.

Reguła podstawiania (b)

Jeśli zdanie złożone P zawiera zdanie Q i jeśli zdanie Q zastąpimy zdaniem logicznym z nim równoważnym Q^*, to otrzymane zdanie złożone P^* jest logicznie równoważne ze zdaniem P.

PRZYKŁAD 3

(a) Weźmy zdanie

$$P = \text{„}\neg[(p \to q) \wedge (p \to r)] \to [q \to (p \to r)]\text{”},$$

które nie jest tautologią. Na podstawie reguły podstawiania (b) otrzymamy zdanie logicznie równoważne P^*, jeśli zastąpimy zdanie $Q = (p \to q)$ zdaniem logicznie z nim równoważnym $Q^* = (\neg p \vee q)$. Podobnie, moglibyśmy zastąpić jedno lub oba wystąpienia zdania $(p \to r)$ zdaniem $(\neg p \vee r)$; por. prawo 10a. Moglibyśmy również zastąpić zdanie $[(p \to q) \wedge (p \to r)]$ zdaniem $[p \to (q \wedge r)]$, zgodnie z prawem 12b. Zatem zdanie P jest logicznie równoważne z każdym z następujących zdań, wśród wielu innych:

$$\neg[(\neg p \vee q) \wedge (p \to r)] \to [q \to (p \to r)],$$
$$\neg[(p \to q) \wedge (\neg p \vee r)] \to [q \to (p \to r)],$$
$$\neg[p \to (q \wedge r)] \to [q \to (\neg p \vee r)].$$

(b) Zobaczmy, dlaczego zdanie $P = \text{„}\neg[(p \to q) \wedge (p \to r)] \to [q \to (p \to r)]\text{”}$ jest równoważne ze zdaniem $P^* = \text{„}\neg[(p \to q) \wedge (\neg p \vee r)] \to [q \to (p \to r)]\text{”}$. Wybieramy do analizy to podstawienie, aby pokazać, że drugie wystąpienie zdania $(p \to r)$, które pozostało nie zmienione, nie ma znaczenia. Wyobraźmy sobie kolumny matrycy logicznej zdań P i P^* odpowiadające wyróżnionym spójnikom $\to$ i $\vee$. Ponieważ zdania $p \to r$ i $\neg p \vee r$ są równoważne, wartości logiczne w obu tych kolumnach są takie same. Pozostałe wartości logiczne w matrycach zdań P i P^* są

identyczne, więc ostateczne wartości logiczne obu zdań są takie same.

Taki sposób rozumowania uogólnia się, aby pokazać, dlaczego reguła podstawiania (b) jest słuszna. ∎

Warto podkreślić, że w przeciwieństwie do reguły podstawiania (b), reguła (a) wymaga, by wszystkie wystąpienia jakiejś zmiennej zostały zastąpione tym samym zdaniem.

PRZYKŁAD 4

(a) Możemy używać reguł (a) i (b) do budowania ciągów zdań równoważnych. Na przykład dowodzimy równoważności $(p \vee q) \vee (p \vee r) \Leftrightarrow (p \vee q) \vee r$, wypisując ciąg zdań równoważnych, zaczynając od zdania $(p \vee q) \vee (p \vee r)$ i kończąc na zdaniu $(p \vee q) \vee r$. Ćwiczenie 23 z § 2.2 pokazuje, że wszystkie zdania w takim ciągu muszą być sobie równoważne.

Zdania równoważne	**Wyjaśnienia**
$(p \vee q) \vee (p \vee r)$	dane zdanie
$[(p \vee q) \vee p] \vee r$	prawo 3a dla $p \vee q$ zamiast p i p zamiast q, korzystając z reguły podstawiania (a); por. przykład 2(a)
$[p \vee (q \vee p)] \vee r$	prawo 3a dla p zamiast r, korzystając z reguł podstawiania (a) i (b)
$[p \vee (p \vee q)] \vee r$	prawo 2a i reguła podstawiania (b)
$[(p \vee p) \vee q] \vee r$	prawo 3a i reguła podstawiania (a) dla p zamiast q i q zamiast r oraz reguła podstawiania (b)
$[p \vee q] \vee r$	prawo 5a i reguła podstawiania (b)

Starannie tu zaznaczaliśmy każde zastosowanie reguł podstawiania (a) i (b), chociaż zazwyczaj wyjaśnia się zastosowanie reguły podstawiania w danym kroku tylko wtedy, gdy Czytelnik mógłby tego zastosowania samodzielnie nie zauważyć.

(b) Udowodnimy tautologię

$$[(p \to q) \vee (p \to r)] \to [p \to (q \vee r)].$$

Na podstawie prawa łączności 3a otrzymujemy tautologię

$$[(p \vee q) \vee r] \to [p \vee (q \vee r)].$$

Przykład (a) pokazywał, że zdanie $(p \vee q) \vee (p \vee r)$ jest logicznie równoważne ze zdaniem $(p \vee q) \vee r$, zatem reguła podstawiania (b) mówi, że zdanie

$$[(p \vee q) \vee (p \vee r)] \to [p \vee (q \vee r)]$$

jest tautologią. Zastępując każde wystąpienie zdania p zdaniem $\neg p$ i stosując regułę podstawiania (a), otrzymujemy tautologię

$$(*) \qquad [(\neg p \vee q) \vee (\neg p \vee r)] \rightarrow [\neg p \vee (q \vee r)].$$

Prawo 10a mówi, że $\neg p \vee q \Leftrightarrow p \rightarrow q$; reguła podstawiania (a) daje więc także

$$\neg p \vee r \Leftrightarrow p \rightarrow r \quad \text{oraz} \quad \neg p \vee (q \vee r) \Leftrightarrow p \rightarrow (q \vee r).$$

Trzykrotne zastosowanie reguły podstawiania (b) do zdania $(*)$ pokazuje, że zdanie

$$[(p \rightarrow q) \vee (p \rightarrow r)] \rightarrow [p \rightarrow (q \vee r)]$$

jest tautologią. ■

Można stosować reguły podstawiania, gdy mamy do czynienia z niezwykle skomplikowanymi zdaniami. Można, ale my nie będziemy tego robić. Chcemy tutaj tylko pokazać, że istnieją metody podobne do metod znanych nam z algebry, które pozwalają nam operować wyrażeniami logicznymi i zapisywać je w wygodniejszej postaci. Określenie „wygodniejsza" zależy oczywiście od tego, do czego zamierzamy ich użyć.

PRZYKŁAD 5 Użyjemy prawa De Morgana 8d i reguły podstawiania do znalezienia zdania logicznie równoważnego ze zdaniem $(p \wedge q) \rightarrow (\neg p \wedge q)$, w którym nie występuje spójnik $\wedge$. Ponieważ zdanie $p \wedge q$ jest logicznie równoważne ze zdaniem $\neg(\neg p \vee \neg q)$, a zdanie $\neg p \wedge q$ jest równoważne ze zdaniem $\neg(\neg\neg p \vee \neg q)$, więc z reguły podstawiania (b) wynika, że nasze zdanie jest równoważne z

$$\neg(\neg p \vee \neg q) \rightarrow \neg(\neg\neg p \vee \neg q),$$

a więc znów na podstawie reguły podstawiania (b) jest równoważne ze zdaniem

$$\neg(\neg p \vee \neg q) \rightarrow \neg(p \vee \neg q).$$

Jeśli chcemy, możemy zastosować prawo 10a, aby otrzymać zdanie równoważne

$$\neg(p \rightarrow \neg q) \rightarrow \neg(q \rightarrow p),$$

w którym nie ma ani $\wedge$ ani $\vee$. Z drugiej strony, możemy uniknąć używania spójnika $\rightarrow$ stosując prawo 10a.

Taki sposób zapisywania w celu wyeliminowania jednego lub więcej spójników ma ważne zastosowania w projektowaniu układów logicznych, o czym przekonamy się w rozdziale 10.

Zauważmy, że nie twierdzimy tutaj, że zdania $(p \wedge q) \rightarrow (\neg p \wedge q)$, $\neg(\neg p \vee \neg q) \rightarrow \neg(\neg\neg p \vee \neg q)$ i $\neg(\neg p \vee \neg q) \rightarrow \neg(p \vee \neg q)$

są tautologiami. My tylko pokazaliśmy, że te trzy zdania są logicznie równoważne. ∎

Tautologie dają jeden ze sposobów patrzenia na zależności logiczne między dwoma różnymi zdaniami — możemy sprawdzić wszystkie możliwe wartości zmiennych zdaniowych i porównać wartości logiczne tych zdań. Inne podejście polega na tym, aby zobaczyć, czy istnieje logiczna metoda wyprowadzenia jednego zdania z drugiego, być może za pomocą małych kroków, co do których każdy się zgadza, że są poprawne. Naszym następnym celem będzie formalizacja tego pomysłu dedukcji. Następnie omówimy związki między prawdą i dedukcją jako dwoma sposobami opisu rozumowania logicznego.

Idea podejścia dedukcyjnego polega na sformalizowaniu pojęcia „dowodu". Przypuśćmy, że mamy dany pewien zbiór zdań, naszych **założeń**, oraz pewien **wniosek** (konkluzję) C. **Dowodem formalnym zdania** C z tych założeń jest ciąg $P_1, P_2, \ldots, P_n, C$ zdań, kończący się zdaniem C, w którym każde zdanie P_i jest albo

(i) założeniem, albo
(ii) tautologią, albo
(iii) wnioskiem z poprzednich wyrazów ciągu, przy użyciu dopuszczalnej reguły wnioskowania.

Twierdzeniem jest zdanie postaci „jeśli H, to C", gdzie H jest zbiorem założeń, a C jest wnioskiem (tezą). Dowód formalny zdania C z założeń H nazywamy dowodem formalnym tego twierdzenia.

Regułami wnioskowania, które dopuścimy w dowodach formalnych, są reguły podstawiania (a) i (b) oraz reguły oparte na implikacjach logicznych postaci $H_1 \wedge H_2 \wedge \ldots \wedge H_m \Rightarrow Q$. Jeśli zdania $H_1, H_2, \ldots, H_m$ pojawiły się już w ciągu, który ma być dowodem formalnym oraz jeśli implikacja $H_1 \wedge H_2 \wedge \ldots \wedge H_m \Rightarrow Q$ jest prawdziwa, to możemy dodać zdanie Q do tego ciągu.

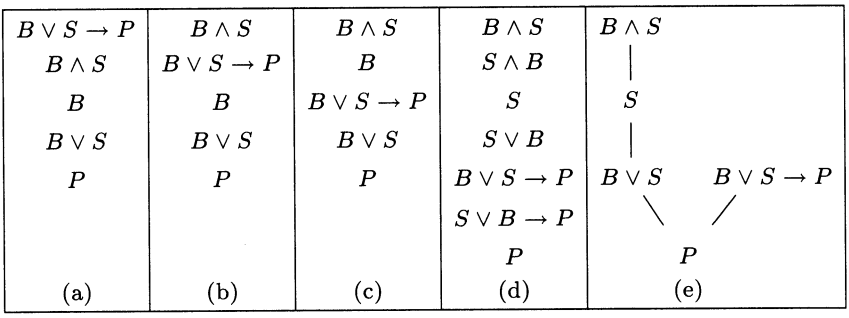
Jednym ze sposobów zapisania ciągu zdań, będącego dowodem formalnym, jest napisanie ich po kolei, jak zwykły tekst prozą. Innym sposobem jest wypisanie ich jednego pod drugim, co daje nam miejsce na wpisanie powodów, dla których każde z tych zdań znalazło się w tym ciągu.

PRZYKŁAD 6 Oto bardzo krótki dowód formalny, zapisany „pionowo". Założeniami są zdania $B \wedge S$ oraz $B \vee S \rightarrow P$, a tezą zdanie P. Będzie to dowód formalny zdania P ze zbioru zdań $\{B \wedge S, B \vee S \rightarrow P\}$. Aby nadać tej symbolice pewne znaczenie, możemy myśleć o zdaniu

B jako o zdaniu „noszę pasek", o zdaniu S jako o zdaniu „noszę szelki", a o zdaniu P jako o zdaniu „moje spodnie nie opadają".

Dowód	**Wyjaśnienia**
1. $B \wedge S$	założenie
2. B	1, prawo opuszczania koniunkcji 17
3. $B \vee S$	2, prawo wprowadzania alternatywy 16
4. $B \vee S \rightarrow P$	założenie
5. P	3, 4, prawo modus ponendo ponens 19

Ten dowód mógłby być również zapisany w innym porządku. Rysunek 2.1a–d pokazuje cztery inne dowody formalne z tymi samymi założeniami i tą samą tezą. Rysunek 2.1e przedstawia zależności logiczne między tymi zdaniami.



Rysunek 2.1

■

W dowodzie w przykładzie 6 korzystaliśmy z trzech reguł wnioskowania, opartych na implikacjach $(p \wedge q) \Rightarrow p$, $p \Rightarrow (p \vee q)$ oraz $[p \wedge (p \rightarrow q)] \Rightarrow q$. W tablicy 2.3 wypisane są te reguły oraz często używane cztery inne reguły. Każda z nich odpowiada implikacji logicznej postaci $H_1 \wedge \ldots \wedge H_m \Rightarrow C$ z tablicy 2.2 w § 2.2, poza regułą 34, która odpowiada implikacji $p \wedge q \Rightarrow p \wedge q$. Stosujemy następujący zapis

$$\begin{array}{l} H_1 \\ H_2 \\ \ldots \\ H_m \\ \hline \therefore\ C \end{array}$$

(symbol $\therefore$ czytamy jako „stąd" lub „zatem"),

aby zaznaczyć, że C jest wnioskiem ze zdania $H_1 \wedge \ldots \wedge H_m$. Kiedy wypisujemy powody, dla których włączamy zdanie do dowodu

formalnego, możemy powołać się albo na prawa z § 2.2, albo na ich odpowiedniki w tablicy 2.3.

Tablica 2.3. Reguły wnioskowania

28.	P	
	$\therefore \overline{P \vee Q}$	reguła wprowadzania alternatywy
29.	$P \wedge Q$	
	$\therefore \overline{P}$	reguła opuszczania koniunkcji
30.	P	
	$P \to Q$	
	$\therefore \overline{Q}$	reguła modus ponendo ponens (w skrócie modus ponens)
31.	$P \to Q$	
	$\neg Q$	
	$\therefore \overline{\neg P}$	reguła modus tollendo tollens (w skrócie modus tollens)
32.	$P \vee Q$	
	$\neg P$	
	$\therefore \overline{Q}$	reguła modus ponendo tollens
33.	$P \to Q$	
	$Q \to R$	
	$\therefore \overline{P \to R}$	reguła sylogizmu hipotetycznego
34.	P	
	Q	
	$\therefore \overline{P \wedge Q}$	reguła wprowadzania koniunkcji

Zauważmy, że równoważności logiczne postaci $H_1 \wedge \ldots \wedge H_m \Leftrightarrow C$, takie jak równoważność $[(p \to r) \wedge (q \to r)] \Leftrightarrow [(p \vee q) \to r]$ z tablicy 2.1 w § 2.2 dają implikacje logiczne $H_1 \wedge \ldots \wedge H_m \Rightarrow C$. Zatem dają one również reguły wnioskowania.

Oto dwa dodatkowe dowody pokazujące, w jaki sposób uzasadniamy kolejne kroki w dowodzie formalnym.

PRZYKŁAD 7 (a) Wyprowadzamy implikację $s \to r$ ze zdań $p \to (q \to r)$, $p \vee \neg s$ i q.

1. $p \to (q \to r)$	założenie
2. $p \vee \neg s$	założenie

3. q	założenie
4. $\neg s \vee p$	3; prawo przemienności 2a
5. $s \to p$	4; określenie implikacji 10a
6. $s \to (q \to r)$	1, 5; reguła sylogizmu hipotetycznego 33
7. $(s \wedge q) \to r$	6; prawo eksportacji 14
8. $q \to [s \to (q \wedge s)]$	prawo 22
9. $s \to (q \wedge s)$	3, 8; reguła modus ponens 30
10. $s \to (s \wedge q)$	9; prawo przemienności 2b
11. $s \to r$	7, 10; reguła sylogizmu hipotetycznego 33

(b) Tutaj wyprowadzamy zdanie $\neg p$ przez sprowadzenie do sprzeczności ze zdań $p \to (q \wedge r)$, $r \to s$ i $\neg(q \wedge s)$.

1. $p \to (q \wedge r)$	założenie
2. $r \to s$	założenie
3. $\neg(q \wedge s)$	założenie
4. $\neg(\neg p)$	zaprzeczenie tezy
5. p	4; prawo podwójnego przeczenia 1
6. $q \wedge r$	1, 5; reguła modus ponens 30
7. q	6; reguła opuszczania koniunkcji 29
8. $r \wedge q$	6; prawo przemienności 2b
9. r	8; reguła opuszczania koniunkcji 29
10. s	2, 9; reguła modus ponens 30
11. $q \wedge s$	7, 10; reguła wprowadzania koniunkcji 34
12. $(q \wedge s) \wedge \neg(q \wedge s)$	3, 11; reguła wprowadzania koniunkcji 34
13. sprzeczność	12; prawo 7b ■

Można się zastanawiać, dlaczego ktoś chciałby to robić w tak formalny sposób. Próbujemy zbudować model formalny, pokazujący, jak się konstruuje dowody logiczne i w jaki sposób można sprawdzić, czy ciąg wnioskowań jest poprawny. Nikt, nawet ktoś zajmujący się logiką formalną czy projektant systemów komputerowych do sprawdzania poprawności dowodów nie lubi sam pisać dowodów formalnych. Chodzi nam o to, aby dowiedzieć się, jakie są możliwe rodzaje dowodów i poznać sposoby stosowania reguł wnioskowania, ponieważ używamy ich w niejawny sposób nawet wtedy, gdy piszemy dowody nieformalnie.

Teraz, gdy znaleźliśmy sposób na opisywanie rozumowań logicznych, możemy zapytać, jaki ma on związek z prawdziwością zdań. Matryce logiczne są wbudowane w nasze reguły wnioskowania, gdyż każda implikacja logiczna $P \Rightarrow Q$ jest naprawdę

stwierdzeniem dotyczącym matryc logicznych. Długie i żmudne rozumowanie pokazuje, że jeśli istnieje dowód formalny zdania C z założeń H i jeśli założenia H są prawdziwe, to prawdziwe jest też zdanie C. Nie można więc wyprowadzić zdania fałszywego ze zdań prawdziwych. Można również pokazać, że jeśli H implikuje logicznie C, to istnieje dowód formalny C z H. Zatem istnieje dowód formalny C z H wtedy i tylko wtedy, gdy prawdziwa jest implikacja logiczna $H \Rightarrow C$.

PRZYKŁAD 8 Przypomnijmy z przykładu 6(b) w § 2.2, że implikacja $A \wedge B \Rightarrow C$ jest prawdziwa wtedy i tylko wtedy, gdy implikacja $A \Rightarrow (B \to C)$ jest prawdziwa. Zatem istnieje dowód formalny zdania C ze zdania $A \wedge B$, tzn. ze zdań A i B, wtedy i tylko wtedy, gdy istnieje dowód zdania $B \to C$ ze zdania A. ∎

ĆWICZENIA DO § 2.4

1. Wykorzystaj regułę podstawiania (a), zastępując zmienną q zdaniem $p \to q$, by otrzymać nowe tautologie z następujących:
(a) $\neg q \to (q \to p)$, (b) $[p \wedge (p \to q)] \to q$,
(c) $p \vee \neg p$, (d) $(p \vee q) \leftrightarrow ((\neg q) \to p)$.

2. Podaj reguły wnioskowania odpowiadające implikacjom logicznym 23, 26a i 27b.

3. Znajdź w tablicy 2.1 w § 2.2 równoważności logiczne, z których za pomocą reguły podstawiania (a) otrzymano następujące równoważności:
(a) $[(p \wedge (q \wedge r)) \to r] \Leftrightarrow [p \to ((q \wedge r) \to r)]$,
(b) $[p \vee (q \wedge (r \wedge s))] \Leftrightarrow [(p \vee q) \wedge (p \vee (r \wedge s))]$,
(c) $\neg[(\neg p \wedge r) \vee (q \to r)] \Leftrightarrow [\neg(\neg p \wedge r) \wedge \neg(q \to r)]$.

4. Znajdź w tablicy 2.2 w § 2.2 implikacje logiczne, z których za pomocą reguły podstawiania (a) otrzymano następujące implikacje:
(a) $[\neg p \vee q] \Rightarrow [q \to ((\neg p \vee q) \wedge q)]$,
(b) $[(p \to q) \wedge (r \to q)] \Rightarrow [(\neg q \vee \neg q) \to (\neg p \vee \neg r)]$,
(c) $[((p \to s) \to (q \wedge s)) \wedge \neg(q \wedge s)] \Rightarrow \neg(p \to s)$.

5. Uzasadnij każdą z następujących równoważności — spróbuj nie korzystać z matryc logicznych:
(a) $(p \vee q) \wedge s \Leftrightarrow (q \wedge p) \wedge s$,
(b) $s \to (\neg(p \vee q)) \Leftrightarrow s \to [(\neg p) \wedge (\neg q)]$,
(c) $(p \to q) \wedge (p \vee q) \Leftrightarrow (\neg p \vee q) \wedge (\neg\neg p \vee q)$,
(d) $t \wedge (s \vee p) \Leftrightarrow t \wedge (p \vee s)$.

6. Wykonaj ćwiczenie 5 dla następujących równoważności:
(a) $s \wedge p \Leftrightarrow s \wedge (p \wedge p)$,
(b) $[(a \vee b) \leftrightarrow (p \to q)] \Leftrightarrow [(a \vee b) \leftrightarrow (\neg p \vee q)]$,
(c) $[(a \vee b) \leftrightarrow \neg(p \wedge q)] \Leftrightarrow [(b \vee a) \leftrightarrow (\neg p \vee \neg q)]$.

7. Podaj uzasadnienie dla każdej równoważności w poniższym ciągu:

$$(p \to s) \vee (\neg s \to t)$$

(a) $\Leftrightarrow (\neg p \vee s) \vee (s \vee t)$,
(b) $\Leftrightarrow [(\neg p \vee s) \vee s] \vee t$,
(c) $\Leftrightarrow [\neg p \vee (s \vee s)] \vee t$,
(d) $\Leftrightarrow (\neg p \vee s) \vee t$,
(e) $\Leftrightarrow \neg p \vee (s \vee t)$,
(f) $\Leftrightarrow p \to (s \vee t)$.

8. Powtórz ćwiczenie 7 dla następującego ciągu równoważności:

$$[(a \wedge p) \vee p] \to p$$

(a) $\Leftrightarrow \neg[(a \wedge p) \vee p] \vee p$,
(b) $\Leftrightarrow [\neg(a \wedge p) \wedge \neg p] \vee p$,
(c) $\Leftrightarrow [(\neg a \vee \neg p) \wedge \neg p] \vee p$,
(d) $\Leftrightarrow p \vee [(\neg a \vee \neg p) \wedge \neg p]$,
(e) $\Leftrightarrow [p \vee (\neg a \vee \neg p)] \wedge (p \vee \neg p)$,
(f) $\Leftrightarrow [(\neg a \vee \neg p) \vee p] \wedge t$ (t jest tu dowolną tautologią),
(g) $\Leftrightarrow (\neg a \vee \neg p) \vee p$,
(h) $\Leftrightarrow \neg a \vee (\neg p \vee p)$,
(i) $\Leftrightarrow \neg a \vee t$,
(j) $\Leftrightarrow t$.

9. Uzasadnij następujące równoważności, korzystając z metody pokazanej w przykładzie 2; uzasadnij kolejne kroki rozumowania:
(a) $[(p \vee r) \wedge (q \to r)] \Leftrightarrow [(p \wedge \neg q) \vee r]$,
(b) $[(p \wedge \neg q) \vee r] \Leftrightarrow [(p \to q) \to r]$,
(c) $p \vee \neg q \Leftrightarrow p \vee (\neg p \wedge \neg q)$.

10. Powtórz ćwiczenie 9 dla następujących implikacji i równoważności:
(a) $\neg q \Rightarrow \neg q \vee p$,
(b) $(p \vee q) \to p \Leftrightarrow p \vee (\neg p \wedge \neg q)$,
(c) $p \vee (\neg p \wedge \neg q) \Leftrightarrow \neg q \vee p$,
(d) $\neg q \Rightarrow [(p \vee q) \to p]$.

11. Niech P będzie zdaniem $[p \wedge (q \vee r)] \vee \neg[p \vee (q \vee r)]$. Zamiana każdego wystąpienia zdania $q \vee r$ na zdanie $q \wedge r$ daje zdanie

$$P^* = \text{„}[p \wedge (q \wedge r)] \vee \neg[p \vee (q \wedge r)]\text{”}.$$

Ponieważ $q \wedge r \Rightarrow q \vee r$, więc można byłoby oczekiwać, że $P \Rightarrow P^*$ lub $P^* \Rightarrow P$. Pokaż, że nie zachodzi żadna z tych implikacji.

12. Pokaż, że jeśli pierwsze wystąpienie p w tautologii $p \to [q \to (p \wedge q)]$ zostanie zastąpione zdaniem $p \vee q$, to otrzymane zdanie nie jest tautologią. To ćwiczenie pokazuje, że należy zachować dużą ostrożność przy korzystaniu z reguły podstawiania (a).

13. Uzasadnij każdy krok w następującym dowodzie formalnym zdania $\neg s$ ze zbioru zdań $\{(s \vee g) \to p, \neg a, p \to a\}$.

1. $(s \vee g) \to p$
2. $\neg a$
3. $p \to a$

4. $s \to (s \vee g)$
5. $s \to p$
6. $s \to a$
7. $\neg s$

14. Zamień na trzy różne sposoby kolejność kroków w dowodzie w ćwiczeniu 13, tak aby w dalszym ciągu były to dowody zdania $\neg s$ ze zbioru $\{(s \vee g) \to p, \neg a, p \to a\}$.

15. (a) Pokaż, że jeśli zdanie A jest tautologią i istnieje dowód formalny zdania C ze zdania A, to istnieje dowód formalny zdania C bez jakichkolwiek założeń.

(b) Pokaż, że jeśli istnieje dowód formalny zdania C ze zdania B, to istnieje dowód formalny zdania $B \to C$ bez jakichkolwiek założeń.

16. Każde zdanie złożone jest równoważne ze zdaniem, w którym występują tylko spójniki $\neg$ i $\vee$. Wynika to z równoważności $(p \to q) \Leftrightarrow (\neg p \vee q)$, $(p \wedge q) \Leftrightarrow \neg(\neg p \vee \neg q)$ i $(p \leftrightarrow q) \Leftrightarrow [(p \to q) \wedge (q \to p)]$. Znajdź zdania logicznie równoważne z następującymi zdaniami, w których będą występować tylko spójniki $\neg$ i $\vee$:
(a) $p \leftrightarrow q$,
(b) $(p \wedge q) \to (\neg q \wedge r)$,
(c) $(p \to q) \wedge (q \vee r)$,
(d) $p \oplus q$.

17. (a) Pokaż, że zdania $p \vee q$ i $p \wedge q$ są logicznie równoważne ze zdaniami, w których występują tylko spójniki $\neg$ i $\to$.

(b) Pokaż, że zdania $p \vee q$ i $p \to q$ są logicznie równoważne ze zdaniami, w których występują tylko spójniki $\neg$ i $\wedge$.

(c) Czy zdanie $p \to q$ jest logicznie równoważne ze zdaniem, w którym występują tylko spójniki $\wedge$ i $\vee$? Odpowiedź uzasadnij.

18. **Kreska Sheffera** jest to spójnik $|$ zdefiniowany za pomocą następującej matrycy logicznej:

p	q	$p\|q$
0	0	1
0	1	1
1	0	1
1	1	0

Zatem $p|q \Leftrightarrow \neg(p \wedge q)$ (to uzasadnia informatyczną nazwę NAND dla tego spójnika). Wszystkie zdania złożone są równoważne ze zdaniami, w których występuje tylko spójnik $|$; jest to przydatny fakt wynikający z uwag poczynionych w ćwiczeniu 16 oraz z poniższych ćwiczeń (a) i (b).
(a) Pokaż, że $\neg p \Leftrightarrow p|p$.
(b) Pokaż, że $p \vee q \Leftrightarrow (p|p)|(q|q)$.
(c) Znajdź zdanie równoważne zdaniu $p \wedge q$, w którym występuje tylko kreska Sheffera.

(d) Zrób to samo ze zdaniem $p \to q$.
(e) Zrób to samo ze zdaniem $p \oplus q$.

19. Uzasadnij następujące **prawa pochłaniania**.
(a) $[p \vee (p \wedge q)] \Leftrightarrow p$ (por. ćwiczenie 8),
(b) $[p \wedge (p \vee q)] \Leftrightarrow p$.

§ 2.5. Analiza rozumowań

W paragrafie 2.4 opisaliśmy formalizację pojęcia „dowodu". Naszym celem w tym paragrafie będzie wykorzystanie tej abstrakcyjnej wersji pojęcia dowodu jako wskazówki do głębszego zrozumienia dowodów i błędów rozumowania, które spotykamy na codzień. Zastanowimy się na przykładach, w jaki sposób można tworzyć dowody i zobaczymy, na jakie błędy należy zwracać uwagę. Taka sama analiza, jaka pozwala nam ocenić poprawność dowodu, pozwoli również rozwikłać skomplikowane konstrukcje logiczne.

Nasze formalne reguły wnioskowania czasami nazywamy **wnioskowaniami poprawnymi**, a dowody formalne czasem nazywamy **dowodami poprawnymi**. Rozszerzymy znaczenie słowa „poprawny" na rozumowania i wnioskowania nieformalne, które odpowiadają formalnym dowodom i wykorzystywanym w nich regułom wnioskowania. Ciąg zdań nie spełniający warunków, które powinien spełniać dowód formalny, nazywamy **rozumowaniem błędnym**. Będziemy używać tej nazwy również na oznaczenie rozumowań, których formalizacja nie jest dowodem poprawnym.

Zobaczymy wiele różnych rozumowań błędnych. Najpierw jednak przeprowadzimy kilka rozumowań poprawnych.

PRZYKŁAD 1 Przeanalizujmy następujące rozumowanie. „Jeśli będę się uczył lub jestem geniuszem, zdam egzamin. Jeśli zdam egzamin, będę mógł uczęszczać na następne wykłady. Zatem, jeśli nie zostanę dopuszczony do następnych wykładów, to nie jestem geniuszem". Niech

s = „będę się uczył,"

g = „jestem geniuszem,"

p = „zdam egzamin,"

n = „zostanę dopuszczony do następnych wykładów."

Mamy tu dwa założenia $s \vee g \to p$ i $p \to n$ i chcemy dowieść zdania $\neg n \to \neg g$. Oto szkic rozumowania. Ponieważ $\neg n \to \neg g$ jest

zdaniem przeciwstawnym do zdania $g \to n$, więc wystarczy podać dowód zdania $g \to n$, a potem powołać się na regułę kontrapozycji, czyli prawo 9. Ze zdania $s \vee g \to p$ możemy oczywiście wywnioskować zdanie $g \to p$ (szczegóły za chwilę), z którego razem ze zdaniem $p \to n$, na podstawie reguły 33 wynika zdanie $g \to n$. Aby otrzymać zdanie $g \to p$ ze zdania $s \vee g \to p$, korzystamy z reguły 33 oraz prawa wprowadzania alternatywy, dającego nam zdanie $g \to s \vee g$. Dokładniej, prawo wprowadzania alternatywy daje nam $g \to g \vee s$, a więc musimy jeszcze skorzystać z prawa przemienności. A oto formalizacja tego rozumowania.

Dowód	**Wyjaśnienia**
1. $s \vee g \to p$	założenie
2. $p \to n$	założenie
3. $g \to g \vee s$	prawo wprowadzania alternatywy 16
4. $g \to s \vee g$	3; prawo przemienności 2a
5. $g \to p$	4, 1; reguła sylogizmu hipotetycznego 33
6. $g \to n$	5, 2; reguła sylogizmu hipotetycznego 33
7. $\neg n \to \neg g$	6; prawo kontrapozycji 9

Zatem istnieje poprawny dowód prowadzący od założeń do tezy. Czy to utwierdza nas w przekonaniu o poprawności rozumowania? Oczywiście tego typu rozumowania zazwyczaj przeprowadzamy codziennie w pamięci i wydaje się, że cały ten bagaż formalizmu raczej komplikuje łatwy problem. ∎

Jednak próbujmy dalej.

PRZYKŁAD 2 „Jeśli będę się uczył lub jestem geniuszem, zdam egzamin. Nie zostanę dopuszczony do następnych wykładów. Jeśli zdam egzamin, zostanę dopuszczony do następnych wykładów. A więc nie będę się uczył". Korzystając z oznaczeń z przykładu 1, chcemy przeprowadzić dowód stwierdzenia: jeśli $s \vee g \to p$, $\neg n$ oraz $p \to n$, to $\neg s$. Jedyną regułą wnioskowania w tablicy 2.3 w § 2.4, z której warto skorzystać, jest reguła 31 (modus tollens). Ta reguła pozwala nam wyprowadzić z założenia $\neg n$ wniosek $\neg s$, o ile wcześniej wyprowadzimy zdanie $s \to n$; to zdanie możemy jednak wyprowadzić z pierwszego i trzeciego założenia dokładnie tak, jak w przykładzie 1. A oto wersja formalna tego rozumowania.

Dowód	**Wyjaśnienia**
1. $s \vee g \to p$	założenie
2. $s \to s \vee g$	prawo wprowadzania alternatywy 16
3. $s \to p$	2, 1; reguła sylogizmu hipotetycznego 33

4. $p \to n$	założenie
5. $s \to n$	3, 4; reguła sylogizmu hipotetycznego 33
6. $\neg n$	założenie
7. $\neg s$	5, 6; reguła modus tollens 31

To rozumowanie można wyrazić słowami w następujący sposób:

Jeśli będę się uczył, to ponieważ zarówno uczenie się, jak i bycie geniuszem powoduje zdanie egzaminu, więc zdam. Ale wtedy, ponieważ zdanie egzaminu pozwoli mi wybrać następne wykłady, będę mógł je wybrać. Zatem, jeśli nie będę mógł wybrać następnych wykładów, coś musi mi przeszkodzić — pewnie to, że się nie będę uczył. Prawo 16 jest tu użyte w milczący sposób, pozostałe trzy wnioskowania są oczywiście „logiczne". ∎

Ten przykład pokazuje ogólną strategię pomocną przy tworzeniu dowodów. Popatrzmy na tezę C. Z czego może ona wynikać? Na przykład, czy umiemy udowodnić coś w rodzaju $B \to C$? Jak trudno byłoby dowieść B? I tak dalej, postępując od końca, od C do B, potem do A, potem .... Można również przyjrzeć się założeniom. Co potrafimy szybko z nich wyprowadzić? Czy któreś z tych wniosków są jakoś związane z C? Jeśli musieliście kiedyś udowodnić tożsamość trygonometryczną, to na pewno rozpoznacie tę strategię: przekształcamy prawą stronę, potem lewą, potem znów prawą itd. próbując doprowadzić obie strony do tej samej postaci.

PRZYKŁAD 3 Możemy również udowodnić tezę z przykładu 2 przez sprowadzenie do sprzeczności, tzn. założyć negację tezy, czyli to, że będę się uczył i dojść do sprzeczności. Dowód formalny wygląda na trochę bardziej skomplikowany. Z założeń $s \vee g \to p$, $\neg n$, $p \to n$ oraz z nowego założenia $\neg(\neg s)$ chcemy wyprowadzić sprzeczność w rodzaju $s \wedge (\neg s)$, $n \wedge (\neg n)$, $g \wedge (\neg g)$ czy $p \wedge (\neg p)$. Ponieważ mamy już $\neg n$ i s (z $\neg(\neg s)$), wydaje się, że najłatwiej będzie uzyskać którąś z dwóch pierwszych sprzeczności. Próbujemy otrzymać sprzeczność $n \wedge (\neg n)$, gdyż mamy już $p \to n$. Mamy też s i możemy wyprowadzić $s \to p$ z $s \vee g \to p$, a więc możemy otrzymać p. A oto jeden z możliwych dowodów formalnych; por. ćwiczenie 11, gdzie zamiast reguły sylogizmu hipotetycznego będzie wykorzystana reguła modus ponens.

Dowód	**Wyjaśnienia**
1. $s \to s \vee g$	prawo wprowadzania alternatywy 16
2. $s \vee g \to p$	założenie

3. $s \to p$	1, 2; reguła sylogizmu hipotetycznego 33
4. $p \to n$	założenie
5. $s \to n$	3, 4; reguła sylogizmu hipotetycznego 33
6. $\neg(\neg s)$	negacja tezy
7. s	6; prawo podwójnego przeczenia 1
8. n	7, 5; reguła modus ponens 30
9. $\neg n$	założenie
10. $n \wedge (\neg n)$	8, 9; reguła wprowadzania koniunkcji 34
11. sprzeczność	10; prawo 7b

W języku potocznym można to rozumowanie wyrazić następująco:

Jak w przykładzie 2, jeśli będę się uczył, to będę mógł wybrać następne wykłady. Przypuśćmy więc, że będę się uczył (tu w ukryty sposób korzystamy z prawa podwójnego przeczenia). Zatem będę mógł wybrać następne wykłady. Ale z założenia nie będę mógł ich wybrać. Ta sprzeczność dowodzi, że moje przypuszczenie było nieprawdziwe.

Na ogół dowód przez sprowadzenie do sprzeczności zaczyna się od przypuszczenia i jest prowadzony dotąd, aż otrzyma się sprzeczność. Wtedy wracamy do ostatniego „przypuszczenia" i mówimy: „no dobrze, to przypuszczenie musiało być złe". Mogliśmy też po prostu wnioskować, że założenia oraz $\neg C$ są ze sobą sprzeczne i jeśli chcemy zachować $\neg C$, to musimy zanegować jedno z założeń.

Formalny dowód metodą sprowadzenia do sprzeczności w tym przykładzie jest dłuższy niż dowód w przykładzie 2, ale jego wersja w języku potocznym jest mniej więcej tej samej długości i może być pojęciowo prostsza. W codziennej praktyce sięgamy po dowody nie wprost wtedy, gdy łatwiej jest użyć $\neg C$ razem z założeniami, niż wyprowadzać C z założeń. Ilustrowały to przykłady w § 2.3. ∎

Gdyby było łatwo konstruować dowody, każdy by to robił i matematycy nie mieliby pracy. Jednak mając odrobinę wprawy i przyglądając się dokładnie dowodom, które czytamy, możemy całkiem dobrze nauczyć się budować własne dowody i znajdować słabe punkty w dowodach innych ludzi.

W tej książce jest wiele przykładów poprawnych dowodów. Popatrzmy teraz na kilka błędnych rozumowań.

PRZYKŁAD 4 Wiemy, że jeśli program komputerowy działa poprawnie, to zaczyna i kończy swoje działanie oraz wiemy, że nasz program rozpoczął działanie i nie działał poprawnie. Wnioskujemy stąd,

że program nie zakończył działania. Czy to rozumowanie jest poprawne? Niech

B = „program rozpoczął działanie"

T = „program zakończył działanie"

F = „program nie działa poprawnie"

Naszymi założeniami są: $\neg F \to (B \wedge T)$ oraz $B \wedge F$, a tezą jest $\neg T$. Oto próba dowodu formalnego.

Dowód?	**Wyjaśnienia**
1. $\neg F \to (B \wedge T)$	założenie
2. $B \wedge F$	założenie
3. $(B \wedge T) \to T$	prawo opuszczania koniunkcji 17
4. $\neg F \to T$	1, 3; reguła sylogizmu hipotetycznego 33
5. F	2; reguła opuszczania koniunkcji 29
6. $\neg T$	4, 5; ??

Jak możemy wyprowadzić zdanie 6 ze zdań 4 i 5? Wydaje się, że największą nadzieję daje reguła modus tollens, ale po bliższym przyjrzeniu się widzimy, że nie można jej tu zastosować. Potrzebujemy reguły $P \to Q$, $\neg P$, $\therefore \neg Q$, a to nie jest poprawna reguła wnioskowania. Powyższy rzekomy dowód nie jest poprawny; jest to rozumowanie błędne. Być może po prostu prowadziliśmy dowód w niewłaściwy sposób? Nie. Tak naprawdę, żaden poprawny dowód w tym przypadku nie istnieje, ponieważ nasza teza nie wynika z przyjętych założeń. To znaczy, zdanie

$$\{[\neg F \to (B \wedge T)] \wedge (B \wedge F)\} \to \neg T$$

nie jest tautologią. Aby się o tym przekonać, popatrzmy na ten wiersz w matrycy logicznej, w którym B, F i T są prawdziwe. Oznacza to, że program może zacząć działanie, zakończyć działanie i nie działać poprawnie z jakiegoś innego powodu.

■

PRZYKŁAD 5 (a) Jestem sławnym koszykarzem. Sławni koszykarze zarabiają dużo pieniędzy. Jeśli zarabiam dużo pieniędzy, to ty powinnaś robić to, co każę. Mówię, że powinnaś kupić buty Pearly Maid. Zatem powinnaś kupić te buty.

Oznaczmy

B = „koszykarz"

M = „zarabia dużo pieniędzy"

D = „robić to, co każę"

S = „kupić te buty!"

Założeniami są B, $B \to M$, $M \to D$ oraz $D \to S$, a teza S wynika oczywiście po zastosowaniu trzy razy reguły modus ponens.

(b) Przypuśćmy, że w przykładzie (a) opuszczamy założenie $M \to D$. Wtedy nie ma poprawnej metody wywnioskowania S z założeń B, $B \to M$ oraz $D \to S$. (Popatrzmy na przykład na przypadek, gdy S i D są fałszywe, a B i M prawdziwe). Rozumowanie wymagałoby uzupełnienia o ukryte założenie, że powinniśmy robić to, co każą nam robić ludzie mający dużo pieniędzy. (Zazwyczaj, gdy przedstawiamy to rozumowanie, to nie wypowiadamy wyraźnie implikacji $B \to M$ i $M \to D$).

(c) Przypuśćmy, że zamiast założenia $M \to D$ w oryginalnej wersji, mamy odwołanie się do współczucia: „zarobię pieniądze tylko wtedy, gdy kupisz nasze buty". Założeniami są teraz B, $B \to M$, $M \to S$ oraz $D \to S$. Znowu S jest poprawnym wnioskiem. Założenie $M \to S$ wydaje się podejrzane, ale jeśli się je uzna za prawdziwe, to rozumowanie będzie bez zarzutu.

(d) Możemy zastąpić założenie $M \to D$ w oryginalnym zadaniu bardziej wiarygodnym stwierdzeniem $B \vee S \to M$, tzn „zarobię pieniądze tylko wtedy, gdy będę sławnym koszykarzem lub gdy kupisz nasze buty". Mieliśmy już mocniejsze założenie $B \to M$, zatem analiza z przykładu (b) pokazuje, że nie możemy wywnioskować S również w tym przypadku. ∎

Czasami sytuacja jest tak skomplikowana, że wcale nie jest jasne na początku, czy powinno się próbować pokazać, że domniemany wniosek wynika z założeń, czy nie wynika. W takim przypadku często warto popatrzeć, co można wywnioskować z założeń, mając nadzieję, że poznamy różne możliwości lub też dojdziemy do sprzeczności. W każdym przypadku coś zyskamy, w najgorszym razie takie podejście zmusi nas do starannej i metodycznej pracy nad problemem, co często wystarcza do rozwiązania go.

PRZYKŁAD 6 (a) Niech

A = „jestem dorosły",
B = „jestem duży i dzielny",
Y = „lata mijają",
L = „życie jest ciężkie",
N = „nikt mnie nie kocha".

Oto fragment rozumowania. Jeśli jestem dorosły, to jestem duży i dzielny. Życie jest ciężkie i nikt mnie nie kocha, jeżeli jestem duży i dzielny lub lata mijają. Jeśli jestem duży i dzielny i nikt mnie nie kocha, to jestem dorosły. Jeśli jestem duży i dzielny

lub nie mijają lata, to też jestem dorosły. Jeżeli jestem duży i dzielny lub lata mijają, to nikt mnie nie kocha. Lata albo mijają, albo nie. Jeśli jestem duży i dzielny i ktoś mnie kocha, to jestem dorosły. A więc jestem dorosły (jest to przykład bardzo dorosłego rozumowania).

Ale mętlik! Czy ten wniosek rzeczywiście jest poprawny? Popatrzmy. Założenia to: $A \to B$, $(B \vee Y) \to (L \wedge N)$, $(B \wedge N) \to A$, $(B \vee \neg Y) \to A$, $(B \vee Y) \to N$ (wniosek z $(B \vee Y) \to (L \wedge N)$), $Y \vee \neg Y$ (tautologia) i $(B \wedge \neg N) \to A$. Aby wszystkie te założenia były prawdziwe, a zdanie A mimo to było fałszywe, to fałszywe muszą być wszystkie zdania p takie, że $p \to A$ jest założeniem. Zatem, jeśli A jest fałszywe, to $B \wedge N$, $B \vee \neg Y$ i $B \wedge \neg N$ muszą być fałszywe. Jeśli $B \vee \neg Y$ jest fałszywe, to B musi być fałszywe a Y prawdziwe. Następnie, ponieważ zdanie $(B \vee Y) \to (L \wedge N)$ jest prawdziwe, więc również prawdziwe jest zdanie $Y \to (L \wedge N)$, a zatem zdania L i N muszą być prawdziwe. W tym momencie możemy stwierdzić, że jeśli zdanie A jest fałszywe, to zdanie B też musi być fałszywe, natomiast zdania Y, L i N muszą być prawdziwe. Możemy sprawdzić, że przy takich wartościach logicznych tych zdań wszystkie założenia są prawdziwe, a więc A nie wynika z założeń przy jakimkolwiek poprawnym rozumowaniu.

Nie musimy dokonywać całej takiej analizy, jak w ostatnim akapicie. Tak naprawdę potrzebujemy zestawu wartości logicznych, przy których wszystkie założenia są prawdziwe, a zdanie A jest fałszywe. Ale jak znaleźć takie wartości? Sposób, w jaki przeprowadziliśmy naszą analizę, pokazuje, że możemy albo znaleźć takie wartości, albo też możemy znaleźć dowód, że takich wartości nie ma. Obie te odpowiedzi są przydatne.

Nasze rozumowanie składało się naprawdę z wielu założeń i wniosku, bez jakichkolwiek kroków pośrednich. Jest całkiem możliwe, że rozumowanie składa się z bardzo wielu poprawnych kroków, ale mimo to jest rozumowaniem błędnym, gdyż gdzieś w środku zawiera lukę.

(b) Przypuśćmy, że chcemy udowodnić, iż życie jest ciężkie. Często spotykane rozumowanie wygląda następująco:

> Załóżmy, że życie jest ciężkie.
> Tratatata ...
> Zatem życie jest ciężkie.

Wyciąganie wniosków w taki sposób jest tak niepoprawne, że wydaje się, iż nie warto nawet o tym mówić, ale doświadczenie uczy, że ten błąd, często dobrze ukryty, jest bardzo popularny.

(c) Przypuśćmy, że chcemy pokazać, że „jeśli jestem dorosły, to życie jest ciężkie". Spróbujmy tak.

Załóżmy L. Ponieważ $L \to (A \to L)$ jest tautologią, więc $A \to L$ musi być zdaniem prawdziwym.

Tu znów zaczynamy rozumowanie od przyjęcia dodatkowego założenia, wystarczająco silnego, by z niego otrzymać tezę. ∎

PRZYKŁAD 7 (a) Wspomnieliśmy już w § 2.3 o błędnym dowodzeniu równoważności $A \Leftrightarrow B$ polegającym na pokazaniu $A \Rightarrow B$ i $\neg B \Rightarrow \neg A$.

(b) Podobny błąd polega na dowodzeniu zdania odwrotnego do tego, którego mamy dowieść, to znaczy na dowodzeniu $B \Rightarrow A$ zamiast $A \Rightarrow B$. Fałszywa reguła wnioskowania A, $B \to A$, $\therefore B$ jest przykładem podobnego błędu. Zdumiewające jest, jak wiele osób wpada w te pułapki. Czasami być może powoduje to mylący szyk zdania. A więc przypuśćmy, że jestem bogaty i że ludzie są bogaci, jeśli tylko mieszkają w dużych domach. Stąd nie wynika, że ja mieszkam w dużym domu. Mogę przecież mieszkać na pokładzie mojego jachtu. Mamy tu R i $H \to R$, a chcemy wywnioskować H. Do tego jednak chcielibyśmy mieć $R \to H$, czyli że wszyscy bogaci ludzie mieszkają w dużych domach. ∎

PRZYKŁAD 8 Spróbujemy uzasadnić zdanie „jeśli dostanę dwóję, to zdam egzamin", prowadząc dowód przez sprowadzenie do sprzeczności. Załóżmy zatem, że nie zdam egzaminu. Znaczyłoby to, że dostanę dwóję (jest to zaprzeczenie tego, czego chcemy dowieść, ale z pewnością jest to prawda). Ale kontrapozycją zdania „jeśli dostanę dwóję, to zdam egzamin" jest zdanie „jeśli nie zdam egzaminu, to nie dostanę dwói". Nie mogę jednocześnie dostać dwói i nie dostać dwói, a więc mamy sprzeczność. Jedyne założenie, mianowicie to, że nie zdam egzaminu, musi być fałszywe. Hurra! Zdam!

W tym przykładzie trochę wikłaliśmy się do momentu, kiedy wypowiedzieliśmy jakieś zdanie i jego zaprzeczenie. Wtedy uznaliśmy, że mamy sprzeczność.

Dowody przez sprowadzenie do sprzeczności są szczególnie pomysłowe. Jeśli są poprawne, to opisują sytuację, która nie może się naprawdę zdarzyć i która powinna przeczyć naszej intuicji. Oczywiście powyższy przykład celowo był taki zagmatwany. Prawdziwe dowody przez sprowadzenie do sprzeczności są często wyjątkowo skomplikowane i polegają na znajdowaniu coraz to nowych argumentów dotąd, aż opisywana sytuacja okazuje się być niemożliwą. Jeden mały błąd może spowodować wystąpienie sprzeczności tam, gdzie naprawdę jej nie ma i może spowodować, że fałszywa argumentacja wyda się poprawną. ∎

Powinniście być wyczuleni na pewne często występujące zwroty. Kiedy ktoś stwierdza, że coś „jest zgodne ze zdrowym rozsądkiem", możecie z dużą dozą prawdopodobieństwa oczekiwać luki w rozumowaniu. Ten ktoś nie potrafi w poprawny sposób uzasadnić tego i próbuje stworzyć na poczekaniu potrzebną tautologię lub regułę wnioskowania, nie precyzując jej dokładnie. Dowody „przez zastraszenie" lub przez „odwołanie się do autorytetu" są szczególnie drastycznymi przykładami tego błędu.

Słowa „jasne" czy „oczywiście" też wskazują na możliwość błędu, nawet w wydrukowanych pracach. Oczywiście nie ma błędów w tej książce, ale poprzednie wydania zawierały jeden czy dwa błędy logiczne.

Kiedy tworzysz dowód, by rozwiązać zadanie, powinieneś sprawdzić, czy zostały wykorzystane wszystkie założenia. W prawdziwych problemach często mamy nadmiar informacji, ćwiczenia w podręcznikach zazwyczaj tego nadmiaru nie mają. Jeśli nie skorzystałeś ze wszystkiego, co zostało podane, to prawdopodobnie twoje rozumowanie zawiera błąd, na przykład przeoczyłeś jakiś przypadek.

Schematy rozumowania, które pozwalają nam sprawdzać poprawność dowodów, mogą również być użyteczne przy rozwikływaniu skomplikowanych logicznie sformułowań.

PRZYKŁAD 9 Popatrzmy na dość typowy wyjątek z podręcznika potężnego systemu operacyjnego. Niech

A = „pojawił się sygnał wysłany do procesu",
P = „sygnał został dodany do zbioru sygnałów oczekujących na odebranie przez proces",
B = „sygnał jest teraz zablokowany przez proces",
D = „sygnał został dostarczony do procesu (proces odebrał sygnał)",
S = „bieżący stan procesu jest zachowany",
M = „oblicza się nową maskę sygnałów",
H = „wywołanie procedury obsługi sygnałów",
N = „procedura obsługi jest wywoływana w zwykły sposób",
R = „proces wznawia wykonanie w poprzednim kontekście",
I = „proces musi sam odtworzyć poprzedni kontekst".

Z podręcznika dowiadujemy się, że „$A \to P$, $(P \wedge \neg B) \to D$, $D \to (S \wedge M \wedge H)$, $(H \wedge N) \to R$, $(H \wedge \neg R) \to I$". Tak jest rzeczywiście. Po prostu przetłumaczyliśmy ten podręcznik z języka polskiego na język liter i symboli i opuściliśmy tylko kilka słów.

Co możemy wywnioskować z tych założeń? W szczególności, co się zdarzy, gdy zdanie $A \wedge \neg B \wedge \neg R$ będzie prawdziwe, tzn. gdy

pojawi się sygnal, ten sygnał nie będzie zablokowany przez proces, ale proces nie będzie chciał wznowić wykonania w poprzednim kontekście?

Możemy wywnioskować $P \wedge \neg B$ z $A \to P$ i $A \wedge \neg B$ (por. ćwiczenie 13(a)). Zatem korzystając z założeń $(P \wedge \neg B) \to D$ i $D \to (S \wedge M \wedge H)$, otrzymujemy $S \wedge M \wedge H$. W szczególności prawdziwe jest zdanie H. Za pomocą innego krótkiego dowodu (zob. ćwiczenie 13(b)) możemy wyprowadzić zdanie $\neg N$ z założeń $H \wedge \neg R$ i $(H \wedge N) \to R$. Zatem, ponieważ $(H \wedge \neg R) \to I$, więc z $H \wedge \neg R$ wynika $I \wedge \neg N$. Moglibyśmy również pokazać, że jeśli prawdziwe jest zdanie A, a zdania B i R są fałszywe, to zdanie I jest prawdziwe, a zdanie N jest fałszywe; tzn. procedura obsługi sygnałów nie jest wywoływana normalnie i proces musi odtworzyć swój poprzedni kontekst. Po drodze pokazaliśmy również, że zdania P, D, S, M i H są prawdziwe.

Oczywiście normalnie nikt nie zapisywałby takiej analizy w postaci dowodu formalnego. Jednak pomocne jest rozbicie opisu słownego na poszczególne stwierdzenia, nadanie tym stwierdzeniom nazw i zapisanie za pomocą symboli założeń i wniosków wynikających z nich. ∎

ĆWICZENIA DO § 2.5

1. Zaobserwowaliśmy C oraz to, że A implikuje C. Wnioskujemy stąd, że to znaczy, iż gdyby A było fałszywe, to C również byłoby fałszywe, a nie jest. Zatem A musi być prawdziwe. Czy to rozumowanie jest poprawne? Odpowiedź uzasadnij.

2. Podaj dwa przykłady rozumowań błędnych z życia codziennego. Wyjaśnij, dlaczego te rozumowania są błędne. Sugestia: reklamy i listy do redakcji są dobrym źródłem takich rozumowań.

3. (a) Pokaż, że nie da się wyprowadzić w poprawny sposób zdania C z założenia $A \to C$.

(b) Czy istnieje poprawny dowód zdania C z założeń $A \to C$ i $B \to C$? Odpowiedź uzasadnij.

(c) Jak mocnym argumentem za prawdziwością zdania C jest prawdziwość implikacji $A_1 \to C$, $A_2 \to C$, ..., $A_{1000000} \to C$? Odpowiedź uzasadnij.

4. (a) Jeśli w przykładzie 5(a) opuścimy założenie $B \to M$, to czy z pozostałych założeń nadal będzie wynikać zdanie S? Odpowiedź uzasadnij.

(b) Czy w ćwiczeniu (a) mocniejszym argumentem za prawdziwością zdania S będzie dodanie założenia $M \to S$, jeśli w dalszym ciągu pominiemy założenie $B \to M$? Odpowiedź uzasadnij.

5. Dla każdego z następujących zbiorów założeń sformułuj wynikający z nich wniosek i podaj reguły wnioskowania, z których skorzystałeś.
(a) Jeśli telewizor nie będzie zepsuty, to nie będę się uczył. Jeśli będę się uczył, to zdam egzamin. Nie zdam egzaminu.
(b) Jeśli zaliczyłem pierwszy i drugi semestr, to zaliczyłem rok. Jeśli zaliczyłem rok, to zaliczyłem drugi semestr. Nie zaliczyłem roku.
(c) Jeśli zaliczyłem pierwszy i drugi semestr, to zaliczyłem rok. Będę studiował na następnym roku tylko wtedy, gdy zaliczę ten rok. Nie będę studiował na następnym roku.

6. Rozważmy następujące założenia. Jeśli pojadę autobusem lub metrem, to spóźnię się na spotkanie. Jeśli pojadę taksówką, to nie spóźnię się na spotkanie, ale zbankrutuję. Nie spóźnię się na spotkanie.
Które z następujących wniosków muszą być prawdziwe, tzn. mogą być wyprowadzone z założeń? Odpowiedź uzasadnij.
(a) Pojadę taksówką.
(b) Zbankrutuję.
(c) Nie pojadę metrem.
(d) Jeśli zbankrutowałem, to pojechałem taksówką.
(e) Jeśli pojadę autobusem, to nie zbankrutuję.

7. Przyjmij założenia z przykładu 6(a). Które z następujących zdań są poprawnie wyprowadzonymi wnioskami? Uzasadnij każdą odpowiedź.
(a) Jestem dorosły wtedy i tylko wtedy, gdy jestem duży i dzielny.
(b) Jeśli jestem duży i dzielny, to ktoś mnie kocha.
(c) Albo lata mijają, albo jestem dorosły.
(d) Życie jest ciężkie i nikt mnie nie kocha.

8. Zrobiła to Ania lub Basia. Basia nie mogła jednocześnie czytać i zrobić tego. Basia czytała. Kto to zrobił? Odpowiedź uzasadnij za pomocą odpowiedniego dowodu formalnego ze zmiennymi A, B i Z.

9. Zapisz każde z poniższych rozumowań za pomocą symboliki logicznej, używając sugerowanych nazw zmiennych. Następnie napisz dowód formalny.
(a) „Jeśli moje obliczenia się zgadzają i zapłacę rachunek za elektryczność, to zabraknie mi pieniędzy. Jeśli nie zapłacę rachunku, to wyłączą mi prąd. Zatem, jeśli nie zabraknie mi pieniędzy i prądu mi nie wyłączą, to moje obliczenia się nie zgadzają. " (o, r, z, p)
(b) „Jeśli meteorolodzy przewidują, że będzie sucho, to pójdę na wycieczkę lub będę pływać. Pójdę pływać wtedy i tylko wtedy, gdy meteorolodzy podadzą, że będzie ciepło. Zatem, jeśli nie idę na wycieczkę, meteorolodzy przewidują, że będzie mokro lub ciepło." (s, w, p, m).
(c) „Jeśli dostanę pracę i będę ciężko pracować, to będę awansować. Jeśli będę awansować, będę zadowolony. Nie będę zadowolony. Zatem albo nie dostanę pracy albo nie będę ciężko pracować." (p, c, a, z)

(d) „Jeśli będę studiował prawo, to będę zarabiał dużo pieniędzy. Jeśli będę studiował archeologię, to będę dużo podróżował. Jeśli będę zarabiał dużo pieniędzy lub dużo podróżował, to nie będę nieszczęśliwy. Zatem, jeśli nie jestem nieszczęśliwy, to nie studiowałem prawa i nie studiowałem archeologii." (sp, d, sa, p, n)

10. (a) Przerób dowód w przykładzie 3 tak, aby dojść do sprzeczności $s \wedge (\neg s)$. *Wskazówka*: wykorzystaj linie 5 i 9 i zmień kolejność kroków.

(b) Przerób dowód w przykładzie 3 tak, aby dojść do sprzeczności $p \wedge (\neg p)$. *Wskazówka*: wykorzystaj linie 3 i 7, a następnie linie 4 i 9 i zmień kolejność kroków.

11. Podaj inny dowód formalny do przykładu 3, w którym nie korzysta się z reguły 33.

12. W każdym z następujących przypadków podaj dowód formalny twierdzenia lub pokaż, że jest ono fałszywe, przedstawiając odpowiedni wiersz w matrycy logicznej:

(a) Jeśli $(q \wedge r) \rightarrow p$ i $q \rightarrow \neg r$, to p.

(b) Jeśli $q \vee \neg r$ i $\neg(r \rightarrow q) \rightarrow \neg p$, to p.

(c) Jeśli $p \rightarrow (q \vee r)$, $q \rightarrow s$ i $r \rightarrow \neg p$, to $p \rightarrow s$.

13. Podaj dowody formalne następujących twierdzeń:

(a) $P \wedge \neg B$ z $A \rightarrow P$ i $A \wedge \neg B$.

(b) $\neg N$ z $H \wedge \neg R$ i $(H \wedge N) \rightarrow R$. *Sugestia*: dowód przez sprowadzenie do sprzeczności.

14. Jeśli jesteś tutaj, to dzisiaj musi być piątek. Jeśli się nie mylę, to dzisiaj jest sobota. Albo dzisiaj to nie wczoraj, albo dzisiaj jest piątek. Nie mogę się mylić, jeśli jesteś tutaj. Piątek to nie sobota. Jeśli dzisiaj jest sobota, to wczoraj był piątek.

(a) Czy jesteś tutaj?

(b) Przypuśćmy, że się nie mylę. Czy możesz wywnioskować, że wczoraj to nie dzisiaj?

(c) Jeśli się nie mylę, jaki dzień był wczoraj? Odpowiedź uzasadnij.

To, co jest najważniejsze w tym rozdziale

Aby sprawdzić, czy dobrze rozumiesz treść tego rozdziału:

1. Przekonaj się, że potrafisz zdefiniować każde pojęcie i oznaczenie oraz możesz opisać każdą metodę.

2. Podaj przynajmniej jeden powód, dla którego dany temat został omówiony w tym rozdziale.

3. Zastanów się nad co najmniej jednym przykładem ilustrującym dane pojęcie oraz co najmniej jedną sytuacją, w której dany fakt czy metoda są przydatne.

Pojęcia i oznaczenia

rachunek zdań
- zdanie
- spójniki logiczne $\neg$, $\vee$, $\wedge$, $\rightarrow$, $\leftrightarrow$
- znaki $\forall$, $\exists$
- zdanie odwrotne, kontrapozycja, kontrprzykład
- warunek konieczny, warunek wystarczający
- zdanie złożone
  - tablica wartości logicznych (matryca wartości logicznych)
  - tautologia, sprzeczność
  - równoważność logiczna $\Leftrightarrow$, implikacja logiczna $\Rightarrow$

metody dowodzenia
- wprost, nie wprost, przez sprowadzenie do sprzeczności, przez rozpatrzenie przypadków
- dowód „w próżni", dowód trywialny
- konstruktywny, niekonstruktywny

dowód formalny
- założenie, teza, twierdzenie
- reguła wnioskowania
- zapis kroków dowodu jeden pod drugim

analiza rozumowań
- rozumowanie poprawne, dowód poprawny, błąd rozumowania

Fakty

Podstawowe równoważności logiczne (tablica 2.1 w § 2.2).
Podstawowe implikacje logiczne (tablica 2.2 w § 2.2).
Reguły podstawiania (a) i (b) w § 2.4.
Podstawowe reguły wnioskowania (tablica 2.3 w § 2.4).
Istnieje dowód formalny C z H wtedy i tylko wtedy, gdy zdanie $H \Rightarrow C$ jest prawdziwe.

Metody

Użycie:
- tablic wartości logicznych, szczególnie do sprawdzania równoważności i implikacji logicznych
- praw De Morgana do eliminowania $\vee$ i $\wedge$
- reguł wnioskowania do konstruowania dowodów formalnych
- symboliki formalnej do analizowania nieformalnych rozumowań.

3. RELACJE

Często chcemy porównywać lub przeciwstawiać sobie różne elementy zbioru, na przykład po to, by ustawić je w jakiejś kolejności lub zgrupować razem elementy o podobnych własnościach. Matematyczne podstawy opisywania takich struktur w zbiorach daje teoria relacji. W tym rozdziale wprowadzimy pojęcie relacji i omówimy ich związek z grafami skierowanymi i macierzami.

§ 3.1. Relacje

Dla danych zbiorów S i T **relacją dwuargumentową** na zbiorze $S \times T$ jest dowolny podzbiór R zbioru $S \times T$. Jest to bardzo ogólna definicja. W tej książce zobaczymy wiele różnych rodzajów interesujących relacji.

PRZYKŁAD 1 Firma wysyłkowa sprzedająca płyty ma listę L swoich klientów. Każdy klient wskazuje rodzaj muzyki, jaka go interesuje: klasyczna, lekka, latynoamerykańska, religijna, pop, rock itp. Niech C będzie zbiorem wszystkich możliwych typów muzyki. Zbiór par uporządkowanych (nazwisko, wybrany rodzaj muzyki) jest relacją R na zbiorze $L \times C$. Relacja ta może zawierać takie pary jak: (K. A. Ross, klasyczna), (C. R. B. Wright, klasyczna) i (C. R. B. Wright, thrash metal). ∎

PRZYKŁAD 2 Uniwersytet może być zainteresowany relacją R_1 składającą się ze wszystkich par uporządkowanych, których poprzednikami są studenci, a następnikami są wykłady, na które ci studenci są aktualnie zapisani. Jest to relacja na zbiorze $S \times C$, gdzie S jest

zbiorem studentów tego uniwersytetu, a C jest zbiorem oferowanych wykładów. Zauważmy, że dla danego studenta s w zbiorze S zbiór $\{c \in C\colon\ (s,c) \in R_1\}$ jest zbiorem wykładów wybranych przez studenta s. Z drugiej strony, dla danego wykładu c w zbiorze C zbiór $\{s \in S\colon\ (s,c) \in R_1\}$ jest zbiorem słuchaczy tego wykładu.

Inna relacja R_2 składa się z par uporządkowanych, których poprzednikami są wykłady, a następnikami są wydziały, których studenci uczęszczają na te wykłady. Zatem R_2 jest relacją na zbiorze $C \times D$, gdzie D jest zbiorem wydziałów uniwersytetu. Dla danego $c \in C$ zbiór $\{d \in D\colon\ (c,d) \in R_2\}$ jest zbiorem wydziałów, których studenci uczęszczają na wykład c. Dla danego $d \in D$ zbiór $\{c \in C\colon\ (c,d) \in R_2\}$ jest zbiorem wykładów dla studentów danego wydziału. Skomputeryzowany system oceniania studentów musiałby używać struktury danych zawierającej wystarczająco dużo informacji do tego, by określić relacje R_1 i R_2. ■

PRZYKŁAD 3 (a) Rozważmy zbiór P programów, które mogą być wykonane na danym komputerze oraz katalog C gotowych programów, które mogą być wykorzystywane przez inne programy. Określamy relację na zbiorze $C \times P$, mówiąc, że gotowy program c jest w relacji z programem $p \in P$, jeśli program p wywołuje program c jako podprogram. Często używany program c jest w relacji z wieloma programami p, natomiast program c, który nigdy nie jest wywoływany, nie jest w relacji z żadnym programem p.

(b) Program przekształcający rozwinięcie dziesiętne liczby na rozwinięcie dwójkowe może być uważany za relację składającą się z par uporządkowanych, których poprzednikami są dopuszczalne rozwinięcia dziesiętne i których następnikami są odpowiednie rozwinięcia dwójkowe. Ta relacja w rzeczywistości jest funkcją. ■

Przypomnijmy, że w § 1.3 pokazaliśmy, w jaki sposób funkcje można utożsamiać z ich wykresami, a zatem ze zbiorami par uporządkowanych. W rzeczywistości, jeśli $f\colon S \to T$, to funkcję f utożsamialiśmy ze zbiorem

$$R_f = \{(x,y) \in S \times T\colon\ y = f(x)\},$$

który jest relacją na zbiorze $S \times T$. Oczywiście nie wszystkie relacje są funkcjami. Jeśli traktujemy funkcje jako relacje, to **funkcja ze zbioru S w zbiór T** jest szczególnym rodzajem relacji R na zbiorze $S \times T$, mianowicie jest relacją taką, że

dla każdego $x \in S$ istnieje dokładnie jeden $y \in T$ taki, że $(x,y) \in R$.

Ta charakteryzacja jest to po prostu inne sformułowanie definicji z § 1.3. Zatem funkcje są to relacje, dla których ma sens zapis funkcyjny: $f(x)$ jest jedynym elementem zbioru T takim, że para $(x, f(x))$ należy do R_f.

W przypadku $S = T$ mówimy, że podzbiór R zbioru $S \times S$ jest **relacją w zbiorze** S.

PRZYKŁAD 4 (a) Każdy zbiór S ma podstawową „relację równości": $E = \{(x, x)\colon\ x \in S\}$. Zatem dwa elementy zbioru S są ze sobą w tej relacji wtedy i tylko wtedy, gdy są identyczne. Relację tę normalnie oznaczamy znakiem $=$. Zatem $(x, y) \in E$ wtedy i tylko wtedy, gdy $x = y$.

(b) W zbiorze $\mathbb{R}$ znana relacja nierówności $\leq$ może być uważana za podzbiór R zbioru $\mathbb{R}\times\mathbb{R}$, mianowicie $R = \{(x, y)\colon\ x \leq y\}$. Ponieważ para $(x, y) \in R$ wtedy i tylko wtedy, gdy $x \leq y$, więc tę relację zazwyczaj oznaczamy przez $\leq$. Zauważmy, że znane własności

(Z) $x \leq x$ dla wszystkich $x \in \mathbb{R}$,
(AS) $x \leq y$ i $y \leq x$ implikują $x = y$,
(P) $x \leq y$ i $y \leq z$ implikują $x \leq z$,

mogą być zapisane w postaci

(Z) $(x, x) \in R$ dla wszystkich $x \in \mathbb{R}$,
(AS) $(x, y) \in R$ i $(y, x) \in R$ implikują $x = y$,
(P) $(x, y) \in R$ i $(y, z) \in R$ implikują $(x, z) \in R$.

Oznaczenia (Z), (AS) i (P) są skrótami określeń zwrotna, antysymetryczna i przechodnia, których będziemy używać w odniesieniu do dowolnych relacji w zbiorze S.

(c) Relacja nierówności ostrej $<$ w zbiorze $\mathbb{R}$ oczywiście też jest relacją i może być uważana za zbiór $R = \{(x, y)\colon\ x < y\}$. Ta relacja ma następujące własności:

(PZ) $x < x$ nie zachodzi dla żadnych $x \in \mathbb{R}$,
(P) $x < y$ i $y < z$ implikują $x < z$.

Mogą one być zapisane w postaci

(PZ) $(x, x) \notin R$ dla wszystkich $x \in \mathbb{R}$,
(P) $(x, y) \in R$ i $(y, z) \in R$ implikują $(x, z) \in R$.

(PZ) jest tu skrótem określenia przeciwzwrotna, a (P) nadal jest skrótem określenia przechodnia. ∎

Niech p będzie ustaloną liczbą całkowitą, większą niż 1. Weźmy liczby całkowite m i n. Mówimy, że liczba m **przystaje do liczby** n **modulo** p i piszemy $m \equiv n \pmod p$, gdy różnica

$m - n$ jest wielokrotnością p. Relację zdefiniowaną w ten sposób nazywamy **relacją kongruencji** w zbiorze liczb całkowitych $\mathbb{Z}$. Zbadamy ją dokładnie w § 3.6. W tym paragrafie zobaczymy, że

(Z) $m \equiv m \pmod{p}$ dla wszystkich $m \in \mathbb{Z}$,
(S) $m \equiv n \pmod{p}$ implikuje $n \equiv m \pmod{p}$,
(P) $m \equiv n \pmod{p}$ i $n \equiv r \pmod{p}$ implikują $m \equiv r \pmod{p}$.

Jeżeli przez R oznaczymy odpowiednią relację

$$R = \{(m,n) \in \mathbb{Z} \times \mathbb{Z}\colon\ m \equiv n \pmod{p}\},$$

to te własności można zapisać w postaci

(Z) $(m,m) \in R$ dla wszystkich $m \in \mathbb{Z}$,
(S) $(m,n) \in R$ implikuje $(n,m) \in R$,
(P) $(m,n) \in R$ i $(n,r) \in R$ implikują $(m,r) \in R$.

(Z), (S) i (P) są tu skrótami określeń relacji — zwrotna, symetryczna i przechodnia. Zauważmy, że własność zwrotności i przechodniości jest taka sama jak w przykładzie 4.

Ogólnie, mówimy, że relacja R w zbiorze S jest: **zwrotna**, **przeciwzwrotna**, **symetryczna**, **antysymetryczna** lub **przechodnia**, jeśli spełnia odpowiedni warunek:

(Z) $(x,x) \in R$ dla wszystkich $x \in S$,
(PZ) $(x,x) \notin R$ dla wszystkich $x \in S$,
(S) $(x,y) \in R$ implikuje $(y,x) \in R$ dla wszystkich $x, y \in S$,
(AS) $(x,y) \in R$ i $(y,x) \in R$ implikują $x = y$,
(P) $(x,y) \in R$ i $(y,z) \in R$ implikują $(x,z) \in R$.

Weźmy znów dowolną relację R na zbiorze $S \times T$, to znaczy $R \subseteq S \times T$. Relacją odwrotną $R^{\leftarrow}$ jest relacja na zbiorze $T \times S$ zdefiniowana wzorem

$$R^{\leftarrow} = \{(y,x) \in T \times S\colon\ (x,y) \in R\}.$$

Ponieważ każda funkcja $f\colon S \to T$ jest relacją, zawsze istnieje relacja odwrotna $f^{\leftarrow}$:

$$\text{jako relacja } f^{\leftarrow} = \{(y,x) \in T \times S\colon\ y = f(x)\}.$$

Relacja ta jest funkcją dokładnie wtedy, gdy f jest funkcją odwracalną, jak zostało to określone w § 1.4 i w tym przypadku mamy $f^{\leftarrow} = f^{-1}$.

PRZYKŁAD 5 (a) Przypomnijmy, że jeśli $f\colon S \to T$ jest funkcją i $A \subseteq S$, to obrazem zbioru A względem funkcji f jest

$$f(A) = \{f(x)\colon\ x \in A\} = \{y \in T\colon\ y = f(x) \text{ dla pewnego } x \in A\}.$$

Jeśli traktujemy funkcję f jako relację R_f, to ten zbiór jest równy

$$\{y \in T\colon\ (x, y) \in R_f \text{ dla pewnego } x \in A\}.$$

Podobnie, dla dowolnej relacji R na zbiorze $S \times T$ możemy zdefiniować

$$R(A) = \{y \in T\colon\ (x, y) \in R \text{ dla pewnego } x \in A\}.$$

Ponieważ $R^{\leftarrow}$ jest relacją na zbiorze $T \times S$, więc dla zbioru $B \subseteq T$ mamy również

$$\begin{aligned} R^{\leftarrow}(B) &= \{x \in S\colon\ (y, x) \in R^{\leftarrow} \text{ dla pewnego } y \in B\} \\ &= \{x \in S\colon\ (x, y) \in R \text{ dla pewnego } y \in B\}. \end{aligned}$$

Jeśli relacja R jest tak naprawdę relacją R_f dla funkcji f ze zbioru S w zbiór T, to mamy

$$\begin{aligned} {R_f}^{\leftarrow}(B) &= \{x \in S\colon\ y = f(x) \text{ dla pewnego } y \in B\} \\ &= \{x \in S\colon\ f(x) \in B\}, \end{aligned}$$

a to jest dokładnie definicja zbioru $f^{\leftarrow}(B)$ podana w § 1.4.

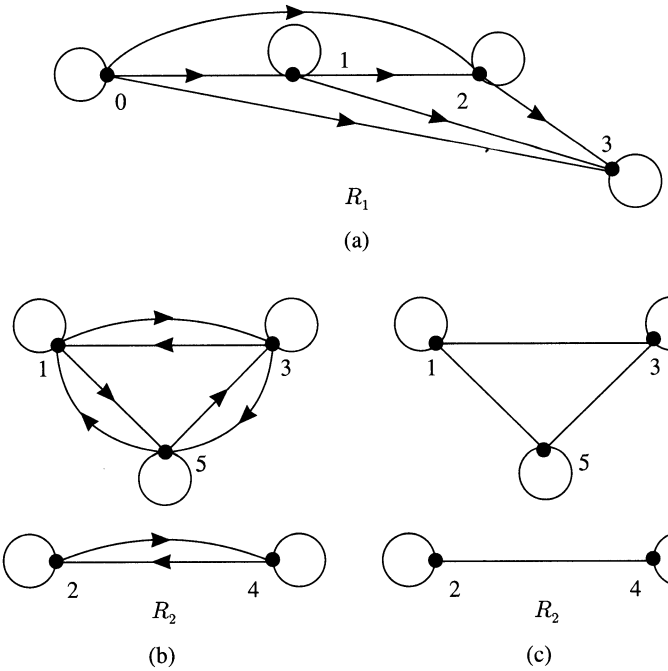
(b) Weźmy konkretny przykład do punktu (a). Niech S będzie zbiorem dostawców, a T zbiorem towarów i określmy, że $(x, y) \in R$, jeśli dostawca x sprzedaje towar y. Dla danego zbioru A dostawców zbiór $R(A)$ jest zbiorem towarów sprzedawanych przez przynajmniej jednego dostawcę ze zbioru A. Dla danego zbioru B towarów $R^{\leftarrow}(B)$ jest zbiorem dostawców, którzy sprzedają przynajmniej jeden towar ze zbioru B. Relacja R jest zbiorem R_f dla funkcji f ze zbioru S w zbiór T wtedy i tylko wtedy, gdy każdy dostawca sprzedaje dokładnie jeden towar. ∎

Jeżeli relacje dotyczą małych zbiorów, czasami wygodnie jest je narysować.

PRZYKŁAD 6 (a) Weźmy relację R_1 na zbiorze $\{0, 1, 2, 3\}$ określoną przez $\leq$. Zatem $(m, n) \in R_1$ wtedy i tylko wtedy, gdy $m \leq n$. Wykres relacji R_1 jest podany na rysunku 3.1(a). Zauważmy, że narysowaliśmy strzałki od m do n, jeśli tylko $(m, n) \in R_1$, jednakże nie narysowaliśmy ich na „pętlach" $0 \longrightarrow 0$, $1 \longrightarrow 1$ itd.

(b) Niech R_2 będzie relacją na zbiorze $\{1, 2, 3, 4, 5\}$ określoną tak, że $(m, n) \in R_2$ wtedy i tylko wtedy, gdy $m - n$ jest liczbą parzystą. Wykres tej relacji jest podany na rysunku 3.1(b).

(c) Wykres relacji odwrotnej ${R^{\leftarrow}}_1$ otrzymujemy odwracając wszystkie strzałki na rysunku 3.1(a). Pętle pozostają nie zmienione.



Rysunek 3.1

(d) Wykres relacji odwrotnej $R^{\leftarrow}{}_2$ też otrzymujemy odwracając wszystkie strzałki (na rysunku 3.1(b)), ale tym razem otrzymujemy ten sam wykres. Jest tak dlatego, że relacja R_2 jest symetryczna, a więc $R^{\leftarrow}{}_2 = R_2$; zob. ćwiczenie 11(a).

■

Jeśli relacja jest symetryczna, tak jak relacja R_2 na rysunku 3.1(b), to dla każdej strzałki istnieje strzałka odwrotna. Zatem nie warto rysować takich par strzałek. Równie dużo informacji daje wykres przedstawiony na rysunku 3.1(c). Tak jak w przypadku planów miast, strzałki oznaczają ulice jednokierunkowe a zwykłe linie oznaczają ulice dwukierunkowe.

ĆWICZENIA DO § 3.1

1. Dla następujących relacji w zbiorze $S = \{0, 1, 2, 3\}$ określ, które z własności (Z), (PZ), (S), (AS) i (P) spełniają te relacje:
(a) $(m, n) \in R_1$, jeśli $m + n = 3$,
(b) $(m, n) \in R_2$, jeśli $m - n$ jest liczbą parzystą,
(c) $(m, n) \in R_3$, jeśli $m \leq n$,
(d) $(m, n) \in R_4$, jeśli $m + n \leq 4$,

(e) $(m,n) \in R_5$, jeśli $max\{m,n\} = 3$.

2. Niech $A = \{0,1,2\}$. Każde z poniższych stwierdzeń określa relację R w zbiorze A w ten sposób, że $(m,n) \in R$, jeśli to stwierdzenie jest prawdziwe dla m i n. Zapisz każdą relację jako zbiór par uporządkowanych.

(a) $m \leq n$	(b) $m < n$	(c) $m = n$
(d) $mn = 0$	(e) $mn = m$	(f) $m + n \in A$
(g) $m^2 + n^2 = 2$	(h) $m^2 + n^2 = 3$	(i) $m = max\{n,1\}$

3. Które z relacji z ćwiczenia 2 są zwrotne, a które symetryczne?

4. W zbiorze $\mathbb{N}$ określone są następujące relacje dwuargumentowe:
(a) Zapisz relację dwuargumentową R_1 określoną wzorem $m + n = 5$ jako zbiór par uporządkowanych.
(b) Zrób to samo dla relacji R_2 określonej wzorem $\max\{m,n\} = 2$.
(c) Relacja dwuargumentowa R_3 określona wzorem $\min\{m,n\} = 2$ zawiera nieskończenie wiele par uporządkowanych. Wypisz pięć z nich.

5. Dla każdej relacji z ćwiczenia 4 określ, które z własności (Z), (PZ), (S), (AS) i (P) spełnia ta relacja.

6. Weźmy relację R w zbiorze $\mathbb{Z}$ określoną w następujący sposób: $(m,n) \in R$ wtedy i tylko wtedy, gdy $m^3 - n^3 \equiv 0 \pmod 5$. Które z własności (Z), (PZ), (S), (AS) i (P) ma ta relacja?

7. (a) Weźmy **relację pustą** $\emptyset$ na niepustym zbiorze S. Które z własności (Z), (PZ), (S), (AS) i (P) spełnia ta relacja?
(b) Powtórz ćwiczenie (a) dla **relacji uniwersalnej** $U = S \times S$ w zbiorze S.

8. Podaj przykład relacji, która jest:
(a) antysymetryczna i przechodnia, ale nie jest zwrotna,
(b) symetryczna, ale nie jest zwrotna ani przechodnia.

9. Niech R_1 i R_2 będą relacjami dwuargumentowymi w zbiorze S.
(a) Pokaż, że relacja $R_1 \cap R_2$ jest zwrotna, jeśli R_1 i R_2 są zwrotne.
(b) Pokaż, że relacja $R_1 \cap R_2$ jest symetryczna, jeśli R_1 i R_2 są symetryczne.
(c) Pokaż, że relacja $R_1 \cap R_2$ jest przechodnia, jeśli R_1 i R_2 są przechodnie.

10. Niech R_1 i R_2 będą relacjami dwuargumentowymi w zbiorze S.
(a) Czy jeśli relacje R_1 i R_2 są zwrotne, to relacja $R_1 \cup R_2$ musi być zwrotna?
(b) Czy jeśli relacje R_1 i R_2 są symetryczne, to relacja $R_1 \cup R_2$ musi być symetryczna?
(c) Czy jeśli relacje R_1 i R_2 są przechodnie, to relacja $R_1 \cup R_2$ musi być przechodnia?

11. Niech R będzie relacją dwuargumentową w zbiorze S.

(a) Udowodnij, że R jest relacją symetryczną wtedy i tylko wtedy, gdy $R = R^{\leftarrow}$.
(b) Udowodnij, że R jest relacją antysymetryczną wtedy i tylko wtedy, gdy $R \cap R^{\leftarrow} \subseteq E$, gdzie $E = \{(x, x)\colon\ x \in S\}$.

12. Niech R_1 i R_2 będą relacjami dwuargumentowymi na zbiorze $S \times T$.
(a) Pokaż, że $(R_1 \cup R_2)^{\leftarrow} = {R^{\leftarrow}}_1 \cup {R^{\leftarrow}}_2$.
(b) Pokaż, że $(R_1 \cap R_2)^{\leftarrow} = {R^{\leftarrow}}_1 \cap {R^{\leftarrow}}_2$.
(c) Pokaż, że jeśli $R_1 \subseteq R_2$, to ${R^{\leftarrow}}_1 \subseteq {R^{\leftarrow}}_2$.

13. Narysuj wykres każdej z relacji z ćwiczenia 1. Nie rysuj strzałek, jeśli relacja jest symetryczna.

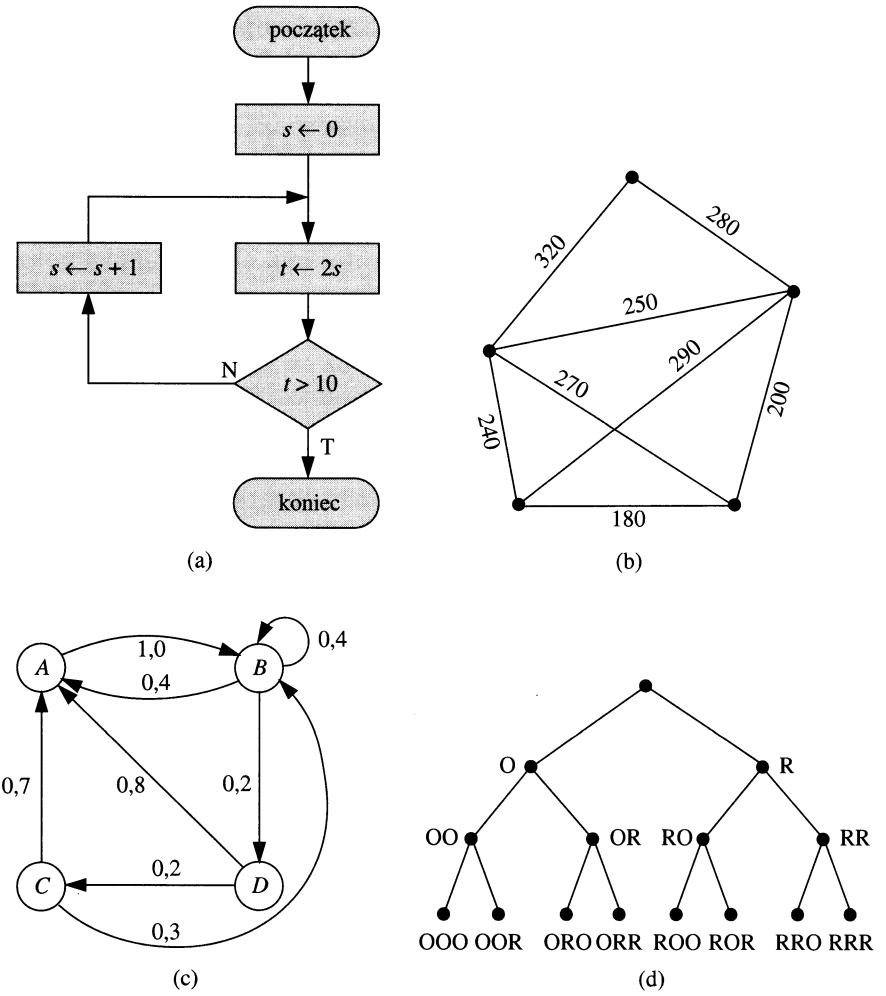
14. Narysuj wykres każdej z relacji z ćwiczenia 2. Nie rysuj strzałek, jeśli relacja jest symetryczna.

§ 3.2. Grafy i grafy skierowane

Idea grafu jako rysunku przedstawiającego funkcję jest nam dobrze znana. Słowa „graf" używa się też do opisu struktury innego rodzaju, która pojawia się w wielu różnych sytuacjach. Mówiąc ogólnie, te nowe grafy są wykresami, które odpowiednio zinterpretowane, mogą zawierać pewne informacje. Grafy, którymi my się zajmujemy, są jak mapy drogowe, rysunki obwodów lub schematy blokowe w tym sensie, że przedstawiają one połączenia lub relacje zachodzące między różnymi fragmentami wykresu.

Wykresy na rysunku 3.2 są bardzo różnorodne. Rysunek 3.2(a) przedstawia zwykły schemat blokowy. Rysunek 3.2(b) mógłby przedstawiać pięć magazynów firmy przewozowej i drogi dla ciężarówek między tymi magazynami wraz z podanymi odległościami. Rysunek 3.2(c) mógłby powiedzieć nam o tym, jakie jest prawdopodobieństwo, że szczur umieszczony w jednej z czterech klatek przebiegnie do którejś z pozostałych trzech klatek lub pozostanie w swojej własnej. Rysunek 3.2(d) mógłby przedstawiać możliwe wyniki powtarzanego eksperymentu, takiego jak rzut monetą (orły i reszki). Wykresy na rysunku 3.1 w § 3.1 mówią, które pary wierzchołków należą do relacji R_1 i R_2. Co mają wspólnego te wszystkie wykresy? Każdy z nich składa się ze zbioru obiektów: prostokątów, okręgów czy kropek oraz pewnych linii, być może krzywych, które łączą te obiekty. Czasami te linie są skierowane, to znaczy mają strzałki.

Istotnymi elementami grafu skierowanego są obiekty i linie skierowane. Dokładniej, **graf skierowany** (digraf) G składa się z dwóch zbiorów, niepustego zbioru $V(G)$ **wierzchołków grafu**



Rysunek 3.2

G i zbioru $E(G)$ **krawędzi grafu** G oraz z funkcji γ (mała grecka litera gamma) ze zbioru $E(G)$ w zbiór $V(G) \times V(G)$. Jeśli e jest krawędzią grafu G i $\gamma(e) = (p, q)$, to p nazywamy **początkiem krawędzi** e, a q **końcem krawędzi** e i mówimy, że e **biegnie od** p **do** q. Definicja ta ma również sens, jeśli zbiory $V(G)$ lub $E(G)$ są nieskończone, ale ponieważ w naszych zastosowaniach mamy do czynienia ze zbiorami skończonymi, przyjmiemy w tym paragrafie, że zbiory $V(G)$ i $E(G)$ są skończone.

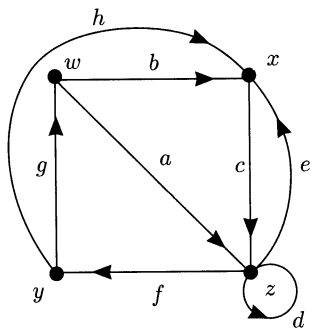
Rysunkiem grafu skierowanego G jest wykres składający się z punktów, odpowiadających elementom zbioru $V(G)$ oraz strzałek, odpowiadających elementom zbioru $E(G)$, takich że, je-

śli $\gamma(e) = (p, q)$, to strzałka odpowiadająca e biegnie od punktu oznaczonego przez p do punktu oznaczonego przez q.

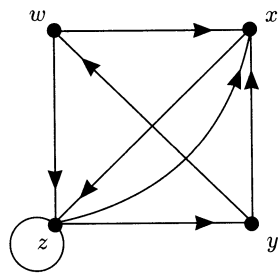
PRZYKŁAD 1 Weźmy graf skierowany G, w którym dane są dwa zbiory: $V(G) = \{w, x, y, z\}$ oraz $E(G) = \{a, b, c, d, e, f, g, h\}$, a funkcja γ jest podana w tabeli na rysunku 3.3(a). Oba wykresy, na rysunkach 3.3(b) i 3.3(c) są rysunkami grafu G. Na rysunku 3.3(b) oznaczyliśmy strzałki literami, aby wyraźniej pokazać przekształcenie zbioru $E(G)$. Na rysunku 3.3(c) oznaczyliśmy po prostu punkty i zaznaczyliśmy łączące je strzałki. Nie prowadzi to do nieporozumień, ponieważ w tym przypadku nie ma **krawędzi wielokrotnych**, tzn. jest co najwyżej jedna krawędź o danym początku i końcu. Innymi słowy, funkcja γ jest funkcją różnowartościową. Zauważmy, że nie narysowaliśmy strzałki na krawędzi d, ponieważ z jest zarówno początkiem, jak i końcem tej krawędzi.■

e	$\gamma(e)$
a	(w, z)
b	(w, x)
c	(x, z)
d	(z, z)
e	(z, x)
f	(z, y)
g	(y, w)
h	(y, x)

(a)



(b)



(c)

Rysunek 3.3

Jeśli γ: $E(G) \to V(G) \times V(G)$ jest przekształceniem różnowartościowym, to możemy utożsamiać krawędzie e z ich obrazami $\gamma(e)$ w zbiorze $V(G) \times V(G)$ i traktować zbiór $E(G)$ jako podzbiór zbioru $V(G) \times V(G)$. Niektórzy definiują grafy skierowane jako grafy, dla których $E(G) \subseteq V(G) \times V(G)$, a bardziej ogólne grafy skierowane, które my tutaj rozważamy, nazywają „multigrafami skierowanymi".

Mając dany rysunek grafu G, możemy odtworzyć sam graf skierowany G, ponieważ strzałki mówią nam wszystko o γ. Będziemy zazwyczaj opisywać grafy skierowane za pomocą rysunku a nie tabeli funkcji γ, choć ten sposób opisu został wybrany ze względu na naszą wygodę. Komputer zapamiętuje grafy skierowane, zapamiętując w jakiś sposób funkcję γ.

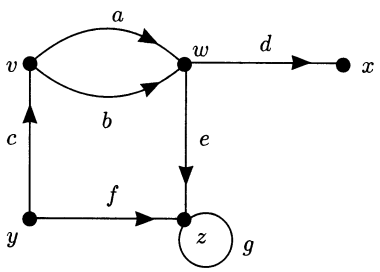
Wiele ważnych pytań związanych z grafami skierowanymi można sformułować używając ciągów krawędzi prowadzących od jednego wierzchołka do drugiego. **Drogą** w grafie skierowanym G nazywamy ciąg krawędzi taki, że koniec jednej krawędzi jest początkiem następnej. Zatem, jeśli $e_1, \ldots, e_n$ należą do zbioru $E(G)$, to $e_1e_2 \ldots e_n$ jest drogą, o ile istnieją wierzchołki $x_1, x_2, \ldots, x_n, x_{n+1}$ takie, że $\gamma(e_i) = (x_i, x_{i+1})$ dla $i = 1, 2, \ldots, n$. Mówimy, że $e_1e_2 \ldots e_n$ jest **drogą** (ścieżką) **długości** n od wierzchołka x_1 do wierzchołka x_{n+1}. Droga jest **zamknięta**, jeśli $x_1 = x_{n+1}$.

PRZYKŁAD 2 W grafie skierowanym G na rysunku 3.3 ciąg $fgae$ jest drogą długości 4 od wierzchołka z do wierzchołka x. Ciągi $cecec$ i $fgafhc$ również są drogami, ale fa nie jest drogą, ponieważ $\gamma(f) = (z, y)$, $\gamma(a) = (w, z)$ i $y \neq w$. Drogi $fgafhc$, $cece$ i d są zamknięte; drogi $fhce$ i df nie są. ∎

Droga $e_1 \ldots e_n$, gdzie $\gamma(e_i) = (x_i, x_{i+1})$, łączy ciąg wierzchołków $x_1x_2 \ldots x_nx_{n+1}$. Jeśli każda krawędź e_i jest jedyną krawędzią od x_i do x_{i+1}, to ten ciąg wierzchołków jednoznacznie określa drogę i możemy opisać tę drogę, wypisując po prostu po kolei te wierzchołki.

PRZYKŁAD 3 (a) Na rysunku 3.3 droga $fgae$ wyznacza ciąg wierzchołków $zywzx$. Zauważmy, że ten ciąg wierzchołków sam określa drogę. Droga może być odtworzona z ciągu $zywzx$ na podstawie rysunku 3.3(b) lub 3.3(c) albo przy użyciu tabeli funkcji γ z rysunku 3.3(a). Ponieważ ten graf skierowany nie ma krawędzi wielokrotnych, wszystkie jego drogi są określone przez ich ciągi wierzchołków.

(b) Dla grafu skierowanego z rysunku 3.4 ciąg wierzchołków $yzzz$ odpowiada tylko drodze fgg, a ciąg $yvwz$ należy zarówno do drogi cae, jak i drogi cbe. ∎



Rysunek 3.4

Drogę zamkniętą długości co najmniej 1 z ciągiem wierzchołków $x_1x_2 \ldots x_nx_1$ nazywamy **cyklem**, jeśli wszystkie wierzchołki $x_1, x_2, \ldots, x_n$ są różne. Język teorii grafów nie ma standardowej wersji — niektórzy autorzy cyklem nazywają drogę zamkniętą. Graf skierowany nie mający cykli nazywamy **grafem acyklicznym**. Droga jest acykliczna, jeśli graf skierowany zawierający jej wierzchołki i krawędzie jest acykliczny.

PRZYKŁAD 4 Na rysunku 3.3 droga afg jest cyklem, ponieważ jej ciągiem wierzchołków jest $wzyw$. Podobnie, drogi cfh i $cfgb$ o ciągach wierzchołków $xzyx$ i $xzywx$ są cyklami. Krótka droga ce i pętla d również są cyklami, ponieważ ich ciągami wierzchołków są odpowiednio xzx oraz zz. Droga $cfgae$ nie jest cyklem, ponieważ jej ciągiem wierzchołków jest $xzywzx$ i wierzchołek z się powtarza.■

Jak widzieliśmy w przykładzie 6 w § 3.1, relacja R w zbiorze S określa w naturalny sposób pewien graf skierowany G: przyjmijmy $V(G) = S$ i przyjmijmy, że wierzchołek v jest połączony krawędzią z wierzchołkiem w, jeśli tylko $(v, w) \in R$. Możemy odwrócić to postępowanie. Dla danego grafu skierowanego G i wierzchołków v i w w zbiorze $V(G)$ mówimy, że wierzchołek v jest **sąsiedni** w stosunku do wierzchołka w, jeśli istnieje krawędź w $E(G)$ od v do w. **Relacja sąsiedztwa** A w zbiorze wierzchołków $V(G)$ jest określona następująco: $(u, v) \in A$ wtedy i tylko wtedy, gdy wierzchołek v jest sąsiedni z wierzchołkiem w. Relacja A jest określona zupełnie ogólnie i nie musi mieć żadnych specjalnych własności. Będzie ona zwrotna tylko wtedy, gdy G ma pętlę w każdym wierzchołku i symetryczna tylko wtedy, gdy istnieje krawędź od wierzchołka v do wierzchołka w, jeśli tylko istnieje krawędź od wierzchołka w do wierzchołka v.

PRZYKŁAD 5 (a) Weźmy graf skierowany z rysunku 3.3. Relacja sąsiedztwa składa się z par uporządkowanych (w, z), (w, x), (x, z), (z, z), (z, x), (z, y), (y, w) i (y, x). Innymi słowy, relacja A zawiera wartości funkcji γ wypisane na rysunku 3.3(a). Ogólnie, $A = \gamma(E(G)) \subseteq V(G) \times V(G)$.

(b) Relacja sąsiedztwa A dla grafu skierowanego z rysunku 3.4 zawiera pary uporządkowane (v, w), (w, x), (w, z), (z, z), (y, z) i (y, v). Nie możemy odtworzyć grafu skierowanego z A, ponieważ A nie podaje informacji o krawędziach wielokrotnych. Ponieważ (v, w) należy do A, wiemy, że graf skierowany ma co najmniej jedną krawędź z wierzchołka v do wierzchołka w, ale nie możemy powiedzieć, że ma on dokładnie dwie krawędzie.

(c) Graf skierowany otrzymany z rysunku 3.4 przez usunięcie krawędzi a ma tę samą relację sąsiedztwa co początkowy graf skierowany. ∎

Ostatni przykład pokazuje, że różne grafy skierowane mogą mieć tę samą relację sąsiedztwa. Jednakże, jeśli ograniczymy nasze zainteresowanie do grafów skierowanych bez krawędzi wielokrotnych, to istnieje wzajemnie jednoznaczna odpowiedniość między grafami skierowanymi i relacjami.

Jeśli opuścimy strzałki na naszych krawędziach, tzn. kierunki, otrzymamy to, co nazywamy grafami (nieskierowanymi). Rysunki 3.2(b) i 3.2(d) w tym paragrafie i rysunek 3.1(c) w § 3.1 przedstawiają grafy. Grafy mogą mieć krawędzie wielokrotne; aby zobaczyć przykład, opuść po prostu strzałki na rysunku 3.4. Idee i pojęcia służące do badania grafów nieskierowanych są podobne do tych, które dotyczyły grafów skierowanych.

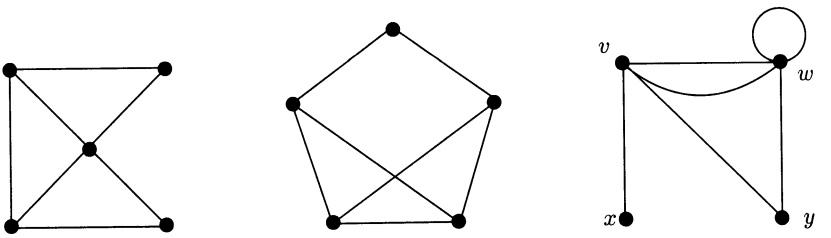
Zamiast uporządkowanych par wierzchołków, które były przypisane krawędziom w grafach skierowanych, krawędzie z nie zaznaczonym kierunkiem mają przypisany nieuporządkowany zbiór wierzchołków. Naśladując to, co zrobiliśmy w przypadku grafów skierowanych, mówimy, że graf (nieskierowany) składa się z dwóch zbiorów, zbioru $V(G)$ wierzchołków grafu G i zbioru $E(G)$ krawędzi grafu G oraz z funkcji γ ze zbioru $E(G)$ w zbiór $\{\{u, v\}\colon\ u, v \in V(G)\}$ wszystkich podzbiorów jedno- lub dwuelementowych zbioru $V(G)$. Dla danej krawędzi e ze zbioru $E(G)$ elementy $\gamma(e)$ nazywamy **wierzchołkami** e lub **końcami** e; mówimy, że krawędź e **łączy** swoje końce. **Pętlą** jest krawędź z tylko jednym końcem. Różne krawędzie e i f takie, że $\gamma(e) = \gamma(f)$ nazywamy **krawędziami wielokrotnymi**.

Te dokładne definicje pokazują wyraźnie, że komputer mógłby traktować graf jako dwa zbiory wraz z funkcją γ określającą końce krawędzi.

To, co przed chwilą opisaliśmy, niektórzy autorzy nazywają **multigrafem**, a termin „graf" rezerwują dla takich grafów, które nie mają pętli czy krawędzi wielokrotnych. Jeśli nie ma krawędzi wielokrotnych, to funkcja γ jest różnowartościowa i zbiory $\gamma(e)$ jednoznacznie określają krawędzie e. To znaczy, istnieje tylko jedna krawędź dla każdego zbioru $\gamma(e)$. W tym przypadku często obywamy się bez zbioru $E(G)$ i funkcji γ i po prostu wypisujemy krawędzie jako zbiory, tak jak $\{u, v\}$ czy $\{u\}$ lub jako ciągi wierzchołków, tak jak uv, vu czy $u\,u$. Zatem będziemy zazwyczaj pisać $e = \{u, v\}$; będziemy też czasem pisać $e = \{u, u\}$ zamiast $e = \{u\}$, jeśli e jest pętlą w wierzchołku u.

Rysunkiem grafu G jest wykres składający się z punktów odpowiadających wierzchołkom zbioru G i łuków czy odcinków odpowiadających krawędziom, tak, że jeśli $\gamma(e) = \{u, v\}$, to łuk dla krawędzi e łączy punkty oznaczone literami u i v.

PRZYKŁAD 6 (a) Grafy na rysunkach 3.5(a) i 3.5(b) mają 5 wierzchołków i 7 krawędzi. Punkt przecięcia dwóch odcinków na rysunku 3.5(b) nie ma znaczenia i jest osobliwością naszego rysunku. Graf na rysunku 3.5(c) ma 4 wierzchołki i 6 krawędzi. Ma on krawędzie wielokrotne łączące v i w oraz pętlę w wierzchołku w.



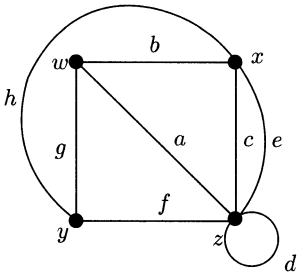
(a) (b) (c)

Rysunek 3.5

(b) Jeśli opuścimy strzałki na krawędziach na rysunku 3.3, otrzymamy graf przedstawiony na rysunku 3.6(a). Tabela funkcji γ dla tego grafu jest podana na rysunku 3.6(b). Graf ten ma krawędzie wielokrotne c i e łączące wierzchołki x i z oraz ma pętlę d w wierzchołku z. Graf narysowany na rysunku 3.6(c) nie ma krawędzi wielokrotnych, więc dla tego grafu powiedzenie „krawędź $\{x, z\}$" jest jednoznaczne. Tego samego zwrotu nie można byłoby użyć do grafu z rysunku 3.6(a). ∎

Ciąg krawędzi, które łączą się ze sobą, nazywamy **drogą**. Aby zilustrować ten pomysł, przedstawiliśmy ponownie rysunek 3.2(c) na rysunku 3.7(a) i oznaczyliśmy krawędzie, jak też wierzchołki. Przykładami dróg są tutaj: $dbfe$ (rysunek 3.7(b)) i $cfeba$ (rysunek 3.7(c)). Zauważmy, że sam rysunek nie mówi nam, którą drogę mamy na myśli: rysunek 3.7(c) jest też rysunkiem dróg $bafec$ i $cafeb$. Krawędzie w drogach mogą się powtarzać: $babefaab$ jest drogą. **Długością** drogi jest liczba krawędzi w tej drodze. Zatem droga $babefaab$ ma długość 8.

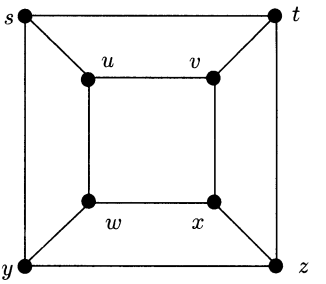
Kolejne dwie krawędzie w drodze muszą mieć wspólny wierzchołek. Zatem droga wyznacza ciąg wierzchołków. Ciągi wierzchołków dla dróg omówionych wyżej są pokazane na rysunku 3.8.



(a)

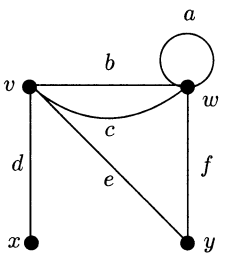
e	$\gamma(e)$
a	$\{w, z\}$
b	$\{w, x\}$
c	$\{x, z\}$
d	$\{z\}$ lub $\{z, z\}$
e	$\{x, z\}$
f	$\{y, z\}$
g	$\{w, y\}$
h	$\{x, y\}$

(b)

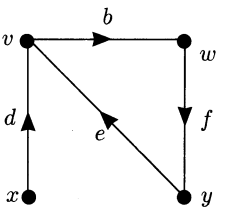


(c)

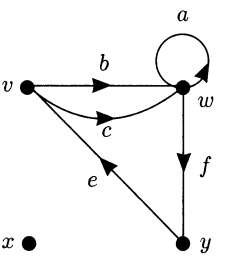
Rysunek 3.6



(a)



(b)



(c)

Rysunek 3.7

droga	ciąg jej wierzchołków
$dbfe$	$xvwyv$
$cfeba$	$vwyvww$
$bafec$	$vwwyvw$
$cafeb$	$vwwyvw$
$babefaab$	$vwwvywwwv$

Rysunek 3.8

Zauważmy, że liczba wierzchołków w takim ciągu wierzchołków jest o 1 większa od liczby krawędzi w drodze. Kiedy droga zawiera pętlę, jej wierzchołek występuje dwukrotnie w tym ciągu wierzchołków. Ciągi wierzchołków nie rozróżniają krawędzi wielokrotnych, zatem różne drogi, takie jak $bafec$ i $cafeb$ dają te same ciągi wierzchołków. Jeśli graf nie ma krawędzi wielokrotnych lub

wielokrotnych pętli, to ciąg wierzchołków wyznacza jednoznacznie drogę. W takim przypadku krawędzie mogą być opisane przez podanie obu wierzchołków, które łączą, a więc możemy opisać drogę za pomocą ciągu jej wierzchołków.

Ogólnie **drogą długości** n od wierzchołka u do wierzchołka v nazywamy ciąg $e_1 \dots e_n$ wraz z ciągiem $x_1 \dots x_{n+1}$ wierzchołków taki, że $\gamma(e_i) = \{x_i, x_{i+1}\}$ dla $i = 1, \dots, n$ oraz $x_1 = u$, $x_{n+1} = v$. Nie będziemy się zajmować drogami długości 0. Jeśli $e_1 e_2 \dots e_n$ wraz z ciągiem wierzchołków $x_1 x_2 \dots x_{n+1}$ jest drogą od wierzchołka u do wierzchołka v, to $e_n \dots e_2 e_1$ wraz z ciągiem wierzchołków $x_{n+1} x_n \dots x_1$ jest drogą od v do u. Obie te drogi możemy nazywać **drogami między wierzchołkami** u i v. Jeśli $u = v$, to drogę nazywamy **zamkniętą**.

Ciąg krawędzi występujących w drodze zazwyczaj określa ciąg wierzchołków, a więc będziemy czasami używali określeń takich jak „droga $e_1 e_2 \dots e_n$" bez wymieniania wierzchołków. Jest tu pewna niejednoznaczność, gdyż drogi takie jak ee, eee itd. nie precyzują, od którego końca krawędzi e zaczynamy. Podobny problem pojawia się w przypadku drogi ef, gdy e i f są krawędziami wielokrotnymi. Jeśli graf nie ma krawędzi wielokrotnych, to ciąg wierzchołków jednoznacznie określa ciąg krawędzi. W takiej sytuacji lub wtedy, gdy konkretny wybór krawędzi jest nieistotny, zazwyczaj opisujemy drogi za pomocą ciągów wierzchołków.

Tak jak w przypadku grafów skierowanych, możemy dla grafu nieskierowanego zdefiniować **relację sąsiedztwa** A: $(u, v) \in A$, gdy zbiór $\{u, v\}$ jest krawędzią w grafie. Do końca tego paragrafu A będzie oznaczać relację sąsiedztwa dla grafu skierowanego lub grafu nieskierowanego. Aby z relacji A otrzymać relację przechodnią, musimy rozważać ciągi krawędzi: od u_1 do u_2, od u_2 do u_3 itd. Otrzymujemy w ten sposób pojęcie osiągalności. Definiujemy **relację osiągalności** R w zbiorze $V(G)$ w następujący sposób:

$(v, w) \in R$, jeżeli istnieje

w G droga długości co najmniej 1 od v do w.

Wówczas relacja R jest przechodnia. Ponieważ żądamy, aby wszystkie drogi były długości co najmniej 1, relacja R może nie być zwrotna.

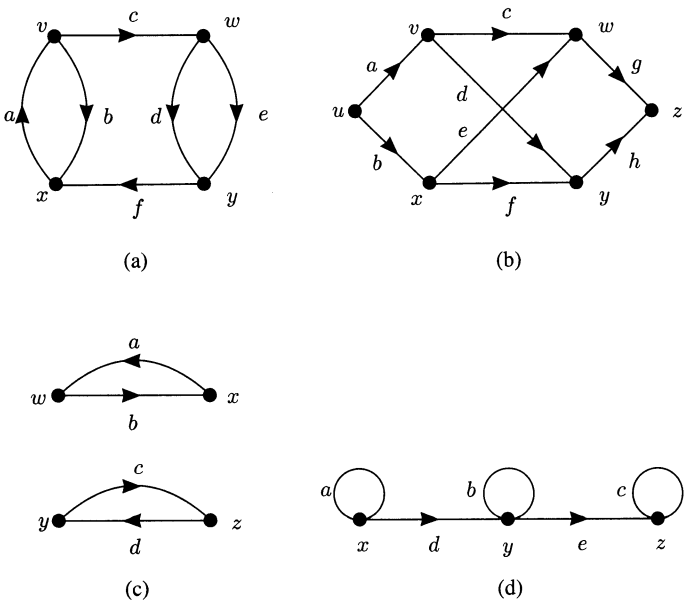
PRZYKŁAD 7 (a) Każdy wierzchołek grafu skierowanego na rysunku 3.3 jest osiągalny z każdego innego wierzchołka. Zatem relacja osiągalności R zawiera wszystkie możliwe pary uporządkowane; jest to tak zwana relacja uniwersalna.

(b) Relacja osiągalności dla grafu skierowanego na rysunku 3.4 składa się z par: (v,w), (v,x), (v,z), (w,x), (w,z), (y,v), (y,w), (y,x), (y,z) i (z,z). Zauważmy, że każdy wierzchołek może być osiągalny z wierzchołka y, poza samym y. Podobnie, z jest jedynym wierzchołkiem, który może być osiągalny z siebie samego.

(c) Wszystkie grafy na rysunkach 3.5 i 3.6 są spójne, w tym sensie, że każdy wierzchołek jest osiągalny z każdego innego wierzchołka. Zatem w każdym z tych przypadków relacja osiągalności jest relacją uniwersalną. ∎

ĆWICZENIA DO § 3.2

1. Podaj tabelę funkcji γ dla każdego z grafów skierowanych z rys. 3.9.

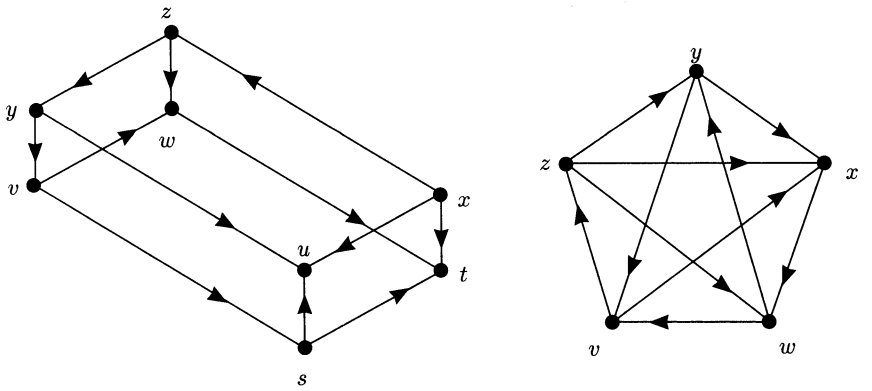


Rysunek 3.9

2. Zrób rysunek grafu skierowanego G, w którym zbiór wierzchołków $V(G) = \{w,x,y,z\}$, zbiór krawędzi $E(G) = \{a,b,c,d,e,f,g\}$, a funkcja γ podana jest w następującej tabeli:

e	a	b	c	d	e	f	g
$\gamma(e)$	(x,w)	(w,x)	(x,x)	(w,z)	(w,y)	(w,z)	(z,y)

3. Które z podanych ciągów wierzchołków opisują drogi w grafie skierowanym przedstawionym na rysunku 3.10(a)?



Rysunek 3.10

(a) $zyvwt$, (b) $xzwt$, (c) $vstx$, (d) $zysu$, (e) $xzyvs$, (f) $suxt$.

4. Znajdź długość najkrótszej drogi z x do w w grafie skierowanym pokazanym na rysunku 3.10(a).

5. Weźmy graf skierowany przedstawiony na rysunku 3.10(b). Opisz drogę acykliczną
(a) od x do y, (b) od y do z, (c) od v do w, (d) od x do z, (e) od z do v.

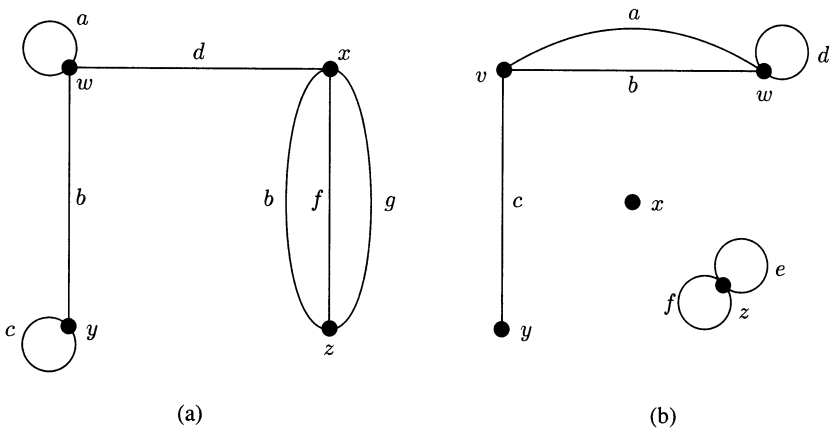
6. Są cztery grupy krwi: A, B, AB i 0. Grupa 0 może być podawana każdemu, grupy A i B mogą być podawane osobom mającym grupę AB lub A lub B odpowiednio, a grupa AB może być podawana tylko osobom z grupą AB. Narysuj graf skierowany, który przedstawia te informacje. Czy ten graf jest acykliczny?

7. Podaj przykład grafu skierowanego o wierzchołkach x, y i z, w którym jest cykl z wierzchołkami x i y i inny cykl z wierzchołkami y i z, ale nie ma cyklu mającego wierzchołki x i z.

8. Podaj relacje osiągalności dla grafów skierowanych z rysunku 3.9 (a), (c) i (d).

9. (a) Które z następujących par należą do relacji osiągalności dla grafu skierowanego z rysunku 3.9(b): (v,u), (v,v), (v,w), (v,x), (v,y), (v,z)?
(b) Które z następujących par należą do relacji osiągalności dla grafu skierowanego z rysunku 3.10(a): (v,s), (v,t), (v,u), (v,v), (v,w), (v,x), (v,y), (v,z)?
(c) Które z następujących par należą do relacji osiągalności dla grafu skierowanego z rysunku 3.10(b): (v,v), (v,w), (v,x), (v,y), (v,z)?

10. Które z następujących ciągów wierzchołków odpowiadają drogom w grafie z rysunku 3.11(a):
(a) zxw, (b) $wxzxwyww$, (c) $wwxz$, (d) $wxzz$, (e) $zxwyywz$, (f) wxw, (g) $yyww$?



(a) (b)

Rysunek 3.11

11. Podaj długość każdej drogi z ćwiczenia 10.

12. Które drogi z ćwiczenia 10 są drogami zamkniętymi?

13. Wypisz krawędzie wielokrotne w grafach z rysunku 3.11.

14. Ile jest pętli w każdym z grafów na rysunku 3.11?

15. Podaj relację sąsiedztwa A oraz relację osiągalności R dla każdego z grafów z rysunku 3.11.

16. Dla grafu przedstawionego na rysunku 3.11(a) podaj przykład każdej z następujących dróg. Sprawdź, czy podałeś ciąg krawędzi i ciąg wierzchołków.
(a) droga długości 2 od w do z,
(b) droga długości 4 od z do z,
(c) droga długości 5 od z do z,
(d) droga długości 3 od w do x.

17. Dla grafu przedstawionego na rysunku 3.11(b) podaj przykład każdej z następujących dróg. Sprawdź, czy podałeś ciąg krawędzi i ciąg wierzchołków.
(a) droga długości 3 od y do w,
(b) droga długości 3 od v do y,
(c) droga długości 4 od v do y,
(d) droga długości 3 od z do z.

§ 3.3. Macierze

W paragrafie 3.2 widzieliśmy bliski związek między grafami skierowanymi i relacjami. Macierze, główny temat tego paragrafu, są ważnym narzędziem do opisywania zarówno grafów skierowanych, jak i relacji, a także mają wiele innych zastosowań. Zobaczymy, że tematy pierwszych trzech paragrafów tego rozdziału pokazują nam trzy różne sposoby patrzenia na struktury równoważne: relacje, grafy skierowane bez krawędzi wielokrotnych i macierze o wyrazach 0 i 1.

Ogólnie, **macierz** jest tablicą prostokątną. Tradycyjnie używa się wielkich liter, takich jak $\mathbf{A}$ na oznaczenie macierzy. Jeśli przez a_{ij} oznaczymy wyraz stojący w i-tym wierszu i j-ej kolumnie, to możemy napisać:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix},$$

lub po prostu $\mathbf{A} = [a_{ij}]$. Macierz ta ma m poziomych wierszy oraz n pionowych kolumn i jest nazywana **macierzą wymiaru** $m \times n$. Jeśli kiedykolwiek w oznaczeniu macierzy występują podwójne indeksy, to zawsze numer wiersza poprzedza numer kolumny. Czasami oznaczamy wyraz w i-tym wierszu i j-ej kolumnie przez $\mathbf{A}[i, j]$; oznaczenie to jest stosowane w informatyce, ponieważ unika się wtedy indeksów. W tej książce wyrazy macierzy będą liczbami rzeczywistymi, chyba że wyraźnie zaznaczymy, że jest inaczej.

PRZYKŁAD 1 (a) Macierz

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 & 3 & 2 \\ 1 & -2 & 1 & -1 & 3 \\ 3 & 0 & 1 & 2 & -3 \end{bmatrix}$$

jest macierzą wymiaru 3×5. Jeśli oznaczymy tę macierz przez $\mathbf{A} = [a_{ij}]$, to $a_{11} = 2$, $a_{31} = 3$, $a_{13} = 0$, $a_{35} = -3$ itd. Jeśli oznaczymy ją przez $\mathbf{A}[i, j]$, to $\mathbf{A}[1, 2] = -1$, $\mathbf{A}[2, 1] = 1$, $\mathbf{A}[2, 2] = -2$ itd.

(b) Jeśli $\mathbf{B}$ jest macierzą wymiaru 3×4, określoną wzorem $\mathbf{B}[i, j] = i - j$, to $\mathbf{B}[1, 1] = 1 - 1 = 0$, $\mathbf{B}[1, 2] = 1 - 2 = -1$ itd.,

a zatem

$$\mathbf{B} = \begin{bmatrix} 0 & -1 & -2 & -3 \\ 1 & 0 & -1 & -2 \\ 2 & 1 & 0 & -1 \end{bmatrix}. \quad \blacksquare$$

Macierzy używa się we wszystkich naukach korzystających z matematyki. Ponieważ są one ściśle związane z relacjami, są wygodnym narzędziem do przechowywania informacji wiążacych dwa zbiory danych, mają więc duże zastosowanie w naukach ekonomicznych i w informatyce. Pojawiają się one również przy rozwiązywaniu układów równań liniowych. Wiele zjawisk fizycznych w przyrodzie daje się opisać przynajmniej w przybliżony sposób za pomocą macierzy. Ponadto, zbiór macierzy wymiaru $n \times n$ ma bardzo bogatą strukturę algebraiczną, którą warto zajmować się dla niej samej i która jest źródłem inspiracji w badaniu bardziej abstrakcyjnych struktur algebraicznych. W tym i w następnym paragrafie wprowadzimy różne działania algebraiczne na macierzach.

Jeśli m i n są liczbami całkowitymi dodatnimi, to przez $\mathfrak{M}_{m,n}$ oznaczamy zbiór wszystkich macierzy wymiaru $m \times n$. Dwie macierze $\mathbf{A}$ i $\mathbf{B}$ ze zbioru $\mathfrak{M}_{m,n}$ są równe, jeśli odpowiadające sobie wyrazy tych macierzy są równe, tzn. $\mathbf{A} = \mathbf{B}$, jeśli $a_{ij} = b_{ij}$ dla wszystkich i i j takich, że $1 \leq i \leq m$ i $1 \leq j \leq n$. Macierze mające tę samą liczbę wierszy i kolumn nazywamy **macierzami kwadratowymi**. Zatem $\mathbf{A}$ jest macierzą kwadratową, jeśli $\mathbf{A}$ należy do zbioru $\mathfrak{M}_{n,n}$ dla pewnego $n \in \mathbb{P}$. **Macierzą transponowaną** $\mathbf{A}^T$ do macierzy $\mathbf{A} = [a_{ij}]$ ze zbioru $\mathfrak{M}_{m,n}$ jest macierz należąca do $\mathfrak{M}_{n,m}$, która w i-tym wierszu i j-ej kolumnie ma wyraz a_{ji}. Inaczej mówiąc, $\mathbf{A}^T[i,j] = \mathbf{A}[j,i]$. Na przykład, jeśli

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 & 4 \\ 3 & 2 & -1 & 2 \\ 4 & 0 & 1 & 3 \end{bmatrix}, \quad \text{to} \quad \mathbf{A}^T = \begin{bmatrix} 2 & 3 & 4 \\ -1 & 2 & 0 \\ 0 & -1 & 1 \\ 4 & 2 & 3 \end{bmatrix}.$$

Pierwszy wiersz macierzy $\mathbf{A}$ staje się pierwszą kolumną macierzy $\mathbf{A}^T$ itd.

Macierze, które mają tylko jeden wiersz, tzn. macierze wymiaru $1 \times n$ często nazywamy **wektorami wierszowymi**, natomiast macierze, które mają tylko jedną kolumnę, to znaczy macierze wymiaru $m \times 1$ nazywamy **wektorami kolumnowymi**. Macierz transponowana do wektora wierszowego jest wektorem kolumnowym, a macierz transponowana do wektora kolumnowego jest wektorem wierszowym. Zatem $[2 \quad 4 \quad -3 \quad -1]$ jest

wektorem wierszowym, a jego macierz transponowana

$$\begin{bmatrix} 2 \\ 4 \\ -3 \\ -1 \end{bmatrix}$$

jest wektorem kolumnowym. Czasami traktujemy macierz wymiaru $m \times n$ jako macierz złożoną z m wektorów wierszowych lub z n wektorów kolumnowych. Dwie macierze $\mathbf{A}$ i $\mathbf{B}$ możemy dodać do siebie, jeśli mają te same wymiary, tzn. jeśli należą do tego samego zbioru $\mathfrak{M}_{m,n}$. Wtedy **sumę macierzy** otrzymujemy dodając do siebie odpowiednie wyrazy. Dokładniej, jeśli obie macierze $\mathbf{A} = [a_{ij}]$ i $\mathbf{B} = [b_{ij}]$ należą do zbioru $\mathfrak{M}_{m,n}$, to sumą macierzy $\mathbf{A} + \mathbf{B}$ jest macierz $\mathbf{C} = [c_{ij}]$ należąca do $\mathfrak{M}_{m,n}$, określona wzorem

$$c_{ij} = a_{ij} + b_{ij} \quad \text{dla} \quad 1 \leq i \leq m \text{ i } 1 \leq j \leq n.$$

Równoważnie, możemy zdefiniować

$$(\mathbf{A} + \mathbf{B})[i, j] = \mathbf{A}[i, j] + \mathbf{B}[i, j] \quad \text{dla} \quad 1 \leq i \leq m \text{ i } 1 \leq j \leq n.$$

Ponieważ m lub n może być równe 1, więc ta definicja dotyczy zwłaszcza wektorów wierszowych i kolumnowych.

PRZYKŁAD 2 Weźmy macierze

$$\mathbf{A} = \begin{bmatrix} 2 & 4 & 0 \\ -1 & 3 & 2 \\ -3 & 1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 & 5 & 3 \\ 2 & 3 & -2 & 1 \\ 4 & -2 & 0 & 2 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 3 & 1 & -2 \\ -5 & 0 & 2 \\ -2 & 4 & 1 \end{bmatrix}.$$

Wtedy mamy

$$\mathbf{A} + \mathbf{C} = \begin{bmatrix} 5 & 5 & -2 \\ -6 & 3 & 4 \\ -5 & 5 & 3 \end{bmatrix},$$

podczas gdy macierze $\mathbf{A}+\mathbf{B}$ i $\mathbf{B}+\mathbf{C}$ nie są określone. Oczywiście sumy $\mathbf{A}+\mathbf{A}$, $\mathbf{B}+\mathbf{B}$ i $\mathbf{C}+\mathbf{C}$ są także dobrze określone; na przykład

$$\mathbf{B} + \mathbf{B} = \begin{bmatrix} 2 & 0 & 10 & 6 \\ 4 & 6 & -4 & 2 \\ 8 & -4 & 0 & 4 \end{bmatrix}.$$

(b) Weźmy wektory wierszowe

$$\mathbf{v}_1 = [-2 \quad 1 \quad 2 \quad 3], \quad \mathbf{v}_2 = [4 \quad 0 \quad 3 \quad -2], \quad \mathbf{v}_3 = [1 \quad 3 \quad 5]$$

oraz wektory kolumnowe

$$\mathbf{v}_4 = \begin{bmatrix} 1 \\ 2 \\ -3 \\ 2 \end{bmatrix}, \quad \mathbf{v}_5 = \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}, \quad \mathbf{v}_6 = \begin{bmatrix} 4 \\ 1 \\ 5 \end{bmatrix}.$$

Jedynymi dobrze określonymi sumami różnych wektorów są:

$$\mathbf{v}_1 + \mathbf{v}_2 = [2 \quad 1 \quad 5 \quad 1] \quad \text{oraz} \quad \mathbf{v}_5 + \mathbf{v}_6 = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}. \qquad \blacksquare$$

Elementy zbioru $\mathbb{R}^n$ są też często nazywane **wektorami**. Dodajemy je tak, jakby były wektorami wierszowymi:

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n).$$

Zanim podamy własności dodawania macierzy, wprowadzimy pewne oznaczenia. Niech 0 oznacza macierz wymiaru $m \times n$, której wszystkie wyrazy są równe 0. (Z kontekstu zawsze będzie wynikało, jakie są wymiary tej macierzy). Jeśli $\mathbf{A}$ jest macierzą należącą do zbioru $\mathfrak{M}_{m,n}$, to macierz $-\mathbf{A}$, nazywaną **macierzą przeciwną** do $\mathbf{A}$, otrzymujemy zastępując każdy wyraz macierzy $\mathbf{A}$ liczbą przeciwną. Zatem, jeśli $\mathbf{A} = [a_{ij}]$, to $-\mathbf{A} = [-a_{ij}]$; równoważnie, $(-\mathbf{A})[i,j] = -\mathbf{A}[i,j]$.

Twierdzenie

Dla dowolnych macierzy $\mathbf{A}$, $\mathbf{B}$ i $\mathbf{C}$ w $\mathfrak{M}_{m,n}$:

(a) $\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$ (prawo łączności),
(b) $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$ (prawo przemienności),
(c) $\mathbf{A}+\mathbf{0}=\mathbf{0}+\mathbf{A}=\mathbf{A}$ (macierz zerowa),
(d) $\mathbf{A}+(-\mathbf{A})=(-\mathbf{A})+\mathbf{A}=\mathbf{0}$ (macierz przeciwna).

Dowód. Powyższe własności dodawania macierzy odzwierciedlają odpowiednie własności dodawania liczb rzeczywistych i są łatwe do sprawdzenia. Sprawdzimy (a) i resztę odłożymy do ćwiczenia 12.

Niech na przykład $\mathbf{A} = [a_{ij}]$, $\mathbf{B} = [b_{ij}]$, $\mathbf{C} = [c_{ij}]$. Wyrazem o współrzędnych (i,j) macierzy $\mathbf{B} + \mathbf{C}$ jest $b_{ij} + c_{ij}$, a więc wyrazem o współrzędnych (i,j) macierzy $\mathbf{A} + (\mathbf{B} + \mathbf{C})$ jest $a_{ij} + (b_{ij} + c_{ij})$. Podobnie, wyrazem o współrzędnych (i,j) macierzy $(\mathbf{A} + \mathbf{B}) + \mathbf{C}$ jest $(a_{ij} + b_{ij}) + c_{ij}$. Ponieważ dodawanie liczb rzeczywistych jest łączne, więc odpowiednie wyrazy macierzy $\mathbf{A} + (\mathbf{B} + \mathbf{C})$ i $(\mathbf{A} + \mathbf{B}) + \mathbf{C}$ są równe, a zatem te macierze są równe. $\blacksquare$

Ponieważ dodawanie macierzy jest łączne, zapis $\mathbf{A} + \mathbf{B} + \mathbf{C}$ nie będzie prowadził do nieporozumień.

Macierze można mnożyć przez liczby rzeczywiste, które w tym kontekście są często nazywane **skalarami**. Jeśli macierz $\mathbf{A}$ należy do zbioru $\mathfrak{M}_{m,n}$ i $c \in \mathbb{R}$, to $c\mathbf{A}$ jest macierzą wymiaru $m \times n$, której wyrazem o współrzędnych (i, j) jest ca_{ij}; zatem $(c\mathbf{A})[i, j] = c\mathbf{A}[i, j]$. To mnożenie nazywa się **mnożeniem macierzy przez skalar**, a macierz $c\mathbf{A}$ nazywa się **iloczynem macierzy przez skalar**.

PRZYKŁAD 3 (a) Jeśli

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & -3 \\ -1 & 0 & 4 \end{bmatrix},$$

to

$$2\mathbf{A} = \begin{bmatrix} 4 & 2 & -6 \\ -2 & 0 & 8 \end{bmatrix} \quad \text{oraz} \quad -7\mathbf{A} = \begin{bmatrix} -14 & -7 & 21 \\ 7 & 0 & -28 \end{bmatrix}.$$

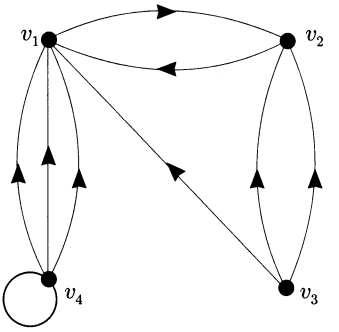
(b) Ogólnie, iloczyn $(-1)\mathbf{A}$ jest macierzą przeciwną $-\mathbf{A}$ do macierzy $\mathbf{A}$. ∎

Na zakończenie tego paragrafu opiszemy pewną macierz przydatną w badaniu relacji skończonych, grafów skierowanych i nieskierowanych. Weźmy najpierw skończony graf skierowany G, którego zbiorem wierzchołków jest $V(G)$. Niech $v_1, v_2, \ldots, v_n$ będzie ciągiem wszystkich wierzchołków ze zbioru $V(G)$. **Macierzą sąsiedztwa** nazywamy macierz $\mathbf{M}$ wymiaru $n \times n$, której każdy wyraz $\mathbf{M}[i, j]$ jest liczbą krawędzi od wierzchołka v_i do wierzchołka v_j. Zatem $\mathbf{M}[i, j] = 0$, jeśli nie istnieje krawędź od v_i do v_j, w przeciwnym przypadku $\mathbf{M}[i, j]$ jest liczbą całkowitą dodatnią.

PRZYKŁAD 4 (a) Macierzą sąsiedztwa grafu skierowanego przedstawionego na rysunku 3.12 jest macierz

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 3 & 0 & 0 & 1 \end{bmatrix}.$$

Zauważmy, że macierz $\mathbf{M}$ zawiera wszystkie informacje o tym grafie skierowanym. Mówi nam ona, że ten graf ma cztery wierzchołki oraz mówi nam, ile krawędzi łączy każdą parę wierzchołków.



Rysunek 3.12

(b) Oto macierz sąsiedztwa pewnego grafu skierowanego z § 3.2 (teraz jednak nie szukaj go):

$$\mathbf{M} = \begin{bmatrix} \mathbf{0} & 1 & 0 & 1 \\ 0 & \mathbf{0} & 0 & 1 \\ 1 & 1 & \mathbf{0} & 0 \\ 0 & 1 & 1 & \mathbf{1} \end{bmatrix}.$$

Popatrzmy, czego możemy dowiedzieć się o tym grafie tylko z tej macierzy. Graf ma cztery wierzchołki, gdyż macierz jest wymiaru 4×4. Ma on osiem krawędzi, gdyż suma wyrazów macierzy wynosi 8. Ponieważ wszystkie wyrazy macierzy są zerami i jedynkami, graf nie ma krawędzi wielokrotnych. Ma on jedną pętlę, gdyż jest jedna jedynka na głównej przekątnej (zaznaczonej półgrubą czcionką). Teraz popatrzmy na rysunek 3.3 w § 3.2, na którym jest narysowany ten graf skierowany, którego wierzchołki w, x, y i z przemianowaliśmy w pamięci na v_1, v_2, v_3 i v_4. Na przykład pętla jest przy wierzchołku $z = v_4$. ∎

Przypomnijmy, że każdej relacji R w zbiorze skończonym S odpowiada skończony graf skierowany G bez krawędzi wielokrotnych. Zatem odpowiada jej również macierz $\mathbf{M}_R$, której wyrazami są tylko 0 i 1. Ponieważ zbiorem wierzchołków grafu G jest zbiór S, więc jeśli $|S| = n$, to ta macierz jest macierzą wymiaru $n \times n$.

PRZYKŁAD 5 Powróćmy do relacji z przykładu 6 w § 3.1.

(a) Relacja R_1 w zbiorze $\{0, 1, 2, 3\}$ jest określona w następujący sposób: $(m, n) \in R_1$, jeśli $m \leq n$. Macierzą odpowiadającą relacji R_1 jest

$$\mathbf{M}_{R_1} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Tu znów przemianowaliśmy w pamięci zbiór wierzchołków; tym razem $0, 1, 2, 3$ odpowiadają wierzchołkom v_1, v_2, v_3, v_4.

(b) Relacja R_2 w zbiorze $S = \{1, 2, 3, 4, 5\}$ jest określona w następujący sposób: $(m, n) \in R_2$, jeśli liczba $m - n$ jest parzysta. Jeśli liczby $1, 2, 3, 4, 5$ będą wzięte w zwykłej kolejności, to macierz relacji R_2 będzie miała postać:

$$\mathbf{M}_{R_2} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Jeśli natomiast zmienimy kolejność tych liczb na $1, 3, 5, 2, 4$, to otrzymamy macierz

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Dzięki tej macierzy jest jasne, że spośród pierwszych trzech elementów zbioru S (mianowicie $1, 3$ i 5) każde dwa są ze sobą w relacji i pozostałe dwa elementy też są ze sobą w relacji.

■

Macierzą relacji odwrotnej $R^{\leftarrow}$ do relacji R jest macierz transponowana do macierzy relacji R. Można to zapisać wzorem $\mathbf{M}_{R^{\leftarrow}} = \mathbf{M}_R^T$.

PRZYKŁAD 6 (a) Macierzą relacji $R_1{}^{\leftarrow}$, dla relacji R_1 takiej, jak w przykładzie 5, jest

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Inaczej mówiąc, jest to macierz relacji określonej za pomocą nierówności $m \geq n$.

(b) Weźmy relację R_2 z przykładu 5. Gdy ustalimy kolejność elementów w zbiorze S, to macierz relacji $R_2{}^{\leftarrow}$ będzie taka sama jak macierz relacji R_2 dlatego, że $R_2 = R_2{}^{\leftarrow}$ (gdyż relacja R_2 jest symetryczna). ■

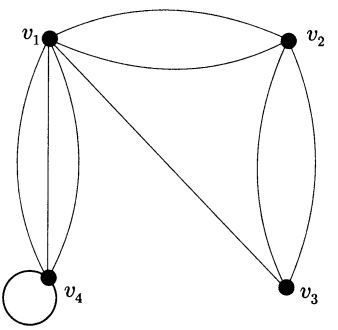
Ogólnie, relacja R jest symetryczna wtedy i tylko wtedy, gdy jej macierz $\mathbf{M}_R$ jest równa macierzy transponowanej do niej.

Macierze $\mathbf{M}$ takie, że $\mathbf{M} = \mathbf{M}^T$ nazywamy **macierzami symetrycznymi**, a więc: relacja jest symetryczna wtedy i tylko wtedy, gdy jej macierz jest symetryczna. Macierz symetryczna, której wyrazy są nieujemnymi liczbami całkowitymi może reprezentować zarówno graf skierowany, jak i nieskierowany.

PRZYKŁAD 7 Na rysunku 3.13 przedstawiono graf otrzymany z grafu z rysunku 3.12 poprzez pominięcie strzałek. Jego macierzą jest

$$\mathbf{M} = \begin{bmatrix} \mathbf{0} & 2 & 1 & 3 \\ 2 & \mathbf{0} & 2 & 0 \\ 1 & 2 & \mathbf{0} & 0 \\ 3 & 0 & 0 & \mathbf{1} \end{bmatrix}.$$

Zauważmy, że $\mathbf{M}^T = \mathbf{M}$, jak tego można było oczekiwać. Tak jak poprzednio, ten graf jest całkowicie wyznaczony przez swoją macierz i z tej macierzy można odczytać wiele informacji o grafie. Na przykład, główna przekątna (zaznaczona półgrubą czcionką) pokazuje, że w tym grafie jest tylko jedna pętla. Zliczanie krawędzi wymaga trochę więcej uwagi, gdyż każda krawędź nie będąca pętlą występuje w tej macierzy dwukrotnie. Liczba krawędzi jest więc sumą wyrazów stojących na i pod główną przekątną. ■



Rysunek 3.13

ĆWICZENIA DO § 3.3

1. Weźmy macierz

$$\mathbf{A} = \begin{bmatrix} 1 & -2 & 5 \\ 3 & -2 & 3 \\ 2 & 0 & 1 \end{bmatrix}.$$

Oblicz:
(a) a_{11}, (b) a_{13}, (c) a_{31}, (d) $\sum_{i=1}^{3} a_{ii}$.

2. Weźmy macierz

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & -2 & 1 \\ 3 & 0 & 1 & 2 \\ 2 & -1 & 4 & 1 \\ 0 & -3 & 1 & 3 \end{bmatrix}.$$

Oblicz: (a) b_{12}, (b) b_{21}, (c) b_{23}, (d) $\sum_{i=1}^{4} b_{ii}$.

3. Weźmy macierze

$$\mathbf{A} = \begin{bmatrix} -1 & 0 & 2 \\ 1 & 3 & -2 \\ 4 & 2 & 3 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 6 & 8 & 5 \\ 4 & -2 & 7 \\ 3 & 1 & 2 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 3 \\ 2 & -4 \\ 5 & -2 \end{bmatrix}.$$

Wyznacz następujące macierze, jeśli tylko one istnieją.
(a) $\mathbf{A}^T$, (b) $\mathbf{C}^T$, (c) $\mathbf{A}+\mathbf{B}$, (d) $\mathbf{A}+\mathbf{C}$, (e) $(\mathbf{A}+\mathbf{B})^T$,
(f) $\mathbf{A}^T+\mathbf{B}^T$, (g) $\mathbf{B}+\mathbf{B}^T$, (h) $\mathbf{C}+\mathbf{C}^T$, (i) $(\mathbf{A}+\mathbf{A})+\mathbf{B}$.

4. Weźmy następujące elementy zbioru $\mathbb{R}^3$:

$$\mathbf{v}_1 = (1,0,0), \quad \mathbf{v}_2 = (0,-1,1), \quad \mathbf{v}_3 = (1,0,-1).$$

Wyznacz:
(a) $\mathbf{v}_1+\mathbf{v}_2$, (b) $\mathbf{v}_1+\mathbf{v}_3$, (c) $\mathbf{v}_3+\mathbf{v}_2$, (d) $(\mathbf{v}_1+\mathbf{v}_2)+\mathbf{v}_1$.

5. Niech $\mathbf{A} = [a_{ij}]$ i $\mathbf{B} = [b_{ij}]$ będą macierzami należącymi do zbioru $\mathfrak{M}_{4,3}$, określonymi za pomocą wzorów: $a_{ij} = (-1)^{i+j}$ oraz $b_{ij} = i+j$. Wyznacz następujące macierze, o ile one istnieją.
(a) $\mathbf{A}^T$, (b) $\mathbf{A}+\mathbf{B}$, (c) $\mathbf{A}^T+\mathbf{B}$, (d) $\mathbf{A}^T+\mathbf{B}^T$, (e) $(\mathbf{A}+\mathbf{B})^T$,
(f) $\mathbf{A}+\mathbf{A}$.

6. Niech $\mathbf{A}$ i $\mathbf{B}$ będą macierzami należącymi do zbioru $\mathfrak{M}_{3,3}$, określonymi za pomocą wzorów: $\mathbf{A}[i,j] = ij$ oraz $\mathbf{B}[i,j] = i+j^2$.
(a) Wyznacz macierz $\mathbf{A}+\mathbf{B}$.
(b) Oblicz $\displaystyle\sum_{i=1}^{3} \mathbf{A}[i,i]$.
(c) Oblicz $\displaystyle\sum_{i=1}^{3}\Big(\sum_{j=1}^{3}\mathbf{B}[i,j]\Big)$ oraz $\displaystyle\sum_{j=1}^{3}\Big(\sum_{i=1}^{3}\mathbf{B}[i,j]\Big)$.
(d) Oblicz $\displaystyle\sum_{i=1}^{2}\Big(\sum_{j=2}^{3}\mathbf{B}[i,j]\Big)$ oraz $\displaystyle\sum_{j=1}^{2}\Big(\sum_{i=2}^{3}\mathbf{B}[i,j]\Big)$.
(e) Oblicz $\displaystyle\prod_{i=1}^{2}\Big(\sum_{j=2}^{3}\mathbf{A}[i,j]\Big)$.
(f) Czy macierz $\mathbf{A}$ jest równa swojej macierzy transponowanej $\mathbf{A}^T$?
(g) Czy macierz $\mathbf{B}$ jest równa swojej macierzy transponowanej $\mathbf{B}^T$?

7. (a) Wypisz wszystkie macierze wymiaru 3×3, których wierszami są następujące wektory wierszowe:

$$[1\ \ 0\ \ 0], \quad [0\ \ 1\ \ 0] \quad \text{oraz} \quad [0\ \ 0\ \ 1].$$

(b) Które macierze z punktu (a) są równe swoim macierzom transponowanym?

8. W tym ćwiczeniu $\mathbf{A}$ i $\mathbf{B}$ oznaczają macierze. Czy następujące zdania są prawdziwe czy fałszywe?
(a) $(\mathbf{A}^T)^T = \mathbf{A}$ dla wszystkich $\mathbf{A}$.
(b) Jeśli $\mathbf{A}^T = \mathbf{B}^T$, to $\mathbf{A} = \mathbf{B}$.
(c) Jeśli $\mathbf{A} = \mathbf{A}^T$, to $\mathbf{A}$ jest macierzą kwadratową.
(d) Jeśli $\mathbf{A}$ i $\mathbf{B}$ są macierzami tego samego wymiaru, to $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$.

9. Dla każdej liczby $n \in \mathbb{N}$ niech

$$\mathbf{A}_n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} \quad \text{oraz} \quad \mathbf{B}_n = \begin{bmatrix} 1 & (-1)^n \\ -1 & 1 \end{bmatrix}.$$

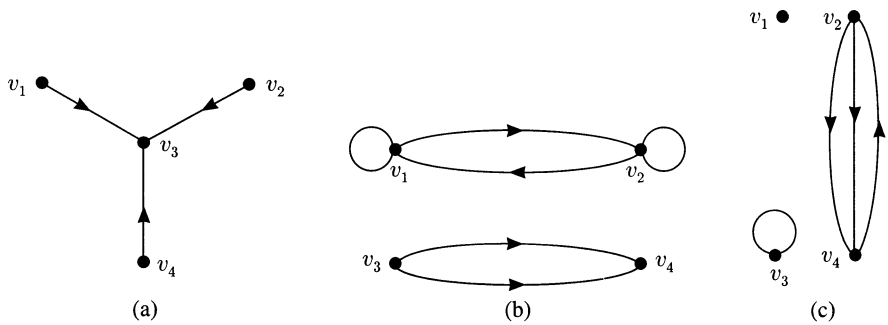
(a) Wyznacz macierze $\mathbf{A}_n^T$ dla wszystkich $n \in \mathbb{N}$.
(b) Wyznacz zbiór $\{n \in \mathbb{N}\colon\ \mathbf{A}_n^T = \mathbf{A}_n\}$.
(c) Wyznacz zbiór $\{n \in \mathbb{N}\colon\ \mathbf{B}_n^T = \mathbf{B}_n\}$.
(d) Wyznacz zbiór $\{n \in \mathbb{N}\colon\ \mathbf{B}_n = \mathbf{B}_0\}$.

10. Dla macierzy $\mathbf{A}$ i $\mathbf{B}$ ze zbioru $\mathfrak{M}_{m,n}$ niech $\mathbf{A}-\mathbf{B} = \mathbf{A}+(-\mathbf{B})$. Wykaż, że
(a) $(\mathbf{A}-\mathbf{B})+\mathbf{B} = \mathbf{A}$,
(b) $-(\mathbf{A}-\mathbf{B}) = \mathbf{B}-\mathbf{A}$,
(c) na ogół $(\mathbf{A}-\mathbf{B})-\mathbf{C} \neq \mathbf{A}-(\mathbf{B}-\mathbf{C})$.

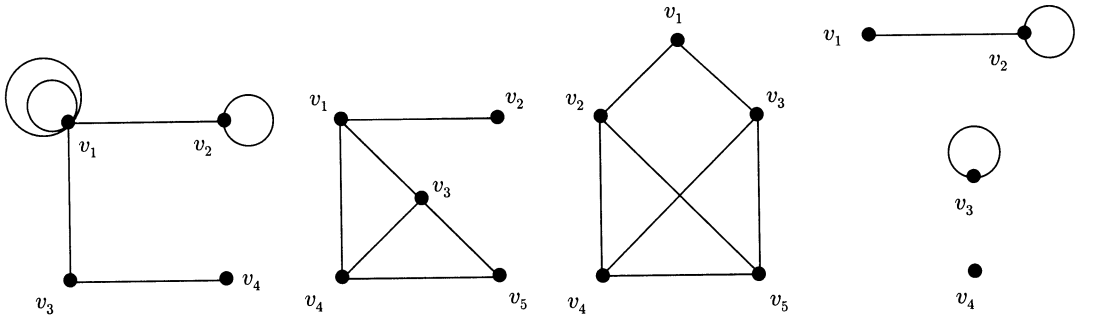
11. Weźmy macierze $\mathbf{A}, \mathbf{B} \in \mathfrak{M}_{m,n}$ oraz $a, b, c \in \mathbb{R}$. Wykaż, że
(a) $c(a\mathbf{A}+b\mathbf{B}) = (ca)\mathbf{A}+(cb)\mathbf{B}$,
(b) $-a\mathbf{A} = (-a)\mathbf{A} = a(-\mathbf{A})$,
(c) $(a\mathbf{A})^T = a\mathbf{A}^T$.

12. Udowodnij punkty (b), (c) i (d) twierdzenia.

13. Wyznacz macierze grafów skierowanych przedstawionych na rysunku 3.14.



Rysunek 3.14



(a) (b) (c) (d)

Rysunek 3.15

14. Wyznacz macierze grafów z rysunku 3.15.

15. Dla każdej macierzy z rysunku 3.16 narysuj graf skierowany mający taką macierz.

$$\begin{bmatrix} 0 & 0 & 2 & 1 \\ 3 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(a) (b) (c)

Rysunek 3.16

16. Dla każdej macierzy z rysunku 3.17 narysuj graf nieskierowany mający taką macierz.

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(a) (b) (c) (d)

Rysunek 3.17

17. Wyznacz macierz każdej relacji z przykładu 1 w § 3.1.

18. Narysuj graf skierowany mający macierz przedstawioną na rysunku 3.17(b).

19. Wyznacz macierz każdej relacji z ćwiczenia 2 w § 3.1.

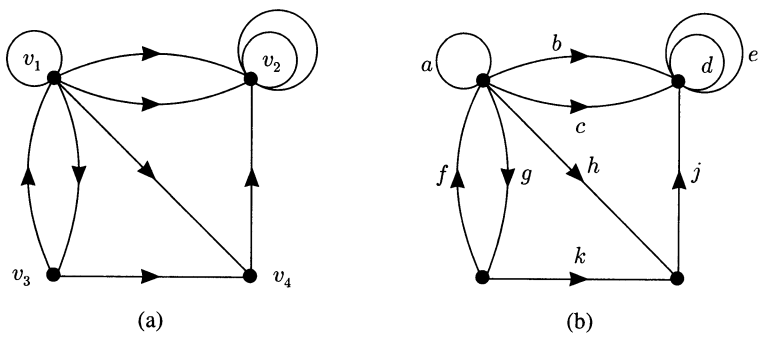
§ 3.4. Mnożenie macierzy

Jak widzieliśmy to w poprzednim paragrafie, dodawanie macierzy i mnożenie macierzy przez liczbę są określone w bardzo naturalny sposób. Jednak, jak zobaczymy, definicja mnożenia macierzy, może wydawać się bardzo dziwna i nienaturalna. Pochodzące z algebry liniowej uzasadnienie przyjęcia tej definicji odwołuje się do tego, że macierze odpowiadają pewnym funkcjom, które nazywamy przekształceniami liniowymi; mnożenie macierzy odpowiada składaniu przekształceń liniowych. Można również podać uzasadnienie odwołujące się do układów równań liniowych. Każde z tych dwóch podejść prowadziłoby zbyt głęboko w metody algebry liniowej. Sięgniemy więc po motywację do teorii grafów.

PRZYKŁAD 1 Weźmy graf skierowany przedstawiony na rysunku 3.18(a). Jego macierz sąsiedztwa ma postać:

$$\mathbf{M} = \begin{bmatrix} 1 & 2 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Zauważmy, że wyraz macierzy $\mathbf{M}$ o współrzędnych (i, j) jest liczbą dróg o długości 1 z wierzchołka v_i do wierzchołka v_j. Spróbujmy policzyć liczbę dróg długości 2 w tym grafie skierowanym. Na rysunku 3.18(b) krawędzie są oznaczone małymi literami. Łatwo zauważyć, że drogami długości 2 z wierzchołka v_1 do wierzchołka v_2 są: ab, ac, bd, be, cd, ce i hj. Jest więc siedem takich dróg. W podobny sposób można znaleźć liczbę dróg długości 2 z dowolnego wierzchołka v_i do dowolnego wierzchołka v_j.



Rysunek 3.18

Musi istnieć lepsza metoda wyznaczania liczby dróg niż ich bezpośrednie zliczanie, zwłaszcza wtedy, gdy mamy do czynienia

z dużymi grafami skierowanymi. Policzmy jeszcze raz drogi długości 2 od wierzchołka v_1 do wierzchołka v_2. Każda taka droga przechodzi w międzyczasie przez jeden z wierzchołków v_1, v_2, v_3 lub v_4, tak więc możemy policzyć drogi z v_1 do v_2 przechodzące przez v_1, drogi przechodzące przez v_2, przez v_3 i przez v_4, a następnie otrzymane liczby dodać do siebie. Jeśli chcemy na przykład policzyć drogi mające ciąg wierzchołków $v_1v_1v_2$, zliczamy krawędzie z v_1 do v_1 (tzn. pętle) i krawędzie z v_1 do v_2 i otrzymane liczby mnożymy przez siebie: $1 \cdot 2 = 2$. Tymi drogami są ab i ac. Liczby, które mnożymy, tzn. $\mathbf{M}[1,1]$ i $\mathbf{M}[1,2]$ są wzięte z macierzy $\mathbf{M}$. Jeszcze jeden przykład. Policzmy drogi mające ciąg wierzchołków $v_1v_2v_2$. W tym celu zliczamy krawędzie z v_1 do v_2 oraz z v_2 do v_2 i te liczby mnożymy przez siebie: $\mathbf{M}[1,2] \cdot \mathbf{M}[2,2] = 2 \cdot 2 = 4$. Tymi drogami są: bd, be, cd i ce. Wszystkie możliwe przypadki potrzebne do zliczenia dróg z v_1 do v_2 są pokazane w tablicy 3.1. Liczba wszystkich dróg długości 2 z v_1 do v_2 wynosi zatem

$$\mathbf{M}[1,1] \cdot \mathbf{M}[1,2] + \mathbf{M}[1,2] \cdot \mathbf{M}[2,2] + \mathbf{M}[1,3] \cdot \mathbf{M}[3,2] + \mathbf{M}[1,4] \cdot \mathbf{M}[4,2] = 2 + 4 + 0 + 1 = 7.$$

Tablica 3.1

Wierzchołek v_i	Liczba krawędzi z v_1 do v_i	Liczba krawędzi z v_i do v_2	Liczba dróg mających ciąg wierzchołków $v_1v_iv_2$
v_1	$\mathbf{M}[1,1] = 1$	$\mathbf{M}[1,2] = 2$	$\mathbf{M}[1,1] \cdot \mathbf{M}[1,2] = 1 \cdot 2 = 2$
v_2	$\mathbf{M}[1,2] = 2$	$\mathbf{M}[2,2] = 2$	$\mathbf{M}[1,2] \cdot \mathbf{M}[2,2] = 2 \cdot 2 = 4$
v_3	$\mathbf{M}[1,3] = 1$	$\mathbf{M}[3,2] = 0$	$\mathbf{M}[1,3] \cdot \mathbf{M}[3,2] = 1 \cdot 0 = 0$
v_4	$\mathbf{M}[1,4] = 1$	$\mathbf{M}[4,2] = 1$	$\mathbf{M}[1,4] \cdot \mathbf{M}[4,2] = 1 \cdot 1 = 1$

Liczba, którą otrzymamy w wyniku, będzie wyrazem stojącym w pierwszym wierszu i drugiej kolumnie macierzy nazywanej „iloczynem macierzy" $\mathbf{MM} = \mathbf{M}^2$. Jest to suma iloczynów wyrazów pierwszego wiersza macierzy $\mathbf{M}$ i drugiej kolumny macierzy $\mathbf{M}$.

Ogólnie, wyraz o współrzędnych (i,j) macierzy $\mathbf{M}^2$ jest liczbą dróg długości 2 od v_i do v_j. Okazuje się, że w tym przykładzie

$$\mathbf{M}^2 = \begin{bmatrix} 2 & 7 & 1 & 2 \\ 0 & 4 & 0 & 0 \\ 1 & 3 & 1 & 1 \\ 0 & 2 & 0 & 0 \end{bmatrix}.$$

Podobnie, wyrazy iloczynu $\mathbf{M}^2\mathbf{M} = \mathbf{M}^3$ podają nam liczbę dróg długości 3 łączących wierzchołki itd. Macierze te łatwo wyzna-

czyć (zob. ćwiczenia 8 i 11) przy użyciu metod, które właśnie zamierzamy opisać. ∎

Dwie macierze **A** i **B** można pomnożyć przez siebie, gdy liczba kolumn macierzy **A** jest równa liczbie wierszy macierzy **B**. Weźmy macierz **A** wymiaru $m \times n$ i macierz **B** wymiaru $n \times p$. Iloczynem **AB** jest macierz wymiaru $m \times p$ określona wzorem

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} \quad \text{dla} \quad 1 \leq i \leq m \text{ oraz } 1 \leq k \leq p.$$

Inaczej możemy to zapisać wzorem

$$(\mathbf{AB})[i,k] = \sum_{j=1}^{n} \mathbf{A}[i,j]\mathbf{B}[j,k].$$

Wyraz o współrzędnych (i,k) macierzy **AB** otrzymujemy mnożąc wyrazy i-tego wiersza macierzy **A** przez odpowiadające im wyrazy k-tej kolumny macierzy **B** i dodając otrzymane iloczyny:

$$\underset{\mathbf{A}}{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \dots & \dots & \dots & \dots \\ \boldsymbol{a_{i1}} & \boldsymbol{a_{i2}} & \boldsymbol{\dots} & \boldsymbol{a_{in}} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}} \underset{\mathbf{B}}{\begin{bmatrix} b_{11} & b_{12} & \dots & \boldsymbol{b_{1k}} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & \boldsymbol{b_{2k}} & \dots & b_{2p} \\ \dots & \dots & \dots & \boldsymbol{\dots} & \dots & \dots \\ b_{n1} & b_{n2} & \dots & \boldsymbol{b_{nk}} & \dots & b_{np} \end{bmatrix}}$$

$$= \underset{\mathbf{AB=C}}{\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1p} \\ c_{21} & c_{22} & \dots & c_{2p} \\ \dots & \dots & \boldsymbol{c_{ik}} & \dots \\ c_{m1} & c_{m2} & \dots & c_{mp} \end{bmatrix}}$$

Możemy sobie wyobrazić, że wyraz c_{ik} otrzymujemy wyjmując i-ty wiersz macierzy **A**, obracając go w prawo o 90°, nakładając go na k-tą kolumnę macierzy **B** i dodając iloczyny odpowiednich wyrazów:

$$c_{ik} = a_{i1}b_{1k} + a_{i2}b_{2k} + \ldots + a_{in}b_{nk}.$$

Aby to obliczenie było wykonalne, wiersze macierzy **A** i kolumny macierzy **B** muszą mieć po tyle samo wyrazów. Jeśli macierz **A** jest wymiaru $m \times n$ i macierz **B** jest wymiaru $r \times p$, to iloczyn macierzy **AB** jest zdefiniowany tylko wtedy, gdy $n = r$, iloczyn **AB** jest wówczas macierzą wymiaru $m \times p$.

PRZYKŁAD 2 Weźmy następujące macierze i wektory

$$\mathbf{A} = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -1 & 0 & 3 \\ 2 & 1 & -5 \end{bmatrix},$$

$$\mathbf{v}_1 = [2 \quad -3 \quad 4] \quad \text{oraz} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}.$$

(a) Aby wyznaczyć iloczyn $\mathbf{AB}$, najpierw obracamy pierwszy wiersz macierzy $\mathbf{A}$ i umieszczamy go kolejno nad pierwszą, drugą i trzecią kolumną macierzy $\mathbf{B}$. Te trzy obliczenia dadzą nam pierwszy wiersz iloczynu $\mathbf{AB}$:

$$\mathbf{AB} = \begin{bmatrix} -3-2 & 0-1 & 9+5 \\ & & \end{bmatrix} = \begin{bmatrix} -5 & -1 & 14 \\ & & \end{bmatrix}.$$

Postępując tak samo z drugim wierszem macierzy $\mathbf{A}$ otrzymamy drugi wiersz iloczynu $\mathbf{AB}$:

$$\mathbf{AB} = \begin{bmatrix} -5 & -1 & 14 \\ 10 & 4 & -26 \end{bmatrix}.$$

(b) Iloczyn $\mathbf{BA}$ nie jest określony, gdyż $\mathbf{B}$ jest macierzą wymiaru 2×3, $\mathbf{A}$ jest macierzą wymiaru 2×2 i $3 \neq 2$. Nasza metoda polegająca na obracaniu wiersza macierzy nie działa, gdyż wiersze macierzy $\mathbf{B}$ składają się z trzech wyrazów, a kolumny macierzy $\mathbf{A}$ mają dwa wyrazy; nie jest zatem jasne, w jaki sposób należałoby nakładać wiersze macierzy $\mathbf{B}$ na kolumny macierzy $\mathbf{A}$.

(c) Mamy

$$\mathbf{A}^2 = \mathbf{AA} = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 11 & -7 \\ -14 & 18 \end{bmatrix}.$$

(d) Mamy

$$\mathbf{Av}_2 = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} 6 \\ -14 \end{bmatrix}.$$

(e) Żaden z iloczynów $\mathbf{Bv}_1$ i $\mathbf{v}_1\mathbf{B}$ nie jest dobrze określony. Natomiast iloczyny $\mathbf{v}_1\mathbf{B}^T$ i $\mathbf{Bv}_1^T$ są równe

$$\mathbf{v}_1\mathbf{B}^T = [2 \quad -3 \quad 4] \begin{bmatrix} -1 & 2 \\ 0 & 1 \\ 3 & -5 \end{bmatrix} = [10 \quad -19]$$

oraz

$$\mathbf{Bv}_1^T = \begin{bmatrix} -1 & 0 & 3 \\ 2 & 1 & -5 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ -19 \end{bmatrix}.$$

Podobieństwo tych dwóch iloczynów nie jest przypadkowe. Przekonamy się o tym w ćwiczeniu 19.

■

Tak jak w przykładzie 1, potęgi macierzy sąsiedztwa mogą być wykorzystane do zliczania dróg w grafach nieskierowanych.

PRZYKŁAD 3 Weźmy graf, którego macierz sąsiedztwa jest równa:

$$\mathbf{M} = \begin{bmatrix} 1 & 2 & 2 & 1 \\ 2 & 2 & 0 & 1 \\ 2 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}.$$

Graf ten można otrzymać z grafu z rysunku 3.18 usuwając strzałki. Otrzymujemy wtedy

$$\mathbf{M}^2 = \begin{bmatrix} 10 & 7 & 3 & 5 \\ 7 & 9 & 5 & 4 \\ 3 & 5 & 5 & 2 \\ 5 & 4 & 2 & 3 \end{bmatrix}.$$

Zatem widzimy, że jest 10 dróg długości 2 z wierzchołka v_1 do niego samego, 7 dróg długości 2 z wierzchołka v_1 do wierzchołka v_2 itd. Aby otrzymać liczbę dróg długości 3, możemy skorzystać z macierzy

$$\mathbf{M}^3 = \begin{bmatrix} 35 & 39 & 25 & 20 \\ 39 & 36 & 18 & 21 \\ 25 & 18 & 8 & 13 \\ 20 & 21 & 13 & 11 \end{bmatrix}.$$

Jest rzeczą oczywistą, że bezpośrednie zliczanie dróg długości 3 w tym grafie byłoby nużące i trudno byłoby uniknąć błędów. ■

PRZYKŁAD 4 Weźmy macierze

$$\mathbf{A} = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}.$$

(a) Mamy

$$\mathbf{AB} = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 5 \\ -14 & 0 \end{bmatrix}$$

oraz

$$\mathbf{BA} = \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 7 \\ -11 & 7 \end{bmatrix}.$$

Przykład ten pokazuje, że mnożenie macierzy nie jest przemienne! Nawet wtedy, gdy oba iloczyny **AB** i **BA** są określone, mogą być sobie równe lub nie.

(b) Ponieważ

$$\mathbf{AB} = \begin{bmatrix} 6 & 5 \\ -14 & 0 \end{bmatrix},$$

mamy

$$(\mathbf{AB})\mathbf{C} = \begin{bmatrix} 6 & 5 \\ -14 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 16 & 15 \\ -14 & 0 \end{bmatrix}.$$

Z drugiej strony,

$$\mathbf{BC} = \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 6 \\ -1 & 3 \end{bmatrix},$$

a zatem

$$\mathbf{A}(\mathbf{BC}) = \begin{bmatrix} 3 & -1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 16 & 15 \\ -14 & 0 \end{bmatrix}.$$

Obliczenia te pokazują, że w tym przypadku $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. Równość nie jest tutaj przypadkowa, jak zobaczymy to w dalszej części tego paragrafu. ∎

PRZYKŁAD 5 Weźmy macierz wymiaru $n \times n$

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix};$$

zatem $\mathbf{I}[i,i] = 1$ dla $i = 1, 2, \dots, n$ oraz $\mathbf{I}[i,j] = 0$ dla $i \neq j$. Tę szczególną macierz nazywamy **macierzą jednostkową** wymiaru $n \times n$. Zawsze, kiedy będziemy chcieli wyraźnie zaznaczyć jej wymiar, będziemy oznaczać ją symbolem $\mathbf{I}_n$. Na przykład

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{oraz} \quad \mathbf{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Weźmy teraz dowolną macierz $\mathbf{A}$ wymiaru $m \times n$. Wtedy iloczyn macierzy $\mathbf{AI}_n$ jest dobrze określony oraz

$$(\mathbf{AI}_n)[i,k] = \sum_{j=1}^{n} \mathbf{A}[i,j]\mathbf{I}_n[j,k]$$

dla $1 \leq i \leq m$ i $1 \leq k \leq n$. Jeśli $j \neq k$, to $\mathbf{I}_n[j,k] = 0$, a więc ta suma ma tylko jeden składnik $\mathbf{A}[i,k]\mathbf{I}_n[k,k] = \mathbf{A}[i,k]$. Jest tak dla wszystkich i oraz k, a więc

(1) $$\mathbf{AI}_n = \mathbf{A} \quad \text{dla wszystkich} \quad \mathbf{A} \in \mathfrak{M}_{m,n}.$$

Weźmy następnie macierz $\mathbf{B}$ wymiaru $n \times p$. Wtedy iloczyn $\mathbf{I}_n\mathbf{B}$ jest dobrze określony oraz

$$(\mathbf{I}_n\mathbf{B})[i,k] = \sum_{j=1}^{n} \mathbf{I}_n[i,j]\mathbf{B}[j,k] = \mathbf{B}[i,k].$$

Zatem

(2) $\mathbf{I}_n\mathbf{B} = \mathbf{B}$ dla wszystkich $\mathbf{B} \in \mathfrak{M}_{n,p}$.

Obie równości (1) i (2) zachodzą jednocześnie dla macierzy wymiaru $n \times n$, a więc

(3) $\mathbf{A}\mathbf{I}_n = \mathbf{I}_n\mathbf{A} = \mathbf{A}$ dla wszystkich $\mathbf{A} \in \mathfrak{M}_{n,n}$. ∎

Weźmy ustaloną liczbę $n \in \mathbb{P}$. Twierdzimy, że istnieje wzajemnie jednoznaczna odpowiedniość między zbiorem $\mathbb{R}^n$, zbiorem $\mathfrak{M}_{n,1}$ wektorów kolumnowych i zbiorem $\mathfrak{M}_{1,n}$ wektorów wierszowych. Istotnie, wzór

$$f(x_1, x_2, \ldots, x_n) = [x_1 \ \ x_2 \ \ \ldots \ \ x_n]$$

określa przekształcenie wzajemnie jednoznaczne f ze zbioru $\mathbb{R}^n$ na zbiór $\mathfrak{M}_{1,n}$, a wzór $\mathrm{TRANS}(\mathbf{A}) = \mathbf{A}^T$ określa przekształcenie wzajemnie jednoznaczne TRANS ze zbioru $\mathfrak{M}_{1,n}$ na zbiór $\mathfrak{M}_{n,1}$. Złożenie tych dwóch przekształceń jest przekształceniem wzajemnie jednoznacznym ze zbioru $\mathbb{R}^n$ na zbiór $\mathfrak{M}_{n,1}$.

Mnożenie macierzy jest działaniem łącznym (tzn. zachodzi równość $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$, jeśli tylko obie strony są dobrze określone), co widzieliśmy w przykładzie 4. Od strony rachunkowej wydaje się to dość tajemnicze; zob. ćwiczenie 22. Ta tajemniczość znika, gdy powiążemy macierze z przekształceniami liniowymi, gdyż złożenie funkcji jest łączne. Jednakże nie wyjaśniliśmy tego związku między macierzami i przekształceniami liniowymi, więc sformułujemy tu ogólne prawo łączności bez dowodu.

Prawo łączności dla macierzy

Jeśli $\mathbf{A}$ jest macierzą wymiaru $m \times n$, $\mathbf{B}$ jest macierzą wymiaru $n \times p$ oraz $\mathbf{C}$ jest macierzą wymiaru $p \times q$, to

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}).$$

Ponieważ mnożenie macierzy jest łączne, więc zapis $\mathbf{ABC}$ nie prowadzi do nieporozumień. Podobnie potęgi takie, jak $\mathbf{A}^3 = \mathbf{AAA}$ są określone w sposób jednoznaczny. Podkreślamy jeszcze raz, że chociaż *mnożenie macierzy jest łączne, to nie jest ono przemienne*. Iloczyn $\mathbf{AB}$ nie musi być równy $\mathbf{BA}$, nawet jeśli

oba iloczyny są dobrze określone, jak widzieliśmy to w przykładzie 4(a). Przykłady innych praw arytmetycznych, których nie spełniają macierze, znajdują się w ćwiczeniach 16 i 20(b).

Dwie macierze $\mathbf{A}$ i $\mathbf{B}$ wymiaru $n \times n$ nazywamy macierzami odwrotnymi do siebie, jeśli $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$. Macierz $\mathbf{A}$ może mieć tylko jedną macierz odwrotną. Aby się o tym przekonać, przypuśćmy, że macierze $\mathbf{B}$ i $\mathbf{C}$ są dwiema macierzami odwrotnymi do macierzy $\mathbf{A}$. Wtedy $\mathbf{BA} = \mathbf{I}_n$ oraz $\mathbf{AC} = \mathbf{I}_n$, zatem $\mathbf{B} = \mathbf{BI}_n = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{I}_n\mathbf{C} = \mathbf{C}$. Macierz $\mathbf{A}$, która ma macierz odwrotną, nazywamy **macierzą odwracalną**, a jedyną macierz odwrotną do niej oznaczamy symbolem $\mathbf{A}^{-1}$. Istnieje wiele metod wyznaczania macierzy odwrotnych i wiele programów potrafi je również znajdować. Nie będziemy rozwijać tego tematu, chociaż w ćwiczeniu 14 podajemy wzór dla macierzy wymiaru 2×2.

ĆWICZENIA DO § 3.4

1. Niech

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 0 & 2 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ -1 & 0 \\ -2 & 3 \end{bmatrix}.$$

Wyznacz następujące macierze, o ile one istnieją:
(a) $\mathbf{AB}$, (b) $\mathbf{BA}$, (c) $\mathbf{ABA}$, (d) $\mathbf{A} + \mathbf{B}^T$, (e) $3\mathbf{A}^T - 2\mathbf{B}$, (f) $(\mathbf{AB})^2$.

2. Niech

$$\mathbf{C} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

i niech macierze $\mathbf{A}$ i $\mathbf{B}$ będą takie jak w ćwiczeniu 1. Wyznacz następujące macierze, o ile one istnieją:
(a) $\mathbf{AC}$, (b) $\mathbf{BC}$, (c) $\mathbf{C}^2$, (d) $\mathbf{C}^T\mathbf{C}$, (e) $\mathbf{CC}^T$, (f) $73\mathbf{C}$.

3. Niech

$$\mathbf{A} = \begin{bmatrix} 3 & -4 & 3 & 1 \\ 2 & 0 & 1 & -2 \\ -1 & 1 & 2 & 0 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

Wyznacz następujące macierze, o ile one istnieją:
(a) $\mathbf{A}^2$, (b) $\mathbf{B}^2$, (c) $\mathbf{AB}$, (d) $\mathbf{BA}$, (e) $\mathbf{BA}^T$, (f) $\mathbf{A}^T\mathbf{B}$.

4. Niech macierze $\mathbf{A}$ i $\mathbf{B}$ będą takie jak w ćwiczeniu 3 i niech macierz $\mathbf{v} = [-2 \quad 1 \quad -1]$. Wyznacz następujące macierze, o ile one istnieją:
(a) $\mathbf{vA}$, (b) $\mathbf{vB}$, (c) $\mathbf{Bv}^T$, (d) $(\mathbf{vB})^T$, (e) $5(\mathbf{vB})^T - 3\mathbf{Bv}^T$.

5. (a) Wyznacz obie macierze $(\mathbf{AB})\mathbf{C}$ i $\mathbf{A}(\mathbf{BC})$, gdy

$$\mathbf{A} = \begin{bmatrix} -1 & 4 \\ 2 & 5 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}.$$

(b) Wyznacz macierze $\mathbf{B}(\mathbf{AC})$ i $(\mathbf{BA})\mathbf{C}$.

6. Niech macierze $\mathbf{A}$, $\mathbf{B}$ i $\mathbf{C}$ będą takie jak w ćwiczeniu 5. Wyznacz
(a) macierze $\mathbf{AB}$ i $\mathbf{BA}$,
(b) macierze $\mathbf{AC}$ i $\mathbf{CA}$,
(c) macierz $\mathbf{A}^2$.

7. Niech

$$\mathbf{A} = \begin{bmatrix} 3 & -1 \\ 2 & 1 \\ -2 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} -1 & 3 \\ 2 & 1 \end{bmatrix}.$$

(a) Wyznacz macierze $\mathbf{A}(\mathbf{BC})$ i $(\mathbf{AB})\mathbf{C}$.
(b) Wyznacz macierze $\mathbf{A}(\mathbf{B}^2)$ i $(\mathbf{AB})\mathbf{B}$.

8. Pokaż, że

$$\mathbf{M}^2 = \begin{bmatrix} 2 & 7 & 1 & 2 \\ 0 & 4 & 0 & 0 \\ 1 & 3 & 1 & 1 \\ 0 & 2 & 0 & 0 \end{bmatrix},$$

gdzie $\mathbf{M}$ jest macierzą sąsiedztwa z przykładu 1.

9. Wykorzystaj macierz $\mathbf{M}^2$ z ćwiczenia 8 do znalezienia liczby dróg długości 2
(a) z wierzchołka v_1 do niego samego,
(b) z wierzchołka v_1 do wierzchołka v_3,
(c) z wierzchołka v_1 do wierzchołka v_4,
(d) z wierzchołka v_2 do wierzchołka v_1.

10. Znajdź bezpośrednio i wypisz drogi długości 2 opisane w ćwiczeniu 9. Wykorzystaj oznaczenia krawędzi z rysunku 3.18(b).

11. (a) Wyznacz macierz $\mathbf{M}^3$ dla macierzy sąsiedztwa $\mathbf{M}$ z przykładu 1.
(b) Znajdź liczbę dróg długości 3 z wierzchołka v_3 do wierzchołka v_2.
(c) Wypisz drogi długości 3 z wierzchołka v_3 do wierzchołka v_2 korzystając z oznaczeń krawędzi na rysunku 3.18(b).

12. To ćwiczenie dotyczy grafu opisanego w przykładzie 3.
(a) Narysuj ten graf; po prostu usuń strzałki z rysunku 3.18(a). Oznacz krawędzie tak jak na rysunku 3.18(b).
(b) Ile dróg długości 2 prowadzi z wierzchołka v_3 do niego samego?
(c) Wypisz drogi długości 2 z wierzchołka v_3 do niego samego.
(d) Ile dróg długości 3 prowadzi z wierzchołka v_3 do niego samego?
(e) Wypisz drogi długości 3 z wierzchołka v_3 do niego samego.

13. Powtórz polecenia (a)–(d) z ćwiczenia 12 dla wierzchołka v_2.

14. Wykaż, że macierz

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

wymiaru 2×2 ma macierz odwrotną wtedy i tylko wtedy, gdy $ad - bc \neq 0$ i w tym przypadku macierzą odwrotną jest

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Wskazówka: spróbuj rozwiązać równanie

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x & y \\ z & w \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

z niewiadomymi x, y, z, w.

15. Skorzystaj z ćwiczenia 14, by sprawdzić, które z następujących macierzy mają macierze odwrotne. Wyznacz macierze odwrotne, o ile istnieją i sprawdź odpowiedzi.

(a) $\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (b) $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ (c) $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

(d) $\mathbf{C} = \begin{bmatrix} 2 & -3 \\ 5 & 8 \end{bmatrix}$ (e) $\mathbf{D} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

16. Znajdź macierze wymiaru 2×2, które pokazują, że następująca równość $(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2$ nie zawsze jest prawdziwa.

17. Niech

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

(a) Wyznacz macierze $\mathbf{A}^n$ dla $n = 1, 2, 3, 4$.

(b) Odgadnij wzór ogólny dla macierzy $\mathbf{A}^n$ i sprawdź, że macierz opisana odgadniętym wzorem spełnia równanie $\mathbf{A}^n \cdot \mathbf{A} = \mathbf{A}^{n+1}$.

18. Weźmy macierze $\mathbf{A}$ i $\mathbf{B}$ ze zbioru $\mathfrak{M}_{m,n}$ oraz $a \in \mathbb{R}$. Pokaż, że $(a\mathbf{A})\mathbf{B} = a(\mathbf{AB}) = \mathbf{A}(a\mathbf{B})$.

19. Pokaż, że jeśli $\mathbf{A}$ jest macierzą wymiaru $m \times n$ i $\mathbf{B}$ jest macierzą wymiaru $n \times p$, to $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$. Zauważ, że po obu stronach tej równości mamy macierze wymiaru $p \times m$.

20. (a) Udowodnij, że w zbiorze $\mathfrak{M}_{m,n}$ zachodzi prawo skracania dla dodawania, tzn. udowodnij, że jeśli macierze $\mathbf{A}$, $\mathbf{B}$ i $\mathbf{C}$ należą do zbioru $\mathfrak{M}_{m,n}$ i $\mathbf{A} + \mathbf{C} = \mathbf{B} + \mathbf{C}$, to $\mathbf{A} = \mathbf{B}$.

(b) Pokaż, że w zbiorze $\mathfrak{M}_{n,n}$ nie zachodzi prawo skracania dla mnożenia, tzn. pokaż, że z równości $\mathbf{AC} = \mathbf{BC}$ nie musi wynikać równość $\mathbf{A} = \mathbf{B}$, nawet jeśli $\mathbf{C} \neq \mathbf{0}$.

21. (a) Niech

$$\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$

dla pewnej ustalonej liczby $a \in \mathbb{R}$. Pokaż, że $\mathbf{AB} = \mathbf{BA}$ dla wszystkich macierzy $\mathbf{B}$ ze zbioru $\mathfrak{M}_{2,2}$.

(b) Weźmy ustaloną macierz $\mathbf{A}$ ze zbioru $\mathfrak{M}_{2,2}$, która spełnia równanie $\mathbf{AB} = \mathbf{BA}$ dla wszystkich macierzy $\mathbf{B} \in \mathfrak{M}_{2,2}$. Pokaż, że

$$\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \quad \text{dla pewnego} \quad a \in \mathbb{R}.$$

Wskazówka: Zapisz macierz $\mathbf{A}$ w postaci

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

i podstaw

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{oraz} \quad \mathbf{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

22. (a) Pokaż bezpośrednio, że $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ dla macierzy $\mathbf{A}$, $\mathbf{B}$ i $\mathbf{C}$ ze zbioru $\mathfrak{M}_{2,2}$.

(b) Czy podobało ci się ćwiczenie (a)? Jeśli tak, to podaj bezpośredni dowód prawa łączności dla dowolnych macierzy.

23. (a) Niech $\mathbf{A}$ i $\mathbf{B}$ będą macierzami wymiaru $m \times n$, a $\mathbf{C}$ macierzą wymiaru $n \times p$. Pokaż, że zachodzi prawo rozdzielności: $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$.

(b) Sprawdź, czy zachodzi prawo rozdzielności $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$. Najpierw określ wymiary macierzy, dla których ma ono sens.

§ 3.5. Relacje równoważności i podziały zbioru

W tym paragrafie zajmiemy się relacjami równoważności; są to relacje, które grupują razem elementy o podobnych cechach lub mające wspólne pewne własności. Relacje równoważności występują w całej matematyce i w innych dziedzinach, chociaż nie zawsze są rozpoznawane jako takie.

PRZYKŁAD 1 (a) Niech S będzie zbiorem szklanych kulek. Kulki s i t uznamy za równoważne, jeśli są tego samego koloru, piszemy wtedy $s \sim t$. Zauważmy, że relacja $\sim$ ma następujące trzy własności:

(Z) $s \sim s$ dla wszystkich kulek s.
(S) Jeśli $s \sim t$, to $t \sim s$.
(P) Jeśli $s \sim t$ i $t \sim u$, to $s \sim u$.

Te własności są prawie że oczywiste. Na przykład, (P) mówi, że jeśli kulki s i t są tego samego koloru oraz kulki t i u są tego

samego koloru, to kulki s i u są tego samego koloru. Zauważmy także, że możemy podzielić zbiór S na rozłączne podzbiory tak, by elementy należały do tego samego podzbioru wtedy i tylko wtedy, gdy są równoważne, tzn. wtedy i tylko wtedy, gdy są tego samego koloru.

(b) W tym samym zbiorze kulek S możemy uznać kulki s i t za równoważne, jeśli są one tej samej wielkości, w tym przypadku piszemy $s \approx t$. Wszystko to, co napisaliśmy w punkcie (a), odnosi się do relacji $\approx$, po wprowadzeniu oczywistych zmian. ∎

Niech S będzie dowolnym zbiorem i załóżmy, że dana jest relacja $\sim$ w zbiorze S, tzn. dla każdej pary $(x, y) \in S \times S$ wiemy, czy zachodzi $x \sim y$. Relację $\sim$ nazywamy **relacją równoważności**, jeśli jest ona zwrotna, symetryczna i przechodnia:

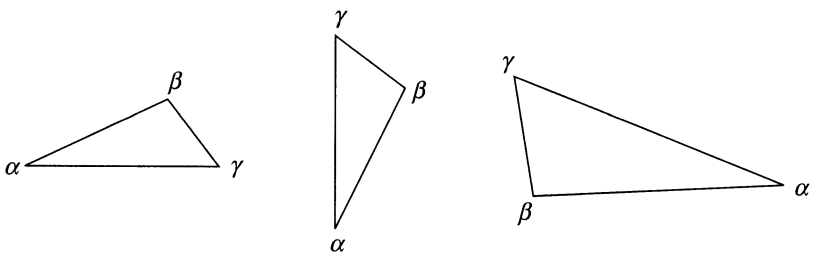
(Z) $s \sim s$ dla każdego $s \in S$.
(S) Jeśli $s \sim t$, to $t \sim s$.
(P) Jeśli $s \sim t$ i $t \sim u$, to $s \sim u$.

Jeśli $s \sim t$, to mówimy, że elementy s i t są równoważne; w zależności od okoliczności możemy również powiedzieć, że elementy s i t są **podobne**, **przystające** lub **izomorficzne**. Czasami relację równoważności oznacza się symbolami: $s \approx t$, $s \cong t$, $s \equiv t$ i $s \leftrightarrow t$. Wszystkie te oznaczenia mają na celu przekazanie informacji, że elementy s i t mają takie samo (lub równoważne) znaczenie. Jest to rozsądne podejście ze względu na własność symetrii.

PRZYKŁAD 2 Trójkąty T_1 i T_2 na płaszczyźnie nazywamy **podobnymi** i piszemy $T_1 \approx T_2$, jeśli ich kąty możemy przyporządkować sobie wzajemnie jednoznacznie tak, by odpowiadające sobie kąty były równe. Jeśli odpowiadające sobie boki są też równe, to trójkąty nazywamy **przystającymi** i piszemy $T_1 \cong T_2$. Na rysunku 3.19 mamy $T_1 \cong T_2$, $T_1 \approx T_3$ i $T_2 \approx T_3$, ale trójkąt T_3 nie jest przystający ani do T_1 ani do T_2. Obie relacje $\approx$ i $\cong$ są relacjami równoważności w zbiorze wszystkich trójkątów na płaszczyźnie. To, że te relacje mają własności (Z), (S) i (P), jest oczywiste. ∎

Zazwyczaj używamy oznaczenia takiego jak $\sim$ lub $\equiv$ tylko wtedy, gdy mamy do czynienia z relacją równoważności. Oczywiście samo użycie takiego oznaczenia nie gwarantuje automatycznie, że mamy relację równoważności.

PRZYKŁAD 3 (a) Dla liczb $m, n \in \mathbb{Z}$ definiujemy $m \sim n$ wtedy, gdy liczba $m-n$ jest nieparzysta. Relacja $\sim$ jest symetryczna, ale jest wysoce



Rysunek 3.19

niezwrotna i nieprzechodnia. W istocie

$$m \not\sim m \quad \text{dla wszystkich} \quad m \in \mathbb{Z}$$

oraz

$$m \sim n \text{ i } m \sim p \text{ zawsze implikują } m \not\sim p.$$

(b) Weźmy zbiór S wszystkich funkcji przekształcających przedział $[0,1]$ w zbiór $\mathbb{R}$. Definiujemy $f \sim g$, jeśli $|f(x) - g(x)| \leq 1$ dla wszystkich $x \in [0,1]$. Można powiedzieć, że $f \sim g$ wtedy, gdy funkcje są „wystarczająco bliskie" lub „w przybliżeniu równe". Relacja $\sim$ jest zwrotna i symetryczna w zbiorze S, ale nie jest przechodnia. Na przykład, jeśli $f(x) = 0$, $g(x) = x$ oraz $h(x) = 2x$ dla $x \in [0,1]$, to $f \sim g$ i $g \sim h$, ale $f \not\sim h$. ∎

PRZYKŁAD 4

(a) Niech G będzie grafem. Definiujemy relację $\sim$ w zbiorze $V(G)$, mówiąc, że $v \sim w$, jeśli wierzchołki v i w są połączone krawędzią. Zatem relacja $\sim$ jest relacją sąsiedztwa. Jest ona zwrotna tylko wtedy, gdy z każdego wierzchołka wychodzi pętla. Możemy określić nową relację $\simeq$ w zbiorze $V(G)$, która jest przechodnia, definiując $v \simeq w$ wtedy i tylko wtedy, gdy $v = w$ lub $v \sim w$. Relacja $\simeq$ jest zwrotna i symetryczna, ale nadal nie musi być przechodnia. Wystarczy sobie wyobrazić przykład wierzchołków u, v, w takich, że $u \simeq v$ i $v \simeq w$, przy czym $u \neq w$ i takich, że wierzchołki u i w nie są połączone krawędzią.

(b) W paragrafie 3.2 zdefiniowaliśmy relację osiągalności R w zbiorze $V(G)$ w ten sposób, że $(v,w) \in R$, jeśli istnieje droga długości co najmniej 1 z wierzchołka v do wierzchołka w. Jest to dokładnie ta relacja, którą otrzymalibyśmy próbując zrobić relację przechodnią z relacji sąsiedztwa $\sim$. Nowa relacja R jest przechodnia i symetryczna, ale jeśli graf G ma izolowane wierzchołki bez pętli, to nie jest ona zwrotna. Możemy utworzyć relację równoważności $\cong$ z relacji R, stosując pomysł opisany w punkcie (a). Mówimy, że $v \cong w$, jeśli $v = w$ lub $(v,w) \in R$. ∎

PRZYKŁAD 5 (a) Weźmy program, który akceptuje jako dane wejściowe ciągi ze zbioru Σ^* dla pewnego alfabetu Σ i generuje ciągi wyjściowe. Możemy zdefiniować relację równoważności $\sim$ w zbiorze Σ^*, przyjmując, że $w_1 \sim w_2$, jeśli ten program generuje te same ciągi wyjściowe zarówno dla danych w_1, jak i dla w_2. Aby sprawdzić, czy dwa słowa są równoważne, czy nie, możemy skorzystać z programu.

(b) Możemy również mówić o równoważnych programach. Określamy relacje równoważności $\approx_1$, $\approx_2$, $\approx_3$, ... pisząc $B \approx_k C$ dla programów B i C, jeśli B i C dają te same wyniki dla każdych danych wejściowych będących słowami długości k. Definiujemy relację $\approx$, przyjmując, że $B \approx C$, jeśli $B \approx_k C$ dla wszystkich $k \in \mathbb{P}$. Wtedy wszystkie relacje $\approx_k$ i relacja $\approx$ są relacjami równoważności w zbiorze programów, a dwa programy są równoważne ze względu na relację $\approx$ wtedy i tylko wtedy, gdy dają one ten sam wynik dla każdych danych wejściowych będących słowami utworzonymi z liter alfabetu Σ. ∎

PRZYKŁAD 6 Niech $\mathcal{S}$ będzie rodziną zbiorów i dla zbiorów $S, T \in \mathcal{S}$ definiujemy $S \sim T$, jeśli istnieje funkcja wzajemnie jednoznaczna ze zbioru S na zbiór T. Relacja $\sim$ jest relacją równoważności w zbiorze $\mathcal{S}$. Istotnie:

(Z) $S \sim S$, gdyż funkcja identycznościowa $1_S\colon S \to S$ jest przekształceniem wzajemnie jednoznacznym ze zbioru S na S.

(S) Jeśli $S \sim T$, to istnieje funkcja wzajemnie jednoznaczna f ze zbioru S na zbiór T. Funkcja f^{-1} odwrotna do niej jest przekształceniem wzajemnie jednoznacznym ze zbioru T na zbiór S, a więc $T \sim S$.

(P) Jeśli $S \sim T$ i $T \sim U$, to istnieją przekształcenia wzajemnie jednoznaczne $f\colon S \to T$ i $g\colon T \to U$. Łatwo sprawdzić, że $g \circ f$ jest przekształceniem wzajemnie jednoznacznym z S na U. Wynika to również z ćwiczenia 9 w § 1.4 oraz z twierdzenia sformułowanego w tamtym paragrafie. Tak czy inaczej, wynika stąd, że $S \sim U$.

Zauważmy, że jeśli zbiór S jest skończony, to $S \sim T$ wtedy i tylko wtedy, gdy zbiory S i T mają tyle samo elementów. Jeśli zbiór S jest nieskończony, to $S \sim T$ dla pewnych zbiorów nieskończonych T należących do $\mathcal{S}$, ale prawdopodobnie nie dla wszystkich takich zbiorów, gdyż nie wszystkie zbiory nieskończone są równoważne. To stwierdzenie nie jest oczywiste. Rzucimy okiem na to fascynujące zagadnienie w ostatnim paragrafie tej książki, § 13.3. ∎

W przykładzie 1 wyraźnie zaobserwowaliśmy, że zbiór S kulek może być traktowany jako suma rozłącznych podzbiorów, przy czym dwie kulki należą do tego samego podzbioru wtedy i tylko wtedy, gdy są równoważne. Cały zbiór kulek został podzielony na rozłączne podzbiory składające się z elementów, które są sobie równoważne. W rzeczywistości podobne zjawisko występuje w każdym z przykładów tego paragrafu, choć w niektórych przypadkach nie jest to tak oczywiste. Taka rodzina zbiorów ma swoją nazwę. **Podziałem** niepustego zbioru S nazywamy zbiór niepustych, rozłącznych podzbiorów tego zbioru, których sumą jest S.

PRZYKŁAD 7 (a) Niech f będzie funkcją ze zbioru S na zbiór T. Wtedy zbiór $\{f^{\leftarrow}(y)\colon\ y \in T\}$ wszystkich przeciwobrazów $f^{\leftarrow}(y)$ jest podziałem zbioru S. Przede wszystkim zauważmy, że każdy zbiór $f^{\leftarrow}(y)$ jest niepusty, ponieważ f jest funkcją na zbiór T. Każdy element x zbioru S znajduje się w dokładnie jednym podzbiorze postaci $f^{\leftarrow}(y)$, mianowicie w zbiorze $f^{\leftarrow}(f(x))$ składającym się ze wszystkich $s \in S$ takich, że $f(s) = f(x)$. Jeśli $y \neq z$, to mamy $f^{\leftarrow}(y) \cap f^{\leftarrow}(z) = \emptyset$. Ponadto suma $\bigcup_{y \in T} f^{\leftarrow}(y)$ wszystkich zbiorów $f^{\leftarrow}(y)$ jest równa S, zatem zbiór $\{f^{\leftarrow}(y)\colon\ y \in T\}$ jest podziałem zbioru S. Taki rodzaj podziału widzieliśmy w przykładzie 7 w § 1.4.

(b) Powróćmy do zbioru S kulek z przykładu 1(a). Niech C będzie zbiorem kolorów kulek i określmy funkcję $f\colon S \to C$ kładąc $f(s) =$ „kolor kulki s" dla $s \in S$. Wtedy podział $\{f^{\leftarrow}(c)\colon\ c \in C\}$ jest dokładnie tym podziałem zbioru S, o którym mówiliśmy w przykładzie 1. Zatem wartości funkcji f dla dwóch kulek są takie same, a więc te kulki należą do tego samego podzbioru wtedy i tylko wtedy, gdy mają one ten sam kolor. Funkcja $g\colon S \to \mathbb{R}$ określona w następujący sposób: $g(s) =$ „średnica kulki s" umieszcza dwie kulki w tym samym zbiorze $g^{\leftarrow}(d)$ wtedy i tylko wtedy, gdy mają one tę samą wielkość. Ten związek między relacjami równoważności i podziałami wyznaczonymi przez przeciwobrazy względem jakiejś funkcji jest zjawiskiem ogólnym, jak się wkrótce przekonamy w twierdzeniu 2. ∎

Weźmy znów jakąś relację równoważności $\sim$ w zbiorze S. Dla każdego $s \in S$ definiujemy

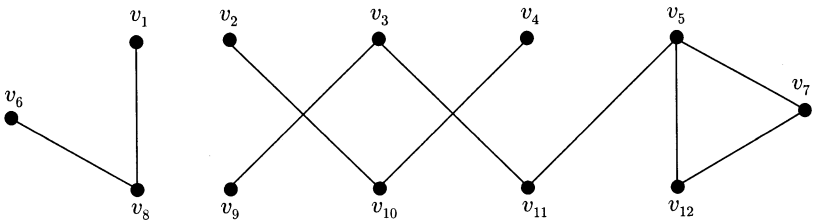
$$[s] = \{t \in S\colon\ s \sim t\}.$$

Zbiór $[s]$ nazywamy **klasą równoważności** (lub klasą abstrakcji) elementu s. Dla nas słowa „klasa" i „zbiór" są synonimami, a więc zbiór $[s]$ mógłby być nazywany „zbiorem równoważności"; nikt

go jednak tak nie nazywa. Zbiór wszystkich klas równoważności w zbiorze S oznaczamy symbolem $[S]$, a więc $[S] = \{[s]\colon\ s \in S\}$. Czasami dodajemy indeksy do symboli $[s]$ i $[S]$ dla wyjaśnienia, o którą z wielu możliwych relacji równoważności chodzi.

PRZYKŁAD 8 (a) W przykładzie 1(a) dotyczącym kulek klasa równoważności $[s]$ danej kulki s jest zbiorem kulek mających ten sam kolor co kulka s; ten zbiór zawiera samą kulkę s. Klasami równoważności są {kulki niebieskie}, {kulki czerwone}, {kulki zielone} itp.

(b) Weźmy relację równoważności $\cong$ w zbiorze V wierzchołków grafu, zdefiniowaną za pomocą relacji osiągalności z przykładu 4(b). Dwa wierzchołki są równoważne dokładnie wtedy, gdy należą do tej samej spójnej części grafu. (W istocie właśnie tak zdefiniujemy pojęcie „spójności" w § 6.2.) Na przykład klasami równoważności dla grafu przedstawionego na rysunku 3.20 są zbiory $\{v_1, v_6, v_8\}$, $\{v_2, v_4, v_{10}\}$ i $\{v_3, v_5, v_7, v_{11}, v_{12}\}$. Jeśli graf jest spójny, to jedyną klasą równoważności jest sam zbiór V. ∎



Rysunek 3.20

Klasy równoważności w przykładzie 8 również tworzą podział zbioru, w którym jest określona odpowiednia relacja równoważności. Zanim udowodnimy twierdzenie 1, które stwierdza, że taki fakt ma zawsze miejsce, udowodnimy lemat, w którym są sformułowane podstawowe własności klas równoważności.

Lemat

Niech $\sim$ będzie relacją równoważności w zbiorze S. Dla s i t należących do zbioru S następujące warunki są logicznie równoważne:

(i) $s \sim t$;
(ii) $[s] = [t]$;
(iii) $[s] \cap [t] \neq \emptyset$.

Dowód. Słowa „logicznie równoważne" oznaczają, że te trzy stwierdzenia są jednocześnie prawdziwe lub jednocześnie fał-

szywe. Jeśli jedno z nich jest prawdziwe, to wszystkie są prawdziwe. Udowodnimy implikacje (i) $\Rightarrow$ (ii), (ii) $\Rightarrow$ (iii) i (iii) $\Rightarrow$ (i).

$(s \sim t) \Rightarrow ([s] = [t])$. Załóżmy $s \sim t$ i weźmy $s' \in [s]$. Wtedy $s \sim s'$. Z symetrii mamy $t \sim s$. Ponieważ $t \sim s$ i $s \sim s'$, więc z przechodniości wynika, że $t \sim s'$. Zatem $s' \in [t]$. Pokazaliśmy, że każdy element s' zbioru $[s]$ należy do zbioru $[t]$, a więc $[s] \subseteq [t]$. W podobny sposób otrzymujemy $[t] \subseteq [s]$.

Implikacja $([s] = [t]) \Rightarrow ([s] \cap [t] \neq \emptyset)$ jest oczywista, gdyż każdy zbiór $[s]$ jest niepusty (dlaczego?).

$([s] \cap [t] \neq \emptyset) \Rightarrow (s \sim t)$. Weźmy element u zbioru $[s] \cap [t]$. Wtedy $s \sim u$ i $t \sim u$. Z symetrii mamy $u \sim t$. Ponieważ $s \sim u$ i $u \sim t$, więc z przechodniości mamy $s \sim t$. ∎

Twierdzenie 1

(a) Jeśli $\sim$ jest relacją równoważności w niepustym zbiorze S, to zbiór $[S]$ jest podziałem zbioru S.

(b) Na odwrót, jeśli rodzina zbiorów $\{A_i\colon\ i \in I\}$ jest podziałem zbioru S, to zbiory A_i są klasami równoważności wyznaczonymi przez pewną relację równoważności w zbiorze S.

Dowód. (a) Aby pokazać, że rodzina zbiorów $[S]$ jest podziałem zbioru S, musimy dowieść, że

$$\bigcup_{s \in S} [s] = S \tag{1}$$

oraz

$$\text{dla } s, t \in S \quad \text{albo} \quad [s] = [t], \quad \text{albo} \quad [s] \cap [t] = \emptyset. \tag{2}$$

Oczywiście $[s] \subseteq S$ dla każdego $s \in S$, a więc $\bigcup_{s \in S} [s] \subseteq S$. Dla danego elementu s zbioru S mamy $s \sim s$, a więc $s \in [s]$; zatem $S \subseteq \bigcup_{s \in S} [s]$. Stąd wynika (1).

Stwierdzenie (2) jest logicznie równoważne ze stwierdzeniem

$$\text{jeśli } [s] \cap [t] \neq \emptyset, \text{ to } [s] = [t], \tag{3}$$

które wynika z lematu.

(b) Dla danego podziału $\{A_i\colon\ i \in I\}$ zbioru S definiujemy relację $\sim$ w zbiorze S kładąc $s \sim t$, jeśli s i t należą do tego samego zbioru A_i. Własności (Z), (S) i (P) są oczywiste, więc $\sim$ jest relacją równoważności w zbiorze S. Dla dowolnego niepustego zbioru A_i mamy $A_i = [s]$ dla wszystkich $s \in A_i$, a więc ten podział składa się dokładnie ze wszystkich klas równoważności $[s]$. ∎

Czasami definiujemy relację równoważności w jakimś zbiorze za pomocą funkcji, której dziedziną jest ten zbiór. W pewnym

sensie każda relacja równoważności powstaje w ten sposób; dokładniej wyjaśni to twierdzenie 2.

PRZYKŁAD 9 Zdefiniujmy relację $\sim$ w zbiorze $\mathbb{N} \times \mathbb{N}$ przyjmując $(m, n) \sim (j, k)$ wtedy i tylko wtedy, gdy $m^2 + n^2 = j^2 + k^2$. Łatwo można sprawdzić bezpośrednio, że relacja $\sim$ jest relacją równoważności. Postąpimy jednak nieco inaczej. Określmy funkcję $f\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ za pomocą wzoru

$$f(m, n) = m^2 + n^2.$$

Wówczas pary uporządkowane są równoważne dokładnie wtedy, gdy funkcja f przyjmuje na nich te same wartości. Klasami równoważności są po prostu niepuste zbiory $f^{\leftarrow}(r)$, gdzie $r \in \mathbb{N}$. Niektóre ze zbiorów $f^{\leftarrow}(r)$ są puste, na przykład $f^{\leftarrow}(3)$, ale to niczemu nie szkodzi. ∎

Twierdzenie 2

(a) Niech S będzie zbiorem niepustym i niech f będzie funkcją o dziedzinie S. Przyjmijmy, że $s_1 \sim s_2$, jeśli $f(s_1) = f(s_2)$. Wtedy $\sim$ jest relacją równoważności w zbiorze S i klasami równoważności są niepuste zbiory $f^{\leftarrow}(t)$ dla t należących do zbioru T, w którym funkcja f przyjmuje swoje wartości.

(b) Każda relacja równoważności $\sim$ w zbiorze S jest wyznaczona przez odpowiednią funkcję o dziedzinie S, tak jak w punkcie (a).

Dowód. Sprawdzamy najpierw, że $\sim$ jest relacją równoważności.

(Z) $f(s) = f(s)$, więc $s \sim s$ dla wszystkich $s \in S$.

(S) Jeśli $f(s_1) = f(s_2)$, to $f(s_2) = f(s_1)$, a więc $s_1 \sim s_2$ implikuje $s_2 \sim s_1$.

(P) Jeśli $f(s_1) = f(s_2)$ oraz $f(s_2) = f(s_3)$, to $f(s_1) = f(s_3)$, a więc relacja $\sim$ jest przechodnia.

Stwierdzenie dotyczące klas równoważności to dokładnie definicja zbiorów $f^{\leftarrow}(t)$.

Aby dowieść (b), definiujemy funkcję $\nu\colon S \to [S]$, nazywaną **przekształceniem naturalnym** (lub kanonicznym) zbioru S na zbiór $[S]$, za pomocą wzoru

$$\nu(s) = [s] \quad \text{dla } s \in S.$$

(Litera ν nie jest literą v; jest to grecka litera ni, od pierwszej litery słowa „naturalne"). Z lematu znajdującego się przed twierdzeniem 1 wynika, że $s \sim t$ wtedy i tylko wtedy, gdy $[s] = [t]$,

a więc $s \sim t$ wtedy i tylko wtedy, gdy $\nu(s) = \nu(t)$. Zatem $\sim$ jest relacją równoważności określoną za pomocą przekształcenia ν. Zauważmy, że funkcja ν przekształca zbiór S na zbiór $[S]$ oraz że $\nu^{\leftarrow}([s]) = [s]$ dla dowolnego $s \in S$. ∎

PRZYKŁAD 10 (a) Jeśli S jest znanym nam zbiorem kulek i $f(s)$ jest kolorem kulki s, to $\nu(s)$ jest klasą równoważności $[s]$ składającą się z kulek tego samego koloru co s. Możemy myśleć o $\nu(s)$ jako o woreczku z kulkami, a o ν jako o funkcji, która wkłada kulkę do właściwego woreczka. Przy tej wymyślnej interpretacji $[S]$ jest zbiorem woreczków, przy czym w każdym woreczku znajduje się co najmniej jedna kulka. Liczba woreczków jest równa liczbie użytych kolorów.

Jeśli weźmiemy pod uwagę inną funkcję g, taką, że $g(s)$ jest średnicą kulki s, to otrzymamy nową funkcję ν i podział $[S]$ składający się z nowych woreczków, po jednym woreczku dla każdego możliwego rozmiaru kulki.

(b) Określmy funkcję $f\colon \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ wzorem $f(x,y) = x^2+y^2$. Wtedy funkcja f określa relację równoważności $\sim$ w taki sposób, że $(x,y) \sim (z,w)$ jeśli $x^2+y^2 = z^2+w^2$. Klasami równoważności są okręgi na płaszczyźnie $\mathbb{R} \times \mathbb{R}$ mające środek w punkcie $(0,0)$, gdyż równość $x^2+y^2 = z^2+w^2$ oznacza $\sqrt{x^2+y^2} = \sqrt{z^2+w^2}$, czyli, że punkty (x,y) i (z,w) leżą w tej samej odległości od punktu $(0,0)$. A więc $[\mathbb{R} \times \mathbb{R}]$ składa się z tych okręgów, włączając w to zbiór $\{(0,0)\}$ (czyli okrąg o promieniu 0). Funkcja ν przeprowadza każdy punkt (x,y) na okrąg, na którym ten punkt leży. Istnieje wzajemnie jednoznaczna odpowiedniość pomiędzy tym zbiorem okręgów i zbiorem wartości funkcji f. ∎

PRZYKŁAD 11 Zbiór $\mathbb{Q}$ liczb wymiernych składa się z liczb postaci $\frac{m}{n}$, gdzie $m, n \in \mathbb{Z}$ i $n \neq 0$. Każda liczba wymierna może być tak zapisana na wiele sposobów, na przykład

$$\frac{2}{3} = \frac{4}{6} = \frac{8}{12}, \quad -5 = \frac{-5}{1} = \frac{-10}{2} \quad \text{oraz} \quad 0 = \frac{0}{1} = \frac{0}{73}.$$

Możemy uważać liczby wymierne za klasy równoważności par liczb całkowitych, tak by liczba $\frac{m}{n}$ odpowiadała klasie równoważności pary (m,n). A oto dlaczego tak jest.

Chcemy, by $(m,n) \sim (p,q)$ dokładnie wtedy, gdy $\frac{m}{n} = \frac{p}{q}$, a więc przy założeniu, że n i q są różne od zera definiujemy

$$(m,n) \sim (p,q), \quad \text{gdy} \quad m \cdot q = n \cdot p.$$

Zauważmy, że w tej definicji wykorzystujemy wyłącznie mnożenie liczb całkowitych. Nie występują w niej ułamki. Można łatwo

sprawdzić (por. ćwiczenie 13), że $\sim$ jest relacją równoważności. Widzimy też, że pary (m,n) i (p,q) są równoważne wtedy, gdy stosunek m do n jest taki sam jak stosunek p do q. Klasa $[(2,3)] = [(4,6)] = [(8,12)]$ odpowiada ułamkowi $\frac{2}{3}$ i ten ułamek możemy nawet uważać za inną nazwę tej klasy równoważności. ∎

Twierdzenie 2(b) mówi, że każda relacja równoważności jest wyznaczona przez funkcję, a twierdzenie 2(a) mówi, że klasy równoważności odpowiadają możliwym wartościom tej funkcji. Przekształcenie θ określone wzorem $\theta([s]) = f(s)$ jest przekształceniem wzajemnie jednoznacznym zbioru $[S]$ klas równoważności na zbiór $f(S)$ wartości funkcji f.

Definiowanie funkcji takich jak θ przed chwilą, określonych na zbiorze klas równoważności, wymaga ostrożności. Musimy zawsze sprawdzić, czy definicja funkcji nie zależy naprawdę od tego, który z reprezentantów klasy równoważności został użyty. W przypadku funkcji θ wszystko jest w porządku, gdyż jeśli $[s] = [t]$, to $f(s) = f(t)$, a więc otrzymujemy tę samą wartość $\theta([s])$ niezależnie od tego, czy myślimy o klasie $[s]$ jako o klasie $[s]$ czy o klasie $[t]$. Jednak następny przykład pokaże, jak można popełnić błąd. Pytanie, czy funkcja f jest **dobrze określona**, jest pytaniem o to, czy każda wartość $f(x)$ zależy od tego, czym jest x, a nie od tego, jaką nazwę mu nadaliśmy. Czy zgodnie z regułą określającą funkcję f nadamy tę samą wartość $f(x)$ i $f(y)$, jeśli $x = y$?

PRZYKŁAD 12 (a) Możemy określić funkcję $f\colon \mathbb{Q} \to \mathbb{Q}$ wzorem $f(m/n) = m^2/n^2$, gdyż jeśli $m/n = p/q$, to $m^2/n^2 = p^2/q^2$. Przy oznaczeniach z przykładu 11, ułamek m/n odpowiada klasie równoważności $[(m,n)]$ i jeśli $(m,n) \sim (p,q)$, to $m \cdot q = n \cdot p$, $m^2 \cdot q^2 = n^2 \cdot p^2$ i stąd $(m^2, n^2) \sim (p^2, q^2)$.

(b) Nie możemy określić funkcji $g\colon \mathbb{Q} \to \mathbb{Q}$ wzorem $g(m/n) = m + n$. Gdyby taką funkcję można było określić, to musielibyśmy mieć zarówno $g(\frac{1}{2}) = 1 + 2 = 3$, jak i $g(\frac{2}{4}) = 2 + 4 = 6$. Problem polega na tym, że mamy dwa różne onaczenia, $\frac{1}{2}$ i $\frac{2}{4}$, tego samego obiektu, a nasza definicja funkcji g korzysta z nazwy obiektu, a nie z samego tego obiektu. Jeśli przyjrzymy się temu problemowi od strony klas równoważności, to zobaczymy, że $[(1,2)] = [(2,4)]$. Nie powstałby ten problem, gdybyśmy chcieli zdefiniować funkcję g wzorem $g(m,n) = m + n$, ale ta definicja nie uwzględniałaby klas równoważności. Mianowicie $(m,n) \sim (p,q)$ nie implikuje $m + n = p + q$. Nasza początkowa funkcja g nie była dobrze określona. ∎

ĆWICZENIA DO § 3.5

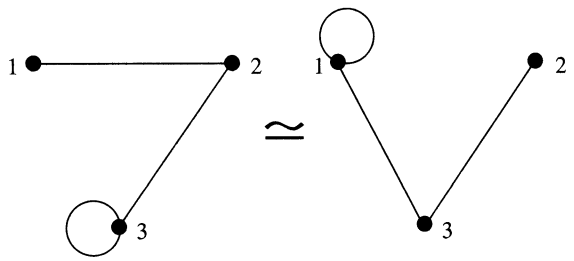
1. Które z poniższych relacji są relacjami równoważności? Jeśli któraś z nich nie jest relacją równoważności, wskaż własności (Z), (S) i (P), których ta relacja nie spełnia i podaj odpowiednie kontrprzykłady.
(a) Relacja $L_1 \parallel L_2$ określona w zbiorze linii prostych na płaszczyźnie, oznaczająca, że proste L_1 i L_2 pokrywają się lub są równoległe.
(b) Relacja $L_1 \perp L_2$ określona w zbiorze linii prostych na płaszczyźnie, oznaczająca, że proste L_1 i L_2 są prostopadłe.
(c) Relacja $p_1 \sim p_2$ określona w zbiorze Amerykanów, oznaczająca, że osoby p_1 i p_2 mieszkają w tym samym stanie.
(d) Relacja $p_1 \approx p_2$ określona w zbiorze Amerykanów, oznaczająca, że osoby p_1 i p_2 mieszkają w tym samym stanie lub w stanach sąsiednich.
(e) Relacja $p_1 \approx p_2$ określona w zbiorze wszystkich ludzi, oznaczająca, że osoby p_1 i p_2 mają wspólnego rodzica.
(f) Relacja $p_1 \cong p_2$ określona w zbiorze wszystkich ludzi, oznaczająca, że osoby p_1 i p_2 mają tę samą matkę.

2. Dla każdej relacji równoważności z ćwiczenia 1 podaj elementy pewnej klasy równoważności.

3. Niech S będzie zbiorem. Czy relacja równości, tzn. „=" jest relacją równoważności?

4. (a) Określmy relację $\equiv$ w zbiorze $\mathbb{Z}$ w następujący sposób: $m \equiv n$ wtedy i tylko wtedy, gdy liczba $m - n$ jest parzysta. Czy relacja $\equiv$ jest relacją równoważności? Odpowiedź uzasadnij.
(b) Powtórz ćwiczenie (a) dla relacji $\approx$ określonej w następujący sposób: $m \approx n$ wtedy i tylko wtedy, gdy $|m - n| \leq 1$.

5. Jeśli G i H są grafami, których zbiorem wierzchołków jest zbiór $\{1, 2, \ldots, n\}$, to mówimy, że graf G jest izomorficzny z grafem H i piszemy $G \simeq H$, wtedy i tylko wtedy, gdy istnieje taki sposób przenumerowania wierzchołków grafu G, by stał się on grafem H. Na przykład dla $n = 3$ dzięki przenumerowaniu $f(1) = 2$, $f(2) = 3$ i $f(3) = 1$ (por. rys. 3.21).



Rysunek 3.21

(a) Podaj rysunek innego grafu izomorficznego z dwoma grafami przedstawionymi na rysunku 3.21.
(b) Pokaż, że relacja $\simeq$ jest relacją równoważności w zbiorze wszystkich grafów mających zbiór wierzchołków $\{1, 2, \ldots, n\}$.

6. Czy możesz podać sytuacje z życia codziennego, w których używasz określenia „równoważne" i w których występuje naturalna relacja równoważności?

7. Określmy relację $\approx$ w zbiorze $\mathbb{Z}$ w następujący sposób: $m \approx n$ wtedy i tylko wtedy, gdy $m^2 = n^2$.
(a) Pokaż, że relacja $\approx$ jest relacją równoważności w zbiorze $\mathbb{Z}$.
(b) Opisz klasy równoważności tej relacji. Ile ich jest?

8. Wykaż, że prawdziwe są stwierdzenia z przykładu 3(a).

9. Weźmy funkcje g i h przekształcające $\mathbb{Z}$ w $\mathbb{N}$, określone wzorami: $g(n) = |n|$ i $h(n) = 1 + (-1)^n$.
(a) Opisz zbiory występujące w podziale $\{g^{\leftarrow}(k)\colon\ k \in \mathbb{N}\}$ zbioru $\mathbb{Z}$. Ile jest tych zbiorów?
(b) Opisz zbiory występujące w podziale $\{h^{\leftarrow}(k)\colon\ k \in \mathbb{N}\}$ zbioru $\mathbb{Z}$. Ile jest tych zbiorów?

10. W zbiorze $\mathbb{N} \times \mathbb{N}$ określamy relację: $(m, n) \sim (k, l)$ wtedy i tylko wtedy, gdy $m + l = n + k$.
(a) Pokaż, że relacja $\sim$ jest relacją równoważności w zbiorze $\mathbb{N} \times \mathbb{N}$.
(b) Zrób szkic zbioru $\mathbb{N} \times \mathbb{N}$ pokazujący klasy równoważności tej relacji.

11. Niech Σ będzie alfabetem i dla w_1 i w_2 w zbiorze Σ^* określamy: $w_1 \sim w_2$ wtedy i tylko wtedy, gdy długość(w_1) = długość(w_2). Wyjaśnij, dlaczego relacja $\sim$ jest relacją równoważności i opisz jej klasy równoważności.

12. Niech P będzie zbiorem programów komputerowych. Programy p_1 i p_2 uważamy za równoważne, jeśli dla takich samych danych dają one zawsze te same wyniki. Czy jest to relacja równoważności? Odpowiedź uzasadnij.

13. Weźmy zbiór $\mathbb{Z} \times \mathbb{P}$ i określmy relację: $(m, n) \sim (p, q)$ wtedy i tylko wtedy, gdy $mq = np$.
(a) Pokaż, że relacja $\sim$ jest relacją równoważności w zbiorze $\mathbb{Z} \times \mathbb{P}$.
(b) Pokaż, że relacja $\sim$ jest relacją równoważności odpowiadającą funkcji $f\colon \mathbb{Z} \times \mathbb{P} \longrightarrow \mathbb{Q}$ danej wzorem $f(m, n) = m/n$; zob. twierdzenie 2(a).

14. W dowodzie twierdzenia 2(b) otrzymaliśmy równość $\nu^{\leftarrow}([s]) = [s]$. Czy to oznacza, że funkcja ν ma funkcję odwrotną i że funkcja odwrotna do ν jest funkcją identycznościową na $[S]$? Wyjaśnij to.

15. Tak jak w ćwiczeniu 7 określamy relację $\approx$ w zbiorze $\mathbb{Z}$ następująco: $m \approx n$ wtedy i tylko wtedy, gdy $m^2 = n^2$.

(a) Co jest złego w następującej „definicji" relacji $\leq$ w zbiorze $[\mathbb{Z}]$? Niech $[m] \leq [n]$ wtedy i tylko wtedy, gdy $m \leq n$.
(b) Co jest złego, jeśli w ogóle jest coś złego, w następującej „definicji" funkcji $f\colon [\mathbb{Z}] \to \mathbb{Z}$? Niech $f([m]) = m^2 + m + 1$.
(c) Powtórz ćwiczenie (b) dla funkcji $g([m]) = m^4 + m^2 + 1$.
(d) Co jest złego, jeśli w ogóle jest coś złego, w następującej „definicji" działania $\oplus$ w zbiorze $[\mathbb{Z}]$? Niech $[m] \oplus [n] = [m+n]$.

15. Które z następujących wyrażeń są dobrze określonymi definicjami funkcji w zbiorze $\mathbb{Q}^+ = \left\{\frac{m}{n}\colon\ m, n \in \mathbb{P}\right\}$?
(a) $f\left(\frac{m}{n}\right) = \frac{n}{m}$.
(b) $g\left(\frac{m}{n}\right) = m^2 + n^2$.
(c) $h\left(\frac{m}{n}\right) = \frac{m^2+n^2}{mn}$.

17. Niech $\sim$ będzie relacją z przykładu 3(b), określoną wzorem: $f \sim g$ wtedy i tylko wtedy, gdy $|f(x) - g(x)| \leq 1$ dla wszystkich $x \in [0,1]$. Określmy relację $\approx$ w następujący sposób: $f \approx g$ wtedy i tylko wtedy, gdy istnieje ciąg funkcji $f = f_1, f_2, \ldots, f_n = g$ taki, że $f_1 \sim f_2$, $f_2 \sim f_3$, $\ldots$, $f_{n-1} \sim f_n$.
(a) Pokaż, że relacja $\approx$ jest relacją równoważności w zbiorze S.
(b) Opisz klasę równoważności funkcji stałej $z(x) = 0$.

18. Niech S będzie zbiorem wszystkich ciągów (s_n) liczb rzeczywistych. Określmy relację: $(s_n) \approx (t_n)$ wtedy i tylko wtedy, gdy zbiór $\{n \in \mathbb{N}\colon\ s_n \neq t_n\}$ jest skończony. Pokaż, że relacja $\approx$ jest relacją równoważności w zbiorze S.

§ 3.6. Algorytm dzielenia i zbiory $\mathbb{Z}_p$

Ten paragraf jest poświęcony tym relacjom równoważności w zbiorze $\mathbb{Z}$, które są powiązane z działaniami algebraicznymi $+$ oraz $\cdot$. Definicje tych relacji wykorzystują dzielenie liczb całkowitych, zaczniemy zatem od zbadania, co dokładnie oznacza dzielenie w zbiorze $\mathbb{Z}$.

Kiedy dzielimy 6 przez 3, by otrzymać w wyniku 2, nie ma problemu; $6 : 3 = 2$ jest innym sposobem powiedzenia, że $6 = 2 \cdot 3$. Jednakże, jeśli dzielimy 7 przez 3, to „nie podzieli się równo". Nie istnieje liczba całkowita q taka, że $7 = q \cdot 3$. W najlepszym razie możemy otrzymać dwie trójki z 7 i resztę 1; $7 = 2 \cdot 3 + 1$. W ogólności, kiedy próbujemy dzielić liczbę całkowitą n przez różną od zera liczbę całkowitą p, możemy co najwyżej oczekiwać podobnego wyniku. Następujące twierdzenie, zapewne dobrze znane,

mówi, że możemy zawsze otrzymać iloraz i resztę i istnieje tylko jedna możliwa odpowiedź.

Algorytm dzielenia

Niech $p \in \mathbb{P}$. Dla każdej liczby całkowitej n istnieje dokładnie jedna para liczb całkowitych q i r spełniających warunki:

$$n = p \cdot q + r \quad \text{oraz} \quad 0 \leq r < p.$$

Liczby q i r są nazywane odpowiednio **ilorazem** i **resztą** w dzieleniu n przez p. Na przykład, jeśli $p = 7$ i $n = 31$, to $31 = 7 \cdot 4 + 3$, a więc $q = 4$ i $r = 3$.

Może wydawać się dziwne, że to twierdzenie nazywamy algorytmem, gdyż jego sformułowanie nie pokazuje żadnej procedury znajdywania q czy r. Jednak to twierdzenie ma tradycyjnie taką nazwę i w większości zastosowań metoda obliczenia nie jest ważna.

Nie podamy teraz dowodu algorytmu dzielenia. Nietrudno udowodnić jednoznaczność q i r (ćwiczenie 19), a ich istnienia można dowieść dość szybko w sposób niekonstruktywny. W paragrafie 4.1 opiszemy algorytm, który dla danych n i p daje w wyniku q i r; ten algorytm oczywiście daje dowód konstruktywny algorytmu dzielenia. Tak czy inaczej, wszyscy wierzymy, że to twierdzenie jest prawdziwe.

W jaki sposób obliczamy q i r? Za pomocą kalkulatora możemy dodawać i mnożyć liczby całkowite i w wyniku otrzymujemy liczby całkowite, ale wynikiem dzielenia na ogół jest pewien ułamek dziesiętny, a nie q i r. To żaden problem. Możemy łatwo otrzymać q i r z tego ułamka dziesiętnego w następujący sposób.

Przepiszmy warunki $n = p \cdot q + r$ i $0 \leq r < p$ w postaci

$$\frac{n}{p} = q + \frac{r}{p} \quad \text{oraz} \quad 0 \leq \frac{r}{p} < 1.$$

Ta nowa wersja mówi, że q jest **częścią całkowitą** liczby n/p, oznaczaną symbolem $\lfloor n/p \rfloor$. Korzystamy tu z symbolu dla funkcji części całkowitej (ang. *floor function*); $\lfloor x \rfloor$ oznacza największą liczbę całkowitą mniejszą lub równą x. Wielkość r/p jest **częścią ułamkową** liczby n/p, tzn. liczbą znajdującą się na prawo od przecinka w liczbie n/p. Zatem, aby otrzymać q, obliczamy $\lfloor n/p \rfloor$ i wtedy możemy obliczyć $r = (n/p - \lfloor n/p \rfloor) \cdot p$.

PRZYKŁAD 1

(a) Niech $n = 31$ i $p = 7$. Kalkulator kieszonkowy daje nam wynik dzielenia 31: 7 $\approx 4{,}429$. Zatem $q = \lfloor 4{,}429 \rfloor = 4$ oraz $r \approx (4{,}429 - 4) \cdot 7 = 0{,}429 \cdot 7 = 3{,}003$, a więc $r = 3$. Sprawdzamy, że $31 = 4 \cdot 7 + 3$.

(b) Niech teraz $p = 7$ i $n = -31$. Wtedy $-31/7 \approx -4,429$. Czy to oznacza, że $q = -4$ i $r = -3$? Nie, gdyż r musi być liczbą nieujemną. Przypomnijmy, że q jest zawsze największą liczbą całkowitą mniejszą lub równą n/p. W naszym przypadku $q = -5$, gdyż $-5 < -4,429 < -4$, a więc $r = -31 - (-5) \cdot 7 = 4$. ∎

Niektóre kalkulatory i większość języków programowania wykonują te czynności za nas, a więc możemy bezpośrednio zażądać podania q i r, używając dwóch zdefiniowanych już funkcji DIV i MOD. Oto ich definicje:

$$n \operatorname{DIV} p = \left\lfloor \frac{n}{p} \right\rfloor \quad \text{oraz} \quad n \operatorname{MOD} p = \left(\frac{n}{p} - n \operatorname{DIV} p \right) \cdot p,$$

a więc

$$n = (n \operatorname{DIV} p) \cdot p + n \operatorname{MOD} p \quad \text{oraz} \quad 0 \leq n \operatorname{MOD} p < p.$$

Będziemy zakładać, że funkcje DIV i MOD spełniają te warunki, nawet jeśli liczba n jest ujemna, ale powinno się starannie sprawdzić to założenie w praktyce, zwłaszcza że definicje tych funkcji w przypadku liczb ujemnych mogą być różne w konkretnych zastosowaniach. Ćwiczenie 18 pokazuje, jak używać funkcji DIV i MOD określonych dla n naturalnych, by móc również posługiwać się nimi dla $n \leq 0$.

Liczby całkowite $n \operatorname{DIV} p$ i $n \operatorname{MOD} p$ są tymi jedynymi liczbami całkowitymi q i r, których istnienie gwarantuje algorytm dzielenia. Dla danej dodatniej liczby całkowitej p, $n \operatorname{MOD} p$ jest resztą otrzymaną z dzielenia n przez p i jest ona nazywana **resztą modulo** p. $\operatorname{MOD} p$ jest funkcją zmiennej n, mimo że piszemy $n \operatorname{MOD} p$ zamiast $(\operatorname{MOD} p)(n)$; zmienna n pojawiała się już w nietypowych miejscach wcześniej, na przykład w oznaczeniach funkcji $|n|$ i $n!$. Wartości funkcji $\operatorname{MOD} p$ znajdują się w zbiorze $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$, a zatem $\operatorname{MOD} p$: $\mathbb{Z} \to \mathbb{Z}_p$. Tak naprawdę, funkcja $\operatorname{MOD} p$ przekształca zbiór $\mathbb{Z}$ na zbiór $\mathbb{Z}_p$, ponieważ $n \operatorname{MOD} p = n$ dla $n \in \mathbb{Z}_p$.

Zbiory $\mathbb{Z}_p$ i przekształcenia $\operatorname{MOD} p$ dla p ze zbioru $\mathbb{P}$ będą szczególnie ważne w rozdziale 12. Odgrywają one również ważną rolę w wielu zastosowaniach, od transmisji sygnałów, poprzez funkcje haszowania (funkcje skrótu) i generatory liczb losowych, do szybkiej grafiki komputerowej.

PRZYKŁAD 2 (a) Mamy $31 \operatorname{MOD} 7 = 3$ oraz $31 \operatorname{DIV} 7 = 4$. Ponadto mamy $(-31) \operatorname{MOD} 7 = 4$ oraz $(-31) \operatorname{DIV} 7 = -5$. Zauważmy, że $(-31) \operatorname{MOD} 7 \neq -(31 \operatorname{MOD} 7)$ i $(-31) \operatorname{DIV} 7 \neq -(31 \operatorname{DIV} 7)$, a więc musimy uważać pisząc $-n \operatorname{MOD} p$ i $-n \operatorname{DIV} p$.

(b) $n \operatorname{MOD} 2$ wynosi 0, jeśli liczba n jest parzysta oraz wynosi 1, jeśli liczba n jest nieparzysta.

(c) $n \operatorname{MOD} 10$ jest ostatnią cyfrą dziesiętną liczby n. ∎

Jesteśmy teraz gotowi, by przyjrzeć się relacjom równoważności w zbiorze $\mathbb{Z}$, związanym z działaniami arytmetycznymi. Dla danej liczby $p \in \mathbb{P}$ mówimy, że liczby całkowite m i n są równoważne (przystają) modulo p, jeśli dają te same reszty przy dzieleniu przez p. Tak więc dla $m, n \in \mathbb{Z}$ określamy

$$m \equiv n \pmod{p} \text{ wtedy i tylko wtedy, gdy } m \operatorname{MOD} p = n \operatorname{MOD} p.$$

Relację $\equiv \pmod{p}$ nazywamy **relacją kongruencji** modulo p. Wyrażenie $m \equiv n \pmod{p}$ czytamy „m przystaje do n modulo p". Aby nie mylić „mod p" z „$\operatorname{MOD} p$", zapamiętaj, że $\operatorname{MOD} p$ jest nazwą funkcji (takiej jak SILNIA), podczas gdy mod p jest częścią oznaczenia relacji.

Kongruencja modulo p jest relacją równoważności w zbiorze $\mathbb{Z}$, określoną przez funkcję $\operatorname{MOD} p\colon \mathbb{Z} \to \mathbb{Z}_p$, zgodnie z twierdzeniem 2 w § 3.5. Zatem

(Z) $m \equiv m \pmod{p}$ dla wszystkich $m \in \mathbb{Z}$.

(S) Jeśli $m \equiv n \pmod{p}$, to $n \equiv m \pmod{p}$.

(P) Jeśli $m \equiv n \pmod{p}$ i $n \equiv r \pmod{p}$, to $m \equiv r \pmod{p}$.

Następne twierdzenie pokazuje, że definicja ta jest zgodna z definicją podaną w § 3.1.

Twierdzenie 1

Niech $p \in \mathbb{P}$. Dla $m, n \in \mathbb{Z}$ mamy

$m \equiv n \pmod{p}$ wtedy i tylko wtedy,

gdy $m - n$ jest wielokrotnością p.

Dowód. Mamy

$$m = (m \operatorname{DIV} p) \cdot p + m \operatorname{MOD} p \quad \text{oraz} \quad n = (n \operatorname{DIV} p) \cdot p + n \operatorname{MOD} p.$$

Jeśli $m \equiv n \pmod{p}$, to mamy $m \operatorname{MOD} p = n \operatorname{MOD} p$, a stąd $m - n = (m \operatorname{DIV} p - n \operatorname{DIV} p) \cdot p$ jest wielokrotnością p. Na odwrót, załóżmy, że $m - n$ jest wielokrotnością p. Wtedy

$$m \operatorname{MOD} p - n \operatorname{MOD} p = (m - n) + (n \operatorname{DIV} p - m \operatorname{DIV} p) \cdot p$$

jest również wielokrotnością p. Ponieważ $m \operatorname{MOD} p$ oraz $n \operatorname{MOD} p$ znajdują się w zbiorze $\{0, 1, \ldots, p-1\}$, więc ich różnica jest równa co najwyżej $p - 1$. Zatem muszą one być równe i stąd mamy $m \equiv n \pmod{p}$. ∎

Ponieważ $m - (m \operatorname{MOD} p) = (m \operatorname{DIV} p) \cdot p$, więc mamy $m \operatorname{MOD} p \equiv m \pmod p$. Ta kongruencja pokazuje historyczne powody dla wprowadzenia oznaczenia MOD. Ponieważ różne elementy $\mathbb{Z}_p$ nie mogą przystawać do siebie modulo p, więc liczba $r = m \operatorname{MOD} p$ jest jedynym elementem $\mathbb{Z}_p$ takim, że $r \equiv m \pmod p$.

Klasę równoważności liczby n ze względu na relację kongruencji $\equiv \pmod p$ nazywamy **klasą reszt modulo** p tej liczby i oznaczamy symbolem $[n]_p$ lub czasami $[n]$, jeśli liczba p jest znana z kontekstu. Zatem

$$[n]_p = \{m \in \mathbb{Z}\colon\ m \equiv n \pmod p\}.$$

Przypadek, gdy $p = 1$ wymaga szczególnego potraktowania i nie jest ciekawy (ćwiczenie 13), tak więc, jeśli nie zaznaczymy tego specjalnie, w dalszym ciągu będziemy zakładać, że $p \geq 2$. Wiele rozumowań zachowa jednak nadal słuszność również w przypadku $p = 1$.

PRZYKŁAD 3 (a) Dwie liczby całkowite przystają do siebie modulo 2, jeśli obie są parzyste lub obie są nieparzyste, tzn. jeśli obie dają tę samą resztę przy dzieleniu przez 2. Klasami równoważności są $[0]_2 = [2]_2 = \ldots = \{n \in \mathbb{Z}\colon$ liczba n jest parzysta$\}$ oraz $[1]_2 = [-3]_2 = [73]_2 = \{n \in \mathbb{Z}\colon$ liczba n jest nieparzysta$\}$.

(b) Liczby całkowite, które są wielokrotnościami 5, mianowicie

$$\ldots, -25, -20, -15, -10, -5, 0, 5, 10, 15, 20, 25, \ldots$$

przystają do siebie modulo 5, ponieważ różnica każdych dwóch liczb z tej listy jest wielokrotnością 5. Wszystkie te liczby dają resztę 0 przy dzieleniu przez 5.

Jeśli dodamy 1 do każdej liczby z tej listy, otrzymamy nową listę:

$$\ldots, -24, -19, -14, -9, -4, 1, 6, 11, 16, 21, 26, \ldots$$

Różnice między liczbami nie zmieniły się, a więc te różnice są nadal wielokrotnościami 5. Na przykład

$$\begin{aligned} 21 - (-14) &= (20 + 1) - (-15 + 1) = 20 + 1 - (-15) - 1 \\ &= 20 - (-15) = 35. \end{aligned}$$

Zatem liczby tej nowej listy również przystają do siebie modulo 5. Wszystkie one dają resztę 1 przy dzieleniu przez 5.

Liczby całkowite

$$\ldots, -23, -18, -13, -8, -3, 2, 7, 12, 17, 22, 27, \ldots$$

tworzą inną klasę równoważności. Podobnie liczby całkowite

$$\ldots, -22, -17, -12, -7, -2, 3, 8, 13, 18, 23, 28, \ldots$$

oraz liczby całkowite

$$\ldots, -21, -16, -11, -6, -1, 4, 9, 14, 19, 24, 29, \ldots$$

Każda liczba całkowita należy do dokładnie jednej z tych pięciu klas, to oznacza, że te klasy tworzą podział zbioru $\mathbb{Z}$, a każda z klas zawiera dokładnie jedną z liczb 0, 1, 2, 3, 4. Możemy wypisać te klasy jako $[0]_5, [1]_5, [2]_5, [3]_5, [4]_5$. ∎

Nasze następne twierdzenie pokazuje związek między przystawaniem modulo p i działaniami arytmetycznymi w zbiorze $\mathbb{Z}$.

Twierdzenie 2

Niech $m, m', n, n' \in \mathbb{Z}$ oraz niech $p \in \mathbb{P}$. Jeśli $m' \equiv m \pmod p$ i $n' \equiv n \pmod p$, to

$$m' + n' \equiv m + n \pmod p \text{ oraz } m' \cdot n' \equiv m \cdot n \pmod p.$$

Dowód. Z założenia $m' = m + k \cdot p$ oraz $n' = n + l \cdot p$ dla pewnych $k, l \in \mathbb{Z}$. Zatem

$$m' + n' = m + n + (k + l) \cdot p \equiv m + n \pmod p \text{ oraz}$$
$$m' \cdot n' = m \cdot n + (k \cdot n + m \cdot l + k \cdot p \cdot l) \cdot p \equiv m \cdot n \pmod p. \quad ∎$$

Przyjmując $m' = m \operatorname{MOD} p$ oraz $n' = n \operatorname{MOD} p$, otrzymujemy następujący przydatny wniosek z twierdzenia 2.

Wniosek

Niech $m, n \in \mathbb{Z}$ i niech $p \in \mathbb{P}$. Wtedy
(a) $m \operatorname{MOD} p + n \operatorname{MOD} p \equiv m + n \pmod p$.
(b) $(m \operatorname{MOD} p) \cdot (n \operatorname{MOD} p) \equiv m \cdot n \pmod p$.

Możemy wykorzystać ten wniosek do przeniesienia działań arytmetycznych ze zbioru $\mathbb{Z}$ do zbioru $\mathbb{Z}_p$. Najpierw w zbiorze $\mathbb{Z}_p$ definiujemy dwa działania $+_p$ oraz $*_p$ w następujący sposób:

$$a +_p b = (a + b) \operatorname{MOD} p \quad \text{oraz} \quad a *_p b = (a \cdot b) \operatorname{MOD} p$$

dla $a, b \in \mathbb{Z}_p$. Ponieważ $m \operatorname{MOD} p \in \mathbb{Z}_p$ dla wszystkich m, więc $a +_p b$ oraz $a *_p b$ należą do zbioru $\mathbb{Z}_p$.

PRZYKŁAD 4

(a) Bardzo prostym, ale ważnym przykładem jest zbiór $\mathbb{Z}_2$. Tablice dodawania i mnożenia w $\mathbb{Z}_2$ podane są na rysunku 3.22.

$+_2$	0	1
0	0	1
1	1	1

$*_2$	0	1
0	0	0
1	0	1

$\mathbb{Z}_2$

Rysunek 3.22

(b) Dla $p = 6$ i $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$ mamy $4 +_6 5 = 3$, ponieważ $(4 + 5) \operatorname{MOD} 6 = 9 \operatorname{MOD} 6 = 3$. Podobnie $4 *_6 4 = 4$, ponieważ $4 \cdot 4 \equiv 4 \pmod 6$. Pełne tablice dodawania i mnożenia w $\mathbb{Z}_6$ podane są na rysunku 3.23. Zauważmy, że iloczyn (w sensie mnożenia $*_6$) elementów różnych od zera może być równy 0.

$+_6$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$*_6$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

$\mathbb{Z}_6$

Rysunek 3.23

(c) Na rysunku 3.24 podane są tablice działań dla zbioru $\mathbb{Z}_5$.

$+_5$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$*_5$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

$\mathbb{Z}_5$

Rysunek 3.24

■

Nowe działania $+_p$ oraz $*_p$ są zgodne ze starymi działaniami $+$ oraz $\cdot$ w zbiorze $\mathbb{Z}$.

Twierdzenie 3

Niech $m, n \in \mathbb{Z}$ oraz $p \in \mathbb{P}$. Wtedy

(a) $(m + n) \operatorname{MOD} p = (m \operatorname{MOD} p) +_p (n \operatorname{MOD} p)$.

(b) $(m \cdot n) \operatorname{MOD} p = (m \operatorname{MOD} p) *_p (n \operatorname{MOD} p)$.

Dowód. (a) Zgodnie z twierdzeniem 2 lub z wnioskiem do tego twierdzenia, $m+n \equiv m \bmod p + n \bmod p \pmod p$. To oznacza, że

$$(m+n) \bmod p = (m \bmod p + n \bmod p) \bmod p,$$

przy czym prawa strona jest z definicji równa $(m \bmod p) +_p (n \bmod p)$. Dowód (b) jest podobny. ∎

Tak więc funkcja $\bmod p$ przeprowadza sumy w zbiorze $\mathbb{Z}$ na sumy (w sensie dodawania $+_p$) w zbiorze $\mathbb{Z}_p$ i iloczyny w zbiorze $\mathbb{Z}$ na iloczyny (w sensie mnożenia $*_p$) w zbiorze $\mathbb{Z}_p$.

PRZYKŁAD 5

(a) $(6+3) \bmod 2 = 9 \bmod 2 = 1$, ale również $6 \bmod 2 +_2 3 \bmod 2 = 0 +_2 1 = 1$. W rzeczywistości, ogólnie (parzysta + nieparzysta) $\bmod 2$ = nieparzysta $\bmod 2 = 1 = 0 +_2 1 =$ parzysta $\bmod 2 +_2$ nieparzysta $\bmod 2$.

(b) $(8 \cdot 3) \bmod 6 = 24 \bmod 6 = 0$ oraz $8 \bmod 6 *_6 3 \bmod 6 = 2 *_6 3 = 0$. ∎

Twierdzenie 3 pozwala również pokazać, że działania $+_p$ oraz $*_p$ spełniają pewne znane prawa algebraiczne.

Twierdzenie 4

Niech $p \in \mathbb{P}$ oraz niech $m, n, r \in \mathbb{Z}_p$. Wtedy
(a) $m +_p n = n +_p m$ oraz $m *_p n = n *_p m$.
(b) $(m +_p n) +_p r = m +_p (n +_p r)$ oraz $(m *_p n) *_p r = m *_p (n *_p r)$.
(c) $(m +_p n) *_p r = (m *_p r) +_p (n *_p r)$.

Dowód. Pokażemy, że zachodzi prawo rozdzielności (c). Inne dowody są podobne (zob. ćwiczenie 21).

Ponieważ $(m+n) \cdot r = m \cdot r + n \cdot r$ w zbiorze $\mathbb{Z}$, więc mamy $(m+n) \cdot r \bmod p = (m \cdot r + n \cdot r) \bmod p$. Na podstawie twierdzenia 3

$$\begin{aligned}((m+n) \cdot r) \bmod p &= (m+n) \bmod p *_p (r \bmod p) \\ &= (m \bmod p +_p n \bmod p) *_p (r \bmod p).\end{aligned}$$

Ponieważ m, n i r należą już do zbioru $\mathbb{Z}_p$, więc $m \bmod p = m$, $n \bmod p = n$ i $r \bmod p = r$. Zatem ostatnia równość po prostu znaczy

$$((m+n) \cdot r) \bmod p = (m +_p n) *_p r.$$

Podobnie

$$(m \cdot r + n \cdot r) \bmod p = (m *_p r) +_p (n *_p r).$$ ∎

Zbiór $\mathbb{Z}_p$ z działaniami $+_p$ oraz $*_p$ zachowuje się, w pewnym sensie, jak skończony model zbioru $\mathbb{Z}$. Musimy jednak być bardziej ostrożni, ponieważ pomimo, że twierdzenie 4 pokazuje, iż wiele praw arytmetyki zachodzi w $\mathbb{Z}_p$, to prawo skracania może nie działać tak jak się tego spodziewamy. Widzieliśmy, że $3 *_6 5 = 3 *_6 3 = 3 *_6 1 = 3$, ale $5 \neq 3 \neq 1$ w $\mathbb{Z}_6$. Ponadto $3 *_6 2 = 0$, przy czym $3 \neq 0$ i $2 \neq 0$. W rozdziale 12 znajduje się dokładniejszy opis $\mathbb{Z}_p$ jako struktury algebraicznej i wyjaśnienie, dlaczego $\mathbb{Z}_5$ i $\mathbb{Z}_7$ zachowują się lepiej niż $\mathbb{Z}_6$.

PRZYKŁAD 6 (a) Moglibyśmy spróbować określić działania + oraz · w rodzinie $[\mathbb{Z}]_p$ klas równoważności $[m]_p$. W naturalny sposób można to zrobić tak:

$$[m]_p + [n]_p = [m+n]_p \text{ oraz } [m]_p \cdot [n]_p = [m \cdot n]_p.$$

Wiemy na podstawie § 3.5, że musimy ostrożnie definiować działania na klasach równoważności. Aby mieć pewność, że działania + oraz · są dobrze zdefiniowane w zbiorze $[\mathbb{Z}]_p$, musimy sprawdzić, że jeśli $[m]_p = [m']_p$ oraz $[n]_p = [n']_p$, to

(1) $[m+n]_p = [m'+n']_p$ oraz

(2) $[m \cdot n]_p = [m' \cdot n']_p$.

Teraz $[m]_p = [m']_p$ wtedy i tylko wtedy, gdy $m' \equiv m \pmod p$ i podobnie $[n]_p = [n']_p$ wtedy i tylko wtedy, gdy $n' \equiv n \pmod p$. Warunek (1) możemy zatem zapisać w postaci $(m'+n') \equiv (m+n) \pmod p$, która wynika z twierdzenia 2. Dowód (2) jest podobny. Zatem nasze nowe działania + oraz · są dobrze zdefiniowane w zbiorze $[\mathbb{Z}]_p$.

(b) Na przykład $[3]_6 + [5]_6 = [8]_6 = [2]_6$ oraz $[3]_5 \cdot [2]_5 = [6]_5 = [1]_5$. Zauważmy dla porównania, że $3 +_6 5 = 2$ oraz $3 *_5 2 = 1$. Nasze działania w zbiorach $[\mathbb{Z}]_6$ i $[\mathbb{Z}]_5$ wyglądają tak jak działania w $\mathbb{Z}_6$ i $\mathbb{Z}_5$. (Zob. ćwiczenie 20, które wszystko wyjaśni.)

Jeśli piszemy PARZYSTA dla oznaczenia klasy $[0]_2$ i NIEPARZYSTA dla oznaczenia klasy $[1]_2$ w zbiorze $[\mathbb{Z}]_2$, to PARZYSTA + NIEPARZYSTA = NIEPARZYSTA, NIEPARZYSTA · NIEPARZYSTA = NIEPARZYSTA itd.

(c) Spróbujmy zdefiniować funkcję $f\colon [\mathbb{Z}]_6 \to \mathbb{Z}$ za pomocą wzoru $f([m]_6) = m^2$. I tak np. $f([2]_6) = 2^2 = 4$, $f([3]_6) = 3^2 = 9$, $f([8]_6) = 64 =$ OJEJ!. Problem polega na tym, że $[8]_6 = [2]_6$, ale $4 \neq 64$. Niedobrze. Nasza funkcja f nie jest dobrze zdefiniowana.

■

ĆWICZENIA DO § 3.6

1. Zastosuj jakąkolwiek metodę do znalezienia q i r takich jak w algorytmie dzielenia dla następujących wartości n i p:
(a) $n = 20$, $p = 3$; (b) $n = 20$, $p = 4$;
(c) $n = -20$, $p = 3$; (d) $n = -20$, $p = 4$;
(e) $n = 371246$, $p = 65$; (f) $n = -371246$, $p = 65$.

2. Znajdź $n \operatorname{DIV} m$ i $n \operatorname{MOD} m$ dla następujących wartości n i m:
(a) $n = 20$, $m = 3$; (b) $n = 20$, $m = 4$;
(c) $n = -20$, $m = 3$; (d) $n = -20$, $m = 4$;
(e) $n = 371246$, $m = 65$; (f) $n = -371246$, $m = 65$.

3. Dla każdej z podanych liczb wypisz trzy liczby całkowite przystające do niej modulo 4:
(a) 0, (b) 1, (c) 2, (d) 3, (e) 4.

4. (a) Podaj wszystkie klasy równoważności relacji przystawania modulo 4 w zbiorze $\mathbb{Z}$.
(b) Ile różnych klas równoważności ma relacja przystawania modulo 73 w zbiorze $\mathbb{Z}$?.

5. Dla każdej z następujących liczb całkowitych m znajdź jedyną liczbę całkowitą r w zbiorze $\{0, 1, 2, 3\}$ taką, że $m \equiv r \pmod 4$:
(a) 17, (b) 7, (c) -7, (d) 2, (e) -88.

6. Oblicz:
(a) $4 +_7 4$, (b) $5 +_7 6$, (c) $4 *_7 4$,
(d) $0 +_7 k$ dla dowolnej liczby $k \in \mathbb{Z}_7$,
(e) $1 *_7 k$ dla dowolnej liczby $k \in \mathbb{Z}_7$.

7. (a) Oblicz $6 +_{10} 7$ i $6 *_{10} 7$.
(b) Opisz słowami $m +_{10} k$ dla dowolnych $m, k \in \mathbb{Z}_{10}$.
(c) Zrób to samo dla $m *_{10} k$.

8. (a) Wypisz elementy zbiorów A_0, A_1 i A_2 określonych wzorem

$$A_k = \{m \in \mathbb{Z}: \ -10 \leq m \leq 10 \quad \text{oraz} \quad m \equiv k \pmod 3\}.$$

(b) Jak wyglądają zbiory A_3, A_4, A_{73}?

9. Wypisz tablice dodawania i mnożenia w zbiorze $\mathbb{Z}_4$.

10. Korzystając z rysunku 3.23 rozwiąż następujące równania z niewiadomą x w zbiorze $\mathbb{Z}_6$:
(a) $1 +_6 x = 0$, (b) $2 +_6 x = 0$, (c) $3 +_6 x = 0$,
(d) $4 +_6 x = 0$, (e) $5 +_6 x = 0$.

11. Korzystając z rysunku 3.24 rozwiąż następujące równania z niewiadomą x w zbiorze $\mathbb{Z}_5$:
(a) $1 *_5 x = 1$, (b) $2 *_5 x = 1$, (c) $3 *_5 x = 1$, (d) $4 *_5 x = 1$.

12. Dla $m, n \in \mathbb{N}$ określamy $m \sim n$ wtedy i tylko wtedy, gdy liczba $m^2 - n^2$ jest wielokrotnością 3.

(a) Pokaż, że $\sim$ jest relacją równoważności w zbiorze $\mathbb{N}$.
(b) Wypisz cztery elementy z klasy równoważności $[0]$.
(c) Wypisz cztery elementy z klasy równoważności $[1]$.
(d) Czy uważasz, że istnieją jakieś inne klasy równoważności?

13. Definicja $m \equiv n \pmod p$ ma sens nawet wtedy, gdy $p = 1$.
(a) Opisz tę relację równoważności dla $p = 1$ i odpowiadające jej klasy równoważności w $\mathbb{Z}$.
(b) Jakie znaczenie możesz nadać liczbom $m \operatorname{DIV} 1$ i $m \operatorname{MOD} 1$?
(c) Co mówi twierdzenie 3, jeśli $p = 1$?

14. (a) Udowodnij, że jeśli $m, n \in \mathbb{Z}$ i $m \equiv n \pmod p$, to $m^2 \equiv n^2 \pmod p$.
(b) Czy funkcja $f\colon [\mathbb{Z}]_p \to [\mathbb{Z}]_p$ określona wzorem $f([n]_p) = [n^2]_p$ jest dobrze określona? Odpowiedź uzasadnij.
(c) Powtórz ćwiczenie (b) dla funkcji $g\colon [\mathbb{Z}]_6 \to [\mathbb{Z}]_{12}$ określonej wzorem $g([n]_6) = [n^2]_{12}$.
(d) Powtórz ćwiczenie (b) dla funkcji $h\colon [\mathbb{Z}]_6 \to [\mathbb{Z}]_{12}$ określonej wzorem $h([n]_6) = [n^3]_{12}$.

15. (a) Pokaż, że czterocyfrowa liczba $n = abcd$ jest podzielna przez 9 wtedy i tylko wtedy, gdy suma jej cyfr $a + b + c + d$ jest podzielna przez 9.
(b) Czy stwierdzenie z ćwiczenia (a) jest prawdziwe dla każdej liczby $n \in \mathbb{P}$, niezależnie od liczby cyfr? Odpowiedź uzasadnij.

16. (a) Pokaż, że liczba $n = abcd$ jest podzielna przez 2 wtedy i tylko wtedy, gdy jej ostatnia cyfra d jest podzielna przez 2.
(b) Pokaż, że liczba $n = abcd$ jest podzielna przez 5 wtedy i tylko wtedy, gdy cyfra d jest podzielna przez 5.

17. Pokaż, że czterocyfrowa liczba $n = abcd$ jest podzielna przez 11 wtedy i tylko wtedy, gdy liczba $a - b + c - d$ jest podzielna przez 11.

18. (a) Pokaż, że jeśli $n \operatorname{MOD} p = 0$, to $(-n) \operatorname{MOD} p = 0$ oraz $(-n) \operatorname{DIV} p = -(n \operatorname{DIV} p)$.
(b) Pokaż, że jeśli $0 < n \operatorname{MOD} p < p$, to $(-n) \operatorname{MOD} p = p - n \operatorname{MOD} p$ oraz $(-n) \operatorname{DIV} p = -(n \operatorname{DIV} p) - 1$.
To ćwiczenie pokazuje, że można łatwo obliczyć $n \operatorname{DIV} p$ i $n \operatorname{MOD} p$ dla ujemnej liczby n za pomocą funkcji $\operatorname{DIV} p$ i $\operatorname{MOD} p$, określonych tylko dla $n \in \mathbb{N}$.

19. Pokaż, że q i r są wyznaczone jednoznacznie przez algorytm dzielenia. To znaczy, pokaż, że jeśli $p, q, r, q', r' \in \mathbb{Z}$, gdzie $p \neq 0$, oraz jeśli

$$q \cdot p + r = q' \cdot p + r', \quad 0 \leq r < p \quad \text{oraz} \quad 0 \leq r' < p,$$

to $q = q'$ i $r = r'$.

20. Niech $p \geq 2$. Określamy przekształcenie $\theta\colon \mathbb{Z}_p \to [\mathbb{Z}]_p$ za pomocą wzoru $\theta(m) = [m]_p$.
(a) Pokaż, że θ jest przekształceniem wzajemnie jednoznacznym zbioru $\mathbb{Z}_p$ na zbiór $[\mathbb{Z}]_p$.

(b) Pokaż, że $\theta(m +_p n) = \theta(m) + \theta(n)$ oraz $\theta(m *_p n) = \theta(m) \cdot \theta(n)$, gdzie działania $+$ oraz $\cdot$ w zbiorze $[\mathbb{Z}]_p$ są określone tak, jak w przykładzie 6(a).

21. (a) Sprawdź, że zachodzi prawo przemienności dla działania $+_p$ w twierdzeniu 4(a).

(b) Sprawdź, że zachodzi prawo łączności dla działania $+_p$ w twierdzeniu 4(b).

(c) Zauważ, że dowody praw przemienności i łączności dla działania $*_p$ są prawie identyczne jak dowody dla działania $+_p$.

To, co jest najważniejsze w tym rozdziale

Aby sprawdzić, czy dobrze rozumiesz treść tego rozdziału:

1. Przekonaj się, że potrafisz zdefiniować i użyć każdego pojęcia i oznaczenia.

2. Podaj przynajmniej jeden powód, dla którego dany temat został omówiony w tym rozdziale.

3. Zastanów się nad co najmniej jednym przykładem ilustrującym dane pojęcie oraz co najmniej jedną sytuacją, w której dany fakt czy metoda są przydatne.

Celem tego przeglądu jest powiązanie każdej idei z możliwie wieloma innymi ideami i możliwie wieloma konkretnymi przykładami. W ten sposób, kiedy zastanawiasz się nad jakimś przykładem czy pojęciem, wszystko, co jest z nim związane, zostanie przypomniane, by mogło być użyte.

Pojęcia i oznaczenia

relacja dwuargumentowa w zbiorze S lub na zbiorze $S \times T$
- relacja zwrotna, przeciwzwrotna, symetryczna, antysymetryczna, przechodnia (dla relacji w zbiorze S)
- relacja odwrotna
- funkcja jako relacja

relacja równoważności
- klasa równoważności (klasa abstrakcji) $[s]$
- podział $[S]$
- przekształcenie naturalne (kanoniczne) ν przeprowadzające element s na $[s]$
- przystawanie modulo p, $\equiv \pmod p$, $[n]_p$

iloraz, DIV p, reszta, MOD p

$\mathbb{Z}_p$, $+_p$, $*_p$

graf skierowany, graf (nieskierowany)
- wierzchołek, krawędź, pętla, krawędzie wielokrotne

początek i koniec krawędzi
droga
długość drogi
droga zamknięta, cykl
droga acykliczna, graf acykliczny
relacja sąsiedztwa
relacja osiągalności
rysunek grafu skierowanego lub grafu nieskierowanego
wykres relacji
macierz
macierz transponowana, suma macierzy, iloczyn macierzy, iloczyn macierzy przez liczbę, macierz przeciwna, macierz odwrotna
macierz kwadratowa, macierz symetryczna
macierz zerowa i jednostkowa
macierz sąsiedztwa grafu lub grafu skierowanego
macierz relacji

Fakty

Relacje równoważności i podziały są dwoma różnymi sposobami podejścia do tego samego pojęcia.

Każda relacja równoważności jest postaci: $a \sim b$ wtedy i tylko wtedy, gdy $f(a) = f(b)$ dla pewnej funkcji f.

Dodawanie macierzy jest łączne i przemienne.

Mnożenie macierzy jest łączne, ale nie jest przemienne.

Potęgi macierzy sąsiedztwa zliczają drogi w grafach skierowanych i nieskierowanych.

Definicje funkcji określonych na klasach równoważności muszą być niezależne od wyboru reprezentantów.

Algorytm dzielenia ($n = p \cdot q + r, 0 \leq r < p$).

Jedynym elementem $x \in \mathbb{Z}_p$ takim, że $x \equiv m \pmod{p}$, jest $m \operatorname{MOD} p$.

Jeśli $m \equiv m' \pmod{p}$ i $n \equiv n' \pmod{p}$, to

$$m + n \equiv m' + n' \pmod{p} \text{ oraz } m \cdot n \equiv m' \cdot n' \pmod{p}.$$

Działania $+_p$ oraz $*_p$ w zbiorze $\mathbb{Z}_p$ naśladują działania $+$ oraz $\cdot$ w zbiorze $\mathbb{Z}$, z wyjątkiem prawa skracania dla mnożenia $*_p$.

4. INDUKCJA I REKURENCJA

Temat tego rozdziału można opisać za pomocą stwierdzenia: „jeśli zaczniemy dobrze i nic złego się nie zdarzy, to zawsze będzie dobrze". Kontekstem, w którym zastosujemy tę obserwację, będzie ciąg zdań, których dowodzimy, następstwo kroków wykonywanych przez algorytm lub ciąg wartości, które obliczamy. W każdym przypadku chcemy być pewni, że wyniki, jakie otrzymujemy, są zawsze poprawne.

W paragrafie 4.1 zbadamy fragmenty algorytmów nazywane pętlami „dopóki", w których powtarza się ciąg kroków tak długo, jak długo spełniony jest pewien określony warunek. W następnym paragrafie zajmujemy się indukcją matematyczną, podstawowym narzędziem służącym do dowodzenia ciągów zdań. W paragrafie 4.3 wprowadzamy definicje rekurencyjne ciągów oraz obliczenia wyrazów ciągów korzystające z poprzednich wyrazów. Następnie w paragrafie 4.4 podajemy metody otrzymywania wzorów jawnych, które można stosować do ciągów określonych za pomocą równań rekurencyjnych pewnych często spotykanych postaci. Zasady indukcji wprowadzone w paragrafie 4.2 mają zastosowanie tylko do pewnych szczególnych sformułowań rekurencyjnych, w paragrafie 4.5 rozszerzymy te metody tak, by dotyczyły bardziej ogólnych problemów rekurencyjnych. Rozdział ten kończymy omówieniem największego wspólnego dzielnika i rekurencyjnego algorytmu Euklidesa służącego do obliczania tego dzielnika. Paragraf 4.6 zawiera również zastosowania tego algorytmu do rozwiązywania kongruencji.

§ 4.1. Niezmienniki pętli

W ostatnim paragrafie rozdziału 3 stwierdziliśmy, że dla każdej pary liczb całkowitych m i n takich, że $m \geq 0$ i $n > 0$ istnieją liczby całkowite q i r (iloraz i reszta), takie, że $m = q \cdot n + r$ i $0 \leq r < n$. W tym paragrafie udowodnimy to stwierdzenie, najpierw pokazując algorytm, którego celem jest skonstruowanie q i r, a następnie dowodząc, że ten algorytm rzeczywiście wykonuje to, co powinien. Algorytm, który pokażemy, ma pewną specjalną cechę, nazywaną pętlą „dopóki"; powoduje ona, że łatwo jest ten algorytm zrozumieć oraz pomaga w przeprowadzeniu dowodu poprawności tego algorytmu. Głównym zadaniem tego paragrafu jest zrozumienie pętli „dopóki" i towarzyszących im oznaczeń. W następnym paragrafie zobaczymy, w jaki sposób zasady logiki związane z tymi pętlami mogą być wykorzystane do wprowadzenia metody indukcji matematycznej, jednej z najsilniejszych technik dowodowych w matematyce.

Oto podstawowy zarys naszego algorytmu dzielenia.

> Odgadnij wartości q i r takie, że $m = q \cdot n + r$.
> Przez cały czas poprawiaj te odgadnięte wartości, nie tracąc własności $m = q \cdot n + r$ dotąd, aż wreszcie zakończysz działanie, gdy $0 \leq r < n$.

Tak naprawdę nasze początkowe przypuszczenie nie jest trudne do wyobrażenia sobie: $q = 0$ i $r = m$ na pewno są dobre. Ta wartość q jest prawdopodobnie za mała, a wartość r jest prawdopodobnie za duża. Od chwili rozpoczęcia działania algorytmu równość $m = q \cdot n + r$ pozostanie prawdziwa, gdy zwiększymy q o 1 i zmniejszymy r o n, gdyż $q \cdot n + r = (q+1) \cdot n + (r-n)$. Mamy nadzieję, że jeśli dokonamy tych zmian wystarczająco wiele razy, to otrzymamy $r < n$, tak jak potrzebowaliśmy.

Możemy uściślić ten szkic algorytmu nadając mu postać, dzięki której rzeczywiście można obliczyć q i r o wymaganych własnościach. Oto ostateczna wersja napisana w stylu naśladującym program komputerowy. Nie przejmuj się. Nie musisz umieć programować, by zrozumieć ideę tego algorytmu.

Algorytm dzielenia

{Dane: liczby całkowite $m \geq 0$ i $n > 0$.}
{Wyniki: liczby całkowite q i r takie, że $q \cdot n + r = m$ oraz $0 \leq r < n$.}
początek
{Zapoczątkowanie.}

$q := 0$
$r := m$
{Zasadnicza część algorytmu.}
dopóki $r \geq n$, wykonuj
  $q := q + 1$
  $r := r - n$
koniec ∎

Zarówno w powyższym algorytmie, jak i w następnych będziemy przyjmować, że zdania zapisane w nawiasach klamrowych {...} są komentarzami i nie są częścią samego algorytmu. Ciąg wierszy algorytmu jest pewnym przepisem. Ten przepis mówi:

1. Nadaj q wartość 0. (Oznaczenie $a := b$ ma następujący sens „nadaj lub określ wartość a równą wartości b".)
2. Nadaj r wartość m.
3. Sprawdź, czy $r \geq n$.
   Jeśli $r \geq n$, to
   zwiększ wartość q o 1,
   zmniejsz wartość r o n oraz
   wróć do punktu 3.
   W przeciwnym przypadku, tzn. jeśli $r < n$,
   pójdź do punktu 4.
4. Stop.

Nadane na początku wartości q i r raczej nie będą dobre, ale powtarzanie kroku 3 ma na celu poprawianie ich dotąd, aż staną się dobre.

Pętla „dopóki" znajduje się tutaj w kroku 3. Ogólnie pętla „dopóki" jest ciągiem kroków w algorytmie lub w programie mającym postać

dopóki g, wykonuj
  S

i oznaczającym

(*) Sprawdź, czy warunek g jest prawdziwy.
Jeśli warunek g jest prawdziwy, to wykonaj to wszystko, co mówi S, a następnie wróć do (*).
W przeciwnym przypadku omiń S i przejdź dalej do tego, co występuje za pętlą w głównym programie.

Warunek g jest pewnym zdaniem, na przykład „$r \geq n$", nazywanym **warunkiem dozoru pętli**, a S jest ciągiem kroków nazywanym **treścią pętli**. Moglibyśmy w S zawrzeć instrukcje

nakazujące programowi opuszczenie pętli i przeskoczenie do innego miejsca w programie, ale bez zmniejszenia ogólności możemy myśleć o S, jako o pewnym fragmencie programu, po wykonaniu którego wracamy do początku pętli, aby ponownie sprawdzić warunek g. Oczywiście S może zmieniać wartość logiczną g; na tym właśnie polega pomysł. Wykonanie kroków S w tej pętli nazywamy **przebiegiem pętli** lub **iteracją** i mówimy, że **pętla kończy się** (lub że **wychodzimy z pętli**), jeśli w pewnym momencie warunek dozoru g nie jest spełniony i kontynuowane jest wykonanie programu za pętlą.

PRZYKŁAD 1 To, co naprawdę wykonuje się w pętli, może zależeć od stanu różnych wielkości w momencie, gdy program lub algorytm wchodzi w pętlę. Poniżej pokazane są pewne proste pętle:

dopóki $n < 5$, wykonuj	dopóki $n \neq 8$, wykonuj	dopóki $A \neq \emptyset$, wykonuj
$n := n + 1$	wypisz n^2	wybierz $x \in A$
wypisz n^2	$n := n + 2$	usuń x z A
(a)	(b)	(c)

Jeśli $n = 0$ w momencie, kiedy program wchodzi w pętlę (a), to liczba n zostaje zastąpiona liczbą 1, wypisana liczba 1, n zastąpiona liczbą 2, wypisana liczba 4, ..., n zastąpiona liczbą 5, wypisana liczba 25, zostaje stwierdzone, że warunek dozoru $n < 5$ nie jest spełniony i wykonuje się następna część programu. Jeśli $n = 4$ na początku pętli, po prostu zostanie wypisana liczba 25 przed wyjściem z pętli. Jeśli na początku $n = 6$, to nic nie zostanie wypisane, ponieważ od razu warunek $6 < 5$ nie jest spełniony.

Jeśli $n = 0$ na początku pętli (b), to zostaną wypisane liczby 0, 4, 16, 36 i program wyjdzie z pętli. Jednakże dla początkowej wartości $n = 1$ w pętli zostaną wypisane liczby 1, 9, 25, 49, 81, 121 itd. Nigdy nie wyjdziemy z pętli, ponieważ mimo, że wartości n się zmieniają, warunek dozoru $n \neq 8$ jest zawsze spełniony. Jak pokazuje ten przykład, wykonywanie pętli nie musi się zakończyć. Pętla (c) jest przykładem nieliczbowym. Jeśli A jest jakimś zbiorem, na przykład zbiorem krawędzi grafu, to w pętli są odrzucane elementy zbioru A, aż nic w nim nie zostanie. Jeśli A jest zbiorem skończonym, wykonywanie pętli zakończy się, w przeciwnym przypadku wykonywanie pętli nie zakończy się. ■

W przypadku algorytmu dzielenia nadaliśmy początkowe wartości q i r zanim weszliśmy w pętlę. Następnie S każe nam zwiększyć q o 1 i zmniejszyć r o n. Poniżej pokazane są kolejne wartości q, r oraz wartość logiczna zdania $r \geq n$ podczas wykonywania tego algorytmu z wartościami początkowymi $m = 17$ i $n = 7$.

$m = 17, n = 7$	q	r	$r \geq n$
Na początku	0	17	Prawda
Po pierwszym przebiegu	1	10	Prawda
Po drugim przebiegu	2	3	Fałsz

Algorytm nie wchodzi w pętlę „dopóki" po raz trzeci, ponieważ warunek dozoru pętli $n \geq r$ nie jest spełniony po drugim przebiegu pętli. Wartościami końcowymi są $q = 2$ i $r = 3$. Ponieważ $17 = 2 \cdot 7 + 3$, więc algorytm daje poprawny wynik dla zadanych wartości początkowych.

W algorytmie dzielenia przedmiotem naszej troski jest to:

1) czy algorytm się zatrzyma po jakimś czasie oraz
2) czy kiedy się zatrzyma, da poprawne wyniki.

Popatrzmy najpierw, dlaczego algorytm się zatrzymuje. Całe działanie algorytmu zawiera się w pętli „dopóki", gdyż zapoczątkowanie nie stwarza żadnych problemów. Każdy przebieg pętli, tzn. każde wykonanie ciągu kroków $q := q + 1$ i $r := r - n$ zabiera tyle samo czasu, a więc musimy pokazać, że algorytm wykonuje tylko skończenie wiele przebiegów pętli. Ważne jest, co się dzieje z r. Na początku $r = m \geq 0$. Każde wykonanie treści pętli zmniejsza wartość r o n. Tak więc kolejnymi wartościami r są m, $m - n$, $m - 2n$ itd. Te liczby całkowite są coraz mniejsze, gdyż $n > 0$. Wcześniej czy później dojdziemy do takiego k, dla którego $r = m - k \cdot n < n$. Wtedy warunek dozoru pętli $r \geq n$ nie będzie spełniony, wyjdziemy z pętli i zakończymy wykonywanie algorytmu.

Skąd wiemy, że osiągniemy takie k? Zbiór liczb naturalnych $\mathbb{N}$ ma następującą ważną własność:

Zasada dobrego uporządkowania

> Każdy niepusty podzbiór zbioru $\mathbb{N}$ ma element najmniejszy.

Z tej zasady wynika w szczególności, że malejące ciągi elementów zbioru $\mathbb{N}$ nie mogą być nieskończone; każdy malejący ciąg $a > b > c > \ldots$ w $\mathbb{N}$ ma skończoną długość, ponieważ żaden z elementów ciągu nie może być mniejszy od najmniejszego elementu. Zatem każdy nieskończony malejący ciąg elementów zbioru $\mathbb{Z}$ musi ostatecznie mieć wyrazy ujemne.

W naszym przypadku ciąg m, $m - n$, $m - 2n$, ... musi zatrzymać się na pewnej wartości $m - k \cdot n \geq 0$ takiej, że $m - (k + 1) \cdot n < 0$, tzn. takiej, że $m - k \cdot n < n$.

Teraz wiemy, że algorytm dzielenia zatrzymuje się, ale dlaczego daje on zawsze poprawne wyniki? Aby odpowiedzieć na to pytanie, zajmiemy się pewnym nowym pomysłem. Powiemy, że zdanie p jest niezmiennikiem pętli

dopóki g, wykonuj
  S

wtedy, gdy spełnia ono następujący warunek:

Jeśli zdania p i g są prawdziwe, zanim wykonamy kroki S, to zdanie p będzie prawdziwe po wykonaniu S.

Następujące twierdzenie wskazuje powód, dla którego niezmienniki pętli są użyteczne.

Twierdzenie o niezmiennikach pętli

Przypuśćmy, że p jest niezmiennikiem pętli „dopóki g, wykonuj S" oraz że zdanie p jest prawdziwe, kiedy wchodzimy w pętlę. Wtedy zdanie p jest prawdziwe po każdej iteracji pętli. Jeśli pętla kończy się, to po jej zakończeniu zdanie p jest prawdziwe, a zdanie g jest fałszywe.

Dowód. Twierdzenie tak naprawdę jest bardzo oczywiste. Z założenia zdanie p jest prawdziwe na początku, tzn. po 0 iteracjach pętli. Przypuśćmy, jeśli to możliwe, że zdanie p jest fałszywe po jakiejś iteracji. Wtedy istnieje pierwszy raz, kiedy zdanie p staje się fałszywe, powiedzmy, po n-tym przebiegu. Oczywiście $n \geq 1$ i zdanie p było prawdziwe po zakończeniu $(n-1)$-go przebiegu, a zatem przed n-tym przebiegiem. Zdanie g również było prawdziwe, gdyż inaczej nie wykonalibyśmy n-tej iteracji. Ponieważ zdanie p jest niezmiennikiem pętli, musi ono także być prawdziwe po n-tym przebiegu, a to jest sprzeczne z wyborem liczby n. Sprzeczność ta pokazuje, że nigdy nie może się zdarzyć taka sytuacja, że zdanie p jest fałszywe na końcu przebiegu pętli, a więc musi być zawsze prawdziwe.

Jeśli pętla zakończy się (ponieważ zdanie g jest fałszywe), zdarzy się to na końcu przebiegu pętli, a więc zdanie p musi być wtedy prawdziwe. ∎

Pomysł, jakiego użyliśmy w tym dowodzie, mógłby być nazwany Zasadą Najmniejszego Przestępcy — jeśli istnieje zły facet, to istnieje najmniejszy zły facet. Jest to odmiana twierdzenia „zawsze jest jakiś pierwszy raz", które tak naprawdę jest innym sposobem wyrażenia zasady dobrego uporządkowania. W naszym dowodzie patrzyliśmy na zbiór

$\{n \in \mathbb{N}$: zdanie p jest fałszywe po n-tym przebiegu$\}$. Z założenia, że ten zbiór jest niepusty, wynika, że ma on element najmniejszy, co w końcu doprowadziło do sprzeczności.

W jaki sposób to twierdzenie pomaga uzasadnić algorytm dzielenia? Poniżej ponownie pokazujemy ten algorytm, dla przypomnienia, z nowym komentarzem dodanym tuż przed pętlą „dopóki".

Algorytm dzielenia

{Dane: liczby całkowite $m \geq 0$ i $n > 0$.}
{Wyniki: liczby całkowite q i r takie, że $q \cdot n + r = m$ oraz $0 \leq r < n$.}
początek
$q := 0$
$r := m$
$\{q \cdot n + r = m$ oraz $r \geq 0.\}$
dopóki $r \geq n$, wykonuj
  $q := q + 1$
  $r := r - n$
koniec

Następnie pokazujemy, że stwierdzenie „$q \cdot n + r = m$ i $r \geq 0$" jest niezmiennikiem pętli. To zdanie jest oczywiście prawdziwe przed wejściem w pętlę „dopóki", ponieważ tak właśnie dobraliśmy q i r. Ponadto, jeśli $q \cdot n + r = m$ i $r \geq n$ oraz jeśli wykonamy treść pętli, aby dostać nowe wartości r' i q', to

$$q' \cdot n + r' = (q + 1) \cdot n + (r - n) = q \cdot n + r = m$$

oraz

$$r' = r - n \geq 0,$$

a więc stwierdzenie to pozostaje prawdziwe po przebiegu pętli. Wiemy również, że pętla się kończy. Twierdzenie o niezmiennikach pętli mówi nam, że po zakończeniu pętli zdania $m = q \cdot n + r$ i $r \geq 0$ są prawdziwe, a zdanie $r \geq n$ jest fałszywe, czyli $r < n$. Są to dokładnie te warunki, które mają spełniać wyniki. Pokazaliśmy więc następujące twierdzenie.

Twierdzenie 1

Dla danych liczb całkowitych m i n, gdzie $m \geq 0$ i $n > 0$, algorytm dzielenia konstruuje liczby całkowite q i r takie, że $m = q \cdot n + r$ oraz $0 \leq r < n$.

Dlaczego wybraliśmy akurat taki niezmiennik pętli? Chcieliśmy otrzymać $m = q \cdot n + r$ i $0 \leq r < n$. Oczywiście każde ze zdań „$m = q \cdot n + r$" oraz „$0 \leq r$ jest niezmiennikiem naszej pętli, ale żadne z nich osobno nie jest wystarczająco dobre. Chcemy mieć je oba, zatem bierzemy oba. Nie możemy użyć zdania „$r < n$" jako niezmiennika naszej pętli, ale jest ono zaprzeczeniem naturalnego warunku dozoru pętli „$r \geq n$", a więc będzie ono prawdziwe wtedy, gdy warunek dozoru pętli nie będzie spełniony po wyjściu z pętli.

Niezmienniki pętli mogą być używane do projektowania algorytmów, jak również pomagają w dowodzeniu ich poprawności. Pokażemy to na przykładzie algorytmu dzielenia.

Na początku tego paragrafu widzieliśmy szkic algorytmu. Nieco poprawiona wersja tego szkicu mogłaby wyglądać następująco:

początek
zapoczątkuj q i r tak, by $m = q \cdot n + r$
dopóki $r \geq n$, wykonuj
  zmierzaj w kierunku zatrzymania oraz
  zachowuj równość $m = q \cdot n + r$
koniec.

Następna wersja mogłaby być nieco dokładniejsza

początek
niech $q := 0$
wybierz r tak, by $m = q \cdot n + r$
dopóki $r \geq n$, wykonuj
  zwiększ q
  zachowaj równość $m = q \cdot n + r$ {zatem r zmniejszy się}
koniec.

Na tym etapie jest oczywiste, że na początku r musi być równe m oraz że jeśli q zwiększa się o 1 w pętli, to r musi zmniejszać się o n. Ostatecznie warunek $r \geq 0$ wynika z naszego żądania, by na początku $m \geq 0$. Projektując algorytm w taki sposób, można łatwo pokazać, używając niezmiennika, że algorytm da poprawny wynik.

Jest to odpowiedni moment, aby podkreślić, że podany przez nas algorytm dzielenia jest wolniejszy niż metoda dzielenia pisemnego, której uczyliśmy się w dzieciństwie. Powodem, dla którego szkolna metoda dzielenia i jej komputerowe wersje są znacznie szybsze, jest to, że wykorzystują one reprezentację danych

wejściowych w systemie pozycyjnym, dziesiętnym lub dwójkowym; pewne prace organizacyjne zostały już wykonane. Algorytm, który podaliśmy, mógłby również być bardzo przyspieszony za pomocą poprawionej strategii odgadywania nowych wartości q i r. Ponieważ jednak chcieliśmy wiedzieć, że właściwe q i r mogą być obliczone, świadomie wybraliśmy nieskomplikowaną metodę obliczenia.

Oczywiście wypisanie po prostu jakiegoś zdania p w nawiasach klamrowych nie czyni go niezmiennikiem, podobnie napisanie programu nie powoduje, że robi on to, co ma robić. Aby skorzystać z twierdzenia o niezmiennikach pętli, musimy sprawdzić, że zdanie p rzeczywiście spełnia definicję niezmiennika.

PRZYKŁAD 2 (a) W paragrafie 1.5 zdefiniowaliśmy $n! = 1 \cdot 2 \cdot \ldots \cdot n$. Jeśli $n > 1$, to $n! = (n-1)! \cdot n$, co pozwala nam obliczać wartości $n!$ za pomocą prostego algorytmu. Poniżej pokazany jest taki algorytm.

```
{Dane: liczba całkowita n > 0}
{Wyniki: liczba całkowita SILNIA taka, że SILNIA = n!}
początek
m := 1
SILNIA := 1
{SILNIA = m!}
dopóki m < n, wykonuj
  m := m + 1
  SILNIA := SILNIA · m
koniec
```

Zamierzonym niezmiennikiem pętli jest „SILNIA $= m!$". Ta równość jest z pewnością prawdziwa na początku oraz jeśli SILNIA $= m!$ na początku pętli „dopóki", to

$$\begin{aligned}(\text{nowa SILNIA}) &= (\text{stara SILNIA}) \cdot (\text{nowe } m) \\ &= m! \cdot (m+1) = (m+1)! = (\text{nowe } m)!,\end{aligned}$$

czego żądamy na końcu pętli. Pętla kończy się, gdy $m = n$, a więc, gdy SILNIA $= n!$. Z chwilą, gdy wpadliśmy na pomysł użycia zdania SILNIA $= m!$ jako niezmiennika pętli, napisanie algorytmu było już proste.

Jeśli skorzystamy z umowy, że $0! = 1$, możemy zapoczątkować algorytm krokiem $m := 0$ zamiast $m := 1$ i nadal otrzymywać poprawne wyniki.

Zwróćmy uwagę na najwyraźniej nieuniknioną własność tego algorytmu; jeśli chcemy obliczyć tylko 150!, to musimy najpierw obliczyć 1!, 2!, ..., 149!.

(b) Poniżej widzimy inny algorytm, służący do obliczania $n!$, tym razem wykonujący mnożenia w odwrotnym porządku, od góry do dołu.

{Dane: liczba całkowita $n \geq 0$}
{Wyniki: liczba całkowita ODWR taka, że ODWR $= n!$}
początek
ODWR := 1
$m := n$
{ODWR $\cdot\, m! = n!$}
dopóki $m > 0$, wykonuj
  ODWR := ODWR $\cdot\, m$
  $m := m - 1$
koniec

Znów sprawdzenie niezmiennika pętli jest łatwe. Na początku jest on oczywiście prawdziwy i jeśli ODWR $\cdot\, m! = n!$ na początku przebiegu pętli, to

$$\begin{aligned}(\text{nowy ODWR}) \cdot (\text{nowe } m)! &= (\text{ODWR} \cdot m) \cdot (m-1)! \\ &= \text{ODWR} \cdot m! = n!\end{aligned}$$

na końcu tego przebiegu. Warunek dozoru pętli powoduje, że pętla kończy się we właściwym momencie, gdy $m = 1$, a więc daje poprawny wynik nawet, jeśli $n = 0$.

Ten algorytm nie oblicza 1!, 2!, ... podczas obliczania $n!$, ale to co robi, jest równie złe, jeśli nie gorsze. Nadal musimy obliczyć wiele iloczynów częściowych, w tym przykładzie od samego początku całkiem sporych, zanim otrzymamy wynik. ∎

Przykłady, które widzieliśmy, zaczynają pokazywać znaczenie niezmiennika pętli. W następnym paragrafie skorzystamy z twierdzenia o niezmiennikach pętli, by stworzyć metodę dowodzenia przez indukcję matematyczną, jedną z najważniejszych technik w tej książce. Późniejsze rozdziały pokażą nam, w jaki sposób korzystanie z niezmienników pętli pomoże nam zrozumieć działanie niektórych dość skomplikowanych algorytmów, jak również sprawdzić, że dają one poprawne wyniki.

Zarówno w algorytmie dzielenia, jak i w naszym pierwszym algorytmie obliczającym $n!$ w przykładzie 2(a) występowały zmienne zwiększające się o 1 w każdym przebiegu pętli. W przypadku algorytmu dzielenia nie wiedzieliśmy na początku, ile wykonamy przebiegów pętli, ale w przykładzie 2(a) wiedzieliśmy, że m przyjmie wartości $1, 2, 3, \ldots, n$ w tej kolejności i że algorytm

wtedy się zatrzyma. Tego rodzaju przewidywalne zwiększanie o 1 można opisać za pomocą wygodnej notacji.

Instrukcja algorytmiczna „dla k od m do n wykonuj S" każe nam podstawiać liczby $m, m+1, \ldots, n$ zamiast k, w tej kolejności, i za każdym razem wykonać S. Zatem fragment algorytmu

SILNIA := 1
dla k od 1 do n wykonuj
  SILNIA := SILNIA $\cdot\, k$

daje ostatecznie w wyniku SILNIA $= n!$.

Używając pętli „dopóki" możemy przepisać fragment „dla k od m do n wykonuj S" jako

$k := m$
dopóki $k \leq n$, wykonuj
  S
  $k := k + 1$,

z jednym ważnym i oczywistym ostrzeżeniem: fragment S nie może zmieniać wartości zmiennej k. Jeśli S zmienia k, powinniśmy zastąpić k w algorytmie jakąś inną zmienną, różną od m i n, której S już nie zmienia. Ćwiczenie 13 pokazuje, co złego może się zdarzyć w przeciwnym przypadku.

W niektórych językach programowania znajduje się konstrukcja „dla k od n w dół do m wykonuj S", oznaczająca podstawianie $n, n-1, \ldots, m$ w miejsce k, w porządku malejącym, i wykonywanie S za każdym razem. Nasz drugi algorytm obliczania silnii mógłby być zapisany przy użyciu tej konstrukcji.

PRZYKŁAD 3 (a) Oto algorytm obliczania sumy

$$\sum_{k=1}^{n} k^2 = 1 + 4 + 9 + \cdots + n^2,$$

gdzie n jest daną liczbą całkowitą dodatnią:

$s := 0$
dla k od 1 do n wykonuj
  $s := s + k^2$

(b) Ogólnie, algorytm

$s := 0$
dla k od m do n wykonuj
  $s := s + a_k$

oblicza $\sum_{k=m}^{n} a_k = a_m + \ldots + a_n$. ∎

ĆWICZENIA DO § 4.1

1. Wypisz pięć pierwszych wartości x dla każdego z następujących fragmentów algorytmów.

(a) $x := 0$
dopóki $0 \leq x$, wykonuj
$x := 2x + 3$

(b) $x := 1$
dopóki $0 \leq x$, wykonuj
$x := 2x + 3$

(c) $x := 1$
dopóki $0 \leq x$, wykonuj
$x := 2x - 1$

2. Znajdź liczbę całkowitą b taką, że ostatnią wypisaną liczbą będzie 6.

(a) $n := 0$
dopóki $n < b$, wykonuj
wypisz n
$n := n + 1$

(b) $n := 0$
dopóki $n < b$, wykonuj
$n := n + 1$
wypisz n

3. (a) Wypisz wartości m i n podczas wykonywania następującego algorytmu, używając sposobu pokazanego po przykładzie 1:
początek
$m := 0$
$n := 0$
dopóki $n \neq 4$, wykonuj
$m := m + 2 \cdot n + 1$
$n := n + 1$
koniec

(b) Zmodyfikuj algorytm z ćwiczenia (a) tak, by $m = 17^2$, kiedy algorytm się zatrzyma.

4. Zamień ze sobą linie „$n := n + 1$" oraz „wypisz n^2" w pętli (a).
(a) Podaj ciąg wypisywanych wartości, jeśli $n = 0$ na początku pętli.
(b) Zrób to samo dla $n = 4$.

5. Zamień ze sobą linie „wypisz n^2" oraz „$n := n + 2$" w pętli (b).
(a) Podaj ciąg wypisywanych wartości, jeśli $n = 0$ na początku pętli.
(b) Zrób to samo dla $n = 1$.

6. (a) Napisz fragment algorytmu
dla i od 1 do 17 wykonuj

$k := k + 2i$
używając pętli „dopóki".

(b) Powtórz ćwiczenie (a) dla fragmentu algorytmu
dla k od 8 w dół do 1 wykonuj
$i := i + 2k$

7. Pokaż, że podane warunki są niezmiennikami pętli
dopóki $1 \leq n$, wykonuj
$m := m + 1$
$n := n + 1$
(a) $m + n$ jest liczbą parzystą.
(b) $m + n$ jest liczbą nieparzystą.

8. Pokaż, że podane warunki są niezmiennikami pętli
dopóki $1 \leq m$, wykonuj
$m := 2m$
$n := 3n$
(a) $n^2 \geq m^3$.
(b) $2m^6 < n^4$.

9. Weźmy pętlę
dopóki $j \geq n$, wykonuj
$i := i + 2$
$j := j + 1$,
gdzie i i j są liczbami całkowitymi.
(a) Czy warunek $i < j^2$ jest niezmiennikiem pętli, jeśli $n = 1$? Odpowiedź uzasadnij.
(b) Czy warunek $i < j^2$ jest niezmiennikiem pętli, jeśli $n = 0$? Odpowiedź uzasadnij.
(c) Czy warunek $i \leq j^2$ jest niezmiennikiem pętli, jeśli $n = 0$? Odpowiedź uzasadnij.
(d) Czy warunek $i \geq j^2$ jest niezmiennikiem pętli, jeśli $n = 0$? Odpowiedź uzasadnij.

10. Weźmy pętlę
dopóki $k \geq 1$, wykonuj
$k := 2k$
(a) Czy warunek $k^2 \equiv 1 \pmod 3$ jest niezmiennikiem pętli? Odpowiedź uzasadnij.
(b) Czy warunek $k^2 \equiv 1 \pmod 4$ jest niezmiennikiem pętli? Odpowiedź uzasadnij.

11. Dla jakich wartości liczby całkowitej b każdy z następujących algorytmów zakończy działanie?
(a) początek
$k := b$
dopóki $k < 5$, wykonuj
$k := 2k - 1$
koniec

(b) początek
  $k := b$
dopóki $k \neq 5$, wykonuj
  $k := 2k - 1$
koniec

(c) początek
  $k := b$
dopóki $k < 5$, wykonuj
  $k := 2k + 1$
koniec

12. Przypuśćmy, że zmienimy treść pętli z przykładu 2(a) na

SILNIA := SILNIA · m

$m := m + 1$.

Jak powinno się zmienić zapoczątkowanie oraz warunek dozoru pętli, aby w wyniku otrzymywać ten sam rezultat, SILNIA = $n!$? Niezmiennik pętli może się zmienić.

13. Czy pokazane tutaj dwa algorytmy dają ten sam wynik? Odpowiedź uzasadnij.

Algorytm A

początek

$k := 1$

dopóki $k \leq 4$, wykonuj

  $k := k^2$

  wypisz k

  $k := k + 1$

koniec

Algorytm B

początek

dla k od 1 do 4 wykonuj

  $k := k^2$

  wypisz k

koniec

14. Czy dwa algorytmy pokazane tutaj dają ten sam wynik? Odpowiedź uzasadnij.

Algorytm C

początek
$m := 1$
$n := 1$
dopóki $1 \leq m \leq 3$, wykonuj
  dopóki $1 \leq n \leq 3$, wykonuj
    $m := 2m$
    $n := n + 1$
  wypisz m
koniec

Algorytm D

początek
$m := 1$
$n := 1$
dopóki $1 \leq m \leq 3$ i $1 \leq n \leq 3$, wykonuj
  $m := 2m$
  $n := n + 1$
  wypisz m
koniec

15. Oto prymitywny generator liczb losowych, generujący liczby należące do zbioru $\{0, 1, 2, \ldots, 72\}$:

początek
$r := c$
dopóki $r > 0$, wykonuj
  $r := 31 \cdot r \text{ MOD } 73$
koniec

Na przykład, jeśli $c = 3$, otrzymujemy 3, 20, 36, 21, 67, 33, ... Które z następujących warunków są niezmiennikami pętli?
(a) $r < 73$.
(b) $r \equiv 0 \pmod 5$.
(c) $r = 0$.

16. (a) Pokaż, że warunek $S = I^2$ jest niezmiennikiem pętli

dopóki $1 \leq I$, wykonuj
  $S := S + 2I + 1$
  $I := I + 1$

(b) Pokaż, że warunek $S = I^2 + 1$ jest również niezmiennikiem pętli z ćwiczenia (a).
(c) Jaka będzie siedemdziesiąta trzecia liczba wypisana przez następujący algorytm? Odpowiedź uzasadnij.

początek
$S := 1$
$I := 1$
dopóki $1 \leq I$, wykonuj
  wypisz S

$S := S + 2l + 1$
$I := I + 1$
koniec

(d) Jaka będzie siedemdziesiąta trzecia liczba wypisana przez algorytm, w którym na początku będzie $S := 2$? Odpowiedź uzasadnij.

17. Które z następujących zbiorów liczb całkowitych mają najmniejszy element na podstawie zasady dobrego uporządkowania? Odpowiedź uzasadnij.

(a) $\mathbb{P}$.
(b) $\mathbb{Z}$.
(c) $\{n \in \mathbb{P}\colon n^2 > 17\}$.
(d) $\{n \in \mathbb{P}\colon n^2 < 0\}$.
(e) $\{n \in \mathbb{Z}\colon n^2 > 17\}$.
(f) $\{n^2\colon\ n \in \mathbb{P} \text{ i } n! > 80^n\}$.

18. (a) Co otrzymalibyśmy, gdybyśmy spróbowali zastosować algorytm z przykładu 2(a) dla $n = -3$ lub dla $n = -73$? Czy otrzymane wyniki są sensowne?

(b) Odpowiedz na to samo pytanie dla algorytmu z przykładu 2(b).

19. Weźmy następującą pętlę, gdzie $a, b \in \mathbb{P}$. (Przypominamy Czytelnikowi, że $m = (m \text{ DIV } n) \cdot n + (m \text{ MOD } n)$, gdzie $0 \leq (m \text{ MOD } n) < n$.)

dopóki $r > 0$, wykonuj
$a := b$
$b := r$
$r := a \text{ MOD } b$

Które z następujących warunków są niezmiennikami tej pętli? Odpowiedź uzasadnij.

(a) a, b i r są wielokrotnościami 5.
(b) a jest wielokrotnością 5.
(c) $r < b$.
(d) $r \leq 0$.

20. (a) Czy warunek $5^k < k!$ jest niezmiennikiem następującej pętli?
dopóki $4 \leq k$, wykonuj
$k := k + 1$

(b) Czy możesz stąd wywnioskować, że $5^k < k!$ dla wszystkich $k \geq 4$?

21. Oto algorytm, który rozkłada liczbę całkowitą na iloczyn liczby nieparzystej i potęgi 2.

{Dane: liczba całkowita dodatnia n}
{Wyniki: nieujemne liczby całkowite k i m takie, że m jest liczbą nieparzystą oraz $m \cdot 2^k = n$}
początek
$m := n$
$k := 0$
dopóki m jest liczbą parzystą, wykonuj

$m := \frac{m}{2}$ oraz $k := k + 1$
koniec

(a) Pokaż, że warunek $m \cdot 2^k = n$ jest niezmiennikiem pętli.
(b) Wyjaśnij dlaczego algorytm się zatrzyma i pokaż, że po wyjściu z pętli m jest liczbą nieparzystą.

22. Przypuśćmy, że zarówno p, jak i q są niezmiennikami pętli
dopóki g, wykonuj S.
(a) Czy warunek $p \wedge q$ jest niezmiennikiem pętli? Uzasadnij to.
(b) Czy warunek $p \vee q$ jest niezmiennikiem pętli? Uzasadnij to.

23. Następujący algorytm podaje szybką metodę podnoszenia liczby do potęgi.

{Dane: liczba a i $n \in \mathbb{N}$}
{Wynik: $p = a^n$}
początek
$p := 1$
$q := a$
$i := n$
dopóki $i > 0$, wykonuj
jeśli i jest liczbą nieparzystą, to $p := p \cdot q$
{Wykonaj następne dwa kroki niezależnie od tego, czy i jest liczbą nieparzystą.}
$q := q \cdot q$
$i := i$ DIV 2
koniec

(a) Podaj tabelę podobną do tabeli znajdującej się po przykładzie 1, pokazującą kolejne wartości p, q oraz i dla danych $a = 2$ i $n = 11$.
(b) Sprawdź, że warunek $q^i \cdot p = a^n$ jest niezmiennikiem pętli oraz że $p = a^n$ po wyjściu z pętli.

24. (a) Weźmy następujący algorytm, gdzie dane są $n \in \mathbb{P}$, $x \in \mathbb{R}$.

początek
$m := n$
$y := x$
dopóki $m \neq 0$, wykonuj
treść pętli
koniec

Przypuśćmy, że „treść pętli" jest dobrana tak, że warunek „$x^n = x^m \cdot y$ oraz $m \geq 0$" jest niezmiennikiem pętli „dopóki". Jaka jest wartość y w momencie, kiedy algorytm kończy działanie, jeśli w ogóle się zatrzyma?
(b) Czy można tak dobrać z, by niezmiennikiem pętli

dopóki $m \neq 0$, wykonuj
$m := m - 1$
$y := z$

był warunek „$x^n = x^m \cdot y$ oraz $m \geq 0$"? Odpowiedź uzasadnij.

§ 4.2. Indukcja matematyczna

W tym paragrafie stworzymy podbudowę do dowodzenia wielu zdań jednocześnie. Przedstawimy główny pomysł, wynikający z zasady dobrego uporządkowania zbioru $\mathbb{N}$, pomysł będący naturalną konsekwencją twierdzenia o niezmiennikach pętli z § 4.1.

PRZYKŁAD 1 Twierdzimy, że liczba $37^{500} - 37^{100}$ jest wielokrotnością 10. Aby to sprawdzić, można byłoby obliczyć $37^{500} - 37^{100}$ i jeśli ostatnią cyfrą byłoby 0, ogłosić zwycięstwo. Trudność polega na tym, że 37^{500} jest całkiem sporą liczbą — ma ona ponad 780 cyfr dziesiętnych — więc obliczenie jej mogłoby trochę potrwać. Ponieważ nie twierdzimy, że znamy dokładną wartość liczby $37^{500} - 37^{100}$, a twierdzimy tylko, że jest ona wielokrotnością 10, być może można się o tym przekonać w prostszy sposób.

Ponieważ 37 jest liczbą nieparzystą, zatem liczby 37^{500} i 37^{100} również są nieparzyste, a więc wiemy przynajmniej, że liczba $37^{500} - 37^{100}$ jest wielokrotnością 2. Aby dowieść naszego twierdzenia, musimy pokazać, że jest ona również wielokrotnością 5. Zauważmy, że $37^{500} = (37^{100})^5$, zatem mamy do czynienie z liczbą $(37^{100})^5 - 37^{100}$. Być może $n^5 - n$ jest zawsze wielokrotnością 5 dla $n \in \mathbb{P}$. Jeśli tak jest, możemy otrzymać nasze twierdzenie podstawiając $n = 37^{100}$.

Eksperymenty na małych liczbach wyglądają obiecująco. Na przykład mamy $1^5 - 1 = 0$, $2^5 - 2 = 30$, $3^5 - 3 = 240$, a nawet $17^5 - 17 = 1419840$. A więc jest jakaś nadzieja.

Teraz czas na główną sztuczkę. Budujemy prostą pętlę, tak naprawdę prostą maszynkę, której zadaniem jest sprawdzenie, czy $n^5 - n$ jest wielokrotnością 5 dla każdego $n \in \mathbb{P}$, poczynając od $n = 1$ i przebiegając kolejne liczby przynajmniej do $n = 37^{100}$. Udzielimy tej maszynce pewnej pomocy. Poniżej jest pokazana ta pętla. Jej wynik jest nieistotny; to, co jest istotne, to jej niezmiennik.

początek
$n := 1$
dopóki $n < 37^{100}$, wykonuj
  jeśli $n^5 - n$ jest wielokrotnością 5, to
    $n := n + 1$
koniec

Ta pętla „dopóki" po prostu sprawdza, czy $n^5 - n$ jest wielokrotnością 5 oraz jeśli jest to prawdą, przechodzi do następ-

nej wartości n. Sprawdziliśmy już, że $1^5 - 1$ jest wielokrotnością 5, a więc przewidywany niezmiennik pętli jest prawdziwy, kiedy wchodzimy w pętlę. Twierdzimy, że algorytm się kończy (dla $n = 37^{100}$) oraz, że warunek „$n^5 - n$ jest wielokrotnością 5" jest niezmiennikiem pętli. Jeśli te twierdzenia są prawdziwe, to oczywiście liczba $37^{500} - 37^{100}$ jest wielokrotnością 5 na podstawie twierdzenia o niezmiennikach pętli.

Rozważmy jakiś przebieg pętli, np. dla $n = k < 37^{100}$. Jeśli $k^5 - k$ nie jest wielokrotnością 5, to w pętli nie wykona się nic i algorytm po prostu powtarza pętlę w nieskończoność, ze stałą wartością n równą k. Z drugiej strony, jeśli $k^5 - k$ jest wielokrotnością 5, to zdanie warunkowe jest prawdziwe, wartość n zwiększa się w pętli do $k + 1$ i algorytm wraca, by sprawdzić, czy warunek dozoru pętli $n < 37^{100}$ jest nadal prawdziwy. Aby być pewnym, że algorytm się zatrzyma, chcemy wiedzieć, że w każdej iteracji zdanie warunkowe jest prawdziwe.

Teraz udzielimy algorytmowi pewnej pomocy. Jeśli wykonuje on $(k + 1)$-szy przebieg pętli, to liczba $k^5 - k$ musiała być wielokrotnością 5. Podamy teraz prosty dowód pokazujący, że ten fakt powoduje, iż również liczba $(k+1)^5 - (k+1)$ jest wielokrotnością 5. Powodem jest to, że (wykonując przekształcenia algebraiczne) mamy

$$\begin{aligned}(k+1)^5 - (k+1) &= k^5 + 5k^4 + 10k^3 + 10k^2 + 5k + 1 - k - 1 \\ &= (k^5 - k) + 5(k^4 + 2k^3 + 2k^2 + k).\end{aligned}$$

Jeśli $k^5 - k$ jest wielokrotnością 5, to ponieważ drugi składnik jest oczywiście wielokrotnością 5, liczba $(k+1)^5 - (k+1)$ musi także być wielokrotnością 5. To wszystko oznacza, że możemy kazać algorytmowi nie przejmować się sprawdzaniem za każdym razem zdania warunkowego; liczba $n^5 - n$ będzie zawsze wielokrotnością 5, ponieważ była taka poprzednim razem.

To rozumowanie wystarczy, by pokazać, że ten warunek jest niezmiennikiem pętli. Ponadto, podczas każdego przebiegu pętli sprawdzane zdanie warunkowe jest prawdziwe, więc n zwiększa się za każdym razem o 1. Ostatecznie $n = 37^{100}$ i pętla kończy się, czego mieliśmy dowieść.

Pętla, którą napisaliśmy, wygląda być może na głupi algorytm. Aby sprawdzić, że $37^{500} - 37^{100}$ jest wielokrotnością 5, wydaje się, że musimy najpierw sprawdzić wszystkie liczby $1^5 - 1$, $2^5 - 2$, $3^5 - 3$, ..., $(37^{100} - 1)^5 - (37^{100} - 1)$. To rzeczywiście byłoby głupie. Ale tak naprawdę jedyną liczbą, którą musieliśmy sprawdzić, była liczba $1^5 - 1$. Dla wszystkich innych podaliśmy

krótkie algebraiczne uzasadnienie, że jeśli $k^5 - k$ jest wielokrotnością 5, to również $(k+1)^5 - (k+1)$ jest wielokrotnością 5. Nigdy nie zamierzaliśmy wykonywać tego algorytmu, był on jedynie stworzony po to, by dać dowód tego, że liczba $37^{500} - 37^{100}$ jest wielokrotnością 5, bez potrzeby obliczania którejkolwiek z tych wielkich potęg. Tak naprawdę, drobna zmiana w algorytmie pokazuje, za pomocą tego samego dowodu, że 5 jest dzielnikiem liczby $73^{5000} - 73^{1000}$. Rzeczywiście, możemy pokazać, że $n^5 - n$ jest wielokrotnością 5 dla każdej liczby $n \in \mathbb{P}$, stosując dokładnie to samo rozumowanie do pokazanej poniżej pętli, która nigdy się nie zakończy. ∎

$n := 1$
dopóki $1 \leq n$, wykonuj
  jeśli $n^5 - n$ jest wielokrotnością 5, to
    $n := n + 1$

Przypuśćmy ogólnie, że mamy skończony ciąg zdań, powiedzmy $p(m)$, $p(m+1)$, ..., $p(n)$, gdzie $m, m+1, \ldots, n$ są kolejnymi liczbami całkowitymi. W przykładzie 1 mieliśmy $p(k)$ =„$k^5 - k$ jest wielokrotnością 5" dla $k = 1, 2, \ldots, 37^{100}$. Przypuśćmy, że wiemy również, tak jak w przykładzie 1, iż:

(P) zdanie $p(m)$ jest prawdziwe oraz
(I) zdanie $p(k+1)$ jest prawdziwe, jeśli tylko zdanie $p(k)$ jest prawdziwe i $m \leq k < n$.

Twierdzimy, że wszystkie zdania $p(m)$, $p(m+1)$, ..., $p(n)$ muszą być wtedy prawdziwe. Uzasadnienie jest takie, jakie przeprowadziliśmy w przykładzie 1. Tworzymy pętlę pokazaną poniżej:

$k := m$
{zdanie $p(k)$ jest prawdziwe}
dopóki $m \leq k < n$, wykonuj
  jeśli $p(k)$ jest prawdziwe, to
    $k := k + 1$

Z (I) wynika, że warunek „$p(k)$ jest prawdziwe" jest niezmiennikiem pętli i pętla kończy się dla $k = n$. Oczywiście warunek $p(k) \Rightarrow p(k+1)$ powoduje, że zdanie $p(k)$ jest prawdziwe na końcu każdego przebiegu pętli. Zasada dobrego uporządkowania mówi, że jeśli zdanie $p(k)$ kiedykolwiek będzie fałszywe, to musi być ten pierwszy raz, kiedy to się stanie, a warunki (P) i (I) pokazują, że żadne zdanie $p(k)$ nie będzie tym pierwszym złym zdaniem.

Udowodniliśmy następujący ważny fakt.

Zasada skończonej indukcji matematycznej

Niech $p(m)$, $p(m+1)$, ..., $p(n)$ będzie skończonym ciągiem zdań. Jeśli

(P) zdanie $p(m)$ jest prawdziwe oraz
(I) zdanie $p(k+1)$ jest prawdziwe, jeśli tylko zdanie $p(k)$ jest prawdziwe i $m \leq k < n$,

to wszystkie te zdania są prawdziwe.

Jest też nieskończona wersja tej zasady. Zastąpmy po prostu warunek „dopóki $m \leq k < n$" przez warunek „dopóki $m \leq k$", aby otrzymać następującą zasadę.

Zasada indukcji matematycznej

Niech $p(m)$, $p(m+1)$, ... będzie ciągiem zdań. Jeśli

(P) zdanie $p(m)$ jest prawdziwe oraz
(I) zdanie $p(k+1)$ jest prawdziwe, jeśli tylko zdanie $p(k)$ jest prawdziwe i $m \leq k$,

to wszystkie te zdania są prawdziwe.

Warunek (P) w każdej z tych zasad indukcji matematycznej nazywamy **warunkiem początkowym**, a (I) **krokiem indukcyjnym**. Dla danego ciągu zdań zasady te pomagają nam dowieść, że wszystkie te zdania są prawdziwe. Warunek początkowy jest zazwyczaj łatwy do sprawdzenia; krok indukcyjny jest czasami trochę trudniejszy do sprawdzenia.

Zasady te mówią nam, że jeśli możemy pokazać (P) oraz (I), to zakończyliśmy dowód, ale nie pokazują nam, w jaki sposób dowieść obu tych warunków. Oczywiście, jeśli zdania $p(k)$ nie są prawdziwe, to zasady indukcji nie mogą pokazać, że one są prawdziwe. Wtedy albo (P), albo (I) musi być fałszywe.

PRZYKŁAD 2

(a) Dla każdej liczby $n \in \mathbb{P}$ niech $p(n)$ będzie zdaniem „$n! > 2^n$", którego dowiedliśmy dla wszystkich $n \geq 4$ w § 1.6, po prostu przyglądając się uważnie obu stronom nierówności. Aby podać dowód indukcyjny, sprawdzamy, że zdanie $p(n)$ jest prawdziwe dla $n = 4$, tzn. sprawdzamy, że $4! > 2^4$, a następnie pokazujemy

(I) jeśli $4 \leq k$ oraz $k! > 2^k$, to $(k+1)! > 2^{k+1}$.

Warunku (I) dowodzi się bezpośrednio:

$$\begin{aligned}(k+1)! &= k! \cdot (k+1) \\ &> 2^k \cdot (k+1) && \text{(z założenia indukcyjnego } k! > 2^k\text{)} \\ &\geq 2^k \cdot 2 && \text{(gdyż } k+1 \geq 5 > 2\text{)} \\ &= 2^{k+1}.\end{aligned}$$

Ponieważ sprawdziliśmy warunek początkowy oraz krok indukcyjny, zdanie $p(n)$ jest prawdziwe dla każdej liczby całkowitej $n \geq 4$, zgodnie z zasadą indukcji matematycznej.

(b) Warto wiedzieć, że

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2} \quad \text{dla wszystkich} \quad n \in \mathbb{P}.$$

Można tego dowieść metodą uśredniania, ale również przez indukcję.

Niech $p(n)$ będzie zdaniem „$\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$". Wtedy $p(1)$ jest zdaniem „$\sum_{i=1}^{1} i = \frac{1(1+1)}{2}$", które jest prawdziwe.

Załóżmy indukcyjnie, że zdanie $p(k)$ jest prawdziwe dla pewnej liczby $k \in \mathbb{P}$, tzn. że

$$\sum_{i=1}^{k} i = \frac{k(k+1)}{2}.$$

Chcemy pokazać, że to implikuje, iż zdanie $p(k+1)$ jest prawdziwe. Mamy

$$\begin{aligned}\sum_{i=1}^{k+1} i &= \sum_{i=1}^{k} i + (k+1) && \left(\text{z definicji } \sum\right) \\ &= \frac{k(k+1)}{2} + (k+1) && \text{(z założenia, że zdanie } p(k) \text{ jest prawdziwe)} \\ &= \left(\frac{k}{2} + 1\right)(k+1) && \text{(wyłączenie } k+1 \text{ przed nawias)} \\ &= \left(\frac{k+2}{2}\right)(k+1) \\ &= \frac{((k+1)+1)(k+1)}{2},\end{aligned}$$

a więc zdanie $p(k+1)$ jest prawdziwe. Z zasady indukcji wynika, że zdanie $p(n)$ jest prawdziwe dla każdej liczby $n \in \mathbb{P}$.

(c) Możemy użyć indukcji matematycznej do uzasadnienia wzoru na sumę wyrazów ciągu geometrycznego:

$$\sum_{i=0}^{n} r^i = \frac{r^{n+1}-1}{r-1}, \quad \text{jeśli} \quad r \neq 0,\ r \neq 1 \quad \text{oraz} \quad n \in \mathbb{N}.$$

Niech $p(n)$ będzie zdaniem „$\sum_{i=0}^{n} r^i = \frac{r^{n+1}-1}{r-1}$". Wtedy $p(0)$ jest zdaniem

$$r^0 = \frac{r^1 - 1}{r-1},$$

które jest prawdziwe, ponieważ $r^0 = 1$. Zatem warunek początkowy indukcji jest spełniony.

Dowodzimy kroku indukcyjnego $p(k) \Rightarrow p(k+1)$ w następujący sposób:

$$\begin{aligned}
\sum_{i=0}^{k+1} r^i &= \sum_{i=0}^{k} r^i + r^{k+1} \\
&= \frac{r^{k+1}-1}{r-1} + r^{k+1} && \text{(z założenia indukcyjnego } p(k)) \\
&= \frac{r^{k+1} - 1 + r^{k+2} - r^{k+1}}{r-1} && \text{(przekształcenia algebraiczne)} \\
&= \frac{r^{k+2}-1}{r-1} && \text{(dalsze przekształcenia algebraiczne)} \\
&= \frac{r^{(k+1)+1}-1}{r-1},
\end{aligned}$$

tzn. zdanie $p(k+1)$ jest prawdziwe. Z zasady indukcji wynika, że zdanie $p(n)$ jest prawdziwe dla każdej liczby $n \in \mathbb{N}$.

(d) Udowodnimy, że wszystkie liczby postaci $8^n - 2^n$ są podzielne przez 6. Dokładniej, pokażemy, że liczba $8^n - 2^n$ jest podzielna przez 6 dla każdej liczby $n \in \mathbb{P}$. Naszym n-tym zdaniem jest zdanie

$$p(n) = \text{„liczba } 8^n - 2^n \text{ jest podzielna przez 6".}$$

Warunek początkowy indukcji, $p(1)$, jest oczywiście spełniony, ponieważ $8^1 - 2^1 = 6$. W kroku indukcyjnym przyjmijmy, że zdanie $p(k)$ jest prawdziwe. Naszym zadaniem jest takie użycie tego założenia, by udowodnić zdanie $p(k+1)$:

$$\text{liczba } 8^{k+1} - 2^{k+1} \text{ jest podzielna przez 6.}$$

Zatem chcielibyśmy zapisać jakoś $8^{k+1} - 2^{k+1}$ za pomocą $8^k - 2^k$, w taki sposób, aby łatwo było widać, że pozostałe składniki są podzielne przez 6. Pomysł polega na tym, by zapisać to wyrażenie jako sumę $8(8^k - 2^k)$ i odpowiednich składników uzupełniających:

$$\begin{aligned} 8^{k+1} - 2^{k+1} &= 8(8^k - 2^k) + 8 \cdot 2^k - 2^{k+1} \\ &= 8(8^k - 2^k) + 8 \cdot 2^k - 2 \cdot 2^k \\ &= 8(8^k - 2^k) + 6 \cdot 2^k. \end{aligned}$$

Liczba $8^k - 2^k$ jest podzielna przez 6 na mocy założenia indukcyjnego $p(k)$, a liczba $6 \cdot 2^k$ jest oczywiście wielokrotnością 6, a więc również liczba $8^{k+1} - 2^{k+1}$ jest wielokrotnością 6. Pokazaliśmy, że krok indukcyjny jest spełniony, a więc zakończyliśmy dowód przez indukcję matematyczną. ∎

Warto podkreślić, że przed ostatnim zdaniem w każdym z tych dowodów nie dowodziliśmy, że „zdanie $p(k+1)$ jest prawdziwe". My tylko dowodziliśmy implikacji: „jeśli zdanie $p(k)$ jest prawdziwe, to zdanie $p(k+1)$ jest prawdziwe". W pewnym sensie dowodziliśmy nieskończenie wielu twierdzeń, mianowicie: $p(1)$; jeśli $p(1)$ jest prawdziwe, to $p(2)$ jest prawdziwe; jeśli $p(2)$ jest prawdziwe, to $p(3)$ jest prawdziwe; jeśli $p(3)$ jest prawdziwe, to $p(4)$ jest prawdziwe itd. Następnie stosowaliśmy zasadę indukcji matematycznej, aby wywnioskować, że $p(1)$ jest prawdziwe; $p(2)$ jest prawdziwe; $p(3)$ jest prawdziwe; $p(4)$ jest prawdziwe itd.

Zauważmy także, że jeśli używamy zasad indukcji, nie musimy pisać żadnych pętli. Używaliśmy pętli do uzasadniania tych zasad, ale kiedy stosujemy te zasady, musimy tylko sprawdzić warunki (P) oraz (I).

PRZYKŁAD 3 Oto kilka przykładów złego użycia indukcji.

(a) Dla $n \in \mathbb{P}$ niech $p(n)$ będzie zdaniem „$n^2 \leq 100$". Wtedy możemy bezpośrednio sprawdzić, że zdania $p(1)$, $p(2)$, ..., $p(10)$ są prawdziwe. Jeśli spróbujemy dowieść przez indukcję, że zdanie $p(n)$ jest prawdziwe dla wszystkich $n \in \mathbb{P}$ lub nawet tylko dla $1 \leq n \leq 20$, to się nam nie uda. Możemy pokazać, że spełniony jest warunek początkowy (P), ale nie da się dowieść kroku indukcyjnego (I), ponieważ wiemy na przykład, że $p(11)$ jest zdaniem fałszywym.

(b) Dla $n \in \mathbb{N}$ niech $r(n)$ będzie zdaniem „$n = n + 5$". Wtedy zdaniem $r(0)$ jest „$0 = 5$", a więc warunek (P) jest fałszywy, natomiast prawdziwa jest implikacja $r(k) \Rightarrow r(k+1)$, bo jeśli $k = k + 5$, to $k + 1 = (k + 1) + 5$. Nie wystarczy po prostu

sprawdzić, że zachodzi warunek (I); musimy również sprawdzić warunek (P).

(c) Dla $m \in \mathbb{P}$ niech $s(m)$ będzie zdaniem „$m \cdot (m+1)$ jest liczbą parzystą". Moglibyśmy oczywiście dowieść, że zdanie $s(m)$ jest prawdziwe dla wszystkich $m \in \mathbb{P}$, używając indukcji (pomyśl, jak wyglądałby taki dowód), ale nie musimy prowadzić tak wymyślnego dowodu. Albo m, albo $m+1$ jest liczbą parzystą, więc w każdym przypadku ich iloczyn jest liczbą parzystą. Czasami tak samo łatwo dowieść bezpośrednio, że wszystkie zdania w jakimś ciągu zdań są prawdziwe, jak zastosować indukcję.

■

PRZYKŁAD 4

(a) Zastosowanie indukcji zaczyna się często od zgadywania. Przypuśćmy na przykład, że ciąg $(s_0, s_1, s_2, \ldots)$ spełnia warunki $s_0 = a$ oraz $s_n = 2s_{n-1} + b$ dla pewnych stałych a i b oraz wszystkich $n \in \mathbb{P}$. Czy możemy podać wzór określający s_n? Obliczamy kilka pierwszych wyrazów tego ciągu, aby zobaczyć, czy istnieje jakaś prawidłowość. W tabelce poniżej pokazane są wyniki tych obliczeń.

n	s_n
0	a
1	$2a + b$
2	$2(2a + b) + b = 4a + 3b$
3	$2(4a + 3b) + b = 8a + 7b$
4	$2(8a + 7b) + b = 16a + 15b$

Wydaje się, że być może ogólnie $s_n = 2^n a + (2^n - 1)b$. (Ten sposób odgadywania faktu ogólnego z kilku obserwacji nazywamy w naukach ścisłych „rozumowaniem indukcyjnym". Znaczenie słowa „indukcyjny" w tym kontekście jest całkiem inne niż znaczenie matematyczne.) Mamy teraz ciąg zdań; dla $n \in \mathbb{N}$, $p(n)$ jest zdaniem „$s_n = 2^n a + (2^n - 1)b$". Okazuje się (ćwiczenie 5), że odgadliśmy dobrze i można dowieść, używając zasady indukcji matematycznej, że zdanie $p(n)$ jest prawdziwe dla każdej liczby $n \in \mathbb{N}$.

(b) Kiedy szukamy prawidłowości, czasami pomaga dokładniejsze przyjrzenie się, w jaki sposób jakiś szczególny przypadek wynika z poprzedniego. Możemy wtedy zobaczyć, jak poprowadzić dowód kroku indukcyjnego.

Niech $\mathcal{P}(S)$ będzie zbiorem potęgowym pewnego skończonego zbioru S. Jeśli zbiór S ma n elementów, to $\mathcal{P}(S)$ ma 2^n elementów. W przykładzie 2 w § 1.1 stwierdziliśmy, że to zdanie może być prawdziwe. Teraz udowodnimy je przez indukcję.

Sprawdziliśmy wcześniej, że to twierdzenie jest prawdziwe dla $n = 0, 1, 2$ i 3. W szczególności przypadek $n = 0$ jest warunkiem początkowym dla indukcji. Zanim udowodnimy krok indukcyjny, chwilę poeksperymentujmy. Porównajmy więc zbiory $\mathcal{P}(S)$ dla $S = \{a, b\}$ i $S = \{a, b, c\}$. Zauważmy, że

$$\mathcal{P}(\{a, b.c\}) = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{c\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}.$$

Pierwsze cztery zbiory tworzą zbiór $\mathcal{P}(\{a, b\})$; każdy z pozostałych zbiorów składa się z elementów zbioru należącego do $\mathcal{P}(\{a, b\})$ i z elementu c. Dlatego zbiór $\mathcal{P}(\{a, b, c\})$ ma dwa razy więcej elementów niż zbiór $\mathcal{P}(\{a, b\})$. Wygląda, że to rozumowanie daje się uogólnić: za każdym razem, kiedy dodawany jest element do zbioru S, liczba elementów zbioru $\mathcal{P}(S)$ podwaja się.

Aby udowodnić krok indukcyjny, załóżmy, że nasze zdanie jest prawdziwe dla jakiegoś n. Weźmy zbiór S składający się z $n + 1$ elementów; dla wygody weźmy $S = \{1, 2, 3, \ldots, n, n + 1\}$. Niech $T = \{1, 2, 3, \ldots, n\}$. Zbiorami należącymi do $\mathcal{P}(T)$ są te podzbiory zbioru S, które nie zawierają $n + 1$. Z założenia dla n w zbiorze $\mathcal{P}(T)$ znajduje się dokładnie 2^n zbiorów. Każdy z pozostałych podzbiorów zbioru S zawiera liczbę $n+1$, a więc jest sumą zbioru należącego do $\mathcal{P}(T)$ i zbioru jednoelementowego $\{n + 1\}$. To znaczy, że $\mathcal{P}(S)$ ma 2^n innych zbiorów, które nie są podzbiorami zbioru T. Wynika stąd, że zbiór $\mathcal{P}(S)$ ma $2^n + 2^n = 2^{n+1}$ elementów. To kończy dowód kroku indukcyjnego, a więc twierdzenie jest prawdziwe dla wszystkich n, na mocy zasady indukcji matematycznej. ∎

Metoda indukcji matematycznej ma zastosowanie w sytuacjach takich, jak opisane w ostatnim przykładzie, w których:

1. Znamy na początku odpowiedź.
2. Wiemy, jak wyprowadzić odpowiedź w danym kroku z odpowiedzi w poprzednim kroku.
3. Odgadujemy ogólne rozwiązanie.

Oczywiście, jeśli odgadnięty wynik jest zły, to fatalnie. Nie będziemy w stanie dowieść, że jest on prawdziwy ani tą metodą, ani jakąkolwiek inną. Ale jeśli nasz wynik jest poprawny, to indukcja matematyczna często daje nam metodę pozwalającą potwierdzić dowodem odgadnięty wynik.

Czasami ma sens rozważanie tylko skończonego ciągu zdań $p(n)$. Na przykład moglibyśmy mieć algorytm iteracyjny, o którym wiemy, że po jakimś czasie się zatrzymuje i chcielibyśmy

wiedzieć, że spełniony jest pewien warunek w czasie, gdy ten algorytm działa. Sprawdzanie niezmienników pętli „dopóki" jest przykładem takiego zadania. Jeśli możemy zapisać nasz warunek w postaci $p(k)$ za pomocą pewnej zmiennej k, której wartość stopniowo rośnie podczas wykonywania algorytmu, powiedzmy $k = m, m+1, \ldots, N$, to być może będziemy mogli zastosować zasadę skończonej indukcji matematycznej, aby dowieść, że zdanie $p(k)$ jest prawdziwe dla wszystkich dopuszczalnych wartości k. Sprawdzenie niezmiennika fragmentu algorytmu „dla i od m do N wykonuj S" sprowadza się do sprawdzenia kroku indukcyjnego dla tej zasady indukcji.

W paragrafie tym omówiliśmy podstawowe idee indukcji matematycznej i pokazaliśmy metody postępowania w wielu często spotykanych sytuacjach. Jak zobaczymy w § 4.5, istnieją inne postaci zasady indukcji, dające się zastosować do rozwiązania wielu zadań, w których zasady indukcji z tego paragrafu nie mają naturalnego zastosowania.

ĆWICZENIA DO § 4.2

1. Udowodnij, że

$$\sum_{i=1}^{n} i^2 = 1 + 4 + 9 + \ldots + n^2 = \frac{n(n+1)(2n+1)}{6} \quad \text{dla} \quad n \in \mathbb{P}.$$

2. Udowodnij, że

$$4 + 10 + 16 + \ldots + (6n-2) = n(3n+1) \quad \text{dla wszystkich} \quad n \in \mathbb{P}.$$

3. Udowodnij, że:
(a) $37^{100} - 37^{20}$ jest wielokrotnością 10.
(b) $37^{20} - 37^4$ jest wielokrotnością 10.
(c) $37^{500} - 37^4$ jest wielokrotnością 10.
(d) $37^4 - 1$ jest wielokrotnością 10.
(e) $37^{500} - 1$ jest wielokrotnością 10.

4. Udowodnij, że

$$\frac{1}{1\cdot 5} + \frac{1}{5\cdot 9} + \frac{1}{9\cdot 13} + \ldots + \frac{1}{(4n-3)(4n+1)} = \frac{n}{4n+1} \quad \text{dla} \quad n \in \mathbb{P}.$$

5. Pokaż przez indukcję, że jeśli $s_0 = a$ oraz $s_n = 2s_{n-1} + b$ dla $n \in \mathbb{P}$, to $s_n = 2^n a + (2^n - 1)b$ dla każdej liczby $n \in \mathbb{N}$.

6. Weźmy następującą procedurę.

początek
$S := 1$

dopóki $1 \leq S$, wykonuj
  wypisz S
  $S := S + 2\sqrt{S} + 1$
koniec

(a) Wypisz cztery pierwsze wartości S.

(b) Zastosuj metodę indukcji matematycznej, by pokazać, że wartość S jest zawsze liczbą całkowitą. (Łatwiej dowieść dużo mocniejszego stwierdzenia, że wartość S jest zawsze kwadratem liczby całkowitej; tak naprawdę $S = n^2$ na początku n-tego przebiegu pętli.)

7. Udowodnij, że liczba $11^n - 4^n$ jest podzielna przez 7 dla wszystkich $n \in \mathbb{P}$.

8. (a) Dobierz m i $p(k)$ w następującym fragmencie algorytmu

$k := m$
dopóki $m \leq k$, wykonuj
  jeśli $p(k)$ jest prawdziwe, to
    $k := k + 1$

tak, aby z tego, że $p(k)$ jest niezmiennikiem pętli, wynikało, iż $2^n < n!$ dla wszystkich liczb całkowitych $n \geq 4$.

(b) Sprawdź, że twój warunek $p(k)$ z ćwiczenia (a) jest niezmiennikiem pętli.

(c) Zdanie $p(k) =$„$8^k < k!$" jest niezmiennikiem tej pętli. Czy stąd wynika, że $8^n < n!$ dla wszystkich $n \geq 4$? Odpowiedź uzasadnij.

9. (a) Pokaż, że warunek $\sum_{i=0}^{k} 2^i = 2^{k+1} - 1$ jest niezmiennikiem pętli w algorytmie

początek
$k := 0$
dopóki $0 \leq k$, wykonuj
  jeśli $\sum_{i=0}^{k} 2^i = 2^{k+1} - 1$, to
    $k := k + 1$
koniec.

(b) Powtórz ćwiczenie (a) dla niezmiennika $\sum_{i=0}^{k} 2^i = 2^{k+1}$.

(c) Czy możesz wykorzystać ćwiczenie (a), aby dowieść, że $\sum_{i=0}^{k} 2^i = 2^{k+1} - 1$ dla każdego $k \in \mathbb{N}$? Odpowiedź uzasadnij.

(d) Czy możesz wykorzystać ćwiczenie (b), aby dowieść, że $\sum_{i=0}^{k} 2^i = 2^{k+1}$ dla każdego $k \in \mathbb{N}$? Odpowiedź uzasadnij.

10. Udowodnij, że $n^2 > n + 1$ dla $n \geq 2$.

11. (a) Oblicz $1 + 3 + \ldots + (2n - 1)$ dla kilku wartości n, a następnie odgadnij wzór ogólny.

(b) Udowodnij przez indukcję, że wzór otrzymany w ćwiczeniu (a) jest prawdziwy.

12. Dla jakich $n \in \mathbb{P}$ zachodzi nierówność $4n \leq n^2 - 7$? Odpowiedź uzasadnij.

13. Weźmy zdanie $p(n) =$„$n^2 + 5n + 1$ jest liczbą parzystą".
(a) Udowodnij, że z prawdziwości zdania $p(k)$ wynika prawdziwość zdania $p(k+1)$ dla każdego $k \in \mathbb{P}$.
(b) Dla jakich wartości n zdanie $p(n)$ jest rzeczywiście prawdziwe? Jaki wynika stąd morał?

14. Udowodnij, że $(2n+1) + (2n+3) + (2n+5) + \ldots + (4n-1) = 3n^2$ dla $n \in \mathbb{P}$. Ta suma może być też zapisana w postaci $\sum_{i=n}^{2n-1} 2i + 1$.

15. Udowodnij, że liczba $5^n - 4n - 1$ jest podzielna przez 16 dla $n \in \mathbb{P}$.

16. Udowodnij, że $1^3 + 2^3 + \ldots + n^3 = (1 + 2 + \ldots + n)^2$, tzn. $\sum_{i=1}^{n} i^3 = (\sum_{i=1}^{n} i)^2$ dla $n \in \mathbb{P}$. *Wskazówka*: wykorzystaj równość z przykładu 2(b).

17. Udowodnij, że

$$\frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{2n} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots + \frac{1}{2n-1} - \frac{1}{2n}$$

dla $n \in \mathbb{P}$. Dla $n = 1$ ta równość mówi, że $\frac{1}{2} = 1 - \frac{1}{2}$, a dla $n = 2$ ta równość mówi, że $\frac{1}{3} + \frac{1}{4} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4}$.

18. Dla $n \in \mathbb{P}$ udowodnij, że
(a) $\sum_{i=1}^{n} \frac{1}{\sqrt{i}} \geq \sqrt{n}$,
(b) $\sum_{i=1}^{n} \frac{1}{\sqrt{i}} \leq 2\sqrt{n} - 1$.

19. Udowodnij, że liczba $5^{n+1} + 2 \cdot 3^n + 1$ jest podzielna przez 8 dla $n \in \mathbb{N}$.

20. Udowodnij, że liczba $8^{n+2} + 9^{2n+1}$ jest podzielna przez 73 dla $n \in \mathbb{N}$.

21. To ćwiczenie wymaga pewnej znajomości tożsamości trygonometrycznych. Udowodnij, że $|\sin nx| \leq n|\sin x|$ dla wszystkich $x \in \mathbb{R}$ i wszystkich $n \in \mathbb{P}$.

§ 4.3. Definicje rekurencyjne

Wyrazy ciągu mogą być podane za pomocą wzorów, na przykład $s_n = n^3 - 73n$, lub opisowo, na przykład „niech t_n będzie wagą n-tej krawędzi w drodze". Wyrazy mogą też być czasami zdefiniowane za pomocą opisów, w których używa się innych wyrazów, wcześniej otrzymanych w ciągu.

Mówimy, że ciąg jest zdefiniowany rekurencyjnie, jeśli:

(P) Określony jest pewien skończony zbiór wyrazów ciągu, zazwyczaj kilka pierwszych lub pierwszy wyraz.
(R) Pozostałe wyrazy ciągu są zdefiniowane za pomocą poprzednich wyrazów ciągu. (Wzór definiujący ciąg w taki sposób nazywamy **wzorem**, **równaniem** lub **zależnością rekurencyjną**.)

Warunek (P) zawiera początek lub krok początkowy definicji. Pozostałe wyrazy ciągu są określane kolejno za pomocą reguły (R) (tzn. rekurencyjnie).

PRZYKŁAD 1 (a) Możemy zdefiniować rekurencyjnie znany ciąg SILNIA w następujący sposób:

(P) $\text{SILNIA}(0) = 1$,
(R) $\text{SILNIA}(n+1) = (n+1) \cdot \text{SILNIA}(n)$ dla $n \in \mathbb{N}$.

Warunek (R) pozwala nam obliczyć SILNIA(1), następnie SILNIA(2), potem SILNIA(3) itd. i szybko widać, że $\text{SILNIA}(n) = n!$ dla pierwszych kilku wartości n. Równości tej można dowieść dla wszystkich n za pomocą indukcji matematycznej. Z warunku (P) mamy $\text{SILNIA}(0) = 1 = 0!$. Zakładając indukcyjnie, że $\text{SILNIA}(m) = m!$ dla pewnego $m \in \mathbb{N}$ i stosując warunek (R), otrzymujemy w kroku indukcyjnym $\text{SILNIA}(m+1) = (m+1) \cdot \text{SILNIA}(m) = (m+1) \cdot m! = (m+1)!$. Ponieważ znamy już ciąg $n!$, powyższa definicja rekurencyjna może wydawać się głupia, ale spróbujemy przekonać Cię w przykładzie (b), że definicje rekurencyjne nawet prostych ciągów są przydatne.

(b) Weźmy ciąg $\text{SUMA}(n) = \sum_{i=0}^{n} \frac{1}{i!}$. Aby napisać program komputerowy obliczający wartości ciągu SUMA dla dużych liczb n, można użyć następującej definicji rekurencyjnej:

(P) $\text{SUMA}(0) = 1$,
(R) $\text{SUMA}(n+1) = \text{SUMA}(n) + \frac{1}{(n+1)!}$.

Dodawany składnik w równości (R) jest odwrotnością $(n+1)!$. Zatem w trakcie działania programu będzie potrzebna wartość $\text{SILNIA}(n+1)$. Dla każdego n można polecić programowi obliczenie $\text{SILNIA}(n+1)$ od początku albo można przechowywać wiele takich wartości. Oczywiście byłoby bardziej efektywnie, gdybyśmy obliczali na zmianę $\text{SILNIA}(n+1)$ i $\text{SUMA}(n+1)$, używając definicji rekurencyjnej z przykładu (a) dla obliczania wartości ciągu SILNIA oraz powyższej definicji rekurencyjnej dla obliczania wartości ciągu SUMA.

(c) Zdefiniujmy ciąg SEQ w następujący sposób:

(P) $\text{SEQ}(0) = 1$,
(R) $\text{SEQ}(n+1) = (n+1)/\text{SEQ}(n)$ dla $n \in \mathbb{N}$.

Dla $n = 0$ z warunku (R) otrzymujemy $\text{SEQ}(1) = 1/1 = 1$. Następnie dla $n = 1$ obliczamy $\text{SEQ}(2) = 2/1 = 2$. Kontynuując w ten sposób widzimy, że pierwszymi wyrazami tego ciągu są 1, 1,

2, 3/2, 8/3, 15/8, 16/5, 35/16. Zupełnie nie widać, jak mógłby wyglądać wzór ogólny na SEQ(n). Jest rzeczą oczywistą, że SEQ(73) istnieje, ale obliczenie tej wartości zajęłoby trochę czasu. ∎

Skąd w przykładzie 1 wiedzieliśmy, że SEQ(73) istnieje? Nasza pewność jest oparta na przekonaniu, że definicje rekurencyjne rzeczywiście definiują ciągi określone na całym zbiorze $\mathbb{N}$, chyba że któryś krok wymagałby wykonania czynności niedozwolonej, takiej jak dzielenie przez 0. Można dowieść przez indukcję, że definicja rekurencyjna w przykładzie 1 określa ciąg, skorzystamy tu jednak z zasady dobrego uporządkowania. Niech

$$S = \{n \in \mathbb{N}: \ \text{SEQ}(n) = 0 \text{ lub SEQ}(n) \text{ nie jest określony}\}.$$

Chcemy pokazać, że zbiór S jest pusty. Gdyby nie był pusty, to miałby najmniejszy element, na przykład m. Na podstawie (P) $m \neq 0$, a więc $m - 1 \in \mathbb{N}$. Ponieważ liczba m jest najmniejszą „złą liczbą", więc $\text{SEQ}(m-1) \neq 0$ i wartość $\text{SEQ}(m-1)$ jest dobrze określona. Ale wtedy z warunku (R) wynika, że $\text{SEQ}(m) = m/\text{SEQ}(m-1) \neq 0$, co przeczy założeniu, że $m \in S$. Zatem S musi być zbiorem pustym.

W dowodzie, że wartość SEQ(n) jest określona dla każdej liczby n, korzystamy z faktu, że SEQ($n+1$) zależy tylko od SEQ(n). W definicjach rekurencyjnych dopuszcza się, by dany wyraz zależał również od innych wyrazów, a nie tylko od tego jednego stojącego bezpośrednio przed nim. W takich przypadkach dowodzimy, że ciąg jest dobrze określony, korzystając albo z zasady dobrego uporządkowania, albo z rozszerzonej wersji zasady indukcji matematycznej, którą omówimy w § 4.5.

Wyrazy ciągu zdefiniowanego rekurencyjnie można obliczać na wiele różnych sposobów. Za pomocą **obliczeń iteracyjnych** znajdujemy wartość s_n obliczając najpierw wszystkie wartości $s_1, s_2, \ldots, s_{n-1}$, tak aby były znane w momencie obliczania s_n. W przykładzie 1 mieliśmy na myśli to, że wyrazy ciągów będą obliczane w sposób iteracyjny. Aby na przykład obliczyć SILNIA(73), obliczalibyśmy najpierw SILNIA(k) dla $k = 1, 2, \ldots, 72$, nawet jeśli te początkowe wartości nie byłyby nam do niczego poza tym potrzebne.

W przypadku ciągu SILNIA tak naprawdę wydaje się, że nie ma lepszej metody. Czasami jednak istnieje sprytniejszy sposób obliczenia danej wartości s_n. W obliczeniach rekurencyjnych znajdujemy wartość s_n patrząc na to, od jakich wyrazów zależy wartość s_n, następnie od jakich wyrazów zależą te wyrazy i tak dalej. Może się okazać, że wartość s_n zależy tylko od wartości stosun-

kowo małego zbioru swoich poprzedników, a inne wyrazy początkowe będzie można pominąć.

PRZYKŁAD 2 (a) Część całkowita liczby rzeczywistej a, oznaczana symbolem $\lfloor a \rfloor$, tak jak w § 3.6, jest największą liczbą całkowitą m taką, że $m \leq a$. Określamy ciąg T w następujący sposób:

(P) $T(1) = 1$,
(R) $T(n) = 2 \cdot T(\lfloor n/2 \rfloor)$ dla $n \geq 2$.

Wtedy

$$\begin{aligned} T(73) &= 2 \cdot T(\lfloor 73/2 \rfloor) = 2 \cdot T(36) = 2 \cdot 2 \cdot T(18) \\ &= 2 \cdot 2 \cdot 2 \cdot T(9) = 2 \cdot 2 \cdot 2 \cdot 2 \cdot T(4) = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot T(2) \\ &= 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot T(1) = 2^6. \end{aligned}$$

W tym obliczeniu potrzebne są tylko wartości $T(36)$, $T(18)$, $T(9)$, $T(4)$, $T(2)$ oraz $T(1)$ i nie ma potrzeby obliczania pozostałych 66 wartości $T(n)$, poprzedzających $T(73)$.

Ten ciąg może być opisany w inny sposób (ćwiczenie 19 w § 4.5):

$T(n)$ jest największą liczbą całkowitą postaci 2^k taką, że $2^k \leq n$.

Używając tego zapisu moglibyśmy obliczyć $T(73)$ wybierając największą liczbę z ciągu potęg 2, mniejszych od 73.

(b) Drobna zmiana w warunku (R) w ćwiczeniu (a) daje ciąg Q taki, że

(P) $Q(1) = 1$,
(R) $Q(n) = 2 \cdot Q(\lfloor n/2 \rfloor) + n$ dla $n \geq 2$.

Teraz nie jest tak oczywiste, jak wygląda wyraz ogólny ciągu, ale nadal możemy obliczyć rekurencyjnie $Q(73)$ za pomocą $Q(36)$, $Q(18)$, ..., $Q(2)$, $Q(1)$. ∎

PRZYKŁAD 3 Ciąg Fibonacciego jest zdefiniowany w następujący sposób:

(P) $\text{FIB}(0) = \text{FIB}(1) = 1$,
(R) $\text{FIB}(n) = \text{FIB}(n-1) + \text{FIB}(n-2)$ dla $n \geq 2$.

Zauważmy, że wzór rekurencyjny nie ma sensu dla $n = 1$, a więc wartość $\text{FIB}(1)$ musi być zdefiniowana oddzielnie w kroku początkowym. Początkowymi wyrazami tego ciągu są

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89.$$

(b) Oto prosty sposób zdefiniowania ciągu

$$0, 0, 1, 1, 2, 2, 3, 3, \ldots.$$

(P) CIĄG(0) = CIĄG(1) = 0,
(R) CIĄG(n) = 1 + CIĄG($n-2$) dla $n \geq 2$.

Porównaj ćwiczenie 11 w § 1.5. ∎

PRZYKŁAD 4 Niech $\Sigma = \{a, b\}$.

(a) Interesuje nas liczba s_n słów długości n, w których nie występują kolejno dwie litery a, tzn. takich, które nie zawierają ciągu aa. Niech A_n oznacza zbiór słów w Σ^n, nie zawierających kolejnych liter a. Wtedy $A_0 = \{\lambda\}$, $A_1 = \Sigma$ i $A_2 = \Sigma^2 \backslash \{aa\}$, a więc $s_0 = 1$, $s_1 = 2$, $s_2 = 4 - 1 = 3$. Aby otrzymać wzór rekurencyjny na s_n, załóżmy, że $n \geq 2$, i spróbujmy zliczyć słowa w A_n za pomocą słów krótszych. Jeśli słowo należące do A_n kończy się literą b, to ta litera b może być poprzedzona dowolnym słowem ze zbioru A_{n-1}. Zatem s_{n-1} słów w zbiorze A_n kończy się literą b. Jeśli słowo należące do zbioru A_n kończy się literą a, to dwiema ostatnimi literami muszą być ba i może je poprzedzać dowolne słowo ze zbioru A_{n-2}. Zatem s_{n-2} słów w zbiorze A_n kończy się literą a. A więc $s_n = s_{n-1} + s_{n-2}$ dla $n \geq 2$. Jest to to samo równanie rekurencyjne, co dla ciągu Fibonacciego, jednak zauważmy, że krok początkowy jest inny: $s_1 = 2$, podczas gdy FIB(1) = 1. W rzeczywistości, s_n = FIB($n+1$) dla $n \in \mathbb{N}$ (ćwiczenie 13).

(b) Ponieważ zbiór Σ^n składa się z 2^n słów, więc istnieje liczba $2^n - s_n = 2^n$ − FIB($n+1$) słów długości n zawierających kolejne litery a. ∎

PRZYKŁAD 5 Niech $\Sigma = \{a, b, c\}$ i niech B_n będzie zbiorem słów w Σ^n, w którym występuje parzysta liczba elementów a oraz niech t_n oznacza liczbę słów w B_n. Wtedy $B_0 = \{\lambda\}$, $B_1 = \{b, c\}$ i $B_2 = \{aa, bb, bc, cb, cc\}$. Zatem $t_0 = 1$, $t_1 = 2$ i $t_2 = 5$. Liczbę słów w B_n zliczamy patrząc na ostatnią literę. Jeśli słowo w B_n kończy się na b, może ono być poprzedzone dowolnym słowem z B_{n-1}. Zatem t_{n-1} w B_n kończy się literą b. Podobnie t_{n-1} słów w B_n kończy się literą c. Jeśli słowo w B_n kończy się literą a, musi ona być poprzedzona słowem z Σ^{n-1}, w którym jest nieparzysta liczba liter a. Ponieważ Σ^{n-1} ma 3^{n-1} słów, to $3^{n-1} - t_{n-1}$ z nich musi mieć nieparzystą liczbę liter a. Stąd $3^{n-1} - t_{n-1}$ słów w B_n kończy się literą a. Zatem

$$t_n = t_{n-1} + t_{n-1} + (3^{n-1} - t_{n-1}) = 3^{n-1} + t_{n-1}$$

dla $n \geq 1$. Stąd $t_3 = 3^2 + t_2 = 9 + 5 = 14$, $t_4 = 3^3 + t_3 = 27 + 14 = 41$ itd.

W tym przypadku jest względnie łatwo znaleźć dokładny wzór na t_n. Po pierwsze zauważmy, że

$$t_n = 3^{n-1} + t_{n-1} = 3^{n-1} + 3^{n-2} + t_{n-2} = \dots$$
$$= 3^{n-1} + 3^{n-2} + \dots + 3^0 + t_0 = 1 + \sum_{k=0}^{n-1} 3^k.$$

Jeśli te trzy kropki Cię denerwują, możesz przeprowadzić dowód przez indukcję. Zastosujemy teraz przykład 2(c) z § 4.2, aby otrzymać

$$t_n = 1 + \frac{3^n - 1}{3 - 1} = 1 + \frac{3^n - 1}{2} = \frac{3^n + 1}{2}.$$

Nasze rozumowanie było przeprowadzone dla $n \geq 1$, ale wzór jest prawdziwy także dla $n = 0$. Wzór na t_n wygląda poprawnie: mniej więcej połowa słów w Σ^n ma parzystą liczbę liter a.

Zwróćmy uwagę na metodę, której używaliśmy, aby otrzymać wzór na t_n. Zapisywaliśmy t_n za pomocą t_{n-1}, następnie za pomocą t_{n-2}, następnie za pomocą t_{n-3}, ... i grupowaliśmy wyrazy po lewej stronie, patrząc, czy sugerują one jakiś wzór. Miały one postać $3^{n-1} + 3^{n-2} + 3^{n-3} + \dots + 3^{n-k} + t_{n-k}$, więc odgadliśmy, że przedłużając to postępowanie wystarczająco daleko, otrzymalibyśmy $t_n = 3^{n-1} + \dots + 3^0 + t_0$.

Pokazaliśmy jeden z profesjonalnych chwytów, ale tak naprawdę nie ma znaczenia, w jaki sposób dochodzimy do wzoru, jeśli umiemy pokazać, że jest on prawdziwy. Mogliśmy po prostu zapytać mamę o rozwiązanie albo będąc natchnionym zgadnąć.

Aby dowieść, że to co zgadliśmy jest poprawne, wystarczy dowieść, że wzór spełnia zależność rekurencyjną, określającą t_n, tzn. $t_0 = 1$ oraz $t_n = 3^{n-1} + t_{n-1}$. Po prostu sprawdzamy:

$$\frac{3^0 + 1}{2} = 1 \text{ oraz } \frac{3^n + 1}{2} = 3^{n-1} + \frac{3^{n-1} + 1}{2} \quad \text{dla} \quad n \in \mathbb{P}.$$

Taka metoda dowodzenia jest poprawna, ponieważ wzór rekurencyjny określa ciąg t_n jednoznacznie. ∎

PRZYKŁAD 6 (a) Definiujemy ciąg S w następujący sposób:

(P) $S(0) = 0$, $S(1) = 1$,
(R) $S(n) = S(\lfloor n/2 \rfloor) + S(\lfloor n/5 \rfloor)$ dla $n \geq 2$.

Opłaca się obliczać wartości S rekurencyjnie, a nie iteracyjnie.

Na przykład

$$\begin{aligned} S(73) &= S(36) + S(14) = [S(18) + S(7)] + [S(7) + S(2)] \\ &= S(18) + 2S(7) + S(2) \\ &= S(9) + S(3) + 2[S(3) + S(1)] + S(1) + S(0) \\ &= \ldots = 8S(1) + 6S(0) = 8. \end{aligned}$$

W obliczeniu wartości $S(73)$ korzystamy z wartości $S(36)$, $S(18)$, $S(14)$, $S(9)$, $S(7)$, $S(4)$, $S(3)$, $S(2)$, $S(1)$ i $S(0)$, ale zawsze to lepiej niż gdybyśmy musieli znajdować wszystkie wartości $S(k)$ dla $k = 1, \ldots, 72$. Można pokazać bardziej ogólnie, że wartość $S(n)$ w tym przykładzie zależy tylko od wartości $S(m)$ dla m postaci $\lfloor n/2^a 5^b \rfloor$.

(b) Obliczanie rekurencyjne wymaga pamięci do przechowywania wartości pośrednich, które były wywoływane, ale jeszcze nie zostały obliczone. Może się jednak zdarzyć, że liczba miejsc pamięci potrzebnych do przechowywania tych wartości będzie dość mała. Na przykład obliczenie rekurencyjne

$$\text{SILNIA}(6) = 6 \cdot \text{SILNIA}(5) = 30 \cdot \text{SILNIA}(4) = 120 \cdot \text{SILNIA}(3) = \ldots$$

Tylko jeden adres jest potrzebny do przechowywania pośredniej (nieznanej) wartości $\text{SILNIA}(k)$ dla $k < 6$. Podobnie obliczenie

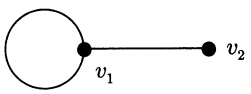
$$\begin{aligned} \text{FIB}(6) &= \text{FIB}(5) + \text{FIB}(4) = (\text{FIB}(4) + \text{FIB}(3)) + \text{FIB}(4) \\ &= 2 \cdot \text{FIB}(4) + \text{FIB}(3) \\ &= 3 \cdot \text{FIB}(3) + 2 \cdot \text{FIB}(2) \\ &= 5 \cdot \text{FIB}(2) + 3 \cdot \text{FIB}(1) \\ &= 8 \cdot \text{FIB}(1) + 5 \cdot \text{FIB}(0) \end{aligned}$$

wymaga tylko dwóch pośrednich adresów. ■

PRZYKŁAD 7

Rysunek 4.1 pokazuje graf i jego macierz sąsiedztwa $\mathbf{M}$. Jak widzieliśmy to w § 3.3, n-ta potęga macierzy $\mathbf{M}$ daje dokładną liczbę dróg długości n łączących każde dwa wierzchołki.

$$\mathbf{M} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$



Rysunek 4.1

Kilka pierwszych potęg możemy obliczyć bezpośrednio.

$$\mathbf{M}^1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{M}^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{M}^3 = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix},$$

$$\mathbf{M}^4 = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}, \quad \mathbf{M}^5 = \begin{bmatrix} 8 & 5 \\ 5 & 3 \end{bmatrix}.$$

Liczby wydają się być znajome. Tak naprawdę (ćwiczenie 12)

$$\mathbf{M}^n = \begin{bmatrix} \text{FIB}(n) & \text{FIB}(n-1) \\ \text{FIB}(n-1) & \text{FIB}(n-2) \end{bmatrix} \quad \text{dla} \quad n \geq 2.$$

Tak więc ten prosty graf prowadzi do macierzy, których wyrazy są zdefiniowane przez określony rekurencyjnie ciąg Fibonacciego z przykładu 3. ∎

Oczywiście możemy podać definicje rekurencyjne nawet wtedy, gdy ciąg ma wartości nie będące liczbami rzeczywistymi.

PRZYKŁAD 8 Niech S będzie zbiorem, a f funkcją z S w S. Określamy

(P) $f^{(0)} = 1_S$ (funkcja identycznościowa na zbiorze S),
(R) $f^{(n+1)} = f^{(n)} \circ f$.

Zatem

(1) $$f^{(1)} = f, \ f^{(2)} = f \circ f, \ f^{(3)} = f \circ f \circ f \text{ itd.}$$

Inaczej mówiąc

(2) $$f^{(n)} = f \circ f \circ \ldots \circ f \quad (n \text{ razy}).$$

$f^{(n)}$ jest po prostu złożeniem funkcji f, n razy. Definicja rekurencyjna jest dokładniejsza niż napis „itd." w (1), czy trzy kropki w (2). ∎

Tak jak w przykładzie 8, często będziemy używać definicji rekurencyjnych, aby dać bardziej zwięzłe definicje obiektów, które już całkiem dobrze znamy.

PRZYKŁAD 9 Niech a będzie niezerową liczbą rzeczywistą. Definiujemy potęgi a w następujący sposób:

(P) $a^0 = 1$,
(R) $a^{n+1} = a^n \cdot a$ dla $n \in \mathbb{N}$.

Równoważnie:

(P) $\text{POT}(0) = 1$,
(R) $\text{POT}(n+1) = \text{POT}(n) \cdot a$ dla $n \in \mathbb{N}$. ∎

PRZYKŁAD 10 Niech $(a_j)_{j\in\mathbb{P}}$ będzie ciągiem liczb rzeczywistych. Możemy zdefiniować iloczyn uogólniony w następujący sposób:

(P) $\prod_{j=1}^{1} a_j = a_1$,
(R) $\prod_{j=1}^{n+1} a_j = a_{n+1} \cdot \prod_{j=1}^{n} a_j$ dla $n \geq 1$.

Równoważnie:

(P) $\text{PROD}(1) = a_1$,
(R) $\text{PROD}(n+1) = a_{n+1} \cdot \text{PROD}(n)$ dla $n \geq 1$.

Te definicje rekurencyjne rozpoczynają się dla $n = 1$. Można także zdefiniować „pusty iloczyn" równy 1, tzn.

(P) $\prod_{j=1}^{0} a_j = 1$ (co wygląda dziwnie),
lub
(P) $\text{PROD}(0) = 1$.

Wtedy te same wzory rekurencyjne (R), co poprzednio, mogą służyć do zdefiniowania pozostałych wyrazów ciągu. ∎

ĆWICZENIA DO § 4.3

1. Definiujemy rekurencyjnie $s_0 = 1$ i $s_{n+1} = 2/s_n$ dla $n \in \mathbb{N}$.
(a) Wypisz kilka pierwszych wyrazów tego ciągu.
(b) Jaki jest zbiór wartości ciągu s?

2. Definiujemy rekurencyjnie $\text{SEQ}(0) = 0$ i $\text{SEQ}(n+1) = 1/[1 + \text{SEQ}(n)]$ dla $n \in \mathbb{N}$. Oblicz $\text{SEQ}(n)$ dla $n = 1, 2, 3, 4$ oraz 6.

3. Weźmy ciąg SEQ: $(1, 3, 9, 27, 81, \ldots)$.
(a) Podaj wzór na n-ty wyraz ciągu $\text{SEQ}(n)$, gdzie $\text{SEQ}(0) = 1$.
(b) Podaj definicję rekurencyjną ciągu SEQ.

4. (a) Podaj definicję rekurencyjną ciągu

$$(2, 2^2, (2^2)^2, ((2^2)^2)^2, \ldots),$$

tzn. ciągu $(2, 4, 16, 256, \ldots)$.
(b) Podaj definicję rekurencyjną ciągu

$$(2, 2^2, 2^{(2^2)}, 2^{(2^{(2^2)})}, \ldots),$$

tzn. ciągu $(2, 4, 16, 65536, \ldots)$.

5. Czy następująca definicja jest definicją rekurencyjną ciągu SEQ? Odpowiedź uzasadnij.
(P) $\text{SEQ}(0) = 1$,
(R) $\text{SEQ}(n+1) = \text{SEQ}(n)/(100 - n)$.

6. (a) Oblicz SEQ(9), gdzie SEQ jest ciągiem z przykładu 1(c).
(b) Oblicz FIB(11), gdzie FIB jest ciągiem z przykładu 3(a).

7. Niech $\Sigma = \{a, b, c\}$ i niech s_n oznacza liczbę słów długości n, które nie mają kolejnych liter a.
(a) Oblicz s_0, s_1 i s_2.
(b) Znajdź wzór rekurencyjny na s_n.
(c) Oblicz s_3 i s_4.

8. Niech $\Sigma = \{a, b\}$ i niech s_n oznacza liczbę słów długości n, nie zawierających ciągu ab.
(a) Oblicz s_0, s_1, s_2 i s_3.
(b) Znajdź wzór na s_n i udowodnij, że jest on poprawny.

9. Niech $\Sigma = \{a, b\}$ i niech t_n oznacza liczbę słów długości n, w których jest parzysta liczba liter a.
(a) Oblicz t_0, t_1, t_2 i t_3.
(b) Znajdź wzór na t_n i udowodnij, że jest on poprawny.
(c) Czy twój wzór na t_n jest prawdziwy dla $n = 0$?

10. Weźmy ciąg SEQ określony w następujący sposób:
(P) $\text{SEQ}(0) = 1$, $\text{SEQ}(1) = 0$,
(R) $\text{SEQ}(n) = \text{SEQ}(n-2)$ dla $n \geq 2$.
(a) Wypisz kilka pierwszych wyrazów tego ciągu.
(b) Jaki jest zbiór wartości tego ciągu?

11. Definiujemy rekurencyjnie ciąg za pomocą wzorów $a_0 = a_1 = 1$ oraz $a_n = a_{n-1} + 2a_{n-2}$ dla $n \geq 2$.
(a) Oblicz rekurencyjnie a_6.
(b) Udowodnij, że wszystkie wyrazy ciągu a_n są nieparzyste.

12. Udowodnij, że ciąg macierzy $\mathbf{M}_1, \mathbf{M}_2, \ldots$ określony wzorami

$$\mathbf{M}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix},$$

$$\mathbf{M}_n = \begin{bmatrix} \text{FIB}(n) & \text{FIB}(n-1) \\ \text{FIB}(n-1) & \text{FIB}(n-2) \end{bmatrix}$$

dla $n \geq 2$ spełnia zależność $\mathbf{M}_{n+1} = \mathbf{M}_1 \cdot \mathbf{M}_n$ dla $n \in \mathbb{P}$.

13. Pokaż, że jeśli ciąg s_n spełnia równania $s_0 = 1$, $s_1 = 2$ i $s_n = s_{n-1} + s_{n-2}$ dla $n \geq 2$, to $s_n = \text{FIB}(n+1)$ dla $n \in \mathbb{N}$. *Wskazówka*: weź najmniejszy element zbioru $S = \{n \in \mathbb{N}: \ s_n \neq \text{FIB}(n+1)\}$.

14. Definiujemy rekurencyjnie ciąg za pomocą wzorów $b_0 = b_1 = 1$ oraz $b_n = 2b_{n-1} + b_{n-2}$ dla $n \geq 2$.
(a) Oblicz b_5 metodą iteracyjną.
(b) Wyjaśnij, dlaczego wszystkie wyrazy b_n są nieparzyste. *Wskazówka*: rozważ pierwszy wyraz parzysty.

15. Niech $\text{SEQ}(0) = 1$ i niech $\text{SEQ}(n) = \sum_{i=0}^{n-1} \text{SEQ}(i)$ dla $n \geq 1$. Okazuje się, że jest to znany, prosty ciąg. Jaki?

16. Definiujemy rekurencyjnie ciąg wzorami $a_0 = 0$, $a_1 = 1$, $a_2 = 2$ oraz $a_n = a_{n-1} - a_{n-2} + a_{n-3}$ dla $n \geq 3$.
(a) Wypisz kilka pierwszych wyrazów tego ciągu, aż pojawi się pewna prawidłowość.
(b) Jaki jest zbiór wartości tego ciągu?

17. Proces przydzielania n dzieci do n miejsc w klasie może być podzielony na dwie fazy: (1) wybór dziecka na pierwsze miejsce i (2) przypisanie pozostałych $n-1$ dzieci do pozostałych miejsc. Niech $A(n)$ będzie liczbą różnych przypisań n dzieci do n miejsc.
(a) Napisz definicję rekurencyjną ciągu A.
(b) Oblicz rekurencyjnie $A(6)$.
(c) Czy ciąg A wydaje Ci się znajomy?

18. Weźmy pod uwagę proces przydzielania $2n$ dzieci do n wagoników kolejki tak, by w każdym wagoniku było dwoje dzieci. Najpierw wybieramy dwoje dzieci do pierwszego wagonika (można to zrobić na $2n(2n-1)/2$ sposobów, jak zobaczymy to w rozdziale 5). Następnie przydzielamy pozostałe dzieci do pozostałych $n-1$ wagoników. Niech $B(n)$ będzie liczbą sposobów, na które można przydzielić $2n$ dzieci do n wagoników.
(a) Napisz definicję rekurencyjną ciągu B.
(b) Oblicz rekurencyjnie $B(3)$.
(c) Oblicz $B(5)$ metodą iteracyjną.
(d) Podaj wzór ogólny na $B(n)$.

19. Weźmy ciąg FOO określony wzorami

$$\text{(P) } \textsc{foo}(0) = 1,\ \textsc{foo}(1) = 1,$$

$$\text{(R) } \textsc{foo}(n) = \frac{10 \cdot \textsc{foo}(n-1) + 100}{\textsc{foo}(n-2)} \quad \text{dla} \quad n \geq 2.$$

(a) Jaki jest zbiór wartości ciągu FOO?
(b) Powtórz ćwiczenie (a) dla ciągu GOO określonego wzorami

$$\text{(P) } \textsc{goo}(0) = 1,\ \textsc{goo}(1) = 2,$$

$$\text{(R) } \textsc{goo}(n) = \frac{10 \cdot \textsc{goo}(n-1) + 100}{\textsc{goo}(n-2)} \quad \text{dla} \quad n \geq 2.$$

Dowiedzieliśmy się o tych interesujących ciągach od naszego kolegi Ivana Nivena.

20. Niech $(a_1, a_2, \ldots)$ będzie ciągiem liczb rzeczywistych.
(a) Podaj definicję rekurencyjną ciągu $\textsc{suma}(n) = \sum_{j=1}^{n} a_j$ dla $n \geq 1$.
(b) Zmień twoją definicję rekurencyjną ciągu $\textsc{suma}(n)$, zaczynając od $n = 0$. Co to jest „pusta suma"?

21. Niech $(A_1, A_2, \ldots)$ będzie ciągiem podzbiorów pewnego zbioru S.
(a) Podaj definicję rekurencyjną dla $\bigcup_{j=1}^{n} A_j$.
(b) Jak zdefiniowałbyś „sumę pustą"?
(c) Podaj definicję rekurencyjną dla $\bigcap_{j=1}^{n} A_j$.
(d) Jak zdefiniowałbyś „puste przecięcie"?

22. Niech $(A_1, A_2, \ldots)$ będzie ciągiem podzbiorów pewnego zbioru S. Definiujemy

(P) $\text{SYM}(1) = A_1$,
(R) $\text{SYM}(n+1) = A_{n+1} \oplus \text{SYM}(n)$ dla $n \geq 1$.

Przypominamy, że $\oplus$ oznacza różnicę symetryczną. Okazuje się, że element $x \in S$ należy do ciągu $\text{SYM}(n)$ wtedy i tylko wtedy, gdy zbiór $\{k\colon\ x \in A_k \text{ i } k \leq n\}$ ma nieparzystą liczbę elementów. Udowodnij ten fakt metodą indukcji matematycznej.

§ 4.4. Zależności rekurencyjne

Ciągi spotykane w matematyce i naukach przyrodniczych są często zdefiniowane w sposób rekurencyjny, a nie za pomocą wzoru. Opracowano wiele metod otrzymywania wzorów jawnych na wyrazy takich ciągów. W tym paragrafie podamy pełne rozwiązanie tego problemu dla ciągów określonych za pomocą wzorów rekurencyjnych postaci:

$$s_n = as_{n-1} + bs_{n-2},$$

gdzie a i b są stałymi, a także uzyskamy istotne informacje na temat ciągów spełniających zależności rekurencyjne postaci

$$s_{2n} = 2s_n + f(n)$$

dla znanych funkcji f.

Zaczniemy od rozpatrzenia zależności

$$s_n = as_{n-1} + bs_{n-2},$$

przy założeniu, że znane są dwie początkowe wartości s_0 i s_1. Szczególnie łatwo zbadać przypadki, gdy $a = 0$ lub $b = 0$.

Jeśli $b = 0$, a więc $s_n = as_{n-1}$ dla $n \geq 1$, to $s_1 = as_0$, $s_2 = as_1 = a^2 s_0$ itd. Proste rozumowanie indukcyjne pokazuje, że $s_n = a^n s_0$ dla wszystkich $n \in \mathbb{N}$.

Teraz przypuśćmy, że $a = 0$. Wtedy $s_2 = bs_0$, $s_4 = bs_2 = b^2 s_0$ itd., a więc $s_{2n} = b^n s_0$ dla wszystkich $n \in \mathbb{N}$. Podobnie $s_3 = bs_1$, $s_5 = b^2 s_1$ itd., a więc $s_{2n+1} = b^n s_1$ dla wszystkich $n \in \mathbb{N}$.

PRZYKŁAD 1 (a) Weźmy zależność rekurencyjną $s_n = 3s_{n-1}$, gdzie $s_0 = 5$. W tym przypadku $a = 3$, więc $s_n = 5 \cdot 3^n$ dla $n \in \mathbb{N}$.

(b) Weźmy zależność rekurencyjną $s_n = 3s_{n-2}$, gdzie $s_0 = 5$ i $s_1 = 2$. W tym przypadku $b = 3$, więc $s_{2n} = 5 \cdot 3^n$ i $s_{2n+1} = 2 \cdot 3^n$ dla $n \in \mathbb{N}$. ∎

W dalszym ciągu będziemy zakładać, że $a \neq 0$ i $b \neq 0$. Wygodnie będzie na początku nie interesować się konkretnymi wartościami s_0 i s_1. Zbadane przez nas szczególne przypadki pokazały, że wydaje się być rozsądnym przypuszczenie, że pewne rozwiązania mają postać $s_n = cr^n$ dla jakiejś stałej c. Gdyby to przypuszczenie było prawdziwe, to musiałaby zachodzić równość

$$r^n = ar^{n-1} + br^{n-2}.$$

Dzieląc obie strony przez r^{n-2}, otrzymalibyśmy wtedy $r^2 = ar+b$, czyli $r^2 - ar - b = 0$. Innymi słowy, jeśli $s_n = cr^n$ dla wszystkich n, to r musi być rozwiązaniem równania kwadratowego $x^2 - ax - b = 0$, które nazywamy **równaniem charakterystycznym** zależności rekurencyjnej. Równanie charakterystyczne ma albo jedno, albo dwa rozwiązania. Będziemy odtąd zakładać, że te rozwiązania są liczbami rzeczywistymi.

Twierdzenie 1

Rozważmy zależność rekurencyjną postaci

$$s_n = as_{n-1} + bs_{n-2},$$

mającą równanie charakterystyczne

$$x^2 - ax - b = 0,$$

gdzie a i b są stałymi niezerowymi.
(a) Jeśli równanie charakterystyczne ma dwa różne rozwiązania r_1 i r_2, to

$$(*) \qquad s_n = c_1 r_1^n + c_2 r_2^n$$

dla pewnych stałych c_1 i c_2. Jeśli s_0 i s_1 są dane, to wartości stałych mogą być wyznaczone przez podstawienie $n = 0$ i $n = 1$ do równania $(*)$ i rozwiązanie układu dwóch równań z niewiadomymi c_1 i c_2.
(b) Jeśli równanie charakterystyczne ma tylko jedno rozwiązanie r, to

$$(**) \qquad s_n = c_1 r^n + c_2 \cdot n \cdot r^n$$

dla pewnych stałych c_1 i c_2. Tak jak w punkcie (a), można wyznaczyć c_1 i c_2, jeśli dane są s_0 i s_1.

Uwaga! To twierdzenie stosuje się tylko do zależności rekurencyjnych postaci

$$s_n = as_{n-1} + bs_{n-2}.$$

PRZYKŁAD 2 Weźmy zależność rekurencyjną $s_n = s_{n-1} + 2s_{n-2}$, w której $s_0 = s_1 = 3$. Mamy więc $a = 1$ i $b = 2$. Równanie charakterystyczne $x^2 - x - 2 = 0$ ma rozwiązania $r_1 = 2$ i $r_2 = -1$, ponieważ $x^2 - x - 2 = (x - 2)(x + 1)$. Zatem ma zastosowanie punkt (a) twierdzenia. Z twierdzenia 1

$$s_n = c_1 \cdot 2^n + c_2 \cdot (-1)^n$$

dla pewnych stałych c_1 i c_2. Podstawiając $n = 0$ i $n = 1$ otrzymujemy

$$s_0 = c_1 \cdot 2^0 + c_2 \cdot (-1)^0$$

oraz

$$s_1 = c_1 \cdot 2^1 + c_2 \cdot (-1)^1,$$

czyli

$$3 = c_1 + c_2 \quad \text{oraz} \quad 3 = 2c_1 - c_2.$$

Rozwiązując ten układ równań otrzymujemy $c_1 = 2$ i $c_2 = 1$. Ostatecznie mamy

$$s_n = 2 \cdot 2^n + 1 \cdot (-1)^n = 2^{n+1} + (-1)^n$$

dla $n \in \mathbb{N}$. ∎

PRZYKŁAD 3 Weźmy ponownie ciąg Fibonacciego z § 4.3. Oznaczając $\text{FIB}(n)$ przez s_n, mamy $s_0 = s_1 = 1$ oraz $s_n = s_{n-1} + s_{n-2}$ dla $n \geq 2$. Mamy tutaj $a = b = 1$, więc rozwiązujemy równanie $x^2 - x - 1 = 0$. Równanie ma dwa rozwiązania:

$$r_1 = \frac{1 + \sqrt{5}}{2} \quad \text{oraz} \quad r_2 = \frac{1 - \sqrt{5}}{2}.$$

Zatem ma zastosowanie punkt (a) twierdzenia, czyli

$$s_n = c_1 \left(\frac{1 + \sqrt{5}}{2} \right)^n + c_2 \left(\frac{1 - \sqrt{5}}{2} \right)^n \quad \text{dla} \quad n \in \mathbb{N}.$$

Wyznaczając c_1 i c_2 wygodniej będzie zachować oznaczenia r_1 i r_2. Podstawiając $n = 0$ i $n = 1$, otrzymujemy

$$1 = c_1 + c_2 \quad \text{oraz} \quad 1 = c_1 r_1 + c_2 r_2.$$

Jeśli podstawimy $c_2 = 1 - c_1$ do drugiego równania, otrzymamy $1 = c_1 r_1 + (1 - c_1) r_2$, więc $1 - r_2 = c_1(r_1 - r_2)$ oraz

$$c_1 = \frac{1 - r_2}{r_1 - r_2}.$$

Ponieważ $r_1 + r_2 = 1$ i $r_1 - r_2 = \sqrt{5}$, więc $c_1 = r_1/\sqrt{5}$. Następnie, $c_2 = 1 - c_1 = (\sqrt{5} - r_1)/\sqrt{5} = -r_2/\sqrt{5}$. Ostatecznie

$$s_n = c_1 r_1^n + c_2 r_2^n = \frac{r_1}{\sqrt{5}} r_1^n - \frac{r_2}{\sqrt{5}} r_2^n$$
$$= \frac{1}{\sqrt{5}}(r_1^{n+1} - r_2^{n+1}),$$

a więc

$$\text{FIB}(n) = s_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right].$$

Trudno było oczekiwać na początku, że wzór na $\text{FIB}(n)$ będzie taki. $\text{FIB}(n)$ jest oczywiście liczbą całkowitą, podczas gdy to wyrażenie z pierwiastkami zupełnie nie wygląda na liczbę całkowitą. Jednakże wzór jest poprawny. Spróbuj obliczyć na kalkulatorze wyrazy ciągu dla kilku wartości n, aby się o tym przekonać. Korzystając z tego wzoru możemy obliczyć dowolną wartość $\text{FIB}(n)$ bez obliczania poprzednich wartości i możemy też oszacować szybkość, z jaką rośnie $\text{FIB}(n)$, kiedy zwiększa się n. Ponieważ

$$\frac{\sqrt{5}-1}{2} < 0{,}62,$$

to wartość bezwzględna wyrażenia

$$\frac{1}{\sqrt{5}} \cdot \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}$$

jest mniejsza od 0,5 dla każdej liczby dodatniej n, a zatem $\text{FIB}(n)$ jest liczbą całkowitą najbliższą liczby

$$\frac{1}{\sqrt{5}} \cdot \left(\frac{1+\sqrt{5}}{2}\right)^{n+1}. \qquad \blacksquare$$

PRZYKŁAD 4 Weźmy ciąg (s_n) określony wzorami $s_0 = 1, s_1 = -3$ i $s_n = 6s_{n-1} - 9s_{n-2}$ dla $n \geq 2$. Równaniem charakterystycznym jest $x^2 - 6x + 9 = 0$, ma ono jedno rozwiązanie, mianowicie $r = 3$. Na podstawie punktu (b) twierdzenia otrzymujemy

$$s_n = c_1 \cdot 3^n + c_2 \cdot n \cdot 3^n \quad \text{dla} \quad n \in \mathbb{N}.$$

Podstawiając $n = 0$ i $n = 1$, otrzymujemy

$$s_0 = c_1 \cdot 3^0 + 0 \quad \text{oraz} \quad s_1 = c_1 \cdot 3^1 + c_2 \cdot 3^1,$$

czyli

$$1 = c_1 \quad \text{oraz} \quad -3 = 3c_1 + 3c_2.$$

Tak więc $c_1 = 1$ i $c_2 = -2$. Stąd

$$s_n = 3^n - 2 \cdot n \cdot 3^n \quad \text{dla} \quad n \in \mathbb{N}. \qquad \blacksquare$$

Dowód twierdzenia 1. (a) Niezależnie od tego, jakie są s_0 i s_1, układ równań

$$s_0 = c_1 + c_2, \qquad s_1 = c_1 r_1 + c_2 r_2$$

z niewiadomymi c_1 i c_2 ma rozwiązanie, ponieważ $r_1 \neq r_2$. Szukany ciąg (s_n) jest określony przez wartości s_0 i s_1 oraz zależność rekurencyjną $s_n = as_{n-1} + bs_{n-2}$, a więc wystarczy pokazać, że ciąg określony przez $(*)$ również spełnia tę zależność rekurencyjną. Ponieważ $x = r_1$ spełnia równanie $x^2 = ax + b$, mamy $r_1^n = ar_1^{n-1} + br_1^{n-2}$, więc ciąg (r_1^n) spełnia zależność $s_n = as_{n-1} + bs_{n-2}$. Podobnie spełnia ją ciąg (r_2^n). Teraz już łatwo sprawdzić, że ciąg określony wzorem $(*)$ również spełnia tę zależność rekurencyjną

$$\begin{aligned} as_{n-1} + bs_{n-2} &= a(c_1 r_1^{n-1} + c_2 r_2^{n-1}) + \\ &\quad + b(c_1 r_1^{n-2} + c_2 r_2^{n-2}) \\ &= c_1(ar_1^{n-1} + br_1^{n-2}) + \\ &\quad + c_2(ar_2^{n-1} + br_2^{n-2}) \\ &= c_1 r_1^n + c_2 r_2^n = s_n. \end{aligned}$$

(b) Jeśli r jest jedynym rozwiązaniem równania charakterystycznego, to równanie charakterystyczne ma postać $(x-r)^2 = 0$. Tak więc $x^2 - 2rx + r^2 = x^2 - ax - b$, czyli $a = 2r$ i $b = -r^2$. Zależność rekurencyjną można teraz zapisać w następujący sposób

$$s_n = 2rs_{n-1} - r^2 s_{n-2}.$$

Podstawiając $n = 0$ i $n = 1$ do wzoru $(**)$, otrzymujemy równania

$$s_0 = c_1, \qquad s_1 = c_1 r + c_2 r.$$

Ponieważ $r \neq 0$, ten układ równań ma rozwiązanie $c_1 = s_0$ oraz $c_2 = -s_0 + s_1/r$. Tak jak w naszym dowodzie punktu (a), wystarczy pokazać, że dowolny ciąg określony wzorem $(**)$ spełnia zależność $s_n = 2rs_{n-1} - r^2 s_{n-2}$. Ale

$$\begin{aligned} 2rs_{n-1} - r^2 s_{n-2} &= 2r[c_1 r^{n-1} + c_2(n-1)r^{n-1}] + \\ &\quad - r^2[c_1 r^{n-2} + c_2(n-2)r^{n-2}] \\ &= 2c_1 r^n + 2c_2(n-1)r^n - c_1 r^n - c_2(n-2)r^n \\ &= c_1 r^n + c_2 \cdot n \cdot r^n = s_n. \end{aligned}$$

$\blacksquare$

Dowód twierdzenia jest również poprawny, jeśli pierwiastki równania charakterystycznego nie są liczbami rzeczywistymi, a więc twierdzenie jest też prawdziwe w tym przypadku. Obliczenie wyrazów ciągu za pomocą wzoru (*) będzie wtedy wymagało działań na liczbach zespolonych, ale wyniki będą oczywiście liczbami rzeczywistymi, jeśli a, b, s_0 i s_1 były liczbami rzeczywistymi. Ta sytuacja jest analogiczna do obliczenia liczb Fibonacciego, które są liczbami całkowitymi, wykorzystującego liczbę $\sqrt{5}$, która nie jest całkowita.

Twierdzenie 1 stosuje się tylko do zależności rekurencyjnych postaci $s_n = as_{n-1} + bs_{n-2}$, ale ponieważ takie zależności rekurencyjne występują dość często, jest ono użytecznym narzędziem. Inny ważny rodzaj zależności rekurencyjnych pojawia się przy szacowaniu czasu działania tzw. **algorytmów „dziel i rządź"**. Algorytm tego typu dzieli problem wejściowy na dwie lub więcej części, rozwiązuje te części oddzielnie, a następnie składa wyniki razem, by otrzymać ostateczny wynik.

Prostym przykładem algorytmu „dziel i rządź" jest proces znajdowania największej liczby w pewnym zbiorze liczb, polegający na rozbiciu tego zbioru na dwie części, mniej więcej tej samej wielkości, znalezieniu największych liczb w każdej części i porównaniu tych dwóch liczb, by otrzymać liczbę, która jest największa w całym zbiorze. Innym przykładem jest sortowanie przez łączenie zbioru liczb, polegające na sortowaniu obu części oddzielnie i łączeniu obu posortowanych list.

Jeśli $T(n)$ oznacza czas, jaki jest potrzebny do wykonania algorytmu dla danych rozmiaru n, to taka struktura algorytmu prowadzi do zależności rekurencyjnej postaci

$$T(n) = T(n/2) + T(n/2) + F(n),$$

gdzie dwa składniki $T(n/2)$ oznaczają czas potrzebny do wykonania algorytmu dla każdej z części danych oddzielnie, a $F(n)$ oznacza czas potrzebny do połączenia obu wyników. Oczywiście to równanie ma sens tylko wtedy, gdy n jest liczbą parzystą.

Nasze następne twierdzenie podaje informacje o zależnościach rekurencyjnych postaci

$$s_{2n} = 2 \cdot s_n + f(n).$$

Sformułowaliśmy to twierdzenie w dość ogólnej postaci, wyróżniając jeden ważny przypadek szczególny. W ćwiczeniu 17 rozważamy dwie wariacje na ten temat.

Twierdzenie 2

Niech (s_n) będzie ciągiem spełniającym zależność rekurencyjną postaci

$$s_{2n} = 2 \cdot s_n + f(n) \quad \text{dla} \quad n \in \mathbb{P}.$$

Wtedy

$$s_{2^m} = 2^m \cdot \left[s_1 + \frac{1}{2} \sum_{i=0}^{m-1} \frac{f(2^i)}{2^i} \right] \quad \text{dla} \quad m \in \mathbb{N}.$$

W szczególności, jeśli

$$s_{2n} = 2 \cdot s_n + A + B \cdot n$$

dla pewnych stałych A i B, to

$$s_{2^m} = 2^m \cdot s_1 + (2^m - 1) \cdot A + \frac{B}{2} \cdot 2^m \cdot m.$$

Zatem, jeśli $n = 2^m$, to w tym przypadku mamy

$$s_n = n s_1 + (n-1)A + \frac{B}{2} \cdot n \cdot \log_2 n.$$

Zanim przeprowadzimy dowód, popatrzmy, jak możemy zastosować to twierdzenie.

PRZYKŁAD 5

(a) Znajdowanie największej liczby w zbiorze za pomocą metody „dziel i rządź" prowadzi do zależności rekurencyjnej

$$T(2n) = 2T(n) + A,$$

gdzie stała A wyraża czas potrzebny do porównania największych liczb w obu połówkach zbioru. Zgodnie z twierdzeniem 2, gdy $B = 0$,

$$T(2^m) = 2^m \cdot T(1) + (2^m - 1) \cdot A.$$

$T(1)$ jest czasem potrzebnym do znalezienia największego elementu w zbiorze jednoelementowym. Wydaje się rozsądnym rozpatrywać $T(1)$ jako koszt sprawdzania jednego elementu, podczas, gdy A jest kosztem jednego porównania dwóch elementów. Jeśli $n = 2^m$, to otrzymujemy

$$T(n) = n \cdot T(1) + (n-1) \cdot A.$$

Sprawdzamy n elementów i dokonujemy $n - 1$ porównań, by znaleźć element największy. W tym przypadku metoda „dziel i rządź" nie daje istotnej poprawy w stosunku do zwykłego przeglądania elementów po kolei i zapamiętywania w każdym kroku największej dotychczas znalezionej liczby.

(b) Sortowanie zbioru metodą „dziel i rządź" daje zależność rekurencyjną

$$T(2n) = 2 \cdot T(n) + B \cdot n,$$

ponieważ czas potrzebny do połączenia dwóch uporządkowanych połówek w jeden zbiór jest proporcjonalny do rozmiaru obu łączonych zbiorów. Twierdzenie 2 dla $A = 0$ mówi, że

$$T(2^m) = 2^m \cdot T(1) + \frac{B}{2} \cdot 2^m \cdot m.$$

Jeśli $n = 2^m$, to

$$T(n) = n \cdot T(1) + \frac{B}{2} \cdot n \cdot \log_2 n = O(n \cdot \log_2 n).$$

Dla dużych n dominujący koszt pochodzi z łączenia zbiorów, a nie ze sprawdzania pojedynczych elementów. Ten algorytm jest w rzeczywistości wystarczająco wydajny dla wielu zastosowań. ■

Dowód twierdzenia 2. Sprawdzimy, że jeśli

$$(*) \qquad s_{2^m} = 2^m \cdot \left[s_1 + \frac{1}{2} \sum_{i=0}^{m-1} \frac{f(2^i)}{2^i} \right]$$

dla $m \in \mathbb{N}$, to

$$s_{2^{m+1}} = 2^{m+1} \cdot \left[s_1 + \frac{1}{2} \sum_{i=0}^{m} \frac{f(2^i)}{2^i} \right].$$

Ponieważ $s_{2^0} = 2^0 \cdot s_1$, gdyż nie ma składników sumy, jeśli $m = 0$, więc z zasady indukcji względem m wynika, że wzór $(*)$ zachodzi dla każdej liczby $m \in \mathbb{N}$. (Chociaż nie jest konieczne sprawdzenie oddzielnie przypadku $m = 1$, Czytelnik może będzie chciał się upewnić). A oto szczegóły dowodu:

$$\begin{aligned}
s_{2^{m+1}} &= s_{2 \cdot 2^m} \\
&= 2 \cdot s_{2^m} + f(2^m) && \text{(zależność rekurencyjna)} \\
&= 2^{m+1} \cdot \left[s_1 + \frac{1}{2} \sum_{i=0}^{m-1} \frac{f(2^i)}{2^i} \right] + \\
&\quad + f(2^m) && \text{(na podstawie } (*)\text{)} \\
&= 2^{m+1} \cdot \left[s_1 + \frac{1}{2} \sum_{i=0}^{m-1} \frac{f(2^i)}{2^i} + \frac{1}{2 \cdot 2^m} f(2^m) \right] \\
&= 2^{m+1} \cdot \left[s_1 + \frac{1}{2} \sum_{i=0}^{m} \frac{f(2^i)}{2^i} \right].
\end{aligned}$$

W przypadku szczególnym, gdy $f(n) = A + Bn$, możemy dość łatwo obliczyć sumę, ale można równie łatwo sprawdzić bezpośrednio (ćwiczenie 16), że

$$s_{2^m} = 2^m \cdot s_1 + (2^m - 1) \cdot A + \frac{B}{2} \cdot 2^m \cdot m,$$

jeśli ciąg (s_n) spełnia zależność rekurencyjną $s_{2n} = 2 \cdot s_n + A + Bn$. ∎

Ten dowód jest technicznie poprawny, ale nie jest pouczający. Po pierwsze, w jaki sposób otrzymaliśmy wzór na s_{2^m}? To łatwe. Po prostu z zależności rekurencyjnej obliczyliśmy kilka pierwszych wyrazów: s_2, s_4, s_8, s_{16}, a następnie odgadliśmy wzór.

Twierdzenie mówi nam tylko, jakie są wartości s_n, gdy n jest potęgą 2. Pozostawia to ogromną lukę w naszej wiedzy. Jednakże często okazuje się, że ciąg jest **monotoniczny**, tzn. $s_k \leq s_n$ dla $k \leq n$. Na przykład tę własność mają na ogół oszacowania czasu działania algorytmów. Jeśli ciąg s jest monotoniczny i jeśli znamy funkcję g taką, że $s_n \leq g(n)$ dla n będących potęgami 2, to dla dowolnej liczby całkowitej k istnieje liczba m taka, że $2^{m-1} < k \leq 2^m$, a więc $s_{2^{m-1}} \leq s_k \leq s_{2^m} \leq g(2^m)$.

Często chcemy mieć oszacowanie wielkości s_n, na ogół postaci $s_n = O(h(n))$ dla pewnej funkcji h. Zastąpienie znaku „=" znakiem „≤" w dowodzie twierdzenia 2 pokazuje, że jeśli

$$s_{2n} \leq 2 \cdot s_n + f(m) \quad \text{dla} \quad n \in \mathbb{P},$$

to

$$s_{2^m} \leq 2^m \cdot \left[s_1 + \frac{1}{2} \sum_{i=0}^{m-1} \frac{f(2^i)}{2^i} \right] \quad \text{dla} \quad m \in \mathbb{N}.$$

W szczególności, jeśli $s_{2n} \leq 2 \cdot s_n + A$, to

$$s_{2^m} \leq 2^m \cdot s_1 + (2^m - 1) \cdot A \ = O(2^m).$$

Jeśli ponadto s jest ciągiem monotonicznym i $2^{m-1} < k \leq 2^m$, to $s_k \leq s_{2^m} \leq C \cdot 2^m$ dla pewnej stałej C (na przykład $C = s_1 + A$), a więc $s_k \leq C \cdot 2 \cdot 2^{m-1} \leq C \cdot 2 \cdot k$. To znaczy, że $s_k = O(k)$.

Nierówność $s_{2n} \leq 2 \cdot s_n + B \cdot n$ daje podobnie $s_{2^m} = O(2^m \cdot m)$, a więc $s_k = O(k \cdot \log_2 k)$. Tak naprawdę szacowanie czasu sortowania przez łączenie w przykładzie 5(b) zrobiliśmy trochę niedokładnie. Jeśli jedna z dwóch list wyczerpie się wcześniej, możemy dołączyć pozostałe elementy drugiej listy, nic z nimi już nie robiąc. To, co naprawdę otrzymaliśmy w tym przykładzie, to nierówność $T(2n) \leq 2T(n) + Bn$, która oczywiście nadal daje $T(n) = O(n \cdot \log_2 n)$.

ĆWICZENIA DO § 4.4

1. Podaj wzór jawny na s_n, gdzie $s_0 = 3$ oraz $s_n = -2s_{n-1}$ dla $n \geq 1$.
2. (a) Podaj wzór jawny na $s_n = 4s_{n-2}$, gdzie $s_0 = s_1 = 1$.
   (b) Powtórz ćwiczenie (a) dla $s_0 = 1$ i $s_1 = 2$.
3. Udowodnij, że jeśli $s_n = as_{n-1}$ dla $n \geq 1$ i $a \neq 0$, to $s_n = a^n \cdot s_0$ dla $n \in \mathbb{N}$.
4. Sprawdź, że ciąg s_n z przykładu 2, dany wzorem $s_n = 2^{n+1} + (-1)^n$ spełnia warunki: $s_0 = s_1 = 3$ oraz $s_n = s_{n-1} + 2s_{n-2}$ dla $n \geq 2$.
5. Sprawdź, że ciąg s_n z przykładu 4, dany wzorem $s_n = 3^n - 2 \cdot n \cdot 3^n$ spełnia warunki: $s_0 = 1$, $s_1 = -3$ oraz $s_n = 6s_{n-1} - 9s_{n-2}$ dla $n \geq 2$.
6. Skorzystaj z wzoru na FIB(n) z przykładu 3 i sprawdź za pomocą kalkulatora, że FIB(5) = 8.
7. Podaj wzór jawny na s_n, gdzie $s_0 = 3$, $s_1 = 6$ i $s_n = s_{n-1} + 2s_{n-2}$ dla $n \geq 2$. *Wskazówka*: powtórz rozumowanie z przykładu 2, ale zauważ, że teraz $s_1 = 6$.
8. Powtórz ćwiczenie 7 dla $s_0 = 3$ i $s_1 = -3$.
9. Podaj wzór jawny na s_n z przykładu 4 z § 4.3, gdzie $s_0 = 1$, $s_1 = 2$ oraz $s_n = s_{n-1} + s_{n-2}$ dla $n \geq 2$. *Wskazówka*: skorzystaj z przykładu 3.
10. Weźmy ciąg s_n, gdzie $s_0 = 2$, $s_1 = 1$ oraz $s_n = s_{n-1} + s_{n-2}$ dla $n \geq 2$.
    (a) Oblicz s_n dla $n = 2, 3, 4, 5$ oraz 6.
    (b) Podaj wzór jawny na s_n.
11. W każdym z następujących przypadków podaj wzór jawny na s_n:
    (a) $s_0 = 2$, $s_1 = -1$ oraz $s_n = -s_{n-1} + 6s_{n-2}$ dla $n \geq 2$.
    (b) $s_0 = 2$ oraz $s_n = 5 \cdot s_{n-1}$ dla $n \geq 1$.
    (c) $s_0 = 1$, $s_1 = 8$ oraz $s_n = 4s_{n-1} - 4s_{n-2}$ dla $n \geq 2$.
    (d) $s_0 = c$, $s_1 = d$ oraz $s_n = 5s_{n-1} - 6s_{n-2}$ dla $n \geq 2$. Liczby c i d są jakimiś stałymi.
    (e) $s_0 = 1$, $s_1 = 4$ oraz $s_n = s_{n-2}$ dla $n \geq 2$.
    (f) $s_0 = 1$, $s_1 = 2$ oraz $s_n = 3 \cdot s_{n-2}$ dla $n \geq 2$.
    (g) $s_0 = 1$, $s_1 = -3$ oraz $s_n = -2s_{n-1} + 3s_{n-2}$ dla $n \geq 2$.
    (h) $s_0 = 1$, $s_1 = 2$ oraz $s_n = -2s_{n-1} + 3s_{n-2}$ dla $n \geq 2$.
12. Przypomnijmy, że jeśli $s_n = bs_{n-2}$ dla $n \geq 2$, to $s_{2n} = b^n s_0$ oraz $s_{2n+1} = b^n s_1$ dla $n \in \mathbb{N}$. Pokaż, że twierdzenie 1 jest prawdziwe dla $a = 0$ i $b > 0$ oraz spróbuj pogodzić ten fakt z poprzednim zdaniem. To znaczy, określ r_1, r_2, c_1 i c_2 za pomocą b, s_0 i s_1.
13. W każdym z następujących przypadków podaj wzór jawny na s_{2^m}:
    (a) $s_{2n} = 2s_n + 3$, $s_1 = 1$.
    (b) $s_{2n} = 2s_n$, $s_1 = 3$.
    (c) $s_{2n} = 2s_n + 5n$, $s_1 = 0$.
    (d) $s_{2n} = 2s_n + 3 + 5n$, $s_1 = 2$.
    (e) $s_{2n} = 2s_n - 7$, $s_1 = 1$.

(f) $s_{2n} = 2s_n - 7$, $s_1 = 5$.
(g) $s_{2n} = 2s_n - n$, $s_1 = 3$.
(h) $s_{2n} = 2s_n + 5 - 7n$, $s_1 = 0$.

14. Przypuśćmy, że ciąg (s_n) spełnia daną nierówność oraz że $s_1 = 7$. Podaj najlepsze oszacowanie tego, jak duża może być liczba s_{2^m}.
(a) $s_{2n} \leq 2s_n + 1$.
(b) $s_{2n} \leq 2(s_n + n)$.

15. Przypuśćmy, że ciąg (s_n) spełnia zależność rekurencyjną

$$s_{2n} = 2s_n + n^2.$$

Podaj wzór na s_{2^m} oraz uzasadnij, że ten wzór jest poprawny. *Sugestia*: skorzystaj z twierdzenia 2 lub odgadnij wzór i sprawdź, że jest on poprawny.

16. Sprawdź wzór z twierdzenia 2 na s_{2^m} dla przypadku $f(n) = A + B \cdot n$.

17. Twierdzenie 2 nie jest szczególnie przydatne w następujących dwóch zależnościach rekurencyjnych, ale przenoszą się jego idee.
(a) Wiedząc, że $t_{2n} = b \cdot t_n + f(n)$ dla pewnej stałej b i funkcji f, znajdź wzór na t_{2^m} wyrażony za pomocą b, t_1 oraz wartości funkcji f.
(b) Wiedząc, że $t_{3n} = 3t_n + f(n)$, znajdź wzór na t_{3^m}.

§ 4.5. Więcej o indukcji

Zasada indukcji matematycznej omawiana w § 4.2 jest czasem nazywana pierwszą zasadą indukcji matematycznej. Przytoczymy ją tutaj w nieco zmienionej postaci.

Pierwsza zasada indukcji matematycznej

Niech m będzie liczbą całkowitą oraz niech $p(n)$ będzie ciągiem zdań zdefiniowanych na zbiorze $\{n \in \mathbb{Z}\colon\ n \geq m\}$. Jeśli

(P) zdanie $p(m)$ jest prawdziwe oraz
(I) dla $k > m$ zdanie $p(k)$ jest prawdziwe, jeśli zdanie $p(k-1)$ jest prawdziwe,

to zdanie $p(n)$ jest prawdziwe dla każdego $n \geq m$.

W kroku indukcyjnym (I) każde zdanie jest prawdziwe przy założeniu, że zdanie bezpośrednio je poprzedzające jest prawdziwe. Aby użyć tej zasady jako schematu do skonstruowania dowodu, musimy sprawdzić, że $p(m)$ jest zdaniem prawdziwym oraz, że każde zdanie jest prawdziwe przy założeniu, że zdanie tuż

przed nim jest prawdziwe. To właśnie to prawo, pozwalające założyć prawdziwość tego poprzedniego przypadku, sprawia, że metoda dowodzenia przez indukcję jest tak silna. Okazuje się, że tak naprawdę możemy założyć, że prawdziwe są wszystkie poprzednie przypadki. To z wyglądu silniejsze stwierdzenie jest wynikiem następującej zasady, której dowód omówimy na końcu tego paragrafu.

Druga zasada indukcji matematycznej

Niech n będzie liczbą całkowitą oraz niech $p(n)$ będzie ciągiem zdań zdefiniowanych na zbiorze $\{n \in \mathbb{Z}\colon\ n \geq m\}$. Jeśli

(P) zdanie $p(m)$ jest prawdziwe oraz
(I) dla $k > m$ zdanie $p(k)$ jest prawdziwe, jeśli wszystkie zdania $p(m), \ldots, p(k-1)$ są prawdziwe,

to zdanie $p(n)$ jest prawdziwe dla każdego $n \geq m$.

Aby sprawdzić warunek (I) dla $k = m+1$, pokazuje się, że ze zdania $p(m)$ wynika zdanie $p(m+1)$. Aby sprawdzić warunek (I) dla $k = m+2$, pokazuje się, że ze zdań $p(m)$ i $p(m+1)$ łącznie wynika $p(m+2)$. I tak dalej. Aby sprawdzić warunek (I) ogólnie, rozpatruje się liczbę $k > m$, zakłada, że zdania $p(n)$ są prawdziwe dla $m \leq n < k$ i pokazuje, że zdanie $p(k)$ jest prawdziwe. Druga zasada indukcji matematycznej jest właśnie tą wersją indukcji, której używa się wtedy, gdy prawdziwość zdań wynika z prawdziwości jakichś zdań poprzednich, a nie zdań bezpośrednio poprzedzających.

PRZYKŁAD 1

Pokażemy, że każda liczba całkowita $n \geq 2$ może być zapisana jako iloczyn liczb pierwszych. Zauważmy, że jeśli n jest liczbą pierwszą, „iloczyn liczb pierwszych" jest po prostu tą liczbą n.

Dla $n \geq 2$ niech $p(n)$ będzie zdaniem „n można zapisać jako iloczyn liczb pierwszych". Zauważmy, że pierwsza zasada indukcji matematycznej jest tutaj naprawdę nieodpowiednia. Pojedynczy fakt, że liczba, powiedzmy, 1311819 jest iloczynem liczb pierwszych, nie pomaga w dowodzie tego, że liczba 1311820 też jest iloczynem liczb pierwszych. Zastosujemy tu drugą zasadę. Oczywiście zdanie $p(2)$ jest prawdziwe, ponieważ 2 jest liczbą pierwszą.

Weźmy liczbę $k \geq 2$ i przyjmijmy, że zdanie $p(n)$ jest prawdziwe dla wszystkich n spełniających nierówności $2 \leq n < k$. Mamy pokazać, że $p(k)$ jest zdaniem prawdziwym. Jeśli k jest liczbą pierwszą, to $p(k)$ jest oczywiście prawdziwe. Jeśli nie jest liczbą pierwszą, to k może być zapisane jako iloczyn $i \cdot j$, gdzie i

i j są liczbami całkowitymi większymi od 1. Zatem $2 \leq i < k$ oraz $2 \leq j < k$. Ponieważ założyliśmy, że oba zdania $p(i)$ oraz $p(j)$ są prawdziwe, możemy zapisać i oraz j jako iloczyny liczb pierwszych. Wtedy $k = i \cdot j$ jest również iloczynem liczb pierwszych. Sprawdziliśmy krok początkowy i krok indukcyjny drugiej zasady indukcji matematycznej, a więc wnioskujemy stąd, że wszystkie zdania $p(n)$ są prawdziwe. ∎

Często ogólny dowód kroku indukcyjnego (I) nie działa dla kilku pierwszych wartości k. W takim przypadku tych kilka pierwszych wartości musi być sprawdzonych oddzielnie, a więc mogą one służyć jako część warunku początkowego (P). Sformułujemy drugą zasadę indukcji matematycznej w ogólniejszej wersji, która będzie miała zastosowanie w takich sytuacjach.

Druga zasada indukcji matematycznej

Niech m będzie liczbą całkowitą, niech $p(n)$ będzie ciągiem zdań zdefiniowanych na zbiorze $\{n \in \mathbb{Z}\colon\ n \geq m\}$ oraz niech l będzie nieujemną liczbą całkowitą. Jeśli

(P) wszystkie zdania $p(m), \ldots, p(m+l)$ są prawdziwe oraz
(I) dla $k > m+l$ zdanie $p(k)$ jest prawdziwe, jeśli wszystkie zdania $p(m), \ldots, p(k-1)$ są prawdziwe,

to zdanie $p(n)$ jest prawdziwe dla wszystkich $n \geq m$.

Jeśli $l = 0$, mamy poprzednią wersję drugiej zasady indukcji.

W paragrafie 4.3 widzieliśmy, że wiele ciągów jest określonych rekurencyjnie za pomocą wyrazów innych niż bezpośrednio poprzedzające dany wyraz. Druga zasada indukcji pozwala w naturalny sposób dowodzić własności takich ciągów.

PRZYKŁAD 2

(a) W zadaniu 14 w § 4.3 zdefiniowaliśmy rekurencyjnie ciąg b wzorami $b_0 = b_1 = 1$ oraz $b_n = 2b_{n-1} + b_{n-2}$ dla $n \geq 2$. W ćwiczeniu (b) prosiliśmy o wyjaśnienie, dlaczego wszystkie b_n są liczbami całkowitymi nieparzystymi. Dowód wymagał wówczas zastosowania zasady dobrego uporządkowania. Bardziej naturalne wydaje się następujące zastosowanie drugiej zasady indukcji matematycznej.

Niech n-tym zdaniem $p(n)$ będzie zdanie „b_n jest liczbą nieparzystą". W kroku indukcyjnym wykorzystamy zależność $b_k = 2b_{k-1} + b_{k-2}$, a więc będziemy musieli założyć, że $k \geq 2$. Zatem przypadki $n = 0$ i $n = 1$ sprawdzimy oddzielnie. Tak więc skorzystamy z drugiej zasady indukcji dla $m = 0$ i $l = 1$.

(P) Zdania $p(0)$ oraz $p(1)$ są oczywiście prawdziwe, ponieważ $b_0 = b_1 = 1$.

(I) Weźmy $k \geq 2$ i załóżmy, że b_n jest liczbą nieparzystą dla wszystkich n spełniających nierówności $0 \leq n < k$. W szczególności, b_{k-2} jest liczbą nieparzystą. Oczywiście $2b_{k-1}$ jest liczbą parzystą, a więc $b_k = 2b_{k-1} + b_{k-2}$ jest sumą liczby parzystej i nieparzystej. Tak więc b_k jest liczbą nieparzystą. Zatem z drugiej zasady indukcji matematycznej wynika, że wszystkie liczby b_n są nieparzyste.

Zauważmy, że w tym dowodzie nieparzystość liczby b_k wynikała z nieparzystości liczby b_{k-2}.

(b) Dla powyższego ciągu dowiedziemy, że $b_n < 6b_{n-2}$ dla $n \geq 4$. Bezpośrednie obliczenia pokazują, że $b_2 = 3$, $b_3 = 7$, $b_4 = 17$ oraz $b_5 = 41$. Dla $n = 4$ nierówność mówi, że $b_4 < 6b_2$, czyli $17 < 6 \cdot 3$, a dla $n = 5$ mówi, że $b_5 < 6b_3$, czyli $41 < 6 \cdot 7$; nierówności te są prawdziwe. Weźmy teraz $k \geq 6$ i przyjmijmy, że

$$b_n < 6b_{n-2} \quad \text{dla} \quad 4 \leq n < k.$$

Ponieważ zarówno $k-1$, jak i $k-2$ są równe co najmniej 4, mamy z założenia $b_{k-1} < 6b_{k-3}$ oraz $b_{k-2} < 6b_{k-4}$. Stąd

$$\begin{aligned} b_k &= && \text{(z definicji } b_k\text{)} \\ &= 2b_{k-1} + b_{k-2} < && \text{(z założenia indukcyjnego)} \\ &< 2(6b_{k-3}) + 6b_{k-4} = && \text{(przekształcenia algebraiczne)} \\ &= 6(2b_{k-3} + b_{k-4}) = && \text{(z definicji } b_{k-2}\text{)} \\ &= 6b_{k-2}. \end{aligned}$$

Zatem z drugiej zasady indukcji matematycznej wynika, że nierówność ta jest prawdziwa dla wszystkich $n \geq 4$.

Zauważmy, że sprawdziliśmy założenie dla $n = 4$ i $n = 5$, zanim przeszliśmy do kroku indukcyjnego. Tak więc zastosowaliśmy drugą zasadę indukcji do $m = 4$ i $l = 1$. Dlaczego, zanim przeszliśmy do kroku indukcyjnego, sprawdzaliśmy, czy prawdziwa jest nierówność zarówno dla $n = 5$, jak i dla $n = 4$? Zanim przeprowadziliśmy dowód, zaobserwowaliśmy, że w kroku indukcyjnym musieliśmy skorzystać z nierówności $b_n < 6b_{n-2}$ dla $n = k - 2$, więc musieliśmy mieć $k - 2 \geq 4$ lub $k \geq 6$. Innymi słowy, krok indukcyjny nie będzie prawdziwy dla $n = 5$: $b_5 = 2b_4 + b_3$, ale b_3 nie jest mniejsze od $6b_1$. ∎

PRZYKŁAD 3 Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = a_1 = a_2 = 1$ oraz $a_n = a_{n-2} + a_{n-3}$ dla $n \geq 3$. Kilkanaście pierwszych wyrazów

tego ciągu to: 1, 1, 2, 2, 3, 4, 5, 7, 9, 12, 16, 21, 28, 37, 49. Udowodnimy, że $a_n \leq (\frac{4}{3})^n$ dla wszystkich $n \in \mathbb{N}$. Nierówność ta jest oczywista dla $n = 0, 1$ oraz 2. Zatem weźmy $k \geq 3$ i przyjmijmy, że $a_n < (\frac{4}{3})^n$ dla $0 \leq n < k$. W szczególności, $a_{k-2} \leq (\frac{4}{3})^{k-2}$ oraz $a_{k-3} \leq (\frac{4}{3})^{k-3}$. Tak więc mamy

$$a_k = a_{k-2} + a_{k-3} \leq \left(\frac{4}{3}\right)^{k-2} + \left(\frac{4}{3}\right)^{k-3}$$
$$= \left(\frac{4}{3}\right)^{k} \cdot \left[\left(\frac{3}{4}\right)^{2} + \left(\frac{3}{4}\right)^{3}\right].$$

Ponieważ $(\frac{3}{4})^2 + (\frac{3}{4})^3 = \frac{63}{64} < 1$, wnioskujemy stąd, że $a_k \leq (\frac{4}{3})^k$. To dowodzi kroku indukcyjnego. Zatem z drugiej zasady indukcji matematycznej (dla $m = 0$ i $l = 2$) wynika, że $a_n \leq (\frac{4}{3})^n$ dla wszystkich $n \in \mathbb{N}$.

W dowodzie tym mieliśmy szczęście, że $(\frac{3}{4})^2 + (\frac{3}{4})^3 < 1$. Gdyby nie zachodziła ta nierówność, musielibyśmy znaleźć inny dowód, dowieść czegoś innego lub zrezygnować z tego zadania. Indukcja daje nam pewien schemat dowodzenia, ale nie dostarcza szczegółów, które wynikają z danego konkretnego problemu, którym się zajmujemy. ∎

Sformułowaliśmy już w § 4.2 pierwszą zasadę indukcji matematycznej dla ciągów skończonych. Obie wersje drugiej zasady mogą też być sformułowane dla ciągów skończonych. Zmiany są proste. Przypuśćmy, że zdania $p(n)$ są określone dla $m \leq n \leq m^*$. Wtedy pierwszą wersję drugiej zasady indukcji można sformułować następująco. Jeśli

(P) zdanie $p(m)$ jest prawdziwe oraz
(I) dla $m < k \leq m^*$ zdanie $p(k)$ jest prawdziwe, jeśli wszystkie zdania $p(m), \ldots, p(k-1)$ są prawdziwe,

to zdanie $p(n)$ jest prawdziwe dla wszystkich n spełniających nierówności $m \leq n \leq m^*$.

Powrócimy do nieskończonych zasad indukcji i zakończymy ten paragraf omówieniem zależności logicznych między tymi dwiema zasadami oraz wyjaśnieniem, dlaczego obie te zasady uważamy za poprawne metody dowodzenia twierdzeń.

Okazuje się, że z każdej z tych dwóch zasad wynika druga, w tym sensie, że jeśli akceptujemy jedną jako poprawną, to druga jest też poprawna. Jest rzeczą oczywistą, że z drugiej zasady indukcji wynika pierwsza zasada, ponieważ, jeśli pozwoliliśmy założyć prawdziwość wszystkich poprzednich przypadków, to na

pewno pozwoliliśmy założyć prawdziwość bezpośrednio poprzedzającego przypadku. Można podać staranny dowód, pokazując, że warunki (P) i (I) w drugiej zasadzie indukcji wynikają z warunków (P) i (I) w pierwszej zasadzie.

Jest chyba bardziej zaskakujące, że z pierwszej zasady indukcji wynika druga. Można podać dowód tego, korzystając ze zdań

$$q(n) = p(m) \wedge \ldots \wedge p(n) \quad \text{dla} \quad n \geq m$$

i pokazując, że jeśli ciąg $p(n)$ spełnia warunki (P) i (I) w drugiej zasadzie, to ciąg $q(n)$ spełnia warunki (P) i (I) w pierwszej zasadzie. Zatem każde zdanie $q(n)$ będzie prawdziwe na podstawie pierwszej zasady indukcji, a więc każde zdanie $p(n)$ będzie także prawdziwe.

Równoważność tych dwóch zasad ma dla nas mniejsze znaczenie niż to, że są one poprawne. Aby to pokazać, odwołamy się do zasady dobrego uporządkowania sformułowanej w § 4.1.

Dowód drugiej zasady indukcji. Przyjmijmy, że

(P) wszystkie zdania $p(m), \ldots, p(m+l)$ są prawdziwe oraz
(I) dla $k > m+l$, zdanie $p(k)$ jest prawdziwe, jeśli zdania $p(m), \ldots, p(k-1)$ są prawdziwe,

ale zdanie $p(n)$ jest fałszywe dla jakiejś liczby $n \geq m$. Wtedy zbiór

$$S = \{n \in \mathbb{Z}\colon\ n \geq m \ \text{ oraz zdanie } p(n) \text{ jest fałszywe}\}$$

jest niepusty. Na mocy zasady dobrego uporządkowania zbiór S ma element najmniejszy n_0. Na mocy warunku (P) $n_0 > m+l$. Ponieważ $p(n)$ jest zdaniem prawdziwym dla $m \leq n < n_0$, to $p(n_0)$ jest też zdaniem prawdziwym na podstawie warunku (I). To przeczy temu, że n_0 należy do zbioru S. Stąd wynika, że jeśli zachodzą warunki (P) i (I), to każde zdanie $p(n)$ jest prawdziwe.■

Można podać podobny dowód pierwszej zasady indukcji, ale ponieważ zasady są równoważne, nie jest to konieczne.

ĆWICZENIA DO § 4.5

Niektóre z ćwiczeń do tego paragrafu wymagają tylko zastosowania pierwszej zasady indukcji matematycznej i zostały włączone tutaj jako dodatkowy materiał do ćwiczeń.

1. Udowodnij, że $3 + 11 + \ldots + (8n-5) = 4n^2 - n$ dla $n \in \mathbb{P}$.

2. Dla $n \in \mathbb{P}$ udowodnij, że

(a) $1 \cdot 2 + 2 \cdot 3 + \ldots + n(n+1) = \frac{1}{3}n(n+1)(n+2)$,

(b) $\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \ldots + \frac{1}{n(n+1)} = \frac{n}{(n+1)}$.

3. Udowodnij, że liczba $n^5 - n$ jest podzielna przez 10 dla wszystkich $n \in \mathbb{P}$.

4. (a) Oblicz b_6 dla ciągu (b_n) z przykładu 2.
(b) Użyj definicji rekurencyjnej ciągu (a_n) z przykładu 3, aby obliczyć a_9.

5. Czy pierwsza zasada indukcji matematycznej jest odpowiednia do tego, aby udowodnić fakt z ćwiczenia 11(b) z § 4.3? Odpowiedź uzasadnij.

6. Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = 1$, $a_1 = 2$ oraz $a_n = \frac{a_{n-1}^2}{a_{n-2}}$ dla $n \geq 2$.
(a) Oblicz kilka pierwszych wyrazów tego ciągu.
(b) Korzystając z punktu (a), odgadnij ogólny wzór na a_n.
(c) Udowodnij wzór z punktu (b).

7. Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = a_1 = 1$ oraz $a_n = \frac{a_{n-1}^2 + a_{n-2}}{a_{n-1} + a_{n-2}}$ dla $n \geq 2$. Powtórz ćwiczenie 6 dla tego ciągu.

8. Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = 1$, $a_1 = 2$ oraz $a_n = \frac{a_{n-1}^2 - 1}{a_{n-2}}$ dla $n \geq 2$. Powtórz ćwiczenie 6 dla tego ciągu.

9. Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = 0$, $a_1 = 1$ oraz $a_n = \frac{1}{4}(a_{n-1} - a_{n-2} + 3)^2$ dla $n \geq 2$. Powtórz ćwiczenie 6 dla tego ciągu.

10. Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = 1$, $a_1 = 2$, $a_2 = 3$ oraz $a_n = a_{n-2} + 2a_{n-3}$ dla $n \geq 3$.
(a) Oblicz a_n dla $n = 3, 4, 5, 6, 7$.
(b) Udowodnij, że $a_n > (\frac{3}{2})^n$ dla wszystkich $n \geq 1$.

11. Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = a_1 = a_2 = 1$ oraz $a_n = a_{n-1} + a_{n-2} + a_{n-3}$ dla $n \geq 3$.
(a) Oblicz kilka pierwszych wyrazów tego ciągu.
(b) Udowodnij, że wszystkie liczby a_n są nieparzyste.
(c) Udowodnij, że $a_n \leq 2^{n-1}$ dla wszystkich $n \geq 1$.

12. Definiujemy rekurencyjnie ciąg a_n wzorami: $a_0 = 1$, $a_1 = 3$, $a_2 = 5$ oraz $a_n = 3a_{n-2} + 2a_{n-3}$ dla $n \geq 3$.
(a) Oblicz a_n dla $n = 3, 4, 5, 6, 7$.
(b) Udowodnij, że $a_n > 2^n$ dla $n \geq 1$.
(c) Udowodnij, że $a_n < 2^{n+1}$ dla $n \geq 1$.
(d) Udowodnij, że $a_n = 2a_{n-1} + (-1)^{n-1}$ dla $n \geq 1$.

13. Definiujemy rekurencyjnie ciąg b_n wzorami: $b_0 = b_1 = b_2 = 1$ oraz $b_n = b_{n-1} + b_{n-3}$ dla $n \geq 3$.
(a) Oblicz b_n dla $n = 3, 4, 5, 6$.
(b) Pokaż, że $b_n \geq 2b_{n-2}$ dla $n \geq 3$.
(c) Udowodnij, że zachodzi nierówność $b_n \geq (\sqrt{2})^{n-2}$ dla $n \geq 2$.

14. Dla ciągu z ćwiczenia 13 pokaż, że $b_n \leq (\frac{3}{2})^{n-1}$ dla $n \geq 1$.

15. Definiujemy rekurencyjnie ciąg $\text{SEQ}(n)$ za pomocą wzorów: $\text{SEQ}(0) = 0$, $\text{SEQ}(1) = 1$ oraz

$$\text{SEQ}(n) = \frac{1}{n} \cdot \text{SEQ}(n-1) + \frac{n-1}{n} \cdot \text{SEQ}(n-2)$$

dla $n \geq 2$. Udowodnij, że $0 \leq \text{SEQ}(n) \leq 1$ dla wszystkich $n \in \mathbb{N}$.

16. Tak jak w ćwiczeniu 15 z § 4.3 niech $\text{SEQ}(0) = 1$ oraz $\text{SEQ}(n) = \sum_{i=0}^{n-1} \text{SEQ}(i)$ dla $n \geq 1$. Udowodnij, że $\text{SEQ}(n) = 2^{n-1}$ dla $n \geq 1$.

17. Przypomnijmy ciąg Fibonacciego z przykładu 3 w § 4.3:

(P) $\text{FIB}(0) = \text{FIB}(1) = 1$,

(R) $\text{FIB}(n) = \text{FIB}(n-1) + \text{FIB}(n-2)$ dla $n \geq 2$.

Udowodnij, że

$$\text{FIB}(n) = 1 + \sum_{k=0}^{n-2} \text{FIB}(k) \quad \text{dla} \quad n \geq 2.$$

18. Ciąg Lucasa definiujemy w następujący sposób:

(P) $\text{LUC}(1) = 1$ i $\text{LUC}(2) = 3$,

(R) $\text{LUC}(n) = \text{LUC}(n-1) + \text{LUC}(n-2)$ dla $n \geq 3$.

(a) Wypisz osiem pierwszych wyrazów ciągu Lucasa.

(b) Udowodnij, że $\text{LUC}(n) = \text{FIB}(n) + \text{FIB}(n-2)$ dla $n \geq 2$, gdzie FIB jest ciągiem Fibonacciego zdefiniowanym w ćwiczeniu 17.

19. W przykładzie 2(a) w § 4.3 zdefiniowaliśmy ciąg T wzorami:

(P) $T(1) = 1$,

(R) $T(n) = 2 \cdot T(\lfloor n/2 \rfloor)$ dla $n \geq 2$.

Pokaż, że $T(n)$ jest największą liczbą postaci 2^k, taką, że $2^k \leq n$. (To znaczy, że $T(n) = 2^{\lfloor \log n \rfloor}$, gdzie podstawą logarytmu jest 2).

20. (a) Pokaż, że jeśli ciąg T jest określony tak jak w ćwiczeniu 19, to $T(n)$ jest $O(n)$.

(b) Pokaż, że jeśli ciąg Q jest zdefiniowany tak jak w przykładzie 2(b) w § 4.3 wzorami

(P) $Q(1) = 1$,

(R) $Q(n) = 2 \cdot Q(\lfloor n/2 \rfloor) + n$ dla $n \geq 2$,

to $Q(n)$ jest $O(n^2)$.

(c) Pokaż, że naprawdę $Q(n)$ jest $O(n \log_2 n)$ dla ciągu Q z ćwiczenia (b).

21. Pokaż, że jeśli ciąg S jest zdefiniowany tak jak w przykładzie 6 w § 4.3 wzorami

(P) $S(0) = 0$, $S(1) = 1$,

(R) $S(n) = S(\lfloor n/2 \rfloor) + S(\lfloor n/5 \rfloor)$ dla $n \geq 2$,

to $S(n)$ jest $O(n)$.

§ 4.6. Algorytm Euklidesa

W tym paragrafie omówimy algorytm, za pomocą którego oblicza się największy wspólny dzielnik. Następnie pokażemy, jak zmodyfikować ten algorytm, aby móc rozwiązywać kongruencje $x \cdot m \equiv a \pmod{n}$ z niewiadomą x.

Przypomnijmy, że liczba całkowita d jest **dzielnikiem** liczby całkowitej m wtedy, gdy m jest wielokrotnością d, tzn. wtedy, gdy $m = d \cdot k$ dla pewnej liczby $k \in \mathbb{Z}$. Mówimy też wtedy, że d dzieli m. Ponieważ $0 = d \cdot 0$, więc każda liczba jest dzielnikiem 0, ale 0 jest tylko dzielnikiem samego siebie. Dzielniki liczb $-m$ i m zawsze są takie same (na przykład dzielnikami liczb -6 i 6 są: -6, -3, -2, -1, 1, 2, 3 i 6). Ponadto d jest dzielnikiem liczby m wtedy i tylko wtedy, gdy $-d$ jest jej dzielnikiem. Tak więc zazwyczaj ograniczamy nasze rozważania do nieujemnych dzielników d nieujemnych liczb m. Jeśli $m > 0$ i $m = d \cdot k$, gdzie $d, k \in \mathbb{Z}$, to $d = m/k$, więc $|d| = m/|k| \leq m$. Zatem wszystkie dzielniki liczby m leżą w przedziale między $-m$ i m.

Wspólnym dzielnikiem liczb m i n jest liczba całkowita, która jest dzielnikiem zarówno m, jak i n. Zauważmy, że 1 i -1 są zawsze wspólnymi dzielnikami m i n. Jeśli liczby m i n nie są obie równe 0, to mogą mieć tylko skończoną liczbę wspólnych dzielników. Wtedy największy z nich nazywamy **największym wspólnym dzielnikiem** liczb m i n i oznaczamy go przez $\mathrm{NWD}(m, n)$.

PRZYKŁAD 1 (a) Dzielnikami 12 są ± 1, ± 2, ± 3, ± 4, ± 6 i ± 12. Znajdujemy je rozkładając 12 na czynniki pierwsze $2 \cdot 2 \cdot 3$ i łącząc te czynniki na wszystkie możliwe sposoby. Dzielnikami liczby $45 = 3 \cdot 3 \cdot 5$ są ± 1, ± 3, ± 5, ± 9, ± 15 i ± 45. Wspólnymi dzielnikami 12 i 45 są więc -3, -1, 1 i 3, z których 3 jest największy. Tak więc $\mathrm{NWD}(12, 45) = 3$.

(b) Wspólnymi dzielnikami $12 = 2 \cdot 2 \cdot 3$ oraz $30 = 2 \cdot 3 \cdot 5$ są -6, -3, -2, -1, 1, 2, 3 i 6, więc $\mathrm{NWD}(12, 30) = 6$.

(c) Wspólnymi dzielnikami $20 = 2 \cdot 2 \cdot 5$ oraz $150 = 2 \cdot 3 \cdot 5 \cdot 5$ są -10, -5, -2, -1, 1, 2, 5 i 10, zatem $\mathrm{NWD}(20, 150) = 10$.

(d) Ponieważ wspólnymi dzielnikami $20 = 2 \cdot 2 \cdot 5$ i $63 = 3 \cdot 3 \cdot 7$ są -1 i 1, więc $\mathrm{NWD}(20, 63) = 1$.

(e) Ponieważ wspólnymi dzielnikami liczby a i 0 są po prostu dzielniki a dla $a > 0$, więc $\mathrm{NWD}(a, 0) = a = \mathrm{NWD}(-a, 0)$. ∎

Przykłady te ilustrują ogólne zjawisko: wszystkie wspólne dzielniki m i n dzielą największy wspólny dzielnik. Moglibyśmy

teraz podać nieco skomplikowany dowód tego faktu, używając algorytmu dzielenia z § 3.6; zamiast tego otrzymamy ten fakt jako wniosek z twierdzenia 3.

Przykład 1 sugeruje jeden ze sposobów znajdowania największego wspólnego dzielnika liczb m i n: rozłóż każdą liczbę na czynniki pierwsze, znajdź najwyższą potęgę każdego czynnika pierwszego, która dzieli zarówno m, jak i n i wymnóż przez siebie wszystkie takie potęgi, aby otrzymać $\mathrm{NWD}(m,n)$. Chociaż nieźle to wygląda w teorii, w praktyce ma dwa słabe punkty. Po pierwsze, jeśli m i n są dużymi liczbami całkowitymi — obecnie w praktycznych zastosowaniach m i n mogą mieć więcej niż 100 cyfr dziesiętnych — to rozłożenie ich na czynniki pierwsze zajmuje niezwykle dużo czasu i wymaga nierealistycznych możliwości sprzętowych. Po drugie, większa część pracy jest zmarnowana na szukanie czynników pierwszych, które nie będą wykorzystane. Na przykład, widzimy w ogóle bez rachunków, że liczby m i $m+1$ nie mają wspólnych dzielników poza 1 i -1 (każdy wspólny dzielnik musi dzielić ich różnicę, czyli 1), więc $\mathrm{NWD}(m, m+1) = 1$ i czynniki pierwsze liczb m i $m+1$ są nieistotne. Przypadek $\mathrm{NWD}(m, m+1) = 1$ jest przypadkiem ważnym w zastosowaniach i może się dość często pojawić, gdy liczby m i n są wybrane mniej więcej losowo.

Na szczęście, istnieje sposób wyliczenia największego wspólnego dzielnika stosunkowo szybko bez rozkładania liczb na czynniki pierwsze, przy użyciu tylko dodawania, odejmowania oraz funkcji DIV i MOD z § 3.6. Naszym następnym celem będzie omówienie algorytmu, który to robi. Kluczem do tego algorytmu jest następujący fakt.

Stwierdzenie

Jeśli m i n są liczbami całkowitymi i $n \neq 0$, to wspólne dzielniki liczb m i n są takie same, jak wspólne dzielniki liczb n i $m \operatorname{MOD} n$. Zatem

$$\mathrm{NWD}(m,n) = \mathrm{NWD}(n, m \operatorname{MOD} n).$$

Dowód. Mamy $m = n \cdot (m \operatorname{DIV} n) + m \operatorname{MOD} n$. Ta równość zachodzi nawet wtedy, gdy $n = 1$, ponieważ $m \operatorname{DIV} 1 = m$ oraz $m \operatorname{MOD} 1 = 0$. Jeśli n i $m \operatorname{MOD} n$ są wielokrotnościami d, to również m jest wielokrotnością d, ponieważ zarówno $n \cdot (m \operatorname{DIV} n)$, jak i $m \operatorname{MOD} n$ dzielą się przez d. W drugą stronę, jeśli m i n są wielokrotnościami e, to ponieważ $m \operatorname{MOD} n = m - n \cdot (m \operatorname{DIV} n)$, więc $m \operatorname{MOD} n$ jest również wielokrotnością e.

Ponieważ pary liczb (m, n) i $(n, m \bmod n)$ mają te same zbiory wspólnych dzielników, to ich największe wspólne dzielniki muszą być takie same. ∎

PRZYKŁAD 2

(a) Mamy $\mathrm{NWD}(45, 12) = \mathrm{NWD}(12, 45 \bmod 12) = \mathrm{NWD}(12, 9) = \mathrm{NWD}(9, 12 \bmod 9) = \mathrm{NWD}(9, 3) = \mathrm{NWD}(3, 9 \bmod 3) = \mathrm{NWD}(3, 0) = 3.$

(b) Również $\mathrm{NWD}(20, 63) = \mathrm{NWD}(63, 20 \bmod 63) = \mathrm{NWD}(63, 20) = \mathrm{NWD}(20, 63 \bmod 20) = \mathrm{NWD}(20, 3) = \mathrm{NWD}(3, 20 \bmod 3) = \mathrm{NWD}(3, 2) = \mathrm{NWD}(2, 3 \bmod 2) = \mathrm{NWD}(2, 1) = \mathrm{NWD}(1, 2 \bmod 1) = \mathrm{NWD}(1, 0) = 1.$

(c) Jako trzeci przykład weźmy $\mathrm{NWD}(12, 6) = \mathrm{NWD}(6, 12 \bmod 6) = \mathrm{NWD}(6, 0) = 6.$ ∎

Przykłady te sugerują ogólną strategię postępowania przy szukaniu $\mathrm{NWD}(m, n)$: zastąp liczby m i n liczbami n i $m \bmod n$ i spróbuj jeszcze raz. Mamy nadzieję, że w ten sposób dojdziemy do przypadku, w którym już znamy odpowiedź. Oto otrzymany w ten sposób algorytm, znany już Euklidesowi ponad 2000 lat temu, a być może jest on jeszcze starszy.

Algorytm NWD

{Dane: $m, n \in \mathbb{N}$, nie równe jednocześnie 0}
{Wyniki: $d = \mathrm{NWD}(m, n)$}
{Zmienne pomocnicze: liczby całkowite a i b}
$a := m; b := n$
{Pary (a, b) i (m, n) mają ten sam największy wspólny dzielnik.}
dopóki $b \neq 0$, wykonuj
  $(a, b) := (b, a \bmod b)$
$d := a$ ∎

PRZYKŁAD 3

W tablicy 4.1 wypisane są kolejne wartości (a, b) otrzymywane podczas działania algorytmu NWD dla liczb z przykładu 2. Wyniki — największe wspólne dzielniki są wyróżnione pismem półgrubym. (Pod koniec środkowej kolumny należy pamiętać, że $2 \bmod 1 = 0$).

Twierdzenie 1

Algorytm NWD daje w wyniku największy wspólny dzielnik danych liczb całkowitych m i n.

Tablica 4.1

(a, b)	(a, b)	(a, b)
$(45, 12)$	$(20, 63)$	$(12, 6)$
$(12, 9)$	$(63, 20)$	$(\mathbf{6}, 0)$
$(9, 3)$	$(20, 3)$	
$(\mathbf{3}, 0)$	$(3, 2)$	
	$(2, 1)$	
	$(\mathbf{1}, 0)$	
$m = 45, n = 12$	$m = 20, n = 63$	$m = 12, n = 6$

Dowód. Musimy sprawdzić, że algorytm zatrzymuje się i wtedy wartość d jest największym wspólnym dzielnikiem liczb m i n.

Jeśli $n = 0$, algorytm nadaje wartości $a := m$ i $b := 0$, nie wykonuje pętli „dopóki" i zatrzymuje się, gdy $d = a = m$. Ponieważ $\mathrm{NWD}(m, 0) = m$ dla $m \in \mathbb{P}$, algorytm zatrzymuje się i daje wynik poprawny w tym przypadku. Tak więc załóżmy, że $n > 0$.

Dopóki b jest liczbą dodatnią, w pętli „dopóki" liczba b jest zastępowana przez $a \bmod b$. Ponieważ $0 \leq a \bmod b < b$, nowa wartość b jest mniejsza niż stara wartość. Malejący ciąg liczb całkowitych dodatnich musi być skończony, więc b musi w końcu stać się zerem i algorytm się zatrzyma.

Warunek „$\mathrm{NWD}(a, b) = \mathrm{NWD}(m, n)$", jak pokazuje stwierdzenie, jest niezmiennikiem pętli „dopóki". Kiedy pętla się zatrzymuje, $a > b = 0$, więc para $(a, b) = (d, 0)$. Ponieważ $d = \mathrm{NWD}(d, 0)$, mamy $d = \mathrm{NWD}(m, n)$, czego należało dowieść.■

Rozumowanie, które pokazuje, że algorytm się zatrzymuje, pokazuje również, że wykonuje on co najwyżej n przebiegów pętli „dopóki". W rzeczywistości algorytm jest jeszcze dużo szybszy.

Twierdzenie 2

Dla danych liczb całkowitych $m > n \geq 0$ algorytm NWD wykonuje co najwyżej $2 \log_2 (m + n)$ przebiegów pętli.

Dowód. Pokażemy ogólnie, że jeśli $a \geq b$, to

$$(*) \qquad b + a \bmod b < \frac{2}{3} \cdot (a + b).$$

Nierówność ta oznacza, że wartość $a + b$ zmniejsza się do co najwyżej $2/3$ swojej wartości przy każdym przebiegu pętli

dopóki $b \neq 0$, wykonuj
$(a, b) := (b, a \text{ MOD } b)$

Na początku zachodzi równość $a+b = m+n$, a po k przebiegach pętli mamy $a+b \leq (\frac{2}{3})^k \cdot (m+n)$. Zawsze zachodzi nierówność $a+b \geq 1+0$, więc $1 \leq (\frac{2}{3})^l \cdot (m+n)$ po ostatnim, l-tym przebiegu pętli. Zatem $m+n \geq (\frac{3}{2})^l$, czyli $\log_2 (m+n) \geq l \cdot \log_2 (\frac{3}{2}) > \frac{1}{2} l$, a więc $l < 2 \cdot \log_2 (m+n)$, czego należało dowieść.

Pozostaje tylko do udowodnienia nierówność $(*)$, którą zapiszemy w postaci

$$3b + 3 \cdot (a \text{ MOD } b) < 2a + 2b$$

lub w postaci $b + 3 \cdot (a \text{ MOD } b) < 2a$. Ponieważ $a = b \cdot (a \text{ DIV } b) + a \text{ MOD } b$, nierówność ta jest równoważna z nierównością

$$b + 3 \cdot (a \text{ MOD } b) < 2b \cdot (a \text{ DIV } b) + 2 \cdot (a \text{ MOD } b),$$

czyli

$$(**) \qquad b + a \text{ MOD } b < 2b \cdot (a \text{ DIV } b).$$

Ponieważ $a \geq b$, mamy $a \text{ DIV } b \geq 1$. Ponadto $a \text{ MOD } b < b$, a więc

$$b + a \text{ MOD } b < 2b \leq 2b \cdot (a \text{ DIV } b),$$

co kończy dowód nierówności $(**)$. ∎

W rzeczywistości algorytm NWD zazwyczaj zatrzymuje się po dużo mniejszej liczbie przebiegów pętli niż $2 \log_2 (m+n)$, chociaż istnieją dane m i n, które wymagają co najmniej $\log_2 (m+n)$ przebiegów (zob. ćwiczenie 17). Ponadto liczby a i b zmniejszają się w każdym przebiegu, a więc zaczynając nawet od bardzo dużych liczb, otrzymujemy szybko liczby rozsądnej wielkości. Zauważmy również, że założenie, że $m > n$ nie jest istotnym ograniczeniem; jeśli $n > m$, to algorytm NWD zamienia parę (m, n) na parę $(n, m \text{ MOD } n) = (n, m)$ w pierwszym przebiegu pętli.

Dlaczego ktoś mógłby się interesować szukaniem największych wspólnych dzielników dużych liczb? Poza czysto matematycznymi problemami, istnieją zastosowania praktyczne w szybkich implementacjach komputerowych arytmetyki „nieskończonej precyzji" oraz w systemach kryptograficznych z publicznym kluczem do bezpiecznego przekazywania danych. W paragrafie 12.8 pokażemy, jak zastosować największy wspólny dzielnik do rozwiązania zadań wykorzystujących chińskie twierdzenie o resztach, odgrywających ważną rolę w projektowaniu szybkich algorytmów dla działań arytmetycznych.

Rozwiązanie tych problemów wymaga rozwiązania kongruencji postaci

$$m \cdot x \equiv a \pmod{m}$$

z niewiadomą x i danymi a, m i n. Gdybyśmy mogli po prostu podzielić a przez m, to liczba $x = a/m$ byłaby rozwiązaniem tej kongruencji. Niestety, dla dowolnych liczb m i a, liczba a nie musi być wielokrotnością m, więc ta metoda zazwyczaj nie da się zastosować. Na szczęście, nie musi wcale zachodzić równość $m \cdot x = a$, ale tylko kongruencja $m \cdot x \equiv a \pmod{n}$, i jest inna metoda, która da się zastosować. Jest ona oparta na własnościach działań modulo n, omówionych w twierdzeniach 2 i 3 w § 3.6.

Dzielenie przez m to to samo, co mnożenie przez $1/m$. Jeśli możemy w jakiś sposób znaleźć liczbę całkowitą s taką, że $m \cdot s \equiv 1 \pmod{n}$, to liczba s będzie zachowywać się, jak $1/m$ modulo n i wtedy $m \cdot s \cdot a \equiv 1 \cdot a \equiv a \pmod{n}$. A więc $x = s \cdot a$ (co wygląda jak $(1/m) \cdot a \pmod{n}$) będzie rozwiązaniem kongruencji, niezależnie od tego, jaka jest liczba a. Znalezienie s jest kluczowe i okazuje się, że pewna modyfikacja algorytmu NWD będzie dawała liczbę s, o ile ona istnieje.

Może się zdarzyć, że nie istnieje rozwiązanie kongruencji $m \cdot x \equiv a \pmod{n}$. Nawet jeśli szczęśliwie rozwiązanie istnieje dla jakiejś szczególnej wartości a, może nie istnieć liczba s taka, że $m \cdot s \equiv 1 \pmod{n}$. Na przykład kongruencja $4 \cdot x \equiv 2 \pmod{6}$ ma rozwiązanie $x = 2$, ale kongruencja $4 \cdot x \equiv 1 \pmod{6}$ nie ma w ogóle rozwiązania, ponieważ kongruencja $4 \cdot x \equiv 1 \pmod{6}$ jest innym sposobem powiedzenia, że $4x = 1 + 6k$ dla jakiejś liczby k. Jednak liczba $4x - 6k$ jest parzysta, więc nie może być równa 1.

To ostatnie rozumowanie pokazuje bardziej ogólnie, że jeśli $\mathrm{NWD}(m, n) = d > 1$, to kongruencja $m \cdot x \equiv a \pmod{n}$ nie ma rozwiązania, chyba że a jest wielokrotnością d (ćwiczenie 15(c) pokazuje, że ten warunek konieczny jest również wystarczający). Jeśli spodziewamy się, że będziemy w stanie rozwiązać kongruencję $m \cdot x \equiv a \pmod{n}$ dla dowolnych wartości a, to musi zachodzić równość $\mathrm{NWD}(m, n) = 1$. Liczby całkowite m i n takie, że $\mathrm{NWD}(m, n) = 1$, nazywamy liczbami **względnie pierwszymi**; nie mają one wspólnych dzielników pierwszych.

Zmodyfikujemy algorytm NWD tak, by dawał w wyniku nie tylko liczbę $d = \mathrm{NWD}(m, n)$, ale również liczby całkowite s i t takie, że $d = s \cdot m + t \cdot n$. Wtedy $d \equiv s \cdot m \pmod{n}$. Jeśli $d = 1$, to liczba s jest szukaną liczbą, a jeśli $d \neq 1$, to przekonamy się, że liczby m i n nie są względnie pierwsze oraz, że nie możemy spodziewać się rozwiązania kongruencji $m \cdot x \equiv a \pmod{n}$ w całej

ogólności. Aby nowy algorytm był bardziej zrozumiały, najpierw nieco zmienimy algorytm NWD, używając funkcji DIV zamiast MOD. Oto rezultat.

Algorytm NWD$^+$

{Dane: $m, n \in \mathbb{N}$, nie równe jednocześnie 0}
{Wyniki: $d = \mathrm{NWD}(m, n)$}
{Zmienne pomocnicze: liczby całkowite a, b oraz q}
$a := m; b := n$
dopóki $b \neq 0$, wykonuj
  $q := a \,\mathrm{DIV}\, b$
  $(a, b) := (b, a - q \cdot b)$
$d := a$ ∎

PRZYKŁAD 4 W tablicy 4.2 podane są kolejne wartości a, b, q i $-q \cdot b$ dla $m = 135$ i $n = 40$. Wynik d jest wyróżniony pismem półgrubym. Oznaczyliśmy b przez a_{nast}, dla przypomnienia, że każda liczba b staje się liczbą a w następnym przebiegu pętli. W rzeczywistości, ponieważ liczby b stają się liczbami a, możemy usunąć z tablicy kolumnę b bez utraty informacji.

Tablica 4.2

a	$b = a_{\mathrm{nast}}$	q	$-q \cdot b$
135	40	3	−120
40	15	2	−30
15	10	1	−10
10	**5**	2	−10
5	0		

W ogólnym przypadku, tak jak w przykładzie 4, jeśli przypisaliśmy $a_0 = m$ i $b_0 = a_1 = n$, to algorytm NWD$^+$ tworzy ciągi $a_0, a_1, \ldots, a_l, a_{l+1} = 0, b_0, b_1, \ldots, b_l$ oraz $q_1, q_2, \ldots, q_l$. Jeśli a_{i-1} i b_{i-1} są wartościami a i b na początku i-tego przebiegu pętli oraz a_i i b_i są wartościami na końcu tego przebiegu, to $a_i = b_{i-1}$, $q_i = a_{i-1} \,\mathrm{DIV}\, a_i$ oraz $a_{i+1} = b_i = a_{i-1} - q_i \cdot b_{i-1} = a_{i-1} - q_i \cdot a_i$. To znaczy, że skończone ciągi (a_i) i (q_i) spełniają równości

$$q_i = a_{i-1} \,\mathrm{DIV}\, a_i \quad \text{oraz} \quad a_{i+1} = a_{i-1} - q_i \cdot a_i$$

dla $i = 1, \ldots, l$. Ponadto $a_l = \mathrm{NWD}(m, n)$. Jesteśmy teraz gotowi do obliczenia s i t takich, że $\mathrm{NWD}(m, n) = s \cdot m + t \cdot n$. Zbudujemy ciągi $s_0, s_1, \ldots, s_l$ oraz $t_0, t_1, \ldots, t_l$ takie, że

$$a_i = s_i \cdot m + t_i \cdot n \quad \text{dla} \quad i = 0, 1, \ldots, l.$$

Przyjmując $i = l$, otrzymamy $\mathrm{NWD}(m,n) = a_l = s_l \cdot m + t_l \cdot n$, czyli to, czego potrzebowaliśmy.

Na początku chcemy, by $m = a_0 = s_0 \cdot m + t_0 \cdot n$; przyjmijmy więc $s_0 = 1$ i $t_0 = 0$. Następnie chcemy, by $n = a_1 = s_1 \cdot m + t_1 \cdot n$; wystarczy przyjąć $s_1 = 0$, $t_1 = 1$. Od tego momentu będziemy wykorzystywać zależność rekurencyjną $a_{i+1} = a_{i-1} - q_i \cdot a_i$. Jeśli mamy już

$$a_{i-1} = s_{i-1} \cdot m + t_{i-1} \cdot n \quad \text{oraz} \quad a_i = s_i \cdot m + t_i \cdot n,$$

to

$$\begin{aligned} a_{i+1} &= a_{i-1} - q_i \cdot a_i \\ &= (s_{i-1} - q_i \cdot s_i) \cdot m + (t_{i-1} - q_i \cdot t_i) \cdot n. \end{aligned}$$

Jeśli zatem weźmiemy

$$s_{i+1} = s_{i-1} - q_i \cdot s_i \quad \text{oraz} \quad t_{i+1} = t_{i-1} - q_i \cdot t_i,$$

to otrzymamy oczekiwaną równość $a_{i+1} = s_{i+1} \cdot m + t_{i+1} \cdot n$.

PRZYKŁAD 5 W tablicy 4.3 pokazano kolejne wartości a_i, q_i, s_i oraz t_i dla $m = 135$ i $n = 40$, tak jak w przykładzie 4. Oczywiście zachodzi żądana równość $5 = 3 \cdot 135 + (-10) \cdot 40$.

Tablica 4.3

i	a_i	q_i	s_i	t_i
0	135		1	0
1	40	3	0	1
2	15	2	1	−3
3	10	1	−2	7
4	**5**	2	**3**	**−10**
5	0			

■

Aby przerobić rekurencyjną definicję ciągów (s_i) i (t_i) na postać odpowiednią dla pętli „dopóki", wprowadzimy nowe zmienne s' i t', odpowiadające s_{i+1} i t_{i+1}. Aby zachować jednolitość oznaczeń, wprowadzimy również oznaczenia $a' = b$. A oto i sam algorytm.

Algorytm Euklidesa

{Dane: $m, n \in \mathbb{N}$, nie równe jednocześnie 0}
{Wyniki: $d = \mathrm{NWD}(m,n)$, liczby s i t takie, że $d = s \cdot m + t \cdot n$}
{Zmienne pomocnicze: liczby całkowite q, a, a' s, s', t, t'}
$a := m; a' := n; s := 1; s' := 0; t := 0; t' := 1$
$\{a = s \cdot m + t \cdot n$ oraz $a' = s' \cdot m + t' \cdot n\}$

dopóki $a' \neq 0$, wykonuj
  $q := a \operatorname{DIV} a'$
  $(a, a') := (a', a - q \cdot a')$
  $(s, s') := (s', s - q \cdot s')$
  $(t, t') := (t', t - q \cdot t')$
$d := a$ ∎

W tej pętli po prostu dodaliśmy dwie nowe linie, nie naruszając przy tym wartości a, zatem ten algorytm wykonuje tyle samo przebiegów pętli, co algorytm NWD. Powyższe rozumowanie pokazuje, że równości $a = s \cdot m + t \cdot n$ oraz $a' = s' \cdot m + t' \cdot n$ są niezmiennikami pętli, tak więc równość $d = a = s \cdot m + t \cdot n$ będzie spełniona po zakończeniu działania algorytmu, które nastąpi po wykonaniu co najwyżej $\log_2 (m + n)$ iteracji. W ten sposób udowodniliśmy następujące twierdzenie.

Twierdzenie 3

Dla danych liczb całkowitych $m > n \geq 0$ algorytm Euklidesa daje w wyniku liczbę $d = \mathrm{NWD}(m, n)$ i liczby całkowite s i t takie, że $d = s \cdot m + t \cdot n$, wykonując przy tym $O(\log_2 m)$ działań arytmetycznych postaci $-$, $\cdot$ oraz DIV.

Wniosek

Największy wspólny dzielnik $\mathrm{NWD}(m, n)$ liczb m i n jest wielokrotnością każdego wspólnego dzielnika m i n.

Dowód. Jeśli c jest wspólnym dzielnikiem liczb m i n, to $m = k \cdot c$ i $n = l \cdot n$ dla pewnych liczb k i l. Zatem $\mathrm{NWD}(m, n) = s \cdot m + t \cdot n = (s \cdot k + t \cdot l) \cdot c$, a więc $\mathrm{NWD}(m, n)$ jest wielokrotnością c. ∎

Rzeczywisty czas działania algorytmu Euklidesa będzie zależał od tego, jak szybko wykonuje się działania arytmetyczne. Dla dużych liczb m i n wąskim gardłem jest działanie DIV. W naszych zastosowaniach do rozwiązywania kongruencji $m \cdot s \equiv 1 \pmod{n}$ potrzebujemy tylko s i nie potrzebujemy t. Możemy zatem pominąć linię algorytmu, w której obliczane są wartości t, przez co zaoszczędzimy trochę czasu. Ponadto, ponieważ w tym przypadku potrzebujemy tylko $s \operatorname{MOD} n$, możemy w obliczeniach zastąpić wyrażenie $s - q \cdot s'$ wyrażeniem $(s - q \cdot s') \operatorname{MOD} n$, dzięki czemu wszystkie liczby s i s' są mniejsze od n.

W paragrafie 3.6 wypowiedzieliśmy uwagę, że chociaż działania $+_p$ i $*_p$ zachowują się w zbiorze $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ podobnie do działań $+$ i $\cdot$ w zbiorze $\mathbb{Z}$, musimy ostrożnie stosować prawo

skracania w $\mathbb{Z}_p$. W tym paragrafie zobaczyliśmy, że jeśli liczby m i p są względnie pierwsze, to możemy znaleźć liczbę s w $\mathbb{Z}$ taką, że $m \cdot s \equiv 1 \pmod{p}$. Jeśli $m \in \mathbb{Z}_p$, to $m *_p (s \text{ MOD } p) = 1$ w $\mathbb{Z}_p$, a więc $s \text{ MOD } p$ zachowuje się tak, jak $1/m$ w $\mathbb{Z}_p$, a $a *_p (s \text{ MOD } p)$ jak a/m w $\mathbb{Z}_p$. Zatem dzielenie przez m jest możliwe w $\mathbb{Z}_p$, jeśli $\text{NWD}(m, p) = 1$. Jeśli liczba p jest pierwsza, to $\text{NWD}(m, p) = 1$ dla wszystkich niezerowych m w $\mathbb{Z}_p$, a więc dopuszczalne jest dzielenie przez wszystkie niezerowe elementy $\mathbb{Z}_p$.

Skracanie czynnika po obu stronach równości jest tym samym, co dzielenie obu stron przez ten skracany czynnik. Jeśli zatem $\text{NWD}(m, p) = 1$ i $a *_p m = b *_p m$ w $\mathbb{Z}_p$, to $a = b$, ponieważ $a = a *_p m *_p (s \text{ MOD } p) = b *_p m *_p (s \text{ MOD } p) = b$, gdzie $(s \text{ MOD } p)$ zachowuje się tak, jak $1/m$ w zbiorze $\mathbb{Z}_p$. W przypadku kongruencji oznacza to, że jeśli m i p są względnie pierwsze oraz jeśli $a \cdot m \equiv b \cdot m \pmod{p}$, to $a \equiv b \pmod{p}$.

PRZYKŁAD 6 (a) Ponieważ 5 jest liczbą pierwszą, więc każda kongruencja $m \cdot x \equiv 1 \pmod{5}$ dla $m = 1, 2, 3, 4$ ma rozwiązanie. Te kongruencje są równoważne z równaniami $m *_5 x = 1$ w $\mathbb{Z}_5$, rozwiązanymi w ćwiczeniu 11 do § 3.6 za pomocą tablicy mnożenia. Jeśli mamy do czynienia z tak małymi liczbami, algorytm Euklidesa nie jest konieczny. Rozwiązaniami są: $1 \cdot 1 \equiv 1 \pmod{5}$, $2 \cdot 3 \equiv 1 \pmod{5}$, $3 \cdot 2 \equiv 1 \pmod{5}$ oraz $4 \cdot 4 \equiv 1 \pmod{5}$.

(b) Rozwiążmy kongruencję $10x \equiv 1 \pmod{37}$. Ponieważ 37 jest liczbą pierwszą, wiemy, że rozwiązanie istnieje, ale tworzenie tablicy mnożenia byłoby niepraktyczne. Zastosowanie algorytmu Euklidesa dla $m = 10$ i $n = 37$ daje nam $s = -11$ i $t = 3$, a więc $(-11) \cdot 10 + 3 \cdot 37 = 1$. Zatem $10 \cdot (-11) \equiv 1 \pmod{37}$, więc rozwiązaniem kongruencji $10x \equiv 1 \pmod{37}$ jest $x = -11$. Każda liczba y taka, że $y \equiv -11 \pmod{37}$ też jest rozwiązaniem, a więc $-11 \text{ MOD } 37 = -11 + 37 = 26$ jest rozwiązaniem w $\mathbb{Z}_{37} = \{0, 1, \ldots, 36\}$. Zauważ, że wartość t była tu zupełnie nieistotna.

(c) Weźmy kongruencję $m \cdot x \equiv 1 \pmod{15}$. Liczbę m traktujemy jako ustaloną i szukamy rozwiązania x w $\mathbb{Z}_{15}$. Jeśli liczby m i 15 nie są względnie pierwsze, to nie istnieje rozwiązanie. W szczególności, nie istnieje rozwiązanie, jeśli liczba m jest wielokrotnością 3 lub 5.

Rozwiążmy teraz kongruencję $13x \equiv 1 \pmod{15}$, która ma rozwiązanie, gdyż $\text{NWD}(13, 15) = 1$. Stosujemy algorytm Euklidesa dla $m = 13$ i $n = 15$ i otrzymujemy $s = 7$ i $t = -6$, a więc $7 \cdot 13 + (-6) \cdot 15 = 1$. (Znów nie interesuje nas wartość t). Zatem $13 \cdot 7 \equiv 1 \pmod{15}$, czyli otrzymanym rozwiązaniem kongruencji $13x \equiv 1 \pmod{15}$ jest $x = 7$.

(d) Kongruencja $8x \equiv 6 \pmod{15}$ jest równoważna kongruencji $4x \equiv 3 \pmod{15}$; możemy skrócić czynnik 2 po obu stronach, gdyż liczba 2 jest względnie pierwsza z 15. Wątpiący mogą zastosować algorytm Euklidesa (lub przeszukiwanie), by stwierdzić, że $2 \cdot 8 \equiv 1 \pmod{15}$, tak więc jeśli $8x \equiv 6 \pmod{15}$, to $4x \equiv (8 \cdot 2) \cdot 4x \equiv 8 \cdot 8x \equiv 8 \cdot 6 \equiv (8 \cdot 2) \cdot 3 \equiv 3 \pmod{15}$. Mnożenie przez 8 jest tym samym, co dzielenie przez 2 modulo 15.

Aby rozwiązać kongruencję $4x \equiv 3 \pmod{15}$, rozwiązujemy najpierw kongruencję $4y \equiv 1 \pmod{15}$ za pomocą algorytmu Euklidesa lub za pomocą przeszukiwania. Rozwiązaniami są liczby y takie, że $y \equiv 4 \pmod{15}$, a zatem $x \equiv 3y \equiv 12 \pmod{15}$ jest rozwiązaniem kongruencji $4x \equiv 3 \pmod{15}$, czyli również kongruencji $8x \equiv 6 \pmod{15}$. ∎

Teraz, gdy już umiemy rozwiązywać kongruencje postaci $m \cdot x \equiv a \pmod{n}$, możemy rozwiązać układ kongruencji

$$\begin{cases} x \equiv a_1 \pmod{n_1}, \\ x \equiv a_2 \pmod{n_2}. \end{cases}$$

Rozwiązaniem ogólnym kongruencji $x \equiv a_1 \pmod{n_1}$ jest $x = a_1 + n_1 \cdot y$, gdzie y jest pewną liczbą całkowitą, a więc chcemy rozwiązać kongruencję

$$a_1 + n_1 \cdot y \equiv a_2 \pmod{n_2},$$

czyli

$$n_1 \cdot y \equiv a_2 - a_1 \pmod{n_2}.$$

Jeśli $\mathrm{NWD}(n_1, n_2) = 1$, to możemy skorzystać z algorytmu Euklidesa, aby otrzymać s takie, że $n_1 \cdot s \equiv 1 \pmod{n_2}$. Wtedy $y = s \cdot (a_2 - a_1)$ spełnia kongruencję $n_1 \cdot y \equiv a_2 - a_1 \pmod{n_2}$, więc liczba $x = a_1 + n_1 \cdot s \cdot (a_2 - a_1)$ spełnia obie kongruencje danego układu. To podejście jest kluczowe w rozwiązywaniu zadań związanych z chińskim twierdzeniem o resztach w § 12.8.

PRZYKŁAD 7 Zilustrujemy tę metodę, rozwiązując układ kongruencji

$$\begin{cases} x \equiv 1 \pmod{13}, \\ x \equiv 4 \pmod{15}. \end{cases}$$

Dla dowolnej liczby całkowitej y, liczba $1 + 13y$ jest rozwiązaniem kongruencji $x \equiv 1 \pmod{13}$, a więc rozwiązujemy kongruencję $1 + 13y \equiv 4 \pmod{15}$, czyli kongruencję $13y \equiv 3 \pmod{15}$. Najpierw musimy rozwiązać kongruencję $13s \equiv 1 \pmod{15}$, ale zrobiliśmy to już w przykładzie 6(c) i stwierdziliśmy tam, że $13 \cdot 7 \equiv 1 \pmod{15}$. Wynika stąd, że $13 \cdot 7 \cdot 3 \equiv 3 \pmod{15}$.

Zatem $y = 21$ jest rozwiązaniem kongruencji $13y \equiv 3 \pmod{15}$. Ostatecznie, $x = 1 + 13y = 1 + 13 \cdot 21 = 274$ spełnia kongruencje $x \equiv 1 \pmod{13}$ oraz $x \equiv 4 \pmod{15}$.

Każda liczba przystająca do 274 modulo $13 \cdot 15 = 195$ będzie również spełniać te dwie kongruencje, a więc możemy znaleźć rozwiązanie x spełniające warunki $0 \leq x \leq 194$. Ma ono postać $274 - 195 = 79$. ∎

ĆWICZENIA DO § 4.6

Zauważ, że wiele z tych odpowiedzi łatwo sprawdzić, kiedy już je znamy.

1. Skorzystaj z jakiejkolwiek metody, aby znaleźć NWD(m, n) dla następujących par:
(a) $m = 20$, $n = 20$; (b) $m = 20$, $n = 10$;
(c) $m = 20$, $n = 1$; (d) $m = 20$, $n = 0$;
(e) $m = 20$, $n = 7$; (f) $m = 20$, $n = -20$;
(g) $m = 120$, $n = 162$; (h) $m = 20$, $n = 27$.

2. Powtórz ćwiczenie 1 dla następujących par:
(a) $m = 17$, $n = 34$; (b) $m = 17$, $n = 72$;
(c) $m = 17$, $n = 850$; (d) $m = 170$, $n = 850$;
(e) $m = 289$, $n = 850$; (f) $m = 2890$, $n = 850$.

3. Wypisz pary liczb (a, b) otrzymane w trakcie działania algorytmu NWD dla następujących liczb m i n oraz znajdź NWD(m, n):
(a) $m = 20$, $n = 14$; (b) $m = 20$, $n = 7$;
(c) $m = 20$, $n = 30$; (d) $m = 2000$, $n = 987$.

4. Powtórz ćwiczenie 3 dla następujących liczb:
(a) $m = 30$, $n = 30$; (b) $m = 30$, $n = 10$;
(c) $m = 30$, $n = 60$; (d) $m = 3000$, $n = 999$.

5. Skorzystaj z algorytmu Euklidesa, aby znaleźć NWD(m, n) oraz liczby s i t takie, że NWD$(m, n) = s \cdot m + t \cdot n$ dla podanych liczb m i n. *Wskazówka*: zrób tablicę taką, jak tabl. 4.3.
(a) $m = 20$, $n = 14$; (b) $m = 72$, $n = 17$;
(c) $m = 20$, $n = 30$; (d) $m = 320$, $n = 30$.

6. Powtórz ćwiczenie 5 dla następujących liczb:
(a) $m = 14259$, $n = 3521$; (b) $m = 8359$, $n = 9373$.

7. Dla każdej wartości m rozwiąż kongruencję $m \cdot x \equiv 1 \pmod{26}$, gdzie $0 \leq x < 26$ lub wyjaśnij, dlaczego nie istnieje rozwiązanie:
(a) $m = 5$, (b) $m = 11$,
(c) $m = 4$, (d) $m = 9$,
(e) $m = 17$, (f) $m = 13$.

8. Powtórz ćwicz. 7 dla kongruencji $m \cdot x \equiv 1 \pmod{24}$, gdzie $0 \leq x < 24$.

9. Rozwiąż kongruencje z niewiadomą x:

(a) $8x \equiv 1 \pmod{13}$, (b) $8x \equiv 4 \pmod{13}$,
(c) $99x \equiv 1 \pmod{13}$, (d) $99x \equiv 5 \pmod{13}$.

10. Rozwiąż kongruencje z niewiadomą x:

(a) $2000x \equiv 1 \pmod{643}$, (b) $643x \equiv 1 \pmod{2000}$,
(c) $1647x \equiv 1 \pmod{788}$, (d) $788x \equiv 24 \pmod{1647}$.

11. Rozwiąż jakąkolwiek metodą układy kongruencji z niewiadomą x, gdzie $0 \leq x < 13 \cdot 99$. *Wskazówka*: prawo skracania pozwoli ci zaoszczędzić trochę pracy w ćwiczeniach (a) i (b).

(a) $\begin{cases} x \equiv 8 \pmod{13}, \\ x \equiv 0 \pmod{99}; \end{cases}$

(b) $\begin{cases} x \equiv 0 \pmod{13}, \\ x \equiv 65 \pmod{99}; \end{cases}$

(c) $\begin{cases} x \equiv 8 \pmod{13}, \\ x \equiv 65 \pmod{99}. \end{cases}$

12. Rozwiąż jakąkolwiek metodą układy kongruencji z niewiadomą x, gdzie $0 \leq x < 1300$.

(a) $\begin{cases} x \equiv 1 \pmod{13}, \\ x \equiv 1 \pmod{99}; \end{cases}$

(b) $\begin{cases} x \equiv -1 \pmod{13}, \\ x \equiv -1 \pmod{99}; \end{cases}$

(c) $\begin{cases} x \equiv -2 \pmod{13}, \\ x \equiv -2 \pmod{99}; \end{cases}$

(d) $\begin{cases} x \equiv 10 \pmod{13}, \\ x \equiv 96 \pmod{99}. \end{cases}$

13. Pokaż, że równości $a = s \cdot m + t \cdot n$ i $a' = s' \cdot m + t' \cdot n$ są niezmiennikami pętli w algorytmie Euklidesa, niezależnie od tego, w jaki sposób jest zdefiniowana liczba q w pętli (więc zrobienie błędu w obliczeniach lub odgadnięciu q nie zaszkodzi na trwałe).

14. Weźmy liczby całkowite m i n nie równe jednocześnie 0. Pokaż, że $\mathrm{NWD}(m, n)$ jest najmniejszą liczbą całkowitą dodatnią, którą można zapisać jako $a \cdot m + b \cdot n$ dla pewnych liczb całkowitych a i b.

15. Przypuśćmy, że $d = \mathrm{NWD}(m, n) = s \cdot m + t \cdot n$ dla pewnych liczb całkowitych s i t.

(a) Pokaż, że liczby m/d i n/d są względnie pierwsze.
(b) Pokaż, że jeśli $d = s' \cdot m + t' \cdot n$ dla $s', t' \in \mathbb{Z}$, to $s' = s + k \cdot n/d$ dla pewnej liczby $k \in \mathbb{Z}$.
(c) Pokaż, że, jeśli a jest wielokrotnością d, to kongruencja $mx \equiv a \pmod{n}$ ma rozwiązanie. *Wskazówka*: $ms \equiv d \pmod{n}$.

16. Przypuśćmy, że $d = \text{NWD}(m,n) = s \cdot m + t \cdot n$ dla pewnych $s, t \in \mathbb{Z}$.
(a) Pokaż, że jeśli $s' = s + k \cdot n/d$ oraz $t' = t - k \cdot m/d$, to $d = s' \cdot m + t' \cdot n$.
(b) Pokaż, że istnieją liczby $s', t' \in \mathbb{Z}$ takie, że $d = s' \cdot m + t' \cdot n$ oraz $0 \leq s' < n/d$.

17. (a) Pokaż, że dla danych $m = \text{FIB}(l+1)$ i $n = \text{FIB}(l)$, gdzie $l \geq 1$, algorytm NWD robi dokładnie l przebiegów pętli. Ciąg Fibonacciego został zdefiniowany w przykładzie 3 w § 4.3. *Wskazówka*: zastosuj indukcję względem l.
(b) Pokaż, że $k \geq \log_2 \text{FIB}(k+2)$ dla $k \geq 3$.
(c) Pokaż, że jeśli $l \geq 3$, $m = \text{FIB}(l+1)$ i $n = \text{FIB}(l)$, to algorytm NWD robi co najmniej $\log_2 (m+n)$ przebiegów pętli.

To, co jest najważniejsze w tym rozdziale

Jak zwykle:

1. Przekonaj się, że potrafisz zdefiniować i użyć każdego pojęcia i każdej metody.

2. Podaj przynajmniej jeden powód, dla którego ten temat został omówiony w tym rozdziale.

3. Zastanów się nad co najmniej jednym przykładem ilustrującym dane pojęcie oraz co najmniej jedną sytuacją, w której dany fakt czy metoda są przydatne.

Pojęcia i oznaczenia

pętla „dopóki"
  warunek dozoru pętli, treść, przebieg = iteracja, zakończenie = wyjście
  niezmiennik pętli
pętla „dla", pętla „dla ... w dół"
indukcja matematyczna
  warunek początkowy, krok indukcyjny
rekurencyjne definiowanie ciągu
  warunek początkowy, wzór rekurencyjny
  obliczenia metodą iteracyjną, metodą rekurencyjną
  ciąg Fibonacciego
równanie charakterystyczne
metoda rekurencyjna „dziel i rządź"
dzielnik, wspólny dzielnik, największy wspólny dzielnik (NWD)
liczby całkowite względnie pierwsze

Fakty i zasady

Twierdzenie o niezmiennikach pętli.

Zasada dobrego uporządkowania zbioru $\mathbb{N}$.

Pierwsza i druga zasada indukcji matematycznej (skończonej) (zasady te są logicznie równoważne).

Twierdzenie 2 z § 4.4 na temat wzorów rekurencyjnych postaci $s_{2n} = 2s_n + f(n)$.

Jeśli p jest liczbą pierwszą, dzielenie przez elementy niezerowe jest dopuszczalne w $\mathbb{Z}_p$.

Metody

Użycie niezmienników pętli do tworzenia algorytmów i sprawdzania ich poprawności.

Rozwiązywanie równań rekurencyjnych $s_n = as_{n-1} + bs_{n-2}$ za pomocą równań charakterystycznych.

Algorytm Euklidesa do wyznaczania $\mathrm{NWD}(m, n)$ i liczb całkowitych s, t takich, że $\mathrm{NWD}(m, n) = s \cdot m + t \cdot n$.

Zastosowanie algorytmu Euklidesa do rozwiązywania kongruencji

$$m \cdot x \equiv a \pmod{n}$$

oraz układów kongruencji

$$\begin{cases} x \equiv a_1 \pmod{n_1}, \\ x \equiv a_2 \pmod{n_2}. \end{cases}$$

5. ZLICZANIE

Głównym celem tego rozdziału jest pokazanie metod zliczania elementów dużych zbiorów skończonych bez faktycznego wypisywania tych elementów. Uzyskaną w ten sposób teorię można zastosować do obliczania prawdopodobieństw w skończonych przestrzeniach zdarzeń elementarnych, które wprowadzimy w § 5.2. Rozdział kończymy omówieniem zasady szufladkowej Dirichleta, która jest jedną z technik dowodzenia twierdzeń dotyczących zbiorów skończonych.

§ 5.1. Podstawowe techniki zliczania

Zaczniemy od pewnych praw zliczania, które są już prawdopodobnie Czytelnikowi lepiej lub gorzej znane. Tak jak w § 1.2 piszemy $|S|$ dla oznaczenia liczby elementów skończonego zbioru S.

Prawa sumy

Niech S i T będą zbiorami skończonymi.

(a) Jeśli zbiory S i T są rozłączne, tzn. $S \cap T = \emptyset$, to $|S \cup T| = |S| \cup |T|$.

(b) Ogólnie, $|S \cup T| = |S| + |T| - |S \cap T|$.

Intuicyjny powód, dla którego zachodzi (b), polega na tym, iż obliczając $|S \cup T|$ jako $|S| + |T|$, zliczamy dwukrotnie elementy zbioru $|S \cap T|$. Aby otrzymać $|S \cup T|$, musimy więc liczbę $|S \cap T|$ odjąć od sumy $|S| + |T|$. Zdanie (b) wynika z (a) w następujący sposób. Stosując (a) dwukrotnie otrzymujemy

$$|S \cup T| = |S| + |T \backslash S| \quad \text{oraz} \quad |T| = |T \backslash S| + |S \cap T|.$$

Zatem

$$|S \cup T| + |S \cap T| = |S| + |T \backslash S| + |S \cap T| = |S| + |T|,$$

co implikuje (b).

Ogólne prawo wyznaczania liczby elementów sumy więcej niż dwóch zbiorów, zwane zasadą włączeń i wyłączeń, przedstawimy w § 5.3.

PRZYKŁAD 1 Ile liczb naturalnych ze zbioru $S = \{1, 2, 3, \ldots, 1000\}$ dzieli się przez 3 lub 5, lub przez obie te liczby?

Niech

$$D_3 = \{n \in S\colon n \text{ dzieli się przez } 3\},$$

$$D_5 = \{n \in S\colon n \text{ dzieli się przez } 5\}.$$

Szukamy liczby elementów w zbiorze $D_3 \cup D_5$ i nie jest wcale oczywiste, ile ona wynosi. Ale łatwo zobaczyć, że $|D_3|$ wynosi $\lfloor 1000/3 \rfloor = 1000 \operatorname{DIV} 3 = 333$. Kto w to wątpi, niech zauważy, że

$$D_3 = \{3m\colon 1 \leq m \leq 333\}.$$

Podobnie $|D_5| = 200$. Ponieważ $D_3 \cap D_5 = \{n \in S\colon n \text{ dzieli się przez } 15\}$ i $1000/15$ równa się $66\frac{2}{3}$, więc $|D_3 \cap D_5|$ równa się 66. Na mocy prawa sumy (b)

$$|D_3 \cup D_5| = |D_3| + |D_5| - |D_3 \cap D_5| = 333 + 200 - 66 = 467.$$

■

Dla skończonych zbiorów S i T mamy $|S \times T| = |S| \cdot |T|$, ponieważ

$$S \times T = \{(s, t)\colon s \in S \text{ i } t \in T\}$$

a jeśli chcemy utworzyć uporządkowaną parę (s, t), to dla każdej spośród $|S|$ możliwości wyboru s mamy $|T|$ możliwości wyboru t. Równość $|S \times T| = |S| \cdot |T|$ ilustrują tablice 1.2 i 1.3 z § 1.2. Podobna równość zachodzi dla iloczynu więcej niż dwóch zbiorów.

Prawa iloczynu

(a) Dla skończonych zbiorów $S_1, S_2, \ldots, S_k$ mamy

$$|S_1 \times S_2 \times \ldots \times S_k| = \prod_{j=1}^{k} |S_j|.$$

(b) Ogólniej, załóżmy, że mamy zbiór ciągów $(s_1, s_2, \ldots, s_k)$ długości k o następującej strukturze. Jest n_1 możliwości wyboru s_1. Dla każdego ustalonego s_1 są n_2 możliwości wyboru s_2, dla ustalonych s_1 i s_2 są n_3 możliwości wyboru s_3, i ogólnie, mając $s_1, s_2, \ldots, s_{j-1}$ możemy wybrać s_j na n_j sposobów. Wówczas nasz zbiór ma $n_1 n_2 \cdot \ldots \cdot n_k$ elementów.

W praktyce będziemy często używać prawa iloczynu w wersji (b), prawie zawsze jednak unikając nieprzyjemnego formalizmu sugerowanego w jego sformułowaniu.

PRZYKŁAD 2 (a) Znajdźmy liczbę sposobów wylosowania ze zwracaniem pięciu kart z talii zawierającej 52 karty. Zliczamy zatem ciągi długości 5, złożone z kart należących do danej talii. **Ze zwracaniem** znaczy, że każda karta wraca do talii przed wyciągnięciem następnej. Zbiór sposobów wyboru pięciu kart ze zwracaniem jest w odpowiedniości wzajemnie jednoznacznej ze zbiorem $D \times D \times D \times D \times D = D^5$, gdzie D oznacza pięćdziesięciodwuelementowy zbiór kart. Zatem zgodnie z prawem iloczynu (a), nasz zbiór ma 52^5 elementów.

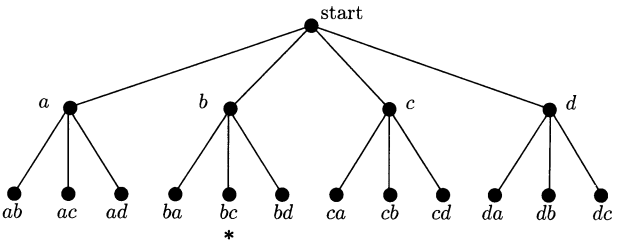
Problem ten można również rozwiązać posługując sie prawem iloczynu (b). Są 52 możliwości wyboru pierwszej karty. Po wybraniu kilku kart i zwróceniu ich do talii nadal mamy 52 możliwości wyboru kolejnej karty. Są więc $52 \cdot 52 \cdot 52 \cdot 52 \cdot 52$ sposoby wyboru pięciu kart ze zwracaniem.

(b) Znajdźmy teraz liczbę możliwych sposobów wylosowania bez zwracania pięciu kart z talii zawierającej 52 karty. **Bez zwracania** znaczy, że raz wyciągnięta karta nie wraca już do talii. W tym przypadku nie można zastosować prawa iloczynu (a), gdyż nie wszystkie ciągi długości 5 należące do zbioru D^5 są dopuszczalnymi wynikami losowania. Dokładniej, zakazane są te ciągi, w których występują powtórzenia kart. Można jednak zastosować prawo iloczynu (b). Pierwszą kartę można wybrać na 52 sposoby. Gdy już zostanie wybrana, jest 51 sposobów wyboru drugiej karty. Trzecią można wybrać na 50 sposobów, czwartą na 49 sposobów, a piątą na 48 sposobów. Tak więc pięć kart można wybrać bez zwracania na $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48$ sposobów.

Dotychczas zliczaliśmy jedynie pięciowyrazowe ciągi kart, a nie podzbiory talii złożone z 5 kart. Do zagadnienia zliczania podzbiorów powrócimy w przykładzie 10. ∎

PRZYKŁAD 3 (a) Niech $\Sigma = \{a, b, c, d, e, f, g\}$. Liczba słów w Σ^* mających długość 5 wynosi $7^5 = 16807$, tzn. $|\Sigma|^5 = 16807$. Można to wykazać stosując którekolwiek z praw iloczynu tak, jak w przykładzie 2(a). Liczba słów w Σ^5, które nie zawierają powtarzających się liter, wynosi $7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 = 2520$. Jest tak dlatego, że pierwsza litera może zostać wybrana na 7 sposobów, następnie druga na 6 sposobów itd.

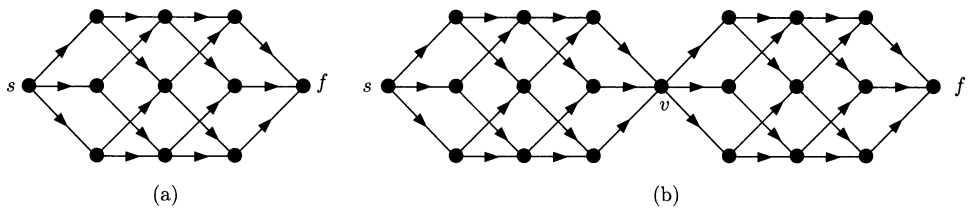
(b) Niech $\Sigma = \{a, b, c, d\}$. Liczba słów w Σ^2, bez powtarzających się liter, wynosi na mocy prawa iloczynu (b) $4 \cdot 3 = 12$.



Rysunek 5.1

Możemy to zilustrować za pomocą rysunku nazywanego drzewem; zob. rysunek 5.1. Każda droga zaczynająca się w punkcie start odpowiada słowu w Σ^2 bez powtórzeń. Na przykład droga kończąca się gwiazdką odpowiada słowu bc. Można sobie wyobrazić podobne, choć bardzo duże drzewo ilustrujące obliczenia z punktu (a). W gruncie rzeczy, takie drzewo można sobie wyobrazić dla każdej sytuacji, w której znajduje zastosowanie prawo iloczynu (b). ∎

PRZYKŁAD 4 (a) Na rysunku 5.2(a) liczba dróg biegnących od s do f wynosi $3 \cdot 2 \cdot 2 \cdot 1 = 12$, są bowiem 3 możliwości wyboru pierwszej krawędzi, a następnie po 2 dla drugiej i trzeciej. Wybór pierwszych trzech krawędzi wyznacza czwartą w sposób jednoznaczny.



Rysunek 5.2

Podobnego rachunku można użyć do obliczenia liczby dróg biegnących od s do f na rysunku 5.2(b). Alternatywne rozwiązanie polega na spostrzeżeniu, iż z powyższego wynika, że jest 12 dróg z s do v i 12 dróg z v do f, są więc $12 \cdot 12 = 144$ drogi z s do f. ∎

PRZYKŁAD 5 (a) Niech S i T będą zbiorami skończonymi. Obliczymy liczbę funkcji $f\colon S \longrightarrow T$. Będzie wygodnie przyjąć, że

$$S = \{s_1, s_2, \ldots, s_m\} \quad \text{oraz} \quad T = \{t_1, t_2, \ldots, t_n\},$$

gdzie $|S| = m$ i $|T| = n$. Dowolną funkcję $f\colon S \longrightarrow T$ można otrzymać określając najpierw $f(s_1)$ jako jeden spośród n elementów

zbioru T, następnie określając $f(s_2)$ jako jeden spośród n elementów zbioru T itd. Proces ten prowadzi do $n^m = n \cdot n \cdot \ldots \cdot n$ (m razy) różnych wyników, z których każdy jednoznacznie opisuje inną funkcję. Stąd wniosek, że funkcji odwzorowujących S w T jest n^m.

(b) W części (a) znaleźliśmy $|\text{FUN}(S,T)|$ dla zbioru $\text{FUN}(S,T)$ złożonego ze wszystkich funkcji ze zbioru S w zbiór T. Niektórzy piszą T^S zamiast $\text{FUN}(S,T)$. Ta notacja wygląda dziwnie, ale pozwala napisać

$$|T^S| = |T|^{|S|},$$

co daje „prawo potęgi" analogiczne do naszych praw sumy i iloczynu. ∎

Weźmy skończony niepusty zbiór S o n elementach i dodatnią liczbę całkowitą $r \leq n$. Ciąg r różnych elementów zbioru S nazywamy **r-wyrazową wariacją** ze zbioru S bez powtórzeń. To znaczy, że r-wyrazowa wariacja ze zbioru S bez powtórzeń jest to odwzorowanie różnowartościowe σ (mała grecka litera sigma) zbioru $\{1,2,\ldots,r\}$ w zbiór S. Dowolna r-wyrazowa wariacja ze zbioru S bez powtórzeń σ jest w pełni określona przez r-wyrazowy ciąg $(\sigma(1),\sigma(2),\ldots,\sigma(r))$ i będziemy czasem używać tego ciągu jako oznaczenia dla samej wariacji. Każdą r-wyrazową wariację ze zbioru S bez powtórzeń można otrzymać przypisując liczbie 1 dowolny spośród n elementów z S, następnie liczbie 2 dowolny spośród pozostałych $n-1$ elementów itd. Zatem na mocy prawa iloczynu (b), istnieje $n(n-1)(n-2)\ldots$ r-wyrazowych wariacji ze zbioru S bez powtórzeń, gdzie w wypisanym iloczynie występuje dokładnie r czynników. Ostatni z czynników jest więc równy $n-r+1$. Iloczyn ten oznaczamy czasami przez $P(n,r)$. Zatem istnieje dokładnie

$$P(n,r) = n(n-1)(n-2)\cdots(n-r+1) = \prod_{j=0}^{r-1}(n-j)$$

r-wyrazowych wariacji ze zbioru S bez powtórzeń. n-wyrazowe wariacje ze zbioru S bez powtórzeń nazywamy **permutacjami** zbioru S. Istnieje dokładnie $P(n,n) = n!$ permutacji zbioru S. Zauważmy, że $P(n,r)\cdot(n-r)! = n!$, a zatem

$$P(n,r) = \frac{n!}{(n-r)!} \quad \text{dla} \quad 1 \leq r \leq n.$$

Wygodnie jest przyjąć dodatkowo, że $P(n,0) = 1$, uważając „pustą permutację" za jedyną 0-wyrazową wariację ze zbioru S bez powtórzeń.

PRZYKŁAD 6 (a) Losowanie pięciu kart bez zwracania w przykładzie 2(b) odpowiada pięciowyrazowej wariacji bez powtórzeń z pięćdziesięciodwuelementowego zbioru wszystkich kart. Zatem liczba wyników takiego losowania wynosi $P(52,5) = 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48$.

(b) Niech Σ będzie siedmioliterowym alfabetem z przykładu 3(a). Słowa z Σ^5, w których nie ma powtarzających się liter, są pięciowyrazowymi wariacjami ze zbioru Σ bez powtórzeń. Takich wariacji jest $P(7,5) = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3$. Zauważmy, że puste słowo λ jest pustą permutacją zbioru Σ.

(c) Dwuliterowe słowa bez powtórzeń w drzewie z przykładu 3(b) są dwuwyrazowymi wariacjami bez powtórzeń z czteroelementowego zbioru Σ. Jest ich $P(4,2) = 4 \cdot 3 = 12$. ∎

PRZYKŁAD 7 Liczba różnych sposobów ułożenia 52 kart w talii wynosi $52! \approx 8{,}07 \cdot 10^{67}$, co jest ogromną liczbą. Pewne bardzo interesujące twierdzenia o tasowaniu kart uzyskane zostały całkiem niedawno — dopiero w latach osiemdziesiątych. Dotyczą one pytań takich jak: ile potrzeba tasowań, by dobrze wymieszać talię? Większość metod ma jedynie znaczenie teoretyczne; nawet największe komputery nie są w stanie przechowywać listy $52!$ pozycji i bezmyślnie sprawdzać przypadek po przypadku. ∎

W zagadnieniach zliczania, w których istotny jest porządek, r-wyrazowe wariacje bez powtórzeń są oczywiście właściwym narzędziem. Często jednak porządek nie jest istotny i wówczas ważna staje się umiejętność zliczania zbiorów. Wiemy już, że dowolny zbiór S o n elementach ma 2^n podzbiorów. Dla $0 \leq r \leq n$, niech $\binom{n}{r}$ będzie liczbą wszystkich r-elementowych podzbiorów zbioru S. Liczbę $\binom{n}{r}$ nazywany **współczynnikiem dwumianowym Newtona** i czytamy „n nad r". Jest ona też czasem nazywana liczbą r-elementowych **kombinacji** z elementów zbioru n-elementowego. Współczynniki dwumianowe zawdzięczają swą nazwę wzorowi dwumianowemu, który zostanie omówiony w § 5.3.

Twierdzenie

Dla $0 \leq r \leq n$ mamy

$$\binom{n}{r} = \frac{n!}{(n-r)!r!}.$$

Dowód. Niech S będzie zbiorem mającym n elementów. Rozważmy następujący dwustopniowy proces wyboru r-wyrazowej

wariacji ze zbioru S bez powtórzeń: najpierw wybierzmy r-elementowy podzbiór zbioru S (na jeden z $\binom{n}{r}$ sposobów), a potem ustawmy jego elementy w pewnej kolejności (na jeden z $r!$ sposobów). Wszystkich możliwych wyników tego procesu będzie $P(n,r)$, a zatem zastosowanie prawa iloczynu (b) daje

$$P(n,r) = \binom{n}{r} \cdot r!$$

Stąd

$$\binom{n}{r} = \frac{P(n,r)}{r!} = \frac{n!}{(n-r)!r!}. \qquad \blacksquare$$

Sztuczka, której tu użyliśmy, polegająca na zastosowaniu prawa iloczynu do sytuacji, gdy znając iloczyn chcemy określić wartość jednego z jego czynników, jest warta zapamiętania.

PRZYKŁAD 8 Znajdziemy liczbę wszystkich ciągów długości n, złożonych z zer i jedynek, w których występuje dokładnie r jedynek. Jest to równoważne z wyznaczeniem liczby wszystkich funkcji ze zbioru $\{1,2,\ldots,n\}$ w $\{0,1\}$, przyjmujących wartość 1 dokładnie r razy. Innymi słowy, chcemy znaleźć liczbę wszystkich funkcji charakterystycznych χ_A, gdzie $|A| = r$. Ale to jest dokładnie liczba wszystkich r-elementowych podzbiorów zbioru $\{1,2,\ldots,n\}$, tzn. $\binom{n}{r}$. $\blacksquare$

PRZYKŁAD 9 Rozważmy graf bez pętli, który jest **pełny** w tym sensie, że każda para wierzchołków jest połączona dokładnie jedną krawędzią. Jeśli graf ten ma n wierzchołków, to ile ma krawędzi? Załóżmy, że $n \geq 2$. Każda krawędź wyznacza pewien dwuelementowy podzbiór zbioru V wszystkich wierzchołków i odwrotnie, każdy dwuelementowy podzbiór V wyznacza pewną krawędź. Innymi słowy, istnieje odpowiedniość wzajemnie jednoznaczna między krawędziami grafu i dwuelementowymi podzbiorami zbioru V. Zatem liczba krawędzi naszego grafu wynosi

$$\binom{n}{2} = \frac{n!}{(n-2)!2!} = \frac{n(n-1)}{2}. \qquad \blacksquare$$

Świetnym sposobem zilustrowania zagadnień omawianych w tym paragrafie jest wyznaczenie liczby różnych rodzajów układów kart w pokerze. Talia kart składa się z czterech kolorów zwanych trefl, karo, kier i pik. Każdy kolor składa się z trzynastu kart: A, 2, 3, 4, 5, 6, 7, 8, 9, 10, W, D, K, wypisanych w kolejności starszeństwa. Tutaj A oznacza asa, W — waleta, D — damę

i K — króla. Są po cztery karty tej samej wysokości, po jednej każdego koloru. Układ kart w pokerze (zwany też figurą lub „ręką" pokerową) jest zbiorem 5 kart z talii 52 kart. Kolejność wyboru kart jest nieistotna. Sekwens składa się z pięciu kolejnych kart takich, jak np. 8, 9, 10, W, D. As A może wystąpić na początku ciągu A, 2, 3, 4, 5 lub na końcu ciągu 10, W, D, K, A. Układy kart w pokerze zalicza się do rozłącznych zbiorów, w zależności od ich rodzaju. Rodzaje te są następujące, a wypisane tu zostały w kolejności odwrotnej do szansy ich otrzymania.

Poker królewski — 10, W, D, K, A w jednym kolorze.
Poker — sekwens w jednym kolorze, nie będący pokerem królewskim.
Czwórka — cztery karty tego układu są tej samej wysokości; na przykład 4 czwórki i jedna dziewiątka.
Ful — trzy karty tej samej wysokości oraz dwie karty innej, lecz między sobą również tej samej wysokości; na przykład trzy walety i dwie ósemki.
Kolor — pięć kart w jednym kolorze, nie tworzących ani pokera królewskiego, ani pokera.
Strit — sekwens, nie będący ani pokerem królewskim, ani pokerem.
Trójka — trzy karty tej samej wysokości, czwarta karta innej wysokości i piąta karta jeszcze innej wysokości.
Dwie pary — dwie karty tej samej wysokości, inne dwie karty innej, lecz między sobą tej samej wysokości i ostatnia karta jeszcze innej wysokości; na przykład dwie królowe, dwie czwórki i siódemka.
Para — dwie karty tej samej wysokości, pozostałe dowolne, ale takie, by układ ten nie był żadnego z opisanych powyżej rodzajów.
Zerówka — układ, który nie jest żadnego z powyższych rodzajów.

PRZYKŁAD 10

(a) Jest $\binom{52}{5}$ układów kart w pokerze. Zauważ, że

$$\binom{52}{5} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 52 \cdot 17 \cdot 10 \cdot 49 \cdot 6 = 2598960.$$

(b) Ile układów kart w pokerze to fule? Układ składający się z trzech waletów i dwóch ósemek nazwijmy fulem typu (W, 8)

i analogicznie oznaczymy pozostałe typy fuli. Kolejność jest tu istotna, gdyż układy typu (8, W) mają po trzy ósemki i po dwa walety. Ponadto typy postaci (W, W) lub (8, 8) są niedopuszczalne. Tak więc typy fuli odpowiadają dwuwyrazowym wariacjom elementów zbioru możliwych wysokości kart; istnieje zatem $13 \cdot 12$ różnych typów fuli.

Znajdziemy teraz liczbę fuli każdego typu; weźmy na przykład typ (W, 8). Liczba sposobów wyboru trzech spośród czterech waletów wynosi $\binom{4}{3} = 4$, a następnie liczba sposobów wyboru dwóch spośród czterech ósemek równa jest $\binom{4}{2} = 6$. Są więc $4 \cdot 6 = 24$ układy typu (W, 8). Rozumowanie to można zastosować do każdego spośród $13 \cdot 12$ typów, a więc liczba fuli wynosi $13 \cdot 12 \cdot 24 = 3744$.

(c) Ile układów kart w pokerze to dwie pary? Powiemy, że układ należący do rodzaju dwie pary jest typu {D, 4}, jeśli zawiera dwie damy i dwie czwórki. Tym razem użyliśmy symbolu oznaczającego parę nieuporządkowaną, ponieważ porządek nie odgrywa tu roli: układy typu {4, D} to układy typu {D, 4} i nie chcemy zliczać ich podwójnie. Istnieje $\binom{13}{2}$ typów dwóch par. Dla każdego typu, powiedzmy, że jest to typ {D, 4}, są $\binom{4}{2}$ sposoby wyboru dwóch dam, $\binom{4}{2}$ sposoby wyboru dwóch czwórek i $52-8 = 44$ sposoby wyboru piątej karty. Zatem liczba układów kart rodzaju dwie pary wynosi

$$\binom{13}{2} \cdot \binom{4}{2} \cdot \binom{4}{2} \cdot 44 = 123552.$$

(d) Ile układów kart w pokerze to strity? Najpierw znajdziemy liczbę wszystkich sekwensów, nie wyłączając pokerów i pokerów królewskich. Sekwens złożony z kart wysokości 8, 9, 10, W, D nazwijmy sekwensem typu D. Ogólnie, przez typ sekwensu rozumiemy wysokość jego najstarszej karty. Ponieważ każda z wysokości 5, 6, 7, 8, 9, 10, W, D, K, A może być wysokością najstarszej karty w sekwensie, to jest 10 typów sekwensów. Dla ustalonego typu każdą z kart danych pięciu wysokości możemy wybrać na 4 sposoby. Jest więc 4^5 sekwensów każdego z typów i $10 \cdot 4^5 = 10240$ wszystkich sekwensów. Istnieją 4 pokery królewskie i 36 pokerów, jest więc 10200 stritów, tzn. sekwensów, które nie są żadnego z tych rzadkich rodzajów.

(e) Ćwiczenie 11 polega na znalezieniu liczby pozostałych układów kart w pokerze; wszystkie odpowiedzi zostały podane na końcu książki.

ĆWICZENIA DO § 5.1

1. Oblicz:
(a) $\binom{8}{3}$, (b) $\binom{8}{0}$, (c) $\binom{8}{5}$,
(d) $\binom{52}{50}$, (e) $\binom{52}{52}$, (f) $\binom{52}{1}$.

2. Niech $A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ i $B = \{2, 3, 5, 7, 11, 13, 17, 19\}$.
(a) Określ liczbę elementów zbiorów $A \cup B$, $A \cap B$ i $A \oplus B$.
(b) Ile jest podzbiorów zbioru A?
(c) Ile jest czteroelementowych podzbiorów zbioru A?
(d) Ile czteroelementowych podzbiorów zbioru A składa się z 3 liczb parzystych i 1 nieparzystej?

3. W pewnej grupie składającej się ze 150 osób 45 regularnie pływa, 40 jeździ na rowerze, a 50 uprawia jogging. Wiemy ponadto, że są 32 osoby, które uprawiają jogging, ale nie jeżdżą na rowerze, 27 takich, które uprawiają jogging i pływają i 10 uprawiających wszystkie trzy rodzaje aktywności.
(a) Ile osób uprawia jogging, ale nie pływa i nie jeździ na rowerze?
(b) Jeśli wiemy dodatkowo, że 21 osób jeździ na rowerze i pływa, to ile nie uprawia żadnej z powyższych aktywności?

4. Pewna grupa studencka składa się z 12 mężczyzn i 16 kobiet. Ile da się z nich utworzyć komisji, składających się z
(a) siedmiu osób?
(b) trzech mężczyzn i czterech kobiet?
(c) siedmiu kobiet lub siedmiu mężczyzn?

5. (a) Ile można utworzyć komisji składających się z 4 osób wybranych z 9-osobowej grupy?
(b) Odpowiedz ponownie na pytanie z części (a) przy dodatkowym założeniu, że są dwie osoby, Anna i Robert, które nie chcą być w tej samej komisji.

6. Ile można utworzyć komisji składających się z 4 mężczyzn i 4 kobiet wybranych z grupy, w której jest 8 mężczyzn i 6 kobiet?

7. Niech $S = \{a, b, c, d\}$ i $T = \{1, 2, 3, 4, 5, 6, 7\}$.
(a) Ile jest funkcji różnowartościowych z T w S?
(b) Ile jest funkcji różnowartościowych z S w T?
(c) Ile jest funkcji z S w T?

8. Niech $P = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ i $Q = \{A, B, C, D, E\}$.
(a) Ile jest czteroelementowych podzbiorów zbioru P?
(b) Ile jest permutacji zbioru Q?
(c) Ile jest numerów rejestracyjnych składających się z trzech liter ze zbioru Q i następujących po nich dwóch cyfr ze zbioru P? Powtórzenia są dozwolone; np. DAD88 jest dopuszczalnym numerem.

9. Z talii złożonej z 52 kart losujemy karty ze zwracaniem.
(a) Na ile sposobów można wylosować kolejno dziesięć kart tak, by dziesiąta z nich nie wystąpiła wcześniej?
(b) Na ile sposobów można wylosować kolejno dziesięć kart tak, by dziesiąta z nich wystąpiła wcześniej?

10. Niech Σ będzie alfabetem $\{a, b, c, d, e\}$ i $\Sigma^k = \{w \in \Sigma^*$: długość$(w) = k\}$. Ile elementów mają następujące zbiory?
(a) Σ^k, dla każdego $k \in \mathbb{N}$,
(b) $\{w \in \Sigma^3$: żadna litera nie występuje w w więcej niż raz$\}$,
(c) $\{w \in \Sigma^4$: litera c występuje w w dokładnie raz$\}$,
(d) $\{w \in \Sigma^4$: litera c występuje w w co najmniej raz$\}$.

11. Znajdź liczbę układów kart w pokerze następujących rodzajów:
(a) czwórka,
(b) kolor,
(c) trójka,
(d) para.

12. (a) Na ile sposobów można ustawić litery a, b, c, d, e, f w takiej kolejności, by litery a i b sąsiadowały ze sobą?
(b) Na ile sposobów można ustawić litery a, b, c, d, e, f w takiej kolejności, by litery a i b nie sąsiadowały ze sobą?
(c) Na ile sposobów można ustawić litery a, b, c, d, e, f w takiej kolejności, by litery a i b sąsiadowały ze sobą, ale litery a i c nie?

13. (a) Znajdź macierz pełnego grafu o n wierzchołkach; zob. przykład 9.
(b) Używając macierzy z części (a) znajdź liczbę krawędzi w tym grafie.
Wskazówka: na ilu miejscach w naszej macierzy stoi 1?

14. (a) Weźmy pełny graf G o n wierzchołkach, gdzie $n \geq 3$. Znajdź liczbę cykli w G długości n.
(b) Ile jest cykli w grafie pełnym o 5 wierzchołkach?

15. Weźmy pełny graf o n wierzchołkach, gdzie $n \geq 4$.
(a) Znajdź liczbę dróg długości 3.
(b) Znajdź liczbę dróg długości 3, w których wierzchołki nie powtarzają się.
(c) Znajdź liczbę dróg prostych długości 3, tzn. dróg złożonych z różnych krawędzi.

§ 5.2. Elementarny rachunek prawdopodobieństwa

W poprzednim paragrafie znaleźliśmy liczbę wszystkich układów kart w pokerze oraz liczby różnych rodzajów układów. Liczby te same przez się nie są może zbyt fascynujące, ale pokerzyści

naprawdę są zainteresowani, jaką część wszystkich układów stanowią kolory, fule itd. Dlaczego? Ponieważ jeśli wszystkie układy są jednakowo prawdopodobne, to ułamki te reprezentują szansę bądź prawdopodobieństwo otrzymania jednego z wymienionych korzystnych układów. Powrócimy do tych układów po uprzednim wprowadzeniu pewnych oznaczeń i terminologii.

Podstawową strukturą w rachunku prawdopodobieństwa jest zbiór nazywany **przestrzenią zdarzeń elementarnych** lub **przestrzenią próbek**, składający się z możliwych wyników jakiegoś eksperymentu, gry losowej, pomiarów itd. Przestrzeń zdarzeń elementarnych tradycyjnie oznacza się wielką grecką literą omega Ω, a jej elementy małymi ω lub innymi greckimi literami. Podzbiory zbioru Ω nazywa się **zdarzeniami**. Prawdopodobieństwem lub **funkcją prawdopodobieństwa** na Ω nazywa się funkcję P, która każdemu zdarzeniu $E \subseteq \Omega$ przypisuje pewną liczbę $P(E)$ z przedziału $[0,1]$ i spełnia następujące warunki:

(P_1) $P(\Omega) = 1$,

(P_2) $P(E \cup F) = P(E) + P(F)$ dla dowolnych rozłącznych zdarzeń E i F.

Na razie myślimy o Ω jak o zbiorze skończonym.

Myśl jest taka, że liczba $P(E)$, zwana **prawdopodobieństwem zdarzenia** E, powinna mierzyć szansę tego, że otrzymany wynik należeć będzie do E. Chcemy, by $P(E) = 0$ znaczyło, że nie ma żadnej szansy na to, że wynik naszego eksperymentu znajdzie się w E, a $P(E) > 0$ znaczyło, że jakaś szansa na to, że wynik należeć będzie do E, istnieje. W takim razie z pewnością chcemy, by $P(\Omega) > 0$ i standardowo przyjmuje się, że $P(\Omega) = 1$, tak jak to zrobiliśmy w (P_1). Warunek (P_2) odzwierciedla nasz sposób myślenia o prawdopodobieństwie zachodzenia zdarzeń w realnym świecie; prawdopodobieństwo, że z worka wyciągniemy kulkę czerwoną lub zieloną jest równe: prawdopodobieństwo wyciągnięcia kulki czerwonej plus prawdopodobieństwo wyciągnięcia kulki zielonej. Po zapoznaniu się z paroma przykładami pokażemy, że warunki (P_1) i (P_2) są wystarczające, by dostarczyć użytecznego matematycznego modelu pojęcia prawdopodobieństwa.

Często, tak jak w przypadku układów kart w pokerze, rozsądnie jest zakładać, że wszystkie wyniki są **jednakowo prawdopodobne**. Ponieważ na mocy (P_2) i (P_1) mamy

$$\sum_{\omega \in \Omega} P(\{\omega\}) = P(\Omega) = 1,$$

więc w takim przypadku $P(\{\omega\}) = 1/|\Omega|$ dla $\omega \in \Omega$. Stosując powtórnie (P_2) widzimy, że

$$P(E) = |E|/|\Omega| \text{ dla } E \subseteq \Omega,$$
jeśli wszystkie wyniki są jednakowo prawdopodobne.

Nasz pierwszy przykład wykorzystuje obliczenia z przykładu 10 z § 5.1.

PRZYKŁAD 1 W grze w pokera zbiorem Ω wszystkich możliwych wyników jest zbiór wszystkich możliwych układów kart. Zatem $|\Omega| = 2598960$. Typowe zdarzenie ma postać

$$F = \{\omega \in \Omega\colon \omega \text{ jest fulem}\}.$$

Zdarzenie „w wyniku rozdania kart dostała fula" odpowiada dokładnie stwierdzeniu: układ kart ω należy do zbioru F. Prawdopodobieństwo dostania fula w rozdaniu wynosi

$$P(F) = \frac{|F|}{|\Omega|} = \frac{3744}{2598960} \approx 0{,}00144,$$

czyli około 1 do 700. W rzeczywistości fule zdarzają się graczom w pokera znacznie częściej, ponieważ zasady gry w pokera pozwalają im na wybiórcze wymienianie kart już po wyjściowym rozdaniu. Powyższa wartość $P(F)$ jest prawdopodobieństwem dostania fula w wyniku rozdania, bez uwzględniania możliwości, jakie dają wyszukane zasady gry.

(b) Innym ważnym zdarzeniem w pokerze jest

$$K = \{\omega \in \Omega\colon \omega \text{ jest kolorem}\}.$$

Możemy je w skrócie zapisać jako K=„ω jest kolorem" lub po prostu K=„kolor". Z ćwiczenia 11 z § 5.1 wiemy, że $|K| = 5108$, więc

$$P(K) = \frac{|K|}{|\Omega|} = \frac{5108}{2598960} \approx 0{,}00197,$$

czyli około 1 do 500.

(c) Prawdopodobieństwo dostania fula lub koloru jest równe $P(F \cup K)$. Układ kart w pokerze nie może być jednocześnie fulem i kolorem, więc zdarzenia F i K są rozłączne. Stąd

$$P(F \cup K) = P(F) + P(K) \approx 0{,}00341.$$

Jest to ilustracja podstawowej własności (P_2) prawdopodobieństwa: prawdopodobieństwo sumy zdarzeń rozłącznych jest sumą ich prawdopodobieństw. Własność (P_1) mówi nam jedynie, że prawdopodobieństwo tego, że w wyniku rozdania kart dostaniemy układ kart, wynosi 1.

Rozważmy ponownie dowolne prawdopododobieństwo P w przestrzeni zdarzeń elementarnych Ω. Ponieważ zbiór Ω składa się ze wszystkich możliwych wyników, prawdopodobieństwo tego, że nie zajdzie żaden spośród nich, powinno z pewnością wynosić zero. To znaczy, że $P(\emptyset) = 0$. Wynika to także łatwo z aksjomatu (P_2):

$$P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset), \text{ co daje } P(\emptyset) = 0.$$

Prawdopodobieństwo, że dane zdarzenie nie zajdzie powinno być równe 1 minus prawdopodobieństwo jego zajścia. Znaczy to, że $P(E^c) = 1 - P(E)$ dla $E \subseteq \Omega$. Jest to znowu konsekwencją aksjomatów, gdyż

$$P(E) + P(E^c) = P(E \cup E^c) = P(\Omega) = 1.$$

W ten sposób uzyskaliśmy już punkty (a) i (b) następującego twierdzenia.

Twierdzenie 1

Niech P będzie prawdopodobieństwem w przestrzeni zdarzeń elementarnych Ω.

(a) $P(\emptyset) = 0$.

(b) $P(E^c) = 1 - P(E)$ dla dowolnego zdarzenia E.

(c) $P(E \cup F) = P(E) + P(F) - P(E \cap F)$ dla dowolnych zdarzeń E i F, bez względu na to, czy są one rozłączne, czy nie.

(d) $P(E_1 \cup E_2 \cup \ldots \cup E_m) = \sum_{k=1}^{m} P(E_k) = P(E_1) + P(E_2) + \ldots + P(E_m)$ dla dowolnych parami rozłącznych zdarzeń $E_1, E_2, \ldots, E_m$.

Przez **parami rozłączne** rozumiemy, że $E_j \cap E_k = \emptyset$, o ile $j \neq k$.

Dowód. (c) Rozumowanie jest w istocie to samo, co dla prawa sumy (b) z § 5.1. Mamy $E \cup F = E \cup (F \setminus E)$ i $F = (F \setminus E) \cup (E \cap F)$, przy czym obie sumy są rozłączne. Zatem $P(E \cup F) = P(E) + P(F \setminus E)$ i $P(F) = P(F \setminus E) + P(E \cap F)$, a stąd $P(E \cup F) = P(E) + P(F) - P(E \cap F)$.

(d) To jest proste rozumowanie indukcyjne. Na mocy aksjomatu (P_2) dowodzony fakt jest prawdziwy dla $m = 2$. Załóżmy, że równość jest prawdziwa dla m zbiorów i niech $E_1, E_2, \ldots, E_{m+1}$ będą parami rozłączne. Wtedy zbiór $E_1 \cup E_2 \cup \ldots \cup E_m$ jest rozłączny ze zbiorem E_{m+1}, a stąd

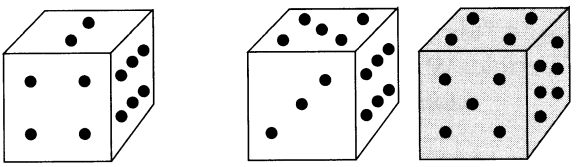
$$\begin{aligned} P(E_1 \cup E_2 \cup \ldots \cup E_{m+1}) &= P(E_1 \cup E_2 \cup \ldots \cup E_m) + P(E_{m+1}) \\ &= P(E_1) + P(E_2) + \ldots + P(E_m) + P(E_{m+1}). \end{aligned}$$

∎

PRZYKŁAD 2 W przykładzie 4 z § 5.1 obliczyliśmy, że na rysunku 5.2(a) jest 12 dróg prowadzących z s do f. Zauważmy, że każda z tych dróg ma długość 4. Przypuśćmy, że losowo wybieramy zbiór złożony z czterech krawędzi. Jakie jest prawdopodobieństwo, że wybrane krawędzie są krawędziami drogi z s do f?

W tym przypadku zbiór Ω składa się ze wszystkich czteroelementowych podzbiorów zbioru wszystkich krawędzi. Wszystkich krawędzi jest 18, więc $|\Omega| = \binom{18}{4} = 3060$. Jedynie 12 spośród tych podzbiorów tworzy drogi z s do f, zatem odpowiedzią na nasze pytanie jest $12/3060 \approx 0{,}00392$. ∎

Teraz czas na łatwy przykład.



(a) 1 kostka (b) 2 kostki

Rysunek 5.3

PRZYKŁAD 3 Zwykła kostka do gry przedstawiona na rysunku 5.3(a) ma sześć ścianek z, odpowiednio, 1, 2, 3, 4, 5 lub 6 oczkami. Kiedy rzucimy kostką, na górnej ściance wypadnie pewna liczba oczek. Jeśli wyrzucenie każdej liczby oczek jest jednakowo prawdopodobne, to mówimy, że kostka jest symetryczna. Załóżmy więc, że nasza kostka jest **symetryczna**. Niech $\Omega = \{1, 2, 3, 4, 5, 6\}$ będzie zbiorem wszystkich możliwych wyników. Ponieważ kostka jest symetryczna, to $P(k) = \frac{1}{6}$ dla każdego $k \in \Omega$; piszemy tu $P(k)$ zamiast $P(\{k\})$. Jeśli E oznacza zdarzenie „k jest liczbą parzystą", tzn. jeśli $E = \{2,\ 4,\ 6\}$, to $P(E) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$. Jeśli F oznacza zdarzenie „cztery lub pięć", tzn. jeśli $F = \{4,\ 5\}$, to $P(F) = \frac{1}{3}$. Mamy również

$$P(E \cup F) = P(\{2,\ 4,\ 5,\ 6\}) = \tfrac{2}{3} \quad \text{oraz} \quad P(E \cap F) = P(\{4\}) = \tfrac{1}{6}.$$

Oczywiście, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$. ∎

Ostatni przykład był zbyt łatwy. Następny będzie w sam raz.

PRZYKŁAD 4 Rozważmy teraz rzuty dwiema symetrycznymi kostkami, białą i szarą. Wynikiem, który nas na ogół interesuje, jest suma liczb oczek, które wypadły na obu kostkach; na przykład dla kostek z rys. 5.3(b) suma ta wynosi 9. Po raz pierwszy nie jest całkiem jasne, jaka powinna być przestrzeń zdarzeń elementar-

nych Ω. W gruncie rzeczy wybór należy do nas, ale niektórymi zbiorami można się będzie łatwiej posługiwać niż innymi. Ponieważ interesującymi nas wynikami są sumy, kuszące jest określenie Ω jako jedenastoelementowego zbioru złożonego z 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 i 12. Kłopot w tym, że te wyniki nie są jednakowo prawdopodobne. Moglibyśmy używać ich prawdopodobieństw (znalezionych poniżej), ale prościej jest przyjąć, że Ω składa się ze wszystkich uporządkowanych par liczb oczek, które mogły wypaść na naszych kostkach:

$$\Omega = \{(k,l)\colon 1 \leq k \leq 6, 1 \leq l \leq 6\}.$$

Pierwsza wartość, k, oznacza tu liczbę z kostki białej, druga wartość, l, jest liczbą z kostki szarej. Sytuacja przedstawiona na rysunku 5.3(b) odpowiada parze $(5,4)$. Wydaje się rozsądne przyjąć, że każdy spośród 36 takich wyników jest jednakowo prawdopodobny; zostanie to dokładniej uzasadnione w § 9.1. Wyniki te są wypisane w tablicy 5.1.

Tablica 5.1

$(1,1)$	$(1,2)$	$(1,3)$	$(1,4)$	$(1,5)$	$(1,6)$
$(2,1)$	$(2,2)$	$(2,3)$	$(2,4)$	$(2,5)$	$(2,6)$
$(3,1)$	$(3,2)$	$(3,3)$	$(3,4)$	$(3,5)$	$(3,6)$
$(4,1)$	$(4,2)$	$(4,3)$	$(4,4)$	$(4,5)$	$(4,6)$
$(5,1)$	$(5,2)$	$(5,3)$	$(5,4)$	$(5,5)$	$(5,6)$
$(6,1)$	$(6,2)$	$(6,3)$	$(6,4)$	$(6,5)$	$(6,6)$

Szczególnie interesują nas zdarzenia postaci „suma wynosi k" dla $k = 2,3,4,\ldots,12$. Na przykład

$$P(\text{suma wynosi } 9) = P(\{(3,6),(4,5),(5,4),(6,3)\})$$
$$= \frac{4}{36} = \frac{1}{9}.$$

Ta oraz pozostałe wartości wyrażenia $P(\text{suma wynosi } k)$ podane są w tablicy 5.2.

Tablica 5.2

Suma	P(Suma)	Suma
2	1/36	12
3	2/36 = 1/18	11
4	3/36 = 1/12	10
5	4/36 = 1/9	9
6	5/36	8
7	6/36 = 1/6	

Zauważmy, że $(4,5)$ i $(5,4)$ to naprawdę dwa różne wyniki. Z drugiej strony, jak widać z tablicy 5.1, jest tylko jeden sposób, by otrzymać dwie piątki, a mianowicie $(5,5)$. Zatem

$$P(\text{suma wynosi } 10) = P(\{(4,6),(5,5),(6,4)\}) = \frac{3}{36} = \frac{1}{12}.$$

Tablicy 5.1 można użyć do znalezienia odpowiedzi na dowolne pytanie z zakresu rachunku prawdopodobieństwa, które odnosi się do naszych dwóch kostek. Możemy skorzystać z tablicy 5.2 zamiast tablicy 5.1 tylko wówczas, gdy pytanie dotyczy jedynie sum liczb oczek.

(a) Jakie jest prawdopodobieństwo, że suma liczb oczek na kostkach będzie większa od 7? Zdarzenie to składa się z par uporządkowanych znajdujących się w tablicy 5.1 poniżej przerywanej linii. Takich par jest 15, a zatem odpowiedzią jest $\frac{15}{36} = \frac{5}{12}$. Możemy również skorzystać z tablicy 5.2 i otrzymamy wówczas:

$$\begin{aligned} P(\text{suma} > 7) &= \sum_{k=8}^{12} P(\text{suma wynosi } k) \\ &= \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{15}{36}. \end{aligned}$$

(b) Jakie jest prawdopodobieństwo, że liczba oczek z kostki białej dzieli liczbę oczek z kostki szarej? To znaczy, ile wynosi $P(E)$, gdy $E = \{(k,l)\colon\ k|l\}$? Tym razem tablica 5.2 nie przyda się do niczego, ale możemy z pomocą lub też bez pomocy tablicy 5.1 wypisać wszystkie elementy zbioru E:

$$\begin{aligned} E = \{&(1,1),\ (1,2),\ (1,3),\ (1,4),\ (1,5),\ (1,6),\ (2,2),\ (2,4),\\ &(2,6),\ (3,3),\ (3,6),\ (4,4),\ (5,5),\ (6,6)\}. \end{aligned}$$

Zatem $P(E) = \frac{14}{36}$. ■

PRZYKŁAD 5

(a) Monetę nazywamy **symetryczną**, jeśli prawdopodobieństwo wypadnięcia orła i prawdopodobieństwo wypadnięcia reszki wynosi $\frac{1}{2}$. Jeśli O oznacza, że wypadł orzeł, a R, że wypadła reszka, to $P(\text{O}) = P(\text{R}) = \frac{1}{2}$. Chcemy rzucić monetą kilkakrotnie, powiedzmy n razy. Zbiór możliwych wyników odpowiada zbiorowi Ω wszystkich ciągów długości n o wyrazach O i R. Na przykład (R, O, O, R, O) odpowiada $n = 5$ rzutom takim, że w pierwszym i czwartym wypadła reszka, a w pozostałych trzech orzeł. Podobnie jak w przykładzie 4, rozsądnie jest zakładać, że każdy spośród wszystkich 2^n ciągów należących do zbioru Ω jest jednakowo prawdopodobny.

Dla $r = 0, 1, \ldots, n$, obliczymy P (orzeł wypadł w dokładnie r rzutach). Liczba wszystkich n-wyrazowych ciągów z r orłami jest to dokładnie liczba wszystkich ciągów złożonych z zer i jedynek, długości n, o r jedynkach. Jak zostało pokazane w przykładzie 8 z § 5.1, liczba ta wynosi $\binom{n}{r}$. Zatem

$$P(\text{orzeł wypadł w dokładnie } r \text{ rzutach}) = \frac{\binom{n}{r}}{2^n}.$$

(b) Gdybyśmy rzucili symetryczną monetą dziesięć razy i otrzymali 8 lub więcej orłów bądź 8 lub więcej reszek, to bylibyśmy zdziwieni. Czy rzeczywiście powinniśmy być? Znajdziemy

$$P(\text{liczba orłów jest } \geq 3 \text{ i } \leq 7) = \sum_{r=3}^{7} P(\text{liczba orłów wynosi } r)$$

$$= \frac{1}{2^{10}}\left[\binom{10}{3} + \binom{10}{4} + \binom{10}{5} + \binom{10}{6} + \binom{10}{7}\right]$$

$$= \frac{1}{1024}[120 + 210 + 252 + 210 + 120] = \frac{912}{1024} \approx 0{,}891.$$

Wynika stąd, że szansa otrzymania 8 lub więcej orłów bądź 8 lub więcej reszek wynosi ponad 10 procent; okazuje się więc, że gdyby się tak zdarzyło, to byłoby to tylko trochę zaskakujące.

(c) Rachunki z części (b) mogłyby się nam wymknąć spod kontroli, gdybyśmy rzucali monetą 100 razy i chcieli, powiedzmy, obliczyć

$$P(\text{liczba orłów jest } \geq 36 \text{ i } \leq 64).$$

W paragrafie 9.4 wskażemy, jak takie wyniki można oszacować. Zobaczymy również, że waga uzyskanych wzorów wykracza daleko poza problem rzucania monetą. Nawiasem mówiąc,

$$P(\text{liczba orłów jest } \geq 36 \text{ i } \leq 64) = 0{,}996,$$

więc to naprawdę byłoby zaskakujące, gdybyśmy otrzymali więcej niż 64 lub mniej niż 36 orłów. ∎

Nasze rozważania z poprzedniego przykładu, dotyczące zdarzeń „zaskakujących", dawały, rzecz jasna, wyraz odczuciom subiektywnym. Jednakże w wielu analizach statystycznych zdarzenia o prawdopodobieństwach mniejszych niż 0,05 rzeczywiście uważa się za zaskakujące lub nieoczekiwane, podczas gdy zdarzeń o prawdopodobieństwach większych niż 0,05 za takie się nie uważa.

Przestrzenie zdarzeń elementarnych często bywają nieskończone. W tym przypadku nasze aksjomaty muszą zostać zmodyfikowane. Podamy najpierw kilka przykładów.

PRZYKŁAD 6 (a) Rozważmy eksperyment polegający na rzucaniu symetryczną monetą aż do momentu, gdy pojawi się orzeł. Naszą przestrzenią zdarzeń elementarnych będzie $\Omega = \mathbb{P} = \{1, 2, 3, \ldots\}$, gdzie wynik k odpowiada k rzutom. Tak jak w przykładzie 5 widzimy, że

$P(1)$ = prawdopodobieństwo wypadnięcia orła w pierwszym rzucie $= \frac{1}{2}$,

$P(2)$ = prawdopodobieństwo, że w pierwszym rzucie wypadnie reszka, a w drugim orzeł $= \frac{1}{2^2}$,

$P(3)$ = prawdopodobieństwo, że w dwóch pierwszych rzutach wypadną reszki, a następnie orzeł $= \frac{1}{2^3}$

itd. Ogólnie $P(k) = 1/2^k$. Dla skończonego zbioru $E \subseteq \Omega$ definiujemy $P(E) = \sum_{k \in E} 1/2^k$. Na przykład prawdopodobieństwo otrzymania orła w mniej niż sześciu rzutach wynosi

$$P(\{1, 2, 3, 4, 5\}) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{32} = \frac{31}{32} \approx 0,969.$$

Nie sprawdziliśmy, czy P spełnia aksjomaty funkcji prawdopodobieństwa, ale tak istotnie jest. Jeśli kiedykolwiek widziałeś szereg nieskończony, to wiesz, że

$$P(\Omega) = \sum_{k \in \mathbb{P}} P(k) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

Jeśli nie widziałeś, to możesz uważać powyższy zapis za skrót stwierdzenia

$$\sum_{k=1}^{n} \frac{1}{2^k} \text{ jest bardzo bliskie 1 dla bardzo dużych } n.$$

Ogólniej, „suma” $\sum_{k \in E} 1/2^k$ ma sens dla każdego zdarzenia $E \subseteq \Omega$ i definiujemy $P(E)$ jako właśnie tę sumę. Na przykład prawdopodobieństwo, że będziemy musieli rzucić monetą nieparzystą liczbę razy, zanim wypadnie orzeł, wynosi

$$P(\{1, 3, 5, 7, 9, \ldots\}) = \sum_{k \text{ nieparzyste}} \frac{1}{2^k},$$

co okazuje się równe $\frac{2}{3}$.

(b) Mówi się czasem o liczbach losowych w zbiorze $\Omega = [0,1)$. Komputery twierdzą, że potrafią generować liczby losowe w $[0,1)$. Co miałoby to znaczyć? Jeśli zapomnimy o tym, że zarówno ludzie, jak i komputery tak naprawdę mają do czynienia wyłącznie ze zbiorami skończonymi, to stwierdzić musimy, że Ω stanowi nieskończony zbiór wyników, i chcemy, by były one jednakowo prawdopodobne. Ale wówczas $P(\omega)$ musi się równać 0 dla każdego ω z $[0,1)$ i definicja $P(E) = \sum_{\omega \in E} P(\omega)$ dla $E \subseteq \Omega$ nie ma sensu.

Nie wszystko jest jeszcze stracone, ale nie możemy oprzeć definicji naszego prawdopodobieństwa wyłącznie na wynikach ω. Chcielibyśmy mieć $P([0, \frac{1}{2})) = \frac{1}{2}$, $P([\frac{3}{4}, 1)) = \frac{1}{4}$ itd. Okazuje się, że istnieje funkcja prawdopodobieństwa P zdefiniowana dla niektórych podzbiorów $[0,1)$, zwanych zdarzeniami, taka, że $P([a,b)) = b - a$, o ile $[a,b) \subseteq [0,1)$. Ta funkcja P spełnia aksjomaty prawdopodobieństwa i co więcej:

$$P(\bigcup_{k=1}^{\infty} E_k) = \sum_{k=1}^{\infty} P(E_k)$$

dla ciągów parami rozłącznych zdarzeń E_1, $E_2, \ldots$ z $[0,1)$.

(c) Jest wiele użytecznych funkcji prawdopodobieństwa P na zbiorze $\Omega = [0,1)$. Prawdopodobieństwo z punktu (b) jest tym jedynym, dla którego $P([a,b)) = b - a$ dla $[a,b) \subseteq [0,1)$. Jest także wiele innych prawdopodobieństw na $\mathbb{P}$, oprócz opisanego w części (a). ∎

A oto obiecane modyfikacje definicji prawdopodobieństwa P na nieskończonej przestrzeni zdarzeń elementarnych Ω. Jak wspomniano w przykładzie 6(b), w ogólnym przypadku tylko niektóre podzbiory zbioru Ω są uważane za zdarzenia. Ponadto (P_2) zostaje wzmocnione do

$$(P_2') \qquad P\Big(\bigcup_{k=1}^{\infty} E_k\Big) = \sum_{k=1}^{\infty} P(E_k)$$

dla ciągów parami rozłącznych zdarzeń E_1, $E_2, \ldots$ z Ω.

ĆWICZENIA DO § 5.2

Ilekroć mowa jest o wyborach losowych, tylekroć zakłada się, że wszystkie możliwe wyniki są jednakowo prawdopodobne.

1. Losowo wybieramy liczbę całkowitą ze zbioru $\{1, 2, 3, \ldots, 25\}$. Znajdź prawdopodobieństwo, że wybrana liczba jest

(a) podzielna przez 3,
(b) podzielna przez 5,
(c) pierwsza.

2. Losowo wybieramy literę polskiego alfabetu. Jakie jest prawdopodobieństwo, że jest to samogłoska (a, ą, e, ę, i, o, ó, u, y)?

3. Losowo wybieramy czteroliterowe słowo należące do zbioru Σ^4, gdzie $\Sigma = \{a, b, c, d, e\}$.
(a) Jakie jest prawdopodobieństwo, że litery w wybranym słowie są różne?
(b) Jakie jest prawdopodobieństwo, że w wybranym słowie nie ma samogłosek?
(c) Jakie jest prawdopodobieństwo, że wybrane słowo zaczyna się od samogłoski?

4. Losowo wybieramy pięcioliterowe słowo należące do zbioru Σ^5, gdzie $\Sigma = \{a, b, c\}$. Powtórz ćwiczenie 3.

5. Urna zawiera trzy czerwone i cztery czarne kule. Losowo wyciągnięto z urny trzy kule (bez zwracania). Podaj prawdopodobieństwo, że wśród tych trzech kul
(a) wszystkie są czerwone,
(b) wszystkie są czarne,
(c) jest jedna czerwona i dwie czarne,
(d) są dwie czerwone i jedna czarna.
(e) Dodaj do siebie wyniki otrzymane w punktach (a)-(d).

6. W urnie są trzy czerwone i dwie czarne kule. Losowo wyciągnięto z urny dwie kule (bez zwracania). Jakie jest prawdopodobieństwo, że te dwie kule są
(a) obie czerwone?
(b) obie czarne?
(c) różnych kolorów?

7. Przypuśćmy, że pewien eksperyment prowadzi do rozpatrywania zdarzeń A, B i C o następujących prawdopodobieństwach: $P(A) = 0,5$, $P(B) = 0,8$, $P(A \cap B) = 0,4$. Znajdź
(a) $P(B^c)$,
(b) $P(A \cup B)$,
(c) $P(A^c \cup B^c)$.

8. Przypuśćmy, że pewien eksperyment prowadzi do rozpatrywania zdarzeń A, B i C o następujących prawdopodobieństwach: $P(A) = 0,6$ i $P(B) = 0,7$. Pokaż, że $P(A \cap B) \geq 0,3$.

9. Jakie jest prawdopodobieństwo, że w wyniku rozdania kart w pokerze dostanie się
(a) czwórkę,
(b) trójkę,
(c) strita,

(d) dwie pary,
(e) parę.
Wskazówka: skorzystaj z przykładu 10 i ćwiczenia 11 z § 5.1.

10. Jakie jest prawdopodobieństwo, że w wyniku rozdania kart w pokerze dostanie się
(a) układ starszy od pary? „Starszy" znaczy tu dowolny inny i różny od zerówki, spośród układów wypisanych w § 5.1.
(b) parę waletów lub starszą od nich? Jedynymi parami starszymi od pary waletów są: para dam, para króli i para asów.

11. Rzucamy białą i szarą kostką tak, jak w przykładzie 4. Jakie jest prawdopodobieństwo, że
(a) suma wartości wyrzuconych oczek jest parzysta?
(b) liczba oczek, która wypadła na kostce szarej, jest większa od liczby oczek, która wypadła na kostce białej?
(c) liczba oczek, która wypadła na kostce szarej, jest dwa razy większa od liczby oczek, która wypadła na kostce białej?

12. Rzucamy dwiema kostkami tak, jak w ćwiczeniu 11. Jakie jest prawdopodobieństwo, że
(a) maksimum z liczb wyrzuconych oczek wynosi 4?
(b) minimum z liczb wyrzuconych oczek wynosi 4?
(c) iloczyn liczb wyrzuconych oczek wynosi 4?

13. Niech P będzie funkcją prawdopodobieństwa na przestrzeni zdarzeń elementarnych Ω. Pokaż, że dla dowolnych zdarzeń E_1, E_2 i E_3 zachodzi

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3).$$

14. Niech P będzie funkcją prawdopodobieństwa na przestrzeni zdarzeń elementarnych Ω. Pokaż, że jeśli E i F są zdarzeniami i $E \subseteq F$, to $P(E) \leq P(F)$.

15. Rzucamy sześć razy symetryczną monetą. Znajdź prawdopodobieństwo otrzymania
(a) samych reszek,
(b) jednego orła,
(c) dwóch orłów,
(d) trzech orłów,
(e) więcej niż trzech orłów.

16. Rzucamy monetą symetryczną aż do momentu, kiedy po raz pierwszy wypadnie orzeł. Jakie jest prawdopodobieństwo, że rzucimy co najmniej cztery razy?

17. Rzucamy monetą symetryczną n razy. Wykaż, że prawdopodobieństwo wypadnięcia parzystej liczby orłów wynosi $\frac{1}{2}$.

18. Prawdopodobieństwo, że wygram pierwszą partię tryktraka wynosi 0,5, że wygram drugą — 0,4, a że wygram obie — 0,3. Ile wynosi prawdopodobieństwo, że przegram obie partie?

19. Losowo wybieramy czteroelementowy podzbiór liczb zbioru $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$ (nie ma zwracania). Jakie jest prawdopodobieństwo, że
(a) dokładnie dwie z nich są parzyste?
(b) żadna z nich nie jest parzysta?
(c) dokładnie jedna z nich jest parzysta?
(d) dokładnie trzy z nich są parzyste?
(e) wszystkie cztery są parzyste?

20. (a) Pewien student odpowiada losowo „tak" lub „nie" na trzy pytania, z kórych składa się test. Jakie jest prawdopodobieństwo, że uzyska co najmniej dwie trzecie poprawnych odpowiedzi?
(b) Powtórz część (a) dla testu składającego się z sześciu pytań.
(c) Powtórz część (a) dla testu składającego się z dziewięciu pytań.

21. Pewien program komputerowy wybiera losowo liczbę ze zbioru $\{k\colon 1 \leq k \leq 1000000\}$ i drukuje wynik. Powtarza to milion razy. Jakie jest prawdopodobieństwo, że wartość $k = 1$ pojawi się na wydruku co najmniej jeden raz? *Wskazówki*:
(a) Trzykrotnie wybieramy losowo liczbę ze zbioru $\{1, 2, 3\}$. Jakie jest prawdopodobieństwo, że liczba 1 została wybrana co najmniej raz? (Znajdź najpierw prawdopodobieństwo, że liczba 1 nie została wybrana.)
(b) Czterokrotnie wybieramy losowo liczbę ze zbioru $\{1, 2, 3, 4\}$. Jakie jest prawdopodobieństwo, że liczba 1 została wybrana co najmniej raz?
(c) n-krotnie wybieramy losowo liczbę ze zbioru $\{1, 2, \ldots, n\}$. Jakie jest prawdopodobieństwo, że liczba 1 została wybrana co najmniej raz?
(d) Weź $n = 1000000$.

§ 5.3. Zasada włączeń i wyłączeń, metody dwumianowe

Paragraf ten zawiera rozwinięcie metod zliczania wprowadzonych w § 5.1. Zasada włączeń i wyłączeń jest uogólnieniem prawa sumy, dającym możliwość znajdowania liczby elementów sum więcej niż dwóch zbiorów. Wzór dwumianowy, jeden z podstawowych faktów algebry, jest blisko związany z zagadnieniami zliczania. Nasz trzeci temat, rozmieszczanie przedmiotów w pudełkach, ukazuje jeszcze jedno zastosowanie współczynników dwumianowych.

Często łatwo jest znaleźć liczbę elementów przecięcia zbiorów, w definicji którego występuje spójnik „i". Z drugiej strony, bezpośrednie obliczenie liczby elementów sumy kilku zbiorów jest często

trudne. Zasada włączeń i wyłączeń pozwoli nam wyrazić liczbę elementów sumy poprzez liczby elementów rozmaitych przecięć.

Niech $A_1, A_2, \ldots, A_n$ będą zbiorami skończonymi. Dla $n = 2$, prawo sumy (b) z § 5.1 stwierdza, że

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|.$$

Dla $n = 3$, sformułowana poniżej zasada włączeń i wyłączeń mówi, że

$$\begin{aligned} |A_1 \cup A_2 \cap A_3| = {} & |A_1| + |A_2| + |A_3| \\ & - \{|A_1 \cap A_2| + |A_1 \cap A_3| + |A_2 \cap A_3|\} \\ & + |A_1 \cap A_2 \cap A_3|, \end{aligned}$$

a dla $n = 4$ daje równość

$$\begin{aligned} |A_1 \cup A_2 \cup A_3 \cup A_4| = {} & |A_1| + |A_2| + |A_3| + |A_4| \\ & - \{|A_1 \cap A_2| + |A_1 \cap A_3| + |A_1 \cap A_4| + |A_2 \cap A_3| + |A_2 \cap A_4| \\ & \qquad + |A_3 \cap A_4|\} \\ & + \{|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| \\ & \qquad + |A_2 \cap A_3 \cap A_4|\} \\ & - |A_1 \cap A_2 \cap A_3 \cap A_4|. \end{aligned}$$

A oto słowne sformułowanie zasady w przypadku ogólnym. Wersja z symbolami podana jest w ćwiczeniu 6.

Zasada włączeń i wyłączeń

> Aby znaleźć liczbę elementów zbioru $A_1 \cup A_2 \cup \ldots \cup A_n$, znajdź liczby elementów wszystkich możliwych przecięć zbiorów spośród $\{A_1, A_2, \ldots, A_n\}$, dodaj do siebie wyniki uzyskane dla przecięć nieparzystej liczby zbiorów, a następnie odejmij wyniki uzyskane dla przecięć parzystej liczby zbiorów.

Używając słów wziętych z nazwy zasady włączeń i wyłączeń powiemy, że należy „włączyć", czyli dodać do siebie liczności poszczególnych zbiorów, następnie „wyłączyć", czyli odjąć liczności wszystkich przecięć po dwa zbiory, potem „włączyć", czyli dodać liczności wszystkich przecięć po trzy zbiory itd.

PRZYKŁAD 1 Obliczymy teraz, ile jest liczb całkowitych w zbiorze $S = \{1, 2, 3, \ldots, 2000\}$, które są podzielne przez 9, 11, 13 lub 15. Dla każdego $k \in \mathbb{P}$ niech

$$D_k = \{n \in S\colon n \text{ jest podzielne przez } k\}.$$

Szukamy $|D_9 \cup D_{11} \cup D_{13} \cup D_{15}|$. Zauważmy, że $|D_k|$ jest największą liczbą całkowitą $\leq \lfloor 2000/k \rfloor$, a więc możemy wykonać następujące obliczenia

$$|D_9| = 222, \quad |D_{11}| = 181, \quad |D_{13}| = 153, \quad |D_{15}| = 133,$$
$$|D_9 \cap D_{11}| = |D_{99}| = 20, \quad |D_9 \cap D_{13}| = |D_{117}| = 17,$$
$$|D_9 \cap D_{15}| = |D_{45}| = 44, \quad |D_{11} \cap D_{13}| = |D_{143}| = 13,$$
$$|D_{11} \cap D_{15}| = |D_{165}| = 12, \quad |D_{13} \cap D_{15}| = |D_{195}| = 10,$$
$$|D_9 \cap D_{11} \cap D_{13}| = |D_{1287}| = 1, \quad |D_9 \cap D_{11} \cap D_{15}| = |D_{495}| = 4,$$
$$|D_9 \cap D_{13} \cap D_{15}| = |D_{585}| = 3, \quad |D_{11} \cap D_{13} \cap D_{15}| = |D_{2145}| = 0,$$
$$|D_9 \cap D_{11} \cap D_{13} \cap D_{15}| = |D_{6435}| = 0.$$

Zauważmy, że $D_9 \cap D_{15} = D_{45}$ (a nie D_{135}), ponieważ NWW $(9, 15) = 45$; podobna ostrożność jest konieczna w przypadku $D_9 \cap D_{11} \cap D_{15}$, $D_9 \cap D_{13} \cap D_{15}$ itd. Teraz na mocy zasady włączeń i wyłączeń mamy

$$\begin{aligned} |D_9 \cup D_{11} \cup D_{13} \cup D_{15}| = {} & 222 + 181 + 153 + 133 \\ & - (20 + 17 + 44 + 13 + 12 + 10) \\ & + (1 + 4 + 3 + 0) - 0 = 581. \end{aligned}$$

■

Zasada włączeń i wyłączeń nadaje się idealnie do sytuacji, w których:

(a) chcemy jedynie znać wielkość zbioru $A_1 \cup \ldots \cup A_n$, bez wypisania jego elementów
oraz
(b) liczby elementów wielokrotnych przecięć łatwo daje się obliczyć.

W przykładzie 1 pokazany został taki właśnie problem.

Jeśli potrzebujemy listy elementów zbioru $A_1 \cup \ldots \cup A_n$, to możemy użyć algorytmu iteracyjnego do wypisania najpierw elementów A_1, następnie $A_2 \backslash A_1$, potem $A_3 \backslash (A_1 \cup A_2)$ i tak dalej, albo też możemy wypisać elementy A_1, a następnie, rekurencyjnie, elementy $(A_2 \backslash A_1) \cup \ldots \cup (A_n \backslash A_1)$. Przyjrzyj się ponownie przykładowi 1, aby zobaczyć, jakich obliczeń wymagałoby posłużenie się każdą z tych metod w celu otrzymania listy elementów zbioru $D_9 \cup D_{11} \cup D_{13} \cup D_{15}$. Z pewnością przydałaby się tu pomoc komputera i przyjaznej struktury danych. Jeśli jednak chcemy znać jedynie liczbę elementów, a nie ich listę, to zasada włączeń i wyłączeń czyni nasze zadanie w miarę bezbolesnym.

PRZYKŁAD 2 Wybieramy losowo liczbę ze zbioru $T=\{1000, 1001, \ldots, 9999\}$. Znajdziemy prawdopodobieństwo tego, że wśród cyfr naszej liczby co najmniej raz występuje 0, co najmniej raz 1 i co najmniej raz 2. Na przykład, takimi liczbami są 1072 i 2101. Ponieważ łatwiej jest obliczyć, ile jest liczb, ktore nie zawierają pewnych cyfr, zajmiemy się dopełnieniami. To znaczy dla $k = 0, 1$ i 2 definiujemy

$$A_k = \{n \in T\colon \text{ w } n \text{ nie występuje cyfra } k\}.$$

Wówczas każdy ze zbiorów A_k^c składa się z tych liczb n ze zbioru T, których co najmniej jedną z cyfr jest k, a zatem zbiór $A_0^c \cap A_1^c \cap A_2^c$ składa się z tych liczb n ze zbioru T, które wśród swych cyfr mają co najmniej raz 0, 1 i 2. A to jest dokładnie ten zbiór, którego wielkość nas interesuje.

Ponieważ na mocy prawa De Morgana mamy $A_0^c \cap A_1^c \cap A_2^c = (A_0 \cup A_1 \cup A_2)^c$, obliczymy najpierw $|A_0 \cup A_1 \cup A_2|$ używając zasady włączeń i wyłączeń. Zgodnie z prawem iloczynu (a) z § 5.1 mamy $|A_1| = 8 \cdot 9 \cdot 9 \cdot 9$, ponieważ jest 8 możliwości wyboru pierwszej cyfry, którą nie może być ani 0 ani 1, i 9 możliwości dla każdej z pozostałych cyfr. Podobne rachunki dają:

$$|A_0| = 9 \cdot 9 \cdot 9 \cdot 9 = 6561, \quad |A_1| = |A_2| = 8 \cdot 9 \cdot 9 \cdot 9 = 5832,$$
$$|A_0 \cap A_1| = |A_0 \cap A_2| = 8 \cdot 8 \cdot 8 \cdot 8 = 4096,$$
$$|A_1 \cap A_2| = 7 \cdot 8 \cdot 8 \cdot 8 = 3584,$$
$$|A_0 \cap A_1 \cap A_2| = 7 \cdot 7 \cdot 7 \cdot 7 = 2401.$$

Na mocy zasady włączeń i wyłączeń,

$$|A_0 \cup A_1 \cup A_2| = 6561 + 5832 + 5832 - (4096 + 4096 + 3584) + 2401 = 8850,$$

a stąd

$$|(A_0 \cup A_1 \cup A_2)^c| = |T| - |A_0 \cup A_1 \cup A_2| = 9000 - 8850 = 150.$$

W zbiorze T jest więc 150 liczb całkowitych, w których zapisie co najmniej raz występuje każda z cyfr 0, 1 i 2. Zatem prawdopodobieństwo rozważanego zdarzenia wynosi $\frac{150}{|T|} = \frac{150}{9000} = \frac{1}{60}$. ■

Uzasadnienie zasady włączeń i wyłączeń. Główną przeszkodą na drodze do dowodu naszej zasady w przypadku ogólnym są oznaczenia (zob. ćwiczenie 6). Dowód można przeprowadzić przez indukcję po n. Pokażemy, jak prawdziwość dowodzonego faktu dla $n = 2$ pociąga za sobą jego prawdziwość dla $n = 3$.

Wykorzystując przypadek $n = 2$, mamy

$$|A \cup B \cup C| = |A \cup B| + |C| - |(A \cup B) \cap C| \tag{1}$$

oraz

$$|A \cup B| = |A| + |B| - |A \cap B|. \tag{2}$$

Stosując prawa rozdzielności dla sum i przecięć (prawo 3b w tablicy 1 z § 1.2), otrzymujemy też

$$\begin{aligned} |(A \cup B) \cap C| &= |(A \cap C) \cup (B \cap C)| \\ &= |A \cap C| + |B \cap C| - |A \cap B \cap C|. \end{aligned} \tag{3}$$

Podstawienie (2) i (3) do (1) daje

$$\begin{aligned} |A \cup B \cup C| = |A| + |B| - |A \cap B| + |C| - |A \cap C| \\ - |B \cap C| + |A \cap B \cap C|, \end{aligned} \tag{4}$$

a to jest nasza zasada dla $n = 3$. ∎

Następne twierdzenie jest Czytelnikowi prawdopodobnie znane z algebry. Ma ono wiele zastosowań, a ponieważ wyrażenie $a + b$ jest dwumianem, wyjaśnia ono też, czemu symbol $\binom{n}{r}$ nazwaliśmy „współczynnikiem dwumianowym".

Wzór dwumianowy

Dla dowolnych liczb rzeczywistych a i b oraz każdego $n \in \mathbb{N}$ mamy

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}.$$

Dowód. Twierdzenie to można udowodnić indukcyjnie, opierając się na zależności rekurencyjnej

$$\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r} \quad \text{dla} \quad 1 \leq r \leq n;$$

zob. ćwiczenie 12. Tej zależności można z kolei dowieść za pomocą pewnych algebraicznych przekształceń. Zamiast tego, podamy wyjaśnienie teoriomnogościowe, w duchu teorii zliczania.

Istnieje $\binom{n+1}{r}$ r-elementowych podzbiorów zbioru $\{1, 2, \ldots, n, n+1\}$. Rozdzielmy je na dwie klasy. Jest $\binom{n}{r}$ podzbiorów, do których należą wyłacznie elementy zbioru $\{1, 2, \ldots, n\}$. Każdy z pozostałych podzbiorów składa się z liczby $n + 1$ i pewnych

$r-1$ elementów zbioru $\{1, 2, \ldots, n\}$. Ponieważ jest $\binom{n}{r-1}$ sposobów wyboru elementów, które są różne od $n+1$, podzbiorów tego rodzaju jest $\binom{n}{r-1}$. Zatem r-elementowych podzbiorów zbioru $\{1, 2, \ldots, n, n+1\}$ jest dokładnie $\binom{n}{r} + \binom{n}{r-1}$, a więc

$$\binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r},$$

tak jak twierdziliśmy.

Jednakże prawdziwym powodem tego, że liczba $\binom{n}{r}$ jest współczynnikiem wyrazu $a^r b^{n-r}$ w rozwinięciu $(a+b)^n$ jest fakt, że jest to liczba tych spośród iloczynów uzyskanych po wykonaniu mnożeń potrzebnych do obliczenia $(a+b)^n$, w których a występuje r-krotnie. Przeanalizuj następujące działania:

$$\begin{aligned}(a+b)^2 &= (a+b)(a+b) = aa + (ab+ba) + bb,\\ (a+b)^3 &= (a+b)(a+b)(a+b)\\ &= aaa + (aab+aba+baa) + (abb+bab+bba) + bbb.\end{aligned}$$

W rozwinięciu $(a+b)^3$ są $3 = \binom{3}{2}$ wyrazy, w których a występuje dwukrotnie, a które odpowiadają wyborom (litery wyróżnione)

$$(\mathbf{a}+b)(\mathbf{a}+b)(a+\mathbf{b}),\ (\mathbf{a}+b)(a+\mathbf{b})(\mathbf{a}+b)\ \text{ i }\ (a+\mathbf{b})(\mathbf{a}+b)(\mathbf{a}+b).$$

W ogólnym przypadku obliczania $(a+b)^n$ każdy wybór tych r spośród czynników $(a+b)$, z których weźmiemy a, daje iloczyn, w którym a występuje r razy (a zatem b występuje $n-r$ razy). Ponieważ jest $\binom{n}{r}$ możliwych wyborów, więc liczba takich wyrazów wynosi $\binom{n}{r}$. ∎

Wzór dwumianowy przydaje się czasem do obliczania wartości sum takich, jak $\sum_{r=0}^{n}(-1)^r\binom{n}{r}$ (ćwiczenie 10), $\sum_{r=0}^{n}\binom{n}{r}$ (ćwiczenie 13) oraz $\sum_{r=0}^{n}\binom{n}{r}2^r$ (ćwiczenie 14). Można go też stosować do obliczania poszczególnych współczynników, często kilku pierwszych, gdy nie chcemy wypisywać dokładnego wyniku potęgowania.

PRZYKŁAD 3 Pierwszych kilka wyrazów wielomianu $(2+x)^{100}$ to

$$2^{100} + \binom{100}{1}2^{99}x + \binom{100}{2}2^{98}x^2 + \ldots = 2^{100} + 100\cdot 2^{99}x + 4950\cdot 2^{98}x^2 + \ldots$$

Jeśli liczba x jest bardzo bliska 0, to kuszące jest, żeby powiedzieć, iż skoro wartości x^3, $x^4, \ldots$ są malutkie, to można na dobrą sprawę pominąć wyrazy występujące po x^2. Uwaga! Współczynniki występujące w środku wyrażenia, takie jak $\binom{100}{50}$, mogą być całkiem duże i pominięte wyrazy, dodane do siebie,

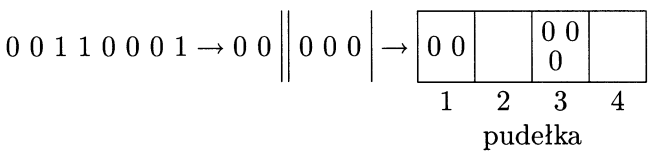
mogą dać duży wynik. W naszym przypadku, jeśli $x = 0{,}01$, to odrzucając te wyższe potęgi, otrzymamy wartość $(2,01)^{100}$ z 1,4-procentowym błędem. Jednakże dla $x = 0,1$ nasz „przybliżony" wynik wynosiłby mniej niż 5 procent prawdziwej wartości.■

Nasza następna zasada zliczania może być stosowana w wielu różnorodnych sytuacjach. Przedstawiamy ją w formie, którą łatwo zapamiętać.

Zasada rozmieszczania przedmiotów w pudełkach

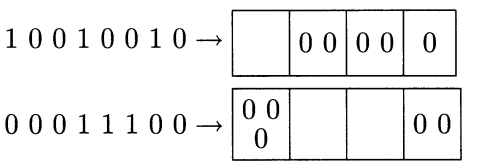
Jest $\binom{n+k-1}{k-1}$ sposobów rozmieszczenia n identycznych przedmiotów w k rozróżnialnych pudełkach.

Dowód. Dowód jest elegancki i zarazem wiele wyjaśniający; przedstawimy go dla przypadku $n = 5$ i $k = 4$. Przedmioty reprezentowane będą przez pięć zer, a trzy jedynki posłużą do rozdzielenia przedmiotów do czterech pudełek. Twierdzimy, że istnieje odpowiedniość wzajemnie jednoznaczna między ciągami złożonymi z pięciu zer i trzech jedynek, a sposobami rozmieszczenia tych pięciu zer w trzech pudełkach. Mianowicie, dany ciąg odpowiada umieszczeniu wszystkich zer występujących przed pierwszą jedynką w pierwszym pudełku, zer, które leżą między pierwszą a drugą jedynką w drugim pudełku, zer, które leżą między drugą a trzecią jedynką w trzecim pudełku, a zer występujących po ostatniej jedynce w pudełku czwartym. Na przykład



W tym przypadku pudełka 2 i 4 są puste, ponieważ nie ma zer między pierwszą a drugą jedynką oraz nie ma zer po ostatniej jedynce.

Inne przykłady:



Ciągów, które mają po pięć zer i trzy jedynki jest $\binom{8}{3}$. Ponieważ $\binom{8}{3} = \binom{5+4-1}{4-1}$, więc otrzymujemy dowodzony rezultat dla $n = 5$ i $k = 4$.

W przypadku ogólnym rozpatrujemy ciągi złożone z n zer i $k-1$ jedynek. Zera odpowiadają przedmiotom, a jedynki służą do ich rozdzielania. Takich ciągów jest $\binom{n+k-1}{k-1}$ i, jak poprzednio, istnieje odpowiedniość wzajemnie jednoznaczna między tymi ciągami a rozmieszczeniami n zer w k pudełkach. ∎

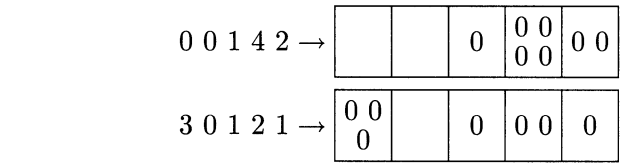
PRZYKŁAD 4 (a) Na ile sposobów dziesięć identycznych czerwonych kulek można umieścić w pięciu rozróżnialnych torbach? Tutaj $n = 10$, $k = 5$ i odpowiedź brzmi

$$\binom{10+5-1}{5-1} = \binom{14}{4} = 1001.$$

(b) Na ile sposobów dziesięć identycznych czerwonych kulek można umieścić w pięciu nierozróżnialnych torbach? To pytanie jest znacznie trudniejsze. Należy zdawać sobie sprawę z tego, że w problemach zliczania drobna modyfikacja może prowadzić do trudnego zagadnienia. Tutaj chciałoby się zastosować w jakiś sposób część (a). Jednakże nawet przy wykorzystaniu metod następnego paragrafu nie ma naturalnego sposobu zrobienia tego. Zostawiamy ten problem. Wszystkie rozwiązania, znane autorom, związane są z rozważaniem wielu przypadków. ∎

Czasami problemy wymagają odpowiedniego przeformułowania, zanim stanie się jasne, że do ich rozwązania można wykorzystać nasze zasady zliczania.

PRZYKŁAD 5 Ile liczb ze zbioru $\{1, 2, 3, \ldots, 100000\}$ ma tę własność, że suma ich cyfr wynosi 7? Możemy pominąć ostatnią z liczb, 100000, i założyć, dopisując zera na początku, jeśli trzeba, że wszystkie rozważane liczby mają po pięć cyfr. Tak więc, na przykład, zastępujemy 1 ciągiem 00001 i 73 ciągiem 00073. Nasze pytanie brzmi teraz tak: ile ciągów złożonych z pięciu cyfr ma tę własność, że suma tych cyfr wynosi 7? Z każdym takim ciągiem możemy w jednoznaczny sposób skojarzyć rozmieszczenie siedmiu kul w pięciu pudełkach; na przykład,



Ponieważ takich rozmieszczeń jest $\binom{11}{4} = 330$, więc jest 330 liczb o żądanej własności. ∎

Podamy teraz trochę inną interpretację zasady rozmieszczania przedmiotów w pudełkach. Rozważmy najpierw k danych pudełek i załóżmy, że każde pudełko zawiera nieograniczoną liczbę przedmiotów zaopatrzonych w etykiety oznaczające to pudełko, w którym przedmioty te leżą. Stosując naszą zasadę do tej odwróconej sytuacji widzimy, że jest $\binom{n+k-1}{k-1}$ sposobów wyjęcia n przedmiotów z naszych k pudełek. Innymi słowy,

> Liczba sposobów wyboru zbioru n przedmiotów k rozróżnialnych typów, przy założeniu, że dopuszczalne są powtórzenia, wynosi
>
> $$\binom{n+k-1}{k-1}.$$

PRZYKŁAD 6 Na ile sposobów można wybrać dziesięć monet, mając nieograniczony zapas groszy oraz pięcio-, dziesięcio- i pięćdziesięciogroszówek? Ten przykład pasuje jak ulał do sformułowanej właśnie zasady. Niech $n = 10$ (ponieważ jest dziesięć monet) i $k = 4$ (gdyż są cztery rodzaje monet). Wtedy odpowiedź brzmi:

$$\binom{10+4-1}{4-1} = \binom{13}{3} = 286.$$

Można uniknąć użycia nowej interpretacji w następujący sposób. Problem jest równoważny z zagadnieniem zliczania czterowyrazowych ciągów złożonych z nieujemnych liczb całkowitych, których suma wynosi 10. Na przykład ciąg (5, 3, 0, 2) odpowiada wyborowi pięciu groszy, trzech pięciogroszówek i dwóch pięćdziesięciogroszówek. Zliczanie tych czterowyrazowych ciągów jest równoważne ze zliczaniem sposobów rozmieszczenia dziesięciu nierozróżnialnych przedmiotów w 4 pudełkach, a takich sposobów jest $\binom{13}{3}$. ∎

ĆWICZENIA DO § 5.3

1. Wśród 200 osób, 150 uprawia pływanie lub jogging, lub i jedno, i drugie. Jeśli 85 osób uprawia pływanie, a 60 pływanie i jogging, to ile uprawia jogging?
2. Niech $S = \{100, 101, 102, \ldots, 999\}$, a więc $|S| = 900$.
   (a) Ile liczb ze zbioru S ma co najmniej jedną z cyfr równą 3 lub 7? Przykłady: 300, 707, 736, 103, 997.

(b) Ile liczb ze zbioru S ma co najmniej jedną z cyfr równą 3 i co najmniej jedną z cyfr równą 7? Przykłady: 736 i 377, ale nie 300, 707, 103, 997.

3. Wybieramy losowo liczbę całkowitą ze zbioru $\{1, 2, 3, \ldots, 1000\}$. Ile wynosi prawdopodobieństwo, że jest ona podzielna przez 4, 5 lub 6?

4. Pewien inwestor ma 7 przekazów gotówkowych, po 1000 złotych każdy, które chce przesłać pocztą 3 funduszom powierniczym.
(a) Na ile sposobów może zainwestować swoje pieniądze?
(b) Na ile sposobów może zainwestować swoje pieniądze, jeśli każdy fundusz wymaga wpłaty co najmniej 1000 zł?

5. Wybieramy losowo liczbę całkowitą ze zbioru $\{1, 2, 3, \ldots, 1000\}$. Ile wynosi prawdopodobieństwo, że liczba ta jest
(a) podzielna przez 7?
(b) podzielna przez 11?
(c) niepodzielna przez 7 lub 11?
(d) podzielna przez 7 lub 11, ale nie przez obie te liczby?

6. Rozważmy zbiory skończone $\{A_1, A_2, \ldots, A_n\}$. Niech $\mathcal{P}_+(n)$ oznacza zbiór wszystkich niepustych podzbiorów I zbioru $\{1, 2, \ldots, n\}$. Wykaż, że zasada włączeń i wyłączeń stwierdza, że

$$\Big|\bigcup_{i=1}^{n} A_i\Big| = \sum_{I\in\mathcal{P}_+(n)} (-1)^{|I|+1} \cdot \Big|\bigcap_{i\in I} A_i\Big|.$$

7. Dwanaście identycznych listów ma zostać wrzuconych do czterech różnych skrzynek pocztowych.
(a) Na ile sposobów można to zrobić?
(b) Ile jest możliwych sposobów, jeśli do każdej ze skrzynek muszą trafić co najmniej dwa listy?

8. Ile można otrzymać różnych mieszanek po 10 cukierków, jeśli mamy do dyspozycji 4 rodzaje cukierków w nieograniczonych ilościach?

9. Posługując się wzorem dwumianowym, znajdź rozwinięcia następujących wyrażeń:
(a) $(x + 2y)^4$,
(b) $(x - y)^6$,
(c) $(3x + 1)^4$,
(d) $(x + 2)^5$.

10. Udowodnij, że $\sum_{r=0}^{n}(-1)^r\binom{n}{r} = 0$ dla $n \in \mathbb{P}$. *Wskazówka*: zastosuj wzór dwumianowy.

11. (a) Udowodnij, że $\binom{n}{r} = \binom{n}{n-r}$ dla $0 \leq r \leq n$.
(b) Podaj teoriomnogościową interpretację tożsamości z części (a).

12. (a) Udowodnij wzór dwumianowy.
(b) Podaj algebraiczny dowód równości $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$ dla $1 \leq r \leq n$.

13. Udowodnij, że $2^n = \sum_{r=0}^{n} \binom{n}{r}$
(a) podstawiając do wzoru dwumianowego $a = b = 1$,
(b) zliczając podzbiory zbioru n-elementowego,
(c) przez indukcję, używając zależności rekurencyjnej z ćwiczenia 12(b).

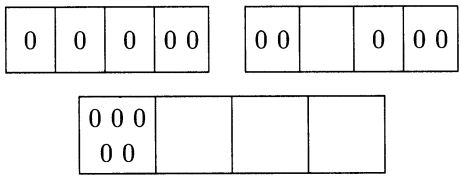
14. Udowodnij, że $\sum_{r=0}^{n} \binom{n}{r} 2^r = 3^n$ dla $n \in \mathbb{P}$.

15. (a) Sprawdź dla kilku małych wartości m i n takich, jak np. $m = 3$ i $n = 5$, że $\sum_{k=m}^{n} \binom{k}{m} = \binom{n+1}{m+1}$.
(b) Udowodnij powyższą tożsamość przez indukcję po n, dla $n \geq m$.
(c) Udowodnij tę tożsamość zliczając $(m+1)$-elementowe podzbiory zbioru $\{1, 2, \ldots, n+1\}$. *Wskazówka*: Ile jest wśród nich takich podzbiorów, których największym elementem jest $k+1$? Jakie wartości może przyjmować k?

16. W dowodzie zasady rozmieszczania przedmiotów w pudełkach wskazaliśmy odpowiedniość wzajemnie jednoznaczną między ciągami złożonymi z pięciu zer i z trzech jedynek, a rozmieszczeniami pięciu przedmiotów w czterech pudełkach.
(a) Znajdź rozmieszczenia, które odpowiadają następującym ciągom:

1 0 1 0 1 0 0 0, 0 1 0 0 1 0 0 1, 1 0 0 0 0 0 1 1, 1 1 1 0 0 0 0 0.

(b) Znajdź ciągi, które odpowiadają następującym rozmieszczeniom:



17. (a) Na ile sposobów można rozmieścić 14 przedmiotów w 3 pudełkach tak, by w jednym z pudełek znalazło się co najmniej 8 przedmiotów?
(b) (a) Na ile sposobów można rozmieścić 14 przedmiotów w 3 pudełkach tak, by w żadnym z pudełek nie znalazło się więcej niż 7 przedmiotów?
(c) Ile jest liczb między 0 a 999, których suma cyfr jest równa 20? *Wskazówka*: każda z cyfr musi być równa co najmniej 2; można zastosować część (b).

§ 5.4. Zliczanie i podziały

Przypomnij sobie, że podział zbioru S to rodzina jego parami rozłącznych podzbiorów, których sumą jest sam zbiór S. W tym paragrafie skoncentrujemy się na problemach związanych z podziałami.

Nasza pierwsza elementarna obserwacja dotyczy tych szczególnych podziałów, których wszystkie bloki mają po tyle samo elementów. Jest ona konsekwencją prawa iloczynu (b), bądź prawa sumy i jest po prostu formalizacją oczywistych metod postępowania. Żeby obliczyć, ile jest owiec w stadzie, po prostu policz ich nogi i podziel przez 4. Inny przykład: jeśli pudełko zawiera 30 różnokolorowych kulek, wśród których jest po 6 kulek każdego z kolorów, to różnych kolorów musi być 5. Tę ostatnią obserwację można traktować jak szczególny przypadek następnego lematu; wystarczy przyjąć, że A jest zbiorem kulek, B — zbiorem kolorów, ψ — funkcją, która każdej kulce przyporządkowuje jej kolor i $r = 6$.

Lemat o zliczaniu

Jeśli $\psi\colon A \to B$ odwzorowuje skończony zbiór A na zbiór B i jeśli wszystkie zbiory postaci

$$\psi^{\leftarrow}(b) = \{a \in A\colon \psi(a) = b\}$$

dla $b \in B$ mają po tyle samo, powiedzmy po r, elementów, to

$$|B| = \frac{|A|}{r}.$$

Dowód. Zbiór A jest sumą parami rozłącznych zbiorów postaci $\psi^{\leftarrow}(b)$, więc na mocy prawa sumy

$$|A| = \sum_{b \in B} |\psi^{\leftarrow}(b)| = \sum_{b \in B} r = r \cdot |B|.$$

Stąd $|B| = |A|/r$.

Alternatywny argument otrzymamy uzmysławiając sobie, że każdy spośród $|A|$ elementów zbioru A można uzyskać wybierając najpierw na $|B|$ sposobów element $b \in B$, a następnie wskazując jeden z r elementów a, dla których $\psi(a) = b$. Stąd, ponownie, $|A| = |B| \cdot r$. ∎

PRZYKŁAD 1

Punktem wyjścia do tego zadania jest kapelusz zawierający karteczki z napisanymi na nich literami. Karteczki są pojedynczo wyjmowane z kapelusza i układane w rzędzie w takiej kolejności, w jakiej zostały wyciągnięte. Pytanie brzmi: ile różnych słów można w ten sposób otrzymać? Dwa słowa uważamy tu za różne, jeśli na co najmniej jednym miejscu mają one różne litery.

Jeśli wszystkie litery w kapeluszu są różne, to zadanie jest łatwe. Na przykład, jeśli kapelusz zawiera dziesięć liter I, M, P, O, R, T, A, N, C, E , to można z nich otrzymać 10! słów,

w większości bezsensownych, a wśród nich dokładnie na jeden sposób słowo IMPORTANCE.

Jeśli jednak kapelusz zawiera osiem liter E, E, N, N, N, O, S, S, to jest 8! sposobów wyciągnięcia ich z kapelusza, ale to samo słowo powstać może na więcej niż jeden sposób. W celu przeanalizowania tej sytuacji, zaopatrzmy powtarzające się litery w indeksy; kapelusz zawiera teraz litery $\mathrm{E}_1, \mathrm{E}_2, \mathrm{N}_1, \mathrm{N}_2, \mathrm{N}_3, \mathrm{O}, \mathrm{S}_1$ i S_2. Jest 8! permutacji tych poindeksowanych liter. Jeśli ψ oznacza funkcję, która każdej z tych permutacji przyporządkowuje ją samą, ale z wymazanymi indeksami, to zbiorem wartości funkcji ψ jest zbiór wszystkich słów złożonych z liter E, E, N, N, N, O, S, S. Na przykład

$$\psi(\mathrm{N_3ON_1S_2E_1N_2S_1E_2}) = \mathrm{NONSENSE}.$$

Ilu permutacjom funkcja ψ przyporządkowuje słowo NONSENSE? Litery $\mathrm{N}_1, \mathrm{N}_2, \mathrm{N}_3$ mogą się pojawić w różnym porządku na 3! sposobów, S_1 i S_2 – na 2! sposobów oraz E_1 i E_2 – też na 2! sposobów. Tak więc są $3! \cdot 2! \cdot 2! = 24$ różne permutacje, które dają słowo NONSENSE. To oznacza, że przeciwobraz $\psi^{\leftarrow}(\mathrm{NONSENSE})$ ma 24 elementy. Podobnie, są 24 permutacje, które dają SNEENONS, bądź dowolne inne słowo złożone z danych liter. Lemat o zliczaniu mówi nam, że całkowita liczba rozróżnialnych słów wynosi

$$\frac{8!}{3! \cdot 2! \cdot 2!} = 1680.$$

Zatem, jeśli losowo wyciągamy karteczki z kapelusza, to prawdopodobieństwo otrzymania słowa NONSENSE wynosi 1/1680 (a stąd prawdopodobieństwo otrzymania słowa bezsensownego wynosi 1679/1680). ∎

Rozumowanie użyte w powyższym przykładzie może posłużyć do dowodu następującego ogólnego faktu.

Zliczanie permutacji

Przypuśćmy, że zbiór złożony z n przedmiotów został podzielony na k podzbiorów mających, odpowiednio, po n_1, $n_2, \ldots$, n_k elementów; tak więc $n = n_1 + n_2 + \ldots + n_k$. Powiemy, że dwa przedmioty są tego samego typu, jeśli należą do tego samego bloku rozważanego podziału. Dwie permutacje naszego zbioru uznajemy za rozróżnialne, jeśli na co najmniej jednym miejscu stoją w nich elementy różnych typów. Wtedy liczba rozróżnialnych permutacji naszego zbioru wynosi

$$\frac{n!}{n_1!n_2! \cdot \ldots \cdot n_k!}.$$

PRZYKŁAD 2 Niech $\Sigma = \{a, b, c\}$. Liczba słów w Σ^* o długości 10, złożonych z 4 liter a, 3 liter b i 3 liter c wynosi

$$\frac{10!}{4!3!3!} = 4200.$$

Liczba słów złożonych z pięciu liter a, trzech liter b i dwóch liter c wynosi

$$\frac{10!}{5!3!2!} = 2520,$$

a liczba słów złożonych z pięciu liter a i pięciu liter b jest równa

$$\frac{10!}{5!5!} = 252.$$

Dla porównania odnotujmy, że w zbiorze Σ^{10} jest $3^{10} = 59049$ słów. ∎

Zliczaliśmy przed chwilą permutacje, rozróżnialne za pomocą pewnego podziału, teraz jednak rozpoczniemy zliczanie samych podziałów. **Podziałem uporządkowanym** zbioru S nazywamy ciąg $(A_1, A_2, \ldots, A_k)$, którego elementy $A_1, A_2, \ldots, A_k$ tworzą podział zbioru S. Nie zakładamy, że elementy zbiorów A_i są ustawione w jakiejś kolejności, istotna jest natomiast kolejność, w jakiej występują same zbiory A_i.

PRZYKŁAD 3 Niech $S = \{1, 2, 3, 4, 5, 6, 7, 8\}$.

(a) Oto niektóre podziały uporządkowane zbioru S:
$(\{1,3,5\}, \{2,4,6,7,8\})$, $(\{2,4,6,7,8\}, \{1,3,5\})$,
$(\{3,6\}, \{2,5,8\}, \{1,4,7\})$, $(\{1\}, \{2,4,6,8\}, \{3,5,7\})$,
$(\{1,6\}, \{2,5,8\}, \{3,4\}, \{7\})$ i $(\{6,1\}, \{2,5,8\}, \{4,3\}, \{7\})$.
Ostatnie dwa podziały uporządkowane są identyczne, gdyż $\{1,6\} = \{6,1\}$ oraz $\{3,4\} = \{4,3\}$, ale wszystkie pozostałe są parami różne.

(b) Znajdziemy liczbę podziałów uporządkowanych zbioru S, które są postaci (A, B, C, D), gdzie $|A| = 2$, $|B| = 3$, $|C| = 1$ i $|D| = 2$. Na zagadnienie to można spojrzeć jak na problem zliczania rozróżnialnych permutacji 8 elementów 4 typów, czyli jak na „problem liter w kapeluszu". A oto jak się to robi.

Rozważmy podział uporządkowany (A, B, C, D) zbioru $\{1, 2, 3, 4, 5, 6, 7, 8\}$ na podzbiory mające, odpowiednio, po 2, 3, 1 i 2 elementy. Weźmy, na przykład, $(\{5,8\}, \{1,3,6\}, \{2\}, \{4,7\})$. Wtedy 5 i 8 są typu A, podczas gdy 1, 3 i 6 są typu B itd. Jeśli wypiszemy po kolei typy liczb $1, 2, \ldots, 8$, to otrzymamy ciąg B, C, B, D, A, B, D, A, który w pełni opisuje nasz podział pod warunkiem, iż wiemy, jak ten ciąg odkodować. Jeśli zamiast

oznaczać bloki podziału literami A, B, C, D, oznaczymy je przez E, N, O i S, to $(\{5,8\}, \{1,3,6\}, \{2\}, \{4,7\})$ odpowiada znajomej sekwencji N, O, N, S, E, N, S, E.

To rozumowanie pokazuje, że podziały uporządkowane zbioru $\{1, 2, 3, 4, 5, 6, 7, 8\}$ na bloki mające po 2, 3, 1 i 2 elementy, są we wzajemnie jednoznacznej odpowiedniości z rozróżnialnymi permutacjami złożonymi z liter A, A, B, B, B, C, D, D. Tak więc liczba takich podziałów uporządkowanych wynosi $\frac{8!}{2!\cdot 3!\cdot 1!\cdot 2!} = 1680$. ∎

Przez łatwe uogólnienie rozumowania z przykładu 3(b) otrzymujemy następującą zasadę zliczania.

Zliczanie podziałów uporządkowanych

Jeśli dany zbiór ma n elementów i jeśli $n_1 + n_2 + \ldots + n_k = n$, to istnieje

$$\frac{n!}{n_1!n_2! \cdot \ldots \cdot n_k!}$$

podziałów uporządkowanych $(A_1, A_2, \ldots, A_k)$ tego zbioru takich, że $|A_i| = n_i$ dla $i = 1, 2, \ldots, k$.

Wyrażenie występujące w powyższym wzorze można też zapisać w postaci

$$\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \ldots \cdot \binom{n-n_1-\ldots-n_{k-1}}{n_k}.$$

Ta postać wzoru ma naturalną interpretację. Wybieramy najpierw z całego n-elementowego zbioru n_1 elementów, które utworzą A_1, następnie n_2 spośród pozostałych $n - n_1$ elementów, które utworzą A_2, i tak dalej, biorąc elementy dla każdego A_i z $(A_1 \cup \ldots \cup A_{i-1})^c$. Ostatnim czynnikiem jest $\binom{n_k}{n_k}$, więc możemy go pominąć. Potencjalną zaletą powyższej iloczynowej postaci wzoru jest to, że jeśli n jest bardzo duże a $n!$ olbrzymie, to współczynniki dwumianowe występujące w iloczynie są znacznie mniejsze i można je pewnie łatwiej obliczyć.

PRZYKŁAD 4

Na ile sposobów da się utworzyć trzy rozłączne komisje z osób wybranych z dwudziestoosobowej grupy, jeśli muszą one mieć, odpowiednio, 3, 5 i 7 członków? Problem ten jest równoważny z zagadnieniem zliczania podziałów uporządkowanych (A, B, C, D) danego zbioru dwudziestu osób, gdzie $|A| = 3$, $|B| = 5$, $|C| = 7$ i $|D| = 5$. Zbiór D składa się z osób, które nie należą do żadnej

komisji. Jest

$$\frac{20!}{3!5!7!5!} = \binom{20}{3} \cdot \binom{17}{5} \cdot \binom{12}{7} \cdot \binom{5}{5}$$
$$= 1140 \cdot 6188 \cdot 792 \cdot 1 \approx 5,587 \cdot 10^9$$

możliwych sposobów utworzenia takich komisji. Zwróć uwagę, że choć $|D| = |B|$, to komisja B i zbiór D odgrywają różne role; kolejność bloków podziału jest istotna i interesuje nas coś więcej, niż tylko sposoby rozbicia naszego zbioru na jeden trzyelementowy, dwa pięcioelementowe i jeden siedmioelementowy podzbiór. ∎

PRZYKŁAD 5 (a) Rozdanie brydżowe to podział uporządkowany talii 52 kart na cztery zbiory, po 13 kart w każdym. Jest zatem

$$\frac{52!}{13!13!13!13!} = \frac{52!}{(13!)^4} \approx 5,3645 \cdot 10^{28}$$

rozdań brydżowych.

(b) Znajdziemy prawdopodobieństwo, że w rozdaniu brydżowym każdy układ składający się z 13 kart zawiera jednego asa. Najpierw rozdajemy asy; można to zrobić na $4! = 24$ sposoby. Tak samo jak w części (a) zbiór pozostałych kart można podzielić na $48!/(12!)^4$ sposobów. Zatem w $24 \cdot 48!/(12!)^4$ rozdaniach brydżowych każdy gracz dostanie asa. Prawdopodobieństwo takiego rozdania wynosi

$$24\frac{48!}{(12!)^4} \cdot \frac{(13!)^4}{52!} = \frac{24 \cdot 13^4}{49 \cdot 50 \cdot 51 \cdot 52} \approx 0{,}1055.$$

(c) Pojedynczy układ kart w brydżu składa się z 13 kart wyciągniętych z talii 52 kart. Takich układów jest $\binom{52}{13} \approx 6,394 \cdot 10^{11}$. Mówimy, że układ kart w brydżu ma rozkład $n_1 - n_2 - n_3 - n_4$, gdzie $n_1 \geq n_2 \geq n_3 \geq n_4$ i $n_1 + n_2 + n_3 + n_4 = 13$, jeśli składa się on z n_1 kart pewnego koloru, n_2 kart jakiegoś innego koloru, n_3 kart jeszcze innego koloru i n_4 kart pozostałego koloru. Dla zilustrowania tego pojęcia znajdziemy liczbę układów kart o rozkładzie $4 - 3 - 3 - 3$. Jest

$$\binom{13}{4}\binom{13}{3}\binom{13}{3}\binom{13}{3}$$

sposobów, by wybrać 4 trefle i po 3 karty każdego z pozostałych kolorów. Ten sam wynik otrzymamy zastępując trefl dowolnym innym kolorem. Wnioskujemy stąd, że jest

$$4\binom{13}{4}\binom{13}{3}^3 \approx 6{,}6906 \cdot 10^{10}$$

układów kart o rozkładzie $4-3-3-3$. Prawdopodobieństwo, że układ kart będzie miał rozkład $4-3-3-3$ wynosi

$$\frac{6{,}6906}{63{,}94} \approx 0{,}1046. \qquad \blacksquare$$

Rozwiązanie niektórych problemów sprowadza się do zliczania podziałów nieuporządkowanych. W takich przypadkach trzeba najpierw znaleźć liczbę podziałów uporządkowanych, a następnie, biorąc pod uwagę, że nie interesuje nas kolejność bloków, podzielić ją przez odpowiednie liczby.

PRZYKŁAD 6 (a) Na ile sposobów można podzielić dwanaścioro studentów na trzy grupy, po czworo studentów w każdej, tak, że jedna grupa studiuje temat T_1, druga temat T_2, a trzecia temat T_3? W tym przypadku porządek jest istotny: jeślibyśmy zmienili kolejność grup, to studenci studiowaliby inne tematy. Tak więc zliczamy podziały uporządkowane, których jest

$$\frac{12!}{4!4!4!} = \binom{12}{4} \cdot \binom{8}{4} = 495 \cdot 70 = 34650.$$

(b) Na ile sposobów można podzielić dwanaścioro studentów na trzy zespoły robocze, po czworo studentów w każdym, z tym, że każdy zespół studiuje ten sam temat? Tym razem chcemy zliczać podziały nieuporządkowane, uważamy bowiem podziały takie jak (A, B, C) i (B, A, C) za równoważne. Odpowiadają one temu samemu podziałowi dwanaściorga naszych studentów na trzy równe ilościowo grupy. Na mocy części (a) jest 34650 podziałów uporządkowanych. Jeśli każdemu z podziałów uporządkowanych (A, B, C) przyporządkujemy odpowiadający mu podział nieuporządkowany $\psi((A, B, C)) = \{A, B, C\}$, to okaże się, że zbiór $\psi^{\leftarrow}(\{A, B, C\})$ ma $3! = 6$ elementów, a mianowicie (A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B) i (C, B, A). Stąd, na mocy lematu o zliczaniu, jest $34650/6 = 5775$ podziałów nieuporządkowanych żądanego typu. A zatem odpowiedzią na nasze pytanie jest 5775. $\blacksquare$

PRZYKŁAD 7 (a) Na ile sposobów można podzielić dziewiętnaścioro studentów na pięć zespołów, w tym dwa zespoły po pięcioro i trzy zespoły po troje osób tak, że każdy zespół studiuje inny spośród pięciu danych tematów? Tak jak w przykładzie 6(a) zliczamy podziały uporządkowane, których jest

$$\frac{19!}{5!5!3!3!3!} \approx 3{,}911 \cdot 10^{10}.$$

(b) Na ile sposobów można podzielić studentów z części (a), jeśli każdy z pięciu zespołów ma studiować ten sam temat? W części (a) zliczaliśmy wszystkie podziały uporządkowane (A, B, C, D, E), gdzie $|A| = |B| = 5$ i $|C| = |D| = |E| = 3$. Jeśli zamienimy miejscami A z B lub zmienimy kolejność w ciągu C, D, E, to otrzymamy ten sam podział na zespoły. Nie wolno nam jednak zamieniać ze sobą miejscami zespołów różnej wielkości takich, jak A i D. Aby znaleźć liczbę podziałów nieuporządkowanych rozważmy funkcję $\psi((A, B, C, D, E)) = (\{A, B\}, \{C, D, E\})$. Każdy przeciwobraz $\psi^{\leftarrow}(\{A, B\}, \{C, D, E\})$ ma $2! \cdot 3!$ elementów (takich jak (B, A, C, E, D)), a stąd na mocy lematu o zliczaniu jest

$$\frac{19!}{5!5!3!3!3!} \cdot \frac{1}{2!3!} \approx 3,26 \cdot 10^9$$

nieuporządkowanych podziałów studentów na zespoły robocze. ■

PRZYKŁAD 8 Na ile sposobów można podzielić grupę 12 zawodników na 4 drużyny po 3 zawodników w każdej? Ogólniej, jaki będzie wynik, jeśli będziemy chcieli podzielić $3n$ zawodników na n drużyn po 3 w każdej? Pytamy o podziały nieuporządkowane. Jest

$$\frac{(3n)!}{(3!) \cdot \ldots \cdot (3!)} = \frac{(3n)!}{6^n}$$

podziałów uporządkowanych $(A_1, A_2, \ldots, A_n)$ takich, że każdy zbiór A_i ma 3 elementy. Dowolna permutacja n danych zbiorów wyznacza ten sam podział nieuporządkowany, a więc jest $\frac{(3n)!}{6^n \cdot n!}$ podziałów nieuporządkowanych zbioru mającego $3n$ elementów na n trzyelementowych bloków. Zauważ, że ta liczba jest równa

$$\begin{aligned}
&\frac{3n(3n-1)(3n-2)}{6n} \cdot \frac{(3n-3)(3n-4)(3n-5)}{6(n-1)} \cdot \\
&\qquad \ldots \cdot \frac{6 \cdot 5 \cdot 4}{6 \cdot 2} \cdot \frac{3 \cdot 2 \cdot 1}{6 \cdot 1} \\
&= \frac{(3n-1)(3n-2)}{2} \cdot \frac{(3n-4)(3n-5)}{2} \cdot \ldots \cdot \frac{5 \cdot 4}{2} \cdot \frac{2 \cdot 1}{2} \\
&= \binom{3n-1}{2} \cdot \binom{3n-4}{2} \cdot \ldots \cdot \binom{5}{2} \cdot \binom{2}{2}.
\end{aligned}$$

Powyższy iloczyn sugeruje inny jeszcze sposób rozwiązania problemu. Ustaw zawodników w pewnej kolejności. Weź pierwszego z nich i wybierz jeszcze dwóch do jego drużyny na $\binom{3n-1}{2}$ sposobów. Następnie weź pierwszego spośród zawodników, którzy dotąd nie zostali wybrani i dobierz do niego jeszcze dwóch na $\binom{3n-4}{2}$ sposobów. I tak dalej. (Lub zastosuj indukcję).

W przypadku dwunastu zawodników będzie

$$\binom{11}{2} \cdot \binom{8}{2} \cdot \binom{5}{2} \cdot \binom{2}{2} = 15400$$

sposobów skompletowania drużyn. ∎

Zasady zliczania przedstawione w tym paragrafie i w § 5.3 mogą wydać się zbiorem trików potrzebnych do rozwiązania problemów, które tu akurat sobie postawiliśmy. Problematyka zliczania jest do pewnego stopnia właśnie taka. Wolimy jednakże myśleć o naszych technikach bardziej jak o narzędziach niż trikach. Znajdują one zastosowanie w wielu różnych, często spotykanych, sytuacjach, ale nie rozwiążą każdego problemu. Procesy myślowe, które przedstawiliśmy w dowodach poszczególnych zasad, są równie cenne jak same zasady, ponieważ podobna analiza może często być użyta w przypadku problemów, do których gotowe narzędzia nie dają się zastosować.

ĆWICZENIA DO § 5.4

1. Z grona 15 osób mają być wybrane 3 komisje składające się, odpowiednio, z 3, 4 i 5 osób.
(a) Ile takich zbiorów komisji można utworzyć, jeśli żadna z osób nie może pracować w więcej niż jednej komisji?
(b) Ile takich zbiorów komisji można utworzyć, jeśli nie ma ograniczeń na liczbę komisji, w których każda z osób może pracować?

2. Porównaj ze sobą liczby:
(a) $\binom{7}{2} \cdot \binom{5}{2}$ oraz $\frac{7!}{2!\cdot 2!\cdot 3!}$,
(b) $\binom{12}{3} \cdot \binom{9}{4}$ oraz $\frac{12!}{3!\cdot 4!\cdot 5!}$,
(c) $\binom{n}{k} \cdot \binom{n-k}{r}$ oraz $\frac{n!}{k!\cdot r!\cdot (n-k-r)!}$.

3. (a) Na ile sposobów można wybrać trzy rozłączne komisje spośród trzynaściorga osób, jeśli muszą one mieć, odpowiednio, 5, 3 i 2 członków.
(b) Powtórz część (a) dla przypadku, gdy komisje muszą mieć, odpowiednio, 4, 3 i 3 członków.
(c) Powtórz część (a) dla przypadku, gdy każda z komisji musi się składać z 3 osób.

4. Ile różnych sygnałów można utworzyć umieszczając obok siebie w pionowej kolumnie dziewięć flag, z których 3 są białe, 2 czerwone a 4 niebieskie?

5. Niech S będzie zbiorem wszystkich ciągów o długości dziesięć, złożonych z cyfr 0, 1 i 2. Na przykład do S należy ciąg 0211012201.

(a) Ile elementów ma zbiór S?
(b) Ile ciągów należących do S ma dokładnie 5 zer i 5 jedynek?
(c) Ile ciągów należących do S ma dokładnie 3 zera i 7 jedynek?
(d) Ile ciągów należących do S ma dokładnie 3 zera?
(e) Ile ciągów należących do S ma dokładnie 3 zera, 4 jedynki i 3 dwójki?
(f) Ile ciągów należących do S ma co najmniej 1 zero, co najmniej 1 jedynkę i co najmniej 1 dwójkę?

6. Znajdź liczbę wszystkich ciągów, które można utworzyć permutując litery w następujących słowach:
(a) FLORYDA,
(b) KALIFORNIA,
(c) MISSISSIPPI,
(d) OHIO.

7. (a) Ile liczb czterocyfrowych można utworzyć używając jedynie cyfr 3, 4, 5, 6 i 7?
(b) Ile spośród liczb z części (a) ma jakieś powtarzające się cyfry?
(c) Ile jest wśród liczb z części (a) liczb parzystych?
(d) Ile jest wśród liczb z części (a) liczb większych niż 5000?

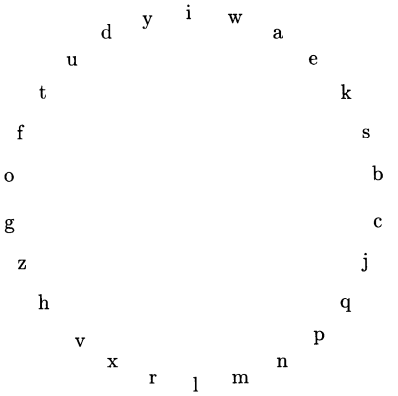
8. Niech $\Sigma = \{a, b, c\}$. Powiemy, że słowo należące do zbioru Σ^6 jest typu $n_1 - n_2 - n_3$, gdzie $n_1 \geq n_2 \geq n_3$ i $n_1 + n_2 + n_3 = 6$, jeśli jedna z liter występuje w tym słowie n_1 razy, inna n_2 razy, a ostatnia n_3 razy. Na przykład słowo *accabc* jest typu $3-2-1$, a *caccca* — typu $4-2-0$. Liczba słów należących do zbioru Σ^6 każdego z typów wynosi odpowiednio

Typ	$6-0-0$	$5-1-0$	$4-2-0$	$4-1-1$	$3-3-0$	$3-2-1$	$2-2-2$
Liczba	3	36	90	90	60	360	90

Sprawdź to dla trzech wybranych typów.

9. Na ile sposobów można podzielić zbiór złożony z $2n$ elementów na dwa zbiory, po n elementów w każdym?

10. Angielski alfabet składa się z 21 spółgłosek i 5 samogłosek. Tymi samogłoskami są a, e, i, o, u.
(a) Udowodnij, że niezależnie od tego, w jakim porządku zostaną wypisane litery alfabetu angielskiego (np. z u v a r q l g h...), to zawsze znajdą się 4 spółgłoski, które wystąpią kolejno, jedna po drugiej.
(b) Wypisz litery w takim porządku, w którym nie występuje 5 spółgłosek z rzędu.
(c) Przypuśćmy teraz, że litery angielskiego alfabetu zostały wypisane w dowolnej kolejności na obwodzie koła; na przykład tak, jak pokazano na rysunku 5.4.
Udowodnij, że musi się znaleźć 5 spółgłosek, które występują kolejno, jedna po drugiej.



Rysunek 5.4

11. Dwoje spośród 12 zawodników z przykładu 8, Anna i Robert, chcą znaleźć się w tej samej drużynie.
(a) Ile jest sposobów wybrania drużyn tak, by Anna i Robert byli w jednej drużynie?
(b) Jeśli wszystkie sposoby wyboru drużyn są jednakowo prawdopodobne, to jakie jest prawdopodobieństwo, że Anna i Robert znajdą się w tej samej drużynie?

12. W turnieju koszykówki bierze udział 16 drużyn. Ile jest sposobów zestawienia tych drużyn w 8 par?

13. (a) Ile jest liczb całkowitych między 1000 a 9999, których suma cyfr wynosi dokładnie 9? Przykłady: 1431, 5121, 9000, 4320.
(b) Ile spośród liczb zliczanych w części (a) ma wszystkie cyfry różne od 0?

14. Weź skończone zbiory, które spełniają równość $\chi_A + \chi_B = \chi_C + \chi_D$. Wykaż, że

$$|A| + |B| = |C| + |D|.$$

15. Ile jest relacji równoważności w zbiorze $\{0, 1, 2, 3\}$? *Wskazówka*: zliczaj podziały nieuporządkowane. Dlaczego to rozwiązuje problem?

16. Kapelusz zawiera litery B, B, B, E, O, O, P, P. Jakie jest prawdopodobieństwo, że wyciągane kolejno z kapelusza litery utworzą słowo B E B O P B O P?

§ 5.5. Zasada szufladkowa Dirichleta

Zwykła zasada szufladkowa Dirichleta stwierdza, że jeśli m przedmiotów umieścimy w n pudełkach lub szufladkach i jeśli

$m > n$, to do jakiegoś pudełka trafi więcej niż jeden przedmiot. A oto nieznaczne uogólnienie tego faktu.

Zasada szufladkowa Dirichleta

> Jeśli skończony zbiór S jest podzielony na k zbiorów, to co najmniej jeden z tych zbiorów ma $|S|/k$ lub więcej elementów.

Dowód. Niech $A_1, \ldots, A_k$ będą blokami danego podziału. Wtedy średnia wartość $|A_i|$ wynosi $\frac{1}{k} \cdot (|A_1| + \ldots + |A_k|) = \frac{1}{k} \cdot |S|$, a więc najliczniejszy ze zbiorów A_i ma co najmniej tyle elementów. ∎

Będziemy często stosować tę zasadę do sytuacji, gdy podział jest określony przez funkcję. W tym przypadku zasadę można sformułować w sposób następujący.

Zasada szufladkowa Dirichleta

> Niech dana będzie funkcja $f\colon S \to T$, gdzie S i T są skończonymi zbiorami spełniającymi nierówność $|S| > r \cdot |T|$. Wówczas co najmniej jeden ze zbiorów $f^{\leftarrow}(t)$ ma więcej niż r elementów.

Dowód. Rodzina $\{f^{\leftarrow}(t)\colon t \in T\}$ tworzy podział zbioru S na k zbiorów, gdzie $k \leq |T|$. Na pomocy dopiero co udowodnionej zasady, pewien zbiór $f^{\leftarrow}(t)$ ma co najmniej $|S|/k$ elementów. Ponieważ z założenia wynika, że $|S|/k \geq |S|/|T| > r$, to zbiór taki ma więcej niż r elementów. ∎

Gdy $r = 1$, zasada ta stwierdza, że jeśli $f\colon S \to T$ i $|S| > |T|$, to co najmniej jeden ze zbiorów $f^{\leftarrow}(t)$ ma więcej niż jeden element. Jest godne uwagi to, jak często ta prosta obserwacja pomaga w rozwiązywaniu problemów.

PRZYKŁAD 1 Wśród dowolnych trzech liczb całkowitych muszą być dwie, których suma jest parzysta, ponieważ albo dwie spośród danych liczb są parzyste, albo dwie z nich są nieparzyste i w każdym przypadku suma takiej pary musi być parzysta. A oto bardziej formalny argument. Niech S będzie zbiorem złożonym z trzech danych liczb. Wtedy $\text{MOD}\,2\colon S \to \{0,1\}$ i na mocy zasady szufladkowej jeden ze zbiorów, $(\text{MOD}\,2)^{\leftarrow}(0)$ lub $(\text{MOD}\,2)^{\leftarrow}(1)$, ma więcej niż jeden element. To znaczy, że do zbioru S należą dwie (lub więcej) liczby parzyste lub też dwie (lub więcej) liczby nieparzyste. ∎

PRZYKŁAD 2 Pokażemy, że jeśli $a_1, a_2, \ldots, a_p$ są liczbami całkowitymi, niekoniecznie różnymi, to suma pewnych spośród nich jest wielokrotnością liczby p. Rozważmy funkcję $\text{MOD}\, p$: $\mathbb{Z} \to \mathbb{Z}_p$ zastosowaną do elementów zbioru

$$S = \{0,\ a_1,\ a_1 + a_2,\ a_1 + a_2 + a_3, \ldots,\ a_1 + a_2 + a_3 + \ldots + a_p\}.$$

Ponieważ $|S| = p + 1 > p = |\mathbb{Z}_p|$, zasada szufladkowa pokazuje, że dwóm różnym liczbom n i m ze zbioru S przyporządkowana jest ta sama wartość w $\mathbb{Z}_p$. Skoro $m \,\text{MOD}\, p = n \,\text{MOD}\, p$, to mamy $m \equiv n \bmod(p)$, a zatem różnice $n-m$ i $m-n$ są wielokrotnościami liczby p. Jedna z tych różnic ma postać $a_k + a_{k+1} + \ldots + a_l$, tzn. jest to szukana suma pewnych liczb całkowitych z naszego ciągu, będąca wielokrotnością liczby p.

Zauważ, że udowodniliśmy więcej niż żądaliśmy; suma pewnej grupy kolejnych elementów a_i jest wielokrotnością p. Dopiero co udowodnionego rezultatu w tym sensie nie można wzmocnić, że istnieją liczby całkowite $a_1, a_2, \ldots, a_{p-1}$ takie, że żadna suma pewnych spośród nich nie jest wielokrotnością liczby p. Wystarczy po prostu przyjąć $a_j = 1$ dla $j = 1, 2, \ldots, p-1$. ∎

PRZYKŁAD 3 Niech A będzie pewnym ustalonym dziesięcioelementowym podzbiorem zbioru $\{1, 2, \ldots, 50\}$. Pokażemy, że zbiór A ma dwa różne pięcioelementowe podzbiory takie, że sumy wszystkich elementów każdego z nich są równe. Niech $\mathcal{S}$ będzie rodziną wszystkich pięcioelementowych podzbiorów B zbioru A. Dla każdego zbioru B należącego do $\mathcal{S}$ niech $f(B)$ będzie sumą wszystkich liczb należących do B. Zauważ, że musi zachodzić $f(B) \geq 1+2+3+4+5 = 15$ i $f(B) \leq 50+49+48+47+46 = 240$, a więc f: $\mathcal{S} \to T$, gdzie $T = \{15, 16, 17, \ldots, 240\}$. Ponieważ $|T| = 226$, a $|\mathcal{S}| = \binom{10}{5} = 252$, zasada szufladkowa pokazuje, że do rodziny $\mathcal{S}$ należą dwa różne zbiory, którym funkcja f przyporządkowuje te same wartości, tzn. różne zbiory takie, że sumy wszystkich elementów każdego z nich są równe. ∎

Niektóre zastosowania zasady szufladkowej wymagają znacznej pomysłowości.

PRZYKŁAD 4 Pokażemy teraz, że jeśli $a_1, a_2, \ldots, a_{n^2+1}$ jest ciągiem złożonym z n^2+1 różnych liczb rzeczywistych, to istnieje jego podciąg długości $n+1$, który jest rosnący bądź malejący. To znaczy, że istnieją indeksy $s(1) < s(2) < \ldots < s(n+1)$ takie, że albo

$$a_{s(1)} < a_{s(2)} < \ldots < a_{s(n+1)},$$

albo

$$a_{s(1)} > a_{s(2)} > \ldots > a_{s(n+1)}.$$

Dla każdego j ze zbioru $\{1, 2, \ldots, n^2+1\}$, niech INC(j) oznacza długość najdłuższego spośród rosnących podciągów, które kończą się na a_j, a DEC(j) oznacza długość najdłuższego spośród malejących podciągów, które kończą się na a_j. Następnie zdefiniujmy $f(j) = (\text{INC}(j), \text{DEC}(j))$. Na przykład, przypuśćmy, że $n = 3$ i wyjściowy ciąg jest określony w następujący sposób:

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
11	3	15	8	6	12	17	2	7	1

W tym przypadku $a_5 = 6$, INC(5)=2, gdyż a_2, a_5 jest najdłuższym rosnącym podciągiem kończącym się na a_5, a DEC(5)=3, gdyż a_1, a_4, a_5 i a_3, a_4, a_5 (tzn. 11, 8, 6 i 15, 8, 6) są najdłuższymi malejącymi podciągami kończącymi się na a_5. Podobnie, INC(6)=3 i DEC(6)=2, a więc $f(5) = (2,3)$ i $f(6) = (3,2)$. Ogólnie, dla tego przykładowego ciągu:

$$\begin{aligned} f(1) &= (1,1), \quad f(2) = (1,2), \quad f(3) = (2,1), \\ f(4) &= (2,2), \quad f(5) = (2,3), \quad f(6) = (3,2), \\ f(7) &= (4,1), \quad f(8) = (1,4), \quad f(9) = (3,3), \\ f(10) &= (1,5). \end{aligned}$$

Ten przykładowy ciąg ma zarówno rosnące podciągi długości 4, takie jak a_2, a_5, a_6, a_7 (zauważ, że INC(7)=4), jak i malejące ciągi długości 4, takie jak a_1, a_4, a_5, a_8 (zauważ, że DEC(8)=4). Ponieważ DEC(10)=5, ma on nawet ciąg malejący długości 5. Zauważ, że w tym przykładzie funkcja f jest różnowartościowa, tak więc nie może ona odwzorowywać dziesięcioelementowego zbioru $\{1, 2, 3, \ldots, 10\}$ w dziewięcioelementowy zbiór $\{1, 2, 3\} \times \{1, 2, 3\}$. Innymi słowy, już sama różnowartościowość funkcji f wymusza, by co najmniej jedna z wartości, INC(j) lub DEC(j), przekroczyła 3, co z kolei zmusza nasz ciąg do posiadania rosnącego bądź malejącego podciągu długości 4.

Dla dowodu ogólnego rezultatu pokażemy najpierw, że funkcja f *zawsze* musi być różnowartościowa. Weź liczby j, k ze zbioru $\{1, 2, 3, \ldots, n^2 + 1\}$ takie, że $j < k$. Jeśli $a_j < a_k$, to INC(j) < INC(k), ponieważ zawsze można dołączyć a_k do najdłuższego rosnącego podciągu kończącego się na a_j i otrzymać dłuższy rosnący podciąg kończący się na a_k. Podobnie, jeśli $a_j > a_k$, to DEC(j) < DEC(k). W obu przypadkach *uporządkowane pary* $f(j)$ i $f(k)$ nie mogą być sobie równe, tzn. $f(j) \neq f(k)$. Ponieważ

funkcja f jest różnowartościowa, zasada szufladkowa mówi, że f nie może odwzorowywać

zbioru $\{1, 2, 3, \ldots, n^2+1\}$ w zbiór $\{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$,

a więc bądź istnieje liczba j taka, że $\text{INC}(j) \geq n+1$, bądź też taka, że $\text{DEC}(j) \geq n+1$. Zatem wyjściowy ciąg ma rosnący lub malejący podciąg o $n+1$ wyrazach. ∎

PRZYKŁAD 5 Prawdopodobnie wiesz, że rozwinięcia dziesiętne liczb wymiernych są okresowe, ale być może nigdy nie widziałeś dowodu tego faktu. Na przykład

$$\frac{29}{54} = 0{,}537037037037037\ldots;$$

sprawdź to za pomocą dzielenia pisemnego zanim będziesz czytać dalej! Ogólny fakt można wyprowadzić z zasady szufladkowej w sposób następujący.

Możemy założyć, że dana liczba wymierna jest postaci m/n, gdzie $0 < m < n$. Przeanalizujmy poszczególne kroki algorytmu dzielenia pisemnego. Gdy dzielimy m przez n, to otrzymujemy $0, d_1 d_2 d_3 \ldots$, gdzie

$$\begin{aligned} 10 \cdot m &= n \cdot d_1 + r_1, \quad 0 \leq r_1 < n, \\ 10 \cdot r_1 &= n \cdot d_2 + r_2, \quad 0 \leq r_2 < n, \\ 10 \cdot r_2 &= n \cdot d_3 + r_3, \quad 0 \leq r_3 < n \end{aligned}$$

itd., a więc $10 \cdot r_j = n \cdot d_{j+1} + r_{j+1}$, gdzie $0 \leq r_{j+1} < n$. To znaczy, że

$$r_{j+1} = (10 \cdot r_j) \text{ MOD } n \quad \text{oraz} \quad d_{j+1} = (10 \cdot r_j) \text{ DIV } n.$$

Na rysunku 5.5 pokazane są szczegóły naszego przykładowego dzielenia.

Wszystkie reszty r_j przyjmują wartości w zbiorze $\{0, 1, 2, \ldots, n-1\} = \mathbb{Z}_n$. Na mocy zasady szufladkowej, w pewnym momencie wartości te muszą się powtórzyć. Ściśle biorąc, dwie spośród liczb $r_1, r_2, \ldots, r_{n+1}$ muszą się sobie równać. Istnieją zatem w zbiorze $\{0, 1, 2, \ldots, n+1\}$ liczby k i l takie, że $k < l$ i $r_k = r_l$. Niech $p = l - k$, wtedy $r_k = r_{k+p}$. (W naszym starannie dobranym przykładzie k może być liczbą 1, a p — liczbą 3.) Pokażemy, że ciągi złożone z liczb r_i i d_i, poczynając od $i = k+1$ powtarzają się co p miejsc.

Najpierw wykażemy przez indukcję, że reszty się powtarzają:

$$(*) \qquad r_j = r_{j+p} \quad \text{dla} \quad j \geq k.$$

$0, d_1d_2d_3d_4d_5 \ldots = 0,53703 \ldots$

$54 \mid 290$	$10 \bullet 29 = 10m$	
270	$54 \bullet 5 = nd_1$	
200	$10 \bullet 20 = 10r_1$	$r_1 = 20$
162	$54 \bullet 3 = nd_2$	
380	$10 \bullet 38 = 10r_2$	$r_2 = 38$
378	$54 \bullet 7 = nd_3$	
20	$10 \bullet 2 = 10r_3$	$r_3 = 2$
0	$54 \bullet 0 = nd_4$	
200	$10 \bullet 20 = 10r_4$	$r_4 = 20$
162	$54 \bullet 3 = nd_5$	
380	$10 \bullet 38 = 10r_5$	$r_5 = 38$
378	itd.	

Rysunek 5.5

Mamy $r_k = r_{k+p}$, gdyż tak zostały wybrane liczby k i p. Zróbmy założenie indukcyjne, że dla pewnego j, $r_j = r_{j+p}$. Wtedy $r_{j+1} = (10 \cdot r_j) \operatorname{MOD} n = (10 \cdot r_{j+p}) \operatorname{MOD} n = r_{j+p+1}$. Zatem pokazaliśmy przez indukcję po j, że równość $(*)$ jest spełniona dla $j \geq k$.

Teraz dla $j > k$, z równości $(*)$ wynika, że

$$d_j = (10 \cdot r_{j-1}) \operatorname{DIV} n = (10 \cdot r_{j+p-1}) \operatorname{DIV} n = d_{j+p}.$$

Zatem wyrazy d_j również powtarzają się cyklicznie, w cyklach długości p i mamy

$$d_{k+1} = d_{k+p+1}, \quad d_{k+2} = d_{k+p+2}, \quad \ldots, \quad d_{k+p} = d_{k+2p},$$

a stąd

$$d_{k+1}d_{k+2} \ldots d_{k+p} = d_{k+p+1}d_{k+p+2} \ldots d_{k+2p}.$$

Tak więc cały ten blok powtarza sie w nieskończoność. Innymi słowy, rozwinięcie dziesiętne liczby m/n jest rozwinięciem okresowym. ∎

W następnym przykładzie zasada szufladkowa nie zostanie bezpośrednio zastosowana, ale problem ma charakter zbliżony do tych, ktore się za jej pomocą rozwiązuje.

PRZYKŁAD 6 Weźmy dziewięć nieujemnych liczb rzeczywistych a_1, a_2, a_3, $\ldots$, a_9, których sumą jest 90.

(a) Pokażemy, że wśród tych liczb muszą istnieć 3 takie, których suma równa się co najmniej 30. To jest łatwe, ponieważ

$$90 = (a_1 + a_2 + a_3) + (a_4 + a_5 + a_6) + (a_7 + a_8 + a_9),$$

więc przynajmniej jedna z sum w nawiasach musi równać się co najmniej 30.

(b) Pokażemy, że wśród tych liczb muszą istnieć 4 takie, których suma równa się co najmniej 40. Można to zrobić na wiele sposobów, ale żaden z nich nie jest tak prosty, jak metoda z części (a). Nasz pierwszy sposób polega na zauważeniu, że sumą wszystkich liczb występujących w tabelce

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9
a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_1
a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_1	a_2
a_4	a_5	a_6	a_7	a_8	a_9	a_1	a_2	a_3

jest 360, gdyż sumą każdego rzędu jest 90. Zatem jedna z dziewięciu kolumn musi mieć sumę równą co najmniej $360/9 = 40$.

Nasz drugi sposób polega na wykorzystaniu części (a), aby wybrać takie trzy spośród danych liczb, których suma wynosi $s \geq 30$. Jedna z pozostałych sześciu liczb musi być równa co najmniej $\frac{1}{6}$ ich sumy, która wynosi $90 - s$. Dodając ją do poprzednio wybranej trójki otrzymujemy cztery liczby, których suma wynosi co najmniej

$$s + \frac{1}{6}(90 - s) = 15 + \frac{5}{6}s \geq 15 + \frac{5}{6} \cdot 30 = 40.$$

Nasz trzeci sposób polega na zauważeniu, że możemy założyć dodatkowo, że $a_1 \geq a_2 \geq \ldots \geq a_9$. Wtedy jasne jest, że $a_1 + a_2 + a_3 + a_4$ jest największą z sum, które składają się z czterech spośród danych liczb i zadanie nasze jest stosunkowo konkretne: pokazać, że $a_1 + a_2 + a_3 + a_4 \geq 40$. Co więcej, to sugeruje, by pokazać, że

$$(*) \qquad a_1 + a_2 + \ldots + a_n \geq 10n$$

dla $1 \leq n \leq 9$. Możemy to zrobić przez skończoną indukcję, tzn. zauważyć, że $(*)$ zachodzi dla $n = 1$ i pokazać, że jeśli $(*)$ zachodzi dla n, gdzie $1 \leq n < 9$, to $(*)$ zachodzi dla $n + 1$. Zastosujemy metodę, której użyliśmy rozwiązując problem sposobem drugim. Załóżmy, że $(*)$ zachodzi dla n i niech $s = a_1 + a_2 + \ldots + a_n$. Ponieważ a_{n+1} jest największą z pozostałych $9-n$ liczb, to mamy $a_{n+1} \geq (90 - s)/(9 - n)$. Stąd

$$a_1 + a_2 + \ldots + a_n + a_{n+1} = s + a_{n+1} \geq s + \frac{90 - s}{9 - n} =$$

$$
\begin{aligned}
&= s + \frac{90}{9-n} - \frac{s}{9-n} \\
&= s\left(1 - \frac{1}{9-n}\right) + \frac{90}{9-n} \\
&\geq 10n\left(1 - \frac{1}{9-n}\right) + \frac{90}{9-n} \quad \text{(założenie indukcyjne)} \\
&= 10n + \frac{90 - 10n}{9-n} = 10n + 10 = 10(n+1).
\end{aligned}
$$

Powyższy dowód przez skończoną indukcję pokazuje, że nierówność $(*)$ jest spełniona dla $1 \leq n \leq 9$.

W przypadku tego problemu sposoby pierwszy i trzeci są znacznie lepsze, gdyż bez żadnych dalszych trików można je w sposób oczywisty uogólnić. ∎

Zasadę szufladkową Dirichleta można uogólnić dopuszczając przypadek, gdy zbiory A_i nie są parami rozłączne.

Uogólniona zasada szufladkowa Dirichleta

Niech $A_1, \ldots, A_k$ będą podzbiorami skończonego zbioru S takimi, że każdy element zbioru S należy do co najmniej t spośród zbiorów A_i. Wówczas średnia arytmetyczna liczb elementów zbiorów A_i wynosi co najmniej $t \cdot |S|/k$.

Dowód. Zastosujemy interesującą i silną technikę, polegającą na zliczaniu par uporządkowanych na dwa sposoby. Niech P będzie zbiorem wszystkich par (s, A_i) takich, że $s \in A_i$. Możemy znaleźć liczbę elementów zbioru P, zliczając dla każdego s ze zbioru S wszystkie pary, w których występuje s, a następnie sumując otrzymane liczby. Otrzymujemy

$$
\begin{aligned}
|P| &= \sum_{s \in S} [\text{liczba takich par } (s, A_i), \text{ że } s \in A_i] \\
&= \sum_{s \in S} [\text{liczba takich } A_i, \text{ że } s \in A_i] \\
&\geq \sum_{s \in S} t \qquad \text{(wynika z założenia)} \\
&= t \cdot |S|.
\end{aligned}
$$

Możemy też znaleźć liczbę elementów zbioru P, zliczając dla każdego A_i wszystkie pary, w których występuje A_i i sumując otrzymane liczby. A zatem

$$|P| = \sum_{i=1}^{k} [\text{liczba takich par } (s, A_i), \text{ że } s \in A_i]$$
$$= \sum_{i=1}^{k} [\text{liczba takich } s, \text{ że } s \in A_i] = \sum_{i=1}^{k} |A_i|.$$

Łącząc te dwa rezultaty, otrzymujemy

$$\sum_{i=1}^{k} |A_i| \geq t \cdot |S|,$$

a więc średnia arytmetyczna z liczb $|A_i|$, czyli

$$\frac{1}{k} \cdot \sum_{i=1}^{k} |A_i|,$$

wynosi co najmniej $t \cdot |S|/k$.

Nasz dowód pokazuje również, że jeśli każdy element s należy do dokładnie t zbiorów A_i, to średnia arytmetyczna z liczb $|A_i|$ jest równa dokładnie $t \cdot |S|/k$. W szczególnym przypadku dla $t = 1$ otrzymujemy zwykłą zasadę szufladkową. ∎

PRZYKŁAD 7 Koło ruletki jest podzielone na 36 sektorów z liczbami 1, 2, 3, ..., 36 wypisanymi w jakiejś kolejności. (Pomijamy sektory oznaczone przez 0 i 00, które istnieją w Las Vegas i dają kasynom przewagę w grze.) Średnia wartość liczb w sektorze to 18,5. Jeśli to nie jest oczywiste, to zauważ, że na mocy przykładu 2(b) z § 4.2,

$$1 + 2 + \ldots + 36 = 666$$

oraz, że $666/36 = 18,5$.

(a) Jest 36 par sektorów, położonych jeden obok drugiego. Pokażemy, że średnia wartość sumy liczb w sektorach z takiej pary wynosi 37. Wyobraź sobie, że każdy sektor zastąpiony został torebką zawierającą tyle szklanych kulek, ile wynosi liczba w tym sektorze. Dla każdej pary kolejnych sektorów utwórz zbiór złożony z kulek z odpowiednich dwóch torebek. Każda kulka należy do dokładnie dwóch takich zbiorów, a liczba wszystkich zbiorów wynosi 36. Ponieważ całkowita liczba kulek wynosi 666, to z uwagi zamieszczonej na końcu dowodu uogólnionej zasady szufladkowej wynika, że średnia liczba kulek w takim zbiorze to $2 \cdot 666/36 = 37$. Liczba kulek w każdym ze zbiorów odpowiada sumie liczb w parze kolejnych sektorów.

Fakt, że każda kulka należy do dokładnie dwóch spośród zbiorów, jest istotny dla rozumowania przedstawionego powyżej.

Przypuśćmy na przykład, że pominiemy w rozważaniach jedną z par kolejnych sektorów. Jeśli liczby z tych dwóch sektorów to, powiedzmy, 5 i 17, to średnia wartość sumy liczb w sektorach z pozostałych 35 par wyniosłaby

$$\frac{2 \cdot 666 - 22}{35} \approx 37,43.$$

(b) Pewna para kolejnych sektorów musi dać sumę równą co najmniej 38. Aby się o tym przekonać, podziel 36 sektorów na 18 rozłącznych par zawierających kolejne sektory. Średnia wartość sumy z takiej pary wynosi 37. Albo jedna z sum przekracza 37, albo też każda para daje sumę równą dokładnie 37. Jeżeli zachodzi ten ostatni przypadek, przesuń rozważania o jeden sektor zgodnie z kierunkiem ruchu wskazówek zegara. Zmieni to wszystkie sumy i jedna z nowych sum będzie musiała być większa niż 37.

(c) Część (a) łatwo można uogólnić. Niech t będzie liczbą całkowitą taką, że $2 \leq t \leq 36$. Jest 36 bloków złożonych z t kolejnych sektorów i średnia wartość sumy liczb w sektorach z takiego bloku wynosi $18,5 \cdot t$. Znów wyobraź sobie kulki z części (a). Każdy blok odpowiada zbiorowi kulek związanych z t sektorami z tego bloku. Ponieważ każda kulka należy do dokładnie t spośród tych zbiorów, to średnia liczba kulek w takim zbiorze wynosi

$$\frac{t \cdot 666}{36} = 18,5 \cdot t.$$

Tak jak poprzednio, liczba kulek w każdym ze zbiorów jest sumą liczb z sektorów w odpowiednim bloku. ∎

ĆWICZENIA DO § 5.5

Większość, choć nie wszystkie, z zamieszczonych tu zadań, dotyczą zasady szufladkowej. Dostarczają one także materiału do ćwiczenia technik z paragrafów 5.1-5.4. Zadania nie są jednakowo trudne i niektóre wymagać mogą dodatkowej pomysłowości.

1. (a) Wyjaśnij, dlaczego wśród dowolnych czterech liczb całkowitych dwie muszą przystawać mod 3.

(b) Udowodnij, że jeśli $a_1, a_2, \ldots, a_{p+1}$ są liczbami całkowitymi, to dwie z nich muszą przystawać mod p.

2. (a) Worek zawiera 50 szklanych kulek w czterech różnych kolorach. Wyjaśnij, dlaczego jest co najmniej 13 kulek tego samego koloru.

(b) Jeśli czerwonych kulek jest dokładnie 8, to wyjaśnij, dlaczego jest co najmniej 14 kulek tego samego koloru.

3. Przypuśćmy, że 73 kulki zostały umieszczone w ośmiu pudełkach.
(a) Wykaż, że jedno z pudełek zawiera co najmniej 10 kulek.
(b) Wykaż, że jeśli dwa pudełka są puste, to jakieś pudełko zawiera co najmniej 13 kulek.

4. (a) Niech B będzie dwunastoelementowym podzbiorem zbioru $\{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$. Wykaż, że B zawiera dwie różne pary uporządkowane, mające równe sumy elementów, poprzednika i następnika.
(b) Ile razy można rzucić parą kostek bez dwukrotnego otrzymania tej samej sumy oczek?

5. Niech A będzie dziesięcioelementowym podzbiorem zbioru $\{1,2,3,\ldots,50\}$. Wykaż, że A ma dwa czteroelementowe podzbiory, mające równe sumy elementów.

6. Niech S będzie trzyelementowym zbiorem złożonym z liczb całkowitych. Wykaż, że S ma dwa różne niepuste podzbiory takie, że sumy liczb w każdym z tych podzbiorów przystają modulo 6.

7. Niech A będzie podzbiorem zbioru $\{1,2,3,\ldots,149,150\}$ złożonym z 25 liczb. Wykaż, że istnieją dwie rozłączne pary elementów zbioru A, mające te same sumy (na przykład pary $\{3,89\}$ i $\{41,51\}$ mają tę samą sumę, a mianowicie 92).

8. W każdym z następujących ciągów znajdź rosnący bądź malejący podciąg długości 5, o ile tylko jest to możliwe.
(a) 4, 3, 2, 1, 8, 7, 6, 5, 12, 11, 10, 9, 16, 15, 14, 13
(b) 17, 13, 14, 15, 16, 9, 10, 11, 12, 5, 6, 7, 8, 1, 2, 3, 4
(c) 10, 6, 2, 14, 3, 17, 12, 8, 7, 16, 13, 11, 9, 15, 4, 1, 5

9. Znajdź rozwinięcia dziesiętne liczb 1/7, 2/7, 3/7, 4/7, 5/7 i 6/7. Porównaj je ze sobą.

10. (a) Wykaż, że jeśli dziesięć nieujemnych liczb całkowitych ma sumę 101, to muszą istnieć wśród nich trzy, których suma wynosi co najmniej 31.
(b) Udowodnij uogólnienie części (a): Jeśli $1 \leq k \leq n$ i jeśli n nieujemnych liczb całkowitych ma sumę m, to musi istnieć wśród nich k liczb, których suma wynosi co najmniej _____.

11. W tym zadaniu dana jest pewna permutacja $(n_1, n_2, n_3, n_4, \ldots, n_{24})$ dwudziestu czterech liczb 1, 2, 3, 4, ..., 24.
(a) Wykaż, że muszą istnieć cztery kolejne wyrazy tej permutacji, mniejsze od 20, tzn. równe co najwyżej 19.
(b) Wykaż, że $n_1 + n_2 + n_3 + \ldots + n_{24} = 300$.
(c) Wykaż, że muszą istnieć trzy kolejne wyrazy tej permutacji, których suma wynosi co najmniej 38.
(d) Wykaż, że musi istnieć pięć kolejnych wyrazów tej permutacji, których suma wynosi co najmniej 61.

12. Rozważmy koło ruletki z przykładu 7.

(a) Wykorzystaj przykład 7(c) do pokazania, że istnieją cztery kolejne sektory o sumie co najmniej 74.
(b) Wykaż, że liczbę 74 z części (a) można w rzeczywistości poprawić na 75.
(c) Wykorzystaj przykład 7(c) do pokazania, że istnieje pięć kolejnych sektorów o sumie co najmniej 93.
(d) Wykaż, że liczbę 93 z części (c) można w rzeczywistości poprawić na 95.

13. Niech n_1, n_2 i n_3 będą różnymi dodatnimi liczbami całkowitymi. Wykaż, że co najmniej jedna z liczb n_1, n_2, n_3, $n_1 + n_2$, $n_2 + n_3$ lub $n_1 + n_2 + n_3$ jest podzielna przez 3. *Wskazówka*: za pomocą funkcji $f = \text{MOD}\,3$ przekształć zbiór $\{n_1, n_1 + n_2, n_1 + n_2 + n_3\}$ w $\mathbb{Z}_3$.

14. Członkami pewnego klubu jest sześciu mężczyzn i dziewięć kobiet. Losowo wybieramy pięcioosobową komisję. Znajdź prawdopodobieństwo, że
(a) w komisji znajdzie się dwóch mężczyzn i trzy kobiety;
(b) w komisji znajdzie się co najmniej jeden mężczyzna i co najmniej jedna kobieta;
(c) komisja składa się z samych mężczyzn bądź z samych kobiet.

15. Chcemy tworzyć liczby sześciocyfrowe, których cyfry należą do zbioru

$$A = \{1, 2, 3, 4, 5, 6, 7, 8\}.$$

(a) Ile takich liczb można utworzyć, jeśli cyfry mogą się powtarzać?
(b) Ile spośród liczb z części (a) ma co najmniej jedną cyfrę 3 i co najmniej jedną cyfrę 5?
(c) Ile sześciocyfrowych liczb można utworzyć, jeśli każda cyfra ze zbioru A może być użyta co najwyżej raz?
(d) Ile sześciocyfrowych liczb da się zapisać za pomocą jednej cyfry 2, dwóch cyfr 4 i trzech cyfr 5?

16. Ile dzielników ma liczba 6000? *Wskazówka*: $6000 = 2^4 \cdot 3 \cdot 5^3$ i każdy dzielnik ma postać $2^m 3^n 5^r$, gdzie $m \leq 4$, $n \leq 1$ i $r \leq 3$.

17. Weź liczbę n ze zbioru $\mathbb{P}$ i niech S będzie podzbiorem zbioru $\{1, 2, \ldots, 2n\}$, złożonym z $n + 1$ liczb.
(a) Wykaż, że do zbioru S należą dwie liczby względnie pierwsze.
(b) Wykaż, że do zbioru S należą dwie liczby, z których jedna dzieli drugą.
(c) Pokaż, że zdanie z części (a) może nie być prawdą, jeśli zbiór S ma tylko n elementów.
(d) Pokaż, że zdanie z części (b) może nie być prawdą, jeśli zbiór S ma tylko n elementów.

18. (a) Weź podzbiór A zbioru $\{0, 1, 2, \ldots, p\}$ taki, że $|A| > \frac{1}{2}p+1$. Wykaż, że do zbioru A należą dwie liczby, których sumą jest p.
(b) Dla $p = 6$ znajdź zbiór A taki, że $|A| = \frac{1}{2}p + 1$, który nie spełnia tezy z części (a).

(c) Dla $p = 7$ znajdź zbiór A taki, że $|A| = \frac{1}{2}(p-1)+1$, który nie spełnia tezy z części (a).

19. Grupa złożona z 21 studentów chce utworzyć 7 zespołów roboczych w taki sposób, by każdy student należał do dokładnie 2 zespołów.
(a) Wykaż, że średnia liczebność zespołu roboczego będzie musiała wynosić 6.
(b) Pokaż, w jaki sposób można przypisać studentów do poszczególnych zespołów tak, by w każdym zespole było dokładnie 6 studentów.

To, co jest najważniejsze w tym rozdziale

Aby sprawdzić, czy dobrze rozumiesz treść tego rozdziału, stosuj się do naszych znanych już porad dotyczących takich przeglądów. Myśl zawsze o przykładach.

Pojęcia

wybór (losowanie) ze zwracaniem lub bez zwracania
r-wyrazowa wariacja bez powtórzeń, permutacja, kombinacja
przestrzeń zdarzeń elementarnych (próbek), zdarzenie, wynik, prawdopodobieństwo (na Ω)

Fakty

$|S \cup T| = |S| + |T| - |S \cap T|$ dla zbiorów skończonych.
$|S_1 \times S_2 \times \ldots \times S_n| = \prod_{k=1}^{n} |S_k|$.
$P(E^c) = 1 - P(E)$.
$P(E \cup F) = P(E) + P(F) - P(E \cap F)$.
$\binom{n}{r} = \frac{n!}{(n-r)!r!}$.
Wzór dwumianowy: $(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}$.
Lemat o zliczaniu.
Wzór $\frac{n!}{n_1!n_2!\cdot\ldots\cdot n_k!}$ na zliczanie nierozróżnialnych permutacji lub podziałów uporządkowanych.
Wzór $\binom{n+k-1}{k-1}$ na liczbę sposobów rozmieszczenia n (nierozróżnialnych) przedmiotów w k pudełkach.
Zasada włączeń i wyłączeń.
Zasada szufladkowa Dirichleta, uogólniona zasada szufladkowa Dirichleta.

Metody

Zliczanie zbioru par na dwa różne sposoby.
Mnóstwo sprytnych pomysłów pokazanych w przykładach.

6. WPROWADZENIE DO GRAFÓW I DRZEW

Z grafami i grafami skierowanymi zapoznaliśmy się już w rozdziale 3; były tam one wykorzystane do ilustrowania relacji. W tym rozdziale omówimy podstawy samej teorii grafów i grafów skierowanych. W paragrafie 6.1 omówimy podstawowe pojęcia i wprowadzimy terminologię, w paragrafach 6.2 i 6.5 omówimy drogi mające pewne specjalne własności. Pozostała część tego rozdziału jest poświęcona drzewom, które będziemy następnie badać w rozdziale 7. Drzewa są grafami, które w naturalny sposób można również traktować jako grafy skierowane. Paragraf 6.4 jest poświęcony drzewom z wyróżnionym korzeniem, które często występują jako struktury danych. W ostatnim paragrafie tego rozdziału zajmujemy się grafami, których krawędzie mają wagi i pokazujemy dwa algorytmy do konstruowania drzew spinających o minimalnej wadze.

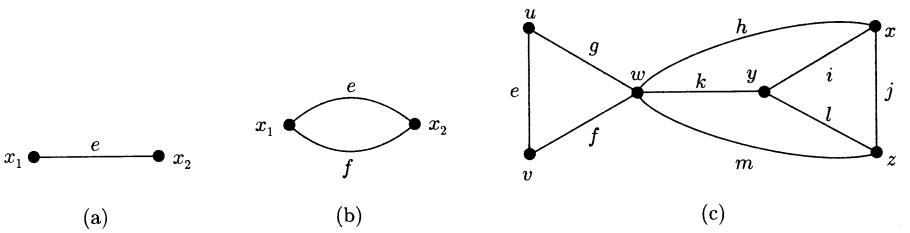
§ 6.1. Grafy

W paragrafie 3.2 wprowadziliśmy pojęcie grafu skierowanego i grafu nieskierowanego. Zajmiemy się teraz bardziej szczegółowym badaniem grafów nieskierowanych. Przypomnijmy z § 3.2, że **drogą długości** n nazywamy ciąg $e_1e_2 \dots e_n$ krawędzi, wraz z ciągiem $v_1v_2 \dots v_{n+1}$ wierzchołków, takim, że $\gamma(e_i) = \{v_i, v_{i+1}\}$ dla $i = 1, 2, \dots, n$. γ jest tu funkcją, która podaje wierzchołki będące końcami każdej krawędzi. Wierzchołki v_i i v_{i+1} mogą być równe, w takim przypadku krawędź e_i jest **pętlą**. Jeśli $v_{n+1} = v_1$, to taką drogę nazywamy **drogą zamkniętą**.

Droga zamknięta może składać się z tego samego ciągu krawędzi, przechodzonych „tam i z powrotem", na przykład $efggfe$. Jednak w naszych dalszych rozważaniach najważniejszymi drogami zamkniętymi będą drogi, w których żadna krawędź się nie powtarza. Drogę nazywamy **drogą prostą**, jeśli jej wszystkie krawędzie są różne. Zatem w drodze prostej żadna krawędź nie jest wykorzystana dwa razy, chociaż taka droga może przechodzić przez ten sam wierzchołek więcej niż jeden raz. Zamkniętą drogę prostą, której ciągiem wierzchołków jest ciąg $x_1 \dots x_n x_1$ nazywamy **cyklem**, jeśli wierzchołki $x_1, \dots, x_n$ są różne. Graf nie zawierający cykli nazywamy **grafem acyklicznym**. Zobaczymy niedługo, że graf jest acykliczny wtedy i tylko wtedy, gdy nie zawiera zamkniętych dróg prostych.

Drogę nazywamy **acykliczną**, jeśli „podgraf" składający się z wierzchołków i krawędzi tej drogi jest acykliczny. W ogólności graf H jest podgrafem grafu G, jeśli $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$ oraz funkcja γ grafu G, określona na $E(G)$, pokrywa się z fukcją γ grafu H, określoną na $E(H)$. Jeśli graf G nie ma krawędzi wielokrotnych i jeśli traktujemy zbiór $E(G)$ jako zbiór jedno- i dwuelementowych podzbiorów $V(G)$, to warunek nałożony na funkcję γ wynika z inkluzji $E(H) \subseteq E(G)$. Wprost z definicji wynika, że jeśli H jest podgrafem grafu G i graf G jest acykliczny, to graf H też jest acykliczny.

PRZYKŁAD 1 (a) Weźmy graf przedstawiony na rysunku 6.1(a). Droga ee wraz z ciągiem wierzchołków $x_1x_2x_1$ jest drogą zamkniętą, ale nie jest cyklem, gdyż nie jest drogą prostą. Podobnie nie jest cyklem droga ee wraz z ciągiem wierzchołków $x_2x_1x_2$. Ten graf jest acykliczny.



Rysunek 6.1

(b) Droga ef z ciągiem wierzchołków $x_1x_2x_1$ w grafie przedstawionym na rysunku 6.1(b) jest cyklem. Podobnie jest cyklem droga ef wraz z ciągiem wierzchołków $x_2x_1x_2$.

(c) W grafie pokazanym na rysunku 6.1(c) droga $efhikg$ długości 6 wraz z ciągiem wierzchołków $uvwxywu$ jest drogą zamkniętą i prostą, ale nie jest cyklem, gdyż nie wszystkie wierzchołki spośród jej pierwszych sześciu wierzchołków u, v, w, x, y i w są różne. Droga, której ciągiem wierzchołków jest $uwvwuvu$ również nie jest cyklem. Graf, jako całość, nie jest acykliczny. Acykliczne nie są również te dwie drogi, ponieważ $uvwu$ jest cyklem w obu ich podgrafach. ∎

Droga $e_1 \dots e_n$ w grafie G wraz z ciągiem wierzchołków $x_1 \dots x_{n+1}$, różnych między sobą, musi oczywiście być drogą prostą, ponieważ żadne dwie krawędzie w niej nie mogą mieć tego samego zbioru końców. Przykład 1(a) pokazuje jednak, że droga zamknięta, w której wierzchołki $x_1, \dots, x_n$ są różne, nie musi być drogą prostą. Ten przykład złej drogi, która jest drogą zamkniętą długości 2, jest tak naprawdę jedynym takim przykładem, jak pokazuje następujące stwierdzenie.

Stwierdzenie 1

Każda droga zamknięta $e_1 \dots e_n$, długości co najmniej 3, o różnych wierzchołkach $x_1, \dots, x_n$, jest cyklem.

Dowód. Musimy tylko pokazać, że krawędzie $e_1, \dots, e_n$ są różne. Ponieważ wierzchołki $x_1, \dots, x_n$ są różne, więc droga $e_1 \dots e_{n-1}$ jest drogą prostą. To znaczy, że wszystkie krawędzie $e_1, \dots, e_{n-1}$ są różne. Ale przecież $\gamma(e_n) = \{x_n, x_1\}$ oraz $\gamma(e_i) = \{x_i, x_{i+1}\}$ dla $i < n$. Ponieważ $n \geq 3$, $e_n \neq e_i$ dla $i < n$, więc ta droga jest drogą prostą. ∎

Dla dróg, które nie są zamknięte, fakt, że wierzchołki są różne, można scharakteryzować w inny sposób.

Stwierdzenie 2

Droga ma wszystkie wierzchołki różne wtedy i tylko wtedy, gdy jest prosta i acykliczna.

Dowód. Weźmy najpierw drogę o różnych wierzchołkach. Musi ona być prosta, jak stwierdziliśmy to wcześniej. Podgraf składający się z wierzchołków i krawędzi tej drogi może być narysowany tak, że wierzchołki leżą na linii prostej, a więc oczywiście jest acykliczny.

W drugą stronę, przypuśćmy, że droga prosta ma ciąg wierzchołków $x_1 \dots x_{n+1}$, przy czym niektóre wierzchołki się powtarzają. Weźmy dwa takie wierzchołki, powiedzmy x_i i x_j, gdzie

$i < j$ oraz różnica $j - i$ jest możliwie najmniejsza. Wtedy wierzchołki $x_i, x_{i+1}, \ldots, x_{j-1}$ są różne, a zatem droga prosta $x_i x_{i+1} \ldots x_{j-1} x_j$ jest cyklem (nawet jeśli $j = i + 1$), więc początkowa droga zawiera cykl. ∎

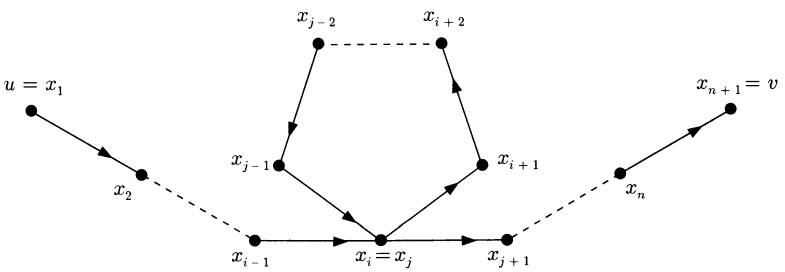
Ze stwierdzenia 2 wynika, że każda prosta droga zamknięta zawiera cykl i tak naprawdę dowód pokazuje, że taki cykl może składać się z kolejnych krawędzi w tej drodze. Inną konsekwencją stwierdzenia 2 jest to, że jeśli graf jest acykliczny, to nie zawiera on cykli i wobec tego nie zawiera prostych dróg zamkniętych. Przyjrzymy się dokładniej takim grafom w następnych paragrafach poświęconych drzewom.

W terminologii z § 3.2, następne twierdzenie mówi, że cykle nie mają wpływu na relację osiągalności.

Twierdzenie 1

Jeśli u i v są różnymi wierzchołkami grafu G i jeśli istnieje w grafie G droga z u do v, to istnieje prosta droga acykliczna z u do v.

Dowód. Spośród wszystkich dróg z u do v w grafie G wybierzmy drogę o najmniejszej długości, powiedzmy o ciągu wierzchołków $x_1 \ldots x_{n+1}$, gdzie $x_1 = u$ i $x_{n+1} = v$. Ze stwierdzenia 2 wynika, że ta droga jest prosta i acykliczna, przy założeniu, że wierzchołki $x_1, \ldots, x_{n+1}$ są różne. Ale w przeciwnym przypadku mielibyśmy $x_i = x_j$ dla pewnych i i j takich, że $1 \leq i < j \leq n+1$. Wtedy droga $x_i x_{i+1} \ldots x_j$ z x_i do x_j byłaby zamknięta (zob. rysunek 6.2), a droga $x_1 \ldots x_i x_{j+1} \ldots x_{n+1}$, otrzymana przez usunięcie fragmentu od x_i do x_j, nadal będzie prowadzić z u do v. Ponieważ droga $x_1 \ldots x_n x_{n+1}$ ma najmniejszą długość, więc ta krótsza droga nie może istnieć. Zatem wierzchołki są różne i najkrótsza droga jest prosta i acykliczna. ∎



Rysunek 6.2

Wniosek

Jeśli e jest krawędzią w zamkniętej drodze prostej w grafie G, to e należy do jakiegoś cyklu.

Dowód. Jeśli e jest pętlą, to nie ma czego dowodzić, więc załóżmy, że krawędź e łączy różne wierzchołki u i v. Usuńmy krawędź e z grafu G. Ponieważ dana droga przechodząca przez u i v jest zamknięta, więc nawet po usunięciu e istnieje droga z u do v; po prostu należy przejść dookoła resztę danej drogi. Na podstawie twierdzenia 1 istnieje prosta droga acykliczna z u do v, nie zawierająca krawędzi e. Połączenie końców takiej drogi krawędzią e utworzy cykl przechodzący przez u i v. ∎

W innych sytuacjach, szczególnie w § 6.3, będziemy usuwać krawędź e z grafu G, by otrzymać nowy graf $G\backslash\{e\}$. Graf $G\backslash\{e\}$ jest podgrafem grafu G, w którym $V(G\backslash\{e\}) = V(G)$ oraz $E(G\backslash\{e\}) = E(G)\backslash\{e\}$.

Następujący fakt będzie potrzebny przy badaniu drzew w § 6.3.

Twierdzenie 2

Jeśli u i v są różnymi wierzchołkami grafu acyklicznego G, to istnieje co najwyżej jedna droga prosta w G prowadząca z u do v.

Dowód. Załóżmy, że twierdzenie jest fałszywe i spośród par wierzchołków połączonych dwiema różnymi drogami prostymi wybierzmy parę (u, v) mającą drogę o najmniejszej długości, łączącą te wierzchołki.

Weźmy dwie drogi proste z u do v, z których jedna jest możliwie najkrótsza. Jeśli te dwie drogi nie mają wspólnych wierzchołków poza u i v, to przejście z u do v wzdłuż jednej drogi i powrót z v do u wzdłuż drugiej drogi da cykl, co jest sprzeczne z założeniem, że graf G jest acykliczny. Zatem obie drogi muszą przechodzić przez co najmniej jeden inny wierzchołek, powiedzmy w. Wtedy drogi z u do v tworzą albo dwie różne drogi proste z u do w, albo dwie różne drogi z w do v. Ale w leży bliżej wierzchołków u i v, niż one względem siebie, co jest sprzeczne z wyborem pary (u, v). ∎

PRZYKŁAD 2

Weźmy graf pokazany na rysunku 6.1(c). Droga $xwyzx$, utworzona z jednej drogi prowadzącej z x do w oraz drugiej prowadzącej z powrotem do x, jest cyklem. Jest też cyklem droga $wuvw$ prowadząca z w do u, a potem z powrotem do w inną drogą.

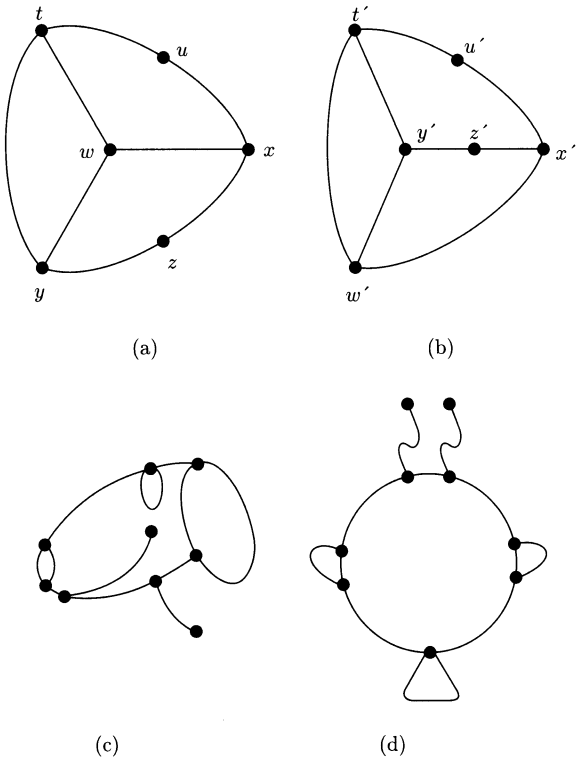
Drogi proste $uwyxwv$ i $uwyzwv$ prowadzą z u do v. Cykl $yxwzy$ jest utworzony z fragmentów tych dróg, jednego z y do w i drugiego prowadzącego z powrotem od w do y. ∎

Często zdarza się, że dwa grafy są „właściwie takie same", pomimo że różnią się nazwami krawędzi i wierzchołków. Zdania ogólne, które można wypowiedzieć o jednym z tych grafów, są również prawdziwe o drugim. Aby te idee uczynić matematycznie ścisłymi, wprowadza się pojęcie izomorfizmu. Mówiąc ogólnie, dwa zbiory wyposażone w pewne struktury matematyczne nazywamy **izomorficznymi**, jeśli istnieje przekształcenie wzajemnie jednoznaczne między tymi zbiorami, zachowujące te struktury (czyli zgodne z nimi). Przypomnijmy, że jeśli G i H są grafami bez krawędzi wielokrotnych, to przyjmujemy, że krawędziami są jedno- i dwuelementowe podzbiory zbiorów wierzchołków. **Izomorfizmem** grafu G na graf H nazywamy przekształcenie wzajemnie jednoznaczne $\alpha : V(G) \to V(H)$ takie, że $\{u, v\}$ jest krawędzią grafu G wtedy i tylko wtedy, gdy $\{\alpha(u), \alpha(v)\}$ jest krawędzią w grafie H. Dwa grafy G i H są **izomorficzne**, co zapisujemy $G \simeq H$, jeśli istnieje izomorfizm α z jednego grafu na drugi; wtedy przekształcenie odwrotne α^{-1} jest również izomorfizmem.

W przypadku grafów mających krawędzie wielokrotne sytuacja jest bardziej skomplikowana. Wymagamy wtedy dwóch przekształceń wzajemnie jednoznacznych α: $V(G) \to V(H)$ i β: $E(G) \to E(H)$ takich, że krawędź e ze zbioru $E(G)$ łączy wierzchołki u i v ze zbioru $V(G)$ wtedy i tylko wtedy, gdy odpowiadająca jej krawędź $\beta(e)$ łączy wierzchołki $\alpha(u)$ i $\alpha(v)$. Zatem dwa grafy są izomorficzne wtedy i tylko wtedy, gdy mają ten sam rysunek, z wyjątkiem oznaczeń krawędzi i wierzchołków. Ta obserwacja przydaje się głównie jako sposób sprawdzania, czy dane przekształcenie jest izomorfizmem, polegający na rysowaniu odpowiadających sobie rysunków obu grafów.

PRZYKŁAD 3 Przekształcenie α takie, że $\alpha(t) = t'$, $\alpha(u) = u', \ldots, \alpha(z) = z'$ jest izomorfizmem grafów przedstawionych na rysunkach 6.3(a) i 6.3(b). Grafy pokazane na rysunkach 6.3(c) i 6.3(d) są również izomorficzne między sobą, ale nie są izomorficzne z grafami na rysunkach 6.3(a) i (b). ∎

Aby rozróżnić grafy przedstawione na rysunkach 6.3(a) i 6.3(c), możemy po prostu policzyć wierzchołki. Grafy izomorficzne mają tyle samo wierzchołków i tyle samo krawędzi. Te dwie liczby są przykładami **niezmienników izomorfizmu** dla gra-



Rysunek 6.3

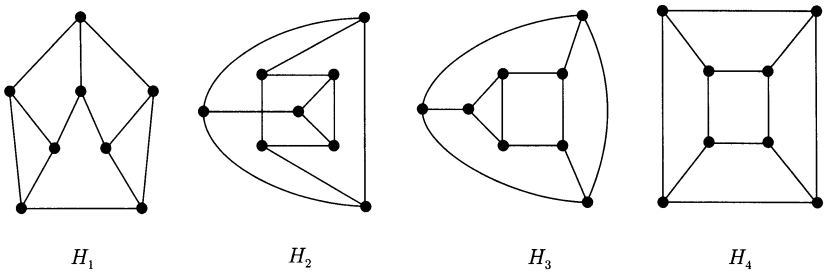
fów. Innymi przykładami niezmienników są: liczba pętli i liczba dróg prostych danej długości.

Często przydaje się zliczenie krawędzi schodzących się w konkretnym wierzchołku. Aby otrzymać właściwą liczbę, musimy odróżnić pętle od krawędzi mających dwa różne końce. Definiujemy **stopień wierzchołka** v, oznaczany symbolem $\deg(v)$, jako liczbę dwuwierzchołkowych krawędzi z v jako jednym z wierzchołków, plus podwojona liczba pętli o wierzchołku v. Liczba wierzchołków stopnia k w grafie G, oznaczana symbolem $D_k(G)$, jest niezmiennikiem izomorfizmu. Niezmiennikiem jest również **ciąg liczb wierzchołków kolejnych stopni** $(D_0(G), D_1(G), D_2(G), \ldots)$.

PRZYKŁAD 4 (a) Każdy z grafów pokazanych na rysunkach 6.3(a) i 6.3(b) ma ciąg liczb wierzchołków kolejnych stopni $(0, 0, 2, 4, 0, 0, \ldots)$. Grafy na rysunkach 6.3(c) i 6.3(d) mają ciąg liczb wierzchołków kolejnych stopni $(0, 2, 0, 6, 1, 0, 0, \ldots)$.

(b) Wszystkie cztery grafy na rysunku 6.4 mają 8 wierzchołków stopnia 3 i nie mają żadnych innych wierzchołków. Okazuje

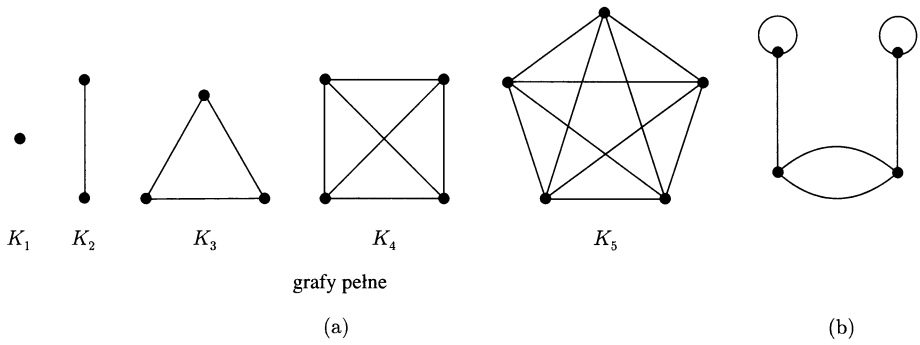
się, że $H_1 \simeq H_2 \simeq H_3$, ale żaden z tych grafów nie jest izomorficzny z grafem H_4. To, że dwa grafy mają ten sam ciąg liczb wierzchołków kolejnych stopni, nie gwarantuje, że są one izomorficzne. ■



Rysunek 6.4

Grafy, w których wszystkie wierzchołki mają ten sam stopień, na przykład takie jak grafy przedstawione na rysunku 6.4, nazywamy **grafami regularnymi**. Jak pokazuje ten przykład, grafy regularne o tej samej liczbie wierzchołków nie muszą być izomorficzne. Grafy nie mające pętli czy krawędzi wielokrotnych i takie, w których każdy wierzchołek jest połączony krawędzią z każdym innym, nazywamy **grafami pełnymi**. Graf pełny o n wierzchołkach ma wszystkie wierzchołki stopnia $n-1$, więc taki graf jest grafem regularnym. Wszystkie grafy pełne o n wierzchołkach są ze sobą izomorficzne, a więc każdy z nich oznaczamy symbolem K_n.

PRZYKŁAD 5 Rysunek 6.5(a) pokazuje pięć pierwszych grafów pełnych. Graf na rysunku 6.5(b) ma cztery wierzchołki, z których każdy jest stopnia 3, ale nie jest on grafem pełnym. ■



Rysunek 6.5

Graf pełny K_n zawiera podgrafy izomorficzne z grafami K_m dla $m = 1, 2, \ldots, n$. Taki podgraf można otrzymać, wybierając dowolnych m spośród n wierzchołków i biorąc wszystkie krawędzie grafu K_n, łączące te wierzchołki. Zatem graf K_5 zawiera $\binom{5}{2} = 10$ podgrafów izomorficznych z K_2, $\binom{5}{3} = 10$ podgrafów izomorficznych z K_3 (tzn. trójkątów) i $\binom{5}{4} = 5$ podgrafów izomorficznych z K_4. W rzeczywistości, każdy graf mający n lub mniej wierzchołków i nie mający pętli i krawędzi wielokrotnych jest izomorficzny z podgrafem grafu K_n.

Grafy pełne są w dużym stopniu symetryczne. Każda permutacja α wierzchołków grafu pełnego jest izomorfizmem tego grafu na ten sam graf, ponieważ zarówno $\{u, v\}$, jak i $\{\alpha(u), \alpha(v)\}$ są krawędziami, gdy tylko $u \neq v$.

Następne twierdzenie pokazuje zależność między stopniami wierzchołków i liczbą krawędzi grafu.

Twierdzenie 3

(a) Suma stopni wierzchołków grafu jest dwa razy większa od liczby krawędzi. To znaczy

$$\sum_{v \in V(G)} \deg(v) = 2 \cdot |E(G)|.$$

(b)

$$D_1(G) + 2D_2(G) + 3D_3(G) + 4D_4(G) + \ldots = 2 \cdot |E(G)|.$$

Dowód. (a) Każda krawędź, niezależnie od tego, czy jest pętlą, czy nie, dodaje 2 do sumy stopni.

(b) W całkowitej sumie stopni udział $D_k(G)$ wierzchołków stopnia k wynosi $k \cdot D_k(G)$. ∎

PRZYKŁAD 6

(a) Graf przedstawiony na rysunku 6.3(c) (i graf z nim izomorficzny przedstawiony na rysunku 6.3(d)) ma wierzchołki stopni 1, 1, 3, 3, 3, 3, 3, 3 oraz 4 i ma 12 krawędzi. Ciągiem liczb wierzchołków kolejnych stopni jest $(0, 2, 0, 6, 1, 0, 0, 0, \ldots)$. Oczywiście

$$1 + 1 + 3 + 3 + 3 + 3 + 3 + 3 + 4 = 2 \cdot 12 = 2 + 2 \cdot 0 + 3 \cdot 6 + 4 \cdot 1.$$

(b) Grafy przedstawione na rysunku 6.4 mają 8 wierzchołków stopnia 3 i 12 krawędzi. Ich ciągiem liczb wierzchołków kolejnych stopni jest $(0, 0, 0, 8, 0, 0, 0, \ldots)$ oraz równość

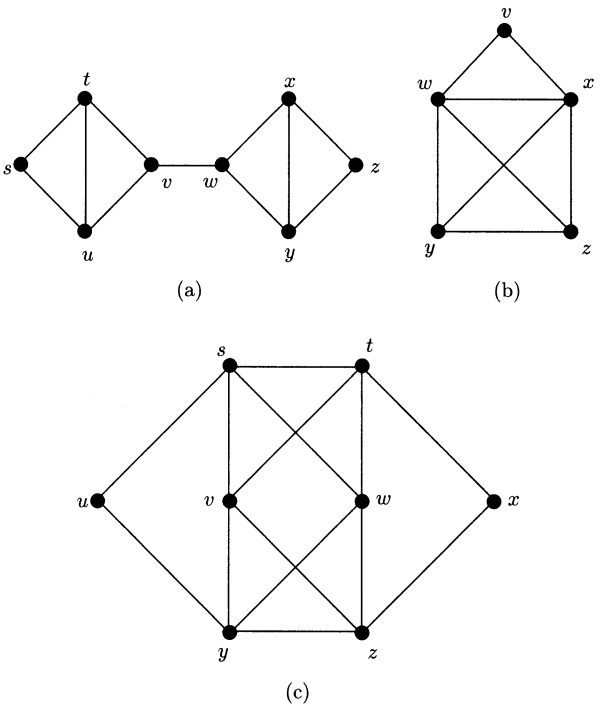
$$2 \cdot 12 = 0 + 2 \cdot 0 + 3 \cdot 8$$

potwierdza prawdziwość twierdzenia 3(b) w tym przypadku.

(c) Graf pełny K_n ma n wierzchołków, każdy z nich jest stopnia $n-1$ i graf ma $n(n-1)/2$ krawędzi. ∎

ĆWICZENIA DO § 6.1

1. Dla grafu przedstawionego na rysunku 6.6(a) podaj ciąg wierzchołków najkrótszej drogi łączącej następujące pary wierzchołków i podaj jej długość:
(a) s i v, (b) s i z, (c) u i y, (d) v i w.



Rysunek 6.6

2. Dla każdej pary wierzchołków z ćwiczenia 1 podaj ciąg wierzchołków najdłuższej łączącej je drogi, w której krawędzie się nie powtarzają. Czy istnieje najdłuższa łącząca je droga, jeśli krawędzie mogą się powtarzać?

3. Czy następujące zdania są prawdziwe czy fałszywe? „Prawdziwe" oznacza „prawdziwe we wszystkich rozważanych przypadkach". Weźmy pod uwagę dowolny graf.
(a) Jeśli istnieje krawędź z wierzchołka u do wierzchołka v, to istnieje krawędź z v do u.

(b) Jeśli istnieje krawędź z wierzchołka u do wierzchołka v i istnieje krawędź z wierzchołka v do wierzchołka w, to istnieje krawędź z u do w.

4. Powtórz ćwiczenie 3, zastępując wszędzie słowo „krawędź" słowem „droga".

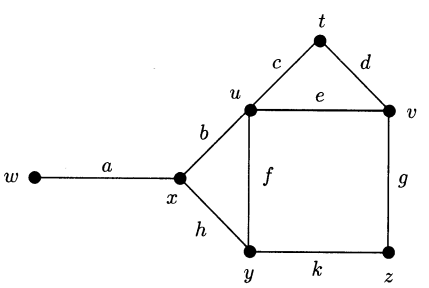
5. Powtórz ćwiczenie 3, zastępując wszędzie słowo „krawędź" słowem „droga o długości parzystej".

6. (a) Sprawdź twierdzenie 3(a) dla każdego z grafów przedstawionych na rysunku 6.6, obliczając
(i) sumę stopni wszystkich wierzchołków,
(ii) liczbę krawędzi.
(b) Czy graf może mieć nieparzystą liczbę wierzchołków nieparzystego stopnia?

7. Podaj przykład grafu z wierzchołkami x, y i z, mającego wszystkie trzy podane własności:
(i) istnieje cykl przechodzący przez wierzchołki x i y;
(ii) istnieje cykl przechodzący przez wierzchołki y i z;
(iii) nie ma cykli przechodzących przez x i z.

8. Przypuśćmy, że cykl zawiera pętlę. Jaka jest jej długość? Czy cykl może zawierać dwie pętle?

9. (a) Podaj tablicę wartości funkcji γ dla grafu G przedstawionego na rysunku 6.7.
(b) Wypisz krawędzie tego grafu, traktowane jako podzbiory zbioru $V(G)$. Na przykład $a = \{w, x\}$.



Rysunek 6.7

10. Narysuj rysunek grafu G, gdzie $V(G) = \{x, y, z, w\}$, $E(G) = \{a, b, c, d, f, g, h\}$ oraz funkcja γ jest dana w postaci tablicy:

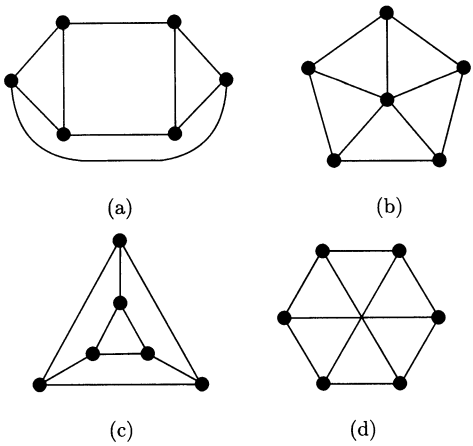
e	a	b	c	d	f	g	h
$\gamma(e)$	$\{x, y\}$	$\{x, y\}$	$\{w, x\}$	$\{w, y\}$	$\{y, z\}$	$\{y, z\}$	$\{w, z\}$

11. W każdej części tego ćwiczenia podane są dwie drogi, które łączą parę punktów w grafie przedstawionym na rysunku 6.7. Użyj idei przedstawionej w dowodzie twierdzenia 2 do skonstruowania cykli z ciągów krawędzi w tych dwóch drogach.
(a) *eba* oraz *gkha*,
(b) *abcdgk* oraz *ah*,
(c) *eba* oraz *dcfha*.

12. (a) Zrób rysunki wszystkich czternastu grafów o trzech wierzchołkach i trzech krawędziach. „Wszystkie" oznacza tutaj, że każdy taki graf jest izomorficzny z jednym z tych czternastu grafów, a żadne dwa z tych czternastu nie są izomorficzne ze sobą.
(b) Zrób rysunki wszystkich grafów o czterech wierzchołkach i czterech krawędziach, które nie mają pętli i krawędzi wielokrotnych.
(c) Wypisz cztery grafy z punktów (a) i (b), które są regularne.

13. (a) Zrób rysunki wszystkich pięciu grafów regularnych o czterech wierzchołkach, z których każdy ma stopień 2.
(b) Zrób rysunki wszystkich grafów regularnych o czterech wierzchołkach, z których każdy ma stopień 3 i nie ma pętli i krawędzi wielokrotnych.
(c) Zrób rysunki wszystkich grafów regularnych o pięciu wierzchołkach, z których każdy ma stopień 3.

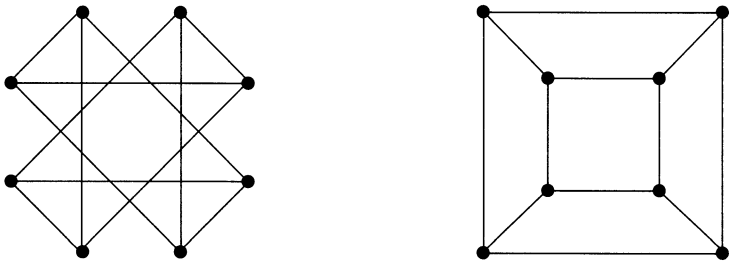
14. Przypuśćmy, że graf H jest izomorficzny z grafem G przedstawionym na rysunku 6.7.
(a) Ile wierzchołków stopnia 1 ma graf H?
(b) Podaj ciąg liczb wierzchołków kolejnych stopni grafu H.
(c) Ile jest różnych izomorfizmów z grafu G na G? Odpowiedź uzasadnij.
(d) Ile jest różnych izomorfizmów z grafu G na graf H?



Rysunek 6.8

15. Które pary grafów z przedstawionych na rysunku 6.8 są izomorficzne, jeśli w ogóle takie istnieją? Uzasadnij swoją odpowiedź opisując izomorfizm lub wyjaśniając, dlaczego taki izomorfizm nie istnieje.

16. Opisz izomorfizm między grafami pokazanymi na rysunku 6.9.



Rysunek 6.9

17. Weźmy graf pełny K_8 o wierzchołkach $v_1, v_2, \ldots, v_8$.
(a) Ile podgrafów grafu K_8 jest izomorficznych z grafem K_5?
(b) Ile istnieje dróg prostych, prowadzących z wierzchołka v_1 do wierzchołka v_2, mających trzy lub mniej krawędzi?
(c) Ile w sumie dróg prostych o trzech lub mniej krawędziach istnieje w grafie K_8?

18. (a) Graf o 21 krawędziach ma 7 wierzchołków stopnia 1, 3 wierzchołki stopnia 2, 7 wierzchołków stopnia 3, a pozostałe wierzchołki mają stopień 4. Ile ma on wierzchołków?
(b) Jaka byłaby odpowiedź w ćwiczeniu (a), gdyby graf miał też 6 wierzchołków stopnia 0?

19. Które z następujacych ciągów są ciągami liczb wierzchołków kolejnych stopni grafów? W każdym przypadku albo narysuj graf o danym ciągu liczb wierzchołków kolejnych stopni tego grafu, albo wyjaśnij, dlaczego taki graf nie istnieje.
(a) $(1, 1, 0, 3, 1, 0, 0, \ldots)$
(b) $(4, 1, 0, 3, 1, 0, 0, \ldots)$
(c) $(0, 1, 0, 2, 1, 0, 0, \ldots)$
(d) $(0, 0, 2, 2, 1, 0, 0, \ldots)$
(e) $(0, 0, 1, 2, 1, 0, 0, \ldots)$
(f) $(0, 1, 0, 1, 1, 1, 0, \ldots)$
(g) $(0, 0, 0, 4, 0, 0, 0, \ldots)$
(h) $(0, 0, 0, 0, 5, 0, 0, \ldots)$

20. Pokaż, że droga w grafie G jest cyklem wtedy i tylko wtedy, gdy jest możliwe przypisanie zwrotów krawędziom grafu G tak, by droga ta była cyklem (skierowanym) w otrzymanym grafie skierowanym.

21. Pokaż, że każdy skończony graf, w którym każdy wierzchołek ma stopień co najmniej 2, zawiera cykl.

22. Pokaż, że każdy graf o n wierzchołkach i co najmniej n krawędziach zawiera cykl. *Wskazówka*: Zastosuj indukcję względem n i skorzystaj z ćwiczenia 21.

23. Pokaż, że

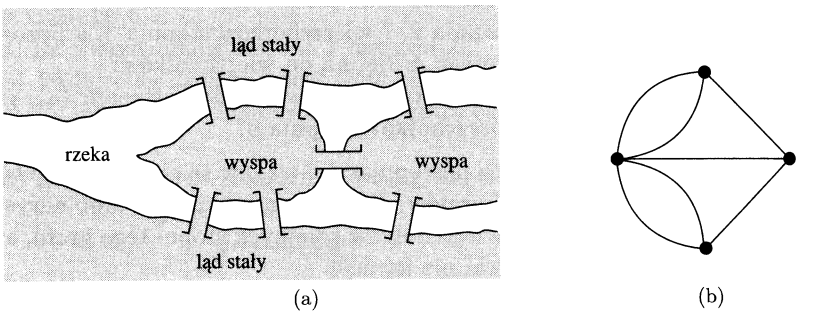
$$2|E(G)| - |V(G)| = -D_0(G) + D_2(G) + 2D_3(G) + \ldots + (k-1)D_k(G) + \ldots$$

24. (a) Niech $\mathcal{S}$ będzie zbiorem grafów. Pokaż, że izomorfizm $\simeq$ jest relacją równoważności w zbiorze $\mathcal{S}$.

(b) Ile klas równoważności istnieje w zbiorze $\mathcal{S}$ składającym się z czterech grafów przedstawionych na rysunku 6.4?

§ 6.2. Zagadnienia związane z poruszaniem się po krawędziach

Jednym z najstarszych problemów dotyczących grafów jest problem mostów królewieckich. Czy można przejść się po mieście pokazanym na rysunku 6.10(a), przechodząc przez każdy most



Rysunek 6.10

dokładnie jeden raz i wrócić do domu? Szwajcarski matematyk Leonhard Euler rozwiązał ten problem w 1736 roku. Zbudował on graf pokazany na rysunku 6.10(b) zastępując obszary lądu wierzchołkami, a mosty łączącymi je krawędziami. Powstało wtedy pytanie: czy istnieje droga zamknięta w tym grafie przechodząca przez każdą krawędź dokładnie jeden raz? Taką drogę nazywamy **cyklem Eulera** w grafie. Ogólniej, droga prosta zawierająca wszystkie krawędzie grafu G jest nazywana **drogą Eulera** w G.

Euler pokazał, że nie ma cyklu Eulera dla grafu mostów królewieckich, w którym wszystkie wierzchołki mają stopień nieparzysty, dowodząc następującego prostego faktu.

Twierdzenie 1

Graf, który ma cykl Eulera, musi mieć wszystkie wierzchołki stopnia parzystego.

Dowód. Wychodząc z dowolnego wierzchołka na cyklu Eulera, poruszamy się po tym cyklu, przechodząc od wierzchołka do wierzchołka i wycierając każdą krawędź, po której przeszliśmy. Kiedy przechodzimy przez wierzchołek, wycieramy jedną krawędź dochodzącą do tego wierzchołka i jedną wychodzącą z niego lub wycieramy pętlę. W każdym przypadku to wycieranie spowoduje zmniejszenie stopnia wierzchołka o 2. Ostatecznie każda krawędź zostanie usunięta i wszystkie wierzchołki będą miały stopień 0. Zatem wszystkie wierzchołki musiały mieć na początku stopień parzysty. ∎

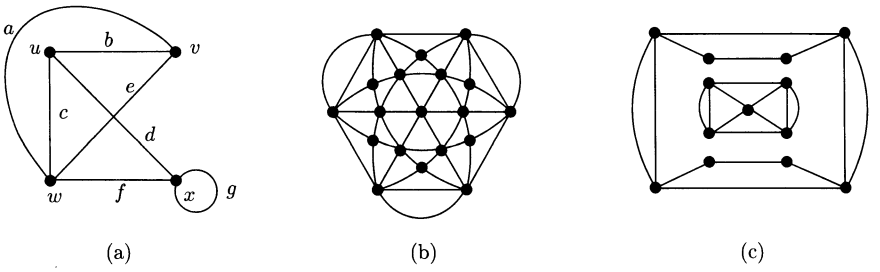
Wniosek

Graf G mający drogę Eulera ma albo dwa wierzchołki stopnia nieparzystego, albo nie ma w ogóle wierzchołków stopnia nieparzystego.

Dowód. Przypuśćmy, że graf G ma drogę Eulera zaczynającą się w wierzchołku u i kończącą się w wierzchołku v. Jeśli $u = v$, to droga jest zamknięta i twierdzenie 1 mówi, że wszystkie wierzchołki mają stopień parzysty. Jeśli $u \neq v$, to tworzymy nową krawędź e łączącą u i v. Nowy graf $G \cup \{e\}$ ma cykl Eulera składający się z drogi Eulera w G i krawędzi e, więc wszystkie wierzchołki grafu $G \cup \{e\}$ mają stopień parzysty. Usuwamy krawędź e. Wtedy u i v są jedynymi wierzchołkami grafu $G = (G \cup \{e\}) \setminus \{e\}$ stopnia nieparzystego. ∎

PRZYKŁAD 1

Graf pokazany na rysunku 6.11(a) nie ma cyklu Eulera, ponieważ u i v mają stopień nieparzysty, ale droga $bacdgfe$ jest drogą Eulera. Graf przedstawiony na rysunku 6.11(b) ma wszystkie wierzchołki stopnia parzystego i rzeczywiście ma cykl Eulera. Graf przedstawiony na rysunku 6.11(c) ma wszystkie wierzchołki stopnia parzystego, ale nie ma cyklu Eulera z oczywistego powodu, ponieważ składa się on z dwóch podgrafów nie połączonych ze sobą. Każdy z tych podgrafów, jednakże, ma swój własny cykl Eulera. ∎



(a) (b) (c)

Rysunek 6.11

Twierdzenie 1 mówi, że warunek parzystości stopni wierzchołków jest warunkiem koniecznym istnienia cyklu Eulera. Głównym osiągnięciem Eulera w rozwiązaniu tego problemu było udowodnienie, że z pominięciem oczywistych trudności, na które natknęliśmy się w przykładzie z rysunku 6.11(c), ten warunek jest również warunkiem wystarczającym istnienia cyklu Eulera.

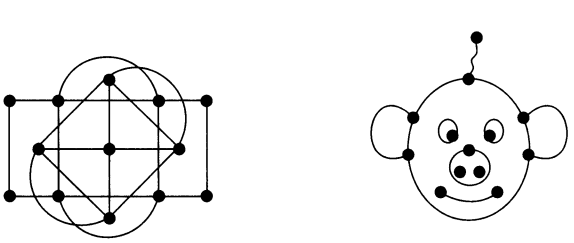
Wprowadzimy terminologię do opisania tych wyjątkowych przypadków. Graf jest **spójny**, jeśli każda para różnych wierzchołków jest połączona drogą w tym grafie. Grafy na rysunkach 6.11(a) i 6.11(b) są spójne, a graf na rysunku 6.11(c) nie jest spójny. Spójny podgraf grafu G, który nie jest zawarty w większym spójnym podgrafie grafu G, nazywamy **składową grafu** G. Składowa zawierająca dany wierzchołek v, wraz z v zawiera wszystkie wierzchołki i krawędzie dróg zaczynających się w wierzchołku v.

PRZYKŁAD 2 (a) Grafy przedstawione na rysunkach 6.11(a) i 6.11(b) są spójne. W takich przypadkach graf ma jedną składową, mianowicie cały graf.

(b) Graf przedstawiony na rysunku 6.11(c) ma dwie składowe, jedną narysowaną na zewnątrz, a drugą wewnątrz. Inne przedstawienie tego grafu jest pokazane na rysunku 6.12(a). Na tym rysunku nie ma składowej „wewnętrznej", ale oczywiście nadal są dwie składowe.

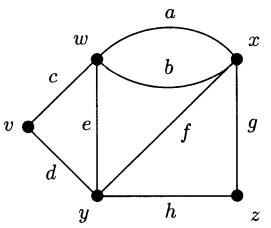
Graf przedstawiony na rysunku 6.12(b) ma siedem składowych, z których dwie są izolowanymi wierzchołkami. ∎

Twierdzenie Eulera mówi, że skończony graf spójny, w którym każdy wierzchołek ma stopień parzysty, ma cykl Eulera. Aby naprawdę zrozumieć to twierdzenie, powinniśmy umieć znaleźć dowód lub opracować algorytm czy procedurę, która zawsze w wyniku dawałaby cykl Eulera. W rzeczywistości te dwa podejścia są ściśle ze sobą związane. Pełne zrozumienie dowodu czę-



(a) (b)

Rysunek 6.12



Rysunek 6.13

sto prowadzi do skonstruowania algorytmu, a za każdym algorytmem stoi dowód. A oto proste wyjaśnienie twierdzenia Eulera, które zilustrujemy rysunkiem 6.13. Zacznijmy od dowolnego wierzchołka, powiedzmy w i dowolnej krawędzi wychodzącej z niego, powiedzmy a. Jej drugi koniec, w tym przypadku x, ma stopień parzysty i był użyty nieparzystą liczbę razy (raz), a więc istnieje nie wykorzystana krawędź wychodząca z x. Wybierzmy taką krawędź, na przykład b. Postępujmy dalej w ten sposób. Ten proces nie zakończy się dotąd, aż nie osiągniemy początkowego wierzchołka w, gdyż za każdym razem, gdy dochodzimy do innego wierzchołka, wykorzystujemy tylko nieparzystą liczbę krawędzi wychodzących z niego. W naszym przykładzie ten algorytm mógł zacząć od wybrania krawędzi abe i wierzchołków $wxwy$. W wierzchołku y możemy wybrać jedną z trzech krawędzi: d, f lub h. Jeśli wybierzemy f, reszta tego procesu jest wyznaczona jednoznacznie. Otrzymujemy cykl Eulera $abefghdc$ wraz z ciągiem wierzchołków $wxwyxzyvw$.

Proste, prawda? Tak, zbyt proste. Co by się stało, gdybyśmy po dojściu pierwszy raz do wierzchołka y wybrali krawędź d? Po przejściu krawędzi c znaleźlibyśmy się w ślepym zaułku w wierzchołku w i nasza droga $abedc$ nie przeszłaby przez krawędzie f, g i h. Nasze wyjaśnienie i nasz algorytm musiały być zbyt proste.

W naszym przykładzie jest jasne, że powinniśmy unikać krawędzi d, gdy po raz pierwszy doszliśmy do wierzchołka y, ale dlaczego? Jaka ogólna zasada mogłaby nas przestrzec przed tym wyborem? Pomyśl o tym. Powrócimy do tej sprawy po podaniu niekonstruktywnego dowodu twierdzenia Eulera.

Twierdzenie 2 (twierdzenie Eulera)

Skończony graf spójny, w którym każdy wierzchołek ma stopień parzysty, ma cykl Eulera.

Dowód. Załóżmy, że G jest skończonym grafem spójnym, w którym każdy wierzchołek ma stopień parzysty. Jeśli graf G ma tylko jeden wierzchołek, to twierdzenie jest trywialne, a więc załóżmy, że $|V(G)| \geq 2$. Ponieważ graf G jest spójny, więc każdy wierzchołek ma stopień co najmniej 1, a zatem co najmniej 2, gdyż wszystkie stopnie są parzyste.

Niech $v_1, v_2, \ldots, v_n$ będzie ciągiem wierzchołków najdłuższej możliwej drogi, w której wszystkie krawędzie są różne. Ponieważ zbiór $E(G)$ jest skończony, taka droga musi istnieć. Twierdzimy, że $v_n = v_1$. Aby tego dowieść, rozważmy graf G' otrzymany przez usunięcie krawędzi tej drogi i zobaczmy, w jaki sposób zmienią się stopnie wierzchołków. Wierzchołek v_n może znajdować się lub nie wśród wierzchołków $v_2, \ldots, v_{n-1}$. Jeśli znajduje się wśród nich, to każde takie wystąpienie zmniejsza stopień v_n w grafie G' o 2: o 1 przez usunięcie krawędzi wchodzącej do v_n i o 1 przez usunięcie krawędzi wychodzącej. Ponieważ usunięcie ostatniej krawędzi na tej drodze zmniejsza stopień v_n o 1, więc gdyby wierzchołek v_n był różny od v_1, jego stopień w G' byłby nieparzysty. W takim przypadku moglibyśmy dodać do naszej drogi nie wykorzystaną krawędź wychodzącą z wierzchołka v_n i otrzymalibyśmy drogę dłuższą, też składającą się z różnych krawędzi, co przeczy naszemu wyborowi najdłuższej drogi. Zatem $v_n = v_1$.

Następnie pokażemy, że nasza najdłuższa droga przechodzi przez każdy wierzchołek grafu G. Gdyby tak nie było, to ze spójności grafu wynikałoby, że istnieje droga od pewnego nie odwiedzonego wierzchołka do jakiegoś wierzchołka zbioru $\{v_1, \ldots, v_n\}$. Pierwsza krawędź tej drogi należy do G'. Ostatnia krawędź e tej drogi, znajdująca się w G', musi dochodzić do któregoś wierzchołka zbioru $\{v_1, \ldots, v_n\}$, powiedzmy do wierzchołka v_i. Jeśli przejdziemy teraz naszą najdłuższą drogę od wierzchołka v_i do v_i, a następnie dodamy krawędź e, otrzymamy dłuższą drogę o różnych krawędziach, co znów jest sprzeczne z naszym wyborem najdłuższej drogi.

Aby zakończyć dowód, pokażemy, że nasza najdłuższa droga przechodzi przez każdą krawędź grafu G, a więc jest cyklem Eulera. Tak naprawdę, jeśli e jest krawędzią, która została pominięta, któryś z wierzchołków v_i musi być jednym z jej końców (a nawet dwa takie v_i, chyba, że e jest pętlą). Tak jak w poprzednim akapicie, krawędź e może być dołączona do najdłuższej drogi i w ten sposób otrzymamy dłuższą drogę, a więc sprzeczność. ∎

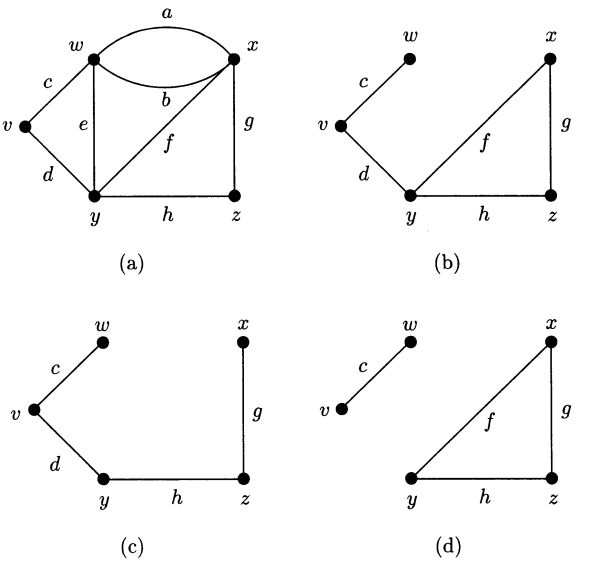
Wniosek

Skończony graf spójny, mający dokładnie dwa wierzchołki stopnia nieparzystego, ma drogę Eulera.

Dowód. Przypuśćmy, że u i v są wierzchołkami stopni nieparzystych. Utwórzmy nową krawędź e, łączącą je. Wtedy graf $G \cup \{e\}$ ma wszystkie wierzchołki stopni parzystych, a więc ma cykl Eulera na podstawie twierdzenia 2. Usuńmy znowu krawędź e. To, co pozostanie z cyklu Eulera, jest drogą Eulera w grafie G. ∎

Powtarzając w zasadzie ten sam dowód, pokażemy w § 8.1, że analogiczne twierdzenie jest również prawdziwe dla grafów skierowanych. Twierdzenie to mówi nam, kiedy graf ma cykl Eulera, ale nie mówi, jak go znaleźć. Chcielibyśmy mieć algorytm, który tworzy cykl lub drogę Eulera, krawędź po krawędzi, zamiast niejasnej procedury, której użyliśmy do grafu z rysunku 6.13. Popatrzmy jeszcze raz na cykl Eulera, który znaleźliśmy na rysunku 6.13, powtórzonym na rysunku 6.14(a). W chwili, gdy wybieramy kolejną krawędź, usuwajmy ją z grafu i rozważajmy otrzymane w ten sposób podgrafy. Rozpoczęliśmy naszą drogę, wybierając krawędzie *abe*. Na rysunku 6.14(b) pokazany jest graf, z którego usunięto te krawędzie. W naszej pomyślnej próbie znalezienia cyklu Eulera wybraliśmy następnie krawędź f oraz zauważyliśmy, że gdybyśmy wybrali krawędź d, skazalibyśmy się na niepowodzenie. Na rysunku 6.14(c) widzimy ten graf po usunięciu również krawędzi f, a na rysunku 6.14(d) widzimy ten graf po usunięciu krawędzi d zamiast f. Widzimy różnicę: po usunięciu krawędzi d graf staje się niespójny, podczas gdy usunięcie krawędzi f nie powoduje tego. Na tym polega istota algorytmu, który działa poprawnie. W każdym wierzchołku algorytm Fleury'ego każe nam wybrać, jeśli to możliwe, taką krawędź, po usunięciu której graf pozostanie spójny. Jeśli nie jest to możliwe, to pozostaje dokładnie jedna krawędź. (Jest to nietrywialny fakt, który wymaga dowodu. Tu właśnie wkracza matematyka).

Wybieramy tę krawędź; wtedy z grafu usuwamy tę krawędź wraz z wierzchołkiem.



Rysunek 6.14

Sformalizujemy teraz ten algorytm. Kiedy algorytm zakończy działanie, ciąg ES będzie ciągiem krawędzi drogi lub cyklu Eulera, a VS będzie ciągiem wierzchołków tej drogi lub cyklu.

Algorytm Fleury'ego

Krok 1. Wybierz dowolny wierzchołek v nieparzystego stopnia, jeśli taki istnieje. W przeciwnym przypadku wybierz dowolny wierzchołek v. Niech $VS = v$ i niech $ES = \lambda$ (ciąg pusty).

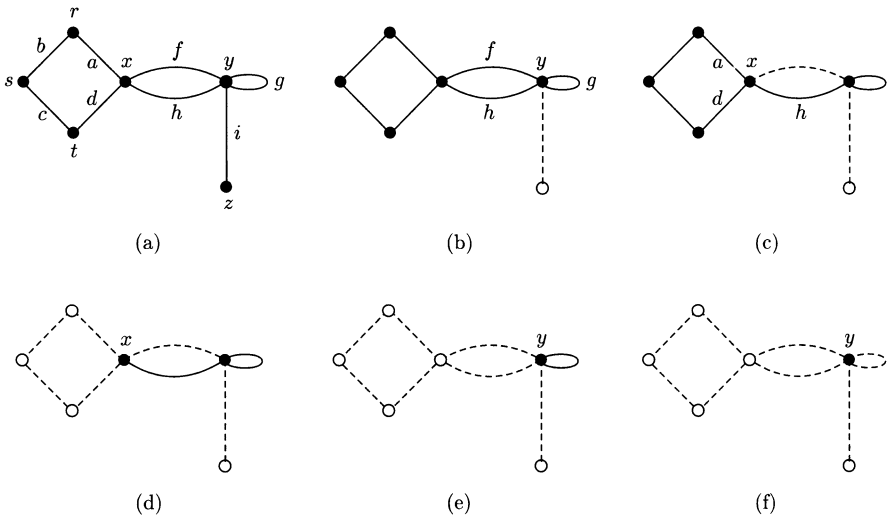
Krok 2. Jeśli z wierzchołka v nie wychodzi już żadna krawędź, zatrzymaj się.

Krok 3. Jeśli pozostała dokładnie jedna krawędź wychodząca z wierzchołka v, powiedzmy krawędź e z wierzchołka v do w, to usuń e z $E(G)$ oraz v z $V(G)$ i przejdź do kroku 5.

Krok 4. Jeśli została więcej niż jedna krawędź wychodząca z wierzchołka v, wybierz taką krawędź, powiedzmy e z v do w, po usunięciu której graf pozostanie spójny; następnie usuń e z $E(G)$.

Krok 5. Dołącz w na końcu ciągu VS, dołącz e na końcu ciągu ES, zastąp v wierzchołkiem w i przejdź do kroku 2. ∎

Zanim wyjaśnimy, dlaczego ten algorytm działa poprawnie, pokażemy przykład ilustrujący jego działanie.



Rysunek 6.15

PRZYKŁAD 3 Weźmy graf pokazany na rysunku 6.15(a). Ten graf nie ma cyklu Eulera, ale ma drogę Eulera łączącą wierzchołki z i y stopni nieparzystych. Wybieramy na początku wierzchołek z, tzn. w kroku 1 kładziemy $v = z$. Zatem $VS = z$ i $ES = \lambda$. Jedyną krawędzią wychodzącą z wierzchołka z jest i. Idziemy zatem do kroku 3, wybieramy $e = i$ i $w = y$, usuwamy i z $E(G)$, usuwamy z z $V(G)$ oraz przechodzimy do kroku 5. Wtedy $VS = zy$ i $ES = i$. Nowy graf, po usunięciu i oraz z jest pokazany na rysunku 6.15(b). Przyjmujemy $v = y$ i powracamy do kroku 2. Z wierzchołka v wychodzą trzy krawędzie, więc przechodzimy do kroku 4.

Teraz jako e możemy wybrać f, g lub h. Wybierzmy w kroku 4 krawędź $e = f$. W kroku 5 otrzymamy $VS = zyx$, $ES = if$ oraz $v = x$ i nowy graf G jest pokazany na rysunku 6.15(c). Powracamy do kroku 2. Znów z wierzchołka v wychodzą trzy krawędzie, więc przechodzimy do kroku 4.

Teraz krawędzią e może być a lub d, nie może natomiast być nią krawędź h, gdyż po usunięciu krawędzi h graf przestałby być spójny. Wybierzmy $e = a$; w kroku 5 otrzymamy $VS = zyxr$, $ES = ifa$ i $v = r$.

Następne trzy ruchy są wymuszone. Każdy prowadzi do kroku 3 i usunięcia kolejnej krawędzi wraz z wierzchołkiem. Otrzymu-

jemy $VS = zyxrstx$, $ES = ifabcd$, $v = x$ oraz graf przedstawiony na rysunku 6.15(d).

Ostatnie dwa ruchy są również wymuszone i prowadzą do grafów pokazanych na rysunkach 6.15(e) i 6.15(f). Ostatecznie ciągami wierzchołków i krawędzi są $VS = zyxrstxyy$ i $ES = ifabcdhg$. ∎

Do dowodu tego, że algorytm Fleury'ego działa poprawnie, potrzebujemy następującego twierdzenia. Będzie ono również przydatne, gdy w § 6.3 będziemy omawiać drzewa.

Twierdzenie 3

Niech e będzie krawędzią grafu spójnego G. Następujące warunki są równoważne:

(a) graf $G\backslash\{e\}$ jest spójny;

(b) e jest krawędzią w pewnym cyklu w grafie G;

(c) e jest krawędzią w pewnej zamkniętej drodze prostej w grafie G.

Dowód. Zauważmy najpierw, że jeśli e jest pętlą, to graf $G\backslash\{e\}$ jest spójny i krawędź e sama w sobie jest cyklem. Ponieważ cykle są zamkniętymi drogami prostymi, więc twierdzenie jest prawdziwe w tym przypadku, możemy zatem przyjąć, że krawędź e nie jest pętlą. Niech więc e łączy dwa różne wierzchołki u i v. Jeśli f jest inną krawędzią łączącą u i v, to oczywiście graf $G\backslash\{e\}$ jest spójny i ciąg ef jest cyklem zawierającym e. A więc twierdzenie jest prawdziwe również w tym przypadku. Możemy zatem założyć, że e jest jedyną krawędzią łączącą u i v.

(a)⇒(b). Przypuśćmy, że graf $G\backslash\{e\}$ jest spójny. Z twierdzenia 1 w § 6.1 wynika, że istnieje prosta droga acykliczna $x_1x_2\ldots x_m$ taka, że $u = x_1$ i $x_m = v$. Ponieważ nie istnieje krawędź z u do v w grafie $G\backslash\{e\}$, więc $x_2 \neq v$, a więc $m \geq 3$. Jak wynika ze stwierdzenia 2 w § 6.1, wierzchołki $x_1, x_2, \ldots x_m$ są różne, a więc $x_1x_2\ldots x_mu$ jest cyklem w grafie G, zawierającym krawędź e.

(b)⇔(c). Oczywiście (b)⇒(c), gdyż cykle są zamkniętymi drogami prostymi, a (c)⇒(b) nawet, jeśli graf G nie jest spójny, na podstawie wniosku z twierdzenia 1 w § 6.1.

(b)⇒(a). Krawędź e jest jedną z dróg między u i v, podczas gdy reszta cyklu zawierającego e jest inną drogą. Nadal jest możliwe dostanie się z dowolnego wierzchołka grafu G do dowolnego innego wierzchołka nawet, jeśli droga prowadzi przez e. Po prostu zastępujemy krawędź e, jeśli trzeba, tą drugą drogą. ∎

Dowód poprawności algorytmu Fleury'ego. Rozważamy skończony graf spójny, którego wszystkie wierzchołki mają stopień parzysty i pokazujemy, że wtedy algorytm Fleury'ego tworzy cykl Eulera. Modyfikacja dowodu, pokazująca, że w przypadku istnienia dwóch wierzchołków stopnia nieparzystego otrzymamy drogę Eulera, jest natychmiastowa.

Każdy przebieg pętli od kroku 2 do kroku 5 powoduje usunięcie jednej krawędzi z grafu G i dodanie jej do ES w taki sposób, że krawędzie w ES tworzą drogę. Ponieważ graf G ma na początku skończenie wiele krawędzi, algorytm wcześniej czy później musi zakończyć działanie — lub musi nastąpić błąd — i żadna krawędź nie wystąpi więcej niż jeden raz w drodze wyznaczonej przez ES. Musimy pokazać, że podczas wykonania algorytmu nie nastąpi błąd oraz że po zakończeniu działania droga ES zawiera wszystkie krawędzie grafu G.

Jedynym miejscem, w którym może pojawić się błąd, jest krok 4. Skąd wiemy, że istnieje krawędź, którą możemy usunąć tak, by graf pozostał spójny? Niech G' będzie aktualną wartością G i v' aktualną wartością v oraz przypuśćmy, że więcej niż jedna krawędź wychodzi z wierzchołka v' w grafie G'. Twierdzimy, że w rzeczywistości możemy wybrać dowolną krawędź wychodzącą z v', z wyjątkiem co najwyżej jednej. Ponieważ krawędzie w ES tworzą drogę, mamy dwa przypadki do rozważenia: albo graf G' ma dwa wierzchołki stopnia nieparzystego, z których jednym jest v', a drugim wierzchołek v_0 wybrany w kroku 1, albo $v' = v_0$ i graf G' nie ma wierzchołków stopnia nieparzystego.

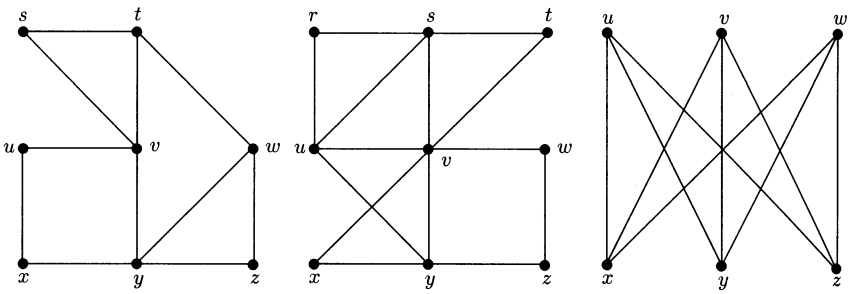
W pierwszym przypadku, z wniosku z twierdzenia 2 wynika, że graf G' ma drogę Eulera prowadzącą z v' do v_0. Za każdym razem, gdy ta droga powraca do v', zatacza ona zamkniętą drogę prostą zawierającą dwie krawędzie wychodzące z v', a więc z wyjątkiem być może krawędzi, którą ta droga opuszcza wierzchołek v' po raz ostatni, wszystkie krawędzie wychodzące z v' należą do pewnych zamkniętych dróg prostych w grafie G'. W drugim przypadku, z twierdzenia 2 wynika, że graf G' ma cykl Eulera i podobne rozumowanie pokazuje, że każda krawędź wychodząca z v' należy do pewnej zamkniętej drogi prostej w G'. Z twierdzenia 3 wynika, że po usunięciu krawędzi należącej do jakiejś zamkniętej drogi prostej graf G' nie przestanie być spójny. Ponieważ jest co najwyżej jedna zła krawędź wychodząca z v', więc jeśli pierwsza krawędź, którą wybierzemy, nie będzie należała do zamkniętej drogi prostej, to każdy następny wybór krawędzi wychodzącej z v' na pewno będzie dobry.

Dlaczego na końcu nie pozostają żadne krawędzie? Na początku graf jest spójny. Kiedy wykonujemy krok 4, graf pozostaje spójny oraz kiedy wykonujemy krok 3, graf też pozostaje spójny, gdyż po usunięciu krawędzi e usuwamy również pozostawiony izolowany wierzchołek v. Zatem przez cały czas graf jest spójny. Kiedy musimy zakończyć działanie, gdyż nie ma już żadnej krawędzi wychodzącej z v', w grafie G' nie może też być żadnych innych wierzchołków. Zatem graf G' nie ma również krawędzi i droga ES zawiera wszystkie krawędzie grafu G. ∎

Większość operacji w algorytmie Fleury'ego, takich jak dodawanie lub usuwanie krawędzi, wymaga ustalonej ilości czasu, niezależnie od tego, ile wierzchołków ma graf G. Operacją, która trwa dłużej, jest testowanie spójności grafu $G\backslash\{e\}$. W paragrafie 8.4 pokażemy, że istnieje test spójności grafu oparty na algorytmie Dijkstry, działający w czasie $O(|V(G)|^2)$. Dowód poprawności algorytmu Fleury'ego pokazuje, że musimy sprawdzić tylko jedną krawędź w kroku 4. Ponieważ algorytm wykonuje jeden przebieg pętli od kroku 2 do kroku 5 dla każdej krawędzi grafu G, a następnie zatrzymuje się, więc całkowity czas działania algorytmu Fleury'ego wynosi $O(|V(G)|^2 \cdot |E(G)|)$.

ĆWICZENIA DO § 6.2

1. Który z grafów przedstawionych na rysunku 6.16 ma cykle Eulera? Podaj ciąg wierzchołków w cyklu Eulera w każdym przypadku, w którym istnieje taki cykl.



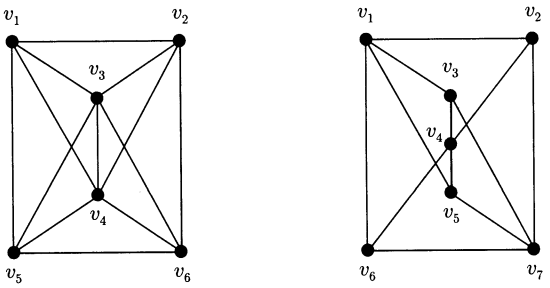
Rysunek 6.16

2. Wykorzystaj algorytm Fleury'ego do znalezienia cyklu Eulera w grafie przedstawionym na rysunku 6.16(b).

3. Zastosuj algorytm Fleury'ego do grafu przedstawionego na rysunku 6.16(a) do momentu, aż algorytm się załamie. Zacznij od wierzchołka w.

4. Powtórz ćwiczenie 3 dla grafu przedstawionego na rysunku 6.16(c).

5. Weźmy graf pokazany na rysunku 6.17(a).



(a) (b)

Rysunek 6.17

(a) Opisz drogę Eulera dla tego grafu lub wyjaśnij, dlaczego taka droga nie istnieje.
(b) Opisz cykl Eulera dla tego grafu lub wyjaśnij, dlaczego taki cykl nie istnieje.

6. Powtórz ćwiczenie 5 dla grafu przedstawionego na rysunku 6.17(b).

7. Czy jest możliwe, aby owad poruszający się wzdłuż krawędzi sześcianu przeszedł każdą krawędź dokładnie raz? Odpowiedź uzasadnij.

8. Zastosuj algorytm Fleury'ego, tak jak w przykładzie 3, aby otrzymać drogę Eulera w grafie przedstawionym na rysunku 6.11(a). Naszkicuj grafy pośrednie otrzymywane podczas działania algorytmu, tak jak zrobiliśmy to na rysunku 6.15.

9. Zbuduj graf mający zbiór wierzchołków $\{0,1\}^3$, w którym wierzchołki v i w są połączone krawędzią, jeśli ciągi v i w różnią się na dokładnie dwóch współrzędnych.
(a) Ile składowych ma ten graf?
(b) Ile wierzchołków danego stopnia ma ten graf?
(c) Czy ten graf ma cykl Eulera?

10. Odpowiedz na te same pytania co w ćwiczeniu 9 dla grafu mającego zbiór wierzchołków równy $\{0,1\}^3$, w którym wierzchołki v i w są połączone krawędzią, jeśli v i w różnią się na dwóch lub trzech współrzędnych.

11. (a) Pokaż, że jeśli graf spójny G ma dokładnie $2k$ wierzchołków stopnia nieparzystego oraz $k > 0$, to zbiór $E(G)$ jest rozłączną sumą

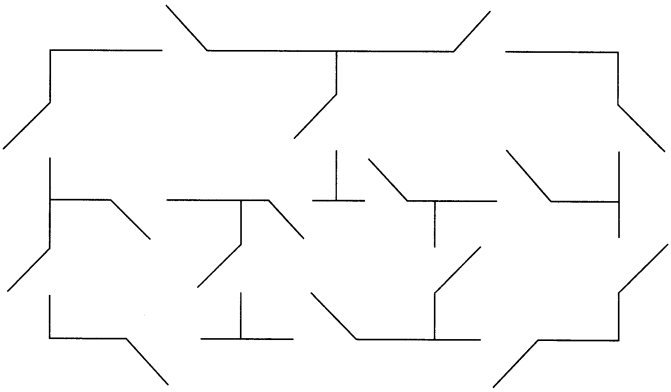
zbiorów krawędzi k dróg prostych. *Wskazówka*: Dodaj więcej krawędzi, tak jak w dowodzie wniosku do twierdzenia 2.

(b) Znajdź dwie rozłączne drogi proste takie, że suma ich zbiorów krawędzi będzie równa $E(G)$ dla grafu mostów królewieckich z rysunku 6.10.

(c) Zrób to samo dla grafu przedstawionego na rysunku 6.17(b) w ćwiczeniu 6.

12. Które grafy pełne K_n mają cykle Eulera?

13. Stara łamigłówka przedstawia pięciopokojowy dom mający szesnaścioro drzwi, jak widać to na rysunku 6.18 Problem polega na tym, aby wyobrazić sobie, jak można obejść cały dom i przez każde drzwi przejść dokładnie raz.



Rysunek 6.18

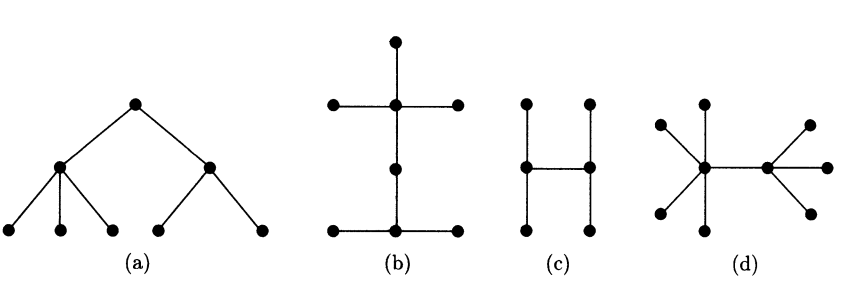
(a) Czy można tak przejść? Odpowiedź uzasadnij.

(b) Jak zmieni się odpowiedź, jeśli drzwi między dwoma dużymi pokojami będą zamknięte?

§ 6.3. Drzewa

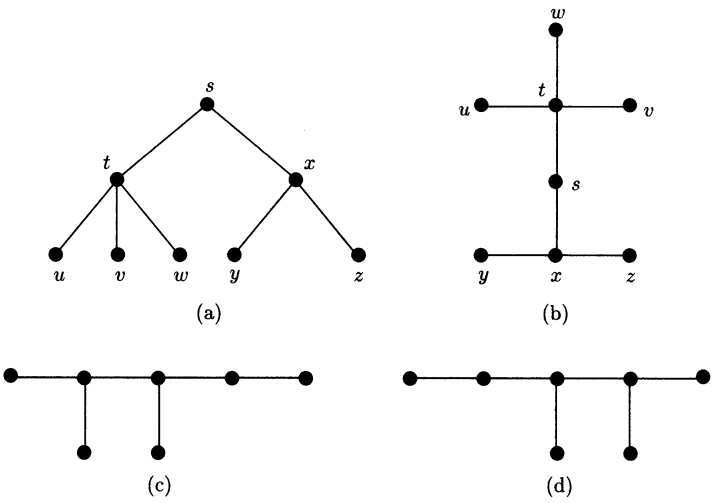
W tym paragrafie będziemy zajmować się grafami, które są acykliczne i spójne; nazywamy je **drzewami**. Ponieważ są one acykliczne, nie mają krawędzi wielokrotnych i pętli. Widzieliśmy już jedno drzewo na rysunku 5.1 w § 5.1. Oto kilka następnych przykładów.

PRZYKŁAD 1 Na rysunku 6.19 pokazanych jest kilka drzew. Drzewa na rysunkach 6.19(a) i 6.19(b) są izomorficzne. Są one narysowane



Rysunek 6.19

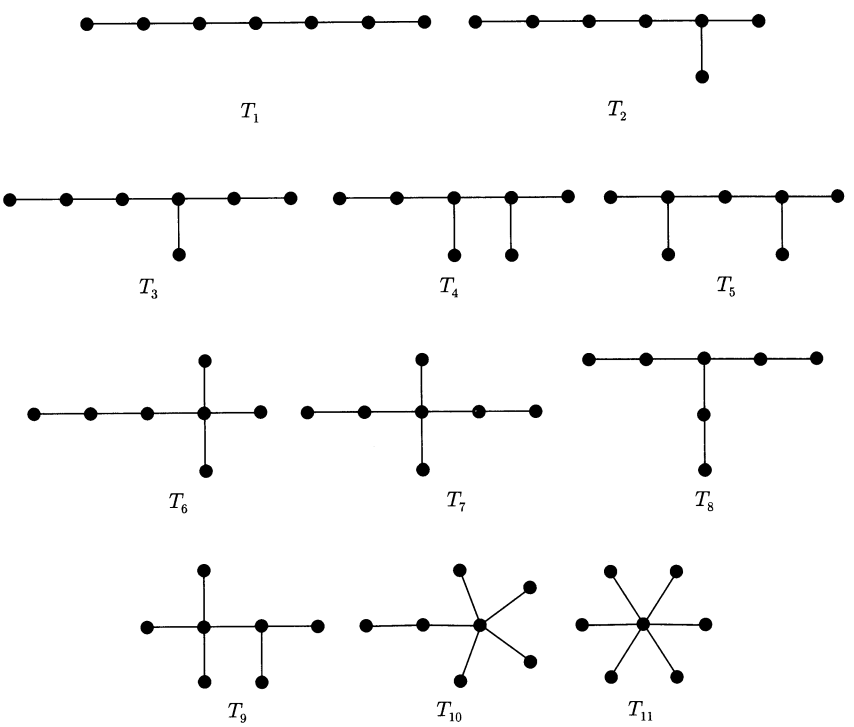
w różny sposób, ale ich zasadnicza struktura (wierzchołki i krawędzie) jest taka sama. Ich własności takie, jak liczba wierzchołków i krawędzi, liczba wierzchołków danego stopnia itd. są takie same. Aby to lepiej uwidocznić, przerysowaliśmy je na rysunkach 6.20(a) i 6.20(b) i oznaczyliśmy odpowiadające sobie wierzchołki. Drzewa na rysunkach 6.20(c) i 6.20(d) również są izomorficzne. ∎



Rysunek 6.20

PRZYKŁAD 2 Na rysunku 6.21 mamy wszystkie drzewa o siedmiu wierzchołkach. Jest ich jedenaście. Inaczej mówiąc, każde drzewo mające siedem wierzchołków jest izomorficzne z jednym z drzew narysowanych na rysunku 6.21 i żadne dwa drzewa z rysunku 6.21 nie są ze sobą izomorficzne. ∎

Dla danego grafu spójnego G interesują nas minimalne podgrafy łączące wszystkie wierzchołki. Taki podgraf musi być acykliczny, gdyż można usunąć jedną krawędź dowolnego cyklu, nie



Rysunek 6.21

tracąc przy tym własności spójności (twierdzenie 3 z § 6.2). Inaczej mówiąc, taki podgraf T jest **drzewem spinającym**: jest to drzewo zawierające każdy wierzchołek grafu G, to znaczy $V(T) = V(G)$. Zatem T jest drzewem otrzymanym przez usunięcie być może niektórych krawędzi grafu G i pozostawienie jednocześnie wszystkich wierzchołków.

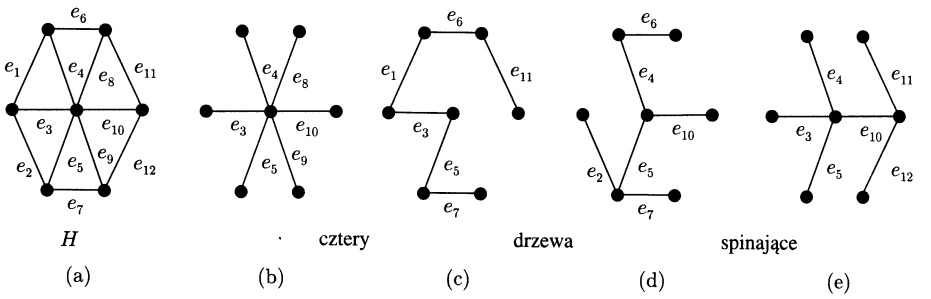
PRZYKŁAD 3 Graf H na rys. 6.22 ma ponad 300 drzew spinających, cztery z nich zostały narysowane. Wszystkie one mają 6 krawędzi. ■

Następne twierdzenie mówi, że wszystkie skończone grafy spójne mają drzewa spinające. W paragrafie 6.6 pokażemy, jak można je skonstruować.

Twierdzenie 1

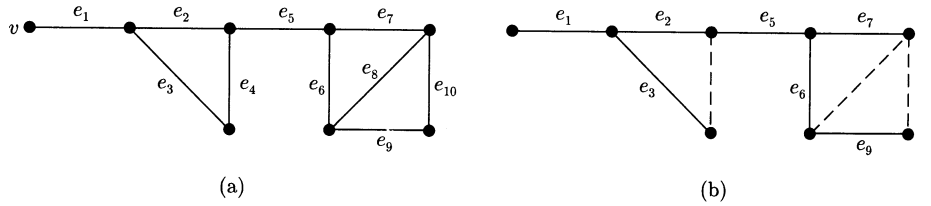
Każdy skończony graf spójny ma drzewo spinające.

Dowód. Weźmy spójny podgraf G' grafu G zawierający wszystkie wierzchołki G i mający najmniejszą możliwą liczbę krawędzi. Przypuśćmy, że graf G' zawiera cykl, do którego należy



Rysunek 6.22

na przykład krawędź e. Z twierdzenia 3 w § 6.2 wynika, że graf $G'\backslash\{e\}$ jest spójnym podgrafem grafu G i ma mniej krawędzi niż G', co jest sprzeczne z wyborem G'. Zatem graf G' nie ma cykli. Ponieważ jest spójny, więc jest drzewem. ∎



Rysunek 6.23

PRZYKŁAD 4 (a) Zilustrujemy twierdzenie 1 za pomocą grafu spójnego przedstawionego na rysunku 6.23(a). Zauważmy, że krawędź e_1 nie należy do żadnego cyklu i że graf $G\backslash\{e_1\}$ nie jest spójny: żadna droga w $G\backslash\{e_1\}$ nie łączy wierzchołka v z innymi wierzchołkami. Podobnie krawędź e_5 nie należy do żadnego cyklu i graf $G\backslash\{e_5\}$ nie jest spójny. Pozostałe krawędzie należą do cykli. Po usunięciu dowolnej z nich graf G pozostanie spójny.

(b) Zauważmy, że graf $G\backslash\{e_{10}\}$ jest nadal spójny, ale ma cykle. Jeśli usuniemy również krawędź e_8, otrzymany graf nadal będzie miał cykl, mianowicie $e_2e_3e_4$. Ale jeśli usuniemy wtedy jedną z krawędzi tego cyklu, np e_4, to otrzymamy acykliczny graf spójny, czyli drzewo spinające. Zob. rys. 6.23(b). Oczywiście w ten sposób można otrzymać różne drzewa spinające. ∎

Charakteryzacja drzew nie traci na ogólności, jeżeli rozważamy tylko drzewa bez pętli i krawędzi wielokrotnych. Nasze pierwsze charakteryzacje dotyczą nawet grafów nieskończonych.

Twierdzenie 2

Niech G będzie grafem bez pętli i krawędzi wielokrotnych, mającym więcej niż jeden wierzchołek. Następujące warunki są równoważne:

(a) G jest drzewem;

(b) każde dwa różne wierzchołki są połączone dokładnie jedną drogą prostą;

(c) graf G jest spójny, ale przestaje być spójny po usunięciu dowolnej krawędzi;

(d) graf G jest acykliczny, ale przestaje być acykliczny po dodaniu jakiejkolwiek krawędzi.

Dowód. Ten dowód składa się z czterech krótkich dowodów.

(a)⇒(b) Załóżmy, że G jest drzewem, a więc G jest grafem spójnym i acyklicznym. Na podstawie twierdzenia 1 z § 6.1 każde dwa wierzchołki są połączone co najmniej jedną drogą prostą, a na podstawie twierdzenia 2 z § 6.1 jest tylko jedna taka droga prosta.

(b)⇒(c) Jeśli zachodzi (b), to G jest oczywiście grafem spójnym. Niech $e = \{u, v\}$ będzie krawędzią grafu G i załóżmy, że graf $G\backslash\{e\}$ jest nadal spójny. Zauważmy, że $u \neq v$, gdyż graf G nie ma pętli. Na podstawie twierdzenia 1 z § 6.1 istnieje droga prosta z u do v w grafie $G\backslash\{e\}$. Ponieważ ta droga i droga składająca się z jednej krawędzi e są dwiema różnymi drogami prostymi z u do v w grafie G, więc mamy sprzeczność z (b).

(c)⇒(d) Przypuśćmy, że zachodzi (c). Jeśli graf G ma cykl, to możemy usunąć jakąś krawędź z grafu G i zachować spójność na podstawie twierdzenia 3 z § 6.2. Zatem G jest grafem acyklicznym. Weźmy teraz krawędź e nie należącą do grafu G i niech G' oznacza graf G wraz z dołączoną tą nową krawędzią. Ponieważ graf $G'\backslash\{e\} = G$ jest grafem spójnym, stosujemy do G' twierdzenie 3 z § 6.2 i stwierdzamy, że krawędź e należy do pewnego cyklu grafu G'. Innymi słowy, dodanie krawędzi e do grafu G niszczy jego acykliczność.

(d)⇒(a) Jeśli zachodzi (d) i graf G nie jest drzewem, to G nie jest grafem spójnym. Wtedy istnieją różne wierzchołki u i v, które nie są połączone żadną drogą w grafie G. Weźmy nową krawędź $e = \{u, v\}$. Zgodnie z założeniem (d) graf $G \cup \{e\}$ ma cykl i e musi być jego częścią. Reszta cyklu to droga w G łącząca u i v. To jest sprzeczne z naszym wyborem u i v. Zatem graf G jest spójny i G jest drzewem. ∎

Aby docenić twierdzenie 2, narysuj drzewo lub popatrz na któreś z drzew przedstawionych na rysunkach od 6.19 do 6.22

i zauważ, że ma ono wszystkie własności (a)-(d) z twierdzenia 2. Następnie narysuj lub popatrz na graf, który nie jest drzewem i zauważ, że nie ma on żadnej z tych własności.

Potrzebujemy dwóch lematów, by móc scharakteryzować drzewa skończone. W drzewie wierzchołki stopnia 1 nazywamy **liśćmi.**

PRZYKŁAD 5 W drzewach przedstawionych na rysunku 6.21: T_1 ma dwa liście; T_2, T_3 i T_8 mają trzy liście; T_4, T_5, T_6 i T_7 mają cztery liście; T_9 i T_{10} mają pięć liści a T_{11} ma sześć liści. ∎

Lemat 1

Drzewo skończone, mające co najmniej jedną krawędź, ma co najmniej dwa liście.

Dowód. Weźmy najdłuższą acykliczną drogę prostą, np. $v_1v_2\ldots v_n$. Wtedy $v_1 \neq v_n$ i oba wierzchołki v_1 oraz v_n są liśćmi. ∎

Lemat 2

Drzewo mające n wierzchołków ma dokładnie $n-1$ krawędzi.

Dowód. Stosujemy indukcję. Dla $n = 2$ lemat jest oczywisty. Załóżmy, że teza jest prawdziwa dla pewnego n i weźmy drzewo T mające $n+1$ wierzchołków. Z lematu 1 wynika, że drzewo T ma liść v_0. Niech T_0 będzie grafem otrzymanym z drzewa T przez usunięcie wierzchołka v_0 i krawędzi wychodzącej z niego. Łatwo sprawdzić, że T_0 jest wtedy drzewem i że ma n wierzchołków. Z założenia indukcyjnego T_0 ma $n-1$ krawędzi, więc T ma n krawędzi. ∎

Twierdzenie 3

Niech G będzie grafem skończonym mającym n wierzchołków i nie mającym pętli i krawędzi wielokrotnych. Wtedy następujące warunki są równoważne:

(a) G jest drzewem.
(b) G jest grafem acyklicznym mającym $n-1$ krawędzi.
(c) G jest grafem spójnym mającym $n-1$ krawędzi.

Innymi słowy, każde dwie spośród własności: „spójność", „acykliczność" i „posiadanie $n-1$ krawędzi" implikują trzecią.

Dowód. Twierdzenie jest oczywiste dla $n = 1$, załóżmy więc, że $n \geq 2$. Obie implikacje (a)⇒(b) i (a)⇒(c) wynikają z lematu 2.

(b)⇒(a) Załóżmy, że zachodzi (b), ale graf G nie jest drzewem.

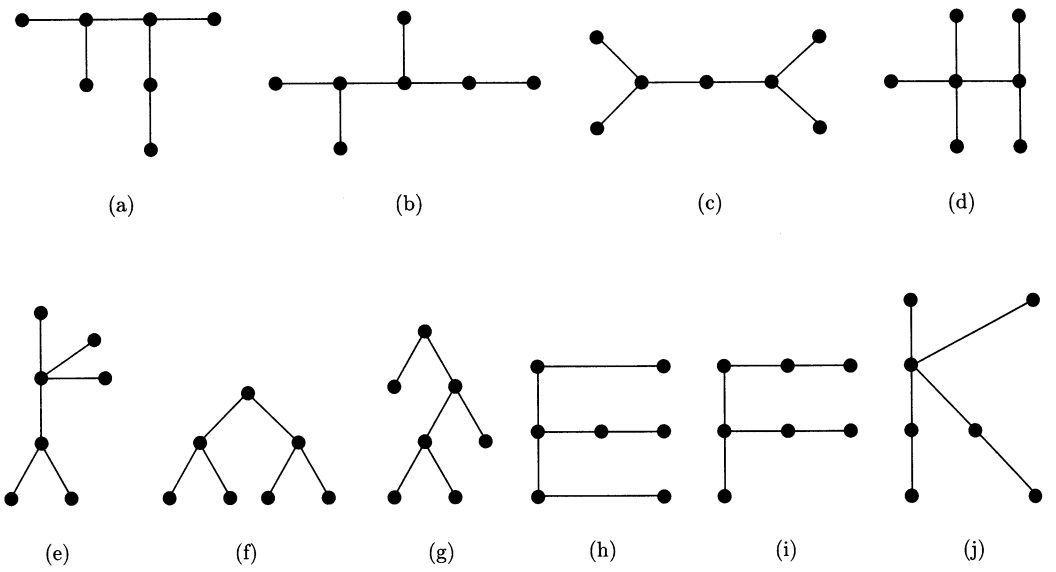
Nie może wtedy zachodzić warunek (d) z twierdzenia 2. Ponieważ G jest grafem acyklicznym, możemy oczywiście dodać trochę krawędzi i zachować acykliczność. Dodajmy tak wiele krawędzi, jak jest to możliwe, by zachować acykliczność. Otrzymany w ten sposób graf G' spełnia warunek (d) z twierdzenia 2, więc jest drzewem. Ale G' ma n wierzchołków i co najmniej n krawędzi, co przeczy lematowi 2. Zatem G jest drzewem.

(c)⇒(a) Załóżmy, że zachodzi (c), ale graf G nie jest drzewem. Z twierdzenia 1 wynika, że graf G ma drzewo spinające T, mające mniej niż $n-1$ krawędzi. To przeczy lematowi 2, a więc G jest drzewem. ∎

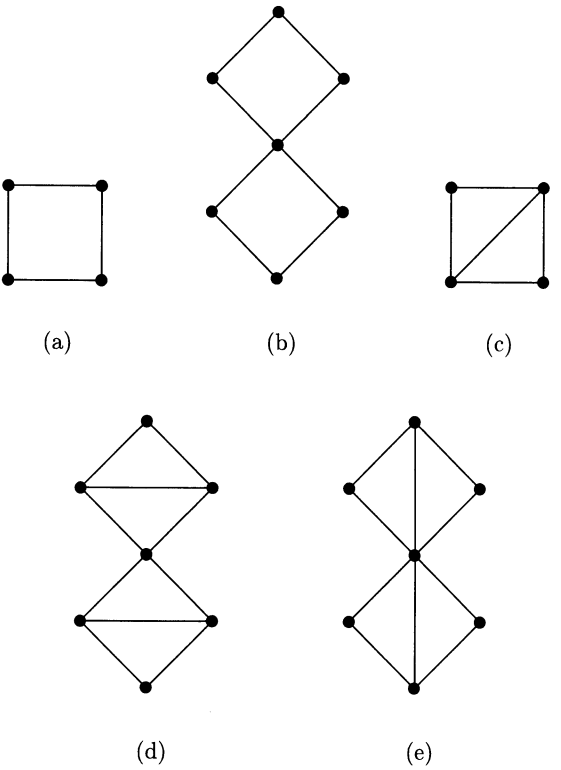
Graf acykliczny, który niekoniecznie jest spójny, czasami nazywamy **lasem**. Oczywiście spójne składowe lasu są drzewami. Można tak uogólnić twierdzenie 3, by scharakteryzować zarówno lasy, jak i drzewa; zob. ćwiczenie 9.

ĆWICZENIA DO § 6.3

1. Znajdź wszystkie drzewa mające sześć wierzchołków.
2. Drzewa przedstawione na rysunku 6.24 mają siedem wierzchołków. Podaj, z którym z drzew przedstawionych na rysunku 6.21 każde z nich jest izomorficzne.



Rysunek 6.24



Rysunek 6.25

3. Podaj liczbę drzew spinających dla grafów przedstawionych na rysunku 6.25.

4. Znajdź dwa nieizomorficzne drzewa spinające dla grafu $K_{3,3}$ przedstawionego na rysunku 6.40(d) w § 6.5.

5. Weźmy drzewo o n wierzchołkach. Ma ono dokładnie $n-1$ krawędzi (lemat 2), więc suma stopni jego wierzchołków wynosi $2n-2$ (twierdzenie 3 z § 6.1).

(a) Pewne drzewo ma dwa wierzchołki stopnia 4, jeden wierzchołek stopnia 3 i jeden wierzchołek stopnia 2. Jeśli inne wierzchołki są stopnia 1, to ile wierzchołków jest w tym grafie? *Wskazówka*: jeśli drzewo ma n wierzchołków, to $n-4$ z nich będą miały stopień 1.

(b) Narysuj drzewo opisane w punkcie (a).

6. Powtórz ćwiczenie 5 dla drzewa mającego dwa wierzchołki stopnia 5, trzy wierzchołki stopnia 3, dwa wierzchołki stopnia 2 i resztę wierzchołków stopnia 1.

7. (a) Pokaż, że istnieje drzewo mające sześć wierzchołków stopnia 1, jeden wierzchołek stopnia 2, jeden wierzchołek stopnia 3, jeden wierzchołek stopnia 5 i żadnych innych.

(b) Dla $n \geq 2$ weźmy n liczb całkowitych dodatnich $d_1, \ldots, d_n$, których suma wynosi $2n - 2$. Pokaż, że istnieje drzewo o n wierzchołkach, którego wierzchołki mają stopnie $d_1, \ldots, d_n$.

(c) Pokaż, że ćwiczenie (a) ilustruje ćwiczenie (b), gdy $n = 9$.

8. Narysuj wszystkie grafy spójne o 4 krawędziach i 4 wierzchołkach. Nie zapomnij o pętlach i krawędziach wielokrotnych.

9. (a) Pokaż, że las o n wierzchołkach i m składowych ma $n-m$ krawędzi.

(b) Pokaż, że graf o n wierzchołkach, m składowych i $n - m$ krawędziach musi być lasem.

10. Pokaż, że graf spójny o n wierzchołkach ma co najmniej $n-1$ krawędzi.

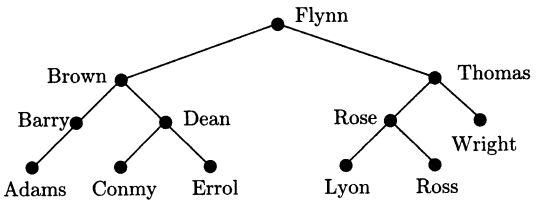
11. Naszkicuj drzewo mające co najmniej jedną krawędź i nie mające liści. *Wskazówka*: zob. lemat 1 do twierdzenia 3.

§ 6.4. Drzewa z wyróżnionym korzeniem

Drzewo z wyróżnionym korzeniem jest to drzewo, w którym jest wyróżniony jeden wierzchołek, nazywany **korzeniem**. Jest to pojęcie proste, ale zadziwiająco użyteczne. Drzewa z wyróżnionym korzeniem mają zastosowanie jako struktury danych, a także pomagają nam opisywać i wizualizować zachodzące relacje w wielu różnorodnych sytuacjach. W tym paragrafie obejrzymy wiele przykładów drzew z wyróżnionym korzeniem i poznamy ich własności.

Drzewa z wyróżnionym korzeniem zazwyczaj rysujemy tak, że ich korzeń znajduje się na górze. Dokładnie na odwrót niż drzewa w lesie. Wśród naszych pierwszych przykładów znajdują się pewne typowe zastosowania wraz z odpowiednimi rysunkami.

PRZYKŁAD 1 (a) Ciąg liczb lub plik uporządkowany alfabetycznie może być zorganizowany w wygodny sposób jako pewien specjalny rodzaj drzewa z wyróżnionym korzeniem, nazywanego **drzewem poszukiwań binarnych**. Rysunek 6.26 pokazuje przykład takiego drzewa, zawierającego zbiór rekordów klientów, uporządkowany w sposób alfabetyczny.



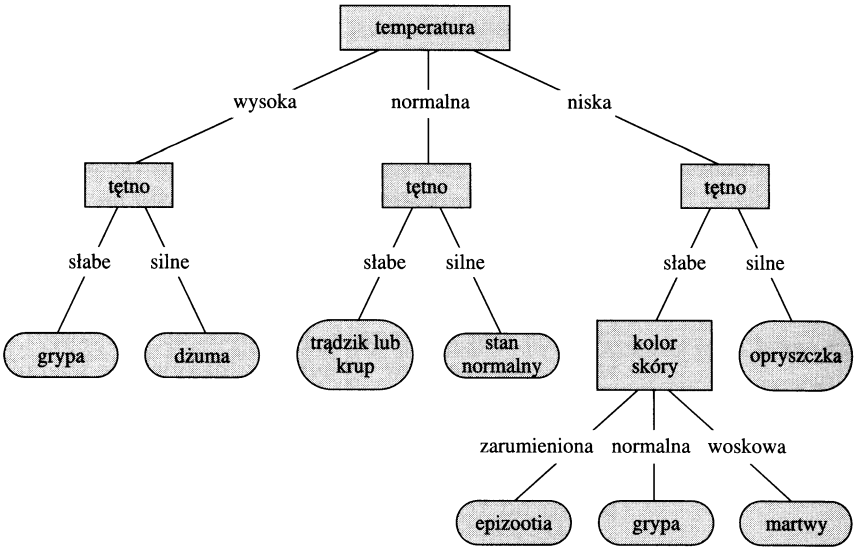
Rysunek 6.26

Aby znaleźć na przykład rekord klienta Conmy, porównujemy nazwisko Conmy z nazwiskiem znajdującym się w korzeniu, Flynn. Conmy jest wcześniej w porządku alfabetycznym, więc wybieramy lewą gałąź i porównujemy nazwiska Conmy i Brown. Conmy jest później w porządku alfabetycznym, więc wybieramy prawą gałąź. Następnie poruszamy się na lewo od rekordu Dean i znajdujemy rekord Conmy. Gdybyśmy szukali rekordu Dean, to zatrzymalibyśmy się wcześniej, gdy dotarliśmy do wierzchołka z tym nazwiskiem. Tym, co powoduje, że ta procedura wyszukiwania jest skuteczna, jest fakt, że z każdego wierzchołka (lub węzła, jak często w tej sytuacji nazywa się wierzchołki) wychodzą co najwyżej dwie krawędzie w dół, co najwyżej jedna w lewo i co najwyżej jedna w prawo. Łatwo też znaleźć miejsce, w którym powinien się znajdować rekord klienta o nazwisku Romanelli: na prawo od rekordu Flynn, na lewo od rekordu Thomas, na lewo od rekordu Rose i na prawo od rekordu Lyon, a więc w nowym wierzchołku poniżej i na prawo od rekordu Lyon.

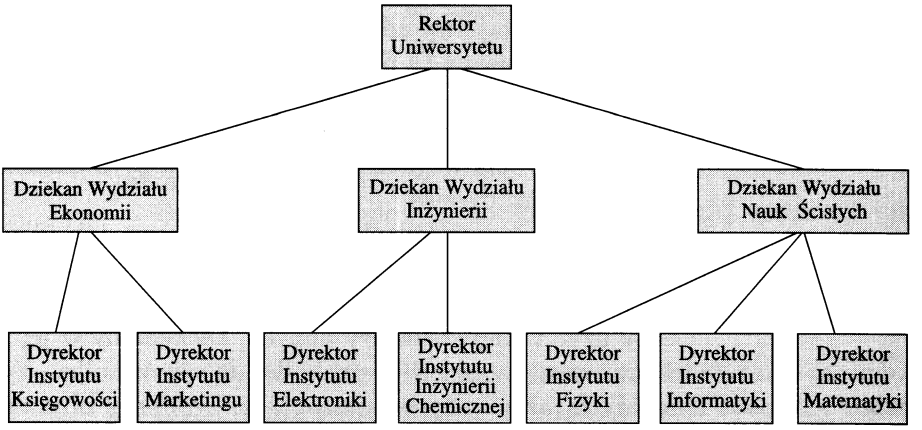
Główną zaletą organizowania danych w postaci drzewa poszukiwań binarnych jest to, że wystarczy tylko kilka porównań, by znaleźć właściwy adres, nawet jeśli całkowita liczba rekordów jest duża. Opisany tu pomysł jest całkiem prosty. Ta metoda jest tak ważna, że zaprojektowano wiele schematów tworzenia i uaktualniania drzew poszukiwań tak, aby średnia długość drogi poszukiwania była względnie mała, rzędu $\log_2 n$, gdzie n jest liczbą wierzchołków drzewa.

(b) Struktury danych do diagnozowania lub identyfikowania często mogą być przedstawione w postaci drzew z wyróżnionym korzeniem. Rysunek 6.27 pokazuje, na czym polega ten pomysł. Aby użyć takiej struktury danych, zaczynamy w jej wierzchołku i przesuwamy się od wierzchołka do wierzchołka, wybierając za każdym razem odpowiednią gałąź, zgodną z objawami choroby występującymi u pacjenta. W liściu znajdującym się na końcu tej drogi znajduje się nazwa najbardziej prawdopodobnego warunku lub warunków, w jakich występują dane objawy. Ten sam rodzaj struktur w postaci drzew z wyróżnionym korzeniem jest podstawą układu klucza do oznaczania grzybów, ptaków, dziko rosnących kwiatów itp. W przypadku rekordów klientów w przykładzie (a) istniał naturalny porządek umieszczania etykiet w węzłach od lewej do prawej. W tym i w następnym przykładzie porządek, w jakim są wymienione wierzchołki, nie ma specjalnego znaczenia.

(c) Zależności służbowe w instytucji można często przedstawić w postaci drzewa z wyróżnionym korzeniem. Na rysunku 6.28 jest pokazana część hierarchii uniwersyteckiej. ■

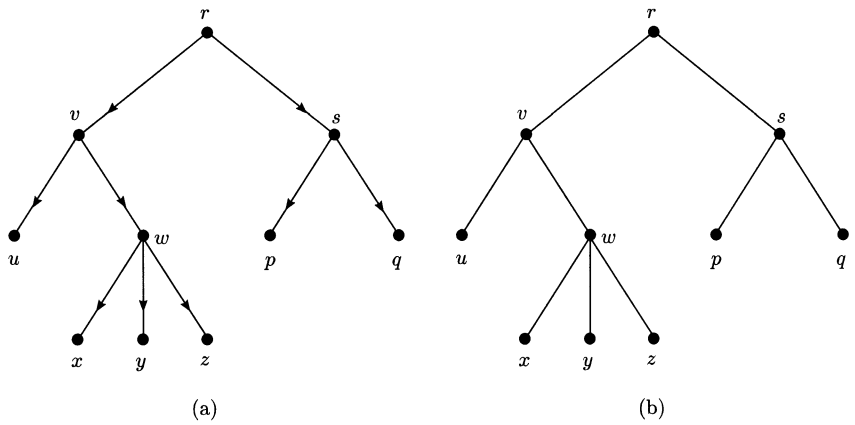


Rysunek 6.27



Rysunek 6.28

W tych przykładach wierzchołki drzew miały wymyślne nazwy lub etykiety, takie jak „Ross" czy „epizootia", a nie u, v, w itp., aby dostarczyć dodatkowych informacji. W rzeczywistości *nazwa* wierzchołka i jego *etykieta* nie muszą być takie same i mogłoby się równie dobrze zdarzyć, że wiele wierzchołków miałoby różne nazwy i tę samą etykietę. Na przykład, etykietą wierzchołka mógłby być związany z tym wierzchołkiem znak dolara. Kiedy mówimy o drzewie z etykietami, mamy na myśli drzewo, w którym wierzchołki mają dołączone dodatkowe informacje. W prak-



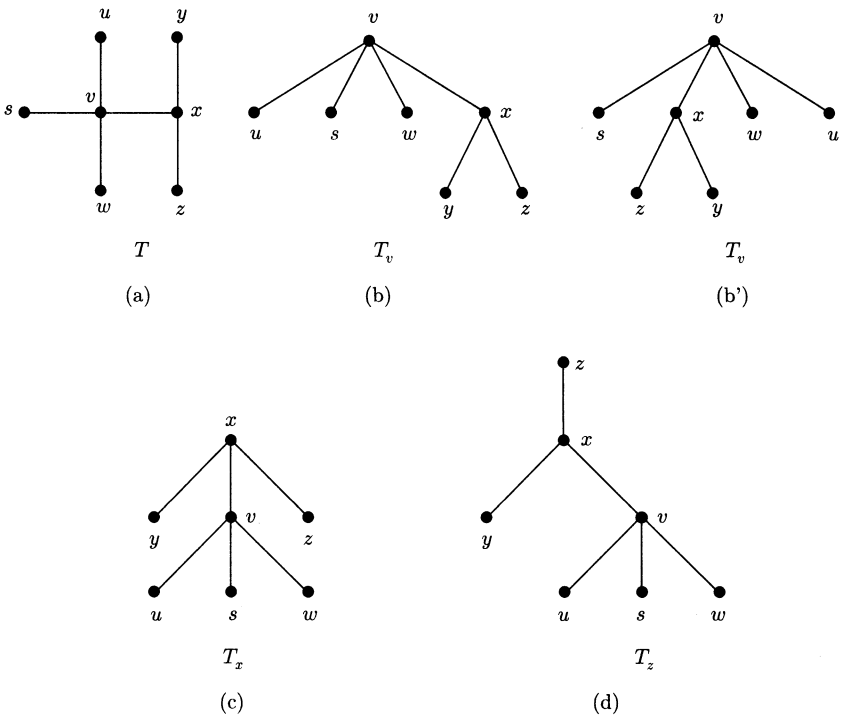
Rysunek 6.29

tycznych zastosowaniach traktujemy etykietę wierzchołka jako informację przechowywaną w nim. Zobaczymy wkrótce, że można również tak dobrać same nazwy wierzchołków, aby za ich pomocą przekazać informacje o położeniu wierzchołków w grafie.

Drzewo z wyróżnionym korzeniem można w naturalny sposób traktować jako graf skierowany. Po prostu złap drzewo za jego korzeń i siła grawitacji skieruje jego krawędzie w dół. Na rysunku 6.29(a) pokazane jest drzewo z korzeniem r skierowane w ten sposób. Zazwyczaj nie rysuje się strzałek na krawędziach, umawiając się jednak, że one wszystkie wskazują w dół. Drzewo z wyróżnionym korzeniem pokazane na rysunku 6.29(b) możemy traktować albo jako drzewo nieskierowane z wyróżnionym wierzchołkiem r, albo jako graf skierowany z rysunku 6.29(a); nie ma istotnej różnicy między tymi dwoma sposobami widzenia.

Aby nadać matematyczną ścisłość tej grawitacyjnej definicji, przypomnijmy twierdzenie 2 z § 6.3, które mówi, że jeśli r i v są dowolnymi wierzchołkami drzewa, to istnieje dokładnie jedna droga prosta łącząca je. Jeśli r jest korzeniem drzewa T i e jest krawędzią w drzewie T łączącą u i w, to albo wierzchołek u znajduje się na jedynej drodze prostej z r do w, albo w znajduje się na jedynej drodze prostej z r do u. W pierwszym przypadku krawędzi e nadajemy kierunek od u do w, a w drugim przypadku nadajemy krawędzi e kierunek od w do u. To właśnie robi grawitacja. Symbolem T_r będziemy oznaczać drzewo z wyróżnionym korzeniem, powstałe z drzewa T przez wyróżnienie korzenia r i kiedy myślimy o T_r jako o grafie skierowanym, uważamy, że ma on tę naturalną strukturę grafu skierowanego, którą przed chwilą opisaliśmy.

PRZYKŁAD 2 Weźmy drzewo (nieskierowane) pokazane na rysunku 6.30(a). Jeśli wybierzemy v, x i z jako korzenie drzewa, to otrzymamy trzy drzewa z wyróżnionym korzeniem pokazane na rysunkach 6.30(b), 6.30(c) i 6.30(d). Dokładne umiejscowienie wierzchołków nie ma znaczenia; rysunki 6.30(b) i 6.30(b') pokazują to samo drzewo z wyróżnionym korzeniem.

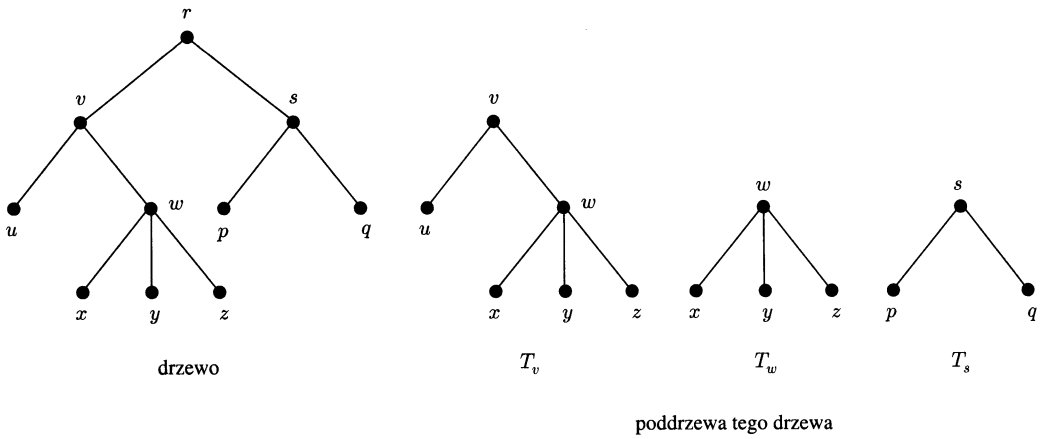


Rysunek 6.30

Zauważmy, że para (v, w) jest krawędzią w drzewie pokazanym na rysunku 6.30(d), gdyż krawędź $\{v, w\}$ znajduje się na jedynej drodze prostej z wierzchołka z do wierzchołka w w grafie z rysunku 6.30(a). Z drugiej strony, para (w, v) nie jest krawędzią w drzewie z wyróżnionym korzeniem T_z; pomimo że $\{w, v\}$ jest krawędzią oryginalnego drzewa, nie znajduje się ona na jedynej drodze prostej z wierzchołka z do wierzchołka v. Podobne uwagi stosują się do wszystkich innych krawędzi. ■

Terminologia stosowana do opisu różnych części drzew jest zadziwiającą mieszaniną terminologii pochodzących zarówno od drzew w lesie, jak i od drzew genealogicznych. Jak poprzednio,

wierzchołki stopnia 1 nazywamy **liśćmi**; jest jeden wyjątek: czasami (na przykład na rysunku 6.30(d)) korzeń będzie miał stopień 1, ale nie będziemy nazywać go liściem. Jeśli popatrzymy na drzewo jak na graf skierowany, to zobaczymy, że korzeń jest jedynym wierzchołkiem nie będącym końcem żadnej krawędzi, a liście są jedynymi wierzchołkami nie będącymi początkami krawędzi. Pozostałe wierzchołki nazywamy czasami „węzłami gałęzi" lub „węzłami wewnętrznymi", a liście nazywamy czasami „węzłami końcowymi". Przyjmujemy umownie, że jeśli para (v, w) jest krawędzią drzewa z wyróżnionym korzeniem, to v jest **rodzicem** w, a w jest **dzieckiem** v. Każdy wierzchołek poza korzeniem ma dokładnie jednego rodzica. Rodzic może mieć kilkoro **dzieci**. Ogólniej, w jest **potomkiem** v, jeśli $w \neq v$ i v jest wierzchołkiem jedynej drogi prostej z korzenia r do wierzchołka w. Wreszcie, dla dowolnego wierzchołka v **poddrzewo** o korzeniu v jest to dokładnie drzewo T_v składające się z wierzchołka v, wszystkich jego potomków i wszystkich krawędzi skierowanych łączących ich. Jeśli v jest liściem, poddrzewo o wierzchołku v jest trywialnym drzewem jednowierzchołkowym.



Rysunek 6.31

PRZYKŁAD 3 Weźmy drzewo z wyróżnionym korzeniem z rysunku 6.29, przedstawione ponownie na rysunku 6.31. Ma ono sześć liści. Rodzic v ma dwoje dzieci u i w oraz pięciu potomków: u, w, x, y i z. Wszystkie wierzchołki, z wyjątkiem samego r, są potomkami r. Całe drzewo jest oczywiście poddrzewem z wyróżnionym korzeniem r i istnieje sześć trywialnych poddrzew składających się z liści. Interesujące poddrzewa są pokazane na rysunku 6.31. ■

Drzewo z wyróżnionym korzeniem jest **drzewem binarnym** wtedy, gdy każdy węzeł ma co najwyżej dwoje dzieci: dziecko lewe, dziecko prawe, dwoje dzieci — lewe i prawe lub w ogóle nie ma dzieci. Takie rozróżnienie na lewe i prawe było istotne w drzewie poszukiwań binarnych w przykładzie 1(a). Podobnie, dla $m > 2$ **drzewem o m rozgałęzieniach** nazywamy drzewo, w którym dzieci każdego rodzica są oznaczone różnymi elementami zbioru $\{1, 2, \ldots, m\}$. Rodzic nie musi mieć całego zbioru m dzieci; mówimy, że i-te dziecko jest **nieobecne**, jeśli nie ma dziecka oznaczonego liczbą i. W terminologii grafów skierowanych, w drzewie o m rozgałęzieniach każdy wierzchołek v ma stopień wyjściowy nie większy niż m (zob. § 8.1). Drzewo o m rozgałęzieniach (lub drzewo binarne) jest **drzewem regularnym o m rozgałęzieniach**, jeśli stopień wyjściowy każdego wierzchołka jest równy m lub 0.

Numerem poziomu wierzchołka v nazywamy długość jedynej drogi prostej od korzenia do v. W szczególności sam korzeń ma numer poziomu równy 0. **Wysokość drzewa** z wyróżnionym korzeniem jest to największy numer poziomu wierzchołka. Tylko liście mogą mieć numer poziomu równy wysokości drzewa. Regularne drzewo o m rozgałęzieniach jest **pełnym drzewem o m rozgałęzieniach**, jeśli wszystkie liście mają ten sam numer poziomu, mianowicie równy wysokości drzewa.

PRZYKŁAD 4 (a) Drzewo z wyróżnionym korzeniem na rysunkach 6.29 i 6.31 jest drzewem o trzech rozgałęzieniach i tak naprawdę drzewem o m rozgałęzieniach dla $m \geq 3$. Nie jest drzewem regularnym o trzech rozgałęzieniach, gdyż wierzchołki r, v i s mają stopień wyjściowy 2. Wierzchołki v i s mają numer poziomu równy 1, wierzchołki u, w, p i q mają numer poziomu równy 2 i liście x, y, z mają numer poziomu równy 3. Drzewo ma wysokość równą 3.

(b) Drzewo z etykietami przedstawione na rysunku 6.33(a) jest pełnym regularnym drzewem binarnym o wysokości 3. Drzewo z etykietami przedstawione na rysunku 6.33(b) jest drzewem regularnym o trzech rozgałęzieniach o wysokości 3. Nie jest pełnym drzewem o trzech rozgałęzieniach, gdyż jeden liść ma numer poziomu równy 1, a pięć liści ma numer poziomu równy 2.

■

PRZYKŁAD 5 Weźmy pełne drzewo o m rozgałęzieniach, wysokości h. Ma ono m wierzchołków na poziomie 1. Każdy rodzic na poziomie pierwszym ma m dzieci, a więc jest m^2 wierzchołków na poziomie 2. Proste rozumowanie indukcyjne pokazuje, że ponieważ

drzewo jest pełne, ma ono m^l wierzchołków na poziomie l dla każdego $l \leq h$. Ma ono zatem $1 + m + m^2 + \ldots + m^h$ wierzchołków. Ponieważ

$$(m-1)(1 + m + m^2 + \ldots + m^h) = m^{h+1} - 1,$$

co można łatwo sprawdzić wymnażając i redukując wyrazy podobne, otrzymujemy

$$1 + m + m^2 + \ldots + m^h = \frac{m^{h+1} - 1}{m - 1}.$$

Zauważmy, że to drzewo ma $p = (m^h - 1)/(m - 1)$ rodziców i $t = m^h$ liści. ∎

W drzewie poszukiwań binarnych w przykładzie 1(a) istotny był porządek alfabetyczny pozwalający porównywać nazwiska klientów. Ten pomysł można uogólnić; jedyne, czego potrzebujemy, to uporządkowanie, które mówi nam, który z dwóch danych obiektów jest wcześniejszy. (Ten rodzaj uporządkowania, którym zajmiemy się w § 11.2, nazywamy **porządkiem liniowym**). Zawsze, kiedy chcemy, możemy zdecydować, w jakim porządku wypisać dzieci danego rodzica w drzewie z wyróżnionym korzeniem. Jeśli uporządkujemy dzieci każdego rodzica w drzewie, otrzymamy tak zwane **uporządkowane drzewo z wyróżnionym korzeniem**. Kiedy rysujemy takie drzewo, rysujemy dzieci w porządku od lewej strony do prawej. Wygodnie jest używać symbolu $v < w$ dla oznaczenia, że v poprzedza w w tym porządku, nawet w przypadku, gdy v i w nie są liczbami.

PRZYKŁAD 6 (a) Jeśli popatrzymy na drzewo przedstawione na rysunku 6.30(b) jak na uporządkowane drzewo z wyróżnionym korzeniem, to dzieci wierzchołka v są uporządkowane: $u < s < w < x$. Dzieci wierzchołka x też są uporządkowane: $y < z$. Rysunek 6.30(b') pokazuje inne uporządkowane drzewo z wyróżnionym korzeniem, gdyż $s < x < w < u$ i $z < y$.

(b) Jeśli narysujemy drzewo z wyróżnionym korzeniem, to wygląda ono jak uporządkowane drzewo z wyróżnionym korzeniem, nawet jeśli nie ma dla nas znaczenia porządek. Na przykład strukturą istotną na rysunku 6.28 jest struktura drzewa z wyróżnionym korzeniem. Uporządkowanie „dzieci" nie ma znaczenia. Dyrektor Instytutu Informatyki poprzedza Dyrektora Instytutu Matematyki po prostu dlatego, że zdecydowaliśmy się wypisać instytuty w porządku alfabetycznym.

(c) Drzewo binarne lub ogólniej drzewo o m rozgałęzieniach jest w naturalny sposób drzewem uporządkowanym, ale jest

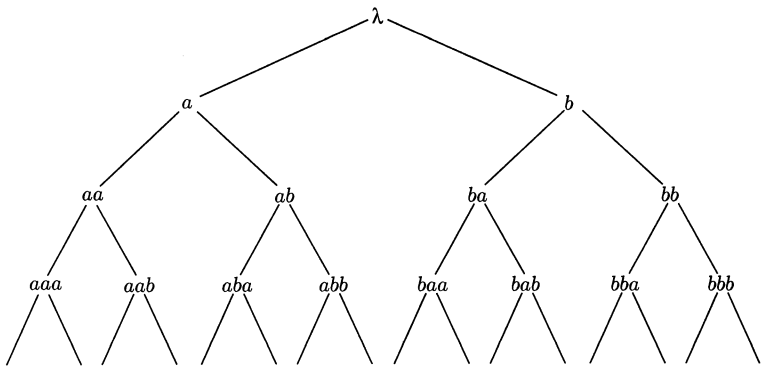
pewna różnica między tymi porządkami. W drzewie binarnym dziecko prawe będzie pierwszym dzieckiem, jeśli nie ma dziecka lewego. W drzewie o trzech rozgałęzieniach, przedstawionym na rysunku 6.31, dzieci w, s i q mogą mieć etykietę 3, nawet wtedy, gdy ich rodzice mają tylko dwoje dzieci. ∎

PRZYKŁAD 7 (a) Weźmy alfabet Σ, uporządkowany w pewien sposób. Ze zbioru Σ^* tworzymy drzewo z wyróżnionym korzeniem w następujący sposób. Słowo puste λ będzie korzeniem. Dla dowolnego słowa $w \in \Sigma^*$ jego zbiorem dzieci będzie zbiór

$$\{wx\colon x \in \Sigma\}.$$

Ponieważ Σ jest zbiorem uporządkowanym, możemy uporządkować każdy zbiór dzieci, aby otrzymać drzewo uporządkowane z wyróżnionym korzeniem Σ^*_λ, przyjmując $wx < wy$ wtedy, gdy $x < y$ w zbiorze Σ.

(b) Niech $\Sigma = \{a, b\}$, gdzie $a < b$. Każdy wierzchołek ma dwoje dzieci. Na przykład dziećmi wierzchołka $abba$ są $abbaa$ i $abbab$. Część takiego nieskończonego uporządkowanego drzewa z wyróżnionym korzeniem Σ^*_λ jest pokazana na rysunku 6.32.



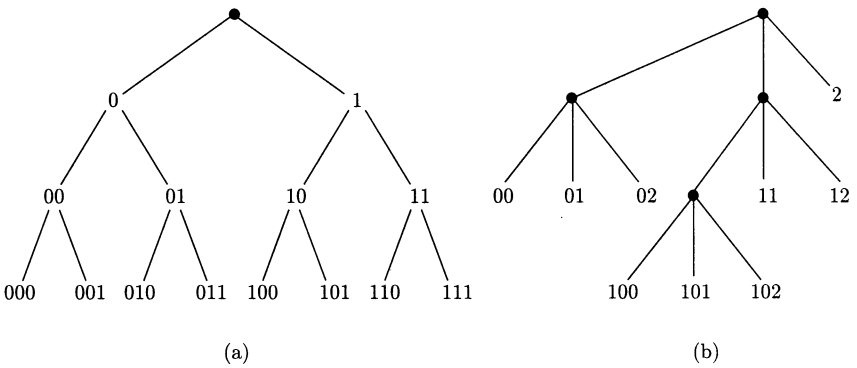
Rysunek 6.32

(c) Rysunek 5.1 w § 5.1 pokazywał część drzewa $\{a, b, c, d\}^*_\lambda$ z naturalnym uporządkowaniem wyznaczonym przez porządek $a < b < c < d$. ∎

Istnieje wiele różnych sposobów nazywania (lub etykietowania) wierzchołków drzewa uporządkowanego z wyróżnionym korzeniem w celu opisania nazwami ich położenia. Jeden taki schemat przypomina przykład 7: wierzchołki drzewa o m rozgałęzieniach mogą być nazwane słowami ze zbioru Σ^*, gdzie

$\Sigma=\mathbb{Z}_m=\{0,1,\ldots,m-1\}$. Uporządkowane dzieci korzenia mają nazwy ze zbioru $\{0,1,2,\ldots,m-1\}$. Jeśli wierzchołek jest nazwany słowem w, to jego dzieci mają nazwy $w0$, $w1$, $w2$ itd. Nazwa wierzchołka podaje nam dokładne położenie wierzchołka w tym drzewie. Na przykład wierzchołek o nazwie 1021 byłby drugim dzieckiem wierzchołka 102, który z kolei byłby trzecim dzieckiem wierzchołka 10 itd. Poziom wierzchołka jest długością jego nazwy; wierzchołek o nazwie 1021 jest na poziomie 4.

PRZYKŁAD 8 Wszystkie wierzchołki oprócz korzenia na rysunku 6.33(a) są nazwane w ten sposób. Na rysunku 6.33(b) są nazwane tylko liście. Nazwy pozostałych wierzchołków powinny być oczywiste. ■



Rysunek 6.33

Drzew używano od dawna do porządkowania i przechowywania informacji. W starych książkach matematycznych pojedyncze akapity były często oznaczane liczbami, tzn. słowami ze zbioru Σ^*, gdzie $\Sigma = \{1,2,3,\ldots\}$. Używano kropek do oddzielania liter w słowach w Σ^*. Zatem 3.4.1.2 oznacza drugi akapit pierwszego punktu czwartego paragrafu rozdziału trzeciego, podczas gdy 3.4.12 oznacza dwunasty punkt czwartego paragrafu rozdziału trzeciego. Ten schemat nie jest zbyt ładny, więc współcześni autorzy zazwyczaj go nie używają, ale ma on pewne zalety, które przenoszą się na dzisiejsze sposoby wykorzystania drzew. Można łatwo wstawić nowe akapity lub paragrafy, nie naruszając systemu numerowania. Można również łatwo usuwać akapity i paragrafy, zwłaszcza jeśli nie przejmujemy się powstałymi lukami. Również etykieta taka jak 3.4.12 mówi nam dokładnie, gdzie dany punkt lub akapit znajduje się w książce. Chociaż sama książka jest wydrukowana liniowo jako jeden długi ciąg znaków, system numerowania pozwala nam wyobrazić sobie związaną z nią strukturę

drzewa. Dla kontrastu, w pewnej znanej książce matematycznej twierdzenia są numerowane od 1 do 460. Etykieta „twierdzenie 303" mówi nam tylko, że twierdzenie prawdopodobnie znajduje się mniej więcej w dwóch trzecich książki.

Oczywiście umiejętność szybkiego zlokalizowania węzła w drzewie jest najbardziej przydatna wtedy, gdy wiemy, który węzeł chcemy znaleźć. Jeśli przechowujemy rekordy pacjentów w drzewie indeksowanym ich nazwiskami i jeśli musimy zajrzeć do każdego węzła, na przykład po to, by stwierdzić, którzy pacjenci zażywają jakieś szczególne lekarstwo, struktura drzewa nie na wiele się przyda. W paragrafie 7.3 zbadamy schematy systematycznego przeszukiwania wszystkich wierzchołków drzewa z wyróżnionym korzeniem.

ĆWICZENIA DO § 6.4

1. (a) Dla drzewa przedstawionego na rysunku 6.29 narysuj drzewo z wyróżnionym nowym korzeniem v.
(b) Jaki jest numer poziomu wierzchołka r?
(c) Jaka jest wysokość tego drzewa?

2. Utwórz drzewo poszukiwań binarnych o wysokości 4 dla zwykłego alfabetu angielskiego $\{a, b, \ldots, z\}$ ze zwykłym porządkiem.

3. (a) Dla każdego drzewa z wyróżnionym korzeniem przedstawionego na rysunku 6.30 podaj numery poziomów wierzchołków oraz wysokość drzewa.
(b) Które drzewa na rysunku 6.30 są drzewami regularnymi o m rozgałęzieniach dla jakiegoś m?

4. Wyjaśnij, dlaczego zwykłe drzewa genealogiczne nie są drzewami z wyróżnionym korzeniem.

5. (a) Istnieje siedem różnych typów drzew z wyróżnionym korzeniem, o wysokości 2, w których każdy węzeł ma co najwyżej dwoje dzieci. Narysuj jedno drzewo każdego typu.
(b) Do którego z typów w ćwiczeniu (a) należą regularne drzewa binarne o wysokości 2.
(c) Które z drzew w ćwiczeniu (a) są pełnymi drzewami binarnymi?
(d) Ile jest różnych typów drzew binarnych o wysokości 2?

6. (a) Powtórz ćwiczenie 5(a) dla siedmiu typów drzew z wyróżnionym korzeniem, o wysokości 3, w których każdy węzeł, który nie jest liściem, ma dwoje dzieci.
(b) Ile jest różnych typów regularnych drzew binarnych o wysokości 3?

7. Dla każdego n narysuj drzewo poszukiwań binarnych, którego węzłami są $1, 2, 3, \ldots, n$ i którego wysokość jest możliwie najmniejsza.

(a) $n = 7$
(b) $n = 15$
(c) $n = 4$
(d) $n = 6$

8. Drzewo z dwoma lub trzema rozgałęzieniami jest takim drzewem z wyróżnionym korzeniem, że każdy wewnętrzny węzeł ma albo dwoje, albo troje dzieci i wszystkie drogi od korzenia do liści mają tę samą długość. Jest siedem różnych typów drzew z dwoma lub z trzema rozgałęzieniami o wysokości 2. Narysuj po jednym drzewie każdego typu. (Drzewa z dwoma lub z trzema rozgałęzieniami stanowią struktury danych, które można stosunkowo łatwo aktualizować).

9. (a) Narysuj pełne drzewo o m rozgałęzieniach, wysokości h dla $m = 2$, $h = 2$; $m = 2$, $h = 3$ oraz $m = 3$, $h = 2$.
(b) Które drzewa w ćwiczeniu (a) mają m^h liści?

10. Weźmy pełne drzewo binarne T o wysokości h.
(a) Ile liści ma T?
(b) Ile wierzchołków ma T?

11. Weźmy pełne drzewo o m rozgałęzieniach, mające p rodziców i t liści. Pokaż, że $t = (m-1)p + 1$, niezależnie od tego, jaka jest wysokość.

12. Podaj kilka przykładów z życia codziennego przechowywania informacji, które można traktować jako drzewa z etykietami.

13. Niech $\Sigma = \{a, b\}$ i weźmy drzewo z wyróżnionym korzeniem Σ_λ^*; zob. przykład 7. Opisz zbiór wierzchołków na poziomie k. Jak duży jest ten zbiór?

14. Narysuj część drzewa z wyróżnionym korzeniem Σ_λ^*, gdzie $\Sigma = \{a, b, c\}$ oraz jak zwykle $a < b < c$.

Następne dwa ćwiczenia ilustrują pewne problemy związane z aktualizacją drzew poszukiwań binarnych.

15. (a) Przypuśćmy, że w przykładzie 1(a) klient Rose przeprowadził się. Jak w naturalny sposób możemy zmienić drzewo, aby usunąć rekordy tego klienta bez zbytniego naruszania reszty drzewa?
(b) Powtórz ćwiczenie (a), jeśli zamiast Rose'a przeprowadził się Brown.

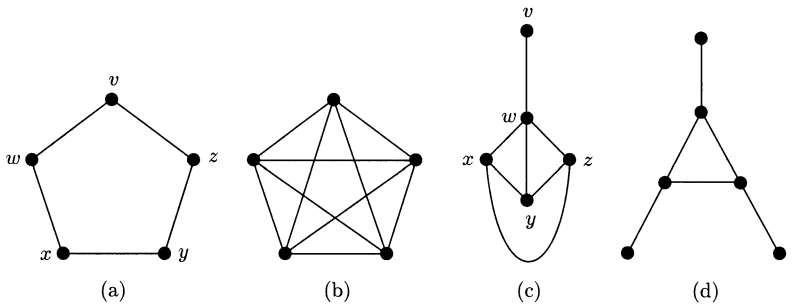
16. (a) Przypuśćmy, że w przykładzie 1(a) nowy klient, Smith musi być dodany do drzewa poszukiwań binarnych. Pokaż, jak to zrobić bez zwiększania wysokości drzewa. Spróbuj nie naruszyć reszty drzewa bardziej niż to będzie konieczne.
(b) Przypuśćmy, że trzech nowych klientów, Smith1, Smith2 i Smith3 musi być dodanych do drzewa z przykładu 1(a). Pokaż, jak to zrobić bez zwiększania wysokości drzewa poszukiwań binarnych.
(c) Co się stanie w ćwiczeniu (b), jeśli będzie czterech nowych klientów?

§ 6.5. Zagadnienia związane z przechodzeniem przez wierzchołki

Twierdzenie Eulera w § 6.2 mówi nam, które grafy mają drogi zamknięte przechodzące przez każdą krawędź dokładnie jeden raz, a algorytm Fleury'ego podaje sposób konstruowania takich dróg, jeśli one istnieją. Z drugiej strony, dużo mniej wiadomo o grafach mających drogi przechodzące przez każdy wierzchołek dokładnie jeden raz. Irlandzki matematyk, Sir William Hamilton był jednym z pierwszych, który badał takie grafy i w pewnym momencie nawet sprzedawał łamigłówkę opartą na tym problemie.

Drogę nazywamy **drogą Hamiltona**, jeśli przechodzi ona przez każdy wierzchołek grafu dokładnie jeden raz. Drogę zamkniętą, która przechodzi przez każdy wierzchołek grafu dokładnie jeden raz, z wyjątkiem ostatniego wierzchołka, którym ponownie jest pierwszy wierzchołek, nazywamy **cyklem Hamiltona**. Graf mający cykl Hamiltona nazywamy **grafem hamiltonowskim**. Droga Hamiltona musi być drogą prostą i na podstawie stwierdzenia 1 w § 6.1, jeśli graf G ma co najmniej trzy wierzchołki, to cykl Hamiltona w grafie G musi być cyklem.

PRZYKŁAD 1 (a) Graf pokazany na rysunku 6.34(a) ma cykl Hamiltona $vwxyzv$.



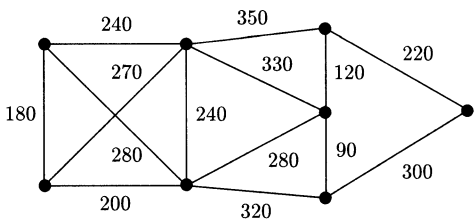
Rysunek 6.34

(b) Graf hamiltonowski, po dodaniu nowych krawędzi nie przestanie być hamiltonowski, więc graf K_5 z rysunku 6.34(b) jest również hamiltonowski. Tak naprawdę, każdy graf pełny K_n dla $n \geq 3$ jest hamiltonowski; możemy przejść wszystkie wierzchołki w dowolnej kolejności.

(c) Graf z rysunku 6.34(c) ma drogę Hamiltona $vwxyz$, ale nie ma cyklu Hamiltona, ponieważ żaden cykl nie przechodzi przez wierzchołek v.

(d) Graf z rysunku 6.34(d) nie ma drogi Hamiltona. ∎

Twierdzenie Eulera powoduje, że teoria cykli Eulera jest zgrabna i kompletna. Czego można dowieść o cyklach Hamiltona? W pewnych warunkach grafy mogą mieć tak wiele krawędzi w stosunku do liczby wierzchołków, że będą musiały one mieć cykle Hamiltona. Ale graf z rysunku 6.34(a) ma bardzo mało krawędzi, a jednak ma cykl Hamiltona. A graf z rysunku 6.34(c) ma mnóstwo krawędzi i nie ma cyklu Hamiltona. Okazuje się, że nie jest znana prosta charakteryzacja tych grafów spójnych, które mają cykle Hamiltona. Pojęcie cyklu Hamiltona wydaje się być bardzo bliskie pojęciu cyklu Eulera, jednak teoria cykli Hamiltona jest znacznie bardziej złożona. W szczególności, nie jest znany żaden efektywny algorytm znajdowania cykli Hamiltona. Ten problem jest szczególnym przypadkiem problemu komiwojażera. W problemie tym mamy do czynienia z grafem, którego krawędziom są przypisane **wagi**, mające oznaczać odległość, koszt, czas pracy komputera, czy jakąś inną wielkość, którą chcemy zminimalizować. Na rysunku 6.35 wagi mogłyby reprezentować odległości między miastami na trasie komiwojażera. Celem jest znalezienie najkrótszej drogi, podczas której odwiedza się każde miasto dokładnie jeden raz. To znaczy, celem jest znalezienie cyklu Hamiltona minimalizującego sumę wag tych krawędzi. Zgrabny algorytm rozwiązujący ten problem będzie w stanie znaleźć cykle Hamiltona w grafie bez wag, ponieważ zawsze możemy każdej krawędzi przypisać wagę 1.



Rysunek 6.35

Oczywiście graf hamiltonowski o n wierzchołkach musi mieć co najmniej n krawędzi. Ten warunek konieczny może nie być wystarczający, jak pokazują to rysunki 6.34(c) i 6.34(d). Oczywiście pętle i krawędzie wielokrotne są bezużyteczne. Następujące twierdzenie podaje prosty warunek wystarczający.

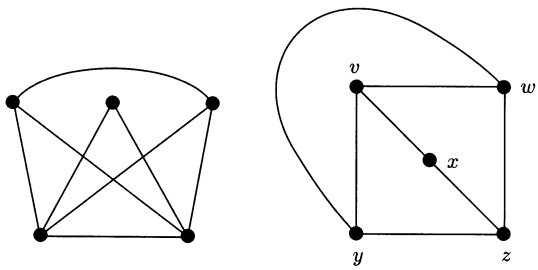
Twierdzenie 1

Jeśli graf G nie ma pętli ani krawędzi wielokrotnych, jeśli $|V(G)| = n \geq 3$ oraz jeśli $\deg(v) \geq n/2$ dla każdego wierzchołka v w grafie G, to graf G jest grafem hamiltonowskim.

PRZYKŁAD 2

(a) W grafie K_5 z rysunku 6.34(b), $\deg(v) = 4$ dla każdego wierzchołka v oraz $|V(G)| = 5$, a więc ten graf spełnia założenia twierdzenia 1.

(b) W każdym z grafów z rysunku 6.36 spełniona jest równość $|V(G)|/2 = 5/2$ oraz istnieje wierzchołek stopnia 2. Nie spełniają one założeń twierdzenia 1, są jednak grafami hamiltonowskimi. ■



Rysunek 6.36

Twierdzenie 1 nakłada jednolite warunki na wszystkie wierzchołki. W naszym następnym twierdzeniu wymagamy tylko, by gdzieś w grafie było dostatecznie wiele krawędzi. Wykażemy to, że te dwa warunki są wystarczające, jako wniosek z twierdzenia 3, podającego pewien warunek nałożony na stopnie par wierzchołków.

Twierdzenie 2

Jeśli graf mający n wierzchołków i nie mający pętli ani krawędzi wielokrotnych ma co najmniej $\frac{1}{2}(n-1)(n-2)+2$ krawędzi, to jest hamiltonowski.

PRZYKŁAD 3

(a) W grafie hamiltonowskim na rysunku 6.36(a) mamy $n = 5$, skąd wynika, że $\frac{1}{2}(n-1)(n-2)+2 = 8$. Ma on 8 krawędzi, a więc spełnia założenia i tezę twierdzenia 2.

(b) W grafie hamiltonowskim na rysunku 6.36(b) mamy $n = 5$, a więc $\frac{1}{2}(n-1)(n-2)+2 = 8$, ale ma on tylko 7 krawędzi. Nie spełnia on założeń twierdzenia 2, ani twierdzenia 1.

Gdyby nie miał wierzchołka w środku, byłby grafem K_4 i $n = 4$, a więc $\frac{1}{2}(n-1)(n-2)+2 = 5$ i 6 krawędzi by w zupełności wystarczyło. W swej obecnej postaci ten graf spełnia założenia następnego twierdzenia. ∎

Twierdzenie 3

Załóżmy, że graf G nie ma pętli ani krawędzi wielokrotnych i $|V(G)| = n \geq 3$. Jeśli

$$\deg(v) + \deg(w) \geq n$$

dla każdej pary wierzchołków v i w nie połączonych krawędzią, to graf G jest hamiltonowski.

PRZYKŁAD 4

W grafie na rysunku 6.36(b) mamy $n = 5$. Są w nim trzy pary różnych wierzchołków, które nie są połączone krawędzią. Sprawdzamy, że spełniają one założenia twierdzenia 3:

dla pary (v, z), $\deg(v) + \deg(z) = 3 + 3 = 6 \geq 5$;
dla pary (w, x), $\deg(w) + \deg(x) = 3 + 2 = 5 \geq 5$;
dla pary (x, y), $\deg(x) + \deg(y) = 2 + 3 = 5 \geq 5$. ∎

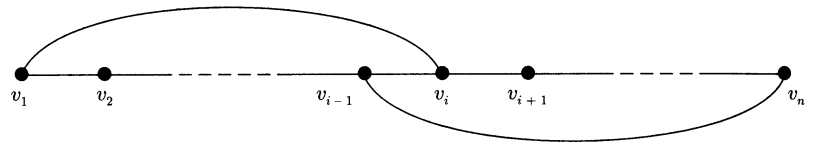
Dowód twierdzenia 3. Przypuśćmy, że twierdzenie jest prawdziwe dla pewnej liczby n i niech G będzie kontrprzykładem takim, że wartość $|E(G)|$ jest jak największa. Graf G jest podgrafem grafu hamiltonowskiego K_n. Dodanie do G krawędzi z grafu K_n daje w wyniku graf, który nadal spełnia warunek nałożony na stopnie wierzchołków i ma więcej niż $|E(G)|$ krawędzi. Ze względu na wybór grafu G każdy taki graf będzie miał cykl Hamiltona. To znaczy, że graf G musi już mieć drogę Hamiltona, której ciągiem wierzchołków jest na przykład $v_1 v_2 \ldots v_n$. Ponieważ graf G nie ma cyklu Hamiltona, wierzchołki v_1 i v_n nie są połączone krawędzią w grafie G, a więc $\deg(v_1) + \deg(v_n) \geq n$.

Zdefiniujmy podzbiory S_1 i S_n zbioru $\{2, \ldots, n\}$ w następujący sposób

$$S_1 = \{i\colon \{v_1, v_i\} \in E(G)\} \quad \text{oraz} \quad S_n = \{i\colon \{v_{i-1}, v_n\} \in E(G)\}.$$

Wtedy $|S_1| = \deg(v_1)$ i $|S_n| = \deg(v_n)$. Ponieważ $|S_1| + |S_n| \geq n$ i zbiór $S_1 \cup S_n$ ma co najwyżej $n-1$ elementów, więc zbiór $S_1 \cap S_n$ musi być niepusty. Istnieje zatem liczba i, dla której $\{v_1, v_i\}$ i $\{v_{i-1}, v_n\}$ są krawędziami w grafie G. Wtedy (zob. rysunek 6.37) droga $v_1 \ldots v_{i-1} v_n \ldots v_i v_1$ jest cyklem Hamiltona w grafie G, co przeczy temu, że graf G został wybrany jako kontrprzykład. ∎

Nasze pierwsze dwa warunki wystarczające wynikają łatwo z twierdzenia 3.



Rysunek 6.37

Dowody twierdzeń 1 i 2. Załóżmy, że graf G nie ma pętli i krawędzi wielokrotnych oraz $|V(G)| = n \geq 3$.

Jeśli $\deg(v) \geq n/2$ dla każdego v, to $\deg(v) + \deg(w) \geq n$ dla każdych v i w, niezależnie od tego, czy są one połączone krawędzią czy nie, a więc założenia twierdzenia 3 są spełnione i graf G jest hamiltonowski.

Załóżmy, że

$$|E(G)| \geq \frac{1}{2}(n-1)(n-2) + 2 = \binom{n-1}{2} + 2$$

i weźmy wierzchołki u i v takie, że $\{u, v\} \notin E(G)$. Usuńmy z grafu G wierzchołki u i v wraz ze wszystkimi krawędziami wychodzącymi z tych wierzchołków. Ponieważ $\{u, v\} \notin E(G)$, więc usunęliśmy $\deg(u) + \deg(v)$ krawędzi i 2 wierzchołki. Graf G', który otrzymaliśmy, jest podgrafem grafu K_{n-2}, a więc

$$\begin{aligned}\binom{n-2}{2} = |E(K_{n-2})| &\geq |E(G')| \geq \\ &\geq \binom{n-1}{2} + 2 - \deg(u) - \deg(v).\end{aligned}$$

Zatem

$$\begin{aligned}\deg(u) + \deg(v) &\geq \binom{n-1}{2} - \binom{n-2}{2} + 2 \\ &= \frac{1}{2}(n-1)(n-2) - \frac{1}{2}(n-2)(n-3) + 2 \\ &= \frac{1}{2}(n-2)[(n-1) - (n-3)] + 2 \\ &= \frac{1}{2}(n-2) \cdot 2 + 2 = n.\end{aligned}$$

Znów graf G spełnia założenia twierdzenia 3. ∎

Twierdzenia 1, 2 i 3 nie zadowalają nas z dwóch powodów. Nie tylko ich warunki wystarczające nie są konieczne, ale te twierdzenia nie dają żadnych wskazówek, jak znaleźć cykl Hamiltona, o którym wiemy, że istnieje. Jak już wspomnieliśmy wcześniej, nie

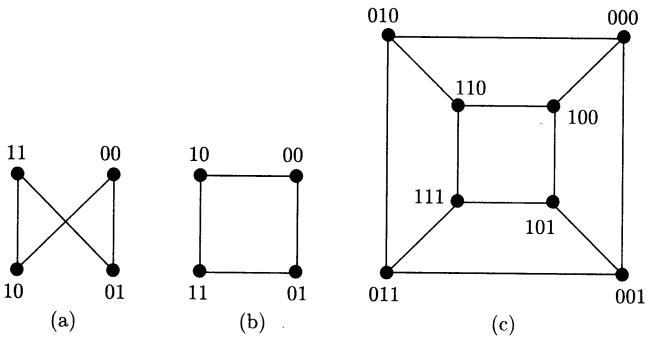
jest znany żaden efektywny algorytm znajdujący drogi lub cykle Hamiltona. Z drugiej strony, graf hamiltonowski musi oczywiście być spójny, zatem te trzy twierdzenia podają warunki wystarczające na to, by graf był spójny.

PRZYKŁAD 5 **Kodem Graya** długości n jest ciąg wszystkich 2^n różnych ciągów n cyfr dwójkowych, ustawionych w taki sposób, że dwa kolejne ciągi różnią się dokładnie jedną cyfrą oraz ostatni ciąg różni się dokładnie jedną cyfrą od pierwszego ciągu. Na przykład, 00, 01, 11, 10 jest ciągiem Graya długości 2.

Możemy popatrzeć na konstrukcję ciągu Graya jak na problem z teorii grafów. Niech $V(G)$ będzie zbiorem $\{0,1\}^n$ wszystkich ciągów cyfr dwójkowych długości n i połączmy ciągi u i v krawędzią, jeśli u i v różnią się dokładnie jedną cyfrą. Kod Graya długości n jest w istocie cyklem Hamiltona w grafie G. Na rysunku 6.38(a) pokazany jest graf G dla $n = 2$. Na rysunku 6.38(b) widzimy ten sam graf narysowany w inny sposób. Ten graf ma dwa cykle Hamiltona, po jednym w każdym kierunku, co pokazuje, że istnieją dwa (w zasadzie równoważne) ciągi Graya długości 2. Na rysunku 6.38(c) pokazany jest taki graf dla $n = 3$. Istnieje 12 kodów Graya długości 3. Na rysunku 6.39 pokazany jest cykl Hamiltona odpowiadający jednemu z tych kodów. Zatem

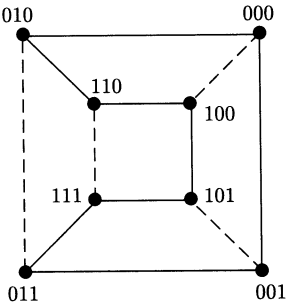
$$000,\ 001,\ 011,\ 111,$$
$$101,\ 100,\ 110,\ 010$$

jest kodem Graya długości 3.



Rysunek 6.38

Kodów Graya można używać do etykietowania pojedynczych procesorów w sieci będącej hiperkostką. (Kwadrat i sześcian na rysunku 6.38 są hiperkostkami odpowiednio wymiarów 2 i 3).



Rysunek 6.39

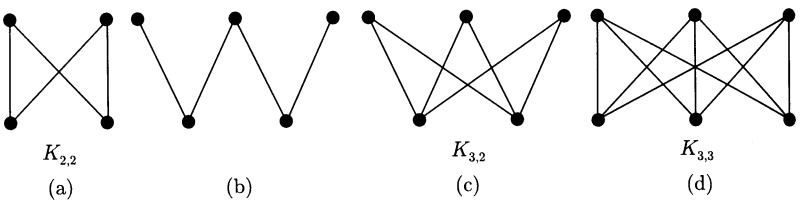
Jeżeli procesory są zaetykietowane zgodnie z tą zasadą, to dwa procesory są połączone wtedy i tylko wtedy, gdy ich etykiety różnią się dokładnie jednym bitem.

■

Wierzchołki grafów skonstruowanych w przykładzie 5 mogą być rozdzielone na dwa zbiory, tych, które mają parzystą liczbę jedynek i tych, które mają nieparzystą liczbę jedynek, w taki sposób, że każda krawędź łączy element jednego zbioru z elementem drugiego. Zakończymy ten paragraf kilkoma obserwacjami dotyczącymi cykli Hamiltona w grafach, w których można dokonać takiego podziału.

Graf G nazywamy **grafem dwudzielnym**, jeśli zbiór $V(G)$ jest sumą dwóch niepustych zbiorów rozłącznych V_1 i V_2 takich, że każda krawędź w grafie G łączy wierzchołek ze zbioru V_1 z wierzchołkiem ze zbioru V_2. Graf nazywamy **pełnym grafem dwudzielnym**, jeśli ponadto każdy wierzchołek zbioru V_1 jest połączony z każdym wierzchołkiem zbioru V_2 dokładnie jedną krawędzią.

PRZYKŁAD 6 Wszystkie grafy pokazane na rysunku 6.40 są grafami dwudzielnymi. Wszystkie oprócz grafu na rysunku 6.40(b) są pełnymi grafami dwudzielnymi. ■



Rysunek 6.40

Dla danych liczb m i n wszystkie pełne grafy dwudzielne takie, że $|V_1| = m$ i $|V_2| = n$, są izomorficzne; oznaczamy je symbolem $K_{m,n}$. Zauważ, że grafy $K_{m,n}$ i $K_{n,m}$ są izomorficzne.

Twierdzenie 4

Niech G będzie grafem dwudzielnym i niech $V(G) = V_1 \cup V_2$ będzie podziałem jego wierzchołków. Jeśli graf G ma cykl Hamiltona, to $|V_1| = |V_2|$. Jeśli graf G ma drogę Hamiltona, to liczby $|V_1|$ i $|V_2|$ różnią się co najwyżej o 1. Dla pełnych grafów dwudzielnych o co najmniej trzech wierzchołkach prawdziwe są również stwierdzenia odwrotne.

Dowód. Wierzchołki na drodze w grafie G należą na przemian do zbiorów V_1 i V_2. Jeśli $x_1x_2 \ldots x_nx_1$ jest drogą zamkniętą przechodzącą dokładnie raz przez każdy wierzchołek, to $x_1, x_3, x_5 \ldots$ muszą należeć do jednego z tych zbiorów, np. do V_1. Ponieważ $\{x_n, x_1\}$ jest krawędzią, liczba n musi być parzysta i wszystkie wierzchołki $x_2, x_4, \ldots, x_n$ należą do V_2. Zatem $|V_1| = |V_2|$. Podobnie postępujemy z drogą niezamkniętą $x_1x_2 \ldots x_n$, z tym tylko, że liczba n może być nieparzysta i wtedy jeden ze zbiorów V_1 i V_2 ma jeden dodatkowy wierzchołek.

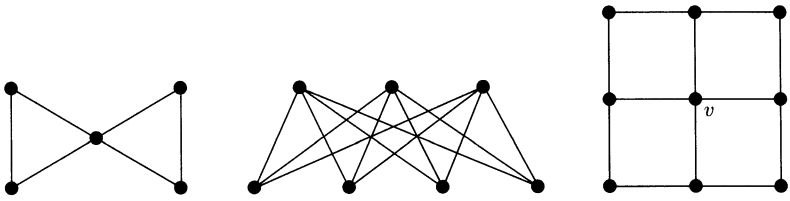
Przypuśćmy teraz, że $G = K_{m,n}$. Jeśli $m = n$, możemy po prostu chodzić tam i z powrotem między V_1 i V_2, gdyż jest wystarczająco dużo krawędzi, byśmy doszli tam, gdzie chcemy. Jeśli $m = n+1$, powinniśmy wyjść z V_1, by utworzyć drogę Hamiltona. ∎

Pewien nasz znajomy informatyk opowiada historyjkę o tym, jak kiedyś spędził ponad dwa tygodnie szukając za pomocą komputera drogi Hamiltona w grafie dwudzielnym o 42 wierzchołkach, zanim uświadomił sobie, że ten graf nie spełnia założeń twierdzenia 4. Historyjka ta ma dwa morały: (1) istnieją praktyczne zastosowania grafów dwudzielnych i dróg Hamiltona oraz (2) *należy najpierw pomyśleć, zanim zacznie się liczyć.*

ĆWICZENIA DO § 6.5

1. (a) Wyjaśnij, dlaczego graf przedstawiony na rysunku 6.34(c) nie ma cyklu Hamiltona.

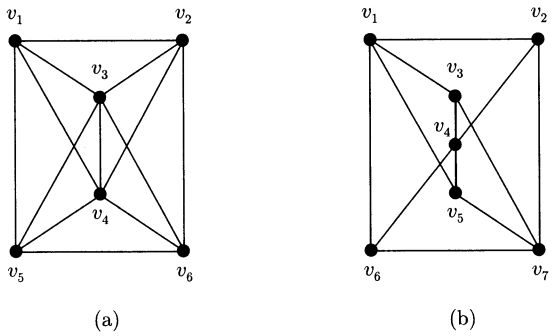
(b) Wyjaśnij, dlaczego graf przedstawiony na rysunku 6.34(d) nie ma drogi Hamiltona.



Rysunek 6.41

2. Dla każdego grafu przedstawionego na rysunku 6.41 podaj cykl Hamiltona lub wyjaśnij, dlaczego żaden taki cykl nie istnieje.

3. Weźmy graf pokazany na rysunku 6.42(a).
(a) Czy jest to graf hamiltonowski?
(b) Czy jest to graf pełny?
(c) Czy jest to graf dwudzielny?
(d) Czy jest to pełny graf dwudzielny?



Rysunek 6.42

4. Odpowiedz na takie same pytania jak w ćwiczeniu 3 dla grafu przedstawionego na rysunku 6.42(b).

5. (a) Ile cykli Hamiltona ma graf $K_{n,n}$ dla $n \geq 2$? (Potraktuj cykle jako różne, jeśli mają one różne punkty początkowe lub ciągi wierzchołków).
(b) Ile jest dróg Hamiltona w $K_{n,n-1}$ dla $n \geq 2$?
(c) Które pełne grafy dwudzielne $K_{m,n}$ mają drogi Eulera?

6. Przerysuj grafy z rysunku 6.38 i zaznacz każdy z podzbiorów V_1 i V_2 podziału zbioru $V(G)$.

7. Rozmieść zera i jedynki na okręgu tak, by każda trzycyfrowa liczba dwójkowa była ciągiem trzech kolejnych symboli na okręgu. *Wskazówka*: znajdź cykl Hamiltona w grafie mającym zbiór wierzchoł-

ków $\{0,1\}^3$ oraz mającym krawędź między wierzchołkami (v_1, v_2, v_3) i (w_1, w_2, w_3), jeśli $(v_1, v_2) = (w_2, w_3)$ lub $(v_2, v_3) = (w_1, w_2)$.

8. Podaj dwa przykłady kodów Graya długości 3 innych niż kod podany w przykładzie 5.

9. Weźmy graf, który ma zbiór wierzchołków $\{0,1\}^3$ i krawędź między wierzchołkami zawsze wtedy, gdy wierzchołki różnią się na dwóch współrzędnych. Czy ten graf ma cykl Hamiltona? Czy ma on drogę Hamiltona?

10. Powtórz ćwiczenie 9 dla grafu, który ma zbiór wierzchołków $\{0,1\}^3$ i krawędź między wierzchołkami, jeśli różnią się one na dwóch lub trzech współrzędnych.

11. Dla $n \geq 4$ zbuduj graf K_n^+ z grafu pełnego K_{n-1} przez dodanie jednego wierzchołka w środku jakiejś krawędzi grafu K_{n-1} (rys. 6.36(b) pokazuje graf K_5^+).
(a) Pokaż, że graf K_n^+ nie spełnia założeń twierdzenia 2.
(b) Skorzystaj z twierdzenia 3, aby pokazać, że K_n^+ jest pomimo to grafem hamiltonowskim.

12. Dla $n \geq 4$ zbuduj graf K_n^{++} z grafu pełnego K_{n-1} przez dodanie jednego wierzchołka i krawędzi od nowego wierzchołka do jakiegoś wierzchołka grafu K_{n-1} (rys. 6.34(c) pokazuje graf K_5^{++}). Pokaż, że graf K_n^{++} nie jest grafem hamiltonowskim. Zauważ, że K_n^{++} ma n wierzchołków i $\frac{1}{2}(n-1)(n-2)+1$ krawędzi. Przykład ten pokazuje, że liczba krawędzi wymagana w twierdzeniu 2 nie może być zmniejszona.

13. **Dopełnieniem** grafu G jest graf mający zbiór wierzchołków $V(G)$ i mający krawędź między różnymi wierzchołkami v i w, jeśli graf G *nie* ma krawędzi łączącej v i w.
(a) Narysuj dopełnienie grafu przedstawionego na rysunku 6.36(b).
(b) Ile składowych ma dopełnienie z ćwiczenia (a)?
(c) Pokaż, że jeśli G nie jest grafem spójnym, to jego dopełnienie jest grafem spójnym.
(d) Podaj przykład grafu, który jest izomorficzny ze swoim dopełnieniem.
(e) Czy zdanie odwrotne do zdania w ćwiczeniu (c) jest prawdziwe?

14. Przypuśćmy, że graf G jest grafem regularnym stopnia $k \geq 1$ (tzn. każdy wierzchołek ma stopień k) i ma co najmniej $2k+2$ wierzchołków. Pokaż, że dopełnienie grafu G jest grafem hamiltonowskim. *Wskazówka*: skorzystaj z twierdzenia 1.

15. Pokaż, że kody Graya długości n zawsze istnieją. *Wskazówka*: zastosuj indukcję względem n i weź graf G_n, w którym cykl Hamiltona odpowiada kodowi Graya długości n, jak opisaliśmy to w przykładzie 5.

16. Wyjaśnij, dlaczego żadne z twierdzeń w tym paragrafie nie może być użyte do rozwiązania ćwiczenia 15.

§ 6.6. Minimalne drzewa spinające

Twierdzenia charakteryzujące drzewa sugerują dwie metody szukania drzew spinających skończonych grafów spójnych. Wykorzystując pomysł dowodu twierdzenia 1 w § 6.3 moglibyśmy po prostu usuwać krawędzie jedną po drugiej, dbając o to, by nie zniszczyć spójności, to znaczy usuwać krawędzie należące do cykli, dotąd, aż będziemy zmuszeni się zatrzymać. Jeżeli graf ma n wierzchołków i więcej niż $2n$ krawędzi, ta procedura zbada i odrzuci więcej niż połowę krawędzi. Można by było szybciej (gdyby się udało) zbudować drzewo spinające wybierając po jednej $n-1$ krawędzi, tak, by w każdym kroku podgraf złożony z wybranych krawędzi był acykliczny. Wszystkie algorytmy w tym paragrafie będą budować drzewa w ten drugi sposób.

W naszym pierwszym algorytmie punktem wyjścia jest wybrany na początku wierzchołek v. Jeśli dany graf jest spójny, to algorytm utworzy jego drzewo spinające. W przeciwnym przypadku algorytm utworzy drzewo spinające spójnej składowej tego grafu, zawierającej wierzchołek v.

Algorytm DRZEWO(v)

{Dane: wierzchołek v w skończonym grafie G}
{Wyniki: zbiór krawędzi E drzewa spinającego składowej grafu G zawierającej v}
{Zmienna pomocnicza: ciąg V odwiedzanych wierzchołków, ostatecznie $V = V(G)$}
Niech $V:=\{v\}$ oraz $E:=\emptyset$.
Dopóki istnieją krawędzie w grafie G łączące wierzchołki ze zbioru V z wierzchołkami, które nie należą do V, wykonuj
  wybierz taką krawędź $\{u,w\}$ łączącą wierzchołek $u \in V$ z wierzchołkiem $w \notin V$,
  dołącz wierzchołek w do V i krawędź $\{u,w\}$ do E. ∎

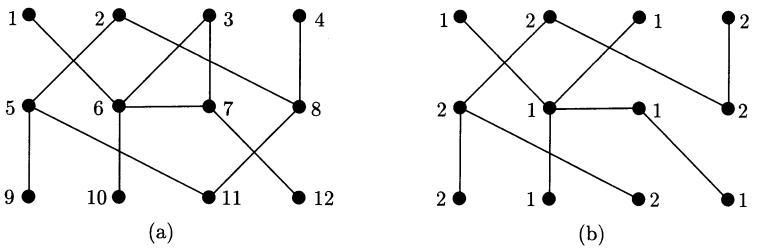
Aby otrzymać las spinający dla grafu G, będziemy sadzić nowe drzewa.

Algorytm LAS

{Dane: skończony graf G}
{Wyniki: zbiór EE krawędzi lasu spinającego dla grafu G}
Niech $VV:=\emptyset$ oraz $EE:=\emptyset$.
Dopóki $VV \neq V(G)$, wykonuj
  wybierz $v \in V(G)\backslash VV$

wykonaj procedurę DRZEWO(v) {po to, by otrzymać zbiór wierzchołków V i zbiór krawędzi E drzewa spinającego składowej zawierającej v}
dołącz wierzchołki zbioru V do zbioru VV i elementy zbioru E do zbioru EE. ■

PRZYKŁAD 1 Możemy zilustrować działanie procedur DRZEWO i LAS na przykładzie grafu pokazanego na rysunku 6.43(a). Poniżej przedstawiono kolejne kroki wykonania algorytmów DRZEWO(1) i DRZEWO(2), a następnie algorytmu LAS, który łączy wyniki ich działania. Wtedy, gdy mamy dokonać wyboru, wybieramy wierzchołki w rosnącym porządku ich numerów i wyczerpujemy wszystkie krawędzie wychodzące z danego wierzchołka, zanim przejdziemy do następnego wierzchołka. Oczywiście są możliwe inne schematy wyboru. Na rysunku 6.43(b) widzimy oba drzewa rosnące w lesie spinającym, numery wierzchołków pokazują, czy dany wierzchołek należy do składowej wierzchołka 1, czy do składowej wierzchołka 2. ■



Rysunek 6.43

DRZEWO(1)
V: $=\{1\}$; E: $=\emptyset$.
Wybierz krawędź $\{1,6\}$.
V: $=\{1,6\}$; E: $=\{\{1,6\}\}$.
Wybierz krawędzie $\{6,3\}, \{6,7\}, \{6,10\}$ (kolejność nie ma tutaj znaczenia).
V: $=\{1,6,3,7,10\}$; E: $=\{\{1,6\},\{6,3\},\{6,7\},\{6,10\}\}$.
Wybierz krawędź $\{7,12\}$ (a nie $\{3,7\}$, ponieważ zarówno 3, jak i 7 należą do V).
V: $=\{1,6,3,7,10,12\}$; E: $=\{\{1,6\},\{6,3\},\{6,7\},\{6,10\},\{7,12\}\}$.

DRZEWO(2)
V: $=\{2\}$; E: $=\emptyset$.

Wybierz krawędzie $\{2,5\},\{2,8\}$.
V: $=\{2,5,8\}$; E: $=\{\{2,5\},\{2,8\}\}$.
Wybierz krawędzie $\{5,9\},\{5,11\}\}$.
V: $=\{2,5,8,9,11\}$; E: $=\{\{2,5\},\{2,8\},\{5,9\},\{5,11\}\}$.
Wybierz krawędź $\{8,4\}$.
V: $=\{2,5,8,9,11,4\}$; E: $=\{\{2,5\},\{2,8\},\{5,9\},\{5,11\},\{8,4\}\}$.

LAS
Połącz oba zbiory E z procedur DRZEWO(1) i DRZEWO(2), aby utworzyć zbiór EE.

Twierdzenie 1

Procedura DRZEWO(v) tworzy drzewo spinające składowej grafu G, zawierającej wierzchołek v. Zatem procedura LAS tworzy las spinający grafu G.

Dowód. Oto dla przypomnienia treść procedury DRZEWO(v):

Algorytm DRZEWO(v)
Niech V: $=\{v\}$ oraz E: $=\emptyset$.
Dopóki istnieją krawędzie w grafie G łączące wierzchołki ze zbioru V z wierzchołkami, które nie należą do V, wykonuj
wybierz taką krawędź $\{u,w\}$ łączącą wierzchołek $u \in V$ z wierzchołkiem $w \notin V$,
dołącz wierzchołek w do V i krawędź $\{u,w\}$ do E.

Każdy przebieg pętli „dopóki" zwiększa wielkość zbioru V, a więc algorytm musi się wreszcie zatrzymać. Zdanie „V jest zbiorem wierzchołków i E jest zbiorem krawędzi drzewa zawierającego wierzchołek v" jest oczywiście prawdziwe wtedy, gdy po raz pierwszy wchodzimy w pętlę. Pokażemy, że jest to niezmiennik pętli, a więc w chwili, gdy algorytm zatrzyma się, da w wyniku drzewo zawierające v. Z twierdzenia 3 w § 6.3 wynika, że dodanie do drzewa nowego wierzchołka wraz z nową krawędzią daje w wyniku drzewo. Ponieważ nasza pętla dokładnie to robi, rozważane zdanie jest niezmiennikiem tej pętli. Kiedy algorytm zatrzyma się, w składowej zawierającej wierzchołek v nie będzie już wierzchołków nie należących do V, gdyż w przeciwnym przypadku warunek dozoru pętli byłby nadal spełniony. Zatem algorytm DRZEWO(v) działa zgodnie z wymaganiami. ∎

Czas, w jakim działa procedura DRZEWO(v), zależy od sposobu dokonywania wyborów i od tego, w jaki sposób są ułożone

listy dostępnych krawędzi. Jeśli zaznaczymy każdy wierzchołek, gdy go wybieramy i jednocześnie powiemy jego sąsiadom, że jest on zaznaczony, to każdy wierzchołek i każda krawędź są rozpatrywane tylko jeden raz. Jeśli graf G jest spójny, to DRZEWO(v) tworzy drzewo spinające w czasie $O(|V(G)| + |E(G)|)$. To samo rozumowanie pokazuje, że w ogólnym przypadku algorytm LAS działa również w czasie $O(|V(G)| + |E(G)|)$.

Liczba składowych grafu G jest równa liczbie drzew w lesie spinającym. Można łatwo śledzić składowe, kiedy tworzymy drzewa za pomocą procedury LAS, korzystając z funkcji C, która przypisuje wartość u każdemu wierzchołkowi w w ciągu V tworzonym przez procedurę DRZEWO(u). Każdy przebieg pętli „dopóki" w procedurze LAS nadaje wtedy funkcji C inną wartość, taką samą dla wszystkich wierzchołków w składowej odpowiadającej temu przebiegowi. Etykiety 1 i 2 na rysunku 6.43(b) zostały utworzone przez taką funkcję C. Aby zmodyfikować odpowiednio algorytm, po prostu dodaj instrukcję $C(v) := v$ na początku procedury DRZEWO(v) i dodaj instrukcję $C(w) := C(u)$ po dołączeniu krawędzi $\{u, w\}$ do zbioru E.

Procedura LAS może być używana do sprawdzania, czy graf jest spójny; po prostu sprawdź, czy ten algorytm tworzy więcej niż jedno drzewo. Algorytm LAS może również być użyty jako względnie szybki test sprawdzający istnienie cykli w grafie. Jeśli graf G jest acykliczny, to utworzony las spinający jest samym grafem G; w przeciwnym przypadku $|EE| < |E(G)|$ po zakończeniu działania algorytmu.

Problem znajdowania drzew spinających jest szczególnie interesujący wtedy, gdy krawędzie mają wagi, tzn. gdy każdej krawędzi e grafu G jest przypisana nieujemna liczba $W(e)$. **Waga** $W(H)$ podgrafu H grafu G jest po prostu sumą wag krawędzi grafu H. Problem polega na znalezieniu **minimalnego drzewa spinającego**, tzn. drzewa spinającego, którego waga jest mniejsza lub równa wadze dowolnego innego drzewa spinającego. Jeśli graf G nie ma wag i przypiszemy każdej krawędzi wagę 1, to wszystkie drzewa spinające będą minimalne, gdyż wszystkie będą miały wagę $|V(G)| - 1$.

Nasz pierwszy algorytm tworzy minimalne drzewo spinające grafu G z wagami, którego krawędzie $e_1, \ldots, e_m$ zostały na początku posortowane w taki sposób, że

$$W(e_1) \leq W(e_2) \leq \ldots \leq W(e_m).$$

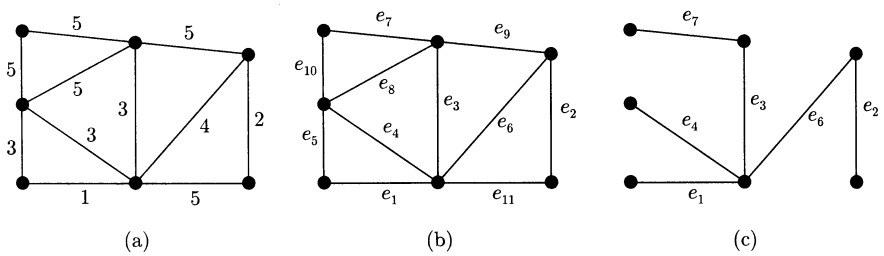
Algorytm przegląda wyraz po wyrazie ciąg krawędzi grafu G, począwszy od najmniejszych wag i wybiera krawędzie, których

dodanie nie tworzy cykli. Kiedy algorytm zatrzyma się, zbiór E okaże się być zbiorem krawędzi minimalnego drzewa spinającego grafu G. Symbolem $E \cup \{e_j\}$ oznaczamy w treści algorytmu podgraf, którego zbiorem krawędzi jest zbiór $E \cup \{e_j\}$, a zbiorem wierzchołków jest $V(G)$.

Algorytm Kruskala

{Dane: skończony graf spójny G z wagami, którego krawędzie są uporządkowane według wzrastających wag}
{Wyniki: zbiór E krawędzi minimalnego drzewa spinającego grafu G}
Niech E: $= \emptyset$.
Dla $j = 1$ do $|E(G)|$
jeśli graf $E \cup \{e_j\}$ jest acykliczny, to dołącz e_j do E. ■

PRZYKŁAD 2 Na rysunku 6.44(a) pokazany jest graf z wagami, wagi zaznaczone są obok krawędzi. Na rysunku 6.44(b) pokazany jest jeden z możliwych sposobów ponumerowania krawędzi tego grafu, tak, by wagi krawędzi tworzyły ciąg niemalejący, tzn. tak, że $W(e_i) \leq W(e_j)$ dla $i < j$.



Rysunek 6.44

Zastosowanie algorytmu Kruskala do tego grafu z wagami daje w wyniku naszkicowane na rysunku 6.44(c) drzewo spinające T, którego krawędziami są $e_1, e_2, e_3, e_4, e_6, e_7$. Krawędź e_5 została odrzucona, gdyż droga $e_1e_4e_5$ byłaby cyklem. Z podobnych powodów zostały odrzucone krawędzie od e_8 do e_{11}. Drzewo spinające T ma wagę 18. ■

Twierdzenie 2

Algorytm Kruskala daje w wyniku minimalne drzewo spinające.

Dowód. Pokażemy najpierw, że zdanie „zbiór E jest zawarty w pewnym minimalnym drzewie spinającym grafu G" jest niezmiennikiem pętli. To zdanie jest oczywiście prawdziwe na początku, gdy zbiór E jest pusty. Przypuśćmy, że jest ono prawdziwe na początku j-ego przebiegu pętli, czyli zbiór E jest zawarty w pewnym minimalnym drzewie spinającym T. Jeśli zbiór $E \cup \{e_j\}$ nie jest acykliczny, to zbiór E nie zmieni się, a więc możemy założyć, że zbiór $E \cup \{e_j\}$ jest acykliczny. Chcemy znaleźć minimalne drzewo spinające T^* zawierające $E \cup \{e_j\}$. Jeśli e_j należy do T, to możemy przyjąć, że $T^* = T$. Zatem przyjmijmy, że e_j nie należy do zbioru T.

Krawędź e_j należy więc do pewnego cyklu C w grafie $T \cup \{e_j\}$ na podstawie twierdzenia 3 z § 6.3. Ponieważ graf $E \cup \{e_j\}$ jest acykliczny, więc cykl C musi zawierać jakąś krawędź f należącą do $T \backslash (E \cup \{e_j\})$. Określmy $T^* = (T \cup C) \backslash \{f\} = (T \cup \{e_j\}) \backslash \{f\}$. Wtedy graf T^* jest spójny, zawiera wszystkie wierzchołki grafu G i ma $|V(G)| - 1$ krawędzi. Z twierdzenia 3 w § 6.3 wynika zatem, że T^* jest drzewem spinającym grafu G, zawierającym $E \cup \{e_j\}$. Ponieważ $E \cup \{f\} \subseteq T$, więc graf $E \cup \{f\}$ jest acykliczny. Ale krawędź f nie była dotychczas wybrana po to, by dołączyć ją do E, więc krawędź e_j była pierwsza w kolejce, co oznacza, że $W(e_j) \leq W(f)$. Ale $W(T^*) = W(T) + W(e_j) - W(f) \leq W(T)$, więc T^* rzeczywiście jest minimalnym drzewem spinającym.

Ponieważ zbiór E jest zawsze zawarty w minimalnym drzewie spinającym, więc wystarczy pokazać, że graf, którego zbiorem krawędzi jest E i którego zbiorem wierzchołków jest $V(G)$, jest spójny w momencie, gdy algorytm się zatrzymuje. Niech u i v będą dwoma wierzchołkami grafu G. Ponieważ graf G jest spójny, istnieje droga z u do v w grafie G. Jeśli jakaś krawędź f na tej drodze nie należy do E, to $E \cup \{f\}$ zawiera cykl (gdyż w przeciwnym przypadku krawędź f byłaby wybrana, gdy przyszła na nią kolej), a więc krawędź f może być w tej drodze zastąpiona tą częścią cyklu, która jest w E. Zastępując w ten sposób inne krawędzie, otrzymamy drogę z u do v leżącą całkowicie w E. ∎

Zauważmy przy okazji, że algorytm Kruskala działa poprawnie nawet, jeśli graf G ma pętle i krawędzie wielokrotne. Ten algorytm nigdy nie wybiera pętli i spośród krawędzi wielokrotnych wybierze pierwszą na liście. Nie jest nawet konieczne założenie spójności grafu G, by móc wykonać algorytm Kruskala. W ogólnym przypadku algorytm daje w wyniku minimalny las spinający składający się z minimalnych drzew spinających poszczególnych składowych grafu G.

Jeśli graf G ma n wierzchołków, to algorytm Kruskala nie może dołączyć więcej niż $n-1$ krawędzi do zbioru E. Ten algorytm może być zaprogramowany w taki sposób, by zatrzymał się, gdy $|E| = n-1$, ale nadal może okazać się, że będzie musiał zbadać wszystkie krawędzie grafu G, zanim się zatrzyma. Należy sprawdzić, czy każda badana gałąź e_j należy do cyklu. Można użyć algorytmu LAS do grafu $E \cup \{e_j\}$, by sprawdzić, czy ten graf zawiera cykl. Można też w inny sposób użyć pomysłu z algorytmu LAS, aby sprawdzić acykliczność; można skorzystać w tym celu z obserwacji dokonanych w twierdzeniu 2 z § 6.3. Przypuśćmy, że G' jest grafem takim, że $V(G') = V(G)$ oraz $E(G') = E$, wtedy, gdy algorytm bada krawędź e_j. Jeśli wiemy, do jakich składowych grafu G' należą końce krawędzi e_j, to możemy dodać e_j do zbioru E, gdy leżą one w różnych składowych i odrzucić e_j w przeciwnym przypadku.

Ten test jest szybki, o ile śledzimy składowe. Na początku każda składowa składa się z pojedynczego wierzchołka, łatwo jest też uaktualnić listę składowych po zaakceptowaniu krawędzi e_j; składowe końców e_j łączą się w jedną składową. Otrzymana w ten sposób wersja algorytmu Kruskala działa w czasie $O(|E(G)| \cdot \log_2 |E(G)|)$, wliczając w to czas potrzebny do posortowania $E(G)$ na początku. Więcej szczegółów na ten temat można znaleźć na przykład w opisach algorytmu Kruskala w książkach: Aho, Hopcroft, Ullman, *Data Structures and Algorithms*; Cormen, Leiserson, Rivest, *Introduction to Algorithms* lub Tarjan, *Data Structures and Network Algorithms*.

W przypadku grafu z przykładu 2 szybciej byłoby usunąć kilka złych krawędzi z grafu G niż budować drzewo T po jednej krawędzi. Istnieje ogólny algorytm, którego działanie polega na usuwaniu krawędzi: dla danego grafu spójnego, którego krawędzie są ustawione w kolejności od najmniejszej wagi do największej, przeglądamy ten ciąg krawędzi, począwszy od krawędzi o największej wadze i odrzucamy krawędź wtedy i tylko wtedy, gdy należy ona do cyklu w aktualnym podgrafie grafu G. Wszystkie podgrafy, które powstają w czasie działania tego algorytmu, są spójne i algorytm zatrzymuje się dopiero wtedy, gdy osiągnie graf acykliczny, a więc daje w wyniku drzewo spinające grafu G. Jest to w istocie minimalne drzewo spinające (por. ćwiczenie 14). Jest to w rzeczywistości to samo drzewo, które otrzymujemy w wyniku działania algorytmu Kruskala. Jeśli $|E(G)| < 2|V(G)| - 1$, to ta procedura może działać krócej niż algorytm Kruskala, ale oczywiście, jeśli graf G ma tak mało krawędzi, to oba algorytmy działają dość szybko.

Algorytm Kruskala daje gwarancję, że budowany podgraf jest zawsze acykliczny, podczas gdy opisana przed chwilą procedura usuwania krawędzi gwarantuje, że wszystkie podgrafy są spójne. Oba algorytmy są algorytmami zachłannymi, co oznacza, że wybierają zawsze najmniejszą krawędź, którą należy dodać lub największą krawędź, którą należy odrzucić. Na szczęście w tym przypadku zachłanność opłaca się.

Algorytm, który teraz opiszemy, jest podwójnie zachłanny; dokonuje on wyborów minimalnych i jednocześnie dba o to, że otrzymany podgraf jest zarówno acykliczny, jak i spójny. Ponadto nie wymaga on, by krawędzie grafu G były na początku posortowane. Ta procedura działa tak jak DRZEWO(v), ale bierze pod uwagę wagi krawędzi. Buduje ona drzewo T wewnątrz grafu G tak, by $V(T) = V$ i $E(T) = E$. W każdym kroku algorytm szuka krawędzi o najmniejszej wadze, łączącej jakiś wierzchołek drzewa T z jakimś nowym wierzchołkiem spoza T. Dodaje wtedy taką krawędź i taki wierzchołek do T i powtarza ten proces.

Algorytm Prima

{Dane: skończony graf spójny G z wagami (z krawędziami wypisanymi w jakimkolwiek porządku)}

{Wyniki: zbiór krawędzi E minimalnego drzewa spinającego grafu G}

Niech E: $= \emptyset$.

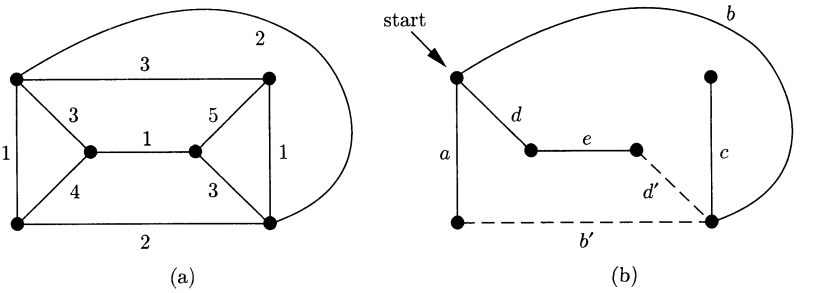
Wybierz w ze zbioru $V(G)$ i niech V: $= \{w\}$.

Dopóki $V \neq V(G)$, wykonuj

wybierz w zbiorze $E(G)$ krawędź $\{u, v\}$ o najmniejszej możliwej wadze, taką że $u \in V$ i $v \in V(G) \backslash V$

dołącz krawędź $\{u, v\}$ do zbioru E i wierzchołek v do zbioru V. ∎

PRZYKŁAD 3 Zastosujemy algorytm Prima do grafu z wagami pokazanego na rysunku 6.45(a). Ponieważ w każdym kroku wykonywania algorytmu można dokonać wielu możliwych wyborów, otrzymane drzewo nie jest wyznaczone jednoznacznie. Linie ciągłe na rysunku 6.45(b) pokazują jeden możliwy sposób wyboru, przy czym krawędzie zostały wybrane w kolejności a, b, c, d, e. Inny wybór polegałby na wybraniu krawędzi b' i d'. Zauważmy, że algorytm Kruskala wybrałby krawędzie c i e przed krawędzią b. ∎



Rysunek 6.45

Twierdzenie 3 Algorytm Prima daje w wyniku minimalne drzewo spinające grafu spójnego z wagami.

Dowód. Dowód twierdzenia 1, a mianowicie to, że algorytm DRZEWO(v) działa poprawnie, pokazuje, że algorytm Prima zatrzymuje się i daje w wyniku drzewo spinające grafu G; to, co jest nowe, to zachłanność. Zdanie „drzewo T jest zawarte w minimalnym drzewie spinającym grafu G" jest oczywiście prawdziwe na początku, gdy drzewo T jest po prostu pojedynczym wierzchołkiem. Twierdzimy, że to zdanie jest niezmiennikiem pętli „dopóki".

Załóżmy, że na początku pewnego przebiegu pętli „dopóki" drzewo T jest zawarte w minimalnym drzewie spinającym T^* grafu G. Przypuśćmy, że algorytm wybiera teraz krawędź $\{u, v\}$. Jeśli krawędź $\{u, v\} \in E(T^*)$, to nowe drzewo T jest nadal zawarte w T^*, co jest wspaniałe. Przypuśćmy więc, że jest przeciwnie. Ponieważ T^* jest drzewem spinającym, istnieje w T^* droga z u do v. Ponieważ $u \in V$ i $v \notin V$, więc na tej drodze musi znajdować się jakaś krawędź łącząca wierzchołek z należący do V z wierzchołkiem w należącym do $V(G) \backslash V$. Ponieważ algorytm Prima wybrał $\{u, v\}$ zamiast $\{z, w\}$, więc $W(u, v) \leq W(z, w)$. Wyrzućmy parę $\{z, w\}$ z $E(T^*)$ i zastąpmy ją parą $\{u, v\}$. Nowy graf T^{**} jest nadal spójny, a więc jest drzewem na podstawie twierdzenia 3 z § 6.3. Ponieważ $W(T^{**}) \leq W(T^*)$, więc T^{**} również jest minimalnym drzewem spinającym i T^{**} zawiera nowy zbiór T. Na końcu przebiegu pętli zbiór T jest nadal zawarty w pewnym minimalnym drzewie spinającym, czego chcieliśmy dowieść. ∎

Algorytm Prima wykonuje $n-1$ przebiegów pętli „dopóki" dla grafu G mającego n wierzchołków. W każdym przebiegu wybie-

rana jest najmniejsza krawędź spełniająca pewien podany warunek. Głupi sposób implementacji wymagałby przejrzenia wszystkich krawędzi w $E(G)$, by znaleźć właściwą. W mądrzejszej implementacji dla każdego wierzchołka $x \in V(G)\backslash V$ zapamiętujemy wierzchołek $u \in V$, dla którego wartość $W(u, v)$ jest najmniejsza i tę wartość również zapamiętujemy. Algorytm wtedy po prostu przegląda listę wierzchołków x należących do $V(G)\backslash V$, znajduje najmniejszą wartość $W(u, x)$, dodaje krawędź $\{u, x\}$ do E i dodaje x do V. Następnie dla każdego $y \in V(G)\backslash V$ sprawdza, czy x jest teraz najbliższym wierzchołka y wierzchołkiem zbioru V i jeśli tak jest, to uaktualnia dane dla y. Czas potrzebny do znalezienia wierzchołka x położonego najbliżej zbioru V oraz do uaktualnienia danych jest rzędu $O(n)$, a więc algorytm Prima z opisaną przed chwilą implementacją działa w czasie $O(n^2)$.

Można łatwo zmodyfikować algorytm Prima (ćwiczenie 13) tak, aby dawał w wyniku minimalny las spinający grafu, niezależnie od tego, czy ten graf jest spójny. Przy powyższym sformułowaniu algorytm załamie się, jeśli dany graf nie jest spójny.

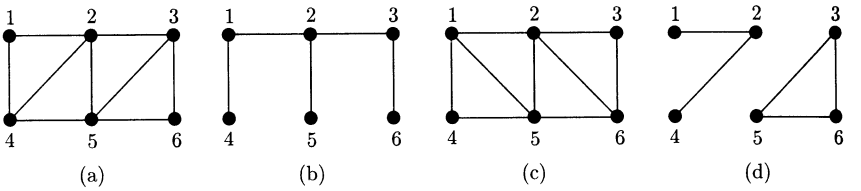
Na zakończenie zauważmy, że waga minimalnego drzewa spinającego pomaga w znalezieniu ograniczenia dolnego dla problemu komiwojażera wspomnianego w § 6.5. Przypuśćmy, że droga $C = e_1 e_2 \ldots e_n$ jest rozwiązaniem problemu komiwojażera dla grafu G, to znaczy jest cyklem Hamiltona o najmniejszej wadze. Wtedy droga $e_2 \ldots e_n$ przechodzi przez każdy wierzchołek dokładnie jeden raz, a więc jest drzewem spinającym grafu G. Jeśli M jest wagą minimalnego drzewa spinającego grafu G (jest to liczba, którą możemy obliczyć za pomocą algorytmu Kruskala lub Prima), to

$$M \leq W(e_2) + \ldots + W(e_n) = W(C) - W(e_1).$$

Zatem $W(C) \geq M +$ (najmniejsza waga krawędzi w G). Jeśli będziemy umieli znaleźć za pomocą jakiejkolwiek metody jakiś cykl Hamiltona C w grafie G, którego waga byłaby bliska $M+$(najmniejsza waga krawędzi), to prawdopodobnie powinniśmy wziąć go i nie tracić czasu na szukanie lepszego.

ĆWICZENIA DO § 6.6

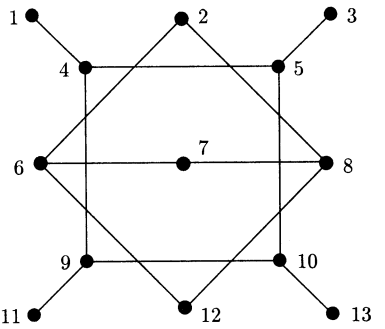
1. (a) Zastosuj procedurę DRZEWO(1) do grafu przedstawionego na rysunku 6.46(a). Zrób rysunek otrzymanego drzewa i oznacz jego krawędzie przez a, b, c, d, e, f w takim porządku, w jakim zostały one wybrane. Skorzystaj ze schematu wyboru przedstawionego w przykładzie 1.



Rysunek 6.46

(b) Powtórz ćwiczenie (a) dla grafu z rysunku 6.46(b).
(c) Powtórz ćwiczenie (a) dla grafu z rysunku 6.46(c).
(d) Powtórz ćwiczenie (a) dla grafu z rysunku 6.46(d).

2. Zastosuj procedurę LAS do grafu przedstawionego na rysunku 6.47.



Rysunek 6.47

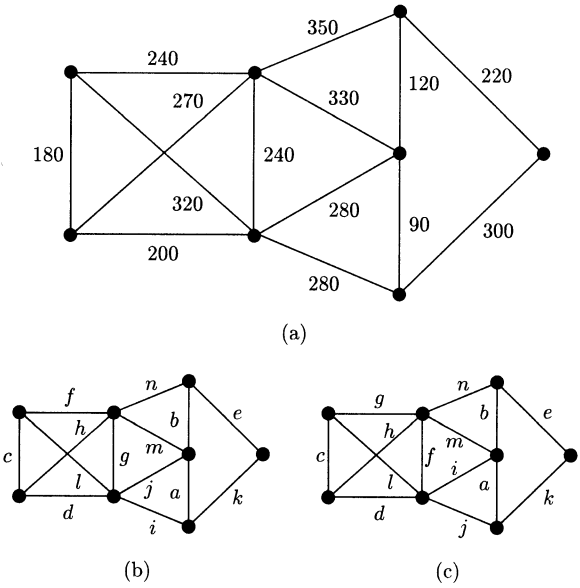
3. Na rysunku 6.48(a) pokazany jest graf z wagami, a na rysunkach 6.48(b) i 6.48(c) pokazane są dwa różne sposoby oznaczenia etykietami jego krawędzi za pomocą liter $a, b, \ldots n$ w kolejności rosnących wag.
(a) Zastosuj algorytm Kruskala do grafu o krawędziach uporządkowanych tak, jak na rysunku 6.48(b). Narysuj otrzymane minimalne drzewo spinające i podaj jego wagę.
(b) Powtórz ćwiczenie (a), kiedy krawędzie są uporządkowane tak, jak na rysunku 6.48(c).

4. Przypuśćmy, że graf przedstawiony na rysunku 6.48(b) ma wagi takie, że

$$W(a) \geq W(b) \geq \ldots \geq W(n).$$

Narysuj minimalne drzewo spinające tego grafu.

5. (a) Zastosuj algorytm Prima do grafu przedstawionego na rysunku 6.48(a), zaczynając w wierzchołku stopnia 4. Narysuj otrzymane minimalne drzewo spinające i oznacz jego krawędzie alfabetycznie w takim porządku, w jakim były wybierane.
(b) Jaka jest waga minimalnego drzewa spinającego tego grafu?
(c) Ile różnych rozwiązań ma ćwiczenie (a)?



Rysunek 6.48

6. Przypuśćmy, że graf przedstawiony na rysunku 6.23(a) w § 6.3 ma wagi takie, że

$$W(e_1) > W(e_2) > \ldots > W(e_{10}).$$

(a) Wypisz krawędzie minimalnego drzewa spinającego tego grafu w takim porządku, w jakim wybierałby je algorytm Kruskala.
(b) Powtórz ćwiczenie (a) dla algorytmu Prima, zaczynając w prawym górnym wierzchołku.

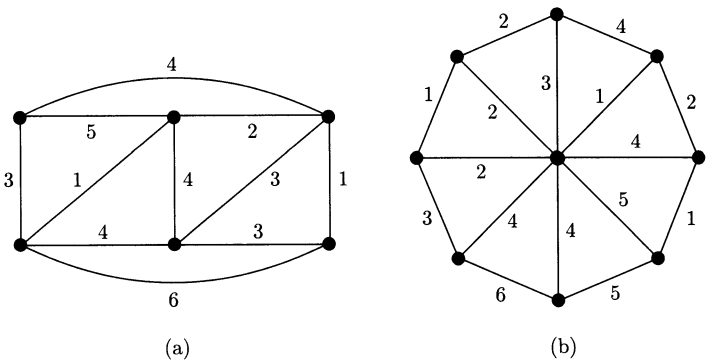
7. Powtórz ćwiczenie 6 dla wag spełniających warunki

$$W(e_1) < W(e_2) < \ldots < W(e_{10}).$$

8. (a) Zastosuj algorytm Kruskala, aby znaleźć minimalne drzewo spinające grafu przedstawionego na rysunku 6.49(a). Oznacz krawędzie w kolejności alfabetycznej, tak jak je wybierałeś. Podaj wagę tego minimalnego drzewa spinającego.
(b) Powtórz ćwiczenie (a) dla algorytmu Prima, zaczynając w środkowym dolnym wierzchołku.

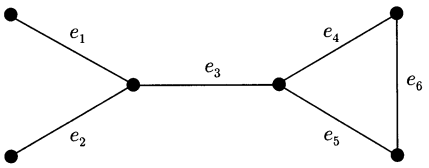
9. (a) Powtórz ćwiczenie 8(a) dla grafu przedstawionego na rysunku 6.49(b).
(b) Powtórz ćwiczenie 8(b) dla grafu przedstawionego na rysunku 6.49(b), zaczynając od górnego wierzchołka.

10. (a) Znajdź wszystkie drzewa spinające grafu przedstawionego na rysunku 6.50.
(b) Które krawędzie należą do każdego drzewa spinającego?



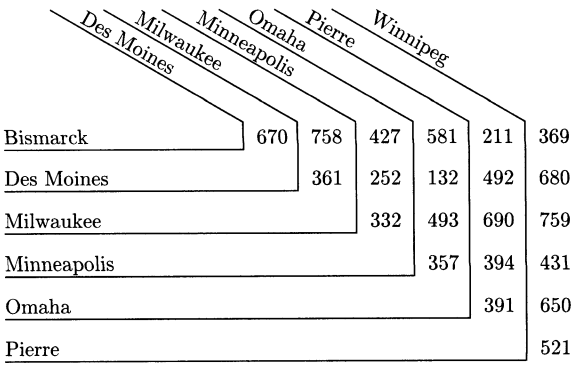
Rysunek 6.49

(c) Dla dowolnego skończonego grafu spójnego scharakteryzuj krawędzie, które należą do każdego drzewa spinającego. Udowodnij swoją tezę.



Rysunek 6.50

11. Na rysunku 6.51 podane są w milach odległości między miastami. Towarzystwo naftowe chce połączyć te miasta rurociągami przechodzącymi bezpośrednio między tymi miastami. Ile co najmniej mil rur do tego potrzeba?



Rysunek 6.51

12. Czy każda krawędź skończonego grafu spójnego bez pętli należy do jakiegoś drzewa spinającego? Uzasadnij swoją odpowiedź.

13. Zmodyfikuj algorytm Prima tak, aby otrzymać minimalny las spinający.

14. (a) Pokaż, że jeśli H jest podgrafem grafu G z wagami, zawierającym minimalne drzewo spinające grafu G, to każde minimalne drzewo spinające grafu H jest minimalnym drzewem spinającym grafu G.

(b) Pokaż, że algorytm polegający na usuwaniu krawędzi opisany po algorytmie Kruskala daje minimalne drzewo spinające. *Wskazówka*: Pokaż, że zbiór krawędzi pozostałych po każdym usunięciu zawiera minimalne drzewo spinające grafu G. *Szkic*: Przypuśćmy, że zbiór E krawędzi pozostałych w chwili, gdy krawędź e ma być właśnie usunięta, zawiera minimalne drzewo spinające grafu G. Zastosuj ćwiczenie (a) do minimalnego drzewa spinającego K utworzonego dla zbioru E przez algorytm Kruskala. Niech C będzie cyklem w E zawierającym krawędź e. Udowodnij, że każda krawędź f w zbiorze $C\backslash(K \cup \{e\})$ poprzedza e na liście krawędzi i tworzy cykl z jakimiś krawędziami w K, które także poprzedzają e. Weźmy drogę w $K\backslash\{e\}$ łączącą końce e i wywnioskujmy, że $e \notin K$.

15. Niech G będzie skończonym grafem spójnym, w którym różne krawędzie mają różne wagi. Pokaż, że graf G ma dokładnie jedno minimalne drzewo spinające. *Wskazówka*: Załóż, że graf G ma więcej niż jedno minimalne drzewo spinające. Weź krawędź o najmniejszej wadze, należącą do pewnych, ale nie do wszystkich minimalnych drzew spinających.

To, co jest najważniejsze w tym rozdziale

Tak jak zwykle, jednym z najlepszych sposobów na to, aby zrobić przegląd materiału, jest postępowanie zgodnie z sugestiami zamieszczonymi na końcu rozdziału 1. Zadaj sobie pytania: Co to znaczy? Dlaczego jest to tutaj? Jak można tego użyć? Zastanów się nad przykładami. Chociaż na poniższej liście znajduje się wiele pozycji, naprawdę nie ma tak wiele nowych pojęć do opanowania, jak wskazywałaby na to długość tej listy. Odwagi!

Pojęcia

droga
  zamknięta, prosta, cykliczna, acykliczna
  droga Eulera, cykl Eulera
  droga Hamiltona, cykl Hamiltona
izomorfizm, niezmiennik izomorfizmu
stopień, ciąg liczb wierzchołków kolejnych stopni
graf

regularny, pełny, dwudzielny, pełny dwudzielny
spójny, składowa
drzewo, liść, las
drzewo spinające, las spinający
drzewo z wyróżnionym korzeniem
korzeń, rodzic, dziecko, potomek, poddrzewo z wyróżnionym korzeniem v
drzewo poszukiwań binarnych, drzewo z etykietami
drzewo binarne (drzewo o m rozgałęzieniach) z wyróżnionym korzeniem
regularne, pełne
numer poziomu, wysokość
uporządkowane drzewo z wyróżnionym korzeniem
waga krawędzi, drogi, podgrafu
minimalne drzewo spinające, minimalny las spinający
kod Graya

Fakty

Droga ma wszystkie wierzchołki różne wtedy i tylko wtedy, gdy jest prosta i acykliczna.

Jeśli istnieje droga między dwoma różnymi wierzchołkami, to istnieje prosta droga acykliczna między nimi.

Istnieje co najwyżej jedna droga prosta między dwoma wierzchołkami w grafie acyklicznym lub w acyklicznym grafie skierowanym. Istnieje dokładnie jedna droga prosta między dwoma wierzchołkami w drzewie.

Jeśli e jest krawędzią w grafie spójnym G, to e należy do pewnego cyklu wtedy i tylko wtedy, gdy graf $G\backslash\{e\}$ jest spójny. Zatem algorytm sprawdzający spójność może sprawdzać, czy są cykle.

Następujące stwierdzenia są równoważne dla grafu G o $n \geq 1$ wierzchołkach, nie mającego pętli:

(a) G jest drzewem.

(b) G jest grafem spójnym, ale nie będzie on spójny po usunięciu krawędzi.

(c) G jest grafem acyklicznym, ale nie będzie on acykliczny po dodaniu krawędzi.

(d) G jest grafem acyklicznym i ma $n-1$ krawędzi (tak dużo, jak tylko jest to możliwe).

(e) G jest grafem spójnym i ma $n-1$ krawędzi (tak mało, jak tylko jest to możliwe).

Wyróżnienie korzenia nadaje drzewu naturalną strukturę grafu skierowanego.

$$\sum_{v \in V(G)} \deg(v) = 2 \cdot |E(G)|.$$

Graf ma cykl Eulera wtedy i tylko wtedy, gdy jest spójny i wszystkie wierzchołki mają stopień parzysty. Drogi Eulera istnieją, jeśli co najwyżej dwa wierzchołki mają stopień nieparzysty.

Jeśli graf nie ma pętli i krawędzi wielokrotnych oraz jeśli $|V(G)| = n \geq 3$, to G jest grafem hamiltonowskim, gdy prawdziwy jest któryś z następujących warunków:

(a) $\deg(v) \geq n/2$ dla każdego wierzchołka v (duże stopnie wierzchołków).

(b) $|E(G)| \geq \frac{1}{2}(n-1)(n-2)+2$ (dużo krawędzi).

(c) $\deg(v)+\deg(w) \geq n$, jeśli v i w nie są połączone krawędzią.

Twierdzenie 4 w § 6.5 podaje informacje na temat dróg Hamiltona w grafach dwudzielnych.

Algorytmy

Algorytm Fleury'ego do konstruowania cyklu Eulera w grafie.

DRZEWO(v), LAS do budowania lasu spinającego lub znajdowania składowych grafu w czasie $O(|V(G)| + |E(G)|)$.

Algorytmy Kruskala i Prima do konstruowania minimalnych drzew spinających (lub lasów) dla grafów z wagami.

7. REKURENCJA, DRZEWA I ALGORYTMY

Jednym z głównych celów tego rozdziału jest zrozumienie, jak działają algorytmy rekurencyjne i w jaki sposób można sprawdzać ich poprawność. Zanim przejdziemy do omawiania algorytmów, przyjrzymy się ogólnemu pojęciu definicji rekurencyjnej, z którym się zetknęliśmy w wersji dla ciągów w rozdziale 4, a także zajmiemy się uogólnieniem indukcji matematycznej. Na algorytmy rekurencyjne można często patrzeć, jak na algorytmy znajdujące drogę w dół jakiegoś drzewa. Aby podkreślić ten punkt widzenia, wybraliśmy przykłady algorytmów, które wręcz badają drzewa lub w naturalny sposób tworzą drzewa. Pomysł uporządkowanego etykietowania grafu skierowanego, tworzonego przez algorytm z § 7.3, będzie odgrywał pewną rolę w rozdziale 8. W ostatnich dwóch paragrafach tego rozdziału pokazujemy zastosowania metod rekurencyjnych do notacji algebraicznej i do kodów prefiksowych, których używa się do kompresji plików oraz do projektowania struktur danych.

§ 7.1. Ogólna postać definicji rekurencyjnych i dowodów indukcyjnych

Za pomocą definicji rekurencyjnych omówionych w § 4.3 można definiować ciągi, podając kilka ich początkowych wyrazów oraz przepis na otrzymywanie późniejszych wyrazów z wcześniejszych. Przy omawianiu algorytmów działających na drzewach wygodniej będzie mieć ogólniejszą wersję definicji rekurencyjnych, którą teraz omówimy. W tym paragrafie są trzy główne tematy:

definicje rekurencyjne zbiorów, uogólnienie indukcji matematycznej i definicje rekurencyjne funkcji. Przekonamy się, że te trzy tematy są ściśle ze sobą związane.

Z grubsza biorąc, zbiór pewnych obiektów jest zdefiniowany rekurencyjnie, jeśli powstał on w wyniku jakiegoś procesu, polegającego na tym, że pewne elementy zostały włożone do tego zbioru na początku, a inne dodane później za przyczyną innych elementów, które już się tam znalazły. Oczywiście taki opis jest zbyt nieprecyzyjny, aby był użyteczny. Za chwilę sprecyzujemy to, ale najpierw popatrzmy na kilka przykładów zbiorów zdefiniowanych rekurencyjnie.

PRZYKŁAD 1 (a) Konstruujemy rekurencyjnie zbiór $\mathbb{N}$ w następujący sposób:

(P) $0 \in \mathbb{N}$;
(R) jeśli $n \in \mathbb{N}$, to $n+1 \in \mathbb{N}$.

Warunek (P) powoduje włożenie liczby 0 do zbioru $\mathbb{N}$, a warunek (R) podaje przepis na generowanie nowych elementów zbioru $\mathbb{N}$ ze starych. Wywołując warunek (P), a następnie wielokrotnie używając warunku (R) otrzymujemy w zbiorze $\mathbb{N}$ elementy: $0, 1, 2, 3 \ldots$.

(b) Za pomocą warunków

(P) $1 \in S$;
(R) jeśli $n \in S$, to $2n \in S$

otrzymujemy elementy podzbioru S zbioru $\mathbb{P}$; w zbiorze S jest 1 (z warunku (P)), 2 (z warunku (R), gdyż $1 \in S$), 4 (gdyż $2 \in S$), 8, 16 i tak dalej. Nietrudno wykazać przez indukcję, że zbiór S zawiera zbiór $\{2^m\colon\ m \in \mathbb{N}\}$. Co więcej, liczby postaci 2^m są jedynymi liczbami, którym warunki (P) i (R) każą należeć do zbioru S (por. ćwiczenie 1). Wydaje się, że rozsądnie jest przyjąć, iż warunki (P) i (R) definiują zbiór $\{2^m\colon\ m \in \mathbb{N}\}$. ∎

PRZYKŁAD 2 Weźmy alfabet Σ. Warunki

(P) $\lambda \in \Sigma^*$;
(R) jeśli $w \in \Sigma^*$ i $x \in \Sigma$, to $wx \in \Sigma^*$

opisują w rekurencyjny sposób zbiór Σ^*. Słowo puste λ należy do zbioru Σ^*, ponieważ tak postanowiliśmy, a wielokrotne stosowanie warunku (R) pozwala nam budować coraz dłuższe słowa. Na przykład, jeśli Σ jest alfabetem angielskim, to słowa λ, b, bi,

big, ..., *bigwor*, *bigword* należą do zbioru opisanego za pomocą warunków (P) i (R). ■

Definicja rekurencyjna zbioru S składa się z dwóch części: **warunku początkowego** postaci

(P) $X \subseteq S$,

gdzie X jest pewnym konkretnym zbiorem, oraz **warunku rekurencyjnego**

(R) jeśli element s powstaje z elementów zbioru S w wyniku zastosowania pewnych reguł, to $s \in S$.

Konkretne reguły w warunku (R) będą oczywiście podane wyraźnie, jak w przykładach 1 oraz 2 i mogą one być dowolnymi regułami mającymi sens. Tak jak w przykładzie 2, mogą one korzystać z obiektów, które same nie należą do zbioru S.

Rozumiemy zawsze, że element należy do zbioru S tylko wtedy, jeśli wymuszają to warunki (P) i (R). Zatem w przykładzie 1(b) tylko potęgi liczby 2 muszą należeć do zbioru S, a więc zbiór S składa się tylko z tych potęg.

Warunki (P) i (R) pozwalają nam budować zbiór S ze zbioru X warstwami. Definiujemy ciąg zbiorów $S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots$ w następujący sposób:

$S_0 = X$ oraz

$S_{n+1} =$ zbiór wszystkich elementów zbioru S_n lub skonstruowanych z elementów zbioru S_n za pomocą reguł podanych w warunku (R).

Zbiór S_n składa się wtedy z tych elementów zbioru S, które mogą być skonstruowane z elementów zbioru X za pomocą n lub mniej zastosowań reguł podanych w warunku (R) oraz $S = \bigcup_{n=0}^{\infty} S_n$.

PRZYKŁAD 3 (a) Dla zbioru $S = \mathbb{N}$ z przykładu 1(a) mamy $S_0 = \{0\}$, $S_1 = \{0, 1\}$, ..., $S_n = \{0, 1, \ldots, n\}$, ...

(b) Dla zbioru $S = \{2^m\colon\ m \in \mathbb{N}\}$ określonego w przykładzie 1(b) mamy $S_0 = \{1\}$, $S_1 = \{1, 2\}$, ..., $S_n = \{1, 2, 4, \ldots, 2^n\}$.

(c) Dla zbioru $\Sigma = \{a, b\}$ definicja zbioru Σ^* w przykładzie 2 prowadzi do ciągu zbiorów $\Sigma_0^* = \{\lambda\}$, $\Sigma_1^* = \{\lambda, a, b\}$, $\Sigma_2^* = \{\lambda, a, b, aa, ab, ba, bb\}$ itd., więc korzystając z dotychczasowych oznaczeń mamy $\Sigma_n^* = \{\lambda\} \cup \Sigma \cup \Sigma^2 \cup \ldots \cup \Sigma^n$.

(d) Dla zbioru $\Sigma = \{a, b\}$ definiujemy zbiór S w następujący sposób:

(P) $\Sigma \subseteq S$;
(R) jeśli $w \in S$, to $awb \in S$.

Wtedy otrzymujemy $S_0 = \Sigma = \{a, b\}$, $S_1 = \{a, b, aab, abb\}$, $S_2 = S_1 \cup \{aaabb, aabbb\}$ itd., a sam zbiór S składa się ze słów postaci $a \ldots ab \ldots b$, w których jest o jedną literę a więcej lub o jedną mniej niż liter b. ∎

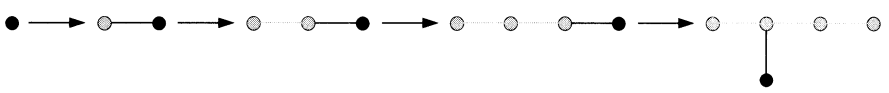
PRZYKŁAD 4 Możemy zdefiniować rekurencyjnie zbiór drzew skończonych. Dla wygody, powiemy, że **graf G' powstaje z grafu G przez dołączenie liścia** v, jeśli

(a) $V(G') = V(G) \cup \{v\} \neq V(G)$ oraz
(b) $E(G') = E(G) \cup \{e\}$, gdzie krawędź e łączy wierzchołek v z pewnym wierzchołkiem grafu G.

Wtedy zbiór drzew jest definiowany w następujący sposób:

(P) Każdy graf mający jeden wierzchołek i nie mający krawędzi jest drzewem (trywialnym);
(R) Jeśli T jest drzewem i graf T' powstaje z drzewa T przez dołączenie liścia, to T' jest drzewem.

Możemy traktować tę definicję rekurencyjną jako sposób budowania drzew poprzez dodawanie liści. Rysunek 7.1 pokazuje typowy ciąg konstrukcji.



Rysunek 7.1

Wcześniej definiowaliśmy drzewa jako spójne grafy acykliczne. Aby pokazać, że ta nowa definicja rekurencyjna jest zgodna ze starą definicją, musimy sprawdzić, że

(1) Za pomocą warunków (P) i (R) tworzy się tylko spójne grafy acykliczne.
(2) Każdy skończony spójny graf acykliczny może być zbudowany przy użyciu tylko warunków (P) i (R).

Dość łatwo można się przekonać, że warunek (1) jest prawdziwy, ponieważ grafy wymienione w (P) są drzewami oraz jeśli graf T jest spójny i acykliczny, to również każdy graf T' otrzymany z grafu T za pomocą warunku (R) jest spójny i acykliczny.

W efekcie, dowodzimy przez indukcję, że jeśli zbiór S_n składa się z drzew, to również zbiór S_{n+1} składa się z drzew.

Nieco trudniej jest udowodnić warunek (2). Wyobraźmy sobie, że istnieją (skończone) drzewa, których nie możemy zbudować za pomocą warunków (P) i (R) oraz przypuśćmy, że T jest takim drzewem mającym możliwie mało wierzchołków. Na podstawie warunku (P) drzewo T ma więcej niż jeden wierzchołek. Wtedy drzewo T ma co najmniej dwa liście na podstawie lematu 1 z § 6.3. Obetnijmy jeden liść z drzewa T, aby otrzymać nowe drzewo T''. Wtedy drzewo T'' daje się skonstruować za pomocą warunków (P) i (R), co wynika z minimalności drzewa T. Ale drzewo T daje się otrzymać z drzewa T'' za pomocą warunku (R); dołączamy po prostu ten liść. Zatem samo drzewo T daje się skonstruować, co jest sprzeczne z tym, w jaki sposób zostało ono wybrane. ∎

PRZYKŁAD 5 (a) Możemy powtórzyć definicję rekurencyjną z przykładu 4, aby otrzymać zbiór (skończonych) drzew z wyróżnionym korzeniem.

(P) Graf mający jeden wierzchołek v i nie mający krawędzi jest (trywialnym) drzewem z wyróżnionym korzeniem v.

(R) Jeśli T jest drzewem z wyróżnionym korzeniem r i graf T' powstaje przez dołączenie liścia do drzewa T, to T' jest drzewem z wyróżnionym korzeniem r.

Tak jak w przykładzie 4, widzimy, że ta definicja nie daje nic oprócz drzew z wyróżnionym korzeniem. Drugie rozumowanie z przykładu 4 pokazuje, że każde drzewo z wyróżnionym korzeniem daje się skonstruować za pomocą warunków (P) i (R); ponieważ każde nietrywialne drzewo ma co najmniej dwa liście, więc możemy odciąć liść, który nie jest korzeniem.

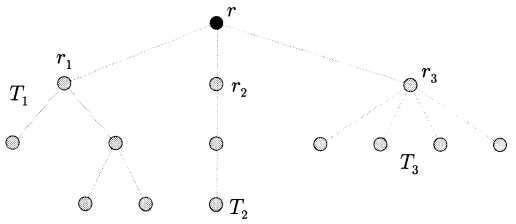
(b) A oto inny sposób zdefiniowania rekurencyjnie zbioru drzew z wyróżnionym korzeniem. Definiujemy zbiór $\mathcal{R}$ par uporządkowanych (T, r) takich, że T jest drzewem i r jest wierzchołkiem drzewa T; wierzchołek r nazywamy **korzeniem drzewa** T. Będziemy mówić, że pary (T_1, r_1) i (T_2, r_2) są **rozłączne**, jeśli drzewa T_1 i T_2 nie mają wspólnych wierzchołków. Jeśli pary $(T_1, r_1), \ldots, (T_k, r_k)$ są rozłączne, to będziemy mówić, że drzewo T powstaje przez **przyczepienie par** $(T_1, r_1), \ldots, (T_k, r_k)$ **do korzenia** r, jeśli:

(1) r nie jest wierzchołkiem żadnego drzewa T_i,

(2) $V(T) = V(T_1) \cup \ldots \cup V(T_k) \cup \{r\}$ oraz

(3) $E(T) = E(T_1) \cup \ldots \cup E(T_k) \cup \{e_1, \ldots, e_k\}$, gdzie krawędź e_i łączy wierzchołki r i r_i.

Na rysunku 7.2 widzimy drzewo otrzymane przez przyczepienie par (T_1, r_1), (T_2, r_2) i (T_3, r_3) do korzenia r.



Rysunek 7.2

A oto definicja $\mathcal{R}$:

(P) Jeśli T jest grafem mającym jeden wierzchołek v i nie mającym krawędzi, to $(T, v) \in \mathcal{R}$;

(R) Jeśli (T_1, r_1), ..., (T_k, r_k) są rozłącznymi elementami zbioru $\mathcal{R}$ i jeśli (T, r) powstaje przez przyczepienie par (T_1, r_1), ..., (T_k, r_k) do korzenia r, to $(T, r) \in \mathcal{R}$.

Tak jak w przykładzie (a) bez trudu pokazujemy, że zbiór $\mathcal{R}$ składa się z par (T, r) takich, że T jest drzewem mającym wierzchołek r, to znaczy drzewem z wyróżnionym korzeniem r. Aby pokazać, że każde drzewo z wyróżnionym korzeniem może być skonstruowane za pomocą warunków (P) i (R), wyobraźmy sobie, że T jest drzewem z wyróżnionym korzeniem, którego nie można tak skonstruować i takim, że liczba $|V(T)|$ jest jak najmniejsza. Ponieważ drzewo T nie spełnia warunku (P), więc ma więcej niż jeden wierzchołek. Rozpatrzmy dzieci korzenia r w drzewie T. Poddrzewa, których korzeniami są te dzieci, są mniejsze od drzewa T, a więc mogą być skonstruowane za pomocą warunków (P) i (R). Ponieważ drzewo T powstaje przez przyczepienie tych poddrzew do korzenia r, więc w efekcie para (T, r) należy do zbioru $\mathcal{R}$, co przeczy założeniu. ∎

W naszych dotychczasowych przykładach wiedzieliśmy, jakie zbiory próbowaliśmy zdefiniować i definicje rekurencyjne były zaprojektowane zgodnie z tym celem. Można też podać definicje rekurencyjne, które są całkowicie poprawne i o których nie wiemy na pewno, co definiują.

PRZYKŁAD 6 (a) Definiujemy zbiór A liczb całkowitych w następujący sposób:

(P) $1 \in A$;

(R) jeśli $n \in A$, to $3n \in A$ i jeśli $2n+1 \in A$, to $n \in A$.

Te warunki opisują zbiór A w sposób jednoznaczny, ale nie jest całkiem jasne, jakie liczby całkowite należą do zbioru A. Chwila eksperymentowania pokazuje, że zbiór A nie zawiera liczb całkowitych n takich, że $n \equiv 2 \pmod 3$ i tak jest w rzeczywistości (ćwiczenie 9). Dalsze eksperymenty sugerują nam, że być może $A = \{n \in \mathbb{P}\colon\ n \not\equiv 2 \pmod 3\}$, tzn. $A = \{1, 3, 4, 6, 7, 9, 10, 12, 13, \ldots\}$. Nie jest jasne, w jaki sposób moglibyśmy udowodnić, że to przypuszczenie jest prawdziwe. Ciąg liczb 1, 3, 9, 27, 13, 39, 19, 57, 171, 85, 255, 127, 63, 31, 15, 7, będący najkrótszą drogą do pokazania, że liczba 7 należy do zbioru A, nie daje nam najmniejszej wskazówki, w jaki sposób przeprowadzić rozumowanie ogólne. Wiele godzin badań nie pozwoliło całkowicie rozstrzygnąć tej kwestii. Nie wiemy, czy nasze przypuszczenie jest prawdziwe, choć być może specjaliści od teorii funkcji rekurencyjnych mogliby podać odpowiedź.

(b) Rozważmy zbiór S określony w następujący sposób:

(P) $1 \in S$;

(R) jeśli $n \in S$, to $2n \in S$ i jeśli $3n+1 \in S$, przy czym liczba n jest nieparzysta, to $n \in S$.

W chwili, gdy to piszemy, *nikt* nie wie, czy $S = \mathbb{P}$, choć wiele osób próbowało rozwiązać ten problem. ■

Ze zbiorami zdefiniowanymi rekurencyjnie wiąże się pewien rodzaj uogólnionej indukcji. Przypuśćmy, że zbiór S został zdefiniowany rekurencyjnie ze zbioru X, danego w warunku (P), za pomocą zbioru reguł tworzenia nowych elementów, opisanych w warunku (R). Przypuśćmy także, że z każdym elementem s zbioru S związane jest zdanie $p(s)$.

Uogólniona zasada indukcji

Przy powyższych założeniach, jeśli

(p) dla każdego $s \in X$ prawdziwe jest zdanie $p(s)$ oraz

(r) zdanie $p(s)$ jest prawdziwe dla każdego elementu s, utworzonego z takich elementów t zbioru S, dla których zdanie $p(t)$ jest prawdziwe,

wtedy zdanie $p(s)$ jest prawdziwe dla każdego $s \in S$.

Dowód. Niech $T = \{s \in S$: zdanie $p(s)$ jest prawdziwe$\}$ i przypomnijmy ciąg zbiorów $X = S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots$, zdefiniowany już wcześniej. Z warunku (p) wynika, że $X = S_0 \subseteq T$. Jeśli $S \not\subseteq T$, to istnieje najmniejsza liczba $n \in \mathbb{P}$, dla której $S_n \not\subseteq T$ i istnieje pewien element $s \in S_n \backslash T$. Ponieważ $S_{n-1} \subseteq T$ oraz $s \notin S_{n-1}$, więc element s musi być skonstruowany z elementów zbioru S_{n-1} za pomocą reguł podanych w warunku (R). Ale warunek (r) mówi, że wtedy zdanie $p(s)$ jest prawdziwe, co jest sprzeczne z wyborem elementu s. Zatem $S \subseteq T$, czego należało dowieść. ∎

PRZYKŁAD 7 (a) Jeśli S jest zbiorem liczb naturalnych $\mathbb{N}$ zdefiniowanym w następujący sposób:

(P) $0 \in S$;

(R) jeśli $n \in S$, to $n + 1 \in S$,

to warunki (p) i (r) będą wyglądały następująco:

(p) zdanie $p(0)$ jest prawdziwe;

(r) jeśli zdanie $p(n)$ jest prawdziwe, to zdanie $p(n + 1)$ jest prawdziwe,

a więc uogólniona zasada indukcji jest w tym przypadku po prostu zwykłą zasadą indukcji matematycznej.

(b) Dla zbioru $\mathcal{R}$ skończonych drzew z wyróżnionym korzeniem, zdefiniowanego tak jak w przykładzie 5(b), otrzymujemy następującą zasadę indukcji:

Przypuśćmy, że

(p) zdanie $p(T)$ jest prawdziwe dla każdego drzewa z wyróżnionym korzeniem, mającego jeden wierzchołek oraz

(r) zdanie $p(T)$ jest prawdziwe wtedy, gdy zdanie p jest prawdziwe dla każdego poddrzewa przyczepionego do korzenia drzewa T.

Wtedy zdanie $p(T)$ jest prawdziwe dla każdego drzewa T z wyróżnionym korzeniem.

(c) Definicja rekurencyjna zbioru drzew z przykładu 4 daje następującą zasadę indukcji:

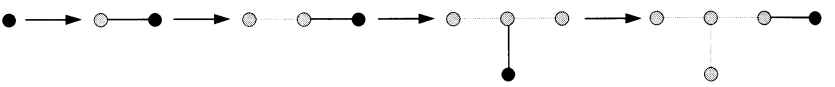
Przypuśćmy, że zdanie $p(T)$ jest prawdziwe dla każdego grafu T mającego jeden wierzchołek i nie mającego krawędzi oraz przypuśćmy, że zdanie $p(T')$ jest prawdziwe wtedy, gdy zdanie $p(T)$ jest prawdziwe i drzewo T' powstało z drzewa T przez dołączenie liścia. Wtedy zdanie $p(T)$ jest prawdziwe dla każdego drzewa T.

(d) Zastosujemy zasadę indukcji sformułowaną w przykładzie (c) do nowego dowodu tego, że drzewo o n wierzchołkach ma $n-1$ krawędzi; por. lemat 2 w § 6.3. Niech zdaniem $p(T)$ będzie „$|E(T)| = |V(T)| - 1$". Jeśli drzewo T ma jeden wierzchołek i nie ma krawędzi, to zdanie $p(T)$ jest prawdziwe. Jeśli zdanie $p(T)$ jest prawdziwe dla pewnego drzewa T i dołączamy liść do drzewa T, to zwiększamy obie strony równości o 1, a więc zdanie $p(T')$ jest nadal prawdziwe. Z uogólnionej zasady indukcji wynika, że zdanie $p(T)$ jest prawdziwe dla każdego drzewa T. ∎

Ogólnie, jeśli zbiór S jest zdefiniowany rekurencyjnie, to może się zdarzyć, że element zbioru S_n daje się skonstruować z elementów zbioru S_{n-1} na więcej niż jeden sposób. Często lepiej jest używać definicji rekurencyjnych, w których nowe elementy zbioru S otrzymuje się z innych elementów w tylko jeden sposób. Takie definicje rekurencyjne będziemy nazywać **definicjami jednoznacznymi**.

PRZYKŁAD 8 Większość podanych dotąd definicji rekurencyjnych to definicje jednoznaczne. Teraz omówimy wyjątki.

(a) Definicja rekurencyjna zbioru drzew skończonych w przykładzie 4 nie jest definicją jednoznaczną. W rzeczywistości drzewo, którego konstrukcja jest pokazana na rysunku 7.1, może być również skonstruowane tak, jak pokazane jest na rysunku 7.3.



Rysunek 7.3

(b) Pierwsza definicja rekurencyjna zbioru skończonych drzew z wyróżnionym korzeniem w przykładzie 5 również nie jest definicją jednoznaczną (zob. ćwiczenie 7). Z drugiej strony, definicja z przykładu 5(b), korzystająca z pojęcia przyczepiania do korzenia, jest definicją jednoznaczną. Drzewo z wyróżnionym korzeniem jest wyznaczone jednoznacznie przez poddrzewa przyczepione do korzenia, każde z tych poddrzew jest określone jednoznacznie w ten sam sposób i tak dalej.

(c) Definicje w przykładzie 6 nie są definicjami jednoznacznymi. Na przykład 1 należy do zbioru A, co wynika z warunku (P), ale także dwukrotne zastosowanie warunku (R) daje kolejno $3 \in A$ oraz $1 \in A$. Podobnie, 1 pojawia się na nowo w zbiorze S w przykładzie 6(b) poprzez ciąg 1, 2, 4, 1. ∎

Jeśli zbiór S jest zdefiniowany rekurencyjnie za pomocą warunków (P) i (R), możemy definiować rekurencyjnie funkcje f określone na zbiorze S, używając następującej procedury:

(1) Definiujemy w wyraźny sposób wartość $f(s)$ dla argumentów s należących do zbioru X, opisanego warunkiem (P);
(2) Podajemy przepis na zdefiniowanie wartości $f(s)$ za pomocą wartości $f(t)$ dla argumentów $t \in S$, otrzymanych wcześniej niż s.

Aby powiedzieć, co mamy na myśli pisząc „otrzymanych wcześniej", musimy przypomnieć nasz ciąg zbiorów występujących przy tworzeniu zbioru S, $X = S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots$. Dla każdego elementu $s \in S$ istnieje najmniejsza liczba n, dla której $s \in S_n$. Wszystkie elementy zbioru S_{n-1} traktujemy jako otrzymane wcześniej niż taki element s, chociaż oczywiście tylko niewiele z nich jest użytych do utworzenia elementu s za pomocą reguł opisanych w warunku (R). Jeśli definicja rekurencyjna zbioru S jest jednoznaczna, to przepis podany w kroku (2) będzie jednoznaczny. W przeciwnym przypadku musielibyśmy się upewnić, że definicja jest poprawna. To znaczy, musielibyśmy sprawdzić, że wartość $f(s)$ nie zależy od sposobu, w jaki element s został skonstruowany z elementów t, otrzymanych wcześniej niż s.

PRZYKŁAD 9 (a) Niech zbiór $\mathbb{N}$ będzie określony w następujący sposób:

(P) $0 \in \mathbb{N}$;
(R) jeśli $n \in \mathbb{N}$, to $n+1 \in \mathbb{N}$.

Jak widzieliśmy wcześniej, $S_m = \{0, \ldots, m\}$, a więc liczba n pojawia się po raz pierwszy w zbiorze S_n, a liczbami otrzymanymi wcześniej niż n są $0, \ldots, n-1$. Aby zdefiniować funkcję f na zbiorze $\mathbb{N}$, możemy określić wartość $f(0)$, a następnie wyjaśnić, jak otrzymać wartość $f(n)$ z wartości $f(0), \ldots, f(n-1)$.

(b) Definiujemy zbiór $\mathbb{N}$ rekurencyjnie w następujący sposób:

(P) $\{0,1\} \subseteq \mathbb{N}$;
(R) $f(0) = 1$, $f(1) = 1$ oraz $f(n+1) = f(n) + f(n-1)$ dla $n \geq 1$. Wtedy funkcja f opisuje ciąg Fibonacciego. Dla tej szczególnej definicji rekurencyjnej zbioru $\mathbb{N}$ mamy: $\{0,1\} = S_0 \subseteq \{0,1,2\} = S_1 \subseteq \{0,1,2,3\} = S_3 \subseteq \ldots$, a więc jeśli tylko $n+1 \in S_m$, to $n, n-1 \in S_{m-1}$. ∎

PRZYKŁAD 10 (a) Zbiór Σ^* był zdefiniowany rekurencyjnie w przykładzie 2. Definiujemy funkcję długości l na zbiorze Σ^* w następujący sposób:

(1) $l(\lambda) = 0$;
(2) jeśli $w \in \Sigma^*$ i $x \in \Sigma$, to $l(wx) = l(w) + 1$.

Wtedy $l(w)$ jest liczbą liter w słowie w.

(b) Możemy także zdefiniować zbiór Σ^* w następujący sposób:

(P) $\{\lambda\} \cup \Sigma \subseteq \Sigma^*$;
(R) jeśli $w, u \in \Sigma^*$, to $wu \in \Sigma^*$.

Wtedy długość mogłaby być zdefiniowana w następujący sposób:

(1) $l'(\lambda) = 0$, $l'(x) = 1$ dla $x \in \Sigma$;
(2) jeśli $w, u \in \Sigma^*$, to $l'(wu) = l'(w) + l'(u)$.

Czy rzeczywiście mogłaby być tak zdefiniowana? Potencjalne trudności mogą wystąpić dlatego, że ta nowa definicja rekurencyjna zbioru Σ^* nie jest definicją jednoznaczną. Jest możliwe skonstruowanie danego słowa wu na więcej niż jeden sposób. Na przykład, jeśli $\Sigma = \{a, b\}$, to ciąg a, ab, $(ab)a$, $((ab)a)b$ jest jedną drogą prowadzącą do słowa $abab$, ale ciąg a, ab, $(ab)(ab)$ jest inną drogą, a ciąg a, b, ba, $(ba)b$, $a((ba)b)$ jest jeszcze inną. Skąd wiemy, że nasza „definicja" funkcji l' daje tę samą wartość $l'(abab)$ dla wszystkich tych dróg?

Okazuje się (ćwiczenie 13), że funkcja l' jest zdefiniowana poprawnie i tak naprawdę jest to funkcja l, którą widzieliśmy w przykładzie (a) powyżej.

(c) Spróbujmy zdefiniować funkcję **głębokości** d na zbiorze Σ^* w następujący sposób:

(1) $d(\lambda) = 0$, $d(x) = 1$ dla $x \in \Sigma$;
(2) jeśli $w, u \in \Sigma^*$, to $d(wu) = \max\{d(w), d(u)\} + 1$.

Tym razem nie udało się. Gdyby funkcja d była zdefiniowana poprawnie, mielibyśmy: $d(a) = 1$, $d(ab) = 1 + 1 = 2$, $d((ab)a) = 2 + 1 = 3$ oraz $d(((ab)a)b) = 3 + 1 = 4$, ale również $d((ab)(ab)) = 2 + 1 = 3$. Tak więc wartość $d(abab)$ nie byłaby określona jednoznacznie. ∎

Przykłady 10(b) i 10(c) wskazują na możliwe trudności w definiowaniu funkcji w sposób rekurencyjny, kiedy definicja rekurencyjna nie jest jednoznaczna. Wolimy jednoznaczne definicje rekurencyjne, kiedy definiujemy rekurencyjnie funkcje. Zatem wolimy definicję zbioru Σ^*, której używaliśmy w przykładzie 10(a) od definicji z przykładu 10(b). Podobnie definicja drzew z wyróżnionym korzeniem w przykładzie 5(b) ma przewagę nad definicją

z przykładu 5(a). Czasami nie mamy wyboru: nie znamy jednoznacznej definicji rekurencyjnej drzew skończonych, ale mimo to definicja z przykładu 4 jest przydatna.

PRZYKŁAD 11 Chcemy zdefiniować rekurencyjnie wysokość skończonego drzewa z wyróżnionym korzeniem tak, by wysokość była równa długości najdłuższej drogi prostej od korzenia do liścia. Do tego celu nasza definicja rekurencyjna drzew z wyróżnionym korzeniem z przykładu 5(a) się nie nadaje, ponieważ dodanie liścia nie musi zwiększyć wysokości drzewa. Natomiast definicja rekurencyjna z przykładu 5(b) nadaje się idealnie do tego celu.

(a) Przypomnijmy, że zbiór $\mathcal{R}$ skończonych drzew z wyróżnionym korzeniem jest zdefiniowany w następujący sposób:

(P) Jeśli T jest grafem mającym jeden wierzchołek v i nie mającym krawędzi, to $(T, v) \in \mathcal{R}$;

(R) Jeśli $(T_1, r_1), \ldots, (T_k, r_k)$ są rozłącznymi elementami zbioru $\mathcal{R}$ i jeśli (T, r) powstaje przez przyczepienie drzew $(T_1, r_1), \ldots, (T_k, r_k)$ do korzenia r, to $(T, r) \in \mathcal{R}$.

Wysokość elementu zbioru $\mathcal{R}$ jest zdefiniowana w następujący sposób:

(1) trywialne jednowierzchołkowe drzewa z wyróżnionym korzeniem mają wysokość 0;

(2) jeśli drzewo (T, r) jest zdefiniowane tak jak w warunku (R) i jeśli drzewa $(T_1, r_1), \ldots, (T_k, r_k)$ mają, odpowiednio, wysokości $h_1, \ldots, h_k$, to wysokość drzewa (T, r) jest równa $1 + \max\{h_1, \ldots, h_k\}$.

(b) Inne pojęcia związane ze skończonymi drzewami z wyróżnionym korzeniem też mogą być definiowane rekurencyjnie. Na przykład, jeśli m jest liczbą całkowitą większą niż 1, to zbiór **drzew z m rozgałęzieniami** jest zdefiniowany w następujący sposób:

(P) trywialne jednowierzchołkowe drzewo z wyróżnionym korzeniem jest drzewem z m rozgałęzieniami;

(R) drzewo z wyróżnionym korzeniem otrzymane przez przyczepienie co najwyżej m drzew z m rozgałęzieniami do korzenia, jest drzewem z m rozgałęzieniami.

(c) Wykorzystamy te definicje rekurencyjne do dowodu, że drzewo z m rozgałęzieniami, wysokości h, ma co najwyżej m^h liści. Jest to oczywiste dla drzew trywialnych, gdyż $m^0 = 1$. Weźmy drzewo z m rozgałęzieniami, (T, r), zdefiniowane tak jak w przykładzie (b), przy czym każde poddrzewo (T_i, r_i) ma wysokość h_i.

Ponieważ (T, r) jest drzewem z m rozgałęzieniami, więc każde (T_i, r_i) jest też drzewem z m rozgałęzieniami, a zatem ma co najwyżej m^{h_i} liści. Niech $h^* = \max\{h_1, \ldots, h_k\}$. Wtedy liczba liści drzewa (T, r) jest ograniczona przez

$$m^{h_1} + \ldots + m^{h_k} \leq m^{h^*} + \ldots + m^{h^*} = k \cdot m^{h^*}.$$

Ponieważ do korzenia jest przyczepionych co najwyżej m poddrzew, więc $k \leq m$, a zatem drzewo (T, r) ma co najwyżej $m \cdot m^{h^*} = m^{h^*+1}$ liści. Ponieważ liczba $h^* + 1$ jest równa wysokości h drzewa (T, r), więc teza została udowodniona. ∎

ĆWICZENIA DO § 7.1

1. Niech S będzie zbiorem zdefiniowanym rekurencyjnie w przykładzie 1(b).
(a) Wykaż, że $2^m \in S$ dla wszystkich $m \in \mathbb{N}$.
(b) Wykaż, że jeśli $n \in S$, to liczba n jest postaci 2^m dla pewnej liczby $m \in \mathbb{N}$.
(c) Wyprowadź stąd wniosek, że $S = \{2^m\colon\ m \in \mathbb{N}\}$.

2. Skorzystaj z definicji w przykładzie 2, by pokazać, że następujące ciągi należą do Σ^*, gdzie Σ jest zwykłym alfabetem angielskim:
(a) *cat*,
(b) *math*,
(c) *zzpq*,
(d) *aint*.

3. (a) Opisz podzbiór S zbioru $\mathbb{N} \times \mathbb{N}$ zdefiniowany rekurencyjnie w następujący sposób:
(P) $(0, 0) \in S$;
(R) jeśli $(m, n) \in S$, gdzie $m < n$, to $(m+1, n) \in S$, i jeśli $(m, m) \in S$, to $(0, m + 1) \in S$.
(b) Skorzystaj z tej definicji rekurencyjnej, aby pokazać, że $(1, 2) \in S$.
(c) Czy ta definicja rekurencyjna jest jednoznaczna?

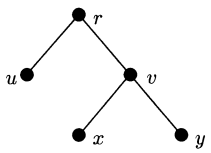
4. (a) Opisz podzbiór T zbioru $\mathbb{N} \times \mathbb{N}$ zdefiniowany rekurencyjnie w następujący sposób:
(P) $(0, 0) \in T$;
(R) jeśli $(m, n) \in T$, to $(m, n + 1) \in T$ oraz $(m + 1, n + 1) \in T$.
(b) Skorzystaj z tej definicji rekurencyjnej, aby pokazać, że $(3, 5) \in T$.
(c) Czy ta definicja rekurencyjna jest jednoznaczna?

5. Niech $\Sigma = \{a, b\}$ oraz niech S będzie zbiorem słów w Σ^*, w których wszystkie a poprzedzają wszystkie b. Na przykład słowa *aab*, *abbb*, *a*, *b*, a nawet λ należą do S, a słowa *bab* i *ba* nie należą.
(a) Podaj definicję rekurencyjną zbioru S.

(b) Skorzystaj z tej definicji rekurencyjnej, aby pokazać, że słowo $abbb \in S$.
(c) Skorzystaj z tej definicji rekurencyjnej, aby pokazać, że słowo $aab \in S$.
(d) Czy twoja definicja rekurencyjna jest jednoznaczna?

6. Niech $\Sigma = \{a, b\}$ oraz niech T będzie zbiorem słów w Σ^*, w których jest dokładnie jedno a.
(a) Podaj definicję rekurencyjną zbioru T.
(b) Skorzystaj z tej definicji rekurencyjnej, aby pokazać, że słowo $bbab \in T$.
(c) Czy twoja definicja rekurencyjna jest jednoznaczna?

7. (a) Opisz dwa różne sposoby konstruowania drzewa z wyróżnionym korzeniem przedstawionego na rysunku 7.4, używając definicji rekurencyjnej z przykładu 5(a).
(b) Opisz konstrukcję drzewa z wyróżnionym korzeniem przedstawionego na rysunku 7.4, używając definicji z przykładu 5(b).



Rysunek 7.4

8. Sprawdź punkt (2) z przykładu 4, pokazując, że S_n zawiera wszystkie drzewa mające co najwyżej $n + 1$ wierzchołków.

9. Niech A będzie zdefiniowanym rekurencyjnie zbiorem z przykładu 6(a). Pokaż, że jeśli $n \in A$, to $n \equiv 0 \pmod 3$ lub $n \equiv 1 \pmod 3$.

10. Pokaż, że liczby 4, 6, 10 i 12 należą do zbioru A określonego w przykładzie 6(a).

10. Pokaż, że liczby 4, 6, 10 i 12 należą do zbioru A określonego w przykładzie 6(a).

11. (a) Pokaż, że zbiór S zdefiniowany w przykładzie 6(b) zawiera liczby 1, 2, 3, 4, 5 oraz 6.
(b) Pokaż, że $7 \in S$.

12. (a) Podaj definicję rekurencyjną zbioru drzew regularnych o m rozgałęzieniach.
(b) Zrób to samo dla drzew pełnych o m rozgałęzieniach.

13. Zdefiniujmy Σ^* tak jak w przykładzie 10(b) oraz niech funkcja l' będzie taka jak w tym przykładzie. Pokaż, że $l'(w)$ jest liczbą liter w słowie w dla każdego $w \in \Sigma^*$. *Wskazówka*: zastosuj uogólnioną zasadę indukcji dla $X = \{\lambda\} \cup \Sigma$.

14. Niech Σ^* będzie zbiorem zdefiniowanym w przykładzie 2. Definiujemy rekurencyjnie **odwrócenie** $\overleftarrow{w}$ słowa $w \in \Sigma^*$ w następujący sposób:
(P) $\overleftarrow{\lambda} = \lambda$;

(R) jeśli $\overleftarrow{w}$ zostało zdefiniowane oraz $x \in \Sigma$, to $\overleftarrow{wx} = x\overleftarrow{w}$.

Jest to inna dobrze określona definicja.

(a) Udowodnij, że $x = \overleftarrow{x}$ dla każdego $x \in \Sigma$.

(b) Skorzystaj z tej definicji, aby znaleźć odwrócenie słowa *cab*.

(c) Skorzystaj z tej definicji, aby znaleźć odwrócenie słowa *abbaa*.

(d) Jeśli w_1 i w_2 są słowami w Σ^*, to jak wyrazić $\overleftarrow{w_1 w_2}$ za pomocą $\overleftarrow{w_1}$ i $\overleftarrow{w_2}$? Co to jest $\overleftarrow{\overleftarrow{w_1}}$?

15. Oto definicja rekurencyjna podzbioru S zbioru $\mathbb{N} \times \mathbb{N}$:

(P) $(0,0) \in S$;

(R) jeśli $(m,n) \in S$, to $(m+2, n+3) \in S$.

(a) Wypisz cztery elementy zbioru S.

(b) Udowodnij, że jeśli $(m,n) \in S$, to $m+n$ dzieli się przez 5.

(c) Czy zdanie odwrotne do zdania z ćwiczenia (b) jest prawdziwe?

16. Oto definicja rekurencyjna innego podzbioru T zbioru $\mathbb{N} \times \mathbb{N}$:

(P) $(0,0) \in T$;

(R) jeśli $(m,n) \in T$, to każda z par $(m+1,n)$, $(m+1,n+1)$ oraz $(m+1,n+2)$ należy do zbioru T.

(a) Wypisz sześć elementów zbioru T.

(b) Udowodnij, że $2m \geq n$ dla wszystkich $(m,n) \in T$.

(c) Czy ta definicja rekurencyjna jest wyznaczona w sposób jednoznaczny?

17. Weźmy następującą definicję rekurencyjną podzbioru A zbioru $\mathbb{N} \times \mathbb{N}$:

(P) $(0,0) \in A$;

(R) jeśli $(m,n) \in A$, to $(m+1,n)$ oraz $(m,n+1)$ należą do zbioru A.

(a) Pokaż, że $A = \mathbb{N} \times \mathbb{N}$.

(b) Niech $p(m,n)$ będzie funkcją określoną w zbiorze $\mathbb{N} \times \mathbb{N}$, której wartościami są zdania. Wykorzystaj ćwiczenie (a), aby zaprojektować ogólną procedurę rekurencyjną pozwalającą dowieść, że zdanie $p(m,n)$ jest prawdziwe dla wszystkich m i n.

18. Niech $\Sigma = \{a,b\}$ i niech B będzie podzbiorem zbioru Σ^* określonym rekurencyjnie w następujący sposób:

(P) a i b należą do zbioru B;

(R) jeśli $w \in B$, to abw i baw należą do zbioru B.

(a) Wypisz sześć elementów zbioru B.

(b) Udowodnij, że jeśli $w \in B$, to długość(w) jest liczbą nieparzystą.

(c) Czy zdanie odwrotne do zdania z ćwiczenia (b) jest prawdziwe?

(d) Czy ta definicja rekurencyjna jest jednoznaczna?

19. Niech Σ będzie alfabetem skończonym. Dla słów $w \in \Sigma^*$, odwrócenie $\overleftarrow{w}$ zostało zdefiniowane w ćwiczeniu 14.

(a) Udowodnij, że długość(w) = długość($\overleftarrow{w}$) dla wszystkich $w \in \Sigma^*$.

(b) Udowodnij, że $\overleftarrow{w_1 w_2} = \overleftarrow{w_2}\,\overleftarrow{w_1}$ dla wszystkich $w_1, w_2 \in \Sigma^*$.

§ 7.2. Algorytmy rekurencyjne

W tym paragrafie opisujemy ogólną strukturę algorytmów rekurencyjnych. Pokazujemy również, w jaki sposób udowodnić, że algorytm rekurencyjny rzeczywiście robi to, co miał robić. Będzie to dalszy ciąg omawiania definicji rekurencyjnych i dowodów indukcyjnych z poprzedniego paragrafu.

Przy omawianiu uogólnionej zasady indukcji i definicji rekurencyjnych funkcji mieliśmy do czynienia ze zbiorem S zdefiniowanym rekurencyjnie. Naszym celem było pokazanie, że podzbiór zbioru S składający się z elementów s, dla których pewne zdanie $p(s)$ jest prawdziwe lub wartość pewnej funkcji $f(s)$ jest dobrze określona, jest całym zbiorem S. Inne ważne pytanie, dotyczące rekurencyjnie zdefiniowanego zbioru S, to „jak można stwierdzić, czy jakiś element jest elementem zbioru S?".

PRZYKŁAD 1 Tak jak w przykładzie 1(b) w § 7.1 definiujemy rekurencyjnie zbiór $S = \{2^m\colon\ m \in \mathbb{N}\}$ w następujący sposób:

(P) $1 \in S$;
(R) $2k \in S$, jeśli tylko $k \in S$.

Wtedy $S \subseteq \mathbb{P}$. Dla danego $n \in \mathbb{P}$ chcemy sprawdzić, czy $n \in S$ i jeśli okaże się, że tak, to chcemy znaleźć liczbę m taką, że $n = 2^m$.

Przypuśćmy, że liczba n jest nieparzysta. Jeśli $n = 1$, to z warunku (P) wynika, że $n \in S$ oraz $n = 2^0$. Jeśli $n > 1$, to warunek (R) nie może powodować, by liczba n należała do S, a więc $n \notin S$. W przypadku przeciwnym, gdy liczba n jest nieparzysta, szybko poznajemy odpowiedź.

Jeśli liczba n jest parzysta, np. $n = 2k$, to $n \in S$, jeśli $k \in S$, zgodnie z warunkiem (R), w przeciwnym przypadku, tzn., gdy $k \notin S$, warunek (R) nie może spowodować, by liczba n należała do S. Musimy zatem sprawdzić, czy liczba $k = n/2$ należy do S. Jeśli $k \in S$ i $k = 2^a$, to liczba n też należy do S i $n = 2^{a+1}$, a jeśli $k \notin S$, to również $n \notin S$:

Ta analiza prowadzi do zaprojektowania następującego algorytmu.

Algorytm TEST $(n;\ b, m)$

{Dane: $n \in \mathbb{P}$}
{Wyniki: b = PRAWDA, jeśli $n \in S$; b = FAŁSZ, jeśli $n \notin S$;
$m = \log_2 n$, jeśli $n \in S$; $m = -\infty$, jeśli $n \notin S$}
Jeśli n jest liczbą nieparzystą, to

jeśli $n = 1$, to
  $b :=$ PRAWDA; $m := 0$
  w przeciwnym przypadku
  $b :=$ FAŁSZ; $m := -\infty$
w przeciwnym przypadku {więc n jest liczbą parzystą}
  TEST$(n/2; b', m')$
  {b' jest prawdziwe wtedy i tylko wtedy, gdy $n/2 \in S$, jeśli $n/2 \in S$, to $n/2 = 2^{m'}$}
  $b := b'$; $m := m' + 1$.

Tak jak zwykle, zdania napisane w klamrach {...} są komentarzami, a nie częścią algorytmu. Wkrótce powiemy więcej o takich komentarzach. Nasz sposób wypisywania danych i wyników w nazwie algorytmu TEST, oddzielonych średnikiem, jest bardzo wygodny. Pomaga nam śledzić działanie takich algorytmów jak ten, wywołujących siebie jako podprogramy.

Algorytm TEST działa w ten sposób, że jeśli $n = 1$, to algorytm informuje o sukcesie, jeśli n jest liczbą nieparzystą, różną od 1, to algorytm informuje o porażce i jeśli n jest liczbą parzystą, to algorytm sprawdza liczbę $n/2$. Na przykład dla danej liczby $n = 4$ algorytm TEST(4;) wywołuje TEST(2;), który z kolei wywołuje TEST(1;), który kończy się sukcesem. Otrzymujemy TEST(1; PRAWDA, 0), TEST(2; PRAWDA, 1) i wreszcie TEST(4; PRAWDA, 2). Te wyniki są odbiciem faktu, że $1 = 2^0 \in S$, $2 = 2^1 \in S$ i $4 = 2^2 \in S$. Dla danej liczby $n = 6$ algorytm TEST(6;) wywołuje TEST(3;), który informuje o porażce. Tak jak wyżej, TEST(3; FAŁSZ, $-\infty$) daje TEST(6; FAŁSZ, $-\infty$). ∎

Algorytm TEST z przykładu 1, sprawdzający przynależność do zbioru, ma tę interesującą własność, że jeśli dana liczba n jest parzysta, to algorytm TEST pyta sam siebie o liczbę $n/2$. Algorytmy, które wywołują same siebie, nazywamy **algorytmami rekurencyjnymi**.

Algorytm TEST przez cały czas wywołuje samego siebie dotąd, aż te wywołania rekurencyjne dojdą do liczby nieparzystej, tak więc pomimo iż liczba 4 należy do zbioru S, algorytm TEST(4;) nie wie od razu, że b=PRAWDA; wywołuje TEST(2;), który wywołuje z kolei TEST(1;) i ten wreszcie podaje wartości b i m, które są następnie przekazywane z powrotem przez cały ciąg wywołań. A oto wykonania algorytmu TEST dla danych liczb 4 i 6.

TEST(4;)
  TEST(2;)
    TEST(1;)
      $b :=$ PRAWDA; $m := 0$
    $b :=$ PRAWDA; $m := 0 + 1 = 1$
  $b :=$ PRAWDA; $m := 1 + 1 = 2$

TEST(6;)
  TEST(3;)
    $b :=$ FAŁSZ; $m := -\infty$
    $b :=$ FAŁSZ; $m := -\infty + 1 = -\infty$

PRZYKŁAD 2 Zbiór drzew został zdefiniowany w przykładzie 4 w § 7.1 w następujący sposób:

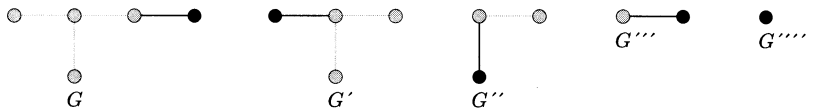
(P) Grafy trywialne (mające jeden wierzchołek) są drzewami;
(R) Jeśli graf T jest otrzymany przez dołączenie liścia do drzewa, to T jest drzewem.

Możemy zaprojektować algorytm rekurencyjny TEST DRZEWA, który sprawdza, czy dany graf jest drzewem.

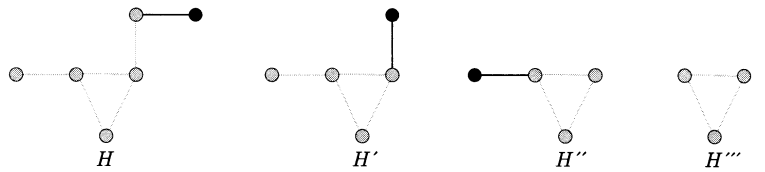
Algorytm TEST DRZEWA$(G; b)$

{Dane: graf skończony G}
{Wyniki: b = PRAWDA, jeśli graf G jest drzewem, b = FAŁSZ, jeśli nie jest}
$b :=$ FAŁSZ
Jeśli graf G ma jeden wierzchołek i nie ma krawędzi, to
  $b :=$ PRAWDA
w przeciwnym przypadku
  jeśli graf G ma liść, to
    odetnij ten liść wraz z wychodzącą z niego krawędzią, by otrzymać graf G'
    TEST DRZEWA$(G'; b')$
    {b' = PRAWDA, jeśli graf G' jest drzewem, b' = FAŁSZ, jeśli nie jest}
    $b := b'$.

Jeśli graf G rzeczywiście jest drzewem, to ten algorytm będzie odcinać z niego krawędzie i wierzchołki dotąd, aż pozostanie jeden wierzchołek i wtedy poinformuje o sukcesie, to znaczy o tym, że b = PRAWDA i prześle tę informację w górę rekurencyjnej drabiny. Na rysunku 7.5 widzimy ciąg G, G', G'', ... danych wejściowych występujących w jednym takim procesie rekurencyjnym.



Rysunek 7.5



Rysunek 7.6

Na rysunku 7.6 pokazany jest taki ciąg dla danego grafu H, który nie jest drzewem. W tym przypadku wywołanie algorytmu TEST DRZEWA(H''';) daje w wyniku $b :=$ FAŁSZ i nie może następnie ani zastosować przypadku dla drzewa trywialnego, ani wywołania rekurencyjnego dla mniejszego grafu. Po prostu informuje, że $b =$ FAŁSZ i następne wywołania TEST DRZEWA(H'';), TEST DRZEWA(H';) oraz TEST DRZEWA(H;) akceptują ten werdykt. ∎

Aby podać motywację dla następnego przykładu, zajmijmy się następującym problemem, z którym prawdopodobnie nie zetknęliśmy się w algebrze: co to jest poprawne wyrażenie? Dlaczego wyrażenie $(x + y)(x - y)$ wygląda dobrze, podczas gdy wyrażenie $(x + -^4/y$ wygląda na bezsensowne? Dlaczego komputer może sobie poradzić z pierwszym wyrażeniem i nie może z drugim? Odpowiedź wymaga wprowadzenia pojęcia **wyrażenia poprawnie zbudowanego** — WPZ (formuły poprawnie zbudowanej, ang. well-formed formula, w skrócie wff).

PRZYKŁAD 3 (a) A oto definicja poprawnie zbudowanego wyrażenia algebraicznego.

(P) Stałe liczbowe i zmienne takie, jak x, y, z są wyrażeniami poprawnie zbudowanymi.

(R) Jeśli f i g są wyrażeniami poprawnie zbudowanymi, to $(f + g)$, $(f - g)$, (fg), (f/g) i (f^g) są wyrażeniami poprawnie zbudowanymi.

Ponieważ x i y są zmiennymi, więc są wyrażeniami poprawnie zbudowanymi. Zatem wyrażeniami poprawnie zbudowanymi są również $(x+y)$ i $(x-y)$. Stąd wynika, że również $((x+y)(x-y))$ jest wyrażeniem poprawnie zbudowanym. Ta definicja nie zado-

wala nas całkowicie, gdyż zewnętrzne nawiasy wydają się zbyteczne. Jednakże bez nich kwadrat $(((x + y)(x - y))^2)$ wyglądałby tak, jak $((x + y)(x - y)^2)$, a te dwa wyrażenia mają różne znaczenia. Problem polega na tym, że w algebrze tradycyjnie opuszczamy nawiasy w pewnych sytuacjach. Gdybyśmy wzięli pod uwagę te wszystkie wyjątkowe przypadki, otrzymalibyśmy skomplikowaną definicję. Zauważmy też, że nasza definicja nie wyklucza dzielenia przez 0. Tak więc $(0/0)$ jest wyrażeniem poprawnie zbudowanym, nawet pomimo że nie możemy przypisać takiemu wyrażeniu wartości liczbowej.

(b) W informatyce symbolu $*$ używa się często do oznaczenia mnożenia, a symbolu $\wedge$ do oznaczenia potęgowania ($a\wedge b$ oznacza a^b). Jeżeli przyjmiemy te oznaczenia, to definicja wyrażeń poprawnie zbudowanych będzie wyglądać następująco:

(P) Stałe liczbowe i zmienne są wyrażeniami poprawnie zbudowanymi.

(R) Jeśli f i g są wyrażeniami poprawnie zbudowanymi, to $(f + g)$, $(f - g)$, $(f * g)$, (f/g) i $(f\wedge g)$ są wyrażeniami poprawnie zbudowanymi.

Na przykład

$$(((((X + Y)\wedge 2) - (2 * (X * Y))) - (X\wedge 2)) - (Y\wedge 2))$$

jest wyrażeniem poprawnie zbudowanym.

(c) W paragrafie 7.4 będziemy omawiać tzw. notację polską, w której nie ma nawiasów. Powyższe przykłady i związane z nimi ćwiczenia pomogą wam docenić zalety tej notacji.

(d) W jaki sposób możemy sprawdzić, czy dany ciąg znaków jest wyrażeniem poprawnie zbudowanym? Definicja rekurencyjna podsuwa nam pewne pomysły dotyczące algorytmu. Przypuśćmy, że umiemy w jakiś sposób rozpoznawać stałe liczbowe i zmienne oraz że nasz dany ciąg znaków składa się ze znaków, których wolno używać. Problem polega na tym, czy te znaki są zestawione ze sobą w sensowny sposób. A oto szkic takiego algorytmu testującego.

Algorytm TEST WPZ $(w; b)$
{Dane: ciąg w dopuszczalnych znaków}
{Wyniki: b = PRAWDA, gdy w jest WPZ, b = FAŁSZ w przeciwnym przypadku}
$b :=$ FAŁSZ
Jeśli w jest stałą liczbową lub zmienną, to
  $b :=$ PRAWDA
w przeciwnym przypadku

znajdź ciągi znaków f i g, dla których ciąg w jest postaci $(f+g)$, $(f-g)$, $(f*g)$, (f/g) lub $(f\hat{}g)$ i dla tych ciągów f i g:
TEST WPZ(f; b')
TEST WPZ(g; b'')
$b := b' \wedge b''$ { a więc b = PRAWDA wtedy i tylko wtedy, gdy b' i b'' są prawdziwe}

Jeśli algorytm znajdzie takie ciągi f i g, które *są* wyrażeniami poprawnie zbudowanymi, to może nie testować już następnych możliwych ciągów f i g. Jest tak dlatego, że z warunku (R) wynika, iż nowe wyrażenia poprawnie zbudowane mogą być skonstruowane tylko w jeden sposób. Wyrażenie $(f*g)$ nie może być również postaci (h/k) dla jakichś wyrażeń poprawnie zbudowanych h i k, a także nie może być postaci $(f'*g')$, gdzie f' i g' są wyrażeniami poprawnie zbudowanymi takimi, że $f' \neq f$ i $g' \neq g$. Wypisanie wszystkich szczegółów w części algorytmu znajdującej się po słowach „w przeciwnym przypadku" byłoby żmudne, ale nie prowadziłoby do żadnych niespodzianek. Można łatwo uwierzyć, że możliwe jest napisanie programu komputerowego, który rozpoznawałby wyrażenia poprawnie zbudowane. ■

Komputer musi być w stanie czytać ciągi danych wejściowych, aby zobaczyć, czy zawierają one jakieś instrukcje lub inne mające znaczenie wyrażenia. Zaprojektowanie kompilatora oznacza zaprojektowanie języka instrukcji i algorytmu rozpoznawania tak, aby komputer mógł dokonać analizy składniowej, to znaczy zrozumieć znaczenie ciągu danych. Ostatnie osiągnięcia w teorii języków formalnych były w znaczny sposób motywowane problemami analizy składniowej.

Algorytmy rekurencyjne przydają się do wielu innych celów, nie tylko do testowania należenia do zbiorów. Przyjrzymy się kilku innym przykładom w tym paragrafie i znajdziemy również ważne przykłady w tym i następnych rozdziałach. Podstawową cechą algorytmu rekurencyjnego jest to, że wywołuje on sam siebie, ale to oczywiście nie miałoby żadnego znaczenia, gdyby wywołania rekurencyjne nie przybliżały nas bardziej do końca działania algorytmu niż wywołanie pierwotne. Aby być pewnym, że algorytm zakończy działanie, musimy wiedzieć, że ciąg wywołań rekurencyjnych nie może ciągnąć się w nieskończoność. Muszą istnieć pewne **przypadki początkowe**, w których wynik jest otrzymywany bez następnych wywołań rekurencyjnych i musimy zagwarantować, że dane dla kolejnych wywołań rekurencyjnych są

za każdym razem w pewnym sensie bliższe przypadkom początkowym. Typowy algorytm rekurencyjny, który ma szanse zakończyć działanie, wygląda mniej więcej następująco:

Algorytm REKUR(I; O)

{Dane I (od ang. input); Wyniki: O (od ang. output)}
jeśli dana I jest jakimś przypadkiem początkowym, to
  wynik O staje się tym, czym być powinien,
w przeciwnym przypadku
  znajdź jakieś dane I_k, bliższe przypadkom początkowym niż I,
  REKUR(I_k; O_k) dla każdego k
  wykorzystaj jakoś wszystkie O_k, aby wyznaczyć O.

PRZYKŁAD 4 W tablicy 7.1 widzimy sześć bardzo prostych algorytmów. Kilka pierwszych linii tych wszystkich algorytmów wygląda podobnie. We wszystkich przypadkach zakładamy, że dana liczba n jest liczbą całkowitą dodatnią i przypadkiem początkowym jest $n = 1$. Czy te algorytmy zakończą działanie? Jeśli tak, to jak się ma wynik r do danych wejściowych n?

Tablica 7.1

Algorytm FOO(n; r)	Algorytm GOO(n; r)	Algorytm BOO(n; r)
Jeśli $n = 1$, to $r := 1$ w przeciwnym przypadku FOO($n - 1$; s) $r := s + 1$.	Jeśli $n = 1$, to $r := 1$ w przeciwnym przypadku GOO($n - 1$; s) $r := n * s$.	Jeśli $n = 1$, to $r := 1$ w przeciwnym przypadku BOO($n + 1$; s) $r := s + 1$.
Algorytm MOO(n; r)	**Algorytm TOO(n; r)**	**Algorytm ZOO(n; r)**
Jeśli $n = 1$, to $r := 1$ w przeciwnym przypadku MOO($n/2$; s) $r := s + 1$.	Jeśli $n = 1$, to $r := 1$ w przeciwnym przypadku TOO(n DIV 2; s) $r := s + 1$.	Jeśli $n = 1$, to $r := 1$ w przeciwnym przypadku ZOO(n DIV 2; s) $r := 2 * s$.

Możemy trochę poeksperymentować. Dla $n = 1$ algorytm FOO daje $r := 1$. Dla $n = 2$ wywołuje on FOO(1;), aby otrzymać $s = 1$, skąd $r := 1 + 1 = 2$. Dla $n = 3$ wywołuje on FOO(2;), by otrzymać $s = 2$, jak to widzieliśmy przed chwilą, skąd $r := 2+1 = 3$. Tak naprawdę wydaje się, że $r = n$ za każdym razem. W każdym przypadku, ponieważ ciąg $n, n - 1, n - 2, \ldots$ nie może się ciągnąć w nieskończoność, nie osiągając przypadku początkowego, algorytm FOO musi zakończyć działanie. Powró-

cimy za chwilę do pytania: w jaki sposób można dowieść, że $r = n$ dla każdego n?

Algorytm GOO również kończy działanie z tych samych powodów co FOO. A oto kilka początkowych wartości r: dla $n = 1$ mamy $r = 1$; dla $n = 2$ mamy $n-1 = 1$, skąd $s = 1$ i $r = 2*1 = 2$; dla $n = 3$ mamy $n-1 = 2$, więc $s = 2$ i $r = 3*2 = 6$. Wygląda na to, że zawsze $r = n!$, a więc za pomocą algorytmu GOO możemy w rekurencyjny sposób obliczać silnie.

Algorytm BOO nigdy nie zakończy działania. Na przykład BOO(2;) wywołuje BOO(3;), który z kolei wywołuje BOO(4;) itd. Wywołania rekurencyjne nie przybliżają się do przypadku początkowego $n = 1$.

Algorytm MOO nie daje wyniku z innego powodu. Pomimo że liczba $n/2$ jest bliższa 1 niż n, to jednak $n/2$ może nie być liczbą całkowitą, a więc MOO($n/2$;) może nie mieć sensu. Jeśli zdecydujemy się dopuścić dane, które nie są liczbami całkowitymi, to algorytm MOO może nie zakończyć działania; na przykład ciąg $3, 3/2, 3/4, 3/8, \ldots$ ciągnie się w nieskończoność.

Problem, który wystąpił w algorytmie MOO, został rozwiązany w algorytmie TOO, gdyż n DIV 2 na pewno jest liczbą całkowitą i jeśli $n > 1$, to $n > n$ DIV $2 \geq 1$. W tablicy 7.2 pokazanych jest kilka pierwszych wartości r dla algorytmu TOO, a także dla pozostałych algorytmów z tablicy 7.1.

Tablica 7.2

	Wartości r					
n	FOO	GOO	BOO	MOO	TOO	ZOO
1	1	1	1	1	1	1
2	2	2		2	2	2
3	3	6			2	2
4	4	24		3	3	4
5	5	120			3	4
6	6	720			3	4
7	7	5040			3	4

Algorytm ZOO również kończy działanie, tak samo jak TOO, ale daje inny wynik. ■

Z chwilą gdy wiemy, że algorytm rekurencyjny zakończy działanie, to chcemy wiedzieć, jakie on daje wyniki. Jeśli chcemy, by ten algorytm dawał jakieś określone wyniki, to musimy sprawdzić, że daje takie wyniki, jakie powinien dać. Znaczy to, że zachowując oznaczenia z naszego algorytmu REKUR(I; O), musimy sprawdzić, iż

(p) Jeśli I jest przypadkiem początkowym, to wynik O ma właściwą wartość;
(r) Jeśli każdy wynik O_k ma właściwą wartość przy wywołaniu rekurencyjnym REKUR(I_k; O_k), to wynik O ma właściwą wartość dla danej I.

Te warunki przypominają warunki, które należy sprawdzić w dowodzie korzystającym z uogólnionej zasady indukcji i oczywiście nie jest to przypadek. Zdefiniujmy rekurencyjnie zbiór S dopuszczalnych danych dla algorytmu REKUR w następujący sposób. W warunku (P) niech X będzie zbiorem danych w przypadkach początkowych. Powiemy, że dane I można otrzymać ze zbioru danych $\{I_1, \ldots, I_m\}$ wtedy, gdy we fragmencie algorytmu REKUR(I; O), występującym po słowach „w przeciwnym przypadku", wynik O może być obliczony z wyników O_1, ..., O_m uzyskanych przez wywołania REKUR(I_1; O_1), ..., REKUR(I_m; O_m). Wtedy warunek rekurencyjny można zapisać następująco:

(R) Jeśli dane I można otrzymać z danych I_1, ..., I_m, należących do zbioru S, to $I \in S$.

Wtedy zbiór S jest to dokładnie zbiór tych danych, dla których algorytm REKUR może obliczać wyniki. Dla każdych danych I ze zbioru S definiujemy zdanie $p(I)$ =„algorytm REKUR daje poprawny wynik dla danych I". Wtedy warunki (p) i (r) przybiorą postać:

(p′) Zdanie $p(I)$ jest prawdziwe dla wszystkich $I \in X$;
(r′) Zdanie $p(I)$ jest prawdziwe dla wszystkich danych I, które można otrzymać z takich danych I_k ze zbioru S, dla których zdania $p(I_k)$ są prawdziwe.

Uogólniona zasada indukcji mówi, że te warunki gwarantują, iż dla każdego $I \in S$ zdanie $p(I)$ jest prawdziwe.

PRZYKŁAD 5 (a) Aby dowieść, że algorytm GOO oblicza wartość $n!$, wystarczy po prostu sprawdzić, że

(p) $1! = 1$ oraz
(r) $n! = n * (n-1)!$.

Te równości są prawdziwe, co wynika z definicji $n!$.

(b) Twierdzimy, że algorytm TEST z przykładu 1 daje wynik b = PRAWDA i $m = \log_2 n$, gdy dana liczba n jest potęgą liczby 2 oraz b = FAŁSZ i $m = -\infty$ w przeciwnym przypadku. A oto przypominamy sam algorytm:

Algorytm TEST $(n;\ b, m)$

Jeśli n jest liczbą nieparzystą, to
  jeśli $n = 1$, to
    $b :=$ PRAWDA; $m := 0$
  w przeciwnym przypadku
    $b :=$ FAŁSZ; $m := -\infty$
w przeciwnym przypadku
  TEST$(n/2;\ b',\ m')$
  $b := b'$; $m := m' + 1$.

Jakie są tu przypadki początkowe i dopuszczalne dane? Oczywiście chcemy, by zbiorem S dopuszczalnych danych był zbiór $\mathbb{P}$. Gdybyśmy dopuścili tylko $n = 1$ jako przypadek początkowy, to ponieważ warunek (R) pozwala wyłącznie otrzymać liczbę $2s$ z liczby s, więc zbiór S składałby się tylko z potęg liczby 2. Ten algorytm traktuje jednak wszystkie liczby nieparzyste ze zbioru $\mathbb{P}$ jako przypadki początkowe.

Aby sprawdzić, że algorytm działa poprawnie, tzn. daje takie wyniki, jak twierdzimy, musimy po prostu sprawdzić, że

(p) jeśli $n = 1$, to $n = 2^0$ oraz jeśli n jest liczbą nieparzystą, większą niż 1, to n nie jest potęgą liczby 2;
(r) jeśli $n/2 = 2^{m'}$, to $n = 2^{m'+1}$.

Oczywiście oba warunki (p) i (r) są prawdziwe.

Kiedy już zajęliśmy się algorytmem TEST, możemy oszacować czas $T(n)$, jakiego ten algorytm potrzebuje do zakończenia obliczeń dla danych n. Do rozpoznania przypadku początkowego i wykonania obliczeń dla danej liczby nieparzystej algorytm potrzebuje stałej ilości czasu C. Jeśli liczba n jest parzysta, to algorytm potrzebuje czasu $T(n/2)$ do wywołania TEST$(n/2,\ )$ oraz pewnego stałego czasu D do obliczenia b i m z b' i m'. Zatem

$$(*) \qquad T(n) \leq C + T(n/2) + D.$$

Niech teraz $n = 2^m \cdot k$, gdzie liczba k jest nieparzysta oraz $m > 0$. Wtedy

$$\begin{aligned} T(n) = T(2^m \cdot k) &\leq T(2^{m-1} \cdot k) + C + D \leq \\ &\leq T(2^{m-2} \cdot k) + C + D + C + D \leq \\ &\leq \ldots \leq \\ &\leq T(k) + m \cdot (C+D) \leq C + m \cdot (C+D) \leq \\ &\leq 2m \cdot (C + D). \end{aligned}$$

Ponieważ $\frac{n}{k} = 2^m$, więc mamy

$$m = \log_2 \frac{n}{k} \leq \log_2 n, \quad \text{a więc} \quad T(n) \leq 2(C+D) \cdot \log_2 n.$$

Zatem $T(n) = O(\log_2 n)$. Wielokropek w trakcie rozumowania niczemu nie szkodzi; z chwilą, gdy odgadniemy nierówność $T(n) \leq 2(C+D)\log_2 n$, możemy dowieść jej przez indukcję, korzystając z nierówności $(*)$.

Ten algorytm jest znacznie szybszy dla dużych n niż którykolwiek z algorytmów postaci „dziel i rządź", do których stosowało się twierdzenie 2 z § 4.4. Zauważmy, że z nierówności $(*)$ mamy

$$T(2n) \leq T(n) + (C+D) \quad \text{dla wszystkich} \quad n.$$

Jest to znacznie mocniejsze oszacowanie niż jakakolwiek nierówność postaci

$$s_{2n} \leq 2s_n + f(n),$$

w której występuje współczynnik 2. Jednakże można tutaj zastosowć odpowiednią postać (z nierównością) ćwiczenia 17 z § 4.4, dla $b = 1$ i $f(n) = C + D$; otrzymamy wtedy

$$T(2^m) \leq C + m \cdot (C+D) \quad \text{dla} \quad m > 0.$$

(c) Wśród przypadków początkowych w algorytmie TEST DRZEWA z przykładu 2 znajdują się nie tylko grafy mające jeden wierzchołek, ale także wszystkie grafy skończone bez liści. Dla każdego takiego grafu algorytm wylicza wartość b bez dalszych wywołań rekurencyjnych.

(d) Podobnie, wśród przypadków początkowych w algorytmie TEST WPZ z przykładu 3 są wszystkie ciągi znaków, które nie są postaci $(f+g)$, $(f-g)$, $(f*g)$, (f/g) czy (f^g). ■

Za cenę przeznaczenia dodatkowej pamięci na przechowywanie wyników częściowych, możemy przekształcić każdą pętlę „dopóki" na procedurę rekurencyjną. Na przykład algorytm rekurencyjny

Algorytm REKUR (;)

Jeśli warunek γ jest fałszywy, to
  nic nie rób
w przeciwnym przypadku
  S
  REKUR (;)

daje ten sam efekt, co pętla

dopóki γ, wykonuj
S.

PRZYKŁAD 6 W paragrafie 4.6 użyliśmy pętli „dopóki" w algorytmie Euklidesa, choć wydaje się, że użycie rekurencji byłoby bardziej naturalne. Podstawowa obserwacja, mianowicie $\mathrm{NWD}(m,n) = \mathrm{NWD}(n,\ m \bmod n)$ dla $n \neq 0$, prowadzi do sformułowania następującego algorytmu rekurencyjnego.

Algorytm EUKLIDES $(m,\ n;\ d)$

{Dane: liczby $m, n \in \mathbb{N}$, nie równe jednocześnie 0}
{Wyniki: $d = \mathrm{NWD}(m,n)$}
Jeśli $n = 0$, to
$d := m$
w przeciwnym przypadku
EUKLIDES$(n,\ m \bmod n;\ d')$
$d := d'$. ∎

Niewielkie zmiany dają algorytm, który wylicza liczby s i t takie, że $d = sm + tn$.

Algorytm EUKLIDES$^+$ $(m,\ n;\ d,\ s,\ t)$

Jeśli $n = 0$, to
$d := m$; $s := 1$; $t := 0$
w przeciwnym przypadku
EUKLIDES$^+(n,\ m \bmod n;\ d',\ s',\ t')$
$d := d'$; $s := t'$; $t := s' - t' \cdot (m \operatorname{div} n)$. ∎

Czasy działania tych algorytmów są w zasadzie takie same, jak czasy działania wersji iteracyjnych, wyznaczone w § 4.6.

Dla przypomnienia powtarzamy warunki, jakie musi spełniać algorytm rekurencyjny.

Warunki poprawności dla algorytmów rekurencyjnych:

(a) Algorytm musi zakończyć działanie. W szczególności, wywołania rekurencyjne muszą prowadzić w kierunku przypadków początkowych.

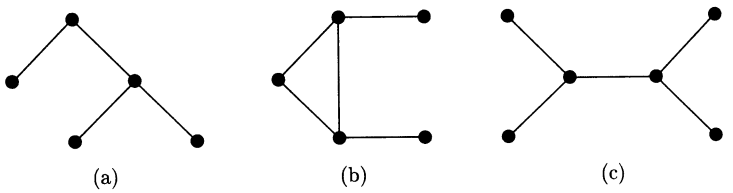
(b) Algorytm musi dawać wyniki poprawne w przypadkach początkowych.

(c) Algorytm musi dawać wyniki poprawne, jeśli wszystkie wywołania rekurencyjne dawały wyniki poprawne.

ĆWICZENIA DO § 7.2

1. (a) Zilustruj wykonanie algorytmu TEST dla danej wejściowej 20.
(b) Zrób to samo dla 8.

2. Zilustruj algorytm TEST DRZEWA dla grafów przedstawionych na rysunku 7.7. Możesz zrobić rysunki takie jak rysunki 7.5 i 7.6.



Rysunek 7.7

3. Dodaj nawiasy do następujących wyrażeń algebraicznych tak, aby były one wyrażeniami poprawnie zbudowanymi, zgodnie z definicją w przykładzie 3:
(a) $x + y + z$, (b) $x + y/z$,
(c) xyz, (d) $(x + y)^{x+y}$.

4. Dodaj nawiasy do następujących wyrażeń algebraicznych tak, aby były one wyrażeniami poprawnie zbudowanymi, tak jak wyrażenia zdefiniowane w przykładzie 3:
(a) $X + Y + Z$, (b) $X * (Y + Z)$,
(c) $X\hat{}2 + 2 * X + 1$, (d) $X + Y/Z - Z * X$.

5. Skorzystaj z definicji rekurencyjnej wyrażenia poprawnie zbudowanego z przykładu 3, aby pokazać, że następujące wyrażenia są wyrażeniami poprawnie zbudowanymi.
(a) $((x^2) + (y^2))$, (b) $(((X\hat{}2) + (Y\hat{}2))\hat{}2)$,
(c) $((X + Y) * (X - Y))$.

6. (a) Niech GOOF będzie algorytmem GOO, w którym zamiast przypisania $r := 1$ mamy $r := 0$ (i wywołaniem rekurencyjnym będzie GOOF$(n - 1; s)$). Jaki będzie wynik r, jeśli daną wejściową jest liczba całkowita dodatnia?
(b) Co się stanie, jeśli przypisanie $r := 1$ pozostanie nie zmienione, a przypisanie $r := n * s$ zmieni się na $r := (n - 1) * s$?

7. Uzupełnij wiersze tablicy 7.2 dla $n = 8$ i $n = 9$.

8. Sprawdź, że algorytm TOO daje w wyniku $k + 1$, jeśli $2^k \leq n < 2^{k+1}$. *Wskazówka*: zastosuj indukcję względem k.

9. Sprawdź, że algorytm ZOO daje w wyniku 2^k, jeśli $2^k \leq n < 2^{k+1}$.

10. Pokaż, że algorytm EUKLIDES$^+$ oblicza $d = \text{NWD}(80, 35)$ i znajduje s i t takie, że $d = 80s + 35t$.

11. Powtórz ćwiczenie 10 dla $d = \text{NWD}(108, 30)$.

12. Powtórz ćwiczenie 10 dla $d = \text{NWD}(56, 21)$.

13. (a) Udowodnij poprawność algorytmu EUKLIDES z przykładu 6.
(b) Udowodnij poprawność algorytmu EUKLIDES$^+$ z przykładu 6.

14. Definiujemy rekurencyjnie **głębokość** algebraicznego wyrażenia poprawnie zbudowanego w następujący sposób (zob. przykład 3):

(P) Stałe liczbowe i zmienne mają głębokość równą 0;
(R) Jeśli głębokość(f) i głębokość(g) są już zdefiniowane, to każde z wyrażeń $(f+g)$, $(f-g)$, $(f*g)$, (f/g) oraz $(f\hat{}g)$ ma głębokość równą $1 + \max\{\text{głębokość}(f), \text{głębokość}(g)\}$.

Okazuje się, że jest to definicja poprawna; por. przykład 10 w § 7.1. Oblicz głębokość następujących wyrażeń algebraicznych:
(a) $((x\hat{}2) + (y\hat{}2))$,
(b) $(((X\hat{}2) + (Y\hat{}2))\hat{}2)$,
(c) $((X + Y) * (X - Y))$,
(d) $(((((X + Y)\hat{}2) - (2 * (X * Y))) - (X\hat{}2)) - (Y\hat{}2))$,
(e) $(((x + (x + y)) + z) - y)$,
(f) $(((X * Y)/X) - (Y\hat{}4))$.

15. Oto jednoznaczna definicja rekurencyjna formuł poprawnie zbudowanych w rachunku zdań.

(P) Zmienne takie, jak p, q, r są formułami poprawnie zbudowanymi;
(R) Jeśli P i Q są formułami poprawnie zbudowanymi, to $(P \vee Q)$, $(P \wedge Q)$, $(P \to Q)$, $(P \leftrightarrow Q)$ i $\neg P$ są formułami poprawnie zbudowanymi.

Zauważ, że nie wymagamy użycia nawiasów przy zaprzeczeniu zdania. Zatem znak negacji $\neg$ obejmuje swoim zasięgiem najkrótszą formułę następującą po tym znaku, która jest poprawnie zbudowana. W praktyce, staramy się opuszczać zewnętrzne nawiasy oraz, aby napisy były bardziej czytelne, możemy zamiast nawiasów zwykłych używać nawiasów [ ] oraz { }. Pokaż, że następujące formuły rachunku zdań są formułami poprawnie zbudowanymi:
(a) $\neg(p \vee q)$,
(b) $(\neg p \wedge \neg q)$,
(c) $((p \leftrightarrow q) \to ((r \to p) \vee q))$.

16. Zmodyfikuj definicję w ćwiczeniu 15 tak, aby dopuścić użycie spójnika $\oplus$ oznaczającego alternatywę wykluczającą.

17. W tym ćwiczeniu niech p i q będą ustalonymi zdaniami. Definiujemy rekurencyjnie rodzinę $\mathcal{F}$ zdań złożonych, używających tylko p, q, $\wedge$ i $\vee$, w następujący sposób:

(P) $p, q \in \mathcal{F}$;
(R) jeśli $P, Q \in \mathcal{F}$, to $(P \wedge Q)$ oraz $(P \vee Q)$ należą do $\mathcal{F}$.

(a) Wykorzystaj tę definicję, aby sprawdzić, że zdanie $(p \wedge (p \vee q))$ należy do zbioru $\mathcal{F}$.
(b) Udowodnij, że jeżeli p i q są zdaniami fałszywymi, to wszystkie zdania należące do zbioru $\mathcal{F}$ są fałszywe.
(c) Pokaż, że zdanie $p \rightarrow q$ nie jest logicznie równoważne z żadnym zdaniem ze zbioru $\mathcal{F}$. To dowodzi nieudowodnionego stwierdzenia w odpowiedzi do ćwiczenia 17(c) w § 2.4.

§ 7.3. Algorytmy przeszukiwania w głąb

Algorytm przechodzenia drzewa jest to algorytm, który wypisuje (odwiedza, przeszukuje lub ustawia w ciąg) wszystkie wierzchołki skończonego uporządkowanego drzewa z wyróżnionym korzeniem. Trzy najbardziej popularne takie algorytmy tworzą ciąg wierzchołków uporządkowany za pomocą porządku prefiksowego (ang. preorder), infiksowego (ang. inorder) — tylko w przypadku drzew binarnych — lub postfiksowego (ang. postorder). Te trzy algorytmy są algorytmami rekurencyjnymi.

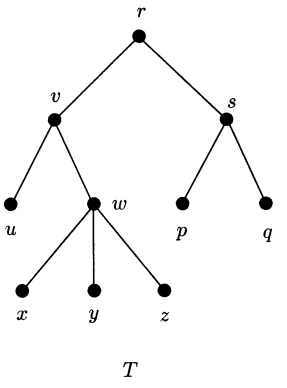
W **porządku prefiksowym** korzeń drzewa jest umieszczony na pierwszym miejscu listy wierzchołków, a dalej znajdują się poddrzewa w kolejności swoich korzeni. Ze względu na sposób rysowania uporządkowanych drzew z wyróżnionym korzeniem, będziemy mówić o kolejności od lewej do prawej. W tym algorytmie korzeń drzewa jest pierwszy, a kolejność potomków zależy od kolejności dzieci. Przypadkami początkowymi są wierzchołki bez dzieci, tzn. liście. Wtedy pierwszy krok algorytmu dołączy liść do listy wierzchołków, a pętla „dla" jest przebiegana „w próżni"; tak więc lista wierzchołków składa się tylko z jednego wierzchołka, mianowicie v.

Algorytm PREORDER(v)
{Dane: skończone, uporządkowane drzewo z wyróżnionym korzeniem v}
{Wyniki: lista (ciąg) $L(v)$ wszystkich wierzchołków tego drzewa, na której rodzice znajdują się zawsze przed swoimi dziećmi}
Umieść wierzchołek v na liście $L(v)$
dla każdego dziecka w wierzchołka v (od lewej do prawej)
  PREORDER(w) {aby otrzymać listę $L(w)$ złożoną z wierzchołka w i jego potomków}
  dołącz otrzymaną listę $L(w)$ na końcu listy $L(v)$. ∎

W **porządku postfiksowym** poddrzewa są umieszczone najpierw i dopiero na końcu znajduje się korzeń. Znów przypadkami początkowymi są liście. W takich przypadkach pierwszy krok jest wykonany „w próżni" i lista składa się z jednego wierzchołka v.

Algorytm POSTORDER(v)
{Dane: skończone, uporządkowane drzewo z wyróżnionym korzeniem v}
{Wyniki: lista (ciąg) $L(v)$ wszystkich wierzchołków tego drzewa, na której rodzice znajdują się zawsze po swoich dzieciach}
Utwórz pustą listę $L(v) = \lambda$.
Dla każdego dziecka w wierzchołka v (od lewej do prawej)
POSTORDER(w) {aby otrzymać listę $L(w)$ złożoną z wierzchołka w i jego potomków}
dołącz otrzymaną listę $L(w)$ na końcu listy dotychczas uzyskanej $L(v)$.
Umieść wierzchołek v na końcu listy $L(v)$. ∎

PRZYKŁAD 1 Zastosujemy algorytmy PREORDER i POSTORDER do drzewa T przedstawionego na rysunku 7.8. Ponieważ zrozumienie wszystkich algorytmów podanych w tym paragrafie zależy od bardzo dobrego zrozumienia tych dwóch prostych algorytmów, wyjaśnimy każdy krok w tym przykładzie.



T

Rysunek 7.8

(a) Algorytm PREORDER działa w następujący sposób:

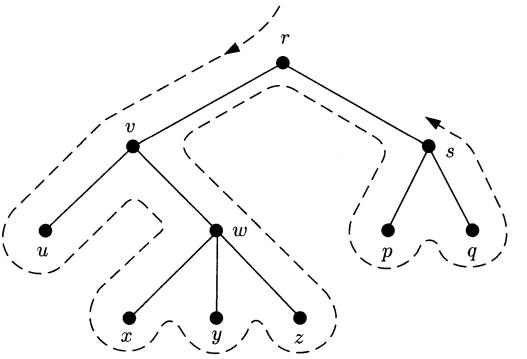
Wpisz r na listę. $L(r) = r$.
Weź pierwsze dziecko v wierzchołka r. Wpisz v na listę. $L(v) = v$.

Weź pierwsze dziecko u wierzchołka v. Wpisz u na listę.
$L(u) = u$.
{u nie ma dzieci.}$L(v) = vu$.
Weź następne dziecko w wierzchołka v. Wpisz w na listę.
$L(w) = w$.
Weź pierwsze dziecko x wierzchołka w. Wpisz x na listę.
$L(x) = x$.
{x nie ma dzieci.}$L(w) = wx$.
Weź następne dziecko y wierzchołka w. Wpisz y na listę.
$L(y) = y$.
{y nie ma dzieci.}$L(w) = wxy$.
Weź następne dziecko z wierzchołka w. Wpisz z na listę.
$L(z) = z$.
{z nie ma dzieci.}$L(w) = wxyz$.
{Wszystkie dzieci wierzchołka w zostały wzięte}
$L(v) = vuwxyz$.
{Wszystkie dzieci wierzchołka v zostały wzięte}
$L(r) = rvuwxyz$.
Weź następne dziecko s wierzchołka r. Wpisz s na listę.
$L(s) = s$.
Weź pierwsze dziecko p wierzchołka s. Wpisz p na listę.
$L(p) = p$.
{p nie ma dzieci.}$L(s) = sp$.
Weź następne dziecko q wierzchołka s. Wpisz q na listę.
$L(q) = q$.
{q nie ma dzieci.}$L(s) = spq$.
{Wszystkie dzieci wierzchołka s zostały wzięte}
$L(r) = rvuwxyzspq$.
{Wszystkie dzieci wierzchołka r zostały wzięte}
Stop.

W wyniku otrzymaliśmy ciąg (listę) wierzchołków $rvuwxyzspq$. Algorytm PREORDER umieszcza każdy wierzchołek na liście wtedy, gdy odwiedza go po raz pierwszy, zatem tę listę można otrzymać z drzewa T tak, jak pokazano na rysunku 7.9. Po prostu należy poruszać się po przerywanej linii, umieszczając każdy wierzchołek na liście, gdy dojdzie się do niego pierwszy raz.

(b) Algorytm POSTORDER tworzy ciąg wierzchołków w taki sposób, że rodzice znajdują się za swoimi dziećmi:

Odwiedź r. $L(r) = \lambda$.
Odwiedź pierwsze dziecko v wierzchołka r. $L(v) = \lambda$.
Odwiedź pierwsze dziecko u wierzchołka v. $L(u) = \lambda$.



Rysunek 7.9

{u nie ma dzieci.} Wpisz u na listę. $L(u)=u; L(v)=u$.
Odwiedź następne dziecko w wierzchołka v. $L(w) = \lambda$.
Odwiedź pierwsze dziecko x wierzchołka w. $L(x) = \lambda$.
{x nie ma dzieci.} Wpisz x na listę. $L(x) = x; L(w) = x$.
Odwiedź następne dziecko y wierzchołka w. $L(y) = \lambda$.
{y nie ma dzieci.} Wpisz y na listę. $L(y) = y; L(w) = xy$.
Odwiedź następne dziecko z wierzchołka w. $L(z) = \lambda$.
{z nie ma dzieci.} Wpisz z na listę. $L(z) = z; L(w) = xyz$.
{Wszystkie dzieci wierzchołka w zostały odwiedzone}
Wpisz w na listę. $L(w) = xyzw; L(v) = uxyzw$.
{Wszystkie dzieci wierzchołka v zostały odwiedzone}
Wpisz v na listę. $L(v) = uxyzwv; L(r) := uxyzwv$.
Odwiedź następne dziecko s wierzchołka r. $L(s) = \lambda$.
Odwiedź pierwsze dziecko p wierzchołka s. $L(p) = \lambda$.
{p nie ma dzieci.} Wpisz p na listę. $L(p) = p; L(s) = p$.
Odwiedź następne dziecko q wierzchołka s. $L(q) = \lambda$.
{q nie ma dzieci.} Wpisz q na listę. $L(q) = q; L(s) = pq$.
{Wszystkie dzieci wierzchołka s zostały odwiedzone}
Wpisz s na listę. $L(s) = pqs; L(r) = uxyzwvpqs$.
{Wszystkie dzieci wierzchołka r zostały odwiedzone}
Wpisz r na listę. $L(r) = uxyzwvpqsr$.
Stop.

Otrzymaliśmy w wyniku ciąg (listę) wierzchołków $uxyzwvpqsr$. Algorytm POSTORDER umieszcza każdy wierzchołek na liście wtedy, gdy odwiedza go po raz ostatni, a więc można tę listę

otrzymać z grafu T przedstawionego na rysunku 7.9, pod warunkiem, że każdy wierzchołek zostanie umieszczony na liście podczas ostatnich odwiedzin. ∎

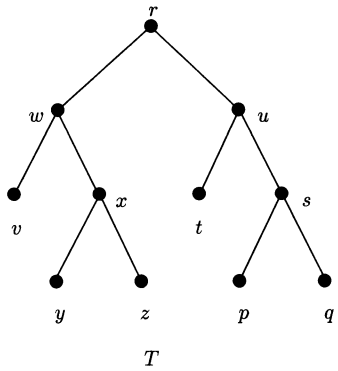
Wierzchołki uporządkowanych drzew *binarnych* z wyróżnionym korzeniem można ustawić w ciąg w jeszcze inny sposób. W porządku infiksowym korzeń v znajduje się między wierzchołkami poddrzewa, którego korzeniem jest lewe dziecko korzenia r, a wierzchołkami poddrzewa, którego korzeniem jest prawe dziecko korzenia r.

Algorytm INORDER(v)
{Dane: skończone, uporządkowane drzewo binarne z wyróżnionym korzeniem v}
{Wyniki: lista (ciąg) $L(v)$ wszystkich wierzchołków tego drzewa, na której lewe dziecko znajduje się przed swoim rodzicem, a prawe dziecko znajduje się po nim.}
Utwórz pustą listę $L(v) = \lambda$.
Jeśli wierzchołek v ma lewe dziecko w, to
  INORDER(w) {aby otrzymać listę $L(w)$ złożoną z wierzchołka w i jego potomków}
  dołącz listę $L(w)$ do listy $L(v)$.
  Dołącz v na końcu listy $L(v)$.
Jeśli wierzchołek v ma prawe dziecko u, to
  INORDER(u) {aby otrzymać listę $L(u)$}
  dołącz listę $L(u)$ na końcu listy $L(v)$. ∎

Jeśli dane drzewo nie jest regularne, to niektóre wierzchołki mają tylko jedno dziecko. Ponieważ każde takie dziecko jest albo lewym dzieckiem, albo prawym dzieckiem, więc algorytm nadal działa. Podobnie jak w przypadku algorytmów PREORDER i POSTORDER, jeśli wierzchołek nie ma dzieci, to w wyniku otrzymuje się ciąg złożony tylko z tego wierzchołka.

PRZYKŁAD 2 (a) Weźmy drzewo binarne przedstawione na rysunku 7.10. Algorytm INORDER działa wtedy w następujący sposób:

Odwiedź r, w i v, dołącz v do listy. $L(w) = L(v) = v$.
  Powróć do w, dołącz w do listy. $L(w) = vw$.
  Następnie odwiedź x i y, dołącz y do listy. $L(x) = L(y) = y$.
    Powróć do x, dołącz x do listy. $L(x) = yx$.
    Następnie odwiedź z, dołącz z do listy. $L(z) = z; L(x) = yxz; L(w) = vwyxz$.



Rysunek 7.10

Powróć do r, dołącz r do listy. $L(r) = vwyxzr$.
Następnie odwiedź u i t, dołącz t do listy. $L(u) = L(t) = t$.
Powróć do u, dołącz u do listy. $L(u) = tu$.
Następnie odwiedź s i p, dołącz p do listy. $L(s) = L(p) = p$.
Powróć do s, dołącz s do listy. $L(s) = ps$.
Następnie odwiedź q, dołącz q do listy. $L(q) = q; L(s) = psq; L(u) = tupsq;$
$L(r) = vwyxzrtupsq$.
Powróć do s, u i r i zatrzymaj się.

Ostatecznym ciągiem wierzchołków jest $vwyxzrtupsq$. W tym przykładzie każdy rodzic miał dwoje dzieci: lewe i prawe. Gdyby niektóre dzieci były nieobecne, to algorytm nadal działałby poprawnie. Na przykład, gdyby dziecko y było nieobecne, to ciąg $L(x)$ byłby postaci xz, a gdyby cała gałąź x, y, z była nieobecna, to ciąg $L(w)$ byłby postaci vw.

(b) Weźmy drzewo etykietowane pokazane na rysunku 6.33(a) w § 6.4. Porządek infiksowy poddrzewa mającego korzeń 00 to: 000, 00 i 001; porządki innych poddrzew można wypisać równie łatwo. Porządek infiksowy całego drzewa jest następujący:

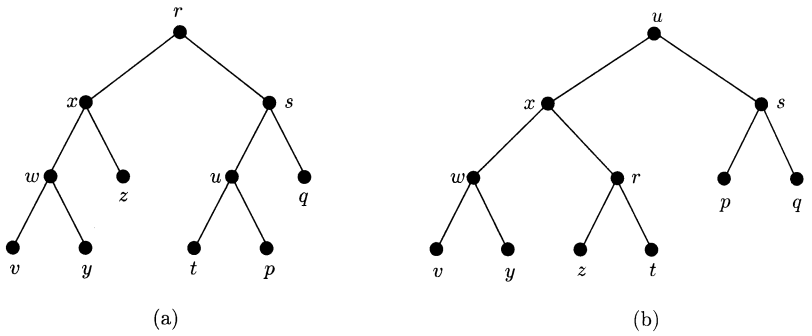
000, 00, 001, 0, 010, 01, 011, korzeń, 100, 10, 101, 1, 110, 11, 111.

■

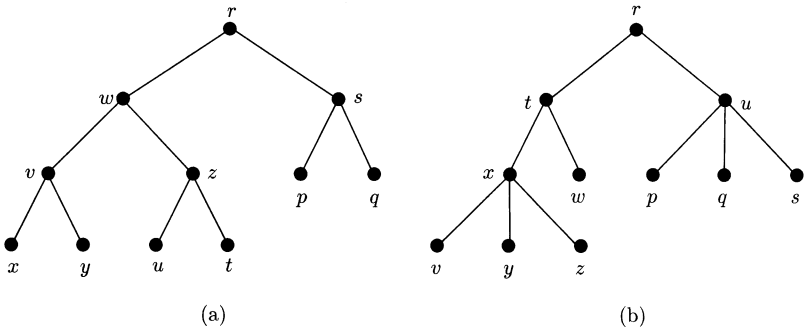
Czy możemy odtworzyć drzewo, jeśli mamy dany porządek jego wierzchołków? Na ogół odpowiedź jest przecząca. Wierzchołki różnych drzew mogą być uporządkowane w taki sam sposób.

PRZYKŁAD 3 Popatrzmy jeszcze raz na drzewo T pokazane na rysunku 7.10. Na rysunku 7.11 widzimy dwa następne drzewa binarne mające

ten sam porządek infiksowy wierzchołków. Drzewo binarne z rysunku 7.12(a) ma ten sam porządek prefiksowy wierzchołków, co drzewo T, a drzewo z rysunku 7.12(b) ma ten sam porządek postfiksowy wierzchołków, co T. Ćwiczenie 8 ma na celu sprawdzenie tych stwierdzeń. ∎



Rysunek 7.11



Rysunek 7.12

Pomimo to, co pokazały powyższe przykłady, istnieją ważne sytuacje, w których można odtworzyć drzewo, mając dany ciąg jego wierzchołków. Będzie to miało zastosowanie do notacji polskiej. Ponieważ wtedy łatwiej będzie uchwycić ideę tego postępowania, powrócimy do tej kwestii na końcu następnego paragrafu.

Możemy przeanalizować, jak długo działają te trzy algorytmy, korzystając z „metody obciążeń". Niech t_1 oznacza czas potrzebny do nadania wierzchołkowi etykiety i niech t_2 oznacza czas potrzebny do dołączenia jakiegoś ciągu wierzchołków do innego ciągu. (Możemy przyjąć, że ten czas potrzebny na połączenie

dwóch ciągów jest stały, jeśli będziemy używać list ze wskaźnikami jako naszych list wierzchołków). Pomysł polega teraz na tym, by całkowite czasy wykonywania różnych czynności w algorytmie traktować jako obciążenia przypisane wierzchołkom i krawędziom drzewa. „Obciążamy" wierzchołek v czasem t_1 potrzebnym do przypisania temu wierzchołkowi etykiety i „obciążamy" krawędź (v, w) czasem t_2 potrzebnym do dołączenia ciągu $L(w)$, utworzonego przy wywołaniu algorytmu dla wierzchołka w, tak aby stał się częścią ciągu $L(v)$. Wtedy czas wykonania każdej czynności staje się obciążeniem któregoś miejsca w drzewie oraz każdy wierzchołek i każda krawędź zostaną obciążone, przy czym żaden wierzchołek i żadna krawędź nie zostaną obciążone dwukrotnie. Dodając wszystkie obciążenia, otrzymujemy całkowity czas działania algorytmu, równy $t_1|V(T)| + t_2|E(T)|$, a więc będący $O(\max\{|V(T)|, |E(T)|\})$. Również $|E(T)| = |V(T)| - 1$ dla drzewa T, więc ten czas działania jest $O(|V(T)|)$.

Te trzy algorytmy są przykładami tzw. **algorytmów przeszukiwania w głąb** lub **poszukiwania z nawrotami**. Działanie takiego algorytmu polega na tym, że idziemy jak najdalej wzdłuż krawędzi drzewa w kierunku od korzenia, następnie powracamy kawałek, znów idziemy najdalej, jak się da i tak dalej. Ten sam pomysł może być wykorzystany w innych okolicznościach, w których nie mamy na początku do czynienia w oczywisty sposób z żadnym drzewem. Pokażemy, na czym polega ten pomysł na przykładzie algorytmu, który wierzchołkom grafu skierowanego G przypisuje w szczególny sposób etykiety będące liczbami całkowitymi. Zanim opiszemy ze szczegółami ten algorytm w przypadku ogólnym, popatrzymy, w jaki sposób mógłby on działać dla danego uporządkowanego drzewa z wyróżnionym korzeniem.

Ciąg wierzchołków utworzony przez algorytm POSTORDER daje naturalny sposób etykietowania wierzchołków drzewa; po prostu k-temu wierzchołkowi w tym ciągu dajemy etykietę k. To etykietowanie ma tę własność, że rodzice mają etykiety większe niż ich dzieci, a więc mają etykiety większe niż ich potomkowie. Przy takim etykietowaniu, nazywanym **etykietowaniem uporządkowanym**, liście mają małe etykiety, a korzeń ma największą etykietę. Można tak przerobić algorytm POSTORDER, by otrzymać algorytm ETYKIETOWANIE DRZEWA dający w wyniku etykietowanie uporządkowane.

Zanim napiszemy ten nowy algorytm, zastanówmy się jeszcze raz nad tym, w jaki sposób algorytm POSTORDER(r) ustawia wierzchołki w ciąg. Wierzchołki, które poprzedzają dany wierz-

chołek w w tym ciągu to nie tylko potomkowie wierzchołka w, ale także wszystkie wierzchołki znajdujące się na gałęziach drzewa zaetykietowane, zanim doszliśmy do wierzchołka w. Jeśli mamy nadać wierzchołkowi w i jego potomkom etykiety zgodne z ich miejscami w tym ciągu, to musimy wiedzieć, jakie etykiety zostały już wykorzystane, gdy dochodzimy do w. Jeśli zostały użyte etykiety $1, 2, \ldots, k$, to etykiety w poddrzewie z wyróżnionym korzeniem w powinny zacząć się od $k+1$. Aby pamiętać, jakie etykiety są dostępne, wśród danych w naszym algorytmie będziemy mieli jedną dodatkową zmienną mówiącą, od jakiej liczby zaczynamy etykietowanie i w wyniku algorytm zwróci ostatnią użytą etykietę. (Później, w algorytmie SORTOWANIE DRZEWA, zobaczymy, w jaki sposób można uniknąć takiego wyraźnego użycia tych zmiennych). Zapamiętajmy, algorytm ETYKIETOWANIE DRZEWA jest to po prostu algorytm POSTORDER, z tą tylko różnicą, że zamiast dołączać kolejne wierzchołki do ciągu, nadaje im etykiety.

Algorytm ETYKIETOWANIE DRZEWA(v, k; n)

{Dane: skończone drzewo z wyróżnionym korzeniem v, liczba całkowita $k \geq 0$}

{Wyniki: liczba całkowita n oraz etykietowanie uporządkowane tego drzewa etykietami $k+1, \ldots, n$}

$n := k$

Dla wszystkich dzieci w wierzchołka v (branych w pewnej ustalonej kolejności) wykonuj

  ETYKIETOWANIE DRZEWA(w, n; m) {etykietujemy potomków wierzchołka w etykietami $n+1, \ldots, m-1$, wierzchołek w otrzymuje etykietę m}

  $n := m$ {ustalamy punkt początkowy etykiet dla potomków następnego dziecka}

Zwiększ n o 1

Nadaj wierzchołkowi v etykietę n. ∎

Aby nadać etykiety wszystkim wierzchołkom drzewa z wyróżnionym korzeniem r, zaczynając od etykiety 1, wywołujemy ETYKIETOWANIE DRZEWA(r, 0;).

Zauważmy, że uporządkowanie dzieci nie jest istotne, jeśli potrzebne nam jest tylko etykietowanie uporządkowane. Zatem nie jest istotne, w jakiej kolejności wywołujemy tę procedurę dla kolejnych dzieci, jeśli tylko w wywołaniach dla dwóch różnych dzieci nie zostaną użyte te same etykiety.

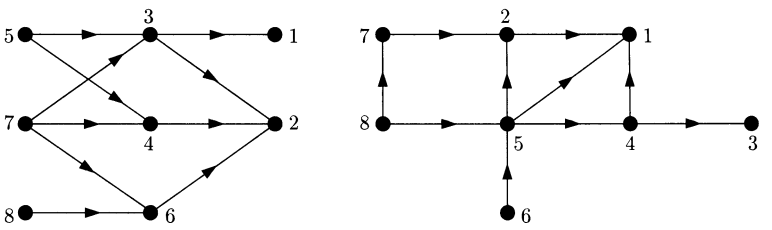
PRZYKŁAD 4 A oto, w jaki sposób algorytm ETYKIETOWANIE DRZEWA będzie działał dla naszego drzewa z przykładu 1.

ETYKIETOWANIE DRZEWA(r, 0;)
  ETYKIETOWANIE DRZEWA(v, 0;)
    ETYKIETOWANIE DRZEWA(u, 0;); wierzchołkowi u nadajemy etykietę **1**
    ETYKIETOWANIE DRZEWA(w, 1;)
      ETYKIETOWANIE DRZEWA(x, 1;); wierzchołkowi x nadajemy etykietę **2**
      ETYKIETOWANIE DRZEWA(y, 2;); wierzchołkowi y nadajemy etykietę **3**
      ETYKIETOWANIE DRZEWA(z, 3;); wierzchołkowi z nadajemy etykietę **4**
    wierzchołkowi w nadajemy etykietę **5**
  wierzchołkowi v nadajemy etykietę **6**
  ETYKIETOWANIE DRZEWA(s, 6;)
    ETYKIETOWANIE DRZEWA(p, 6;); wierzchołkowi p nadajemy etykietę **7**
    ETYKIETOWANIE DRZEWA(q, 7;); wierzchołkowi q nadajemy etykietę **8**
  wierzchołkowi s nadajemy etykietę **9**
wierzchołkowi r nadajemy etykietę **10**
Koniec ■

Etykietowania uporządkowane mogą być definiowane i rozważane dla dowolnego grafu skierowanego. Weźmy skończony graf skierowany G i podzbiór L zbioru $V(G)$. Etykietowaniem uporządkowanym zbioru L nazywamy taką numerację wierzchołków należących do L liczbami $1, 2, \ldots, |L|$, że jeśli tylko istnieje droga z wierzchołka i do wierzchołka j w grafie G, to $i > j$.

PRZYKŁAD 5 Na rysunku 7.13 widzimy dwa grafy skierowane, których zbiory wierzchołków mają etykietowania uporządkowane. Zauważmy, że jeśli $i > j$, to nie musi istnieć droga od wierzchołka i do wierzchołka j; wymagamy tylko, by $i > j$, jeśli taka droga istnieje. ■

Graf skierowany, którego zbiór wierzchołków ma etykietowanie uporządkowane, nie może mieć cykli, ponieważ gdyby wierzchołki cyklu miały etykiety $ij \ldots ki$, to mielibyśmy nierówności $i > j > \ldots > k > i$. Jeśli ograniczymy się do grafów skierowanych acyklicznych, to znaczy do takich grafów, które nie zawierają cykli, to zawsze będzie można znaleźć etykietowanie



Rysunek 7.13

uporządkowane zbioru $V(G)$. Udowodnimy to, przedstawiając algorytm ETYKIETOWANIE, który tworzy takie etykietowanie uporządkowane. Terminy „dziecko" i „potomek", których używaliśmy w odniesieniu do drzew, w ogólnym przypadku grafów skierowanych zostaną zastąpione terminami „bezpośredni następnik" i „wierzchołek dostępny", ale idea algorytmu pozostanie taka sama. Symbolem NAST(v) oznaczymy zbiór bezpośrednich następników wierzchołka v, czyli tych wierzchołków w, dla których istnieje krawędź od v do w. Symbolem DOST(v) oznaczymy zbiór tych wierzchołków w, dla których istnieje droga z v do w, łącznie z samym wierzchołkiem v. Ogólnie, jeśli $L \subseteq V(G)$, to zbiór DOST(L) będzie sumą zbiorów DOST(v) dla $v \in L$.

PRZYKŁAD 6 Zbiory bezpośrednich następników i wierzchołków dostępnych dla grafu z rysunku 7.13(a) są przedstawione w tablicy 7.3. ∎

Tablica 7.3

Wierzchołek	1	2	3	4	5	6	7	8
NAST( )	Ø	Ø	{1}	{2}	{3, 4}	{2}	{3, 4, 6}	{6}
DOST( )	{1}	{2}	{1, 2, 3}	{2, 4}	{1, 2, 3, 4, 5}	{2, 6}	{1, 2, 3, 4, 6, 7}	{2, 6, 8}

Istnieją dwa powody, dla których algorytm ETYKIETOWANIE będzie bardziej skomplikowany niż algorytm ETYKIETOWANIE DRZEWA. W drzewie z wyróżnionym korzeniem wszystkie wierzchołki są dostępne z korzenia, a więc algorytm może rozpocząć działanie w korzeniu i posuwać się w dół drzewa. Natomiast w acyklicznym grafie skierowanym może nie istnieć jeden taki wierzchołek, z którego dostępne są wszystkie inne wierzchołki. Przykładami są oba grafy skierowane przedstawione na rysunku 7.13. Zatem algorytm ETYKIETOWANIE będzie musiał korzystać z podprogramów, które etykietują zbiory takie jak DOST(v). Drugi powód to ten, że dzieci tego samego rodzica w drzewie z wyróżnionym korzeniem mają rozłączne zbiory

potomków, a więc żadne z tych dzieci nie będzie nigdy próbowało nadać etykiety potomkowi innego dziecka. Z drugiej strony, wierzchołek w acyklicznym grafie skierowanym może być dostępny z różnych bezpośrednich następników tego samego wierzchołka. Na przykład wierzchołek 1 na rysunku 7.13(b) jest dostępny z obu bezpośrednich następników wierzchołka 8. Zatem algorytm będzie musiał zwracać uwagę na to, które wierzchołki zostały już zaetykietowane, by uniknąć nadania im etykiety jeszcze raz.

Algorytm ETYKIETOWANIE wybiera najpierw wierzchołek v, nadaje etykiety wszystkim wierzchołkom dostępnym z v (włącznie z v), a następnie wywołuje sam siebie rekurencyjnie, by zająć się wierzchołkami, które nie mają jeszcze etykiet. W celu zwiększenia przejrzystości procedurę nadającą etykiety wierzchołkom dostępnym z wierzchołka v zapiszemy jako oddzielny algorytm, któremu damy nazwę SORTOWANIE DRZEWA z powodów, które wkrótce staną się jasne.

Algorytm SORTOWANIE DRZEWA również jest algorytmem rekurencyjnym. Jeśli zastosujemy algorytm SORTOWANIE DRZEWA do drzewa z wyróżnionym korzeniem i uruchomimy go w korzeniu r, to najpierw znajdzie on jakieś dziecko korzenia r i nada etykiety wszystkim potomkom tego dziecka, następnie rekurencyjnie nada etykiety pozostałym potomkom korzenia i wreszcie nada etykietę samemu korzeniowi r. Zatem w tym przypadku algorytm SORTOWANIE DRZEWA działa tak, jak algorytm ETYKIETOWANIE DRZEWA(r). Dzięki temu, że zapisaliśmy wywołanie rekurencyjne nieco inaczej, udało nam się uniknąć pewnych niezręczności związanych z użyciem dodatkowych zmiennych k i n w algorytmie ETYKIETOWANIE DRZEWA, ale kolejność przechodzenia przez wierzchołki jest dokładnie taka sama w obu algorytmach.

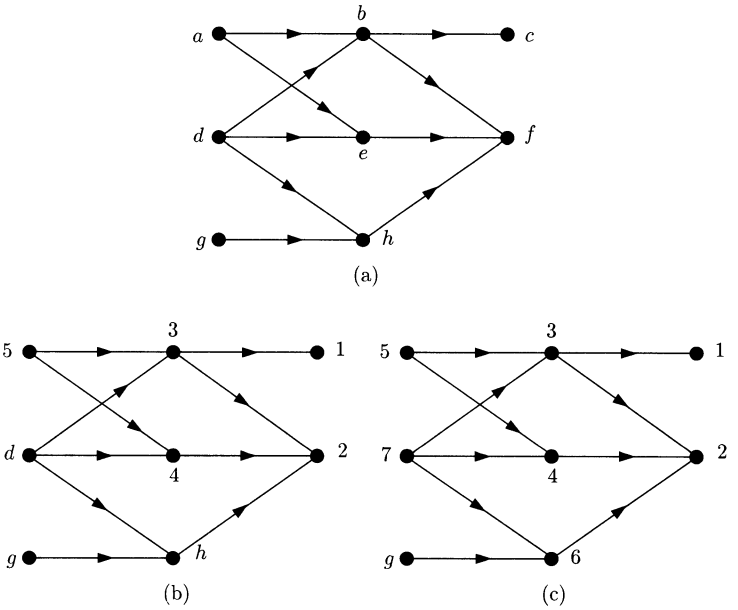
Kiedy któryś z algorytmów ETYKIETOWANIE lub SORTOWANIE DRZEWA jest wywoływany rekurencyjnie, niektóre wierzchołki mogą już mieć etykiety. Tę informację należy przekazać tym algorytmom, aby nie próbowały nadać etykiet tym wierzchołkom po raz drugi. Aby ułatwić sprawdzanie poprawności algorytmów, przyjmiemy jeszcze jedną definicję. Powiemy, że podzbiór L zbioru $V(G)$ jest **zaetykietowany**, jeśli ma on etykietowanie uporządkowane oraz $L = \text{DOST}(L)$. A oto pierwszy algorytm.

Algorytm SORTOWANIE DRZEWA (v, L)
{Dane: skończony graf acykliczny G, zaetykietowany podzbiór

L zbioru $V(G)$, wierzchołek $v \in V(G) \backslash L$}
{Wyniki: zaetykietowany zbiór $L \cup \text{DOST}(v)$, którego etykietowanie uporządkowane jest zgodne z danym etykietowaniem zbioru L}
Jeśli $\text{NAST}(v) \subseteq L$ {czyli $L \cup \text{DOST}(v) = L \cup \{v\}$}, to
zaetykietuj zbiór L tak jak przedtem i wierzchołkowi v nadaj etykietę $|L| + 1$
w przeciwnym przypadku
wybierz wierzchołek $w \in \text{NAST}(v) \backslash L$
SORTOWANIE DRZEWA (w, L)
{zbiór $L \cup \text{DOST}(w)$ zostaje zaetykietowany w sposób zgodny z etykietowaniem zbioru L}
SORTOWANIE DRZEWA (v, $L \cup \text{DOST}(w)$)
{zbiór $L \cup \text{DOST}(w) \cup \text{DOST}(v)$ zostaje zaetykietowany w sposób zgodny z etykietowaniem zbioru $L \cup \text{DOST}(w)$} ∎

Pierwsza instrukcja dotyczy przypadku początkowego, kiedy wszystkie wierzchołki w zbiorze $\text{DOST}(v)$, z wyjątkiem wierzchołka v, mają już etykiety. Zauważmy, że ten przypadek początkowy jest spełniony w próżni, jeśli wierzchołek v nie ma w ogóle bezpośrednich następników. Tak jak w algorytmie POSTORDER (gdzie bezpośrednie następniki nazywaliśmy dziećmi), część algorytmu znajdująca się po słowach „w przeciwnym przypadku" przenosi zadanie etykietowania na bezpośrednie następniki, z tą tylko komplikacją, że trzeba coś zrobić z wierzchołkami mającymi już etykiety. Po tym, jak pierwszy wybrany bezpośredni następnik w_1 i wierzchołki dostępne z w_1 dostały etykiety, wywołanie algorytmu SORTOWANIE DRZEWA(v, $L \cup \text{DOST}(w_1)$) prowadzi do jednej z dwóch możliwości. Jeśli wszystkie bezpośrednie następniki wierzchołka v są w zbiorze $L \cup \text{DOST}(w_1)$, to v jest jedynym wierzchołkiem zbioru $L \cup \text{DOST}(v)$, który nie ma etykiety, a więc otrzymuje ją. W przeciwnym przypadku zostaje wybrany kolejny bezpośredni następnik w_2 wierzchołka v; wierzchołek w_2 i wszystkie wierzchołki dostępne z w_2 i nie mające jeszcze etykiet, otrzymują je. I tak dalej.

PRZYKŁAD 7 Zilustrujemy działanie algorytmu SORTOWANIE DRZEWA na przykładzie grafu skierowanego G z rysunku 7.13(a), przerysowanego na rysunku 7.14(a) z wierzchołkami oznaczonymi literami. Jak zwykle, wierzchołki będziemy wybierać w kolejności alfabetycznej.



Rysunek 7.14

(a) Wywołamy najpierw algorytm SORTOWANIE DRZEWA(v, L) dla $v = a$ i $L = \emptyset$, by nadać etykiety wierzchołkom ze zbioru DOST(a).

Weź wierzchołek a, wybierz z kolei jego bezpośredni następnik b, wybierz jego bezpośredni następnik c i nadaj wierzchołkowi $\boldsymbol{c}$ etykietę **1**.

Powróć do b, wybierz jego bezpośredni następnik f i nadaj wierzchołkowi $\boldsymbol{f}$ etykietę **2**.

Powróć do b i nadaj wierzchołkowi $\boldsymbol{b}$ etykietę **3**.

Powróć do a, wybierz jego bezpośredni następnik e nie mający już bezpośrednich następników bez etykiet i nadaj wierzchołkowi $\boldsymbol{e}$ etykietę **4**.

Powróć do a i nadaj wierzchołkowi $\boldsymbol{a}$ etykietę **5**.

Na rysunku 7.14(b) widzimy etykietowanie zbioru DOST(a). Jeśli opisane powyżej wykonanie algorytmu jest dla ciebie jasne, to możesz przejść do punktu (b). Jeśli nie jest, to popatrz teraz na szczegółowe wyjaśnienie pokazujące kolejne wywołania podprogramów.

SORTOWANIE DRZEWA (a, $\emptyset$)

Wybierz bezpośredni następnik b wierzchołka a {gdyż wierzchołek a ma niezaetykietowane bezpośrednie następniki}

SORTOWANIE DRZEWA (b, $\emptyset$)
Wybierz bezpośredni następnik c wierzchołka b {gdyż wierzchołek b ma niezaetykietowane bezpośrednie następniki}
SORTOWANIE DRZEWA (c, $\emptyset$)
Nadaj wierzchołkowi $\boldsymbol{c}$ etykietę **1** {gdyż wierzchołek c nie ma żadnych bezpośrednich następników}
SORTOWANIE DRZEWA (b, $\{c\}$)
Wybierz bezpośredni następnik f wierzchołka b {gdyż wierzchołek f jest ciągle niezaetykietowany}
Nadaj wierzchołkowi $\boldsymbol{f}$ etykietę **2** {gdyż wierzchołek f nie ma bezpośrednich następników}
Nadaj wierzchołkowi $\boldsymbol{b}$ etykietę **3** {gdyż wierzchołek b nie ma już niezaetykietowanych bezpośrednich następników}
{Zbiór DOST(b) = $\{c, f, b\}$ jest już zaetykietowany}
SORTOWANIE DRZEWA (a, $\{c, f, b\}$)
Wybierz bezpośredni następnik e wierzchołka a {wierzchołek a ma jeszcze niezaetykietowany bezpośredni następnik}
SORTOWANIE DRZEWA (e, $\{c, f, b\}$)
Nadaj wierzchołkowi $\boldsymbol{e}$ etykietę **4** {wierzchołek f ma już etykietę}
SORTOWANIE DRZEWA (a, $\{c, f, b, e\}$)
Nadaj wierzchołkowi $\boldsymbol{a}$ etykietę **5** {wierzchołek a nie ma niezaetykietowanych bezpośrednich następników}

(b) Następnie wywołamy algorytm SORTOWANIE DRZEWA(v, L) dla $v = d$ i $L = \{a, b, c, e, f\}$, by nadać etykiety wierzchołkom ze zbioru $L \cup \text{DOST}(d)$. Popatrz na rysunek 7.14(b).

Weź wierzchołek d, wybierz jego bezpośredni następnik h nie mający etykiety i następnie nadaj wierzchołkowi $\boldsymbol{h}$ etykietę **6**, gdyż nie ma on bezpośrednich następników bez etykiet.
Powróć do d i nadaj wierzchołkowi $\boldsymbol{d}$ etykietę **7**, gdyż d nie ma już bezpośrednich następników bez etykiet.

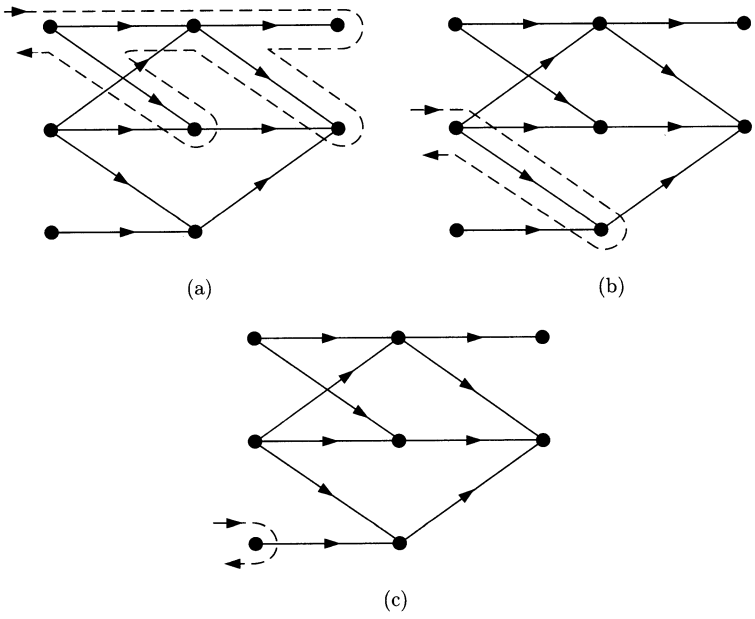
Na rysunku 7.14(c) pokazane jest etykietowanie zbioru $\{a, b, c, e, f\} \cup \text{DOST}(d)$.

(c) Wreszcie wywołujemy algorytm SORTOWANIE DRZEWA(v, L) dla $v = g$ i $L = \{a, b, c, d, e, f, h\}$, by nadać etykiety wierzchołkom ze zbioru $V(G) = L \cup \text{DOST}(g)$. Popatrz na rysunek 7.14(b). To będzie łatwe.

Weź wierzchołek $\boldsymbol{g}$ i nadaj mu etykietę **8**, gdyż nie ma on bezpośrednich następników bez etykiet.

Na rysunku 7.13(a) widzimy etykietowanie grafu G. Moglibyśmy oczywiście otrzymać inne etykietowania, gdybyśmy zdecydowali się na inną kolejność wybierania wierzchołków, niekoniecznie alfabetyczną (por. ćwiczenie 13). ∎

W przykładzie 7 algorytm dokonywał przeszukiwań w głąb, podążając wzdłuż dróg zaznaczonych przerywaną linią na rysunku 7.15. Każdy wierzchołek otrzymuje etykietę podczas ostatnich odwiedzin. Zbiór wierzchołków i krawędzi odwiedzanych w trakcie każdego przeszukiwania jest drzewem, o ile nie będziemy zwracać uwagi na zwroty krawędzi. Zatem ten algorytm sortuje pewne drzewo, które wybiera z grafu G; dlatego właśnie daliśmy mu nazwę SORTOWANIE DRZEWA. W paragrafie 6.6 widzieliśmy algorytm DRZEWO, który w podobny sposób budował drzewo spinające wewnątrz grafu spójnego.



Rysunek 7.15

Aby sprawdzić, że algorytm SORTOWANIE DRZEWA(v, L) we właściwy sposób etykietuje zbiór $L \cup \text{DOST}(v)$, musimy sprawdzić, że daje on właściwy wynik w przypadku początkowym, gdy $\text{NAST}(v) \subseteq L$ oraz że jeśli wywołania rekurencyjne SORTOWANIE DRZEWA(w, L) i SORTOWANIE DRZEWA(v, $L \cup \text{DOST}(w)$) dają właściwe wyniki, to również właściwy wynik daje wywołanie SORTOWANIE DRZEWA(v, L). Powinniśmy

również sprawdzić, że parametry w wywołaniach rekurencyjnych są poprawne i są bliższe warunków początkowych niż parametry (v, L).

Przypuśćmy więc, że $\text{DOST}(v) \subseteq L$. Każdy wierzchołek dostępny z wierzchołka v, z wyjątkiem samego v, należy do pewnego zbioru $\text{DOST}(w)$ dla jakiegoś bezpośredniego następnika w wierzchołka v. Ponieważ wszystkie takie wierzchołki w należą do zbioru L oraz $L = \text{DOST}(L)$, więc wszystkie wierzchołki dostępne z wierzchołka w też należą do L. A więc w tym przypadku $\text{DOST}(L \cup \{v\}) = \text{DOST}(L) \cup \text{DOST}(v) = L \cup \{v\}$. Ponieważ algorytm SORTOWANIE DRZEWA etykietuje zbiór L tak jak przedtem liczbami $1, \ldots, |L|$ i nadaje wierzchołkowi v etykietę $|L| + 1$, więc zbiór $L \cup \{v\}$ staje się rzeczywiście zbiorem zaetykietowanym.

Jeśli wywołania rekurencyjne dają wyniki poprawne, tzn. jeśli spełnione są warunki podane w komentarzach w treści algorytmu (w nawiasach klamrowych) po tych wywołaniach, to zbiór $L \cup \text{DOST}(v) = L \cup \text{DOST}(w) \cup \text{DOST}(v)$ jest zbiorem zaetykietowanym i jego etykietowanie jest zgodne z danym etykietowaniem zbioru L, a więc wynik działania algorytmu SORTOWANIE DRZEWA(v, L) jest poprawny.

Jako miarę oddalenia parametrów algorytmu od przypadków początkowych możemy przyjąć liczbę $|\text{DOST}(v) \backslash L|$ wierzchołków należących do zbioru $\text{DOST}(v)$ i nie mających jeszcze etykiety. Jeśli $w \in \text{NAST}(v) \backslash L$, to $\text{DOST}(w) \subset \text{DOST}(v)$ oraz $|\text{DOST}(v) \backslash L| > |\text{DOST}(w) \backslash L|$. Ponadto, $w \in \text{DOST}(v) \backslash L$, więc $|\text{DOST}(v) \backslash L| > |\text{DOST}(v) \backslash (L \cup \text{DOST}(w))|$. Zatem wywołania rekurencyjne przybliżają nas do przypadków początkowych (w tym miejscu korzystamy z acykliczności grafu). A więc algorytm SORTOWANIE DRZEWA działa poprawnie, czego należało dowieść.

Z chwilą, gdy zakończyliśmy przykład 7, cały acykliczny graf skierowany G został zaetykietowany. Stosowaliśmy wielokrotnie algorytm SORTOWANIE DRZEWA do wierzchołków nie mających etykiet dotąd, aż wszystkie otrzymały te etykiety, tzn. wykonaliśmy następujący algorytm ETYKIETOWANIE(L) dla $L = \emptyset$.

Algorytm ETYKIETOWANIE (G, L)

{Dane: skończony acykliczny graf skierowany G, zaetykietowany podzbiór L zbioru $V(G)$}

{Wyniki: zbiór $V(G)$ z etykietowaniem uporządkowanym,

które jest zgodne z etykietowaniem uporządkowanym zbioru L}
Jeśli $L = V(G)$, to
koniec
w przeciwnym przypadku
wybierz $v \in V(G) \backslash L$
SORTOWANIE DRZEWA (v, L)
{zbiór $L \cup$ DOST(v) zostaje zaetykietowany w sposób zgodny z etykietowaniem zbioru L}
ETYKIETOWANIE (G, $L \cup$ DOST(v))
{zaetykietuj to, co pozostało ze zbioru $V(G)$, zachowując etykiety w zbiorze $L \cup$ DOST(v)} ■

Sprawdzenie, że algorytm ETYKIETOWANIE działa poprawnie, jest podobne do dowodu poprawności algorytmu SORTOWANIE DRZEWA. Właściwą miarą oddalenia od przypadku początkowego $L = V(G)$ jest liczba $|V(G) \backslash L|$ wierzchołków nie mających jeszcze etykiet.

PRZYKŁAD 8 Jak już zauważyliśmy, jeśli wykonamy algorytm ETYKIETOWANIE dla grafu z rysunku 7.14(a) i dla $L = \emptyset$, to otrzymamy etykietowanie uporządkowane pokazane na rysunku 7.13(a). Podsumujmy.

Wybierz a
Zaetykietuj zbiór DOST(a) $= \{c, b, f, e, a\}$ za pomocą algorytmu SORTOWANIE DRZEWA (a, $\emptyset$) {por. przykład 7(a)}
Wywołaj rekurencyjnie dla zbioru $L = \{c, b, f, e, a\}$
Wybierz d
Zaetykietuj zbiór $L \cup$ DOST(d) $= L \cup \{h, d\}$ za pomocą algorytmu SORTOWANIE DRZEWA (d, L) {por. przykład 7(b)}
Wywołaj rekurencyjnie dla zbioru $L = \{c, b, f, e, a, h, d\}$
Wybierz g
Zaetykietuj zbiór $L \cup$ DOST(g) $= L \cup \{g\} = V(G)$ {por. przykład 7(c)}
Wywołaj rekurencyjnie dla zbioru $L = V(G)$
Koniec ■

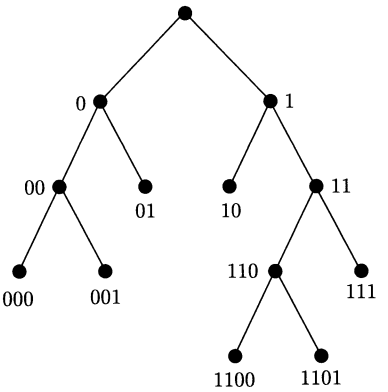
Wprawdzie wszystkie grafy w naszych przykładach były spójne, ale algorytmy SORTOWANIE DRZEWA i ETYKIETOWANIE działałyby równie dobrze dla grafów skierowanych mających więcej niż jedną składową.

Korzystając z metody obciążeń, możemy otrzymać oszacowanie czasu działania algorytmu SORTOWANIE DRZEWA. Obciążamy każdy wierzchołek czasem potrzebnym do nadania mu etykiety. Krawędzie obciążamy na jeden z następujących dwóch sposobów. Jeśli krawędź (u, w) nie została wybrana, gdyż wierzchołek w był już w zbiorze L, gdy dokonywaliśmy wyboru, to obciążamy tę krawędź czasem potrzebnym do usunięcia wierzchołka w z listy możliwych do wyboru wierzchołków należących do zbioru NAST(u). Jeśli wierzchołek w został wybrany ze zbioru NAST(u), to krawędź (u, w) obciążamy czasem potrzebnym do przejścia do wierzchołka w, by dokonać następnego wywołania rekurencyjnego, i powrotu do u po zakończeniu tego wywołania. Na przykład w grafie pokazanym na rysunku 7.14(a) krawędź (e, f) nie została wybrana, tzn. algorytm SORTOWANIE DRZEWA nie przechodził od wierzchołka e do wierzchołka f, gdyż wierzchołek f miał już etykietę, gdy rozpatrywany był wierzchołek e. Z drugiej strony, widzimy, że przerywana linia przechodzi w obie strony przez krawędź (a, e); w jedną stronę po to, by wywołać algorytm SORTOWANIE DRZEWA w wierzchołku e i z powrotem po to, by przekazać otrzymane wyniki wywołaniu algorytmu SORTOWANIE DRZEWA w wierzchołku a.

Ponieważ algorytm ETYKIETOWANIE tylko wywołuje algorytm SORTOWANIE DRZEWA i uaktualnia zbiór L, więc te same obciążenia mogą być zastosowane także do algorytmu ETYKIETOWANIE i pokażą wtedy, że czas potrzebny do zaetykietowania całego zbioru $V(G)$, gdy zaczynamy od $L = \emptyset$, jest rzędu $O(|V(G)| + |E(G)|)$. Można mieć wątpliwości i powiedzieć, że ponieważ wykorzystywane były tylko krawędzie (u, w) dla $w \in$ NAST$(V)\backslash L$, więc tylko one zabierają czas. Ale tak naprawdę każda krawędź grafu G musiała być rozważana chociaż przez chwilę, by stwierdzić, czy powinna być wykorzystana, czy nie.

ĆWICZENIA DO § 7.3

1. Zastosuj pomysł z rysowaniem „przerywanej linii" takiej jak na rysunku 7.9, by utworzyć porządki prefiksowy i postfiksowy wierzchołków drzewa przedstawionego na rysunku 7.11(a).
2. Powtórz ćwiczenie 1 dla rysunku 7.12(a).
3. Powtórz ćwiczenie 1 dla rysunku 7.12(b).
4. Wyznacz porządek infiksowy wierzchołków drzewa z etykietami pokazanego na rysunku 7.16.



Rysunek 7.16

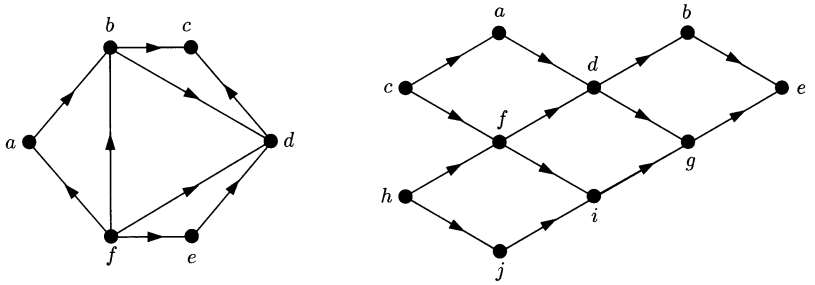
5. (a) Zastosuj algorytm PREORDER, by utworzyć ciąg wierzchołków drzewa T pokazanego na rysunku 7.10. Wyznacz ciągi $L(w)$ i $L(u)$ utworzone przez ten algorytm.
(b) Powtórz ćwiczenie (a) dla algorytmu POSTORDER.

6. Wypisz zbiory bezpośrednich następników i wierzchołków dostępnych dla grafu z rysunku 7.13(b); zrób to tak, jak w tablicy 7.3.

7. (a) Za pomocą algorytmu POSTORDER utwórz ciąg wierzchołków drzewa z rysunku 7.11(b).
(b) Powtórz ćwiczenie (a) dla algorytmu PREORDER.
(c) Powtórz ćwiczenie (a) dla algorytmu INORDER.

8. Udowodnij stwierdzenia wypowiedziane w przykładzie 3.

9. (a) Zastosuj algorytm SORTOWANIE DRZEWA(a, $\emptyset$) do grafu skierowanego pokazanego na rysunku 7.17(a). Bezpośrednie następniki wierzchołków wybieraj w kolejności alfabetycznej. Narysuj linię przerywaną pokazującą sposób przeszukiwania i zaznacz etykiety wierzchołków, tak jak na rysunku 7.14(c).



(a) (b)

Rysunek 7.17

(b) Wypisz kroki algorytmu z ćwiczenia (a) w mniej szczegółowy sposób z przykładu 7(a).

(c) Zastosuj algorytm ETYKIETOWANIE do grafu skierowanego pokazanego na rysunku 7.17(a) i zaznacz na rysunku otrzymane etykiety wierzchołków.

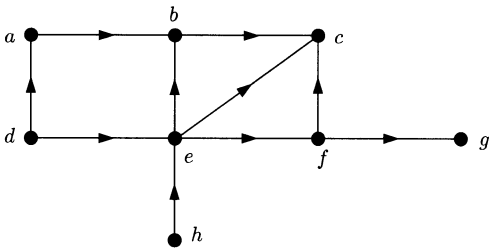
10. (a) Powtórz ćwiczenie 9(a) dla grafu skierowanego z rysunku 7.17(a) i algorytmu SORTOWANIE DRZEWA(a, $\emptyset$).

(b) Powtórz ćwiczenie 9(c) dla tego grafu skierowanego.

11. Zastosuj algorytm SORTOWANIE DRZEWA(v, L) do grafu skierowanego pokazanego na rysunku 7.18, gdzie

(a) $v = a$, $L = \emptyset$. (b) $v = d$, $L = \{a, b, c\}$.

(c) $v = h$, $L = \{a, b, c, d, e, f, g\}$.

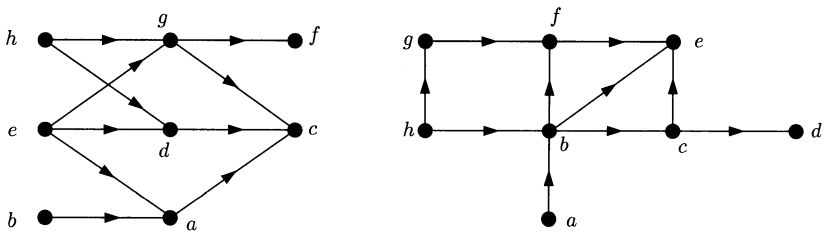


Rysunek 7.18

12. Skorzystaj z ćwiczenia 11 i algorytmu ETYKIETOWANIE, by utworzyć etykietowanie uporządkowane grafu skierowanego z rysunku 7.18.

13. (a) Graf skierowany z rysunku 7.14(a) jest ponownie przedstawiony na rysunku 7.19(a) z innymi oznaczeniami wierzchołków. Użyj algorytmu ETYKIETOWANIE, wybierając wierzchołki w kolejności alfabetycznej, by utworzyć etykietowanie uporządkowane tego grafu.

(b) Powtórz ćwiczenie (a) dla grafu skierowanego z rysunku 7.19(b).



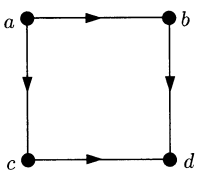
Rysunek 7.19

14. Algorytm SORTOWANIE DRZEWA przypomina algorytm POSTORDER w tym, że nadaje wierzchołkowi etykietę podczas ostatnich odwiedzin. Oto algorytm SORTOWANIE KRZAKA, który ma naśladować algorytm PREORDER.

Algorytm SORTOWANIE KRZAKA(v, L)

{Dane: skończony graf acykliczny G, podzbiór L zbioru $V(G)$ zaetykietowany w dowolny sposób, wierzchołek $v \in V(G) \backslash L$}
{Wyniki: zbiór $L \cup \text{DOST}(v)$ zaetykietowany w sposób zgodny z etykietowaniem zbioru L}
Nadaj etykietę $|L| + 1$ wierzchołkowi v {gdy po raz pierwszy znajdujemy się w wierzchołku v}
Zastąp zbiór L zbiorem $L \cup \{v\}$
Dla wierzchołków $w \in \text{NAST}(v) \backslash L$ {następniki wierzchołka v brane w dowolnej kolejności}
  SORTOWANIE KRZAKA(w, L)
  zastąp zbiór L zbiorem $L \cup \text{DOST}(w)$

(a) Wykorzystaj algorytm SORTOWANIE KRZAKA$(a, \emptyset)$ do zaetykietowania grafu z rys. 7.20.



Rysunek 7.20

(b) Czy etykietowanie otrzymane w ćwiczeniu (a) jest etykietowaniem uporządkowanym tego grafu skierowanego? Czy jest to etykietowanie odwrotnie uporządkowane, tzn. czy wierzchołki mają liczby mniejsze niż ich bezpośrednie następniki? Odpowiedź uzasadnij.

15. (a) Jakie są przypadki początkowe dla algorytmu ETYKIETOWANIE DRZEWA?
(b) Jaki otrzymamy wynik n, gdy wywołamy algorytm ETYKIETOWANIE DRZEWA$(v, k;\ )$ dla przypadku początkowego?
(c) Podaj miarę odległości od przypadku początkowego dla algorytmu ETYKIETOWANIE DRZEWA i pokaż, że wywołanie rekurencyjne z parametrami (w, n) jest bliższe przypadków początkowych, niż wywołanie z parametrami (v, k), które je wywołuje.

16. (a) Podaj przykład acyklicznego grafu skierowanego bez krawędzi wielokrotnych mającego 4 wierzchołki i $4(4-1)/2 = 6$ krawędzi. Ile różnych etykietowań uporządkowanych ma twój graf?

(b) Pokaż, że acykliczny graf skierowany bez krawędzi wielokrotnych, mający n wierzchołków i $n(n-1)/2$ krawędzi, ma dokładnie jedno etykietowanie uporządkowane.

17. Pokaż, że acykliczny graf skierowany bez krawędzi wielokrotnych, mający n wierzchołków, ma nie więcej niż $n(n-1)/2$ krawędzi. (Zatem algorytm ETYKIETOWANIE działa w czasie nie dłuższym niż $O(n^2)$.)

§ 7.4. Notacja polska

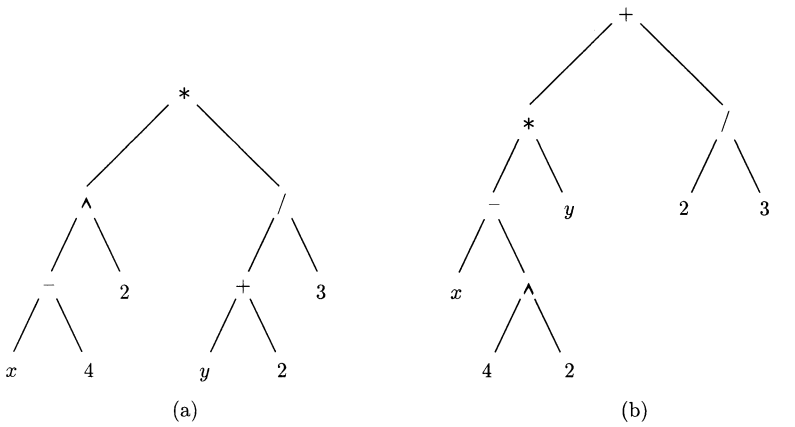
Porządki prefiksowy, postfiksowy i infiksowy pozwalają utworzyć listę wierzchołków uporządkowanego drzewa z wyróżnionym korzeniem. Jeśli te wierzchołki mają etykiety, które mogą być liczbami, znakami dodawania, znakami mnożenia itp., to takiej liście można nadać pewne znaczenie. Na przykład, jeśli używamy zwykłej symboliki algebraicznej, to liście $4 * 3$: 2 przypisuje się liczbę 6. Lista $4 + 3 * 2$ wydaje się niejednoznaczna; czy oznacza ona liczbę 14 czy 10? Notacja polska, którą opiszemy poniżej, jest to metoda tworzenia wyrażeń algebraicznych bez użycia nawiasów, korzystająca z list otrzymanych z drzew. Ważne jest przy tym to, że te listy w pełni określają odpowiadające im drzewa etykietowane i związane z nimi wyrażenia. Po omówieniu notacji polskiej udowodnimy, że przy całkiem ogólnych założeniach te listy rzeczywiście wyznaczają jednoznacznie drzewa.

Notacja polska może być używana do zapisywania wyrażeń, w których występują obiekty pewnego systemu (liczby, macierze, zdania w rachunku zdań itp.) oraz pewne działania na tych obiektach. Działania te zazwyczaj, choć nie zawsze, są **dwuargumentowe**, to znaczy takie, które wykonuje się na dwóch obiektach, lub **jednoargumentowe**, tzn. takie, które wykonuje się tylko na jednym obiekcie. Przykładami działań dwuargumentowych są $+$, $*$, $\wedge$, $\rightarrow$; działaniem jednoargumentowym jest $\neg$. Liście uporządkowanych drzew z wyróżnionym korzeniem odpowiadających wyrażeniom będą miały etykiety, które mogą być obiektami tego systemu (takimi jak liczby) lub zmiennymi reprezentującymi obiekty tego systemu (takimi jak x). Etykietami innych wierzchołków będą symbole działań.

PRZYKŁAD 1 Wyrażenie algebraiczne

$$((x-4)\hat{}2) * ((y+2)/3)$$

jest reprezentowane przez drzewo pokazane na rysunku 7.21(a). W tym wyrażeniu występuje wiele znanych działań dwuargumentowych w zbiorze $\mathbb{R}$: $+$, $-$, $*$, $/$, $\hat{}$. Przypomnijmy, że $*$ oznacza



Rysunek 7.21

mnożenie, a ^ oznacza potęgowanie: $a\hat{}b$ oznacza a^b. Nasze wyrażenie jest zatem równoważne z wyrażeniem

$$(x-4)^2\left(\frac{y+2}{3}\right).$$

Zauważmy, że to drzewo jest drzewem *uporządkowanym*; gdybyśmy na przykład zamienili x i 4, to drzewo opisywałoby inne wyrażenie algebraiczne.

Jasne jest, że drzewo uporządkowane z wyróżnionym korzeniem wyznacza wyrażenie algebraiczne. Zauważmy, że porządek infiksowy wierzchołków daje wyrażenie $x-4\hat{}2*y+2/3$, czyli dokładnie nasze wyjściowe wyrażenie, z tym tylko, że bez nawiasów. Co więcej, nawiasy są konieczne, gdyż to ostatnie wyrażenie nie określa jednoznacznie ani drzewa, ani początkowego wyrażenia algebraicznego. Tę samą listę moglibyśmy równie dobrze otrzymać z wyrażenia algebraicznego

$$((x-(4\hat{}2))*y)+(2/3),$$

któremu odpowiada drzewo pokazane na rysunku 7.21(b). To wyrażenie algebraiczne jest równoważne z wyrażeniem $(x-16)y+\frac{2}{3}$, różniącym się znacznie od

$$(x-4)^2\left(\frac{y+2}{3}\right).$$

Powróćmy do naszego wyjściowego wyrażenia algebraicznego, któremu odpowiada drzewo z rysunku 7.21(a). Porządek prefiksowy daje

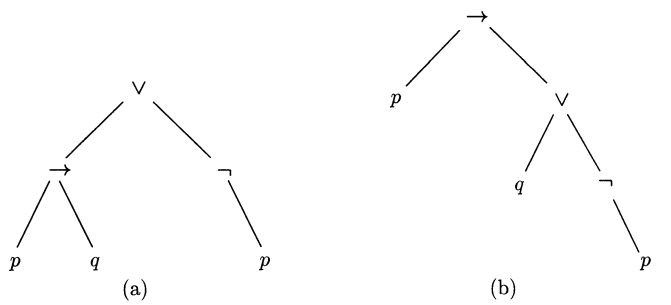
$$*\hat{}-x42/+y23,$$

a uporządkowanie postfiksowe daje

$$x4 - 2\hat{}y2 + 3/*.$$

Okazuje się, że każda z tych dwóch ostatnich list wyznacza jednoznacznie drzewo, a więc i wyjściowe wyrażenie algebraiczne. Zatem te wyrażenia są jednoznaczne bez nawiasów. Tej niezwykle użytecznej obserwacji dokonał polski logik Łukasiewicz. Listę uporządkowaną prefiksowo nazywamy **wyrażeniem** zapisanym **w notacji polskiej** lub **notacji prefiksowej**. Listę uporządkowaną postfiksowo nazywamy wyrażeniem zapisanym **w odwrotnej notacji polskiej** lub **w notacji postfiksowej**. Naszą zwykłą notację algebraiczną wraz z koniecznymi nawiasami nazywamy **notacją infiksową**. ∎

PRZYKŁAD 2 Weźmy zdanie złożone $(p \to q) \vee (\neg p)$. Traktujemy działania dwuargumentowe $\to$ i $\vee$ tak jak przedtem. Natomiast $\neg$ jest działaniem jednoargumentowym. Przyjmujemy, że jego dziecko jest prawym dzieckiem, gdyż znak działania poprzedza zdanie, którego dotyczy. Odpowiednie drzewo binarne przedstawione na rysunku 7.22(a) można przejść na wszystkie trzy sposoby. Porządek prefiksowy daje $\vee \to pq\neg p$, a postfiksowy daje $pq \to p\neg\vee$. Porządek infiksowy daje wyjściowe wyrażenie, z którego usunięto nawiasy. Inne drzewo, mające ten sam porządek infiksowy, jest pokazane na rysunku 7.22(b). Tak jak w przykładzie 1 porządki prefiksowy i postfiksowy wyznaczają jednoznacznie drzewo i wyjściowe zdanie złożone. ∎



Rysunek 7.22

Jak już pokazaliśmy to w przykładzie 3 w § 7.3, na ogół prefiksowy lub postfiksowy porządek wierzchołków nie wyznacza drzewa, nawet jeśli jest to regularne drzewo binarne. Potrzebne

są dodatkowe informacje. Okazuje się, że jeśli znamy poziom każdego wierzchołka, to porządek prefiksowy lub postfiksowy wyznacza drzewo. Nie będziemy tym się zajmować. Okazuje się również, że jeśli wiemy, ile dzieci ma każdy wierzchołek, to drzewo jest wyznaczone jednoznacznie przez porządek prefiksowy lub postfiksowy.

PRZYKŁAD 3 Zilustrujemy ostatnie zdanie na przykładzie wyrażenia

$$x4 - 2\hat{}y2 + 3/*,$$

zapisanego w notacji postfiksowej. Każdy wierzchołek z etykietą $*$, $\hat{}$, $-$, / lub + odpowiada działaniu dwuargumentowemu, a więc ma dwoje dzieci. Pozostałe wierzchołki nie mają dzieci, tak więc wiemy dokładnie, ile dzieci ma każdy wierzchołek.

Odtwórzmy drzewo, ale zamiast rysować poddrzewa, będziemy wyznaczać odpowiadające im podwyrażenia. Aby zrekonstruować drzewo, przypomnijmy, że na liście w porządku postfiksowym każdy wierzchołek jest poprzedzony bezpośrednio listami poddrzew, których korzeniami są dzieci tego wierzchołka. Naszym zadaniem jest rozpoznanie list tych poddrzew w danym wyrażeniu.

Zaczynając od lewej strony, pierwsze działanie dwuargumentowe $-$ musi mieć dwoje dzieci, mianowicie liście x i 4. Poddrzewo, którego korzeniem jest $-$, ma listę uporządkowaną postfiksowo postaci $x4-$, odpowiadającą zwykłemu (w porządku infiksowym) wyrażeniu $x-4$. Zastępujemy listę $x4-$ wyrażeniem $(x-4)$ i otrzymujemy zmodyfikowany ciąg $(x-4)2\hat{}y2+3/*$. Następne działanie dwuargumentowe $\hat{}$ ma dwa poddrzewa przyczepione do niego. Jednym jest poddrzewo $(x-4)$, które przed chwilą znaleźliśmy, drugim jest liść 2. Odpowiadającym mu wyrażeniem (infiksowym) jest $((x-4)\hat{}2)$, daje ono nowy ciąg $((x-4)\hat{}2)y2+3/*$. Argumentami następnego działania + są bezpośrednio poprzedzające je poddrzewa, mianowicie liście y i 2. Daje ono poddrzewo $(y+2)$ i odpowiedni zmodyfikowany ciąg $((x-4)\hat{}2)(y+2)3/*$. Argumentami działania / są jego dwa bezpośrednie poprzedniki $(y+2)$ i 3, otrzymujemy więc $((y+2)/3)$ i $((x-4)\hat{}2)((y+2)/3)*$. Wreszcie mamy działanie $*$, poprzedzone dwoma wyrażeniami reprezentującymi poddrzewa, a więc otrzymujemy

$$((x-4)\hat{}2)*((y+2)/3).$$

Odtworzyliśmy więc wyrażenie, którego drzewo jest pokazane na rysunku 7.21(a).

Podsumujmy krótko procedurę z ostatniego akapitu:

$$x4 - 2\hat{}y2 + 3/*$$
$$(x - 4)2\hat{}y2 + 3/*$$
$$((x - 4)\hat{}2)y2 + 3/*$$
$$((x - 4)\hat{}2)(y + 2)3/*$$
$$((x - 4)\hat{}2)((y + 2)/3)*$$
$$((x - 4)\hat{}2) * ((y + 2)/3).$$

Można przedstawić tę procedurę w inny sposób, który posłuży jako ilustracja dowodu twierdzenia, które podamy po przykładzie 6. Przesuwając się z lewa na prawo, wybieramy dzieci każdego wierzchołka w następujący sposób: każde działanie dwuargumentowe dostaje dwoje najbliższych osiągalnych dzieci ze stosu dzieci na lewo od niego. Dziecko jest osiągalne, jeśli nie zostało dotychczas przydzielone jakiemuś rodzicowi. Pozostałe wierzchołki nie otrzymują dzieci. Widzimy to w tablicy 7.4.

Tablica 7.4

Wierzchołek	x	4	$-$	2	$\hat{}$	y	2	$+$	3	$/$	$*$
Dzieci osiągalne		x	x, 4	$-$	$-$, 2	$\hat{}$	$\hat{}$, y	$\hat{}$, y, 2	$\hat{}$, $+$	$\hat{}$, $+$, 3	$\hat{}$, $/$
Dzieci przypisane			x, 4		$-$, 2			y, 2		$+$, 3	$\hat{}$, $/$

■

PRZYKŁAD 4 Można użyć tej samej metody dla zdań złożonych zapisanych w odwrotnej notacji polskiej, ale kiedy natkniemy się na działanie jednoargumentowe $\neg$, to jego argumentem będzie bezpośrednio poprzedzające je podwyrażenie. Na przykład

$$pq \to pq \wedge \neg\vee$$
$$(p \to q)pq \wedge \neg\vee$$
$$(p \to q)(p \wedge q)\neg \vee \quad \text{[argumentem } \neg \text{ jest } (p \wedge q)\text{]}$$
$$(p \to q)(\neg(p \wedge q))\vee$$
$$(p \to q) \vee (\neg(p \wedge q)).$$

Czytelnik może narysować drzewo odpowiadające temu zdaniu złożonemu. Możemy również wybierać dzieci, przesuwając się z lewa na prawo, jak w przykładzie 3. Pokazuje to tablica 7.5.

Tablica 7.5

Wierzchołek	p	q	$\rightarrow$	p	q	$\wedge$	$\neg$	$\vee$
Dzieci osiągalne		p	p, q	$\rightarrow$	$\rightarrow, p$	$\rightarrow, p, q$	$\rightarrow, \wedge$	$\rightarrow, \neg,$
Dzieci przypisane			p, q			p, q	$\wedge$	$\rightarrow, \neg$

■

Nie wszystkie ciągi działań i innych symboli dają w wyniku wyrażenia sensowne.

PRZYKŁAD 5 Przypuśćmy, że chcemy sprawdzić, czy wyrażenia

$$y + 2x *^{\wedge} 4 \quad \text{oraz} \quad q\neg pq \vee \wedge \rightarrow$$

są zapisane poprawnie w odwrotnej notacji polskiej. Natychmiast okazuje się, że pierwsze z tych wyrażeń jest złe, gdyż znak + nie jest poprzedzony dwoma wyrażeniami. Niepoprawność drugiego wychodzi na jaw, gdy próbujemy je odkodować (tzn. dokonać jego analizy składniowej), tak jak w przykładzie 4:

$$\begin{aligned} &q\neg pq \vee \wedge \rightarrow \\ &(\neg q)pq \vee \wedge \rightarrow \\ &(\neg q)(p \vee q)\wedge \rightarrow \\ &((\neg q) \wedge (p \vee q)) \rightarrow . \end{aligned}$$

Niestety znak działania $\rightarrow$ jest poprzedzony tylko jednym podwyrażeniem. Wynika stąd, że żaden z powyższych ciągów symboli nie jest wyrażeniem sensownym. ■

Dokładnie tak samo, jak to zrobiliśmy w § 7.2 ze zwykłymi wyrażeniami algebraicznymi, możemy zdefiniować rekurencyjnie **wyrażenia poprawnie zbudowane** dla **notacji polskiej** i dla **odwrotnej notacji polskiej**. Pokażemy to na przykładzie definicji algebraicznych wyrażeń poprawnie zbudowanych w odwrotnej notacji polskiej:

(P) Stałe liczbowe i zmienne są wyrażeniami poprawnie zbudowanymi;

(R) Jeśli f i g są wyrażeniami poprawnie zbudowanymi, to $fg+$, $fg-$, $fg*$, $fg/$ i $fg^{\wedge}$ są wyrażeniami poprawnie zbudowanymi.

PRZYKŁAD 6 Pokażemy, że $x2^{\wedge}y - xy * /$ jest wyrażeniem poprawnie zbudowanym. Z warunku (P) wynika, że wszystkie stałe i zmienne są

wyrażeniami poprawnie zbudowanymi. Zatem na mocy (R) wyrażenie $x2\hat{}$ jest poprawnie zbudowane. Stąd wynika, że $x2\hat{}y-$ jest wyrażeniem poprawnie zbudowanym, gdyż możemy zastosować warunek (R) do wyrażeń $f = x2\hat{}$ i $g = y$. Podobnie z warunku (R) wynika, że $xy*$ jest wyrażeniem poprawnie zbudowanym. Wreszcie całe nasze wyrażenie jest poprawnie zbudowane na mocy warunku (R), gdyż ma postać $fg/$, gdzie $f = x2\hat{}y-$ i $g = xy*$. ∎

Zakończymy ten paragraf dowodem twierdzenia mówiącego, że wyrażenia w notacji polskiej i w odwrotnej notacji polskiej wyznaczają jednoznacznie pierwotne wyrażenie. Dowód jest oparty na drugim z algorytmów pokazanych w przykładach 3 i 4.

Twierdzenie

Niech T będzie skończonym uporządkowanym drzewem z wyróżnionym korzeniem, którego wierzchołki zostały wypisane w porządku prefiksowym lub postfiksowym. Przypuśćmy, że znana jest liczba dzieci każdego wierzchołka. Wtedy to drzewo jest wyznaczone jednoznacznie, to znaczy, że to drzewo może być odtworzone z ciągu wierzchołków.

Dowód. Rozważymy tylko przypadek porządku postfiksowego. Mamy dany ciąg $v_1v_2 \ldots v_n$ wierzchołków drzewa T w porządku postfiksowym oraz mamy dany ciąg $c_1 \ldots c_n$ liczb dzieci wierzchołków $v_1, \ldots, v_n$. Wykażemy, że dla każdego wierzchołka v_m drzewa T zbiór $S(v_m)$ dzieci tego wierzchołka oraz porządek tego zbioru są wyznaczone jednoznacznie.

Weźmy pewien wierzchołek v_m. Wtedy v_m jest korzeniem poddrzewa T_m złożonego z v_m i jego potomków. Kiedy algorytm POSTORDER dołączał wierzchołek v_m do listy wierzchołków, wierzchołki poddrzew drzewa T_m były już umieszczone na tej liście i listy ich wierzchołków bezpośrednio poprzedzały v_m w kolejności wyznaczonej przez porządek zbioru $S(v_m)$. Ponadto ten algorytm nie umieszczał późniejszych danych wewnątrz tej listy wierzchołków drzewa T_m, a więc lista wierzchołków T_m występuje w całej liście $v_1v_2 \ldots v_n$ wierzchołków drzewa T jako nieprzerwany ciąg zakończony wierzchołkiem v_m.

Ponieważ wierzchołek v_1 nie ma poprzedników, więc v_1 jest liściem. Zatem $S(v_1) = \emptyset$ i porządek tego zbioru jest wyznaczony jednoznacznie „w próżni". Załóżmy indukcyjnie, że dla każdej liczby k takiej, że $k < m$ zbiór $S(v_k)$ i jego porządek są wyznaczone jednoznacznie oraz weźmy wierzchołek v_m. (Korzystamy

tu z drugiej zasady indukcji w zbiorze skończonym $\{1, 2, \ldots, n\}$). Zatem zbiór $U(m) = \{v_k\colon\ k < m\} \setminus \bigcup_{k<m} S(v_k)$ jest wyznaczony jednoznacznie. Składa się on z tych wierzchołków v_k występujących na lewo od v_m, których rodzice nie są na lewo od v_m. Dzieci wierzchołka v_m są tymi elementami zbioru $U(m)$, które znajdują się w drzewie T_m. Ponieważ lista wierzchołków drzewa T_m znajduje się bezpośrednio na lewo od wierzchołka v_m, więc dzieci wierzchołka v_m to ostatnie c_m elementów z prawej strony zbioru $U(m)$. Ponieważ zbiór $U(m)$ jest wyznaczony jednoznacznie i znamy liczbę c_m, więc zbiór $S(v_m)$ dzieci wierzchołka v_m jest wyznaczony jednoznacznie. Co więcej, jego porządek pokrywa się z porządkiem występowania na liście $v_1 v_2 \ldots v_n$.

Z zasady indukcji wynika, że każdy zbiór uporządkowany $S(v_m)$ dzieci wierzchołka v_m jest wyznaczony przez listę wierzchołków uporządkowaną postfiksowo i przez ciąg $c_1, c_2, \ldots, c_n$. Korzeniem drzewa jest oczywiście ostatni wierzchołek v_n. Zatem cała struktura uporządkowanego drzewa z wyróżnionym korzeniem jest wyznaczona jednoznacznie. ∎

PRZYKŁAD 7 Lista wierzchołków $uxyzwvpqsr$ i ciąg $0, 0, 0, 0, 3, 2, 0, 0, 2, 2$ wyznaczają zbiory podane w tablicy 7.6.

Tablica 7.6

v_k	u	x	y	z	w	v	p	q	s	r
$U(k)$	$\emptyset$	$\{u\}$	$\{u, x\}$	$\{u, x, y\}$	$\{u, x, y, z\}$	$\{u, w\}$	$\{v\}$	$\{v, p\}$	$\{v, p, q\}$	$\{v, s\}$
$S(v_k)$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	$\{x, y, z\}$	$\{u, w\}$	$\emptyset$	$\emptyset$	$\{p, q\}$	$\{v, s\}$

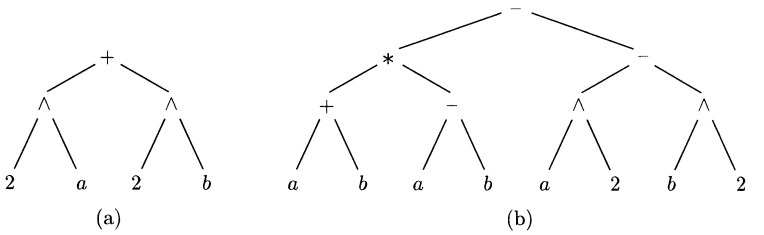
Z uporządkowanych zbiorów $S(v_k)$ można rekurencyjnie utworzyć drzewo T przedstawione na rysunku 7.8 w § 7.3. ∎

ĆWICZENIA DO § 7.4

1. Napisz wyrażenie algebraiczne podane na rysunku 7.21(b) w odwrotnej notacji polskiej i w notacji polskiej.

2. Dla uporządkowanego drzewa z wyróżnionym korzeniem, przedstawionego na rysunku 7.23(a) napisz odpowiadające mu wyrażenie algebraiczne w odwrotnej notacji polskiej oraz w zwykłej algebraicznej notacji infiksowej.

3. (a) Dla uporządkowanego drzewa z wyróżnionym korzeniem, przedstawionego na rysunku 7.23(b) napisz odpowiadające mu wyrażenie algebraiczne w odwrotnej notacji polskiej oraz w zwykłej algebraicznej notacji infiksowej.



Rysunek 7.23

(b) Uprość wyrażenie algebraiczne otrzymane w ćwiczeniu (a), a następnie narysuj odpowiadające mu drzewo.

4. Oblicz wartości następujących wyrażeń podanych w odwrotnej notacji polskiej:
(a) $33451 - * + +$
(b) $33 + 4 + 5 * 1 -$
(c) $334 + 5 * 1 - +$

5. Oblicz wartości następujących wyrażeń podanych w odwrotnej notacji polskiej:
(a) $63/3 + 73 - *$ (b) $32\hat{}42\hat{} + 5/2 *$

6. Oblicz wartości następujących wyrażeń podanych w notacji polskiej:
(a) $- * 3 \hat{} 522$ (b) $\hat{} * 35 - 22$
(c) $-\hat{} * 3522$ (d) $/ * 2 + 25\hat{} + 342$
(e) $* + /633 - 73$

7. Napisz następujące wyrażenia algebraiczne w odwrotnej notacji polskiej:
(a) $(3x - 4)^2$ (b) $(a + 2b)/(a - 2b)$
(c) $x - x^2 + x^3 - x^4$

8. Napisz wyrażenia z ćwiczenia 7 w notacji polskiej.

9. (a) Napisz wyrażenia algebraiczne $a(bc)$ oraz $(ab)c$ w odwrotnej notacji polskiej.
(b) Zrób to samo dla $a(b + c)$ oraz $ab + ac$.
(c) Jak wygląda prawo łączności i prawo rozdzielności mnożenia względem dodawania w odwrotnej notacji polskiej?

10. Napisz wyrażenie $xy + 2\hat{}xy - 2\hat{} - xy * /$ w zwykłej algebraicznej notacji infiksowej i uprość je.

11. Weźmy zdanie złożone przedstawione na rysunku 7.22(b).
(a) Napisz to zdanie w zwykłej notacji infiksowej (z nawiasami).
(b) Napisz to zdanie w odwrotnej notacji polskiej i w notacji polskiej.

12. Następujące zdania złożone podane są w notacji polskiej. Narysuj odpowiadające im drzewa z wyróżnionym korzeniem i przepisz te wyrażenia w zwykłej notacji infiksowej.

(a) $\leftrightarrow \neg \wedge \neg p \neg q \vee pq$
(b) $\leftrightarrow \wedge pq\neg \rightarrow p\neg q$
(Są to prawa z tablicy 2.1 w § 2.2.)

13. Powtórz ćwiczenie 12 dla następujących zdań.
(a) $\rightarrow \wedge p \rightarrow pqq$ (b) $\rightarrow \wedge\wedge \rightarrow pq \rightarrow rs \vee pr \vee qs$

14. Napisz następujące zdania złożone w odwrotnej notacji polskiej.
(a) $[(p \rightarrow q) \wedge (q \rightarrow r)] \rightarrow (p \rightarrow r)$ (b) $[(p \vee q) \wedge \neg p] \rightarrow q$

15. Pokaż niejednoznaczność „notacji infiksowej bez nawiasów", zapisując następujące pary wyrażeń w notacji infiksowej bez nawiasów.
(a) $(a/b) + c$ oraz $a/(b + c)$ (b) $a + (b^3 + c)$ oraz $(a + b)^3 + c$

16. Skorzystaj z definicji rekurencyjnej dla wyrażenia poprawnie zbudowanego w odwrotnej notacji polskiej, aby pokazać, że następujące wyrażenia są wyrażeniami poprawnie zbudowanymi.
(a) $3x2\hat{}*$ (b) $xy + 1x/1y/ + *$
(c) $4x2\hat{}yz + 2\hat{}/-$

17. (a) Zdefiniuj wyrażenie poprawnie zbudowane w notacji polskiej dla wyrażeń algebraicznych.
(b) Użyj definicji z ćwiczenia (a), aby pokazać, że $\hat{} + x/4x2$ jest wyrażeniem poprawnie zbudowanym.

18. Niech $S_1 = x_1 2\hat{}$ oraz $S_{n+1} = S_n x_{n+1} 2\hat{} +$ dla $n \geq 1$. $x_1, x_2, \ldots$ oznaczają tu zmienne.
(a) Pokaż, że każde S_n jest wyrażeniem poprawnie zbudowanym w odwrotnej notacji polskiej. *Wskazówka*: zastosuj indukcję.
(b) Jak wygląda S_n w zwykłej notacji infiksowej?

19. (a) Zdefiniuj wyrażenie poprawnie zbudowane w odwrotnej notacji polskiej dla rachunku zdań; zob. definicję znajdującą się tuż przed przykładem 6.
(b) Użyj definicji z ćwiczenia (a), aby pokazać, że $pq\neg \wedge \neg pq\neg \rightarrow \vee$ jest wyrażeniem poprawnie zbudowanym.
(c) Zdefiniuj wyrażenie poprawnie zbudowane w notacji polskiej dla rachunku zdań.
(d) Użyj definicji z ćwiczenia (c), aby pokazać, że $\vee\neg \wedge p\neg q \rightarrow p\neg q$ jest wyrażeniem poprawnie zbudowanym.

20. (a) Narysuj drzewo o uporządkowanym postfiksowo ciągu wierzchołków

$$stvyrzwuxq$$

i ciągu liczb dzieci 0, 0, 0, 2, 2, 0, 0, 0, 2, 3.
(b) Czy istnieje drzewo o uporządkowanym prefiksowo ciągu wierzchołków

$$stvyrzwuxq$$

i ciągu liczb dzieci 0, 0, 0, 2, 2, 0, 0, 0, 2, 3? Odpowiedź uzasadnij.

§ 7.5. Drzewa z wagami

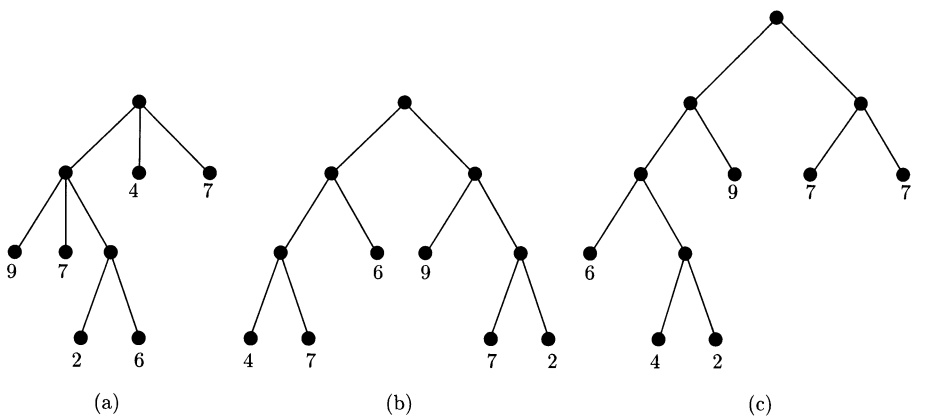
Drzewo z wagami jest to skończone drzewo z wyróżnionym korzeniem, w którym każdemu liściowi jest przyporządkowana nieujemna liczba rzeczywista, nazywana **wagą** tego **liścia**. W tym paragrafie będziemy zajmować się drzewami z wagami w ogólności oraz podamy ich zastosowania do otrzymywania kodów prefiksowych i posortowanych list.

W celu ustalenia oznaczeń przyjmijmy, że nasze drzewo T z wagami ma t liści, których wagami są liczby $w_1, w_2, \ldots, w_t$. Bez straty ogólności możemy założyć, że $w_1 \leq w_2 \leq \ldots \leq w_t$. Wygodnie będzie przypisać liściom etykiety będące ich wagami, tak więc będziemy często używać wagi jako nazwy wierzchołka. Niech $l_1, l_2, \ldots, l_t$ oznaczają odpowiednio numery poziomów liści, tak więc l_i jest długością drogi od korzenia do liścia w_i. **Wagą drzewa** T jest liczba

$$W(T) = \sum_{i=1}^{t} w_i l_i,$$

w której każda waga liścia jest pomnożona przez numer poziomu tego liścia.

PRZYKŁAD 1 (a) Sześć liści drzewa z wagami przedstawionego na rysunku 7.24(a) ma wagi 2, 4, 6, 7, 7 i 9. Zatem $w_1 = 2$, $w_2 = 4$, $w_3 = 6$, $w_4 = 7$, $w_5 = 7$ oraz $w_6 = 9$. Dwa liście mają wagę 7 i obojętne jest, który z nich nazwiemy w_4, a który w_5. Dla ustalenia uwagi przyjmijmy, że w_4 oznacza liść mający wagę 7 na



Rysunek 7.24

poziomie 2. Wtedy numery poziomów wynoszą: $l_1 = 3$, $l_2 = 1$, $l_3 = 3$, $l_4 = 2$, $l_5 = 1$ i $l_6 = 2$. Zatem

$$W(T) = \sum_{i=1}^{6} w_i l_i = 2 \cdot 3 + 4 \cdot 1 + 6 \cdot 3 + 7 \cdot 2 + 7 \cdot 1 + 9 \cdot 2 = 67.$$

(b) Te same wagi mogą być umieszczone na drzewie binarnym w taki, na przykład, sposób, jak na rysunku 7.24(b). Teraz numerami poziomów są $l_1 = 3$, $l_2 = 3$, $l_3 = 2$, $l_4 = l_5 = 3$ i $l_6 = 2$, a więc

$$W(T) = 2 \cdot 3 + 4 \cdot 3 + 6 \cdot 2 + 7 \cdot 3 + 7 \cdot 3 + 9 \cdot 2 = 90.$$

(c) Na rysunku 7.24(c) widzimy jeszcze inne drzewo binarne z tymi samymi wagami. Jego waga wynosi

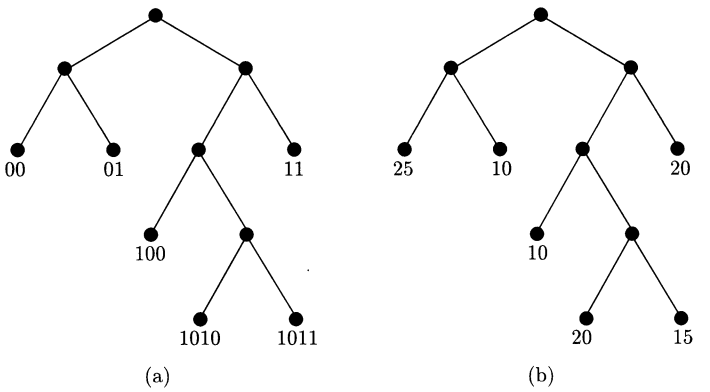
$$W(T) = 2 \cdot 4 + 4 \cdot 4 + 6 \cdot 3 + 7 \cdot 2 + 7 \cdot 2 + 9 \cdot 2 = 88.$$

Całkowita waga tego drzewa jest mniejsza niż drzewa z przykładu (b), gdyż cięższe liście znajdują się bliżej korzenia, a lżejsze są od niego dalej. W dalszym ciągu tego paragrafu omówimy algorytm, za pomocą którego można otrzymać drzewo binarne o minimalnej wadze dla zadanego ciągu wag $w_1, w_2, \ldots, w_t$. ■

PRZYKŁAD 2 (a) Jak wyjaśnimy dalej w tym paragrafie, pewne zbiory liczb zapisanych w systemie dwójkowym mogą być używane jako kody. Na przykład takim zbiorem jest zbiór $\{00, 01, 100, 1010, 1011, 11\}$. Te liczby są etykietami liści w drzewie binarnym pokazanym na rysunku 7.25(a). Można użyć tego zbioru do kodowania liter alfabetu Σ, który ma sześć liter. Przypuśćmy, że wiemy, jak często poszczególne litery alfabetu Σ występują w przesyłanych wiadomościach. Na rysunku 7.25(a) umieściliśmy w każdym liściu drzewa wagi oznaczające procentowy udział symboli kodowych znajdujących się w tych liściach. Na przykład, litera zakodowana za pomocą liczby 00 występuje w 25 procentach przypadków, litera zakodowana za pomocą liczby 1010 występuje w 20 procentach przypadków itd. Ponieważ długość każdego symbolu kodowego, jako słowa złożonego z zer i jedynek, jest dokładnie równa numerowi poziomu w drzewie binarnym, więc przeciętna długość zakodowanej wiadomości składającej się ze 100 liter alfabetu Σ będzie równa po prostu wadze tego drzewa, w tym przypadku

$$25 \cdot 2 + 10 \cdot 2 + 10 \cdot 3 + 20 \cdot 4 + 15 \cdot 4 + 20 \cdot 2 = 280.$$

Waga drzewa jest więc miarą efektywności kodu. Jak zobaczymy w przykładzie 7, istnieją bardziej efektywne kody w tym przykładzie, tzn. dla zbioru częstości 10, 10, 15, 20, 20 i 25 procent.



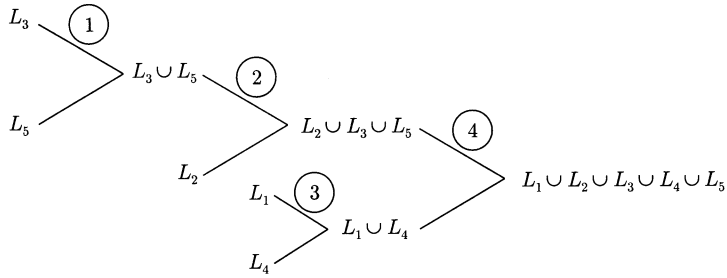
Rysunek 7.25

(b) Możemy przyjąć, że zamiast liter alfabetu Σ, słowa kodowe w przykładzie (a) mogą oznaczać adresy liści w drzewie poszukiwań binarnych z rysunku 7.25(a). Przypuśćmy, że czas dostępu do danego liścia jest proporcjonalny do liczby decyzji podejmowanych w wierzchołkach (w rozgałęzieniach) na drodze do tego liścia. Zatem aby zminimalizować średni czas dostępu, rekordy, do których zaglądamy najczęściej, powinny być umieszczone w liściach znajdujących się najbliżej korzenia, a rekordy mniej popularne dalej od korzenia. Jeśli wagi na rysunku 7.25(b) pokazują, jak często (w procentach) zagląda się do poszczególnych rekordów, to średni czas dostępu do rekordu będzie proporcjonalny do wagi tego drzewa z wagami. ■

PRZYKŁAD 3 Weźmy zbiór posortowanych list, $L_1, L_2, \ldots, L_n$. Na przykład każda lista może składać się z ułożonych alfabetycznie adresów pocztowych klientów lub może być stosem prac egzaminacyjnych ułożonych w rosnącej kolejności ocen. Dla ustalenia uwagi przyjmijmy, że każda lista jest ciągiem liczb rzeczywistych, uporządkowanych za pomocą zwykłej relacji porządku $\leq$. Przypuśćmy, że umiemy łączyć ze sobą po dwie listy i otrzymywać w ten sposób nowe listy. Naszym zadaniem jest znalezienie najbardziej efektywnego sposobu, w jaki powinniśmy łączyć ze sobą n list, aby otrzymać jedną posortowaną listę.

Dwie listy łączymy ze sobą porównując pierwsze elementy obu list i wybierając mniejszy z nich (lub którykolwiek z nich, jeśli są równe). Wybrana liczba zostaje usunięta i staje się pierwszym elementem listy połączonej, a następnie powtarzamy ten proces. Następna wybrana liczba zostaje umieszczona na drugim miejscu listy połączonej i tak dalej. Proces kończy się, gdy jedna z list jest pusta.

Na przykład, jeśli chcemy połączyć ze sobą listy 4, 8, 9 i 3, 6, 10, to: porównujemy 3 i 4, wybieramy 3 i redukujemy listy do 4, 8, 9 i 6, 10; porównujemy 4 i 6, wybieramy 4 i redukujemy listy do 8, 9 i 6, 10; porównujemy 8 i 6, wybieramy 6 i redukujemy listy do 8, 9 i 10; porównujemy 8 i 10, wybieramy 8 i redukujemy listy do 9 i 10; porównujemy 9 i 10, wybieramy 9 i redukujemy listy do listy pustej i 10; wybieramy 10. Połączoną listą jest 3, 4, 6, 8, 9, 10. W tym przykładzie potrzebnych było 5 porównań. Jeśli listy zawierają j i k elementów, to w ogólności ten proces musi zakończyć się po co najwyżej $j + k - 1$ porównaniach. Naszym celem jest tak łączyć listy $L_1, L_2, \ldots, L_n$ w pary, by w najgorszym przypadku liczba porównań była jak najmniejsza.

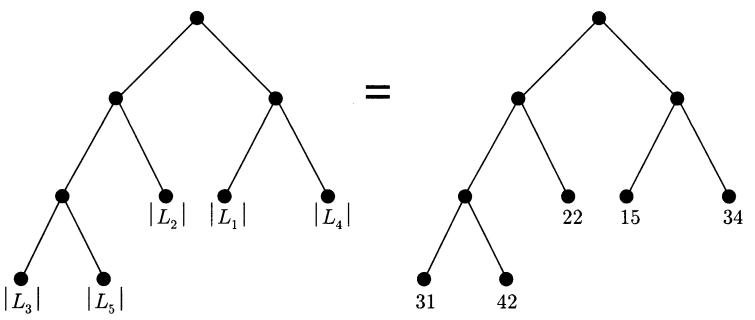


Rysunek 7.26

Przypuśćmy, na przykład, że mamy pięć list L_1, L_2, L_3, L_4, L_5 mających odpowiednio 15, 22, 31, 34 i 42 elementy i przypuśćmy, że łączymy je tak, jak na rysunku 7.26. Mamy cztery operacje łączenia list, zaznaczone liczbami w kółeczkach. Pierwsze łączenie wymaga co najwyżej $|L_3|+|L_5|-1 = 72$ porównań. Drugie łączenie wymaga co najwyżej $|L_2|+|L_3|+|L_5|-1 = 94$ porównań. Trzecie i czwarte łączenia wymagają co najwyżej $|L_1| + |L_4| - 1 = 48$ i $|L_1| + |L_2| + |L_3| + |L_4| + |L_5| - 1 = 143$ porównań. Cały ten proces będzie wymagał więc co najwyżej 357 porównań. Ta liczba sama w sobie nie ma znaczenia, ale zauważmy, że

$$357 = 2 \cdot |L_1| + 2 \cdot |L_2| + 3 \cdot |L_3| + 2 \cdot |L_4| + 3 \cdot |L_5| - 4.$$

Jest to dokładnie o 4 mniej niż wynosi waga drzewa na rysunku 7.27. Zauważmy ścisły związek między rysunkami 7.26 i 7.27. Niezależnie od tego, w jaki sposób będziemy łączyć ze sobą w pary te pięć list, zawsze będziemy musieli wykonać cztery połączenia. Obliczenia takie jak wyżej pokazują, że wszystkie te łączenia będą wymagały co najwyżej $W(T)-4$ porównań, gdzie T jest drzewem ilustrującym sposób łączenia. Zatem znalezienie sposobu łączenia



Rysunek 7.27

minimalizującego w najgorszym przypadku liczbę porównań jest równoważne ze znalezieniem drzewa binarnego z wagami 15, 22, 31, 34, 42, mającego minimalną wagę. Powrócimy do tego problemu w przykładzie 8.

Łączenie ze sobą parami n list wymaga $n-1$ połączeń. Zwykle łączenie n list będzie wymagało co najwyżej $W(T)-(n-1)$ porównań, gdzie T jest drzewem z wagami odpowiadającym sposobowi łączenia. ∎

Przykłady 2 i 3 sugerują następujący ogólny problem. Mamy dany co najmniej dwuelementowy ciąg $L = (w_1, \ldots, w_t)$ liczb nieujemnych i chcemy skonstruować drzewo binarne T z wagami będącymi wyrazami ciągu L, którego waga $W(T)$ byłaby jak najmniejsza. Takie drzewo T nazywamy **optymalnym drzewem binarnym** dla ciągu wag $w_1, \ldots, w_t$. Następujący algorytm rekurencyjny rozwiązuje ten problem, budując optymalne drzewo binarne $T(L)$.

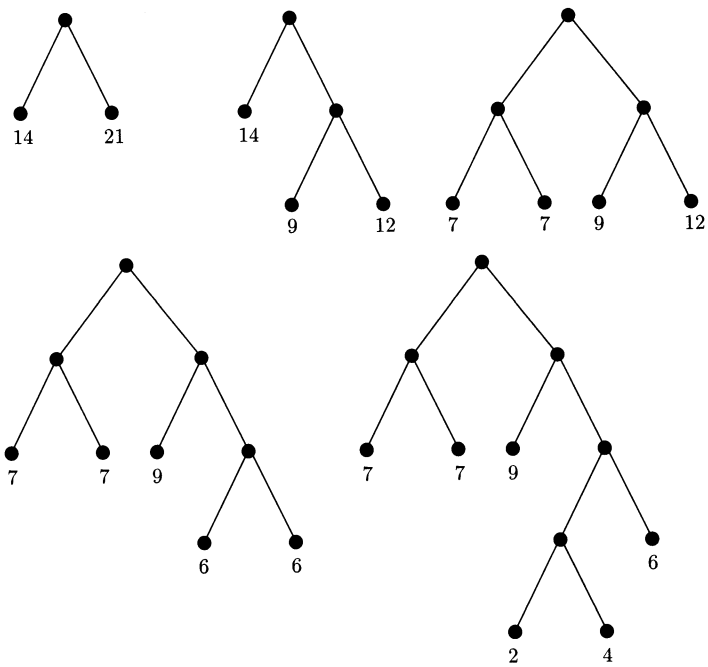
Algorytm HUFFMAN (L)

{Dane: ciąg $L = (w_1, \ldots, w_t)$ liczb nieujemnych, gdzie $t \geq 2$}
{Wyniki: optymalne drzewo binarne $T(L)$ dla ciągu L}
Jeśli $t = 2$, to niech $T(L)$ będzie drzewem z dwoma liśćmi mającymi wagi w_1 i w_2,
w przeciwnym przypadku, znajdźmy dwa najmniejsze wyrazy ciągu L, np. u i v,
  niech L' będzie listą otrzymaną z L przez usunięcie u i v i wstawienie $u+v$,
  wykonaj algorytm HUFFMAN (L'), by otrzymać drzewo $T(L')$,
  utwórz drzewo $T(L)$ z drzewa $T(L')$ zastępując liść wagi $u+v$ w drzewie $T(L')$ poddrzewem mającym dwa liście o wagach u i v. ∎

Ten algorytm redukuje w końcu problem do problemu w przypadku początkowym, mianowicie znalezienia optymalnego drzewa binarnego mającego dwa liście; rozwiązanie tego problemu jest już trywialne. Pokażemy wkrótce, że algorytm HUFFMAN(L) zawsze tworzy optymalne drzewo binarne. Popatrzymy jednak najpierw na kilka przykładów działania tego algorytmu i zastosujemy go do rozwiązania problemów, które były motywacją naszego zainteresowania drzewami optymalnymi.

PRZYKŁAD 4 Weźmy wagi 2, 4, 6, 7, 7, 9. Algorytm najpierw wielokrotnie zastępuje dwie najmniejsze wagi ich sumą i w ten sposób tworzy coraz krótsze ciągi. A oto ciąg wywołań rekurencyjnych:

HUFFMAN (2, 4, 6, 7, 7, 9) zastępuje wagi 2 i 4 sumą 2+4 = 6 i wywołuje
HUFFMAN (6, 6, 7, 7, 9), który zastępuje wagi 6 i 6 sumą 12 i wywołuje
HUFFMAN (7, 7, 9, 12), który wywołuje
HUFFMAN (9, 12, 14), który wywołuje
HUFFMAN (14, 21), który buduje pierwsze drzewo $T(14, 21)$ pokazane na rysunku 7.28.



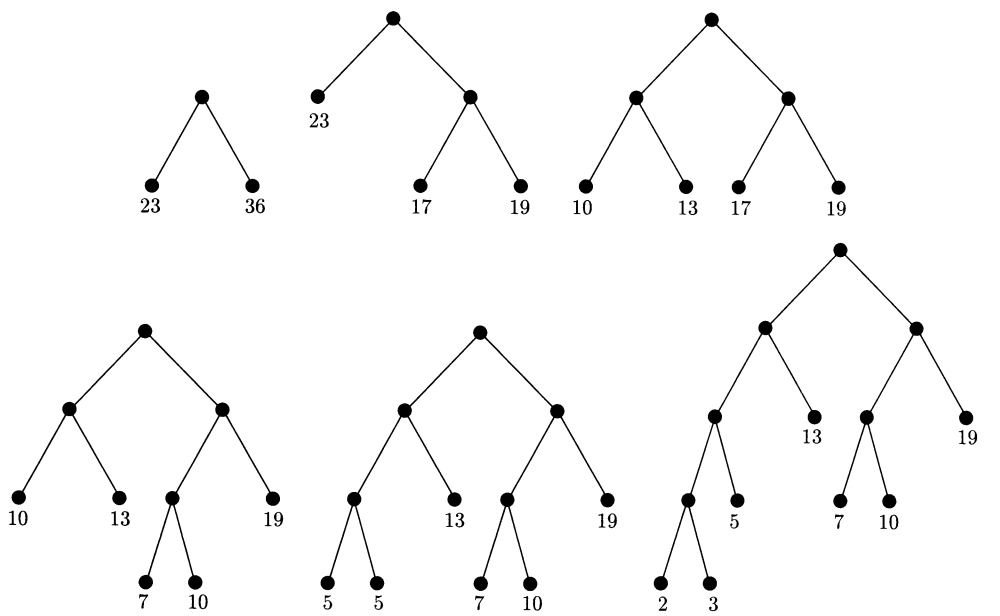
Rysunek 7.28

Następnie każde z poprzednich wywołań rekurencyjnych konstruuje swoje drzewo i przekazuje je z powrotem w górę ciągu wywołań. Na rysunku 7.28 widzimy cały ciąg drzew $T(14, 21)$, $T(9, 12, 14)$, ..., $T(2, 4, 6, 7, 7, 9)$. Zauważmy na przykład, że trzecie drzewo $T(7, 7, 9, 12)$ jest otrzymane z drugiego drzewa $T(9, 12, 14)$ przez zastąpienie liścia wagi $14 = 7 + 7$ poddrzewem mającym dwa liście, każdy o wadze 7. Ostatecznym drzewem z wagami jest w zasadzie to samo drzewo co drzewo przedstawione na rysunku 7.24(c), tak więc to drzewo jest optymalnym drzewem binarnym. Jak zauważyliśmy w przykładzie 1(c), ma ono wagę 88. ∎

PRZYKŁAD 5 Znajdźmy optymalne drzewo binarne z wagami 2, 3, 5, 7, 10, 13, 19. Kilkakrotnie dodajemy najmniejsze dwie wagi, otrzymując w ten sposób ciągi

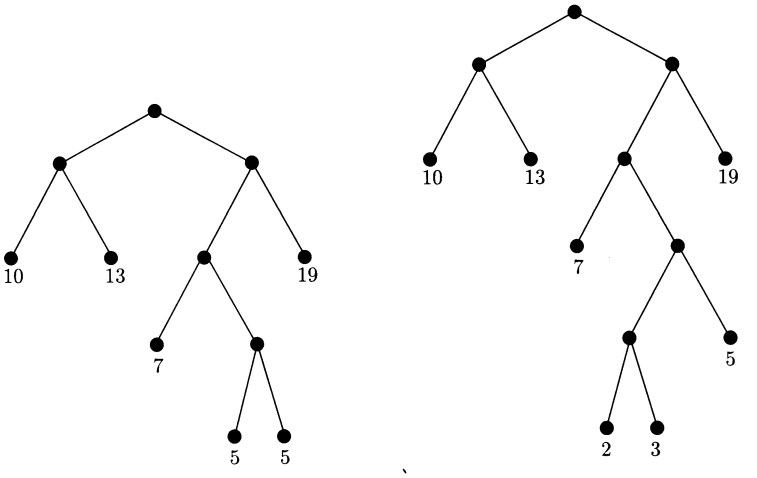
$$2, 3, 5, 7, 10, 13, 19 \to 5, 5, 7, 10, 13, 19 \to$$
$$\to 7, 10, 10, 13, 19 \to 10, 13, 17, 19 \to 17, 19, 23 \to 23, 36.$$

Następnie używamy algorytmu Huffmana do budowania optymalnych drzew binarnych przedstawionych na rysunku 7.29. Po otrzymaniu czwartego drzewa każdy liść wagi 10 mógł być zastąpiony poddrzewem z wagami 5 i 5. Zatem ostatnie dwa drzewa



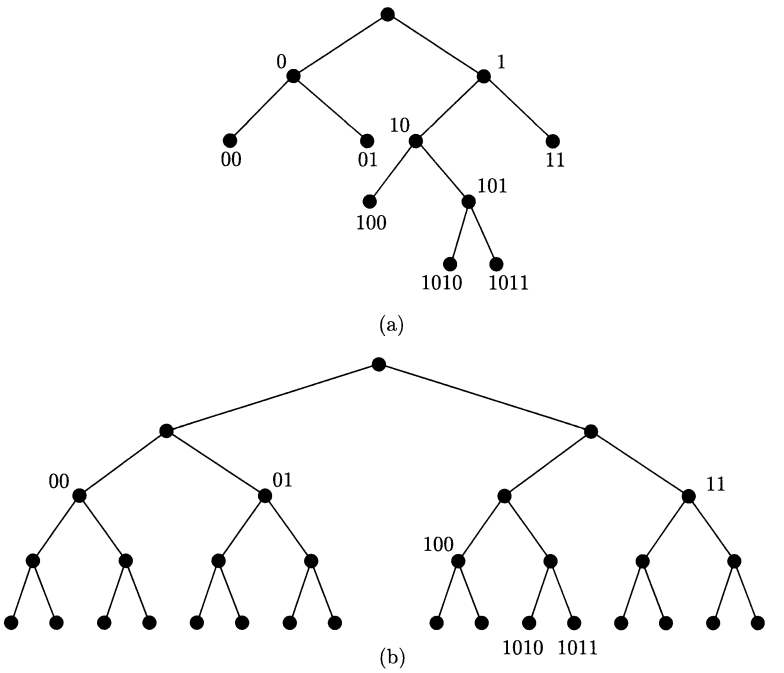
Rysunek 7.29

mogą wyglądać tak jak na rysunku 7.30. Niezależnie od sposobu wyboru ostateczne drzewo ma wagę 150 (por. ćwiczenie 2). Zauważmy, że optymalne drzewo binarne nie jest wyznaczone jednoznacznie; drzewo na rysunku 7.29 ma wysokość 4, a drzewo z rysunku 7.30 ma wysokość 5. ■



Rysunek 7.30

Powrócimy teraz do kodowania liczb w zapisie dwójkowym. Aby uniknąć niejednoznaczności, musimy nałożyć pewne ograniczenia; na przykład w kodowaniu nie można używać wszystkich trzech ciągów 10, 01 i 0110, gdyż ciąg taki jak 011010 mógłby być odczytany na różne sposoby. Jeśli traktujemy te ciągi zer i jedynek jako etykiety wierzchołków w drzewie binarnym, to każda etykieta zawiera instrukcje, w jaki sposób można znaleźć ten wierzchołek, idąc od korzenia: idziemy w lewo, gdy napotkamy 0, w prawo, gdy 1. Motywowani częściowo drzewem poszukiwań z przykładu 2(b) nakładamy warunek, że żaden symbol kodowy nie może być odcinkiem początkowym innego symbolu kodowego. Zatem na przykład 01 i 0110 nie mogą być jednocześnie symbolami kodowymi. Jeśli wyrazimy to w terminologii drzew binarnych, ten warunek oznacza, że żaden wierzchołek mający etykietę będącą symbolem kodowym nie może leżeć poniżej innego takiego wierzchołka w tym drzewie. Narzucimy jeszcze jeden warunek na kody: drzewo binarne, którego liście są etykietowane słowami kodowymi, musi być regularne, tak więc każdy wierzchołek nie będący liściem ma dwoje dzieci. Kod spełniający te dwa warunki nazywamy **kodem prefiksowym**.



Rysunek 7.31

PRZYKŁAD 6 Zbiór $\{00, 01, 100, 1010, 1011, 11\}$ jest kodem prefiksowym. Jest to zbiór liści etykietowanego drzewa binarnego przedstawionego na rysunku 7.25(a), przerysowanego na rysunku 7.31(a) w taki sposób, że wszystkie wierzchołki oprócz korzenia mają etykiety. Każdy ciąg zer i jedynek długości 4 zaczyna się od jednego z tych symboli kodowych, gdyż każda droga długości 4, wychodząca z korzenia, w pełnym drzewie binarnym z rysunku 7.31(b) przechodzi przez jeden z tych wierzchołków kodowych. To znaczy, że możemy próbować odkodować dowolny ciąg zer i jedynek przesuwając się od lewej do prawej w tym ciągu i znajdując pierwszy podciąg będący symbolem kodowym, potem drugi taki podciąg itd. Ta procedura albo zużyje cały ciąg, albo pozostawi na końcu co najwyżej trzy nie odkodowane zera lub jedynki.

Weźmy na przykład ciąg

$$11101011011000100111110010.$$

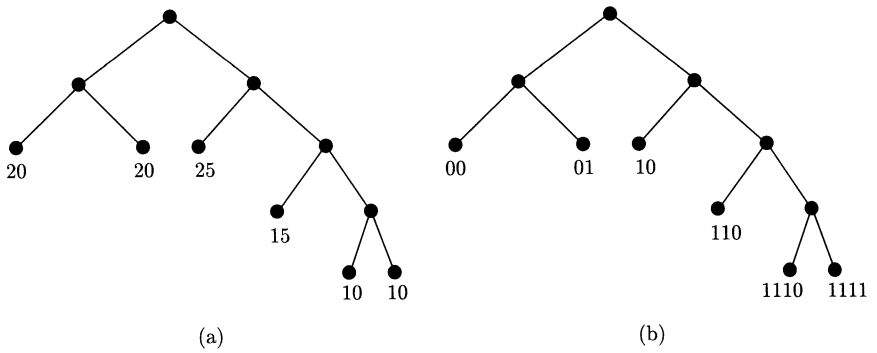
Odwiedzamy wierzchołek 1, następnie wierzchołek 11. Ponieważ wierzchołek 11 jest liściem, zapisujemy 11 i powracamy do korzenia. Następnie odwiedzamy wierzchołki 1, 10, 101 i 1010. Ponieważ wierzchołek 1010 jest liściem, zapisujemy 1010 i znów powracamy do korzenia. Postępując dalej w ten sposób otrzymamy

ciąg symboli kodowych

$$11,\ 1010,\ 11,\ 01,\ 100,\ 01,\ 00,\ 11,\ 11,\ 100,$$

po którym pozostały cyfry 10. Taki system odkodowywania dowolnych ciągów zer i jedynek będzie działał poprawnie dla każdego kodu mającego tę własność, że każda droga od korzenia w pełnym drzewie binarnym przechodzi przez dokładnie jeden wierzchołek kodowy. Kody prefiksowe mają tę własność z definicji. ∎

PRZYKŁAD 7 (a) Rozwiążemy teraz problem postawiony w przykładzie 2(a), to znaczy znajdziemy najbardziej efektywny kod prefiksowy dla zbioru częstości 10, 10, 15, 20, 20 i 25 procent. Chcemy zminimalizować średnią długość zakodowanej wiadomości składającej się ze 100 liter alfabetu Σ. Mamy zatem znaleźć optymalne drzewo binarne dla tych wag. Korzystając z procedury pokazanej w przykładach 4 i 5 otrzymamy drzewo z wagami pokazane na rysunku 7.32(a). Etykietujemy to drzewo liczbami w zapisie dwójkowym, tak jak na rysunku 7.32(b). Wtedy zbiór



Rysunek 7.32

$\{00, 01, 10, 110, 1110, 1111\}$ będzie najbardziej efektywnym kodem dla alfabetu Σ, pod warunkiem, że dopasujemy litery alfabetu Σ do symboli kodowych tak, aby częstości liter były zgodne z rysunkiem 7.32(a). Przy takim kodowaniu średnia długość zakodowanej wiadomości mającej 100 liter alfabetu Σ wynosi

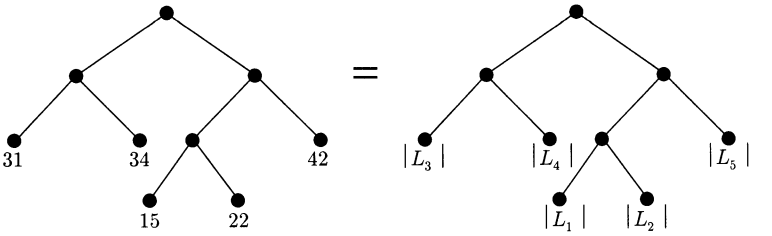
$$20 \cdot 2 + 20 \cdot 2 + 25 \cdot 2 + 15 \cdot 3 + 10 \cdot 4 + 10 \cdot 4 = 255,$$

czyli lepiej niż średnia długość 280, otrzymana w przykładzie 2.

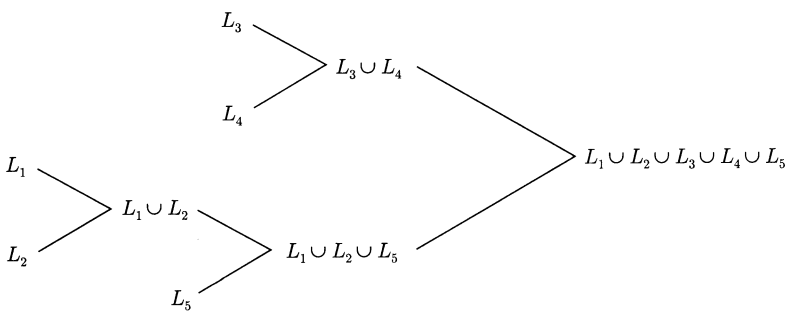
(b) Otrzymane przed chwilą rozwiązanie daje również najbardziej efektywne drzewo binarnych poszukiwań rekordów, których

częstości dostępu wynoszą 10, 10, 15, 20, 20 i 25 procent, czyli tak jak w przykładzie 2(b). ■

PRZYKŁAD 8 Dokończymy teraz omawianie łączenia posortowanych list, rozpoczęte w przykładzie 3. Widzieliśmy tam, że szukamy optymalnego drzewa binarnego dla ciągu wag 15, 22, 31, 34, 42. Korzystając z procedur z przykładów 4 i 5 znajdujemy drzewo pokazane na rysunku 7.33. To drzewo ma wagę 325. Odpowiadające mu łączenie list parami pokazane na rysunku 7.34 będzie wymagało co najwyżej $325 - 4 = 321$ porównań. ■



Rysunek 7.33



Rysunek 7.34

Aby pokazać, że algorytm Huffmana działa poprawnie, udowodnimy najpierw lemat, który mówi, że w optymalnym drzewie binarnym ciężkie liście znajdują się blisko korzenia. Dowody lematu i wniosku z tego lematu są dość oczywiste, jeśli wszystkie wagi są różne (por. ćwiczenie 14). Jednakże jest nam potrzebny przypadek bardziej ogólny. Jeśli nawet na początku wszystkie wagi są różne, to algorytm Huffmana może doprowadzić do przypadku, w którym nie wszystkie wagi są różne, tak jak to się stało w przykładzie 5.

Lemat

Niech T będzie optymalnym drzewem binarnym dla ciągu wag $w_1, w_2, \ldots, w_t$. Niech l_i dla $i = 1, 2, \ldots, t$ oznacza numer poziomu wierzchołka w_i. Jeśli $w_j < w_k$, to $l_j \geq l_k$.

Dowód. Załóżmy, że $w_j < w_k$ i $l_j < l_k$ dla pewnych j i k. Niech T' będzie drzewem otrzymanym przez zamianę wag w_j i w_k. Przy obliczaniu wagi $W(T)$ udział wierzchołków w_j i w_k w całej sumie wynosi $w_j l_j + w_k l_k$, natomiast przy obliczaniu wagi $W(T')$ ich udział wynosi $w_j l_k + w_k l_j$. Ponieważ udział innych wierzchołków w obydwu sumach $W(T)$ i $W(T')$ jest taki sam, więc

$$\begin{aligned} W(T) - W(T') &= w_j l_j + w_k l_k - w_j l_k - w_k l_j \\ &= (w_k - w_j)(l_k - l_j) > 0. \end{aligned}$$

Zatem $W(T') < W(T)$, a więc T nie jest optymalnym drzewem binarnym, co przeczy założeniu. ∎

Wniosek

Istnieje optymalne drzewo binarne T, w którym dwie najmniejsze wagi w_1 i w_2 znajdują się na tym samym najniższym poziomie l.

Dowód. Istnieją co najmniej dwa liście znajdujące się na najniższym poziomie, niech będą to w_j i w_k. Jeśli $w_1 < w_j$, to na mocy lematu $l_1 \geq l_j = l$, a więc $l_1 = l$, czyli w_1 znajduje się na poziomie l. Jeśli $w_1 = w_j$, to może się zdarzyć, że $l_1 < l_j$, ale wtedy możemy zamienić w_1 i w_j, nie zmieniając całkowitej wagi drzewa T. Podobnie, zamieniając miejscami w_2 i w_k, jeśli jest to konieczne, możemy założyć, że w_2 znajduje się na poziomie l. ∎

Następujące twierdzenie pokazuje, że algorytm HUFFMAN(L) działa poprawnie.

Twierdzenie

Załóżmy, że $0 \leq w_1 \leq w_2 \leq \ldots \leq w_t$. Niech T' będzie optymalnym drzewem binarnym dla ciągu wag $w_1 + w_2, w_3, \ldots, w_t$ i niech T będzie drzewem binarnym z wagami, otrzymanym z drzewa T' przez zastąpienie liścia wagi $w_1 + w_2$ poddrzewem mającym dwa liście o wagach w_1 i w_2. Wtedy T jest optymalnym drzewem binarnym dla ciągu wag $w_1, w_2, \ldots, w_t$.

Dowód. Ponieważ istnieje tylko skończenie wiele drzew binarnych mających t liści, więc musi istnieć optymalne drzewo

binarne T_0 dla ciągu wag $w_1, w_2, \ldots, w_t$. Naszym zadaniem jest wykazanie, że $W(T) = W(T_0)$. Na podstawie wniosku z lematu możemy założyć, że wagi w_1 i w_2 w drzewie T_0 znajdują się na tym samym poziomie. Całkowita waga drzewa T_0 nie zmieni się, jeśli zamienimy ze sobą wagi na tym samym poziomie. Możemy więc założyć, że w_1 i w_2 są dziećmi tego samego rodzica p. Te trzy wierzchołki tworzą małe poddrzewo T_p, którego korzeniem jest p.

Niech teraz T_0 będzie drzewem z wagami $w_1 + w_2, w_3, \ldots, w_t$, otrzymanym z drzewa T_0 przez zastąpienie poddrzewa T_p liściem $\bar{p}$ wagi $w_1 + w_2$. Niech l będzie numerem poziomu wierzchołka p. Zauważmy, że przy obliczaniu wagi $W(T_0)$ udział poddrzewa T_p wynosi $w_1(l+1) + w_2(l+1)$, podczas gdy przy obliczaniu wagi $W(T_0')$ udział wierzchołka $\bar{p}$ wagi $w_1 + w_2$ wynosi $(w_1 + w_2)l$. Zatem

$$W(T_0) = W(T_0') + w_1 + w_2.$$

Podobne rozumowanie pokazuje, że

$$W(T) = W(T') + w_1 + w_2.$$

Ponieważ drzewo T' jest drzewem optymalnym dla ciągu wag $w_1 + w_2, w_3, \ldots, w_t$, więc $W(T') \leq W(T_0')$, a zatem

$$W(T) = W(T') + w_1 + w_2 \leq W(T_0') + w_1 + w_2 = W(T_0).$$

Oczywiście $W(T_0) \leq W(T)$, gdyż drzewo T_0 jest drzewem optymalnym dla ciągu wag $w_1, w_2, \ldots, w_t$, a zatem $W(T) = W(T_0)$, czego należało dowieść. Tak więc drzewo T jest optymalnym drzewem binarnym dla ciągu wag $w_1, w_2, \ldots, w_t$. ∎

Działanie algorytmu HUFFMAN(L) polega na $t-1$ wyborach rodziców w drzewie $T(L)$. Każdy wybór wymaga znalezienia dwóch najmniejszych wyrazów w aktualnej liście wag, co może być zrobione w czasie $O(t)$ po prostu poprzez przejrzenie całej listy. Tak więc całkowity czas działania algorytmu wynosi $O(t^2)$.

Istnieją co najmniej dwa sposoby przyspieszenia działania tego algorytmu. Można znaleźć dwa najmniejsze elementy listy w czasie $O(\log_2 t)$, korzystając z drzewa binarnego jako struktury danych dla tej listy. Można również użyć algorytmów, które sortują początkową listę L w porządku niemalejącym w czasie $O(t \log_2 t)$. Następnie zauważamy, że dwa najmniejsze elementy są po prostu pierwszymi dwoma elementami listy, a po usunięciu ich możemy zachować niemalejący porządek całej listy, wstawiając sumę tych elementów we właściwe miejsce, tak jak to robiliśmy

w przykładach 4 i 5. Właściwe miejsce, w które należy wstawić tę sumę, można znaleźć w czasie $O(\log_2 t)$, a więc tak zmodyfikowany algorytm działa również w czasie $O(t \log_2 t)$.

Kody prefiksowe o wadze minimalnej, nazywane **kodami Huffmana**, mają duże znaczenie praktyczne, ze względu na zastosowania do efektywnego przesyłania wiadomości oraz do projektowania struktur danych będących drzewami poszukiwań, tak jak to opisaliśmy w przykładach 2(a) i 2(b). W obu tych zastosowaniach faktyczne częstości występowania symboli w wiadomościach czy dostępów do rekordów są wyznaczane doświadczalnie i mogą zmieniać się z upływem czasu. W przypadku drzew poszukiwań może się nawet zmieniać liczba symboli kodowych, gdy dodaje się lub usuwa rekordy. Problem dynamicznego modyfikowania kodów Huffmana, nadążającego za zmieniającymi się okolicznościami, jest interesujący i pociągający, ale niestety wykracza poza zakres tej książki.

ĆWICZENIA DO § 7.5

1. (a) Oblicz wagi wszystkich drzew z rysunku 7.28.
(b) Oblicz wagi wszystkich drzew z rysunku 7.29.

2. Oblicz wagi dwóch drzew z rysunków 7.29 i 7.30 z wagami 2, 3, 5, 7, 10, 13, 19.

3. Zbuduj optymalne drzewo binarne dla następujących zbiorów wag i oblicz wagę tego optymalnego drzewa binarnego.
(a) $\{1, 3, 4, 6, 9, 13\}$
(b) $\{1, 3, 5, 6, 10, 13, 16\}$
(c) $\{2, 4, 5, 8, 13, 15, 18, 25\}$
(d) $\{1, 1, 2, 3, 5, 8, 13, 21, 34\}$

4. Znajdź optymalne drzewo binarne dla wag 10, 10, 15, 20, 20, 25 i porównaj swoją odpowiedź z rysunkiem 7.32(a).

5. Które z następujących zbiorów ciągów są kodami prefiksowymi? Jeśli zbiór jest kodem prefiksowym, zbuduj drzewo binarne, którego liście reprezentują ten kod binarny. Jeśli zbiór nie jest kodem prefiksowym, wyjaśnij, dlaczego.
(a) $\{000, 001, 01, 10, 11\}$
(b) $\{00, 01, 110, 101, 0111\}$
(c) $\{00, 0100, 0101, 011, 100, 101, 11\}$

6. Oto kod prefiksowy: $\{00, 010, 0110, 0111, 10, 11\}$.
(a) Zbuduj drzewo binarne, którego liście reprezentują ten kod binarny.

(b) Odkoduj ciąg

0010000110010000100111110110,

jeśli 00 = A, 10 = D, 11 = E, 010 = H, 0110 = M oraz 0111 reprezentuje apostrof '. Otrzymasz króciutki wierszyk zatytułowany „Fleas".

(c) Odkoduj 0101101100010110110110110010.

(d) Odkoduj następujący wierszyk: 1 0 0 0 1 0 0 1 0 0 0 1 0 0 1 1 0 0 0 1 0 0 0 0 1 1 0. 0 1 0 1 1 0 1 1 0 0 0 1 0 1 1 0 1 1 0 0 0 0 1 1 0 0 0 1 0. 0 1 1 0 0 0 0 1 1 0 0 0 1 0 1 1 1 0 0 0 1 0 1 0 1 1 0 0 1 0.

7. Przypuśćmy, że dany jest fikcyjny alfabet Σ składający się z siedmiu liter a, b, c, d, e, f i g z następującymi częstościami liter (podanymi w procentach): $a - 11$, $b - 20$, $c - 4$, $d - 22$, $e - 14$, $f - 8$, $g - 21$.

(a) Zaprojektuj optymalny binarny kod prefiksowy dla tego alfabetu.

(b) Jaka jest średnia długość zakodowanej wiadomości składającej się ze 100 liter alfabetu Σ?

8. Powtórz ćwiczenie 7 dla częstości (wyrażonych w procentach): $a - 25$, $b - 2$, $c - 15$, $d - 10$, $e - 38$, $f - 4$, $g - 6$.

9. (a) Pokaż, że kod $\{000, 001, 10, 110, 111\}$ spełnia wszystkie wymagania dotyczące kodów prefiksowych, z wyjątkiem tego, że odpowiadające mu drzewo binarne nie jest regularne.

(b) Pokaż, że pewne ciągi cyfr dwójkowych nie mają sensu dla tego kodu.

(c) Pokaż, że ciąg $\{00, 01, 10, 110, 111\}$ jest kodem prefiksowym i porównaj jego drzewo binarne z drzewem z ćwiczenia (a).

10. Powtórz ćwiczenie 7 dla częstości (wyrażonych w procentach): $a - 31$, $d - 31$, $e - 12$, $h - 6$, $m - 20$.

11. Niech L_1, L_2, L_3, L_4 będą posortowanymi listami mającymi odpowiednio 23, 31, 61 i 73 elementy. Ile potrzeba co najwyżej porównań, aby połączyć te listy tak jak to pokazano na rysunku 7.35?

(e) Jak powinny być łączone te cztery listy, aby całkowita liczba porównań była minimalna? Nie wystarczy po prostu sprawdzić części od (a) do (d) rysunku 7.35, ponieważ istnieją inne sposoby łączenia tych list.

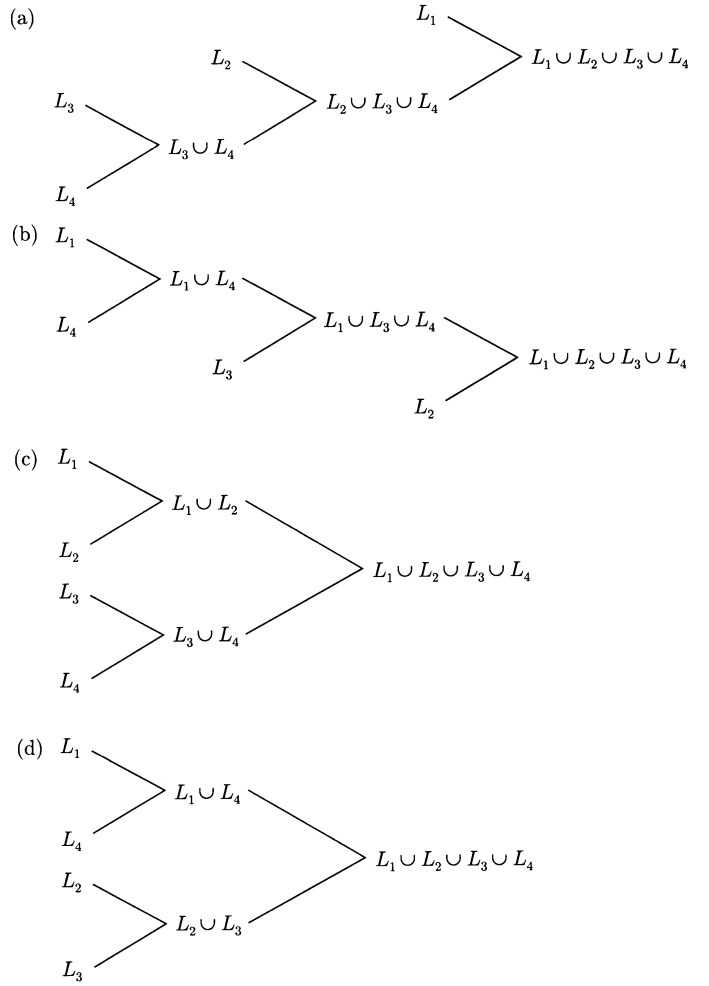
12. Niech L_1, L_2, L_3, L_4, L_5, L_6 będą posortowanymi listami mającymi odpowiednio 5, 6, 9, 22, 29, 34 elementy.

(a) Pokaż, jak powinny być łączone te listy, by całkowita liczba porównań była jak najmniejsza.

(b) Ile porównań może być potrzebnych w twojej procedurze?

13. Powtórz ćwiczenie 12 dla siedmiu list mających odpowiednio 2, 5, 8, 12, 16, 22 i 24 elementy.

14. Niech T będzie optymalnym drzewem binarnym, którego wagi spełniają nierówności $w_1 < w_2 < \ldots < w_t$. Pokaż, że odpowiadające



Rysunek 7.35

numery poziomów spełniają nierówności

$$l_1 \geq l_2 \geq l_3 \geq \ldots \geq l_t.$$

15. Popatrz jeszcze raz na ćwiczenie 1 i zauważ, że jeśli tylko wierzchołek wagi $w_1 + w_2$ w drzewie T' jest zastąpiony poddrzewem o wagach w_1 i w_2, to waga wzrasta o $w_1 + w_2$. To znaczy, że nowe drzewo T ma wagę $W(T') + w_1 + w_2$, dokładnie tak jak w dowodzie twierdzenia.

To, co jest najważniejsze w tym rozdziale

Sugestie na temat tego, jak korzystać z tego materiału, znajdziesz w podobnym punkcie na końcu rozdziału 1.

Pojęcia i oznaczenia

definicja rekurencyjna zbioru, funkcji
    warunek początkowy, warunek rekurencyjny
    definicja jednoznaczna
wyrażenie algebraiczne poprawnie zbudowane, wyrażenie poprawnie zbudowane w odwrotnej notacji polskiej
algorytm rekurencyjny
    przypadek początkowy
    dane otrzymane z innych danych
porządek prefiksowy, postfiksowy i infiksowy
etykietowanie uporządkowane drzewa z wyróżnionym korzeniem, acyklicznego grafu skierowanego
NAST(V), DOST(V)
notacja polska, odwrotna notacja polska, notacja infiksowa
działania dwuargumentowe, jednoargumentowe
drzewo z wagami (wagi liści), waga drzewa
optymalne drzewo binarne
    łączenie list
    kod prefiksowy
kod Huffmana = kod prefiksowy o wadze minimalnej

Fakty

Zbiory drzew skończonych i skończonych drzew z wyróżnionym korzeniem mogą być zdefiniowane rekurencyjnie.

Algorytmy rekurencyjne mogą sprawdzać przynależność do zbiorów zdefiniowanych rekurencyjnie i obliczać wartości funkcji zdefiniowanych rekurencyjnie.

Uogólniona zasada indukcji.

Warunki poprawności dla algorytmów rekurencyjnych.

Uporządkowane drzewo z wyróżnionym korzeniem nie zawsze może być odtworzone z listy wierzchołków uporządkowanej prefiksowo, infiksowo lub postfiksowo.

Może ono być odtworzone z listy uporządkowanej prefiksowo lub postfiksowo, jeśli wiemy, ile dzieci ma każdy wierzchołek.

Metody i algorytmy

Przeszukiwanie w głąb, aby przejść drzewo z wyróżnionym korzeniem.

Algorytmy PREORDER, POSTORDER i INORDER, służące do tworzenia list wierzchołków uporządkowanego drzewa z wyróżnionym korzeniem.

Metoda obciążeń do szacowania czasu działania algorytmu.

Algorytm ETYKIETOWANIE DRZEWA tworzący etykietowanie uporządkowane drzewa z wyróżnionym korzeniem.

Algorytmy SORTOWANIE DRZEWA i ETYKIETOWANIE tworzące uporządkowane etykietowanie acyklicznego grafu skierowanego w czasie $O(|V(G)| + |E(G)|)$.

Użycie binarnych drzew z wagami do znalezienia efektywnych sposobów łączenia list i efektywnych kodów prefiksowych.

Algorytm Huffmana do znajdowania optymalnych drzew binarnych z danymi wagami.

8. GRAFY SKIEROWANE

Ten rozdział jest poświęcony grafom skierowanym, które zdefiniowaliśmy w rozdziale 3. W pierwszym paragrafie zajmujemy się przede wszystkim acyklicznymi grafami skierowanymi. Omawiamy tam etykietowania uporządkowane, w sposób niezależny od metody użytej w rozdziale 7, a także zajmujemy się wersją twierdzenia Eulera dla grafów skierowanych. W pozostałej części tego rozdziału zajmujemy się grafami skierowanymi z wagami, kładąc główny nacisk na drogi o najmniejszej wadze między parami wierzchołków. W paragrafie 8.2 wprowadzamy pojęcie grafu skierowanego z wagami oraz omawiamy sieci zdarzeń, w których duże znaczenie mają drogi o największej wadze. Pomimo że w tym rozdziale odwołujemy się kilka razy do pojęcia etykietowania uporządkowanego i do paragrafu 7.3, można ten rozdział czytać niezależnie od rozdziału 7.

§ 8.1. Grafy skierowane

Drogi zamknięte i cykle w grafach skierowanych i nieskierowanych zostały zdefiniowane w § 3.2. W tym paragrafie będziemy badać acykliczność grafów skierowanych oraz podamy wersję twierdzenia Eulera dla grafów skierowanych.

Najpierw zauważymy, że w pewnym sensie cykle w drogach nie mają znaczenia.

Twierdzenie 1

Jeśli u i v są różnymi wierzchołkami grafu skierowanego G i jeśli istnieje w grafie G droga z wierzchołka u do wierzchołka v, to istnieje droga acykliczna z u do v.

Dowód jest dokładnie taki sam jak dowód twierdzenia 1 w § 6.1. Jedyna różnica polega na tym, że teraz krawędzie dróg są skierowane.

Wniosek 1

Jeśli istnieje droga zamknięta z wierzchołka v do v, to istnieje cykl z v do v.

Dowód. Jeśli istnieje w tym grafie krawędź e prowadząca z wierzchołka v do niego samego, to jednoelementowy ciąg e jest cyklem z v do v. W przeciwnym przypadku istnieje droga zamknięta z v do v mająca postać $vx_2 \ldots x_n v$, gdzie $x_n \neq v$. Wtedy z twierdzenia 1 wynika, że istnieje droga acykliczna z v do x_n. Dołączając do niej ostatnią krawędź z x_n do v otrzymamy żądany cykl. ∎

Wniosek 2

Droga jest acykliczna wtedy i tylko wtedy, gdy wszystkie jej wierzchołki są różne.

Dowód. Jeśli droga nie ma powtarzających się wierzchołków, to oczywiście jest acykliczna. Jeśli jakiś wierzchołek powtarza się na drodze, to zawiera ona drogę zamkniętą, a więc zawiera cykl na mocy wniosku 1. ∎

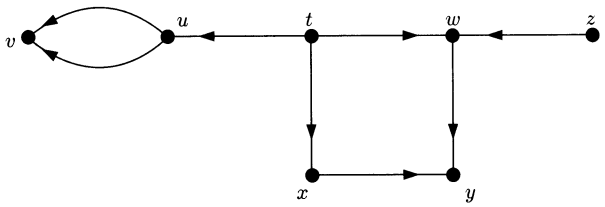
Zdefiniujemy teraz pewne szczególne typy wierzchołków, które nie występują w grafach nieskierowanych. Wierzchołek grafu skierowanego nazwiemy **ujściem**, jeśli nie jest początkiem żadnej krawędzi. Na wykresie ujścia odpowiadają punktom, z których nie wychodzą żadne strzałki. Punkty, do których nie dochodzą żadne strzałki, mają również znaczenie. Wierzchołek grafu skierowanego nazywamy **źródłem**, jeśli nie jest on końcem żadnej krawędzi.

PRZYKŁAD 1

Weźmy graf pokazany na rysunku 8.1. Wierzchołki v i y są ujściami, a wierzchołki t i z są źródłami. Ten graf jest acykliczny. Następne twierdzenie pokaże, że nie przypadkiem ma on co najmniej jedno ujście i co najmniej jedno źródło. ∎

Twierdzenie 2

Każdy skończony acykliczny graf skierowany ma co najmniej jedno ujście i co najmniej jedno źródło.



Rysunek 8.1

Podamy trzy dowody dla ujść, gdyż analogiczne fakty dotyczące źródeł mogą być udowodnione przez odwrócenie wszystkich strzałek (por. ćwiczenie 13).

Dowód 1. Ponieważ graf jest acykliczny, więc każda droga w nim jest acykliczna. Ponieważ graf jest skończony, więc długości dróg są ograniczone i musi istnieć droga o największej długości, na przykład $v_1 v_2 \ldots v_n$. Wierzchołek v_n musi wtedy być ujściem. (Oczywiście, jeśli graf skierowany nie ma wcale krawędzi, to każdy wierzchołek jest ujściem). ∎

Ten dowód jest krótki i elegancki, ale nie mówi nam, jak znaleźć v_n lub jakiekolwiek inne ujście. Nasz następny dowód jest konstruktywny.

Dowód 2. Weźmy dowolny wierzchołek v_1. Jeśli v_1 jest ujściem, to koniec konstrukcji. Jeśli nie, to z wierzchołka v_1 wychodzi jakaś krawędź do pewnego wierzchołka v_2. Jeśli v_2 jest ujściem, to koniec konstrukcji. Jeśli nie, to ... i tak dalej. Otrzymamy w ten sposób ciąg $v_1, v_2, v_3 \ldots$ taki, że dla każdego k ciąg $v_1 v_2 \ldots v_k$ jest drogą. Tak jak w pierwszym dowodzie, takie drogi nie mogą być dowolnie długie, a więc w którymś momencie osiągniemy ujście. ∎

Dowód 3. Opuść ten dowód, jeśli nie przestudiowałeś § 7.3. Za pomocą algorytmu ETYKIETOWANIE można utworzyć etykietowanie uporządkowane skończonego acyklicznego grafu skierowanego. Znaczy to, że jeśli istnieje droga z wierzchołka o etykiecie i do wierzchołka o etykiecie j, to $i > j$. Zwłaszcza wierzchołek o najmniejszej etykiecie nie może być początkiem żadnej krawędzi, a więc musi być ujściem. ∎

Podamy algorytm oparty na konstrukcji z drugiego dowodu, który dla danego skończonego acyklicznego grafu skierowanego G daje w wyniku ujście. Ten algorytm korzysta ze zbiorów **bezpośrednich następników**, zdefiniowanych w następujący sposób:

NAST(v) = $\{u \in V(G)$: istnieje krawędź z v do $u\}$. Te zbiory występowały również w algorytmie SORTOWANIE DRZEWA, który był podprogramem algorytmu ETYKIETOWANIE, wspomnianego wyżej w dowodzie 3. Zauważmy, że wierzchołek v jest ujściem wtedy i tylko wtedy, gdy NAST(v) jest zbiorem pustym. Te zbiory NAST(v) będą danymi dostarczonymi wraz z opisem grafu skierowanego G, gdy będzie wykonywany ten algorytm.

Algorytm UJŚCIE

{Dane: skończony acykliczny graf skierowany G}
{Wyniki: UJŚCIE(G), tzn. ujście grafu G}
Wybierz dowolny wierzchołek v w zbiorze $V(G)$.
Niech UJŚCIE(G) := v.
Dopóki NAST(v) $\neq \emptyset$, wykonuj
  wybierz wierzchołek u w zbiorze NAST(v)
  niech UJŚCIE(G) := u
  niech $v := u$. ■

PRZYKŁAD 2 Weźmy acykliczny graf skierowany G pokazany na rysunku 8.1. Zbiorami bezpośrednich następników są: NAST(t) = $\{u, w, x\}$, NAST(u) = $\{v\}$, NAST(v) = $\emptyset$, NAST(w) = $\{y\}$, NAST(x) = $\{y\}$, NAST(y) = $\emptyset$, NAST(z) = $\{w\}$. Jednym z możliwych ciągów wyborów w algorytmie UJŚCIE dla danego grafu G jest ciąg t, w, y. Innymi ciągami zaczynającymi się od t są t, x, y i t, u, v. Inny pierwszy wybór mógłby dać ciąg z, w, y. Moglibyśmy nawet mieć szczęście i wybrać ujście za pierwszym razem. Za każdym razem jednak algorytm UJŚCIE da jako wynik albo v, albo y. ■

Czas działania algorytmu UJŚCIE jest proporcjonalny do liczby wierzchołków, które on wybierze, zanim dotrze do ujścia, zatem dla grafu skierowanego o n wierzchołkach ten algorytm działa w czasie $O(n)$.

Jeśli wszystkie n wierzchołków grafu skierowanego zostało ponumerowanych liczbami od 1 do n w taki sposób, że $i > j$, jeśli tylko istnieje droga z wierzchołka i do wierzchołka v, to takie ponumerowanie wierzchołków nazywamy **etykietowaniem uporządkowanym**. Algorytm ETYKIETOWANIE z § 7.3 pokazywał, że acykliczny graf skierowany zawsze ma etykietowanie uporządkowane. Podamy teraz niezależny dowód tego faktu wykorzystujący ujścia. Ten dowód da inny algorytm, który nie będzie jednak tak efektywny, jak algorytm ETYKIETOWANIE.

Twierdzenie 3

Skończony acykliczny graf skierowany ma etykietowanie uporządkowane.

Dowód. Użyjemy indukcji ze względu na liczbę n wierzchołków; zauważmy, że twierdzenie jest oczywiste dla $n = 1$. Załóżmy indukcyjnie, że acykliczne grafy skierowane, mające mniej niż n wierzchołków, mają etykietowanie uporządkowane i weźmy acykliczny graf skierowany G o n wierzchołkach. Z twierdzenia 2 wynika, że graf G ma ujście, na przykład s. Przyporządkujmy wierzchołkowi s liczbę 1. Utwórzmy nowy graf H taki, że $V(H) = V(G)\backslash\{s\}$ i taki, że jego krawędziami są te krawędzie grafu G, których końcem nie jest wierzchołek s. Ponieważ graf G nie ma cykli, więc i graf H nie ma cykli. Ponieważ graf H ma tylko $n - 1$ wierzchołków, więc z założenia indukcyjnego ma etykietowanie uporządkowane. Zwiększmy wartość każdej etykiety o 1, tak by wierzchołki grafu H były ponumerowane liczbami 2, 3, ..., n. Wierzchołek s ma numer 1, a więc każdy wierzchołek grafu G ma numer między 1 i n.

Załóżmy teraz, że istnieje w grafie G droga od wierzchołka i do wierzchołka j. Jeśli ta droga leży całkowicie w grafie H, to $i > j$, gdyż graf H był ponumerowany we właściwy sposób. W przeciwnym przypadku którymś wierzchołkiem na tej drodze jest s, ale ponieważ s jest ujściem, musi on być ostatnim wierzchołkiem, to znaczy j. Ale wtedy $j = 1$, a więc również w tym przypadku $i > j$. Stąd wynika, że graf G, mający n wierzchołków, ma etykietowanie uporządkowane. Z zasady indukcji matematycznej wynika, że twierdzenie zachodzi dla wszystkich n. ∎

Ideę dowodu twierdzenia 3 można wykorzystać do zaprojektowania procedury, która tworzy etykietowanie uporządkowane danego acyklicznego grafu skierowanego.

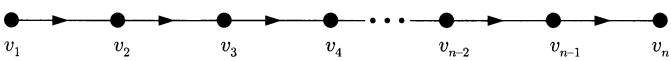
Algorytm NUMEROWANIE WIERZCHOŁKÓW

{Dane: skończony acykliczny graf skierowany G mający n wierzchołków}
{Wyniki: etykietowanie uporządkowane wierzchołków grafu G}
Niech $V := V(G)$ i $E := E(G)$.
Dopóki $V \neq \emptyset$, wykonuj
  niech H będzie grafem, którego zbiorem wierzchołków jest V i zbiorem krawędzi E

zastosuj algorytm UJŚCIE do grafu H {otrzymując ujście grafu H}
nadaj wierzchołkowi UJŚCIE(H) etykietę $n - |V| + 1$
usuń wierzchołek UJŚCIE(H) ze zbioru V i wszystkie krawędzie wchodzące do niego ze zbioru E. ∎

W każdym przebiegu pętli „dopóki" zostaje usunięty jeden wierzchołek z wyjściowego zbioru wierzchołków $V(G)$, a więc algorytm musi zakończyć działanie i kiedy zatrzyma się, każdy wierzchołek będzie miał etykietę. Etykietami są liczby 1, 2, ..., n. Ten algorytm wywołuje algorytm UJŚCIE jako podprogram. Pouczające może być zastosowanie tego algorytmu do ponumerowania wierzchołków grafu z rysunku 8.1. Zobacz również ćwiczenie 14, by znaleźć procedurę, która numeruje wierzchołki zaczynając od liczby n.

Jeżeli będziemy szacować czas działania algorytmu NUMEROWANIE WIERZCHOŁKÓW, zakładając, że dla znalezienia potrzebnych ujść wywołuje on algorytm UJŚCIE, najlepsze oszacowanie, jakie otrzymamy, będzie postaci $O(n^2)$, gdzie n jest liczbą wierzchołków grafu G. Ten algorytm nie jest efektywny, gdyż podprogram UJŚCIE może wielokrotnie badać ten sam wierzchołek. Algorytm ETYKIETOWANIE z § 7.3 jest bardziej efektywny, chociaż trudniej wykonać go ręcznie.



Rysunek 8.2

PRZYKŁAD 3 Weźmy graf skierowany przedstawiony na rysunku 8.2. Algorytm NUMEROWANIE WIERZCHOŁKÓW mógłby zaczynać od wierzchołka v_1 za każdym razem, gdy wywołuje algorytm UJŚCIE. Za pierwszym razem znalazłby ujście v_n, przydzieliłby mu etykietę i usunął je. Za drugim razem znalazłby ujście v_{n-1}, przydzieliłby mu etykietę i usunął je itd. Liczba wierzchołków, które by zbadał, wynosi

$$n + (n-1) + \ldots + 2 + 1 = \frac{1}{2}n(n+1) > \frac{1}{2}n^2.$$

Przypuśćmy dla kontrastu, że algorytm ETYKIETOWANIE z § 7.3 rozpoczął działanie od wierzchołka v_1. Zamiast rozpoczynać od niego za każdym razem, po osiągnięciu wierzchołka v_n powróciłby do wierzchołka v_{n-1}, przydzielił mu etykietę, powrócił do wierzchołka v_{n-2}, przydzielił mu etykietę itd., badając w ten sposób $2n$ wierzchołków. ∎

Drugi dowód twierdzenia 2 polegał na skonstruowaniu drogi od danego wierzchołka v do ujścia. Wierzchołek u nazwiemy **wierzchołkiem osiągalnym** z wierzchołka v w grafie G, jeśli istnieje w grafie G droga długości co najmniej 1 z v do u. Definiujemy wtedy

$$R(v) = \{u \in V(G)\text{: wierzchołek } u \text{ jest osiągalny z wierzchołka } v\}.$$

Zatem $R(v) = \emptyset$ wtedy i tylko wtedy, gdy wierzchołek v jest ujściem, a drugi dowód twierdzenia 2 pokazywał naprawdę, że w acyklicznym grafie skierowanym każdy niepusty zbiór $R(v)$ zawiera co najmniej jedno ujście. Pojęcie osiągalności pojawiło się już w tej książce w trochę innym kontekście. Para wierzchołków (v, w) należy do relacji osiągalności zdefiniowanej w § 3.2 wtedy i tylko wtedy, gdy wierzchołek w jest osiągalny z wierzchołka v. Również w § 7.3 korzystaliśmy ze zbioru DOST(v) wierzchołków dostępnych z wierzchołka v, tzn. zbioru $R(v)$ wraz z samym wierzchołkiem v.

PRZYKŁAD 4 (a) W grafie skierowanym przedstawionym na rysunku 8.2 mamy

$$R(v_k) = \{v_j\text{: } k < j \leq n\} \quad \text{dla} \quad k = 1, 2, \ldots, n.$$

Jest to prawdą również dla $k = n$, gdyż wtedy zbiory po obu stronach znaku równości są puste.

(b) W grafie skierowanym pokazanym na rysunku 8.1 mamy $R(t) = \{u, v, w, x, y\}$, $R(u) = \{v\}$, $R(v) = \emptyset$, $R(w) = \{y\}$, $R(y) = \emptyset$ i $R(z) = \{w, y\}$. ∎

Nawet jeżeli graf G nie jest acykliczny, zbiory $R(v)$ mogą mieć znaczenie. Jak zobaczymy w § 11.4, wyznaczenie wszystkich zbiorów $R(v)$ sprowadza się do znalezienia domknięcia przechodniego pewnej relacji. W paragrafie 8.4 będziemy badać algorytmy znajdujące zbiory $R(v)$ oraz odpowiadające na inne pytania dotyczące grafów.

Podamy następnie wersję twierdzenia Eulera z § 6.2 dla grafów skierowanych. Ponieważ graf skierowany może być traktowany jako graf nieskierowany, pojęcie **stopnia wierzchołka** nadal ma sens. Aby wziąć pod uwagę to, że krawędzie są skierowane, uściślimy to pojęcie. Jeśli v jest wierzchołkiem grafu skierowanego G, to **stopniem wejściowym** v (indeg(v)) nazywamy liczbę krawędzi grafu G, których końcem jest wierzchołek v, a **stopniem wyjściowym** v (outdeg(v)) nazywamy liczbę krawędzi, których początkiem jest v. Wtedy

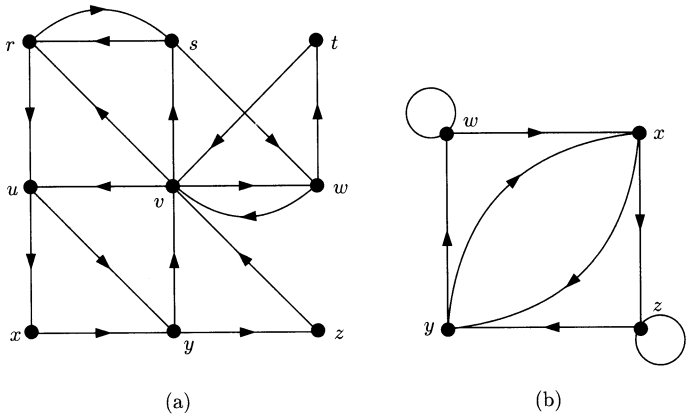
$$\text{indeg}(v) + \text{outdeg}(v) = \deg(v)$$

dla wszystkich wierzchołków v. Pętla jest w tym wzorze liczona dwukrotnie, raz przy wejściu, raz przy wyjściu z wierzchołka.

PRZYKŁAD 5

(a) Weźmy ponownie acykliczny graf skierowany przedstawiony na rysunku 8.1. Ujścia v i y mają stopień wyjściowy 0, a źródła t i z mają stopień wejściowy 0. Inne stopnie wejściowe i wyjściowe są łatwe do odczytania. Na przykład $\text{indeg}(u) = 1$ i $\text{outdeg}(u) = 2$. Suma wszystkich stopni wejściowych jest dokładnie równa liczbie krawędzi, mianowicie 8 i jest także równa sumie wszystkich stopni wyjściowych.

(b) Grafy skierowane przedstawione na rysunku 8.3 mają następującą szczególną własność: stopień wejściowy i stopień wyjściowy każdego wierzchołka są równe. ■



(a) (b)

Rysunek 8.3

Dowód następnego twierdzenia jest bardzo podobny do dowodów twierdzeń 1 i 2 z § 6.2 dla grafów nieskierowanych. Jednakże napisaliśmy ten dowód dla wygody Czytelnika, a także dlatego, że wykorzystamy to twierdzenie w § 12.1.

Twierdzenie 4. Twierdzenie Eulera dla grafów skierowanych

Załóżmy, że skończony graf skierowany G jest spójny, gdy traktujemy go jako graf nieskierowany. Wówczas istnieje droga (skierowana) zamknięta w grafie G, przechodząca przez wszystkie krawędzie grafu G wtedy i tylko wtedy, gdy $\text{indeg}(v) = \text{outdeg}(v)$ dla każdego wierzchołka v.

Dowód. Przypuśćmy, że taka droga zamknięta istnieje. Wychodzimy z pewnego wierzchołka i podążamy wzdłuż tej drogi,

usuwając za każdym razem krawędź, którą przeszliśmy. W chwili przechodzenia przez wierzchołek usuwamy jedną krawędź wchodzącą do niego i jedną wychodzącą z niego lub też usuwamy pętlę. W każdym przypadku to usuwanie zmniejsza zarówno stopień wejściowy, jak i wyjściowy o 1. W końcu wszystkie krawędzie zostaną usunięte, a więc wszystkie stopnie wejściowe i wyjściowe będą równe 0. Zatem na początku stopień wejściowy i stopień wyjściowy każdego wierzchołka musiały być równe.

Przypuśćmy teraz, że stopień wejściowy i stopień wyjściowy każdego wierzchołka są równe. Możemy założyć, że istnieje więcej niż jeden wierzchołek, gdyż w przeciwnym przypadku twierdzenie jest trywialne. Zatem każdy wierzchołek ma dodatni stopień wejściowy i wyjściowy. Niech $v_1 v_2 \ldots v_n$ będzie ciągiem wierzchołków najdłuższej możliwej drogi skierowanej, mającej różne krawędzie. Zobaczmy teraz, w jaki sposób zmienią się stopień wejściowy i stopień wyjściowy wierzchołka v_n, gdy usuniemy z grafu krawędzie tej drogi. Za każdym razem, gdy wierzchołek v_n pojawia się w ciągu $v_2, \ldots, v_{n-1}$, stopień wejściowy i stopień wyjściowy zmniejszają się o 1, a przy ostatnim pojawieniu się v_n, tylko stopień wejściowy zmniejsza się o 1. Gdyby $v_n \neq v_1$, musiałaby pozostać nie wykorzystana krawędź, której początkiem jest v_n i którą moglibyśmy dodać do naszej najdłuższej drogi, otrzymując w ten sposób jeszcze dłuższą drogę z różnymi krawędziami. Ta sprzeczność pokazuje, że $v_n = v_1$.

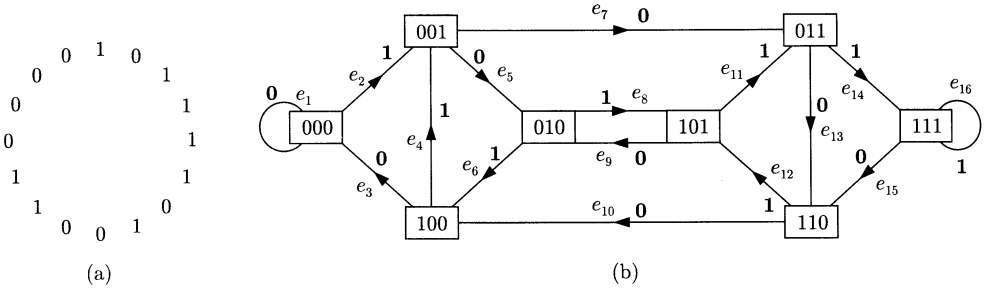
Wyjaśnimy następnie, dlaczego nasza najdłuższa droga przechodzi przez każdy wierzchołek grafu G. Gdyby tak nie było, to ponieważ graf G jest spójny jako graf nieskierowany, istnieje *nieskierowana* droga od pewnego nie odwiedzonego wierzchołka do jakiegoś wierzchołka w zbiorze $\{v_1, \ldots, v_n\}$. Pierwsza krawędź tej drogi nie należy do naszej najdłuższej drogi. Ostatnia krawędź e tej drogi, która nie należy do naszej najdłuższej drogi, musi mieć swój początek lub koniec w zbiorze $\{v_1, \ldots, v_n\}$, niech będzie to na przykład wierzchołek v_i. Oczywiście ta krawędź jest w rzeczywistości skierowana. Jeśli v_i jest początkiem krawędzi e, to przechodzimy tę naszą najdłuższą drogę od v_i do v_i i na końcu dodajemy krawędź e, aby otrzymać w ten sposób drogę dłuższą, o różnych krawędziach, co daje nam sprzeczność. Jeśli v_i jest końcem krawędzi e, to możemy zacząć naszą drogę od krawędzi e i następnie przejść dookoła naszą najdłuższą drogę od v_i do v_i, znów otrzymać drogę dłuższą z różnymi krawędziami — znów mamy sprzeczność.

Wreszcie pokażemy, że ta najdłuższa droga przechodzi przez wszystkie krawędzie grafu G. Jeśli nie przechodzi przez pewną

krawędź, to bierzemy jej początek lub koniec v_i. Dokładnie tak, jak w poprzednim akapicie, możemy dodać tę drogę na początku lub na końcu najdłuższej drogi z v_i do v_i, aby otrzymać w ten sposób drogę dłuższą, co znów daje sprzeczność. ∎

PRZYKŁAD 6 (a) Każdy z grafów skierowanych pokazanych na rysunku 8.3 ma drogę zamkniętą przechodzącą przez wszystkie krawędzie (ćwiczenie 9).

(b) **Ciągiem de Bruijna rzędu** n nazywamy cykliczne ustawienie 2^n cyfr 0 i 1, takie, że każdy ciąg długości n występuje dokładnie jeden raz w tym ustawieniu jako n kolejnych cyfr. Na rysunku 8.4(a) widzimy takie ustawienie dla $n = 4$.



Rysunek 8.4

Ustawienie z rysunku 8.4(a) otrzymaliśmy korzystając z grafu skierowanego pokazanego na rysunku 8.4(b); wymaga to pewnego wyjaśnienia. Zbiór $2^{n-1} = 8$ wierzchołków tego grafu skierowanego składa się ze wszystkich ciągów 0 i 1, długości $n - 1 = 3$. Krawędź skierowana łączy dwa takie ciągi wtedy, gdy ostatnie dwie cyfry początkowego wierzchołka pokrywają się z pierwszymi dwiema cyframi końcowego wierzchołka. Przypisujemy takiej krawędzi etykietę będącą ostatnią cyfrą wierzchołka końcowego. Zatem krawędzie e_1, e_3, e_5, e_6 itd. mają etykietę 0, a krawędzie e_2, e_4 itp. mają etykietę 1. Mówiąc inaczej, etykieta krawędzi na drodze długości 1 jest ostatnią cyfrą końcowego wierzchołka tej drogi. Wynika stąd, że etykiety dwóch krawędzi na drodze długości 2 są ostatnimi dwiema cyframi wierzchołka końcowego tej drogi, w tej samej kolejności. Podobnie, etykiety krawędzi na drodze długości 3 są wszystkimi trzema cyframi wierzchołka końcowego, w tej samej kolejności.

Zauważmy, że każdy wierzchołek tego grafu skierowanego ma stopień wejściowy równy 2 i stopień wyjściowy równy 2. Ponieważ ten graf jest spójny, więc z twierdzenia Eulera wynika, że istnieje

droga zamknięta przechodząca przez każdą krawędź dokładnie jeden raz. Twierdzimy, że etykiety tych $2^n = 16$ krawędzi tworzą ciąg de Bruijna rzędu $n = 4$. Na przykład etykiety krawędzi drogi zamkniętej

$$e_2, e_5, e_8, e_{11}, e_{14}, e_{16}, e_{15}, e_{12}, e_9, e_6, e_4, e_7, e_{13}, e_{10}, e_3, e_1$$

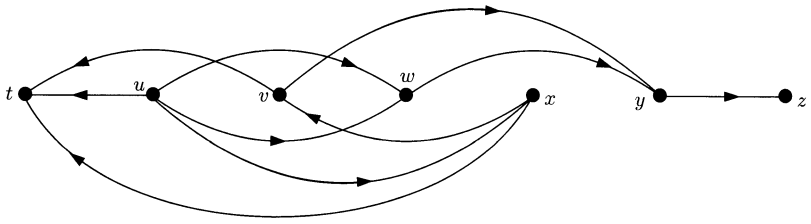
dają ustawienie cykliczne z rysunku 8.4(a), zaczynające się u góry i obiegane zgodnie z ruchem wskazówek zegara. Ponieważ istnieje w tym ustawieniu tylko 16 różnych ciągów kolejnych cyfr, więc wystarczy pokazać, iż każdy ciąg $d_1d_2d_3d_4$ zer i jedynek występuje w tym ustawieniu cyklicznym. Ponieważ ciąg $d_1d_2d_3$ jest początkiem krawędzi mającej etykietę 0, a także krawędzi mającej etykietę 1, więc pewna krawędź na tej drodze ma jako początek wierzchołek $d_1d_2d_3$ i etykietę d_4. Jak zauważyliśmy w poprzednim akapicie, trzy krawędzie poprzedzające ją mają etykiety d_1, d_2 i d_3 w tej kolejności. Zatem etykietami tych czterech kolejnych krawędzi są d_1, d_2, d_3 i d_4, czego należało dowieść. ∎

Zauważmy, że jeśli skończony graf skierowany nie jest spójny, ale spełnia warunek dotyczący stopni, sformułowany w twierdzeniu Eulera, to twierdzenie Eulera stosuje się do każdej z jego spójnych składowych.

ĆWICZENIA DO § 8.1

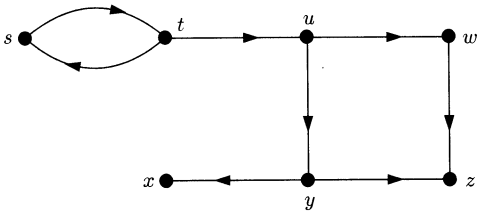
1. Znajdź ujścia i źródła w grafie skierowanym przedstawionym na rysunku 8.5.

2. (a) Podaj zbiory bezpośrednich następników NAST(v) dla wszystkich wierzchołków grafu skierowanego pokazanego na rysunku 8.5.
(b) Jaką wartość UJŚCIE(G) poda algorytm UJŚCIE, jeśli wierzchołkiem początkowym był wierzchołek w?
(c) Które ujścia grafu G znajdują się w zbiorze $R(x)$?



Rysunek 8.5

3. Weźmy graf G przedstawiony na rysunku 8.6.
(a) Znajdź $R(v)$ dla każdego wierzchołka v w zbiorze $V(G)$.
(b) Znajdź wszystkie ujścia grafu G.
(c) Czy G jest grafem acyklicznym?



Rysunek 8.6

4. Czy algorytm UJŚCIE można stosować do grafów skierowanych, które mają cykle? Wyjaśnij to.

5. Weźmy graf skierowany G z następującymi zbiorami bezpośrednich następników: $\text{NAST}(r) = \{s, u\}$, $\text{NAST}(s) = \emptyset$, $\text{NAST}(t) = \{r, w\}$, $\text{NAST}(u) = \emptyset$, $\text{NAST}(w) = \{r, t, x, y\}$, $\text{NAST}(x) = \emptyset$, $\text{NAST}(y) = \{w, z\}$ oraz $\text{NAST}(z) = \emptyset$.
(a) Narysuj taki graf skierowany.
(b) Czy te zbiory bezpośrednich następników $\text{NAST}(v)$ wyznaczają w sposób jednoznaczny $E(G)$? Wyjaśnij to.
(c) Znajdź wszystkie ujścia w grafie G.
(d) Znajdź drogi z wierzchołka w do trzech różnych ujść w tym grafie skierowanym.

6. Podaj dwa etykietowania uporządkowane dla acyklicznego grafu skierowanego przedstawionego na rysunku 8.1.

7. Podaj dwa etykietowania uporządkowane dla acyklicznego grafu skierowanego przedstawionego na rysunku 8.5.

8. Turniejem jest graf skierowany, w którym każde dwa wierzchołki łączy dokładnie jedna krawędź. (Traktuj (x, y) jako krawędź, zakładając, że x pokonał y).
(a) Podaj przykład turnieju z czterema wierzchołkami.
(b) Pokaż, że turniej nie może mieć dwóch ujść.
(c) Czy turniej z cyklem ma ujście? Wyjaśnij to.
(d) Czy chciałbyś być ujściem w turnieju?

9. Dla każdego grafu skierowanego z rysunku 8.3 podaj drogę zamkniętą przechodzącą wszystkie jego krawędzie.

10. Użyj grafu skierowanego z rysunku 8.4(b), aby podać dwa istotnie różne ciągi de Bruijna, które są też istotnie różne od ciągu przedstawionego na rysunku 8.4(a). (Dwa ciągi ustawione cyklicznie traktujemy

jako te same, jeśli jeden można otrzymać z drugiego przez obrót lub odbicie symetryczne).

11. (a) Utwórz graf skierowany, podobny do grafu z rysunku 8.4(b), którego można użyć do znalezienia ciągów de Bruijna rzędu 3.
(b) Wykorzystaj swój graf skierowany do narysowania dwóch ciągów de Bruijna rzędu 3.

12. (a) Wyjaśnij, jak musi być zmodyfikowane wyjaśnienie w przykładzie 6(b), aby pokazać, że istnieją ciągi de Bruijna każdego rzędu $n \geq 3$.
(b) Łatwo narysować ciąg de Bruijna rzędu 2. Zrób to.

13. **Odwróceniem** grafu skierowanego G jest graf skierowany $\widehat{G}$ otrzymany przez odwrócenie wszystkich strzałek w grafie G. To znaczy, że $V(\widehat{G}) = V(G)$, $E(\widehat{G}) = E(G)$ oraz jeśli $\gamma(e) = (x, y)$, to $\widehat{\gamma}(e) = (y, x)$. Wykorzystaj graf $\widehat{G}$ oraz twierdzenie 2, aby pokazać, że jeśli graf G jest acykliczny (i skończony), to G ma źródło.

14. (a) Zmodyfikuj algorytm NUMEROWANIE WIERZCHOŁKÓW używając źródeł zamiast ujść, aby otrzymać algorytm, który numeruje elementy zbioru $V(G)$ w porządku malejącym.
(b) Użyj swojego algorytmu do ponumerowania grafu skierowanego przedstawionego na rysunku 8.1.

15. (a) Przypuśćmy, że skończony acykliczny graf skierowany ma po prostu jedno ujście. Pokaż, że istnieje droga do ujścia od każdego wierzchołka.
(b) Jak brzmi odpowiednie stwierdzenie dotyczące źródeł?

16. Niech G będzie grafem skierowanym. Definiujemy w zbiorze $V(G)$ relację $\sim$ w następujący sposób: $x \sim y$, jeśli $x = y$ lub wierzchołek x jest osiągalny z wierzchołka y i y jest osiągalny z x.
(a) Pokaż, że relacja $\sim$ jest relacją równoważności.
(b) Znajdź klasy równoważności dla grafu skierowanego przedstawionego na rysunku 8.6.
(c) Opisz relację $\sim$, gdy graf G jest acykliczny.

17. (a) Pokaż, że w twierdzeniu 1 i wniosku 1 droga, w której wierzchołki się nie powtarzają, może być zbudowana z krawędzi danej drogi. Zatem każda droga zamknięta zawiera co najmniej jeden cykl.
(b) Pokaż, że jeśli u i v są wierzchołkami grafu skierowanego i jeśli istnieje droga z u do v, to istnieje droga z u do v, w której żadna krawędź się nie powtarza. (Rozpatrz przypadki $u = v$ i $u \neq v$).

18. Niech G będzie grafem skierowanym.
(a) Pokaż, że jeśli wierzchołek u jest osiągalny z wierzchołka v, to $R(u) \subseteq R(v)$.
(b) Podaj inny dowód twierdzenia 2, wybierając wierzchołek $v \in V(G)$ tak, by wielkość $|R(v)|$ była możliwie mała.

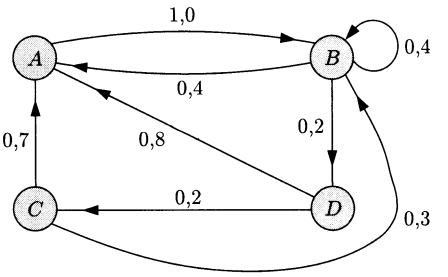
(c) Czy twój dowód z ćwiczenia (b) prowadzi do jakiejś przydatnej procedury konstruktywnej? Odpowiedź uzasadnij.

§ 8.2. Grafy skierowane z wagami

W wielu zastosowaniach grafów skierowanych chcemy wiedzieć, czy dany wierzchołek v można osiągnąć z innego wierzchołka u, tzn. czy można dostać się z u do v podążając za strzałkami. Na przykład przypuśćmy, że każdy wierzchołek odpowiada pewnemu stanowi automatu, takiemu jak POBIERZ, ODROCZ lub WYKONAJ, a krawędź od s do t odpowiada temu, że automat może przejść od stanu s do stanu t w odpowiedzi na pewne dostarczone dane. Jeśli automat znajduje się w stanie u, to czy może kiedyś później być w stanie v? Odpowiedź jest twierdząca wtedy i tylko wtedy, gdy w tym grafie skierowanym istnieje droga z u do v.

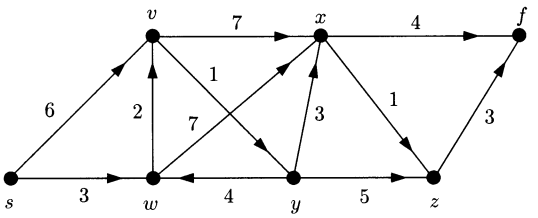
Przypuśćmy teraz, że każde przejście od jednego stanu do drugiego, tzn. przejście każdej krawędzi w grafie skierowanym wymaga poniesienia pewnych kosztów. Ten koszt może być kosztem finansowym, może to być miara czasu potrzebnego do przeprowadzenia tej zmiany lub też może mieć jakieś inne znaczenie. Możemy teraz zapytać o drogę z u do v mającą najmniejszy łączny koszt, otrzymany przez dodanie kosztów wszystkich krawędzi na tej drodze.

Jeśli wszystkie krawędzie kosztują tyle samo, to najtańsza droga jest po prostu drogą najkrótszą. Na ogół jednak koszty krawędzi mogą się różnić. Graf skierowany bez krawędzi wielokrotnych nazywamy **grafem skierowanym z wagami**, jeśli każdej krawędzi jest przyporządkowana pewna liczba, nazywana **wagą** tej krawędzi. W konkretnych zastosowaniach może ona być trafniej nazywana „kosztem", „długością" czy „pojemnością" lub też może mieć jeszcze inną interpretację. Na ogół zakłada się, że wagi są nieujemne, ale wiele faktów dotyczących grafów skierowanych z wagami jest prawdziwych również bez tego ograniczenia. Wagi krawędzi w grafie skierowanym G będą wartościami pewnej funkcji W o dziedzinie $E(G)$ i wartościach w zbiorze $\mathbb{R}$; przez $W(e)$ będziemy oznaczać wagę krawędzi e. **Wagą drogi** $e_1e_2\ldots e_m$ w grafie G nazwiemy wtedy sumę $\sum_{i=1}^{m} W(e_i)$. Ponieważ graf skierowany z wagami nie ma krawędzi wielokrotnych, więc możemy zakładać, że $E(G) \subseteq V(G) \times V(G)$ i oznaczać symbolem $W(u,v)$ wagę krawędzi (u,v) prowadzącej z wierzchołka u do wierzchołka v.



Rysunek 8.7

PRZYKŁAD 1 (a) Graf skierowany pokazany na rysunku 8.7 został wzięty z rysunku 3.2(c) w § 3.2, gdzie opisywał szczura poruszającego się między klatkami. Mógłby on równie dobrze opisywać automat mający stany A, B, C i D, a liczby stojące obok krawędzi (wagi krawędzi) mogłyby oznaczać liczby mikrosekund potrzebnych do przejścia od stanu początkowego do stanu końcowego. Przy takiej interpretacji, przejście od stanu C do stanu B trwa 0,3 mikrosekundy, przejście ze stanu D do stanu A poprzez stan C trwa 0,9 mikrosekund, przejście bezpośrednio ze stanu D do stanu A trwa 0,8 mikrosekund i pozostawanie w stanie B w odpowiedzi na pewne wprowadzone dane trwa 0,4 mikrosekundy. W jakim najkrótszym czasie można przejść od stanu D do stanu B?



Rysunek 8.8

(b) Na rysunku 8.8 pokazany jest bardziej skomplikowany przykład. W tym przypadku najkrótsze drogi $svxf$ i $swxf$ z wierzchołka s do wierzchołka f mają odpowiednio wagi $6+7+4=17$ i $3+7+4=14$, podczas gdy dłuższa droga $swvyxzf$ ma wagę $3+2+1+3+1+3=13$, mniejszą od dwóch poprzednich wag. Nie ma bezpośredniego związku między długością i wagą drogi. W tym przykładzie widzimy też drogę swv z wierzchołka s do wierzchołka v, która ma wagę mniejszą od wagi krawędzi z s do v.

Ten graf ma cykl $wvyw$. Oczywiście cały ten cykl nie może wchodzić w skład drogi o najmniejszej wadze, ale jego fragmenty

mogą. Na przykład wvy jest drogą o najmniejszej wadze z wierzchołka w do wierzchołka y i krawędź yw jest drogą o najmniejszej wadze z wierzchołka y do wierzchołka w. ∎

Jeśli chcemy, możemy przedstawić funkcję wagi W w postaci tablicy, oznaczając wiersze i kolumny tablicy za pomocą elementów zbioru $V(G)$ i pisząc wartość $W(u, v)$ na przecięciu wiersza u z kolumną v.

PRZYKŁAD 2 Tablica grafu skierowanego z rysunku 8.8 jest pokazana na rysunku 8.9(a). Liczby znajdują się w miejscach odpowiadających krawędziom tego grafu. Tablica pokazana na rysunku 8.9(a) zawiera wystarczająco wiele informacji, by móc odtworzyć ten graf skierowany z wagami, gdyż widzimy w tej tablicy, skąd i dokąd prowadzą krawędzie i jakie mają wagi. Na rysunku 8.9(b) mamy tablicę funkcji wagi W^*, gdzie $W^*(u, v)$ oznacza najmniejszą wagę drogi prowadzącej z wierzchołka u do wierzchołka v, o ile taka droga istnieje. ∎

W	s	v	w	x	y	z	f
s		6	3				
v				7	1		
w		2		7			
x						1	4
y			4	3		5	
z							3
f							

(a)

W^*	s	v	w	x	y	z	f
s		5	3	9	6	10	13
v		7	5	4	1	5	8
w		2	7	6	3	7	10
x						1	4
y		6	4	3	7	4	7
z							3
f							

(b)

Rysunek 8.9

Najmniejszą wagę drogi (długości co najmniej 1) prowadzącej z wierzchołka u do wierzchołka v będziemy nazywać **wagą minimalną** z u do v i oznaczać symbolem $W^*(u, v)$, jak w przykładzie 2. Będziemy także nazywać drogę z u do v, mającą tę minimalną wagę, **drogą minimalną**.

Założenie, że nasze grafy skierowane z wagami nie mają pętli, nie jest istotnym ograniczeniem (dlaczego?), a więc moglibyśmy w ogóle nie przejmować się wartościami $W(u, u)$ lub przyjąć, że $W(u, u) = 0$ dla wszystkich wierzchołków u. Okazuje się jednak, że uzyskamy więcej informacji, jeśli postąpimy inaczej, korzystając z następującego pomysłu.

Weźmy wierzchołki u i v nie połączone krawędzią w grafie G. Możemy utworzyć nową fikcyjną krawędź z u do v, jakiejś niezwykle dużej wagi, tak dużej, że ta krawędź nigdy nie zo-

stanie wybrana w żadnej drodze minimalnej, o ile tylko istnieje prawdziwa droga z u do v. Przypuśćmy, że utworzymy takie fikcyjne krawędzie dla każdej krawędzi, której brak w grafie G. Jeśli kiedykolwiek stwierdzimy, że jakaś droga w tym rozszerzonym grafie ma ogromną wagę, to będziemy wiedzieć, że ta droga zawiera co najmniej jedną krawędź fikcyjną, a więc w pierwotnym grafie G nie było drogi łączącej te same dwa wierzchołki. Wygodnie jest przyjąć oznaczenie $W(u, v) = \infty$, jeśli nie ma krawędzi z u do v w grafie G; symbol $W^*(u, v) = \infty$ oznacza, że w grafie G nie ma drogi z wierzchołka u do wierzchołka v. Mamy następujące reguły wykonywania działań na symbolu ∞: $\infty + x = x + \infty = \infty$ dla każdego x oraz $a < \infty$ dla każdej liczby rzeczywistej a.

Korzystając z tych oznaczeń możemy napisać $W(u, u)=a$, jeśli istnieje pętla w wierzchołku u, mająca wagę a, oraz $W(u, u) = \infty$ w przeciwnym przypadku. Wtedy $W^*(u, u) < \infty$ znaczy, że istnieje droga (długości co najmniej 1) z wierzchołka u do tego samego wierzchołka w grafie G i $W^*(u, u) = \infty$ znaczy, że takiej drogi nie ma. Graf skierowany G jest acykliczny wtedy i tylko wtedy, gdy $W^*(v, v) = \infty$ dla każdego wierzchołka v.

PRZYKŁAD 3 Korzystając z tych oznaczeń, możemy wpisać znak ∞ w puste miejsca w tablicach na rysunkach 8.9(a) i 8.9(b). ∎

W przykładzie 2 po prostu podaliśmy wartości funkcji W^*, a przykład był na tyle mały, by można było łatwo sprawdzić te wartości. Dla bardziej skomplikowanych grafów skierowanych wyznaczenie wartości W^* oraz dróg minimalnych może być nietrywialnym zadaniem. W paragrafach 8.3 i 8.4 podamy algorytmy, za pomocą których można znaleźć zarówno W^*, jak i odpowiednie drogi minimalne, ale do tego czasu będziemy się tak długo przyglądać rysunkowi grafu, aż odpowiedź będzie oczywista. Dla małych grafów skierowanych ta metoda jest równie dobra jak każda inna.

W wielu sytuacjach opisanych za pomocą grafów skierowanych z wagami szczególnie nas interesują wagi i drogi minimalne. Jednakże, jedną klasą problemów, w których stosuje się grafy skierowane z wagami i w których drogi minimalne nie mają znaczenia, jest szeregowanie zadań składających się z pojedynczych kroków. Takie problemy szeregowania zadań mają dość duże znaczenie w przemyśle i biznesie. Następujący przykład pokazuje, w jakich sytuacjach mamy z nimi do czynienia.

PRZYKŁAD 4 Przypuśćmy, że kucharz chce przygotować proste danie ryż z curry. Przepis na curry wymaga wykonania następujących czynności:

(a) Pokroić mięso w kawałki — ok. 10 minut.
(b) Posiekać cebulę w kostkę — ok. 2 minut za pomocą robota.
(c) Obrać ziemniaki i pokroić je na ćwiartki — ok. 5 minut.
(d) Zamarynować mięso z cebulą i przyprawami — ok. 30 minut.
(e) Rozgrzać olej — 4 minuty. Usmażyć ziemniaki — 15 minut. Podsmażyć kminek — 2 minuty.
(f) Usmażyć zamarynowane mięso — 4 minuty.
(g) Zapiec podsmażone mięso i ziemniaki — 60 minut.

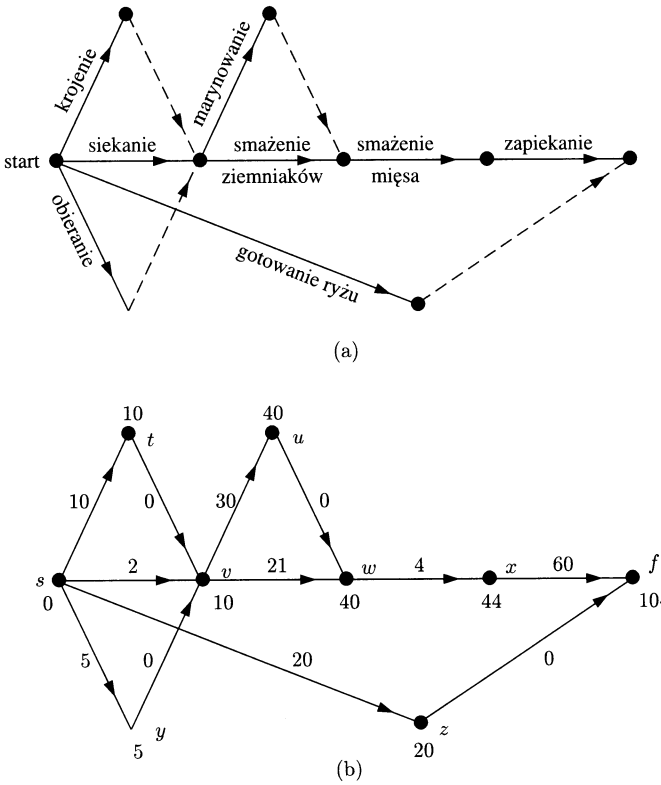
Ponadto,

(h) Ugotować ryż — 20 minut.

Zgrupowaliśmy trzy czynności w punkcie (e), ponieważ muszą one być wykonane po kolei. Niektóre z innych czynności mogą być wykonane jednocześnie, jeśli kucharzowi ktoś pomaga. Przypuśćmy, że nasz kucharz ma tylu pomocników, ilu potrzebuje.

Na rysunku 8.10(a) pokazany jest graf skierowany, który przedstawia kolejność kroków i możliwości działania jednoczesnego. Krojenie mięsa, siekanie cebuli, obieranie ziemniaków i gotowanie ryżu mogą być wykonane w tym samym czasie. Przerywane strzałki po krojeniu i obieraniu oznaczają, że marynowanie mięsa i smażenie ziemniaków nie mogą się rozpocząć, zanim nie zakończy się krojenie i obieranie. Pozostałe dwie przerywane strzałki mają podobne znaczenie. Graf ten został ponownie przerysowany na rysunku 8.10(b) wraz z wagami krawędzi oznaczającymi czas wykonania (na chwilę nie zajmujmy się liczbami przy wierzchołkach). Wierzchołki oznaczają etapy częściowego wykonania całego procesu, od lewej do prawej. W tym przypadku droga minimalna od lewej do prawej ma wagę 20, ale potrzeba dużo więcej czasu, aby przygotować ten posiłek niż 20 minut, w czasie których gotuje się ryż. Waga minimalna nic tu nie daje. Ważnym pytaniem jest tutaj: jaki jest najmniejszy całkowity czas potrzebny do wykonania wszystkich czynności w tym procesie?

Aby odpowiedzieć na to pytanie, najpierw zbadamy wierzchołki od lewej do prawej. Przypuśćmy, że rozpoczynamy w wierzchołku s w chwili 0. W którym najwcześniejszym momencie możemy zakończyć krojenie, siekanie i obieranie oraz dotrzeć do wierzchołka v? Oczywiście po 10 minutach, ponieważ musimy poczekać na pokrojenie niezależnie od tego, jak wcześnie rozpocz-



Rysunek 8.10

niemy siekanie i obieranie. Tak naprawdę, 10 jest *największą* wagą drogi z s do v. Teraz, w którym najwcześniejszym momencie możemy dotrzeć do wierzchołka w? Najkrótszym czasem, w którym możemy się dostać z v do w, jest 30 minut (największa waga drogi z v do w), więc najkrótszym czasem, w którym można dostać się z wierzchołka s do w, jest $10+30 = 40$ minut. Podobnie, najkrótszym czasem, w którym możemy wykonać cały proces i dotrzeć do wierzchołka f, jest $40 + 4 + 60 = 104$ minuty od momentu rozpoczęcia.

W każdym przypadku, najkrótszym czasem dojścia do danego wierzchołka jest największa waga drogi z s do tego wierzchołka. Liczby występujące obok wierzchołków na rysunku 8.10(b) podają te najkrótsze czasy. ■

Acykliczny graf skierowany z nieujemnymi wagami, jednym źródłem i jednym ujściem, taki jak graf skierowany przedstawiony na rysunku 8.10, nazywamy **siecią zdarzeń**. W dalszej części tego paragrafu przyjmiemy, że mamy do czynienia z siecią

zdarzeń G ze źródłem s (start) i ujściem f (ang. finish — meta). Dla wierzchołków u i v sieci G **drogą maksymalną** z u do v jest droga o największej wadze, a jej wagę nazywamy **wagą maksymalną** z u do v i oznaczamy symbolem $M(u, v)$. Wagi maksymalne i drogi maksymalne mogą być badane właściwie w taki sam sposób, jak wagi minimalne i drogi minimalne. W paragrafie 8.3 pokażemy, jak można zmodyfikować algorytm wyznaczający W^*, aby można było otrzymać algorytm wyznaczający wartości M. Na razie jednak będziemy wyznaczać wartości M przyglądając się grafowi skierowanemu. Drogę maksymalną z wierzchołka s do wierzchołka f nazywamy **drogą krytyczną**, a każdą krawędź należącą do takiej drogi nazywamy **krawędzią krytyczną**.

Wprowadzimy trzy funkcje określone na zbiorze $V(G)$, które pomogą nam zrozumieć sieci zdarzeń. Jeśli wyjdziemy z wierzchołka s w chwili 0, to najkrótszy czas, w jakim możemy przybyć do wierzchołka v po wykonaniu wszystkich zadań poprzedzających v, wynosi $M(s, v)$. Ten najwcześniejszy moment przybycia do wierzchołka v będziemy oznaczać przez $A(v)$. W szczególności $A(f) = M(s, f)$ jest czasem, w jakim można zakończyć cały proces. Niech $L(v) = M(s, f) - M(v, f) = A(f) - M(v, f)$. Ponieważ $M(v, f)$ oznacza najkrótszy czas potrzebny do wykonania wszystkich czynności od v do f, więc $L(v)$ jest najpóźniejszym momentem, w którym możemy opuścić wierzchołek v i nadal wykonać wszystkie pozostałe kroki do chwili $A(f)$. Aby obliczyć $L(v)$, możemy posuwać się w tył od wierzchołka f. Pojęcia te są zilustrowane w przykładzie 5(a).

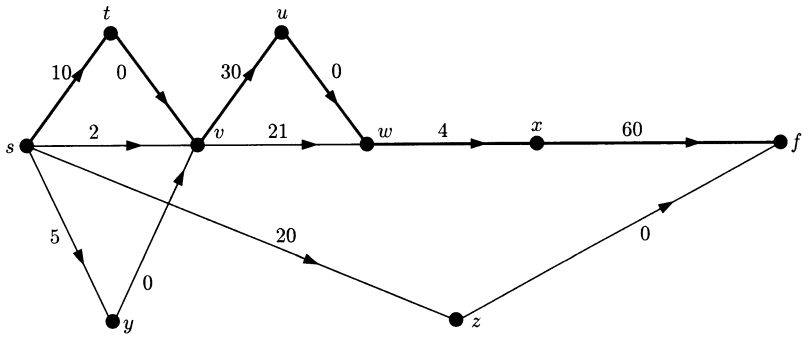
Rezerwę czasową wierzchołka (zdarzenia) v definiujemy wzorem $S(v) = L(v) - A(v)$. Jest to maksymalny czas, w jakim wszystkie zadania zaczynające się w wierzchołku v mogą nie być wykonywane, nie opóźniając całego procesu. Oczywiście $S(v) \geq 0$ dla wszystkich wierzchołków v. Możemy tego dowieść formalnie, zauważając, że

$$S(v) = L(v) - A(v) = M(s, f) - M(v, f) - M(s, v) \geq 0,$$

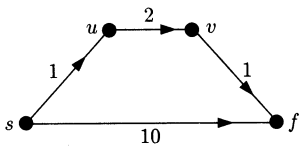
ponieważ $M(s, v) + M(v, f) \leq M(s, f)$. Ta ostatnia nierówność wynika z tego, że możemy połączyć drogi maksymalne z s do v i z v do f, otrzymując w ten sposób drogę z s do f o długości $M(s, v) + M(v, f)$.

PRZYKŁAD 5 (a) Sieć zdarzeń pokazana na rysunku 8.10(b) ma tylko jedną drogę krytyczną: $stvuwxf$. Zatem czynności (a), (d), (f) i (g) są krytyczne. Ta sieć została przerysowana na rysunku 8.11(a), na którym droga krytyczna jest zaznaczona pogrubioną

linią. Zauważmy, że gdyby wagi krawędzi niekrytycznych zostały zmniejszone dzięki zwiększeniu efektywności, to cały proces nadal trwałby 104 minuty. Nawet całkowite wyeliminowanie zadań niekrytycznych nie przyspieszyłoby procesu.



(a)



(b)

Rysunek 8.11

Wartości funkcji A, L i S są podane w poniższej tabeli. Zauważmy, że wartości funkcji A są napisane obok wierzchołków na rysunku 8.10(b).

	s	t	u	v	w	x	y	z	f
A	0	10	40	10	40	44	5	20	104
L	0	10	40	10	40	44	10	104	104
S	0	0	0	0	0	0	5	84	0

Oto kilka przykładowych obliczeń. Jeśli wiemy, że $A(u) = 40$, $W(u, v) = 0$, $A(v) = 10$ oraz $W(v, w) = 21$, to

$$A(w) = \max\{40 + 0, 10 + 21\} = 40.$$

Jeśli wiemy, że $L(w) = 40$, $W(v, w) = 21$, $L(u) = 40$ oraz $W(v, w) = 30$, to

$$L(v) = \min\{40 - 21, 40 - 30\} = 10.$$

Te obliczenia pokazują, co mieliśmy na myśli, mówiąc wcześniej o prowadzeniu obliczeń w tył od wierzchołka f, aby obliczyć $L(v)$.

Wartości funkcji $A(v)$ w pierwszym wierszu są obliczane od lewej do prawej i z chwilą, gdy znana jest wartość $A(f)$, wartości funkcji $L(v)$ w drugim wierszu są obliczane od prawej do lewej. W ćwiczeniu 21 prosimy o podanie dowodu, że te metody obliczania wartości $A(v)$ i $L(v)$ są poprawne.

Zauważmy, że rezerwa czasowa wynosi 0 w każdym wierzchołku na drodze krytycznej. Tak jest zawsze (por. ćwiczenie 18). Ponieważ $S(y) = 5$, więc zadanie obierania ziemniaków może być opóźnione o 5 minut bez opóźnienia obiadu. Ponieważ $S(z) = 84$, więc można opóźnić gotowanie ryżu o 84 minuty; tak naprawdę zwykle odczekalibyśmy ok. 74 minut, zanim zaczęlibyśmy gotować ryż i pozwolilibyśmy mu „odczekać" 10 minut na końcu.

Rzut oka na krawędź (v, w) pokazuje, że moglibyśmy opóźnić to zadanie 9 $(= 30 - 21)$ minut, ale nie możemy wywnioskować tego z rezerw czasowych. Z tego, że $S(v) = 0$ możemy wywnioskować tylko, że nie możemy opóźnić wszystkich zadań zaczynających się w wierzchołku v. Z tego powodu zdefiniujemy bardziej przydatną funkcję, określoną na zbiorze $E(G)$ krawędzi, nazywaną **rezerwą czasową krawędzi** (czynności) lub **luzem pełnym**.

(b) Rezerwa czasowa $S(v)$ wierzchołka v jest to ilość czasu, o jaką można opóźnić wszystkie zadania zaczynające się w wierzchołku v, przy założeniu, że wszystkie pozostałe zadania są wykonywane w sposób efektywny. Zatem jeśli ta rezerwa zostanie wykorzystana w więcej niż jednym wierzchołku, to całkowity czas potrzebny do zakończenia procesu może się wydłużyć. Aby się o tym przekonać, weźmy sieć przedstawioną na rysunku 8.11(b). Mamy tu $S(u) = L(u) - A(u) = 7 - 1 = 6$ oraz $S(v) = L(v) - A(v) = 9 - 3 = 6$. Opóźnienie o 6 minut może wystąpić w wierzchołku u przed wykonaniem zadania (u, v) lub w wierzchołku v po wykonaniu zadania (u, v), ale nie w obu wierzchołkach u i v. ∎

Jeśli (u, v) jest krawędzią w sieci zdarzeń, to

$$M(s, u) + W(u, v) + M(v, f) \leq M(s, f),$$

gdyż droga $s \ldots uv \ldots f$ oczywiście nie ma większej wagi niż droga krytyczna z wierzchołka s do wierzchołka f. Zatem

$$W(u, v) \leq [M(s, f) - M(v, f)] - M(s, u) = L(v) - A(u).$$

Ponieważ $L(v) - A(u)$ jest maksymalnym czasem, w jakim może być wykonane zadanie przypisane do krawędzi (u, v) bez zwiększenia całkowitego czasu wykonania procesu, więc definiujemy

rezerwę czasową $F(u,v)$ krawędzi (u,v) wzorem

$$F(u,v) = L(v) - A(u) - W(u,v).$$

Jest to maksymalne możliwe opóźnienie zadania odpowiadającego krawędzi (u,v). Zależność między rezerwami czasowymi wierzchołków i krawędzi będzie podana w następnym twierdzeniu.

PRZYKŁAD 6 Wszystkie krawędzie krytyczne na rysunku 8.11(a) mają rezerwę czasową równą 0. Można to sprawdzić bezpośrednio, ale następne twierdzenie pokazuje, że tak jest zawsze. Również

$$\begin{aligned} F(s,v) &= L(v) - A(s) - W(s,v) = 10 - 0 - 2 = 8, \\ F(v,w) &= L(w) - A(v) - W(v,w) = 40 - 10 - 21 = 9, \\ F(s,y) &= L(y) - A(s) - W(s,y) = 10 - 0 - 5 = 5, \\ F(y,v) &= L(v) - A(y) - W(y,v) = 10 - 5 - 0 = 5 \end{aligned}$$

itd. ∎

Twierdzenie

Niech dana będzie sieć zdarzeń.

(a) Rezerwa czasowa krawędzi $F(u,v)$ jest równa 0 wtedy i tylko wtedy, gdy (u,v) jest krawędzią krytyczną.

(b) $F(u,v) \geq \max\{S(u), S(v)\}$ dla wszystkich krawędzi sieci.

Dowód. (a) Przypuśćmy, że $F(u,v) = 0$; wtedy

$$M(s,f) = M(s,u) + W(u,v) + M(v,f).$$

Jeśli dołączymy krawędź (u,v) do dróg maksymalnych z s do u i z v do f, to otrzymamy drogę krytyczną dla tej sieci. Zatem (u,v) jest krawędzią krytyczną.

Jeśli (u,v) jest krawędzią drogi krytycznej $s \ldots uv \ldots f$, to droga $s \ldots u$ musi mieć wagę $M(s,u)$, ponieważ w przeciwnym przypadku drogę $s \ldots u$ można byłoby zastąpić drogą o większej wadze. Podobnie droga $v \ldots f$ ma wagę $M(v,f)$, a więc

$$M(s,u) + W(u,v) + M(v,f) = M(s,f).$$

Tak więc mamy $F(u,v) = 0$.

(b) Nierówność $S(u) \leq F(u,v)$ jest równoważna każdej z na-

stępujących nierówności:

$$L(u) - A(u) \leq L(v) - A(u) - W(u,v),$$
$$W(u,v) \leq L(v) - L(u),$$
$$W(u,v) \leq [M(s,f) - M(v,f)] - [M(s,f) - M(u,f)],$$
$$W(u,v) \leq M(u,f) - M(v,f),$$
$$W(u,v) + M(v,f) \leq M(u,f).$$

Ostatnia nierówność jest oczywista, ponieważ krawędź (u,v) można dołączyć do drogi maksymalnej z v do f i otrzymać drogę z u do f o wadze $W(u,v) + M(v,f)$.

Nierówności $S(v) \leq F(u,v)$ można dowieść nieco łatwiej (ćwiczenie 18). ∎

Jak widzieliśmy, skrócenie czasu wymaganego przez krawędź niekrytyczną nie powoduje zmniejszenia całkowitego czasu $M(s,f)$ potrzebnego do zakończenia procesu. Rozpoznanie krawędzi krytycznych zwraca uwagę na te kroki w procesie, w których ulepszenia mogą sprawić różnicę i w których opóźnienia na pewno będą kosztować. Od jej wprowadzenia w latach pięćdziesiątych metoda analizy dróg krytycznych, czasami nazywana PERT (od ang. Program Evaluation and Review Technique), stała się popularną metodą rozwiązywania problemów szeregowania w zarządzaniu w przemyśle.

ĆWICZENIA DO § 8.2

Używaj znaku ∞ we wszystkich tablicach W i W^*.

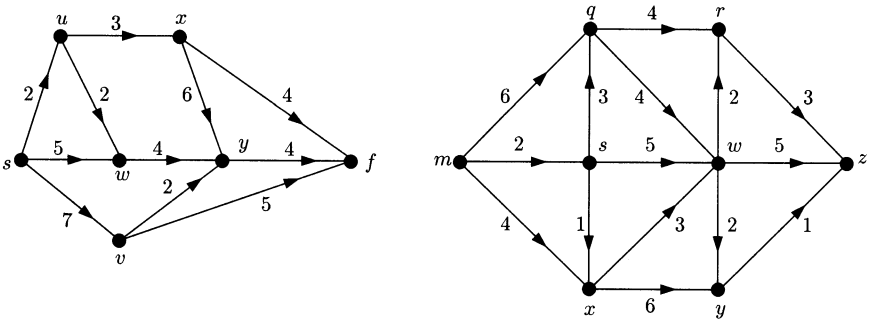
1. Wypisz tablice wartości W i W^* dla grafu skierowanego z rysunku 8.7 z usuniętą pętlą w wierzchołku B.

2. Podaj tablicę wartości W^* dla grafu skierowanego z rysunku 8.12(a).

3. Podaj tablice wartości W i W^* dla grafu skierowanego z rysunku 8.12(b).

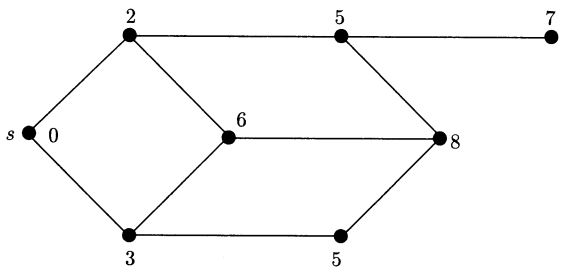
4. Droga $swvyxzf$ jest drogą minimalną z s do f w grafie skierowanym przedstawionym na rysunku 8.8. Znajdź inną drogę minimalną z s do f w tym grafie skierowanym.

5. Na rysunku 8.13 pokazany jest graf skierowany z wagami. Strzałki i wagi zostały usunięte z krawędzi, a liczbą wypisaną w każdym wierzchołku v jest $W^*(s,v)$.



(a) (b)

Rysunek 8.12



Rysunek 8.13

(a) Podaj trzy różne funkcje wagi W, które dają te wartości $W^*(s, v)$. (Rozwiązaniem mogą być trzy rysunki z odpowiednimi liczbami na krawędziach).

(b) Czy różne przypisania wag dają różne drogi minimalne między wierzchołkami? Wyjaśnij to.

6. Przypuśćmy, że u, v i w są wierzchołkami grafu skierowanego z wagami z funkcją minimalnej wagi W^* oraz że $W^*(u, v) + W^*(v, w) = W^*(u, w)$. Wyjaśnij, dlaczego istnieje droga minimalna z u do w przechodząca przez v.

7. (a) Znajdź drogę krytyczną dla grafu skierowanego z rysunku 8.8 z usuniętą krawędzią z y do w.

(b) Dlaczego metoda drogi krytycznej stosuje się tylko do acyklicznych grafów skierowanych?

8. Oblicz rezerwę czasową krawędzi (s, z) i (z, f) z rysunku 8.11(a). *Wskazówka*: zobacz przykład 5(a), gdzie znajdziesz kilka przydatnych liczb.

9. (a) Podaj tablicę wartości A i L dla sieci z rysunku 8.12(a).

(b) Znajdź rezerwy czasowe wierzchołków tej sieci.

(c) Znajdź drogi krytyczne dla tej sieci.

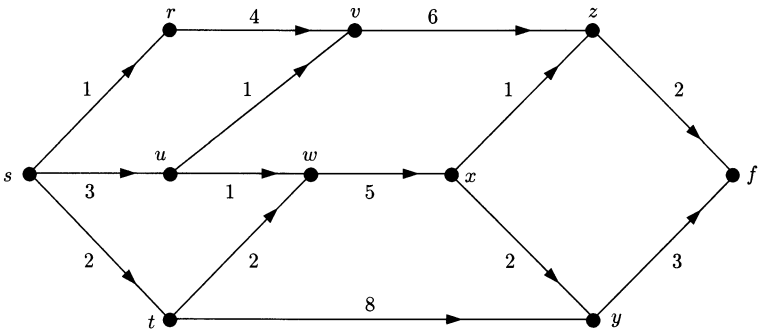
(d) Znajdź rezerwy czasowe krawędzi tej sieci.

10. Powtórz ćwiczenie 9 dla grafu skierowanego przedstawionego na rysunku 8.12(a), gdzie każda krawędź ma wagę 1.

11. Powtórz ćwiczenie 9 dla grafu skierowanego z rysunku 8.12(b).

12. (a) Oblicz rezerwy czasowe krawędzi sieci z rysunku 8.11(b).
(b) Czy każde zadanie w sieci może być opóźnione o jego rezerwę czasową bez opóźniania całego procesu? Odpowiedź uzasadnij.

13. Weźmy sieć przedstawioną na rysunku 8.14.



Rysunek 8.14

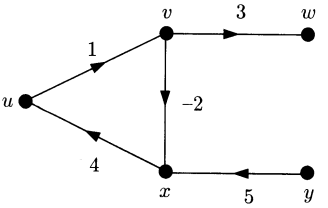
(a) Ile dróg krytycznych ma ten graf skierowany?
(b) Jaka jest największa rezerwa czasowa krawędzi w tym grafie skierowanym?
(c) Które krawędzie mają największą rezerwę czasową?

14. Znajdź rezerwy czasowe wierzchołków sieci z rysunku 8.14.

15. W przykładzie 4 użyliśmy krawędzi o wadze 0, aby zaznaczyć, że pewne kroki nie mogą się rozpocząć, zanim inne nie będą zakończone.
(a) Wyjaśnij, jak uniknąć takich krawędzi o wadze 0, jeśli dopuszczamy krawędzie wielokrotne.
(b) Narysuj graf skierowany dla procesu z przykładu 4, aby zilustrować swoją odpowiedź.

16. Jeśli kucharz w przykładzie 4 nie ma pomocników, to kroki (a), (b) i (c) muszą być wykonywane po kolei, a w przeciwnym przypadku sytuacja przedstawia się tak, jak w tym przykładzie.
(a) Narysuj sieć zdarzeń dla procesu, w którym kucharz nie ma pomocników.
(b) Znajdź drogę krytyczną dla tego procesu.
(c) Które kroki w tym procesie są niekrytyczne?

17. (a) Podaj tablice wartości W i W^* dla grafu skierowanego przedstawionego na rysunku 8.15.
(b) Wyjaśnij, jak na podstawie tablicy wartości W^* można stwierdzić, czy ten graf jest acykliczny czy nie.



Rysunek 8.15

(c) Jak rozwiązałbyś ćwiczenie (a), gdyby krawędź o wadze -2 miała wagę -6?

(d) Wyjaśnij, jak na podstawie tablicy wartości W można stwierdzić, które wierzchołki są źródłami i ujściami tego grafu skierowanego?

18. (a) Uzupełnij dowód twierdzenia, pokazując, że $S(v) \leq F(u,v)$.

(b) Pokaż, że jeśli (u,v) jest krawędzią krytyczną, to $S(u) = S(v) = 0$.

(c) Czy jeśli (u,v) jest krawędzią, dla której $S(u) = S(v) = 0$, to (u,v) musi być krawędzią krytyczną? Odpowiedź uzasadnij.

19. Rezerwę czasową krawędzi $F(u,v)$ można traktować jako tę ilość czasu, o którą możemy opóźnić przejście wzdłuż krawędzi (u,v) bez opóźniania wykonania całego procesu. Zdefiniuj **luz swobodny** $FF(u,v)$ jako tę ilość czasu, o którą możemy opóźnić przejście wzdłuż krawędzi (u,v) bez zwiększenia $A(v)$.

(a) Spróbuj wyrazić $FF(u,v)$ za pomocą funkcji A i W.

(b) Znajdź $FF(u,v)$ dla wszystkich krawędzi grafu skierowanego przedstawionego na rysunku 8.12(a).

(c) Jaka jest różnica między $F(u,v)$ i $FF(u,v)$?

20. (a) Pokaż, że zwiększenie wagi krawędzi krytycznej w sieci zdarzeń powoduje zwiększenie wagi maksymalnej od źródła do ujścia.

(b) Czy istnieje jakakolwiek sytuacja, w której zmniejszenie ilości czasu w kroku krytycznym procesu nie zmniejszy całkowitego czasu wymaganego dla wykonania tego procesu? Odpowiedź uzasadnij.

21. (a) Pokaż, że $A(u) = \max\{A(w) + W(w,u)\colon\ (w,u) \in E(G)\}$.

(b) Pokaż, że $L(u) = \min\{L(v) - W(u,v)\colon\ (u,v) \in E(G)\}$.

§ 8.3. Algorytmy na grafach skierowanych

Nasze badania grafów skierowanych i nieskierowanych doprowadziły nas do pewnej liczby konkretnych pytań. Jaka jest długość najkrótszej drogi od jednego wierzchołka do drugiego w danym grafie skierowanym? Jeśli ten graf skierowany ma wagi, to

jaka jest waga minimalna lub waga maksymalna takiej drogi? Czy w ogóle istnieje jakakolwiek droga?

W tym paragrafie opiszemy pewne algorytmy, za pomocą których można odpowiedzieć na tego typu pytania; są to algorytmy, które można zaprogramować na komputer, ale których można też używać do obliczeń ręcznych. Algorytmy, które wybraliśmy, są względnie szybkie i dość proste do wykonania. Więcej informacji na temat takich algorytmów można znaleźć w książkach specjalistycznych, takich jak *Data Structures and Algorithms*, Aho, Hopcrofta i Ullmana.

Problem wagi minimalnej jest właściwie problemem dotyczącym grafów skierowanych bez pętli i krawędzi wielokrotnych, tak więc ograniczymy rozważania do tego przypadku. Zatem $E(G) \subseteq V(G) \times V(G)$ i możemy opisać graf za pomocą tablicy podającej wartości funkcji wagi $W(u, v)$, tak jak to robiliśmy w § 8.2. Naszym zadaniem będzie znalezienie wag minimalnych $W^*(u, v)$.

Wszystkie algorytmy znajdujące wagi minimalne, które będziemy rozpatrywać, najpierw przeglądają drogi długości 1, tzn. pojedyncze krawędzie, a potem systematycznie przeglądają coraz to dłuższe drogi między wierzchołkami. W czasie działania te algorytmy znajdują drogi o coraz mniejszych wagach między wierzchołkami, w chwili zakończenia te wagi są najmniejszymi możliwymi wagami.

Nasz pierwszy algorytm po prostu znajduje wagi minimalne dróg z wybranego wierzchołka do innych wierzchołków grafu skierowanego G. Przy opisie działania tego algorytmu wygodnie będzie założyć, że $V(G) = \{1, \ldots, n\}$ oraz że 1 jest tym wybranym wierzchołkiem. Poczynając od wierzchołka 1, algorytm przegląda po kolei pozostałe wierzchołki, wybiera nowy wierzchołek w, dla którego najlepsza znana droga od wierzchołka 1 ma możliwie małą wagę oraz uaktualnia najlepsze znane wagi dróg od wierzchołka 1 do innych wierzchołków, biorąc pod uwagę drogi przechodzące przez wierzchołek w. Wierzchołki przejrzane umieszcza w zbiorze L i nie przegląda ich ponownie. W każdej chwili liczba $D(j)$ jest najmniejszą wagą drogi od wierzchołka 1 do wierzchołka j, której wierzchołki znajdują się w zbiorze L. A oto sam algorytm.

Algorytm Dijkstry

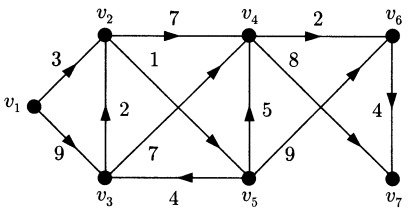
{Dane: graf skierowany bez pętli i krawędzi wielokrotnych, którego zbiorem wierzchołków jest $\{1, \ldots, n\}$, funkcja W wag krawędzi o wartościach nieujemnych}

{Wyniki: wagi minimalne $W^*(1,j)$ dla $j = 2,\ldots,n$}
{Zmienne pomocnicze: zbiory L, V oraz funkcja D}
Niech $L := \emptyset$ oraz $V := \{2,\ldots,n\}$.
Dla $i \in V$ wykonuj
  $D(i) := W(1,i)$.
Dopóki $V \backslash L \neq \emptyset$, wykonuj
  wybierz $k \in V \backslash L$ o najmniejszej wartości $D(k)$
  dołącz k do zbioru L
  dla każdego $j \in V \backslash L$ wykonuj
    jeśli $D(j) > D(k) + W(k,j)$, to
      zastąp $D(j)$ sumą $D(k) + W(k,j)$.
Dla $j \in V$ wykonuj
  $W^*(1,j) := D(j)$. ∎

Musimy sprawdzić, że ten algorytm kończy działanie oraz że ostateczną wartością $D(j)$ dla każdego j jest $W^*(1,j)$, tzn. jest waga minimalna drogi z wierzchołka 1 do wierzchołka j. Chcemy również oszacować czas działania tego algorytmu. Najpierw jednak popatrzmy, jak ten algorytm działa.

PRZYKŁAD 1 Weźmy graf skierowany G z wagami, pokazany na rysunku 8.16(a). Tablica wag jego krawędzi jest przedstawiona na rysunku 8.16(b).

Tablica 8.1 pokazuje, w jaki sposób zmieniają się wartości L oraz $D(2),\ldots,D(7)$ w czasie działania algorytmu, rozpoczynającego swoje działanie w wierzchołku 1. Zauważmy, że wartości w kolumnach $D(j)$ zmniejszają się z upływem czasu, a od momentu, gdy wierzchołek j znajdzie się w zbiorze L, już się nie zmieniają. ∎



(a)

W	v_1	v_2	v_3	v_4	v_5	v_6	v_7
v_1	∞	3	9	∞	∞	∞	∞
v_2	∞	∞	∞	7	1	∞	∞
v_3	∞	2	∞	7	∞	∞	∞
v_4	∞	∞	∞	∞	∞	2	8
v_5	∞	∞	4	5	∞	9	∞
v_6	∞	∞	∞	∞	∞	∞	4
v_7	∞	∞	∞	∞	∞	∞	∞

(b)

Rysunek 8.16

Tablica 8.1

L	$D(2)$	$D(3)$	$D(4)$	$D(5)$	$D(6)$	$D(7)$	Komentarz
$\emptyset$	3	9	∞	∞	∞	∞	dane początkowe
$\{2\}$	3	9	10	4	∞	∞	znaleziono drogi $v_1v_2v_4$ oraz $v_1v_2v_5$
$\{2,5\}$	3	8	9	4	13	∞	znaleziono drogi $v_1v_2v_5v_3$, $v_1v_2v_5v_4$ oraz $v_1v_2v_5v_6$
$\{2,5,3\}$	3	8	9	4	13	∞	brak poprawy
$\{2,5,3,4\}$	3	8	9	4	11	17	znaleziono drogi $v_1v_2v_5v_4v_6$ oraz $v_1v_2v_5v_4v_7$
$\{2,5,3,4,6\}$	3	8	9	4	11	15	znaleziono drogi $v_1v_2v_5v_4v_6v_7$

Twierdzenie 1 Jeśli wagi krawędzi grafu skierowanego G są nieujemne, to algorytm Dijkstry kończy działanie i po zakończeniu $D(j) = W^*(1,j)$ dla $j = 2,\ldots,n$.

Dowód. Ponieważ w każdym przebiegu pętli „dopóki" algorytm dodaje jeden wierzchołek więcej do zbioru L, więc wykona on $n-1$ przebiegów tej pętli i zatrzyma się. Nie jest wcale oczywiste, że na końcu $D(j) = W^*(1,j)$, gdyż za każdym razem algorytm dokonuje zachłannego wyboru nowych wierzchołków. Zachłanność nie zawsze się opłaca — spróbuj wybrać 40 centów ze stosu dziesięciocentówek i dwudziestopięciocentówek wybierając za pierwszym razem dwudziestopięciocentówkę — ale tym razem się opłaca.

Twierdzimy, że następujące dwa zdania są niezmiennikami pętli:

(a) $D(l) \leq D(j)$, jeśli tylko $l \in L$ i $j \in V\backslash L$;
(b) dla każdego $j \in V$ wartość $D(j)$ jest najmniejszą wagą drogi prowadzącej z wierzchołka 1 do wierzchołka j przez wierzchołki należące do zbioru L.

Drogę, której pośrednie wierzchołki należą do zbioru L, będziemy w skrócie nazywać ***L*-drogą**, a taką drogę o minimalnej wadze będziemy nazywać ***L*-drogą minimalną**.

Załóżmy, że zbiór L i funkcja D spełniają warunki (a) i (b) na początku przebiegu pętli, w którym wybrany jest wierzchołek k. Chcemy pokazać, że również zbiór $L \cup \{k\}$ i uaktualniona funkcja D spełniają te warunki na końcu tego przebiegu pętli.

Przypuśćmy najpierw, że $l \in L \cup \{k\}$ oraz $j \in V\backslash(L \cup \{k\})$. Wtedy $D(l) \leq D(k)$ z założenia (a) dla L i D. Ponieważ $j \in V\backslash L$,

więc $D(j) \geq D(k)$ na początku tego przebiegu, ze względu na wybór wierzchołka k; jeśli zaś wartość $D(j)$ zmieni się, to zostanie zastąpiona przez $D(k) + W(k,j)$, a więc będzie nadal równa co najmniej $D(k)$, gdyż $W(k,j) \geq 0$. Zatem $D(l) \leq D(k) \leq D(j)$, a więc zbiór $L \cup \{k\}$ i nowa funkcja D nadal spełniają warunek (a) na końcu tego przebiegu pętli.

Aby wykazać, że warunek (b) jest niezmiennikiem pętli, weźmy $j \in V$. Na początku przebiegu pętli wartość $D(j)$ jest najmniejszą wagą L-drogi prowadzącej z wierzchołka 1 do wierzchołka j. Tę początkową wartość będziemy oznaczać symbolem $D_L(j)$. Chcemy wykazać, że na końcu przebiegu pętli wartość $D(j)$ będzie równa $D_{L\cup\{k\}}(j)$, czyli wadze $(L \cup \{k\})$-drogi minimalnej z wierzchołka 1 do wierzchołka j.

Przyjmijmy najpierw, że $j = k$. W drogach minimalnych wierzchołki się nie powtarzają, więc każda $(L \cup \{k\})$-droga minimalna z wierzchołka 1 do wierzchołka k jest L-drogą minimalną z 1 do k. Zatem $D_{L\cup\{k\}}(k) = D_L(k)$. Ponieważ $W(k,k) \geq 0$, więc $D(k) + W(k,k) \geq D(k)$. A więc w tym przebiegu pętli wartość $D(k)$ nie zmieni się, czyli rzeczywiście $D(k) = D_{L\cup\{k\}}(k)$ po zakończeniu tego przebiegu.

Przypuśćmy, że $j \in V$ oraz $j \neq k$. Jeśli $D_{L\cup\{k\}}(j) = D_L(j)$, to waga każdej $(L \cup \{k\})$-drogi z wierzchołka 1 do wierzchołka j, przechodzącej przez wierzchołek k, jest równa co najmniej $D_L(j) = D(j)$. W szczególności $D(k) + W(k,j) \geq D(j)$ w tym przypadku, a więc wartość (właściwa) $D(j)$ nie zmieni się podczas tego przebiegu pętli.

Wreszcie załóżmy, że $D_{L\cup\{k\}}(j) < D_L(j)$. Niech P będzie $(L \cup \{k\})$-drogą minimalną z wierzchołka 1 do wierzchołka j. Wtedy droga P musi przechodzić przez wierzchołek k. Niech l będzie ostatnim wierzchołkiem na drodze P znajdującym się przed wierzchołkiem j. Pokażemy, że $l = k$. Gdyby tak nie było, to $l \in L$ i z warunku (a) dla zbioru L i funkcji D otrzymalibyśmy $D_L(l) \leq D_L(k)$. Ale część drogi P od wierzchołka 1 do wierzchołka k musi być L-drogą minimalną z 1 do k, mającą wagę $D_L(k)$, gdyż w przeciwnym przypadku moglibyśmy zastąpić ją L-drogą minimalną i otrzymać w ten sposób drogę całkowicie zawartą w zbiorze $L \cup \{k\}$, prowadzącą z wierzchołka 1 do wierzchołka j i mającą wagę mniejszą niż waga drogi P. A więc

$$D_{L\cup\{k\}}(j) = \text{waga}(P) \geq D_L(k) + W(l,j) \geq D_L(l) + W(l,j).$$

Zatem L-droga minimalna z wierzchołka 1 do wierzchołka l, uzupełniona krawędzią z l do j, jest L-drogą z wierzchołka 1 do wierzchołka j, o wadze $D_L(l) + W(l,j)$. Stąd $D_L(l) + W(l,j) \geq D_L(j)$,

a więc $D_{L\cup\{k\}}(j) \geq D_L(j)$, co przeczy przyjętemu założeniu. Zatem $l = k$, nasza droga P ma postać $1 \ldots kj$ oraz

$$D_{L\cup\{k\}}(j) = \text{waga}(P) = D(k) + W(k, j).$$

Ponieważ ta ostatnia wartość jest mniejsza niż $D(j) = D_L(j)$, więc wartość $D(j)$ jest uaktualniona poprawnie jako $D_{L\cup\{k\}}(j)$ w tym przebiegu pętli.

Oba warunki (a) i (b) są spełnione przed pierwszym przebiegiem pętli. Kiedy algorytm kończy działanie, zachodzi równość $L = V$ oraz spełniony jest warunek (b), a więc ostateczną wartością $D(j)$ jest $W^*(1, j)$. ∎

Jak długo działa algorytm Dijkstry dla danego grafu skierowanego mającego n wierzchołków? Najwięcej czasu pochłania pętla „dopóki", w której za każdym przebiegiem zostaje usunięty jeden z początkowych n wierzchołków. Czas potrzebny do znalezienia najmniejszej wartości $D(k)$ wynosi $O(n)$, jeśli będziemy badać po kolei wierzchołki zbioru $V(G)$, choć w rzeczywistości istnieją algorytmy sortujące, które pozwalają znaleźć k szybciej. Dla danego wybranego wierzchołka k trzeba dokonać co najwyżej n porównań i podstawień, a więc całkowity czas jednego przebiegu pętli wynosi $O(n)$. A zatem, ponieważ algorytm dokonuje n przebiegów, całkowity czas działania wynosi $O(n^2)$.

Jeśli graf skierowany jest reprezentowany za pomocą list bezpośrednich następników (lub inaczej list sąsiedztwa), to można tak przepisać ten algorytm, by w kroku polegającym na uaktualnianiu algorytm przeglądał tylko bezpośrednie następniki wierzchołka k. Podczas działania algorytmu każda krawędź jest rozpatrywana tylko jeden raz w tym kroku uaktualniania. Taka modyfikacja algorytmu powoduje jego przyspieszenie, jeśli liczba krawędzi $|E(G)|$ jest znacznie mniejsza od n^2.

Algorytm Dijkstry znajduje wagi minimalnych dróg prowadzących z danego wierzchołka. Aby znaleźć wartość $W^*(v_i, v_j)$ dla wszystkich par wierzchołków v_i oraz v_j, możemy po prostu wykonać ten algorytm n razy, za każdym razem zaczynając od innego wierzchołka. Istnieje jednak inny algorytm, pochodzący od Warshalla i ulepszony przez Floyda, który daje na końcu wszystkie wartości $W^*(v_i, v_j)$ i który jest łatwy do zaprogramowania. Podobnie jak algorytm Dijkstry, tworzy on powiększającą się listę zbadanych wierzchołków i patrzy na drogi przechodzące przez wierzchołki z tej listy.

Przypuśćmy, że $V(G) = \{v_1, \ldots, v_n\}$. Algorytm Warshalla używa macierzy $\mathbf{W}$ wymiaru $n \times n$, którą na początku jest macierz

wag krawędzi $\mathbf{W}_0$: $\mathbf{W}_0[i,j] = W(v_i, v_j)$ dla wszystkich i oraz j, a na końcu macierz $\mathbf{W}_n = \mathbf{W}^*$ taka, że $\mathbf{W}^*[i,j] = W^*(v_i, v_j)$.

Algorytm Warshalla

{Dane: macierz $\mathbf{W}_0$ nieujemnych wag krawędzi grafu skierowanego bez pętli i krawędzi wielokrotnych}
{Wyniki: macierz $\mathbf{W}^*$ wag minimalnych tego grafu}
{Zmienne pomocnicze: macierz $\mathbf{W}$}
$\mathbf{W} := \mathbf{W}_0$
dla $k = 1$ do n wykonuj
  dla $i = 1$ do n wykonuj
    dla $j = 1$ do n wykonuj
      jeśli $\mathbf{W}[i,j] > \mathbf{W}[i,k] + \mathbf{W}[k,j]$, to
        zastąp $\mathbf{W}[i,j]$ sumą $\mathbf{W}[i,k] + \mathbf{W}[k,j]$.
$\mathbf{W}^* := \mathbf{W}$. ∎

Twierdzenie 2

Algorytm Warshalla daje w wyniku macierz $\mathbf{W}^*$ wag minimalnych.

Dowód. Zapisaliśmy ten algorytm używając zagnieżdżonych pętli. Możemy również zapisać go w następującej postaci, przypominającej algorytm Dijkstry.

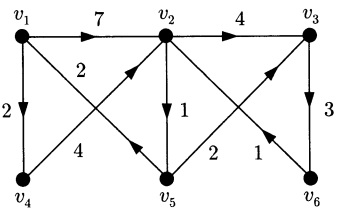
Niech $\mathbf{W} := \mathbf{W}_0$, $L := \emptyset$, $V := \{1, \ldots, n\}$
dopóki $V \setminus L \neq \emptyset$, wykonuj
  wybierz $k \in V \setminus L$
  dołącz k do zbioru L
  dla wszystkich $i, j \in V$ wykonuj
    niech $\mathbf{W}[i,j] := \min\{\mathbf{W}[i,j], \mathbf{W}[i,k] + \mathbf{W}[k,j]\}$
niech $\mathbf{W}^* := \mathbf{W}$.

Przypisanie $\mathbf{W}[i,j] := \min\{\mathbf{W}[i,j], \mathbf{W}[i,k] + \mathbf{W}[k,j]\}$ daje ten sam efekt, co poprzednia instrukcja „jeśli ..., to zastąp ...”; wartość $\mathbf{W}[i,j]$ zmienia się tylko wtedy, gdy $\mathbf{W}[i,j] > \mathbf{W}[i,k] + \mathbf{W}[k,j]$.

Twierdzimy, że dla każdych i oraz j warunek „$\mathbf{W}[i,j]$ jest wagą minimalnej L-drogi z wierzchołka i do wierzchołka j” jest niezmiennikiem pętli „dopóki”. Załóżmy, że ten warunek jest spełniony na początku przebiegu pętli, w którym została wybrana liczba k. Wtedy najmniejsza waga $(L \cup \{k\})$-drogi z i do j przechodzącej przez k wynosi $\mathbf{W}[i,k] + \mathbf{W}[k,j]$. Jeśli ta liczba jest mniejsza od $\mathbf{W}[i,j]$, to staje się ona nową wartością $\mathbf{W}[i,j]$. W przeciwnym przypadku możemy otrzymać równie do-

brą $(L \cup \{k\})$-drogę, ale nie przechodzącą przez k, tzn. L-drogę. Zatem aktualna wartość $\mathbf{W}[i,j]$, która nie zmienia się, jest właściwą wagą minimalnej $(L \cup \{k\})$-drogi z i do j.

To rozumowanie pokazuje, że warunek ten jest niezmiennikiem pętli. Jest on spełniony na początku, gdyż $\mathbf{W}_0[i,j]$ jest minimalną wagą drogi z i do j nie przechodzącej przez żadne pośrednie wierzchołki. Wykonanie pętli kończy się, gdy wszystkie wierzchołki znajdują się w zbiorze L, a więc ostateczna wartość $\mathbf{W}[i,j]$ jest wagą $\mathbf{W}^*[i,j]$ drogi minimalnej z wierzchołka i do wierzchołka j. ∎



Rysunek 8.17

PRZYKŁAD 2 Zastosujemy algorytm Warshalla do grafu skierowanego przedstawionego na rysunku 8.17. Obliczenia ręczne przeprowadzone za pomocą algorytmu Warshalla dają n nowych macierzy, po jednej dla każdej wartości k. W tym przykładzie mamy następujące macierze:

$$\mathbf{W} = \mathbf{W}_0 = \begin{bmatrix} \infty & 7 & \infty & 2 & \infty & \infty \\ \infty & \infty & 4 & \infty & 1 & \infty \\ \infty & \infty & \infty & \infty & \infty & 3 \\ \infty & 4 & \infty & \infty & \infty & \infty \\ 2 & \infty & 2 & \infty & \infty & \infty \\ \infty & 1 & \infty & \infty & \infty & \infty \end{bmatrix},$$

$$\mathbf{W}_1 = \begin{bmatrix} \infty & 7 & \infty & 2 & \infty & \infty \\ \infty & \infty & 4 & \infty & 1 & \infty \\ \infty & \infty & \infty & \infty & \infty & 3 \\ \infty & 4 & \infty & \infty & \infty & \infty \\ 2 & 9 & 2 & 4 & \infty & \infty \\ \infty & 1 & \infty & \infty & \infty & \infty \end{bmatrix},$$

$$\mathbf{W}_2 = \begin{bmatrix} \infty & 7 & 11 & 2 & 8 & \infty \\ \infty & \infty & 4 & \infty & 1 & \infty \\ \infty & \infty & \infty & \infty & \infty & 3 \\ \infty & 4 & 8 & \infty & 5 & \infty \\ 2 & 9 & 2 & 4 & 10 & \infty \\ \infty & 1 & 5 & \infty & 2 & \infty \end{bmatrix},$$

$$\mathbf{W}_3 = \begin{bmatrix} \infty & 7 & 11 & 2 & 8 & 14 \\ \infty & \infty & 4 & \infty & 1 & 7 \\ \infty & \infty & \infty & \infty & \infty & 3 \\ \infty & 4 & 8 & \infty & 5 & 11 \\ 2 & 9 & 2 & 4 & 10 & 5 \\ \infty & 1 & 5 & \infty & 2 & 8 \end{bmatrix},$$

$$\mathbf{W}_4 = \begin{bmatrix} \infty & 6 & 10 & 2 & 7 & 13 \\ \infty & \infty & 4 & \infty & 1 & 7 \\ \infty & \infty & \infty & \infty & \infty & 3 \\ \infty & 4 & 8 & \infty & 5 & 11 \\ 2 & 8 & 2 & 4 & 9 & 5 \\ \infty & 1 & 5 & \infty & 2 & 8 \end{bmatrix},$$

$$\mathbf{W}_5 = \begin{bmatrix} 9 & 6 & 9 & 2 & 7 & 12 \\ 3 & 9 & 3 & 5 & 1 & 6 \\ \infty & \infty & \infty & \infty & \infty & 3 \\ 7 & 4 & 7 & 9 & 5 & 10 \\ 2 & 8 & 2 & 4 & 9 & 5 \\ 4 & 1 & 4 & 6 & 2 & 7 \end{bmatrix},$$

$$\mathbf{W}^* = \mathbf{W}_6 = \begin{bmatrix} 9 & 6 & 9 & 2 & 7 & 12 \\ 3 & 7 & 3 & 5 & 1 & 6 \\ 7 & 4 & 7 & 9 & 5 & 3 \\ 7 & 4 & 7 & 9 & 5 & 10 \\ 2 & 6 & 2 & 4 & 7 & 5 \\ 4 & 1 & 4 & 6 & 2 & 7 \end{bmatrix}.$$

Zilustrujemy obliczenia na przykładzie $\mathbf{W}_4[5,2]$. W tym przypadku $k = 4$. Wartość $\mathbf{W}_3[5,2]$ wynosi 9, co odpowiada długości najkrótszej drogi $v_5v_1v_2$, której pośrednie wierzchołki należą do zbioru $\{v_1, v_2, v_3\}$. Aby znaleźć $\mathbf{W}_4[5,2]$, patrzymy na liczbę $\mathbf{W}_3[5,4]+\mathbf{W}_3[4,2]$. Wynosi ona $4+4 = 8$ i odpowiada parze dróg $v_5v_1v_4$ i v_4v_2. Ponieważ $9 > 8$, więc zastępujemy wartość $\mathbf{W}_3[5,2]$ sumą $\mathbf{W}_3[5,4] + \mathbf{W}_3[4,2] = 8$, co odpowiada drodze $v_5v_1v_4v_2$.

Dany wyraz macierzy, taki jak $\mathbf{W}[5,2]$, może zmieniać się wiele razy w czasie obliczeń, gdy k przebiega wszystkie możliwe wartości. ∎

Łatwo sprawdzić, jak długo działa algorytm Warshalla. Jeden krok polegający na porównaniu i ewentualnej zmianie wartości wewnątrz pętli „dla $j = 1$ do n wykonuj" trwa co najwyżej pewną stałą ilość czasu, powiedzmy t. Ten krok jest wykonywany dokładnie n^3 razy, po jednym razie dla każdej możliwej trójki (k, i, j), tak więc całkowity czas działania algorytmu wynosi n^3t, tzn. $O(n^3)$.

Ten krok polegający na porównaniu i zmianie wartości jest w algorytmie Warshalla taki sam, jak w algorytmie Dijkstry, który jednak zawiera również inne kroki. Ponieważ algorytm Dijkstry działa w czasie $O(n^2)$, więc wykonanie go dla wszystkich n wierzchołków pozwala znaleźć wszystkie wagi dróg minimalnych w czasie $O(n^3)$. Jednak algorytm Dijkstry i algorytm Warshalla różnią się stałymi przy n^3, a więc wybór algorytmu może zależeć od dostępnej implementacji komputerowej. Jeśli liczba $|E(G)|$ jest mała w porównaniu z n^2, to użycie reprezentacji grafu skierowanego za pomocą list bezpośrednich następników sugeruje raczej wybór algorytmu Dijkstry.

W tym momencie należy przyznać, że celowo nie wspominaliśmy o jednej możliwej komplikacji. Nawet jeśli graf G ma niewiele wierzchołków, to wagi $W(i,j)$ mogą być tak duże, że wypisanie ich trwałoby całe lata i oba algorytmy Dikstry i Warshalla działałyby bardzo długo. Jednak w praktyce liczby, z którymi mamy do czynienia, gdy stosujemy te dwa algorytmy, są rozsądnej wielkości. W każdym razie, jeśli mamy dany ten sam zbiór wag, to nasze porównanie czasów działania obu algorytmów jest słuszne.

Nieujemność wag krawędzi została wykorzystana w dowodzie twierdzenia 2 w subtelny sposób. Jeśli $\mathbf{W}[k,k] < 0$ dla pewnego k, to $\mathbf{W}[k,k] + \mathbf{W}[k,j] < \mathbf{W}[k,j]$, ale nie istnieje $(L \cup \{k\})$-droga o najmniejszej wadze z wierzchołka k do wierzchołka j. Możemy obejść jakąś drogę o ujemnej wadze prowadzącą z wierzchołka k z powrotem do k dowolnie wiele razy, zanim wyruszymy w kierunku wierzchołka j. Jednym ze sposobów wyeliminowania takich sytuacji jest przyjęcie założenia, że wszystkie wagi są nieujemne — jak to właśnie zrobiliśmy — tak by każda wartość, którą podstawiamy, była również nieujemna. Można również dopuścić możliwość wag ujemnych w algorytmie Warshalla, żądając jednak, by graf był acykliczny, tak aby nie było możliwe powracanie do wierzchołka k.

Można zaadaptować algorytm Warshalla tak, by dawał w wyniku wagi maksymalne w grafie acyklicznym. Zastąpmy wszystkie wystąpienia ∞ wartościami $-\infty$, przy czym $-\infty + x = -\infty = x + (-\infty)$ dla dowolnego x oraz $-\infty < a$ dla wszystkich liczb rzeczywistych a. Zastąpmy nierówność w kroku, w którym dokonujemy podstawienia, nierównością $\mathbf{W}[i,j] < \mathbf{W}[i,k] + \mathbf{W}[k,j]$. Otrzymany w ten sposób algorytm wyznacza wartości $\mathbf{W}_n[i,j] = M(v_i, v_j)$, gdzie M jest funkcją wag maksymalnych z § 8.2.

Jeśli potrzebne są nam tylko wagi maksymalne dróg wychodzących z jednego źródła, jak na przykład w przypadku sieci zdarzeń, to możemy uprościć nieco ten algorytm, ale kosztem

przenumerowania wierzchołków grafu skierowanego za pomocą algorytmu takiego jak NUMEROWANIE WIERZCHOŁKÓW z § 8.1 lub ETYKIETOWANIE z § 7.3. Wydaje się naturalne, by w tym przypadku użyć etykietowania uporządkowanego odwrotnie, w którym etykiety wierzchołków są im przypisane tak, że $i < j$, jeśli istnieje droga z wierzchołka v_i do wierzchołka v_j. Takie etykietowanie można łatwo otrzymać ze zwykłego etykietowania uporządkowanego; dla $i = 1, \ldots, n$ wystarczy po prostu przenumerować wierzchołek v_i na v_{n+1-i}.

Aby znaleźć wagi maksymalne dróg wychodzących z danego wierzchołka v_s, na przykład z v_1, po prostu ustalamy $i = s$ w odpowiednio zmodyfikowanym algorytmie Warshalla. Etykietowanie uporządkowane odwrotnie daje równość $\mathbf{W}[k, j] = -\infty$ dla $j \leq k$, a więc w pętli „dla j ..." zmienna j nie musi przebiegać wszystkich wartości od 1 do n. Otrzymany w ten sposób algorytm wygląda następująco, dla $i = s = 1$:

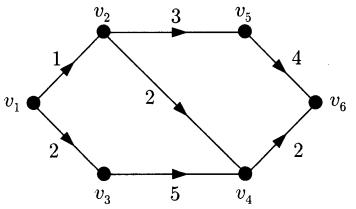
Algorytm WAGI MAKSYMALNE

{Dane: macierz $\mathbf{W}_0$ wag krawędzi acyklicznego grafu skierowanego G, mającego etykietowanie uporządkowane odwrotnie}
{Wyniki: wagi maksymalne $M(1, j)$ dla $j = 2, \ldots, n$}
Dla $k = 2$ do $n - 1$ wykonuj
  dla $j = k + 1$ do n wykonuj
    jeśli $\mathbf{W}[1, j] < \mathbf{W}[1, k] + \mathbf{W}[k, j]$, to
      zastąp $\mathbf{W}[1, j]$ sumą $\mathbf{W}[1, k] + \mathbf{W}[k, j]$.
Dla $j = 2$ do n wykonuj
  $M(1, j) := \mathbf{W}[1, j]$. ∎

Podkreślamy, że ten algorytm jest przeznaczony tylko dla acyklicznych grafów skierowanych, mających etykiety uporządkowane odwrotnie. Dowód tego, że czas działania tego algorytmu wynosi $O(n^2)$ (por. ćwiczenie 12) jest podobny do odpowiedniego rozumowania dla algorytmu Warshalla. Można trochę przyspieszyć ten algorytm, jeśli graf skierowany będzie reprezentowany przez listy bezpośrednich następników i zmienna j będzie przebiegać zbiór bezpośrednich następników NAST(k). Wtedy każdą krawędź badamy tylko raz, by sprawdzić, czy nie zwiększa ona wagi maksymalnej. Czas działania algorytmu wynosi wtedy $O(\max\{|V(G)|, |E(G)|\})$, co jest porównywalne z czasem potrzebnym do początkowego posortowania wierzchołków za pomocą algorytmu ETYKIETOWANIE.

Algorytm Dijkstry nie działa dla wag ujemnych (por. ćwiczenie 10). Ponadto (por. ćwiczenie 11) wydaje się, że nie można w naturalny sposób zmodyfikować go tak, by znajdował wagi maksymalne.

PRZYKŁAD 3 Zastosujemy algorytm WAGI MAKSYMALNE do grafu skierowanego pokazanego na rysunku 8.18(a). Na rysunku



(a)

$$\mathbf{W}_0 = \begin{bmatrix} -\infty & 1 & 2 & -\infty & -\infty & -\infty \\ -\infty & -\infty & -\infty & 2 & 3 & -\infty \\ -\infty & -\infty & -\infty & 5 & -\infty & -\infty \\ -\infty & -\infty & -\infty & -\infty & -\infty & 2 \\ -\infty & -\infty & -\infty & -\infty & -\infty & 4 \\ -\infty & -\infty & -\infty & -\infty & -\infty & -\infty \end{bmatrix}$$

(b)

Rysunek 8.18

8.18(b) przedstawiona jest macierz początkowa $\mathbf{W}_0$. Dla wygody będziemy używać macierzy wierszowej $\mathbf{D}$ takiej, że $\mathbf{D}[j] = \mathbf{W}[1, j]$ dla $j = 1, \ldots, 6$. Otrzymujemy następujący ciąg macierzy:

$$\mathbf{D}_0 = \mathbf{D}_1 = [-\infty, 1, 2, -\infty, -\infty, -\infty],$$
$$\mathbf{D}_2 = [-\infty, 1, 2, 3, 4, -\infty],$$
$$\mathbf{D}_3 = [-\infty, 1, 2, 7, 4, -\infty],$$
$$\mathbf{D}_4 = \mathbf{D}_5 = [-\infty, 1, 2, 7, 4, 9].$$

Dla ilustracji, obliczymy wartość $\mathbf{D}_4$ przyjmując, że macierz $\mathbf{D}_3$ zawiera właściwe wartości $\mathbf{W}[1, j]$ dla $k = 3$. Pętla „dla $j \ldots$" dla $k = 4$ ma postać:

dla $j = 5$ do 6 wykonuj
  jeśli $\mathbf{D}[j] < \mathbf{D}[4] + \mathbf{W}[4, j]$, to
    zastąp $\mathbf{D}[j]$ sumą $\mathbf{D}[4] + \mathbf{W}[4, j]$.

Ponieważ $\mathbf{D}_3[5] = 4 > -\infty = 7 + (-\infty) = \mathbf{D}_3[4] + \mathbf{W}[4, 5]$, więc nie dokonujemy podstawienia i otrzymujemy $\mathbf{D}_4[5] = \mathbf{D}_3[5] = 4$. Ponieważ $\mathbf{D}_3[6] = -\infty < 9 = 7 + 2 = \mathbf{D}_3[4] + \mathbf{W}[4, 6]$, więc dokonujemy podstawienia i otrzymujemy $\mathbf{D}_4[6] = 9$. Wartości $\mathbf{D}_3[1], \ldots, \mathbf{D}_3[4]$ oczywiście nie zmieniają się w $\mathbf{D}_4$. ∎

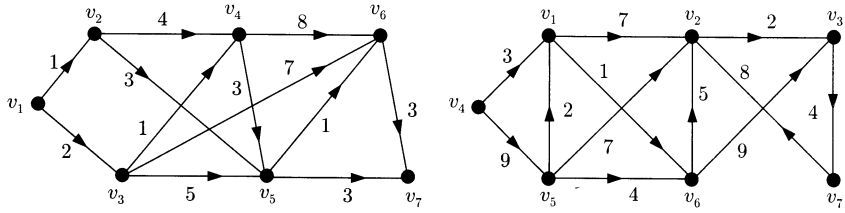
W następnym paragrafie zastanowimy się nad tym, jak zmodyfikować opisane tutaj algorytmy, by móc odpowiedzieć na kilka podstawowych pytań dotyczących grafów skierowanych, w tym na

pytanie, jak znaleźć drogi minimalne czy drogi maksymalne odpowiadające wagom minimalnym lub maksymalnym, które właśnie wyznaczyliśmy.

ĆWICZENIA DO § 8.3

1. (a) Podaj macierz wag minimalnych $\mathbf{W}^*$ dla grafu skierowanego pokazanego na rysunku 8.19(a). Dozwolona jest każda metoda, w tym również wpatrywanie się w rysunek.

(b) Powtórz ćwiczenie (a) dla grafu skierowanego przedstawionego na rysunku 8.19(b).



Rysunek 8.19

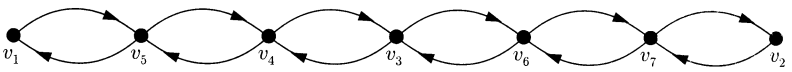
2. Znajdź macierz wag maksymalnych dla grafu skierowanego przedstawionego na rysunku 8.19(a).

3. (a) Zastosuj algorytm Dijkstry do grafu skierowanego przedstawionego na rysunku 8.19(a). Zacznij od wierzchołka v_1 i użyj zapisu takiego jak w tablicy 8.1.

(b) Powtórz ćwiczenie (a) dla grafu skierowanego przedstawionego na rysunku 8.19(b).

4. Zastosuj algorytm Dijkstry do grafu skierowanego przedstawionego na rysunku 8.17, zaczynając od wierzchołka v_1. Porównaj odpowiedź z odpowiedzią otrzymaną przez algorytm Warshalla w przykładzie 2.

5. (a) Użyj algorytmu Warshalla do znalezienia długości dróg minimalnych w grafie skierowanym przedstawionym na rysunku 8.20. Podaj macierz $\mathbf{W}$ na początku każdej pętli „dla k ...".



Rysunek 8.20

(b) Użyj algorytmu Dijkstry do tego grafu skierowanego, aby znaleźć długości dróg minimalnych z wierzchołka v_1. Zapisz swoją odpowiedź w takiej postaci jak jest to zrobione w tablicy 8.1.

6. (a) Użyj algorytmu Warshalla do znalezienia macierzy $\mathbf{W}^*$ dla grafu skierowanego przedstawionego na rysunku 8.18(a).

(b) Znajdź wagi maksymalne dla tego samego grafu skierowanego, używając zmodyfikowanego algorytmu Warshalla.

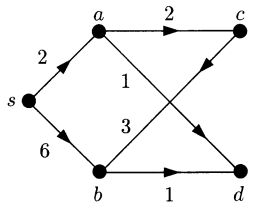
7. Podaj macierz końcową $\mathbf{W}$, jeśli zastosujemy algorytm Warshalla do następującej macierzy:

$$\mathbf{W}_0 = \begin{bmatrix} \infty & 11 & 9 & \infty & 2 \\ \infty & \infty & \infty & \infty & \infty \\ \infty & 1 & \infty & \infty & \infty \\ \infty & \infty & 2 & \infty & \infty \\ \infty & \infty & 6 & 3 & \infty \end{bmatrix}.$$

8. Zastosuj algorytm Dijkstry do znalezienia wag minimalnych z wierzchołka v_3 w grafie skierowanym przedstawionym na rysunku 8.19(a).

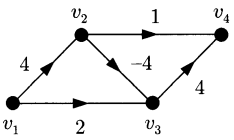
9. (a) Użyj algorytmu WAGI MAKSYMALNE do znalezienia wag maksymalnych z wierzchołka v_1 do innych wierzchołków w grafie skierowanym przedstawionym na rysunku 8.19(a).

(b) Użyj algorytmu WAGI MAKSYMALNE do znalezienia wag maksymalnych z wierzchołka s do innych wierzchołków w grafie skierowanym przedstawionym na rysunku 8.21. (Zacznij od uporządkowania grafu).



Rysunek 8.21

10. (a) Pokaż, że algorytm Dijkstry nie daje w wyniku poprawnych wag minimalnych dla acyklicznego grafu skierowanego pokazanego na rysunku 8.22. (A więc wagi ujemne mogą powodować pewne trudności).



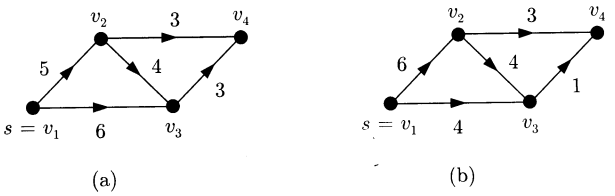
Rysunek 8.22

(b) Czy algorytm Dijkstry dawałby poprawne wagi minimalne dla tego grafu skierowanego, gdyby linia „dla każdego $j \in V \backslash L$ wykonuj" została zastąpiona linią „dla każdego $j \in V$ wykonuj"? Odpowiedź uzasadnij.

11. To ćwiczenie pokazuje pewne trudności, jakie napotykamy, próbując zmodyfikować algorytm Dijkstry, tak aby dawał on wagi maksymalne. Zastąp krok, w którym dokonujemy podstawienia w algorytmie Dijkstry, następującym krokiem:

jeśli $D(j) < D(k) + W(k, j)$, to
zastąp $D(j)$ sumą $D(k) + W(k, j)$

(a) Przypuśćmy, że tak zmodyfikowany algorytm wybiera wierzchołek $k \in V \backslash L$ tak, by wartość $D(k)$ była jak największa. Pokaż, że ten nowy algorytm nie daje poprawnej odpowiedzi dla grafu skierowanego przedstawionego na rysunku 8.23(a). (Zacznij od wierzchołka v_1).



Rysunek 8.23

(b) Przypuśćmy, że zmodyfikowany algorytm wybiera wierzchołek $k \in V \backslash L$ tak, by wartość $D(k)$ była nieujemna i przy tym możliwie najmniejsza. Pokaż, że ten nowy algorytm nie daje poprawnej odpowiedzi dla grafu skierowanego pokazanego na rysunku 8.23(b).

(c) Czy coś pomoże, jeśli w ćwiczeniu (a) lub (b) zastąpimy „dla każdego $j \in V \backslash L$ wykonuj" linią „dla każdego $j \in V$ wykonuj"?

12. (a) Pokaż, że algorytm WAGI MAKSYMALNE działa poprawnie; znajduje on wagi maksymalne dróg z wierzchołka v_1 do wszystkich innych wierzchołków.

(b) Pokaż, że algorytm WAGI MAKSYMALNE działa w czasie $O(n^2)$.

§ 8.4. Modyfikacje i zastosowania algorytmów na grafach skierowanych

Wersje algorytmów Dijkstry i Warshalla, które widzieliśmy w § 8.3, wyznaczały wagi minimalne i maksymalne. Możemy łatwo zmodyfikować te algorytmy tak, aby otrzymać również od-

powiadające im drogi minimalne i maksymalne. Podstawowy pomysł polega na tym, by drogi traktować jak listy powiązanych ze sobą wierzchołków.

Aby opisać drogę, związemy z każdym wierzchołkiem **wskaźnik** do następnego wierzchołka na drodze, aż dojdziemy do końca. Spróbujmy sobie wyobrazić turystę w nieznanym mieście, który pyta o drogę do dworca; idzie on od skrzyżowania do skrzyżowania, na każdym ponownie pytając o kierunek. Aby opisać drogę $xyzw$, potrzebujemy wskaźników od x do y, od y do z i od z do w, możemy je opisać za pomocą funkcji p takiej, że $p(x) = y$, $p(y) = z$ oraz $p(z) = w$. Dla innych wierzchołków $v \in V(G)$ możemy zdefiniować wartość $p(v)$ w dowolny sposób lub też możemy zdefiniować funkcję p tylko na zbiorze $\{x, y, z\}$.

Okazuje się, że w przypadku algorytmu Dijkstry znajdującego wagi minimalne dróg prowadzących od pewnego wybranego wierzchołka do wszystkich innych wierzchołków technicznie łatwiej będzie tworzyć wskaźniki skierowane w tył odpowiadających dróg minimalnych, a nie w przód. Jeśli znany jest ciąg wierzchołków od końca do początku drogi, to można łatwo odwrócić ten porządek i wypisać wierzchołki we właściwej kolejności.

Aby opisać zmodyfikowany algorytm Dijkstry, przyjmiemy, że zbiorem $V(G)$ jest $\{1, 2, \ldots, n\}$ oraz 1 jest tym wybranym wierzchołkiem. Tworzymy funkcję P o wartościach będących wskaźnikami, zdefiniowaną na zbiorze $\{2, \ldots, n\}$, taką że za każdym razem albo para $(P(j), j)$ jest ostatnią krawędzią na drodze o najmniejszej możliwej wadze, prowadzącej od wierzchołka 1 do wierzchołka j, albo $P(j) = 0$, jeśli nie została znaleziona droga od 1 do j. Kiedy algorytm kończy działanie, równość $P(j) = 0$ zachodzi, jeśli nie istnieje w grafie G droga od wierzchołka 1 do wierzchołka j. W przeciwnym przypadku w ciągu

$$j, P(j), P(P(j)), P(P(P(j))), \ldots$$

występują wierzchołki drogi minimalnej od 1 do j w odwrotnej kolejności.

Na początku niech $P(j) = 0$, jeśli nie istnieje krawędź z wierzchołka 1 do wierzchołka j; w przeciwnym razie niech $P(j) = 1$. W oryginalnym algorytmie dodajemy linię w pętli, w której dokonujemy podstawienia, taką, że gdy tylko zostanie znaleziona droga $1 \ldots kj$ lepsza od dotychczas najlepszej znanej drogi z wierzchołka 1 do wierzchołka j, to wskaźnik $P(j)$ przyjmie wartość k. Zatem wartość $P(j)$ jest zawsze przedostatnim przystankiem na najlepszej znanej drodze do wierzchołka j. A oto ten poprawiony algorytm.

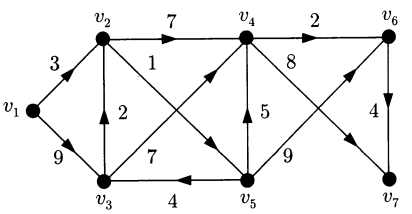
Algorytm Dijkstry ze wskaźnikami

{Dane: graf skierowany bez pętli i krawędzi wielokrotnych, zbiór jego wierzchołków $\{1,\ldots,n\}$, funkcja W wag krawędzi o wartościach nieujemnych}
{Wyniki: wagi minimalne $W^*(1,j)$ dla $j=2,\ldots,n$, wskaźniki $P(j)$ dla $j=2,\ldots,n$}
Niech $L := \emptyset$ oraz $V := \{2,\ldots,n\}$.
Dla $i \in V$ wykonuj
  $D(i) := W(1,i)$
  jeśli $W(1,i) = \infty$, to niech $P(i) := 0$,
  w przeciwnym przypadku niech $P(i) := 1$.
Dopóki $V \setminus L \neq \emptyset$, wykonuj
  wybierz $k \in V \setminus L$ o najmniejszej wartości $D(k)$
  dołącz k do zbioru L
  dla każdego $j \in V \setminus L$ wykonuj
    jeśli $D(j) > D(k) + W(k,j)$, to
      zastąp $D(j)$ sumą $D(k) + W(k,j)$
      zastąp $P(j)$ liczbą k.
Dla $j \in V$ wykonuj
  $W^*(1,j) := D(j)$. ∎

PRZYKŁAD 1 Weźmy graf skierowany z wagami przedstawiony na rysunku 8.24, jest to ten sam graf, który widzieliśmy w przykładzie 1 w § 8.3. W tablicy 8.2 widzimy kolejne wartości funkcji D i P w chwili, gdy nowy wierzchołek v_k został zbadany i dodany do zbioru L. Z ostatniego wiersza w tablicy 8.2 możemy odczytać drogę minimalną z wierzchołka 1 do wierzchołka 7, wypisaną w odwrotnym porządku:

$$7,\ P(7) = 6,\ P(6) = 4,\ P(4) = 5,\ P(5) = 2,\ P(2) = 1.$$

Zatem szukaną drogą jest $v_1v_2v_5v_4v_6v_7$. ∎



Rysunek 8.24

Tablica 8.2

k	$D(2)$	$D(3)$	$D(4)$	$D(5)$	$D(6)$	$D(7)$	$P(2)$	$P(3)$	$P(4)$	$P(5)$	$P(6)$	$P(7)$
1	3	9	∞	∞	∞	∞	1	1	0	0	0	0
2	3	9	10	4	∞	∞	1	1	2	2	0	0
5	3	8	9	4	13	∞	1	5	5	2	5	0
3	3	8	9	4	13	∞	1	5	5	2	5	0
4	3	8	9	4	11	17	1	5	5	2	4	4
6	3	8	9	4	11	15	1	5	5	2	4	6

W algorytmie Warshalla możemy również mieć funkcję wskaźnikową. W tym przypadku jednak można równie łatwo wybrać wskaźniki tak, by wskazywały kierunek naprzód, a więc tak, że za każdym razem para $(v_i, v_{P(i,j)})$ jest pierwszą krawędzią na drodze o najmniejszej znanej wadze z wierzchołka v_i do wierzchołka v_j, jeśli taka droga została znaleziona. Kiedy ten zmodyfikowany algorytm zakończy działanie, równość $P(i,j) = 0$ będzie zachodziła wtedy, gdy nie istnieje droga z wierzchołka v_i do wierzchołka v_j. W przeciwnym razie ciąg

$$i,\ P(i,j),\ P(P(i,j),j),\ P(P(P(i,j),j),j),\ \ldots$$

jest ciągiem indeksów wierzchołków na drodze minimalnej z wierzchołka v_i do wierzchołka v_j. Ponieważ po każdym indeksie k w tym ciągu następuje $P(k,j)$, ten ciąg można łatwo zdefiniować rekurencyjnie za pomocą funkcji P.

Możemy traktować P jako macierz o wymiarach $n \times n$, której wyrazy należą do zbioru $\{0,\ldots,n\}$. Na początku kładziemy $\mathbf{P}[i,j] = j$, jeśli istnieje krawędź z wierzchołka v_i do wierzchołka v_j i kładziemy $\mathbf{P}[i,j] = 0$ w przeciwnym przypadku. W algorytmie Warshalla dodajemy linię w pętli, w której dokonujemy podstawienia, w taki sposób, by za każdym razem, gdy algorytm znajdzie drogę $v_i \ldots v_k v_j$ z wierzchołka v_i do wierzchołka v_j, lepszą od dotychczas znalezionej najlepszej drogi, wskaźnikowi $\mathbf{P}[i,j]$ została nadana wartość $\mathbf{P}[i,k]$; aby skierować się od v_i do v_j, należy skierować się najpierw do v_k. Zmodyfikowany algorytm Warshalla wygląda następująco.

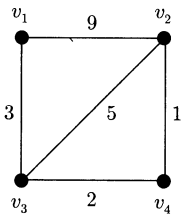
Algorytm Warshalla ze wskaźnikami

{Dane: macierz $\mathbf{W}_0$ nieujemnych wag krawędzi grafu skierowanego bez pętli i krawędzi wielokrotnych}
{Wyniki: macierz $\mathbf{W}^*$ wag minimalnych i macierz wskaźników $\mathbf{P}^*$ dla tego grafu}
$\mathbf{W} := \mathbf{W}_0$
dla $i = 1$ do n wykonuj

dla $j = 1$ do n wykonuj
jeśli istnieje krawędź z v_i do v_j, to
$\mathbf{P}[i,j] := j$
w przeciwnym przypadku
$\mathbf{P}[i,j] := 0$
dla $k = 1$ do n wykonuj
dla $i = 1$ do n wykonuj
dla $j = 1$ do n wykonuj
jeśli $\mathbf{W}[i,j] > \mathbf{W}[i,k] + \mathbf{W}[k,j]$, to
zastąp $\mathbf{W}[i,j]$ sumą $\mathbf{W}[i,k] + \mathbf{W}[k,j]$.
zastąp $\mathbf{P}[i,j]$ wartością $\mathbf{P}[i,k]$.
$\mathbf{W}^* := \mathbf{W}$, $\mathbf{P}^* := \mathbf{P}$. ∎

Za pomocą takich samych modyfikacji otrzymamy drogi maksymalne w przypadku, gdy algorytm Warshalla został zmieniony tak, aby znajdował wagi maksymalne.

Zarówno algorytm Dijkstry, jak i algorytm Warshalla mogą być stosowane do grafów nieskierowanych, po prostu wystarczy zastąpić każdą krawędź nieskierowaną $\{u,v\}$ taką, że $u \neq v$ dwoma krawędziami skierowanymi (u,v) i (v,u). Jeśli krawędź nieskierowana ma wagę w, to obu krawędziom skierowanym przypisuje się tę samą wagę w. Jeśli graf nie ma wag, to każdej krawędzi przypisuje się wagę 1. Jeśli poszukuje się dróg minimalnych w grafach nieskierowanych, to pętle nie mają żadnego znaczenia, a zatem wygodnie jest przyjąć, że $W(i,i) = 0$ dla wszystkich i.



Rysunek 8.25

PRZYKŁAD 2 Weźmy graf z wagami pokazany na rysunku 8.25. Za pomocą algorytmu Warshalla ze wskaźnikami znajdujemy wagi i drogi minimalne między wierzchołkami, przy czym dopuszczamy podróżowanie wzdłuż krawędzi w obie strony. Poniżej widzimy kolejne wartości macierzy $\mathbf{W}$ i $\mathbf{P}$. Pokażemy przykładowe obliczenie $\mathbf{P}_4[2,1]$. Ponieważ $\mathbf{W}_3[2,4]+\mathbf{W}_3[4,1] = 1+5 = 6 < 8 = \mathbf{W}_3[2,1]$, więc $\mathbf{W}_4[2,1] = 6$ oraz $\mathbf{P}_4[2,1] = \mathbf{P}_3[2,4] = 4$. Droga minimalna

z wierzchołka v_2 do wierzchołka v_1 jest opisana za pomocą ciągu 2, $\mathbf{P}^*[2,1] = 4$, $\mathbf{P}^*[4,1] = 3$, $\mathbf{P}^*[3,1] = 1$.

Wartości na przekątnej macierzy $\mathbf{P}$ nie mają w tym przykładzie specjalnego znaczenia. ∎

$$\mathbf{W}_0 = \mathbf{W}_1 = \begin{bmatrix} 0 & 9 & 3 & \infty \\ 9 & 0 & 5 & 1 \\ 3 & 5 & 0 & 2 \\ \infty & 1 & 2 & 0 \end{bmatrix}, \quad \mathbf{P}_0 = \mathbf{P}_1 = \begin{bmatrix} 0 & 2 & 3 & 0 \\ 1 & 0 & 3 & 4 \\ 1 & 2 & 0 & 4 \\ 0 & 2 & 3 & 0 \end{bmatrix},$$

$$\mathbf{W}_2 = \begin{bmatrix} 0 & 9 & 3 & 10 \\ 9 & 0 & 5 & 1 \\ 3 & 5 & 0 & 2 \\ 10 & 1 & 2 & 0 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 0 & 2 & 3 & 2 \\ 1 & 0 & 3 & 4 \\ 1 & 2 & 0 & 4 \\ 2 & 2 & 3 & 0 \end{bmatrix},$$

$$\mathbf{W}_3 = \begin{bmatrix} 0 & 8 & 3 & 5 \\ 8 & 0 & 5 & 1 \\ 3 & 5 & 0 & 2 \\ 5 & 1 & 2 & 0 \end{bmatrix}, \quad \mathbf{P}_3 = \begin{bmatrix} 0 & 3 & 3 & 3 \\ 3 & 0 & 3 & 4 \\ 1 & 2 & 0 & 4 \\ 3 & 2 & 3 & 0 \end{bmatrix},$$

$$\mathbf{W}^* = \mathbf{W}_4 = \begin{bmatrix} 0 & 6 & 3 & 5 \\ 6 & 0 & 3 & 1 \\ 3 & 3 & 0 & 2 \\ 5 & 1 & 2 & 0 \end{bmatrix}, \quad \mathbf{P}^* = \mathbf{P}_4 = \begin{bmatrix} 0 & 3 & 3 & 3 \\ 4 & 0 & 4 & 4 \\ 1 & 4 & 0 & 4 \\ 3 & 2 & 3 & 0 \end{bmatrix}.$$

Można używać algorytmu Warshalla do znajdowania zbiorów $R(v)$ wierzchołków osiągalnych z wierzchołka v. Po prostu nadajemy wszystkim krawędziom, włączając pętle, wagi równe 1. Ostateczna wartość $\mathbf{W}^*[i,j]$ jest równa ∞, jeśli nie istnieje droga z wierzchołka v_i do wierzchołka v_j, i jest liczbą całkowitą dodatnią, jeśli taka droga istnieje. W szczególności $\mathbf{W}^*[i,i] < \infty$ wtedy i tylko wtedy, gdy wierzchołek v_i należy do pewnego cyklu w grafie G. Możemy sprawdzić, czy graf skierowany G jest acykliczny, stosując algorytm Warshalla i sprawdzając wyrazy stojące na przekątnej w macierzy $\mathbf{W}^*$.

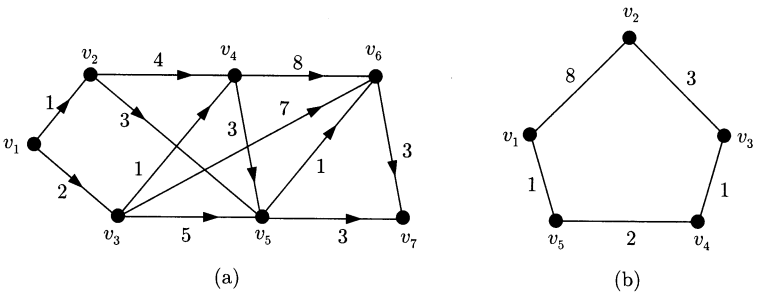
W algorytmie Fleury'ego z § 6.2 trzeba było sprawdzać, czy usunięcie danej krawędzi e z grafu nieskierowanego G powodowało zwiększenie liczby składowych, tzn. czy końce krawędzi e były osiągalne jeden z drugiego w grafie $G\backslash\{e\}$. Można sprawdzić osiągalność za pomocą algorytmu Warshalla, ale algorytm Dijkstry jest szybszy; jeśli krawędź e łączy wierzchołki v_i i v_j, to możemy zastosować algorytm Dijkstry do grafu $G\backslash\{e\}$ i wierzchołka v_i jako wierzchołka początkowego. Co wię-

cej, na mocy twierdzenia 3 z § 6.2, w ten sposób sprawdzamy po prostu, czy krawędź e należy do jakiegoś cyklu w grafie G, a więc sprawdzając w ten sposób po kolei każdą krawędź, możemy sprawdzić, czy graf nieskierowany G jest acykliczny. W paragrafie 6.6 widzieliśmy jeszcze szybsze testy acykliczności grafu, oparte na algorytmie LAS i algorytmie Kruskala.

ĆWICZENIA DO § 8.4

1. (a) Znajdź macierze $\mathbf{P}$ wskaźników początkowej i końcowej drogi minimalnej dla algorytmu Warshalla zastosowanego do grafu skierowanego przedstawionego na rysunku 8.26(a). (Por. ćwiczenie 1 w § 8.3). Możesz zastosować dowolną metodę, aby znaleźć odpowiedź, nawet możesz wpatrywać się w rysunek.

(b) Powtórz ćwiczenie (a) dla dróg maksymalnych zamiast dla dróg minimalnych.



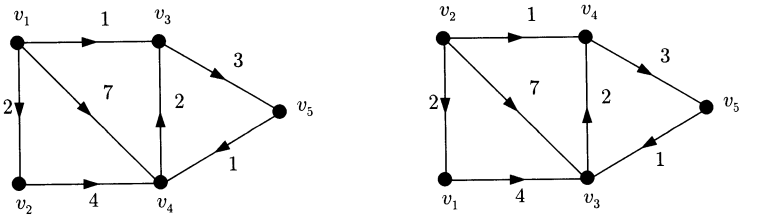
Rysunek 8.26

2. Użyj algorytmu Warshalla ze wskaźnikami, aby znaleźć macierze $\mathbf{W}^*$ i $\mathbf{P}^*$ dla grafu przedstawionego na rysunku 8.26(b).

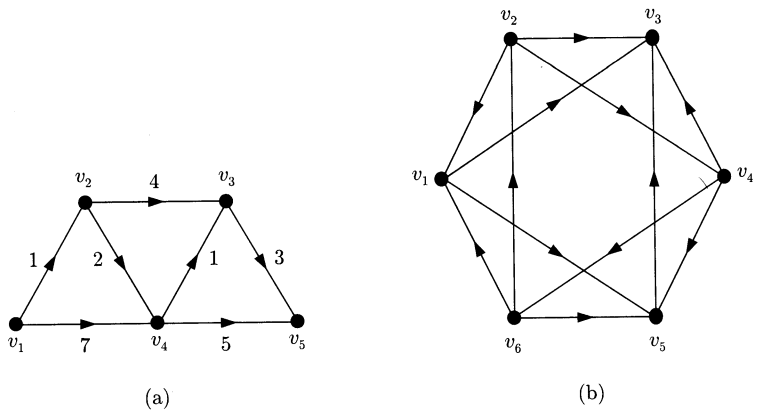
3. (a) Zastosuj algorytm Dijkstry ze wskaźnikami do grafu skierowanego pokazanego na rysunku 8.27(a), zaczynając od wierzchołka v_1. Swoją odpowiedź napisz w takiej postaci, jaka jest przedstawiona w tablicy 8.2.

(b) Powtórz ćwiczenie (a) dla grafu skierowanego przedstawionego na rysunku 8.27(b).

4. Wypisz wierzchołki dróg minimalnych z wierzchołka v_1 do wierzchołka v_2, z v_1 do v_4 i z v_1 do v_6 dla grafu skierowanego przedstawionego na rysunku 8.24. *Wskazówka*: zobacz tablicę 8.2.



Rysunek 8.27



Rysunek 8.28

5. (a) Użyj algorytmu Warshalla ze wskaźnikami, aby znaleźć macierz $\mathbf{P}^*$ dla grafu przedstawionego na rysunku 8.28(a).

(b) Powtórz ćwiczenie (a) dla dróg maksymalnych zamiast dróg minimalnych.

6. **Macierz osiągalności** $\mathbf{M}_R$ dla grafu skierowanego jest określona w następujący sposób: $\mathbf{M}_R[i,j] = 1$, jeśli $v_j \in R(v_i)$ i $\mathbf{M}_R[i,j] = 0$ w przeciwnym przypadku.

(a) Znajdź macierz osiągalności $\mathbf{M}_R$ dla grafu skierowanego przedstawionego na rysunku 8.28(b), używając algorytmu Warshalla.

(b) Czy ten graf skierowany jest acykliczny?

7. (a) Zmodyfikuj algorytm WAGI MAKSYMALNE z § 8.3, aby otrzymać algorytm, za pomocą którego znajduje się wskaźniki wzdłuż dróg maksymalnych z pojedynczego wierzchołka.

(b) Zastosuj swój algorytm z ćwiczenia (a), aby znaleźć wskaźniki drogi maksymalnej z wierzchołka v_1 w grafie skierowanym przedstawionym na rysunku 8.18(a) w § 8.3.

8. (a) Narysuj graf skierowany, którego macierzą wag jest

$$\mathbf{W}_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

(b) Czy ten graf skierowany jest acykliczny?
(c) Znajdź macierz osiągalności $\mathbf{M}_R$ (zob. ćwiczenie 6) dla tego grafu skierowanego.

9. Powtórz ćwiczenie 8 dla grafu skierowanego, którego macierzą wag jest macierz

$$\mathbf{W}_0 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

10. Zmodyfikuj algorytm Warshalla tak, by $\mathbf{W}[i,j] = 0$, jeśli nie została znaleziona droga z v_i do v_j i $\mathbf{W}[i,j] = 1$, jeśli taka droga jest znana. *Wskazówka*: dokonaj wygodnego zapoczątkowania i użyj $\min\{\mathbf{W}[i,k], \mathbf{W}[k,j]\}$.

11. (a) Pokaż, że algorytm Dijkstry ze wskaźnikami potrzebuje czasu rzędu $O(n^2)$ dla grafu skierowanego o n wierzchołkach.
(b) Podaj najlepsze, jakie umiesz, oszacowanie czasu działania algorytmu Warshalla ze wskaźnikami.

To, co jest najważniejsze w tym rozdziale

Jak zwykle: Co to znaczy? Dlaczego znajduje się to tutaj? Jak można tego użyć? Zastanów się nad przykładami.

Pojęcia i oznaczenia

ujście, źródło
etykietowanie uporządkowane, etykietowanie uporządkowane odwrotnie
stopień wejściowy, stopień wyjściowy
graf skierowany z wagami
waga minimalna, droga minimalna, $W^*(u,v)$
waga maksymalna, droga maksymalna, $M(u,v)$
sieć zdarzeń

droga krytyczna, krawędź krytyczna
$A(v), L(v)$
rezerwa czasowa $S(v)$ wierzchołka
rezerwa czasowa $F(u, v)$ krawędzi
wskaźnik

Fakty

Droga w grafie skierowanym jest acykliczna wtedy i tylko wtedy, gdy jej wierzchołki są różne.

W każdym skończonym acyklicznym grafie skierowanym istnieje etykietowanie uporządkowane (wiadomo to też z rozdziału 7).

Graf skierowany ma cykl Eulera wtedy i tylko wtedy, gdy jest spójny jako graf i każdy wierzchołek ma ten sam stopień wejściowy i stopień wyjściowy.

Algorytmy

Algorytm UJŚCIE do znajdowania ujścia w skończonym acyklicznym grafie skierowanym.

Algorytm NUMEROWANIE WIERZCHOŁKÓW do znajdowania etykietowania uporządkowanego acyklicznego grafu skierowanego w czasie $O(|V(G)|^2)$.

Algorytm Dijkstry do obliczania wag minimalnych z wybranego wierzchołka w czasie $O(|V(G)|^2)$ (lub lepszym) dla danego etykietowania uporządkowanego.

Algorytm Warshalla do obliczania wag minimalnych lub wag maksymalnych między wszystkimi parami wierzchołków, jak również do określania osiągalności w czasie $O(|V(G)|^3)$.

Algorytm WAGI MAKSYMALNE do znajdowania wag maksymalnych z wybranego wierzchołka w czasie $O(|V(G)|^2)$ lub $O(|V(G)| + |E(G)|)$ dla danego etykietowania uporządkowanego odwrotnie.

Algorytmy Dijkstry i Warshalla oraz WAGI MAKSYMALNE ze wskaźnikami do opisywania dróg odpowiadających wagom minimalnym lub maksymalnym.

Algorytmy Dijkstry i Warshalla do znajdowania wag minimalnych i dróg minimalnych dla grafów nieskierowanych (gdzie każda waga $W(i, i)$ na początku równa jest 0).

9. RACHUNEK PRAWDOPODOBIEŃSTWA

Rozdział ten stanowi rozwinięcie krótkiego wstępu do rachunku prawdopodobieństwa z rozdziału 5. W pierwszym paragrafie wprowadzone są pojęcia niezależności i prawdopodobieństwa warunkowego zdarzeń. Począwszy od § 9.2 przenosimy uwagę z przestrzeni zdarzeń elementarnych na zmienne losowe i ich rozkłady. W paragrafie 9.3 podane są podstawowe własności wartości oczekiwanej i wariancji, jak również zawarte jest omówienie niezależności zmiennych losowych. W ostatnim paragrafie tego rozdziału przyglądamy się bliżej rozkładom dwumianowym i wskazujemy, bez dowodu, na ich związek z rozkładem normalnym.

§ 9.1. Niezależność

Tak jak w § 5.2, nasz rachunek prawdopodobieństwa związany jest z przestrzenią zdarzeń elementarnych Ω i funkcją prawdopodobieństwa P; dla zdarzenia $E \subseteq \Omega$, $P(E)$ jest pewną liczbą z przedziału $[0, 1]$, która reprezentuje prawdopodobieństwo zajścia zdarzenia E.

Prawdopodobieństwa zdarzeń mogą się zmieniać, jeśli mamy o nich więcej informacji, to znaczy, jeśli wiemy, że interesujące nas wyniki należą do jakiegoś danego podzbioru przestrzeni Ω. Na przykład, jeśli dowiem się, że mój przeciwnik w grze w pokera ma asa, to teraz będę uważał za bardziej prawdopodobne, iż ma on dwa asy, niż sądziłbym tak wcześniej, przed uzyskaniem tej informacji. Z drugiej strony, jeśli dowiem się, że wygrał on właśnie na loterii, to fakt ten wydaje się nie mieć żadnego związku z szansą

posiadania przez niego dwóch asów. W tym paragrafie naszym celem jest nadanie matematycznego sensu wyrażeniom takim, jak „prawdopodobieństwo zdarzenia A pod warunkiem zdarzenia B" oraz „zdarzenie A jest niezależne od zdarzenia B".

PRZYKŁAD 1 W tablicy 9.1 wypisane są elementy zbioru Ω złożonego z 36 jednakowo prawdopodobnych wyników rzutu dwiema symetrycznymi kostkami, białą i szarą. Sytuacja ta była omawiana w przykładzie 4 z § 5.2. Niech B oznacza zdarzenie „liczba oczek na kostce białej jest ≤ 3", R oznacza „liczba oczek na kostce szarej jest ≥ 5" i S oznacza „suma liczb oczek na kostkach jest ≥ 8.

Tablica 9.1

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

B (obejmuje trzy górne rzędy)

R (obejmuje dwie ostatnie kolumny)

Wyniki należące do zbioru B znajdują się w trzech górnych rzędach tabl.9.1, wyniki należące do zbioru R zajmują jej dwie ostatnie kolumny, a te poniżej przerywanej linii należą do zbioru S.

(a) Zauważmy, że $P(R) = \frac{1}{3}$. Prawdopodobieństwo tego, że na szarej kostce wypadło 5 lub 6 oczek, byłoby wyższe, gdybyśmy wiedzieli, że suma oczek jest ≥ 8, tzn., gdybyśmy wiedzieli, że zaszło zdarzenie S. Zbiór S składa się z 15 możliwych wyników, znajdujących się w tablicy poniżej przerywanej linii. Liczba oczek na kostce szarej jest ≥ 5 w 9 spośród 15 wyników należących do S, a więc wiedza o tym, że zaszło zdarzenie S sprawia, że prawdopodobieństwo zdarzenia R jest równe $\frac{9}{15} = 0{,}6$, gdyż wszystkie wyniki należące do S są jednakowo prawdopodobne.

(b) Zauważmy, że $P(B) = \frac{1}{2}$, ale prawdopodobieństwo tego, że liczba oczek na białej kostce jest ≤ 3, maleje do $\frac{3}{15} = 0{,}2$, jeśli wiemy, że suma liczb oczek na obu kostkach jest ≥ 8.

(c) Jak zmieniłaby się wartość $P(B) = \frac{1}{2}$, jeśli wiedzielibyśmy, że zaszło zdarzenie R? To znaczy, ile wynosi prawdopodobieństwo tego, że liczba oczek na kostce białej jest ≤ 3, jeśli wiemy, że na kostce szarej jest ona ≥ 5? Dwanaście wyników z tablicy 9.1 należy do zbioru R, a sześć spośród nich należy również do zbioru B, a więc prawdopodobieństwo to wynosi nadal $\frac{1}{2}$. Brzmi to rozsądnie; nie oczekujemy, aby informacje dotyczące kostki szarej

powiedziały nam coś na temat kostki białej. Fakt, iż wiemy, że wynik należy do R, zmniejsza zbiór możliwych rezultatów, ale nie wpływa na prawdopodobieństwo tego, by wynik był w B.

Poprzedni akapit jest trochę mylący. Wygląda na to, że użyliśmy rachunku prawdopodobieństwa do pokazania, że wiedza dotycząca szarej kostki nie ma wpływu na prawdopodobieństwa zdarzeń związanych z kostką białą. Tymczasem prawdą jest coś odwrotnego. *Wierzymy*, że wynik uzyskany na szarej kostce nie wpływa na prawdopodobieństwa zdarzeń związanych z kostką białą. Dlatego, jak wyjaśnimy w przykładzie 7, *określiliśmy* funkcję prawdopodobieństwa P tak, by fakt ten znajdował swoje odbicie w teorii. Innymi słowy, uważamy każdy z 36 wyników z tablicy 9.1 za jednakowo prawdopodobny, *ponieważ* taka definicja funkcji prawdopodobieństwa ma tę własność, iż informacje dotyczące kostki szarej nie wpływają na prawdopodobieństwa zdarzeń dotyczących jedynie kostki białej i odwrotnie. ∎

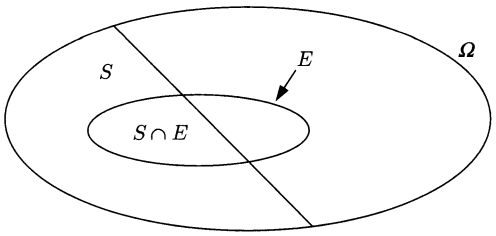
Nadamy teraz formalny kształt idei zawartej w przykładzie 1. Przypuśćmy, że wiemy, iż wynik ω należy do pewnego zdarzenia $S \subseteq \Omega$. Wówczas wynik ten należy do zdarzenia E wtedy i tylko wtedy, gdy należy on do zdarzenia $E \cap S$, a więc prawdopodobieństwo tego, że ω należy do zdarzenia E, jeśli wiemy, że ω należy do zdarzenia S, powinno zależeć jedynie od prawdopodobieństwa zdarzenia $E \cap S$. Jeśli wszystkie wyniki należące do zbioru S są jednakowo prawdopodobne, to prawdopodobieństwo zdarzenia E pod warunkiem zdarzenia S, tzn. prawdopodobieństwo zdarzenia $E \cap S$ pod warunkiem S, jest po prostu równe ułamkowi pokazującemu, ile spośród wyników należących do zbioru S znajduje się w zbiorze $E \cap S$, a więc wynosi ono $|E \cap S|/|S|$. Ogólniej, prawdopodobieństwo zdarzenia $E \cap S$ pod warunkiem zdarzenia S powinno równać się ułamkowi wartości $P(S)$, odpowiadającemu zdarzeniu $E \cap S$. Jeśli $P(S) > 0$, to definiujemy **prawdopodobieństwo warunkowe** $P(E|S)$, co czytamy „prawdopodobieństwo warunkowe zdarzenia E pod warunkiem (lub: „przy warunku"), że zaszło zdarzenie S"(lub: „przy danym S"), wzorem:

$$P(E|S) = \frac{P(E \cap S)}{P(S)} \quad \text{dla} \quad E \subseteq \Omega.$$

Wtedy $P(\Omega|S)=P(\Omega \cap S)/P(S)=P(S)/P(S)=1$, $P(E|S)=P(E \cap S|S)$ dla $E \subseteq \Omega$ i nietrudno jest stwierdzić, że funkcja $E \to P(E|S)$ spełnia warunki definicji prawdopodobieństwa na Ω.

Ponieważ $P(S|S) = 1$, to można by także uważać zbiór S, zamiast zbioru Ω , za przestrzeń zdarzeń elementarnych, do któ-

rej odnosi się prawdopodobieństwo warunkowe, definiując w takim przypadku $P(E|S)$ jedynie dla zdarzeń $E \subseteq S$. Rysunek 9.1 pokazuje, w jaki sposób traktować możemy zbiór S jako swego rodzaju „obciętą" wersję przestrzeni Ω, gdzie zdarzeniu E w Ω odpowiada zdarzenie $S \cap E$ w S. Jeśli $E \subseteq S$ i $P(S) < 1$, to, zgodnie z oczekiwaniami, $P(E|S) = P(E)/P(S) > P(E)$; fakt, że zaszło zdarzenie E czyni zajście zdarzenia S bardziej prawdopodobnym, niż byłoby ono bez tej dodatkowej informacji. Jeśli zdarzenie E nie jest zawarte w S, to liczba $P(E|S)$ może być większa, mniejsza bądź równa liczbie $P(E)$, w zależności od okoliczności.



Rysunek 9.1

PRZYKŁAD 2 Wracamy do zagadnienia rzutu dwiema kostkami, a więc zbiory Ω, B, R i S są określone tak, jak w przykładzie 1 i tablicy 9.1.

(a) Stosując nasze nowe oznaczenia, stwierdzenie z przykładu 1(a) można zapisać po prostu jako równość $P(R|S) = \frac{9}{15}$. Istotnie,

$$P(R|S) = \frac{P(R \cap S)}{P(S)} = \frac{9/36}{15/36} = \frac{9}{15}.$$

W przykładzie 1(b) zauważyliśmy, że $P(B|S) = \frac{3}{15}$; zgadza się to z faktem, iż

$$P(B|S) = \frac{P(B \cap S)}{P(S)} = \frac{3/36}{15/36} = \frac{3}{15}.$$

Może się wydawać, że nasza nowa definicja prawdopodobieństwa warunkowego $P(E|S)$ skomplikowała jedynie sytuację, powodując wielokrotne pojawienie się liczby 36, którą i tak eliminuje się przez skracanie. Ta nowa definicja ma jednak sens nawet wtedy, gdy wyniki nie są jednakowo prawdopodobne, a w takim przypadku zliczanie możliwych wyników już nie wystarcza.

(b) W przykładzie 1(c) widzieliśmy, że $P(B|R) = \frac{1}{2} = P(B)$.

Podobnie,

$$P(R|B) = \frac{P(R \cap B)}{P(B)} = \frac{6/36}{18/36} = \frac{1}{3} = P(R).$$

Powtarzamy: wiedza na temat zdarzenia R nie wpływa na prawdopodobieństwo zdarzenia B i odwrotnie. ∎

Być może najważniejsze pytanie w rachunku prawdopodobieństwa i statystyce brzmi: czy zajście jednego z dwóch zdarzeń zmienia szanse zajścia drugiego z nich? Jeśli nie, to zdarzenia te z punktu widzenia rachunku prawdopodobieństwa uważamy za niezależne. Jeśli tak, to jedno ze zdarzeń ma wpływ na drugie i, jeśli jest to ważne, szukać można przyczyn tego wpływu.

PRZYKŁAD 3 Niech Ω będzie zbiorem wszystkich dorosłych Amerykanów płci męskiej. Zakładamy, że w dowolnym badaniu sondażowym istnieją jednakowe szanse wylosowania każdego z nich. Jeśli S jest zdarzeniem „palił od co najmniej 10 lat", a C jest zdarzeniem „ma raka płuc", to badania sondażowe wykazują, że liczba $P(C|S)$ jest znacznie większa od liczby $P(C)$. Wydaje się, że mamy tu do czynienia z przyczyną i jej skutkiem. Z punktu widzenia rachunku prawdopodobieństwa z pewnością tak jest. Palenie zwiększa prawdopodobieństwo zachorowania na raka płuc, choć nie każdy palacz na raka płuc choruje. Rak płuc nie „zależy" od palenia w tym znaczeniu, że musisz palić, aby na raka płuc zachorować, ale palenie zwiększa szanse zachorowania. Rak płuc jest „zależny" od palenia w sensie rachunku prawdopodobieństwa.

Niech M oznacza zdarzenie „studiował matematykę w szkole wyższej". O ile wiemy, $P(C|M) = P(C)$. W tym przypadku, zachorowanie na raka płuc jest z punktu widzenia rachunku prawdopodobieństwa niezależne od studiowania matematyki. W skrócie, zdarzenia C i M są (probabilistycznie) niezależne.

Przypuśćmy, że odkryliśmy, iż $P(C|M) > P(C)$. Zdrowy rozsądek mówi nam, że matematyka nie jest bezpośrednią przyczyną raka. Szukalibyśmy więc rozsądniejszego wyjaśnienia. Być może ludzie z wyższym wykształceniem prowadzą bardziej stresujące życie i rak szybciej rozwija się u takich ludzi. I tak dalej. ∎

Mówimy o zdarzeniach A i B, gdzie $P(B) > 0$, że są **niezależne**, jeśli $P(A|B) = P(A)$. Warunek ten jest równoważny z warunkami: $P(A \cap B)/P(B) = P(A)$ i $P(A \cap B) = P(A) \cdot P(B)$,

oraz jeśli $P(A) > 0$, to również z warunkiem

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = P(B).$$

Zatem jeśli zdarzenia A i B są niezależne, to zdarzenia B i A są niezależne; to znaczy, że niezależność zdarzeń jest relacją symetryczną. Równoważne sformułowanie

(I) $P(A \cap B) = P(A) \cdot P(B)$ dla niezależnych zdarzeń A i B

przydaje się w obliczeniach. Ponadto ma ono sens nawet wtedy, gdy $P(A)$ lub $P(B)$ wynosi 0. Przyjmiemy więc warunek (I) za definicję **niezależności zdarzeń**. Zgodnie z tradycją mówić będziemy o parach zdarzeń niezależnych, ale będziemy przy tym używać oznaczenia $\{A, B\}$ dla podkreślenia, że kolejność zdarzeń jest nieistotna.

PRZYKŁAD 4 Powróćmy do problemu dwóch kostek z przykładów 1 i 2. Zdarzenia R i S nie są niezależne, gdyż $P(R) = \frac{1}{3}$, podczas gdy $P(R|S) = \frac{9}{15} = \frac{3}{5}$. Ponadto,

$$P(S) = \frac{15}{36} = \frac{5}{12}, \text{ podczas gdy } P(S|R) = \frac{9}{12} = \frac{3}{4}.$$

Podobnie, zdarzenia B i S nie są niezależne (sprawdź to).

Zdarzenia B i R są natomiast niezależne:

$$P(B|R) = \frac{1}{2} = P(B) \quad \text{oraz} \quad P(R|B) = \frac{1}{3} = P(R). \qquad \blacksquare$$

Ludzie mówią czasem, że dwa zdarzenia są od siebie niezależne, jeśli nie mogą one zajść jednocześnie, tzn. gdy są one rozłączne. Jest to nie do pogodzenia z naszą definicją i należy tego unikać. Jeśli $A \cap B = \emptyset$, to informacja, że zaszło zdarzenie B mówi nam coś na temat zdarzenia A, a mianowicie, że zdarzenie A nie zaszło: $P(A|B) = 0$. Zdarzenia rozłączne nie są niezależne, chyba, że $P(A)$ lub $P(B)$ równa się 0.

Zanim wyjaśnimy, co to znaczy, że ciąg $A_1, A_2, \ldots, A_n$ składa się ze zdarzeń wzajemnie niezależnych, zastanówmy się, co to powinno znaczyć dla trzech zdarzeń A_1, A_2, A_3. Nie wystarczy, by każde dwa spośród tych zdarzeń były niezależne (mówi się wtedy, że zdarzenia te są **parami niezależne**). Na przykład, aby mieć naprawdę do czynienia z niezależnością, prawdopodobieństwo zdarzenia A_1 nie powinno się zmienić, jeśli wiemy, że zaszły oba zdarzenia A_2 i A_3 (tzn. $P(A_1|A_2 \cap A_3)$ powinno się równać $P(A_1)$).

PRZYKŁAD 5 Ponownie rozważymy problem dwóch kostek. Niech B_0 będzie zdarzeniem „liczba oczek na kostce białej jest nieparzysta", R_0 oznacza „liczba oczek na kostce szarej jest nieparzysta" i E oznacza „suma oczek jest parzysta". Łatwo jest sprawdzić, że zdarzenia należące do każdej z par $\{B_0, R_0\}$, $\{B_0, E\}$ i $\{R_0, E\}$ są niezależne (ćwiczenie 3). Z drugiej strony, jeśli zajdą oba zdarzenia B_0 i R_0, to zajście zdarzenia E jest pewne. Znaczy to, że

$$P(E|B_0 \cap R_0) = 1 \neq P(E) = \frac{1}{2}.$$

Zatem nie uznalibyśmy zdarzeń B_0, R_0, E za wzajemnie niezależne, mimo iż są one parami niezależne. ∎

Zatem, by zdarzenia A_1, A_2, A_3 były wzajemnie niezależne, żądamy, aby

$$P(A_1) = P(A_1|A_2 \cap A_3) = \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2 \cap A_3)},$$

a więc, by

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3),$$

dodatkowo, oprócz niezależności parami. Krótko mówiąc, wymagamy by

$$\text{(I')} \qquad P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$

dla wszystkich niepustych podzbiorów J zbioru $\{1, 2, 3\}$. To jest właściwa definicja i działa ona w sytuacji ogólnej. Zdarzenia $A_1, A_2, \ldots, A_n$ są wzajemnie niezależne, jeśli warunek (I') jest spełniony dla wszystkich niepustych podzbiorów J zbioru $\{1, 2, \ldots, n\}$.

PRZYKŁAD 6 Rzucamy n razy monetą symetryczną tak, jak w przykładzie 5 z § 5.2. Dla $k = 1, 2, \ldots, n$ niech E_k oznacza zdarzenie „orzeł w k-tym rzucie". Zakładamy, że zdarzenia te są wzajemnie niezależne, gdyż informacja o tym, co stało się w, powiedzmy, pierwszym i piątym rzucie, nie powinna wpływać na prawdopodobieństwo wypadnięcia orła w drugim lub siódmym rzucie. Oczywiście, $P(E_k) = \frac{1}{2}$ dla każdego k, gdyż nasza moneta jest symetryczna. Z niezależności zdarzeń wynika równość

$$P\left(\bigcap_{k=1}^{n} E_k\right) = \prod_{k=1}^{n} P(E_k) = \frac{1}{2^n}.$$

Innymi słowy, prawdopodobieństwo uzyskania samych orłów wynosi $1/2^n$. Podobne rozumowanie pokazuje, że prawdopodobieństwo uzyskania każdego ustalonego ciągu złożonego z n orłów i reszek wynosi $1/2^n$. Zatem założenie, że wyniki różnych rzutów są zdarzeniami wzajemnie niezależnymi, prowadzi do wniosku, że prawdopodobieństwo każdego ciągu długości n powinno wynosić $1/2^n$ tak, jak w przykładzie 5 z § 5.2. ∎

PRZYKŁAD 7 Pokażemy teraz, że funkcja prawdopodobieństwa, której używaliśmy w zagadnieniu rzutu dwiema kostkami, jest określona przez następujące założenia: prawdopodobieństwo wypadnięcia każdej danej liczby oczek na kostce białej wynosi $\frac{1}{6}$, prawdopodobieństwo wypadnięcia każdej danej liczby oczek na kostce szarej wynosi $\frac{1}{6}$ i wyniki uzyskane na kostce szarej są niezależne od wyników uzyskanych na kostce białej. Skoncentrujmy naszą uwagę na parze (4,5). Jest to jedyny wynik należący do zbioru $B_4 \cap R_5$, gdzie B_4 oznacza „na białej kostce 4 oczka" i R_5 oznacza „na szarej kostce 5 oczek". Wówczas $P(B_4) = P(R_5) = \frac{1}{6}$ i na mocy niezależności

$$P(4,5) = P(B_4 \cap R_5) = P(B_4) \cdot P(R_5) = \frac{1}{36}.$$

To dlatego założyliśmy, że $P(k,l) = \frac{1}{36}$ dla wszystkich par (k,l) znajdujących się na rysunku 9.1. ∎

Jeśli zdarzenia A i B nie są niezależne, to równość $P(A \cap B) = P(A) \cdot P(B)$ *nie zachodzi*. Natomiast, jeśli $P(A) \neq 0$ i $P(B) \neq 0$, to mamy zawsze równości

$$P(A \cap B) = P(A) \cdot P(B|A) \quad \text{oraz} \quad P(A \cap B) = P(B) \cdot P(A|B),$$

które są użyteczne, gdy jedna z wartości $P(B|A)$ lub $P(A|B)$ jest łatwa do wyznaczenia.

PRZYKŁAD 8 Wyciągamy dwie karty z talii 52 kart.

(a) Jakie jest prawdopodobieństwo, że obie karty są asami, jeśli pierwsza karta jest zwracana do talii (która następnie zostaje potasowana) przed wyciągnięciem drugiej karty, tzn. jeśli jest to losowanie ze zwracaniem? Zdarzenia: A_1, które polega na tym, że „pierwsza karta jest asem" i A_2, polegające na tym, że „druga karta jest asem", są niezależne, zatem

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) = \frac{1}{13} \cdot \frac{1}{13} = \frac{1}{169}.$$

Ułamek $\frac{1}{13}$ bierze się stąd, że w talii 52 kart są cztery asy.

(b) Jakie jest prawdopodobieństwo, że obie karty są asami, jeśli pierwszej karty nie zwracamy do talii (losowanie bez zwracania)? Zdarzenia nie są już niezależne: jeśli zachodzi zdarzenie A_1, to wśród pozostałych kart jest o jednego asa mniej. Stwierdzamy, że

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}.$$

Możemy także potraktować to zagadnienie, jak problem losowego wyboru dwuelementowego podzbioru z talii 52 kart. Prawdopodobieństwo wyciągnięcia dwóch asów wynosi wówczas

$$\frac{\binom{4}{2}}{\binom{52}{2}} = \frac{\frac{4 \cdot 3}{2 \cdot 1}}{\frac{52 \cdot 51}{2 \cdot 1}} = \frac{1}{221}. \quad \blacksquare$$

PRZYKŁAD 9 Przedsiębiorstwo kupuje pewne części od trzech firm i rejestruje, ile części jest wadliwych. Wyniki podsumowuje tablica 9.2. Tak więc 30 procent części kupowanych jest w firmie C i 2 procent z nich jest wadliwych. W języku rachunku prawdopodobieństwa, jeśli zakupiona część jest wybierana losowo, to

$$P(A) = 0,50, \quad P(B) = 0,20, \quad P(C) = 0,30,$$

gdzie A oznacza zdarzenie „część pochodzi z firmy A" itd. Jeśli D oznacza zdarzenie „część jest wadliwa", to mamy też

$$P(D|A) = 0,01, \quad P(D|B) = 0,04, \quad P(D|C) = 0,02.$$

Tablica 9.2

Firma	A	B	C
Części kupione	0,50	0,20	0,30
Części wadliwe	0,01	0,04	0,02

(a) Prawdopodobieństwo, że dana część była kupiona w firmie A i jest wadliwa, wynosi

$$P(A \cap D) = P(A) \cdot P(D|A) = (0,50) \cdot (0,01) = 0,005.$$

Analogicznie,

$$P(B \cap D) = P(B) \cdot P(D|B) = (0,20) \cdot (0,04) = 0,008$$

oraz

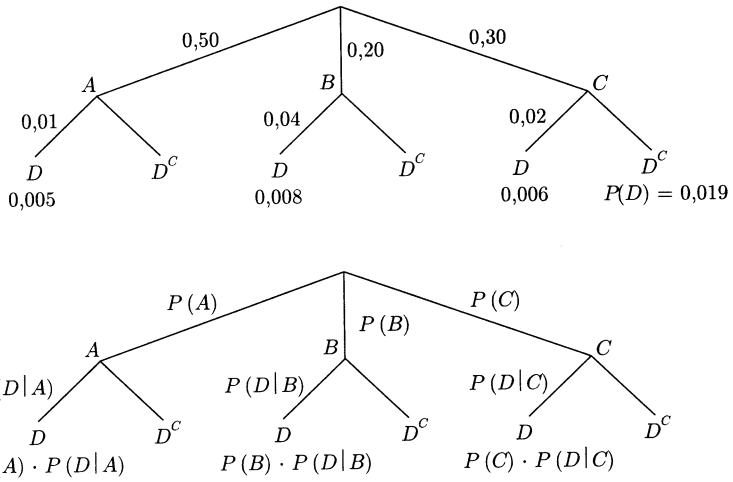
$$P(C \cap D) = P(C) \cdot P(D|C) = (0,30) \cdot (0,02) = 0,006.$$

(b) Ile wynosi prawdopodobieństwo, że losowo wybrana część jest wadliwa? Ponieważ zdarzenie D jest sumą rozłączną zdarzeń

$A \cap D$, $B \cap D$ i $C \cap D$, to odpowiedź brzmi

$$P(D) = P(A) \cdot P(D|A) + P(B) \cdot P(D|B) + P(C) \cdot P(D|C)$$
$$= 0{,}005 + 0{,}008 + 0{,}006 = 0{,}019.$$

Drzewo przedstawione na rysunku 9.2 może pomóc w zorientowaniu się, o co tu chodzi. ■



Rysunek 9.2

Obliczenie wartości $P(D)$ w przykładzie 9(b) stanowi ilustrację następującej obserwacji; w cytowanym przykładzie blokami odpowiedniego podziału są zdarzenia A, B, C.

Wzór na prawdopodobieństwo całkowite

Jeśli zdarzenia $A_1, A_2, \ldots, A_n$ tworzą podział przestrzeni zdarzeń elementarnych Ω oraz $P(A_i) > 0$ dla każdego i, to dla dowolnego zdarzenia B mamy

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \ldots + P(A_k) \cdot P(B|A_k)$$
$$= \sum_{i=1}^{k} P(A_i) \cdot P(B|A_i).$$

Czasami wiemy, że zaszło zdarzenie B, a chcemy znać prawdopodobieństwa poszczególnych zdarzeń A_j. To znaczy, chcemy znaleźć liczby $P(A_1|B), \ldots, P(A_k|B)$ i nie jest od początku jasne,

ile one wynoszą. Ponieważ

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j) \cdot P(B|A_j)}{P(B)},$$

to ze wzoru na prawdopodobieństwo całkowite wynika następujący fakt.

Wzór Bayesa

Przypuśćmy, że zdarzenia $A_1, \ldots, A_k$ tworzą podział przestrzeni Ω, oraz $P(A_i) > 0$ dla każdego i. Przypuśćmy też, że B jest dowolnym zdarzeniem takim, że $P(B) > 0$. Wtedy dla każdego j mamy

$$P(A_j|B) = \frac{P(A_j) \cdot P(B|A_j)}{P(B)},$$

gdzie

$$P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + \ldots + P(A_k) \cdot P(B|A_k).$$

PRZYKŁAD 10 Przypuśćmy, że w przykładzie 9 losowo wybrana część okazała się wadliwa. Jakie jest prawdopodobieństwo, że pochodzi ona z firmy A? Chcemy znać liczbę $P(A|D)$ i wzór Bayesa daje

$$P(A|D) = \frac{P(A) \cdot P(D|A)}{P(D)} = \frac{(0{,}50) \cdot (0{,}01)}{0{,}019} = \frac{0{,}005}{0{,}019} \approx 0{,}263.$$

Mimo, że połowa części kupowana jest od firmy A, tylko około jedna czwarta wadliwych części pochodzi z tej firmy. Podobnie,

$$P(B|D) = \frac{0{,}08}{0{,}019} \approx 0{,}421 \quad \text{oraz} \quad P(C|D) = \frac{0{,}006}{0{,}019} \approx 0{,}316.$$

Zauważ, że ułamki te można odczytać z rysunku 9.2. ∎

Wzór $P(A \cap B) = P(A) \cdot P(B|A)$ ma następujące uogólnienie:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

W ćwiczeniu 26 należy przedstawić dowód tego faktu.

PRZYKŁAD 11 Wyciągamy z talii cztery karty bez zwracania. Jakie jest prawdopodobieństwo, że pierwsze dwie z nich są asami, a następne dwie są królami? Używając sugestywnych oznaczeń, rozwiązanie

możemy zapisać w sposób następujący:

$$\begin{aligned} P(A_1 \cap A_2 \cap K_3 \cap K_4) &= P(A_1) \cdot P(A_2|A_1) \cdot P(K_3|A_1 \cap A_2) \\ &\quad \times P(K_4|A_1 \cap A_2 \cap K_3) \\ &= \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{4}{50} \cdot \frac{3}{49} \approx 0{,}000022. \end{aligned}$$ ■

ĆWICZENIA DO § 9.1

Ilekroć pojawia się prawdopodobieństwo warunkowe, takie jak $P(A|B)$, tylekroć zakładamy, że $P(B) > 0$.

1. W pudełku znajdują się trzy kulki białe i osiem kulek czarnych. Losowo wyciągamy dwie kulki bez zwracania. Znajdź prawdopodobieństwo tego, że
(a) obie kulki są białe,
(b) obie są czarne,
(c) jedna z nich jest czarna, a druga biała.

2. Powtórz ćwiczenie 1 przy założeniu, że pierwsza kulka po wyciągnieciu i obejrzeniu jest odkładana do pudełka, po czym dopiero wyciągamy drugą kulkę (tzn. jest to losowanie ze zwracaniem). Porównaj odpowiedzi z tymi, które uzyskane zostały w ćwiczeniu 1.

3. (a) Wykaż, że zdarzenia B_0, R_0 i E z przykładu 5 są parami niezależne.
(b) Znajdź $P(B_0|E \cap R_0)$ i porównaj z $P(B_0)$.

4. Rzucamy dwiema kostkami, białą i szarą. Rozważ zdarzenia

S, polegające na tym, że „suma oczek na obu kostkach jest ≥ 8",
L, polegające na tym, że „liczba oczek na kostce białej jest mniejsza niż liczba oczek na kostce szarej",
E, polegające na tym, że „liczby oczek na obu kostkach są równe",
G, polegające na tym, że „liczba oczek na kostce białej jest większa od liczby oczek na kostce szarej".

Które z następujących par składają się ze zdarzeń niezależnych? $\{S, L\}$, $\{S, E\}$, $\{L, E\}$, $\{L, G\}$. Nie wykonuj obliczeń, chyba że jest to niezbędne (ale zob. ćwiczenie 5).

5. Obliczając odpowiednie prawdopodobieństwa stwierdź, które pary z ćwiczenia 4 składają się ze zdarzeń niezależnych.

6. Ponownie rzucamy dwiema kostkami z ćwiczenia 4.
(a) Znajdź prawdopodobieństwo, że liczba oczek na szarej kostce jest ≥ 5 pod warunkiem, że suma wartości na obu kostkach wynosi 9.
(b) Znajdź prawdopodobieństwo tego samego zdarzenia pod warunkiem, że powyższa suma jest ≥ 9.

7. Rzucamy cztery razy symetryczną monetą. Niech A oznacza zdarzenie „w pierwszych dwóch rzutach dokładnie raz wypadł orzeł", B oznacza „w czterech rzutach orzeł wypadł dokładnie dwa razy". Czy zdarzenia A i B są niezależne? Uzasadnij swą odpowiedź.

8. Rzucamy cztery razy monetą symetryczną.
(a) Jakie jest prawdopodobieństwo, że orzeł wypadnie (co najmniej) dwa razy z rzędu?
(b) Jakie jest prawdopodobieństwo wypadnięcia orła dwa razy z rzędu pod warunkiem, że orzeł wypadł w co najmniej dwóch rzutach?

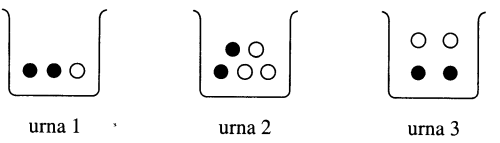
9. Przypuśćmy, że pewien eksperyment prowadzi do zdarzeń A, B i C o prawdopodobieństwach $P(A) = 0{,}3$, $P(B) = 0{,}4$, $P(A \cap B) = 0{,}1$ i $P(C) = 0{,}8$.
(a) Znajdź $P(A|B)$.
(b) Znajdź $P(A^c)$.
(c) Czy zdarzenia A i B są niezależne? Odpowiedź uzasadnij.
(d) Czy zdarzenia A^c i B są niezależne?

10. Mając dane niezależne zdarzenia A i B o prawdopodobieństwach $P(A) = 0{,}4$ i $P(B) = 0{,}6$, znajdź
(a) $P(A|B)$,
(b) $P(A \cup B)$,
(c) $P(A^c \cap B)$.

11. Wyciągamy trzy karty bez zwracania ze zwykłej talii 52 kart. Określ prawdopodobieństwo tego, że
(a) zostają wyciągnięte trzy asy,
(b) wyciągamy kolejno asa, króla i królową, w tym właśnie porządku,
(c) wyciągamy co najmniej jednego asa.

12. Przypomnij sobie, że prawdopodobieństwo tego, że układ kart w pokerze jest kolorem, wynosi około 0,00197 (przykład 1, § 5.2). Jakie jest prawdopodobieństwo, że figura pokerowa jest kolorem pod warunkiem, że każda z należących do niej pięciu kart jest czerwona?

13. W trzech urnach znajdują się szklane kulki, jak pokazuje rysunek 9.3. Wybieramy losowo jedną z tych urn, a następnie wybieramy z niej losowo kulkę.



Rysunek 9.3

(a) Jakie jest prawdopodobieństwo, że wyciągnięta kulka jest czarna?
(b) Jeśli wiadomo, że wyciągnięta kulka jest czarna, to jakie jest prawdopodobieństwo, że pochodzi ona
z urny 1?
z urny 2?
z urny 3?
(c) Jakie jest prawdopodobieństwo, że wyciągnięta kulka jest czarna i pochodzi z urny 1?

14. Losowo wybieramy jedną z urn z rysunku 9.3, a następnie losowo wybieramy z niej dwie kulki (bez zwracania). Znajdź prawdopodobieństwo, że
(a) obie kulki są białe,
(b) obie kulki są czarne.

15. (a) Jeśli wiadomo, że obie kulki wylosowane w ćwiczeniu 14 są czarne, to jakie jest prawdopodobieństwo, że wybrana była urna 1?
(b) Odpowiedz na to samo pytanie, jeśli obie kulki są białe.

16. Z każdej z urn z rysunku 9.3 wybieramy losowo po jednej kulce. Jakie jest prawdopodobieństwo, że:
wszystkie trzy kulki są białe?
wszystkie trzy są czarne?

17. Wszystkie zamówienia w barze Burger Queen przyjmowane są przez Helenę, Franciszka i Grzegorza. Tablica 9.3 pokazuje, jaką część wszystkich zamówień każda z zatrudnionych osób przyjmuje, oraz dla każdej z nich podaje, jaka część zrealizowanych przez nią zamówień powoduje zażalenia.

Tablica 9.3

Pracownicy	Helena	Franciszek	Grzegorz
Zamówienia	0,25	0,35	0,40
Zażalenia	0,04	0,06	0,03

(a) Jakie jest prawdopodobieństwo, że losowo wybrane zamówienie spowodowało zażalenie?
(b) Jeśli dane zamówienie spowodowało zażalenie, to ile wynosi prawdopodobieństwo, że zostało ono zrealizowane przez
Helenę?
Franciszka?
Grzegorza?

18. W pudełku znajdują się dwie monety, jedna symetryczna i druga mająca po obu stronach orły. Rzucamy losowo wybraną monetą.
(a) Jakie jest prawdopodobieństwo, że wypadnie orzeł?
(b) Jeśli wypadł orzeł, to jakie jest prawdopodobieństwo, że rzucona moneta ma dwa orły?

(c) Jeśli wybraną monetą rzucimy dwa razy, to jakie jest prawdopodobieństwo, że za każdym razem wypadnie orzeł?

19. Przeprowadzając testy mające wykryć pewną rzadką chorobę dokonano następujących obserwacji. Niech D oznacza zdarzenie „badana osoba ma tę chorobę", N oznacza zdarzenie „wynik testu u badanej osoby jest ujemny" i N^c będzie zdarzeniem „wynik testu u badanej osoby jest dodatni". Wówczas $P(N^c \cap D) = 0{,}004$, $P(N \cap D) = 0{,}0001$, $P(N^c) = 0{,}044$. Udowodnij, że następujące zdania są prawdziwe.

(a) Test daje wynik zgodny z prawdą u ponad 97,5% tych badanych osób, które mają daną chorobę. *Wskazówka*: znajdź $P(N^c|D)$.

(b) Test daje wynik zgodny z prawdą u prawie 96% ogólnej populacji badanych osób. *Wskazówka*: znajdź $P((N^c \cap D) \cup (N \cap D^c))$.

(c) Jednakże test wprowadza w błąd w ponad 90% przypadków, w których daje wynik dodatni! *Wskazówka*: znajdź $P(D|N^c)$.

Morał: trzeba być bardzo ostrożnym interpretując testy na rzadkie choroby.

20. Wykaż, że jeśli zdarzenia A i B są niezależne, to niezależne są też zdarzenia należące do par $\{A, B^c\}$, $\{A^c, B\}$ i $\{A^c, B^c\}$.

21. Pewne urządzenie elektroniczne składa się z n części. Prawdopodobieństwo zepsucia się przed upływem okresu gwarancji dla każdej z tych części wynosi q. (Możesz też myśleć o żarówkach w swoim mieszkaniu).

(a) Jakie jest prawdopodobieństwo, że jedna z części zepsuje się przed upływem okresu gwarancji? Jakie przyjmujesz założenia odpowiadając na to pytanie?

(b) Jak brzmi odpowiedź, jeśli $n = 100$ i $q = 0{,}01$?

(c) Jak brzmi odpowiedź, jeśli $n = 100$ i $q = 0{,}001$?

(d) Jak brzmi odpowiedź, jeśli $n = 100$ i $q = 0{,}1$?

22. Udowodnij, że jeśli $P(A|B) > P(A)$, to $P(B|A) > P(B)$.

23. W pudełku jest 20 sztuk towaru. Kontrolę jego jakości przeprowadza się sprawdzając 5 sztuk, losowo wybranych z pudełka. Jeśli żadna z nich nie jest wadliwa, to akceptuje się całą zawartość pudełka.

(a) Jakie jest prawdopodobieństwo, że zostanie zaakceptowane pudełko, w którym są dwie wadliwe sztuki towaru?

(b) Powtórz część dla przypadku, gdy sprawdza się nie 5, lecz 10 sztuk towaru.

24. Udowodnij, że jeśli $P(A) = P(B) = \frac{2}{3}$, to $P(A|B) \geq \frac{1}{2}$.

25. Przypuśćmy, że zdarzenia A, B i C są wzajemnie niezależne. Czy zdarzenia $A \cap B$ i $A \cap C$ muszą być niezależne?

26. Udowodnij, że

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) =$$
$$= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap A_2 \cap \ldots \cap A_{n-1}).$$

Jeśli krok indukcyjny okaże się zbyt skomplikowany, to udowodnij ten rezultat jedynie dla $n = 3$ i $n = 4$.

27. Dla każdego z następujących zdań udowodnij je lub wykaż, że jest ono fałszywe.
(a) Jeśli zdarzenia A i B są niezależne i zdarzenia B i C są niezależne, to zdarzenia A i C są niezależne.
(b) Każde zdarzenie A jest niezależne samo od siebie.
(c) Jeśli zdarzenia A i B są rozłączne, to są niezależne.

28. Jest oczywiście nieprawdą, że $(b \longrightarrow a) \wedge a \Rightarrow b$. (Rozważ przypadek, kiedy a jest fałszem, a b prawdą). Z drugiej strony, jeśli zdania $b \longrightarrow a$ oraz a są prawdziwe, to prawdziwość zdania b staje się bardziej prawdopodobna w następującym sensie.
(a) Wykaż, że jeśli $P(A|B) = 1$ i $P(A) < 1$, to $P(B|A) > P(B)$.
(b) Co będzie, gdy $P(A) = 1$?

29. Niech S będzie zdarzeniem w przestrzeni Ω oraz $P(S) > 0$.
(a) Wykaż, że jeśli $P^*(E) = P(E|S)$ dla $E \subseteq \Omega$, to funkcja P^* zachowuje ilorazy w następujący sposób:

$$\frac{P^*(E)}{P^*(F)} = \frac{P(E \cap S)}{P(F \cap S)} \quad \text{ilekroć} \quad P(F \cap S) \neq 0.$$

(b) Wykaż, że jeśli P^* jest funkcją prawdopodobieństwa w przestrzeni Ω, zachowującą ilorazy tak, jak w części (a), to $P^*(E) = P(E|S)$ dla każdego $E \subseteq \Omega$.

§ 9.2. Zmienne losowe

W tym paragrafie zmieniamy trochę nasz punkt widzenia. Zamiast koncentrować uwagę na zdarzeniach, tzn. podzbiorach przestrzeni Ω, przyjrzymy się funkcjom o wartościach liczbowych, zdefiniowanym na zbiorze Ω. Ponieważ każde zdarzenie E odpowiada swojej funkcji charakterystycznej χ_E: $\Omega \longrightarrow \{0, 1\}$, więc nic na takiej zmianie nie tracimy, zyskamy natomiast bardzo wiele.

Funkcję ze zbioru Ω w zbiór $\mathbb{R}$ nazywamy **zmienną losową**. Historycznie terminologia ta bierze się z tego, że wartości takiej funkcji zmieniają się, gdy przechodzimy od jednego losowego wyniku ω w przestrzeni Ω do drugiego. Tradycyjnie zmienne losowe oznacza się wielkimi literami z końca alfabetu takimi, jak X, Y, Z.

Jeśli dana jest zmienna losowa X na przestrzeni Ω i pewien warunek C, który może być spełniony przez niektóre wartości tej zmiennej, to używamy zapisu $\{X \text{ spełnia } C\}$ dla skrótowego oznaczenia zdarzenia $\{\omega \in \Omega\colon X(\omega) \textbf{ spełnia } C\}$. Na przykład:

$$\{5 < X \leq 7\} = \{\omega \in \Omega\colon 5 < X(\omega) \leq 7\},$$
$$\{X = a \text{ lub } X^2 = b\} = \{\omega \in \Omega\colon X(\omega) = a \text{ lub } X(\omega)^2 = b\},$$
$$\{|X - 4| < 3\} = \{\omega \in \Omega\colon |X(\omega) - 4| < 3\} \text{ itd.}$$

Dla dowolnego zbioru $A \subseteq \mathbb{R}$, zapis $\{X \in A\}$ oznacza zdarzenie $\{\omega \in \Omega\colon X(\omega) \in A\}$ i będziemy go często używać. Zapisując prawdopodobieństwa tego rodzaju zdarzeń, będziemy na ogół opuszczali nawiasy klamrowe. Tak na przykład,

$$P(X = 2) = P(\{X = 2\}) = P(\{\omega \in \Omega\colon X(\omega) = 2\}),$$

podczas gdy

$$P(X \in A) = P(\{\omega \in \Omega\colon X(\omega) \in A\}) \quad \text{dla} \quad A \subseteq \mathbb{R}.$$

Zbiór wartości przyjmowanych przez zmienną losową X, tzn. zbiór $X(\Omega) = \{X(\omega)\colon \omega \in \Omega\}$ będzie nazywać **zbiorem wartości zmiennej** X. Jeśli Ω jest zbiorem skończonym, to zbiór wartości zmiennej X jest skończonym podzbiorem zbioru $\mathbb{R}$.

PRZYKŁAD 1 Niech zbiór Ω składa się z 36 jednakowo prawdopodobnych wyników rzutu dwiema symetrycznymi kostkami.

(a) Najbardziej interesująca zmienna losowa na przestrzeni Ω podaje sumę wartości oczek na obu kostkach, to znaczy

$$X_S(k,l) = k + l \quad \text{dla} \quad (k,l) \in \Omega.$$

Zbiorem wartości zmiennej X_S jest $\{2,3,4,5,6,7,8,9,10,11,12\}$. Wszystkie liczby $P(X_S \in A)$ mogą być wyznaczone na podstawie tablicy 9.4(a), która zawiera te same informacje, co tablica 5.2 z § 5.2. Na przykład,

$$P(X_S \leq 5) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} = \frac{10}{36}$$

oraz

$$P(4 < X_S < 10) = \frac{4}{36} + \frac{5}{36} + \frac{6}{36} + \frac{5}{36} + \frac{4}{36} = \frac{2}{3}.$$

Tablica 9.4(a)

k	2	3	4	5	6	7	8	9	10	11	12
$P(X_S = k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Tablica 9.4(b)

k	1	2	3	4	5	6
$P(X_B = k) = P(X_R = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Jeśli E jest zbiorem wszystkich parzystych liczb całkowitych, to

$$\begin{aligned} P(\text{liczba } X_S \text{ jest parzysta}) &= P(X_S \in E) \\ &= P(X_S = 2) + P(X_S = 4) + \ldots + P(X_S = 12) \\ &= \frac{1}{36}(1+3+5+5+3+1) = \frac{1}{2}. \end{aligned}$$

Z punktu widzenia rachunku prawdopodobieństwa, wszystko co chcemy wiedzieć o funkcji X_S, można odczytać z tablicy 9.4(a).

(b) Dla każdej pary (k,l) definiujemy $X_B(k,l) = k$ oraz $X_R(k,l) = l$. Zatem zmienna X_B podaje liczbę oczek na białej kostce, a zmienna X_R podaje liczbę oczek na kostce szarej. Zbiorem wartości obu zmiennych losowych X_B i X_R jest $\{1,2,3,4,5,6\}$. Przyjęcie dowolnej wartości z tego zbioru przez którąkolwiek z rozpatrywanych zmiennych losowych jest jednakowo prawdopodobne, zob. tablica 9.4(b). Te zmienne losowe są nieodróżnialne w tym sensie, że przyjmują te same wartości z tymi samymi prawdopodobieństwami.

(c) Zmienna losowa X_S jest sumą zmiennych losowych X_B i X_R. Zdarzenia rozważane w przykładzie 1 z § 9.1 można opisać za pomocą tych zmiennych:

$$\begin{aligned} B &= \text{„liczba oczek na kostce białej jest } \leq 3\text{”} = \{X_B \leq 3\}, \\ R &= \text{„liczba oczek na kostce szarej jest } \geq 5\text{”} = \{X_R \geq 5\}, \\ S &= \text{„suma liczb oczek na kostkach jest } \geq 8\text{”} = \{X_S \geq 8\}. \end{aligned}$$

Wyniki obliczeń z przykładów 1 i 2 z § 9.1 można teraz zapisać w następujący sposób

$$\begin{aligned} P(X_R \geq 5) &= \frac{1}{3}, \quad P(X_R \geq 5 | X_S \geq 8) = \frac{9}{15}, \\ P(X_B \leq 3) &= \frac{1}{2}, \quad P(X_B \leq 3 | X_S \geq 8) = \frac{3}{15}, \\ P(X_B \leq 3 | X_R \geq 5) &= \frac{1}{2} = P(X_B \leq 3), \\ P(X_R \geq 5 | X_B \leq 3) &= \frac{1}{3} = P(X_R \geq 5). \end{aligned}$$

Z dwóch ostatnich równości wynika, że zdarzenia $\{X_R \geq 5\}$ i $\{X_B \leq 3\}$ są niezależne. Ten fakt nie jest zaskakujący i nie jest związany ani ze szczególnym wyborem wartości 5 i 3, ani z sensem nierówności $\geq$ oraz $\leq$. Spodziewamy się, że zmienne losowe X_R i X_B są niezależne, tzn. że informacje odnoszące się do X_R nie

mają wpływu na prawdopodobieństwa zdarzeń związanych z X_B i odwrotnie. Podamy teraz formalną definicję tego pojęcia. ∎

Chciałoby się powiedzieć, że zmienne losowe X i Y, określone na przestrzeni zdarzeń elementarnych Ω, są niezależne, jeśli zdarzenia $\{X \in A\}$ i $\{X \in B\}$ są niezależne dla dowolnego wyboru podzbiorów A i B zbioru $\mathbb{R}$. Wiedza o tym, że $X(\omega)$ należy do pewnego zbioru A, nie powinna mieć wpływu na prawdopodobieństwo tego, że $Y(\omega)$ należy do pewnego zbioru B. Zatem zmienne X i Y powinny być niezależne, jeśli

$$(\mathrm{I}_1) \qquad P(\{X \in A\} \cap \{Y \in B\}) = P(X \in A) \cdot P(Y \in B)$$
$$\text{dla wszystkich} \quad A, B \subseteq \mathbb{R}.$$

Lewą stronę powyższej równości będziemy na ogół zapisywać w skrócie jako $P(X \in A \text{ i } X \in B)$. Niestety w sytuacji najbardziej ogólnej, kiedy zbiór Ω może być nieskończony, w sformułowaniu warunku (I_1) nie wszystkie podzbiory A i B zbioru $\mathbb{R}$ mogą zostać użyte, chociaż wolno nam użyć wszystkich takich, które Czytelnik potrafi sobie wyobrazić, jak na przykład przedziały. Ponieważ nie chcemy tu wnikać w tę techniczną kwestię, zastąpimy (I_1) wymaganiem, które staje się równoważne z warunkiem (I_1), jeśli tylko wspomniana kwestia techniczna zostanie odpowiednio wyjaśniona. Zatem dwie zmienne losowe X i Y na Ω są **niezależne**, jeśli

$$(\mathrm{I}_2) \qquad P(X \in I \text{ i } Y \in J) = P(X \in I) \cdot P(Y \in J)$$

dla wszystkich przedziałów I i J w $\mathbb{R}$. Jeśli zbiór Ω jest skończony, to warunek (I_2) jest równoważny z warunkiem

$$(\mathrm{I}_3) \qquad P(X = x \text{ i } Y = y) = P(X = x) \cdot P(Y = y)$$
$$\text{dla wszystkich} \quad x, y \in \mathbb{R};$$

zob. ćwiczenie 19. Oczywiście wystarczy sprawdzać warunek (I_3) dla liczb x i y należących, odpowiednio, do zbiorów wartości zmiennych X i Y.

PRZYKŁAD 2 (a) Zmienne losowe X_B i X_R, podające liczby oczek, odpowiednio, na białej i szarej kostce, są niezależne. Istotnie,

$$P(X_B = k \text{ i } X_R = l) = \frac{1}{36} = P(X_B = k) \cdot P(X_R = l)$$
$$\text{dla wszystkich } (k, l).$$

(b) Niech zmienna losowa X_S podaje sumę liczb oczek na obu kostkach. Wtedy zmienne X_S i X_B nie są niezależne; jeśli

wartość X_B jest duża, to jest bardziej prawdopodobne, że wartość X_S będzie duża. To odwołujące się do intuicji uzasadnienie *nie jest dowodem* i musi być poparte matematycznym argumentem. Zamiast bezmyślnie sprawdzać warunek (I_3), posłużymy się powyższą uwagą jak wskazówką. Jeśli wartość X_B jest duża i wynosi na przykład 6, to suma X_S z pewnością nie może być mała; w istocie nie może ona być mniejsza niż 7. A więc, na przykład,

$$P(X_B = 6 \text{ i } X_S = 2) = 0 \neq P(X_B = 6) \cdot P(X_S = 2).$$

To wystarczy do dowodu, że zmienne X_B i X_S nie są niezależne.■

Zmienne losowe ciągu $X_1, X_2, \ldots, X_n$, określone na przestrzeni zdarzeń elementarnych Ω, są **niezależne**, jeśli

$$(I_2') \qquad P(X_i \in J_i \quad \text{dla} \quad i = 1, 2, \ldots, n) = \prod_{i=1}^{n} P(X_i \in J_i)$$

dla wszystkich przedziałów $J_1, J_2, \ldots, J_n$ w $\mathbb{R}$. Jeśli zbiór Ω jest skończony, to warunek ten jest równoważny z warunkiem

$$(I_3') \qquad P(X_i = x_i \quad \text{dla} \quad i = 1, 2, \ldots, n) = \prod_{i=1}^{n} P(X_i = x_i)$$

dla $x_1, x_2, \ldots, x_n \in \mathbb{R}$.

PRZYKŁAD 3 (a) Rzucamy n-krotnie symetryczną monetą, tak jak w przykładzie 6 z § 9.1 (i przykładzie 5 z § 5.2). Dla $i = 1, 2, \ldots, n$, niech $X_i = 1$, jeśli w i-tym rzucie wypadł orzeł, i $X_i = 0$ w przeciwnym razie. Zatem X_i jest funkcją charakterystyczną zdarzenia E_i oznaczającego „orzeł w i-tym rzucie". W przykładzie 6 z § 9.1 założyliśmy, że dla każdego ciągu $x_1, x_2, \ldots, x_n$, złożonego z wyrazów 0 i 1, zdarzenia $\{X_i = x_i\}$ są wzajemnie niezależne, a więc

$$P(X_i = x_i \quad \text{dla} \quad i = 1, 2, \ldots, n) = \prod_{i=1}^{n} P(X_i = x_i) = \frac{1}{2^n}.$$

Innymi słowy założyliśmy, że zmienne losowe $X_1, X_2, \ldots, X_n$ są niezależne.

(b) Suma $S_n = X_1 + X_2 + \ldots + X_n$ jest bardzo przydatną zmienną losową. Ponieważ każdy element rozpatrywanej przestrzeni zdarzeń elementarnych jest ciągiem długości n, złożonym z wyrazów O i R, oraz $X_i = 1$, jeśli i-tym wyrazem tego ciągu jest O, a w przeciwnym przypadku $X_i = 0$, więc zmienna S_n podaje liczbę wyrazów O w całym ciągu. Innymi słowy, zmienna S_n

podaje liczbę orłów w n rzutach. Zbiorem wartości zmiennej S_n jest $\{0, 1, 2, \ldots, n\}$. Jak wyjaśniliśmy w przykładzie 5(a) z § 5.2

$$P(S_n = k) = \frac{1}{2^n} \cdot \binom{n}{k} \quad \text{dla} \quad k \in \{0, 1, 2, \ldots, n\}. \qquad \blacksquare$$

Zmienną losową nazywamy **dyskretną**, jeśli jej zbiór wartości jest skończony lub składa się z wyrazów jakiegoś nieskończonego ciągu. Wszystkie zmienne losowe na skończonej przestrzeni zdarzeń elementarnych Ω są dyskretne. Jeśli X jest dyskretną zmienną losową, to wszystkie informacje o niej uzyskać można znając liczby $P(X = x)$. Funkcję f_X, zdefiniowaną na zbiorze $\mathbb{R}$ wzorem

$$f_X(x) = f(x) = P(X = x) \quad \text{dla} \quad x \in \mathbb{R},$$

nazywamy **rozkładem prawdopodobieństwa zmiennej losowej** X. Mówi nam ona, jak wartości zmiennej X są rozłożone w zbiorze $\mathbb{R}$ i podaje nam ich prawdopodobieństwa.

PRZYKŁAD 4 (a) Niech X_S, X_B i X_R będą zmiennymi losowymi, które rozpatrywaliśmy w związku z rzutami dwiema monetami symetrycznymi. Ich rozkłady prawdopodobieństwa można odczytać z tablicy 9.4. Zmienne X_B i X_R mają ten sam rozkład prawdopodobieństwa, a mianowicie

$$f(x) = \begin{cases} 1/6, & \text{jeśli } x = 1, 2, 3, 4, 5 \text{ lub } 6, \\ 0 & \text{dla innych wartości } x. \end{cases}$$

Rozkład prawdopodobieństwa f_S zmiennej X_S jest trudniejszy do zapisania, ale z tablicy 9.4(a) jasno wynika, że

$$f_S(x) = \begin{cases} 1/36, & \text{jeśli } x = 2 \text{ lub } 12, \\ 2/36, & \text{jeśli } x = 3 \text{ lub } 11, \\ \text{itd.} & \\ 0, & \text{jeśli } x \notin \{2, 3, \ldots, 12\}. \end{cases}$$

(b) Niech $X_1, X_2, \ldots, X_n$ i S_n będą zmiennymi losowymi zliczającymi orły w przykładzie 3. Wszystkie zmienne X_i mają jednakowy rozkład prawdopodobieństwa f, gdzie $f(0) = f(1) = \frac{1}{2}$ i $f(x) = 0$ dla wszystkich innych wartości x z $\mathbb{R}$. Rozkład prawdopodobieństwa zmiennej S_n jest dany wzorem $f(k) = \frac{1}{2^n} \cdot \binom{n}{k}$ dla $k = 0, 1, \ldots, n$ i $f(x) = 0$ dla pozostałych wartości x. Rozkład ten nazywamy rozkładem dwumianowym z parametrem $\frac{1}{2}$, choć właściwie należałoby nazwać go jednym z **rozkładów dwumianowych**, gdyż rozkładów takich jest nieskończenie wiele, po

jednym dla każdego $n \geq 1$. O bardziej ogólnym rozkładzie dwumianowym mowa będzie w § 9.4, gdzie moneta nie będzie już musiała być symetryczna. ∎

Nieskończone przestrzenie zdarzeń elementarnych wprowadzone zostały w przykładzie 6 z § 5.2 oraz bezpośrednio po nim. Zmienne losowe na nieskończonych przestrzeniach zdarzeń elementarnych mogą, ale nie muszą być dyskretne.

PRZYKŁAD 5 (a) W przykładzie 6(a) z § 5.2 rzucamy symetryczną monetą aż do momentu, gdy wypadnie orzeł; niech W będzie „czasem oczekiwania" na pojawienie się pierwszego orła. Wówczas

$$P(W = k) = \frac{1}{2^k} \quad \text{dla} \quad k = 1, 2, \ldots$$

Zbiór wartości zmiennej W jest zbiorem wyrazów pewnego nieskończonego ciągu, a więc W jest dyskretną zmienną losową. Jej rozkładem prawdopodobieństwa jest funkcja f, gdzie $f(k) = 1/2^k$ dla $k \in \mathbb{P}$ i $f(x) = 0$ dla $x \in \mathbb{R} \backslash \mathbb{P}$.

(b) Tak jak w przykładzie 6(b) z § 5.2, losowo wybieramy liczbę ze zbioru $\Omega = [0, 1)$. Niech U oznacza wybraną wartość, tzn. niech $U(x) = x$ dla $x \in [0, 1)$. Taką funkcję U nazywa się czasem **jednostajną zmienną losową**; przedziały tej samej długości traktujemy w ten sam „jednostajny" sposób, ponieważ $P(U \in [a, b)) = b - a$, gdzie $[a, b) \subseteq [0, 1)$. Zbiorem wartości zmiennej U jest cały przedział [0,1), którego elementy nie mogą być ustawione w ciąg (§ 13.3). Zatem U nie jest dyskretną zmienną losową. Ponadto, jeśli spróbujemy określić jej „rozkład prawdopodobieństwa" wzorem

$$f(x) = P(U = x) = 0 \quad \text{dla} \quad x \in \mathbb{R},$$

to otrzymany obiekt będzie bezużyteczny. Za pomocą tej funkcji f z pewnością nie uda nam się uzyskać informacji na temat zmiennej U. Przykład ten pokazuje, dlaczego zdefiniowaliśmy rozkład prawdopodobieństwa jedynie dla dyskretnych zmiennych losowych. ∎

Kolejny obiekt, który teraz zdefiniujemy, jest jeszcze ważniejszy niż rozkład zmiennej losowej. Dla dowolnej zmiennej losowej X określonej na dowolnej przestrzeni zdarzeń elementarnych, **dystrybuantą tej zmiennej** nazywamy funkcję F_X, zdefiniowaną na zbiorze $\mathbb{R}$ wzorem

$$F_X(y) = F(y) = P(X \leq y) \quad \text{dla} \quad y \in \mathbb{R}.$$

„Akumuluje" ona wartości rozkładu prawdopodobieństwa zmiennej X dla argumentów $\leq y$. Rzeczywiście, jeśli X jest dyskretną zmienną losową, to jej dystrybuanta jest sumą tych wartości jej rozkładu prawdopodobieństwa:

$$F_X(y) = \sum_{x \leq y} f_X(x).$$

PRZYKŁAD 6 (a) Rozważmy zmienne losowe X_S, X_B i X_R o rozkładach prawdopodobieństwa podanych w przykładzie 4(a) i w tablicy 9.4. Dystrybuantą zmiennych X_B i X_R jest funkcja F dana wzorem

$$F(y) = \begin{cases} 0 & \text{dla } y < 1, \\ 1/6 & \text{dla } 1 \leq y < 2, \\ 2/6 & \text{dla } 2 \leq y < 3, \\ 3/6 & \text{dla } 3 \leq y < 4, \\ 4/6 & \text{dla } 4 \leq y < 5, \\ 5/6 & \text{dla } 5 \leq y < 6, \\ 1 & \text{dla } y \geq 6. \end{cases}$$

Dystrybuanta zmiennej X_S jest dana wzorem

$$F_S(y) = \sum_{x \leq y} f_S(x).$$

Te dystrybuanty przedstawione są na rysunku 9.4.

(b) Rozważmy zmienne losowe $X_1, X_2, \ldots, X_n$ i S_n, zliczające liczbę orłów w rzutach symetryczną monetą. Dystrybuanta F dowolnej zmiennej X_i jest bardzo prosta:

$$F(y) = \begin{cases} 0, & \text{jeśli } y < 0, \\ 1/2, & \text{jeśli } 0 \leq y < 1, \\ 1, & \text{jeśli } y \geq 1. \end{cases}$$

Dystrybuanta zmiennej S_n jest dana wzorem

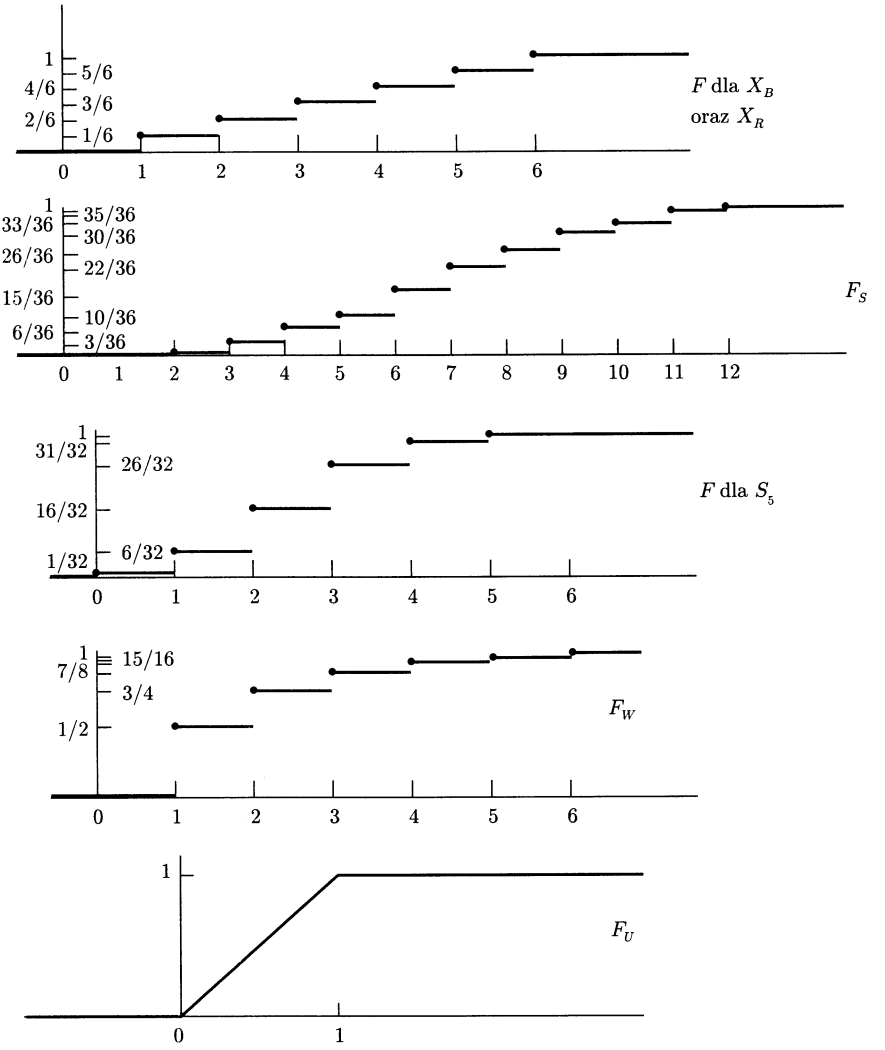
$$F(y) = \sum_{x \leq y} \frac{1}{2^n} \binom{n}{x}.$$

Dla $n = 5$ przedstawia ją rysunek 9.4.

(c) Dystrybuanta F_W zmiennej losowej czasu oczekiwania W z przykładu 5(a) dana jest wzorem

$$F_W(y) = \sum_{k \leq y} \frac{1}{2^k}.$$

Jest ona naszkicowana na rysunku 9.4.



Rysunek 9.4

(d) Dystrybuantę F_U jednostajnej zmiennej losowej z przykładu 5(b) nazywamy **dystrybuantą rozkładu jednostajnego**. Ponieważ $U(x) \in [0,1)$ dla każdego x, to stwierdzamy, że

$$\begin{aligned}
&\text{jeśli } y < 0, && \text{to } F_U(y) = P(U \leq y) = 0,\\
&\text{jeśli } y \geq 1, && \text{to } F_U(y) = P(U \leq y) = P(U \leq 1) = 1,\\
&\text{jeśli } 0 \leq y < 1, && \text{to } F_U(y) = P(U \leq y)\\
& && \qquad = P(0 \leq U \leq y) = P([0,y]) = y.
\end{aligned}$$

To znaczy, że

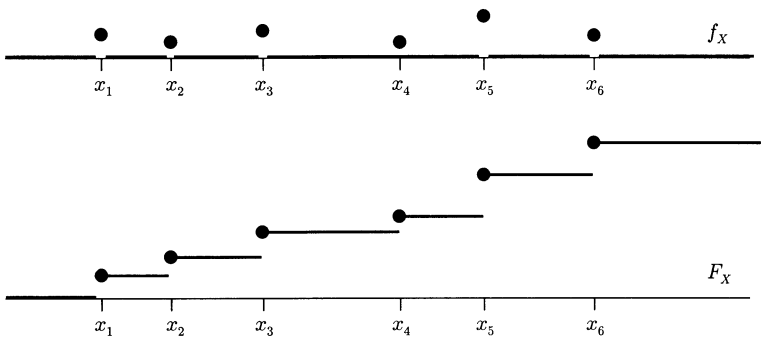
$$F_U(y) = \begin{cases} 0 & \text{dla } y < 0, \\ y & \text{dla } 0 \leq y < 1, \\ 1 & \text{dla } y \geq 1. \end{cases}$$

Zobacz rysunek 9.4. Dystrybuanty dyskretnych zmiennych losowych skokowo rosną w punktach należących do zbiorów wartości tych zmiennych. W przeciwieństwie do nich, dystrybuanta zmiennej F_U jest przyjemną funkcją ciągłą bez „skoków". ∎

Dystrybuantę F_X dyskretnej zmiennej losowej X można zdefiniować za pomocą jej rozkładu prawdopodobieństwa f_X, gdyż $F_X(y) = \sum_{x \leq y} f_X(x)$. Możemy również znając funkcję F_X odtworzyć funkcję f_X. Jeśli zbiór wartości zmiennej X daje się zapisać w postaci $\{x_1, x_2, \ldots\}$, gdzie $x_1 < x_2 < x_3 < \ldots$, jak to ma miejsce we wszystkich naszych przykładach, to mamy

$$\begin{aligned} f_X(x_1) &= F_X(x_1), \\ f_X(x_k) &= F_X(x_k) - F_X(x_{k-1}) \text{ dla } k \geq 2, \\ f_X(x) &= 0 \text{ dla innych } x. \end{aligned}$$

Każda liczba $f_X(x)$ jest wartością, o jaką skokowo zwiększa się $F_X(x)$ w punkcie x. Zobacz rysunek 9.5.



Rysunek 9.5

Podsumujmy. Dla dyskretnej zmiennej losowej X określony jest jej rozkład prawdopodobieństwa f_X i dystrybuanta F_X. Każda z tych funkcji zawiera wszystkie istotne z punktu widzenia rachunku prawdopodobieństwa informacje na temat zmiennej X i każdą z nich można wyznaczyć za pomocą drugiej. Rzeczywiście, jeśli w podręczniku rachunku prawdopodobieństwa lub statystyki znajduje się tablica przedstawiająca pewne ważne rozkłady prawdopodobieństwa, takie jak rozkład dwumianowy, to

w zależności od podręcznika, podane są w niej bądź rozkłady prawdopodobieństw f_X, bądź dystrybuanty F_X. Oczekuje się, że Czytelnik będzie umiał znaleźć jedną z tych funkcji za pomocą drugiej.

My uważamy, że dystrybuanty powinny się znaleźć we wszystkich podręcznikach i dystrybuantę powinno się uważać za pojęcie bardziej fundamentalne. Jeśli chodzi o obliczenia, to łatwiej jest nam obliczać różnice postaci $F_X(x_k) - F_X(x_{k-1})$, niż sumy typu $\sum_{x \leq y} f_X(x)$. To znaczy, że łatwiej jest przejść od f_X do F_X niż otrzymać F_X z f_X. Po drugie, interesujące nas prawdopodobieństwa mają zwykle postać $P(a \leq X \leq b)$, $P(a < X < b)$, $P(X < b)$ itd. Na przykład, jeśli rzucamy 1000 razy monetą i zmienna losowa $X = S_{1000}$ podaje, ile razy wypadł orzeł, to liczby takie, jak $P(470 \leq X \leq 530)$ są znacznie bardziej interesujące od liczb typu $P(X = 501)$. Równości

$$P(470 \leq X \leq 530) = \sum_{x=470}^{530} f_X(x)$$

oraz

$$P(470 \leq X \leq 530) = F_X(530) - F_X(469)$$

pokazują, że dystrybuanty są łatwiejsze w użyciu i bardziej nadają się do tego rodzaju obliczeń. Trzecim powodem, dla którego wolimy dystrybuanty, jest to, że rozkłady prawdopodobieństwa zdefiniowane są jedynie dla dyskretnych zmiennych losowych, podczas gdy dystrybuanty są zdefiniowane (i użyteczne!) dla wszystkich zmiennych. Najważniejszą ze wszystkich dystrybuant jest dystrybuanta rozkładu normalnego, która, podobnie jak dystrybuanta rozkładu jednostajnego, nie jest dystrybuantą dyskretnej zmiennej losowej. Zajmiemy się nią pokrótce w § 9.4. Czytelnikowi, który spodziewa się poświęcić podczas swoich studiów więcej czasu rachunkowi prawdopodobieństwa, radzimy, by przyzwyczaił się do używania dystrybuant.

Czy zauważyłeś, jak w tym paragrafie przestrzenie zdarzeń elementarnych stopniowo znikały z pola widzenia? Są one tam oczywiście obecne, gdyż zmienne losowe określone są na przestrzeniach zdarzeń elementarnych. Ale jednym z wczesnych osiągnięć rachunku prawdopodobieństwa było zdanie sobie sprawy z tego, że praktycznie cała ta teoria związana jest ze zmiennymi losowymi i że dystrybuanty zmiennych losowych zawierają pełną informację o nich. Rzeczywiście, matematycy zajmujący się rachunkiem prawdopodobieństwa uważają dwie zmienne losowe za równoważne, jeśli mają one te same dystrybuanty; mówi się wtedy,

że są to **zmienne losowe o jednakowym rozkładzie**. Przestrzenie, na których są one określone, często bywają nieistotne i nie poświęca się im uwagi.

PRZYKŁAD 7 Rzucamy jeden raz monetą, a więc przestrzenią zdarzeń elementarnych Ω jest zbiór $\{O, R\}$. Jeśli przyjmiemy $X(O) = 1$ i $X(R) = 0$, to $P(X = 0) = P(X = 1) = \frac{1}{2}$. Ta zmienna losowa ma tę samą dystrybuantę F, co każda ze zmiennych $X_1, X_2, \ldots, X_n$ z przykładu 6(b).

Rzucamy dwiema kostkami i definiujemy zmienną losową Y, kładąc 1, jeśli suma wyrzuconych oczek jest parzysta i 0 w przeciwnym razie. To znaczy, że $Y(k, l) = 1$, jeśli liczba $k + l$ jest parzysta i $Y(k, l) = 0$ w przeciwnym razie. Wtedy $P(Y = 0) = P(Y = 1) = \frac{1}{2}$ (zob. przykład 1(a)) i zmienna Y ma tę samą dystrybuantę, co rozpatrywana przed chwilą zmienna X mimo, że odpowiednie przestrzenie zdarzeń elementarnych są całkiem różne.

Niech Ω będzie nieskończoną przestrzenią zdarzeń elementarnych równą przedziałowi [0,1). Dla losowo wybranej liczby x z przedziału [0,1) kładziemy $W(x) = 0$, jeśli x należy do przedziału $[0, \frac{1}{2})$, i $W(x) = 1$, jeśli x należy do przedziału $[\frac{1}{2}, 1)$. Ponownie mamy $P(W = 0) = P(W = 1) = \frac{1}{2}$, a więc zmienna losowa W ma tę samą dystrybuantę, co zmienne X i Y. ∎

Na koniec trzeba przyznać, że dystrybuanty (a także rozkłady prawdopodobieństwa w przypadku dyskretnych zmiennych losowych) nie mówią nam, jakie zachodzą związki między różnymi zmiennymi losowymi określonymi na tej samej przestrzeni zdarzeń elementarnych. Do tego potrzebne są tak zwane łączne rozkłady prawdopodobieństwa zmiennych losowych; zagadnienia te omawiane są w bardziej kompletnych opracowaniach dotyczących rachunku prawdopodobieństwa.

ĆWICZENIA DO § 9.2

1. (a) Rzucamy trzema kostkami symetrycznymi. Jaki jest zbiór wartości zmiennej losowej podającej sumę oczek wyrzuconych na wszystkich kostkach?

(b) Powtórz część (a) dla przypadku n kostek.

2. Rzucamy dwiema symetrycznymi kostkami. Znajdź

(a) prawdopodobieństwo, że suma wyrzuconych oczek jest mniejsza lub równa 7,

(b) $P(5 \leq$ rozpatrywana suma $\leq 10)$,
(c) prawdopodobieństwo, że suma ta jest wielokrotnością 3.

3. Rozważmy następujące dwie zmienne losowe na przestrzeni Ω, złożonej z 36 jednakowo prawdopodobnych wyników rzutu dwiema symetrycznymi kostkami: $D(k,l) = |k-l|$ i $M(k,l) = \max\{k,l\}$.
(a) Znajdź zbiory wartości zmiennych D i M.
(b) Sporządź tablice rozkładów prawdopodobieństwa zmiennych D i M takie jak tablica 9.4. *Wskazówka*: wykorzystaj tablicę 9.1 z § 9.1.
(c) Oblicz $P(D \leq 1)$, $P(M \leq 3)$ i $P(D \leq 1$ i $M \leq 3)$.
(d) Czy zmienne losowe D i M są niezależne? Odpowiedź uzasadnij.

4. Znajdź dystrybuanty zmiennych losowych D i M z ćwiczenia 3. Naszkicuj ich wykresy.

5. Niech T będzie zmienną losową zdefiniowaną na zbiorze Ω z przykładu 3 wzorem: $T(k,l) = k \cdot l$ (k razy l).
(a) Znajdź zbiór wartości zmiennej T.
(b) Oblicz $P(T \leq 2)$.
(c) Oblicz $P(T = 12)$.

6. Przypuśćmy, że niezależne zmienne losowe X i Y mają ten sam rozkład prawdopodobieństwa f, gdzie $f(0) = f(1) = \frac{1}{4}$, $f(2) = \frac{1}{2}$ i $f(x) = 0$ dla pozostałych wartości x. Oblicz
(a) $P(X = 0$ i $Y = 2)$,
(b) $P(X = 0$ lub $Y = 2)$,
(c) $P(X \leq 1$ i $Y \geq 1)$.

7. (a) Znajdź zbiór wartości sumy $X + Y$ zmiennych losowych z ćwiczenia 6.
(b) Oblicz $P(X + Y = 2)$.
(c) Znajdź rozkład prawdopodobieństwa zmiennej $X + Y$.

8. (a) Dla zmiennej X z ćwiczenia 6 znajdź rozkład prawdopodobieństwa zmiennej losowej $3X + 2$.
(b) Zrób to samo dla zmiennej $2 - X$.

9. W urnie znajduje się 5 białych i 5 niebieskich kulek.
(a) Wybieramy losowo cztery kulki (bez zwracania). Znajdź rozkład prawdopodobieństwa zmiennej losowej X, która podaje liczbę wylosowanych kulek białych.
(b) Powtórz część (a) dla przypadku, gdy wybieramy losowo siedem kulek.

10. Gracz rzuca jeden raz symetryczną kostką i wygrywa 5 dolarów, jeśli wypadnie 5, a traci dolara, w przeciwnym razie. Znajdź rozkład prawdopodobieństwa zmiennej losowej W, która podaje wygrane gracza (strata jest ujemną wygraną).

11. (a) Rozważ funkcję $f\colon \mathbb{R} \to \mathbb{R}$, gdzie

$$f(x) = \begin{cases} 1/4, & \text{jeśli } x = 1, 3, 5, \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

Czy funkcja f jest rozkładem prawdopodobieństwa jakiejś zmiennej losowej? Uzasadnij odpowiedź.

(b) Powtórz część (a) dla funkcji

$$f(x) = \begin{cases} 1/5, & \text{jeśli } x = 1, 2, 3, 4, 5, \\ 0, & \text{w przeciwnym razie.} \end{cases}$$

12. Narysuj wykres dystrybuanty dowolnej zmiennej losowej X spełniającej warunki $P(X = 0) = P(X = 1) = \frac{1}{2}$. Będzie to dystrybuanta zmiennych losowych $X_1, \ldots, X_n$ z przykładu 6(b) i zmiennych losowych z przykładu 7.

13. Niech F będzie dystrybuantą pewnej zmiennej losowej X. Udowodnij, że

(a) $0 \leq F(y) \leq 1$ dla wszystkich $y \in \mathbb{R}$.

(b) Funkcja F jest niemalejąca, tzn. $y_1 < y_2$ pociąga za sobą $F(y_1) \leq F(y_2)$.

14. Rzucamy n-krotnie symetryczną monetą. Niech X będzie zmienną losową, która podaje liczbę wyrzuconych orłów, a Y — liczbę wyrzuconych reszek.

(a) Czy zmienne X i Y mają te same rozkłady prawdopodobieństwa? Uzasadnij odpowiedź.

(b) Czemu jest równa zmienna losowa $X + Y$?

15. Rzucamy symetryczną kostką. Niech W będzie zmienną losową podającą liczbę rzutów do momentu, gdy po raz pierwszy pojawi się 6. Znajdź rozkład prawdopodobieństwa $f(x) = P(W = x)$ zmiennej W.

16. Rzucamy n-krotnie symetryczną kostką.

(a) Znajdź prawdopodobieństwo tego, że w każdym rzucie wypadnie 4, 5 lub 6.

(b) Znajdź prawdopodobieństwo tego, że w każdym rzucie wypadnie 5 lub 6.

(c) Znajdź prawdopodobieństwo tego, że w każdym rzucie wypadnie 6.

17. Losowo wybieramy punkt x z odcinka [0,1). Znajdź prawdopodobieństwo, tego że

(a) $|x - \frac{1}{2}| \leq \frac{1}{3}$,

(b) $\min\{|x|, |x - 1|\} \leq \frac{1}{3}$.

18. Wykaż, że dwa zdarzenia E i F są niezależne wtedy i tylko wtedy, gdy ich funkcje charakterystyczne χ_E i χ_F są niezależnymi zmiennymi losowymi.

19. Udowodnij, że warunki (I_2) i (I_3), sformułowane przed przykładem 2, są równoważne dla zmiennych losowych X i Y mających skończone zbiory wartości.

§ 9.3. Wartość oczekiwana i odchylenie standardowe

Rozważmy zmienną losową X na skończonej przestrzeni zdarzeń elementarnych Ω, w której wszystkie wyniki (tzn. elementy przestrzeni) są jednakowo prawdopodobne. Wówczas średnia wartość

$$A = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} X(\omega),$$

zmiennej X na Ω ma probabilistyczną interpretację: jeśli wielokrotnie dokonujemy losowego wyboru elementów ω przestrzeni Ω, zapisując za każdym razem wartość $X(\omega)$, to średnia z takich wartości prawdopodobnie będzie bliska liczby A. Zdanie to jest w gruncie rzeczy twierdzeniem, które wymaga dowodu, ale mamy nadzieję, że zaakceptujesz je jako w miarę zgodne z intuicją.

PRZYKŁAD 1 Rzucamy symetryczną kostką i zmienna X podaje liczbę oczek, które wypadły na górnej ściance kostki. Każdy z sześciu wyników 1, 2, 3, 4, 5 i 6 jest jednakowo prawdopodobny:

$$P(X = 1) = P(X = 2) = \ldots = P(X = 6) = \frac{1}{6}.$$

Średni wynik wynosi

$$\frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2} = 3,5.$$

Jeśli wielokrotnie rzucalibyśmy kostką, to spodziewalibyśmy się, że średnia z uzyskanych wyników będzie bliska 3,5. Zatem „oczekiwana średnia" wynosi 3,5. ∎

Jeśli wyniki nie są jednakowo prawdopodobne, to interesuje nas średnia, która będzie związana z prawdopodobieństwami tych wyników, a więc wartości przyjmowane przez zmienną losową X pomnożymy przez odpowiednie współczynniki. Dla zmiennej losowej X na skończonej przestrzeni zdarzeń elementarnych Ω definiujemy jej wartość **średnią** albo **wartość oczekiwaną** wzorem

$$E(X) = \mu = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}).$$

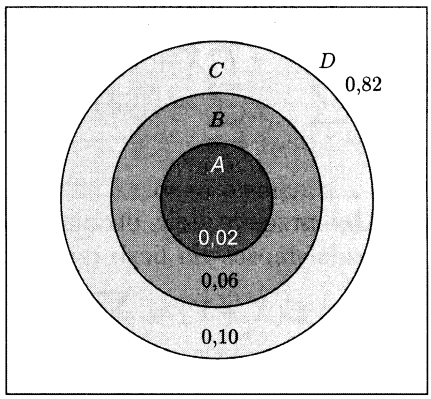
Jeśli wszystkie wyniki są jednakowo prawdopodobne, to $P(\{\Omega\}) = 1/|\Omega|$ dla każdego wyniku ω z przestrzeni Ω, a więc wartość oczekiwana $E(X)$ jest dokładnie równa średniej A, którą omawialiśmy przed przykładem 1. Zauważmy, że używamy dwóch terminów i symboli dla tego samego pojęcia: wartość oczekiwaną bądź

wartość średnią zmiennej X oznaczamy przez $E(X)$ lub μ (mała grecka litera mi) albo przez μ_X, jeśli chcemy zwrócić uwagę, o jaką zmienną losową chodzi.

PRZYKŁAD 2 Pewien miłośnik gry w rzutki wie z doświadczenia, że w każdym rzucie do tarczy przedstawionej na rysunku 9.6, trafia w jej odpowiednie obszary ze wskazanymi prawdopodobieństwami. Zatem $P(A) = 0{,}02$, $P(B) = 0{,}06$ itd. Właściciel salonu gry oferuje 10 dolarów za każde trafienie strzałką w obszar A, 5 dolarów za każde trafienie w obszar B i 3 dolary ilekroć strzałka trafi w obszar C. Ile wynosi spodziewana średnia wygrana naszego gracza? Nie jest to z pewnością średnia $\frac{1}{4}(10 + 3 + 1 + 0) = 3,50$. Jest to probabilistyczna średnia

$$\begin{aligned}\mu &= 10 \cdot (0{,}02) + 3 \cdot (0{,}06) + 1 \cdot (0,10) + 0 \cdot (0,82) \\ &= 0{,}20 + 0,18 + 0,10 = 0{,}48.\end{aligned}$$

Jest to wartość oczekiwana $E(X)$ zmiennej losowej X określonej na przestrzeni zdarzeń elementarnych $\Omega = \{A, B, C, D\}$ za pomocą równości $X(A) = 10$, $X(B) = 3$, $X(C) = 1$ i $X(D) = 0$.



Rysunek 9.6

Właściciel lokalu pobiera zwykle opłaty za prawo do gry. Opłata wynosząca 48 centów za jeden rzut byłaby „sprawiedliwa" w tym sensie, że spodziewana wygrana gracza wynosiłaby wtedy 0. Ale z punktu widzenia właściciela rozsądne byłoby, gdyby opłata za jedną próbę wynosiła trochę więcej niż 48 centów.

Wyjaśnijmy, dlaczego spodziewana wygrana gracza wynosiłaby 0, gdyby opłata za jeden rzut była równa μ. Nową zmienną losową, reprezentującą wygraną netto naszego gracza, byłaby funkcja $X - \mu$; zobacz tablica 9.5. Można sprawdzić bezpośrednio,

że $E(X-\mu)=0$, ale — jak dowiemy się z następnego twierdzenia — równość ta zawsze zachodzi. ∎

Tablica 9.5

Obszar	X [bez opłaty]	$X-\mu$ [z opłatą μ]
A	10,00	9,52
B	3,00	2,52
C	1,00	0,52
D	0	−0,48

Twierdzenie 1

Niech X i Y będą zmiennymi losowymi na skończonej przestrzeni zdarzeń elementarnych Ω. Wówczas
(a) $E(X+Y)=E(X)+E(Y)$.
(b) $E(aX)=a\cdot E(X)$ dla $a\in\mathbb{R}$.
(c) $E(a)=a$ dla $a\in\mathbb{R}$.

W części (c), litera a w nawiasie oznacza funkcję stałą na Ω; jest to w pełni dopuszczalna zmienna losowa.

Dowód. Mamy

$$E(X+Y)=\sum_{\omega\in\Omega}(X+Y)(\omega)\cdot P(\{\omega\})$$
$$=\sum_{\omega\in\Omega}[X(\omega)+Y(\omega)]\cdot P(\{\omega\}).$$

Aby przekształcić otrzymane wyrażenie, użyjemy teraz prawa rozdzielności dla liczb rzeczywistych:

$$E(X+Y)=\sum_{\omega\in\Omega}X(\omega)\cdot P(\{\omega\})$$
$$+\sum_{\omega\in\Omega}Y(\omega)\cdot P(\{\omega\})=E(X)+E(Y).$$

Zatem równość (a) jest prawdziwa; dowód równości (b) jest jeszcze łatwiejszy. W końcu,

$$E(a)=\sum_{\omega\in\Omega}a\cdot P(\{\omega\})=a\cdot\sum_{\omega\in\Omega}P(\{\omega\})=a\cdot P(\Omega)=a.$$ ∎

Wniosek

Dla dowolnej zmiennej losowej X mamy $E(X-\mu)=0$.

Dowód. $E(X-\mu)=E(X)-E(\mu)=\mu-\mu=0$. ∎

Łatwo pokazuje się przez indukcję, że

$$E(X_1 + X_2 + \ldots + X_n) = E(X_1) + E(X_2) + \ldots + E(X_n)$$

dla zmiennych losowych $X_1, X_2, \ldots, X_n$. Następne twierdzenie pokazuje, w jaki sposób można wyznaczyć wartość oczekiwaną za pomocą rozkładu prawdopodobieństwa. Ponownie usuwamy z pola widzenia przestrzeń zdarzeń elementarnych.

Twierdzenie 2

Jeśli X jest zmienną losową na skończonej przestrzeni zdarzeń elementarnych, to

$$E(X) = \sum_x x \cdot P(X = x) = \sum_x x \cdot f_X(x).$$

W powyższym wyrażeniu występują sumy skończone, gdyż składniki odpowiadające elementom x spoza zbioru wartości zmiennej X są równe 0.

Dowód. Niech A będzie zbiorem wartości zmiennej X. Wówczas rodzina zbiorów $\{\{X = x\}\colon x \in A\}$ tworzy podział przestrzeni Ω. (Zbiór $\{X = x\}$ można również zapisać jako $X^{\leftarrow}(x)$). Zatem

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) = \sum_{x \in A} \Big\{ \sum_{\omega \in \{X = x\}} X(\omega) \cdot P(\{\omega\}) \Big\}.$$

Ponieważ $X(\omega) = x$ dla każdego elementu ω ze zbioru $\{X = x\}$ (na mocy definicji tego zbioru!), więc wewnętrzna suma równa się

$$\sum_{\omega \in \{X = x\}} X(\omega) \cdot P(\{\omega\}) = x \cdot \sum_{\omega \in \{X = x\}} P(\{\omega\}) = x \cdot P(X = x).$$

W konsekwencji mamy

$$E(X) = \sum_{x \in A} x \cdot P(X = x). \qquad \blacksquare$$

Wzór przedstawiony w twierdzeniu 2 mógł zostać przyjęty za definicję wartości oczekiwanej i jest to istotnie definicja, którą można stosować w przypadku wszystkich dyskretnych zmiennych losowych. Nie można jej stosować w ogólnej sytuacji, gdyż rozkład prawdopodobieństwa f_X może nie być do niczego przydatny, jeśli zmienna X nie jest dyskretna. Okazuje się, że dla wielu zmiennych losowych, w twierdzeniu 2 można zastąpić sumy sumami uogólnionymi, zwanymi całkami i oznaczanymi symbolem

$\int$, przypominającym wydłużoną literę S. Rozkład prawdopodobieństwa zastępuje się pewną funkcją f_X, określoną na zbiorze $\mathbb{R}$, zwaną gęstością rozkładu. Wartość oczekiwana zmiennej X wyraża się wówczas wzorem $E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x)\mathrm{d}x$, a gęstość rozkładu wyznacza jej dystrybuantę F_X za pomocą wzoru $F_X(y) = \int_{-\infty}^{y} f_X(x)\mathrm{d}x$. Zauważ, że wzory te wyglądają na uogólnienia przyjętych przez nas wcześniej definicji. Nawet wtedy, gdy nie można zastosować tych uogólnionych wzorów, wartość oczekiwaną $E(X)$ daje się zdefiniować za pomocą dystrybuanty F_X zmiennej X.

PRZYKŁAD 3 Zilustrujemy poprzednie twierdzenia, rozpatrując ponownie zmienne losowe X_S, X_B i X_R na przestrzeni Ω, która jest modelem zagadnienia rzutu dwiema kostkami (przykłady 1 i 2, § 9.2).

(a) Aby obliczyć $E(X_S)$ wprost z definicji, musimy dodać do siebie 36 liczb:

$$E(X_S) = \sum_{(k,l)\in\Omega} X_S(k,l)\cdot P(k,l) = \frac{1}{36}(2+3+3+4+\ldots+11+12).$$

Przejdźmy raczej do punktu (b).

(b) Obliczymy $E(X_S)$ wykorzystując twierdzenie 2 i wartości $P(X_S = k)$, podane w tablicy 9.4(a) w § 9.2:

$$\begin{aligned} E(X_S) &= 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + 5\cdot\frac{4}{36} + 6\cdot\frac{5}{36} + 7\cdot\frac{6}{36} \\ &\quad + 8\cdot\frac{5}{36} + 9\cdot\frac{4}{36} + 10\cdot\frac{3}{36} + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36} \\ &= \frac{1}{36}[2+6+12+20+30+42+40+36+30+22+12] \\ &= \frac{252}{36} = 7. \end{aligned}$$

Jeszcze lepsze rozwiązanie podajemy w punkcie (c).

(c) Przypomnij sobie, że $X_S = X_B + X_R$. Z przykładu 1 mamy $E(X_B) = E(X_R) = 3,5$, a więc $E(X_S) = E(X_B) + E(X_R) = 7$. Innymi słowy, wartość oczekiwana liczby oczek na kostce białej wynosi 3,5 i wartość oczekiwana liczby oczek na kostce szarej wynosi 3,5, więc wartość oczekiwana sumy oczek na obu kostkach jest równa 7. ∎

PRZYKŁAD 4 Rzucamy n-krotnie symetryczną monetą. Jaka jest oczekiwana liczba orłów? Jeśli Twoja intuicja mówi Ci, że $n/2$, to masz rację. Prosimy jednak, byś czytał dalej.

(a) Nasze pytanie brzmi: ile wynosi $E(S_n)$, gdzie S_n jest zmienną losową, która podaje, ile wypadło orłów? Korzystając z twierdzenia 2 otrzymujemy

$$E(S_n) = \sum_{k=0}^{n} k \cdot P(S_n = k) = \sum_{k=0}^{n} k \cdot \frac{1}{2^n} \binom{n}{k}.$$

Powyższe wyrażenie jest zbyt skomplikowane, ale możemy wykorzystać doświadczenia zdobyte w przykładzie 3. Zmienna losowa S_n jest sumą $X_1 + X_2 + \ldots + X_n$, gdzie X_i równa się 1, jeśli w i-tym rzucie wypadł orzeł i 0 — w przeciwnym razie. Zatem $P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$. Ponieważ dla każdego i mamy

$$E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = \frac{1}{2},$$

to widzimy, że

$$E(S_n) = \sum_{i=1}^{n} E(X_i) = \frac{n}{2},$$

co jest zgodne z naszymi przewidywaniami.

(b) Możesz być zaintrygowany sumą, która pojawiła się na początku punktu (a). Skoro jednak równość, w której ona występuje, jest prawdziwa, to otrzymujemy stąd interesujący dowód „dwumianowej tożsamości":

$$\sum_{k=0}^{n} k \cdot \binom{n}{k} = n \cdot 2^{n-1}.$$

Wiele podobnych intrygujących tożsamości wykorzystuje się i bada w rachunku prawdopodobieństwa i kombinatoryce. ∎

PRZYKŁAD 5 (a) Niech W będzie zmienną losową, która podaje, w którym z kolejnych rzutów symetryczną monetą po raz pierwszy wypadł orzeł. Ile wynosi oczekiwany średni czas czekania na wypadnięcie orła? Możesz spróbować zbadać to eksperymentalnie. Odpowiedź jest dana za pomocą nieskończonej sumy

$$\sum_{k=1}^{\infty} k \cdot P(W = k) = \sum_{k=1}^{\infty} k \cdot \frac{1}{2^k},$$

która, jak się okazuje, jest równa 2.

Czy odpowiedź: „2" jest zgodna z intuicją? Sądzimy, że dla niektórych osób tak, ale nie dla większości. Oto niedorzeczny argument, który nie jest jednak całkiem bezsensowny: w jednym

rzucie możemy oczekiwać połowy orła, a więc powinniśmy otrzymać całego orła w dwóch rzutach!

(b) Okazuje się, że jeśli losowo wybierzesz liczbę z przedziału $[0,1)$, to jej oczekiwana średnia wartość jest równa $\frac{1}{2}$. To znaczy, jeśli będziesz wielokrotnie losowo wybierał liczby z tego przedziału, to ich średnia będzie bliska $\frac{1}{2}$. Nie będziemy rozwijać tego tematu, gdyż dla tego rodzaju zmiennych losowych nie zdefiniowaliśmy nawet pojęcia wartości oczekiwanej $E(X)$. ∎

Oba wzory

$$E(X) = \sum_{\omega \in \Omega} X(\omega) \cdot P(\{\omega\}) \quad \text{oraz} \quad E(X) = \sum_{x} x \cdot P(X = x)$$

stanowią szczególne przypadki następującego lematu.

Lemat

Niech Z będzie zmienną losową na przestrzeni zdarzeń elementarnych Ω. Przypuśćmy, że dany jest podział przestrzeni Ω taki, że zmienna Z jest stała na każdym bloku tego podziału. Wtedy wartość oczekiwana $E(Z)$ jest sumą wszystkich składników następującej postaci: prawdopodobieństwo bloku podziału pomnożone przez (stałą) wartość, przyjmowaną przez zmienną Z na każdym z elementów tego bloku.

Dowód. Niech $E_1, \ldots, E_k$ będą blokami danego podziału. Dla każdego i, niech $v(i)$ będzie wartością przyjmowaną przez Z na elementach zbioru E_i. Wówczas $A = \{v(i)\colon\ i = 1, \ldots, k\}$ jest zbiorem wartości zmiennej Z; niektóre wartości mogły zostać wypisane kilkakrotnie. Dla dowolnego elementu x zbioru A, zbiory postaci $\{E_i\colon\ v(i) = x\}$ tworzą podział zbioru $\{Z = x\}$, stąd

$$P(Z = x) = \sum_{\{i\colon\ v(i)=x\}} P(E_i).$$

Zatem

$$x \cdot P(Z = x) = \sum_{\{i\colon\ v(i)=x\}} v(i) \cdot P(E_i),$$

a więc

$$E(Z) = \sum_{x \in A} x \cdot P(Z = x) = \sum_{x \in A} \sum_{\{i\colon\ v(i)=x\}} v(i) \cdot P(E_i).$$

Ponieważ zbiór $\{1, 2, \ldots, k\}$ jest rozłączną sumą zbiorów postaci $\{i\colon v(i) = x\}$, gdzie $x \in A$, to ostatecznie otrzymujemy

$$E(Z) = \sum_{i=1}^{k} v(i) \cdot P(E_i).$$ ∎

Użyteczne jest umieć wyrażać wartości oczekiwane $E(X^2)$, $E(|X|)$ itd. za pomocą wielkości związanych ze zmienną losową X. Zmienne losowe takie jak X^2 i $|X|$ są postaci $\varphi \circ X$, gdzie $\varphi\colon \mathbb{R} \to \mathbb{R}$. Istotnie, $X^2 = \varphi \circ X$, gdzie $\varphi(x) = x^2$ i $|X| = \varphi \circ X$, gdzie $\varphi(x) = |x|$.

Twierdzenie 3

Dla zmiennej losowej X o skończonym zbiorze wartości oraz funkcji $\varphi\colon \mathbb{R} \to \mathbb{R}$ mamy

$$E(\varphi \circ X) = \sum_{x} \varphi(x) \cdot P(X = x).$$

Powyższe sumowanie odbywa się po wszystkich elementach x ze zbioru wartości zmiennej X.

Dowód. Ponieważ zmienna $\varphi \circ X$ jest stała na zbiorach, na których stała jest zmienna X, to możemy zastosować ostatni lemat do zmiennej losowej $Z = \varphi \circ X$ i zbiorów postaci $\{X = x\}$, które tworzą podział przestrzeni Ω. Na elementach zbioru $\{X = x\}$, zmienna $\varphi \circ X$ przyjmuje wartość $\varphi(x)$. ∎

Wniosek

Dla zmiennej losowej X o skończonym zbiorze wartości mamy

$$E(X^2) = \sum_{x} x^2 \cdot P(X = x)$$

oraz

$$E((X - \mu)^2) = \sum_{x} (x - \mu)^2 \cdot P(X = x).$$

Potrzebny nam też będzie wzór na $E(XY)$; wyniknie on z następnego twierdzenia. Dla zmiennych losowych X i Y o skończonych zbiorach wartości oraz funkcji $\psi\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, symbolem $\psi(X, Y)$ będziemy oznaczać zmienną losową zdefiniowaną wzorem $\psi(X, Y)(\omega) = \psi(X(\omega), Y(\omega))$ dla $\omega \in \Omega$. Na przykład, jeśli $\psi(x, y) = xy$ dla $x, y \in \mathbb{R}$, to $\psi(X, Y)(\omega) = X(\omega)Y(\omega)$ dla $\omega \in \Omega$, tzn. $\psi(X, Y) = XY$.

Twierdzenie 4

Dla X, Y i ψ jak wyżej mamy

$$E(\psi(X,Y)) = \sum_x \sum_y \psi(x,y) \cdot P(X = x \text{ i } Y = y).$$

Sumowanie odbywa się po elementach (skończonych) zbiorów wartości zmiennych X i Y.

Dowód. Zbiory $\{\omega \in \Omega\colon X(\omega) = x \text{ i } Y(\omega) = y\}$ tworzą podział przestrzeni Ω. Dla argumentów z takiego zbioru zmienna losowa $\psi(X,Y)$ przyjmuje wartość $\psi(x,y)$. Twierdzenie wynika więc z lematu udowodnionego przed twierdzeniem 3. ∎

Wniosek

$$E(XY) = \sum_x \sum_y xy \cdot P(X = x \text{ i } Y = y).$$

Uwaga: Następny fakt nie musi być prawdziwy dla dowolnych zmiennych losowych; musimy zakładać, że są one niezależne.

Twierdzenie 5

Dla niezależnych zmiennych losowych X i Y mamy

$$E(XY) = E(X) \cdot E(Y).$$

Dowód. Twierdzenie jest ogólnie prawdziwe, ale nasz dowód obejmuje jedynie przypadek, gdy zmienne X i Y mają skończone zbiory wartości. Na mocy ostatniego wniosku

$$E(XY) = \sum_x \sum_y xy \cdot P(X = x \text{ i } Y = y).$$

Ponieważ zmienne X i Y są niezależne, to

$$E(XY) = \sum_x \sum_y x \cdot y \cdot P(X = x) \cdot P(Y = y).$$

Ponieważ powyższe sumy są skończone, to możemy posłużyć się prawem rozdzielności dla liczb rzeczywistych i zapisać tę równość w postaci

$$\begin{aligned} E(XY) &= \Big\{\sum_x x \cdot P(X = x)\Big\} \cdot \Big\{\sum_y y \cdot P(Y = y)\Big\} \\ &= E(X) \cdot E(Y). \end{aligned}$$

∎

Wartość oczekiwana zmiennej losowej X podaje nam jej „probabilistyczną średnią". Jednakże nie możemy z niej odczytać, na ile wartości tej zmiennej będą bliskie wartości średniej. Potrzebna nam jest inna jeszcze wielkość, która to właśnie będzie mierzyć. Naturalnym kandydatem jest „probabilistyczna średnia" odległości zmiennej X od jej wartości średniej μ. Mamy na myśli „średnie odchylenie" $E(|X-\mu|)$, tzn. wartość średnią wszystkich odchyleń $|X(\omega)-\mu|$ dla $\omega \in \Omega$. Wielkość ta jest czasami wykorzystywana, ale okazuje się, że podobna wielkość, zwana odchyleniem standardowym, jest znacznie wygodniejsza i bardziej użyteczna. **Odchyleniem standardowym** σ (lub σ_X) nazywamy pierwiastek kwadratowy z liczby $E((X-\mu)^2)$. Podobnie jak „średnie odchylenie", jest to liczba leżącą między najmniejszymi i największymi odchyleniami $|X(\omega)-\mu|$, będąc czymś w rodzaju średniej z tych odchyleń. Trudność z $E(|X-\mu|)$ leży w tym, że posługiwanie się funkcją $|x|$, w przeciwieństwie do funkcji x^2, sprawia kłopoty. Studenci, którzy zetknęli się z analizą matematyczną, pewnie przypominają sobie, że trudności w posługiwaniu się funkcją $|x|$ związane są przede wszystkim z jej nieprzyjemnym zachowaniem w punkcie 0, w którym jej wykres skręca pod kątem prostym.

Jeśli X jest zmienną losową o wartości średniej μ, to kwadrat jej odchylenia standardowego σ nazywamy **wariancją zmiennej losowej** X i oznaczamy symbolem $V(X)$. Zatem

$$V(X) = \sigma^2 = \sigma_X^2 = E((X-\mu)^2).$$

Na mocy wniosku z twierdzenia 3 mamy

$$V(X) = \sum_x (x-\mu)^2 \cdot P(X=x),$$

o ile zbiór wartości zmiennej X jest skończony. Wariancja i odchylenie standardowe mierzą rozproszenie wartości zmiennej X. Im liczba $V(X)$ jest mniejsza, tym większą możemy mieć pewność, że losowo wybrana wartość $X(\omega)$ jest bliska średniej μ. W szczególności, $V(X) = 0$ wtedy i tylko wtedy, gdy zmienna X jest stała i równa μ.

PRZYKŁAD 6 (a) Niech X będzie zmienną losową podającą liczbę oczek, które wypadły w wyniku rzutu symetryczną kostką. Ponieważ z przykładu 1 wiemy, że $\mu = 3,5$, to mamy

$$\begin{aligned} V(X) &= \sum_{k=1}^{6} (k-\mu)^2 \cdot \frac{1}{6} \\ &= \frac{1}{6}[(1-3{,}5)^2 + (2-3,5)^2 + (3-3{,}5)^2 + \end{aligned}$$

$$+ (4-3{,}5)^2 + (5-3{,}5)^2 + (6-3{,}5)^2]$$
$$= \frac{1}{6} \cdot 17{,}5 = \frac{35}{12}.$$

Odchylenie standardowe wynosi $\sigma = \sqrt{\frac{35}{12}} \approx 1{,}71$.

(b) Niech X będzie zmienną losową związaną z rzutem symetryczną monetą, tzn. $P(X=0) = P(X=1) = \frac{1}{2}$. Wtedy

$$\mu = \frac{1}{2}, \quad V(X) = \left(0 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}$$

oraz

$$\sigma = \sqrt{V(X)} = \frac{1}{2}. \qquad \blacksquare$$

Kolejny wzór na $V(X)$ pozwala uprościć nieco obliczenia, gdyż wartość średnia μ występuje w nim tylko raz.

Twierdzenie 6

Dla zmiennej losowej X o wartości średniej μ mamy $V(X) = E(X^2) - \mu^2$.

Dowód. Ponieważ $(X-\mu)^2 = X^2 - 2\mu X + \mu^2$, to z twierdzenia 1 wynika, że

$$V(X) = E(X^2) - 2\mu \cdot E(X) + \mu^2.$$

Ale $E(X) = \mu$, więc

$$V(X) = E(X^2) - 2\mu^2 + \mu^2 = E(X^2) - \mu^2. \qquad \blacksquare$$

PRZYKŁAD 6 raz jeszcze

(a) Mamy

$$E(X^2) = \sum_{k=1}^{6} k^2 \cdot \frac{1}{6} = \frac{1}{6}[1+4+9+16+25+36] = \frac{91}{6},$$

a więc

$$V(X) = E(X^2) - \mu^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

(b) Mamy $E(X^2) = 1 \cdot \frac{1}{2} = \frac{1}{2}$, więc

$$V(X) = E(X^2) - \mu^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}.$$

W ogólnej sytuacji wzór na $V(X+Y)$ jest skomplikowany, ale dla niezależnych zmiennych losowych mamy następujący rezultat.

Twierdzenie 7

Dla niezależnych zmiennych losowych $X_1, X_2, \ldots, X_n$ mamy

$$V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n).$$

Dowód. Można to najpierw udowodnić dla dwu zmiennych losowych, a następnie zastosować rozumowanie indukcyjne. Nie jest to jednakże całkiem proste, dopóki nie wiemy, że zmienne $X_1 + X_2, X_3, \ldots, X_n$ są niezależne. Zob. ćwiczenie 21. Dowód, który podajemy, obywa się bez indukcji.

Przypadek 1: Załóżmy, że $E(X_i) = 0$ dla każdego i, a więc dla każdego i mamy $V(X_i) = E(X_i^2)$. Wtedy

$$V\Big(\sum_{i=1}^{n} X_i)\Big) = E\Big(\Big(\sum_{i=1}^{n} X_i\Big)^2\Big)$$

$$= E\Big(\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\Big) = \sum_{i=1}^{n}\sum_{j=1}^{n} E(X_i X_j).$$

Ponieważ zmienne X_i są niezależne, to z twierdzenia 5 wynika, że $E(X_i X_j) = E(X_i) \cdot E(X_j) = 0 \cdot 0$ dla $i \neq j$. Zatem w tym przypadku

$$V\Big(\sum_{i=1}^{n} X_i\Big) = \sum_{i=1}^{n} E(X_i^2) = \sum_{i=1}^{n} V(X_i).$$

Przypadek 2: W przypadku ogólnym dla każdego i oznaczmy $E(X_i)$ przez μ_i. Ponieważ mamy $E(X_i - \mu_i) = 0$ dla każdego i, to możemy zastosować to, co udowodniliśmy w przypadku 1, do zmiennych losowych $X_i - \mu_i$. Otrzymujemy wówczas

$$V\Big(\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i\Big) = V\Big(\sum_{i=1}^{n} (X_i - \mu_i)\Big) = \sum_{i=1}^{n} V(X_i - \mu_i).$$

Ponieważ dla dowolnej zmiennej losowej Y i każdej liczby $c \in \mathbb{R}$ mamy $V(Y + c) = V(Y)$ (ćwiczenie 17), to pomijając stałe otrzymujemy $V(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} V(X_i)$. ∎

PRZYKŁAD 7

(a) Szukamy wariancji zmiennej losowej X_S, podającej sumy liczb oczek, które wypadły w rzucie dwiema symetrycznymi kostkami. Próba bezpośredniego skorzystania z definicji lub z twierdzenia 6 zmusiłaby nas do ciężkiej pracy. Ale wiemy, że $X_S = X_B + X_R$, gdzie zmienne X_B i X_R są niezależne. Z przykładu 6

wiemy też, że $V(X_B) = V(X_R) = \frac{35}{12}$. Stąd na mocy twierdzenia 7 otrzymujemy $V(X_S) = \frac{35}{6}$, a więc $\sigma_S = \sqrt{\frac{35}{6}} \approx 2,42$.

(b) Aby obliczyć wariancję zmiennej losowej S_n podającej liczbę orłów, które wypadły w wyniku n rzutów symetryczną monetą, przypomnijmy sobie, że S_n jest sumą wzajemnie niezależnych zmiennych losowych $X_1, \ldots, X_n$, gdzie $P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$. Wiemy z przykładu 6, że $V(X_i) = \frac{1}{4}$ dla każdego i. Z twierdzenia 7 wynika teraz, że

$$V(S_n) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} \frac{1}{4} = \frac{n}{4}.$$

Odchylenie standardowe zmiennej S_n wynosi $\frac{1}{2}\sqrt{n}$; liczba ta zwiększa się wraz ze wzrostem n, ale dla dużych wartości n rośnie znacznie wolniej niż n. ∎

Twierdzenia 1, 6 i 7 są prawdziwe w ogólności, ale nasze dowody odnoszą się jedynie do przypadku, w którym rozpatrywane zmienne losowe określone są na skończonych przestrzeniach zdarzeń elementarnych; wynika to ze sposobu, w jaki udowodniliśmy twierdzenie 1.

ĆWICZENIA DO § 9.3

1. Tak jak w ćwiczeniu 9(a) z § 9.2, z urny, w której jest 5 czerwonych i 5 niebieskich kulek, wyciągamy losowo (bez zwracania) cztery kulki.
(a) Jaka jest wartość oczekiwana liczby wyciągniętych czerwonych kulek?
(b) Ile wynosi odchylenie standardowe?

2. W artykule zamieszczonym w gazecie podano, że przeciętna rodzina ma 2,1 dzieci i 1,8 samochodów. Ile jest przeciętnych rodzin? Omów to zagadnienie.

3. Oblicz średnie odchylenia zmiennych losowych z przykładu 6 i porównaj otrzymane wyniki z odchyleniami standardowymi tych zmiennych.

4. Przypuśćmy, że każdy los pewnej loterii kosztuje 1 dolar, a jedyna nagroda wynosi 1000 000 dolarów. Jeśli dla każdego losu prawdopodobieństwo wygranej wynosi 0,0000005, to jaka jest wartość oczekiwana wygranej w przypadku nabycia jednego losu?

5. Rozważmy niezależne zmienne losowe X i Y o tym samym rozkładzie prawdopodobieństwa f, gdzie $f(0) = f(1) = \frac{1}{4}$ i $f(2) = \frac{1}{2}$. Znajdź wartości oczekiwane i odchylenia standardowe zmiennych X, Y i $X+Y$.

6. Znajdź wartości oczekiwane i odchylenia standardowe zmiennych losowych D i M z ćwiczenia 3 z § 9.2, określonych na przestrzeni Ω, złożonej z 36 jednakowo prawdopodobnych wyników rzutu dwiema symetrycznymi kostkami, wzorami: $D(k,l) = |k-l|$ i $M(k,l) = \max\{k,l\}$.

7. Niech X będzie zmienną losową o rozkładzie prawdopodobieństwa f, gdzie $f(-2) = f(0) = f(1) = \frac{1}{5}$, $f(2) = \frac{2}{5}$ i $f(x) = 0$ dla pozostałych wartości x. Znajdź wartości oczekiwane następujących zmiennych losowych:
(a) X, (b) $|X|$, (c) X^2, (d) 3X+2.

8. Znajdź odchylenia standardowe zmiennych losowych X oraz $|X|$ z ćwiczenia 7.

9. Znajdź odchylenie standardowe zmiennej losowej X^2 z ćwiczenia 7.

10. Czy wolałbyś mieć 50% szans na wygranie 1000 000 dolarów, czy 20% szans na wygranie 3000 000 dolarów? Omów to zagadnienie.

11. Jaka jest oczekiwana liczba asów w układzie pięciu kart w pokerze? *Wskazówka*: można uniknąć obliczeń, do których prowadzi bezpośrednie zastosowanie twierdzenia 2.

12. Rzucamy symetryczną kostką aż do momentu, gdy po raz pierwszy wypadnie 5 oczek; zob. ćwiczenie 15 do § 9.2. Zastosuj „niedorzeczny" argument z przykładu 5, aby określić, jaki jest spodziewany czas oczekiwania na pierwsze wyrzucenie 5 oczek.

13. W urnie znajdują się 4 kulki czerwone i 1 niebieska.
(a) Wyciągamy kulki z urny bez zwracania aż do momentu wyciągnięcia po raz pierwszy kulki niebieskiej. Jaki jest spodziewany czas oczekiwania na wyciągnięcie niebieskiej kulki?
(b) Powtórz część (a) dla przypadku, w którym po każdym losowaniu zwracamy kulkę do urny. *Wskazówka*: zob. ćwiczenie 12.

14. Wykaż, że $\sum_{k=1}^{n} k^2 \cdot \binom{n}{k} = n(n+1)\cdot 2^{n-2}$ dla $n \geq 0$. *Wskazówka*: Niech S_n będzie zmienną losową z przykładu 7. Zatem $V(S_n) = n/4$ oraz $\mu = E(S_n) = n/2$. Oblicz $E(S_n^2)$ dwoma sposobami: stosując twierdzenie 6 i korzystając z wniosku z twierdzenia 3.

15. Niech X będzie zmienną losową wybierającą losowo liczbę całkowitą ze zbioru $\{1, 2, \ldots, n\}$.
(a) Zauważ, że zmienna losowa $Y = n + 1 - X$ ma taki sam rozkład prawdopodobieństwa jak zmienna X.
(b) Wykorzystaj część (a) do wykazania, że $E(X) = \frac{1}{2}(n+1)$.
(c) Zastosuj twierdzenie 2 do wyprowadzenia wzoru $1 + 2 + \ldots + n = \frac{1}{2}n(n+1)$.

16. Wykaż, że liczba $E(XY)$ nie zawsze jest równa liczbie $E(X) \cdot E(Y)$.

17. Wykaż, że $V(X + c) = V(X)$ i $V(cX) = c^2 \cdot V(X)$, gdzie c należy do $\mathbb{R}$, a X jest zmienną losową o skończonym zbiorze wartości.

18. Niech X będzie zmienną losową o wartości średniej μ i odchyleniu standardowym σ. Znajdź wartość średnią i odchylenie standardowe zmiennej $-X$.

19. (a) Załóżmy, że $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi, z których każda ma wartość średnią μ i odchylenie standardowe σ. Znajdź wartość średnią i odchylenie standardowe zmiennej $S = X_1 + X_2 + \ldots + X_n$.

(b) Zrób to samo co w części (a) tego ćwiczenia dla średniej arytmetycznej $\frac{1}{n}S = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$.

20. (a) Wykaż, że jeśli zmienne losowe X i Y na skończonej przestrzeni zdarzeń elementarnych Ω spełniają warunek $X(\omega) \leq Y(\omega)$ dla wszystkich elementów $\omega \in \Omega$, to $E(X) \leq E(Y)$.

(b) Wykaż, że jeśli $|X(\omega) - \mu| \leq |X(\omega_0) - \mu|$ dla wszystkich $\omega \in \Omega$, to $|X(\omega_0) - \mu| \geq \sigma_X$. To znaczy, że największa wartość odchylenia $|X - \mu|$ jest większa lub równa standardowemu odchyleniu zmiennej X.

21. Niech $X_1, X_2, \ldots, X_n$ będą niezależnymi zmiennymi losowymi o skończonych zbiorach wartości.

(a) Wykaż, że zmienne $X_1 + X_2, X_3, \ldots, X_n$ są niezależne.

(b) Skorzystaj z twierdzenia 7 dla $n = 2$ i z części (a) tego ćwiczenia do przeprowadzenia indukcyjnego dowodu twierdzenia 7.

§ 9.4. Rozkład dwumianowy i inne rozkłady z nim związane

Punktem wyjścia do rozpatrywania ogólnego, mającego liczne zastosowania, rozkładu dwumianowego jest następujący model. Wyobraźmy sobie eksperyment, którego tylko jeden rezultat, spośród wielu możliwych, nas interesuje. Jest on tradycyjnie zwany **sukcesem**; zdarzenie doń przeciwne nazywamy **porażką**. Zakładamy, że $P(\text{sukces}) = p$, dla pewnej liczby p, gdzie $0 < p < 1$. Niech $q = P(\text{porażka})$, skąd mamy $p+q = 1$. Zakładamy ponadto, że nasz eksperyment jest powtarzany wielokrotnie, powiedzmy n razy, oraz że wyniki różnych eksperymentów są wzajemnie niezależne; sukces pierwszego eksperymentu nie zmienia szansy na sukces w drugim eksperymencie i tak dalej.

PRZYKŁAD 1 (a) Rzucanie n razy symetryczną monetą można traktować jak n-krotne powtarzanie eksperymentu, w którym wypadnięcie orła uważamy za sukces, a reszki — za porażkę. Wówczas $p =$

$P(\text{sukces}) = \frac{1}{2}$. Wiemy już, że

$$P(k \text{ orłów w } n \text{ rzutach}) = \frac{1}{2^n} \cdot \binom{n}{k}$$

dla $k = 0, 1, \ldots, n$. Ogólnie,

$$P(k \text{ sukcesów w } n \text{ eksperymentach}) = \frac{1}{2^n} \cdot \binom{n}{k},$$

o ile $p = \frac{1}{2}$.

(b) Monetę nazywamy **niesymetryczną**, jeśli prawdopodobieństwo p wypadnięcia orła jest różne od $\frac{1}{2}$. Większość monet jest troszkę niesymetrycznych. Możemy nadal pytać o prawdopodobieństwo wypadnięcia k orłów w n rzutach, ale wzór z części (a) nie będzie już prawdziwy. ∎

PRZYKŁAD 2 (a) Pewne elektroniczne urządzenie ma n części, a prawdopodobieństwo, że dana część zepsuje się przed upływem okresu gwarancji, wynosi q. Zatem sukces takiej części polega na tym, że działa ona aż do momentu wygaśnięcia gwarancji (jeśli zepsuje się następnego dnia, będziemy to nadal uważali za sukces!). Jeśli przyjmiemy, że $p = 1 - q$, to opisana sytuacja pasuje do naszego schematu mimo, że n „eksperymentów" przebiega jednocześnie. Tak jak w ćwiczeniu 21 z § 9.1 zakładamy, że części działają dobrze niezależnie od siebie, co może być rozsądnym założeniem lub nie. Poza przypadkiem $p = \frac{1}{2}$, nie znamy jeszcze wzoru na

$$P(k \text{ sukcesów w } n \text{ eksperymentach}),$$

to znaczy na

$$P(k \text{ części działa w momencie wygaśnięcia gwarancji}).$$

(b) Jeśli prawdopodobieństwo wyleczenia pewnej choroby wynosi p i jeśli kuracją obejmuje się n osób chorych na tę chorobę, to zbiór takich „eksperymentów" pasuje do naszego ogólnego modelu. Ponownie zakładamy, że szanse poszczególnych pacjentów na wyleczenie są wzajemnie niezależne. ∎

Dla danych liczb p i n obliczymy teraz prawdopodobieństwo uzyskania k sukcesów w n niezależnych eksperymentach, gdzie p jest prawdopodobieństwem sukcesu w każdym z nich. Tutaj $k = 0, 1, \ldots, n$. Przestrzeń zdarzeń elementarnych Ω składa się ze wszystkich n-wyrazowych ciągów złożonych z liter S i P (oznaczających sukces bądź porażkę). Jest 2^n takich ciągów, ale nie są one jednakowo prawdopodobne. Porównaj prawdopodobieństwo

uzyskania ciągu złożonego z samych liter S i prawdopodobieństwo uzyskania ciągu złożonego z samych liter P, jeśli $p = 0{,}001$. Dla ustalonego k, jest $\binom{n}{k}$ ciągów długości n, w których litera S występuje dokładnie k razy i okazuje się, że takie ciągi są jednakowo prawdopodobne. Dla zilustrowania tego, rozważmy pewien szczególny przypadek.

PRZYKŁAD 3 Dla dowolnego p oraz dla $n = 5$ znajdziemy prawdopodobieństwo dokładnie 3 sukcesów. Ustalmy najpierw jakiś ciąg złożony z 5 liter S i P, w którym litera S występuje dokładnie 3 razy, na przykład (S,S,P,S,P). Prawdopodobieństwo takiego właśnie wyniku wynosi

$$P(\text{pierwszym wyrazem jest S}) \cdot P(\text{drugim wyrazem jest S}) \cdot P(\text{trzecim wyrazem jest P}) \cdot P(\text{czwartym wyrazem jest S}) \cdot P(\text{piątym wyrazem jest P})$$
$$p \cdot p \cdot q \cdot p \cdot q = p^3 q^2.$$

Dla innego takiego ciągu kolejność czynników będzie inna, ale wynik p^3q^2 się nie zmieni, gdyż prawdopodobieństwo każdego z trzech sukcesów wynosi p, a każdej z dwóch porażek — q. Mamy zatem $\binom{5}{3}$ wyników, każdy o prawdopodobieństwie p^3q^2, skąd otrzymujemy

$$P(3 \text{ sukcesy w } 5 \text{ eksperymentach}) = \binom{5}{3} p^3 q^2. \quad \blacksquare$$

Dokładnie to samo rozumowanie odnosi się do ogólnego przypadku; zatem

$$P(k \text{ sukcesów w } n \text{ eksperymentach}) = \binom{n}{k} p^k q^{n-k}.$$

Ogólnie, **rozkładem dwumianowym** nazywamy rozkład prawdopodobieństwa określony wzorem

$$f(k) = \binom{n}{k} p^k q^{n-k}, \qquad k = 0, 1, \ldots, n.$$

(Zakłada się, że $f(x) = 0$ dla wszystkich innych wartości x). Zauważmy, że każda para dopuszczalnych wartości parametrów p i n określa pewien rozkład dwumianowy.

Jeśli $p = \frac{1}{2}$, tak jak w przypadku eksperymentu polegającego na n rzutach symetryczną monetą, to

$$f(k) = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \frac{1}{2^n} \cdot \binom{n}{k}.$$

Wzór ten jest zgodny z wynikiem, który uzyskaliśmy wcześniej.

Dystrybuanta F odpowiadająca rozkładowi f jest określona wzorem

$$F(y) = \sum_{k \le y} f(k) \quad \text{dla} \quad y \in \mathbb{R}.$$

Tę dystrybuantę nazywamy **dystrybuantą rozkładu dwumianowego**. Ponieważ funkcja F jest stała między punktami, w których zmienia skokowo swą wartość, to jest ona w zupełności wyznaczona przez swe wartości dla $k = 0, 1, \ldots, n$. Ponadto $f(0) = F(0)$ i $f(k) = F(k) - F(k-1)$ dla $k = 1, 2, \ldots, n$, a więc funkcję f można łatwo określić za pomocą F.

Dla zilustrowania tych pojęć weźmy $n = 10$ i rozważmy przypadki, gdy $p = \frac{1}{2}$, $p = \frac{1}{3}$ i $p = \frac{1}{10}$. Tablica 9.6 podaje wartości dystrybuanty rozkładu dwumianowego dla tak wybranych liczb n i p. Tablice zawierające o wiele więcej informacji znaleźć można w książkach z rachunku prawdopodobieństwa i statystyki.

Tablica 9.6 Dystrybuanta rozkładu dwumianowego F dla $n = 10$

p \ k	0	1	2	3	4	5	6	7	8	9	10
1/2	0,001	0,011	0,055	0,172	0,377	0,623	0,828	0,945	0,989	0,999	1,00
1/3	0,017	0,104	0,299	0,559	0,787	0,923	0,980	0,997	1,00	1,00	1,00
1/10	0,349	0,736	0,930	0,987	0,998	1,00	1,00	1,00	1,00	1,00	1,00

Dystrybuanta F najlepiej nadaje się do przeprowadzania obliczeń; dla nabrania wyobrażenia o rozpatrywanych prawdopodobieństwach przedstawiliśmy na rysunku 9.7 wykresy danych rozkładów prawdopodobieństwa (jednostki na osiach poziomej i pionowej są różne).

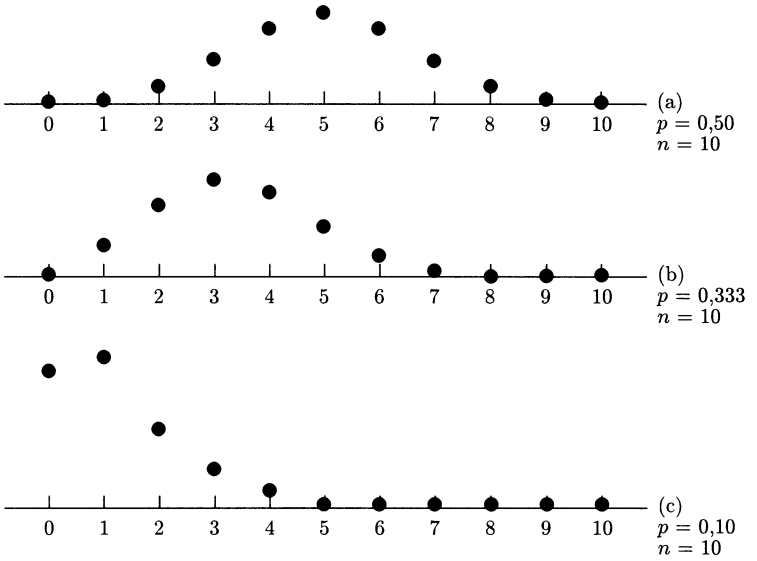
PRZYKŁAD 4 Rzucamy 10 razy monetą symetryczną. Wtedy

$$P(k \text{ orłów w 10 rzutach}) = \frac{1}{2^{10}} \cdot \binom{10}{k} \quad \text{dla} \quad k = 0, 1, \ldots, 10.$$

Korzystając z tablicy 9.6 można szybko obliczyć wiele prawdopodobieństw. I tak na przykład,

$$P(\text{najwyżej 5 orłów}) = F(5) \approx 0{,}623,$$
$$P(\text{dokładnie 7 orłów}) = F(7) - F(6) \approx 0{,}945 - 0{,}828 = 0{,}117,$$
$$P(3 \le \text{liczba orłów} \le 7) = F(7) - F(2) \approx 0{,}945 - 0{,}055 = 0{,}890.$$

(b) Załóżmy, że mamy do czynienia z dość niesymetryczną monetą, dla której prawdopodobieństwo wypadnięcia orła w każ-



Rysunek 9.7

dym z rzutów wynosi $\frac{1}{3}$. Wówczas

$$P(k \text{ orłów w 10 rzutach}) = \binom{10}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{10-k}.$$

Ponownie możemy skorzystać z tablicy 9.6, żeby obliczyć niektóre prawdopodobieństwa:

$$P(\text{co najmniej 5 orłów}) = F(10) - F(4) \approx 1 - 0{,}787 = 0{,}213,$$
$$P(3 \leq \text{liczba orłów} \leq 7) = F(7) - F(2) \approx 0{,}997 - 0{,}299 = 0{,}698.$$

Oczekiwana liczba orłów wynosi $np = 3{,}33$, jak wyniknie z twierdzenia 1. Tak więc można by oczekiwać, że liczba orłów będzie bliska 3. Istotnie,

$$P(\text{dokładnie 3 orły}) = F(3) - F(2) \approx 0{,}599 - 0{,}299 = 0{,}260$$

oraz

$$P(2 \leq \text{liczba orłów} \leq 5) = F(5) - F(1) \approx 0{,}923 - 0{,}104 = 0{,}819.$$

(c) Nawet jeśli wiele pojedynczych wyników naszego wyjściowego eksperymentu jest dla nas interesujących, to możemy skoncentrować uwagę na jakimś wybranym zdarzeniu i określić je jako „sukces". Na przykład, przypuśćmy, że rzucamy n razy symetryczną kostką i chcemy wiedzieć, jak często liczba wyrzuconych oczek jest równa 1 lub 2. Nazywamy więc to zdarzenie sukcesem

i mamy $P(\text{sukces}) = \frac{1}{3} = p$. Zatem

$$P(k \text{ sukcesów w } n \text{ eksperymentach}) = \binom{n}{k} \cdot \left(\frac{1}{3}\right)^k \cdot \left(\frac{2}{3}\right)^{n-k}.$$

Jeśli rzucamy kostką 10 razy, to możemy skorzystać z tablicy 9.6. I tak na przykład prawdopodobieństwo wyrzucenia 1 lub 2 oczek co najmniej 3 razy wynosi

$$P(3 \leq \text{liczba sukcesów} \leq 10) = F(10) - F(2) \approx 1 - 0{,}299 = 0{,}701.$$

(d) Slim Hulk jest koszykarzem, który wykorzystuje $\frac{2}{3}$ swoich rzutów wolnych. Jeśli jego sukcesy w trafianiu do kosza są od siebie niezależne, to ile wynosi prawdopodobieństwo, że wykorzysta on co najmniej 7 spośród 10 następnych rzutów wolnych? Mamy tutaj $n = 10$ i $p = \frac{2}{3}$. Nie możemy bezpośrednio skorzystać z tablicy 9.6, ale uda się to nam, jeśli (ku przerażeniu Slima) nazwiemy nietrafienie do kosza sukcesem i obliczymy $P(\text{liczba rzutów nietrafionych} \leq 3)$. Mamy teraz $P(\text{brak trafienia}) = P(\text{sukces}) = \frac{1}{3}$; korzystamy więc z tablicy 9.6 dla $p = \frac{1}{3}$ i otrzymujemy

$$\begin{aligned} P(\text{co najmniej 7 trafień}) &= P(\text{liczba rzutów nietrafionych} \leq 3) \\ &= F(3) \approx 0{,}559. \end{aligned}$$

Choć Slim wykorzystuje tylko $\frac{2}{3}$ swoich rzutów wolnych, ma on więcej niż 50% szans na to, że w następnych 10 próbach trafi do kosza 7 lub więcej razy! ∎

PRZYKŁAD 5 Powróćmy do rozpatrywanej wcześniej sytuacji, w której mamy do czynienia z n częściami, z których każda psuje się z prawdopodobieństwem q przed upływem okresu gwarancji. Prawdopodobieństwo, że w tym czasie zepsuje się co najmniej jedna część, wynosi

$$1 - P(n \text{ sukcesów}) = 1 - \binom{n}{n} \cdot p^n = 1 - p^n = 1 - (1-q)^n.$$

Wynik ten jest zgodny z odpowiedzią do ćwiczenia 21 z § 9.1.

Prawdopodobieństwo zepsucia się co najwyżej dwu części wynosi

$$\begin{aligned} &P(n-2 \text{ sukcesy}) + P(n-1 \text{ sukcesów}) + P(n \text{ sukcesów}) = \\ &\quad = \binom{n}{n-2} \cdot p^{n-2}q^2 + \binom{n}{n-1} \cdot p^{n-1}q + \binom{n}{n} \cdot p^n \\ &\quad = \frac{n}{2}(n-1) \cdot p^{n-2}q^2 + n \cdot p^{n-1}q + p^n. \end{aligned}$$

Jeśli sukces i porażkę zamienimy rolami, to otrzymamy

$$P(2 \text{ porażki}) + P(1 \text{ porażka}) + P(0 \text{ porażek}) = \\ = \binom{n}{2} \cdot q^2 p^{n-2} + \binom{n}{1} \cdot qp^{n-1} + \binom{n}{n} \cdot p^n,$$

co oczywiście jest tą samą liczbą. ∎

Obliczona w następnym twierdzeniu wartość oczekiwana nie powinna stanowić niespodzianki.

Twierdzenie 1

Jeśli S_n jest zmienną losową o rozkładzie dwumianowym dla pewnych parametrów n i p, to

$$E(S_n) = np \quad \text{oraz} \quad V(S_n) = npq.$$

Odchylenie standardowe tej zmiennej wynosi $\sqrt{npq}$.

Dowód. Możemy założyć, że zmienna S_n podaje liczbę sukcesów uzyskiwanych w n niezależnych eksperymentach, przy czym dla każdego z nich $P(\text{sukces}) = p$. Tak jak w przypadku, gdy $p = \frac{1}{2}$, omówionym w § 9.3, nie byłoby najrozsądniej próbować obliczać $E(S_n)$ i $V(S_n)$ wprost z definicji.

Dla każdego $i = 1, 2, \ldots, n$ połóżmy $X_i = 1$, jeśli i-ty eksperyment kończy się sukcesem i $X_i = 0$ — w przeciwnym przypadku. Wówczas $P(X_i = 1) = p$, $P(X_i = 0) = q$ oraz

$$E(X_i) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = p.$$

Mamy też $E(X_i^2) = p$, a więc

$$V(X_i) = E(X_i^2) - [E(X_i)]^2 = p - p^2 = p(1-p) = pq.$$

Jasne jest, że $S_n = X_1 + X_2 + \ldots + X_n$; stąd

$$E(S_n) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} p = np.$$

Ponieważ zmienne X_i są wzajemnie niezależne, to z twierdzenia 7 z § 9.3 wynika, że

$$V(S_n) = \sum_{i=1}^{n} V(X_i) = \sum_{i=1}^{n} pq = npq. \qquad \blacksquare$$

W paragrafie 9.3 omawialiśmy „czas oczekiwania" na pojawienie się pierwszego orła w ciągu kolejnych rzutów symetryczną monetą. Możemy rozważać tę samą zmienną losową również wtedy,

gdy prawdopodobieństwo p, gdzie $0 < p < 1$, wypadnięcia orła w jednym rzucie jest różne od $\frac{1}{2}$. (Prawdopodobieństwo, które obliczymy, będzie się również odnosić do zagadnienia oczekiwania na pierwszy w ciągu niezależnych eksperymentów, eksperyment zakończony sukcesem). Niech W będzie zmienną losową, która podaje, w którym z kolejnych rzutów po raz pierwszy wypadł orzeł. Wówczas $P(W = 1) = p$. Jeśli $W = 2$, to w pierwszym rzucie wypadła reszka, a w drugim — orzeł; zatem $P(W = 2) = qp$. Jeśli $W = 73$, to w pierwszych 72 rzutach wypadła reszka, a w 73 rzucie — orzeł; zatem $P(W = 73) = q^{72}p$. Ogólnie,

$$f(k) = P(W = k) = p \cdot q^{k-1} \quad \text{dla} \quad k = 1, 2, \ldots$$

Tę funkcję f nazywamy **rozkładem geometrycznym**. Zmienna losowa W jest dyskretna; jej zbiorem wartości jest $\{1, 2, \ldots\}$. Jej wartość oczekiwaną można wyrazić za pomocą nieskończonego szeregu

$$\sum_{k=1}^{\infty} k \cdot p \cdot q^{k-1};$$

okazuje się, że suma tego szeregu jest równa $1/p$. Zatem $E(W) = 1/p$; porównaj ten wynik z przykładem 5(a) z § 9.3, gdzie $p = \frac{1}{2}$.

PRZYKŁAD 6 (a) Przypuśćmy, że wiemy iż prawdopodobieństwo tego, że pewien element (na przykład część urządzenia elektronicznego lub żarówka) przepali się w ciągu godziny, wynosi q. Jeśli liczba q jest mała, to dla otrzymania przybliżonego modelu sytuacji uzasadnione jest przyjęcie, że zdarzenia „element przepali się w ciągu k-tej godziny" są wzajemnie niezależne. Wówczas prawdopodobieństwo, że element przepali się w ciągu k-tej godziny wynosi $p^{k-1}q$, gdzie $p = 1 - q$. Spodziewany czas oczekiwania na jego przepalenie się jest równy $1/q$. Zatem jeśli $q = 0{,}01$, to oczekiwany czas działania tego elementu wynosi 100 godzin. Innymi słowy, w przypadku dużej liczby elementów, średnia liczba godzin funkcjonowania elementu będzie bliska 100.

(b) Powróćmy do naszego elektronicznego urządzenia, mającego n części, z których każda psuje się z prawdopodobieństwem q przed upływem okresu gwarancji. Ponieważ wszystkie części działają jednocześnie, to zmienna losowa czasu oczekiwania i jej rozkład geometryczny (z parametrem q) nie znajdują tu zastosowania! Można by pytać o czas oczekiwania na moment, kiedy po raz pierwszy nastąpi awaria którejś z części, ale jest to całkiem inne zagadnienie i nie mamy ani dostatecznych danych, ani metod do jego rozwiązania. ∎

Czasami, mimo że może nie życzylibyśmy sobie tego, rozpatrywanie przyjemnych skończonych dyskretnych obiektów prowadzi do bardziej skomplikowanych obiektów teoretycznych. Tak właśnie jest w przypadku rozkładu dwumianowego. W zastosowaniach liczba n eksperymentów bądź zdarzeń elementarnych bywa duża i obliczenia związane z rozkładem dwumianowym stają się trudne. Z pewnością byłoby wygodne, gdyby dla dużych wartości n dystrybuanty F_n rozkładów dwumianowych były bliskie dystrybuanty jakiejś jednej zmiennej losowej mającej dobre własności. (Później sprecyzujemy, co mamy tu na myśli używając słowa „bliskie"). W tak sformułowanej postaci nie jest to jednak prawdą. Wartości średnie i odchylenia standardowe zmiennych losowych związanych z tymi rozkładami (odpowiednio: np i $\sqrt{npq}$) rosną wraz ze wrostem wartości n, a więc dystrybuanty F_n przesuwają się coraz bardziej na prawo. I tak na przykład, jeśli $n = 1000\,000$ i $p = \frac{1}{2}$, to wartość średnia dla rozkładu dwumianowego wynosi 500 000 i jego wykres jest symetryczny względem prostej $x = 500\,000$. Odpowiadająca mu dystrybuanta nie osiąga wartości $\frac{1}{2}$ zanim $x = 500\,000$.

Pojawia się przypuszczenie, że jeśli dystrybuanty F_n dla dużych wartości n są blisko siebie, to także ich wartości średnie i odchylenia standardowe niewiele się od siebie różnią. Ta obserwacja prowadzi do następującego pojęcia. Dla danej zmiennej losowej X o wartości średniej μ i odchyleniu standardowym $\sigma > 0$, **zmienną unormowaną** odpowiadającą zmiennej X nazywamy zmienną losową zdefiniowaną wzorem $\widetilde{X} = (X - \mu)/\sigma$. Ta nowa zmienna losowa różni się od zmiennej X jedynie stałym czynnikiem $1/\sigma$ i pewną stałą addytywną, więc informację dotyczącą $\widetilde{X}$ łatwo jest przekształcić w informację odnoszącą się do X. Ładne własności zmiennej $\widetilde{X}$ sformułowane są w punkcie (a) następnego twierdzenia. Pozostała jego część ukazuje prosty związek między dystrybuantami zmiennych X i $\widetilde{X}$.

Twierdzenie 2

Niech X będzie zmienną losową o wartości średniej μ, odchyleniu standardowym $\sigma > 0$ i dystrybuancie F. Niech $\widetilde{X}$ będzie zmienną unormowaną odpowiadającą zmiennej X i niech $\widetilde{F}$ będzie dystrybuantą zmiennej $\widetilde{X}$.

(a) $E(\widetilde{X}) = 0$, $V(\widetilde{X}) = 1$ i $\sigma_{\widetilde{X}} = 1$.

(b) $F(y) = \widetilde{F}((y - \mu)/\sigma)$ dla $y \in \mathbb{R}$.

(c) $\widetilde{F}(y) = F(\sigma y + \mu)$ dla $y \in \mathbb{R}$.

Dowód. (a) $E(\widetilde{X})=\frac{1}{\sigma}E(X-\mu)=\frac{1}{\sigma}[E(X)-\mu]=\frac{1}{\sigma}[\mu-\mu]=0$. Ćwiczenie 17 z § 9.3 pokazuje, że ogólnie mamy $V(X+c)=V(X)$ i $V(cX)=c^2V(X)$. Zatem

$$V(\widetilde{X})=V\left(\frac{X}{\sigma}\right)=\frac{1}{\sigma^2}V(X),$$

a ponieważ $V(X)=\sigma^2$, to otrzymujemy $V(\widetilde{X})=1$.

(b) Z definicji mamy $F(y)=P(X\leq y)$. Dalej

$$X(\omega)\leq y \Leftrightarrow X(\omega)-\mu\leq y-\mu \Leftrightarrow \frac{X(\omega)-\mu}{\sigma}\leq\frac{y-\mu}{\sigma},$$

a więc

$$\{X\leq y\}=\left\{\frac{X(\omega)-\mu}{\sigma}\leq\frac{y-\mu}{\sigma}\right\}=\left\{\widetilde{X}\leq\frac{y-\mu}{\sigma}\right\}.$$

Wynika stąd, że

$$F(y)=P\left(\widetilde{X}\leq\frac{y-\mu}{\sigma}\right)=\widetilde{F}\left(\frac{y-\mu}{\sigma}\right).$$

(c) W punkcie (b) wystarczy zastąpić y sumą $\sigma y+\mu$. ∎

PRZYKŁAD 7 Niech S_n będzie zmienną losową o rozkładzie dwumianowym z parametrami n i p, gdzie p jest pewną ustaloną liczbą taką, że $0<p<1$. Odpowiadającą jej zmienną unormowaną określa wzór

$$\widetilde{S}_n=\frac{S_n-np}{\sqrt{npq}}.$$ ∎

Ponieważ wszystkie zmienne losowe $\widetilde{S}_n$ mają te same wartości średnie i odchylenia standardowe, to można mieć nadzieją, że ich dystrybuanty są dla dużych wartości n bliskie pewnej jednej ustalonej dystrybuanty. Jedno z głównych twierdzeń rachunku prawdopodobieństwa głosi, że istnieje dystrybuanta Φ, którą nazywamy **dystrybuantą rozkładu normalnego** lub **rozkładu Gaussa** taka, że

(*) $\widetilde{F}_n(y)\approx\Phi(y)$ dla dużych wartości n oraz dla $y\in\mathbb{R}$.

Dystrybuanta Φ nie zależy od parametru p; bez względu na to, ile wynosi liczba p, dystrybuanty $\widetilde{F}_n$ dla dużych wartości n są bliskie Φ. Ten zadziwiający rezultat i pewne jego silne uogólnienia znane są jako „centralne twierdzenie graniczne". Jego dowód podany jest w bardziej zaawansowanych podręcznikach.

Zanim wyjaśnimy, czym jest dystrybuanta Φ, pokażemy, jak ten rezultat można stosować. Najpierw jednak zauważmy, że

stwierdzenie (*) jest bardzo nieprecyzyjne. Jak dobre przybliżenie można uzyskać? Jak duża musi być wartość n? Te subtelne kwestie pozostawiamy statystykom.

PRZYKŁAD 8 Przypuśćmy, że pewien eksperyment powtarzany jest $n = 10\,000$ razy oraz że za każdym razem $P(\text{sukces}) = 0,1 = p$. Oczekiwana liczba sukcesów wynosi $\mu = np = 1000$, a chcielibyśmy znać

$$P(950 \leq \text{ liczba sukcesów } \leq 1050).$$

Liczba ta jest równa $F_n(1050) - F_n(949)$. Ponieważ

$$\sigma = \sqrt{npq} = \sqrt{10000 \cdot \frac{1}{10} \cdot \frac{9}{10}} = 30,$$

to mamy

$$F_n(1050) = \widetilde{F}_n\left(\frac{1050 - 1000}{30}\right) = \widetilde{F}_n\left(\frac{5}{3}\right) \approx \Phi(1,667) \approx 0,952$$

oraz

$$\begin{aligned} F_n(949) &= \widetilde{F}_n\left(\frac{949 - 1000}{30}\right) \\ &= \widetilde{F}_n\Big(-1,700\Big) \approx \Phi(-1,700) \approx 0,048. \end{aligned}$$

Przybliżone wartości funkcji Φ zostały wzięte z tablicy jej wartości. Ostatecznie otrzymujemy

$$P(950 \leq \text{ liczba sukcesów } \leq 1050) \approx 0,904.$$

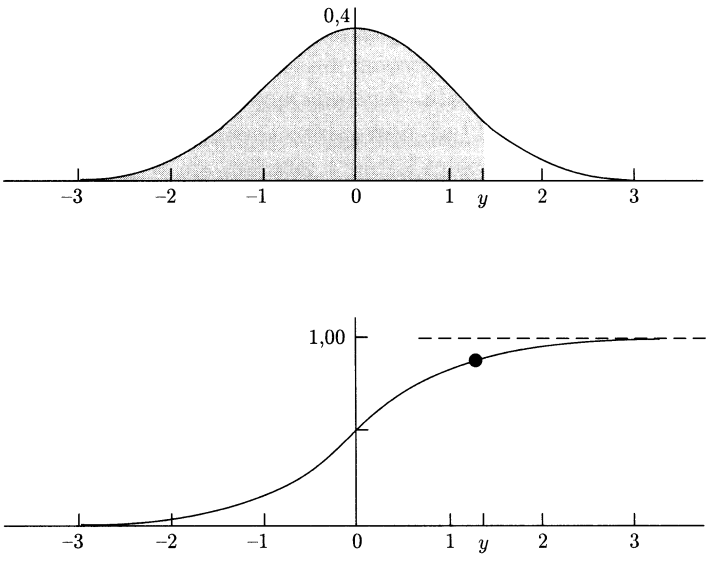
Zwróć uwagę na siłę centralnego twierdzenia granicznego. Dla dużych wartości n wystarczy, że dobrze znamy własności jednej tylko dystrybuanty. Jeśli chodzi o obliczenia, znaczy to, że potrzebujemy jednej tylko tablicy wartości bądź też jednego klawisza funkcyjnego na kalkulatorze. ∎

Odpowiedzmy w końcu na pytanie, czym jest funkcja Φ. Rysunek 9.8(a) przedstawia tzw. **krzywą dzwonową**. Jest ona wykresem funkcji danej wzorem

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2},$$

gdzie $\mathrm{e} \approx 2,71828$. Obszar ograniczony tą krzywą i osią x ma pole równe 1. Dla każdego $y \in \mathbb{R}$

$$\Phi(y) = \text{pole obszaru pod wykresem } \varphi \text{ na lewo od } y;$$



Dystrybuanta rozkładu normalnego Φ

Rysunek 9.8

zobacz rysunek 9.8(b). Dla wartości y, zaznaczonej na rysunku 9.8(a), chodzi tu o pole zacieniowanego obszaru. Studenci analizy matematycznej napisaliby, że

$$\Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \mathrm{e}^{-x^2/2}\,\mathrm{d}x,$$

i mieliby nadzieję, że całkę tę można wyrazić jakimś prostym wzorem. Niestety nie ma prostego wzoru na $\Phi(y)$, ale wartości funkcji Φ można przybliżać z dowolną dokładnością i znaleźć je można za pomocą tablic matematycznych lub kalkulatora. Na koniec zauważmy, że jedna z wartości jest oczywista: $\Phi(0) = 0{,}500$, ponieważ połowa obszaru ograniczonego wykresem funkcji φ leży na lewo od liczby 0.

Wiele różnych zagadnień rachunku prawdopodobieństwa i statystyki prowadzi do rozkładu normalnego. Okazuje się, że dające się obserwować i mierzyć wielkości takie, jak wzrost wszystkich studentów czwartego roku bądź wszystkich dorosłych kobiet mają rozkład zbliżony do rozkładu normalnego. Znaczy to, że odpowiednie unormowane zmienne losowe w przybliżeniu mają rozkład normalny.

PRZYKŁAD 9 (a) Bierzemy losową próbę dorosłych kobiet i mierzymy ich wzrost. Mamy więc do czynienia ze zmienną losową X okre-

śloną na danym zbiorze kobiet. Stwierdzamy, że jej wartość średnia μ wynosi 66 cali, a odchylenie standardowe σ jest równe 2,5 cala. Zakładając, że ta zmienna losowa ma w przybliżeniu rozkład normalny, oszacujemy prawdopodobieństwo, że wzrost dorosłej kobiety zawiera się w granicach 61 i 71 cali. Pominiemy kobiety, których wzrost wynosi dokładnie 61 cali i oszacujemy $P(61 < X \leq 71)$. Liczba ta jest równa $F(71) - F(61)$, gdzie F jest dystrybuantą zmiennej X. Jeśli $\widetilde{F}$ jest dystrybuantą unormowanej zmiennej losowej $\widetilde{X}$ odpowiadającej zmiennej X, to z twierdzenia 2(b) wynika, że

$$F(71) - F(61) = \widetilde{F}\left(\frac{71-66}{2{,}5}\right) - \widetilde{F}\left(\frac{61-66}{2{,}5}\right) = \widetilde{F}(2) - \widetilde{F}(-2).$$

Na mocy założenia dystrybuanta $\widetilde{F}$ jest w przybliżeniu równa dystrybuancie rozkładu normalnego Φ, a więc

$$P(61 < X \leq 71) \approx \Phi(2) - \Phi(-2).$$

Korzystając z tablic matematycznych stwierdzamy, że ostatnia wartość jest w przybliżeniu równa $0{,}9773 - 0{,}0227 = 0{,}9546$.

(b) Obliczenia z części (a) odnoszą się do ogólniejszej sytuacji. Niech X będzie zmienną losową o rozkładzie normalnym; Φ jest więc dystrybuantą zmiennej $\widetilde{X}$. Jeśli μ i σ oznaczają, odpowiednio, wartość średnią i odchylenie standardowe zmiennej X, to

$$\begin{aligned} P(\mu - 2\sigma < X \leq \mu + 2\sigma) &= F(\mu + 2\sigma) - F(\mu - 2\sigma) \\ &= \Phi\left(\frac{\mu + 2\sigma - \mu}{\sigma}\right) - \Phi\left(\frac{\mu - 2\sigma - \mu}{\sigma}\right) \\ &= \Phi(2) - \Phi(-2) \approx 0{,}9546. \end{aligned}$$

W części (a), $\mu = 66$ i $\sigma = 2{,}5$. Probabiliści powiedzieliby, że „prawdopodobieństwo, że wartości zmiennej losowej o rozkładzie normalnym odchylają się od wartości średniej co najwyżej o podwojenie odchylenia standardowego jest w przybliżeniu równe 0,95."

(c) Prawdopodobieństwo, że wartości zmiennej losowej o rozkładzie normalnym różnią od wartości średniej co najwyżej o odchylenie standardowe jest w przybliżeniu równe

$$\Phi(1) - \Phi(-1) \approx 0{,}8413 - 0{,}1587 = 0{,}6826.$$

Podobnie, prawdopodobieństwo, że wartości takiej zmiennej losowej odchylają się od wartości średniej co najwyżej o potrojenie odchylenia standardowego, jest w przybliżeniu równe

$$\Phi(3) - \Phi(-3) \approx 0{,}9987 - 0{,}0013 = 0{,}9974. \qquad \blacksquare$$

ĆWICZENIA DO § 9.4

1. W którym miejscu w przykładzie 3 wykorzystany został fakt, że wyniki kolejnych eksperymentów są wzajemnie niezależne?

2. Pewien baseballista ma średnią uderzeń równą 0,333. Znaczy to, że w $\frac{1}{3}$ prób jego uderzenia są skuteczne. Zakładamy, że próby są niezależne.
   (a) Jaka jest wartość oczekiwana liczby skutecznych uderzeń w przypadku 3 prób?
   (b) Jakie jest prawdopodobieństwo, że w kolejnych trzech próbach nasz baseballista choć raz uderzy skutecznie?

3. Przypuśćmy, że baseballista z ćwiczenia 2 ma przed sobą 10 prób.
   (a) Jaka jest wartość oczekiwana liczby skutecznych uderzeń?
   (b) Jakie jest prawdopodobieństwo, że uderzy on skutecznie co najwyżej 3 razy? Wskazówka: wykorzystaj tablicę 9.6.
   (c) Jakie jest prawdopodobieństwo, że uderzy on skutecznie co najmniej 3 razy?

4. Połowa klientów baru Burger Queen zamawia frytki. Załóżmy, że zamówienia są od siebie niezależne.
   (a) Jakie jest prawdopodobieństwo, że dokładnie pięcioro spośród następnych dziesięciorga klientów zamówi frytki?
   (b) Jakie jest prawdopodobieństwo, że co najmniej troje spośród następnych dziesięciorga klientów zamówi frytki?

5. Pewne urządzenie elektroniczne składa się z 10 części. Prawdopodobieństwo przepalenia się w ciągu następnego roku dla każdej z nich wynosi 0,10. Załóżmy, że części przepalają się niezależnie od siebie nawzajem.
   (a) Jakie jest prawdopodobieństwo, że w ciągu następnego roku nie przepali się żadna z części?
   (b) Jakie jest prawdopodobieństwo, że w ciągu następnego roku przepalą się co najwyżej dwie części?

6. Bierzemy losową próbę studentów płci męskiej i mierzymy ich wzrost. Wartość średnia μ uzyskanej w ten sposób zmiennej losowej wynosi 69 cali, a odchylenie standardowe σ jest równe 3 cale.
   (a) Określ w przybliżeniu prawdopodobieństwo, że losowo wybrany student płci męskiej będzie miał między 66 a 72 cale wzrostu. *Wskazówka*: wykorzystaj przybliżone obliczenia związane z rozkładem normalnym zawarte w przykładzie 9.
   (b) Określ w przybliżeniu prawdopodobieństwo, że losowo wybrany student płci męskiej będzie miał między 63 a 75 cali wzrostu.
   (c) Określ w przybliżeniu prawdopodobieństwo, że losowo wybrany student płci męskiej będzie miał między 60 a 78 cali wzrostu.

7. (a) Jakie jest prawdopodobieństwo, że przy założeniach z ćwiczenia 6 losowo wybrany student płci męskiej będzie miał co najmniej 72 cale wzrostu?

(b) Ile wynosi prawdopodobieństwo, że ma on mniej niż 63 cale wzrostu?

8. Pewien eksperyment powtarzany jest 30 000 razy, przy czym prawdopodobieństwo sukcesu wynosi za każdym razem $\frac{1}{4}$. Oczekiwana liczba sukcesów wynosi więc 7500. Określ w przybliżeniu, posługując się dystrybuantą Φ rozkładu normalnego, prawdopodobieństwo, że liczba X sukcesów będzie należeć do przedziału $(7400, 7600]$.

9. Pewien eksperyment powtarzany jest 1800 razy, przy czym prawdopodobieństwo sukcesu wynosi za każdym razem $\frac{1}{3}$.
(a) Ile wynosi oczekiwana liczba sukcesów?
(b) Ile wynosi odchylenie standardowe?
(c) Przypomnij sobie z przykładu 9, że $\Phi(2)-\Phi(-2) \approx 0,9546$. Znajdź taki przedział, że możemy być pewni na około 95%, że liczba sukcesów będzie do niego należeć.

10. Na duże wieczorne przyjęcie wysłanych zostaje 1000 zaproszeń. Szanse przyjścia na przyjęcie każdej poszczególnej zaproszonej osoby gospodarz ocenia na 60% wierząc jednocześnie, że decyzje zaproszonych są wzajemnie niezależne. Na przyjście ilu osób powinien on być przygotowany, jeśli chce być pewny na 97%, że ma wystarczająco dużo miejsc przy stole?

11. Rzucamy 1000 razy symetryczną monetą.
(a) Określ w przybliżeniu, posługując się rozkładem normalnym, prawdopodobieństwo, że orzeł wypadnie w 49 do 51 procent wszystkich rzutów. *Wskazówka:* Określ w przybliżeniu prawdopodobieństwo, że liczba X wszystkich orłów należy do przedziału $(490, 510]$; zob. przykład 9.
(b) Powtórz część (a) dla przypadku rzucania monetą 10 000 razy.
(c) Powtórz część (a) dla przypadku rzucania monetą 1 000 000 razy.

12. Podczas testowego egzaminu państwowego, w którym wzięło udział 1 000 000 studentów, w jednym z pytań spośród dwu możliwych odpowiedzi (tak, nie) tylko 51% zdających wybrało odpowiedź właściwą. Czy ktokolwiek ze zdających zrozumiał treść pytania? A może czystemu przypadkowi można przypisać, że było 51% poprawnych odpowiedzi?

13. Przeprowadzamy kolejno trzy niezależne eksperymenty. Prawdopodobieństwo sukcesu w pierwszym eksperymencie wynosi $\frac{1}{2}$, w drugim $\frac{1}{3}$, a w trzecim jest równe tylko $\frac{1}{4}$.
(a) Znajdź wartość oczekiwaną liczby sukcesów X.
(b) Znajdź odchylenie standardowe.
(c) Czy X jest zmienną losową o rozkładzie dwumianowym?

14. Niech X_1 i X_2 będą zmiennymi losowymi o wartościach średnich równych, odpowiednio, μ_1 i μ_2 i odchyleniach standardowych σ_1 i σ_2.

Przypuśćmy, że zmiennym X_1 i X_2 odpowiada ta sama zmienna unormowana $\widetilde{X}$. Znajdź wzór wyrażający dystrybuantę F_2 zmiennej X_2 za pomocą F_1, μ_1, μ_2, σ_1 i σ_2.

15. Niech Φ będzie dystrybuantą rozkładu normalnego.
(a) Uzasadnij, dlaczego $\Phi(y) + \Phi(-y) = 1$ dla $y \in \mathbb{R}$.
(b) Wykaż, że $\Phi(y) - \Phi(-y) = 2 \cdot \Phi(y) - 1$ dla $y \in \mathbb{R}$.

16. Niech X będzie zmienną losową o rozkładzie normalnym; niech μ będzie jej wartością średnią, a σ — odchyleniem standardowym.
(a) Wykaż, że $P(\mu - \sigma < X \leq \mu + \sigma) \approx \Phi(1) - \Phi(-1) \approx 0{,}6826$.
(b) Dla każdej liczby $c > 0$ wyraź $P(\mu - c\sigma < X \leq \mu + c\sigma)$ za pomocą Φ.

To, co jest najważniejsze w tym rozdziale

Jak zwykle: Co to znaczy? Dlaczego tutaj się znalazło? Jak mogę to zastosować? Myśl o przykładach. Wymyślaj nowe przykłady.

Pojęcia i oznaczenia

prawdopodobieństwo warunkowe, $P(E|S)$
zdarzenia niezależne, zdarzenia parami niezależne
zmienna losowa
  zbiór wartości
  dyskretna
  niezależne zmienne losowe
  rozkład prawdopodobieństwa zmiennej losowej, f_X
  dystrybuanta zmiennej losowej, F_X
  rozkład jednostajny
  wartość oczekiwana = wartość średnia , $E(X) = \mu$
  odchylenie standardowe, wariancja, $V(X) = \sigma^2$
rozkład dwumianowy, S_n
  sukces, p, porażka, q
rozkład geometryczny
zmienna unormowana, $\widetilde{X}$, $\widetilde{F}$
rozkład normalny = rozkład Gaussa, Φ

Fakty

Wzór na prawdopodobieństwo całkowite w odniesieniu do zdarzeń tworzących podział przestrzeni Ω.
Wzór Bayesa

$$P(A_j|B) = \frac{P(A_j) \cdot P(B|A_j)}{P(B)},$$

gdzie

$$P(B) = \sum_i P(A_i) \cdot P(B|A_i).$$

Jeśli zmienna losowa X jest dyskretna, to każdą z funkcji f_X i F_X można określić za pomocą drugiej z nich.

Jeśli wszystkie wyniki są jednakowo prawdopodobne, to $E(X)$ jest średnią wartości przyjmowanych przez zmienną X.

$E(X+Y) = E(X) + E(Y)$, $E(aX) = aE(X)$ i $E(a) = a$ dla $a \in \mathbb{R}$.

Jeśli zbiór $X(\Omega)$ jest skończony i $\varphi\colon \mathbb{R} \to \mathbb{R}$, to

$$E(\varphi \circ X) = \sum_x \varphi(x) \cdot P(X = x).$$

Twierdzenie 4 z § 9.3 podaje podobny wzór dla $\psi\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$.

Jeśli X i Y są niezależnymi zmiennymi losowymi, to

$$E(XY) = E(X) \cdot E(Y).$$

$V(X) = E(X^2) - \mu^2$.

Jeśli zmienne losowe $X_1, X_2, \ldots, X_n$ są wzajemnie niezależne, to

$$V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2) + \ldots + V(X_n).$$

Jeśli S_n jest zmienną losową o rozkładzie dwumianowym, to

$$E(S_n) = np \quad \text{oraz} \quad V(S_n) = npq.$$

Zmienna losowa unormowana odpowiadająca zmiennej X ma wartość średnią 0 i wariancję 1; jej dystrybuanta $\widetilde{F}$ jest związana z dystrybuantą F zmiennej X następującymi wzorami:

$$F(y) = \widetilde{F}\left(\frac{y-\mu}{\sigma}\right) \quad \text{oraz} \quad \widetilde{F}(y) = F(\sigma y + \mu) \quad \text{dla} \quad y \in \mathbb{R}.$$

Dystrybuanta unormowanej zmiennej losowej odpowiadającej zmiennej losowej o rozkładzie dwumianowym dla dużych wartości n jest równa w przybliżeniu dystrybuancie Φ rozkładu normalnego.

Metody

Wykorzystanie tablic wartości dystrybuant do obliczania prawdopodobieństw zdarzeń.

Wykorzystanie tablicy wartości dystrybuanty Φ do przybliżonego znajdowania wartości $P(a \leq S_n \leq b)$ dla zmiennych losowych S_n o rozkładzie dwumianowym.

10. ALGEBRY BOOLE'A

Określenie algebra Boole'a ma dwa różne, ale powiązane ze sobą znaczenia. Z jednej strony, oznacza ono pewien rodzaj arytmetyki symbolicznej, opracowanej po raz pierwszy przez George'a Boole'a w dziewiętnastym wieku do wykonywania obliczeń na wartościach logicznych w sposób algebraiczny. Jest to więc dział algebry odpowiednio dostosowany do opisu logiki dwuwartościowej. W szerszym sensie, algebra Boole'a zawiera w sobie wszystkie rodzaje metod matematycznych służących do opisu działania układów logicznych.

Algebra Boole'a jest również nazwą pewnej szczególnej struktury matematycznej, w której działania spełniają pewne konkretne prawa. Te prawa i działania zostały wybrane w taki sposób, by dostarczyć konkretnych modeli arytmetyki logiki.

Rozpoczniemy ten rozdział od opisu algebr Boole'a jako struktur algebraicznych, pokażemy powiązania między tymi algebrami i ich interpretacją logiczną, zastosujemy otrzymaną teorię do sieci logicznych i zakończymy go pokazaniem jednej metody upraszczania skomplikowanych wyrażeń logicznych.

§ 10.1. Algebry Boole'a

W tym paragrafie podajemy definicje i podstawowe własności algebr Boole'a oraz pokazujemy pewne ważne przykłady. Motywacja pochodzi z logiki symbolicznej, z którą zapoznaliśmy się w rozdziale 2. Jednym z głównych twierdzeń w tym paragrafie jest twierdzenie mówiące, że każda skończona algebra Boole'a jest w zasadzie algebrą podzbiorów pewnego zbioru skończonego, ro-

dzajem algebry, którą przedstawialiśmy za pomocą diagramów Venna w § 1.2. To twierdzenie daje teoretyczną podstawę do zaobserwowanego przedtem powiązania między działaniami teoriomnogościowymi $\cup$ i $\cap$ oraz ich odpowiednikami logicznymi $\vee$ i $\wedge$. Dobrze jest studiować ten paragraf, opuszczając dowody twierdzeń w pierwszym czytaniu i skupiając się na pojęciach i sformułowaniach twierdzeń.

Będziemy tworzyć teorię opartą na jej własnych podstawach, ale analogie z prawami działań na zbiorach oraz z matrycami logicznymi będą od razu widoczne. Zanim jednak podamy formalną definicję, pokażemy kilka przykładów algebr Boole'a.

PRZYKŁAD 1 (a) W zbiorze $\mathcal{P}(S)$ wszystkich podzbiorów zbioru S mamy znane działania $\cup$ i $\cap$. Dwóm elementom A i B zbioru $\mathcal{P}(S)$, czyli dwóm podzbiorom zbioru S, przyporządkowują one elementy $A \cup B$ i $A \cap B$ zbioru $\mathcal{P}(S)$. Działanie dopełnienia ^{c} przyporządkowuje zbiorowi A jego dopełnienie $A^c = S \backslash A$, będące innym elementem zbioru $\mathcal{P}(S)$. Działania $\cup$, $\cap$ oraz ^{c} mają wiele własności, z których niektóre są wyszczególnione w tablicy 1.1 w § 1.2. Zbiory S i $\emptyset$ mają specjalne własności, takie jak $A \cap S = A$ i $A \cap \emptyset = \emptyset$ dla wszystkich $A \in \mathcal{P}(S)$:

(b) W zbiorze $\mathbb{B} = \{0, 1\}$ mamy zwykłe działania logiczne $\vee$ i $\wedge$ oraz działanie $'$ zdefiniowane wzorami $0' = 1$ i $1' = 0$. Liczby 0 i 1 można interpretować odpowiednio jako "fałsz" i „prawdę", a działanie $'$ odpowiada negacji $\neg$. Poniżej mamy podany opis działań $\vee$ i $\wedge$:

$\vee$	0	1
0	0	1
1	1	1

$\wedge$	0	1
0	0	0
1	0	1

Korzystając ze zwykłych działań i funkcji liczbowych możemy napisać: $a \vee b = \max\{a, b\}$, $a \wedge b = \min\{a, b\} = a \cdot b$ oraz $a' = 1 - a$ dla $a, b \in \mathbb{B}$.

(c) W zbiorze $\mathbb{B}^n = \mathbb{B} \times \ldots \times \mathbb{B}$ (z n czynnikami $\mathbb{B}$) ciągów zerojedynkowych długości n mamy działania booleowskie $\vee$, $\wedge$ i $'$, pochodzące od odpowiednich działań w zbiorze $\mathbb{B}$. Na przykład

$$(a_1, \ldots, a_n) \vee (b_1, \ldots, b_n) = (a_1 \vee b_1, \ldots, a_n \vee b_n)$$

i podobnie definiujemy po współrzędnych działania $\wedge$ oraz $'$.

(d) Możemy traktować zbiór $\mathbb{B}^n$ jako zbiór funkcji f ze zbioru $\{1, 2, \ldots, n\}$ do zbioru $\mathbb{B}$, utożsamiając funkcję f z cią-

giem $(f(1), \ldots, f(n))$. Przy tej identyfikacji mamy

$$\begin{aligned}(f \vee g)(k) &= f(k) \vee g(k),\\ (f \wedge g)(k) &= f(k) \wedge g(k) \quad \text{oraz}\\ f'(k) &= (f(k))'\end{aligned}$$

dla $k = 1, 2, \ldots, n$ i $f, g \in \mathbb{B}^n$. Ogólniej, dla dowolnego zbioru S możemy zdefiniować działania $\vee$, $\wedge$ oraz $'$ w zbiorze $\text{FUN}(S, \mathbb{B})$ wszystkich funkcji ze zbioru S do zbioru $\mathbb{B}$ za pomocą wzorów $(f \vee g)(x) = f(x) \vee g(x)$, $(f \wedge g)(x) = f(x) \wedge g(x)$ oraz $f'(x) = (f(x))'$ dla wszystkich funkcji f i g ze zbioru S do $\mathbb{B}$. ∎

W ogólności, definiujemy **algebrę Boole'a** jako zbiór z dwoma działaniami dwuargumentowymi $\vee$ i $\wedge$, działaniem jednoargumentowym $'$ oraz różnymi elementami 0 i 1, spełniającymi następujące prawa:

1Ba. $x \vee y = y \vee x$ b. $x \wedge y = y \wedge x$	}	prawa przemienności
2Ba. $(x \vee y) \vee z = x \vee (y \vee z)$ b. $(x \wedge y) \wedge z = x \wedge (y \wedge z)$	}	prawa łączności
3Ba. $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$ b. $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$	}	prawa rozdzielności
4Ba. $x \vee 0 = x$ b. $x \wedge 1 = x$	}	prawa identyczności
5Ba. $x \vee x' = 1$ b. $x \wedge x' = 0$	}	prawa dopełnienia

Działanie $\vee$ nazywamy **sumą**, $\wedge$ — **iloczynem**, a działanie jednoargumentowe $'$ — **dopełnieniem**.

Jeśli zamienimy ze sobą działania $\vee$ i $\wedge$ w prawach definiujących algebrę Boole'a i jednocześnie zamienimy ze sobą 0 i 1, to otrzymamy z powrotem te same prawa. W każdym przypadku prawa a i b zamienią się ze sobą. W szczególności nie zmienią się własności dopełnienia. Własności, które zachodzą we wszystkich algebrach Boole'a, są to własności będące konsekwencjami praw definiujących, a więc mamy następującą podstawową **zasadę dualności**:

Jeśli zamienimy ze sobą znaki $\wedge$ i $\vee$ oraz 0 i 1 wszędzie we wzorze prawdziwym we wszystkich algebrach Boole'a, to otrzymany wzór będzie też prawdziwy we wszystkich algebrach Boole'a.

A oto kilka przykładów wniosków z praw definiujących algebry Boole'a.

Twierdzenie 1

Następujące własności zachodzą w każdej algebrze Boole'a:

6Ba. $x \vee x = x$ b. $x \wedge x = x$	prawa idempotentności
7Ba. $x \vee 1 = 1$ b. $x \wedge 0 = 0$	następne prawa identyczności
8Ba. $(x \wedge y) \vee x = x$ b. $(x \vee y) \wedge x = x$	prawa pochłaniania

Dowód. Oto wyprowadzenie prawa 6Ba:

$x \vee x = (x \vee x) \wedge 1$	prawo identyczności 4Bb
$= (x \vee x) \wedge (x \vee x')$	prawo dopełnienia 5Ba
$= x \vee (x \wedge x')$	prawo rozdzielności 3Ba
$= x \vee 0$	prawo dopełnienia 5Bb
$= x$	prawo identyczności 4Ba.

Dla dowodu prawa 7Ba zauważmy, że

$x \vee 1 = x \vee (x \vee x')$	prawo dopełnienia 5Ba
$= (x \vee x) \vee x'$	prawo łączności 2Ba
$= x \vee x'$	prawo idempotentności 6Ba, przed chwilą udowodnione
$= 1$	prawo dopełnienia 5Ba.

A dla dowodu prawa 8Ba mamy:

$(x \wedge y) \vee x = (x \wedge y) \vee (x \wedge 1)$	prawo identyczności 4Bb
$= x \wedge (y \vee 1)$	prawo rozdzielności 3Bb
$= x \wedge 1$	prawo identyczności 7Ba, przed chwilą udowodnione
$= x$	prawo identyczności 4Bb.

Prawa 6Bb, 7Bb i 8Bb wynikają teraz z zasady dualności. ∎

Okazuje się, że prawa łączności wynikają z innych praw definiujących algebry Boole'a, a więc są zbyteczne. Ponadto twierdzenie 1 może być dowiedzione bez odwoływania się do praw łączności. Dowody tych faktów są żmudne i nie są bardzo pouczające, więc je pominiemy.

Twierdzenie 2

W każdej algebrze Boole'a spełnione są następujące prawa De Morgana:

$$\left.\begin{array}{ll} \text{9Ba.} & (x \vee y)' = x' \wedge y' \\ \text{b.} & (x \wedge y)' = x' \vee y' \end{array}\right\}.$$

Dowód. Pokażemy najpierw, że jeśli $w \vee z = 1$ i $w \wedge z = 0$, to $z = w'$; zatem dwie własności $w \vee w' = 1$ i $w \wedge w' = 0$ charakteryzują w'. Istotnie:

$$\begin{array}{ll} z = z \vee 0 & \text{prawo identyczności 4Ba} \\ = z \vee (w \wedge w') & \text{prawo dopełnienia 5Bb} \\ = (z \vee w) \wedge (z \vee w') & \text{prawo rozdzielności 3Ba} \\ = (w \vee z) \wedge (w' \vee z) & \text{prawo przemienności 1Ba} \\ = 1 \wedge (w' \vee z) & \text{założenie} \\ = (w \vee w') \wedge (w' \vee z) & \text{prawo dopełnienia 5Ba} \\ = (w' \vee w) \wedge (w' \vee z) & \text{prawo przemienności 1Ba} \\ = w' \vee (w \wedge z) & \text{prawo rozdzielności 3Ba} \\ = w' \vee 0 & \text{założenie} \\ = w' & \text{prawo identyczności 4Ba.} \end{array}$$

Z tej charakteryzacji dopełnienia wynika, że aby dowieść własności 9Ba wystarczy pokazać $(x \vee y) \vee (x' \wedge y') = 1$ oraz $(x \vee y) \wedge (x' \wedge y') = 0$. Mamy

$$\begin{array}{ll} (x \vee y) \vee (x' \wedge y') & \\ \quad = [(x \vee y) \vee x'] \wedge [(x \vee y) \vee y'] & \text{rozdzielność} \\ \quad = [y \vee (x \vee x')] \wedge [x \vee (y \vee y')] & \text{łączność i przemienność} \\ \quad = (y \vee 1) \wedge (x \vee 1) = 1 \wedge 1 = 1. & \end{array}$$

Podobnie pokazuje się, że $(x \vee y) \wedge (x' \wedge y') = 0$, skąd wynika wzór 9Ba. Wzór 9Bb wynika z zasady dualności. ∎

PRZYKŁAD 2 Nietrudno sprawdzić, że zbiory z działaniami z przykładu 1 są algebrami Boole'a. Z twierdzenia 2 wynika, że muszą one spełniać również prawa De Morgana. W przypadku algebry $\mathcal{P}(S)$ z działaniami $\cup$, $\cap$ oraz ^{c} prawa De Morgana przybierają postać

$$(A \cup B)^c = A^c \cap B^c \quad \text{oraz} \quad (A \cap B)^c = A^c \cup B^c.$$

Te prawa w dwuelementowej algebrze Boole'a $\mathbb{B}$ oraz w algebrze $\mathbb{B}^n$ odpowiadają znanym prawom logicznym De Morgana

$$\neg(p \vee q) \Leftrightarrow (\neg p) \wedge (\neg q) \quad \text{oraz} \quad \neg(p \wedge q) \Leftrightarrow (\neg p) \vee (\neg q). \qquad \blacksquare$$

W algebrze Boole'a $\mathcal{P}(S)$ jest relacja $\subseteq$, która zachodzi między niektórymi jej elementami. Ta relacja może być wyrażona za pomocą działania $\cup$, gdyż $A \subseteq B$ wtedy i tylko wtedy, gdy $A \cup B = B$. Ten fakt sugeruje następującą definicję w ogólnym przypadku.

Definiujemy relację $\leq$ w algebrze Boole'a wzorem

$$x \leq y \quad \text{wtedy i tylko wtedy, gdy} \quad x \vee y = y.$$

Z prawa idempotentności $x \vee x = x$ wynika, że $x \leq x$ dla każdego x. Przyjmujemy, że $x < y$, jeśli $x \leq y$ oraz $x \neq y$. Niech $x \geq y$ oznacza $y \leq x$ oraz niech $x > y$ oznacza $y < x$.

PRZYKŁAD 3 (a) W przypadku algebry Boole'a $\mathcal{P}(S)$ relacje $\leq$, $<$, $\geq$ oraz $>$ są odpowiednio relacjami inkluzji $\subseteq$, $\subset$, $\supseteq$ oraz $\supset$.

(b) W algebrze Boole'a $\mathbb{B}$ mamy $0 \leq 0$, $0 \leq 1$, $1 \leq 1$ oraz $0 < 1$, a więc również $0 \geq 0$, $1 \geq 0$, $1 \geq 1$ oraz $1 > 0$. Nie ma tu żadnych niespodzianek!

(c) W algebrze $\mathbb{B}^n$ $(a_1, \ldots, a_n) \leq (b_1, \ldots, b_n)$ wtedy i tylko wtedy, gdy $(a_1 \vee b_1, \ldots, a_n \vee b_n) = (a_1, \ldots, a_n) \vee (b_1, \ldots, b_n) = (b_1, \ldots, b_n)$, tzn. wtedy i tylko wtedy, gdy $a_k \vee b_k = b_k$ dla każdego k. Zatem $(a_1, \ldots, a_n) \leq (b_1, \ldots, b_n)$ wtedy i tylko wtedy, gdy $a_k \leq b_k$ dla każdego k. $\blacksquare$

Chociaż relację $\leq$ zdefiniowaliśmy za pomocą działania $\vee$, to moglibyśmy równie dobrze użyć działania $\wedge$.

Lemat 1

W algebrze Boole'a:

$$x \vee y = y \quad \text{wtedy i tylko wtedy, gdy} \quad x \wedge y = x.$$

Dowód. Jeśli $x \vee y = y$, to

$$\begin{aligned} x &= (x \vee y) \wedge x && \text{prawo pochłaniania 8Bb} \\ &= y \wedge x && \text{założenie} \\ &= x \wedge y && \text{prawo przemienności 1Bb.} \end{aligned}$$

Implikacja odwrotna wynika z rozumowania dualnego. ∎

Może być nieco zaskakujące to, że relacje $\leq$ oraz $<$ zdefiniowane abstrakcyjnie za pomocą działań $\vee$ oraz $\wedge$ mają pewne znane własności relacji $\leq$ oraz $<$ dla liczb.

Lemat 2

W algebrze Boole'a:
(a) jeśli $x \leq y$ i $y \leq z$, to $x \leq z$;
(b) jeśli $x \leq y$ i $y \leq x$, to $x = y$;
(c) jeśli $x < y$ i $y < z$, to $x < z$.

Własności (a) i (b) mówią, że zgodnie z terminologią, którą wprowadzimy w § 11.1, relacja $\leq$ jest relacją częściowego porządku. Własność **przechodniości** (a) jest znana z badania relacji równoważności w § 3.5, ale relacja $\leq$ różni się bardzo od relacji równoważności.

Dowód lematu 2. (a) Jeśli $x \leq y$ i $y \leq z$, to

$$\begin{aligned} z &= y \vee z && \text{gdyż } y \leq z \\ &= (x \vee y) \vee z && \text{gdyż } x \leq y \\ &= x \vee (y \vee z) && \text{prawo przemienności 2Ba} \\ &= x \vee z && \text{gdyż } y \leq z. \end{aligned}$$

Zatem $x \leq z$.

(b) Jeśli $x \leq y$ i $y \leq x$, to $x \vee y = y$ i $y \vee x = x$. Z prawa przemienności 1Ba wynika, że $x = y$.

(c) Jeśli $x < y$ i $y < z$, to z punktu (a) mamy $x \leq z$. Przypadek $x = z$ jest niemożliwy ze względu na (b), gdyż wtedy mielibyśmy $x \leq y$, $y \leq z = x$ oraz $x \neq y$. ∎

Relacja $\leq$ jest również powiązana z działaniami $\vee$ i $\wedge$ oraz z elementami wyróżnionymi 0 i 1 w sposób naśladujący strukturę algebry $\mathcal{P}(S)$.

Lemat 3

W algebrze Boole'a:
(a) $x \wedge y \leq x \leq x \vee y$ dla każdych x i y;
(b) $0 \leq x \leq 1$ dla każdego x.

Dowód. (a) Ponieważ $(x \wedge y) \vee x = x$ na podstawie prawa 8Ba, więc $x \wedge y \leq x$. Podobnie z prawa 8Bb wynika, że $(x \vee y) \wedge x = x$, więc z lematu 1 otrzymujemy $x \leq x \vee y$.

(b) Teza wynika z praw 4Ba, $x \vee 0 = x$ oraz 4Bb, $x \wedge 1 = x$. ∎

Zbiory skończone mogą być tworzone jako sumy zbiorów jednoelementowych. Nierozkładalne cegiełki odgrywają również ważną rolę przy analizowaniu bardziej ogólnych algebr Boole'a. **Atomem** w algebrze Boole'a nazywamy niezerowy element a, który nie może być przedstawiony w postaci $a = b \vee c$, gdzie $a \neq b$ i $a \neq c$, tzn. nie może być przedstawiony w postaci sumy dwóch elementów różnych od niego.

PRZYKŁAD 4

(a) Atomami w algebrze $\mathcal{P}(S)$ są zbiory jednoelementowe $\{s\}$; każdy zbiór A należący do $\mathcal{P}(S)$, mający więcej niż jeden element, może być rozłożony na sumę $(A \backslash \{s\}) \cup \{s\}$, gdzie $s \in A$.

(b) Jedynym atomem algebry $\mathbb{B}$ jest 1.

(c) Atomami algebry $\mathbb{B}^n$ są te ciągi n-elementowe, w których dokładnie jeden wyraz jest równy 1, a pozostałe są równe 0. ∎

Atomy przypominają liczby pierwsze pod tym względem, że nie mają nietrywialnych rozkładów. Zobaczymy, że można ich używać, tak jak liczb pierwszych, do tworzenia innych elementów naszych algebr. Najpierw jednak podamy inną charakteryzację atomów, jako minimalnych elementów niezerowych.

Stwierdzenie

Niezerowy element x algebry Boole'a jest atomem wtedy i tylko wtedy, gdy nie istnieje element y taki, że $0 < y < x$.

Dowód. Przypuśćmy, że x jest atomem oraz $y < x$. Wtedy mamy $x = x \wedge 1 = (y \vee x) \wedge (y \vee y') = y \vee (x \wedge y')$. Ponieważ x jest atomem, więc jeden ze składników y i $x \wedge y'$ musi być równy x. Ale $y \neq x$ z założenia, więc $x \wedge y' = x$. Jednak wtedy $y = x \wedge y = (x \wedge y') \wedge y = x \wedge (y' \wedge y) = x \wedge 0 = 0$.

Z drugiej strony, jeśli x nie jest atomem, to $x = y \vee z$ dla pewnych y i z różnych od x. Ponieważ $y \leq y \vee z = x$ na podstawie lematu 3(a), więc $y < x$. Również $0 < y$, gdyż w przeciwnym razie mielibyśmy $y = 0$, a zatem $x = 0 \vee z = z \neq x$. ∎

Możemy teraz pokazać, w jaki sposób atomy można uważać za podstawowe cegiełki, z których są zbudowane skończone algebry Boole'a. Popatrzmy najpierw na kilka przykładów.

PRZYKŁAD 5

(a) Weźmy skończony zbiór S. Atomami algebry $\mathcal{P}(S)$ są jednoelementowe zbiory $\{s\}$. Jeśli $T = \{t_1, \ldots, t_m\}$ jest dowolnym m-elementowym podzbiorem zbioru S, to atomami $\{s\}$ algebry $\mathcal{P}(S)$ spełniającymi warunek $\{s\} \subseteq T$ są atomy $\{t_1\}, \ldots, \{t_m\}$ i zbiór T jest ich sumą $\{t_1\} \cup \ldots \cup \{t_m\}$.

(b) Atomami algebry $\mathbb{B}^n$ są ciągi n-elementowe mające dokładnie jeden wyraz równy 1. Niech na przykład a_i będzie atomem mającym jedynkę na i-tym miejscu. Przykład 3(c) pokazuje, że jeśli $x = (x_1, \ldots, x_n) \in \mathbb{B}^n$, to $a_i \leq x$ wtedy i tylko wtedy, gdy $x_i = 1$ oraz x jest sumą tych atomów a_i, dla których $x_i = 1$. Na przykład, jeśli $x = (1,1,0,1,0) \in \mathbb{B}^5$, to $x = (1,0,0,0,0) \vee (0,1,0,0,0) \vee (0,0,0,1,0) = a_1 \vee a_2 \vee a_4$. ∎

Następne twierdzenie pokazuje, że to, co się zdarzyło w tych dwóch przykładach, jest zjawiskiem ogólnym.

Twierdzenie 3

Niech B będzie skończoną algebrą Boole'a, której zbiorem atomów jest zbiór $A = \{a_1, \ldots, a_n\}$. Każdy niezerowy element x algebry B może być przedstawiony w postaci sumy różnych atomów:

$$x = a_{i_1} \vee \ldots \vee a_{i_k}.$$

Ponadto ten sposób przedstawienia jest jednoznaczny, z dokładnością do kolejności atomów.

Dowód. Pokażemy najpierw, że każdy niezerowy element może być przedstawiony w żądanej postaci, przy czym same atomy są przedstawione w postaci sum mających tylko jeden składnik. Przypuśćmy, że tak nie jest, i niech S będzie zbiorem niezerowych elementów algebry B nie będących sumami atomów. Jeśli $x \in S$, to element x sam nie jest atomem, więc tak jak w dowodzie stwierdzenia, może być przedstawiony w postaci $x = y \vee z$, gdzie $0 < y < x$ i $0 < z < x$. Ponadto, co najmniej jeden z elementów y i z musi należeć do zbioru S, gdyż w przeciwnym przypadku y i z byłyby sumami atomów, a więc element x, będący sumą y i z, też byłby sumą atomów. Zatem dla każdego $x \in S$ istnieje pewien element y (czy też z) w zbiorze S taki, że $x > y$. Wynika stąd, że dla dowolnego elementu x ze zbioru S istnieje ciąg $x = x_0 > x_1 > x_2 > \ldots$ elementów zbioru S. Ponieważ

algebra B jest skończona, więc wszystkie elementy $x_0, x_1, x_2, \ldots$ nie mogą być różne; wcześniej czy później $x_k = x_m$ dla pewnych liczb $k < m$. Wtedy z przechodniości nierówności $x_k > \ldots > x_m$ i lematu 2(c) wynika, że $x_k > x_m$, co przeczy temu, że $x_k = x_m$. (W tym rozumowaniu jest ukryta indukcja). Założenie, że zbiór S jest niepusty, prowadzi do sprzeczności, a więc każdy niezerowy element x musi być sumą atomów. (Na tym rozumowaniu można oprzeć algorytm rekurencyjny znajdowania rozkładu na sumę atomów).

Pokażemy następnie, że każdy element x algebry B jest sumą wszystkich atomów a, dla których $a \leq x$:

$$(*) \qquad x = \bigvee \{a \in A\colon\ a \leq x\}.$$

Przed chwilą pokazaliśmy, że element 1 jest sumą pewnego zbioru atomów. Wynika stąd, że

$$1 = \bigvee \{a \in A\colon\ a \leq 1\} = a_1 \vee \ldots \vee a_n,$$

gdyż możemy dopisać więcej atomów do sumy, która nadal będzie równa 1. (Tak naprawdę, twierdzenie, którego dowodzimy, pokazuje, że nie ma więcej atomów, ale tego jeszcze nie wiemy). Teraz

$$x = x \wedge 1 = x \wedge (a_1 \vee \ldots \vee a_n) = (x \wedge a_1) \vee \ldots \vee (x \wedge a_n).$$

Ponieważ $0 \leq x \wedge a_i \leq a_i$ oraz a_i jest atomem, więc ze stwierdzenia wynika, że $x \wedge a_i = a_i$, jeśli $a_i \leq x$ oraz $x \wedge a_i = 0$ w przeciwnym przypadku. Stąd wynika równość $(*)$.

Aby dowieść jednoznaczności, przypuśćmy, że $x = b_1 \vee \ldots \vee b_k$ jest pewnym sposobem przedstawienia x w postaci sumy atomów. Wtedy $b_i \leq x$ dla wszystkich i, a więc wszystkie b_i należą do zbioru $\{a \in A\colon\ a \leq x\}$. Z drugiej strony, jeśli $a \in A$ oraz $a \leq x$, to

$$0 \neq a = a \wedge x = a \wedge (b_1 \vee \ldots \vee b_k) = (a \wedge b_1) \vee \ldots \vee (a \wedge b_k).$$

Pewien iloczyn $a \wedge b_i$ musi być różny od 0, a więc $a \wedge b_i = a = b_i$, gdyż a i b_i są atomami. Zatem a jest jednym z b_i. Tak więc zbiór $\{b_1, \ldots, b_k\}$ jest dokładnie zbiorem atomów mniejszych lub równych x. ∎

Przekształcenie wzajemnie jednoznaczne φ z algebry Boole'a B_1 w algebrę Boole'a B_2, spełniające warunki

$$(1) \qquad \varphi(x \vee y) = \varphi(x) \vee \varphi(y),$$

$$(2) \qquad \varphi(x \wedge y) = \varphi(x) \wedge \varphi(y)$$

oraz

$$\varphi(x') = \varphi(x)' \tag{3}$$

dla wszystkich $x, y \in B_1$ nazywamy **izomorfizmem algebr Boole'a**. Dwie algebry Boole'a nazywamy **izomorficznymi**, jeśli istnieje izomorfizm z jednej algebry na drugą. Wtedy ich struktury algebraiczne są w zasadzie takie same.

Następne twierdzenie mówi nam, że skończona algebra Boole'a jest całkowicie wyznaczona, z dokładnością do izomorfizmu, przez liczbę atomów, które posiada.

Twierdzenie 4

Jeśli B_1 jest skończoną algebrą Boole'a, mającą zbiór atomów $A_1 = \{a_1, \ldots, a_n\}$ oraz jeśli B_2 jest inną skończoną algebrą Boole'a, której zbiorem atomów jest $A_2 = \{b_1, \ldots, b_n\}$, to istnieje izomorfizm algebr Boole'a φ z algebry B_1 na algebrę B_2 taki, że $\varphi(a_i) = b_i$ dla każdego i.

Dowód. Z twierdzenia 3 wynika, że każdy element x algebry B_1 może być jednoznacznie przedstawiony w postaci

$$x = a_{i_1} \vee \ldots \vee a_{i_k}.$$

Definiujemy $\varphi(a_i) = b_i$ dla $i = 1, 2, \ldots, n$ oraz ogólniej,

$$\varphi(a_{i_1} \vee \ldots \vee a_{i_k}) = b_{i_1} \vee \ldots \vee b_{i_k}.$$

Z naszej definicji i twierdzenia 3 wynika, że

$$\begin{aligned}\varphi(a_{i_1} \vee \ldots \vee a_{i_k}) &= \varphi(a_{i_1}) \vee \ldots \vee \varphi(a_{i_k}) \\ &= \bigvee \{\varphi(a)\colon\ a \in A \ \text{ oraz } \ a \leq x\}.\end{aligned}$$

Ale również

$$\varphi(x) = \bigvee \{b\colon\ b \in A_2 \ \text{ oraz } \ b \leq \varphi(x)\}.$$

Ponieważ przedstawienie $\varphi(x)$ w postaci sumy atomów jest jednoznaczne, więc dla $a \in A_1$ otrzymujemy

$$a \leq x \ \text{ wtedy i tylko wtedy, gdy } \ \varphi(a) \leq \varphi(x).$$

Aby sprawdzić warunek (1) z definicji izomorfizmu, weźmy elementy x i y algebry B_1 i zauważmy, że dla $a \in A_1$:

$$\begin{aligned}&\varphi(a) \leq \varphi(x \vee y) \\ &\quad \Leftrightarrow a \leq x \vee y && \text{z powyższego, dla } x \vee y \text{ zamiast } x \\ &\quad \Leftrightarrow a \leq x \ \text{ lub } \ a \leq y && \text{(por. ćwiczenie 11(a))} \\ &\quad \Leftrightarrow \varphi(a) \leq \varphi(x) \ \text{ lub } \ \varphi(a) \leq \varphi(y).\end{aligned}$$

Zatem dla $b \in A_2$ mamy:

$$b \leq \varphi(x \vee y) \Leftrightarrow b \leq \varphi(x) \text{ lub } b \leq \varphi(y) \Leftrightarrow b \leq \varphi(x) \vee \varphi(y).$$

Z twierdzenia 3 zastosowanego do algebry B_2 wynika, że $\varphi(x \vee y) = \varphi(x) \vee \varphi(y)$. Warunku (2) definicji izomorfizmu dowodzi się podobnie. Wreszcie

$$\varphi(x) \vee \varphi(x') = \varphi(x \vee x') = \varphi(1)$$

oraz

$$\varphi(x) \wedge \varphi(x') = \varphi(x \wedge x') = \varphi(0),$$

a więc $\varphi(x') = \varphi(x)'$ na podstawie faktu pokazanego na początku dowodu twierdzenia 2. ∎

Jeśli zbiór S ma n elementów, to algebra Boole'a $\mathcal{P}(S)$ (z działaniami $\cup$, $\cap$ i działaniem dopełnienia) ma dokładnie n atomów; są to jednoelementowe podzbiory zbioru S. Mamy więc następujący wniosek.

Wniosek

Każda skończona algebra Boole'a mająca n atomów jest izomorficzna z algebrą Boole'a $\mathcal{P}(S)$ wszystkich podzbiorów n-elementowego zbioru S, a więc ma dokładnie 2^n elementów. W szczególności algebra $\mathbb{B}^n$ jest izomorficzna z algebrą $\mathcal{P}(\{1, 2, \ldots, n\})$.

Funkcją booleowską n-argumentową nazywamy funkcję

$$f\colon \mathbb{B}^n \longrightarrow \mathbb{B}.$$

Zbiór wszystkich n-argumentowych funkcji booleowskich będziemy oznaczać symbolem $\text{BOOL}(n)$. Jeśli $f \in \text{BOOL}(n)$ oraz $(x_1, \ldots, x_n) \in \mathbb{B}^n$, to wartość $f((x_1, \ldots, x_n))$ oznaczamy symbolem $f(x_1, \ldots, x_n)$.

PRZYKŁAD 6

Trójargumentowa funkcja booleowska jest to funkcja f taka, że $f(x, y, z) = 0$ lub $f(x, y, z) = 1$ dla każdego z 2^3 wyborów x, y i z ze zbioru $\mathbb{B} \times \mathbb{B} \times \mathbb{B}$. Można przyjąć, że trzy zmienne x, y i z odpowiadają trzem przełącznikom, przy czym każdy może znajdować się w jednym z dwóch położeń. Wtedy funkcja f zachowuje się jak czarna skrzynka, która daje wynik 0 lub 1 w zależności od ustawienia przełączników i wewnętrznej struktury skrzynki. Ponieważ istnieje 8 sposobów ustawienia przełączników i każdy sposób może dać jeden z dwóch wyników, w zależności od funkcji,

więc istnieje $2^8 = 256$ trójargumentowych funkcji booleowskich. A więc $|\,\mathrm{BOOL}\,(3)| = 256$.

Trójargumentową funkcję booleowską możemy również traktować jako kolumnę w matrycy logicznej. Poniżej w tabelce widzimy jednoznacznie określoną funkcję f, jedną z $2^8 = 256$ możliwych funkcji, gdyż w odpowiadającej jej kolumnie może znajdować się dowolne ustawienie ośmiu zer lub jedynek. Każda kolumna odpowiadająca x, y lub z też określa pewną funkcję ze zbioru $\mathrm{BOOL}\,(3)$, tzn. funkcję z $\mathbb{B}^3$ do $\mathbb{B}$. Na przykład kolumna odpowiadająca zmiennej x określa funkcję g taką, że $g(a,b,c) = a$ dla $(a,b,c) \in \mathbb{B}^3$. ∎

x	y	z	f
0	0	0	1
0	0	1	1
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	0
1	1	0	0
1	1	1	1

Sposób zliczania przedstawiony w przykładzie 6 pokazuje ogólnie, że $|\,\mathrm{BOOL}\,(n)| = 2^{(2^n)}$; jest to bardzo duża liczba, chyba że liczba n jest bardzo mała. Tak jak w przykładzie 1(d), $\mathrm{BOOL}\,(n)$ jest algebrą Boole'a z działaniami booleowskimi zdefiniowanymi po współrzędnych.

PRZYKŁAD 7 Poniżej w tabelce widzimy matrycę logiczną ilustrującą działania booleowskie dla zadanych funkcji f i g. Zauważmy, że ponieważ funkcja $f \wedge g$ przyjmuje wartość 1 w dokładnie jednym punkcie $\mathbb{B}^3$, więc jest atomem w algebrze $\mathrm{BOOL}\,(3)$. Istnieje siedem innych atomów w algebrze $\mathrm{BOOL}\,(3)$. W następnym paragrafie pokażemy, w jaki sposób można przedstawić dowolny element skończonej algebry Boole'a, w szczególności algebry $\mathrm{BOOL}\,(n)$, w postaci sumy atomów. ∎

x	y	z	f	g	$f \vee g$	$f \wedge g$	f'	$f' \wedge g$
0	0	0	1	0	1	0	0	0
0	0	1	1	1	1	1	0	0
0	1	0	0	0	0	0	1	0
0	1	1	1	0	1	0	0	0
1	0	0	0	1	1	0	1	1
1	0	1	0	0	0	0	1	0
1	1	0	0	1	1	0	1	1
1	1	1	1	0	1	0	0	0

ĆWICZENIA DO § 10.1

1. (a) Wykaż, że algebra $\mathbb{B} = \{0, 1\}$ z przykładu 1(b) jest algebrą Boole'a, sprawdzając niektóre z praw od 1Ba do 5Bb.
(b) Zrób to samo z algebrą $\text{FUN}(S, \mathbb{B})$ z przykładu 1(d).

2. (a) Niech $S = \{a, b, c, d, e\}$. Przedstaw zbiór $\{a, c, d\}$ w postaci sumy atomów w algebrze $\mathcal{P}(S)$.
(b) Przedstaw $(1, 0, 1, 1, 0)$ w postaci sumy atomów w algebrze $\mathbb{B}^5$.
(c) Niech f będzie funkcją należącą do zbioru $\text{FUN}(S, \mathbb{B})$, przekształcającą a, c i d na 1 oraz b i e na 0. Przedstaw funkcję f w postaci sumy atomów w algebrze $\text{FUN}(S, \mathbb{B})$.

3. Znajdź taki zbiór S, by $\mathcal{P}(S)$ i $\mathbb{B}^5$ były izomorficznymi algebrami Boole'a. Wskaż izomorfizm algebry Boole'a $\mathbb{B}^5$ na algebrę $\mathcal{P}(S)$.

4. Opisz atomy algebry Boole'a $\text{FUN}(S, \mathbb{B})$ z przykładu 1(d). Czy twój opis jest poprawny, jeśli zbiór S jest nieskończony?

5. (a) Sporządź tabelę dla atomów algebry Boole'a $\text{BOOL}(2)$.
(b) Przedstaw funkcję g: $\mathbb{B}^2 \to \mathbb{B}$ zdefiniowaną wzorem $g(x, y) = x$ w postaci sumy atomów w algebrze $\text{BOOL}(2)$.
(c) Przedstaw funkcję h: $\mathbb{B}^2 \to \mathbb{B}$ zdefiniowaną wzorem $h(x, y) = x' \vee y$ w postaci sumy atomów w algebrze $\text{BOOL}(2)$.

6. (a) Ile atomów ma algebra Boole'a $\text{BOOL}(4)$?
(b) Weźmy funkcję należącą do algebry $\text{BOOL}(4)$ opisaną za pomocą kolumny mającej pięć jedynek i resztę zer. Ile atomów występuje w przedstawieniu tej funkcji w postaci sumy atomów?
(c) Ile różnych elementów algebry $\text{BOOL}(4)$ jest sumami pięciu atomów?

7. (a) Czy istnieje algebra Boole'a mająca 6 elementów? Odpowiedź uzasadnij.
(b) Czy każda skończona algebra Boole'a jest izomorficzna z algebrą Boole'a $\text{BOOL}(n)$ funkcji booleowskich? Odpowiedź uzasadnij.

8. (a) Opisz atomy algebry Boole'a $\mathcal{P}(\mathbb{N})$.
(b) Czy każdy niepusty element algebry $\mathcal{P}(\mathbb{N})$ jest sumą atomów? Przeanalizuj różne przypadki.

9. Istnieje naturalny sposób rysowania skończonych algebr Boole'a. Jeśli x i y są elementami algebry Boole'a B, to mówimy, że x pokrywa y wtedy, gdy $x > y$ i nie istnieją elementy z takie, że $x > z > y$. (Zatem atomy są elementami pokrywającymi 0). **Diagramem Hassego** algebry B nazywamy rysunek grafu skierowanego, którego wierzchołkami są elementy algebry B i w którym istnieje krawędź od wierzchołka x do wierzchołka y wtedy i tylko wtedy, gdy x pokrywa y.
Narysuj diagram Hassego dla następujących algebr Boole'a. Narysuj element 1 na górze i staraj się kierować krawędzie w dół.
(a) $\mathcal{P}(\{1, 2\})$, (b) $\mathbb{B}$, (c) $\mathbb{B}^2$, (d) $\mathbb{B}^3$.

10. Dla danej liczby naturalnej n, większej niż 1, niech D_n będzie zbiorem dzielników liczby n. Definiujemy działania $\vee$, $\wedge$ oraz $'$ w zbiorze D_n wzorami: $a \vee b = \mathrm{NWW}(a,b)$, $a \wedge b = \mathrm{NWD}(a,b)$ oraz $a' = n/a$.
(a) Zbiór $D_6 = \{1,2,3,6\}$ z działaniami $\vee$, $\wedge$ oraz $'$ jest algebrą Boole'a. Które elementy są zerem i jednością?
(b) Znajdź zbiór S taki, że algebry D_6 i $\mathcal{P}(S)$ są izomorficzne i wskaż izomorfizm między nimi.
(c) Pokaż, że zbiór D_4 z tymi działaniami nie jest algebrą Boole'a. (Spróbuj znaleźć oczywisty powód).
(d) Pokaż, że zbiór D_8 z tymi działaniami nie jest algebrą Boole'a.

11. Niech x i y będą elementami algebry Boole'a i niech a będzie atomem.
(a) Pokaż, że $a \leq x \vee y$ wtedy i tylko wtedy, gdy $a \leq x$ lub $a \leq y$.
(b) Pokaż, że $a \leq x \wedge y$ wtedy i tylko wtedy, gdy $a \leq x$ i $a \leq y$.
(c) Pokaż, że albo $a \leq x$, albo $a \leq x'$, ale nie mogą zajść obie te możliwości.

12. Niech x i y będą elementami skończonej algebry Boole'a, przy czym każdy jest przedstawiony w postaci sumy atomów:

$$x = a_1 \vee \ldots \vee a_n \quad \text{oraz} \quad y = b_1 \vee \ldots \vee b_m.$$

(a) Wyjaśnij, w jaki sposób można przedstawić $x \vee y$ i $x \wedge y$ w postaci sumy różnych atomów. Pokaż to na przykładach.
(b) W jaki sposób można przedstawić x' w postaci sumy różnych atomów?

13. Wykaż, że jeśli φ jest izomorfizmem algebry Boole'a B_1 na algebrę Boole'a B_2, to $x \leq y$ wtedy i tylko wtedy, gdy $\varphi(x) \leq \varphi(y)$.

14. Niech $S = [0,1)$ i niech rodzina zbiorów $\mathcal{A}$ składa się ze zbioru pustego $\emptyset$ i wszystkich podzbiorów zbioru S, które można przedstawić w postaci sumy skończonej liczby przedziałów postaci $[a,b)$.
(a) Wykaż, że każdy element rodziny $\mathcal{A}$ można przedstawić w postaci sumy skończonej liczby rozłącznych przedziałów postaci $[a,b)$.
(b) Pokaż, że rodzina $\mathcal{A}$ z działaniami $\cup$, $\cap$ i działaniem dopełnienia jest algebrą Boole'a.
(c) Wykaż, że algebra $\mathcal{A}$ nie ma ani jednego atomu.

§ 10.2. Wyrażenia booleowskie

Zasadniczym celem tego paragrafu jest wprowadzenie terminologii matematycznej i pojęć, których używa się w zastosowaniach metod algebry Boole'a do projektowania układów elektronicznych i analizy logicznej. W następnym paragrafie zajmiemy się tymi zastosowaniami znacznie dokładniej.

Wyrażenie booleowskie jest to ciąg symboli, wśród których występują stałe 0 i 1, pewne zmienne oraz działania booleowskie.

Dokładniej, definiujemy rekurencyjnie **wyrażenie booleowskie** n **zmiennych** $x_1, x_2, \ldots, x_n$ w następujący sposób:

(P) Symbole 0, 1 oraz $x_1, x_2, \ldots, x_n$ są wyrażeniami booleowskimi zmiennych $x_1, \ldots, x_n$.

(R) Jeśli E_1 i E_2 są wyrażeniami booleowskimi zmiennych $x_1, \ldots, x_n$, to $(E_1 \vee E_2)$, $(E_1 \wedge E_2)$ i E_1' też są takimi wyrażeniami.

Jak zwykle, w praktyce będziemy opuszczać zewnętrzne nawiasy oraz będziemy korzystać z praw łączności.

PRZYKŁAD 1 (a) Oto cztery wyrażenia booleowskie trzech zmiennych x, y, z:

$$(x \vee y) \wedge (x' \vee z) \wedge 1; \quad (x' \wedge z) \vee (x' \wedge y) \vee z'; \quad x \vee y; \quad z.$$

Pierwsze dwa z nich oczywiście zawierają wszystkie trzy zmienne. Ostatnie dwa nie zawierają. To, czy wyrażenie $x \vee y$ rozpatrujemy jako wyrażenie dwóch, trzech czy większej liczby zmiennych, na ogół nie ma znaczenia. Wtedy, gdy ma to znaczenie i zmienne nie są znane z kontekstu, będziemy starannie informowali, w jaki sposób należy traktować to wyrażenie.

Wyrażenia booleowskie 0 i 1 można traktować jako wyrażenia dowolnej liczby zmiennych, tak samo jak funkcje stałe można traktować jako funkcje jednej lub wielu zmiennych.

(b) Wyrażenie

$$(x_1 \wedge x_2 \wedge \ldots \wedge x_n) \vee (x_1' \wedge x_2 \wedge \ldots \wedge x_n) \vee (x_1 \wedge x_2' \wedge \ldots \wedge x_n)$$

jest przykładem wyrażenia booleowskiego n zmiennych. ∎

Używanie obu symboli $\vee$ i $\wedge$ prowadzi do powstawania długich i nieczytelnych wyrażeń booleowskich, więc zazwyczaj zastępujemy spójnik $\wedge$ kropką lub pomijamy całkowicie.

PRZYKŁAD 2 (a) Stosując tę nową umowę dotyczącą symbolu $\wedge$, pierwsze dwa wyrażenia booleowskie z przykładu 1(a) możemy zapisać w postaci

$$(x \vee y) \cdot (x' \vee z) \cdot 1 \quad \text{oraz} \quad (x'z) \vee (x'y) \vee z'$$

lub jeszcze prościej w postaci

$$(x \vee y)(x' \vee z)1 \quad \text{oraz} \quad x'z \vee x'y \vee z';$$

dokładnie tak, jak w przypadku zwykłej algebry, „iloczyn" $\wedge$ czy $\cdot$ ma pierwszeństwo przed „sumą" $\vee$.

(b) Wyrażenie booleowskie z przykładu 1(b) ma postać

$$x_1x_2 \dots x_n \vee x_1'x_2 \dots x_n \vee x_1x_2' \dots x_n.$$

(c) Wyrażenie $xyz \vee xy'z \vee x'z$ jest skróconą postacią wyrażenia

$$(x \wedge y \wedge z) \vee (x \wedge y' \wedge z) \vee (x' \wedge z).$$ ∎

Jeśli podstawimy 0 lub 1 w każdym miejscu, w którym występuje zmienna w wyrażeniu booleowskim, to otrzymamy wyrażenie składające się tylko z symboli 0, 1, $\vee$, $\wedge$ oraz $'$, które ma wartość w algebrze Boole'a $\mathbb{B} = \{0, 1\}$. Na przykład, jeśli zastąpimy x zerem, y jedynką i z jedynką w wyrażeniu booleowskim $x'z \vee x'y \vee z'$, to otrzymamy

$$0'1 \vee 0'1 \vee 1' = (1 \wedge 1) \vee (1 \wedge 1) \vee 0 = 1 \vee 1 \vee 0 = 1.$$

Ogólnie, jeśli E jest wyrażeniem booleowskim n zmiennych $x_1, x_2, \dots, x_n$, to E definiuje funkcję booleowską przekształcającą algebrę $\mathbb{B}^n$ w algebrę $\mathbb{B}$, której wartością dla argumentu $(a_1, a_2, \dots, a_n)$ jest element algebry $\mathbb{B}$ otrzymany przez zastąpienie w wyrażeniu E zmiennej x_1 wartością a_1, zmiennej x_2 wartością a_2, ..., zmiennej x_n wartością a_n.

PRZYKŁAD 3 Wartości funkcji booleowskiej przekształcającej algebrę $\mathbb{B}^3$ w algebrę $\mathbb{B}$, odpowiadające wyrażeniu $x'z \vee x'y \vee z'$ są przedstawione w poniższej tabeli. Dokładnie tak, jak w przypadku matryc logicznych dla zdań, obliczamy najpierw wartości funkcji booleowskich dla wyrażeń prostszych. Czwarta wartość w ostatniej kolumnie jest to wartość, którą obliczyliśmy przed chwilą. Zauważmy, że wyrażenie booleowskie z' odpowiada funkcji określonej w zbiorze $\mathbb{B}^3$, przeprowadzającej każdą trójkę (a, b, c) na c', gdzie $a, b, c \in \{0, 1\}$. Podobnie, wyrażenie z odpowiada funkcji przeprowadzającej trójkę (a, b, c) na c.

x	y	z	$x'z$	$x'y$	z'	$x'z \vee x'y \vee z'$
0	0	0	0	0	1	1
0	0	1	1	0	0	1
0	1	0	0	1	1	1
0	1	1	1	1	0	1
1	0	0	0	0	1	1
1	0	1	0	0	0	0
1	1	0	0	0	1	1
1	1	1	0	0	0	0

∎

Dwa wyrażenia booleowskie będziemy nazywać **równoważnymi**, jeśli odpowiadające im funkcje booleowskie są takie same. Na przykład wyrażenia $x(y \vee z)$ i $(xy) \vee (xz)$ są równoważne, gdyż każde odpowiada funkcji przyjmującej wartość 1 dla każdej trójki (a, b, c) spośród $(1, 1, 0)$, $(1, 0, 1)$ i $(1, 1, 1)$ i przyjmującej wartość 0 dla pozostałych trójek. Będziemy to zapisywać jako $x(y \vee z) = (xy) \vee (xz)$ i ogólnie będziemy pisać $E = F$, jeśli dwa wyrażenia booleowskie E i F są równoważne. Użycie znaku równości „=" na oznaczenie tej relacji równoważności jest powszechne i wydaje się, że nie prowadzi do nieporozumień.

Ten sposób oznaczania wyrażeń równoważnych jest dobrze znany z praktyki dotyczącej wyrażeń algebraicznych i funkcji rzeczywistych. Z formalnego punktu widzenia, wyrażenia $x^2 - 1$ i $(x + 1)(x - 1)$ są różne (gdyż wyglądają inaczej), ale funkcje f i g określone wzorami

$$f(x) = (x + 1)(x - 1) \quad \text{oraz} \quad g(x) = x^2 - 1$$

są równe. Uważamy te dwa wyrażenia za równoważne i zazwyczaj przyjmujemy albo wyrażenie $(x+1)(x-1)$, albo $x^2 - 1$ za nazwę funkcji, którą one definiują. Podobnie, będziemy często używać wyrażeń booleowskich jako nazw funkcji booleowskich, które one definiują.

PRZYKŁAD 4 Funkcja należąca do zbioru BOOL (3), mająca nazwę xy, jest zdefiniowana wzorem $xy(a, b, c) = ab$ dla wszystkich $(a, b, c) \in \mathbb{B}^3$, a więc

$$xy(a, b, c) = \begin{cases} 1, & \text{jeśli } a = b = 1, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Podobnie, funkcje mające nazwy $x \vee z'$ i $xy'z$ spełniają warunki

$$(x \vee z')(a, b, c) = a \vee c' = \begin{cases} 1, & \text{jeśli } a = 1 \text{ lub } c = 0, \\ 0, & \text{w przeciwnym przypadku} \end{cases}$$

oraz

$$xy'z(a, b, c) = ab'c = \begin{cases} 1, & \text{jeśli } a = 1,\ b = 0,\ c = 1, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Ponieważ funkcja $xy'z$ przyjmuje wartość 1 tylko w jednym punkcie algebry $\mathbb{B}^3$, więc jest atomem algebry Boole'a BOOL (3). Pozostałymi siedmioma atomami algebry BOOL (3) są

$$xyz,\ xyz',\ xy'z',\ x'yz,\ x'yz',\ x'y'z \quad \text{oraz} \quad x'y'z'. \qquad \blacksquare$$

Przypuśćmy, że E_1, E_2 i E_3 są wyrażeniami booleowskimi n zmiennych. Ponieważ BOOL(n) jest algebrą Boole'a, więc wyrażenia booleowskie $E_1(E_2 \vee E_3)$ i $(E_1E_2) \vee (E_1E_3)$ definiują tę samą funkcję. Zatem te dwa wyrażenia są równoważne i możemy zapisać to w postaci prawa rozdzielności

$$E_1(E_2 \vee E_3) = (E_1E_2) \vee (E_1E_3).$$

W taki sam sposób wyrażenia booleowskie spełniają również wszystkie pozostałe prawa algebry Boole'a, jeśli tylko równoważności zapisujemy w postaci równości.

Wyrażenia booleowskie składające się tylko z jednej zmiennej lub jej dopełnienia, takie jak x czy y', nazywamy **symbolami atomowymi**. Funkcje odpowiadające im przyjmują wartość 1 dla połowy elementów algebry $\mathbb{B}^n$. Na przykład symbol atomowy y' dla $n = 3$ odpowiada funkcji przyjmującej wartość 1 dla wszystkich argumentów $(a, 0, c) \in \mathbb{B}^3$ i przyjmuje wartość 0 dla wszystkich argumentów $(a, 1, c) \in \mathbb{B}^3$.

Tak jak w przykładzie 7 w § 10.1, atomami algebry BOOL(n) są funkcje przyjmujące wartość 1 dla dokładnie jednego elementu $\mathbb{B}^n$. Każdy atom odpowiada wyrażeniu booleowskiemu pewnej szczególnej postaci, nazywanemu **iloczynem minimalnym** (ang. minterm). Iloczyn minimalny n zmiennych jest to iloczyn dokładnie n symboli atomowych, z których każdy zawiera inną zmienną.

PRZYKŁAD 5 (a) Wyrażenia $xy'z'$ i $x'yz'$ są iloczynami minimalnymi trzech zmiennych x, y, z. Odpowiadające im funkcje ze zbioru BOOL(3) przyjmują wartość 1 odpowiednio w punktach $(1, 0, 0)$ i $(0, 1, 0)$.

(b) Wyrażenie xz' jest iloczynem minimalnym dwóch zmiennych x, z. Nie jest natomiast iloczynem minimalnym trzech zmiennych x, y, z; odpowiadająca mu funkcja ze zbioru BOOL(3) przyjmuje wartość 1 w dwóch punktach $(1, 0, 0)$ i $(1, 1, 0)$.

(c) Wyrażenie $xyx'z$ nie jest iloczynem minimalnym, gdyż zawiera zmienną x w więcej niż jednym symbolu atomowym. W rzeczywistości to wyrażenie jest równoważne z zerem. Wyrażenie $xy'zx$ również nie jest iloczynem minimalnym; jednakże jest równoważne z iloczynem minimalnym $xy'z$ trzech zmiennych x, y, z.

(d) W poniższej tabeli wypisanych jest osiem elementów algebry $\mathbb{B}^3$ wraz z odpowiednimi iloczynami minimalnymi przyjmującymi wartość 1 dla wskazanych argumentów. Zauważmy, że symbole atomowe odpowiadające 0 w argumencie są dopełnieniami zmiennych, pozostałe symbole atomowe nie są dopełnieniami.

(a, b, c)	iloczyn minimalny przyjmujący wartość 1 w punkcie (a, b, c)
$(0, 0, 0)$	$x'y'z'$
$(0, 0, 1)$	$x'y'z$
$(0, 1, 0)$	$x'yz'$
$(0, 1, 1)$	$x'yz$
$(1, 0, 0)$	$xy'z'$
$(1, 0, 1)$	$xy'z$
$(1, 1, 0)$	xyz'
$(1, 1, 1)$	xyz

■

Z twierdzenia 3 z § 10.1 wynika, że każdy element algebry BOOL (n) można przedstawić w postaci sumy atomów. Ponieważ atomy algebry BOOL (n) odpowiadają iloczynom minimalnym, więc każde wyrażenie booleowskie n zmiennych jest równoważne z sumą różnych iloczynów minimalnych. Ponadto takie przedstawienie w postaci sumy jest jednoznaczne, z dokładnością do porządku, w jakim są wypisane te iloczyny minimalne. Taką sumę iloczynów minimalnych, równoważną z danym wyrażeniem booleowskim E, będziemy nazywać **postacią kanoniczną** wyrażenia E. (Używa się również nazwy **postać normalna alternatywno-koniunkcyjna**, ang. *disjunctive normal form*, *DNF*). Części (b) i (c) następnego przykładu ilustrują dwie różne procedury znajdowania postaci kanonicznej.

PRZYKŁAD 6 (a) Wyrażenie booleowskie

$$x'yz' \vee xy'z' \vee xy'z \vee xyz'$$

jest już sumą iloczynów minimalnych zmiennych x, y, z, a więc jest swoją własną postacią kanoniczną. Odpowiadająca mu funkcja booleowska przyjmuje wartości pokazane w prawej kolumnie poniżej. Jedynki w tej kolumnie wskazują, które atomy algebry BOOL (3) są dodawane, a więc wskazują również odpowiednie iloczyny minimalne. Na przykład 1 w wierszu odpowiadającym argumentowi $(1, 1, 0)$ odpowiada iloczynowi minimalnemu xyz'.

x	y	z	$x'yz'$	$xy'z'$	$xy'z$	xyz'	$x'yz' \vee xy'z' \vee xy'z \vee xyz'$
0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0
0	1	0	1	0	0	0	1
0	1	1	0	0	0	0	0
1	0	0	0	1	0	0	1
1	0	1	0	0	1	0	1
1	1	0	0	0	0	1	1
1	1	1	0	0	0	0	0

(b) Wyrażenie booleowskie $(x \vee yz')(yz)'$ nie jest zapisane w postaci sumy iloczynów minimalnych. Aby otrzymać jego postać kanoniczną, możemy obliczyć wartości odpowiadającej mu funkcji booleowskiej. Na przykład, dla $x = 0$, $y = 0$, $z = 0$ otrzymujemy wartość

$$(0 \vee 01)(00)' = (0 \vee 0)0' = 01 = 0,$$

a dla $x = 1$, $y = 0$, $z = 1$ otrzymujemy wartość

$$(1 \vee 00)(01)' = (1 \vee 0)0' = 11 = 1.$$

Kiedy obliczymy wszystkie osiem wartości tej funkcji, otrzymamy prawą kolumnę tabeli z części (a). Zatem wyrażenie $(x \vee yz')(yz)'$ jest równoważne z sumą iloczynów minimalnych z części (a), czyli jego postacią kanoniczną jest $x'yz' \vee xy'z' \vee xy'z \vee xyz'$.

(c) Możemy zająć się samym wyrażeniem $(x \vee yz')(yz)'$ i spróbować przekształcić je na sumę iloczynów minimalnych, korzystając z praw algebry Boole'a. Przypomnijmy, że piszemy $E = F$, jeśli wyrażenia booleowskie E i F są równoważne. Prawa algebry Boole'a dają

$$\begin{aligned}
&(x \vee yz')(yz)' \\
&= (x \vee yz')(y' \vee z') && \text{prawo De Morgana} \\
&= (x(y' \vee z')) \vee ((yz')(y' \vee z')) && \text{prawo rozdzielności} \\
&= (xy' \vee xz') \vee (yz'y' \vee yz'z') && \text{dwukrotnie prawo rozdzielności} \\
&= (xy' \vee xz') \vee (0 \vee yz') && yy' = 0,\ z'z' = z' \\
&= xy' \vee xz' \vee yz' && \text{prawo łączności i własności 0}
\end{aligned}$$

Zastosowaliśmy najpierw prawa De Morgana, by przenieść wszystkie dopełnienia na poziom symboli atomowych. Następnie rozdzieliliśmy mnożenie względem dodawania tak dalece, jak tylko było to możliwe.

Mamy teraz wyrażenie będące sumą iloczynów symboli atomowych, ale nie jest ono sumą iloczynów minimalnych zmiennych x, y, z. Weźmy podwyrażenie xy', w którym brakuje zmiennej z. Ponieważ $z \vee z' = 1$, więc mamy $xy' = xy'1 = xy'(z \vee z') = xy'z \vee xy'z'$, a to ostatnie wyrażenie jest już sumą iloczynów minimalnych. Możemy postąpić podobnie z pozostałymi dwoma iloczynami i otrzymamy wyrażenie

$$xy' \vee xz' \vee yz' = (xy'z \vee xy'z') \vee (xyz' \vee xy'z') \vee (xyz' \vee x'yz'),$$

będące już sumą iloczynów minimalnych. Opuszczając składniki powtarzające się otrzymamy postać kanoniczną

$$xy'z \vee xy'z' \vee xyz' \vee x'yz'$$

początkowego wyrażenia $(x \vee yz')(yz)'$. Jest to oczywiście ten sam wynik, który otrzymaliśmy w części (b). ∎

Metody pokazane w tym przykładzie mają ogólne zastosowanie. Dla danego wyrażenia booleowskiego możemy obliczyć wartości funkcji booleowskiej, którą ono definiuje — tak naprawdę znaleźć jego matrycę logiczną. Wtedy każda wartość 1 odpowiada iloczynowi minimalnemu w postaci kanonicznej tego wyrażenia. Jest to metoda z przykładu 6(b). Z tego punktu widzenia postać kanoniczna daje po prostu inny sposób patrzenia na funkcje booleowskie.

Inaczej, możemy otrzymać postać kanoniczną tak, jak w przykładzie 6(c). Najpierw korzystamy z praw De Morgana, aby przenieść dopełnienia do symboli atomowych. Następnie rozdzielamy mnożenie względem dodawania wszędzie tam, gdzie jest to możliwe. Wtedy zastępujemy xx symbolem x i xx' zerem tam, gdzie jest to potrzebne oraz wstawiamy dodatkowe zmienne korzystając z równości $x \vee x' = 1$. Wreszcie usuwamy powtarzające się składniki.

Nie jest zawsze jasne, który sposób jest lepszy dla danego wyrażenia booleowskiego. Nie chcielibyśmy wykonywać ręcznie zbyt wielu obliczeń, używając którejkolwiek metody. Na szczęście, postać kanoniczna jest przede wszystkim używana jako narzędzie teoretyczne, a kiedy w praktyce musimy wykonać obliczenia, to możemy je wykonać na komputerze za pomocą prostych algorytmów.

Z teoretycznego punktu widzenia postać kanoniczna wyrażenia booleowskiego jest bardzo przydatna, gdyż pokazuje, jak to wyrażenie jest zbudowane z podstawowych składników, mianowicie iloczynów minimalnych, czyli atomów. Jak pokażemy w § 10.3, wyrażenia booleowskie mogą być realizowane jako układy elektroniczne i wyrażenia równoważne odpowiadają układom elektronicznym, które działają identycznie, tzn. dają te same wyniki dla takich samych danych. Zatem interesuje nas „upraszczanie" wyrażeń booleowskich, aby otrzymać odpowiadające im „uproszczone" układy elektroniczne.

Prostotę można mierzyć w różny sposób. Nie byłoby możliwe przedstawienie tutaj wszystkich metod mających znaczenie praktyczne, ale możemy przynajmniej zająć się jednym prostym

kryterium. Powiemy, że suma iloczynów symboli atomowych jest optymalna, jeśli nie istnieje równoważne wyrażenie booleowskie będące sumą mniejszej liczby iloczynów oraz jeśli wśród wszystkich równoważnych sum tej samej liczby iloczynów nie istnieją wyrażenia z mniejszą liczbą symboli atomowych. Naszym zadaniem jest znalezienie optymalnej sumy iloczynów równoważnej z danym wyrażeniem booleowskim. Możemy założyć, że znaleźliśmy już jedną równoważną sumę iloczynów, mianowicie postać kanoniczną.

PRZYKŁAD 7 (a) Weźmy wyrażenie $(xy)'z$. W tabeli poniżej pokazane są wartości funkcji booleowskiej, którą to wyrażenie definiuje. Postacią kanoniczną jest zatem $x'y'z \vee x'yz \vee xy'z$. To wyrażenie nie jest optymalne. Korzystając z praw algebry Boole'a otrzymujemy $(xy)'z = (x' \vee y')z = x'z \vee y'z$; otrzymane wyrażenie jest sumą tylko dwóch iloczynów i składa się tylko z czterech symboli atomowych. W przykładzie 2(d) w § 10.4 będziemy mogli pokazać, że wyrażenie $x'z \vee y'z$ jest optymalne (por. też ćwiczenie 13).

x	y	z	xy	$(xy)'$	$(xy)'z$
0	0	0	0	1	0
0	0	1	0	1	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	0	1	0
1	0	1	0	1	1
1	1	0	1	0	0
1	1	1	1	0	0

Ten przykład ilustruje problem, który pojawia się w praktyce. Wydaje się prawdopodobne, że układ realizujący wyrażenie $x'z \vee y'z$ będzie prostszy od układu realizującego wyrażenie $x'y'z \vee x'yz \vee xy'z$, ale może okazać się, że układ realizujący oryginalne wyrażenie $(xy)'z$ byłby najprostszy. Powrócimy do tej kwestii w § 10.3.

(b) Weźmy sumę iloczynów $E = x'z' \vee x'y \vee xy' \vee xz$. Czy jest ona optymalna? Prowadzimy obliczenia, korzystając z praw algebry Boole'a, włącznie ze sztuczką $x \vee x' = 1$, by znaleźć postać kanoniczną:

$$\begin{aligned} E &= x'yz' \vee x'y'z' \vee x'yz \vee x'yz' \vee xy'z \vee xy'z' \vee xyz \vee xy'z \\ &= x'yz' \vee x'y'z' \vee x'yz \vee xy'z \vee xy'z' \vee xyz. \end{aligned}$$

To oczywiście pogorszyło sprawę — mamy więcej iloczynów i więcej symboli atomowych. Chcemy przegrupować nasze wyrażenie w jakiś sprytny sposób. Zauważmy, że możemy zgrupować

te sześć iloczynów minimalnych parami, $x'yz'$ z $x'y'z'$, $x'yz$ z xyz i $xy'z$ z $xy'z'$, tak aby dwa iloczyny minimalne w jednej parze różniły się dokładnie jednym symbolem atomowym. Ponieważ

$$x'yz' \vee x'y'z' = x'(y \vee y')z' = x'z',$$
$$x'yz \vee xyz = yz \quad \text{oraz} \quad xy'z \vee xy'z' = xy',$$

więc otrzymujemy $E = x'z' \vee yz \vee xy'$. Inny sposób grupowania daje

$$x'yz' \vee x'yz = x'y, \quad x'y'z' \vee xy'z' = y'z' \quad \text{oraz} \quad xy'z \vee xyz = xz,$$

a więc $E = x'y \vee y'z' \vee xz$. W przykładzie 2(c) w § 10.4 pokażemy, że obie te sumy iloczynów $x'z' \vee yz \vee xy'$ i $x'y \vee y'z' \vee xz$ są optymalne. Zatem żadna suma iloczynów równoważna z wyrażeniem E nie ma mniej niż trzy iloczyny i żadna suma trzech iloczynów nie ma mniej niż sześć symboli atomowych. Niezależnie od tego, czy uwierzymy już teraz w te stwierdzenia, czy nie, oba te wyrażenia wyglądają prościej niż suma czterech iloczynów, od której zaczęliśmy. ∎

Istnieje metoda, zwana **procedurą Quine'a–McCluskeya**, tworzenia wyrażeń optymalnych za pomocą systematycznego grupowania iloczynów różniących się jednym symbolem atomowym. Ten algorytm jest żmudny, gdy wykonuje się go ręcznie, ale łatwo daje się zaprogramować do wykonania za pomocą komputera. Oprócz innych źródeł, podręczniki *Applications-Oriented Algebra* J. L. Fishera i *Modern Applied Algebra* G. Birkhoffa i T. C. Bartee'go zawierają czytelny opis tej metody.

Inna metoda znajdowania wyrażeń optymalnych, zwana **metodą tablic Karnaugha**, przypomina korzystanie z diagramów Venna. Ta metoda działa zupełnie dobrze dla wyrażeń booleowskich trzech lub czterech zmiennych, gdzie rozwiązanie problemu i tak jest dość łatwe, ale jest mniej użyteczna dla więcej niż czterech zmiennych. W podręczniku *Computer Hardware and Organization* M. E. Sloana wiele paragrafów jest poświęconych tablicom Karnaugha oraz omówione są zalety i wady zastosowań tej metody. Pokażemy tę metodę w § 10.4, po dokonaniu wprowadzenia do układów logicznych.

ĆWICZENIA DO § 10.2

1. Niech $f\colon \mathbb{B}^3 \to \mathbb{B}$ będzie funkcją booleowską taką, że $f(0,0,0) = f(0,0,1) = f(1,1,0) = 1$ oraz $f(a,b,c) = 0$ dla pozostałych trójek $(a,b,c) \in \mathbb{B}^3$. Zapisz odpowiadające jej wyrażenie booleowskie w postaci kanonicznej.

2. Wskaż funkcję booleowską odpowiadającą wyrażeniu booleowskiemu z przykładu 7(b).

3. Dla każdego z następujących wyrażeń booleowskich zmiennych x, y, z opisz odpowiadającą mu funkcję booleowską i zapisz to wyrażenie w postaci kanonicznej.
(a) xy, (b) z', (c) $xy \vee z'$, (d) 1.

4. Weźmy wyrażenie booleowskie $x \vee yz$ zmiennych x, y, z.
(a) Podaj tabelę wartości funkcji booleowskiej $f\colon \mathbb{B}^3 \to \mathbb{B}$ odpowiadającej temu wyrażeniu.
(b) Zapisz to wyrażenie w postaci kanonicznej.

5. Znajdź postać kanoniczną dla następujących wyrażeń booleowskich czterech zmiennych:
(a) $(x_1x_2x_3') \vee (x_1'x_2x_3x_4')$, (b) $(x_1 \vee x_2)x_3'x_4$.

6. Skorzystaj z metody z przykładu 6(c), by znaleźć postać kanoniczną wyrażenia booleowskiego $((x \vee y)' \vee z)'$ trzech zmiennych.

7. (a) Zapisz wyrażenie

$$xz \vee (y' \vee y'z) \vee xy'z'$$

w równoważnej postaci, jako sumę iloczynów, w której występują tylko trzy symbole atomowe.
(b) Powtórz ćwiczenie (a) dla wyrażenia $((xy \vee xyz) \vee xz) \vee z$.

8. Funkcja booleowska $f\colon \mathbb{B}^3 \to \mathbb{B}$ jest dana wzorem $f(a,b,c) = a+_2b+_2c$ dla $(a,b,c) \in \mathbb{B}^3$. Przypominamy, że symbol $+_2$ oznacza dodawanie modulo 2, zdefiniowane w § 3.6.
(a) Znajdź wyrażenie booleowskie odpowiadające funkcji f.
(b) Zapisz to wyrażenie zmiennych x, y, z w postaci kanonicznej.

9. Znajdź wyrażenie optymalne równoważne z wyrażeniem

$$(x \vee y)' \vee z \vee x(yz \vee y'z').$$

10. Pogrupuj trzy iloczyny minimalne wyrażenia $xyz \vee xyz' \vee xy'z$ tak, aby powstały dwie pary i następnie zapisz to wyrażenie w postaci równoważnej jako sumę dwóch iloczynów, z których każdy składa się z dwóch symboli atomowych.

11. Istnieje pojęcie sumy maksymalnej, dualne do pojęcia iloczynu minimalnego. **Sumą maksymalną zmiennych** $x_1, \ldots, x_n$ nazywamy sumę n symboli atomowych, z których każdy zawiera inną zmienną spośród $x_1, \ldots, x_n$.
(a) Skorzystaj z praw De Morgana, by pokazać, że każde wyrażenie booleowskie zmiennych $x_1, \ldots, x_n$ jest równoważne z iloczynem sum maksymalnych.
(b) Zapisz wyrażenie $xy' \vee x'y$ w postaci iloczynu sum maksymalnych zmiennych x i y.

12. Niech E będzie iloczynem k symboli atomowych wybranych spośród symboli $x_1, x_1', \ldots, x_n, x_n'$, zawierającym k różnych zmiennych x_i. Wykaż, że wyrażenie E określa funkcję należącą do zbioru BOOL (n), przyjmującą wartość 1 dla argumentów z pewnego podzbioru $\mathbb{B}^n$ mającego 2^{n-k} elementów. *Wskazówka*: znajdź postać kanoniczną wyrażenia E, stosując sztuczkę $x \vee x'$ z przykładu 6(c) dla zmiennych $w_1, \ldots, w_{n-k}$ nie występujących w wyrażeniu E.

13. (a) Wykaż, że wyrażenie $x'z \vee y'z$ nie jest równoważne z iloczynem symboli atomowych. *Wskazówka*: skorzystaj z ćwiczenia 12.

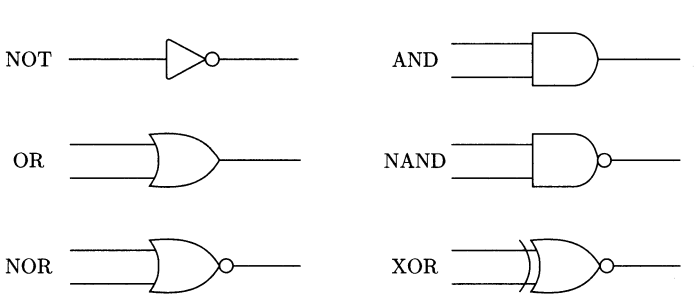
(b) Wykaż, że wyrażenie $x'z \vee y'z$ nie jest równoważne z sumą iloczynów symboli atomowych, w której jeden z „iloczynów" jest symbolem atomowym. (Ćwiczenia (a) i (b) pokazują razem, że wyrażenie $x'z \vee y'z$ jest optymalne)).

14. Udowodnij, że jeśli E_1 i E_2 są wyrażeniami booleowskimi zmiennych $x_1, \ldots, x_n$, to wyrażenia $E_1 \vee E_2$ i $E_2 \vee E_1$ są równoważne.

§ 10.3. Sieci logiczne

Na poziomie sprzętowym informatyka zajmuje się projektowaniem urządzeń, które mają dawać właściwe wyniki dla zadanych danych wejściowych. Jeśli dane wejściowe i wyniki na wyjściu są zerami lub jedynkami, to problem ten polega na zaprojektowaniu układu przetwarzającego dane wejściowe zgodnie z regułami określonymi za pomocą funkcji booleowskich. W tym paragrafie przyjrzymy się pobieżnie sposobom zastosowania metod algebry Boole'a do projektowania układów logicznych. Niektóre metody omówione w tym i w następnym paragrafie mają również zastosowanie w logice oprogramowania dla procesorów działających równolegle.

Podstawowymi cegiełkami, z których buduje się nasze sieci logiczne, są małe jednostki zwane **bramkami**, odpowiadające prostym funkcjom booleowskim. Sprzętowe odpowiedniki tych jednostek są dostępne na rynku. Są one połączone ze sobą, tworząc wiele konfiguracji. Na rysunku 10.1 widzimy symbole dla sześciu najbardziej podstawowych bramek, przedstawione zgodnie ze standardem ANSI/IEEE. Przyjmuje się, że linie dochodzące do danego symbolu z lewej strony są liniami wejściowymi, a linia po prawej stronie jest linią wyjściową. Umieszczenie małego kółka na linii wejściowej lub wyjściowej oznacza sygnał dopełniający do sygnału na tej linii. W tabeli poniżej pokazane są wartości funkcji booleowskich związanych z tymi sześcioma bramkami oraz podane są odpowiednie nazwy tych funkcji dwóch zmiennych x i y.



Rysunek 10.1

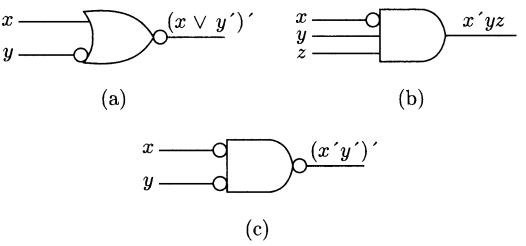
Bramki AND, OR, NAND i NOR dostępne są również z większą liczbą linii wejściowych.

x	y	x' NOT	$x \vee y$ OR	$(x \vee y)'$ NOR	xy AND	$(xy)'$ NAND	$x \oplus y$ XOR
0	0	1	0	1	0	1	0
0	1	1	1	0	0	1	1
1	0	0	1	0	0	1	1
1	1	0	1	0	1	0	0

PRZYKŁAD 1 (a) Bramka pokazana na rysunku 10.2(a) odpowiada funkcji booleowskiej $(x \vee y')'$, czyli równoważnie $x'y$.

(b) Bramka AND z trzema wejściami pokazana na rysunku 10.2(b) odpowiada funkcji $x'yz$.

(c) Bramka na rysunku 10.2(c) odpowiada funkcji $(x'y')'$, czyli $x \vee y$, a więc działa tak samo, jak bramka OR. ∎



Rysunek 10.2

Zajmiemy się problemem projektowania sieci bramek, która ma dawać wyniki zgodne z zadaną skomplikowaną funkcją booleowską wielu zmiennych. Jednym z ważnych wymagań jest, by liczba bramek była możliwie mała. Innym, by długość najdłuższego ciągu bramek była mała. W konkretnych zastosowaniach praktycznych spotykamy jeszcze inne wymagania.

PRZYKŁAD 2 Popatrzmy na nierozsądnie zaprojektowaną sieć pokazaną na rysunku 10.3(a). (Kropkami zaznaczone zostały miejsca, w których linie wejściowe rozdzielają się). W tej sieci są cztery bramki. Poruszając się od lewej do prawej mamy dwa ciągi składające się z trzech bramek. Obliczmy wartość funkcji booleowskich w punktach A, B, C i D:

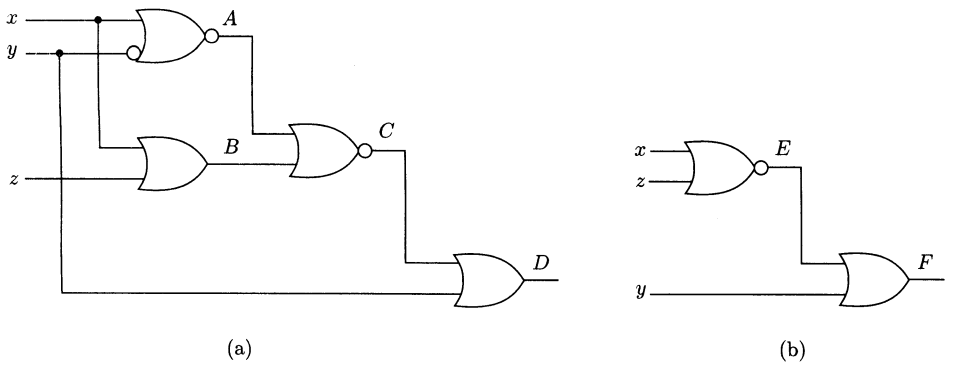
$$A = (x \vee y')'; \quad B = x \vee z;$$
$$C = (A \vee B)' = ((x \vee y')' \vee (x \vee z))';$$
$$D = C \vee y = ((x \vee y')' \vee (x \vee z))' \vee y.$$

Prawa algebry Boole'a dają

$$D = ((x \vee y')(x \vee z)') \vee y = ((x \vee y')x'z') \vee y =$$
$$= (xx'z' \vee y'x'z') \vee y = y'x'z' \vee y =$$
$$= y'x'z' \vee yx'z' \vee y = (y' \vee y)x'z' \vee y = x'z' \vee y.$$

Sieć pokazana na rysunku 10.3(b) daje te same wyniki, gdyż

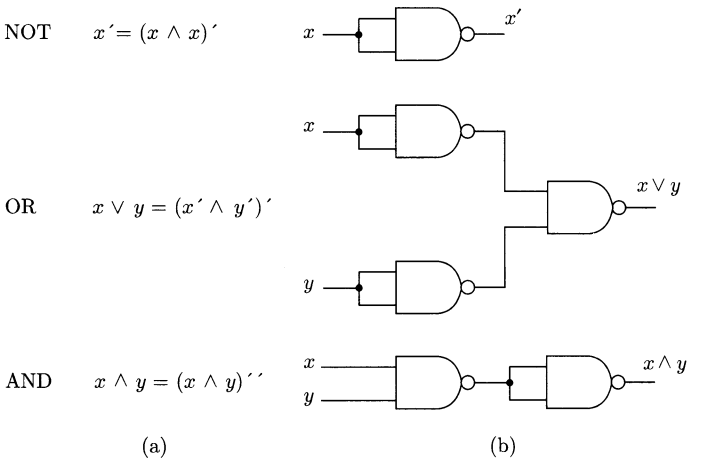
$$E = (x \vee z)' = x'z' \quad \text{oraz} \quad F = E \vee y = x'z' \vee y. \qquad \blacksquare$$



Rysunek 10.3

Ten prosty przykład pokazuje, w jaki sposób można czasem inaczej zaprojektować skomplikowaną sieć tak, aby zużyć mniej bramek. Jednym z powodów, dla którego staramy się zredukować długość ciągu kolejnych bramek, jest to, że w wielu sytuacjach, np. przy programowanych symulacjach konkretnych układów elektronicznych, działanie każdej bramki zabiera określoną ilość czasu, a bramki ustawione jedna za drugą wykonują swoje czynności po kolei. Długie ciągi bramek oznaczają więc wolniejsze działanie całej sieci.

Wyrażenie $x'z' \vee y$, które otrzymaliśmy po uproszczeniu skomplikowanego wyrażenia D w ostatnim przykładzie, jest wyrażeniem optymalnym dla D w sensie § 10.2. Wyrażenia optymalne nie zawsze dają najprostsze sieci. Na przykład można pokazać (por. ćwiczenie 7(a) w § 10.4), że wyrażenie $xz \vee yz$ jest wyrażeniem optymalnym zmiennych x, y, z. Ale wyrażenie $xz \vee yz = (x \vee y)z$ może być zaimplementowane za pomocą jednej bramki OR i jednej bramki AND, podczas gdy bezpośrednia implementacja wyrażenia $xz \vee yz$ wymaga dwóch bramek AND, by utworzyć xz i yz oraz bramki OR, by je połączyć. W praktyce należy zmienić naszą definicję wyrażeń „optymalnych" tak, by dostosować ją do dostępnych urządzeń sprzętowych.



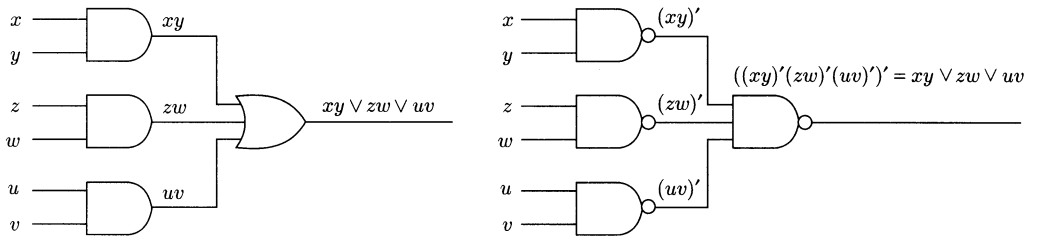
Rysunek 10.4

W pewnych sytuacjach chcemy mieć wszystkie bramki tego samego typu lub co najwyżej dwóch typów. Okazuje się, że zawsze wystarczą tylko bramki NAND lub tylko bramki NOR. To, których wygodniej jest użyć, może zależeć od stosowanych konkretnych rozwiązań elektronicznych. Na rysunku 10.4(a) pokazano, w jaki sposób zapisać NOT, OR i AND za pomocą NAND. Na rysunku 10.4(b) widzimy odpowiednie sieci. Ten rysunek pokazuje również rozwiązanie ćwiczenia 18 z § 2.4, gdyż NAND jest inną nazwą kreski Sheffera, wspomnianej w tym ćwiczeniu. W ćwiczeniu 2 prosimy o zrobienie analogicznej tabeli i rysunku sieci dla bramki NOR. Sieć dla bramki OR na rysunku 10.4 może być utworzona za pomocą pojedynczej bramki NAND, w której w obu wejściach mamy dopełnienie. Dopełnienie może wymagać oddzielnej bramki w konkretnych zastosowaniach, w zależności

od stosowanej technologii i od źródła, z którego pochodzą dane wejściowe. Będziemy na ogół zakładać, że dopełnienie odbywa się bez dodatkowych kosztów.

Kombinacje bramek AND i OR, takie jak te, które powstają przy okazji sum iloczynów, mogą być łatwo zrealizowane wyłącznie za pomocą bramek NAND.

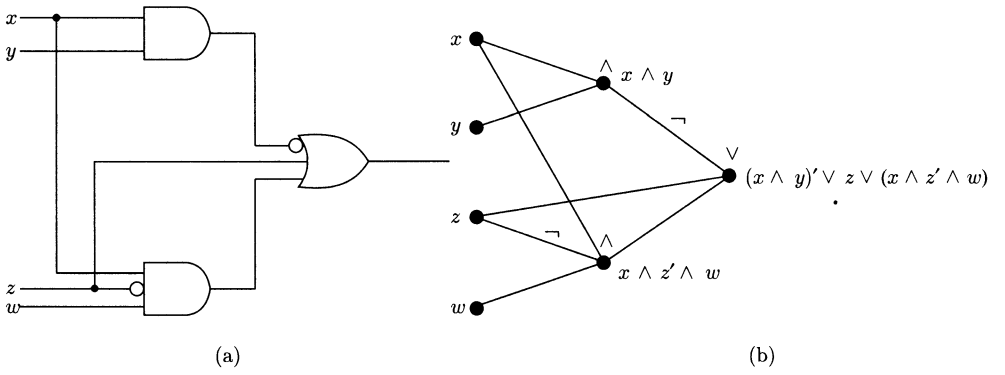
PRZYKŁAD 3 Na rysunku 10.5 widzimy prostą tego ilustrację. Po prostu zastąpmy wszystkie bramki AND i OR bramkami NAND w tej dwustopniowej sieci AND-OR, by otrzymać sieć równoważną. Dwustopniowa sieć OR-AND może być w podobny sposób zastąpiona siecią NOR (por. ćwiczenie 4). ∎



Rysunek 10.5

Na sieci logiczne można patrzeć jak na grafy skierowane, których źródła są oznaczone etykietami $x_1, x_2, \ldots$ i pozostałe wierzchołki mają etykiety $\vee$, $\wedge$ i $\oplus$ oraz których niektóre krawędzie mają etykietę $\neg$ oznaczającą dopełnienie. Z każdym wierzchołkiem można wtedy związać wyrażenie booleowskie zmiennych będących etykietami źródeł.

PRZYKŁAD 4 Z siecią na rysunku 10.6(a) związany jest graf pokazany na rysunku 10.6(b), w którym wszystkie krawędzie są skierowane z lewa na prawo. Jeśli do tego grafu dołączymy jeden wierzchołek $\wedge$ mający tylko jedno wejście tak, by znalazł się w środku krawędzi prowadzącej z wierzchołka z do wierzchołka $(x \wedge y)' \vee z \vee (x \wedge z' \wedge w)$, to nie zmieni się układ logiczny tej sieci, ale w otrzymanym grafie skierowanym wierzchołki będą się układały kolumnami — najpierw kolumna zmiennych, potem kolumna wierzchołków $\wedge$ i wreszcie kolumna wierzchołków $\vee$ — a krawędzie będą prowadziły wyłącznie od jednej kolumny do następnej. ∎



Rysunek 10.6

Etykietowany graf skierowany z przykładu 4 pokazuje sposób obliczania wartości funkcji $(x \wedge y)' \vee z \vee (x \wedge z' \wedge w)$. Podobnie każdy taki etykietowany graf skierowany opisuje sposób obliczania wartości funkcji booleowskiej związanej z jego ujściami. Ponieważ każde wyrażenie booleowskie może być zapisane w postaci sumy iloczynów symboli atomowych, więc wartości każdej funkcji booleowskiej mogą być obliczane w sposób opisany za pomocą grafu skierowanego takiego, jak w przykładzie 4: z kolumną zmiennych, kolumną wierzchołków $\wedge$ i kolumną składającą się z jednego wierzchołka $\vee$. Co więcej, jak widzieliśmy w przykładzie 3, wszystkie bramki mogą być bramkami NAND, a więc wierzchołek $\vee$ może być zastąpiony wierzchołkiem $\wedge$.

W takim grafie skierowanym żadna droga nie ma długości większej niż 2. Znaczy to, że związane z nim obliczenia trwają równo dwie jednostki czasu. Ceną, jaką się za to płaci, może być ogromna liczba bramek.

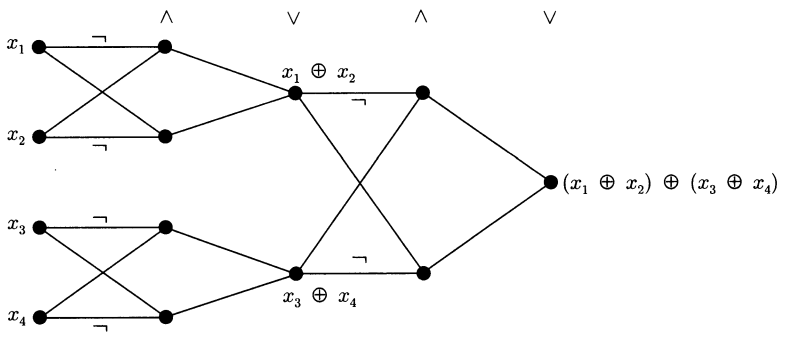
PRZYKŁAD 5 Weźmy wyrażenie booleowskie $E = x_1 \oplus x_2 \oplus \ldots \oplus x_n$. Odpowiadająca mu funkcja booleowska określona na algebrze $\mathbb{B}^n$ przyjmuje wartość 1 w punkcie $(a_1, a_2, \ldots, a_n)$ wtedy i tylko wtedy, gdy wśród wartości $a_1, a_2, \ldots, a_n$ jest nieparzysta liczba jedynek. Odpowiadające jej iloczyny minimalne mają nieparzystą liczbę symboli atomowych nie będących dopełnieniami. Zatem postać kanoniczna wyrażenia E jest sumą połowy wszystkich możliwych iloczynów minimalnych, a więc jest sumą 2^{n-1} składników.

Pokażemy następnie, że postać kanoniczna wyrażenia E jest optymalna, tzn. jeśli wyrażenie E zostanie zapisane w postaci sumy iloczynów symboli atomowych, to te iloczyny muszą być iloczynami minimalnymi wymienionymi w ostatnim akapicie. Gdyby mianowicie tak nie było, to pewien składnik byłby ilo-

czynem mniej niż n symboli atomowych, na przykład takim iloczynem, w którym nie występuje ani x_k, ani x'_k. Przy pewnych wyborach $a_1, a_2, \ldots, a_n$ ten składnik przyjmie wartość 1 i wtedy wśród wartości $a_1, a_2, \ldots, a_n$ jest nieparzysta liczba jedynek. Jeśli zmienimy wartość a_k z 1 na 0 lub z 0 na 1, to ten składnik będzie nadal miał wartość 1, ale wśród wartości $a_1, \ldots, a_n$ będzie teraz parzysta liczba jedynek. Żaden składnik występujący w wyrażeniu E nie może mieć tej własności, a więc każdy iloczyn wyrażenia E zawiera wszystkie n zmiennych.

Fakty zauważone w ostatnich dwóch akapitach pokazują, że graf skierowany długości 2 związany z wyrażeniem $E = x_1 \oplus x_2 \oplus \ldots \oplus x_n$ musi mieć co najmniej $2^{n-1} + 1$ wierzchołków $\wedge$ i $\vee$; ta liczba wierzchołków rośnie wykładniczo wraz z n. ∎

Jeśli godzimy się na to, by rosła długość dróg, możemy użyć metody „dziel i rządź", by utrzymać liczbę wierzchołków w rozsądnych granicach. Na rysunku 10.7 widzimy, jak to zrobić dla wyrażenia $x_1 \oplus x_2 \oplus x_3 \oplus x_4$. Ten graf skierowany ma 9 wierzchołków $\wedge$ i $\vee$. Tyle samo wierzchołków ma graf skierowany związany z przedstawieniem wyrażenia $x_1 \oplus x_2 \oplus x_3 \oplus x_4$ w postaci sumy iloczynów, gdyż $2^3 + 1 = 9$; nic więc nie zyskaliśmy. Ale jak będzie dla wyrażenia $x_1 \oplus x_2 \oplus \ldots \oplus x_8$? Graf pochodzący od sumy iloczynów ma $2^7 + 1 = 129$ wierzchołków $\wedge$ i $\vee$, podczas gdy graf analogiczny do tego z rysunku 10.7 ma tylko $9 + 9 + 2 + 1 = 21$ wierzchołków $\wedge$ i $\vee$ (por. ćwiczenie 11).



Rysunek 10.7

Ogólnie, dla wyrażenia $x_1 \oplus x_2 \oplus \ldots \oplus x_n$ mamy $2^{n-1}+1$ bramek w dwustopniowej sieci w porównaniu z tylko $3(n-1)$ bramkami w schemacie „dziel i rządź", podczas gdy długość najdłuższej drogi rośnie od 2 do co najwyżej $2\log_2 n$ (por. ćwiczenie 13).

Zatem podwojenie liczby danych wejściowych zwiększa długość dróg co najwyżej o 2.

PRZYKŁAD 6 Ważną klasą sieci logicznych są sieci służące do wykonywania działań arytmetycznych, np. dodawania, na liczbach zapisanych w systemie dwójkowym. Zilustrujemy pewne kwestie związane z takimi sieciami na przykładzie dodawania liczb 25 i 13, zapisanych w systemie dwójkowym odpowiednio jako 11001 i 1101. Ta reprezentacja po prostu znaczy to, że

$$25 = 16 + 8 + 1 = \mathbf{1} \cdot 2^4 + \mathbf{1} \cdot 2^3 + \mathbf{0} \cdot 2^2 + \mathbf{0} \cdot 2 + \mathbf{1} \cdot 1$$

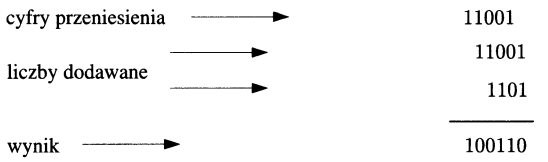
oraz

$$13 = 8 + 4 + 1 = \mathbf{1} \cdot 2^3 + \mathbf{1} \cdot 2^2 + \mathbf{0} \cdot 2 + \mathbf{1} \cdot 1.$$

Nasze zadanie wygląda więc następująco

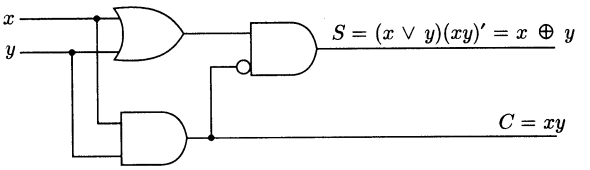
$$\begin{array}{r} 11001 \\ +\quad 1101 \\ \hline ? \end{array}$$

Dodawanie w systemie dwójkowym jest podobne do zwykłego dodawania w systemie dziesiętnym. Postępując od prawej do lewej, możemy dodawać cyfry w każdej kolumnie. Jeśli suma wynosi 0 lub 1, to zapisujemy tę sumę w linii wyniku i przenosimy cyfrę 0 do następnej kolumny z lewej strony. Jeśli suma wynosi 2 lub 3 (tzn. 10 lub 11 w systemie dwójkowym), to zapisujemy odpowiednio 0 lub 1 i przechodzimy do następnej kolumny, przenosząc cyfrę 1. Na rysunku 10.8 widzimy szczegóły naszego przykładu, dodany jest górny wiersz pokazujący cyfry przeniesienia. Odpowiedź jest rozwinięciem liczby $1 \cdot 2^5 + 0 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2 + 0 \cdot 1 = 38$, tak jak powinno być.



Rysunek 10.8

Kolumna znajdująca się najbardziej z prawej strony zawiera tylko dwie cyfry x i y (w naszym przykładzie $x = y = 1$). W tej kolumnie cyfrą wyniku jest $(x + y) \bmod 2$, czyli $x \oplus y$, a cyfrą przeniesienia do następnej kolumny jest $(x + y) \operatorname{DIV} 2$, czyli xy. Prosta sieć logiczna pokazana na rysunku 10.9, nazywana **półsumatorem**, daje dwa wyniki $S = x \oplus y$ i $C = xy$ dla danych

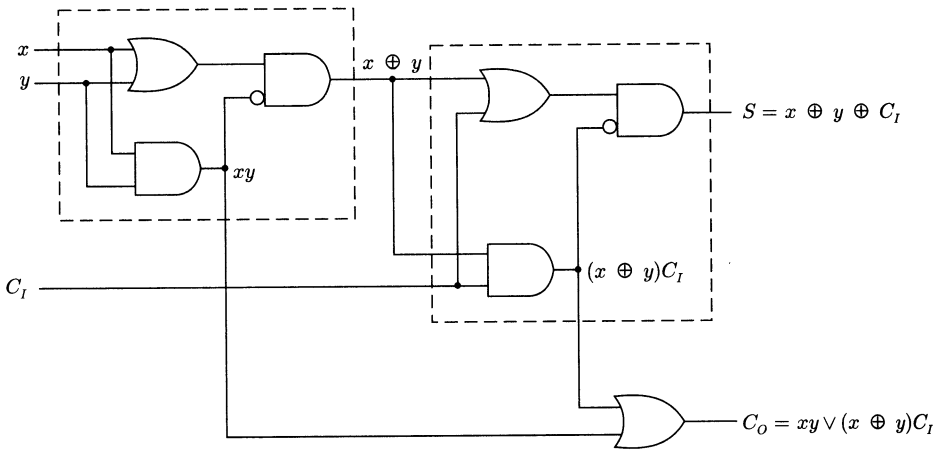


półsumator

Rysunek 10.9

wejściowych x i y. Litera S oznacza tu sumę, a litera C przeniesienie (ang. *carry*).

W bardziej ogólnym przypadku, gdy mamy przeniesienie C_I jako daną wejściową oprócz przeniesienia C_O jako wyniku, możemy połączyć dwa półsumatory i bramkę OR, by otrzymać sieć pokazaną na rysunku 10.10 i nazywaną **pełnym sumatorem**. Wartość C_O jest tu równa 1 wtedy i tylko wtedy, gdy co najmniej dwie spośród wartości x, y i C_I są równe 1, tzn. wtedy i tylko wtedy, gdy obie wartości x i y są równe 1 lub dokładnie jedna z nich jest równa 1 i przeniesienie C_I jest też równe 1.



pełny sumator

Rysunek 10.10

Wiele pełnych sumatorów można połączyć w sieć, która dodaje n-cyfrowe liczby zapisane w systemie dwójkowym lub też pojedynczy pełny sumator może być użyty wielokrotnie wraz z odpowiednimi urządzeniami opóźniającymi, wprowadzającymi dane cyfry dwójkowe (bity) po kolei. W praktyce każde z tych dwóch rozwiązań jest zbyt wolne. Zaprojektowano bardziej złożone sieci,

które dodają znacznie szybciej i wykonują inne działania arytmetyczne. ∎

Podaliśmy przykład 6 w celu pokazania zastosowania sieci logicznych do projektowania sprzętu komputerowego, a także wskazania, w jaki sposób można ze sobą łączyć wyniki częściowe pochodzące z procesów wykonywanych równolegle. Pełny sumator jest przykładem tego, w jaki sposób można łączyć ze sobą sieci realizujące dwie lub więcej funkcji booleowskich. Postacią kanoniczną wyrażenia $S = x \oplus y \oplus C_I$ jest

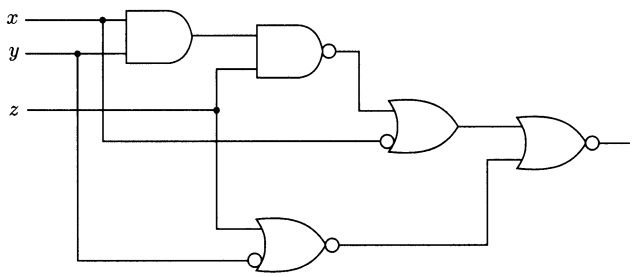
$$xyC_I \vee xy'C_I' \vee x'yC_I' \vee x'y'C_I$$

i okazuje się, że jest to wyrażenie optymalne (por. przykład 5 lub ćwiczenie 7(c) w § 10.4). Może ono być zaimplementowane za pomocą sieci logicznej składającej się z czterech bramek AND i jednej bramki OR, jeśli dopuścimy cztery linie wejściowe. Optymalnym wyrażeniem w postaci sumy iloczynów dla C_O jest $xy \vee xC_I \vee yC_I$, które można zrealizować za pomocą trzech bramek AND i jednej bramki OR. Utworzenie sieci dla S i C_O oddzielnie wymagałoby $4+1+3+1 = 9$ bramek, podczas gdy rysunek 10.10 pokazuje, że wystarczy 7 bramek, jeśli chcemy otrzymać S i C_O jednocześnie. Ponadto każda bramka na rysunku 10.10 ma tylko dwie linie wejściowe. Ta dyskusja pokazuje, że projektowanie oszczędnych sieci logicznych nie jest zadaniem łatwym.

ĆWICZENIA DO § 10.3

Uwaga. W tych ćwiczeniach można używać dopełnień danych na liniach wejściowych, chyba, że wyraźnie zaznaczyliśmy, iż jest inaczej.

1. (a) Podaj wzór funkcji booleowskiej odpowiadającej sieci logicznej pokazanej na rysunku 10.11.

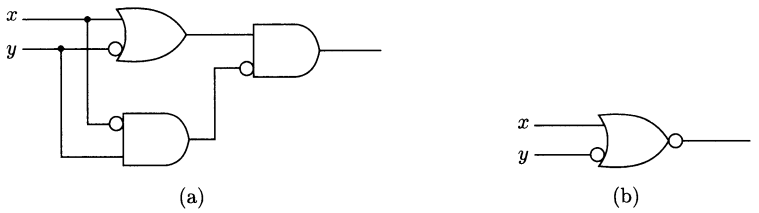


Rysunek 10.11

(b) Naszkicuj równoważną sieć logiczną składającą się z dwóch bramek mających po dwie linie wejściowe.

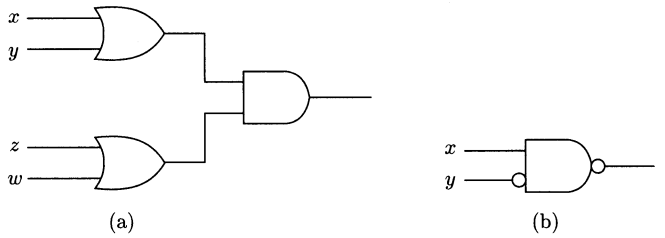
2. Napisz równania logiczne i naszkicuj sieci logiczne, takie jak na rysunku 10.4, pokazujące, w jaki sposób można wyrazić bramki NOT, OR i AND za pomocą bramki NOR, nie stosując dopełnień na liniach wejściowych.

3. Naszkicuj sieci logiczne równoważne z sieciami pokazanymi na rysunku 10.12, ale składające się wyłącznie z bramek NAND.



Rysunek 10.12

4. Naszkicuj sieci logiczne równoważne z sieciami pokazanymi na rysunku 10.13, ale składające się wyłącznie z bramek NOR.



Rysunek 10.13

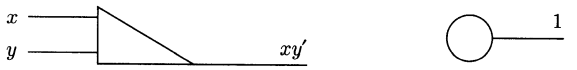
5. Naszkicuj sieć logiczną dla funkcji XOR używając
(a) dwóch bramek AND i jednej bramki OR.
(b) dwóch bramek OR i jednej bramki AND.

6. Naszkicuj sieć logiczną, która daje wynik 1 wtedy i tylko wtedy, gdy
(a) dokładnie jedna wartość wejściowa x, y, z jest równa 1,
(b) co najmniej dwie wartości wejściowe x, y, z, w są równe 1.

7. Oblicz wartości S i C_O w pełnym sumatorze dla następujących danych wejściowych:
(a) $x = 1$, $y = 0$, $C_I = 0$;
(b) $x = 1$, $y = 1$, $C_I = 0$;
(c) $x = 0$, $y = 1$, $C_I = 1$;
(d) $x = 1$, $y = 1$, $C_I = 1$.

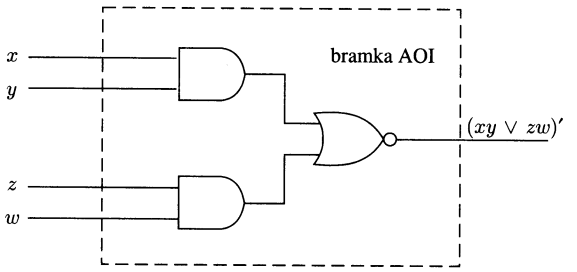
8. Znajdź wszystkie wartości x, y i C_I, dla których pełny sumator daje następujące wyniki:
(a) $S = 0$, $C_O = 0$;
(b) $S = 0$, $C_O = 1$;
(c) $S = 1$, $C_O = 1$.

9. Weźmy bramki „trójkątną" i „okrągłą", które dają wyniki pokazane na rysunku 10.14. Pokaż, w jaki sposób można zbudować z tych bramek sieci logiczne, nie używając dopełnień na liniach wejściowych i wyjściowych, odpowiadające następującym funkcjom booleowskim:
(a) x', (b) xy, (c) $x \vee y$.



Rysunek 10.14

10. Można kupić bramkę AND-OR-INVERT (AOI), dającą ten sam efekt, co sieć logiczna pokazana na rysunku 10.15. Jakich danych wejściowych należy użyć, by przekształcić tę bramkę w bramkę XOR?



Rysunek 10.15

11. Narysuj graf skierowany taki jak na rysunku 10.7 w celu obliczenia wartości $x_1 \oplus x_2 \oplus \ldots \oplus x_8$ metodą „dziel i rządź".

12. (a) Narysuj graf skierowany do obliczenia wartości $x_1 \oplus x_2 \oplus x_3 \oplus x_4$ za pomocą sieci dwustopniowej, wykorzystującej przedstawienie tego wyrażenia w postaci sumy iloczynów.
(b) Ile wierzchołków $\wedge$ ma graf skierowany do obliczenia $x_1 \oplus x_2 \oplus \ldots \oplus x_8$, wykorzystującego przedstawienie tego wyrażenia w postaci sumy iloczynów?
(c) Czy chciałbyś narysować graf skierowany z ćwiczenia (b)?

13. Wykaż przez indukcję, że dla $n \geq 2$ istnieje graf skierowany do obliczenia $x_1 \oplus x_2 \oplus \ldots \oplus x_n$, mający $3(n-1)$ wierzchołków $\wedge$ i $\vee$ i taki, że jeśli $2^m \geq n$, to każda droga w tym grafie ma długość co najwyżej $2m$. *Sugestia*: weź liczbę k taką, że $2^{k-1} < n \leq 2^k$ i połącz grafy skierowane dla 2^{k-1} zmiennych i dla $n - 2^{k-1}$ zmiennych.

14. Narysuj graf skierowany do obliczenia $x_1 \oplus x_2 \oplus \ldots \oplus x_6$, który ma 15 wierzchołków $\wedge$ i $\vee$ i w którym wszystkie drogi mają długość nie przekraczającą 6. *Wskazówka*: skorzystaj z ćwiczenia 13.

§ 10.4. Tablice Karnaugha

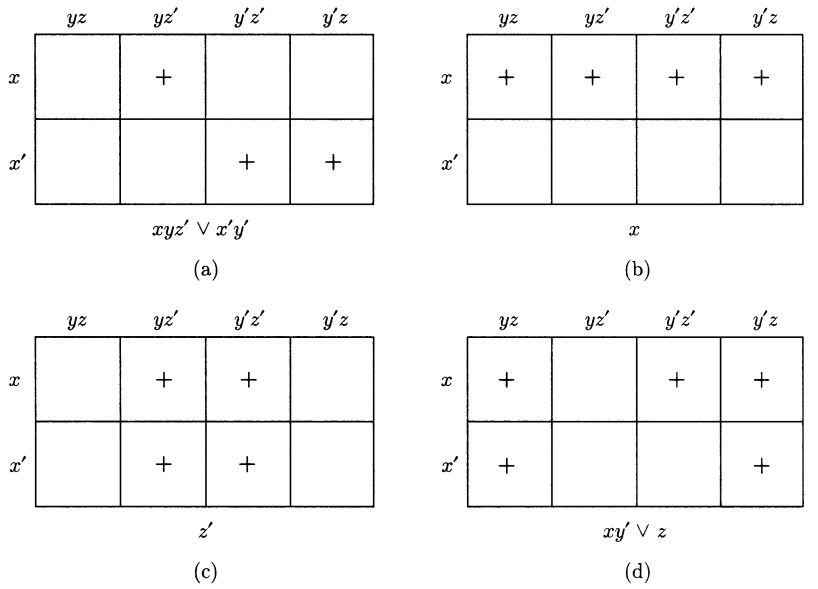
Zamiast szukać najbardziej opłacalnej czy „najlepszej" sieci logicznej, możemy postanowić szukać rozwiązania, które wydaje się całkiem niezłe. Rozwiązania optymalne w rozumieniu § 10.2 mogą być uważane za rozwiązania w przybliżeniu najlepsze, a więc warto znać metody szukania rozwiązań optymalnych. Taką metodą jest metoda tablic Karnaugha, którą teraz przedstawimy w skrócie. Możemy uważać ją za połączenie w teorii algebr Boole'a metody diagramów Venna i matryc logicznych, używanych wcześniej do zilustrowania relacji między zbiorami i między zdaniami.

Rozważymy najpierw przypadek funkcji booleowskich trzech zmiennych x, y i z. Tablica Karnaugha takiej funkcji jest tabelą o wymiarach 2×4, taką jak na rysunku 10.16. Każdy z ośmiu kwadratów w tej tabeli odpowiada iloczynowi minimalnemu. Znak „plus" wskazuje, które iloczyny minimalne występują w funkcji opisanej tą tabelą. Kolumny w tablicy Karnaugha są ustawione tak, że sąsiednie kolumny różnią się tylko jednym symbolem atomowym. Jeśli zwiniemy tę tabelę i skleimy lewą krawędź z prawą, to otrzymamy walec, którego kolumny nadal mają tę samą własność.

PRZYKŁAD 1 (a) Postacią kanoniczną wyrażenia z rysunku 10.16(a) jest $xyz' \vee x'y'z' \vee x'y'z$. Ponieważ $x'y'z' \vee x'y'z = x'y'(z' \vee z) = x'y'$, więc tę funkcję można również zapisać wzorem $xyz' \vee x'y'$.

(b) Tablice Karnaugha dla symboli atomowych są szczególnie proste. Na tablicy dla x na rysunku 10.16(b) wszystkie kwadraty w pierwszym wierszu są zaznaczone; na tablicy dla x' zaznaczone byłyby kwadraty w drugim wierszu. Na tablicy dla y zaznaczone byłyby kwadraty w lewym bloku o wymiarach 2×2, a na tablicy dla y' zaznaczone byłyby kwadraty prawego bloku o wymiarach 2×2. Na tablicy dla z' przedstawionej na rysunku 10.16(c) zaznaczone są kwadraty w środkowym bloku 2×2. Jeśli skleimy ze sobą lewą i prawą krawędź, to kolumny zawierające z też utworzą blok o wymiarach 2×2.

(c) Na tablicy z rysunku 10.16(d) zaznaczone są wszystkie kwadraty zawierające z oraz kwadrat odpowiadający iloczynowi



Rysunek 10.16

minimalnemu $xy'z'$. Ponieważ oba kwadraty zawierające xy' są zaznaczone, więc ta funkcja może być zapisana wzorem $xy' \vee z$. ■

Mamy więc tablicę narysowaną na walcu, w której symbolom atomowym x i x' odpowiadają bloki o wymiarach 1×4, symbolom atomowym y, z, y' i z' odpowiadają bloki o wymiarach 2×2, iloczynom dwóch symboli atomowych odpowiadają bloki o wymiarach 1×2 lub 2×1 oraz iloczynom trzech symboli atomowych odpowiadają bloki o wymiarach 1×1.

Aby znaleźć wyrażenie optymalne dla danej funkcji booleowskiej zmiennych x, y i z, zaznaczamy bloki odpowiadające iloczynom, wykonując następujące kroki.

Krok 1. Zaznacz na tablicy Karnaugha kwadraty odpowiadające danej funkcji.

Krok 2. (a) Zakreśl każdy blok składający się z ośmiu zaznaczonych kwadratów. (Jeśli wszystkich 8 kwadratów zostało zaznaczonych, to naszą funkcją booleowską jest 1. Możesz odpocząć.)

(b) Zakreśl każdy blok składający się z czterech zaznaczonych kwadratów, nie zawierający się w większym zakreślonym bloku.

(c) Zakreśl każdy blok składający się z dwóch zazna-

czonych kwadratów, nie zawierający się w większym zakreślonym bloku.

(d) Zakreśl każdy zaznaczony kwadrat, nie zawierający się w większym zakreślonym bloku.

Krok 3. Wybierz zbiór zakreślonych bloków tak, by

(a) każdy zaznaczony kwadrat znalazł się w co najmniej jednym wybranym bloku,

(b) liczba wybranych bloków była jak najmniejsza,

(c) wśród wszystkich zbiorów spełniających warunek (b) wyrażenie odpowiadające wybranemu zbiorowi zawierało jak najmniej symboli atomowych.

Powiemy jeszcze więcej o tym, jak spełnić warunki (b) i (c) podane w kroku 3, ale najpierw popatrzymy na kilka przykładów.

PRZYKŁAD 2 (a) Weźmy funkcję booleowską, której tablica Karnaugha jest przedstawiona na rysunku 10.17(a). „Prostokąty z zaokrąglonymi rogami" zakreślają trzy bloki, jeden składający się z czterech kwadratów i odpowiadający zmiennej y oraz dwa składające się z dwóch kwadratów, odpowiadające wyrażeniom xz' i $x'z$. Blok odpowiadający wyrażeniu $x'z$ składa się z dwóch kwadratów leżących po obu stronach linii, wzdłuż której sklejaliśmy ze sobą prawą i lewą krawędź. Ponieważ wszystkie trzy zakreślone bloki są potrzebne, by pokryć wszystkie zaznaczone kwadraty, więc musimy użyć tych trzech bloków w kroku 3. Otrzymanym wyrażeniem optymalnym jest więc $y \vee xz' \vee x'z$.

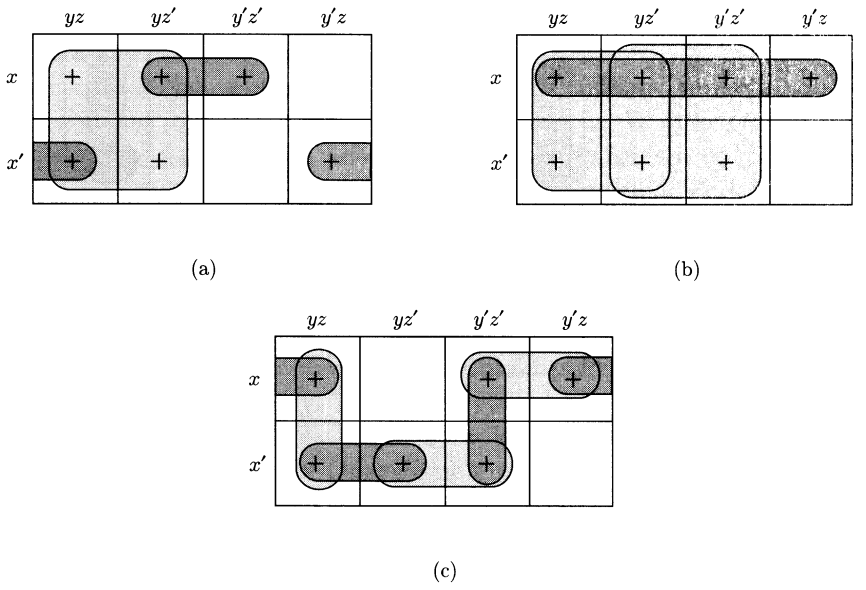
(b) Funkcja booleowska $(x'y'z)'$ jest przedstawiona na rysunku 10.17(b). Zakreślone bloki odpowiadają tu symbolom atomowym x, y i z'. Znów potrzebne są wszystkie trzy bloki, by pokryć zaznaczone kwadraty, więc wyrażeniem optymalnym jest $x \vee y \vee z'$.

(c) Na tablicy Karnaugha przedstawionej na rysunku 10.17(c) mamy sześć zakreślonych bloków, każdy składający się z dwóch kwadratów. Kwadraty zaznaczone mogą być pokryte każdym z dwóch zbiorów składających się z trzech bloków, odpowiadających wyrażeniom

$$x'y \vee xz \vee y'z' \quad \text{oraz} \quad x'z' \vee yz \vee xy'.$$

Ponieważ żaden zbiór składający się z mniejszej liczby bloków nie może pokryć sześciu kwadratów, więc oba te wyrażenia są optymalne. Tę funkcję booleowską widzieliśmy już w przykładzie 7(b) w § 10.2.

(d) Narysujmy tablicę Karnaugha dla funkcji booleowskiej $x'z \vee y'z$ i zakreślmy bloki, postępując zgodnie z krokami od 1

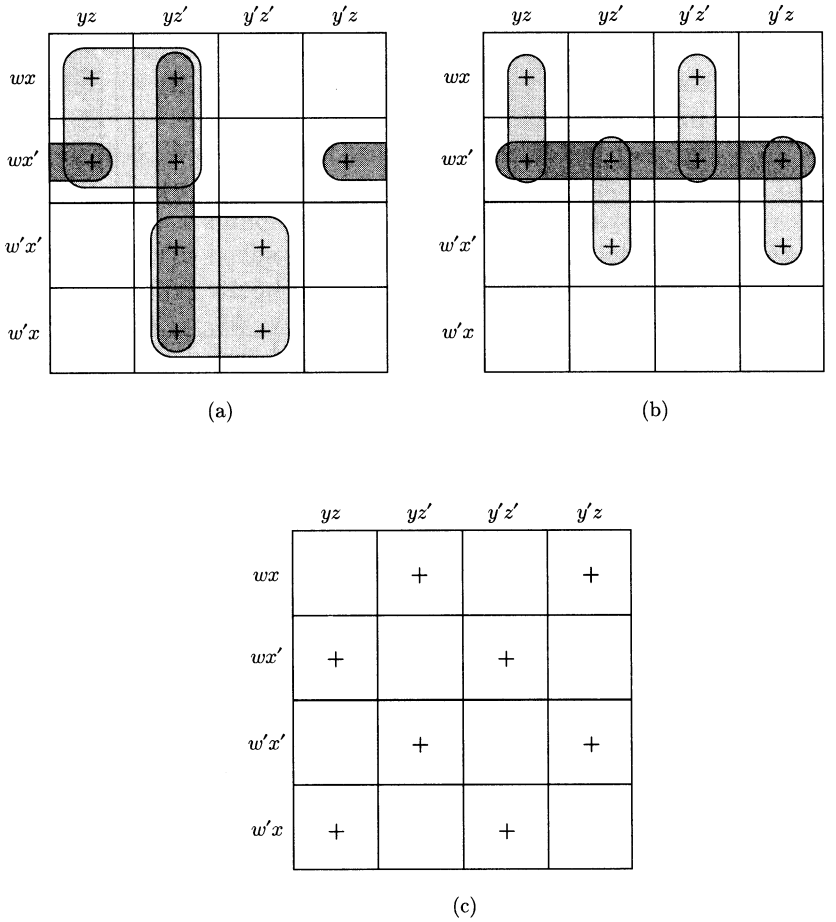


Rysunek 10.17

do 3. Zaznaczone zostaną tylko bloki odpowiadające iloczynom $x'z$ i $y'z$. Oba będą potrzebne do pokrycia zaznaczonych kwadratów, więc wyrażenie booleowskie $x'z \vee y'z$ jest samo dla siebie wyrażeniem optymalnym. ■

W przykładzie 2(c) pokazana jest sytuacja, w której więcej niż jeden wybór jest możliwy. Aby pokazać, na czym polega problem, gdy przy wyborze bloków w kroku 3 mamy do wyboru więcej niż jedną możliwość, zwiększymy liczbę zmiennych do czterech, np. w, x, y, z. Teraz tablica ma wymiary 4×4, jak na rysunku 10.18 i możemy wyobrazić sobie, że sklejamy ze sobą krawędzie górną i dolną, otrzymując rurkę, a następnie sklejamy razem lewą i prawą krawędź, by otrzymać powierzchnię w kształcie obwarzanka. Procedura polegająca na wykonaniu naszych trzech kroków jest taka sama jak przedtem, z tą tylko różnicą, że w kroku 2 zaczynamy od poszukiwania bloków składających się z 16 kwadratów.

PRZYKŁAD 3 (a) Na tablicy na rysunku 10.18(a) mamy zakreślone cztery bloki, trzy z nich składają się z czterech kwadratów odpowiadających iloczynom wy, yz' i $w'z'$, a jeden składa się z dwóch kwadratów i odpowiada iloczynowi $wx'z$. Ten blok mający dwa kwadraty jest jedynym blokiem zawierającym zaznaczony kwadrat $wx'y'z$, a bloki odpowiadające iloczynom wy i $w'z'$ są



Rysunek 10.18

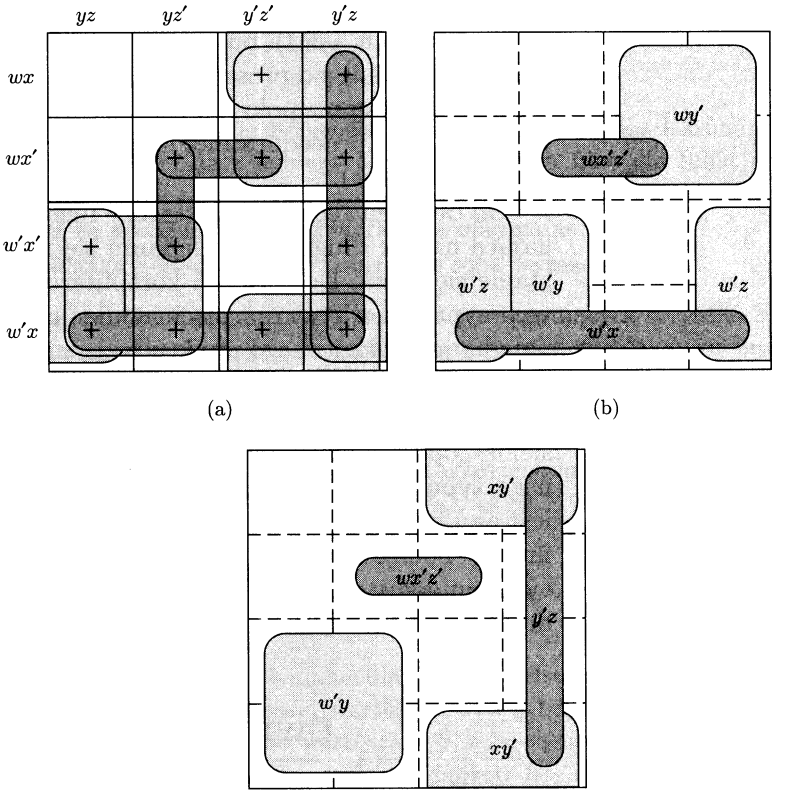
jedynymi blokami zawierającymi odpowiednio kwadraty $wxyz$ i $w'x'y'z'$, a więc te trzy bloki muszą być użyte. Ponieważ pokrywają one wszystkie zaznaczone kwadraty, więc spełniają warunki sformułowane w kroku 3. Optymalnym wyrażeniem jest $wx'z \vee wy \vee w'z'$.

(b) Na tablicy przedstawionej na rysunku 10.18(b) mamy pięć bloków. Każdy blok składający się z dwóch kwadratów jest potrzebny, gdyż każdy z nich jest jedynym blokiem zawierającym jeden z zaznaczonych kwadratów. Duży blok składający się z czterech kwadratów, odpowiadający iloczynowi wx', jest zbyteczny, gdyż kwadraty, które on zawiera są już pokryte przez inne bloki. Wyrażeniem optymalnym jest $wyz \vee x'yz' \vee wy'z' \vee x'y'z$. Zachłanność nie opłaca się tutaj, gdyż każdy algorytm, który wy-

biera najpierw największe bloki, wybrałby niepotrzebnie ten blok mający cztery kwadraty.

(c) Szachownica na rysunku 10.18(c) opisuje funkcję boolеowską $w \oplus x \oplus y \oplus z$ zmiennych w, x, y, z. W tym przypadku wszystkie bloki mają wymiary 1×1 i wyrażeniem optymalnym jest po prostu postać kanoniczna tego wyrażenia. Podobnie jest dla wyrażenia ogólnego $x_1 \oplus x_2 \oplus \ldots \oplus x_n$, jak zauważyliśmy to w przykładzie 5 w § 10.3. ■

PRZYKŁAD 4 Tablice w przykładzie 3 nie dawały nam żadnych prawdziwych możliwości wyboru; istotne bloki pokrywały już wszystkie zaznaczone kwadraty. Tablica na rysunku 10.19(a) daje nam takie możliwości. Każdy zaznaczony kwadrat występuje w co najmniej dwóch blokach. Oczywiście musimy wybrać co najmniej jeden blok składający się z dwóch kwadratów, by pokryć kwadrat



Rysunek 10.19

$wx'yz'$. Przypuśćmy, że wybraliśmy blok $wx'z'$. Możemy teraz wybrać cztery dodatkowe bloki pokazane na rysunku 10.19(b). Otrzymanym wyrażeniem jest

$$wx'z' \vee wy' \vee w'y \vee w'z \vee w'x.$$

Na rysunku 10.19(c) widzimy inny sposób wyboru bloków, tym razem mamy tylko cztery bloki. Odpowiadającym im wyrażeniem jest

$$wx'z' \vee w'y \vee xy' \vee y'z.$$

To wyrażenie jest lepsze, ale czy jest optymalne? Jedyne możliwe ulepszenie mogłoby polegać na ograniczeniu się do jednego bloku składającego się z dwóch kwadratów i dwóch mających po cztery kwadraty. Ponieważ mamy do pokrycia 12 kwadratów, więc takie ulepszenie nie jest możliwe i otrzymane wyrażenie jest optymalne. ∎

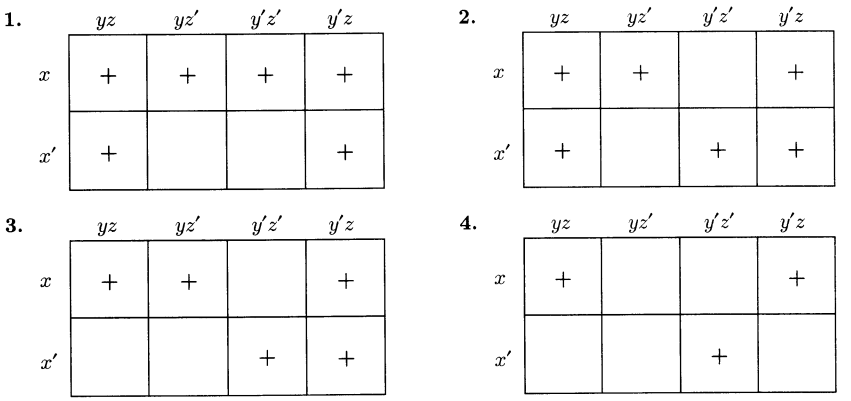
Zasady, jakimi należy się kierować przy wyborze bloków w takich sytuacjach, jak ta powyżej, są dość skomplikowane. Nie wystarczy wybranie po prostu koniecznych bloków, gdyż te musimy wybrać, a następnie pokryć pozostałe kwadraty jak największymi blokami. Taka procedura mogłaby dać rozwiązanie nieoptymalne z przykładu 4(b).

Do obliczeń ręcznych wystarczy metoda prób i błędów, polegająca na tak długim przyglądaniu się rysunkowi, aż odpowiedź stanie się jasna. Do obliczeń komputerowych, i tak potrzebnych w przypadku więcej niż pięciu zmiennych, metoda tablic Karnaugha jest z logicznego punktu widzenia taka sama, jak metoda Quine'a–McCluskeya, dla której istnieje gotowe oprogramowanie.

Zakończymy to omówienie podkreśleniem jeszcze raz, że określenie „optymalny" odnosi się tylko do złożoności wyrażeń pewnego typu, takich jak sumy iloczynów symboli atomowych. Nie jest ono synonimem słowa „najlepszy". Wyrażenie optymalne dla funkcji booleowskiej daje metodę konstruowania odpowiadających mu dwustopniowych sieci AND-OR o możliwie małej liczbie bramek, ale inne rodzaje sieci mogą być tańsze w budowie.

ĆWICZENIA DO § 10.4

Dla każdej z tablic Karnaugha z ćwiczeń 1-4 pokazanych na rysunku 10.20 znajdź odpowiadającą jej postać kanoniczną i wyrażenie optymalne.



Rysunek 10.20

5. Narysuj tablice Karnaugha i zakreśl bloki dla następujących funkcji booleowskich zmiennych x, y i z, stosując podaną metodę trzech kroków:
(a) $x \vee x'yz$, (b) $(x \vee yz)'$,
(c) $y'z \vee xyz$, (d) $y \vee z$.

6. Przypuśćmy, że funkcje booleowskie E i F mają tablice Karnaugha składające się każda z jednego bloku, przy czym blok dla funkcji E zawiera blok dla funkcji F.
(a) Jak są powiązane ze sobą wyrażenia optymalne dla E i F?
(b) Podaj przykład tak powiązanych ze sobą wyrażeń E i F.

7. Narysuj tablice Karnaugha dla każdego z następujących wyrażeń booleowskich zmiennych x, y i z i pokaż, że te wyrażenia są optymalne:
(a) $xz \vee yz$, (b) $xy \vee xz \vee yz$,
(c) $xyz \vee xy'z' \vee x'yz' \vee x'y'z$.

8. Powtórz ćwiczenie 7 dla następujących wyrażeń zmiennych x, y, z, w:
(a) $x' \vee yzw$, (b) $x'z' \vee xy'z \vee w'xy$,
(c) $wxz \vee wx'z' \vee w'x'z \vee w'xz'$.

9. Znajdź wyrażenia optymalne dla funkcji booleowskich, mających tablice Karnaugha pokazane na rysunku 10.21.

10. (a) Znajdź wyrażenie optymalne dla funkcji booleowskiej E zmiennych x, y, z, w, przyjmującej wartość 1 wtedy i tylko wtedy, gdy co najmniej dwie spośród wartości x, y, z, w są równe 1.
(b) Podaj wyrażenie booleowskie dla funkcji E z ćwiczenia (a), mające osiem symboli działań $\vee$ i $\wedge$. (Wyrażenie optymalne z ćwiczenia (a) nie minimalizuje liczby bramek w sieci logicznej dla wyrażenia E).

(a)

	yz	yz'	$y'z'$	$y'z$
wx	+	+	+	
wx'		+	+	+
$w'x'$	+	+	+	+
$w'x$	+	+	+	

(b)

	yz	yz'	$y'z'$	$y'z$
wx	+			
wx'				
$w'x'$	+	+	+	+
$w'x$		+	+	+

(c)

	yz	yz'	$y'z'$	$y'z$
wx	+	+	+	
wx'	+		+	+
$w'x'$		+	+	
$w'x$			+	+

(d)

	yz	yz'	$y'z'$	$y'z$
wx	+			+
wx'	+			+
$w'x'$		+	+	
$w'x$	+			+

Rysunek 10.21

To, co jest najważniejsze w tym rozdziale

Tak jak zwykle: Co to znaczy? Dlaczego jest to tutaj? Jak można tego użyć? Zastanów się nad przykładami.

Pojęcia i oznaczenia

Algebra Boole'a
- suma, iloczyn, dopełnienie
- zasada dualności
- porządek $\leq$
- atom
- izomorfizm

$\mathbb{B}$, $\mathbb{B}^n$

funkcja booleowska, $\text{BOOL}(n)$

wyrażenie booleowskie
- wyrażenia równoważne
- iloczyn minimalny, postać kanoniczna
- optymalna suma iloczynów symboli atomowych

sieć logiczna
- bramki NOT, AND, OR, NAND, NOR, XOR
- sieci równoważne

tablice Karnaugha, bloki

Fakty

Prawa algebry Boole'a podane w twierdzeniach 1 i 2 w § 10.1.
Własności relacji $\leq$ podane w lematach 1, 2 i 3 w § 10.1.
Niezerowe elementy skończonej algebry Boole'a wyrażają się jednoznacznie jako sumy atomów.
Każde dwie skończone algebry Boole'a mające n atomów są izomorficzne.
Każda sieć logiczna jest równoważna z siecią, w której występują tylko bramki NAND lub tylko bramki NOR.
Wyrażeniom booleowskim i sieciom logicznym odpowiadają etykietowane acykliczne grafy skierowane (por. § 10.3).
Optymalne wyrażenia booleowskie nie muszą odpowiadać najprostszym sieciom.
Wybór bloków koniecznych najpierw na tablicy Karnaugha, a potem zachłanne wybory największych pozostałych pokrywających bloków, nie muszą dać wyrażenia optymalnego.

Metody

Wyznaczanie postaci kanonicznej za pomocą odpowiedniej funkcji booleowskiej lub korzystając z praw algebry Boole'a.
Wykorzystanie tablic Karnaugha do znajdowania wyrażeń optymalnych równoważnych z danym wyrażeniem booleowskim.

11. WIĘCEJ O RELACJACH

W tym rozdziale kontynuujemy badanie relacji rozpoczęte dawno temu, w rozdziale 3. Rozsądnie byłoby pokrótce go przejrzeć dla przypomnienia sobie terminologii. Pierwsze dwa paragrafy tego rozdziału zajmują się relacjami, które porządkują elementy danego zbioru. Zaczniemy od częściowych porządków w przypadku ogólnym, a następnie zajmiemy się pewnymi szczególnymi relacjami porządku na zbiorach $S_1 \times \ldots \times S_n$ i Σ^*. W paragrafie 11.3 omawiamy pojęcie złożenia dowolnych relacji i pokazujemy, jak stwierdzenia dotyczące relacji można wyrażać w języku macierzy. W ostatnim paragrafie zajmujemy się pojęciem najmniejszej relacji, spełniającej różnorodne własności i zawierającej daną relację R w zbiorze S. Opisana jest tam, w szczególności, najmniejsza relacja równoważności zawierająca R.

Paragrafy 11.3 i 11.4 są niezależne od pierwszych dwóch paragrafów tego rozdziału i mogą być studiowane osobno.

§ 11.1. Zbiory częściowo uporządkowane

W tym paragrafie zajmiemy się zbiorami, których elementy można ze sobą w pewien sposób porównywać. W typowej sytuacji jeden z elementów będziemy traktować jako mniejszy od drugiego bądź jako występujący przed nim w sensie pewnej kolejności.

PRZYKŁAD 1 (a) Przyzwyczajeni jesteśmy do porównywania liczb rzeczywistych. Na przykład liczba 3 jest mniejsza od liczby 5, liczba -1 jest mniejsza od liczby 4 i liczba -1 jest większa od liczby -3.

Porównujemy dwie liczby stwierdzając, która z nich jest większa, a która mniejsza.

(b) Jeśli elementy jakiegoś zbioru S są wypisane z użyciem indeksów ze zbioru $\mathbb{P}$ lub $\mathbb{N}$ w taki sposób, że różne elementy mają różne indeksy, to możemy porównać dwa elementy zbioru S stwierdzając, który z nich ma mniejszy indeks. Różne sposoby indeksowania elementów zbioru S prowadzą do różnych sposobów ustalenia kolejności elementów S. Dany element raz może mieć najmniejszy indeks, innym razem zaś może być poprzedzony wieloma innymi elementami. ∎

Zbiór, którego elementy można w ten sposób porównywać, nazywamy **uporządkowanym**, a strukturę, która niesie informację, jak jego elementy ze sobą porównywać, nazywamy **relacją porządku** w zbiorze S. Aby móc powiedzieć cokolwiek pożytecznego o zbiorach uporządkowanych, musimy uczynić te definicje bardziej precyzyjnymi. Na początek wszakże zauważmy, że w przypadku wielu pojawiających się w sposób naturalny zbiorów wiemy, jak porównywać pewne elementy z innymi, ale jednocześnie mamy do czynienia z parami elementów nieporównywalnych.

PRZYKŁAD 2

(a) Gdybyśmy chcieli porównywać ze sobą marki samochodów, to moglibyśmy się, być może, zgodzić co do tego, że marka RR jest lepsza od marki H, ponieważ jest od niej lepsza pod każdym względem. Moglibyśmy jednak nie być w stanie powiedzieć, czy lepsza jest marka F, czy marka C, ponieważ każda z nich może pod pewnymi względami drugą przewyższać.

(b) Możemy się umówić, że dwie liczby ze zbioru $\{1, 2, 3, \ldots, 73\}$ są porównywalne, jeśli jedna z nich jest dzielnikiem drugiej. Wówczas liczby 6 i 72, a także 6 i 3, są porównywalne. Natomiast liczby 6 i 8 nie są porównywalne, gdyż żadna z nich nie jest dzielnikiem drugiej.

(c) Możemy porównać dwa podzbiory danego zbioru S (tzn. elementy zbioru $\mathcal{P}(S)$), jeśli jeden z nich jest zawarty w drugim. Jeśli zbiór S ma więcej niż jeden element, to zawiera jakieś nieporównywalne podzbiory. Na przykład, jeśli $s_1 \neq s_2$ oraz s_1 i s_2 należą do S, to zbiory $\{s_1\}$ i $\{s_2\}$ są nieporównywalne.

(d) Możemy porównywać funkcje punktowo. Jeśli, na przykład, funkcje f i g są określone na zbiorze S i przyjmują wartości w $\{0, 1\}$, to moglibyśmy traktować funkcję f jako mniejszą lub równą g, gdy $f(s) \leq g(s)$ dla każdego $s \in S$. To jest w gruncie rzeczy ten porządek, który nadaliśmy zbiorowi $\mathbb{B}^n$ w przykładzie 3

z § 10.1. Jest on podobny do sposobu porównywania samochodów z punktu (a). ∎

Zbiory, takie jak te z przykładu 2, z relacjami pozwalającymi porównywać ze sobą elementy, ale dopuszczającymi istnienie elementów nieporównywalnych, nazywamy częściowo uporządkowanymi. Tworzą one ważną klasę, którą teraz zdefiniujemy w ścisły sposób.

Przypomnij sobie, że relacja R w zbiorze S jest to podzbiór zbioru $S \times S$. **Częściowy porządek** w zbiorze S to relacja R, która jest zwrotna, antysymetryczna i przechodnia. Jeśli będziemy pisać $x \preceq y$ zamiast $(x, y) \in R$, to warunki te stwierdzają, że częściowy porządek spełnia:

(Z) $s \preceq s$ dla każdego s w S;
(AS) $s \preceq t$ i $t \preceq s$ implikują $s = t$;
(P) $s \preceq t$ i $t \preceq u$ implikują $s \preceq u$.

To są dokładnie te własności zwykłego porządku $\leq$, które wyróżniliśmy w przykładzie 4 z § 3.1.

Jeśli $\preceq$ jest częściowym porządkiem w zbiorze S, to parę $(S, \preceq)$ nazywamy **zbiorem częściowo uporządkowanym**. Używamy symbolu „$\preceq$" jako ogólnego oznaczenia dla częściowego porządku. Jeśli dla danego szczególnego częściowego porządku istnieje już oznaczenie, takie jak „$\leq$" lub „$\subseteq$", to będziemy raczej używali właśnie jego zamiast „$\preceq$".

W przykładzie 2 chodziło o relacje „jest nie tak dobry jak", „jest dzielnikiem", „jest podzbiorem" i „w żadnym punkcie nie jest większa". Mogliśmy równie dobrze rozważać relacje „jest co najmniej tak dobry jak", „jest wielokrotnością", „zawiera", i „w każdym punkcie jest co najmniej równa", ponieważ relacje te dostarczają tych samych informacji dotyczących porównywania elementów, co relacje przez nas wybrane. Każdy częściowy porządek w danym zbiorze wyznacza **relację odwrotną**, która wiąże x z y wtedy i tylko wtedy, gdy y pozostaje w relacji wyjściowej z x. Relacja odwrotna do częściowego porządku $\preceq$ jest na ogół oznaczana przez $\succeq$. Zatem $x \succeq y$ znaczy to samo, co $y \preceq x$. Ta relacja odwrotna również jest częściowym porządkiem (ćwiczenie 7(a)). Jeśli spojrzymy na relację $\preceq$ w zbiorze S jako na podzbiór R zbioru $S \times S$, to relacja $\succeq$ odpowiada relacji odwrotnej $R^{\leftarrow}$, zdefiniowanej w § 3.1.

Mając dany częściowy porządek $\preceq$ w zbiorze S, możemy zdefiniować jeszcze jedną relację, $\prec$, w S w następujący sposób:

$$x \prec y \quad \text{wtedy i tylko wtedy, gdy} \quad x \preceq y \text{ i } x \neq y.$$

Jeśli, na przykład, $\preceq$ oznacza inkluzję, to $A \prec B$ znaczy, że A jest właściwym podzbiorem B, tzn. $A \subset B$. Relacja $\prec$ jest przeciwzwrotna i przechodnia:

(PZ) $s \prec s$ nie zachodzi dla żadnego s w S;
(P) $s \prec t$ i $t \prec u$ implikują $s \preceq u$.

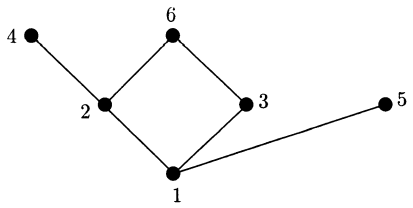
Relację przeciwzwrotną i przechodnią nazywamy **quasi-porządkiem**. Każdy częściowy porządek w zbiorze S wyznacza pewien quasi-porządek i, na odwrót, jeśli $\prec$ jest quasi-porządkiem w S, to relacja $\preceq$ zdefiniowana formułą

$$x \preceq y \quad \text{wtedy i tylko wtedy, gdy} \quad x \prec y \text{ lub } x = y$$

jest częściowym porządkiem w S (ćwiczenie 7(b)). To, czy do porównywania elementów danego zbioru częściowo uporządkowanego wybierze się częściowy porządek, czy też związany z nim quasi-porządek, zależy od problemu, z którym mamy w danym momencie do czynienia. Będziemy na ogół używali częściowego porządku, ale jeśli tak będzie wygodnie, będziemy również używać na zmianę obu relacji.

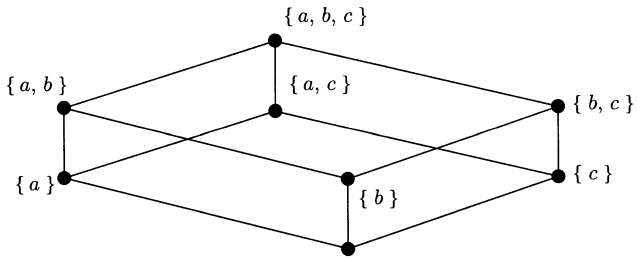
Możliwe jest, przynajmniej teoretycznie, narysowanie diagramu, który pozwala jednym rzutem oka objąć całą relację porządku w skończonym zbiorze częściowo uporządkowanym. Mając dany częściowy porządek $\preceq$ w S, powiemy, że element t **nakrywa** element s, gdy $s \prec t$ i nie ma w S elementu u takiego, że $s \prec u \prec t$. **Diagram Hassego** zbioru częściowo uporządkowanego $(S, \preceq)$ jest to rysunek grafu skierowanego, którego wierzchołkami są elementy zbioru S i w którym od wierzchołka t do wierzchołka s krawędź biegnie wtedy i tylko wtedy, gdy t nakrywa s. Diagramy Hassego, podobnie jak drzewa z wyróżnionym korzeniem, są zazwyczaj rysowane z krawędziami skierowanymi w dół i bez strzałek.

PRZYKŁAD 3 (a) Niech $S = \{1, 2, 3, 4, 5, 6\}$. Będziemy pisać $m|n$, gdy m dzieli n, tzn. gdy n jest całkowitą wielokrotnością m. Diagram z rysunku 11.1 jest diagramem Hassego zbioru częściowo upo-



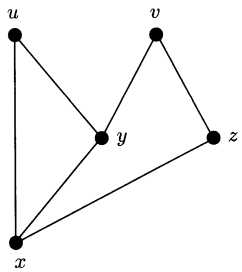
Rysunek 11.1

rządkowanego $(S, |)$. Między 1 i 6 nie ma krawędzi, ponieważ 6 nie nakrywa 1. Możemy wszakże wywnioskować z tego diagramu, że $1|6$, ponieważ relacja $|$ jest przechodnia i istnieje w diagramie łańcuch krawędzi odpowiadających temu, że $1|2$ i $2|6$. Podobnie, możemy zobaczyć, że $1|4$ biorąc pod uwagę drogę $1|2|4$. Zauważ, że, ogólnie, w przypadku relacji przechodnich stosować można takie łączone zapisy i nie prowadzi to do nieporozumień: $x \preceq y \preceq z$ znaczy to samo, co $x \preceq y$, $y \preceq z$ i $x \preceq z$.



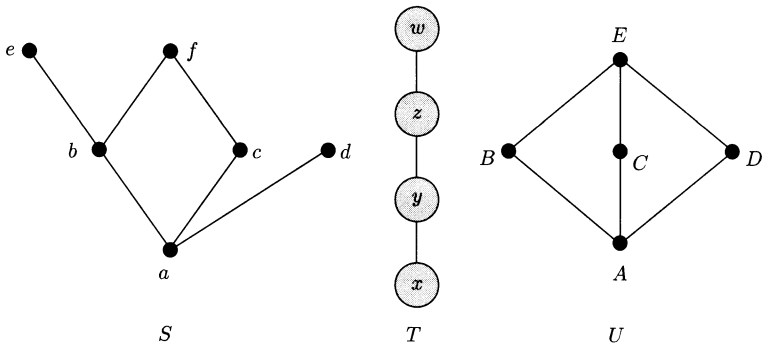
Rysunek 11.2

(b) Rozważmy zbiór potęgowy $\mathcal{P}(\{a, b, c\})$ z częściowym porządkiem $\subseteq$. Rysunek 11.2 przedstawia diagram Hassego zbioru $(\mathcal{P}(\{a, b, c\}), \subseteq)$. Zauważ, że linia łącząca $\{a, c\}$ z $\{a\}$ przecina sie z linią łączącą $\{a, b\}$ z $\{b\}$, ale przecięcie to jest po prostu cechą tego rysunku i jest bez znaczenia z punktu widzenia własności przedstawianego częściowego porządku. W szczególności, przecięcie tych dwóch linii nie reprezentuje elementu naszego zbioru.



Rysunek 11.3

(c) Diagram przedstawiony na rysunku 11.3 nie jest diagramem Hassego, ponieważ u nie może nakrywać x, skoro u nakrywa także y, a y nakrywa x. Gdyby którakolwiek z krawędzi łączących u, x i y została usunięta, rysunek stałby się diagramem Hassego.



Rysunek 11.4

(d) Diagramy przedstawione na rys. 11.4 są diagramami Hassego pewnych zbiorów częściowo uporządkowanych. Ich relacje porządku mogą być odczytane bezpośrednio z odpowiednich diagramów. Wszystkie elementy są w relacji same ze sobą. Ponadto:

Dla zbioru $S=\{a,b,c,d,e,f\}$ mamy $a \preceq b$, $a \preceq c$, $a \preceq d$, $a \preceq e$, $a \preceq f$, $b \preceq e$, $b \preceq f$ i $c \preceq f$. Widzieliśmy ten rysunek już wcześniej, w części (a) tego przykładu.

Dla zbioru $T = \{x,y,z,w\}$ mamy $x \preceq y$, $x \preceq z$, $x \preceq w$, $y \preceq z$, $y \preceq w$ i $z \preceq w$. Taki właśnie rysunek otrzymalibyśmy dla zbioru wszystkich dzielników liczby 8, albo 27 lub 125 z relacją porządku $|$.

Dla zbioru $U = \{A,B,C,D,E\}$ mamy $A \preceq B$, $A \preceq C$, $A \preceq D$, $A \preceq E$, $B \preceq E$, $C \preceq E$ i $D \preceq E$. Ten rysunek jest diagramem Hassego zbioru częściowo uporządkowanego składającego się ze zbiorów $\{1\}$, $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{1,2,3,4\}$ z inkluzją jako relacją porządku.

(e) Relacja $<$, którą zdefiniowaliśmy w algebrze Boole'a w § 10.1 jest quasi-porządkiem. Atomy są to dokładnie te elementy algebry, które nakrywają element 0. Zbiór częściowo uporządkowany $(\mathcal{P}(\{a,b,c\}),\subseteq)$, rozważany powyżej w części (b), stanowi przykład algebry Boole'a traktowanej jako zbiór częściowo uporządkowany. ∎

Ogólnie, jeśli dany jest diagram Hassego pewnego zbioru częściowo uporządkowanego, to widzimy, że $s \preceq t$, gdy bądź $s = t$, bądź istnieje (biegnąca w dół) droga od t do s. Pamiętając, że porządek jest zwrotny i przechodni, dowiemy się o nim wszystkiego na podstawie informacji dotyczących wzajemnego nakrywania elementów.

Fakt, że każdy skończony zbiór częściowo uporządkowany ma diagram Hassego, jest może intuicyjnie oczywisty, ale mimo to podamy jego dowód, opierając sie na własnościach acyklicznych grafów skierowanych.

Twierdzenie

Każdy skończony zbiór częściowo uporządkowany ma diagram Hassego.

Dowód. Dla danego zbioru częściowo uporządkowanego $(P, \preceq)$ niech H będzie grafem skierowanym o zbiorze wierzchołków P i w którym od wierzchołka x do wierzchołka y biegnie krawędź wtedy i tylko wtedy, gdy x nakrywa y. Typowa droga w grafie H ma ciąg wierzchołków $x_1x_2 \ldots x_{n+1}$, w którym x_1 nakrywa x_2, x_2 nakrywa x_3 itd., a więc $x_1 \succ x_2 \succ \ldots \succ x_{n+1}$. Na mocy przechodniości i antysymetrii, $x_1 \succ x_{n+1}$; w szczególności $x_1 \neq x_{n+1}$ i nasza droga nie jest zamknięta. Zatem H jest acyklicznym grafem skierowanym. Pokazaliśmy w paragrafach 7.3 i 8.1, że każdy acykliczny graf skierowany ma etykietowanie uporządkowane. Wybierając dla grafu H takie etykietowanie i robiąc jego rysunek w ten sposób, by wierzchołki z większymi liczbami były wyżej, otrzymamy diagram Hassego dla $(P, \preceq)$. ∎

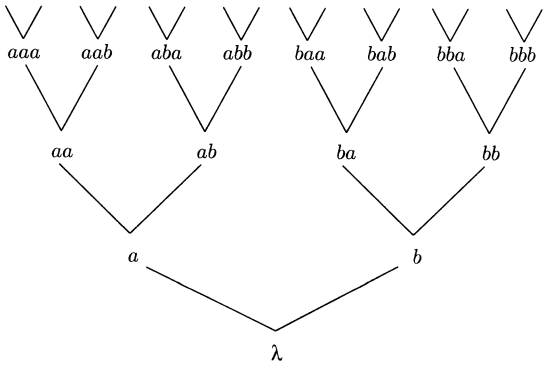
Niektóre nieskończone zbiory częściowo uporządkowane również mają diagramy Hassego. Diagram Hassego zbioru $\mathbb{Z}$ ze zwykłym porządkiem $\leq$ składa się z kropek rozmieszczonych w jednakowych odstępach wzdłuż pionowej prostej. Z drugiej strony, ponieważ żadna liczba rzeczywista nie nakrywa żadnej innej w sensie zwykłego porządku $\leq$, więc zbiór częściowo uporządkowany $(\mathbb{R}, \leq)$ nie ma diagramu Hassego.

PRZYKŁAD 4

(a) Mając dany alfabet Σ, możemy określić pewien częściowy porządek w nieskończonym zbiorze Σ^* wszystkich słów używających liter z Σ, w następujący sposób. Jeśli w_1 i w_2 są słowami z Σ^*, to $w_1 \preceq w_2$, jeśli w_1 jest **odcinkiem początkowym** w_2, tzn., jeśli w Σ^* istnieje słowo w takie, że $w_1w = w_2$. Na przykład mamy $ab \preceq abbaa$, gdyż $w_1w = w_2$, gdzie $w_1 = ab$, $w = baa$ i $w_2 = abbaa$. Zachodzi też $\lambda \preceq w$ dla każdego słowa w, gdyż $w = \lambda w$. Zauważmy, że $abbaa$ nie nakrywa ab, gdyż zarówno $u = abb$, jak i $u = abba$ spełniają $ab \prec u \prec abbaa$. Jednakże $abbaa$ nakrywa $abba$, $abba$ nakrywa abb, a abb nakrywa ab. Ogólnie, jeśli w_2 nakrywa w_1, to długość(w_2) = 1 + długość(w_1).

Dla Σ=$\{a, b\}$, część diagramu Hassego dla $(\Sigma^*, \preceq)$ jest przedstawiona na rysunku 11.5. Ten diagram Hassego jest drzewem.

W paragrafie 6.4 (przykład 7) potraktowaliśmy ten diagram jak drzewo z wyróżnionym korzeniem i obowiązująca tradycja zmusiła nas do narysowania go do góry nogami.



Rysunek 11.5

(b) W dowolnym skończonym drzewie T z wyróżnionym korzeniem istnieje naturalny porządek, w którym jego korzeń r jest elementem największym. Zdefiniujmy relację $\preceq$ w zbiorze V wszystkich wierzchołków T mówiąc, że $v \preceq w$, gdy $v = w$ lub w należy do (jedynej) drogi biegnącej z r do v. Łatwo jest sprawdzić, że $\preceq$ jest częściowym porządkiem w V; zob. ćwiczenie 8. Jednym z diagramów Hassego dla $(V, \preceq)$ jest wyjściowe drzewo T, narysowane, jak zazwyczaj, z korzeniem na górze i gałęziami biegnacymi w dół. Rysunki drzew z wyróżnionymi korzeniami można traktować jak rysunki pewnych dość szczególnych zbiorów częściowo uporządkowanych.

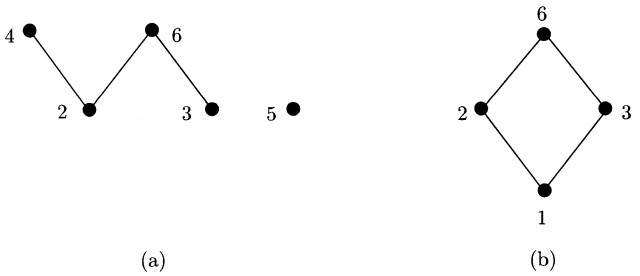
(c) Przykłady z punktów (a) i (b) opisywały drzewa z korzeniami na samym dole lub na samej górze. Istnieje naturalny związek między tymi dwoma sposobami opisu. Diagram Hassego relacji $\succeq$, odwrotnej do relacji $\preceq$ w zbiorze częściowo uporządkowanym S, jest to po prostu diagram Hassego dla $(S, \preceq)$, odwrócony do góry nogami. Powodem tego jest fakt, że y nakrywa x w sensie relacji wyjściowej wtedy i tylko wtedy, gdy x nakrywa y w sensie relacji odwrotnej. A zatem krawędzie w diagramie dla $(S, \succeq)$ skierowane są przeciwnie niż krawędzie w diagramie dla $(S, \preceq)$. ∎

Elementy odpowiadające punktom znajdującym się na diagramie Hassego w pobliżu samej góry bądź samego dołu okazują się ważne. Jeśli $(P, \preceq)$ jest zbiorem częściowo uporządkowanym, to element x zbioru P nazywamy **elementem maksymalnym**,

gdy w P nie istnieje element y taki, że $x \prec y$. Element x nazywamy **elementem minimalnym**, jeśli w P nie istnieje element y taki, że $y \prec x$. W zbiorach częściowo uporządkowanych, których diagramy Hassego przedstawia rysunek 11.4, elementy d, e, f, w i E są maksymalne, podczas gdy a, x i A są minimalne. Nieskończony zbiór częściowo uporządkowany z rysunku 11.5 nie ma elementów maksymalnych; puste słowo λ jest jego jedynym elementem minimalnym.

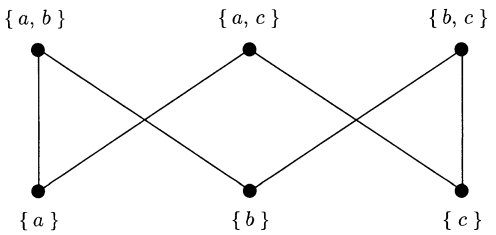
Dowolny podzbiór S częściowo uporządkowanego zbioru P dziedzicząc częściowy porządek z P sam staje się zbiorem częściowo uporządkowanym, gdyż prawa (Z), (AS) i (P) dotyczą wszystkich elementów zbioru P. Ten częściowy porządek zbioru S **jest zawarty** w częściowym porządku zbioru P.

PRZYKŁAD 5 (a) Zbiory $\{2,3,4,5,6\}$ i $\{1,2,3,6\}$ traktowane jako podzbiory zbioru częściowo uporządkowanego $\{1,2,3,4,5,6\}$, rozważanego w przykładzie 3(a), mają diagramy Hassego przedstawione na rysunku 11.6. (Zauważ, gdzie umieszczone są na rysunku 11.6(a) liczby pierwsze.)



Rysunek 11.6

(b) Zbiór wszystkich niepustych właściwych podzbiorów zbioru $\{a,b,c\}$, traktowany jako podzbiór zbioru $\mathcal{P}(\{a,b,c\})$ z częściowym porządkiem $\subseteq$, ma diagram Hassego przedstawiony na rysunku 11.7. Porównaj go z rysunkiem 11.2. ∎



Rysunek 11.7

Jeśli S jest podzbiorem zbioru częściowo uporządkowanego $(P, \preceq)$, to może się zdarzyć, że S ma element M taki, że $s \preceq M$ dla każdego s z S. Na rysunku 11.6(b) $s \preceq 6$ dla każdego s, podczas gdy na rysunkach 11.6(a) i 11.7 żadnego takiego elementu M nie ma. Element M o tej własności nazywamy **elementem największym** w zbiorze S lub **maksimum** zbioru S i oznaczamy go przez $\max(S)$.(Istnieje co najwyżej jeden taki element M; dlaczego?) Jest to zgodne z naszym dotychczasowym oznaczaniem przez $\max\{m, n\}$ większej z dwóch liczb m i n. Podobnie, jeśli zbiór S ma element m taki, że $m \preceq s$ dla każdego s w S, to m nazywamy **elementem najmniejszym** zbioru S lub **minimum** zbioru S i oznaczamy go przez $\min(S)$.

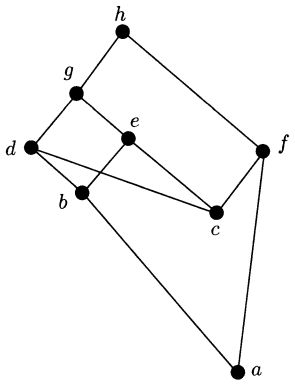
PRZYKŁAD 6 (a) Rozważmy ponownie zbiór częściowo uporządkowany $(\{1, 2, 3, 4, 5, 6\}, |)$, przedstawiony na rysunku 11.1. Zbiór ten nie ma elementu największego, tzn. maksimum, pomimo, że 4, 6 i 5 są jego elementami maksymalnymi. Liczba 1 jest jego minimum i jest to zarazem jedyny element minimalny. Podzbiór $\{2, 3\}$ nie ma elementu największego; liczba 3 jest większa od liczby 2 w sensie zwykłego porządku, ale nie w sensie porządku, który rozpatrujemy.

(b) Jeśli algebrę Boole'a traktujemy jako zbiór częściowo uporządkowany, to jej elementem największym jest 1, a najmniejszym 0. ∎

Niezależnie od tego, czy podzbiór S zbioru częściowo uporządkowanego $(P, \preceq)$ ma element największy, czy nie, może istnieć element x w zbiorze P taki, że $s \preceq x$ dla każdego s z S. (Na przykład, oba elementy zbioru $\{2, 3\}$ z przykładu 6 dzielą liczbę 6.) Taki element x nazywamy **ograniczeniem górnym** zbioru S w zbiorze P. Jeśli x jest ograniczeniem górnym S w P takim, że $x \preceq y$ dla każdego ograniczenia górnego y zbioru S w P, to x nazywamy **kresem górnym** lub **supremum** S w P i piszemy $x = \sup(S)$. Podobnie, element z w P taki, że $z \preceq s$ dla każdego s w S, jest **ograniczeniem dolnym** zbioru S w zbiorze P. Ograniczenie dolne z takie, że $w \preceq z$ dla każdego ograniczenia dolnego w nazywamy **kresem dolnym** lub **infimum** S w P i oznaczamy przez $\inf(S)$. Na mocy prawa antysymetryczności (AS), podzbiór zbioru P nie może mieć dwóch różnych kresów górnych, ani dwóch różnych kresów dolnych.

PRZYKŁAD 7 (a) W zbiorze częściowo uporządkowanym $(\{1, 2, 3, 4, 5, 6\}, |)$, podzbiór $\{2, 3\}$ ma dokładnie jedno ograniczenie górne, a mianowicie 6; zatem $\sup\{2, 3\} = 6$. Podobnie $\inf\{2, 3\} = 1$. Podzbiór

$\{4,6\}$ nie ma w ogóle ograniczeń górnych w naszym zbiorze; jego ograniczeniami dolnymi są 2 i 1, a więc $\inf\{4,6\} = 2$. Jedynym ograniczeniem górnym podzbioru $\{3,6\}$ jest 6, a ograniczeniami dolnymi są 3 i 1; stąd $\sup\{3,6\} = 6$ i $\inf\{3,6\} = 3$. Zatem kres górny lub kres dolny podzbioru może istnieć lub nie, a jeśli istnieje, to może do danego podzbioru należeć lub też nie. Dla rozważanego zbioru częściowo uporządkowanego kresy dolne to największe wspólne dzielniki, a kresy górne to najmniejsze wspólne wielokrotności, o ile należą one do naszego zbioru.



Rysunek 11.8

(b) W zbiorze częściowo uporządkowanym P przedstawionym na rysunku 11.8 ograniczeniami górnymi podzbioru $\{b,c\}$ są elementy d, e, g i h. Element h jest też ograniczeniem górnym podzbioru $\{d,f\}$. Zbiór $\{b,c\}$ nie ma kresu górnego w P (dlaczego?), ale $h = \sup\{d,f\}$. Elementy a i c są ograniczeniami dolnymi zbioru $\{d,e,f\}$, który nie ma kresu dolnego, ponieważ elementy a i c są nieporównywalne. Element a jest kresem dolnym zbioru $\{b,d,e,f\}$. ∎

Wiele spośród pojawiających się w praktyce zbiorów częściowo uporządkowanych ma tę własność, że dla każdego dwuelementowego podzbioru istnieje zarówno kres górny, jak i kres dolny. **Krata** jest to taki zbiór częściowo uporządkowany, w którym $\sup\{x,y\}$ i $\inf\{x,y\}$ istnieją dla każdego x i y. Dla kraty $(P,\preceq)$ równości

$$x \vee y = \sup\{x,y\} \quad \text{oraz} \quad x \wedge y = \inf\{x,y\}$$

definiują dwuargumentowe działania $\vee$ i $\wedge$ w zbiorze P. Jak zobaczymy w następnym przykładzie, oznaczenie to jest zgodne z tym, które wprowadzone zostało w § 10.1 dla algebr Boole'a.

Zauważ, że $\inf\{x,y\} = x \wedge y = x$ wtedy i tylko wtedy, gdy $x \preceq y$ wtedy i tylko wtedy, gdy $\sup\{x,y\} = x \vee y = y$. W szczególności, możemy odtworzyć relację porządku $\preceq$, jeśli tylko znamy którekolwiek z dwuargumentowych działań $\wedge$ lub $\vee$; zob. ćwiczenie 12. Można dowieść indukcyjnie (ćwiczenie 19(b)), że każdy skończony podzbiór dowolnej kraty ma zarówno kres górny, jak i kres dolny.

PRZYKŁAD 8 Zbiór częściowo uporządkowany $(\mathcal{P}(\{a,b,c\}),\subseteq)$ przedstawiony na rysunku 11.2 jest kratą. Na przykład,

$$\sup(\{a\},\{c\}) = \{a\} \vee \{c\} = \{a,c\},$$
$$\sup(\{a,b\},\{a,c\}) = \{a,b\} \vee \{a,c\} = \{a,b,c\},$$
$$\inf(\{a,b\},\{c\}) = \{a,b\} \wedge \{c\} = \emptyset$$

oraz

$$\inf(\{a,b\},\{b,c\}) = \{a,b\} \wedge \{b,c\} = \{b\}.$$

Ogólnie, dla każdego zbioru S, zbiór częściowo uporządkowany $(\mathcal{P}(S),\subseteq)$ jest kratą, przy czym $\sup(A,B) = A \cup B$ i $\inf(A,B) = A \cap B$, a stąd

$$\sup(A,B,\ldots,Z) = A \cup B \cup \ldots \cup Z$$

oraz

$$\inf(A,B,\ldots,Z) = A \cap B \cap \ldots \cap Z.$$

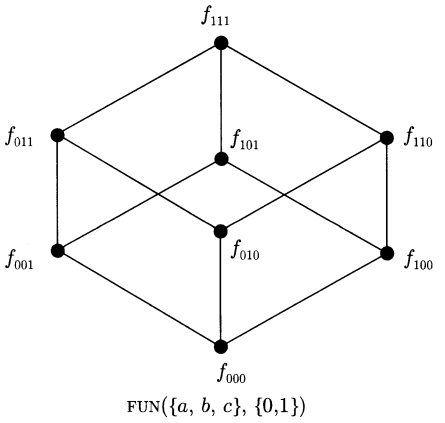
Zbiór częściowo uporządkowany przedstawiony na rysunku 11.7 nie jest kratą; na przykład elementy $\{a,b\}$ i $\{a,c\}$ nie mają kresu górnego w tym zbiorze.

(b) Zdefiniujmy częściowy porządek $|$ w zbiorze $\mathbb{P}$ mówiąc, że $m|n$ wtedy i tylko wtedy, gdy m dzieli n. Zbiór $S = \{1,2,3,4,5,6\}$, traktowany jako podzbiór zbioru częściowo uporządkowanego $\mathbb{P}$, nie jest kratą, gdyż zbiór $\{3,4\}$ nie ma kresu górnego w S. Jednakże cały zbiór częściowo uporządkowany $(\mathbb{P},|)$ jest kratą. Ograniczeniem górnym dla $\{m,n\}$ jest każda liczba k w $\mathbb{P}$ taka, że m dzieli k i n dzieli k, tzn. każda wspólna wielokrotność liczb m i n. Kresem górnym $\sup(m,n)$ jest **najmniejsza wspólna wielokrotność** m i n. Podobnie, kresem dolnym $\inf(m,n)$ jest **największy wspólny dzielnik** liczb m i n, tzn. największa dodatnia liczba całkowita, która dzieli zarówno m, jak i n. I tak na przykład, $\sup\{12,10\} = 60$ i $\inf\{12,10\} = 2$. Liczby $\sup(m,n)$ i $\inf(m,n)$ można znaleźć posługując się rozkładami m i n na iloczyny liczb pierwszych. Liczby pierwsze są minimalnymi elementami w zbiorze częściowo uporządkowanym $\mathbb{P}\backslash\{1\}$, tzn. są to liczby, które nakrywają 1 w $\mathbb{P}$.

(c) Rozważmy zbiór FUN$(\{a,b,c\},\{0,1\})$ wszystkich funkcji z trzyelementowego zbioru $\{a,b,c\}$ w $\{0,1\}$. Definiujemy pewien częściowy porządek $\leq$ w tym zbiorze mówiąc, że

$$f \leq g \text{ wtedy i tylko wtedy, gdy } f(x) \leq g(x) \text{ dla } x = a,b,c.$$

Wygodnie będzie oznaczyć osiem funkcji, z których składa się nasz zbiór, indeksami takimi, jak 101, które ukazują kolejno wartości przyjmowane przez daną funkcję w punktach a, b i c. Na przykład f_{101} reprezentuje funkcję taką, że $f_{101}(a) = 1$, $f_{101}(b) = 0$ i $f_{101}(c) = 1$. Diagram Hassego zbioru częściowo uporządkowanego $(\text{FUN}(\{a,b,c\},\{0,1\}),\leq)$ przedstawiony jest na rysunku 11.9. Ten zbiór częściowo uporządkowany jest kratą. Jest to w gruncie rzeczy algebra Boole'a BOOL(3) funkcji boleowskich trzech zmiennych.



Rysunek 11.9

(d) W paragrafie 10.1 zaczęliśmy od algebry Boole'a $(B,\vee,\wedge,')$ a następnie wprowadziliśmy relację $\leq$ mówiąc, że $x \leq y$ wtedy i tylko wtedy, gdy $x \vee y = y$ wtedy i tylko wtedy, gdy $x \wedge y = x$. Lemat 2 z tamtego paragrafu mówi, że $\leq$ jest częściowym porządkiem. Pokażemy teraz, że w sensie porządku $\leq$, element $a \wedge b$ jest kresem dolnym zbioru $\{a,b\}$; podobnie, element $a \vee b$ jest równy $\sup\{a,b\}$. Skorzystamy tu po prostu z algebraicznych własności działań $\wedge$ i $\vee$.

Po pierwsze, $a \wedge b \leq a$, ponieważ $(a \wedge b) \wedge a = a \wedge b$; podobnie, $a \wedge b \leq b$. Zatem $a \wedge b$ jest ograniczeniem dolnym dla $\{a,b\}$. Jeśli jednocześnie $c \leq a$ i $c \leq b$, to $c \wedge a = c$ i $c \wedge b = c$. Wynika stąd, że $c \wedge (a \wedge b) = (c \wedge a) \wedge b = c \wedge b = c$, a więc $c \leq a \wedge b$. Wobec tego $a \wedge b$ jest kresem dolnym $\{a,b\}$. ∎

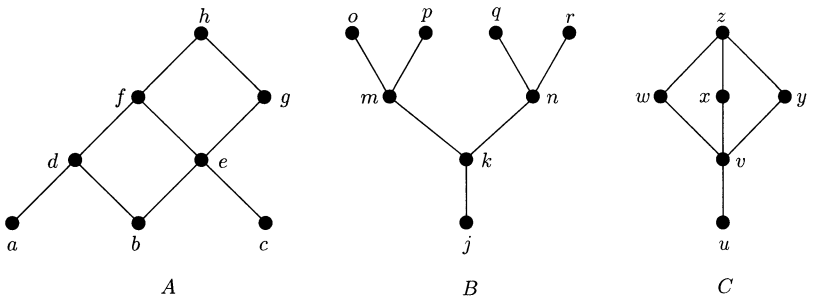
W większości zbiorów częściowo uporządkowanych, które napotkaliśmy w tym paragrafie istniały pary elementów nieporównywalnych. W następnym paragrafie zajmiemy się takimi zbiorami częściowo uporządkowanymi, w których każdy element daje się porównać z dowolnym innym elementem. Pokażemy również w jaki sposób za pomocą porządków na stosunkowo prostych zbiorach można skonstruować porządki na zbiorach bardziej skomplikowanych.

ĆWICZENIA DO § 11.1

1. Narysuj diagramy Hassego następujących zbiorów częściowo uporządkowanych:
(a) $(\{1, 2, 3, 4, 6, 8, 12, 24\}, |)$, gdzie $m|n$ oznacza, że m jest dzielnikiem (tzn. dzieli) n.
(b) Zbiór wszystkich podzbiorów zbioru $\{3, 7\}$ z relacją $\subseteq$ jako częściowym porządkiem.

2. (a) Podaj przykłady dwóch zbiorów częściowo uporządkowanych wziętych bądź z codziennego życia, bądź z innych wykładów.
(b) Czy zbiory z twoich przykładów mają elementy maksymalne lub minimalne? Jeśli tak, to jakie?
(c) Jak wyglądają relacje odwrotne do częściowych porządków z twoich przykładów?

3. Rysunek 11.10 przedstawia diagramy Hassego trzech zbiorów częściowo uporządkowanych.



Rysunek 11.10

(a) Jakie elementy maksymalne mają te zbiory?
(b) W których spośród tych zbiorów istnieją elementy minimalne?
(c) Które spośród tych zbiorów mają elementy najmniejsze?
(d) Które elementy nakrywają element e?

(e) Znajdź następujące elementy, o ile istnieją:

$$\sup\{d, c\}, \quad \sup\{w, y, v\}, \quad \sup\{p, m\}, \quad \inf\{a, g\}.$$

(f) Które z tych zbiorów częściowo uporządkowanych są kratami?

4. Znajdź maksymalne podzbiory właściwe trzyelementowego zbioru $\{a, b, c\}$. To znaczy, znajdź elementy maksymalne podzbioru zbioru $\mathcal{P}(\{a, b, c\})$, częściowo uporządkowanego przez $\subseteq$, złożonego z właściwych podzbiorów zbioru $\{a, b, c\}$.

5. Rozważmy zbiór $\mathbb{R}$ wraz ze zwykłym porządkiem $\leq$.
(a) Czy $\mathbb{R}$ jest kratą? Jeśli jest, to jak wyglądają w $\mathbb{R}$ działania $a \vee b$ i $a \wedge b$?
(b) Podaj przykład niepustego podzbioru zbioru $\mathbb{R}$, który nie ma ograniczenia górnego w $\mathbb{R}$.
(c) Znajdź $\sup\{x \in \mathbb{R}\colon x < 73\}$.
(d) Znajdź $\sup\{x \in \mathbb{R}\colon x \leq 73\}$.
(e) Znajdź $\sup\{x \in \mathbb{R}\colon x^2 < 73\}$.
(f) Znajdź $\inf\{x \in \mathbb{R}\colon x^2 < 73\}$.

6. Niech S będzie zbiorem procedur pewnego programu komputerowego. Dla A i B ze zbioru S będziemy pisać, że $A \prec B$, jeśli procedura A musi zostać zakończona, zanim można będzie zakończyć procedurę B. Jakie ograniczenia trzeba nałożyć na wywoływanie procedur w tym programie, by relacja $\prec$ stała się quasi-porządkiem w zbiorze S?

7. (a) Wykaż, że jeśli $\preceq$ jest częściowym porządkiem w zbiorze S, to jest nim również relacja $\succeq$ odwrotna do $\preceq$.
(b) Wykaż, że jeśli $\prec$ jest quasi-porządkiem w zbiorze S, to relacja $\preceq$ zdefiniowana w następujący sposób:

$$x \preceq y \quad \text{wtedy i tylko wtedy, gdy} \quad x \prec y \text{ lub } x = y$$

jest częściowym porządkiem w zbiorze S.

8. Niech G będzie acyklicznym grafem skierowanym. Wykaż, że relacja $\prec$, zdefiniowana w następujący sposób: $u \prec v$, jeśli istnieje droga z u do v, jest quasi-porządkiem w zbiorze $V(G)$. Wobec tego, na mocy ćwiczenia 7(b), relacja $\preceq$ z przykładu 4(b) jest częściowym porządkiem.

9. Sprawdź, że częściowy porządek $\preceq$ w zbiorze Σ^*, z przykładu 4(a), jest relacją zwrotną i przechodnią.

10. Niech Σ będzie pewnym alfabetem. Dla w_1, $w_2 \in \Sigma^*$ powiemy, że $w_1 \preceq w_2$, jeśli w Σ^* istnieją słowa w i w' takie, że $w_2 = ww_1w'$. Czy $\preceq$ jest częściowym porządkiem w zbiorze Σ^*? Uzasadnij swoją odpowiedź.

11. Niech Σ będzie pewnym alfabetem. Dla w_1, $w_2 \in \Sigma^*$ niech $w_1 \preceq w_2$ znaczy, że długość$(w_1) \leq$ długość(w_2). Czy $\preceq$ jest częściowym porządkiem w zbiorze Σ^*? Uzasadnij swoją odpowiedź.

12. Tabelka przedstawiona poniżej została częściowo wypełniona. Podane są w niej wartości działań $x \vee y$ dla niektórych elementów x i y pewnej kraty $(L, \preceq)$. Na przykład $b \vee c = d$.
(a) Wypełnij pozostałą część tabelki.
(b) Wskaż element największy i element najmniejszy w L?
(c) Wykaż, że $f \preceq c \preceq d \preceq e$.
(d) Narysuj diagram Hassego dla L.

$\vee$	a	b	c	d	e	f
a		e	a	e	e	a
b			d	d	e	b
c				d	e	c
d					e	d
e						e
f						

13. Niech $\mathcal{F}(\mathbb{N})$ oznacza rodzinę wszystkich skończonych podzbiorów zbioru $\mathbb{N}$. Wówczas $(\mathcal{F}(\mathbb{N}), \subseteq)$ jest zbiorem częściowo uporządkowanym.
(a) Czy zbiór $\mathcal{F}(\mathbb{N})$ ma elementy maksymalne? Jeśli tak, to wskaż jeden z nich. Jeśli nie, to uzasadnij dlaczego.
(b) Czy zbiór $\mathcal{F}(\mathbb{N})$ ma elementy minimalne? Jeśli tak, to wskaż jeden z nich. Jeśli nie, to uzasadnij dlaczego.
(c) Czy dla dowolnych elementów A, B z $\mathcal{F}(\mathbb{N})$ istnieje kres górny zbioru $\{A, B\}$ w $\mathcal{F}(\mathbb{N})$? Jeśli tak, to określ go. Jeśli nie, to podaj konkretny kontrprzykład.
(d) Czy dla dowolnych elementów A, B z $\mathcal{F}(\mathbb{N})$ istnieje kres dolny zbioru $\{A, B\}$ w $\mathcal{F}(\mathbb{N})$? Jeśli tak, to określ go. Jeśli nie, to podaj konkretny kontrprzykład.
(e) Czy $\mathcal{F}(\mathbb{N})$ jest kratą? Uzasadnij swoją odpowiedź.

14. Powtórz ćwiczenie 13 dla rodziny $\mathcal{T}(\mathbb{N})$ wszystkich nieskończonych podzbiorów zbioru $\mathbb{N}$.

15. Określamy relacje $<$, $\leq$ oraz $\preceq$ w zbiorze $\mathbb{R} \times \mathbb{R}$ wszystkich punktów płaszczyzny w następujący sposób:

$$(x, y) < (z, w), \quad \text{jeśli} \quad x^2 + y^2 < z^2 + w^2;$$
$$(x, y) \leq (z, w), \quad \text{jeśli} \quad (x, y) < (z, w) \text{ lub } (x, y) = (z, w);$$
$$(x, y) \preceq (z, w), \quad \text{jeśli} \quad x^2 + y^2 \leq z^2 + w^2.$$

(a) Które z tych relacji są częściowymi porządkami? Odpowiedź uzasadnij.
(b) Które z nich są quasi-porządkami? Odpowiedź uzasadnij.
(c) Wyznacz graficznie zbiór $\{(x, y)\colon (x, y) \leq (3, 4)\}$.
(d) Wyznacz graficznie zbiór $\{(x, y)\colon (x, y) \preceq (3, 4)\}$.

16. Niech $\mathcal{E}(\mathbb{N})$ będzie zbiorem tych wszystkich skończonych podzbiorów zbioru $\mathbb{N}$, które mają parzystą liczbę elementów. Rozważamy $\mathcal{E}(\mathbb{N})$ z częściowym porządkiem $\subseteq$.

(a) Niech $A = \{1, 2\}$ i $B = \{1, 3\}$. Znajdź cztery ograniczenia górne zbioru $\{A, B\}$ w $\mathcal{E}(\mathbb{N})$.
(b) Czy zbiór $\{A, B\}$ ma kres górny w $\mathcal{E}(\mathbb{N})$? Odpowiedź uzasadnij.
(c) Czy $\mathcal{E}(\mathbb{N})$ jest kratą?

17. Czy każdy podzbiór kraty jest kratą? Odpowiedź uzasadnij.

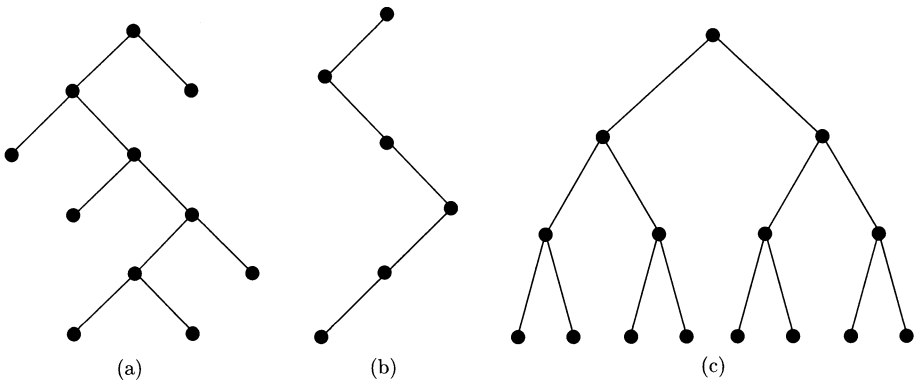
18. (a) Wykaż, że w każdym skończonym zbiorze częściowo uporządkowanym istnieje element minimalny. *Wskazówka*: zastosuj indukcję.
(b) Podaj przykład zbioru częściowo uporządkowanego, który ma element maksymalny, ale nie ma elementu minimalnego.

19. (a) Weź elementy x, y, z danego zbioru częściowo uporządkowanego. Wykaż, że jeśli $\sup\{x, y\} = a$ i $\sup\{a, z\} = b$, to $\sup\{x, y, z\} = b$.
(b) Wykaż, że każdy skończony niepusty podzbiór dowolnej kraty ma kres górny.
(c) Wykaż, że jeśli x, y i z są elementami kraty, to $(x \vee y) \vee z = x \vee (y \vee z)$.

20. Rozważmy zbiór częściowo uporządkowany C, którego diagram Hassego jest przedstawiony na rysunku 11.10. Wykaż, że $w \vee (x \wedge y) \neq (w \vee x) \wedge (w \vee y)$ oraz $w \wedge (x \vee y) \neq (w \wedge x) \vee (w \wedge y)$. Ten przykład pokazuje, że kraty nie muszą spełniać praw „rozdzielności" dla $\vee$ i $\wedge$.

§ 11.2. Szczególne porządki

Zbiory częściowo uporządkowane pojawiają się w różnych sytuacjach i w wielu przypadkach fakt istnienia par elementów, których nie można porównać ze sobą jest ich istotną cechą. Jak widzieliśmy w poprzednim paragrafie, drzewa z wyróżnionym korzeniem można uważać za diagramy Hassego, takie jak te przedstawione na rysunku 11.11, w których $\sup(x, y)$ istnieje dla każdych



Rysunek 11.11

elementów x i y, ale $\inf(x, y)$ istnieje jedynie wtedy, gdy $x \preceq y$ lub $y \preceq x$. Drzewa są użytecznymi strukturami danych, pomimo że istnieją w nich pary elementów nieporównywalnych, ponieważ zaczynając od korzenia i poruszając się zgodnie z porządkiem można dotrzeć dość szybko do dowolnego elementu.

Listy, czyli ciągi stanowią inną ważną klasę struktur danych. Niezależnie od tego, które dwa elementy wybierzemy z listy, zawsze jeden z nich występuje na liście przed drugim. Struktura taka jest przykładem **łańcucha** lub zbioru liniowo uporządkowanego, który definiujemy jako zbiór częściowo uporządkowany, w którym każde dwa elementy są porównywalne. Częściowy porządek nazywamy **liniowym porządkiem**, jeśli dla wszystkich elementów s i t z S, bądź $s \preceq t$, badź też $t \preceq s$. Zatem łańcuch jest to zbiór częściowo uporządkowany przez liniowy porządek. Terminy „łańcuch " i „zbiór liniowo uporządkowany" będą używane wymiennie.

PRZYKŁAD 1 (a) Zbiór częściowo uporządkowany z rysunku 11.11(b) jest łańcuchem, ale pozostałe zbiory częściowo uporządkowane, przedstawione na tym rysunku, nie są łańcuchami.

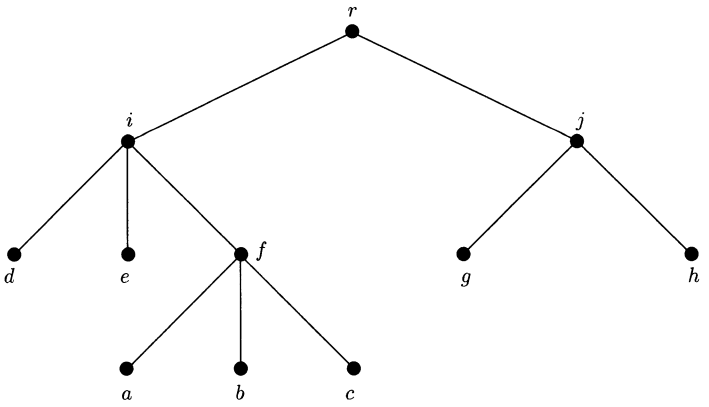
(b) Zbiór $\mathbb{R}$ ze zwykłym porządkiem jest zbiorem liniowo uporządkowanym.

(c) Lista nazwisk w książce telefonicznej lub słów w słowniku jest łańcuchem, jeśli powiemy, że $w_1 \preceq w_2$ znaczy, że $w_1 = w_2$ lub w_1 występuje wcześniej niż w_2. ∎

Każdy podzbiór łańcucha sam jest łańcuchem. Na przykład zbiory $\mathbb{Z}$ i $\mathbb{Q}$ jako podzbiory zbioru $\mathbb{R}$ są liniowo uporządkowane przez porządki odziedziczone z $\mathbb{R}$. Słowa w słowniku, znajdujące się między słowem „start" a słowem „stop" tworzą łańcuch, będący fragmentem łańcucha wszystkich słów z przykładu 1(c).

Każdy zbiór częściowo uporządkowany, niezależnie od tego, czy sam jest liniowo uporządkowany, czy nie, zawiera podzbiory będące łańcuchami. Pewna znajomość tych podzbiorów często okazuje się przydatna.

PRZYKŁAD 2 (a) Niech S będzie zbiorem osób zgromadzonych na pewnym rodzinnym zjeździe; powiemy, że $m \prec n$, jeśli m jest potomkiem n. Wówczas $\prec$ jest quasi-porządkiem i wyznacza on częściowy porządek $\preceq$ zdefiniowany w następujący sposób: $m \preceq n$ wtedy i tylko wtedy, gdy $m \prec n$ lub $m = n$. Łańcuch w zbiorze częściowo uporządkowanym $(S, \preceq)$ będzie zbiorem postaci $\{m, n, p, \ldots, r\}$, w którym m jest potomkiem n, n jest potomkiem p itd. Byłoby



Rysunek 11.12

niezwykłe, gdyby łańcuch taki miał więcej niż 5 elementów, chociaż jednocześnie sam zbiór S mógłby być całkiem duży.

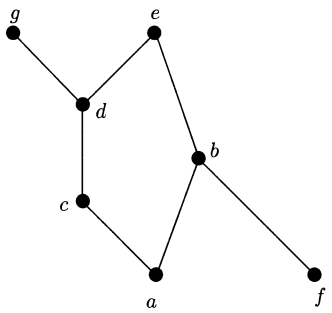
(b) Diagram Hassego pokazany na rysunku 11.12 przedstawia zbiór częściowo uporządkowany ze znaczną liczbą łańcuchów (49, wliczając łańcuchy jednoelementowe, ale nie wliczając łańcucha pustego). ■

Czasem interesują nas takie łańcuchy w zbiorach częściowo uporządkowanych, których nie da się już powiększyć. Zauważmy, że jeśli $(S, \preceq)$ jest zbiorem częściowo uporządkowanym i $\mathcal{C}(S)$ jest zbiorem wszystkich łańcuchów w S, to $(\mathcal{C}(S), \subseteq)$ jest również zbiorem częściowo uporządkowanym. **Łańcuch maksymalny** w zbiorze S jest zdefiniowany jako element maksymalny w zbiorze $\mathcal{C}(S)$, tzn. taki łańcuch, który nie jest właściwym podzbiorem żadnego innego łańcucha.

PRZYKŁAD 3 (a) W zbiorze częściowo uporządkowanym z rysunku 11.12 łańcuchami maksymalnymi są zbiory $\{a, f, i, r\}$, $\{b, f, i, r\}$, $\{c, f, i, r\}$, $\{d, i, r\}$, $\{e, i, r\}$, $\{g, j, r\}$ i $\{h, j, r\}$. Zauważmy, że te łańcuchy maksymalne nie mają tej samej liczby elementów.

(b) W zbiorze częściowo uporządkowanym przedstawionym na rysunku 11.13, jedyne dwa łańcuchy maksymalne $\{a, c, d, e\}$ i $\{a, b, e\}$, do których należą elementy a i e, mają różne liczby elementów. Ten zbiór częściowo uporządkowany ma w sumie cztery łańcuchy maksymalne.

(c) W drzewie z wyróżnionym korzeniem drogi wiodące od korzenia do liści odpowiadają maksymalnym łańcuchom, w odniesieniu do zwykłego częściowego porządku w tym drzewie. ■



Rysunek 11.13

Skończony łańcuch musi mieć element najmniejszy i to samo dotyczy również każdego jego niepustego podzbioru. Z drugiej strony, łańcuchy nieskończone mogą się zachowywać w różnorodny sposób. Nieskończone zbiory liniowo uporządkowane $(\mathbb{R}, \leq)$ i $(\mathbb{Z}, \leq)$, z ich zwykłymi porządkami, nie mają elementów najmniejszych. Łańcuch $(\{x \in \mathbb{R}\colon 0 \leq x\}, \leq)$ ma element najmniejszy, a mianowicie 0, ale zawiera podzbiory takie, jak $\{x \in \mathbb{R}\colon 1 < x\}$, które elementów najmniejszych nie mają. Nieskończony zbiór liniowo uporządkowany $(\mathbb{N}, \leq)$ ma element najmniejszy i własność dobrego uporządkowania, sformułowana w § 4.1, mówi, że każdy niepusty podzbiór zbioru $\mathbb{N}$ ma również element najmniejszy. To była ta własność, której użyliśmy do pokazania, że każdy malejący ciąg o wyrazach w $\mathbb{N}$ jest skończony, a więc, że algorytm dzielenia zakończy działanie.

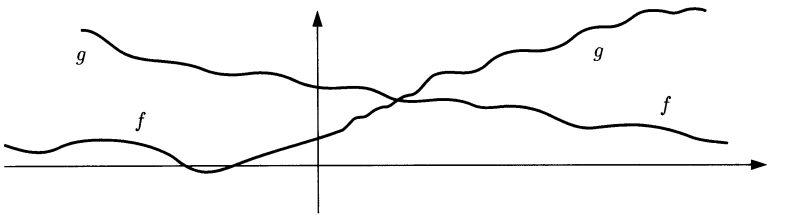
Mówimy, że zbiór liniowo uporządkowany C jest **dobrze uporządkowany**, gdy każdy niepusty podzbiór zbioru C ma element najmniejszy. Jeśli zbiór C jest dobrze uporządkowany i dla każdego elementu c z C mamy zdanie $p(c)$, to możemy mieć nadzieję, że uda nam się dowieść prawdziwości wszystkich zdań $p(c)$ poprzez przypuszczenie, że podzbiór $\{c \in C\colon p(c)$ jest fałszywe$\}$ zbioru C jest niepusty, a następnie wzięcie najmniejszego elementu c, dla którego zdanie $p(c)$ jest fałszywe i doprowadzenie do sprzeczności. Ten pomysł leżał u podstaw naszego uzasadnienia zasad indukcji w § 4.2 i § 4.5.

W pozostałej części tego paragrafu będziemy badać metody tworzenia nowych częściowych porządków z porządków, które są dane. Przypuśćmy najpierw, że dany jest zbiór częściowo uporządkowany $(S, \preceq)$ i że T jest pewnym niepustym zbiorem. Możemy zdefiniować częściowy porządek, który oznaczymy przez $\preceq$, w zbiorze wszystkich funkcji z T do S mówiąc, że

$$f \preceq g, \quad \text{jeśli} \quad f(t) \preceq g(t) \text{ dla wszystkich } t \text{ w } T.$$

Tego właśnie częściowego porządku użyliśmy w przykładach 2(d) i 8(c) z § 11.1. Bezpośrednio sprawdza się (ćwiczenie 7(a)), że ta nowa relacja jest częściowym porządkiem w zbiorze $\text{FUN}(T, S)$. Jeśli porządek w zbiorze S jest oznaczony przez $\leq$, to również odpowiadający mu porządek w zbiorze funkcji oznaczać będziemy przez $\leq$ zamiast $\preceq$.

PRZYKŁAD 4 (a) Jeśli $S = T = \mathbb{R}$ ze zwykłym porządkiem, to $f \leq g$ znaczy, że wykres funkcji f leży na wykresie lub poniżej wykresu funkcji g, tak jak na rysunku 11.14.



Rysunek 11.14

(b) Rozważmy zbiór $\{0, 1\}$ z porządkiem $0 < 1$. Funkcje należące do zbioru $\text{FUN}(T, \{0, 1\})$ są funkcjami charakterystycznymi podzbiorów zbioru T. Każdy podzbiór A zbioru T ma odpowiadającą mu funkcję χ_A w zbiorze $\text{FUN}(T, \{0, 1\})$ taką, że $\chi_A(x) = 1$, jeśli $x \in A$ i $\chi_A(x) = 0$, jeśli $x \notin A$. Stąd $\chi_A \leq \chi_B$ wtedy i tylko wtedy, gdy $x \in B$, o ile tylko $x \in A$, tzn. wtedy i tylko wtedy, gdy $A \subseteq B$. Zatem diagramy Hassego dla zbiorów $(\text{FUN}(T, \{0, 1\}), \leq)$ i $(\mathcal{P}(T), \subseteq)$ wyglądają tak samo. Rysunki 11.2 i 11.9 z § 11.1 przedstawiają te diagramy dla $T = \{a, b, c\}$. Powyższe dwa zbiory częściowo uporządkowane są w rzeczywistości izomorficznymi algebrami Boole'a. ∎

Przykład 4(b) pokazuje, że $(\text{FUN}(T, S), \preceq)$ nie musi być zbiorem liniowo uporządkowanym nawet, gdy zbiór S jest liniowo uporządkowany. Zbiór częściowo uporządkowany $(\text{FUN}(T, S), \preceq)$ dziedziczy jednakże pewne własności zbioru S. Jeśli zbiór S ma element największy bądź najmniejszy, to to samo jest prawdą dla $\text{FUN}(T, S)$, i jeśli S jest kratą, to jest nią również $\text{FUN}(T, S)$. Więcej na ten temat znajdziesz w ćwiczeniu 7.

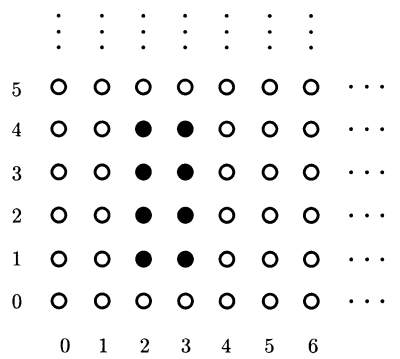
Innym sposobem utworzenia nowego zbioru z dwóch danych zbiorów jest wzięcie ich **iloczynu kartezjańskiego**. Przypu-

ściemy, że $(S, \preceq_1)$ i $(T, \preceq_2)$ są zbiorami częściowo uporządkowanymi, przy czym użyliśmy indeksów, by móc odróżnić jeden częściowy porządek od drugiego. Istnieje więcej niż jeden naturalny sposób nadania zbiorowi $S \times T$ struktury zbioru częściowo uporządkowanego. Od danego problemu zależeć będzie, który porządek wybierzemy.

Pierwszy częściowy porządek, który opiszemy, nazywany jest **porządkiem produktowym**. Dla $s, s' \in S$ oraz $t, t' \in T$ definiujemy:

$$(s,t) \preceq (s',t'), \quad \text{jeśli} \quad s \preceq_1 s' \text{ oraz } t \preceq_2 t'.$$

PRZYKŁAD 5 Niech $S = T = \mathbb{N}$, w obu przypadkach wraz ze zwykłym porządkiem $\leq$. Wtedy $(2,5) \preceq (3,7)$, gdyż $2 \leq 3$ i $5 \leq 7$. Również $(2,5) \preceq (3,5)$, ponieważ $2 \leq 3$ i $5 \leq 5$. Ale pary $(2,7)$ i $(3,5)$ są nieporównywalne; $(2,7) \preceq (3,5)$ znaczyłoby, że $2 \leq 3$ oraz $7 \leq 5$, podczas gdy $(3,5) \preceq (2,7)$ znaczyłoby, że $3 \leq 2$ oraz $5 \leq 7$. Rysunek 11.15 wskazuje te pary (m,n) w zbiorze $S \times T = \mathbb{N} \times \mathbb{N}$, dla których $(2,1) \preceq (m,n) \preceq (3,4)$. ∎



Rysunek 11.15

Rozważmy ponownie dwa zbiory częściowo uporządkowane $(S, \preceq_1)$ i $(T, \preceq_2)$. Fakt, że porządek produktowy $\preceq$ jest częściowym porządkiem wynika prawie natychmiast z definicji. Na przykład, jeśli $(s,t) \preceq (s',t')$ i $(s',t') \preceq (s,t)$, to $s \preceq_1 s'$, $t \preceq_2 t'$, $s' \preceq_1 s$ i $t' \preceq_2 t$. Ponieważ relacje $\preceq_1$ i $\preceq_2$ są antysymetryczne, to $s = s'$ i $t = t'$. Zatem relacja $\preceq$ jest antysymetryczna.

Rozszerzenie tego pomysłu na iloczyn kartezjański skończenie wielu zbiorów częściowo uporządkowanych $S_1 \times S_2 \times \ldots \times S_n$ i zdefiniowanie w nim odpowiedniego częściowego porządku nie

przedstawia żadnych trudności. Definiujemy:

$(s_1, s_2, \ldots, s_n) \preceq (s'_1, s'_2, \ldots, s'_n)$, jeśli $s_i \preceq s'_i$ dla wszystkich i.

W przykładzie 5 zbiór $\mathbb{N}$ jest liniowo uporządkowany, ale zbiór $\mathbb{N} \times \mathbb{N}$ nie; na przykład pary $(2,4)$ i $(3,1)$ nie są porównywalne w sensie porządku produktowego. W gruncie rzeczy porządek produktowy prawie nigdy nie jest liniowym porządkiem (ćwiczenie 12). Z drugiej strony, jeśli $S_1, S_2, \ldots, S_n$ są zbiorami liniowo uporządkowanymi, to istnieje inny naturalny porządek w zbiorze $S_1 \times S_2 \times \ldots \times S_n$, który jest liniowy. Zajmiemy się nim teraz.

PRZYKŁAD 6 Rozważmy zbiór $S = \{0, 1, 2, \ldots, 9\}$ wraz ze zwykłym porządkiem. Zbiór $S \times S$ składa się z par (m, n), które możemy utożsamić z liczbami całkowitymi 00, 01, 02, ..., 98, 99 od 0 do 99. Aby wprowadzić w zbiorze $S \times S$ liniowy porządek, możemy po prostu powiedzieć, że $(m, n) \prec (m', n')$, jeśli liczby całkowite odpowiadające danym parom spełniają odpowiednią nierówność, tzn. jeśli $m < m'$ lub jeśli $m = m'$ i $n < n'$. Na przykład, mamy $(5, 7) \prec (6, 3)$, gdyż $57 < 63$ oraz $(3, 5) \prec (3, 7)$, ponieważ $35 < 37$.

W podobny sposób możemy utożsamić zbiór $S \times S \times S$ ze zbiorem liczb całkowitych od 0 do 999 i wprowadzić w $S \times S \times S$ liniowy porządek mówiąc, że $(m, n, p) \prec (m', n', p')$, jeśli $m < m'$ lub $m = m'$ i $n < n'$ lub jeśli $m = m'$, $n = n'$ i $p < p'$. ∎

Pomysł, leżący u podstaw powyższego przykładu, daje się zastosować w przypadku ogólnym. Jeśli $(S_1, \preceq_1), \ldots, (S_n, \preceq_n)$ są zbiorami częściowo uporządkowanymi, to możemy zdefiniować relację $\prec$ w zbiorze $S_1 \times \ldots \times S_n$ w następujący sposób:

$(s_1, s_2, \ldots, s_n) \prec (t_1, t_2, \ldots, t_n)$, jeśli $s_1 \prec t_1$ lub jeśli istnieje r w $\{2, \ldots, n\}$ takie, że $s_1 = t_1, \ldots, s_{r-1} = t_{r-1}$ i $s_r \prec_r t_r$.

Wówczas relacja $\prec$ jest quasi-porządkiem (ćwiczenie 19), indukującym w zbiorze $S_1 \times S_2 \times \ldots \times S_n$ częściowy porządek $\preceq$, który będziemy nazywać **porządkiem leksykograficznym**.

PRZYKŁAD 7 Rozważmy zbiory $S = \{a, b, c, \ldots, z\}$ i $T = \{0, 1, \ldots, 9\}$ wraz ze zwykłymi porządkami. Elementy zbioru $S \times T$ możemy utożsamić z dwuelementowymi ciągami takimi, jak $a5$ i $x3$. Wyobraź sobie urządzenie mające oznaczone literami podzespoły, z których każdy składa się z co najwyżej dziesięciu części. Byłoby rozsądnie oznaczyć części zapasowe do tego urządzenia kombinacjami typu „litera–liczba" i umieścić je w skrzyniach ustawionych zgodnie z porządkiem leksykograficznym w zbiorze $S \times T$. Wówczas

skrzynia $a5$ stałaby bliżej niż skrzynia $x3$, gdyż a poprzedza x, natomiast dalej niż skrzynia $a3$, ponieważ $3 < 5$. ∎

Porządek leksykograficzny jest użyteczny głównie wtedy, gdy każdy ze zbiorów S_i jest uporządkowany liniowo.

Twierdzenie 1

Niech $(S_1, \preceq_1), \ldots, (S_n, \preceq_n)$ będą zbiorami liniowo uporządkowanymi. Wówczas porządek leksykograficzny liniowo porządkuje iloczyn kartezjański $S_1 \times \ldots \times S_n$.

Dowód. Wiemy już, że relacja $\preceq$ jest częściowym porządkiem w zbiorze $S_1 \times \ldots \times S_n$. Niech $s = (s_1, \ldots, s_n)$ i $t = (t_1, \ldots, t_n)$ będą różnymi elementami zbioru $S_1 \times \ldots \times S_n$. Ponieważ $s \neq t$, więc istnieje najmniejsza liczba r, dla której $s_r \neq t_r$. Ponieważ $(S_r, \preceq_r)$ jest zbiorem liniowo uporządkowanym, to albo $s_r \prec t_r$, albo $t_r \prec s_r$. W pierwszym przypadku $s \prec t$, w drugim zaś $t \prec s$. W każdym z przypadków dwa rozważane elementy zbioru $S_1 \times \ldots \times S_n$ są porównywalne. ∎

Przypadek szczególny, gdy wszystkie zbiory $(S_i, \preceq_i)$ są identyczne, jest wystarczająco ważny, by uzasadnić wprowadzenie specjalnego oznaczenia. Jeśli $(S, \preceq)$ jest zbiorem częściowo uporządkowanym i $k \in \mathbb{P}$, to przez $\preceq^k$ będziemy oznaczać porządek leksykograficzny w zbiorze $S^k = S \times \ldots \times S$.

W pozostałej części tego paragrafu będzie nas głównie interesował przypadek, w którym S jest alfabetem. W związku z tym, od tego momentu będziemy pisać Σ zamiast S. Będziemy również stosować naturalne utożsamienie ciągów długości k $(a_1, \ldots, a_k)$ należących do iloczynu kartezjańskiego Σ^k ze słowami $a_1 \ldots a_k$ długości k w Σ^*. Zakładamy nadal, że w zbiorze Σ dany jest pewien częściowy porządek $\preceq$. Zbiór Σ^0 jest równy $\{\lambda\}$ i porządek $\preceq^0$ w Σ^0 definiujemy w jedyny możliwy sposób: $\lambda \preceq^0 \lambda$.

PRZYKŁAD 8

Rozważmy $\Sigma = \{a, b, c, \ldots, z\}$ ze zwykłym liniowym porządkiem liter w alfabecie. Wówczas, zbiór Σ^k składa się z literowych ciągów długości k i $\preceq^k$ jest zwykłym porządkiem alfabetycznym w Σ^k. Jeśli, na przykład, $k = 3$, to

$$kos \preceq^3 kot \preceq^3 las \preceq^3 nos \preceq^3 rak \preceq^3 rok \preceq^3 sok.$$ ∎

Ponieważ $\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \ldots$, to możemy zebrać razem częściowe porządki $\preceq^0, \preceq^1, \preceq^2, \ldots$ i utworzyć porządek $\preceq^*$ w Σ^*, który będziemy nazywać **porządkiem standardowym**.

a w którym na początku jest λ, potem wszystkie słowa długości 1, następnie słowa długości 2 i tak dalej. Dokładniej,

$$w_1 \preceq^* w_2, \text{ jeśli bądź } w_1 \in \Sigma^k \text{ i } w_2 \in \Sigma^r \text{ i } k < r,$$

$$\text{bądź } w_1 \in \Sigma^k \text{ i } w_2 \in \Sigma^k \text{ dla tego samego } k \text{ i } w_1 \preceq^k w_2.$$

Jeśli, tak jak to było w przykładzie 8, $(\Sigma, \preceq)$ jest zbiorem liniowo uporządkowanym, to $(\Sigma^*, \preceq^*)$ też jest zbiorem liniowo uporządkowanym.

PRZYKŁAD 9 (a) Niech Σ będzie alfabetem angielskim wraz ze zwykłym porządkiem. Wówczas kilka pierwszych słów z Σ^*, w sensie porządku standardowego, to:

$$\lambda, a, b, \ldots, z, aa, ab, \ldots, az, ba, bb, \ldots, bz, ca, cb, \ldots, cz,$$
$$da, db, \ldots, dz, \ldots, za, zb, \ldots, zz, aaa, aab, aac, \ldots$$

(b) Niech $\Sigma = \{0, 1\}$, gdzie $0 < 1$. Wówczas kilka pierwszych elementów z Σ^*, w sensie porządku standardowego, to:

$$\lambda, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011,$$
$$100, 101, 110, 111, 0000, 0001, 0010, \ldots \quad \blacksquare$$

Zauważmy, że gdyby słowa w słowniku ułożone zostały w kolejności zgodnej z porządkiem standardowym z przykładu 9(a), to wszystkie krótkie słowa występowałyby na początku tego słownika i dla znalezienia jakiegoś słowa istotna byłaby znajomość jego długości. (W istocie niektóre słowniki przeznaczone dla wielbicieli rozwiązywania krzyżówek są tak właśnie napisane.) Aby znaleźć jakieś słowo w słowniku napisanym zgodnie ze zwykłym porządkiem alfabetycznym, badamy kolejne słowa, sprawdzając ich litery od lewej strony do prawej w poszukiwaniu różnic i nie zważając na ich długość. W ten sposób „abdykacja" występuje wcześniej niż „alt", a „komis" poprzedza „komisariat". Porządek alfabetyczny oparty jest na zwykłym alfabecie, uporządkowanym zwykłą metodą. Pomysł, leżący u jego podstaw, uogólnia się w naturalny sposób na dowolny alfabet Σ z częściowym porządkiem $\preceq$.

Porządek leksykograficzny lub **słownikowy** $\preceq_L$ w zbiorze Σ^* definiuje się w następujący sposób. Dla liter $a_1, \ldots, a_m$ i $b_1, \ldots, b_n$ z alfabetu Σ niech $k = \min\{m, n\}$. Wówczas

$$a_1 \ldots a_m \prec_L b_1 \ldots b_n, \quad \text{jeśli} \quad a_1 \ldots a_k \prec^k b_1 \ldots b_k$$
$$\text{lub jeśli} \quad k = m < n \quad \text{i} \quad a_1 \ldots a_k = b_1 \ldots b_k.$$

Wówczas relacja $\preceq_L$ jest quasi-porządkiem, który w zbiorze Σ^* definiuje częściowy porządek $\preceq_L$. W ten sposób *abdykacja* $\prec_L$

alt, gdyż $abd \prec^3 alt$, podczas gdy *komis* $\prec_L$ *komisariat*, gdyż *komis* jest postaci $a_1 \dots a_5$, *komisariat* jest postaci $b_1 \dots b_{10}$ i $a_1 \dots a_5 = b_1 \dots b_5$.

Porządek $\preceq_L$ możemy opisać w inny jeszcze sposób. Dla słów w i z w Σ^*, $w \preceq_L z$ wtedy i tylko wtedy, gdy bądź

(a) w jest początkowym fragmentem z, tzn. $z = wu$ dla pewnego słowa $u \in \Sigma^*$, bądź

(b) $w = xu$ a $z = xv$ dla pewnych słów u i v z Σ^* takich, że pierwsza litera słowa u poprzedza pierwszą literę słowa v w sensie danego porządku w Σ.

Zauważmy, że w punkcie (b) x może być słowem dowolnym, w szczególności także pustym.

Jeśli rozpatrujemy słowa pewnej ustalonej długości k, to porządek leksykograficzny jest identyczny z wcześniej wprowadzonym porządkiem leksykograficznym $\preceq^k$ w Σ^k. Ponadto porządek ten dla elementów z Σ^k zgadza się z porządkiem standardowym, ale różni się od niego, jeśli rozpatrujemy słowa różnej długości.

PRZYKŁAD 10 Niech $\Sigma = \{a, b\}$, gdzie $a \prec b$. Kilka pierwszych elementów zbioru Σ^*, w sensie porządku leksykograficznego, to:

$$\lambda, a, aa, aaa, aaaa, aaaaa, \dots$$

Każde słowo, w którym występuje litera b, jest poprzedzone przez nieskończenie wiele słów, wśród których są wszystkie słowa używające jedynie litery a. Co więcej, zbiór Σ^* zawiera nieskończone malejące ciągi słów; na przykład,

$$b \succ_L ab \succ_L aab \succ_L aaab \succ_L \dots.$$

Ponadto, między dwoma elementami powyższego ciągu znaleźć można nieskończenie wiele słów; na przykład, wszystkie słowa

$$aaab, aaaba, aaabaa, aaabaaa, aaabaaaa, \dots$$

poprzedzają *aab*. Zatem porządek leksykograficzny w nieskończonym zbiorze Σ^* jest bardzo skomplikowany i trudno jest go sobie wyobrażać. Jest on jednakże liniowym porządkiem, jak pokażemy w następnym twierdzeniu. ∎

Twierdzenie 2

Jeśli $(\Sigma, \preceq)$ jest zbiorem liniowo uporządkowanym, to $(\Sigma^*, \preceq^*)$ jest zbiorem dobrze uporządkowanym, a $(\Sigma^*, \preceq_L)$ jest zbiorem liniowo uporządkowanym.

Dowód. Wiemy z twierdzenia 1, że każdy ze zbiorów $(\Sigma^k, \preceq^k)$ jest liniowo uporządkowany, czyli jest łańcuchem. Porządek standardowy $\preceq^*$ łączy po prostu te łańcuchy kolejno dla $k = 0, 1, 2, \ldots$ ze sobą tak, że koniec poprzedniego bezpośrednio poprzedza początek następnego, a więc $(\Sigma^*, \preceq^*)$ jest łańcuchem. Żeby sprawdzić, że $\preceq^*$ dobrze porządkuje Σ^*, weźmy dowolny niepusty podzbiór A zbioru Σ^*. Niech k będzie długością najkrótszego słowa ze zbioru A. Ponieważ zbiór $A \cap \Sigma^k$ jest niepusty i skończony, to ma on element najmniejszy w_0, w sensie porządku $\preceq^k$. Wynika stąd, że $w_0 \preceq^* w$ dla wszystkich elementów w ze zbioru A, a więc w_0 jest najmniejszym elementem w A.

Rozważmy teraz porządek leksykograficzny $\preceq_L$. Weźmy ze zbioru Σ^* dwa elementy $a_1 \ldots a_m$ i $b_1 \ldots b_n$, gdzie $m \leq n$. Jeśli $a_1 \ldots a_m = b_1 \ldots b_m$, to na mocy definicji, $a_1 \ldots a_m \preceq_L b_1 \ldots b_n$. W przeciwnym razie, ponieważ $(\Sigma^m, \preceq^m)$ jest łańcuchem, to jeden z elementów $a_1 \ldots a_m$ i $b_1 \ldots b_m$ poprzedza drugi w zbiorze $(\Sigma^m, \preceq^m)$, a więc elementy $a_1 \ldots a_m$ i $b_1 \ldots b_n$ są porównywalne w zbiorze $(\Sigma^*, \preceq_L)$. Zatem każde dwa elementy zbioru Σ^* są porównywalne w sensie porządku $\preceq_L$, a to znaczy, że jest on liniowy. ∎

Jeśli zbiór Σ ma więcej niż jeden element, to $(\Sigma^*, \preceq_L)$ nie jest zbiorem dobrze uporządkowanym. Na przykład zbiór $\{b, ab, aab, aaab, aaaab, \ldots\}$ z przykładu 10 nie ma elementu najmniejszego. Oczywiście dowolny skończony podzbiór zbioru Σ^* ma element najmniejszy, gdyż sam jest skończonym łańcuchem.

ĆWICZENIA DO § 11.2

1. Niech P będzie zbiorem wszystkich podzbiorów zbioru $\{1, 2, 3, 4, 5\}$.
(a) Podaj dwa przykłady łańcuchów maksymalnych w $(P, \subseteq)$.
(b) Ile jest łańcuchów maksymalnych w $(P, \subseteq)$?

2. Weźmy zbiór $A = \{1, 2, 3, 4\}$ ze zwykłym porządkiem i zbiór $S = A \times A$ z porządkiem produktowym.
(a) Znajdź łańcuch w S mający 7 elementów.
(b) Czy łańcuch w S może mieć 8 elementów? Odpowiedź uzasadnij.

3. Niech $(S, |)$ będzie zbiorem $\{2, 3, 4, \ldots, 999, 1000\}$ z częściowym porządkiem „jest dzielnikiem".
(a) W $(S, |)$ jest dokładnie 500 elementów maksymalnych. Wskaż je.
(b) Podaj dwa przykłady łańcuchów maksymalnych w $(S, |)$.
(c) Czy do każdego łańcucha maksymalnego należy jakiś element minimalny w zbiorze S? Odpowiedź uzasadnij.

4. (a) Załóżmy, że żaden łańcuch w zbiorze częściowo uporządkowanym $(S, \preceq)$ nie ma więcej niż 73 elementy. Czy dowolny łańcuch w zbiorze S, który ma 73 elementy, musi być maksymalny? Odpowiedź uzasadnij.

(b) Podaj przykład zbioru częściowo uporządkowanego, który ma dwa łańcuchy maksymalne o czterech elementach i cztery łańcuchy maksymalne o dwóch elementach.

5. Czy każdy zbiór liniowo uporządkowany jest kratą? Odpowiedź uzasadnij.

6. Niech $(C_1, \preceq_1), (C_2, \preceq_2), \ldots, (C_n, \preceq_n)$ będzie rodziną parami rozłącznych zbiorów liniowo uporządkowanych. Opisz, w jaki sposób określić można liniowy porządek w zbiorze $C_1 \cup C_2 \cup \ldots \cup C_n$.

7. Niech $(S, \preceq)$ będzie zbiorem częściowo uporządkowanym i niech T będzie dowolnym zbiorem. Zdefiniuj w zbiorze $\text{FUN}\,(T, S)$ relację $\preceq$ w następujący sposób:

$$f \preceq g, \quad \text{jeśli} \quad f(t) \preceq g(t) \text{ dla wszystkich } t \text{ z } T.$$

(a) Wykaż, że relacja $\preceq$ jest częściowym porządkiem.

(b) Wykaż, że jeśli m jest elementem maksymalnym w zbiorze S, to funkcja f_m zdefiniowana wzorem $f_m(t) = m$ dla wszystkich $t \in T$ jest elementem maksymalnym w zbiorze $\text{FUN}\,(T, S)$.

(c) Wykaż, że jeśli S jest kratą i $f, g \in \text{FUN}\,(T, S)$, to funkcja h zdefiniowana wzorem $h(t) = f(t) \vee g(t)$ dla $t \in T$ jest kresem górnym zbioru $\{f, g\}$.

(d) Opisz quasi-porządek $\prec$ związany z porządkiem $\preceq$.

8. Rozważmy zbiór $\mathbb{N} \times \mathbb{N}$ wraz z porządkiem produktowym. Zrób rysunek, taki jak rysunek 11.15, który będzie przedstawiał zbiór $\{(m, n)\colon (m, n) \preceq (5, 2)\}$.

9. Dane są zbiory $S = \{0, 1, 2\}$ wraz ze zwykłym porządkiem i $T = \{a, b\}$, gdzie $a < b$.

(a) Narysuj diagram Hassego zbioru częściowo uporządkowanego $(\text{FUN}\,(T, S), \preceq)$ z porządkiem $\preceq$ opisanym w ćwiczeniu 7. *Wskazówka*: zobacz przykład 8(c) z § 11.1.

(b) Narysuj diagram Hassego zbioru częściowo uporządkowanego $(S \times S, \preceq)$ z porządkiem produktowym.

(c) Narysuj diagram Hassego zbioru częściowo uporządkowanego $S \times T$ z porządkiem produktowym.

10. Przypuśćmy, że $(S, \preceq_1)$ i $(T, \preceq_2)$ są zbiorami częściowo uporządkowanymi. Definiujemy relację $\preceq$ w zbiorze $S \times T$ w następujący sposób:

$$(s, t) \preceq (s', t'), \quad \text{jeśli} \quad s \preceq_1 s' \text{ lub } t \preceq_2 t'.$$

Czy relacja $\preceq$ jest częściowym porządkiem? Odpowiedź uzasadnij.

11. Dane są zbiory $S = \{0, 1, 2\}$ i $T = \{3, 4\}$ wraz ze zwykłymi porządkami. Wypisz elementy następujących zbiorów w rosnącym leksykograficznym porządku:
(a) $S \times S$, (b) $S \times T$, (c) $T \times S$.

12. Przypuśćmy, że $(S, \preceq_1)$ i $(T, \preceq_2)$ są zbiorami częściowo uporządkowanymi, przy czym każdy z nich ma więcej niż jeden element. Wykaż, że zbiór $S \times T$ nie jest liniowo uporządkowany przez porządek produktowy.

13. Dany jest zbiór $\mathbb{B} = \{0, 1\}$ ze zwykłym porządkiem. Wypisz elementy 101, 010, 11, 000, 10, 0010, 1000 zbioru $\mathbb{B}^*$ w porządku rosnącym
(a) w sensie porządku leksykograficznego,
(b) w sensie porządku standardowego.

14. Niech $(\Sigma, \preceq)$ będzie niepustym zbiorem liniowo uporządkowanym.
(a) Czy zbiór $(\Sigma^*, \preceq^*)$ ma element maksymalny? Odpowiedź uzasadnij.
(b) Czy zbiór $(\Sigma^*, \preceq_L)$ ma element maksymalny? Odpowiedź uzasadnij.

15. Niech Σ będzie polskim alfabetem uporządkowanym w zwykły sposób.
(a) Wypisz słowa, z których składa się to zdanie w rosnącym porządku standardowym.
(b) Wypisz słowa, z których składa się to zdanie w rosnącym porządku leksykograficznym.

16. Jakie warunki musi spełniać zbiór Σ, by porządki standardowy i leksykograficzny w zbiorze Σ^* były identyczne?

17. Udowodnij, że w skończonym zbiorze częściowo uporządkowanym $(S, \preceq)$, w każdym łańcuchu maksymalnym istnieje element minimalny zbioru S.

18. Niech $(S, \preceq_1)$ i $(T, \preceq_2)$ będą zbiorami częściowo uporządkowanymi. Rozważmy zbiór $S \times T$ wraz z porządkiem leksykograficznym.
(a) Wykaż, że jeśli m_1 jest elementem maksymalnym w zbiorze S, a m_2 jest elementem maksymalnym w zbiorze T, to para (m_1, m_2) jest elementem maksymalnym w zbiorze $S \times T$.
(b) Czy w zbiorze $S \times T$ istnieją elementy maksymalne postaci innej, niż opisana w części (a)? Odpowiedź uzasadnij.
(c) Przypuśćmy, że zbiór $S \times T$ ma element największy. Czy wówczas któryś ze zbiorów S lub T musi mieć element największy? Odpowiedź uzasadnij.

19. Niech $(S_1, \preceq_1), \ldots, (S_n, \preceq_n)$ będą zbiorami częściowo uporządkowanymi. W zbiorze $S_1 \times \ldots \times S_n$ definiujemy relację $\prec$ w sposób następujący:

$(s_1, \ldots, s_n) \prec (t_1, \ldots, t_n)$, jeśli $s_1 \prec_1 t_1$ lub jeśli istnieje r w $\{2, \ldots, n\}$ takie, że $s_1 = t_1, \ldots, s_{r-1} = t_{r-1}$ i $s_r \prec_r t_r$.

Wykaż, że relacja $\prec$ jest quasi-porządkiem.

§ 11.3. Ogólne własności relacji

W rozdziale 3 badaliśmy relacje równoważności, a w dwóch poprzednich paragrafach zajmowaliśmy sie częściowymi porządkami. Relacje obydwu typów są zwrotne i przechodnie, ale różnica między symetrią a antysymetrią sprawia, że są one zupełnie różne. Częściowe porządki dają wrażenie kierunku w zbiorze, od elementów małych do dużych. Natomiast relacje równoważności dzielą zbiór na nie związane ze sobą bloki, których elementy są zebrane razem, ponieważ mają jakieś wspólne cechy.

Relacje obydwu typów wprowadzają w danym zbiorze S pewną zorganizowaną strukturę. Natomiast teoria dowolnych relacji dwuargumentowych w S, tj. podzbiorów zbioru $S \times S$, jest, przeciwnie, tak abstrakcyjna i mało konkretna, że nie ma tu właściwie struktury, badaniu której teoria ta byłaby poświęcona. W tym paragrafie skoncentrujemy się na stwierdzeniach, które są prawdziwe dla wszystkich relacji. Nic dziwnego więc, że ograniczymy się do zagadnień bardzo ogólnych. Zaczniemy od zbadania, w jaki sposób dwie relacje można złożyć tak, by otrzymać trzecią. Wyrazimy następnie uzyskane rezultaty za pomocą macierzy i zajmiemy się ponownie związkami między relacjami, macierzami i grafami skierowanymi, które omawialiśmy po raz pierwszy w rozdziale 3.

Jak widzieliśmy w § 3.1, funkcje można uważać za relacje. Utożsamiamy funkcję $f\colon S \to T$ z relacją R_f na zbiorze $S \times T$ zdefiniowaną w następujący sposób:

$$R_f = \{(s,t) \in S \times T\colon f(s) = t\}.$$

Jeśli dana jest też funkcja $g\colon T \to U$, to złożenie $g \circ f\colon S \to U$ daje relację

$$R_{g\circ f} = \{(s,u) \in S \times U\colon (g \circ f)(s) = g(f(s)) = u\},$$

którą możemy uważać za złożenie relacji R_f i R_g. Para (s,u) należy do zbioru $R_{g\circ f}$ wtedy i tylko wtedy, gdy $u = g(t)$, gdzie $t = f(s) \in T$, a więc

$$R_{g\circ f} = \{(s,u) \in S \times U\colon (s,t) \in R_f \text{ i } (t,u) \in R_g \text{ dla pewnego } t \in T\}.$$

Uogólnimy powyższy fakt, dotyczący relacji związanych z funkcjami, i zdefiniujemy złożenie dwóch dowolnych relacji. Jeśli R_1 jest relacją na zbiorze $S \times T$, a R_2 jest relacją na zbiorze $T \times U$, to **złożeniem relacji** R_1 i R_2 nazywamy relację $R_2 \circ R_1$ zdefiniowaną w następujący sposób:

$$R_2 \circ R_1 = \{(s,u) \in S \times U: \text{ dla pewnego } t \in T, (s,t) \in R_1 \text{ i } (t,u) \in R_2\}.$$

Dla relacji określonych przez funkcje, z definicji tej wynika, że

$$R_g \circ R_f = R_{g \circ f}.$$

Ponieważ myślimy o R_1 jako o pierwszej relacji, a o R_2 jako o drugiej, to z czysto estetycznego punktu widzenia ta ogólna definicja wydaje się być sformułowana na odwrót. Co więcej, w pewnym sensie taką się ona rzeczywiście okaże, gdy zaobserwujemy związek zachodzący między składaniem relacji a mnożeniem ich macierzy. Wiele osób używa oznaczenia $R_1 \circ R_2$ na to, co my określiliśmy jako $R_2 \circ R_1$. Taka notacja jest niespójna z notacją dotyczącą składania funkcji i bywa źródłem omyłek. Aby uniknąć nieporozumień, będziemy używać R_1R_2, a nie $R_1 \circ R_2$, jako alternatywnego oznaczenia dla $R_2 \circ R_1$. Podsumujmy:

Jeśli dane są relacje: R_1 na zbiorze $S \times T$ i R_2 na zbiorze $T \times U$, to **złożenie** tych relacji, czyli relacja

$$\{(s,u) \in S \times U: (s,t) \in R_1 \text{ i } (t,u) \in R_2 \text{ dla pewnego } t \in T\}$$

będzie oznaczana albo przez R_1R_2, albo przez $R_2 \circ R_1$. Zatem $(s,u) \in R_1R_2$ wtedy i tylko wtedy, gdy istnieje w zbiorze T element t taki, że $(s,t) \in R_1$ i $(t,u) \in R_2$.

PRZYKŁAD 1 Przykład 2 z § 3.1 dotyczy studentów, wykładów i wydziałów uniwersytetu. Rozważane są relacje

$$R_1 = \{(s,c) \in S \times C: s \text{ jest zapisany na wykład } c\}$$

oraz

$$R_2 = \{(c,d) \in C \times D: \text{wykład } c \text{ odbywa się na wydziale } d\}.$$

Zauważmy, że

$$R_1R_2 = \{(s,d) \in S \times D: (s,c) \in R_1 \text{ i } (c,d) \in R_2 \text{ dla pewnego } c \in C\}.$$

Zatem para (s,d) należy do R_1R_2, o ile student s jest zapisany na wykład prowadzony przez wydział d.

Zauważmy, że zapis R_2R_1 nie ma sensu, gdyż następniki par należących do relacji R_2 są elementami zbioru D, podczas gdy poprzedniki par należących do relacji R_1 są elementami zbioru S; nie może się zdarzyć, że $(c,t) \in R_2$ i $(t,c') \in R_1$. Zapis $R_2 \circ R_1$ ma oczywiście sens; jest po prostu innym oznaczeniem relacji R_1R_2. ■

PRZYKŁAD 2 Rozważmy relacje R_1 i R_2 na zbiorze $S \times T$ i relacje R_3 i R_4 na zbiorze $T \times U$.

(a) Jeśli $R_1 \subseteq R_2$ i $R_3 \subseteq R_4$, to $R_1R_3 \subseteq R_2R_4$. Aby to pokazać, weźmy $(s,u) \in R_1R_3$. Wówczas dla pewnego $t \in T$ mamy $(s,t) \in R_1$ i $(t,u) \in R_3$. Ponieważ $R_1 \subseteq R_2$ i $R_3 \subseteq R_4$, więc mamy także $(s,t) \in R_2$ i $(t,u) \in R_4$. Zatem $(s,u) \in R_2R_4$. To rozumowanie pokazuje, że $R_1R_3 \subseteq R_2R_4$.

(b) Suma dwóch relacji na zbiorze $A \times B$, tzn. dwóch podzbiorów zbioru $A \times B$, jest pewną relacją na zbiorze $A \times B$. Pokażemy, że dla relacji R_1, R_2, R_3 i R_4 spełniających powyższe założenia,

$$(R_1 \cup R_2)R_3 = R_1R_3 \cup R_2R_3.$$

Ponieważ $R_1 \subseteq R_1 \cup R_2$, to na mocy części (a) mamy $R_1R_3 \subseteq (R_1 \cup R_2)R_3$; podobnie, $R_2R_3 \subseteq (R_1 \cup R_2)R_3$, a zatem

$$R_1R_3 \cup R_2R_3 \subseteq (R_1 \cup R_2)R_3.$$

Dla sprawdzenia inkluzji odwrotnej, weźmy $(s,u) \in (R_1 \cup R_2)R_3$. Dla pewnego $t \in T$ mamy $(s,t) \in R_1 \cup R_2$ i $(t,u) \in R_3$. Stąd $(s,t) \in R_1$ i wówczas $(s,u) \in R_1R_3$ lub też $(s,t) \in R_2$ i wówczas $(s,u) \in R_2R_3$. W obu przypadkach $(s,u) \in R_1R_3 \cup R_2R_3$, a więc

$$(R_1 \cup R_2)R_3 \subseteq R_1R_3 \cup R_2R_3. \qquad \blacksquare$$

W paragrafie 1.3 odnotowaliśmy, że składanie funkcji jest łączne. To samo jest prawdą i dla relacji.

Prawo łączności dla relacji

Jeśli R_1 jest relacją na zbiorze $S \times T$, R_2 jest relacją na zbiorze $T \times U$, a R_3 jest relacją na zbiorze $U \times V$, to

$$(R_1R_2)R_3 = R_1(R_2R_3).$$

Dowód. Pokażemy, że para uporządkowana (s,v) ze zbioru $S \times V$ należy do zbioru $(R_1R_2)R_3$ wtedy i tylko wtedy, gdy

(*) dla pewnych $t \in T$ i $u \in U$ mamy
$$(s,t) \in R_1, (t,u) \in R_2 \text{ i } (u,v) \in R_3.$$

Podobne rozumowanie pokazuje, że (s,v) należy do $R_1(R_2R_3)$ wtedy i tylko wtedy, gdy spełniony jest warunek (*).

Weźmy parę (s,v) ze zbioru $(R_1R_2)R_3$. Ponieważ R_1R_2 jest relacją na zbiorze $S \times U$, znaczy to, że istnieje element $u \in U$ taki, że $(s,u) \in R_1R_2$ i $(u,v) \in R_3$. Skoro $(s,u) \in R_1R_2$, to istnieje element $t \in T$ taki, że $(s,t) \in R_1$ i $(t,u) \in R_2$. Zatem spełniony jest warunek (*).

Załóżmy teraz, że dla elementu $(s,v) \in S \times V$ spełniony jest warunek (*). Wówczas $(s,t) \in R_1$ i $(t,u) \in R_2$, skąd

$(s,u) \in R_1R_2$. Ponieważ jednocześnie $(u,v) \in R_3$, to otrzymujemy ostatecznie $(s,v) \in (R_1R_2)R_3$. ∎

Wobec prawa łączności możemy pisać $R_1R_2R_3$ dla oznaczenia zarówno $(R_1R_2)R_3$, jak i $R_1(R_2R_3)$. Jak wynika z powyższego dowodu, para (s,v) należy do $R_1R_2R_3$, o ile istnieją elementy $t \in T$ i $u \in U$ takie, że $(s,t) \in R_1$, $(t,u) \in R_2$ i $(u,v) \in R_3$.

Zbiory i relacje, które rozpatrywaliśmy do tej pory, mogły być zarówno skończone, jak i nieskończone. Teraz ograniczymy się do skończonych zbiorów S, T i U, i będziemy rozważać relacje: R_1 na zbiorze $S \times T$ oraz R_2 na zbiorze $T \times U$. Tak jak to robiliśmy w § 3.3, z relacjami tymi możemy powiązać macierze, których wyrazami są tylko 0 i 1. Wypisujemy elementy każdego ze zbiorów S, T i U w pewnej kolejności. Wtedy wyraz o współrzędnych (s,t) macierzy $\mathbf{A}_1$ odpowiadającej relacji R_1 jest równy 1, jeśli $(s,t) \in R_1$, a 0 w przeciwnym przypadku. Macierz $\mathbf{A}_2$ relacji R_2 jest zdefiniowana analogicznie. Mając dane macierze $\mathbf{A}_1$ i $\mathbf{A}_2$ chcemy znaleźć macierz odpowiadającą złożeniu R_1R_2.

PRZYKŁAD 3 Niech $S = \{1,2,3,4,5\}$, $T = \{a,b,c\}$ i $U = \{e,f,g,h\}$. Rozważmy relacje

$$R_1 = \{(1,a),(2,a),(2,c),(3,a),(3,b),(4,a),(4,b),(4,c),(5,b)\},$$
$$R_2 = \{(a,e),(a,g),(b,f),(b,g),(b,h),(c,e),(c,g),(c,h)\}.$$

Wówczas

$$R_1R_2 = \{(1,e),(1,g),(2,e),(2,g),(2,h),(3,e),(3,f),(3,g),$$
$$(3,h),(4,e),(4,f),(4,g),(4,h),(5,f),(5,g),(5,h)\}.$$

Macierze $\mathbf{A}_1$, $\mathbf{A}_2$ i $\mathbf{A}$, odpowiadające tym relacjom, pokazane są na rysunku 11.16. Porównaj na tym rysunku macierz $\mathbf{A}$ z iloczynem $\mathbf{A}_1\mathbf{A}_2$. W macierzy $\mathbf{A}$ jedynki występują tam, gdzie w macierzy $\mathbf{A}_1\mathbf{A}_2$ są wyrazy niezerowe. Związek ten jest nieprzypadkowy, co zaraz wyjaśnimy.

$$\begin{array}{c|ccc} & a & b & c \\ \hline 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 \\ 3 & 1 & 1 & 0 \\ 4 & 1 & 1 & 1 \\ 5 & 0 & 1 & 0 \end{array}$$
$\mathbf{A}_1$

$$\begin{array}{c|cccc} & e & f & g & h \\ \hline a & 1 & 0 & 1 & 0 \\ b & 0 & 1 & 1 & 1 \\ c & 1 & 0 & 1 & 1 \end{array}$$
$\mathbf{A}_2$

$$\begin{array}{c|cccc} & e & f & g & h \\ \hline 1 & 1 & 0 & 1 & 0 \\ 2 & 1 & 0 & 1 & 1 \\ 3 & 1 & 1 & 1 & 1 \\ 4 & 1 & 1 & 1 & 1 \\ 5 & 0 & 1 & 1 & 1 \end{array}$$
$\mathbf{A}$

$$\begin{array}{c|cccc} & e & f & g & h \\ \hline 1 & 1 & 0 & 1 & 0 \\ 2 & 2 & 0 & 2 & 1 \\ 3 & 1 & 1 & 2 & 1 \\ 4 & 2 & 1 & 3 & 2 \\ 5 & 0 & 1 & 1 & 1 \end{array}$$
$\mathbf{A}_1\mathbf{A}_2$

Rysunek 11.16

Weźmy $s \in \{1,2,3,4,5\}$ i $u \in \{e,f,g,h\}$. Wyraz macierzy $\mathbf{A}_1\mathbf{A}_2$ o współrzędnych (s,u) jest równy

$$\sum_{t\in\{a,b,c\}} \mathbf{A}_1[s,t]\mathbf{A}_2[t,u].$$

Suma ta jest dodatnia, o ile każdy jej składnik jest dodatni. Iloczyn $\mathbf{A}_1[s,t]\mathbf{A}_2[t,u]$ jest różny od 0 jedynie wówczas, gdy zarówno $\mathbf{A}_1[s,t]$, jak i $\mathbf{A}_2[t,u]$ równe są 1, a w takim przypadku $(s,t) \in R_1$ i $(t,u) \in R_2$. Zatem nasza suma wynosi 0, jeśli nie istnieje element t spełniający powyższe warunki, tzn. jeśli $(s,u) \notin R_1R_2$, jest natomiast większa od 0, jeśli $(s,u) \in R_1R_2$. Ściślej, suma ta jest dokładnie równa liczbie

$$|\{t \in \{a,b,c\}\colon (s,t) \in R_1 \text{ i } (t,u) \in R_2\}|.$$

Na przykład wyraz o współrzędnych $(2,e)$ jest równy 2, gdyż

$$\{t \in \{a,b,c\}\colon (2,t) \in R_1 \text{ i } (t,e) \in R_2\} = \{a,c\}.$$

Wyraz o współrzędnych $(2,f)$ jest równy 0, gdyż

$$\{t \in \{a,b,c\}\colon (2,t) \in R_1 \text{ i } (t,f) \in R_2\} = \emptyset,$$

tzn. $(2,f) \notin R_1R_2$. ∎

Jak pokazuje przykład 3, wyraz o współrzędnych (s,u) w macierzy będącej iloczynem macierzy relacji R_1 i R_2 równy jest liczbie takich elementów t, dla których $(s,t) \in R_1$ i $(t,u) \in R_2$. Aby znaleźć macierz relacji R_1R_2 wystarczy tylko wiedzieć, czy liczba ta jest różna od zera; jeśli tak, to wyraz o współrzędnych (s,u) w macierzy relacji R_1R_2 jest równy 1, natomiast w przeciwnym razie wynosi 0.

Aby otrzymać macierz relacji R_1R_2, moglibyśmy zdefiniować nowe działanie mnożenia macierzy mówiąc, że $\mathbf{A}_1 * \mathbf{A}_2$ jest macierzą powstałą z macierzy $\mathbf{A}_1\mathbf{A}_2$ przez zastąpienie każdego niezerowego wyrazu liczbą 1. Podejście równoważne, ale lepsze, polega na zastąpieniu działań arytmetycznych na liczbach całkowitych przez pewne wygodniejsze działania. Wyrazami macierzy odpowiadających relacjom są elementy zbioru $\mathbb{B} = \{0,1\}$, na którym określone są **działania booleowskie** $\vee$ i $\wedge$ w sposób następujący:

$\vee$	0	1
0	0	1
1	1	1

$\wedge$	0	1
0	0	0
1	0	1

Są to te same działania, które napotkaliśmy zajmując się tablicami wartości logicznych w rozdziale 2 i algebrami Boole'a w rozdziale 10. Zauważ, że $m \vee n = \max\{m,n\}$ i $m \wedge n = \min\{m,n\}$.

Macierze o wyrazach w zbiorze $\mathbb{B}$ nazywane są **macierzami booleowskimi**. Definiujemy **iloczyn booleowski** $\mathbf{A}_1 * \mathbf{A}_2$ macierzy booleowskiej $\mathbf{A}_1$ wymiaru $m \times n$ i macierzy booleowskiej $\mathbf{A}_2$ wymiaru $n \times p$ używając normalnej definicji iloczynu macierzy, z tym, że działania dodawania i mnożenia liczb zastąpione są działaniami, odpowiednio, $\vee$ i $\wedge$. Znaczy to, że wyraz o współrzędnych (i,k) macierzy $\mathbf{A}_1 * \mathbf{A}_2$ jest równy

$$\begin{aligned}(\mathbf{A}_1 * \mathbf{A}_2)[i,k] = \\ &= ((\mathbf{A}_1[i,1] \wedge \mathbf{A}_2[1,k]) \vee (\mathbf{A}_1[i,2] \wedge \mathbf{A}_2[2,k]) \\ &\vee \ldots \vee (\mathbf{A}_1[i,n] \wedge \mathbf{A}_2[n,k]),\end{aligned}$$

co można zapisać w bardziej zwartej formie jako

$$\bigvee_{j=1}^{n} (\mathbf{A}_1[i,j] \wedge \mathbf{A}_2[j,k]).$$

PRZYKŁAD 4 Iloczyn booleowski $\mathbf{A}_1 * \mathbf{A}_2$ macierzy przedstawionych na rysunku 11.16 jest macierzą $\mathbf{A}$ z tego rysunku. Na przykład wyraz macierzy $\mathbf{A}_1 * \mathbf{A}_2$, mający współrzędne $(3,g)$ jest równy $(1 \wedge 1) \vee (1 \wedge 1) \vee (0 \wedge 1) = 1 \vee 1 \vee 0 = 1$. Wyraz o współrzędnych $(5,e)$ jest równy $(0 \wedge 1) \vee (1 \wedge 0) \vee (0 \wedge 1) = 0 \vee 0 \vee 0 = 0$. ■

Na mocy definicji działań $\vee$ i $\wedge$, wyraz macierzy $\mathbf{A}_1 * \mathbf{A}_2$, o współrzędnych (i,k) jest równy 1 wtedy i tylko wtedy, gdy wartość co najmniej jednego z wyrażeń $\mathbf{A}_1[i,j] \wedge \mathbf{A}_2[j,k]$ wynosi 1, a to zachodzi wtedy i tylko wtedy, gdy zarówno $\mathbf{A}_1[i,j]$, jak i $\mathbf{A}_2[j,k]$ są równe 1. Fakt ten, wraz z przeprowadzoną po przykładzie 3 dyskusją, prowadzą do następującego rezultatu.

Twierdzenie 1

> Rozważmy relacje R_1 na zbiorze $S \times T$ i R_2 na zbiorze $T \times U$, gdzie S, T i U są zbiorami skończonymi. Jeśli $\mathbf{A}_1$ i $\mathbf{A}_2$ są macierzami booleowskimi tych relacji, to iloczyn booleowski $\mathbf{A}_1 * \mathbf{A}_2$ jest macierzą złożenia $R_1 R_2$.

Iloczyn booleowski $*$ jest działaniem łącznym. Fakt ten można wykazać bądź bezpośrednio, bądź, jak sugeruje ćwiczenie 17, wykorzystując twierdzenie 1 i łączność składania odpowiednich relacji.

W pozostałej części tego paragrafu rozważać będziemy relacje w jednym ustalonym zbiorze S. Ponieważ relacje w S są po prostu

podzbiorami zbioru $S \times S$, to rodziną wszystkich relacji w S jest zbiór $\mathcal{P}(S \times S)$.

Twierdzenie 2

Składanie relacji jest działaniem łącznym w zbiorze $\mathcal{P}(S \times S)$. Ponadto w zbiorze $\mathcal{P}(S \times S)$ istnieje element neutralny E, tzn. taki element E, że $RE = ER = R$ dla wszystkich $R \in \mathcal{P}(S \times S)$.

Dowód. Zauważmy najpierw, że jeśli relacje R_1 i R_2 należą do zbioru $\mathcal{P}(S \times S)$, to ich złożenie $R_1 R_2$ również. Łączność składania została już sprawdzona. Elementem neutralnym jest **relacja równości**

$$E = \{(x, x) \in S \times S \colon x \in S\}. \qquad \blacksquare$$

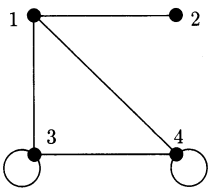
Będziemy stosować zwykłe oznaczenia dotyczące działań łącznych. A zatem, jeśli R jest relacją w zbiorze S, to $R^0 = E$ oraz, dla $n \in \mathbb{P}$, R^n oznacza n-krotne złożenie relacji R ze sobą. Zauważmy, że jeśli $n > 1$, to para (x, z) należy do R^n, o ile istnieją elementy y_1, $y_2, \ldots, y_{n-1}$ zbioru S takie, że wszystkie pary (x, y_1), $(y_1, y_2), \ldots$, (y_{n-1}, z) należą do R. Innymi słowy, $(x, z) \in R^n$, jeśli x i z są R-powiązane łańcuchem długości n.

PRZYKŁAD 5

(a) Rozważmy graf bez krawędzi wielokrotnych. Tak jak w § 3.2, zdefiniujmy relację sąsiedztwa R w zbiorze V wierzchołków grafu, mówiąc, że $(u, v) \in R$, o ile $\{u, v\}$ jest krawędzią grafu.

$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

A= macierz dla R



graf dla R

Rysunek 11.17

Rysunek 11.17 przedstawia przykład takiej sytuacji. Mamy wówczas $(u, v) \in R^n$, o ile istnieją wierzchołki $v_1, \ldots, v_{n-1}$ takie, że wszystkie pary (u, v_1), $(v_1, v_2), \ldots$, (v_{n-1}, w) należą do R, tzn. jeśli istnieje droga długości n, łącząca u z w. Macierz booleowska **A** relacji R jest macierzą sąsiedztwa grafu G, zdefiniowaną w § 3.3. Jej potęga booleowska

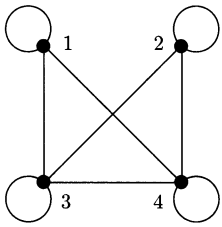
$$\mathbf{A} * \mathbf{A} * \ldots * \mathbf{A} \quad (n \text{ razy}),$$

będąca macierzą booleowską relacji R^n, pokazuje nam, które pary wierzchołków grafu połączone są drogami długości n. Gdybyśmy chcieli także znać liczbę takich dróg, to obliczylibyśmy zwykłą potęgę $\mathbf{A}^n$. Robiliśmy to już w § 3.4 i zrobimy ponownie teraz.

(b) Jeśli relacja R w zbiorze $\{1,2,3,4\}$ ma macierz i graf takie, jak na rysunku 11.17, to relacja R^2 ma macierz i graf takie jak na rysunku 11.18. Wyraz o współrzędnych $(3,4)$ w macierzy $\mathbf{A} * \mathbf{A}$ mówi nam, że w grafie z rysunku 11.17 istnieje co najmniej jedna droga długości 2 biegnąca od wierzchołka 3 do wierzchołka 4. Właśnie dlatego w grafie relacji R^2 istnieje krawędź łącząca wierzchołek 3 z wierzchołkiem 4. Wyraz o współrzędnych $(3,4)$ w macierzy $\mathbf{A}^2$ jest równy 3, więc w grafie z rysunku 11.17 istnieją dokładnie 3 drogi długości 2 biegnące od wierzchołka 3 do wierzchołka 4. Znajdź je! ∎

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$\mathbf{A} * \mathbf{A} =$ macierz dla R^2



graf dla R^2

$$\begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 2 & 1 & 3 & 3 \\ 2 & 1 & 3 & 3 \end{bmatrix}$$

$\mathbf{A}^2$

Rysunek 11.18

Jeśli ustaliliśmy porządek, zgodnie z którym będziemy wypisywać elementy danego skończonego zbioru S, to każda relacja R w S ma odpowiadającą jej macierz booleowską $\mathbf{A}$. Macierz tę można uważać za macierz sąsiedztwa grafu skierowanego, którego wierzchołkami są elementy zbioru S i w którym od wierzchołka x do wierzchołka y biegnie krawędź wtedy i tylko wtedy, gdy w macierzy $\mathbf{A}$ na miejscu o współrzędnych (x,y) stoi 1. Zatem w tym grafie skierowanym istnieje krawędź z x do y wtedy i tylko wtedy, gdy $(x,y) \in R$, tzn. relacja R jest relacją sąsiedztwa dla tego grafu skierowanego. Rysunek rozważanego grafu skierowanego nazywać będziemy **rysunkiem relacji** R. Zauważ, że choć macierz $\mathbf{A}$ zależy od porządku, zgodnie z którym wypisujemy elementy zbioru S, to nasz graf skierowany od niego nie zależy.

Jeśli relacja R jest symetryczna, to pary przeciwnie skierowanych krawędzi w związanym z nią grafie skierowanym można zastąpić pojedynczymi krawędziami w pewnym grafie nieskierowanym, dla którego relacja R jest relacją sąsiedztwa. W przykładzie 5 punktem wyjścia był graf nieskierowany; relacja, którą z niego otrzymaliśmy, była oczywiście symetryczna.

Następne twierdzenie pokazuje związek między przechodniością a składaniem relacji w zbiorze S.

Twierdzenie 3

Jeśli R jest relacją w zbiorze S, to R jest przechodnia wtedy i tylko wtedy, gdy $R^2 \subseteq R$.

Dowód. Załóżmy najpierw, że relacja R jest przechodnia i weźmy $(x, z) \in R^2$. Na mocy definicji relacji R^2, istnieje w zbiorze S element y taki, że $(x, y) \in R$ i $(y, z) \in R$. Ponieważ relacja R jest przechodnia, to para (x, z) również należy do R. Tym samym pokazaliśmy, że każdy element (x, z) z R^2 należy do R, tzn., że $R^2 \subseteq R$.

Dla dowodu implikacji odwrotnej załóżmy, że $R^2 \subseteq R$. Weźmy pary (x, y) i (y, z) należące do R. Wtedy para (x, z) należy do R^2, a więc także i do R. Dowodzi to, że relacja R jest przechodnia. ∎

Dla macierzy booleowskich $\mathbf{A}_1$ i $\mathbf{A}_2$ wymiaru $m \times n$ będziemy pisać $\mathbf{A}_1 \leq \mathbf{A}_2$, jeśli każdy wyraz macierzy $\mathbf{A}_1$ jest nie większy od odpowiadającego mu wyrazu macierzy $\mathbf{A}_2$, tzn. gdy

$$\mathbf{A}_1[i, j] \leq \mathbf{A}_2[i, j] \quad \text{dla} \quad 1 \leq i \leq m \text{ oraz } 1 \leq j \leq n.$$

Jeśli R_1 i R_2 są relacjami na zbiorze $S \times T$, o macierzach $\mathbf{A}_1$ i $\mathbf{A}_2$, to

$$R_1 \subseteq R_2 \quad \text{wtedy i tylko wtedy, gdy} \quad \mathbf{A}_1 \leq \mathbf{A}_2;$$

zastanów się, gdzie w macierzach $\mathbf{A}_1$ i $\mathbf{A}_2$ występują jedynki i zera (ćwiczenie 16(a)). Wynika stąd, że relacja $\leq$ jest częściowym porządkiem w zbiorze macierzy booleowskich. Ponadto, relacja R w zbiorze S spełnia warunek $R^2 \subseteq R$ wtedy i tylko wtedy, gdy jej macierz $\mathbf{A}$ spełnia warunek $\mathbf{A} * \mathbf{A} \leq \mathbf{A}$.

PRZYKŁAD 6

Rozważmy relację R w zbiorze $\{1, 2, 3\}$ o macierzy

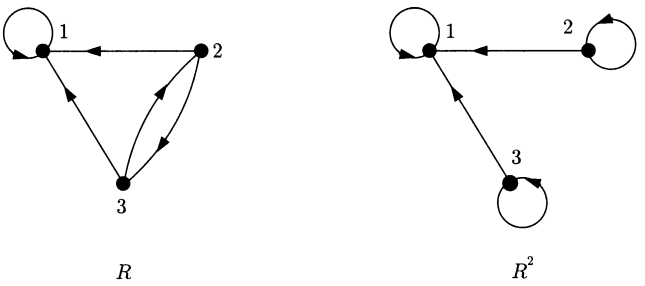
$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

Ponieważ

$$\mathbf{A} * \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} * \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \leq \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix} = \mathbf{A},$$

to relacja R jest przechodnia. Przechodniość relacji R można też wykazać posługując się odpowiadającym jej grafem skierowanym, przedstawionym na rysunku 11.19. Jeśli droga długości 2 łączy którekolwiek dwa wierzchołki tego grafu, to są one także połączone pojedynczą krawędzią. I tak na przykład 3 2 1 jest drogą i istnieje też krawędź od 3 do 1. ∎



R R^2

Rysunek 11.19

Ponieważ relacje na zbiorach skończonych odpowiadają macierzom booleowskim, własności relacji mogą być opisywane za pomocą macierzy. Aby pokazać niektóre ważniejsze macierzowe odpowiedniki tych własności, wprowadzimy trochę więcej oznaczeń. Dla macierzy booleowskich $\mathbf{A}_1$ i $\mathbf{A}_2$ definiujemy $\mathbf{A}_1 \vee \mathbf{A}_2$ wzorem

$$(\mathbf{A}_1 \vee \mathbf{A}_2)[i,j] = \mathbf{A}_1[i,j] \vee \mathbf{A}_2[i,j] \text{ dla } 1 \leq i \leq m \text{ oraz } 1 \leq j \leq n.$$

Macierz $\mathbf{A}_1 \wedge \mathbf{A}_2$ ma analogiczną definicję. I tak na przykład, jeśli

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \quad \text{oraz} \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix},$$

to

$$\mathbf{A}_1 \vee \mathbf{A}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \text{oraz} \quad \mathbf{A}_1 \wedge \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Poniższe zestawienie przedstawia macierzowe wersje niektórych znanych własności relacji. Przypomnij sobie, że $\mathbf{A}^T$ oznacza macierz transponowaną do macierzy $\mathbf{A}$. Poniższe równoważności są łatwe do sprawdzenia (zob. ćwiczenia 15 i 16).

Zestawienie

Niech R będzie relacją w skończonym zbiorze S mającą macierz booleowską $\mathbf{A}$. Wówczas

(Z) Relacja R jest zwrotna wtedy i tylko wtedy, gdy macierz $\mathbf{A}$ ma na przekątnej głównej same jedynki.

(PZ) Relacja R jest przeciwzwrotna wtedy i tylko wtedy, gdy macierz $\mathbf{A}$ ma na przekątnej głównej same zera.

(S) Relacja R jest symetryczna wtedy i tylko wtedy, gdy $\mathbf{A} = \mathbf{A}^T$.

(AS) Relacja R jest antysymetryczna wtedy i tylko wtedy, gdy $\mathbf{A} \wedge \mathbf{A}^T \leq \mathbf{I}$, gdzie $\mathbf{I}$ jest macierzą jednostkową

(P) Relacja R jest przechodnia, wtedy i tylko wtedy, gdy $\mathbf{A} * \mathbf{A} \leq \mathbf{A}$.

Niech R_1 i R_2 będą relacjami na zbiorze $S \times T$, o macierzach booleowskich $\mathbf{A}_1$ i $\mathbf{A}_2$, gdzie S i T są danymi skończonymi zbiorami. Wówczas:

(a) $R_1 \subseteq R_2$ wtedy i tylko wtedy, gdy $\mathbf{A}_1 \leq \mathbf{A}_2$;

(b) macierzą booleowską relacji $R_1 \cup R_2$ jest $\mathbf{A}_1 \vee \mathbf{A}_2$;

(c) macierzą booleowską relacji $R_1 \cap R_2$ jest $\mathbf{A}_1 \wedge \mathbf{A}_2$.

I na koniec, złożenie relacji odpowiada iloczynowi booleowskiemu odpowiadających im macierzy, jak to zostało wyjaśnione w twierdzeniu 1.

ĆWICZENIA DO § 11.3

1. Dla każdej z następujących macierzy booleowskich weź odpowiadającą jej relację R w zbiorze $\{1, 2, 3\}$. Znajdź macierz booleowską relacji R^2 i stwierdź, czy relacja R jest przechodnia.

(a) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, (b) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, (c) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

2. Zrób rysunki relacji z ćwiczenia 1.

3. Niech $S = \{1, 2, 3\}$ i $R = \{(2, 1), (2, 3), (3, 2)\}$.
(a) Znajdź macierze relacji R i R^2.
(b) Zrób rysunki macierzy z części (a).
(c) Czy relacja R jest przechodnia?
(d) Czy relacja R^2 jest przechodnia?
(e) Czy relacja $R \cup R^2$ jest przechodnia?

4. Niech $S = \{1, 2, 3\}$ i $R = \{(1, 1), (1, 2), (1, 3), (3, 2)\}$.
(a) Znajdź macierze relacji R i R^2.
(b) Zrób rysunki relacji z części (a).

(c) Wykaż, że relacja R jest przechodnia, tzn. że $R^2 \subseteq R$, ale $R^2 \neq R$.
(d) Znajdź relacje R^n dla $n = 1, 2, 3, \ldots$.

5. Niech R będzie relacją w zbiorze $\{1, 2, 3\}$ o macierzy booleowskiej

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

(a) Znajdź macierze booleowskie relacji R^n dla $n \geq 0$.
(b) Czy relacja R jest zwrotna, symetryczna lub przechodnia?

6. Powtórz ćwiczenie 5 dla

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

7. Weźmy funkcje f i g, ze zbioru $\{1, 2, 3, 4\}$ w ten zbiór, zdefiniowane wzorami $f(m) = \max\{2,\ 4 - m\}$ i $g(m) = 5 - m$.
(a) Znajdź macierze booleowskie $\mathbf{A}_f$ i $\mathbf{A}_g$ relacji R_f i R_g odpowiadających funkcjom f i g.
(b) Znajdź macierze booleowskie relacji $R_f R_g$ i $R_{f \circ g}$; porównaj je ze sobą.
(c) Znajdź macierze booleowskie relacji odwrotnych $R_f^{\leftarrow}$ i $R_g^{\leftarrow}$. Czy relacje te odpowiadają funkcjom?

8. Znajdź macierze booleowskie następujących relacji w zbiorze $S = \{0, 1, 2, 3\}$.
(a) $(m, n) \in R_1$, jeśli $m + n = 3$.
(b) $(m, n) \in R_2$, jeśli $m \equiv n \pmod 2$.
(c) $(m, n) \in R_3$, jeśli $m \leq n$.
(d) $(m, n) \in R_4$, jeśli $m + n \leq 4$.
(e) $(m, n) \in R_5$, jeśli $\max\{m, n\} = 3$.

9. Dla każdej relacji z ćwiczenia 8 określ, które spośród własności (Z), (PZ), (S), (AS) i (P) ona spełnia.

10. (a) Które z relacji z ćwiczenia 8 są częściowymi porządkami?
(b) Które z relacji z ćwiczenia 8 są relacjami równoważności?

11. Niech R_1 i R_2 będą dowolnymi relacjami w zbiorze S. Dla każdego z następujących zdań udowodnij je lub wykaż, że jest ono fałszywe.
(a) Jeśli relacje R_1 i R_2 są zwrotne, to relacja $R_1 R_2$ też jest zwrotna.
(b) Jeśli relacje R_1 i R_2 są przechodnie, to relacja $R_1 R_2$ też jest przechodnia.
(c) Jeśli relacje R_1 i R_2 są symetryczne, to relacja $R_1 R_2$ też jest symetryczna.

12. Jak wygląda macierz booleowska relacji równości E w skończonym zbiorze S? Czy relacja ta jest zwrotna, symetryczna lub przechodnia?

13. Weźmy relacje R_1 i R_2 na zbiorze $S \times T$ i relacje R_3 i R_4 na zbiorze $T \times U$.

(a) Wykaż, że $R_1(R_3 \cup R_4) = R_1R_3 \cup R_1R_4$.
(b) Wykaż, że $(R_1 \cap R_2)R_3 \subseteq R_1R_3 \cap R_2R_3$, a równość nie musi zachodzić.
(c) Jaki związek zachodzi między relacjami $R_1(R_3 \cap R_4)$ oraz $R_1R_3 \cap R_1R_4$?

14. Niech R_1 będzie relacją na zbiorze $S \times T$, a R_2 — relacją na zbiorze $T \times U$. Wykaż, że relacją odwrotną do relacji R_1R_2 jest relacja $R_2^{\leftarrow}R_1^{\leftarrow}$.

15. Wykaż prawdziwość stwierdzeń (Z), (PZ), (S), (AS) i (P) z zestawienia zamieszczonego na końcu tego paragrafu.

16. Wykaż prawdziwość stwierdzeń (a), (b) i (c) z zestawienia zamieszczonego na końcu tego paragrafu.

17. Wykorzystaj prawo łączności składania relacji do wykazania, że iloczyn booleowski jest działaniem łącznym.

18. Niech R będzie relacją na zbiorze $S \times T$.
(a) Udowodnij, że relacja $RR^{\leftarrow}$ w zbiorze S jest symetryczna. Nie używaj macierzy booleowskich, gdyż któryś ze zbiorów S, T może być nieskończony.
(b) Zastosuj część (a) do szybkiego wykazania, że relacja $R^{\leftarrow}R$ w zbiorze T jest symetryczna.
(c) Przy jakich założeniach relacja $RR^{\leftarrow}$ jest zwrotna?

19. Niech R będzie antysymetryczną i przechodnią relacją w zbiorze S.
(a) Udowodnij, że relacja $R \cup E$ jest częściowym porządkiem w zbiorze S.
(b) Udowodnij, że relacja $R \backslash E$ jest quasi-porządkiem w zbiorze S.

§ 11.4. Domknięcia relacji

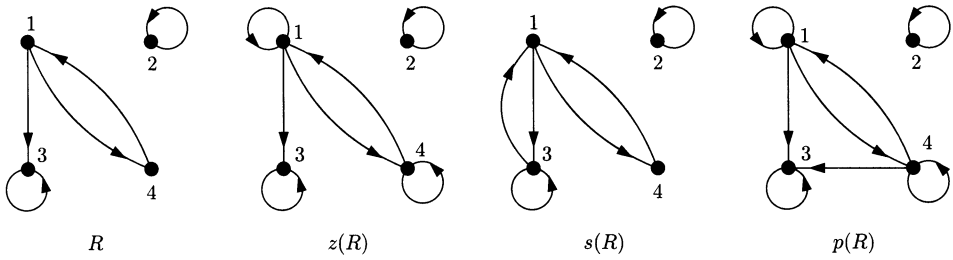
Czasami chcemy tworzyć nowe relacje za pomocą relacji, które już mamy. Na przykład, możemy mieć dwie relacje równoważności, tzn. relacje zwrotne, symetryczne i przechodnie, R_1 i R_2 w zbiorze S, i chcemy znaleźć relację równoważności, która zawiera je obie. Ponieważ R_1 i R_2 są podzbiorami zbioru $S \times S$, to oczywistym kandydatem jest relacja $R_1 \cup R_2$. Niestety jednak, relacja $R_1 \cup R_2$ nie musi być relacją równoważności; kłopot polega na tym, że niekoniecznie jest ona przechodnia. W takim razie, jak wygląda najmniejsza relacja przechodnia zawierająca $R_1 \cup R_2$? Okazuje się, że jest to bardzo dobre pytanie. Skąd wiemy, że taka relacja w ogóle istnieje? W dalszej części tego paragrafu zobaczymy, że jeśli R jest dowolną relacją w zbiorze S, to zawsze istnieje najmniejsza relacja przechodnia zawierająca R,

którą oznaczać będziemy przez $p(R)$; nauczymy się, jak ją znajdować. Istnieją również: najmniejsza relacja zwrotna zawierająca R i najmniejsza relacja symetryczna zawierająca R; oznaczać je będziemy, odpowiednio, przez $z(R)$ i $s(R)$.

PRZYKŁAD 1 Weźmy relację R w zbiorze $\{1,2,3,4\}$, której macierzą booleowską jest

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Na rysunku 11.20 znaleźć można rysunek relacji R.



Rysunek 11.20

(a) Relacja R nie jest zwrotna, ponieważ ani element 1, ani element 4 nie są same z sobą w relacji. Aby uzyskać relację zwrotną $z(R)$, musimy po prostu dołączyć do R pary uporządkowane $(1,1)$ i $(4,4)$. Macierzą booleowską $\mathbf{z}(\mathbf{A})$ relacji $z(R)$ jest po prostu macierz $\mathbf{A}$, w której wszystkie wyrazy na głównej przekątnej zostały zastąpione jedynkami:

$$\mathbf{z}(\mathbf{A}) = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Na rysunku 11.20, aby otrzymać rysunek relacji $z(R)$, dorysowaliśmy brakujące strzałki, łączące punkty same z sobą.

(b) Relacja R nie jest symetryczna, ponieważ $(1,3) \in R$, ale $(3,1) \notin R$. Jeśli dołączymy do R parę uporządkowaną $(3,1)$, to otrzymamy relację symetryczną $s(R)$. Jej macierzą booleowską jest

$$\mathbf{s}(\mathbf{A}) = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Aby otrzymać z rysunku relacji R rysunek relacji $s(R)$, dorysowaliśmy brakujące strzałki, skierowane przeciwnie do strzałek narysowanych wcześniej.

(c) Relacja R nie jest także przechodnia. Mamy, przykładowo, $(4,1) \in R$ i $(1,3) \in R$, ale $(4,3) \notin R$. Schemat znajdowania relacji $p(R)$ (i jej macierzy booleowskiej) nie jest tak prosty, jak pokazane metody dla relacji $z(R)$ i $s(R)$. Ponieważ $(4,1) \in R$ i $(1,3) \in R$, pary $(4,1)$ i $(1,3)$ będą również należeć do relacji $p(R)$. Ponieważ relacja $p(R)$ ma być przechodnia, to do $p(R)$ musi należeć para $(4,3)$, a więc zmuszeni jesteśmy parę $(4,3)$ do relacji $p(R)$ zaliczyć. Ogólnie, jeśli w grafie skierowanym relacji R istnieje droga biegnąca od wierzchołka x do wierzchołka y, tzn., jeśli istnieją punkty $x_1, x_2, \ldots, x_{m-1}$, takie, że wszystkie pary $(x, x_1), (x_1, x_2), \ldots, (x_{m-1}, y)$ należą do relacji R, to para (x, y) musi należeć do $p(R)$. Jeśli istnieje droga od x do y, a także droga od y do z, to istnieje droga od x do z. A więc zbiór wszystkich par uporządkowanych (x, y), których poprzedniki są połączone z następnikami drogami w naszym grafie skierowanym, jest relacją przechodnią i jest to najmniejsza relacja przechodnia zawierająca relację R, czyli relacja $p(R)$. W terminologii z § 3.2 relacja ta jest relacją osiągalności dla tego grafu skierowanego.

Na rysunku 11.20, aby otrzymać rysunek relacji $p(R)$ z rysunku relacji R, dorysowaliśmy krawędź od punktu x do punktu y, ilekroć punkty x i y łączyła jakaś droga w R, ale nie było wcześniej łączącej ich krawędzi. Na przykład dodaliśmy krawędź $(4,3)$ z uwagi na istnienie w R drogi 4 1 3 oraz dodaliśmy pętlę $(1,1)$, ponieważ w R istnieje droga 1 4 1. ∎

Następne stwierdzenie jest niemal oczywiste. Pomyśl, dlaczego jest ono prawdziwe, zanim przeczytasz dowód.

Stwierdzenie

Niech R będzie relacją. Wtedy $R = z(R)$ wtedy i tylko wtedy, gdy relacja R jest zwrotna, $R = s(R)$ wtedy i tylko wtedy, gdy relacja R jest symetryczna i $R = p(R)$ wtedy i tylko wtedy, gdy relacja R jest przechodnia. Ponadto,

$$z(z(R)) = z(R), \quad s(s(R)) = s(R) \quad \text{oraz} \quad p(p(R)) = p(R).$$

Dowód. Jeśli relacja R jest zwrotna, to jasne jest, że R jest najmniejszą relacją zwrotną zawierającą R, tzn. $R = z(R)$. Odwrotnie, jeśli $R = z(R)$, to relacja R jest zwrotna, ponieważ zwrotna jest relacja $z(R)$. Ponieważ $z(R)$ jest relacją zwrotną, to na mocy tego, co właśnie pokazaliśmy, $z(R) = z(z(R))$.

Dowody dla $s(R)$ i $p(R)$ są podobne. ∎

Możemy traktować r, s i t jak funkcje, które relacjom przyporządkowują relacje. Czasem funkcje takie, jak te nazywane są „operatorami". Nasze stwierdzenie pokazuje, że powtórne zastosowanie któregokolwiek z tych trzech operatorów nie daje niczego nowego; operatory o tej własności nazywane są **operatorami domknięcia**. Z drugiej strony, kolejne zastosowanie dwóch lub więcej spośród tych operatorów, może prowadzić do nowych relacji. Następne twierdzenie podaje ścisły opis relacji $z(R)$, $s(R)$ i $p(R)$ nazywanych, odpowiednio, **zwrotnym**, **symetrycznym** i **przechodnim domknięciem relacji** R.

Twierdzenie 1

Jeśli R jest relacją w zbiorze S i jeśli $E = \{(x,x)\colon x \in S\}$, to

$$\text{(z)} \quad z(R) = R \cup E;$$

$$\text{(s)} \quad s(R) = R \cup R^{\leftarrow};$$

$$\text{(p)} \quad p(R) = \bigcup_{k=1}^{\infty} R^k.$$

Dowód. (z) Relacja w zbiorze S jest zwrotna wtedy i tylko wtedy, gdy zawiera E. Wynika stąd, że relacja $R \cup E$ jest zwrotna oraz że każda relacja zwrotna, która zawiera R, musi także zawierać $R \cup E$. A więc $R \cup E$ jest najmniejszą relacją zwrotną zawierającą relację R. To dowodzi, że $z(R) = R \cup E$.

(s) Relacja jest symetryczna wtedy i tylko wtedy, gdy jest równa relacji do niej odwrotnej (ćwiczenie 11 z § 3.1). Jeśli $(x,y) \in R \cup R^{\leftarrow}$, to $(y,x) \in R^{\leftarrow} \cup R = R \cup R^{\leftarrow}$; zatem relacja $R \cup R^{\leftarrow}$ jest symetryczna. Weźmy dowolną relację symetryczną R' zawierającą R. Jeśli $(x,y) \in R^{\leftarrow}$, to $(y,x) \in R \subseteq R'$ i, ponieważ relacja R' jest symetryczna, $(x,y) \in R'$. Argument ten pokazuje, że $R^{\leftarrow} \subseteq R'$. Ale $R \subseteq R'$, więc ostatecznie otrzymujemy, że $R \cup R^{\leftarrow} \subseteq R'$. Zatem $R \cup R^{\leftarrow}$ jest najmniejszą relacją symetryczną zawierającą relację R, czyli $s(R) = R \cup R^{\leftarrow}$.

(p) Pokażemy najpierw, że suma $U = \bigcup_{k=1}^{\infty} R^k$ jest relacją przechodnią. Weźmy dowolne elementy x, y, z zbioru S takie, że $(x,y) \in U$ i $(y,z) \in U$. Wówczas musimy mieć $(x,y) \in R^k$ i $(y,z) \in R^j$, dla pewnych k i j ze zbioru $\mathbb{P}$. Stąd wynika, że para (x,z) należy do $R^k R^j = R^{k+j}$, a więc $(x,z) \in U$. Zatem U jest przechodnią relacją zawierającą relację R.

Weźmy teraz dowolną relację przechodnią R^* zawierającą R. Aby wykazać, że $U \subseteq R^*$, udowodnimy przez indukcję, że dla każdego $k \in \mathbb{P}$, $R^k \subseteq R^*$. Dla $k = 1$ inkluzja ta zachodzi na mocy wyboru R^*. Jeśli nasza inkluzja zachodzi dla pewnego k, to

$$R^{k+1} = R^k R \subseteq R^* R \subseteq R^* R^* \subseteq R^*;$$

ostatnie zawieranie wynika z przechodniości relacji R^* (twierdzenie 3 z § 11.3). Na mocy zasady indukcji $R^k \subseteq R^*$ dla wszystkich $k \in \mathbb{P}$, a więc $U \subseteq R^*$. Zatem U jest najmniejszą relacją przechodnią zawierającą relację R oraz

$$p(R) = U = \bigcup_{k=1}^{\infty} R^k. \qquad \blacksquare$$

Jeśli zbiór S jest skończony, to opis relacji $p(R)$ przedstawiony w twierdzeniu 1 można ulepszyć.

Twierdzenie 2

Jeśli R jest relacją w zbiorze S mającym n elementów, to

$$p(R) = \bigcup_{k=1}^{n} R^k.$$

Dowód. Rozważmy graf skierowany relacji R. Para (x, y) należy do $p(R)$ wtedy i tylko wtedy, gdy w tym grafie skierowanym istnieje droga z x do y. Jeśli droga taka istnieje, to istnieje też droga z x do y, która nie przechodzi dwukrotnie przez ten sam wierzchołek, wyjąwszy przypadek, gdy $x = y$. Nie może ona przechodzić przez więcej niż n wierzchołków, a więc jej długość nie może być większa niż n. A zatem $(x, y) \subset R^k$ dla pewnego $k \leq n$. (To rozumowanie jest w gruncie rzeczy dowodem twierdzenia 1 z § 6.1). $\blacksquare$

PRZYKŁAD 2

(a) Załóżmy, że R jest relacją w zbiorze S mającym n elementów i $\mathbf{A}$ jest macierzą booleowską relacji R. Twierdzenie 2 oraz zebrane w § 11.3 przykłady macierzowych odpowiedników różnych własności relacji i działań na nich pokazują, że macierzami booleowskimi relacji $p(R)$, $s(R)$ i $z(R)$ są

$$\mathbf{p}(\mathbf{A}) = \mathbf{A} \vee \mathbf{A}^2 \vee \ldots \vee \mathbf{A}^n,$$
$$\mathbf{s}(\mathbf{A}) = \mathbf{A} \vee \mathbf{A}^T$$

oraz

$$\mathbf{z}(\mathbf{A}) = \mathbf{A} \vee \mathbf{I},$$

gdzie $\mathbf{I}$ jest macierzą jednostkową wymiaru $n \times n$, a potęgi $\mathbf{A}^k$ oznaczają potęgi booleowskie.

(b) Dla relacji R z przykładu 1 łatwo widzieć, że $\mathbf{s}(\mathbf{A}) = \mathbf{A} \vee \mathbf{A}^T$ i $\mathbf{z}(\mathbf{A}) = \mathbf{A} \vee \mathbf{I}$, gdzie $\mathbf{I}$ jest macierzą jednostkową wymiaru 4×4. Można też sprawdzić, że

$$\mathbf{p}(\mathbf{A}) = \mathbf{A} \vee \mathbf{A}^2 = \mathbf{A} \vee \mathbf{A}^2 \vee \mathbf{A}^3 \vee \mathbf{A}^4 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Oczywiście dla tak prostej relacji łatwiej jest znaleźć $p(R)$ korzystając z rysunku relacji R. ∎

Dla danej macierzy booleowskiej $\mathbf{A}$ łatwo jest znaleźć macierze $\mathbf{s}(\mathbf{A})$ i $\mathbf{z}(\mathbf{A})$ posługując się wzorami z przykładu 2(a). Możemy również używać przedstawionego w tym przykładzie wzoru na obliczanie macierzy $\mathbf{p}(\mathbf{A})$, znajdując potęgi booleowskie $\mathbf{A}^k = \mathbf{A} * \ldots * \mathbf{A}$, a następnie sumując otrzymane wyniki. Dla dużych wartości n takie obliczenia wymagają wielokrotnego mnożenia przez siebie dużych macierzy i przebiegają dość wolno.

Jeżeli S jest dużym zbiorem, to relację $p(R)$ możemy szybciej znaleźć posługując się grafem skierowanym relacji R. Zauważyliśmy w § 3.2 oraz w przykładzie 1, że $p(R)$ jest relacją osiągalności dla grafu skierowanego relacji R, którego relacją sąsiedztwa jest R. Jeśli każdej krawędzi przypiszemy wagę 1, to algorytm WARSHALLA z § 8.4 znajdzie macierz $\mathbf{W}^*$, której wyraz o współrzędnych (i, j) jest dodatnią liczbą całkowitą, o ile w naszym grafie skierowanym istnieje droga z v_i do v_j, a w przeciwnym razie równy jest ∞. Zastąpienie w otrzymanym wyniku symboli ∞ zerami, a liczb całkowitych jedynkami, daje macierz booleowską $\mathbf{p}(\mathbf{A})$. W rzeczywistości (ćwiczenie 18), niewielka zmiana algorytmu pozwala nam obliczać $\mathbf{p}(\mathbf{A})$ dla macierzy booleowskich stosując działania booleowskie bez potrzeby obliczeń na dużych liczbach całkowitych i wprowadzania symbolu ∞.

PRZYKŁAD 3 Dla relacji R z przykładu 1 otrzymaliśmy

$$\mathbf{z}(\mathbf{A}) = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{oraz} \quad \mathbf{s}(\mathbf{A}) = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Macierzą booleowską relacji $sz(R) = s(z(R))$ jest

$$\mathbf{sz(A)} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$

Jest to także macierz $\mathbf{zs(A)}$ relacji $zs(R)$, a więc $zs(R) = sz(R)$. Ta równość nie jest przypadkowa (ćwiczenie 11(b)). Przechodnie domknięcie relacji $sz(R) = zs(R)$ ma macierz

$$\mathbf{psz(A)} = \mathbf{pzs(A)} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix},$$

będącą także macierzą relacji równoważności w zbiorze $\{1, 2, 3, 4\}$, której klasami równoważności są zbiory $\{2\}$ i $\{1, 3, 4\}$. Zatem relacja $psz(R)$ jest przechodnia, symetryczna i zwrotna. Fakt ten może wydawać się oczywisty z powodu użytego przez nas oznaczenia psz, ale musimy tu być ostrożni. Jest na przykład do pomyślenia, że domknięcie przechodnie relacji symetrycznej mogłoby nie być relacją symetryczną. Następny nasz przykład pokazuje, że domknięcie symetryczne relacji przechodniej nie musi być relacją przechodnią; musimy więc być ostrożni z pochopnym wyciąganiem wniosków na podstawie oznaczeń. ∎

PRZYKŁAD 4 Niech R będzie relacją w zbiorze $\{1, 2, 3\}$ o macierzy booleowskiej

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Wówczas relacja R jest zwrotna (gdyż $\mathbf{I} \leq \mathbf{A}$) i przechodnia (gdyż $\mathbf{A}*\mathbf{A} = \mathbf{A}$), a więc $\mathbf{A} = \mathbf{z(A)} = \mathbf{p(A)} = \mathbf{pz(A)} = \mathbf{zp(A)}$. Relacja $s(R)$ ma macierz

$$\mathbf{s(A)} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Ponieważ pary $(2, 1)$ i $(1, 3)$ należą do $s(R)$, ale para $(2, 3)$ doń nie należy, to relacja $s(R)$ nie jest przechodnia. Mamy

$$\mathbf{sp(A)} = \mathbf{spz(A)} = \mathbf{szp(A)} = \mathbf{s(A)} \neq \mathbf{ps(A)} = \mathbf{psz(A)} = \mathbf{pzs(A)}.$$

Zatem kolejność stosowania różnych operatorów domknięcia jest istotna. ∎

Następny lemat pokazuje, że kłopot, z jakim zetknęliśmy się w przykładzie 4, jest jedynym, jaki może się zdarzyć. Poza przypadkiem operatora symetrycznego domknięcia, który może zniszczyć przechodniość relacji, tworzenie domknięć zachowuje posiadane już przez relację własności.

Lemat

(a) Jeśli relacja R jest zwrotna, to zwrotne są też relacje $s(R)$ i $p(R)$.
(b) Jeśli relacja R jest symetryczna, to symetryczne są też relacje $z(R)$ i $p(R)$.
(c) Jeśli relacja R jest przechodnia, to przechodnia jest też relacja $z(R)$.

Dowód. (a) To jest oczywiste, ponieważ $E \subseteq R$, a stąd mamy $E \subseteq s(R)$ i $E \subseteq p(R)$. Część (b) zostawiamy jako ćwiczenie 10.

(c) Załóżmy, że relacja R jest przechodnia i weźmy pary (x, y), i (y, z) ze zbioru $z(R) = R \cup E$. Jeśli $(x, y) \in E$, to $x = y$, a więc para $(x, z) = (y, z)$ należy do $R \cup E$. Jeśli $(y, z) \in E$, to $y = z$, a więc para $(x, z) = (x, y)$ należy do $R \cup E$. Jeśli ani para (x, y), ani para (y, z) nie należą do E, to obie należą do R, a więc $(x, z) \in R \subseteq R \cup E$ na mocy przechodniości relacji R. Zatem w każdym z przypadków $(x, z) \in R \cup E$. ∎

Kolejne twierdzenie daje odpowiedź na zasadnicze pytanie, od którego rozpoczęliśmy ten paragraf.

Twierdzenie 3

Dla dowolnej relacji R w zbiorze S, $psz(R)$ jest najmniejszą relacją równoważności zawierającą R.

Dowód. Ponieważ relacja $z(R)$ jest zwrotna, dwukrotne zastosowanie punktu (a) powyższego lematu pokazuje, że relacja $psz(R)$ jest zwrotna. Ponieważ relacja $sz(R)$ jest symetryczna, to jednokrotne zastosowanie punktu (b) lematu pokazuje, że relacja $psz(R)$ jest symetryczna. W końcu, ponieważ relacja $psz(R)$ jest przechodnia, to jest ona relacją równoważności.

Weźmy dowolną relację równoważności R' taką, że $R \subseteq R'$. Wówczas $z(R) \subseteq z(R') = R'$, a więc $sz(R) \subseteq s(R') = R'$, a stąd $psz(R) \subseteq p(R') = R'$. Zatem $psz(R)$ jest najmniejszą relacją równoważności zawierającą R. ∎

PRZYKŁAD 5

(a) W przykładzie 3 relacja $psz(R)$ okazała się relacją równoważności o klasach abstrakcji $\{2\}$ i $\{1, 3, 4\}$.

(b) Niech R będzie relacją w zbiorze $\{1,2,3\}$ opisaną w przykładzie 4. Wówczas

$$\mathbf{z}(\mathbf{A}) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{sz}(\mathbf{A}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix},$$

$$\mathbf{psz}(\mathbf{A}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Najmniejszą relacją równoważności zawierającą R jest relacja uniwersalna $\{1,2,3\}\times\{1,2,3\}$. Te obliczenia można dodatkowo sprawdzić przez wykonanie rysunków odpowiednich relacji. ∎

Twierdzenie 3 ma interpretację w języku teorii grafów. Zaczynając od grafu skierowanego, którego relacją sąsiedztwa jest R, tworzymy graf skierowany relacji $z(R)$ przez dodanie pętli we wszystkich wierzchołkach. Następnie tworzymy graf skierowany relacji $sz(R)$ przez dodanie, jeśli trzeba, krawędzi tak, by dla dowolnych wierzchołków x, y, z istnienia krawędzi od x do y wynikało istnienie krawędzi od y do x. Wynikiem tego jest w gruncie rzeczy graf nieskierowany z pętlą w każdym wierzchołku. W końcu, krawędzie grafu relacji $psz(R)$ łączą te wierzchołki x, y, które są połączone drogą w grafie relacji $sz(R)$. Zatem para (x,y) należy do relacji $psz(R)$ wtedy i tylko wtedy, gdy wierzchołki x i y należą do tej samej składowej spójnej grafu relacji $sz(R)$. Dowolny algorytm, taki jak LAS z § 6.6, który znajduje składowe spójne, potrafi znaleźć także $psz(R)$.

ĆWICZENIA DO § 11.4

1. Weźmy relację R w zbiorze $\{1,2,3\}$ o macierzy booleowskiej

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Znajdź macierze booleowskie relacji:
(a) $z(R)$, (b) $s(R)$, (c) $zs(R)$, (d) $sz(R)$, (e) $psz(R)$.

2. Powtórz ćwiczenie 1 dla $\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

3. Dla relacji R z ćwiczenia 1 wypisz klasy abstrakcji relacji $psz(R)$.

4. Dla relacji R z ćwiczenia 2 wypisz klasy abstrakcji relacji $psz(R)$.

5. Powtórz ćwiczenie 1 dla relacji R w zbiorze $\{1,2,3,4,5\}$ o macierzy booleowskiej

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

6. Dla relacji R z ćwiczenia 5, wypisz klasy abstrakcji relacji $psz(R)$.

7. Niech R będzie zwykłym quasi-porządkiem w zbiorze $\mathbb{P}$: $(m,n) \in R$, jeśli $m < n$. Znajdź lub opisz relacje:
(a) $z(R)$, (b) $sz(R)$, (c) $zs(R)$, (d) $psz(R)$, (e) $p(R)$, (f) $sp(R)$.

8. Powtórz ćwiczenie 7, gdy teraz $(m,n) \in R$ znaczy, że m dzieli n.

9. Zakon Wrogich Sobie Pustelników jest interesującą organizacją. Każdy pustelnik zna sam siebie. Ponadto, każdy zna Najwyższego Pustelnika, ale ani on, ani żaden inny członek Zakonu nie zna nikogo spośród pozostałych członków. Definiujemy relację R w zbiorze Z.W.S.P. mówiąc, że $(p_1,p_2) \in R$, jeśli p_1 zna p_2. Określ i porównaj ze sobą relacje $sp(R)$ i $ps(R)$. (Najwyższy Pustelnik pełni funkcję podobną trochę do roli serwera plików w sieci komputerowej.)

10. (a) Wykaż, że jeśli (R_k) jest ciągiem relacji symetrycznych w zbiorze S, to suma $\bigcup_{k=1}^{\infty} R_k$ jest relacją symetryczną.
(b) Niech R będzie relacją symetryczną w zbiorze S. Wykaż, że dla każdego $n \in \mathbb{P}$ relacja R^n jest symetryczna.
(c) Wykaż, że jeśli relacja R jest symetryczna, to symetryczne są też relacje $z(R)$ i $p(R)$.

11. Weźmy relację R w zbiorze S.
(a) Wykaż, że $pz(R) = zp(R)$.
(b) Wykaż, że $sz(R) = zs(R)$.

12. Niech R_1 i R_2 będą dwuargumentowymi relacjami w zbiorze S.
(a) Wykaż, że $z(R_1 \cup R_2) = z(R_1) \cup z(R_2)$.
(b) Wykaż, że $s(R_1 \cup R_2) = s(R_1) \cup s(R_2)$.
(c) Czy zawsze jest prawdą, że $z(R_1 \cap R_2) = z(R_1) \cap z(R_2)$? Uzasadnij odpowiedź.
(d) Czy zawsze jest prawdą, że $s(R_1 \cap R_2) = s(R_1) \cap s(R_2)$? Uzasadnij odpowiedź.

13. Weźmy dwie relacje równoważności R_1 i R_2 w zbiorze S.
(a) Wykaż, że $p(R_1 \cup R_2)$ jest najmniejszą relacją równoważności zawierającą jednocześnie R_1 i R_2. *Wskazówka*: skorzystaj z ćwiczenia 12(a) i (b).
(b) Opisz największą relację równoważności zawartą jednocześnie w R_1 i R_2.

14. Wykaż, że $sp(R) \neq ps(R)$, gdzie R jest relacją z przykładu 4.

15. Wykaż, że nie istnieje najmniejsza relacja przeciwzwrotna zawierająca relację R w zbiorze $\{1,2\}$ o macierzy booleowskiej $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$.

16. Mówimy, że relacja R w zbiorze S jest **relacją „na"**, jeśli dla każdego $y \in S$ istnieje $x \in S$ taki, że $(x,y) \in R$. Wykaż, że nie istnieje najmniejsza relacja „na", zawierająca relację R w zbiorze $\{1,2\}$, opisaną w ćwiczeniu 15.

17. Załóżmy, że w jest pewną własnością, jaką może mieć relacja w niepustym zbiorze S, spełniającą następujące warunki:
(i) relacja uniwersalna $S \times S$ ma własnośc w;
(ii) własność w jest **zamknięta na przecięcia**, tzn. jeśli $\{R_i\colon i \in I\}$ jest niepustą rodziną relacji w S, mających własność w, to przecięcie $\bigcap_{i\in I} R_i$ też ma własność w.
(a) Udowodnij, że dla dowolnej relacji R istnieje najmniejsza relacja zawierająca R i mająca własność w.
(b) Przekonaj się, że własności: zwrotności, symetrii i przechodniości spełniają warunki (i) i (ii).
(c) Którego z warunków (i), (ii) nie spełnia własność przeciwzwrotności?
(d) Którego z warunków (i), (ii) nie spełnia własność bycia relacją „na" z ćwiczenia 16?

18. Zmodyfikuj algorytm WARSHALLA tak, by macierz $\mathbf{p}(\mathbf{A})$ można było obliczać zaczynając bezpośrednio od $\mathbf{A}$ i używając działań booleowskich.

To, co jest najważniejsze w tym rozdziale

Jak zwykle: Co to znaczy? Dlaczego tutaj się znalazło? Jak mogę to zastosować? Myśl o przykładach.

Pojęcia

częściowy porządek, $\preceq$, zbiór częściowo uporządkowany, podzbiór zbioru częściowo uporządkowanego
  porządek odwrotny, $\succeq$
  quasi-porządek, $\prec$
  diagram Hassego
  element maksymalny, minimalny, największy, najmniejszy
  ograniczenie górne, dolne
    kres górny, $x \vee y = \sup\{x,y\}$
    kres dolny, $x \wedge y = \inf\{x,y\}$

krata
zbiór liniowo uporządkowany = łańcuch
zbiór dobrze uporządkowany
porządek produktowy w zbiorze $S_1 \times \ldots \times S_n$
porządek leksykograficzny w zbiorze $S_1 \times \ldots \times S_n$, $\preceq^k$ w zbiorze S^k
porządek standardowy w zbiorze Σ^*
leksykograficzny = słownikowy porządek $\preceq_L$ w zbiorze Σ^*
złożenie relacji $R_2 \circ R_1 = R_1 R_2$
relacja równości E
macierz booleowska, iloczyn booleowski $*$

Fakty

Każdy skończony zbiór częściowo uporządkowany ma diagram Hassego.

Jeśli każdy ze zbiorów S_i jest liniowo uporządkowany, to porządek leksykograficzny w zbiorze $S_1 \times \ldots \times S_n$ jest liniowy.

Jeśli zbiór Σ jest liniowo uporządkowany, to porządek standardowy w zbiorze Σ^* jest dobrym porządkiem, a porządek leksykograficzny w zbiorze Σ^* jest liniowy, ale nie jest dobrym porządkiem.

Składanie relacji jest łączne.

Dla danych relacji R_1 i R_2 macierz ich złożenia $R_1 R_2$ jest iloczynem booleowskim $\mathbf{A}_1 * \mathbf{A}_2$ ich macierzy.

Relacja R w zbiorze S jest przechodnia wtedy i tylko wtedy, gdy $R^2 \subseteq R$ wtedy i tylko wtedy, gdy $\mathbf{A} * \mathbf{A} \leq \mathbf{A}$.

Odpowiedniki macierzowe niektórych często spotykanych własności relacji są zebrane na końcu § 11.3.

Operatory z, s i p zdefiniowane wzorami $z(R) = R \cup E$, $s(R) = R \cup R^{\leftarrow}$ i $p(R) = \bigcup_{k=1}^{\infty} R^k$ są operatorami domknięcia w klasie wszystkich relacji.

$p(R) = \bigcup_{k=1}^{n} R^k$, o ile $|S| = n$.

Relacja $sp(R)$ może nie być przechodnia.

Najmniejszą relacją równoważności zawierającą relację R jest $psz(R)$.

Metody

Algorytm WARSHALLA z wagami wszystkich krawędzi równymi 1 może znaleźć $p(R)$.

Algorytm LAS może wyznaczyć $psz(R)$, znajdując składowe spójne grafu relacji $sz(R)$.

12. STRUKTURY ALGEBRAICZNE

Koncepcja przyjęta w tym rozdziale polega na zmierzaniu od konkretu do abstrakcji. Najpierw badamy permutacje i grupy permutacji. Dowodzimy paru twierdzeń dotyczących zliczania i stosujemy je do pewnych nietrywialnych problemów kolorowania. Po takim przygotowaniu, wprowadzamy w paragrafach 12.5–12.8 ogólną definicję grupy i innych systemów algebraicznych. Te późniejsze paragrafy napisane są tak, by można było czytać je niezależnie od paragrafów 12.1–12.4, opuszczając przykłady odnoszące się do grup permutacji. Uważamy jednak, że większość Czytelników lepiej opanuje ten materiał, badając najpierw sytuacje konkretne.

§ 12.1. Permutacje

Permutacja jest to wzajemnie jednoznaczne przekształcenie pewnego zbioru na siebie. W tym paragrafie będziemy badać zbiór S_n wszystkich permutacji zbioru n-elementowego. (S_n oznacza grupę symetryczną zbioru n-elementowego). Można przyjąć, że tym zbiorem n-elementowym jest zbiór $\{1, 2, \ldots, n\}$, ale w zastosowaniach może to być zbiór X wierzchołków pewnego grafu, krawędzi jakiegoś grafu skierowanego itd. W takich przypadkach zbiór wszystkich permutacji zbioru X będziemy również oznaczać przez $\mathrm{PERM}(X)$.

Stwierdziliśmy dawno temu, że jeśli f i g są permutacjami, to są nimi także funkcje $f \circ g$ oraz f^{-1} i g^{-1}. Istotnie, ćwiczenia 9 i 10 z § 1.4 stwierdzały, że złożenie przekształceń wzajemnie jednoznacznych jest przekształceniem wzajemnie jednoznacz-

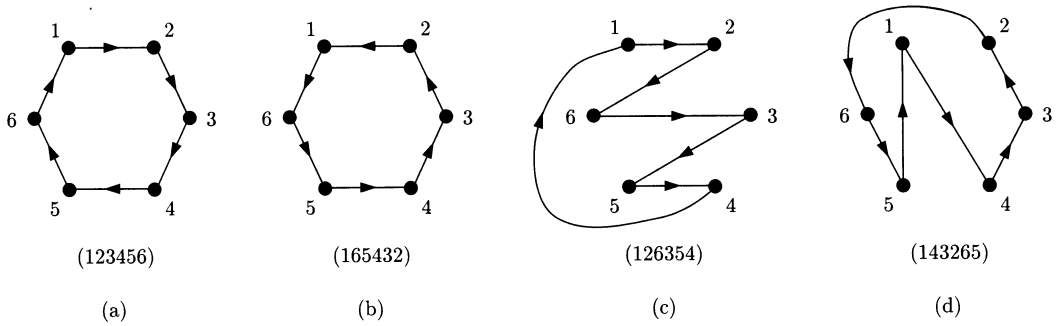
nym oraz że funkcja odwrotna do przekształcenia wzajemmnie jednoznacznego jest też przekształceniem wzajemnie jednoznacznym. W szczególności, $f \circ f^{-1} = f^{-1} \circ f = e$ dla wszystkich $f \in S_n$, gdzie e jest **permutacją identycznościową**: $e(k) = k$ dla $k \in \{1, 2, \ldots, n\}$.

Dla $n \geq 3$, w zbiorze S_n nie jest spełnione prawo przemienności, tzn. że, jak zobaczymy w przykładzie 2(c), $f \circ g$ nie musi równać się $g \circ f$. Dla każdego n grupa S_n ma $n!$ elementów; ta liczba wraz ze wzrostem n rośnie bardzo szybko. W naszych pierwszych przykładach zobaczymy kilka elementów zbioru S_6, mającego 720 elementów.

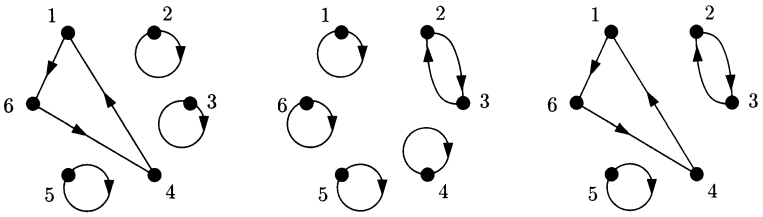
Aby zajmować się permutacjami, musimy wprowadzić jakieś wygodne oznaczenia. Weźmy, na przykład, permutację ze zbioru S_6, przedstawioną na rysunku 12.1(a). Tę permutację oznaczymy przez (1 2 3 4 5 6). Zatem (1 2 3 4 5 6) jest nazwą permutacji, która przeprowadza 1 na 2, 2 na 3, 3 na 4, 4 na 5, 5 na 6 i 6 na 1. To oznaczenie pozwala łatwo, nawet bez potrzeby robienia rysunku, zdać sobie sprawę, o jaką permutację chodzi; pomyśl o strzałkach biegnących od 1 do 2, od 2 do 3 itd., a na koniec od 6 do 1. Innymi słowy wyobraź sobie

$$(1 \to 2 \to 3 \to 4 \to 5 \to 6)$$

Możemy również oznaczyć tę permutację przez (2 3 4 5 6 1), (4 5 6 1 2 3) itd. Permutacja (1 6 5 4 3 2) = (6 5 4 3 2 1), przedstawiona na rysunku 12.1(b), jest permutacją odwrotną do permutacji (1 2 3 4 5 6). Rysunki 12.1(c) i 12.1(d) przedstawiają, odpowiednio, permutacje (1 2 6 3 5 4) i (1 4 3 2 6 5).



Rysunek 12.1

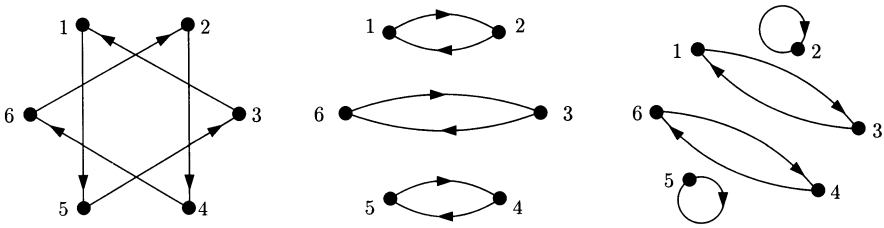


Rysunek 12.2

Permutacja niekoniecznie musi ruszać wszystkie elementy zbioru $\{1,2,3,4,5,6\}$. Jeśli permutacja p nie rusza pewnego elementu s, tzn. jeśli $p(s) = s$, to mówimy, że s jest **punktem stałym permutacji** p. Dla permutacji przedstawionej na rysunku 12.2(a), która przeprowadza 1 na 6, 6 na 4 i 4 na 1, a liczby 2, 3 i 5 są jej punktami stałymi, używamy oznaczenia (1 6 4). Rysunek 12.2(b) przedstawia permutację (2, 3), a rysunek 12.2(c) — iloczyn (1 6 4)(2 3) poprzednich dwóch permutacji. Iloczyn znaczy tu po prostu tyle, co złożenie tych dwóch permutacji. Widzimy, że ponieważ zbiory $\{1,6,4\}$ i $\{2,3\}$ są rozłączne, to nie ma znaczenia, czy napiszemy (1 6 4)(2 3), czy też (2 3)(1 6 4).

Permutacja, którą możemy zapisać w postaci $(a_1 a_2 \ldots a_n)$, gdzie $a_1, a_2, \ldots, a_n$ są parami różne, nazywana jest **cyklem**. Zatem permutacje (1 6 4) i (2 3) są cyklami. Zobaczymy, że nieprzypadkowo rysunki tych permutacji są cyklami w sensie teorii grafów. Permutacja (1 6 4)(2 3) nie jest cyklem; przemieszczane przez nią cyklicznie elementy tworzą dwie rozłączne klasy. Mówimy, że dwa cykle $(a_1 a_2 \ldots a_n)$ i $(b_1 b_2 \ldots b_m)$ są **rozłączne**, jeśli zbiory $\{a_1, a_2, \ldots, a_n\}$ i $\{b_1, b_2, \ldots, b_m\}$ nie mają wspólnych elementów. Permutacja (1 6 4)(2 3) jest iloczynem dwóch rozłącznych cykli.

PRZYKŁAD 1 (a) Rozważmy permutację z S_6, przedstawioną na rysunku 12.3(a). Postępując zgodnie z naturalnym porządkiem, zaczynamy od liczby 1 i poruszamy się zgodnie z naszą permutacją, aż do momentu, gdy powrócimy do liczby 1. Otrzymujemy (1 5 3), gdyż permutacja przeprowadza 1 na 5, 5 na 3, a 3 z powrotem na 1. Postępując nadal zgodnie z naturalnym porządkiem, wybieramy najmniejszą z liczb, których jeszcze nie odwiedziliśmy i powtarzamy poprzednią procedurę. Otrzymujemy (2 4 6) i wnosimy stąd, że rysunek 12.3(a) przedstawia permutację (1 5 3)(2 4 6).



Rysunek 12.3

(b) Na rysunku 12.3(b), zaczynamy od liczby 1 i otrzymujemy (1 2). Ponieważ 3 jest najmniejszą z liczb dotychczas nie odwiedzonych, więc bierzemy 3 i otrzymujemy (3 6). Następnie otrzymujemy (4 5), więc ostatecznie nasza permutacja jest równa (1 2)(3 6)(4 5).

(c) Na rysunku 12.3(c), zaczynamy od liczby 1 i otrzymujemy (1 3). Następnie bierzemy liczbę 2, ale jest ona punktem stałym, tzn. jest przeprowadzana sama na siebie. Aby to odnotować, napiszemy (2). Następnie bierzemy 4 i otrzymujemy (4 6). Na koniec, 5 jest punktem stalym, więc piszemy (5). Nasza permutacja jest zatem równa (1 3)(2)(4 6)(5), ale (2) i (5) to po prostu permutacje identycznościowe, choć zapisane nieco inaczej. W takim razie możemy napisać (1 3)(2)(4 6)(5) = (1 3)(4 6). ∎

Twierdzenie 1

Każda permutacja ze zbioru S_n jest iloczynem rozłącznych cykli.

Dowód. Niech g należy do S_n. Rozważmy graf skierowany, którego zbiorem wierzchołków jest zbiór $V = \{1, 2, \ldots, n\}$, a zbiorem krawędzi zbiór $\{(k, g(k))\colon k \in V\}$. Rysunki 12.1, 12.2 i 12.3 są grafami skierowanymi tego typu i mogą posłużyć za modele dla tego dowodu. Każdy wierzchołek k ma stopień wyjściowy 1, ponieważ funkcja g przyjmuje w punkcie k dokładnie jedną wartość $g(k)$. Każdy wierzchołek k ma stopień wejściowy 1, gdyż $k = g(l)$ dla dokładnie jednej liczby l; w istocie $l = g^{-1}(k)$.

Teraz zapomnijmy na chwilę o strzałkach na krawędziach naszego grafu skierowanego i rozważmy składowe spójne otrzymanego tą drogą grafu nieskierowanego. Przyjrzyjmy się naszym przykładom. Każdy graf z rysunku 12.1 jest spójny, więc ma tylko jedną składową: cały graf. Grafy z rysunku 12.2 mają, odpowiednio, po 4, 5 i 3 składowe, podczas gdy grafy z rysunku 12.3 mają, odpowiednio, po 2, 3 i 4 składowe. Twierdzenie Eulera w wersji

dla grafów skierowanych (twierdzenie 4, § 8.1) zapewnia, że w każdej składowej istnieje zamknięta droga skierowana, przechodząca przez wszystkie jej krawędzie. Ponieważ stopień wejściowy i wyjściowy każdego z wierzchołków wynosi 1, więc wierzchołki się nie powtarzają i te zamknięte drogi skierowane są cyklami w sensie teorii grafów. Cykle te odpowiadają cyklom w sensie odnoszącym się do permutacji. Wyjściowa permutacja jest iloczynem wszystkich takich cykli. ∎

Mnożąc permutacje musimy pamiętać, że w rzeczywistości składamy funkcje, takie jak np. $g_1 \circ g_2 \circ g_3$. By znaleźć $g_1 \circ g_2 \circ g_3(k)$, oblicza się najpierw $g_3(k)$, potem do wyniku stosuje się g_2, co daje $g_2(g_3(k))$, a następnie stosuje się g_1 otrzymując $g_1(g_2(g_3(k)))$. Permutacje stosuje się w kolejności od prawej strony do lewej. Ponieważ złożenie permutacji jest także permutacją, aby znaleźć jego reprezentację w postaci iloczynu cykli, postępuje się z poszczególnymi elementami tak, jak w zamieszczonym powyżej przykładzie 1.

PRZYKŁAD 2 (a) Zilustrujemy szczegółowo jak mnoży (składa!) się permutacje z S_6. Znajdźmy wynik mnożenia permutacji (1 6 4)(2 3) (z rysunku 12.2(c)) przez permutację (2 5 3 4 6). Chcemy zatem znaleźć

$$(1\,6\,4)(2\,3)(2\,5\,3\,4\,6).$$

Proces mnożenia, po nabraniu pewnej wprawy, jest łatwy i nie wymaga wcale pisania, ale jego wyjaśnienie jest długie; przeczytaj je starannie, a następnie zrób ćwiczenia 5 i 6 dla nabrania wprawy.

Krok 1. Zaczynamy od znalezienia wartości w punkcie 1, biorąc pod uwagę kolejne cykle od strony prawej do lewej. Liczba 1 nie jest ruszona ani przez permutację (2 5 3 4 6), ani przez permutację (2 3), a w końcu jest przeprowadzona na 6 przez permutację (1 6 4). Stwierdzamy, że 1 przechodzi na 6 i odnotowujemy to, pisząc „(1 6".

Krok 2. Aby kontynuować rozpoczęty cykl, znajdziemy teraz wartość w punkcie 6. Najpierw permutacja (2 5 3 4 6) przeprowadza 6 na 2, następnie permutacja (2 3) przeprowadza 2 na 3, a w końcu permutacja (1 6 4) nie rusza 3. Stwierdzamy, że 6 przechodzi na 3 i dopisujemy to do naszych dotychczasowych ustaleń: „(1 6 3".

Krok 3. Aby znaleźć wartość w punkcie 3 zauważmy, że permutacja (2 5 3 4 6) przeprowadza go na 4, potem per-

mutacja (2 3) nie rusza 4, a następnie permutacja (1 6 4) przeprowadza 4 na 1. To zamyka cykl: „(1 6 3)". Zauważ, że zamknęliśmy po prostu nawias dla odnotowania, że mamy już cały cykl.

Krok 4. W kroku 3 otrzymaliśmy jeden cykl z rozkładu naszego iloczynu na cykle. Aby znaleźć kolejny cykl, wybieramy element, którego w uzyskanym cyklu nie ma i śledzimy jego losy. Żeby postępować metodycznie, bierzemy liczbę 2, najmniejszą z liczb wchodzących w grę. Najpierw permutacja (2 5 3 4 6) przeprowadza 2 na 5, potem permutacje (2 3) i permutacje (1 6 4) już 5 nie ruszają. Zatem 2 przechodzi na 5 i odnotowujemy: „(1 6 3)(2 5".

Krok 5. Aby znaleźć wartość w punkcie 5 zauważmy, że permutacja (2 5 3 4 6) przeprowadza 5 na 3, potem permutacja (2 3) przeprowadza 3 na 2, a w końcu permutacja (1 6 4) nie rusza 2. Zatem 5 przechodzi na 2, zamykając kolejny cykl: „(1 6 3)(2 5)".

Krok 6. Ponieważ wiemy już, co się dzieje ze wszystkimi liczbami z wyjątkiem 4, to 4 musi oczywiście przejść na 4. Jednakże, rozsądnie będzie wykonać dodatkowe sprawdzenie: permutacja (2 5 3 4 6) przeprowadza 4 na 6, permutacja (2 3) nie rusza 6, a w końcu permutacja (1 6 4) przeprowadza 6 z powrotem na 4. Tak, jak się spodziewaliśmy, 4 przechodzi na 4. Dopisujemy to do naszych ustaleń: „(1 6 3)(2 5)(4)".

Krok 7. Wiemy już, co się dzieje ze wszystkimi wierzchołkami i zadanie jest zakończone. Zatem

$$(1\,6\,4)(2\,3)(2\,5\,3\,4\,6) = (1\,6\,3)(2\,5)(4) = (1\,6\,3)(2\,5).$$

(b) Podobne obliczenia pokazują, że

$$(2\,5\,3\,4\,6)(1\,6\,4)(2\,3) = (1\,2\,4)(3\,5).$$

Porównanie tego z wynikiem z punktu (a) pokazuje, że grupa S_6 nie jest przemienna.

(c) W rzeczywistości grupa S_n nie jest przemienna dla żadnego $n \geq 3$. Na przykład,

$$(1\,2)(1\,3) = (1\,3\,2), \quad \text{podczas gdy} \quad (1\,3)(1\,2) = (1\,2\,3). \qquad \blacksquare$$

Dla bardzo małych n pouczające jest napisanie tablicy mnożenia dla S_n.

PRZYKŁAD 3 Rysunek 12.4 przedstawia tablice mnożenia dla S_2 i S_3. Zrezygnowaliśmy z podawania tablicy dla S_4, która miałaby wymiar 24×24. Zaobserwuj, że każdy rząd i każda kolumna tablicy mnożenia dla S_3 zawiera każdy element zbioru S_3 dokładnie jeden raz. Ta własność tablicy znaczy, że mnożenie z prawej bądź lewej strony przez ustalony element zbioru S_3 jest przekształceniem wzajemnie jednoznacznym S_3 na S_3.

	e	(12)
e	e	(12)
(12)	(12)	e

S_2

	e	(123)	(132)	(23)	(12)	(13)
e	e	(123)	(132)	(23)	(12)	(13)
(123)	(123)	(132)	e	(12)	(13)	(23)
(132)	(132)	e	(123)	(13)	(23)	(12)
(23)	(23)	(13)	(12)	e	(132)	(123)
(12)	(12)	(23)	(13)	(123)	e	(132)
(13)	(13)	(12)	(23)	(132)	(123)	e

S_3

Rysunek 12.4

Ogólniej, jeśli h jest elementem S_n, to funkcje $g \to g \circ h$ i $g \to h \circ g$ są przekształceniami wzajemnie jednoznacznymi S_n na S_n. Fakt ten jest szczególnym przypadkiem twierdzenia 5 w § 12.5. ∎

Powodem, dla którego zbiór S_n nazywany jest często **grupą symetryczną** zbioru n-elementowego jest to, że spełnia on definicję pojęcia „grupa", którą podamy w § 12.5. W szczególności,

(i) jeśli $f, g \in S_n$, to $f \circ g \in S_n$,
(ii) jeśli $f \in S_n$, to $f^{-1} \in S_n$.

Interesować nas będą podzbiory zbioru S_n, spełniające powyższe warunki. Niepusty podzbiór G zbioru S_n będziemy nazywać **podgrupą grupy** S_n, o ile

(i) jeśli $f, g \in G$, to $f \circ g \in G$,
(ii) jeśli $f \in G$, to $f^{-1} \in G$.

W rzeczywistości warunku (ii) można było nie wypisywać: twierdzenie 2 w § 12.5 pokazuje, że dla podzbiorów G zbioru S_n z własności (i) wynika własność (ii). Własności (i) i (ii) implikują, że do zbioru G należy permutacja identycznościowa e:

(iii) $e \in G$.

Istotnie, jeśli f jest dowolnym elementem zbioru G, to z warunku (ii) wynika, że $f^{-1} \in G$, a więc na mocy warunku (i) otrzymujemy $e = f \circ f^{-1} \in G$.

PRZYKŁAD 4 (a) Jest rzeczą oczywistą, że zbiory S_n i $\{e\}$ są **podgrupami grupy** S_n. Podgrupę S_n nazywamy czasem **podgrupą niewłaściwą**. Podgrupę $\{e\}$ nazywamy **podgrupą trywialną**.

(b) Z rysunku 12.4 możemy odczytać wszystkie nietrywialne podgrupy właściwe grupy S_3. Są nimi: $G_1 = \{e, (1\,2)\}$, $G_2 = \{e, (1\,3)\}$, $G_3 = \{e, (2\,3)\}$ i $G_4 = \{e, (1\,2\,3), (1\,3\,2)\}$. Podgrupy również mają swoje tablice mnożenia. Tablice mnożenia dla podgrup G_1, G_2 i G_3 wyglądają tak, jak tablica mnożenia dla S_2. W istocie podgrupa G_1 to po prostu S_2. Żeby otrzymać z S_2 podgrupę G_2, zamieniamy 2 na 3, a by otrzymać podgrupę G_3 zamieniamy 1 na 3. Tablica mnożenia dla podgrupy G_4 przedstawiona jest na rysunku 12.5.

	e	$(1\,2\,3)$	$(1\,3\,2)$
e	e	$(1\,2\,3)$	$(1\,3\,2)$
$(1\,2\,3)$	$(1\,2\,3)$	$(1\,3\,2)$	e
$(1\,3\,2)$	$(1\,3\,2)$	e	$(1\,2\,3)$

G_4

	e	$(1\,2\,3\,4)$	$(1\,3)(2\,4)$	$(1\,4\,3\,2)$
e	e	$(1\,2\,3\,4)$	$(1\,3)(2\,4)$	$(1\,4\,3\,2)$
$(1\,2\,3\,4)$	$(1\,2\,3\,4)$	$(\mathbf{1\,3})(\mathbf{2\,4})$	$(1\,4\,3\,2)$	e
$(1\,3)(2\,4)$	$(1\,3)(2\,4)$	$(1\,4\,3\,2)$	e	$(1\,2\,3\,4)$
$(1\,4\,3\,2)$	$(1\,4\,3\,2)$	e	$(1\,2\,3\,4)$	$(\mathbf{1\,3})(\mathbf{2\,4})$

G_5

	e	$(1\,2)$	$(3\,4)$	$(1\,2)(3\,4)$
e	e	$(1\,2)$	$(3\,4)$	$(1\,2)(3\,4)$
$(1\,2)$	$(1\,2)$	e	$(1\,2)(3\,4)$	$(3\,4)$
$(3\,4)$	$(3\,4)$	$(1\,2)(3\,4)$	e	$(1\,2)$
$(1\,2)(3\,4)$	$(1\,2)(3\,4)$	$(3\,4)$	$(1\,2)$	e

G_6

Rysunek 12.5

(c) Rysunek 12.5 przedstawia również tablicę mnożenia dla G_5, najmniejszej spośród tych podgrup grupy S_4, do których należy permutacja $(1\,2\,3\,4)$. Zauważmy, że podgrupa G_5 jest przemienna: $f \circ g = g \circ f$ dla wszystkich $f, g \in G_5$. Wynika to z tego, że jej tablica jest symetryczna względem przekątnej (jej elementy są na rysunku wyróżnione pismem półgrubym).

(d) Tablica mnożenia dla G_6, najmniejszej spośród tych podgrup grupy S_4, do których należą permutacje $(1\,2)$ i $(3\,4)$, znajduje się również na rysunku 12.5. Podobnie jak G_5, jest to podgrupa przemienna mająca 4 elementy, ale te dwie podgrupy są

całkiem inne. Przykładowo, $g \circ g = e$ dla wszystkich $g \in G_6$ (zwróć uwagę na wyróżnioną przekątną tablicy podgrupy G_6), natomiast podgrupa G_5 własności takiej nie ma. Inna cecha odróżniająca od siebie te grupy omówiona jest w przykładzie 7.

(e) Zauważmy, że tablice z rysunku 12.5 mają własności związane z różnowartościowością mnożenia przez ustalony element, wspomniane w przykładzie 3: każdy element pojawia się dokładnie raz w każdym wierszu i każdej kolumnie. ∎

PRZYKŁAD 5 Rysunek 12.6 przedstawia tablicę mnożenia dla J, najmniejszej spośród tych podgrup grupy S_6, do których należą permutacje (1 6 4) i (2 3). Zauważmy, że tablica ta jest symetryczna względem wyróżnionej przekątnej, a więc podgrupa J jest przemienna. Ta tablica nie jest taka sama, jak tablica z rysunku 12.4, przedstawiająca nieprzemienną sześcioelementową grupę S_3. ∎

	e	(1 6 4)(2 3)	(1 4 6)	(2 3)	(1 6 4)	(1 4 6)(2 3)
e	e	(1 6 4)(2 3)	(1 4 6)	(2 3)	(1 6 4)	(1 4 6)(2 3)
(1 6 4)(2 3)	(1 6 4)(2 3)	**(1 4 6)**	(2 3)	(1 6 4)	(1 4 6)(2 3)	e
(1 4 6)	(1 4 6)	(2 3)	**(1 6 4)**	(1 4 6)(2 3)	e	(1 6 4)(2 3)
(2 3)	(2 3)	(1 6 4)	(1 4 6)(2 3)	e	(1 6 4)(2 3)	(1 4 6)
(1 6 4)	(1 6 4)	(1 4 6)(2 3)	e	(1 6 4)(2 3)	**(1 4 6)**	(2 3)
(1 4 6)(2 3)	(1 4 6)(2 3)	e	(1 6 4)(2 3)	(1 4 6)	(2 3)	**(1 6 4)**

J

Rysunek 12.6

W miarę posuwania się w głąb tego rozdziału, będziemy stosować bardziej abstrakcyjne podejście do grup. Jak pokazuje kolejny przykład, abstrakcyjne podejście może pewne sprawy lepiej wyjaśnić.

PRZYKŁAD 6 (a) Ponownie rozważmy podgrupę J z rysunku 12.6. Przez g oznaczmy permutację (1 6 4)(2 3), przez g^2 — permutację $g \circ g =$ (1 4 6), przez g^3 – permutację $g \circ g \circ g =$ (2 3) itd. Rysunek 12.7(a) przedstawia tablicę z rysunku 12.6, zapisaną za pomocą tych oznaczeń. Jest teraz jasne, że podgrupa J składa się z permutacji g i wszystkich jej potęg. Występujące w tablicy iloczyny elementów podgrupy uzyskuje się w oczywisty sposób, jeśli się tylko zauważy, że $g^6 = e$. Przez analogię do algebraicznej równości $x^0 = 1$ dla $x \neq 0$, przyjmujemy, że $g^0 = e$.

Jeśli w naszej tablicy mnożenia zastąpimy każdy wyraz g^k liczbą k (i $e = g^0$ zerem), to otrzymamy tablicę mnożenia przedstawioną na rysunku 12.7(b). To jest po prostu tablica dodawania dla $\mathbb{Z}_6$, która znajduje się też w § 3.6. Chociaż $\mathbb{Z}_6$ nie jest zbiorem

	e	g	g^2	g^3	g^4	g^5
e	e	g	g^2	g^3	g^4	g^5
g	g	g^2	g^3	g^4	g^5	e
g^2	g^2	g^3	g^4	g^5	e	g
g^3	g^3	g^4	g^5	e	g	g^2
g^4	g^4	g^5	e	g	g^2	g^3
g^5	g^5	e	g	g^2	g^3	g^4

(a) J

$+_6$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

(b) $\mathbb{Z}_6$

Rysunek 12.7

permutacji, okazuje się, że $\mathbb{Z}_6$ wraz z dodawaniem tworzy grupę i że J i $\mathbb{Z}_6$ to w gruncie rzeczy te same grupy; ich tablice dla, odpowiednio, mnożenia i dodawania są ze sobą zgodne.

(b) Tablicom z rysunku 12.5 można także nadać bardziej abstrakcyjną postać; zob. rysunek 12.8. ∎

	e	g	g^2
e	e	g	g^2
g	g	g^2	e
g^2	g^2	e	g

G_4

	e	g	g^2	g^3
e	e	g	g^2	g^3
g	g	g^2	g^3	e
g^2	g^2	g^3	e	g
g^3	g^3	e	g	g^2

G_5

	e	g	h	gh
e	e	g	h	gh
g	g	e	gh	h
h	h	gh	e	g
gh	gh	h	g	e

G_6

Rysunek 12.8

Podgrupa J jest szczególnym rodzajem podgrupy, ponieważ składa się z pojedynczej permutacji i wszystkich jej potęg. Dla dowolnej permutacji g w dowolnej grupie S_n, przez $\langle g\rangle$ oznaczać będziemy podgrupę $\{e, g, g^2, \ldots\}$; $\langle g\rangle$ nazywać będziemy **grupą generowaną** przez g. Zatem $J = \{e, g, g^2, g^3, g^4, g^5\}$ jest grupą generowaną w S_6 przez $g = (1\,6\,4)(2\,3)$. Dla każdego $g \in S_n$ istnieje najmniejsza dodatnia liczba całkowita m taka, że $g^m = e$ (sceptycy mogą zajrzeć do dowodu twierdzenia 2 w § 12.5). Liczbę m nazywamy **rzędem elementu** g; mamy $\langle g\rangle = \{e, g, g^2, \ldots, g^{m-1}\}$. Zauważmy, że rząd cyklu $(k_1 k_2 \ldots k_m)$ to po prostu jego długość m.

PRZYKŁAD 7 (a) Podgrupa G_4 z rysunku 12.5 jest generowana przez permutację $g = (1\,2\,3)$, a podgrupa G_5 jest generowana przez permutację $g = (1\,2\,3\,4)$; zob. rysunek 12.8.

(b) Podgrupa G_6 nie jest postaci $\langle g\rangle$, tzn. grupa G_6 nie jest generowana przez żaden pojedynczy element. Wynika to stąd, że dla wszystkich $g \in G_6$ mamy $g^2 = e$, tak więc każda podgrupa

$\langle g \rangle$ ma dwa elementy lub tylko jeden, gdy $g = e$. Jednakże grupa G_6 jest generowana przez dwie permutacje: $g = (1\,2)$ i $h = (3\,4)$; zob. rysunek 12.8. ∎

Jeśli permutację g zapiszemy w postaci iloczynu rozłącznych cykli, to łatwo jest określić jej rząd, a co za tym idzie, liczbę elementów grupy $\langle g \rangle$. Zauważyliśmy już, że cykle długości m mają rząd m. Jeśli permutacja g jest iloczynem $c_1 c_2 \ldots c_k$ rozłącznych cykli, to korzystając z przemienności, wynikającej z rozłączności cykli, widzimy, że

$$g^j = (c_1 c_2 \ldots c_k)^j = c_1^j c_2^j \ldots c_k^j \quad \text{dla } j \geq 1.$$

Jeśli potęga j jest wielokrotnością każdego z rzędów m_i cyklu c_i, to dla każdego i, $c_i^j = e$, a więc $g^j = e$. Najmniejsza taka liczba j jest najmniejszą wspólną wielokrotnością rzędów wszystkich cykli. Rozumowanie to sugeruje, że prawdziwe jest następujące twierdzenie.

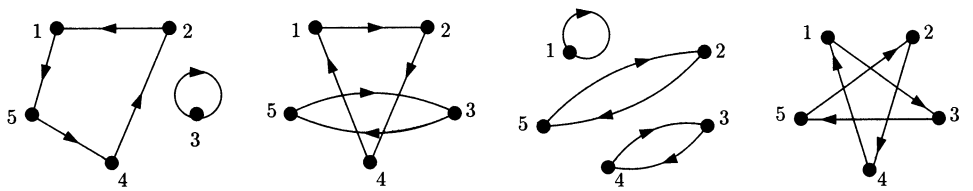
Twierdzenie 2

Jeśli permutacja jest przedstawiona w postaci iloczynu rozłączych cykli, to jej rząd jest najmniejszą wspólną wielokrotnością rzędów (tzn. długości) wszystkich cykli.

Rozważania poprzedzające sformułowanie twierdzenia pokazują, że $g^j = e$ dla $j = \text{NWW}(m_1, m_2, \ldots, m_k)$. Dla zakończenia dowodu musielibyśmy pokazać, że jeśli $g^j = e$ i $j > 0$, to $j \geq \text{NWW}(m_1, m_2, \ldots, m_k)$. Szkic dowodu podany jest w ćwiczeniu 19.

ĆWICZENIA DO § 12.1

1. Każdą z permutacji przedstawionych na rysunku 12.9 zapisz w postaci cyklu lub iloczynu rozłącznych cykli.



Rysunek 12.9

2. Każdą z permutacji odwrotnych do permutacji przedstawionych na rysunku 12.9 zapisz w postaci cyklu lub iloczynu cykli rozłącznych.

3. Zrób rysunki następujących permutacji należących do S_4:
(a) (1 2 3 4), (b) (1 4 2),
(c) (1 2)(3 4), (d) (2 3).

4. Zrób rysunki następujących permutacji należących do S_6:
(a) (1 4)(3 6 5), (b) (1 5)(2 6)(3 4),
(c) (1 3)(2 5 4 6), (d) (1 2 5 3 4).

5. Zapisz każdą z następujących permutacji w postaci cyklu bądź iloczynu rozłącznych cykli, tzn. wykonaj mnożenia tak, jak w przykładzie 2:
(a) (1 2)(1 4), (b) (1 2)(1 4 3),
(c) (1 3)(2 3)(3 4), (d) (1 3 2)(1 2 4),
(e) (1 2 3)(1 2 4), (f) (1 3)(2 3 4)(1 4)(3 4).

6. Powtórz ćwiczenie 5 dla następujących iloczynów:
(a) (1 3 6 4)(1 4 2), (b) (2 5 4)(1 4)(1 3),
(c) (1 5 3 4 6)(2 5)(1 3 5), (d) (1 2 3 4 5)(1 2 3 4 5),
(e) (1 4 2)(1 4 2)(1 4 2), (f) (2 6 5 4 3)(4 6).

7. Znajdź permutacje odwrotne do następujących cykli:
(a) (1 2), (b) (1 4 2 6),
(c) (1 3 6), (d) (1 6 2 5 4 3).

8. Każdą z permutacji odwrotnych do permutacji z ćwiczenia 5 zapisz w postaci cyklu lub iloczynu rozłącznych cykli. *Wskazówka*: wykorzystaj swoje odpowiedzi z ćwiczenia 5.

9. Cykl, który zamienia miejscami dwa elementy, taki jak (1 3), nazywany jest **transpozycją**.
(a) Wykaż, że (2 3 4 6 5) = (2 5)(2 6)(2 4)(2 3).
(b) Przedstaw permutację (1 7 2 5 3 4) w postaci iloczynu transpozycji.
(c) Sformułuj uogólnienie części (a) i (b) dotyczące dowolnego cyklu $(k_1 k_2 \ldots k_m)$.
(d) Wyjaśnij, dlaczego dowolną permutację można przedstawić w postaci iloczynu transpozycji.

10. Udowodnij przez indukcję względem m stwierdzenie, które należało sformułować w ćwiczeniu 9(c).

11. Rozważmy podgrupę J z rysunku 12.7. Ile spośród jej elementów ma rząd równy
(a) 1? (b) 2? (c) 3? (d) 6?

12. Powtórz ćwiczenie 11 dla grupy S_3.

13. Wypisz elementy podgrupy grupy S_6 generowanej przez każdą z następujących permutacji:
(a) (2 5), (b) (1 5 2 4),
(c) (1 6)(2 4 3), (d) (1 3 5 2)(4 6).

14. Znajdź rzędy następujących permutacji:
(a) (1 6)(3 6), (b) (1 3 5)(2 5 1), (c) (1 3)(2 6 3).

15. (a) Dla każdej z wartości $m = 1, 2, 3, 4, 5, 6$ znajdź permutację należącą do S_6 mającą rząd m. Unikaj, jeśli to będzie możliwe, cykli.
(b) Wykaż, że każda permutacja należąca do S_6 ma rząd 1, 2, 3, 4, 5 lub 6. *Wskazówka*: wykorzystaj twierdzenie 2. *Ostrzeżenie*: w ogólnym przypadku grupa S_n może mieć elementy rzędu większego niż n. Zobacz następne ćwiczenie.

16. (a) Podaj przykład permutacji należącej do S_5 i mającej rząd 6.
(b) Podaj przykłady permutacji należących do S_7 i mających rzędy 10 i 12.

17. Czy każda podgrupa grupy S_n, mająca postać $\langle g \rangle$, jest przemienna? Odpowiedź uzasadnij.

18. Niech $g = (1\,2\,3)$ i $h = (2\,3)$ będą elementami grupy S_3.
(a) Wykaż, że $S_3 = \{e, g, g^2, h, gh, hg\}$.
(b) Czy permutacje ghg i hgh należą do S_3?

19. (a) Wykaż, że jeśli element g grupy S_n ma rząd m, to $g^j = e$ wtedy i tylko wtedy, gdy j jest wielokrotnością m. *Wskazówka*: algorytm dzielenia może pomóc.
(b) Wykaż, że jeśli $c_1, c_2, \ldots, c_k$ są cyklami rozłącznymi i $c_1^j c_2^j \ldots c_k^j = e$, to $c_i^j = e$ dla $i = 1, 2, \ldots, k$.
(c) Wykorzystaj części (a) i (b), aby zakończyć dowód twierdzenia 2.

20. Podaj tablice mnożenia dla podgrup z ćwiczeń 13(a) i (b).

§ 12.2. Działania grup na zbiorach

Rozważmy grupę G, której elementy są permutacjami pewnego zbioru X, tzn. podgrupę grupy PERM (X). Pojęcia i fakty z § 12.1 dotyczą również grupy G, mimo że X jest tu dowolnym skończonym zbiorem zamiast $\{1, 2, \ldots, n\}$. Czasami mówi się, że G **działa na zbiorze** X. Określenie to stosuje się zwłaszcza wtedy, gdy interesuje nas przede wszystkim, co permutacje z grupy G robią z poszczególnymi elementami bądź podzbiorami zbioru X.

Jeśli przedstawimy permutację f zbioru X w postaci iloczynu rozłącznych cykli, to otrzymamy podział zbioru X na rozłączne podzbiory, po jednym dla każdego cyklu, uwzględniając cykle długości 1. Przykładowo, permutacja $f = (1\,7\,4\,2)(6\,8) = (1\,7\,4\,2)(3)(5)(6\,8)$ z S_8 daje podział zbioru $\{1, 2, 3, 4, 5, 6, 7, 8\}$ na bloki $\{\{1, 2, 4, 7\}, \{3\}, \{5\}, \{6, 8\}\}$. Każdy blok takiego podziału składa się z elementów, które permutacja f permutuje cyklicznie

między sobą i dowolny element danego bloku może być przeprowadzony na jakikolwiek inny jego element przez odpowiednią potęgę permutacji f. W naszym przykładzie f przeprowadza 7 na 4, permutacja f^2 przeprowadza 7 na 2, permutacja f^3 przeprowadza 7 na 1, a permutacja $e = f^0$ przeprowadza 7 na 7.

Bloki dowolnego podziału można zawsze traktować jak klasy abstrakcji pewnej relacji równoważności. Permutacja f pozwala zdefiniować relację równoważności $\sim$ w zbiorze X w następujący sposób:

$$x \sim y, \text{ gdy } y = f^j(x) \text{ dla pewnej permutacji } f^j \text{ z grupy} \langle f \rangle.$$

Definicję tę można rozszerzyć z grupy $\langle f \rangle$ na dowolną grupę G, niekoniecznie generowaną przez pojedynczy element.

Stwierdzenie 1

Niech G będzie grupą działającą na zbiorze X. Zdefiniujmy relację $\sim$ mówiąc, że dla $x, y \in X$, $x \sim y$, jeśli $y = g(x)$ dla pewnej permutacji $g \in G$. Wówczas $\sim$ jest relacją równoważności w zbiorze X. Dla dowolnego $x \in X$, klasą abstrakcji, do której należy element x, jest zbiór

$$Gx = \{g(x) \colon g \in G\}.$$

W przypadku ogólnych relacji równoważności, klasy abstrakcji oznaczaliśmy zwykle przez $[x]$, ale bardziej naturalnym oznaczeniem jest tutaj Gx.

Dowód. Wystarczy sprawdzić, że relacja $\sim$ jest zwrotna, symetryczna i przechodnia.

(Z) Niech x będzie elementem zbioru X. Ponieważ $e \in G$ i $x = e(x)$, więc mamy $x \sim x$.

(S) Załóżmy, że $x \sim y$, gdzie $x, y \in X$. Znaczy to, że $y = g(x)$ dla pewnej permutacji $g \in G$. Ponieważ $g^{-1} \in G$ oraz $x = g^{-1}(y)$, to mamy $y \sim x$.

(P) Załóżmy, że $x \sim y$ i $y \sim z$. Wtedy $y = g(x)$ i $z = f(y)$ dla pewnych permutacji g i f z grupy G. Ponieważ $f \circ g \in G$ i $z = f \circ g(x)$, to wnosimy stąd, że $x \sim z$. ∎

Te szczególne klasy abstrakcji, Gx, nazywane są **orbitami działania grupy** G na zbiorze X lub G-**orbitami**. Każda relacja równoważności w zbiorze X określa jego podział na klasy abstrakcji (twierdzenie 1 w § 3.5), a więc otrzymujemy następujący:

Wniosek

Jeśli grupa G działa na zbiorze X, to G-orbity tworzą podział zbioru X.

PRZYKŁAD 1

Niech $g \in S_6$, $g = (1\,6\,4)(2\,3)$. Wówczas $\langle g \rangle = \{e, g, g^2, g^3, g^4, g^5\}$, a więc jedną z orbit jest $\langle g \rangle 1 = \{e(1), g(1), g^2(1), g^3(1), g^4(1), g^5(1)\} = \{1, 6, 4, 1, 6, 4\} = \{1, 4, 6\}$. Podobnie, $\langle g \rangle 2 = \{2, 3\}$ i $\langle g \rangle 5 = \{5\}$, tak jak można się było spodziewać. Krótko mówiąc, $\langle g \rangle$-orbity mogą być odczytane bez żadnych obliczeń, na podstawie przedstawienia permutacji g w postaci iloczynu rozłącznych cykli: $g = (1\,6\,4)(2\,3)$. Tego rodzaju rozumowania użyć można do dowodu następnego wyniku. ∎

Stwierdzenie 2

Jeśli permutacja g jest iloczynem rozłącznych cykli, to zbiory wyrazów tych cykli, uwzględniając cykle długości 1, tworzą rodzinę wszystkich $\langle g \rangle$-orbit.

Często interesować nas będą G-orbity wtedy, gdy grupa G nie jest generowana przez pojedynczą permutację. W dalszej części tego paragrafu będziemy zakładać, że nasze grupy są skończone i działają na zbiorach skończonych.

PRZYKŁAD 2

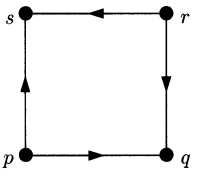
(a) Jeśli $G = \text{PERM}(X)$, to każdy element zbioru X jest równoważny dowolnemu innemu elementowi zbioru X, gdyż dla danych elementów x i y z X znajdzie się permutacja przeprowadzająca x na y. Wynika stąd, że istnieje tylko jedna, duża orbita, a mianowicie sam zbiór X. Ta grupa G jest zbyt duża, by mieć interesujące orbity.

(b) Niech G_6 będzie grupą przedstawioną na rysunku 12.5 z § 12.1. Łatwo jest sprawdzić, że orbitami są zbiory $\{1, 2\}$ i $\{3, 4\}$. Zauważmy, że nie możemy tu zastosować stwierdzenia 2, ponieważ grupa G_6 nie jest generowana przez pojedynczą permutację. ∎

G-orbity interesować nas będą najbardziej, gdy G jest grupą permutacji opisującą symetrie jakiegoś obiektu, takiego jak graf bądź graf skierowany. Weźmy graf skierowany D bez krawędzi wielokrotnych. Permutacja g zbioru $V(D)$ wszystkich wierzchołków jest **automorfizmem** naszego **grafu skierowanego**, jeśli zachowuje strukturę krawędzi, tzn. jeśli dla dowolnych $x, y \in V(G)$ zachodzi równoważność: (x, y) jest krawędzią wtedy i tylko wtedy, gdy $(g(x), g(y))$ jest krawędzią. O automorfizmie grafu skierowanego możemy myśleć jak o permutacji, która zmienia

etykiety wierzchołków, nie zmieniając zasadniczo struktury naszego grafu. Zbiór wszystkich automorfizmów grafu skierowanego D będziemy oznaczać przez $\mathrm{AUT}(D)$. Łatwo jest sprawdzić, że $\mathrm{AUT}(D)$ jest grupą permutacji działającą na zbiorze $V(D)$ (ćwiczenie 1). Mówiąc bardzo nieściśle, im większa jest grupa $\mathrm{AUT}(D)$, tym więcej symetrii ma graf D.

PRZYKŁAD 3 (a) Graf skierowany D przedstawiony na rysunku 12.10(a) ma dwa źródła i dwa ujścia i każdy jego automorfizm musi przeprowadzać źródła na źródła, a ujścia na ujścia. Można sprawdzić, że istnieją cztery automorfizmy, a mianowicie e, $f = (q\,s)$, $g = (p\,r)$ i $f\,g = (p\,r)(q\,s)$. Dla wygody, wartości przyjmowane przez każdy z nich wypisane są na rysunku 12.10(b). Zauważ na przykład, że automorfizm g zamienia ze sobą etykiety p i r, ale moglibyśmy osiągnąć to samo, odwracając po prostu ten graf na drugą stronę, bez zmiany jego struktury. Tablica mnożenia dla naszej grupy automorfizmów znajduje się na rysunku 12.10(c). Jest to w gruncie rzeczy ta sama grupa, co grupa G_6 z rysunku 12.5 w § 12.1.



D

(a)

	p	q	r	s	Cykle
e	p	q	r	s	e
f	p	s	r	q	(qs)
g	r	q	p	s	(pr)
fg	r	s	p	q	(pr)(qs)

AUT(D)

(b)

	e	f	g	fg
e	e	f	g	fg
f	f	e	fg	g
g	g	fg	e	f
fg	fg	g	f	e

AUT(D)

(c)

Rysunek 12.10

(b) Aby w jakiś systematyczny sposób znaleźć orbity działania grupy $\mathrm{AUT}(D)$, zaczynamy od wierzchołka p i otrzymujemy

$$\mathrm{AUT}(D)p = \{e(p), f(p), g(p), fg(p)\} = \{p, p, r, r\} = \{p, r\};$$

zob. pierwszą kolumnę tablicy z rysunku 12.10(b). Bierzemy teraz jeden z wierzchołków, które dotąd się nie pojawiły i powtarzamy powyższy proces. Korzystając z drugiej kolumny tablicy z rysunku 12.10(b), otrzymujemy

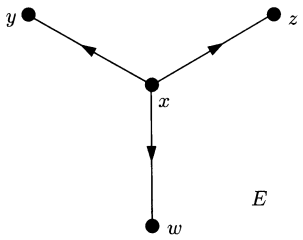
$$\mathrm{AUT}(D)q = \{q, s, q, s\} = \{q, s\}.$$

Wnioskujemy stąd, że istnieją dwie orbity, $\{p, r\}$ i $\{q, s\}$. ∎

PRZYKŁAD 4 (a) Graf skierowany E z rysunku 12.11(a) ma inny rodzaj symetrii i inną grupę automorfizmów. Każdy automorfizm grafu E

musi przeprowadzać x na x, ale pozostałe wierzchołki mogą być ze sobą w dowolny sposób permutowane. A więc grupa $\text{AUT}(E)$ to w gruncie rzeczy grupa $\text{PERM}(\{y, z, w\})$ lub grupa S_3. Na rysunku 12.11(b) wypisane są wartości przyjmowane przez poszczególne automorfizmy; $g = (y\,z\,w)$ i $h = (z\,w)$.

(b) Orbity działania grupy $\text{AUT}(E)$ mogą być odczytane z tablicy z rysunku 12.11(b) lub z rysunku 12.11(a). Są nimi zbiory $\{x\}$ i $\{y, z, w\}$. ∎



(a)

	x	y	z	w	Cykle
e	x	y	z	w	e
g	x	z	w	y	$(y\,z\,w)$
g^2	x	w	y	z	$(y\,w\,z)$
h	x	y	w	z	$(z\,w)$
gh	x	z	y	w	$(y\,z)$
hg	x	w	z	y	$(y\,w)$

$\text{AUT}(E)$

(b)

	e	g	g^2	h	gh	hg
e	e	g	g^2	h	gh	hg
g	g	g^2	e	gh	hg	h
g^2	g^2	e	g	hg	h	gh
h	h	hg	gh	e	g^2	g
gh	gh	h	hg	g	e	g^2
hg	hg	gh	h	g^2	g	e

$\text{AUT}(E)$

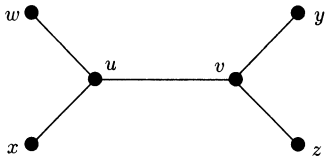
(c)

Rysunek 12.11

W ten sam sposób można badać grupy automorfizmów grafów nieskierowanych. Permutacja g zbioru $V(H)$ wszystkich wierzchołków grafu nieskierowanego H bez krawędzi wielokrotnych jest **automorfizmem tego grafu**, jeśli dla dowolnych jego wierzchołków x i y, $\{x, y\}$ jest krawędzią wtedy i tylko wtedy, gdy $\{g(x), g(y)\}$ jest krawędzią. Grupa wszystkich takich automorfizmów oznaczana jest przez $\text{AUT}(H)$.

PRZYKŁAD 5 (a) Symetrie grafu H, przedstawionego na rysunku 12.12(a), opisane są przez jego grupę automorfizmów $\text{AUT}(H)$, której elementy wypisane są na rysunku 12.12(b).

(b) Orbitami działania grupy $\text{AUT}(H)$ są zbiory $\{u, v\}$ i $\{w, x, y, z\}$. Zobacz rysunek 12.12(b); orbity te można by również znaleźć patrząc na rysunek 12.12(a) i wyobrażając sobie automorfizmy naszego grafu. Zauważmy, że niektóre automorfizmy nie są wynikiem obracania lub odwracania grafu w całości.



H

(a)

	u	v	w	x	y	z	Cykle
e	u	v	w	x	y	z	e
g	u	v	x	w	y	z	$(w\,x)$
h	u	v	w	x	z	y	$(y\,z)$
gh	u	v	x	w	z	y	$(w\,x)(y\,z)$
f	v	u	y	z	w	x	$(u\,v)(w\,y)(x\,z)$
fh	v	u	y	z	x	w	$(u\,v)(w\,y\,x\,z)$
fg	v	u	z	y	w	x	$(u\,v)(w\,z\,x\,y)$
fgh	v	u	z	y	x	w	$(u\,v)(w\,z)(x\,y)$

$\mathrm{AUT}(H)$

(b)

Rysunek 12.12

(c) W tablicy 12.1 przedstawione są orbity tych podgrup grupy $\mathrm{AUT}(H)$, które generowane są przez pojedyncze permutacje. ∎

Tablica 12.1

Podgrupa	Orbity	Liczba orbit
$\langle e\rangle$	$\{u\}$, $\{v\}$, $\{w\}$, $\{x\}$, $\{y\}$, $\{z\}$	6
$\langle g\rangle$	$\{u\}$, $\{v\}$, $\{w,x\}$, $\{y\}$, $\{z\}$	5
$\langle h\rangle$	$\{u\}$, $\{v\}$, $\{w\}$, $\{x\}$, $\{y,z\}$	5
$\langle gh\rangle$	$\{u\}$, $\{v\}$, $\{w,x\}$, $\{y,z\}$	4
$\langle f\rangle$	$\{u,v\}$, $\{w,y\}$, $\{x,z\}$	3
$\langle fh\rangle$	$\{u,v\}$, $\{w,x,y,z\}$	2
$\langle fg\rangle$	$\{u,v\}$, $\{w,x,y,z\}$	2
$\langle fgh\rangle$	$\{u,v\}$, $\{w,z\}$, $\{x,y\}$	3

G-orbita dowolnego elementu x zbioru X ma bliski związek z pewną podgrupą grupy G, a mianowicie z podgrupą $\mathrm{FIX}_G(x)$, złożoną a tych permutacji należących do G, które nie ruszają punktu x. To znaczy

$$\mathrm{FIX}_G(x) = \{g \in G\colon g(x) = x\}.$$

Przeprowadzenie dowodu tego, że $\mathrm{FIX}_G(x)$ jest naprawdę podgrupą grupy G, pozostawiamy jako ćwiczenie 10. Gdy grupa G jest znana z kontekstu, to możemy zamiast $\mathrm{FIX}_G(x)$ pisać $\mathrm{FIX}(x)$.

PRZYKŁAD 6 (a) W grupie $\mathrm{AUT}(D)$ automorfizmów grafu skierowanego D z rysunku 12.10 mamy $\mathrm{FIX}(p) = \{e, f\}$, $\mathrm{FIX}(q) = \{e, g\}$, $\mathrm{FIX}(r) = \{e, f\}$ i $\mathrm{FIX}(s) = \{e, g\}$. Zbiory te można odczytać z tablicy przedstawionej na rysunku 12.10(b), znajdując wyrazy równe p w pierwszej kolumnie, wyrazy równe q w drugiej kolumnie itd.

(b) Dla grupy $\text{AUT}(E)$ z rysunku 12.11 mamy $\text{FIX}(x) = \text{AUT}(E)$, $\text{FIX}(y) = \{e, h\}$, $\text{FIX}(z) = \{e, hg\}$ i $\text{FIX}(w) = \{e, gh\}$.

(c) Dla grupy $\text{AUT}(H)$ z rysunku 12.12 otrzymujemy $\text{FIX}(u) = \text{FIX}(v) = \{e, g, h, gh\}$, $\text{FIX}(w) = \text{FIX}(x) = \{e, h\}$ i $\text{FIX}(y) = \text{FIX}(z) = \{e, g\}$. Zauważmy, że podgrupa $\text{FIX}(u)$ ma 4 elementy, a orbita $\text{AUT}(H)u = \{u, v\}$ ma 2 elementy, podczas gdy podgrupa $\text{FIX}(w)$ ma 4 elementy, a orbita $\text{AUT}(H)w = \{w, x, y, z\}$ ma 4 elementy. Zauważmy także, że $2 \cdot 4 = 4 \cdot 2 = 8 = |\text{AUT}(H)|$. Następne twierdzenie pokazuje, że nie jest to przypadek. ∎

Twierdzenie

Niech G będzie grupą działającą na zbiorze X i niech $x \in X$. Liczba permutacji, z których składa się grupa G, jest iloczynem liczby elementów orbity Gx przez liczbę permutacji należących do podgrupy $\text{FIX}(x)$. To znaczy,

$$|G| = |Gx| \cdot |\text{FIX}_G(x)|.$$

W szczególności, $|Gx|$ dzieli $|G|$.

Dowód. Każdy element orbity Gx jest postaci $g(x)$ dla pewnej permutacji $g \in G$. Istnieją zatem w grupie G permutacje $g_1, g_2 \ldots, g_k$ takie, że $Gx = \{g_1(x), \ldots, g_k(x)\}$, przy czym wypisane elementy nie powtarzają się, a $k = |Gx|$. Twierdzimy, że

$$(1) \qquad G = \bigcup_{j=1}^{k} (g_j \circ \text{FIX}(x)),$$

gdzie $g_j \circ \text{FIX}(x)$ oznacza zbiór $\{g_j \circ g\colon g \in \text{FIX}(x)\}$. Twierdzimy również, że

(2) zbiory $g_j \circ \text{FIX}(x)$ są parami rozłączne

oraz

(3) każdy zbiór $g_j \circ \text{FIX}(x)$ ma tyle samo elementów co zbiór $\text{FIX}(x)$.

Korzystając ze stwierdzeń (1)-(3), łatwo jest zakończyć dowód:

$$|G| = \sum_{j=1}^{k} |g_j \circ \text{FIX}(x)| = \qquad \text{na mocy (1) i (2)}$$

$$= \sum_{j=1}^{k} |\text{FIX}(x)| = \qquad \text{na mocy (3)}$$

$$= k \cdot |\text{FIX}(x)| = |Gx| \cdot |\text{FIX}(x)|.$$

Dowody zdań (1)-(3) są nietrudnymi ćwiczeniami (ćwiczenie 14). Ponadto stwierdzenie (3) wynika z ogólnej własności podgrup w grupach, którą pokażemy w twierdzeniu 5 w § 12.5.

ĆWICZENIA DO § 12.2

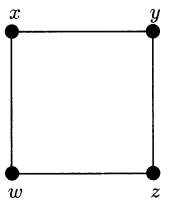
1. Niech D będzie grafem skierowanym.
(a) Wykaż, że $\text{AUT}(D)$ jest podgrupą grupy $\text{PERM}(V(D))$.
(b) Automorfizmy grafu skierowanego zachowują jego własności. Przykładowo, wykaż, że automorfizmy przeprowadzają źródła na źródła i ujścia na ujścia.

2. (a) Znajdź *kilka* wyrazów tablicy mnożenia grupy z przykładu 5 i rysunku 12.12.
(b) Czy grupa ta jest przemienna?

3. Weźmy grupę $\text{AUT}(D)$ z rysunku 12.10. Jej orbity podane są w przykładzie 3(b), a podgrupy postaci $\text{FIX}(x)$ — w przykładzie 6(a). Sprawdź, że w tych przypadkach zachodzą równości, wynikające z twierdzenia udowodnionego w tym paragrafie.

4. Powtórz ćwiczenie 3 dla grupy $\text{AUT}(E)$ z rysunku 12.11. Orbity podane są w przykładzie 4(b), a podgrupy postaci $\text{FIX}(x)$ — w przykładzie 6(b).

5. Rozważmy graf z rysunku 12.13(a).



(a)

	x	y	z	w
e	x	y	z	w
r	y	z	w	x
r^2	z	w	x	y
r^3	w	x	y	z
h	w	z	y	x
v	y	x	w	z
d	x	w	z	y
f	z	y	x	w

(b)

Rysunek 12.13

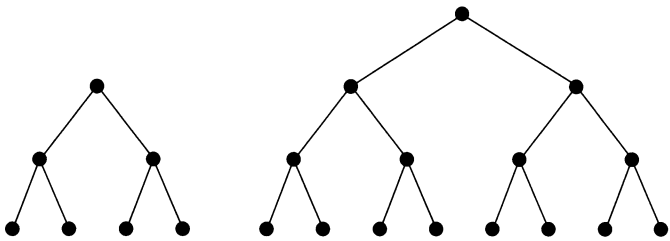
(a) Przekonaj się, że grupa automorfizmów tego grafu składa się z ośmiu permutacji wypisanych na rysunku 12.13(b). Zauważ, że r jest obrotem, h — symetrią względem poziomej osi, v — symetrią względem pionowej osi, a d i f — symetriami względem przekątnych.
(b) Znajdź orbity wyznaczone przez każdą z ośmiu podgrup generowanych przez pojedyncze permutacje.

(c) Zapisz permutacje należące do tej grupy automorfizmów w postaci iloczynów rozłącznych cykli.

6. Co możesz powiedzieć o rozmiarach orbit wyznaczonych przez działanie grupy mającej 27 elementów?

7. Załóżmy, że grupa G działa na zbiorze X. Wykaż, że jeśli dla pewnego k, $|G| = 2^k$ oraz $|X|$ jest liczbą nieparzystą, to pewien element zbioru X musi być punktem stałym dla wszystkich permutacji z grupy G. *Wskazówka*: ponieważ orbity działania grupy G tworzą podział zbioru X, to możemy wybrać po jednym elemencie z każdej orbity; niech $x_1, \ldots, x_m$ będą wybranymi elementami. Wtedy $|X| = \sum_{j=1}^{m} |Gx_j|$. Zastosuj twierdzenie z tego paragrafu do każdego elementu x_j.

8. (a) Ile automorfizmów ma drzewo binarne T_1 z rysunku 12.14?



Rysunek 12.14

(b) Powtórz część (a) dla drzewa T_2 z rysunku 12.14.

(c) Wykorzystaj ćwiczenie 7 dla pokazania, że pewien wierzchołek jest przeprowadzany sam na siebie przez każdy automorfizm drzewa T_2.

(d) Znajdź wierzchołek o własności opisanej w części (c).

9. (a) Pokaż, jak skonstruować, dla każdego $n \in \mathbb{P}$, takie drzewo T, dla którego $|\operatorname{AUT}(T)| = 2^n$. *Wskazówka*: dołączaj kolejno odpowiednie grafy, mające po dwa automorfizmy.

(b) Pokaż, jak skonstruować, dla każdego $n \in \mathbb{P}$, taki graf skierowany D, dla którego $|\operatorname{AUT}(D)| = n$.

10. Niech G będzie grupą działającą na zbiorze X.

(a) Udowodnij bezpośrednio, że każdy zbiór postaci $\operatorname{FIX}(x)$ jest podgrupą grupy G lub zrób ćwiczenie (b) i zauważ, że mamy tu do czynienia z jego szczególnym przypadkiem.

(b) Dla podzbioru Y zbioru X niech $\operatorname{FIX}(Y) = \{g \in G\colon g(Y) = Y\}$. Wykaż, że $\operatorname{FIX}(Y)$ jest podgrupą grupy G.

11. Niech $G = \operatorname{AUT}(H)$ będzie grupą z rysunku 12.12. Dla każdego podzbioru Y zbioru H niech $\operatorname{FIX}(Y)$ oznacza podgrupę zdefiniowaną w ćwiczeniu 10(b). Znajdź $\operatorname{FIX}(\{w, y\})$. Czy $\operatorname{FIX}(\{w, y\}) = \operatorname{FIX}(w) \cap \operatorname{FIX}(y)$?

12. Niech grupa G działa na zbiorze X i dla $Y \subseteq X$, niech $\operatorname{FIX}(Y)$ będzie podgrupą zdefiniowaną w ćwiczeniu 10(b).
(a) Wykaż, że $\bigcap_{y \in Y} \operatorname{FIX}(y) \subseteq \operatorname{FIX}(Y)$.
(b) Zauważ (ćwiczenie 11), że w części (a) równość nie musi zachodzić.

13. Rozważmy grupę $\operatorname{AUT}(H)$ z przykładu 6(c) i rysunku 12.12.
(a) Zgodnie z twierdzeniem z tego paragrafu,

$$|\operatorname{AUT}(H)| = |\operatorname{AUT}(H)u| \cdot |\operatorname{FIX}(u)|,$$

tzn. $8 = 2 \cdot 4$. Zilustruj dowód wspomnianego twierdzenia, używając odpowiednich elementów g_1 i g_2 zbioru $\operatorname{AUT}(H)$ takich, że $\operatorname{AUT}(H)u = \{g_1(u), g_2(u)\}$.
(b) Czy elementy g_1 i g_2 można wybrać w jeden tylko sposób?
(c) Biorąc x zamiast u, zilustruj dowód twierdzenia podobnie, jak w części (a).

14. Udowodnij zdania (1)-(3) z dowodu jedynego twierdzenia tego paragrafu.

15. Mówimy, że grupa G działająca na zbiorze X działa na X **przechodnio**, jeśli $X = Gx$ dla pewnego elementu x zbioru X. Załóżmy, że G działa przechodnio na X.
(a) Wykaż, ze $X = Gx$ dla każdego elementu x ze zbioru X.
(b) Wykaż, że jeśli grupa G jest skończona, to zbiór X także jest skończony oraz

$$|G| = |\operatorname{FIX}(x)| \cdot |X| \text{ dla każdego } x \in X.$$

16. Załóżmy, że grupa G działa na zbiorze X i K jest podgrupą grupy G. Wówczas podgrupa K również działa na zbiorze X. Wykaż, że każda orbita działania grupy G w X jest sumą pewnych orbit działania grupy K.

17. Niech G będzie grupą działającą na zbiorze X oraz niech

$$R = \{(x, y) \in X \times X \colon g(x) = y \text{ dla pewnego } g \in G\}.$$

(a) Wykaż, że R jest relacją równoważności w zbiorze X.
(b) Opisz podział zbioru X odpowiadający relacji R.

§ 12.3. Działania grup na zbiorach, część 2

Kontynuujemy nasze przygotowania do tego, by w następnym paragrafie móc pokazać interesujące zastosowania omawianych zagadnień do problemów zliczania. Rozpoczniemy od przedstawienia w dość abstrakcyjnej formie pewnego twierdzenia dotyczącego zliczania. Później zajmiemy się jego szczególnymi przypadkami, które łatwo będzie stosować. Będziemy też zmuszeni uogólnić pojęcie grupy działającej na zbiorze.

Rozważmy ponownie grupę G działającą na zbiorze X. W paragrafie 12.2 wyróżniliśmy grupę $\text{FIX}_G(x)$, złożoną z tych permutacji należących do G, które nie ruszają danego elementu x zbioru X. Teraz odwróćmy punkt widzenia i dla każdej permutacji g z grupy G spójrzmy na zbiór

$$\text{FIX}_X(g) = \{x \in X\colon g(x) = x\}$$

składający się z tych wszystkich elementów zbioru X, które są punktami stałymi permutacji g. Indeks, występujący w oznaczeniu FIX_X, powinien nam pomóc w pamiętaniu, że chodzi tu o podzbiór zbioru X; podobnie, dobrze nam znane zbiory FIX_G są podzbiorami grupy G. Zbiory postaci $\text{FIX}_X(g)$ wiążą się w zaskakujący sposób z orbitami grupy G.

Twierdzenie 1

Niech G będzie (skończoną) grupą działającą na zbiorze X. Liczba orbit działania grupy G na zbiorze X jest równa

$$\frac{1}{|G|}\Big(\sum_{g\in G} |\,\text{FIX}_X(g)|\Big).$$

W powyższej sumie, każdemu elementowi g z grupy G odpowiada dokładnie jeden składnik. Jeśli w miarę dobrze zrozumiałeś twierdzenie z § 12.2, to przeczytaj teraz poniższy dowód, a następnie zobacz, co twierdzenie to mówi w szczególnych przypadkach. Jeśli nie, to radzimy, byś teraz dowód ten pominął i powrócił doń po przestudiowaniu przykładów.

Dowód. Wykorzystamy pewien pomysł z dowodu uogólnionej zasady szufladkowej z § 5.5; polega on na zliczaniu par uporządkowanych na dwa sposoby. Znalezienie zbioru, którego elementy należy zliczać, może czasem wymagać pomysłowości, ale w naszym przypadku wybór jest całkiem naturalny:

$$S = \{(g, x) \in g \times X\colon g(x) = x\}.$$

Najpierw dla każdej permutacji g z grupy G zliczamy pary (g, x) takie, że $g(x) = x$ — jest ich $|\,\text{FIX}_X(g)|$ — a następnie otrzymane wyniki sumujemy. Otrzymujemy

$$|S| = \sum_{g\in G} |\,\text{FIX}_X(g)|. \tag{1}$$

Liczbę elementów zbioru S możemy także znaleźć zliczając, dla każdego punktu x ze zbioru X, pary (g, x) takie, że $g(x) = x$.

Otrzymujemy wtedy

$$|S| = \sum_{x \in X} |\operatorname{FIX}_G(x)|. \tag{2}$$

Na mocy twierdzenia z § 12.2 $|\operatorname{FIX}_G(x)| = \frac{|G|}{|Gx|}$ dla każdego $x \in X$, a więc

$$|S| = \sum_{x \in X} \frac{|G|}{|Gx|} = |G| \sum_{x \in X} \frac{1}{|Gx|}.$$

Zgrupujmy teraz wszystkie te wyrazy powyższej sumy, które pochodzą od elementów ustalonej orbity Gx_0. Ponieważ $Gx = Gx_0$ dla każdego x z Gx_0, to

$$\sum_{x \in Gx_0} \frac{1}{|Gx|} = \sum_{x \in Gx_0} \frac{1}{|Gx_0|} = 1.$$

Znaczy to, że elementy należące do ustalonej orbity wnoszą wspólnie do ogólnej sumy $\sum_{x \in X} \frac{1}{|Gx|}$ wartość 1. Jeśli więc jest m orbit, to

$$\sum_{x \in X} \frac{1}{|Gx|} = 1 + 1 + \ldots + 1 = m.$$

Wynika stąd, że $|S| = |G| \cdot m$, a więc na mocy (1) mamy

$$m = \frac{1}{|G|} \cdot |S| = \frac{1}{G}\Big(\sum_{g \in G} |\operatorname{FIX}_X(g)|\Big),$$

co należało pokazać. ∎

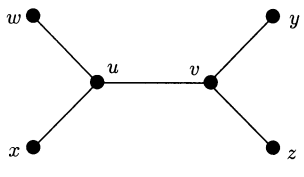
W naszych pierwszych przykładach łatwo będzie obliczyć wartości obu stron równości z twierdzenia 1, tak więc sens udowodnionego twierdzenia może wzbudzić wątpliwości. W nietrywialnych zastosowaniach, takich jak w § 12.4, bezpośrednie obliczenie liczby orbit może być trudne, natomiast wartości $|\operatorname{FIX}_X(g)|$ można określić stosunkowo łatwo.

PRZYKŁAD 1 Wracamy do grupy $G = \operatorname{AUT}(H)$ automorfizmów grafu H z rysunku 12.15, z którą zetknęliśmy się w przykładzie 5 z § 12.2. Grupa ta działa na zbiorze $V = \{u, v, w, x, y, z\}$ wierzchołków grafu H. Aby zobaczyć, które elementy zbioru V są punktami stałymi poszczególnych automorfizmów, przyjrzymy się tablicy (b) z rysunku 12.15. Stwierdzamy, że $\operatorname{FIX}_V(e) = \{u, v, w, x, y, z\}$, $\operatorname{FIX}_V(g) = \{u, v, y, z\}$, $\operatorname{FIX}_V(h) = \{u, v, w, x\}$, $\operatorname{FIX}_V(gh) = \{u, v\}$ i $\operatorname{FIX}_V(f) = \operatorname{FIX}_V(fh) = \operatorname{FIX}_V(fg) = \operatorname{FIX}_V(fgh) = \emptyset$. Zbiory te składają się z, odpowiednio, 6, 4, 4, 2,

0, 0, 0 i 0 wierzchołków. Ponadto, $|G| = 8$, a więc twierdzenie 1 głosi, że liczba orbit działania grupy G wynosi

$$\frac{1}{8}(6+4+4+2+0+0+0+0) = 2.$$

Istotnie, z rysunku grafu H lub ze wspomnianej tablicy widzimy, że istnieją dokładnie dwie orbity działania grupy G, a mianowicie $\{u, v\}$ i $\{w, x, y, z\}$; twierdzenie 1 potwierdza więc nasze obserwacje. ∎



H

(a)

	u	v	w	x	y	z	Cykle
e	u	v	w	x	y	z	e
g	u	v	x	w	y	z	$(w\,x)$
h	u	v	w	x	z	y	$(y\,z)$
gh	u	v	x	w	z	y	$(w\,x)(y\,z)$
f	v	u	y	z	w	x	$(u\,v)(w\,y)(x\,z)$
fh	v	u	y	z	x	w	$(u\,v)(w\,y\,x\,z)$
fg	v	u	z	y	w	x	$(u\,v)(w\,z\,x\,y)$
fgh	v	u	z	y	x	w	$(u\,v)(w\,z)(x\,y)$

AUT (H)

(b)

Rysunek 12.15

Rozważmy dowolny graf H o zbiorze wierzchołków V i załóżmy, że H nie ma krawędzi wielokrotnych. Dotychczas uważaliśmy grupę $G = \text{AUT}\,(H)$ za grupę działającą na zbiorze V, tzn. za podzbiór zbioru PERM (V). Możemy również potraktować G jak grupę działającą na zbiorze E krawędzi naszego grafu. Mianowicie, dla każdej permutacji $g \in G \subseteq \text{PERM}\,(V)$, definiujemy element g^* zbioru PERM (E), kładąc $g^*(\{u, v\}) = \{g(u), g(v)\}$ dla każdej krawędzi $\{u, v\}$. Zauważ, że jeśli $f, g \in G$, to

$$(*) \qquad (f \circ g)^* = f^* \circ g^*;$$

pierwsze złożenie odbywa się w zbiorze PERM (V), a drugie — w PERM (E). Własność $(*)$ zachodzi dlatego, że

$$(f \circ g)^*(\{u, v\}) = \{f \circ g(u), f \circ g(v)\}$$

oraz

$$f^* \circ g^*(\{u, v\}) = f^*(\{g(u), g(v)\}) = \{f(g(u)), f(g(v))\}.$$

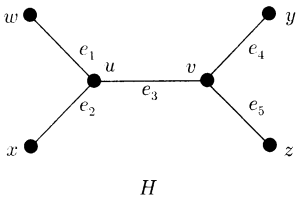
Na ogół funkcja $g \longrightarrow g^*$ jest różnowartościowa i wówczas warunek $(*)$ pokazuje, że grupy G i $G^* = \{g^*\colon g \in G\}$ są w gruncie

rzeczy takie same. Będziemy mówili, że grupa G działa na zbiorze E, mimo że formalnie rzecz biorąc to grupa G^* działa na E. W rzeczywistości będziemy mówili, że grupa G działa na zbiorze X nawet wtedy, gdy funkcja $g \to g^*$ nie będzie różnowartościowa, pod warunkiem jednak, że spełniona będzie własność $(*)$.

PRZYKŁAD 2 (a) Rysunek 12.16(a) przedstawia ponownie graf H z tym, że jego krawędzie również otrzymały swoje etykiety. Tablica z rysunku 12.16(b) pokazuje, w jaki sposób grupa G działa na zbiorze E. Aby zobaczyć, jak ta tablica powstała, sprawdźmy, że wartości funkcji f^* zostały w niej wypisane poprawnie: f jest automorfizmem, który przekształca graf przez symetrię względem prostopadłej osi przechodzącej przez krawędź e_3; zatem f^* zamienia miejscami krawędzie e_1 z e_4 oraz e_2 z e_5 i nie rusza krawędzi e_3. Możemy te stwierdzenia sprawdzić formalnie, pisząc

$$f^*(e_1) = f^*(\{u,w\}) = \{f(u), f(w)\} = \{v,y\} = e_4,$$
$$f^*(e_3) = f^*(\{u,v\}) = \{f(u), f(v)\} = \{v,u\} = e_3$$

itd., ale nie jest to konieczne w przypadku grafów, które potrafimy narysować.



H

(a)

	e_1	e_2	e_3	e_4	e_5	Cykle
e^*	e_1	e_2	e_3	e_4	e_5	e
g^*	e_2	e_1	e_3	e_4	e_5	(e_1e_2)
h^*	e_1	e_2	e_3	e_5	e_4	(e_4e_5)
$(gh)^*$	e_2	e_1	e_3	e_5	e_4	$(e_1e_2)(e_4e_5)$
f^*	e_4	e_5	e_3	e_1	e_2	$(e_1e_4)(e_2e_5)$
$(fh)^*$	e_4	e_5	e_3	e_2	e_1	$(e_1e_4e_2e_5)$
$(fg)^*$	e_5	e_4	e_3	e_1	e_2	$(e_1e_5e_2e_4)$
$(fgh)^*$	e_5	e_4	e_3	e_2	e_1	$(e_1e_5)(e_2e_4)$

(b)

Rysunek 12.16

Teraz posługując się rysunkiem 12.16(b) znajdujemy $\text{FIX}_E(e^*) = \{e_1, e_2, e_3, e_4, e_5\}$, $\text{FIX}_E(g^*) = \{e_3, e_4, e_5\}$, $\text{FIX}_E(h^*) = \{e_1, e_2, e_3\}$ i $\text{FIX}_E(a^*) = \{e_3\}$ dla każdego innego automorfizmu a. Zbiory te składają się z, odpowiednio, 5, 3, 3, 1, 1, 1, 1, 1 krawędzi. Zgodnie z twierdzeniem 1, grupa $G = \text{AUT}(H)$ wyznacza liczbę orbit w zbiorze E:

$$\frac{1}{8}(5+3+3+1+1+1+1+1) = 2.$$

Oczywiście orbitami tymi są zbiory $\{e_3\}$ i $\{e_1, e_2, e_4, e_5\}$.

(b) Grupa G działa również na zbiorze T wszystkich dwuelementowych podzbiorów zbioru V, przy czym definicja g^* jest taka sama, jak poprzednio. Zauważmy, że zbiór T zawiera zbiór krawędzi E i jeszcze wiele innych elementów. Istotnie, zbiór T ma $\binom{6}{2} = 15$ elementów. Tym razem sytuacja jest bardziej abstrakcyjna i trudniej ją sobie wyobrazić niż wtedy, gdy rozważaliśmy działanie grupy G na zbiorze wierzchołków bądź krawędzi. Nie mamy rysunku zbioru T, który mógłby nam pomóc, a tablica taka jak na rysunku 12.16(b), byłaby bardzo zagmatwana. Będziemy więc po prostu musieli posłużyć się matematycznym rozumowaniem. Jak zawsze, $\text{FIX}_T(e^*)$ jest całym zbiorem, na którym działa grupa G, a więc tym razem jest to zbiór T. Permutacja g przeprowadza każdy z elementów zbioru $\{u, v, y, z\}$ na siebie, a więc permutacja g^* przeprowadza każdy z dwuelementowych podzbiorów zbioru $\{u, v, y, z\}$ z powrotem na siebie. Nie rusza ona też pary $\{w, x\}$, gdyż $g^*(\{w, x\}) = \{g(w), g(x)\} = \{x, w\}$. Zatem zbiór $\text{FIX}_T(g^*)$ ma $\binom{4}{2} + 1 = 7$ elementów. Podobnie, $|\text{FIX}_T(h^*)| = 7$. Ponadto, $\text{FIX}_T((gh)^*) = \{\{u, v\}, \{w, x\}, \{y, z\}\}$, $\text{FIX}_T(f^*) = \{\{u, v\}, \{w, y\}, \{x, z\}\}$, $\text{FIX}_T((fh)^*) = \text{FIX}_T((fg)^*) = \{\{u, v\}\}$ i $\text{FIX}_T((fgh)^*) = \{\{u, v\}, \{w, z\}, \{x, y\}\}$. A więc liczby elementów ośmiu zbiorów postaci $\text{FIX}_T(\ )$ wynoszą, odpowiednio, 15, 7, 7, 3, 3, 1, 1, 3.

Twierdzenie 1 mówi zatem, że zbiór T dzieli się na

$$\frac{1}{8}(15 + 7 + 7 + 3 + 3 + 1 + 1 + 3) = 5$$

G-orbit. Nie było to od początku takie oczywiste. Teraz możemy zauważyć, że zbiory $\{w, u\}$, $\{w, v\}$, $\{w, x\}$, $\{w, y\}$ i $\{u, v\}$ należą do pięciu różnych orbit. (Spójrz na rysunek 12.15(a), aby przekonać się, że żaden z tych zbiorów nie może być przekształcony na żaden inny spośród nich przez automorfizm grafu H). Gdyby chodziło o znalezienie reprezentanta każdej orbity, to w momencie, w którym wskazalibyśmy pięć powyższych zbiorów, wiedzielibyśmy, że zadanie zostało wykonane. ∎

Rozważmy ponownie grupę G działającą na zbiorze X: $G \subseteq \text{PERM}(X)$. Czasami chcemy ograniczyć działanie grupy do podzbioru Y zbioru X, nie biorąc po prostu pod uwagę tego, co permutacje z grupy G robią z elementami spoza zbioru Y. Jednakże żeby można było tak zrobić, musimy mieć pewność, że permutacje z grupy G przeprowadzają elementy zbioru Y na elementy zbioru Y. Weźmy zatem taki podzbiór Y zbioru X, który jest bądź orbitą działania grupy G, bądź sumą takich orbit. Dla każdej permutacji $g \in G$ definiujemy g^*: $Y \to Y$ wzorem $g^*(y) = g(y)$; funkcja g^*

jest nazywana **obcięciem funkcji** g do zbioru Y. Sprawdź, jak pojęcie to było używane w § 1.4. Ponieważ zbiór Y jest skończony i g^* jest różnowartościowym przekształceniem zbioru Y w Y, to g^* przekształca Y na Y. Wynika stąd, że każda z funkcji g^* jest permutacją zbioru Y. Zauważ również, że równość

$$(*) \qquad (f \circ g)^* = f^* \circ g^*$$

zachodzi dla wszystkich permutacji $f, g \in G$. Przekształcenie $g \longrightarrow g^*$ może być różnowartościowe lub nie, jak zobaczymy w następnym przykładzie.

PRZYKŁAD 3 (a) Wracamy do przykładu z rysunku 12.15 i obcinamy działanie grupy $G = \text{AUT}(H)$ do orbity $\{u, v\}$. Żeby otrzymać tablicę wartości obcięć permutacji z G do zbioru $\{u, v\}$, wystarczy po prostu pominąć cztery ostatnie kolumny tablicy z rysunku 12.15(b) i przyjąć, że każda z permutacji e, g, h itd. ma dorysowaną gwiazdkę $*$. Mamy osiem różnych nazw dla otrzymanych obcięć, ale są wśród nich tylko dwa naprawdę różne: e^* i f^*. Przekształcenie $g \longrightarrow g^*$ nie jest różnowartościowe; przykładowo, $e^* = g^* = h^* = (gh)^*$. Pomimo to mówimy, że grupa G działa na orbicie $\{u, v\}$.

(b) Używamy tej samej grupy G, ale obcinamy jej działanie do drugiej orbity $\{w, x, y, z\}$. Tablicę wartości otrzymujemy pomijając kolumny u i v. Zauważmy, że każda z ośmiu obciętych permutacji jest inna, a więc przekształcenie $g \longrightarrow g^*$ jest tym razem różnowartościowe. ∎

Powróćmy do twierdzenia 1. Ponieważ jest $|G|$ wyrazów w sumie

$$\sum_{g \in G} |\,\text{FIX}_X(g)|,$$

to dzieląc tę sumę przez $|G|$ otrzymujemy średnią arytmetyczną liczb $|\,\text{FIX}_X(g)|$ dla $g \in G$. Jeśli pewne wyrazy sumy są większe od tej średniej, to jakieś inne wyrazy muszą być od niej mniejsze. Ta obserwacja prowadzi do następujących niespodziewanych wniosków.

Wniosek 1

> Jeśli X jest jedyną orbitą działania grupy G na zbiorze X i $|X| > 1$, to istnieje element g w G taki, że $g(x) \neq x$ dla wszystkich $x \in X$.

Dowód. Na mocy twierdzenia 1 średnia arytmetyczna liczb $|\,\text{FIX}_X(g)|$ wynosi 1, ponieważ istnieje tylko jedna orbita. Po-

nadto, $|\text{FIX}_X(e)| = |X| > 1$, a więc $|\text{FIX}_X(g)| < 1$ dla co najmniej jednego elementu g z G. Dla takiego elementu g zbiór $\text{FIX}_X(g) = \{x \in X\colon g(x) = x\}$ musi być zbiorem pustym. ∎

Ćwiczenie 15 w § 12.2 dotyczy działań takich, jak we wniosku 1.

Wniosek 2

Jeśli grupa G działa na zbiorze X i Y jest orbitą działania G na X taką, że $|Y| > 1$, to istnieje element g grupy G taki, że $g(x) \neq x$ dla wszystkich $x \in Y$.

Dowód. Zbiór Y jest jedyną orbitą grupy złożonej z obcięć g^* elementów g grupy G do zbioru Y. Zatem na mocy wniosku 1, istnieje permutacja g^* zbioru Y taka, że $g^*(x) \neq x$ dla wszystkich $x \in Y$. Wynika stąd, że $g(x) \neq x$ dla wszystkich $x \in Y$. ∎

PRZYKŁAD 4

Wróćmy do sytuacji z przykładu 3. Zgodnie z wnioskiem 2, pewien automorfizm musi poruszyć każdy element orbity $\{u, v\}$. W istocie każdy z automorfizmów f, fh, fg i fgh ma tę własność.

Ten sam wniosek zapewnia nas, że pewien automorfizm porusza każdy element orbity $\{w, x, y, z\}$. Rzut oka na rysunek 12.15(b) wystarczy, by zobaczyć, że każdy z automorfizmów gh, f, fh, fg i fgh ma tę własność. ∎

Kilkakrotnie w tym paragrafie uważaliśmy jakąś grupę G za grupę działającą na danym zbiorze, mimo że zbiór ten nie występował w wyjściowej definicji grupy G. Było to zgodne z następującą formalną definicją. Jeśli G jest grupą, a $g \to g^*$ jest funkcją z G w $\text{PERM}(X)$ spełniającą warunek

$$(*) \qquad (f \circ g)^* = f^* \circ g^*,$$

to mówimy, że **grupa G działa na zbiorze X**. Ponieważ zbiór $G^* = \{g^*\colon g \in G\}$ jest grupą permutacji zbioru X, to wyniki tego i poprzedniego paragrafu pozostają w mocy dla tak rozszerzonego pojęcia działania grupy na zbiorze X.

Czasami grupy działają na pewnych użytecznych i pojawiających się w naturalny sposób zbiorach funkcji. Przypuśćmy, że grupa G działa na zbiorze X. Niech $\text{FUN}(X, K)$ będzie zbiorem wszystkich funkcji ze zbioru X w pewien skończony zbiór, który oznaczyliśmy przez K, z uwagi na planowane zastosowania do zagadnień kolorowania, które omówimy w następnym paragrafie. Dla funkcji φ należącej do zbioru $\text{FUN}(X, K)$ definiujemy

$g^*(\varphi) = \varphi \circ g^{-1}$. Wówczas dla $f, g \in G$,

$$(f \circ g)^*(\varphi) = \varphi \circ (f \circ g)^{-1} = \varphi \circ g^{-1} \circ f^{-1},$$

podczas gdy

$$f^* \circ g^*(\varphi) = f^*(\varphi \circ g^{-1}) = (\varphi \circ g^{-1}) \circ f^{-1}.$$

Znaczy to, że

$$(*) \qquad (f \circ g)^* = f^* \circ g^* \quad \text{dla} \quad f, g \in G,$$

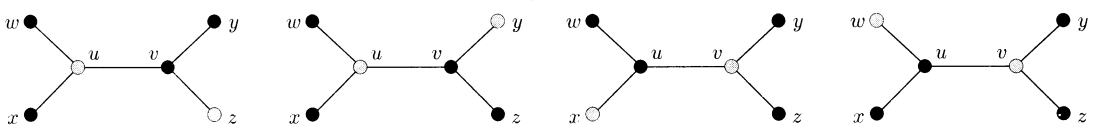
a więc grupa G działa na zbiorze $\text{FUN}(X, K)$.

G-orbitą danej funkcji φ należącej do zbioru $\text{FUN}(X, K)$ jest zbiór $G\varphi = \{g^*(\varphi)\colon g \in G\} = \{\varphi \circ g^{-1}\colon g \in G\}$. Ponieważ każdy element h grupy G jest odwrotnością pewnego elementu tej grupy, a mianowicie elementu $h = (h^{-1})^{-1}$, to zbiór ten jest po prostu zbiorem $\{\varphi \circ h\colon h \in G\}$, który moglibyśmy oznaczyć przez $\varphi \circ G$.

PRZYKŁAD 5 W następnym paragrafie będziemy używać działań grup na zbiorach postaci $\text{FUN}(X, K)$ do rozwiązywania problemów takich, jak zagadnienia związane z kolorowaniem sześcianu i znajdywanie liczby istotnie różnych układów logicznych. Dla zilustrowania, jak w przydatny sposób interpretować można zbiór $\text{FUN}(X, K)$, zaczniemy tutaj od pewnego prostszego zadania dotyczącego kolorowania:

Na ile sposobów można pokolorować wierzchołki grafu przedstawionego na rysunku 12.15, używając szarego, czarnego lub obydwu kolorów?

Każdy z tych sześciu wierzchołków może być pokolorowany na szaro lub czarno, a więc są $2^6 = 64$ możliwe sposoby przyporządkowania każdemu wierzchołkowi jednego z kolorów. Jednakże to obliczenie pomija strukturę grafu. Chcemy uważać dwa kolorowania za równoważne, jeśli pewien automorfizm naszego grafu przeprowadza jedno z nich na drugie. Naszym zadaniem jest przetłumaczenie ostatniego zdania na język matematyki.



Rysunek 12.17

Aby zobaczyć, o co chodzi, rozważmy pokolorowane grafy z rysunku 12.17. Odpowiadają one czterem różnym sposobom wybrania kolorów dla wierzchołków, ale na podstawie rysunku widzimy,

że możemy z każdego z tych pokolorowanych grafów przejść do dowolnego innego za pomocą odpowiedniego automorfizmu grafu. Przykładowo, automorfizm h z tablicy z rysunku 12.15(b) zamienia miejscami prawy górny wierzchołek z prawym dolnym wierzchołkiem i przeprowadza rysunek 12.17(a) na rysunek (b); obrót rysunku o 180^0 za pomocą automorfizmu fgh przeprowadza rysunek (a) na rysunek (d) i rysunek (b) na rysunek (c).

Każde kolorowanie grafu z rysunku 12.17 odpowiada pewnej funkcji φ: $V \to K$, gdzie V jest zbiorem wierzchołków grafu, a $K = \{C, S\}$ jest zbiorem kolorów: czarnego C i szarego S; niech φ_a, φ_b, φ_c, φ_d będą nazwami tych czterech kolorowań. Kolorowania φ_a i φ_b są rzeczywiście równoważne, gdyż $\varphi_b = \varphi_a \circ h$, tzn. możemy otrzymać kolorowanie (b), zamieniając najpierw miejscami odpowiednie wierzchołki za pomocą automorfizmu h, a następnie dokonując kolorowania (a). Podobnie, $\varphi_a \circ f = \varphi_c$ i $\varphi_a \circ fgh = \varphi_d$.

Gdy mówimy, że **kolorowania ψ i φ są równoważne**, to rozumiemy przez to, że $\psi = \varphi \circ f$ dla pewnej funkcji f należącej do grupy automorfizmów G. Innymi słowy, kolorowania ψ i φ są równoważne wtedy, gdy należą do tej samej orbity $\varphi \circ G$ działania grupy G na zbiorze $\text{FUN}\,(V, K)$. ∎

Liczba istotnie różnych kolorowań w przykładzie 5 jest równa liczbie G-orbit w zbiorze $\text{FUN}\,(V, K)$. Odpowiedź, że jest ich 21, nie jest oczywista. Aby znaleźć tę liczbę, trzeba będzie zastosować twierdzenie 1 do zbioru $X = \text{FUN}\,(V, K)$, a więc trzeba będzie umieć znaleźć liczby takie, jak $|\,\text{FIX}_{\,\text{FUN}\,(V,K)}(g^*)|$.

PRZYKŁAD 6 Automorfizm fh grafu z przykładu 5 można przedstawić w postaci iloczynu rozłącznych cykli jako $(uv)(wyxz)$, z orbitami $\{u, v\}$ i $\{w, x, y, z\}$. Aby dane kolorowanie było punktem stałym permutacji $(fh)^*$, wierzchołki należące do tej samej orbity muszą być jednakowego koloru; w przeciwnym przypadku zastosowanie automorfizmu fh spowodowałoby widoczną zmianę kolorowania. Istnieją cztery kolorowania spełniające ten warunek: wszystkie wierzchołki czarne, wszystkie szare, wierzchołki u i v czarne, ale w, x, y, z szare oraz wierzchołki u i v szare, ale w, x, y, z czarne. Te cztery kolorowania są jedynymi elementami zbioru $\text{FIX}_{\,\text{FUN}\,(V,K)}(fh)^*$. Innymi słowy, zbiór $\text{FIX}_{\,\text{FUN}\,(V,K)}(fh)^*$ składa się z tych funkcji należących do zbioru $\text{FUN}\,(V, K)$, które są stałe na orbitach automorfizmu fh. ∎

Następujące twierdzenie, które wykorzystywać będziemy w następnym paragrafie, jest uogólnieniem powyższego przykładu. Jak zwykle, $\langle g \rangle$ oznacza tu grupę generowaną przez g.

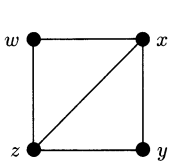
Twierdzenie 2

Załóżmy, że grupa G działa na zbiorze X, a więc również na zbiorze $\text{FUN}(X, K)$. Jeśli g jest elementem grupy G, to zbiór $\text{FIX}_{\text{FUN}(X,K)}(g^*)$ składa się ze wszystkich funkcji $\varphi\colon X \to K$, które są stałe na $\langle g\rangle$-orbitach.

Dowód. Nasze zadanie polega na pokazaniu, że $g^*(\varphi) = \varphi$ wtedy i tylko wtedy, gdy funkcja φ jest stała na $\langle g\rangle$-orbitach. Zauważmy, że $g^*(\varphi) = \varphi$ wtedy i tylko wtedy, gdy $\varphi \circ g^{-1}(x) = \varphi(x)$ dla każdego $x \in X$. Podstawiając y w miejsce $g^{-1}(x)$, co daje $x = g(y)$, stwierdzamy, że $g^*(\varphi) = \varphi$ wtedy i tylko wtedy, gdy $\varphi(y) = \varphi(g(y))$ dla każdego $y \in X$. To znaczy, że $g^*(\varphi) = \varphi$ wtedy i tylko wtedy, gdy dla każdego y wszystkie wartości $\varphi(y)$, $\varphi(g(y))$, $\varphi(g^2(y))$, $\varphi(g^3(y)), \ldots$ są takie same. Wynika stąd, że $g^*(\varphi) = \varphi$ wtedy i tylko wtedy, gdy funkcja φ jest stała na każdej orbicie $\{g^n(y)\colon n \in \mathbb{P}\}$ grupy $\langle g\rangle$. ∎

ĆWICZENIA DO § 12.3

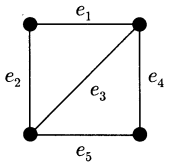
1. Weźmy grupę $G = \text{AUT}(H)$ z przykładu 1, działającą na zbiorze V wierzchołków grafu H.
   (a) Dla każdego automorfizmu a z grupy G znajdź $|\text{FIX}_V(a)|$, a otrzymane wyniki dodaj do siebie.
   (b) Dla każdego wierzchołka p grafu H znajdź $|\text{FIX}_G(p)|$, a otrzymane wyniki dodaj do siebie.
   (c) Czy sumy otrzymane w punktach (a) i (b) są równe? Omów to zagadnienie.
2. Weźmy graf z rysunku 12.13 z § 12.2. Przekonaj się, że dla tego przykładu zachodzi wzór z twierdzenia 1.



(a)

	w	x	y	z
e	w	x	y	z
g	y	x	w	z
h	w	z	y	x
gh	y	z	w	x

(b)



(c)

Rysunek 12.18

3. Weźmy graf z rysunku 12.18 i niech G będzie grupą automorfizmów tego grafu, działającą na zbiorze $\{w, x, y, z\}$.

(a) Przekonaj się, że $G = \{e, g, h, gh\}$, gdzie automorfizmy te są określone na rysunku 12.18(b).
(b) Przekonaj się, że dla tego przykładu zachodzi wzór z twierdzenia 1.

4. Niech grupa G z ćwiczenia 3 działa na pięcioelementowym zbiorze E wszystkich krawędzi naszego grafu. Zobacz rysunek 12.18(c).
(a) Podaj tablicę wartości permutacji z G^* analogiczną do tej, którą przedstawia rysunek 12.16(b).
(b) Przekonaj się, że dla tego przykładu zachodzi wzór z twierdzenia 1.

5. Wykaż bezpośrednio, że wniosek 2 jest prawdziwy w odniesieniu do każdej orbity działania grupy z ćwiczenia 3.

6. Wykaż bezpośrednio, że wniosek 2 jest prawdziwy w odniesieniu do każdej orbity działania grupy z ćwiczenia 4.

7. (a) Pokoloruj wierzchołki grafu z rysunku 12.18(a) w następujący sposób: w i x na czarno, y i z na szaro. Znajdź wszystkie równoważne kolorowania.
(b) Powtórz to samo dla kolorowania, przy którym wierzchołek w jest czarny, a wszystkie pozostałe wierzchołki są szare.

8. (a) Pokoloruj krawędzie grafu z rysunku 12.18(c) w następujący sposób: e_1, e_3, e_5 na szaro, e_2, e_4 na czarno. Znajdź wszystkie równoważne kolorowania.
(b) Powtórz to samo dla kolorowania, przy którym krawędzie e_1, e_2, e_3 są szare, a e_4, e_5 są czarne.

9. Weźmy graf z rysunku 12.13 z § 12.2.
(a) Pokoloruj jego wierzchołki w następujący sposób: x, y na szaro, a w, z na czarno. Znajdź wszystkie równoważne kolorowania.
(b) Zrób to samo w przypadku, gdy wierzchołki x, z są szare, a w, y są czarne.
(c) Zrób to samo w przypadku, gdy wierzchołek x jest szary, a wszystkie pozostałe wierzchołki są czarne.
(d) Czy potrafisz zgadnąć, ile jest istotnie różnych kolorowań zbioru wierzchołków, przy użyciu szarego, czarnego lub obu tych kolorów?

10. W przykładzie 3(b) grupa G^* złożona z obcięć automorfizmów grafu H do orbity $\{w, x, y, z\}$ składa się z 8 elementów, podczas gdy grupa $\text{PERM}(\{w, x, y, z\}$ ma 24 elementy. Wskaż dwie permutacje zbioru $\{w, x, y, z\}$, które nie należą do grupy G^*.

11. Sprawdź, że w przykładzie 2(b) $|\text{FIX}_T(h^*)| = 7$.

12. (a) Wykaż, że twierdzenie z ćwiczenia 7 w § 12.2 można zastosować do grupy G działającej na zbiorze E z przykładu 2(a).
(b) Wskaż konkretny przykład na poparcie tezy z części (a).

13. (a) Wykaż, że twierdzenie z ćwiczenia 7 w § 12.2 można zastosować do grupy z rysunku 12.18, działającej na zbiorze krawędzi danego grafu.
(b) Wskaż konkretny przykład na poparcie tezy z części (a).

14. Dlaczego, pokazując w jaki sposób grupę $G \subseteq \text{PERM}(X)$ uważać można za grupę działającą na zbiorze $\text{FUN}(X, K)$, nie zdefiniowaliśmy g^* wzorem $g^*(\varphi) = \varphi \circ g$?

15. Niech H będzie grafem spójnym bez krawędzi wielokrotnych, V — zbiorem jego wierzchołków, a E — zbiorem jego krawędzi. Ponadto, niech $G \subseteq \text{PERM}(V)$ będzie grupą wszystkich automorfizmów grafu H, działającą na zbiorze V, i niech $g \longrightarrow g^*$ będzie przekształceniem grupy G w grupę $\text{PERM}(E)$, opisanym przed przykładem 2.

(a) Wykaż, że przekształcenie $g \longrightarrow g^*$ nie jest różnowartościowe, jeśli graf H ma dokładnie jedną krawędź łączącą jego jedyne dwa wierzchołki u i v.

(b) Wykaż, że jeśli graf H ma więcej niż jedną krawędź, to przekształcenie $g \longrightarrow g^*$ jest różnowartościowe. *Wskazówka*: wystarczy pokazać, że jeśli $g^* = e^*$, to $g = e$.

16. Weźmy grupę G działającą na pewnym n-elementowym zbiorze X. Wykaż, że jeśli dla każdego $g \in G$, $|\text{FIX}_X(g)| \geq 1$, to działanie grupy G wyznacza co najmniej $1 + \frac{n-1}{|G|}$ orbit w zbiorze X. Dla $n > 1$ wynika stąd wniosek 1 z twierdzenia 1. *Wskazówka*: przyjrzyj się oddzielnie elementowi e grupy G.

§ 12.4. Zastosowania działań grup na zbiorach do problemów kolorowania

Rozważmy zagadnienie tworzenia układów logicznych, które mają służyć do obliczania wszystkich możliwych funkcji booleowskich czterech zmiennych. Z jednej strony, ponieważ tablica logiczna każdej z tych funkcji ma $2^4 = 16$ wierszy, funkcji takich jest $2^{16} = 65536$. Z drugiej strony, zamieniając po prostu ze sobą przewody na wejściu, możemy sprawić, że jeden układ obliczać będzie wiele różnych funkcji, a więc nie potrzebujemy aż tylu układów, ile jest funkcji. Jeśli chcemy używać zewnętrznego urządzenia, które wartości na wejściu i wyjściu zamieniać będzie na ich dopełnienia, to wystarczy wytworzyć jeszcze mniej układów. Ile?

Rozważmy też zagadnienie kolorowania ścian sześcianu przy użyciu tylko trzech kolorów. Jeśli jedna ze ścian jest czerwona, a pozostałe są niebieskie, to w gruncie rzeczy nie gra roli, która ze ścian jest czerwona, bo możemy obrócić sześcian tak, by czerwona ściana znalazła się tam, gdzie chcemy. Jeśli są dane dwa kolorowania ścian w taki sposób, że jedna jest czerwona, trzy są niebieskie i dwie — zielone, to może być, ale może też i nie być możliwe obrócenie sześcianu tak, by te kolorowania okazały się w gruncie rzeczy takie same. Ile jest zatem istotnie różnych spo-

sobów pokolorowania sześcianu? Naturalne jest, by w tym przypadku rozważać grupę tych wszystkich obrotów przestrzeni, które przekształcają nasz sześcian na siebie, gdyż ta właśnie grupa opisuje własności sześcianu, które są niezmiennicze ze względu na obroty. Grupa ta działa na zbiorze ścian sześcianu, ale działa też na zbiorze $\text{FUN}(X, K)$ wszystkich kolorowań tych ścian, gdzie $K = \{\text{czerwony}, \text{niebieski}, \text{zielony}\}$. Klasy równoważnych kolorowań to po prostu orbity działania grupy G na zbiorze $\text{FUN}(X, K)$ i mamy nadzieję zliczyć je za pomocą twierdzenia 1 z poprzedniego paragrafu. Metoda, której użyjemy, ma dość szerokie zastosowania, więc zamiast już teraz rozwiązywać problem liczby kolorowań sześcianu, udowodnimy najpierw twierdzenie ogólne, a następnie wykorzystamy je w przykładzie 3 do szczególnego przypadku, jakim jest zagadnienie kolorowania sześcianu.

Twierdzenie 1

Weźmy skończoną grupę G działającą na zbiorze X. Jeśli K jest dowolnym zbiorem, to grupa G działa również na zbiorze $\text{FUN}(X, K)$ zgodnie z następującą definicją: $g^*(\varphi) = \varphi \circ g^{-1}$ dla $g \in G$ i $\varphi\colon X \to K$. Dla każdego elementu g grupy G niech $m(g)$ będzie liczbą orbit działania grupy $\langle g \rangle$ w zbiorze X. Wówczas liczba orbit działania grupy G w zbiorze $\text{FUN}(X, K)$ wynosi

$$\frac{1}{|G|} \sum_{g \in G} |K|^{m(g)}.$$

Dowód. Zgodnie z twierdzeniem 1 z § 12.3, aby wykazać prawdziwość powyższego wzoru, musimy po prostu pokazać, że dla każdego $g \in G$ mamy

$$|\,\text{FIX}_{\,\text{FUN}(X,K)}(g^*)| = |K|^{m(g)}.$$

Ale twierdzenie 2 z § 12.3 mówi, że zbiór $\text{FIX}_{\,\text{FUN}(X,K)}(g^*)$ składa się z tych funkcji $\varphi\colon X \to K$, które są stałe na wszystkich $\langle g \rangle$-orbitach. Aby określić taką funkcję φ, podamy po prostu, jaką wartość przyjmuje na każdej z $\langle g \rangle$-orbit. Jest $|K|$ możliwych wartości funkcji i $m(g)$ orbit działania grupy $\langle g \rangle$, a więc jest $|K|^{m(g)}$ funkcji φ, które są stałe na wszystkich $\langle g \rangle$-orbitach. Stąd wynika teza dowodzonego twierdzenia. ∎

Kolorowanie zbioru X za pomocą kolorów ze zbioru K jest po prostu funkcją ze zbioru X w zbiór K; dwa kolorowania uważamy za równoważne ze względu na działanie grupy G, jeśli należą one do tej samej G-orbity $\{\varphi \circ g^{-1}\colon g \in G\} = \varphi \circ G$ w zbiorze

FUN (X, K). Te G-orbity są klasami G-równoważnych kolorowań. Używając tej terminologii twierdzenie 1 można zapisać w następujący sposób.

Twierdzenie 2

Weźmy grupę G działającą na zbiorze X oraz zbiór k kolorów. Niech $C(k)$ będzie liczbą klas złożonych z G-równoważnych kolorowań zbioru X przy użyciu niektórych bądź wszystkich spośród danych k kolorów. Wtedy

$$C(k) = \frac{1}{|G|} \sum_{g \in G} k^{m(g)},$$

gdzie $m(g)$ jest liczbą orbit działania grupy $\langle g \rangle$ w zbiorze X.

PRZYKŁAD 1

Najpierw kolorujemy wierzchołki grafu z rysunku 12.12 z § 12.2 i rysunku 12.15 z § 12.3. Mamy $|G| = 8$, a liczby $m(g)$ podane są w tablicy 12.1 z § 12.2. Wynoszą one 6, 5, 5, 4, 3, 2, 2, 3, a więc na mocy twierdzenia 2 mamy

$$\begin{aligned} C(k) &= \frac{1}{8}(k^6 + k^5 + k^5 + k^4 + k^3 + k^2 + k^2 + k^3) \\ &= \frac{1}{8}(k^6 + 2k^5 + +k^4 + 2k^3 + 2k^2). \end{aligned}$$

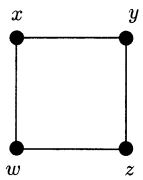
Kilka pierwszych wartości $C(k)$ podanych jest w tablicy 12.2. Problem ten dla $k = 2$ rozpatrywany był w przykładzie 5 z § 12.3, gdzie okazał się dość trudny do bezpośredniego rozwiązania. Dla $k \geq 3$ zagadnienie to wydaje się niemożliwe do rozwiązania bez pomocy rozwiniętej właśnie przez nas teorii.

Tablica 12.2

k	$C(k)$
1	1
2	21
3	171
4	820
5	2850
6	8001
7	19306

PRZYKŁAD 2

Zanim powrócimy do kolorowania sześcianu, pokolorujemy wierzchołki kwadratu za pomocą k kolorów. Dwa kolorowania uważamy za jednakowe, jeśli możemy jedno z nich otrzymać z drugiego za pomocą odpowiedniego obrotu bądź odwrócenia kwa-



(a)

	x	y	z	w	Cykle
e	x	y	z	w	e
r	y	z	w	x	$(x\,y\,z\,w)$
r^2	z	w	x	y	$(x\,z)(y\,w)$
r^3	w	x	y	z	$(x\,w\,z\,y)$
h	w	z	y	x	$(x\,w)(y\,z)$
v	y	x	w	z	$(x\,y)(z\,w)$
d	x	w	z	y	$(y\,w)$
f	z	y	x	w	$(x\,z)$

(b)

Rysunek 12.19

dratu na drugą stronę. Rysunek 12.19(a) przedstawia nasz kwadrat, a na rysunku 12.19(b) wypisane są elementy grupy permutacji zbioru wierzchołków, o którą chodzi w tym zagadnieniu. W tablicy 12.3 wypisane są orbity podgrup tej grupy, generowanych przez pojedyncze permutacje; stanowi to rozwiązanie części (b) i (c) ćwiczenia 5 z § 12.2.

Tablica 12.3

$\langle e\rangle$	$\{x\}$, $\{y\}$, $\{z\}$, $\{w\}$	$m(e)=4$
$\langle r\rangle$	$\{x,\ y,\ z,\ w\}$	$m(r)=1$
$\langle r^2\rangle$	$\{x,\ z\}$, $\{y,\ w\}$	$m(r^2)=2$
$\langle r^3\rangle$	$\{x,\ y,\ z,\ w\}$	$m(r^3)=1$
$\langle h\rangle$	$\{x,\ w\}$, $\{y,\ z\}$	$m(h)=2$
$\langle v\rangle$	$\{x,\ y\}$, $\{w,\ z\}$	$m(v)=2$
$\langle d\rangle$	$\{x\}$, $\{z\}$, $\{y,\ w\}$	$m(d)=3$
$\langle f\rangle$	$\{x,\ z\}$, $\{y\}$, $\{w\}$	$m(f)=3$

Można sprawdzić, że $\langle e\rangle = \{e\}$, $\langle r\rangle = \langle r^3\rangle = \{e, r, r^2, r^3\}$, $\langle r^2\rangle = \{e, r^2\}$, $\langle h\rangle = \{e, h\}$, $\langle v\rangle = \{e, v\}$, $\langle d\rangle = \{e, d\}$ i $\langle f\rangle = \{e, f\}$. I tak na przykład, ponieważ $\langle f\rangle = \{e, f\}$, to orbitami działania grupy $\langle f\rangle$ są zbiory $\{e(s), f(s)\}$ dla s ze zbioru $\{x, y, z, w\}$, a więc są to zbiory $\{e(x), f(x)\} = \{x, z\}$, $\{e(y), f(y)\} = \{y\}$, $\{e(z), f(z)\} = \{z, x\}$ i $\{e(w), f(w)\} = \{w\}$. Są tylko trzy różne orbity, zatem $m(f) = 3$. Orbity te można także odczytać z rozkładu permutacji na cykle, przy czym nie należy zapomnieć o orbitach długości 1.

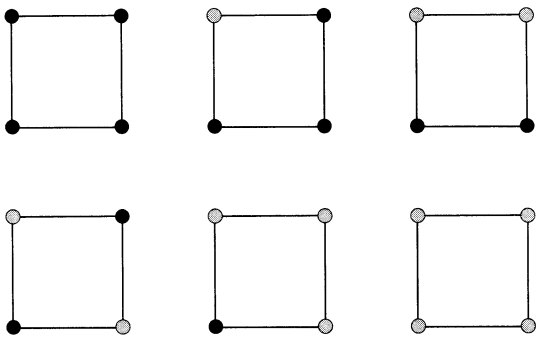
Zgodnie z twierdzeniem 2 istnieje

$$C(k) = \frac{1}{8}(k^4 + k + k^2 + k + k^2 + k^2 + k^3 + k^3)$$
$$= \frac{1}{8}(k^4 + 2k^3 + 3k^2 + 2k)$$

różnych sposobów pokolorowania wierzchołków naszego kwadratu przy użyciu k kolorów. Dla $k = 1$ liczba ta wynosi oczywiście 1. W tabeli z rysunku 12.20(a) podane są liczby możliwych kolorowań dla kilku pierwszych wartości k. Rysunek 12.20(b) przedstawia wszystkie, tzn. sześć, możliwości w przypadku 2 kolorów, wliczając w to oba kolorowania z użyciem jednego koloru. ■

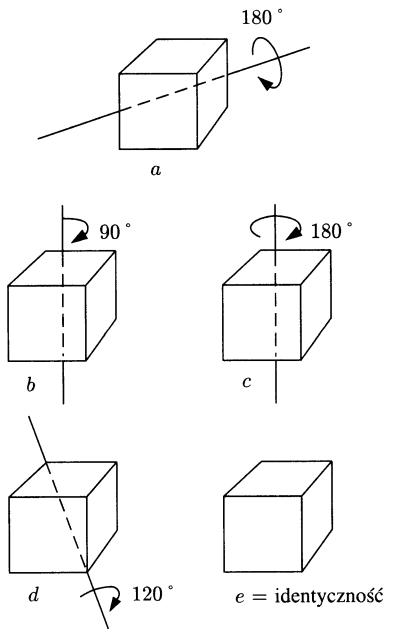
k	Liczba możliwych kolorowań
1	1
2	6
3	21
4	55
5	120
6	231
7	406

(a)



(b)

Rysunek 12.20



(a)

Typ	Liczba obrotów danego typu	$m(g)$
a	6	3
b	6	3
c	3	4
d	8	2
e	1	6

(b)

k	$C(k)$
1	1
2	10
3	57
4	240
5	800
6	2226

(c)

Rysunek 12.21

PRZYKŁAD 3 Pokolorujmy teraz ściany sześcianu, rozpatrywanego na początku tego paragrafu. Są 24 obroty, które przekształcają ten sześcian z powrotem na siebie. Wypisanie ich wszystkich wymagałoby znacznego nakładu pracy, ale w rzeczywistości musimy jedynie znać długości wyznaczonych przez nie orbit, a do tego wystarczy zająć się obrotami każdego z pięciu typów, które pokazuje rysunek 12.21(a). Z tabeli (b) z rysunku 12.21 dowiadujemy się, ile jest obrotów każdego typu. Podane są w niej również wartości $m(g)$, gdzie $m(g)$ jest liczbą orbit działania na zbiorze wszystkich ścian sześcianu grupy cyklicznej generowanej przez dany obrót g. Weźmy, na przykład, obrót o 90^0 typu b. Każdy taki obrót odbywa się wokół osi przechodzącej przez środki przeciwległych ścian sześcianu. Jest 6 ścian, a więc są 3 pary ścian przeciwległych, a stąd 3 takie osie. W tabeli (b) wypisanych jest 6 obrotów typu b: po dwa obroty o 90^0 wokół każdej z osi, po jednym w każdą stronę. Każdy obrót g typu b daje 3 $\langle g\rangle$-orbity: każda ze ścian, przez które przechodzi oś obrotu, tworzy orbitę długości 1, a zbiór pozostałych czterech ścian jest orbitą długości 4. Wynika stąd, że $m(g) = 3$. Pozostałe wartości z tabeli (b) zostały znalezione w drodze podobnego rozumowania. Twierdzenie 2 daje następujący wzór na liczbę kolorowań ścian za pomocą k kolorów:

$$C(k) = \frac{1}{24}(6k^3 + 6k^3 + 3k^4 + 8k^2 + k^6).$$

W tabeli (c) na rysunku 12.21 wypisanych jest kilka pierwszych wartości funkcji $C(k)$. ∎

Twierdzenie 2 podaje liczbę kolorowań za pomocą co najwyżej k kolorów. W niektórych zastosowaniach chcemy znać liczbę kolorowań przy użyciu dokładnie k kolorów. Przyda się do tego zasada włączeń i wyłączeń z § 5.3.

PRZYKŁAD 4 Znajdźmy liczbę (nierównoważnych) kolorowań wierzchołków kwadratu z przykładu 2 za pomocą dokładnie czterech danych kolorów, którymi są, powiedzmy, kolory czerwony, niebieski, zielony i żółty. W tabeli (a) z rysunku 12.20 podana jest liczba $C(k)$ sposobów pokolorowania przy użyciu co najwyżej k kolorów. Dla $i = 1, 2, 3, 4$ niech A_i będzie zbiorem wszystkich kolorowań, w których nie jest użyty kolor i. Wówczas $A_1 \cup A_2 \cup A_3 \cup A_4$ jest zbiorem tych kolorowań, które wykorzystują trzy lub mniej spośród danych kolorów, a szukaną przez nas odpowiedzią jest $C(4) - |A_1 \cup A_2 \cup A_3 \cup A_4|$. Teraz na mocy zasady włączeń i wy-

łączeń otrzymujemy

$$\begin{aligned}|A_1 \cup A_2 \cup A_3 \cup A_4| &= |A_1| + |A_2| + |A_3| + |A_4| - \{|A_1 \cap A_2| + |A_1 \cap A_3| \\ &\quad + |A_1 \cap A_4| + |A_2 \cap A_3| + |A_2 \cap A_4| + |A_3 \cap A_4|\} \\ &\quad + \{|A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + |A_1 \cap A_3 \cap A_4| \\ &\quad + |A_2 \cap A_3 \cap A_4|\} - |A_1 \cap A_2 \cap A_3 \cap A_4|.\end{aligned}$$

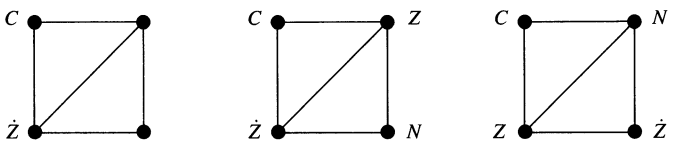
Zbior A_1 składa się ze wszystkich kolorowań za pomocą kolorów niebieskiego, zielonego lub żółtego, a więc $|A_1|$ to $C(3)$. Analogicznie, to samo jest prawdą dla A_2, A_3 i A_4. Zbiór $A_1 \cap A_2$ składa się z kolorowań za pomocą kolorów zielonego i żółtego, a więc ma $C(2)$ elementów; analogiczna obserwacja dotyczy każdego przecięcia dwóch spośród naszych zbiorów. Przecięcia takie jak $A_1 \cap A_2 \cap A_3$ mają po jednym kolorowaniu, tzn. mają po $C(1)$ elementów. W końcu, $A_1 \cap A_2 \cap A_3 \cap A_4$ jest zbiorem pustym. Wyciągamy stąd wniosek, że

$$|A_1 \cup A_2 \cup A_3 \cup A_4| = 4C(3) - 6C(2) + 4C(1),$$

a więc liczba kolorowań przy użyciu dokładnie czterech danych kolorów wynosi

$$C(4) - 4C(3) + 6C(2) - 4C(1) = 55 - 4 \cdot 21 + 6 \cdot 6 - 4 \cdot 1 = 3.$$

Teraz, kiedy znamy już odpowiedź, łatwo możemy przedstawić na rysunku 12.22 wszystkie trzy różne kolorowania, których szukaliśmy. ■

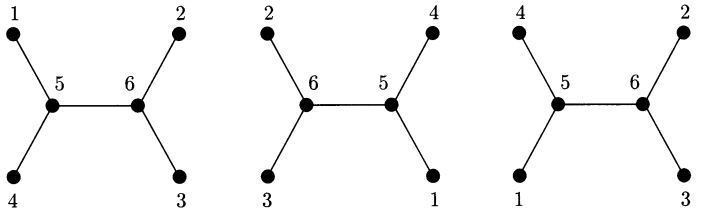


Rysunek 12.22

Następny przykład wykorzystuje dokładnie te same pomysły, co przykład 4, ale szczegóły są bardziej skomplikowane.

PRZYKŁAD 5 Rozważmy teraz problem przypisywania różnych etykiet 1, 2, 3, 4, 5, 6 wierzchołkom grafu omawianego ostatnio w przykładzie 1, co ilustruje rysunek 12.23. Dwa takie przyporządkowania

są uważane za identyczne, jeśli istnieje automorfizm grafu, przeprowadzający jedno z nich na drugie. Tak więc wszystkie sposoby etykietowania, przedstawione na rysunku 12.23 uważamy za takie same.



Rysunek 12.23

Możemy to zagadnienie potraktować jak problem kolorowania z użyciem dokładnie 6 kolorów. W przykładzie 1 znaleźliśmy już liczbę $C(k)$ sposobów pokolorowania wierzchołków naszego grafu przy użyciu co najwyżej k kolorów. Tak jak w przykładzie 4, dla $i = 1, 2, \ldots, 6$ niech A_i będzie zbiorem wszystkich kolorowań, w których nie jest użyty kolor i. Wówczas $A_1 \cup A_2 \cup \ldots \cup A_6$ jest zbiorem kolorowań z użyciem 5 lub mniej spośród danych 6 kolorów i odpowiedzią na nasze pytanie jest liczba $C(6) - |A_1 \cup A_2 \cup \ldots \cup A_6|$. Zasada włączeń i wyłączeń daje wzór na $|A_1 \cup A_2 \cup \ldots \cup A_6|$, a mianowicie

$$\sum_{i=1}^{6} |A_i| - \sum_{1 \le i < j \le 6} |A_i \cap A_j| + \sum_{1 \le i < j < k \le 6} |A_i \cap A_j \cap A_k| - \ldots,$$

gdzie, na przykład, w trzeciej sumie dodane są do siebie liczby elementów wszystkich przecięć trzech różnych zbiorów spośród $A_1, \ldots, A_6$. Teraz

$$\begin{aligned} |A_i| &= C(5) \quad \text{dla każdego } i, \\ |A_i \cap A_j| &= |\{\text{kolorowania bez użycia kolorów } i \text{ oraz } j\}| \\ &= C(4) \quad \text{dla} \quad i < j, \\ |A_i \cap A_j \cap A_k| &= C(3) \quad \text{dla} \quad i < j < k \end{aligned}$$

itd. Wynika stąd, że

$$\begin{aligned} &|A_1 \cup A_2 \cup \ldots \cup A_6| \\ &\quad = \tbinom{6}{1}C(5) - \tbinom{6}{2}C(4) + \tbinom{6}{3}C(3) - \tbinom{6}{4}C(2) + \tbinom{6}{5}C(1), \end{aligned}$$

a więc liczba kolorowań przy użyciu dokładnie 6 kolorów wynosi

$$\begin{aligned} &C(6) - \tbinom{6}{1}C(5) + \tbinom{6}{2}C(4) - \tbinom{6}{3}C(3) + \tbinom{6}{4}C(2) - \tbinom{6}{5}C(1) \\ &\quad = 8001 - 6 \cdot 2850 + 15 \cdot 820 - 20 \cdot 171 + 15 \cdot 21 - 6 \cdot 1 = 90. \end{aligned}$$

Podobnie, liczba kolorowań przy użyciu dokładnie dwóch kolorów wynosi po prostu $C(2)-\binom{2}{1}C(1) = 21-2 = 19$, a liczba kolorowań z użyciem dokładnie siedmiu kolorów jest równa

$$19306-7\cdot 8001+21\cdot 2850-35\cdot 820+35\cdot 171-21\cdot 21+7\cdot 1 = 0. \blacksquare$$

Wróćmy teraz do problemu tworzenia układów logicznych, na początek dla dwóch tylko wejść. Jest $2^4 = 16$ funkcji booleowskich dwóch zmiennych, tzn. elementów zbioru $\text{FUN}(\mathbb{B}\times\mathbb{B},\mathbb{B})$, gdzie $\mathbb{B} = \{0,1\}$. Możemy myśleć o układzie logicznym jako o czarnej skrzynce z dwoma przewodami wejściowymi, po jednym dla każdej ze zmiennych x_1 i x_2, oraz jednym przewodem wyjściowym. Każdy element (a_1,a_2) zbioru $\mathbb{B}\times\mathbb{B}$ odpowiada wyborowi wartości a_1 dla x_1 i a_2 dla x_2.

Zamienienie ze sobą przewodów dla x_1 i x_2 sprowadza się do zastąpienia (a_1,a_2) przez (a_2,a_1) i odpowiada permutacji g zbioru $\mathbb{B}\times\mathbb{B}$, która zamienia ze sobą $(0,1)$ z $(1,0)$. Dwa układy logiczne chcemy uważać za równoważne, jeśli dają one zawsze te same wyniki lub też jeśli będą dawać zawsze te same wyniki, o ile w jednym z nich zamienimy ze sobą przewody wejściowe. To znaczy, że dwie czarne skrzynki są równoważne, jeśli odpowiadające im funkcje booleowskie f i f' są bądź identyczne, bądź spełniają równość $f' = f\circ g$. Ponieważ $|\mathbb{B}| = 2$, to nasz problem wygląda dokładnie tak jak problem kolorowania za pomocą 2 kolorów czteroelementowego zbioru $\mathbb{B}\times\mathbb{B}$, w którym dwa elementy można zamieniać ze sobą miejscami. Stosujemy twierdzenie 1 do zbiorów $X = \mathbb{B}\times\mathbb{B}$, $K = \mathbb{B}$ i $G = \langle g\rangle = \{e,g\}$. Liczba G-orbit wynosi

$$\frac{1}{2}(2^4+2^3) = 12.$$

Prawdziwość tego wyniku potwierdza tablica z rysunku 12.24, w której wypisanych jest wszystkich szesnaście funkcji booleowskich z $\mathbb{B}\times\mathbb{B}$ w $\mathbb{B}$. Wartości funkcji 2 i 4 mogą być obliczane przy użyciu tej samej czarnej skrzynki i to samo ma miejsce dla par 3 i 5, 10 i 12 oraz 11 i 13, a więc liczba orbit wynosi $16-4 = 12$.

	Numery funkcji															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(0,0)	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
(0,1)	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
(1,0)	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
(1,1)	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

Rysunek 12.24

	(0,0)	(0,1)	(1,0)	(1,1)
c	(0,0)	(0,1)	(1,0)	(1,1)
c_1	(1,0)	(1,1)	(0,0)	(0,1)
c_2	(0,1)	(0,0)	(1,1)	(1,0)
$c_1 \circ c_2$	(1,1)	(1,0)	(0,1)	(0,0)
g	(0,0)	(1,0)	(0,1)	(1,1)
$c_1 \circ g$	(1,0)	(0,0)	(1,1)	(0,1)
$c_2 \circ g$	(0,1)	(1,1)	(0,0)	(1,0)
$c_1 \circ c_2 \circ g$	(1,1)	(0,1)	(1,0)	(0,0)

(a)

	x	y	w	z
e	x	y	w	z
h	w	z	x	y
v	y	x	z	w
r^2	z	w	y	x
d	x	w	y	z
r^3	w	x	z	y
r	y	z	x	w
f	z	y	w	x

(b)

Rysunek 12.25

Załóżmy teraz, że dopuszczamy zamianę wartości na wejściu na ich dopełnienia. Takie dopełnianie wartości na pierwszym przewodzie oznacza, że zamieniamy ze sobą pary $(0,0)$ i $(1,0)$ oraz $(0,1)$ i $(1,1)$. Tę permutację zbioru $\mathbb{B} \times \mathbb{B}$ oznaczmy przez c_1, a permutację odpowiadającą dopełnianiu drugiej z wartości na wejściu — przez c_2. Permutacje g, c_1 i c_2 generują wspólnie grupę G permutacji zbioru $\mathbb{B} \times \mathbb{B}$, opisaną na rysunku 12.25(a). (Nie oczekujemy, że fakt ten wyda się oczywisty; ręczymy jednak za niego.) Grupa ta działa na czteroelementowym zbiorze $\mathbb{B} \times \mathbb{B}$ w taki sam sposób, w jaki grupa z przykładu 2 działa na zbiorze wierzchołków kwadratu. Można się o tym przekonać porównując rysunek 12.25(a) z rysunkiem 12.25(b), który jest po prostu rysunkiem 12.19(b), zrobionym powtórnie w taki sposób, że niektóre wiersze i kolumny zostały zamienione miejscami. Odpowiedniości $(0,0) \to x$, $(0,1) \to y$, $(1,0) \to w$, $(1,1) \to z$ pozwalają z jednej z tablic z rysunku 12.25 uzyskać drugą. Wiemy z rysunku 12.20, że jest $C(2) = 6$ sposobów pokolorowania wierzchołków kwadratu za pomocą 2 kolorów, a więc istnieje sześć orbit działania grupy G w zbiorze $\text{FUN}(\mathbb{B}\times\mathbb{B}, \mathbb{B})$ naszych funkcji boolеowskich, tzn. sześć istotnie różnych czarnych skrzynek. Używając numerów funkcji z rysunku 12.24, orbity te można tak zapisać:

$$\{0\},\quad \{1,2,4,8\},\quad \{3,5,10,12\},\quad \{6,9\},\quad \{7,11,13,14\},\quad \{15\}.$$

Aby tworzyć układy logiczne, wystarczyłoby wybrać po jednej funkcji z każdej orbity, tzn. np. 0, 1, 3, 6, 7 i 15 i dla każdej z nich mieć układ, który oblicza jej wartości.

Jeśli dopuścimy również zastępowanie wartości na wyjściu układu przez ich dopełnienia, to każdy układ, który oblicza wartości funkcji o numerze n będzie też obliczał wartości funkcji o numerze $15 - n$ i liczba niezbędnych czarnych skrzynek jeszcze się zmniejszy. Układ dobry dla funkcji 0 będzie też dobry dla funk-

cji 15. Układ dla funkcji 1 będzie dobry dla funkcji 14, a więc również i dla funkcji 7, 11 lub 13. Układ dla funkcji 3 będzie też, o czym już wiemy, dobry dla funkcji 12 i podobnie, układ dla funkcji 6 będzie dobry dla funkcji 9. Klasy funkcji równoważnych wyglądają teraz tak:

$$\{0,15\}, \quad \{1,2,4,8,7,11,13,14\}, \quad \{3,5,10,12\}, \quad \{6,9\}.$$

Jeśli więc dopuszczamy dopełnianie wartości na przewodach wejściowych i wyjściowych, to ciągle jeszcze potrzebujemy czterech różnych układów, by móc obliczać wartości wszystkich dwuargumentowych funkcji booleowskich.

Nasze metody teoretycznie można uogólnić tak, by móc znaleźć liczbę niezbędnych czarnych skrzynek w przypadku funkcji booleowskich o n argumentach na wejściu. W praktyce, szczegółowe opisanie orbit wyznaczonych przez wszystkie elementy odpowiedniej grupy G staje się niesłychanie skomplikowane. Dla funkcji czteroargumentowych odpowiedź brzmi, że nawet jeśli dopuścimy branie dopełnień na wejściu i wyjściu, to i tak potrzebne są 222 różne układy. Liczba ta jest znacząco mniejsza od $2^{16} = 65536$. Znajomość niezbędnej liczby układów nie pomaga w znajdowaniu reprezentantów poszczególnych klas układów równoważnych, ale mówi nam, w którym momencie znaleźliśmy ich już wystarczająco dużo.

Nasze metody nie korzystały w żaden uporządkowany sposób z symetrii samej grupy G. Wykorzystując takie symetrie uzyskać można wzór na liczbę wszystkich G-orbit w zbiorze $\text{FUN}(\mathbb{B}^n, \mathbb{B})$, których elementy przyjmują wartość 0 dla dokładnie k argumentów, gdzie $k = 1, 2, \ldots$. Istnieje bogata literatura dotycząca tematu wykorzystywania grup w zagadnieniach zliczania. Po więcej informacji w duchu tego paragrafu sięgnij do książek z algebry stosowanej szukając nazwisk Polyà i Burnside.

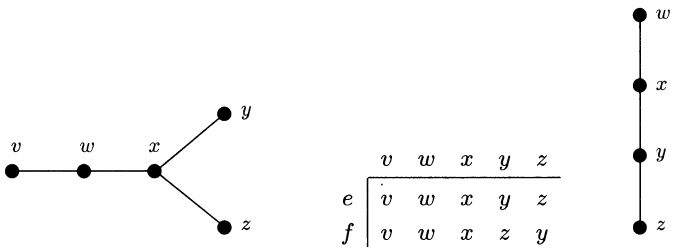
ĆWICZENIA DO § 12.4

1. Graf przedstawiony na rysunku 12.26(a) ma dwa automorfizmy, które opisane są na rysunku 12.26(b).

(a) Ile wynosi średnia arytmetyczna liczb wierzchołków, które są punktami stałymi poszczególnych automorfizmów tego grafu?

(b) Dla których automorfizmów grafu liczba punktów stałych jest równa średniej wartości z części (a)?

(c) Znajdź liczbę sposobów pokolorowania wierzchołków tego grafu za pomocą k kolorów.



(a) (b) (c)

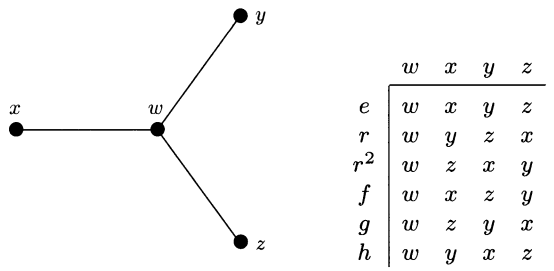
Rysunek 12.26

2. Sprawdź prawdziwość twierdzenia 1 z § 12.3 dla
   (a) grupy automorfizmów grafu z rysunku 12.26(a) działającej na zbiorze wierzchołków tego grafu.
   (b) grupy z części (a) działającej na zbiorze krawędzi grafu z rysunku 12.26(a).
   (c) grupy obrotów z przykładu 3, działającej na zbiorze ścian danego sześcianu.

3. Graf z rysunku 12.26(c) ma dwa automorfizmy.
   (a) Ile jest sposobów pokolorowania wierzchołków tego grafu przy pomocy k kolorów?
   (b) Ile jest sposobów przypisania wierzchołkom tego grafu czterech różnych etykiet?

4. Znajdź $g \circ c_1 \circ g$ dla elementów grupy opisanej w tablicy na rysunku 12.25(a). Z Twojego rozwiązania wyniknie, że tablicę tę można utworzyć używając jedynie permutacji c_1 i g.

5. Graf przedstawiony na rysunku 12.27(a) ma sześć automorfizmów, które opisane są na rysunku 12.27(b).



(a) (b)

Rysunek 12.27

   (a) Znajdź liczbę sposobów pokolorowania wierzchołków tego grafu za pomocą k kolorów.

(b) Znajdź liczbę sposobów pokolorowania krawędzi tego grafu za pomocą k kolorów.

6. (a) Wykorzystaj zasadę włączeń i wyłączeń oraz rozwiązanie ćwiczenia 5(a) do znalezienia liczby sposobów pokolorowania wierzchołków grafu z rysunku 12.26(a) za pomocą dokładnie trzech kolorów.
(b) Opisz wszystkie różne kolorowania z części (a) używając kolorów czerwonego, niebieskiego i zielonego.
(c) Znajdź liczbę sposobów pokolorowania wierzchołków naszego grafu za pomocą dokładnie czterech kolorów.

7. (a) Ile jest sposobów pokolorowania wierzchołków kwadratu z przykładu 2 przy użyciu dokładnie czterech kolorów?
(b) Wypisz wszystkie różne kolorowania z użyciem wszystkich czterech kolorów: czerwonego, niebieskiego, zielonego i żółtego.

8. Ile różnych naszyjników można zrobić z pięciu paciorków, mając do dyspozycji paciorki w k różnych kolorach? Dwa naszyjniki uznajemy za identyczne, jeśli jeden z nich w wyniku obracania lub odwracania na drugą stronę wygląda tak, jak drugi. *Wskazówka*: Odpowiednia grupa składa się w tym przypadku z e, czterech nietrywialnych obrotów i pięciu odwróceń na drugą stronę. Zobacz przykład 2, gdzie omówiony jest przypadek czterech paciorków.

9. Znajdź liczbę kolorowań wierzchołków kwadratu z przykładu 2 z użyciem
(a) dokładnie trzech kolorów,
(b) dokładnie dwóch kolorów,
(c) dokładnie pięciu kolorów.

10. Wykonaj rysunki przedstawiające różne kolorowania każdego z typów opisanych w ćwiczeniu 9.

11. Weźmy graf i grupę G działającą na zbiorze $V = \{w, x, y, z\}$ z ćwiczenia 3 z § 12.3.
(a) Dla każdego elementu g grupy G znajdź liczbę $\langle g \rangle$-orbit w zbiorze V.
(b) Znajdź liczbę sposobów pokolorowania wierzchołków tego grafu za pomocą co najwyżej k kolorów.
(c) Wykorzystaj rozwiązanie części (b) do znalezienia liczby sposobów pokolorowania wierzchołków za pomocą czerwonego, czarnego lub obu tych kolorów. Przedstaw na rysunku kolorowania reprezentujące każdą z klas równoważnych kolorowań.

12. Grupa G z ćwiczenia 11 działa również na zbiorze $E = \{e_1, e_2, e_3, e_4, e_5\}$ krawędzi rozważanego grafu; zob. ćwiczenie 4, § 12.3.
(a) Dla każdego elementu g grupy G znajdź liczbę $\langle g \rangle$-orbit w zbiorze E.
(b) Znajdź liczbę sposobów pokolorowania krawędzi tego grafu za pomocą co najwyżej k kolorów.

(c) Ile jest sposobów pokolorowania krawędzi tego grafu za pomocą dokładnie dwóch kolorów?

13. (a) Ile jest sposobów pokolorowania wierzchołków sześcianu za pomocą k kolorów?

(b) Ile jest sposobów pokolorowania krawędzi sześcianu za pomocą k kolorów?

W ćwiczeniach (a) i (b) dwa kolorowania uważa się za jednakowe, jeśli jedno z nich można uzyskać z drugiego w wyniku pewnego obrotu sześcianu. *Sugestia*: Wykorzystaj rysunek 12.21(a) do znalezienia nowych tablic, takich jak ta na rysunku 12.21(b), opisujących działanie danej grupy na zbiorze wierzchołków bądź krawędzi sześcianu.

14. Rozważmy problem kolorowania ścian sześcianu za pomocą czerwonych, zielonych i niebieskich kredek, tak jak w przykładzie 3.

(a) Ile jest różnych kolorowań wykorzystujących dokładnie dwa spośród trzech danych kolorów?

(b) Stosując dowolną metodę znajdź liczbę takich kolorowań, w wyniku których cztery ściany są czerwone, a dwie niebieskie.

(c) Ile jest kolorowań, w wyniku których dokładnie cztery ściany są czerwone?

(d) Czy miałbyś ochotę szukać liczby kolorowań, w wyniku których są po dokładnie dwie ściany każdego z kolorów, drogą badania kolejnych przypadków?

15. (a) Ile istnieje różnych układów logicznych o dwóch argumentach na wejściu, jeśli dwa układy uważamy za takie same jedynie wówczas, gdy wartości przyjmowane na wyjściu są w obu przypadkach dla wszystkich par argumentów takie same bądź dla wszystkich par argumentów są swoimi dopełnieniami?

(b) Ile ich jest, jeśli dodatkowo uznajemy dwa układy za takie same wtedy, gdy dają tę samą funkcję w wyniku zamiany miejscami przewodów wejściowych w jednym z nich?

§ 12.5. Grupy

Paragrafy od 12.1 do 12.4 zawierają wstęp do teorii grup permutacji. Grupy pojawiają się jednak również w wielu różnorodnych sytuacjach, które z permutacjami nie wiążą się w żaden oczywisty sposób. A oto ogólna definicja. **Grupa** $(G, \cdot)$ jest to zbiór G wraz z dwuargumentowym działaniem $\cdot$ (tzn. funkcją $(g, h) \longrightarrow g \cdot h$ ze zbioru $G \times G$ w zbiór G), która ma następujące własności:

(i) $(g \cdot h) \cdot k = g \cdot (h \cdot k)$ dla wszystkich $g, h, k \in G$;
[**prawo łączności**]

(ii) do G należy element e, nazywany **elementem neutralnym**, taki, że $g \cdot e = e \cdot g = g$ dla wszystkich $g \in G$;
(iii) każdemu elementowi g z G odpowiada element g^{-1}, nazywany **elementem doń odwrotnym**, taki, że $g \cdot g^{-1} = g^{-1} \cdot g = e$.

Mówiąc o dowolnej grupie, dla oznaczenia dwuargumentowego działania z powyższej definicji używać będziemy symboli $\cdot$ i $\bullet$, ale czasem też będziemy te symbole w ogóle pomijać, tzn. będziemy pisać $g \cdot h$, $g \bullet h$ lub po prostu gh i taki element nazywać będziemy „iloczynem" elementów g i h. Elementy odwrotne są wyznaczone jednoznacznie w następującym silnym sensie. Jeśli $h \cdot g = e$, to

$$h = h \cdot e = h \cdot (g \cdot g^{-1}) = (h \cdot g) \cdot g^{-1} = e \cdot g^{-1} = g^{-1}.$$

Analogicznie, równość $g \cdot h = e$ implikuje, że element h musi być równy g^{-1}. Zatem dla sprawdzenia, że element h jest równy g^{-1}, wystarczy pokazać, że zachodzi którykolwiek z warunków, $h \cdot g = e$ bądź $g \cdot h = e$. Wtedy drugi z nich będzie też spełniony.

Niepusty podzbiór H grupy G jest automatycznie grupą, o ile jest **zamknięty** ze względu na dane działanie w G oraz branie elementów odwrotnych: jeśli $g, h \in H$, to $g \cdot h \in H$ i $g^{-1} \in H$. Podzbiór taki nazywamy **podgrupą grupy** G. Podgrupy oznaczać będziemy zwykle przez H, J lub K.

PRZYKŁAD 1 (a) Grupa symetryczna zbioru n-elementowego, S_n, omawiana w § 12.1 jest grupą, w której dwuargumentowym działaniem jest składanie permutacji, $\circ$. Nawiasem mówiąc, zbiór $\mathrm{PERM}(X)$ wszystkich permutacji jakiegokolwiek zbioru X jest grupą ze względu na składanie $\circ$. Jak widzieliśmy w poprzednich paragrafach, podgrupy grupy $\mathrm{PERM}(X)$ są bogatym źródłem przykładów.

(b) Zbiór $\mathrm{GL}(n)$ składający się ze wszystkich odwracalnych macierzy wymiaru $n \times n$ jest grupą ze względu na zwykłe mnożenie macierzy; zob. § 3.4. Zauważ, że iloczyn macierzy odwracalnych jest macierzą odwracalną; w istocie $(AB)^{-1} = B^{-1}A^{-1}$. Na pierwszy rzut oka grupa ta wygląda całkiem inaczej niż grupy, w których działaniem jest składanie, ale istnieje tu jednak pewien związek. W algebrze liniowej pokazuje się, że macierze wymiaru $n \times n$ są w odpowiedniości wzajemnie jednoznacznej z przekształceniami liniowymi przestrzeni $\mathbb{R}^n$, a iloczyn macierzy odpowiada składaniu przekształceń liniowych. ∎

W pewnym sensie przykład 1(a) zawiera w sobie wszystkie istniejące przykłady. Dwie grupy są **izomorficzne**, jeśli ist-

nieje wzajemnie jednoznaczne przekształcenie (nazywane **izomorfizmem**) jednej z nich na drugą, które zachowuje działania grupowe. Używając symboli można napisać, że jeśli dane są grupy $(G_1, \cdot)$ i $(G_2, \bullet)$, to przekształcenie wzajemnie jednoznaczne φ zbioru G_1 na zbiór G_2 jest izomorfizmem, o ile $\varphi(g \cdot h) = \varphi(g) \bullet \varphi(h)$ dla wszystkich $g, h \in G$. Grupy izomorficzne to w gruncie rzeczy te same grupy; ich podstawowa struktura jest identyczna. Twierdzenie, znane jako twierdzenie Cayleya, mówi, że każda grupa jest izomorficzna z podgrupą grupy permutacji pewnego zbioru; w rzeczywistości jako zbiór ten wziąć można samą grupę G. (Zob. ćwiczenie 24.) Nie zważając na twierdzenie Cayleya, wiele grup będziemy badać w ramach określonych przez ich naturalne definicje, w których grupy permutacji nie będą występować.

Grupa G jest **przemienna**, jeśli $g \cdot h = h \cdot g$ dla wszystkich $g, h \in G$. Dla grup przemiennych stosuje się często notację addytywną, oznaczając działanie grupowe przez $+$, ponieważ notacja taka jest tradycyjnie używana w przypadku szeregu ważnych grup przemiennych. **Element neutralny** działania $+$ oznacza się zwykle przez 0. Element odwrotny do elementu g ze względu na działanie $+$ nazywamy **elementem przeciwnym** do g i piszemy $-g$ zamiast g^{-1}. Zatem w grupie przemiennej $(G, +)$ określone jest dwuargumentowe działanie $(g, h) \to g+h$, które spełnia następujące warunki:

(i) $(g + h) + k = g + (h + k)$ dla wszystkich $g, h, k \in G$; [**prawo łączności**]

(ii) do G należy element 0, nazywany **elementem zerowym** lub **zerem**, taki, że $g + 0 = g$ dla wszystkich $g \in G$;

(iii) każdemu elementowi g z G odpowiada element przeciwny $-g$ taki, że $g + (-g) = 0$;

(iv) $g + h = h + g$ dla wszystkich $g, h \in G$. [**prawo przemienności**]

PRZYKŁAD 2 (a) $(\mathbb{R}, +)$ jest grupą przemienną. Elementem neutralnym dla dodawania jest oczywiście 0, a prawa od (i) do (iv) są dobrze znane.

(b) $(\mathbb{Z}, +)$ jest także grupą przemienną; jest to podgrupa grupy $(\mathbb{R}, +)$.

(c) Zbiór $\mathfrak{M}_{m,n}$ wszystkich macierzy wymiaru $m \times n$ jest grupą przemienną ze względu na dodawanie. Fakt ten sformułowany jest w twierdzeniu z § 3.3.

(d) Zbiór $\mathbb{R}$ nie jest grupą ze względu na mnożenie $\cdot$ mimo że spełnia warunki (i), (ii) i (iv). Kłopot w tym, że nie ma elementu

odwrotnego do liczby 0. Ponieważ iloczyn niezerowych liczb rzeczywistych jest różny od zera, to $\cdot$ jest dwuargumentowym działaniem w zbiorze $\mathbb{R}\setminus\{0\}$, a więc $(\mathbb{R}\setminus\{0\},\cdot)$ jest grupą. Jest to przykład grupy przemiennej, której działania grupowego nie oznacza się symbolem $+$.

(e) Zbiór $\mathbb{N}$ nie jest grupą ze względu na dodawanie, ponieważ z faktu, że $n \in \mathbb{N}$ nie wynika, iż $-n \in \mathbb{N}$. W rzeczywistości, jeśli zarówno liczba n, jak i $-n$ należą do $\mathbb{N}$, to $n = 0$.

(f) Zbiór $\mathfrak{M}_{n,n}$ macierzy kwadratowych nie jest grupą ze względu na mnożenie, ponieważ wiele macierzy, tak jak macierz zerowa, nie ma elementów odwrotnych. Pewien pożyteczny podzbiór $\mathfrak{M}_{n,n}$, który jest grupą ze względu na mnożenie, został wprowadzony w przykładzie 1(b). ∎

Teoria grup jest dziedziną zdumiewająco bogatą, jeśli się zważy, że wszystkie jej rezultaty opierają się na trzech sformułowanych powyżej prawach od (i) do (iii). My dotkniemy jedynie jej powierzchni. Zaczniemy od pewnych łatwych konsekwencji wspomnianych praw. Zauważmy, że każda równość w przeprowadzonym dowodzie wynika z jednego z nich.

Twierdzenie 1

Niech $(G,\cdot)$ będzie grupą.

(a) Jeśli $g\cdot h = g\cdot k$ lub $h\cdot g = k\cdot g$, gdzie $g,h,k \in G$, to $h = k$. (To są tzw. **prawa skracania równości**.)

(b) $(g\cdot h)^{-1} = h^{-1}\cdot g^{-1}$ dla wszystkich $g,h \in G$.

Dowód. (a) Jeśli $g\cdot h = g\cdot k$, to $h = e\cdot h = (g^{-1}\cdot g)\cdot h = g^{-1}\cdot(g\cdot h) = g^{-1}\cdot(g\cdot k) = (g^{-1}\cdot g)\cdot k = e\cdot k = k$. Analogicznie, jeśli $h\cdot g = k\cdot g$, to $h = k$.

(b) Ponieważ element odwrotny do $g\cdot h$ jest wyznaczony jednoznacznie, więc wystarczy dowieść, że dla elementu $h^{-1}\cdot g^{-1}$ spełniona jest równość $(g\cdot h)\cdot(h^{-1}\cdot g^{-1}) = e$. Ale $(g\cdot h)\cdot(h^{-1}\cdot g^{-1}) = g\cdot(h\cdot(h^{-1}\cdot g^{-1})) = g\cdot((h\cdot h^{-1})\cdot g^{-1}) = g\cdot(e\cdot g^{-1}) = g\cdot g^{-1} = e$. ∎

Tak jak w przypadku grup permutacji, dla $g \in G$ przez $\langle g\rangle$ oznaczać będziemy grupę $\{g^n\colon n \in \mathbb{Z}\}$ generowaną przez element g. Jak zwykle, $g^0 = e$. Dla dodatnich liczb n, symbol g^n reprezentuje n-krotny iloczyn elementu g przez siebie, a g^{-n} oznacza $(g^{-1})^n$. Przy tych oznaczeniach można wykazać (ćwiczenie 14), że dla wszystkich $m,n \in \mathbb{Z}$ spełnione jest znajome prawo $g^{m+n} = g^m\cdot g^n$. Wynika stąd natychmiast, że powyżej zdefi-

niowany zbiór $\langle g \rangle$ rzeczywiście jest podgrupą grupy G. Grupy postaci $\langle g \rangle$ nazywane są często **grupami cyklicznymi**. Ponieważ $g^m \cdot g^n = g^{m+n} = g^{n+m} = g^n \cdot g^m$, to grupy cykliczne są przemienne. Mówimy, że element g grupy G ma (skończony) **rząd** m, jeśli zbiór $\langle g \rangle$ ma m elementów; jeśli element g nie ma skończonego rzędu, to mówimy, że jest on **rzędu nieskończonego**. Zobaczymy dalej, że jeśli g ma rząd m, to $g^m = e$ i m jest najmniejszą dodatnią liczbą całkowitą o tej własności.

PRZYKŁAD 3 (a) Twierdzenie 2 z § 12.1 podaje przepis na znajdywanie rzędów permutacji należących do grupy S_n. Ćwiczenia od 11 do 16 z tamtego paragrafu polegają na określaniu rzędów danych permutacji.

(b) W paragrafie 3.6 zdefiniowaliśmy działania $+_p$ oraz $*_p$ w zbiorze $\mathbb{Z}_p = \{0, \ldots, p-1\}$. Twierdzenie 4 z § 3.6 pokazało, że oba te działania są łączne i przemienne. Ponadto, 0 jest elementem neutralnym dodawania, a 1 jest elementem neutralnym mnożenia w $\mathbb{Z}_p$. Każdy element m ma element przeciwny, a mianowicie $(-m) \bmod p$, gdyż na mocy twierdzenia 3 z § 3.6

$$\begin{aligned} m +_p (-m) \bmod p &= m \bmod p +_p (-m) \bmod p \\ &= (m + (-m)) \bmod p = 0 \bmod p = 0. \end{aligned}$$

Zatem $(\mathbb{Z}_p, +_p)$ jest grupą przemienną. Ponieważ element 1 jest jej generatorem, grupa ta jest cykliczna. (Nawiasem mówiąc, k jest generatorem grupy $\mathbb{Z}_p$ wtedy i tylko wtedy, gdy NWD $(k, p) = 1$.)

Choć system algebraiczny $(\mathbb{Z}_p, *_p)$ ma wiele przyjemnych własności, to nie jest jednak grupą; nie ma elementu m takiego, że $0 *_p m = 1$, a więc 0 nie ma elementu odwrotnego ze względu na mnożenie.

(c) Nieskończona grupa $(\mathbb{Z}, +)$ jst cykliczna; zarówno liczba 1, jak i liczba -1 są generatorami tej grupy. Wszystkie niezerowe elementy zbioru $\mathbb{Z}$ są rzędu nieskończonego. Rzeczywiście, jeśli $n \in \mathbb{Z}$, to grupą cykliczną generowaną przez n jest dokładnie zbiór $n\mathbb{Z} = \{nk\colon k \in \mathbb{Z}\}$, a jest on nieskończony, chyba, że $n = 0$. ∎

W następnym paragrafie zobaczymy, że grupa $(\mathbb{Z}, +)$ i grupy $(\mathbb{Z}_p, +_p)$ są jedynymi grupami cyklicznymi w tym sensie, że każda grupa cykliczna jest izomorficzna z jedną z nich. Świadczące o tym izomorfizmy określa się w naturalny sposób. Jeśli grupa $\langle g \rangle$ jest nieskończona, to funkcja $g^k \longrightarrow k$ przekształca zbiór $\langle g \rangle$ różnowartościowo na $\mathbb{Z}$ i oczywiście fakt, że $g^m \cdot g^n = g^{m+n} \longrightarrow m+n$

pokazuje, że przekształcenie to jest izomorfizmem. Jeśli grupa $\langle g\rangle$ ma p elementów, tzn. jeśli element g jest rzędu p, to funkcja $g^k \longrightarrow k \operatorname{MOD} p$ jest izomorfizmem grupy $\langle g\rangle$ na $\mathbb{Z}_p$. W tym przypadku wszystkie elementy $e, g, \ldots, g^{p-1}$ są różne, $g^p = e$ i $g^{-1} = g^{p-1}$.

Normalnie, aby sprawdzić, że jakiś niepusty podzbiór H danej grupy G jest jej podgrupą, musielibyśmy zbadać, czy zbiór H jest zamknięty zarówno ze względu na branie elementów odwrotnych, jak też i na mnożenie. Jeśli zbiór H jest skończony, a więc w szczególności, jeśli grupa G jest skończona, to następujący fakt pozwala nam zaoszczędzić czas, który poświęcilibyśmy na badanie elementów odwrotnych.

Twierdzenie 2

Jeśli niepusty skończony podzbiór H grupy G jest zamknięty ze względu na działanie grupowe, tzn., jeśli $g \cdot h \in H$, o ile elementy g i h należą do H, to H jest podgrupą grupy G.

Dowód. Mając dany element $g \in H$ chcemy pokazać, że $g^{-1} \in H$. Jeśli uwierzyłeś w nasze ostatnie stwierdzenie, że zbiór $\langle g\rangle$ jest dla pewnego p równy $\{e, g, g^2, \ldots, g^{p-1}\}$, to $g^{-1} = g^{p-1} \in H$, ponieważ element g^{p-1} jest iloczynem elementów z H.

A oto inny argument, przeznaczony dla sceptyków. Elementy $e, g, g^2, \ldots$ należą do skończonego zbioru H, a więc nie mogą być wszystkie różne. Istnieją zatem liczby całkowite $k, n \in \mathbb{P}$ takie, że $k < n$ i $g^k = g^n$. Stąd

$$e = g^k \cdot (g^{-1})^k = g^n \cdot (g^{-1})^k = g^{n-k}.$$

Jeśli $n - k = 1$, to $g^{-1} = e = g \in H$. W przeciwnym razie $n - k \geq 2$; stąd $g^{-1} = e \cdot g^{-1} = g^{n-k} \cdot g^{-1} = g^{n-k-1} \in H$. ∎

Podgrupy mogą być generowane przez podzbiory składające się z więcej niż jednego elementu. Niech A będzie niepustym podzbiorem grupy G. **Podgrupą generowaną** przez A nazywamy zbiór $\langle A\rangle$ złożony ze wszystkich iloczynów elementów należących do zbioru $A \cup A^{-1}$, gdzie $A^{-1} = \{g^{-1} : g \in A\}$. Można wykazać, że zbiór $\langle A\rangle$ jest podgrupą grupy G i że jest to najmniejsza podgrupa grupy G zawierająca zbiór A. Zbiór $\langle A\rangle$ można też zdefiniować rekurencyjnie w następujący sposób:

(P) $A \subseteq \langle A\rangle$;
(R_1) Jeśli $g, h \in \langle A\rangle$, to $g \cdot h \in \langle A\rangle$;
(R_2) Jeśli $g \in \langle A\rangle$, to $g^{-1} \in \langle A\rangle$.

Warunki (R_1) i (R_2) stwierdzają, że zbiór $\langle A\rangle$ jest zamknięty ze względu na operacje $\cdot$ i odwracania. Weźmy element $g \in A$. Wówczas $g \in \langle A\rangle$ na mocy (P), a więc $g^{-1} \in \langle A\rangle$ na mocy (R_2), skąd na mocy (R_1) element neutralny $e = g \cdot g^{-1}$ także należy do $\langle A\rangle$. Zatem zbiór $\langle A\rangle$ jest podgrupą grupy G.

PRZYKŁAD 4

(a) Rozważmy podgrupę $\langle\{4,-6\}\rangle$ grupy $\mathbb{Z}$. Na mocy (R_1) należy do niej liczba $4+(-6) = -2$, a więc także, na mocy (R_2), liczba $-(-2) = 2$. Wynika stąd, że elementami tej podgrupy są wszystkie całkowite wielokrotności liczby 2. Znaczy to, że $\langle\{4,-6\}\rangle \supseteq \langle\{2\}\rangle = 2\mathbb{Z}$. Jednocześnie każda liczba, którą można otrzymać z liczb 4 i -6 za pomocą dodawania dwóch liczb bądź brania liczby przeciwnej do danej, będzie zawsze parzysta, a więc grupa $\langle\{4,-6\}\rangle$ składa się wyłącznie z liczb parzystych. Wynika stąd, że $\langle\{4,-6\}\rangle \subseteq 2\mathbb{Z}$, a więc $\langle\{4,-6\}\rangle = 2\mathbb{Z}$.

(b) Podgrupa $\langle\{3,5\}\rangle$ grupy $(\mathbb{Z},+)$ jest całym zbiorem $\mathbb{Z}$. Mianowicie, ponieważ wiemy, że do podgrupy $\langle\{3,5\}\rangle$ należy liczba $3+3-5=1$, to $\langle\{3,5\}\rangle \supseteq \langle 1\rangle = \mathbb{Z}$. Zauważmy, że żadna z liczb 3 lub 5 samodzielnie nie generuje $\mathbb{Z}$. ∎

W rzeczywistości okazuje się, że wszystkie podgrupy grupy $(\mathbb{Z},+)$ są cykliczne, co jest najprostszą możliwą sytuacją.

Twierdzenie 3

Każda podgrupa grupy $(\mathbb{Z},+)$ jest postaci $n\mathbb{Z}$ dla pewnej liczby $n \in \mathbb{N}$.

Dowód. Weźmy podgrupę H grupy $(\mathbb{Z},+)$. Jeśli $H = \{0\}$, to $H = 0\mathbb{Z}$, a więc jest żądanej postaci. Załóżmy, że $H \neq \{0\}$. Jeśli $0 \neq m \in H$, to również $-m \in H$. Stąd wynika, że zbiór $H \cap \mathbb{P}$ jest niepusty, a więc ma element najmniejszy n. Pokażemy, że $H = n\mathbb{Z}$. Ponieważ $n \in H$ i H jest podgrupą, to mamy $n\mathbb{Z} \subseteq H$. Weźmy element m z H. Na mocy algorytmu dzielenia $m = qn+r$, gdzie $0 \leq r < n$. Skoro $n\mathbb{Z} \subseteq H$, to $qn \in H$, a stąd mamy $r = m-nq \in H$. Ponieważ $r < n$ i n jest najmniejszym dodatnim elementem zbioru H, więc $r = 0$. To znaczy, że $m = qn \in n\mathbb{Z}$. Ponieważ liczba m była dowolnym elementem zbioru H, więc $H \subseteq n\mathbb{Z}$, jak twierdziliśmy. ∎

W ogólnym przypadku można się spodziewać, że podzbiór grupy, mający więcej niż jeden element, będzie generował podgrupę, która nie jest cykliczna.

PRZYKŁAD 5 (a) Dla dowolnej grupy G mamy $\langle G \rangle = G$. Aby jednak zbiór generujący grupę był użyteczny, powinien być względnie mały, a już z pewnością mniejszy od G.

(b) Paragraf 12.2 zawiera wiele przykładów grup o małym zadanym zbiorze generatorów. Grupa z rysunku 12.10 z tamtego paragrafu jest generowana przez zbiór $\{f, g\}$; jest ona przemienna, ale nie jest cykliczna. Grupa z rysunku 12.11 jest generowana przez zbiór $\{g, h\}$ i nawet nie jest przemienna. Grupa z rysunku 12.12 jest generowana przez zbiór $\{f, g, h\}$. W gruncie rzeczy jest ona generowana przez zbiór $\{f, g\}$, ponieważ $h = fgf$. Aby to stwierdzić, pokaż bezpośrednio, że $gf = fh$, a następnie zauważ, że $fgf = ffh = h$, gdyż $f^2 = e$. Grupa ta nie jest przemienna: przykładowo, $gf \neq fg$. ∎

Rozważmy podgrupę H grupy $(G, \cdot)$, której elementem neutralnym jest e. **Warstwą lewostronną** podgrupy H w grupie G nazywamy podzbiór postaci

$$g \cdot H = gH = \{gh\colon h \in H\}$$

dla pewnego elementu $g \in G$. Warstwa eH to sama podgrupa H, a ponadto $hH = H$ dla każdego elementu h z H (ćwiczenie 21). Ponieważ $e \in H$, to do warstwy gH należy element $ge = g$. Wynika stąd, że każdy element g grupy G należy do co najmniej jednej warstwy lewostronnej podgrupy H. W rzeczywistości, każdy element g należy do dokładnie jednej warstwy lewostronnej.

Twierdzenie 4

Zbiór wszystkich warstw lewostronnych dowolnej podgrupy danej grupy jest podziałem tej grupy.

Dowód. Pokazaliśmy przed chwilą, że zbiór G jest sumą warstw gH, a więc musimy jedynie pokazać, że warstwy, mające elementy wspólne, są identyczne. Pokażemy najpierw, że jeśli $k \in gH$, to $kH = gH$. Ponieważ $HH = H$, to mamy $kH \subseteq gHH = gH$. Jednocześnie, $k = gh$ dla pewnego $h \in H$, a więc $g = kh^{-1} \in kH$. Zatem $gH \subseteq kHH = kH$.

Załóżmy teraz, że warstwy gH i $g'H$ mają niepuste przecięcie; niech, powiedzmy, $k \in gH \cap g'H$. Wtedy z tego, co właśnie pokazaliśmy, wynika, że $gH = kH = g'H$. ∎

Pod koniec tego paragrafu omówimy alternatywny dowód twierdzenia 4.

Zamiast zajmować się warstwami lewostronnymi moglibyśmy równie dobrze badać **warstwy prawostronne**, tzn. zbiory po-

staci $Hg = \{hg\colon h \in H\}$. Jeśli grupa G jest przemienna, to $gH = Hg$ i można mówić po prostu o **warstwach**. Ogólnie, warstwa lewostronna gH i warstwa prawostronna Hg mogą być różnymi podzbiorami grupy G. Przedstawiony dowód twierdzenia 4 jest, po oczywistej zamianie „lewo–prawo", prawdziwy także dla warstw prawostronnych.

PRZYKŁAD 6 Rozważmy podgrupę $3\mathbb{Z}$ grupy $(\mathbb{Z}, +)$. Jej warstwy są zbiorami postaci $3\mathbb{Z} + r = \{3k + r\colon k \in \mathbb{Z}\}$. Jest ich tylko trzy, a mianowicie są to klasy reszt modulo 3: $[0]_3 = 3\mathbb{Z}$, $[1]_3 = 3\mathbb{Z}+1$ i $[2]_3 = 3\mathbb{Z} + 2$. Każdą liczbę n ze zbioru $\mathbb{Z}$ można zapisać w postaci $n = 3q + r$, gdzie $r \in \{0, 1, 2\}$ i $q \in \mathbb{Z}$, a więc $\mathbb{Z}$ jest sumą tych trzech rozłącznych zbiorów. Analogicznie, dla dowolnej liczby $p \in \mathbb{P}$, warstwami podgrupy $p\mathbb{Z}$ w grupie $(\mathbb{Z}, +)$ są zbiory $[r]_p = p\mathbb{Z} + r$, gdzie $r = 0, 1, 2, \ldots, p - 1$. ∎

PRZYKŁAD 7 Rozważmy grupę permutacji $G = \mathrm{PERM}(X)$ złożoną ze wszystkich różnowartościowych funkcji z niepustego zbioru X na samego siebie, ze składaniem jako działaniem grupowym. Dla dowolnego elementu x_0 zbioru X można zdefiniować podgrupę $\mathrm{FIX}_G(x_0) = \{f \in G\colon f(x_0) = x_0\}$; podgrupa ta była wprowadzana i używana w § 12.2. Warstwa lewostronna $e \circ \mathrm{FIX}_G(x_0)$ jest po prostu samą grupą $\mathrm{FIX}_G(x_0)$. Weźmy teraz dowolną warstwę lewostronną $g \circ \mathrm{FIX}_G(x_0)$. Dowolna funkcja należąca do tej warstwy jest postaci $g \circ f$, gdzie $f \in \mathrm{FIX}_G(x_0)$, a więc spełnia ona równość $g \circ f(x_0) = g(x_0)$. W istocie twierdzimy, że

$$g \circ \mathrm{FIX}_G(x_0) = \{h \in G\colon h(x_0) = g(x_0)\}.$$

Załóżmy bowiem, że $h(x_0) = g(x_0)$. Wówczas $g^{-1} \circ h(x_0) = g^{-1} \circ g(x_0) = x_0$; zatem funkcja $g^{-1} \circ h$ należy do zbioru $\mathrm{FIX}_G(x_0)$ i funkcja $h = g \circ (g^{-1} \circ h)$ należy do zbioru $g \circ \mathrm{FIX}_G(x_0)$.

Jeśli $g \notin \mathrm{FIX}_G(x_0)$, a zbiór X ma co najmniej trzy elementy, to warstwa prawostronna $\mathrm{FIX}_G(x_0) \circ g$ nie jest identyczna z warstwą lewostronną $g \circ \mathrm{FIX}_G(x_0)$. By się o tym przekonać, zauważmy, że $g(x_0) \neq x_0$, gdyż $g \notin \mathrm{FIX}_G(x_0)$. Wybierzmy permutację h z grupy G taką, że $h(x_0) = x_0$, ale $h(g(x_0)) \neq g(x_0)$. Wówczas $h \in \mathrm{FIX}_G(x_0)$, a stąd $h \circ g \in \mathrm{FIX}_G(x_0) \circ g$, ale $(h \circ g)(x_0) \neq g(x_0)$, a więc $h \circ g \notin g \circ \mathrm{FIX}_G(x_0)$.

W ćwiczeniu 17 szczegółowo zajmujemy się tym przykładem dla przypadku, w którym zbiór X ma trzy elementy. ∎

Warstwy podgrupy H nie dość, że tworzą podział grupy G, to są wszystkie tej samej wielkości, a mianowicie takiej, jak zbiór H.

Twierdzenie 5

Niech H będzie podgrupą grupy G i niech $g \in G$. Funkcja $h \to gh$ jest wzajemnie jednoznacznym przekształceniem podgrupy H na warstwę gH.

Dowód. Funkcja ta z pewnością przekształca H na gH oraz jest różnowartościowa dzięki prawu skracania: $gh = gh'$ implikuje, że $h = h'$. ∎

Nasz następny rezultat jest jednym z najczęściej wykorzystywanych faktów teorii grup skończonych. Do sformułowania go potrzebne nam będzie nowe oznaczenie. Dla podgrupy H grupy G, niech G/H oznacza zbiór wszystkich warstw lewostronnych podgrupy H w grupie G. Jak zwykle, $|H|$, $|G|$ i $|G/H|$ oznaczają liczby elementów zbiorów H, G i G/H.

Twierdzenie Lagrange'a

Niech H będzie podgrupą skończonej grupy G. Wówczas:

$$|G| = |G/H| \cdot |H|.$$

W szczególności, liczby $|H|$ i $|G/H|$ dzielą $|G|$.

Dowód. Jest $|G/H|$ warstw lewostronnych podgrupy H, z których każda, na mocy twierdzenia 5, ma $|H|$ elementów. Z twierdzenia 4 wynika, że tworzą one podział zbioru G. ∎

Twierdzenie 5 oraz twierdzenie Lagrange'a mają swoje odpowiedniki dla warstw prawostronnych, a więc liczba warstw prawostronnych podgrupy H w grupie G jest równa liczbie jej warstw lewostronnych. Ćwiczenie 23 polega na znalezieniu bezpośredniego dowodu tego faktu.

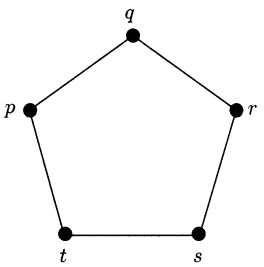
PRZYKŁAD 8

(a) Podgrupy dowolnej grupy mającej 10 elementów mogą mieć jedynie 1, 2, 5 lub 10 elementów.

(b) Podgrupy dowolnej grupy mającej 81 elementów mogą mieć jedynie 1, 3, 9, 27 lub 81 elementów.

(c) Załóżmy, że zbiór X ma n elementów. Grupa $G = \text{PERM}(X)$ z przykładu 7 ma wówczas $n!$ elementów, a podgrupa $\text{FIX}_G(x_0)$ ma $(n-1)!$ elementów. Jest n jej warstw lewostronnych $g \circ \text{FIX}_G(x_0)$, po jednej dla każdej możliwej wartości $g(x_0)$. W tym przypadku twierdzenie Lagrange'a przyjmuje postać

$$n! = n \cdot (n-1)!.$$



Rysunek 12.28

(d) Niech G będzie grupą automorfizmów grafu przedstawionego na rysunku 12.28, ze złożeniem jako działaniem grupowym. Twierdzenie Lagrange'a może pomóc w sprawdzeniu, że grupa G ma 10 elementów. Od razu dostrzegamy pięć „obrotów" o kąty 0°, 72°, 144°, 216° i 288°, tworzących podgrupę, którą oznaczymy przez R. Przekształcenie g, zdefiniowane za pomocą równości $g(p) = p$, $g(q) = t$, $g(r) = s$, $g(s) = r$ i $g(t) = q$, należy do grupy G, ale nie należy do podgrupy R, a więc $|G| > |R| = 5$. Z twierdzenia Lagrange'a wynika, że liczba $|G|$ jest wielokrotnością 5, a więc jest równa co najmniej 10. Ale dla elementu f z G, wartość $f(p)$ można wybrać tylko na pięć sposobów, następnie wartość $f(q)$ jedynie na co najwyżej dwa sposoby (w grę wchodzą wierzchołki z każdej strony $f(p)$) i wtedy pozostałe wartości funkcji f są już w pełni określone. Zatem grupa G ma co najwyżej $5 \cdot 2 = 10$ elementów, skąd ostatecznie wynika, że ma ich dokładnie 10.

Połowa elementów grupy G należy do podgrupy R. Pozostałe tworzą jedną warstwę $g \circ R$, gdzie g jest funkcją zdefiniowaną powyżej. W tym przypadku $R \circ g = g \circ R$. (Zob. też ćwiczenie 22).

Podgrupa $\langle g \rangle$ składa się jedynie z elementów e i g, ponieważ $g \circ g = e$. A więc $|\langle g \rangle| = 2$. Skoro $10 = |G/\langle g \rangle| \cdot |\langle g \rangle|$, to istnieje pięć warstw podgrupy $\langle g \rangle$ w grupie G. W rzeczywistości $\langle g \rangle$ jest podgrupą grupy G złożoną z tych wszystkich automorfizmów, dla których p jest punktem stałym. Podobnie jak w punkcie (c), warstwami podgrupy $\langle g \rangle$ jest po prostu pięć podzbiorów grupy G, z których każdy składa się z automorfizmów przeprowadzających p na jedną spośród pięciu możliwych wartości. Są to więc zbiory $\{f \in G\colon f(p) = p\}$, $\{f \in G\colon f(p) = q\}$, $\{f \in G\colon f(p) = r\}$ itd. ■

Paragraf ten zakończymy pokazując, w jaki sposób niektóre z naszych rezultatów można uważać za szczególne przypadki faktów dotyczących grup permutacji. Po pierwsze, można rozpatrywać działanie grupy G na niej samej, dla każdego $g_0 \in G$ definiu-

jąc funkcję g_0^* wzorem $g_0^*(g) = g_0 \cdot g$ dla $g \in G$. Każda z funkcji g_0^* jest permutacją zbioru G (sprawdź to). Mamy również

$$(g_0 \cdot g_1)^*(g) = g_0 \cdot g_1 \cdot g = g_0 \cdot (g_1^*(g)) = g_0^*(g_1^*(g)) = g_0^* \circ g_1^*(g)$$

dla wszystkich elementów $g \in G$, a więc $(g_0 \circ g_1)^* = g_0^* \circ g_1^*$. W rzeczywistości funkcja $g \to g^*$ ustala izomorfizm grupy G z pewną podgrupą grupy PERM (G); dowodzi to twierdzenia Cayleya, które zostało wspomniane po przykładzie 1.

Dowolna podgrupa H grupy G również działa na G; dla $h \in H$ definiujemy $h^*(g) = h \cdot g$. Orbitą danego elementu $g \in G$ (grupa H działa na zbiorze G) względem działania grupy H jest zbiór

$$\{h^*(g)\colon h \in H\} = \{h \cdot g\colon h \in H\} = H \cdot g = Hg.$$

Innymi słowy, orbitami działania grupy H są po prostu warstwy lewostronne podgrupy H w grupie G. Wniosek ze stwierdzenia 1 z § 12.2 mówi nam, że H-orbity tworzą podział zbioru G. A więc warstwy prawostronne podgrupy H w grupie G tworzą podział zbioru G, co dowodzi twierdzenia 4 w wersji dla warstw prawostronnych. Podobnie, równość $h^*(g) = g \cdot h^{-1}$ definiuje inne działanie grupy H na zbiorze G, dla którego orbitami są po prostu warstwy lewostronne podgrupy H w grupie G (ćwiczenie 24). A więc warstwy lewostronne także tworzą podział zbioru G; stanowi to kolejne potwierdzenie prawdziwości twierdzenia 4.

ĆWICZENIA DO § 12.5

1. Opisz elementy każdej z następujących podgrup grupy $(\mathbb{Z}, +)$:
(a) $\langle 1\rangle$, (b) $\langle 0\rangle$, (c) $\langle\{-1, 2\}\rangle$,
(d) $\langle\mathbb{Z}\rangle$, (e) $\langle\{2, 3\}\rangle$, (f) $\langle 6\rangle \cap \langle 9\rangle$.

2. Które z podgrup z ćwiczenia 1 są grupami cyklicznymi? Uzasadnij swoje odpowiedzi.

3. Które z następujących podzbiorów zbioru $\mathbb{Z}$ są podgrupami grupy $(\mathbb{Z}, +)$? Przedstaw te podgrupy w postaci $n\mathbb{Z}$; zob. twierdzenie 3.
(a) $\{0, \pm 3, \pm 6, \pm 9, \ldots\}$, (b) $\{0, 5, 10, 15, 20, \ldots\}$,
(c) $\{0, \pm 2, \pm 4, \pm 8, \pm 16, \ldots\}$, (d) $\{k \in \mathbb{Z}\colon k$ jest wielokrotnością $4\}$,
(e) $\mathbb{N} \cup (-\mathbb{N})$.

4. Które z następujących podzbiorów zbioru $\mathbb{R}$ są podgrupami grupy $(\mathbb{R}, +)$?
(a) $\mathbb{Z}$, (b) $\mathbb{N}$, (c) $\mathbb{Q}$,
(d) $\{n\sqrt{2}\colon n \in \mathbb{Z}\}$, (e) $\{m + n\sqrt{2}\colon m, n \in \mathbb{Z}\}$.

5. Uzupełnij dowód twierdzenia 1(a). To znaczy, wykaż prawdziwość stwierdzenia zawartego w zdaniu zaczynającym się od „Analogicznie".

6. Sformułuj i udowodnij twierdzenie 1 dla grupy przemiennej używając notacji addytywnej.

7. (a) Znajdź wszystkie generatory grupy $(\mathbb{Z}_5, +_5)$.
(b) Znajdź wszystkie generatory grupy $(\mathbb{Z}_6, +_6)$.

8. (a) Znajdź część wspólną wszystkich podgrup $n\mathbb{Z}$ grupy $(\mathbb{Z}, +)$, gdzie $n \in \mathbb{P}$.
(b) Czy podgrupa znaleziona w części (a) jest cykliczna? Wyjaśnij to.

9. (a) Podaj przykład przekształcenia wzajemnie jednoznacznego podgrupy $4\mathbb{Z}$ na jej warstwę $4\mathbb{Z} + 3$ w grupie $(\mathbb{Z}, +)$.
(b) Podaj inny jeszcze przykład takiego przekształcenia.

10. Weź grupę $(\mathbb{Z}, +)$. Przedstaw grupę $\mathbb{Z}$ w postaci rozłącznej sumy pięciu warstw pewnej jej podgrupy.

11. Zbiór S_4 wszystkich permutacji zbioru $\{1, 2, 3, 4\}$ jest grupą ze względu na działanie składania funkcji. Które z następujących podzbiorów są podgrupami grupy $(S_4, \circ)$? Uzasadnij swoje odpowiedzi.
(a) $\{g \in S_4\colon g(4) = 4\}$,
(b) $\{g \in S_4\colon g(1) = 2\}$,
(c) $\{g \in S_4\colon g(1) \in \{1, 2\}\}$,
(d) $\{g \in S_4\colon g(1) \in \{1, 2\}$ i $g(2) \in \{1, 2\}\}$.

12. (a) Wyjaśnij, dlaczego dowolny trzynastoelementowy podzbiór grupy $(S_4, \circ)$ z ćwiczenia 11 generuje tę grupę. (W rzeczywistości istnieją też pewne dwuelementowe zbiory generujące grupę S_4, ale trudniej jest je wskazać.)
(b) Czy grupa $(S_4, \circ)$ jest cykliczna? Uzasadnij swoją odpowiedź.

13. (a) Udowodnij, że $(g_1 \cdot g_2 \cdot g_3)^{-1} = g_3^{-1} \cdot g_2^{-1} \cdot g_1^{-1}$ dla wszystkich elementów g_1, g_2, g_3 dowolnej grupy G.
(b) Udowodnij uogólnienie powyższej równości dla $(g_1 \cdot g_2 \cdot \ldots \cdot g_n)^{-1}$.

14. Dla elementu g dowolnej grupy G zdefiniowaliśmy przed przykładem 3 potęgi g^n dla wszystkich liczb $n \in \mathbb{Z}$. Udowodnij, że
(a) $(g^{-1})^{-k} = g^k$ dla wszystkich $g \in G$ i $k \in \mathbb{Z}$.
(b) $g^m \cdot g^1 = g^{m+1}$ dla wszystkich $g \in G$ i $m \in \mathbb{Z}$. (Przypadek, gdy $m \in \mathbb{N}$ jest łatwy.)
(c) $g^m \cdot g^n = g^{m+n}$ dla wszystkich $g \in G$, $m \in \mathbb{Z}$ i $n \in \mathbb{N}$.
(d) $g^m \cdot g^n = g^{m+n}$ dla wszystkich $g \in G$, $m \in \mathbb{Z}$ i $n \in \mathbb{Z}$. (*Wskazówka*: wykorzystaj część (c) z g^{-1} w miejsce g).

15. (a) Wykaż, że część wspólna dowolnej rodziny podgrup grupy G jest podgrupą grupy G.
(b) Podaj przykład grupy G i jej podgrup H i K takich, że zbiór $H \cup K$ nie jest podgrupą grupy G.

16. (a) Wykaż, że podgrupa R z przykładu 8(d) jest cykliczna i wskaż jakiś jej generator.

(b) Wykaż, że grupa G z przykładu 8(d) ma elementy rzędów 1, 2 i 5, ale nie ma elementu rzędu 10.

17. Załóżmy, że w przykładzie 7 mamy $X = \{1, 2, 3\}$ i $x_0 = 1$.

(a) Znajdź $|\,\mathrm{PERM}\,(X)|$.

(b) Wypisz permutacje tworzące podgrupę $\mathrm{FIX}_G(1)$.

(c) Permutacja (1 2 3) nie należy do zbioru $\mathrm{FIX}_G(1)$. Wypisz elementy warstwy $\mathrm{FIX}_G(1) \circ (1\,2\,3)$.

(d) Wykaż, że $\mathrm{FIX}_G(1) \circ (1\,2\,3) \neq (1\,2\,3) \circ \mathrm{FIX}_G(1)$.

(e) Wykaż, że w rzeczywistości zbiór $\mathrm{FIX}_G(1) \circ (1\,2\,3)$ w ogóle nie jest żadną warstwą lewostronną podgrupy $\mathrm{FIX}_G(1)$.

(f) Ile warstw ma podgrupa $\mathrm{FIX}_G(1)$ w grupie $\mathrm{PERM}\,(X)$?

18. Tablica zamieszczona poniżej opisuje dwuargumentowe działanie $\bullet$ w zbiorze $G = \{a, b, c, d, e\}$, z elementem neutralnym e.

$\bullet$	e	a	b	c	d
e	e	a	b	c	d
a	a	e	c	d	b
b	b	d	a	e	c
c	c	b	d	a	e
d	d	c	e	b	a

(a) Wykaż, że zbiór $\{e, a\}$ jest grupą ze względu na działanie $\bullet$.

(b) Bez wykonywania jakichkolwiek obliczeń, wywnioskuj z wyniku części (a) oraz twierdzenia Lagrange'a, że $(G, \bullet)$ nie jest grupą.

19. Następująca tablica opisuje działanie grupowe w pewnej grupie $(G, \bullet)$ o elementach a, b, c, d, e, f.

$\bullet$	e	a	b	c	d	f
e	e	a	b	c	d	f
a	a	b	e	d	f	c
b	b	e	a	f	c	d
c	c	f	d	e	b	a
d	d	c	f	a	e	b
f	f	d	c	b	a	e

(a) Wypisz elementy podgrupy $\langle a \rangle$.

(b) Wykaż, że $\langle a \rangle \bullet c = c \bullet \langle a \rangle$.

(c) Znajdź wszystkie podgrupy tej grupy, mające dwa elementy.

(d) Znajdź $|G/\langle d \rangle|$.

(e) Opisz warstwy prawostronne podgrupy $\langle d \rangle$.

20. Powtórz ćwiczenie 19 dla grupy opisanej następującą tablicą.

•	e	a	b	c	d	f
e	e	a	b	c	d	f
a	a	b	e	d	f	c
b	b	e	a	f	c	d
c	c	d	f	a	b	e
d	d	f	c	b	e	a
f	f	c	d	e	a	b

21. Wykaż, że jeśli H jest podgrupą grupy $(G, \cdot)$ oraz $g \in G$, to $g \cdot H = H$ wtedy i tylko wtedy, gdy $g \in H$. Spróbuj mądrze wykorzystać twierdzenie 4.

22. Weź dowolną skończoną grupę $(G, \cdot)$ wraz z jej podgrupą H taką, że $|G| = 2|H|$. Wykaż, że $g \cdot H = H \cdot g$ dla każdego elementu $g \in G$. *Sugestia*: rozważ osobno przypadki: $g \in H$ i $g \notin H$.

23. Niech H będzie podgrupą grupy $(G, \cdot)$.
(a) Wykaż, że dla każdego elementu g grupy G warstwa prawostronna $H \cdot g^{-1}$ składa się z odwrotności elementów warstwy lewostronnej $g \cdot H$.
(b) Znajdź przekształcenie wzajemnie jednoznaczne zbioru warstw lewostronnych podgrupy H w G na zbiór jej warstw prawostronnych.

24. Niech będzie podgrupą grupy G. Dla dowolnego elementu $h \in H$ definiujemy funkcję $h^*\colon G \to G$ wzorem $h^*(g) = g \cdot h^{-1}$ dla $g \in G$.
(a) Udowodnij, że w ten sposób zdefiniowane zostało działanie grupy H na zbiorze G.
(b) Wykaż, że funkcja $h \to h^*$ jest różnowartościowa.
(c) Zauważ, że w przypadku, gdy $G = H$, część (b) kończy dowód twierdzenia Cayleya, ponieważ pokazuje, że grupa G jest izomorficzna z pewną podgrupą grupy PERM (G).

25. Niech H będzie podgrupą grupy $(G, \cdot)$ i dla dowolnych elementów $g_1, g_2 \in G$ niech $g_1 \sim g_2$, jeśli $g_2^{-1} \cdot g_1 \in H$.
(a) Wykaż, że $\sim$ jest relacją równoważności w zbiorze G.
(b) Wykaż, że podział, o którym mowa w twierdzeniu 4, jest to dokładnie podział na klasy abstrakcji względem relacji $\sim$, opisany w twierdzeniu 1 z § 3.5.

26. Wykaż, że jeśli H jest podgrupą grupy $(G, \cdot)$, to dla każdego elementu g grupy G, gHg^{-1} jest również podgrupą grupy G.

§ 12.6. Twierdzenie o izomorfizmie

W tym paragrafie wprowadzone zostaną homomorfizmy, które są najważniejszymi funkcjami z punktu widzenia badania

grup. Przypomnijmy sobie, że izomorfizm między grupą $(G, \cdot)$ a grupą $(G_0, \bullet)$ jest to wzajemnie jednoznaczne przekształcenie $\varphi\colon G \to G_0$ zbioru G na zbiór G_0 spełniające warunek $\varphi(g \cdot h) = \varphi(g) \bullet \varphi(h)$ dla wszystkich elementów $g, h \in G$. Izomorfizmy zachowują strukturę grupy we wzajemnie jednoznaczny sposób. Homomorfizmy również zachowują strukturę grupy, ale nie muszą być funkcjami różnowartościowymi ani funkcjami „na".

Homomorfizm grupy $(G, \cdot)$ w grupę $(G_0, \bullet)$ jest to funkcja $\varphi\colon G \to G_0$ spełniająca warunek $\varphi(g \cdot h) = \varphi(g) \bullet \varphi(h)$ dla wszystkich elementów $g, h \in G$. Ponieważ element neutralny i operacja brania elementu odwrotnego są częściami składowymi struktury grupowej, więc wydaje się, że powinniśmy także żądać, by funkcja φ zachowywała element neutralny i operację odwracania. To jednak, szczęśliwie, są automatyczne konsekwencje przyjętej definicji.

Stwierdzenie

Niech $\varphi\colon G \to G_0$ będzie homomorfizmem.

(a) Jeśli e oznacza element neutralny grupy G, to $\varphi(e)$ jest elementem neutralnym, e_0, grupy G_0.

(b) $\varphi(g^{-1}) = \varphi(g)^{-1}$ dla wszystkich elementów $g \in G$.

(c) $\varphi(G)$ jest podgrupą grupy G_0.

Dowód. (a) $\varphi(e) = \varphi(e) \bullet [\varphi(e) \bullet \varphi(e)^{-1}] = [\varphi(e) \bullet \varphi(e)] \bullet \varphi(e)^{-1} = \varphi(e \cdot e) \bullet \varphi(e)^{-1} = \varphi(e) \bullet \varphi(e)^{-1} = e_0$.

(b) Zwróć uwagę, że pierwsza operacja brania elementu odwrotnego ma miejsce w grupie G, podczas gdy druga odbywa się w grupie G_0. Ponieważ $\varphi(g) \bullet \varphi(g^{-1}) = \varphi(g \cdot g^{-1}) = \varphi(e) = e_0$ oraz elementy odwrotne są wyznaczone jednoznacznie, więc $\varphi(g^{-1}) = \varphi(g)^{-1}$.

(c) Skoro funkcja φ zachowuje operację odwracania, to do zbioru $\varphi(G)$ należą odwrotności wszystkich jego elementów. Ponieważ zbiór $\varphi(G)$ jest również zamknięty ze względu na mnożenie, to jest podgrupą grupy G. ∎

PRZYKŁAD 1

(a) Niech zarówno $(G, \cdot)$, jak i $(G_0, \bullet)$ oznaczają grupę $(\mathbb{Z}, +)$. Warunek definiujący homomorfizm przyjmuje postać

$$\varphi(m + n) = \varphi(m) + \varphi(n) \quad \text{dla} \quad m, n \in \mathbb{Z}.$$

Funkcja φ zdefiniowana dla wszystkich n wzorem $\varphi(n) = 5n$ jest przykładem homomorfizmu, ponieważ

$$\varphi(m + n) = 5 \cdot (m + n) = 5m + 5n = \varphi(m) + \varphi(n).$$

W liczbie 5 nie ma nic szczególnego; dla każdej innej liczby całkowitej można w ten sam sposób zdefiniować homomorfizm $(\mathbb{Z}, +)$ w $(\mathbb{Z}, +)$.

(b) Niech $(G, \cdot)$ oznacza grupę $(\mathbb{Z}, +)$, $(G_0, \bullet)$ — grupę $(\mathbb{R}\setminus\{0\}, \cdot)$ i niech $\varphi(m) = 2^m$ dla każdej liczby m ze zbioru $\mathbb{Z}$. Ponieważ

$$\varphi(m+n) = 2^{m+n} = 2^m \cdot 2^n = \varphi(m) \cdot \varphi(n),$$

więc φ jest homomorfizmem $(\mathbb{Z}, +)$ w $(\mathbb{R}\setminus\{0\}, \cdot)$. Zauważ, że $\varphi(0) = 1$.

(c) Przypomnij sobie, że $a = \log_2 b$ wtedy i tylko wtedy, gdy $b = 2^a$. Funkcja φ, przekształcająca grupę $(\{x \in \mathbb{R}\colon x > 0\}, \cdot)$ w grupę $(\mathbb{R}, +)$, zdefiniowana wzorem $\varphi(x) = \log_2 x$, jest homomorfizmem, ponieważ $\log_2(xy) = \log_2(x) + \log_2(y)$.

(d) Równość

$$(f \circ g)^* = f^* \circ g^*$$

widzieliśmy wielokrotnie, zajmując się w § 12.3 automorfizmami grafów. W każdym z rozpatrywanych przypadków mieliśmy do czynienia z homomorfizmem. Po raz pierwszy, gdy chodziło o przekształcenie $f \to f^*$ grupy $\mathrm{AUT}(H)$ w grupę permutacji zbioru krawędzi grafu, zdefiniowane wzorem $f^*(\{u, v\}) = \{f(u), f(v)\}$. Inny przykład homomorfizmu $f \to f^*$ stanowiła operacja obcinania automorfizmów z grupy $\mathrm{AUT}(H)$ do zbioru będącego sumą pewnych orbit działania grupy $\mathrm{AUT}(H)$, rozważana w przykładzie 3 z § 12.3.

Stwierdzenie, że grupa G „działa na" zbiorze X w sensie, który został zdefiniowany w § 12.3, znaczy po prostu, że dany jest pewien homomorfizm grupy G w grupę $\mathrm{PERM}(X)$, tzn. pewien sposób przyporządkowania elementom zbioru G permutacji tak, by zachowana była struktura grupowa. ∎

PRZYKŁAD 2 (a) Weźmy dodatnią liczbę całkowitą $p \geq 2$. Tak, jak w § 3.6, niech $n \bmod p$ będzie resztą z dzielenia n przez p. Przypomnijmy, że $\bmod p$ jest funkcją ze zbioru $\mathbb{Z}$ w zbiór $\mathbb{Z}_p$. W rzeczywistości, na mocy twierdzenia 3 z § 3.6, $\bmod p$ jest homomorfizmem grupy $(\mathbb{Z}, +)$ na grupę $(\mathbb{Z}_p, +_p)$. To znaczy,

$$(m+n) \bmod p = (m \bmod p) +_p (n \bmod p) \quad \text{dla} \quad m, n \in \mathbb{Z}.$$

(b) Zacytowane twierdzenie głosi także, że

$$(m \cdot n) \bmod p = (m \bmod p) *_p (n \bmod p).$$

Kuszące jest wywnioskowanie stąd, że MOD p jest również homomorfizmem $(\mathbb{Z}, \cdot)$ na $(\mathbb{Z}_p, *_p)$. Ale $(\mathbb{Z}_p, *_p)$ *nie* jest grupą! Dla elementu 0 nie istnieje element odwrotny ze względu na mnożenie. Przykład ten pokazuje, że w stwierdzeniu, udowodnionym na początku tego paragrafu, założenie, że G_0 jest grupą, jest istotne. ■

PRZYKŁAD 3 Niech $(G, \cdot)$ będzie grupą i niech $g \in G$. W ćwiczeniu 14 z § 12.5 stwierdza się, że $g^{m+n} = g^m \cdot g^n$ dla wszystkich liczb $m, n \in \mathbb{Z}$. Znaczy to, że funkcja φ, przekształcająca grupę $(\mathbb{Z}, +)$ w grupę $(G, \cdot)$, dana wzorem $\varphi(n) = g^n$, jest homomorfizmem. Jego zbiorem wartości $\varphi(\mathbb{Z})$ jest podgrupa cykliczna $\langle g \rangle$ grupy G. ■

Z każdym homomorfizmem φ, określonym na grupie G, związana jest pewna szczególna podgrupa grupy G, nazywana jego **jądrem**, którą definiujemy jako zbiór $\{g \in G\colon \varphi(g) = \varphi(e)\}$. W pewnym sensie jądro homomorfizmu mówi nam, co zaniedbujemy, kiedy przechodzimy od G do $\varphi(G)$. Punkt (c) następnego twierdzenia można uważać za stwierdzenie, że homomorfizm φ przyjmuje tę samą wartość dla dwóch elementów grupy G wtedy i tylko wtedy, gdy różnią się one o element jądra homomorfizmu φ.

Twierdzenie 1

Niech $\varphi\colon G \to G_0$ będzie homomorfizmem, przy czym działaniem grupowym w grupie G jest $\cdot$, a działaniem grupowym w grupie G_0 jest $\bullet$. Niech e i e_0 będą elementami neutralnymi, odpowiednio, grup G i G_0, i niech K będzie jądrem homomorfizmu φ. Wtedy:

(a) K jest podgrupą grupy G.
(b) $g \cdot K = K \cdot g$ dla każdego elementu g grupy G.
(c) $g \cdot K = \{h \in G\colon \varphi(h) = \varphi(g)\}$ dla każdego elementu g grupy G.

Dowód. (a) Jeśli $g, h \in K$, to $\varphi(g \cdot h) = \varphi(g) \bullet \varphi(h) = \varphi(e) \bullet \varphi(e) = e_0 \bullet e_0 = e_0$, a więc $g \cdot h \in K$. Ponadto, $\varphi(g^{-1}) = \varphi(g)^{-1} = e_0^{-1} = e_0$, a więc element g^{-1} należy do K. Wynika stąd, że zbiór K jest zamknięty ze względu na iloczyny i branie elementów odwrotnych, a więc jest podgrupą grupy G.

(b) Pokażemy najpierw, że $K \cdot g \subseteq g \cdot K$. Wystarczy wziąć element $k \in K$ i pokazać, że $k \cdot g \in g \cdot K$. Ponieważ $k \cdot g = g \cdot (g^{-1} \cdot k \cdot g)$, to wystarczy pokazać, że $g^{-1} \cdot k \cdot g \in K$. Ale

$\varphi(g^{-1} \cdot k \cdot g) = \varphi(g^{-1}) \bullet \varphi(k) \bullet \varphi(g) = \varphi(g)^{-1} \bullet \varphi(e) \bullet \varphi(g) = \varphi(g^{-1} \cdot e \cdot g) = \varphi(e)$. Podobny argument pokazuje, że $g \cdot K \subseteq K \cdot g$.

(c) Dla $k \in K$ mamy $\varphi(g \cdot k) = \varphi(g) \bullet \varphi(k) = \varphi(g) \bullet e_0 = \varphi(g)$, a więc $g \cdot K \subseteq \{h \in G\colon \varphi(h) = \varphi(g)\}$. Odwrotnie, jeśli $\varphi(h) = \varphi(g)$, to $\varphi(g^{-1} \cdot h) = \varphi(g)^{-1} \bullet \varphi(h) = e_0$, co znaczy, że $g^{-1} \cdot h \in K$, a więc $h = g \cdot (g^{-1} \cdot h) \in g \cdot K$. ∎

Wniosek 1

Przy oznaczeniach z twierdzenia 1, warstwy postaci $g \cdot K$ są klasami abstrakcji relacji równoważności wyznaczonej przez homomorfizm φ w następujący sposób: $g \sim h$ wtedy i tylko wtedy, gdy $\varphi(g) = \varphi(h)$.

Dowód. Jest to po prostu przeformułowanie punktu (c). ∎

W twierdzeniu 5 z § 12.5 pokazaliśmy, że wszystkie warstwy $g \cdot K$ mają tę samą liczbę elementów, a mianowicie $|K|$. Z faktu tego wynika przydatny test na różnowartościowość homomorfizmów.

Wniosek 2

Homomorfizm jest różnowartościowy wtedy i tylko wtedy, gdy jego jądro składa się wyłącznie z elementu neutralnego danej grupy.

Dowód. Ponieważ K jest podgrupą, więc z pewnością należy doń element neutralny e. Co więcej, na mocy twierdzenia 1(c), homomorfizm φ jest różnowartościowy wtedy i tylko wtedy, gdy każda z warstw $g \cdot K$ ma dokładnie jeden element, a to zachodzi wtedy i tylko wtedy, gdy samo jądro K ma dokładnie jeden element. ∎

PRZYKŁAD 4

(a) Homomorfizm φ grupy $(\mathbb{Z}, +)$ w $(\mathbb{Z}, +)$ zdefiniowany wzorem $\varphi(n) = 5n$ jest różnowartościowy, a jego jądrem jest zbiór $\{n \in \mathbb{Z}\colon 5n = 0\} = \{0\}$. Homomorfizm z części (b) przykładu 1 jest także różnowartościowy, a jego jądrem jest zbiór $\{n \in \mathbb{Z}\colon 2^n = 1\} = \{0\}$. Homomorfizm $\varphi(x) = \log_2 x$ z części (d) przykładu 1 jest różnowartościowy, a jego jądrem jest zbiór $\{x \in \mathbb{R}\colon \log_2 x = 0\} = \{1\}$.

(b) Homomorfizm $\text{MOD}\, p$ (przykład 2(a)) grupy $(\mathbb{Z}, +)$ w $(\mathbb{Z}_p, +_p)$ nie jest różnowartościowy. Jego jądrem jest zbiór $\{n \in \mathbb{Z}\colon n \,\text{MOD}\, p = 0\} = \{n \in \mathbb{Z}\colon n \text{ jest wielokrotnością } p\} = p\mathbb{Z}$.

Dwie liczby całkowite są w tej samej warstwie jądra wtedy i tylko wtedy, gdy ich różnica jest wielokrotnością p, a warstwami są po prostu klasy reszt modulo p, czyli zbiory $[r]_p = p\mathbb{Z} + r$. ∎

PRZYKŁAD 5 Homomorfizm $n \to g^n$ grupy $(\mathbb{Z}, +)$ w grupę $(\langle g\rangle, \cdot)$, z którym zetknęliśmy się w przykładzie 3, stanowi klucz do zrozumienia grup cyklicznych. Jego jądrem jest zbiór $\{n \in \mathbb{Z}\colon\ g^n = e\}$. Ta podgrupa grupy $\mathbb{Z}$ jest równa po prostu $\{0\}$, jeśli zbiór $\langle g\rangle$ jest nieskończony, ale w przeciwnym razie jest równa $p\mathbb{Z}$, dla pewnej, różnej od zera liczby p. W pierwszym przypadku homomorfizm jest różnowartościowy, w drugim natomiast nie. Fakty te wykorzystamy w przykładzie 7, ale przedtem rozbudujemy trochę nasz aparat pojęciowy. ∎

Podgrupę K grupy G mającą tę własność, że $g \cdot K = K \cdot g$ dla każdego elementu g z G, nazywamy **dzielnikiem normalnym** lub **podgrupą normalną** grupy G. Twierdzenie 1(b) pokazuje, że jądra homomorfizmów są dzielnikami normalnymi. Jeśli grupa G jest przemienna, to każda jej podgrupa jest dzielnikiem normalnym, ale w ogólnym przypadku istnieją podgrupy, które nie są dzielnikami normalnymi. Jeśli podgrupa K jest dzielnikiem normalnym, to jej warstwy lewo- i prawostronne pokrywają się i będziemy je nazywać po prostu **warstwami**.

Gdy K jest dzielnikiem normalnym grupy G, to w zbiorze G/K jej warstw lewostronnych $g \cdot K$ określone jest jego własne, naturalne i użyteczne działanie grupowe $*$. Elementy zbioru G/K, które chcemy przez siebie mnożyć, są zbiorami (a mianowicie warstwami podgrupy K) i iloczyn dwóch warstw będzie ponownie jedną z warstw. Nie po raz pierwszy spotykamy się z takim pomysłem. Jesteśmy przyzwyczajeni do tego, że mając dwa zbiory A i B, można z nich utworzyć nowy zbiór, na przykład biorąc $A \cap B$ lub $A \cup B$. Również w przykładzie 6(a) z § 3.6 dodawaliśmy i mnożyliśmy przez siebie klasy reszt modulo p. Nasz iloczyn warstw jest uogólnieniem działań z tamtego przykładu.

Jeśli chcemy spróbować pomnożyć przez siebie warstwy $g \cdot K$ i $h \cdot K$, to naturalne jest przyjęcie następującej definicji:

$$(g \cdot K) * (h \cdot K) = g \cdot K \cdot h \cdot K = \{g \cdot k_1 \cdot h \cdot k_2\colon\ k_1, k_2 \in K\}.$$

W przypadku dowolnej podgrupy K iloczyn ten mógłby nie być warstwą (zob. ćwiczenie 19), ale musi nią być, jeśli K jest dzielnikiem normalnym.

Twierdzenie 2

Niech K będzie dzielnikiem normalnym grupy $(G, \cdot)$. Wtedy

(a) $g \cdot K \cdot h \cdot K = (g \cdot h) \cdot K$ dla wszystkich elementów $g, h \in G$.

(b) Zbiór G/K wszystkich warstw podgrupy K w grupie G jest grupą ze względu na działanie $*$ zdefiniowane wzorem

$$(g \cdot K) * (h \cdot K) = g \cdot h \cdot K.$$

(c) Funkcja $\nu\colon G \to G/K$ zdefiniowana wzorem $\nu(g) = g \cdot K$ jest homomorfizmem o jądrze K (ν jest małą grecką literą ni).

Dowód. (a) Ponieważ K jest podgrupą grupy G, to $K = K \cdot e \subseteq K \cdot K \subseteq K$, a więc $K = K \cdot K$. Ponieważ K jest dzielnikiem normalnym, to mamy $h \cdot K = K \cdot h$, a więc

$$g \cdot K \cdot h \cdot K = g \cdot h \cdot K \cdot K = g \cdot h \cdot K \in G/K.$$

(b) Zgodnie z punktem (a), $(g \cdot K) * (h \cdot K)$ jest zbiorem $(g \cdot K) \cdot (h \cdot K)$, skąd wynika, że $*$ jest poprawnie zdefiniowanym dwuargumentowym działaniem w zbiorze G/K. Łatwo sprawdzić, że jest ono łączne, że K jest elementem neutralnym oraz że $(g \cdot K)^{-1} = g^{-1} \cdot K$.

(c) Zgodnie z definicją działania $*$ mamy $\nu(g \cdot h) = g \cdot h \cdot K = (g \cdot K) * (h \cdot K) = \nu(g) * \nu(h)$, a więc ν jest homomorfizmem. Jeśli $\nu(g) = \nu(e)$, to $g \in g \cdot K = e \cdot K = K$, a jeśli $g \in K$, to $\nu(g) = g \cdot K = K = \nu(e)$ (ćwiczenie 21, § 12.5). Zatem K jest jądrem homomorfizmu ν. ∎

Przekształcenie ν nazywane jest **naturalnym** (lub **kanonicznym**) **homomorfizmem** grupy G na grupę G/K. Oznaczenie go grecką literą ni ma pomóc w pamiętaniu o tym, że jest on naturalny. Twierdzenia 1 i 2 mówią nam, że jądra homomorfizmów są dzielnikami normalnymi i odwrotnie, każdy dzielnik normalny jest jądrem pewnego homomorfizmu. Jeżeli K jest jądrem homomorfizmu φ, to naturalny homomorfizm ν, o którym mowa w twierdzeniu 2, jest tym samym naturalnym przekształceniem $s \to [s]$, z którym zapoznaliśmy się w § 3.5. Warstwa $g \cdot K$ jest bowiem, na mocy wniosku 1 z twierdzenia 1, klasą równoważności $[g]$ elementu g.

PRZYKŁAD 6

Niech $(G, \cdot) = (\mathbb{Z}, +)$ i niech $K = 6\mathbb{Z}$. Twierdzenie 2(b) mówi nam, że zbiór $\mathbb{Z}/6\mathbb{Z}$ jest grupą ze względu na działanie

$$(k + 6\mathbb{Z}) * (m + 6\mathbb{Z}) = k + m + 6\mathbb{Z}.$$

Elementem neutralnym grupy $\mathbb{Z}/6\mathbb{Z}$ jest zbiór $6\mathbb{Z}$.

Twierdzenia 2(c) mówi nam, że jeśli $\nu(k) = k + 6\mathbb{Z}$, to homomorfizm ν przekształca grupę $\mathbb{Z}$ na grupę $\mathbb{Z}/6\mathbb{Z}$, a jego jądrem jest zbiór $6\mathbb{Z}$, tzn.

$$\{k \in \mathbb{Z}\colon \nu(k) = 6\mathbb{Z}\} = \{k \in \mathbb{Z}\colon k + 6\mathbb{Z} = 6\mathbb{Z}\} = 6\mathbb{Z}.$$

W przykładzie 4(b) zwróciliśmy uwagę na to, że zbiór $6\mathbb{Z}$ jest także jądrem homomorfizmu $\varphi = \text{MOD}\,6$, przekształcającego grupę $(\mathbb{Z}, +)$ na grupę $(\mathbb{Z}_6, +_6)$. Jeśli położymy

$$\varphi^*(k + 6\mathbb{Z}) = k \qquad \text{dla} \quad k \in \{0, 1, 2, 3, 4, 5\} = \mathbb{Z}_6,$$

to otrzymamy wzajemnie jednoznaczne przekształcenie zbioru $\mathbb{Z}/6\mathbb{Z}$ na zbiór $\mathbb{Z}_6$. Korzystając z oznaczeń z § 3.6 możemy napisać, że jest to odpowiedniość $[m]_6 \leftrightarrow m \text{ MOD } 6$. Przekształcenie φ^* jest homomorfizmem, ponieważ

$$\begin{aligned}\varphi^*((k + 6\mathbb{Z}) * (m + 6\mathbb{Z})) &= \varphi^*(k + m + 6\mathbb{Z}) = \varphi^*(k +_6 m + 6\mathbb{Z}) \\ &= k +_6 m = \varphi^*(k + 6\mathbb{Z}) +_6 \varphi^*(m + 6\mathbb{Z}).\end{aligned}$$

Wynika stąd, że grupy $\mathbb{Z}/6\mathbb{Z}$ i $\mathbb{Z}_6$ są izomorficzne. Jeśli zapiszemy to używając naszych oznaczeń G i K, to okaże się, że izomorficzne są grupy G/K i $\varphi(G)$, co jest szczególnym przypadkiem kolejnego twierdzenia. ∎

Twierdzenie o izomorfizmie

Niech φ będzie homomorfizmem grupy $(G, \cdot)$ w grupę $(G_0, \bullet)$ i niech K będzie jego jądrem. Wówczas grupy G/K i $\varphi(G)$ są izomorficzne, a ich izomorfizm φ^* jest dany wzorem

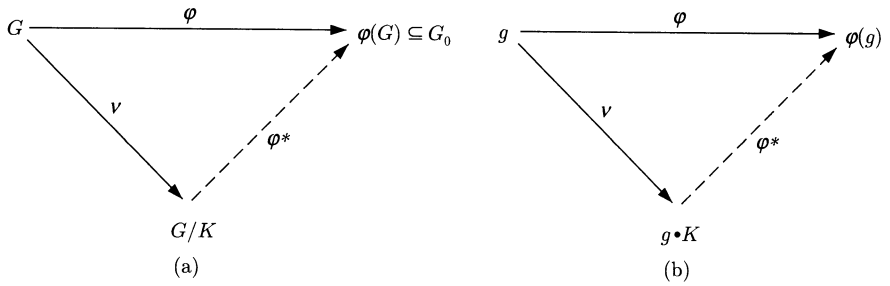
$$\varphi^*(g \cdot K) = \varphi(g).$$

Dowód. Twierdzenie 1(c) mówi, że $g \cdot K$ jest zbiorem tych wszystkich elementów h grupy G, dla których $\varphi(h) = \varphi(g)$. Wynika stąd, że $\varphi^*(g \cdot K)$ jest wspólną wartością przyjmowaną przez homomorfizm φ dla wszystkich elementów ze zbioru $g \cdot K$. Dla elementów z różnych warstw funkcja φ przyjmuje różne wartości, a więc przekształcenie φ^* jest różnowartościowe. Jego przeciwdziedziną jest oczywiście zbiór $\varphi(G)$, pozostaje nam zatem sprawdzić, że φ^* jest homomorfizmem:

$$\begin{aligned}&\varphi^*((g \cdot K) * (h \cdot K)) = \\ &\quad = \varphi^*(g \cdot h \cdot K) && \text{definicja } * \\ &\quad = \varphi(g \cdot h) && \text{definicja } \varphi^* \\ &\quad = \varphi(g) \bullet \varphi(h) && \varphi \text{ jest homomorfizmem} \\ &\quad = \varphi^*(g \cdot K) \bullet \varphi^*(h \cdot K) && \text{ponownie definicja } \varphi^*.\end{aligned}$$

∎

Twierdzenie o izomorfizmie ilustrują diagramy przedstawione na rysunku 12.29. Na rysunku 12.29(a) pokazane są grupy i przekształcenia. Twierdzenie mówi, że homomorfizm $\varphi\colon G \to G_0$ indukuje izomorfizm $\varphi^*\colon G/K \to \varphi(G)$ taki, że $\varphi = \varphi^* \circ \nu$. Tutaj ν oznacza naturalny homomorfizm G w G/K z twierdzenia 2. Na rysunku 12.29(b) pokazane jest, jakie wartości przyjmują rozpatrywane homomorfizmy na poszczególnych elementach.



Rysunek 12.29

PRZYKŁAD 7 Pokażemy, że każda grupa cykliczna jest izomorficzna bądź z grupą $(\mathbb{Z}_p, +_p)$, dla pewnej liczby $p \in \mathbb{P}$, bądź z grupą $(\mathbb{Z}, +)$.

Dla danej grupy cyklicznej $\langle g\rangle = \{g^n\colon n \in \mathbb{Z}\}$ zdefiniujmy funkcję $\varphi\colon \mathbb{Z} \to \langle g\rangle$ wzorem $\varphi(n) = g^n$. Ponieważ $\varphi(m+n) = g^{m+n} = g^m \cdot g^n$ (ćwiczenie 14 z § 12.5), więc φ jest homomorfizmem grupy $(\mathbb{Z}, +)$ na grupę $(\langle g\rangle, \cdot)$. Zatem homomorfizm φ ma jądro K i, na mocy twierdzenia o izomorfizmie, grupa $\langle g\rangle$, będąca obrazem $\mathbb{Z}$, jest izomorficzna z grupą $\mathbb{Z}/K$, a izomorfizm dany jest wzorem $\varphi^*(n+K) = g^n$.

Ponieważ K jest podgrupą grupy $\mathbb{Z}$, więc zgodnie z twierdzeniem 3 z § 12.5 $K = p\mathbb{Z}$ dla pewnej liczby $p \in \mathbb{N}$. Wynika stąd, że grupa $\langle g\rangle$ jest izomorficzna z grupą $\mathbb{Z}/p\mathbb{Z}$.

Jeśli $p > 0$, to $|\mathbb{Z}/p\mathbb{Z}| = |\{p\mathbb{Z}, 1+p\mathbb{Z}, \ldots, p-1+p\mathbb{Z}\}| = p$, a więc $|\langle g\rangle| = p$. Przekształcenie MOD $p\colon \mathbb{Z} \to \mathbb{Z}_p$ także jest homomorfizmem o jądrze $p\mathbb{Z}$, zatem grupa $\mathbb{Z}/p\mathbb{Z}$ jest również izomorficzna z grupą $\mathbb{Z}_p$. Wynika stąd, że złożenie funkcji $k \to k + p\mathbb{Z} \to g^k$ jest izomorfizmem grupy $(\mathbb{Z}_p, +_p)$ na grupę $(\langle g\rangle, \cdot)$. (Zobacz przykład 6 dla przypadku $p = 6$).

Jeśli $p = 0$, to jądrem homomorfizmu φ jest zbiór $\{0\}$, a więc homomorfizm φ jest różnowartościowy (wniosek 2 z twierdzenia 1) i grupa $\langle g\rangle$ jest izomorficzna z $\mathbb{Z}$, a izomorfizm ten ustala funkcja dana wzorem $\varphi(n) = g^n$. ∎

Jeśli grupa G jest skończona, to twierdzenie Lagrange'a z § 12.5 mówi, że $|G/K| = |G|/|K|$. Ponieważ, zgodnie z twier-

dzeniem o izomorfizmie, $|\varphi(G)| = |G/K|$, więc otrzymujemy następujący wniosek.

Wniosek

Niech φ będzie homomorfizmem określonym na skończonej grupie G, mającym jądro K. Wówczas $|\varphi(G)| = |G|/|K|$. W szczególności, $|\varphi(G)|$ dzieli $|G|$.

PRZYKŁAD 8

(a) Funkcja φ z grupy $\mathbb{Z}_{30}$ w grupę $\mathbb{Z}_5$ zdefiniowana wzorem $\varphi(n) = n \bmod 5$, jak się okazuje, jest poprawnie zdefiniowanym homomorfizmem. Przyczyna tego leży w tym, że liczby mające te same reszty z dzielenia przez 30 przystają modulo 5, a więc $(m +_{30} n) \bmod 5 = (m + n) \bmod 5 = (m \bmod 5) +_5 (n \bmod 5)$. Jądrem K homomorfizmu φ jest

$$\{n \in \mathbb{Z}_{30}\colon n \bmod 5 = 0\} = \{0, 5, 10, 15, 20, 25\}.$$

Oczywiście, $|\mathbb{Z}_5| = 5 = 30/6 = |\mathbb{Z}_{30}|/|K|$.

(b) Jeśli $(G, \cdot)$ i $(H, \bullet)$ są grupami, to zbiorowi $G \times H$ możemy nadać strukturę grupy przyjmując, że $(g_1, h_1)\square(g_2, h_2) = (g_1 \cdot g_2, h_1 \bullet h_2)$ (zobacz ćwiczenie 11). Zdefiniujmy funkcję $\varphi\colon \mathbb{Z}_6 \to \mathbb{Z}_3 \times \mathbb{Z}_2$ wzorem $\varphi(n) = (n \bmod 3, n \bmod 2)$. Można sprawdzić, podobnie jak w części (a), że φ jest homomorfizmem. Mamy

$$\begin{aligned} &\varphi(0) = (0,0) = \text{element neutralny}, && \varphi(3) = (0,1),\\ &\varphi(1) = (1,1), && \varphi(4) = (1,0),\\ &\varphi(2) = (2,0), && \varphi(5) = (2,1). \end{aligned}$$

Jądrem homomorfizmu φ jest po prostu zbiór $\{0\}$, homomorfizm φ jest różnowartościowy i przekształca $\mathbb{Z}_6$ na $\mathbb{Z}_3 \times \mathbb{Z}_2$, a więc jest izomorfizmem. Grupa $\mathbb{Z}_6$ jest cykliczna, generowana przez 1. Skoro $\varphi(\mathbb{Z}_6) = \varphi(\langle 1\rangle) = \langle\varphi(1)\rangle$, to grupa $\mathbb{Z}_3 \times \mathbb{Z}_2$ także jest cykliczna. Czy to jest niespodzianka? Niezupełnie. Inną postać tego przykładu widzieliśmy w przykładach 5 i 6 z § 12.1.

(c) Przekształcenie $\theta\colon \mathbb{Z}_6 \to \mathbb{Z}_3 \times \mathbb{Z}_2$ dane wzorem $\theta(n) = (n \bmod 3, 0)$ również jest homomorfizmem. Jego jądrem jest zbiór $\{0, 3\}$, a $\theta(\mathbb{Z}_6) = \mathbb{Z}_3 \times \{0\} = \{(m, 0)\colon m \in \mathbb{Z}_3\}$. Jak można się było spodziewać, $|\mathbb{Z}_3 \times \{0\}| = 3 = 6/2 = |\mathbb{Z}_6|/|\{0, 3\}|$. Grupa $\mathbb{Z}_6/\{0, 3\}$ składa się z trzech warstw: $\{0, 3\}$, $\{1, 4\}$ oraz $\{2, 5\}$ i jest w naturalny sposób izomorficzna z grupą $\mathbb{Z}_3$. ∎

Twierdzenie o izomorfizmie przekazuje nam informację, że aby badać homomorficzne obrazy grupy G, wystarczy brać pod uwagę różne grupy postaci G/K, utworzone z warstw pewnych

podgrup K grupy G. Twierdzenie 2 mówi, jak wyróżnić interesujące z tego punktu widzenia podgrupy; są to te podgrupy, które spełniają równość $g \cdot K = K \cdot g$ dla wszystkich elementów g grupy G. Jeśli dobrze znasz grupę i jej dzielniki normalne, to znasz też dobrze jej wszystkie homomorficzne obrazy.

Zamiast porównywać warstwy lewo- i prawostronne, możemy sprawdzić, czy dana podgrupa jest dzielnikiem normalnym przez przyjrzenie się podgrupom z nią **sprzężonym**, tzn. zbiorom postaci $g \cdot K \cdot g^{-1}$. Jeśli K jest podgrupą grupy G, to dla każdego elementu $g \in G$ jest nią również zbiór $g \cdot K \cdot g^{-1}$ (zob. ćwiczenie 26 z § 12.5). Ponadto, $g \cdot K = K \cdot g$ wtedy i tylko wtedy, gdy $g \cdot K \cdot g^{-1} = K \cdot g \cdot g^{-1} = K$, a więc K jest dzielnikiem normalnym wtedy i tylko wtedy, gdy dla każdego g mamy $K = g \cdot K \cdot g^{-1}$, tzn., wtedy i tylko wtedy, gdy wszystkie podgrupy sprzężone z K są równe K. W rzeczywistości, aby udowodnić, że podgrupa K jest dzielnikiem normalnym, wystarczy pokazać, że $g \cdot K \cdot g^{-1} \subseteq K$ dla wszystkich elementów $g \in G$, gdyż jeśli jest to prawdą, to mamy

$$K = g^{-1} \cdot (g \cdot K \cdot g^{-1}) \cdot g \subseteq g^{-1} \cdot K \cdot g \subseteq K,$$

ponieważ g^{-1} też jest elementem grupy G.

ĆWICZENIA DO § 12.6

1. Które z następujących funkcji φ z $(\mathbb{Z}, +)$ w $(\mathbb{Z}, +)$ są homomorfizmami?
(a) $\varphi(n) = 6n$, (b) $\varphi(n) = n + 1$,
(c) $\varphi(n) = -n$, (d) $\varphi(n) = n^2$,
(e) $\varphi(n) = (6n^2 + 3n)/(2n + 1)$.

2. Które z następujących funkcji φ są homomorfizmami $(\mathbb{Z}, +)$ w $(\mathbb{R}\backslash\{0\}, \cdot)$?
(a) $\varphi(n) = 6^n$, (b) $\varphi(n) = n$,
(c) $\varphi(n) = (-6)^n$, (d) $\varphi(n) = n^2$,
(e) $\varphi(n) = 2^{n+1}$.

3. Które z homomorfizmów z ćwiczenia 1 są izomorfizmami? Odpowiedzi krótko uzasadnij.

4. Które z homomorfizmów z punktów (a), (b) i (c) przykładu 1 są izomorfizmami? Odpowiedzi krótko uzasadnij.

5. Niech $F = \text{FUN}(\mathbb{R}, \mathbb{R})$ i niech działanie $+$ w zbiorze F będzie zdefiniowane wzorem

$$(f + g)(x) = f(x) + g(x) \quad \text{dla} \quad x \in \mathbb{R}.$$

(a) Wykaż, że $(F, +)$ jest grupą.

(b) Czy F jest grupą przemienną?
(c) Zdefiniujmy funkcję φ z F w $\mathbb{R}$ wzorem $\varphi(f) = f(73)$. Wykaż, że φ jest homomorfizmem $(F, +)$ w $(\mathbb{R}, +)$.

6. Uzupełnij dowód twierdzenia 2(b).

7. Znajdź jądro każdego z następujących homomorfizmów φ.
(a) $(\mathbb{Z}, +)$ w $(\mathbb{Z}, +)$ dany wzorem $\varphi(n) = 73n$.
(b) $(\mathbb{Z}, +)$ w $(\mathbb{Z}, +)$ dany wzorem $\varphi(n) = 0$ dla wszystkich n.
(c) $(\mathbb{Z}, +)$ w $(\mathbb{Z}_5, +_5)$ dany wzorem $\varphi(n) = n \bmod 5$.
(d) $(\mathbb{Z}, +)$ w $(\mathbb{Z}, +)$ dany wzorem $\varphi(n) = n$.

8. Dla każdego z homomorfizmów z ćwiczenia 7 znajdź tę warstwę jego jądra, do której należy liczba 73.

9. Przypuśćmy, że φ jest pewnym homomorfizmem określonym na grupie G, $|G| = 12$ oraz $|\varphi(G)| = 3$.
(a) Znajdź $|K|$, gdzie K jest jądrem homomorfizmu φ.
(b) Ile elementów grupy G homomorfizm φ przeprowadza na każdy z elementów grupy $\varphi(G)$?
(c) Ile wynosi $|G/K|$?

10. (a) Sprawdź, że przekształcenie φ grupy $(\mathbb{Z}_{p \cdot q})$ w grupę $\mathbb{Z}_p$ zdefiniowane wzorem $\varphi(n) = n \bmod p$ jest homomorfizmem.
(b) Co jest przeciwdziedziną przekształcenia φ?
(c) Co jest jądrem homomorfizmu φ?

11. **Iloczynem prostym** dwu grup $(G, \cdot)$ i $(H, \bullet)$ nazywamy grupę $(G \times H, \square)$ z działaniem $\square$ zdefiniowanym wzorem

$$(g_1, h_1) \square (g_2, h_2) = (g_1 \cdot g_2, h_1 \bullet h_2) \quad \forall g_1, g_2 \in G, \ \forall h_1, h_2 \in H.$$

(a) Znajdź element neutralny grupy $G \times H$ oraz element odwrotny do elementu (g, h) w grupie $G \times H$.
(b) Sprawdź, że przekształcenie π: $G \times H \longrightarrow G$ zdefiniowane wzorem $\pi(g, h) = g$ jest homomorfizmem. (Oznaczenie pi ma związek z angielskim słowem „projection", czyli po polsku „rzut").
(c) Znajdź jądro homomorfizmu π z części (b).
(d) Znajdź dzielnik normalny grupy $G \times H$, który jest izomorficzny z grupą H.

12. Zdefiniujmy przekształcenie φ grupy $(\mathbb{Z}, +)$ w grupę $\mathbb{Z}_2 \times \mathbb{Z}_2$ (zob. ćwiczenie 11) wzorem $\varphi(k) = (k \bmod 2, k \bmod 2)$.
(a) Sprawdź, że φ jest homomorfizmem.
(b) Sprawdź, że $\varphi(\mathbb{Z})$ jest podgrupą grupy $\mathbb{Z}_2 \times \mathbb{Z}_2$.
(c) Znajdź jądro homomorfizmu φ.
(d) Czy grupa $\mathbb{Z}_2 \times \mathbb{Z}_2$ jest izomorficzna z grupą $\mathbb{Z}_4$? Odpowiedź uzasadnij.

13. Niech H będzie zbiorem wszystkich macierzy wymiaru 2×2 postaci $\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$ wraz z iloczynem macierzy jako działaniem grupowym.

(a) Sprawdź, że H jest grupą, w której operacja brania elementu odwrotnego jest określona następująco:

$$\begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -x \\ 0 & 1 \end{bmatrix}.$$

(b) Sprawdź, że

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot H \neq H \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

To pokazuje, że H nie jest dzielnikiem normalnym grupy G wszystkich odwracalnych macierzy wymiaru 2×2 wraz z mnożeniem macierzy jako działaniem grupowym.

(c) Wykaż, że H jest dzielnikiem normalnym grupy T wszystkich macierzy wymiaru 2×2 postaci

$$\begin{bmatrix} y & z \\ 0 & 1/y \end{bmatrix}, \quad y \neq 0,$$

wraz z mnożeniem jako działaniem grupowym.

(d) Przekształcenie φ grupy T w grupę $(\mathbb{R}\backslash\{0\}, \cdot)$ zdefiniowane wzorem

$$\varphi\left(\begin{bmatrix} y & z \\ 0 & 1/y \end{bmatrix}\right) = y$$

jest homomorfizmem. Znajdź jego jądro.

(e) Wykaż, że grupa T/H jest izomorficzna z grupą niezerowych liczb rzeczywistych wraz z mnożeniem jako działaniem grupowym.

14. **Antyhomomorfizmem** grupy $(G, \cdot)$ w grupę $(G_0, \bullet)$ nazywamy funkcję ψ taką, że

$$\psi(g \cdot h) = \psi(h) \bullet \psi(g) \quad \text{dla wszystkich} \quad g, h \in G.$$

(a) Wykaż, że odwzorowanie $g \to g^{-1}$ zawsze jest antyhomomorfizmem danej grupy na siebie. (To przekształcenie ψ można by nazwać antyizomorfizmem, ponieważ jest różnowartościowe i „na").

(b) Kiedy antyizomorfizm z części (a) jest zarazem izomorfizmem?

(c) Wykaż, że jeśli ψ_1 i ψ_2 są antyhomomorfizmami, dla których można określić złożenie $\psi_1 \circ \psi_2$, to funkcja $\psi_1 \circ \psi_2$ jest homomorfizmem.

15. Wykaż, że jeśli φ jest homomorfizmem określonym na grupie G i dla pewnego $g \in G$ przeciwobraz zbioru złożonego z $\varphi(g)$ ma tylko jeden element, to homomorfizm φ jest różnowartościowy.

16. Wykaż, że jeśli J i K są dzielnikami normalnymi grupy $(G, \cdot)$, to zbiór $J \cap K$ też jest jej dzielnikiem normalnym. *Sugestia*: weź pod uwagę $g \cdot (J \cap K) \cdot g^{-1}$.

17. Niech H będzie podgrupą grupy $(G, \cdot)$.

(a) Wykaż, że zbiór $\{g \in G\colon g \cdot H \cdot g^{-1} = H\}$ jest podgrupą grupy G.

(b) Wyciągnij stąd wniosek, że jeśli grupa G jest generowana przez podzbiór A oraz dla wszystkich elementów $g \in A$, $g \cdot H \cdot g^{-1} = H$, to H jest dzielnikiem normalnym grupy G.

18. Weźmy skończoną grupę G oraz jej podgrupę H taką, że $|G| = 2|H|$. Wykaż, że H musi być dzielnikiem normalnym grupy G. *Sugestia*: zob. ćwiczenie 22 z § 12.5.

19. Niech G oznacza grupę symetryczną S_3 i niech $K = S_2 = \langle(12)\rangle$.

(a) Wykaż, że zbiór $e \cdot K \cdot (13) \cdot K$ nie jest warstwą podgrupy K w grupie G.

(b) Czy K jest dzielnikiem normalnym grupy G? Odpowiedź swoją uzasadnij.

§ 12.7. Półgrupy

Często spotyka się struktury mające niektóre, lecz nie wszystkie własności grup. Na przykład, ważne prawo łączności może być spełnione, choćby nawet dla pewnych elementów nie istniały elementy odwrotne. Może nawet nie być elementu neutralnego. Przy omawianiu tego typu struktur, na oznaczenie dwuargumentowego działania używać będziemy małego kwadracika $\square$. Działaniem tym może być dodawanie lub mnożenie, bądź jakieś inne znajome działanie, w rodzaju sumy lub przecięcia, ale też nie musi tak być. Zbiór S wraz z dwuargumentowym działaniem $\square$, które jest łączne, nazywamy **półgrupą**. Zatem $(S, \square)$ jest półgrupą w przypadku, gdy $s_1 \square s_2$ należy do zbioru S dla wszystkich elementów $s_1, s_2 \in S$ oraz

$$s_1 \square (s_2 \square s_3) = (s_1 \square s_2) \square s_3 \quad \text{dla wszystkich} \quad s_1, s_2, s_3 \in S.$$

Wobec łączności działania $\square$, wyrażenia takie, jak $s_1 \square s_2 \square s_3$ są w półgrupie jednoznacznie określone. O działaniach, które nie spełniają prawa łączności, niewiele da się powiedzieć; dotychczas, ilekroć spotykaliśmy się z działaniem łącznym, tylekroć w pobliżu była jakaś półgrupa. Jedyna nowość obecnego podejścia polega na zwróceniu uwagi na ogólne własności takiej struktury.

Tak jak w przypadku grup, jeśli półgrupa jest **przemienna** (tzn., jeśli $s_1 \square s_2 = s_2 \square s_1$ dla wszystkich elementów $s_1, s_2 \in S$), to zamiast $\square$ używamy czasem zapisu addytywnego $+$. Element e zbioru S jest **elementem neutralnym** lub **jedynką półgrupy** S, o ile

$$s \square e = e \square s = s \quad \text{dla wszystkich} \quad s \in S.$$

Półgrupę z jedynką nazywamy **monoidem**. **Grupa** jest więc monoidem, w którym dla każdego elementu istnieje element odwrotny. W tym paragrafie skoncentrujemy się przede wszystkim na półgrupach, które nie są grupami. **Podpółgrupą** danej półgrupy nazywamy jej dowolny podzbiór zamknięty ze względu na związane z tą półgrupą działanie.

PRZYKŁAD 1 (a) $(\mathbb{N}, +)$ jest półgrupą, gdyż suma dwóch dowolnych elementów z $\mathbb{N}$ należy do $\mathbb{N}$ oraz dodawanie jest łączne. Ma ona element neutralny, a mianowicie 0, ponieważ $n + 0 = 0 + n = n$ dla wszystkich liczb $n \in \mathbb{N}$; jest więc monoidem. Żadna z dodatnich liczb należących do $\mathbb{N}$ nie ma w $\mathbb{N}$ elementu przeciwnego (którym w $\mathbb{Z}$ jest liczba do niej przeciwna), zatem $(\mathbb{N}, +)$ nie jest grupą. Monoid $(\mathbb{N}, +)$ jest przemienny.

(b) $(\mathbb{P}, +)$ również jest półgrupą przemienną, będącą podpółgrupą półgrupy $(\mathbb{N}, +)$, ale nie ma elementu neutralnego. Nie ma więc sensu zastanawianie się, czy istnieją elementy odwrotne. ∎

PRZYKŁAD 2 (a) Mnożenie $\cdot$ w zbiorze $\mathbb{R}$ jest łączne i przemienne, a więc dowolny podzbiór zbioru $\mathbb{R}$ zamknięty ze względu na mnożenie jest półgrupą. Zatem $(\mathbb{R}, \cdot)$, $(\mathbb{Q}, \cdot)$, $(\mathbb{Z}, \cdot)$ i $(\mathbb{P}, \cdot)$ są półgrupami przemiennymi. Do każdej z nich należy jedynka, którą jest liczba 1, a więc wszystkie są monoidami. Żadna z tych półgrup nie jest grupą: do każdego ze zbiorów $\mathbb{R}$, $\mathbb{Q}$ i $\mathbb{Z}$ należy 0, a 0 nie ma elementu odwrotnego ze względu na mnożenie. Do półgrupy $\mathbb{P}$ należy liczba 2, ale nie należy liczba do niej odwrotna $\frac{1}{2}$.

(b) Ponieważ iloczyn niezerowych liczb rzeczywistych jest liczbą różną od zera, to mnożenie $\cdot$ jest dwuargumentowym działaniem także w zbiorze $\mathbb{R}\backslash\{0\}$. Co więcej, $(\mathbb{R}\backslash\{0\})$ jest grupą przemienną.

(c) Zbiór $\mathbb{Z}_p$ jest półgrupą przemienną ze względu na działanie $*_p$. Fakt ten jest udowodniony w twierdzeniu 4 z § 3.6. Zauważyliśmy już wcześniej w przykładzie 3(b) z § 12.5, że zbiór Z_p jest grupą przemienną ze względu na działanie $+_p$. ∎

PRZYKŁAD 3 Niech Σ będzie dowolnym alfabetem. Zbiór Σ^* wszystkich słów używających liter alfabetu Σ został nieformalnie zdefiniowany w § 1.1, a następnie zdefiniowany rekurencyjnie w przykładzie 2 z § 7.1. Dwa słowa w_1 i w_2, należące do Σ^*, mnoży się przez **konkatenację**; tzn. w_1w_2 jest słowem otrzymanym w wyniku umieszczenia ciągu w_2 zaraz za ciągiem w_1. Zatem jeśli $w_1 = a_1a_2 \ldots a_m$, a $w_2 = b_1b_2 \ldots b_n$, gdzie wszystkie wyrazy a_j i b_k należą do alfabetu Σ, to $w_1w_2 = a_1a_2 \ldots a_mb_1b_2 \ldots b_n$.

Na przykład, jeśli $w_1 = cat$, a $w_2 = nip$, to $w_1w_2 = catnip$, a $w_2w_1 = nipcat$. Mnożenie przez słowo puste λ nie zmienia słów:

$$w\lambda = \lambda w = w \quad \text{dla wszystkich} \quad w \in \Sigma^*.$$

Wobec przyjętej definicji oczywiste jest, że konkatenacja jest dwuargumentowym działaniem łącznym. Ponieważ puste słowo λ odgrywa rolę jedynki w Σ^*, zbiór Σ^* jest monoidem. Z pewnością nie jest to grupa; jedynie słowo puste ma element odwrotny. ∎

PRZYKŁAD 4 (a) Niech $\mathfrak{M}_{m,n}$ będzie zbiorem wzystkich macierzy wymiaru $m \times n$. Dodawanie macierzy jest dwuargumentowym działaniem w zbiorze $\mathfrak{M}_{m,n}$, gdyż

$$\mathbf{A} + \mathbf{B} \in \mathfrak{M}_{m,n} \quad \text{dla wszystkich} \quad \mathbf{A}, \mathbf{B} \in \mathfrak{M}_{m,n}.$$

Zbiór $\mathfrak{M}_{m,n}$ wszystkich macierzy wymiaru $m \times n$ jest grupą przemienną ze względu na dodawanie. Fakt ten jest sformułowany w twierdzeniu z § 3.3.

(b) Mnożenie macierzy również jest dwuargumentowym działaniem w zbiorze $\mathfrak{M}_{n,n}$ wszystkich macierzy wymiaru $n \times n$. Ze względu na mnożenie zbiór $\mathfrak{M}_{n,n}$ jest monoidem, z jedynką $\mathbf{I}_n$. Prawo łączności omówione jest pod koniec § 3.4. Poza trywialnym przypadkiem $n = 1$, monoid ten nie jest przemienny: macierze $\mathbf{AB}$ i $\mathbf{BA}$ nie muszą być identyczne. Niektóre macierze ze zbioru $\mathfrak{M}_{n,n}$ są odwracalne, lecz nie wszystkie. Zob. ćwiczenia 14 i 15 § 3.4. Podpółgrupa złożona z macierzy odwracalnych jest grupą (przykład 1(b) z § 12.5). ∎

PRZYKŁAD 5 (a) Niech $\mathcal{P}(U)$ będzie zbiorem wszystkich podzbiorów pewnego zbioru U. Zbiór $\mathcal{P}(U)$ wraz z działaniem $\cup$ jest półgrupą przemienną, której jedynką jest $\emptyset$; zob. prawa 1a, 2a i 5a w tablicy 1.1 z § 1.2. Tylko sam zbiór pusty ma element odwrotny, gdyż $A \cup B \neq \emptyset$, jeśli tylko $A \neq \emptyset$.

(b) Zbiór $\mathcal{P}(U)$ jest przemienną półgrupą ze względu na działanie $\cap$, z jedynką U; zob. prawa 1b, 2b oraz 5d w tablicy 1.1 z § 1.2.

(c) Zbiór $\mathcal{P}(U)$ jest także półgrupą z różnicą symetryczną $\div$ jako działaniem; zob. ćwiczenie 2. ∎

PRZYKŁAD 6 Weźmy dowolny zbiór $T = \{a, b, \ldots\}$ mający co najmniej dwa elementy. Zbiór $\text{FUN}(T, T)$ wszystkich funkcji z T w T jest półgrupą ze składaniem funkcji jako działaniem; prawo przemienności zostało omówione w § 1.3. Funkcja identycznościowa 1_T jest jedynką tej półgrupy, ponieważ

$$(1_T \circ f)(t) = 1_T(f(t)) = f(t) \quad \text{dla wszystkich} \quad t \in T,$$

czyli $1_T \circ f = f$ i, analogicznie, $f \circ 1_T = f$. Monoid ten nie jest przemienny. Na przykład, niech f i g będą funkcjami stałymi, zdefiniowanymi wzorami $f(t) = a$ i $g(t) = b$ dla wszystkich elementów t zbioru T. Wtedy $(f \circ g)(t) = f(g(t)) = a$ dla wszystkich t, a $(g \circ f)(t) = b$ dla wszystkich t. To znaczy, że $f \circ g = f \neq g = g \circ f$.

Grupa $\text{PERM}(T)$ wszystkich permutacji zbioru T jest ważną podpółgrupą półgrupy $\text{FUN}(T, T)$. ∎

PRZYKŁAD 7 (a) Niech T będzie niepustym zbiorem. Zbiór $\text{FUN}(T, \mathbb{N})$ wszystkich funkcji przekształcających zbiór T w zbiór $\mathbb{N}$ jest półgrupą ze względu na działanie $+$ zdefiniowane wzorem

$$(f + g)(t) = f(t) + g(t) \quad \text{dla wszystkich} \quad t \in T.$$

(Zauważ, że, na przykład, funkcja $(f + g) + h$ jest zdefiniowana wzorem $((f + g) + h)(t) = (f + g)(t) + h(t) = (f(t) + g(t)) + h(t)$ dla wszystkich $t \in T$). Jedynką półgrupy $\text{FUN}(T, \mathbb{N})$ jest funkcja stała dana wzorem $z(t) = 0$ dla wszystkich $t \in T$. Elementem odwrotnym do funkcji f byłaby funkcja g taka, że $f(t) + g(t) = 0$ dla wszystkich t, tzn. funkcja dana wzorem $g(t) = -f(t)$ dla wszystkich t. Ponieważ jednak zarówno f, jak i g przyjmują wyłącznie wartości w zbiorze $\mathbb{N}$, więc jeśli $f(t) > 0$ dla pewnego t, to takiej funkcji g w naszej półgrupie nie ma.

(b) Zastąpmy w części (a) zbiór $\mathbb{N}$ zbiorem $\mathbb{Z}$. Teraz elementy odwrotne istnieją; elementem odwrotnym do funkcji f jest funkcja $-f$ dana wzorem $(-f)(t) = -f(t)$ dla wszystkich $t \in T$. Zbiór $\text{FUN}(T, \mathbb{Z})$ jest więc grupą ze względu na działanie $+$.

(c) Zbiorowi $\text{FUN}(T, \mathbb{N})$ można także nadać strukturę półgrupy wprowadzając działanie $*$ zdefiniowane wzorem $(f * g)(t) = f(t) \cdot g(t)$. Łączność działania $*$ wynika z łączności mnożenia $\cdot$ w zbiorze $\mathbb{N}$. Jedynką jest funkcja stała e, dana wzorem $e(t) = 1$ dla wszystkich t.

(d) Zwykłe definicje dodawania i mnożenia funkcji, a mianowicie

$$(f + g)(x) = f(x) + g(x) \quad \text{dla wszystkich} \quad x,$$

oraz

$$(f \cdot g)(x) = f(x) \cdot g(x) \quad \text{dla wszystkich} \quad x,$$

nadają zbiorowi $\text{FUN}(\mathbb{R}, \mathbb{R})$ strukturę półgrupy na dwa różne sposoby. ∎

Rozważmy raz jeszcze dowolną półgrupę $(S, \square)$. Dla $s \in S$ i $n \in \mathbb{P}$, stosując znaną już notację, przez s^n oznaczać będziemy

n-krotny $\square$-iloczyn elementu s przez siebie. Jeśli półgrupa S ma jedynkę, to s^0 oznacza e. Następne twierdzenie jest oczywiste, a jego formalny dowód indukcyjny można przeprowadzić korzystając z definicji rekurencyjnej, zgodnie z którą $s^{n+1} = s^n \square s$.

Twierdzenie 1

Niech $(S, \square)$ będzie półgrupą. Dla $s \in S$ oraz $m, n \in \mathbb{P}$, mamy
(a) $s^m \square s^n = s^{m+n}$,
(b) $(s^m)^n = s^{mn}$.
Jeśli $(S, \square)$ jest monoidem, to wzory te są prawdziwe dla $m, n \in \mathbb{N}$.

Potraktowaliśmy działanie $\square$ tak jak zrobilibyśmy to z mnożeniem. Zobaczmy, jak powyższe wzory wyglądałyby, gdybyśmy symbol $\square$ zastąpili symbolem $+$. Iloczyn s^n zastąpiony jest teraz sumą $ns = s + s + \ldots + s$ i twierdzenie mówi, że

(a) $ms + ns = (m + n)s$,
(b) $n(ms) = (mn)s$

dla $s \in S$ i $m, n \in \mathbb{P}$.

Biorąc za punkt wyjścia niepusty podzbiór A zbioru S, rekurencyjnie definiujemy **zbiór A^+ generowany przez A**, w następujący sposób:

(P) $A \subseteq A^+$;
(R) Jeśli $s, t \in A^+$, to $s \square t \in A^+$.

Z twierdzenia 3 dowiemy się, że zbiór A^+ składa sie ze wszystkich iloczynów elementów zbioru A. Mówimy, że **zbiór A generuje półgrupę** S, jeśli $A^+ = S$. Niezależnie od tego, czy zbiór A generuje S, czy nie, na mocy następującego faktu zbiór A^+ jest podpółgrupą półgrupy S.

Twierdzenie 2

Niech A będzie niepustym podzbiorem półgrupy $(S, \square)$. Wówczas A^+ jest najmniejszą podpółgrupą półgrupy S, zawierającą zbiór A.

Dowód. Na mocy warunku (P), A^+ zawiera zbiór A. Wobec warunku (R), zbiór A^+jest zamknięty ze względu na działanie $\square$, zatem zgodnie z definicją, A^+ jest podpółgrupą półgrupy S.

Weźmy teraz dowolną podpółgrupę T półgrupy S zawierającą zbiór A. Pokażemy, że $A^+ \subseteq T$, skąd będzie wynikało, że A^+ jest najmniejszą podpółgrupą zawierającą zbiór A. Chcemy

wykazać, że $s \in T$ dla każdego elementu $s \in A^+$. Ponieważ zbiór A^+ jest zdefiniowany rekurencyjnie, więc naturalne będzie zastosowanie uogólnionej zasady indukcji matematycznej, wprowadzonej w § 7.1. Niech $p(s)$ będzie zdaniem „$s \in T$". Aby wykazać, że zdanie $p(s)$ jest prawdziwe dla wszystkich elementów s zbioru A^+, musimy dowieść, że:

(P′) Zdanie $p(s)$ jest prawdziwe dla tych elementów zbioru A^+, które określa warunek (P).

(I) Jeśli element u zbioru A^+ jest opisany przez warunek (R) za pomocą elementów już wcześniej zdefiniowanych, tj. $u = s \,\square\, t$, to $p(s) \wedge p(t) \Rightarrow p(u)$.

Warunek (P′) jest spełniony, gdyż $A \subseteq T$, ponieważ taki zbiór T wybraliśmy. Warunek (I) stwierdza po prostu, że „jeśli $u = s \,\square\, t$ i $s \in T$ i $t \in T$, to $u \in T$", co jest prawdą, gdyż T jest podpółgrupą. Założenia uogólnionej zasady indukcji są spełnione, a więc $s \in T$ dla wszystkich elementów s zbioru A^+. ∎

Twierdzenie 2 w następujący sposób pomaga nam bez pomocy rekurencji opisać elementy zbioru A^+.

Twierdzenie 3

Niech A będzie niepustym podzbiorem półgrupy $(S, \square)$. Wówczas zbiór A^+ składa się z tych wszystkich elementów półgrupy S, które są postaci $a_1 \square \ldots \square a_n$, gdzie $n \in \mathbb{P}$ oraz $a_1, \ldots, a_n \in A$.

Dowód. Niech X będzie zbiorem złożonym ze wszystkich iloczynów $a_1 \square \ldots \square a_n$. Chcemy wykazać, że $A^+ = X$. Jeśli elementy $a_1, \ldots, a_n, b_1, \ldots, b_m$ należą do zbioru A, to $a_1 \square \ldots \square a_n \square b_1 \square \ldots \square b_m$ jest iloczynem rozpatrywanej postaci. Zatem zbiór X jest zamknięty ze względu na działanie $\square$, tzn. jest podpółgrupą półgrupy S. Dla przypadku, gdy $n = 1$ otrzymujemy że $A \subseteq X$, zatem na mocy twierdzenia 2, $A^+ \subseteq X$.

Aby wykazać, że $X \subseteq A^+$ użyjemy zwykłej indukcji po n. Ponieważ $A \subseteq A^+$, to $a_1 \in A^+$, jeśli tylko $a_1 \in A$. Zróbmy założenie indukcyjne, że $a_1 \square \ldots \square a_n \in A^+$ dla pewnego n i pewnych elementów $a_1, \ldots, a_n$ ze zbioru A. Niech $a_{n+1} \in A$. Ponieważ $a_{n+1} \in A^+$, więc wobec warunku (R) rekurencyjnej definicji zbioru A^+ mamy $a_1 \square \ldots \square a_n \square a_{n+1} \in A^+$. Na mocy zasady indukcji, każdy element zbioru X należy do A^+. ∎

PRZYKŁAD 8

(a) Podpółgrupa $\{2\}^+$ półgrupy $(\mathbb{Z}, +)$ składa się ze wszystkich „iloczynów" elementów zbioru $\{2\}$. Ponieważ naszym dzia-

łaniem jest tutaj $+$, to zbiór $\{2\}^+$ składa się ze wszystkich sum $2+2+\ldots+2$, tzn. ze wszystkich dodatnich parzystych liczb całkowitych. Zatem $\{2\}^+ = 2\mathbb{P} = \{2n\colon n \in \mathbb{P}\}$.

(b) Ogólniej, podpółgrupa półgrupy $(S, \square)$ generowana przez pojedynczy element s jest zbiorem

$$\{s\}^+ = \{s^n\colon n \in \mathbb{P}\}$$

(lub $\{s\}^+ = \{ns\colon n \in \mathbb{P}\}$, jeśli $\square$ to $+$). Taką półgrupę generowaną przez pojedynczy element nazywamy **półgrupą cykliczną**.

(c) Podpółgrupa $\{2\}^+$ półgrupy $(\mathbb{Z}, \cdot)$ jest zbiorem $\{2^n\colon n \in \mathbb{P}\} = \{2, 4, 8, 16, \ldots\}$. Zauważmy, że nasze oznaczenie jest nieprecyzyjne; z samego zapisu $\{2\}^+$, nie możemy wywnioskować, czy chodzi nam o tę półgrupę, czy też o tę z części (a).

(d) Podpółgrupa $\{2,7\}^+$ półgrupy $(\mathbb{Z}, \cdot)$ generowana przez parę $\{2,7\}$ jest zbiorem $\{2^m 7^n\colon m, n \in \mathbb{N} \text{ i } m+n \geq 1\}$.

(e) Pokażemy, że podpółgrupa cykliczna półgrupy $(\mathfrak{M}_{2,2}, \cdot)$ generowana przez macierz

$$\mathbf{M} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

jest równa

$$\{\mathbf{M}\}^+ = \left\{ \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} : n \in \mathbb{P} \right\}.$$

Na mocy części (b) wystarczy wykazać, że $\mathbf{M}^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$ dla $n \in \mathbb{P}$. Jest to jasne dla $n = 1$. Zróbmy założenie indukcyjne, że $\mathbf{M}^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$ dla pewnego $n \in \mathbb{P}$. Wówczas mnożenie macierzy daje

$$\mathbf{M}^{n+1} = \mathbf{M}^n \cdot \mathbf{M} = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1+n \\ 0 & 1 \end{bmatrix}.$$

Dowodzony rezultat wynika teraz z zasady indukcji matematycznej. ∎

Przecięcie dowolnej rodziny podpółgrup dowolnej półgrupy S jest bądź puste, bądź samo jest podpółgrupą półgrupy S. Istotnie, jeśli elementy s i t należą do tego przecięcia, to należą do każdej z półgrup danej rodziny, a więc to samo odnosi się i do ich iloczynu $s \square t$.

PRZYKŁAD 9 (a) Zarówno $2\mathbb{Z}$, jak i $3\mathbb{Z}$ są podpółgrupami półgrupy $(\mathbb{Z}, \cdot)$, a więc ich przecięcie $6\mathbb{Z}$ też. Analogicznie, część wspólna $\mathbb{P} \cap 2\mathbb{Z} = \{2k\colon k \in \mathbb{P}\}$ jest podpółgrupą półgrupy $(\mathbb{Z}, \cdot)$.

(b) Zarówno $\mathbb{P}$, jak i $\{-k\colon k \in \mathbb{P}\}$ są podpółgrupami półgrupy $(\mathbb{Z}, +)$. Ich część wspólna jest pusta.

(c) Rozważmy niepusty podzbiór A półgrupy S. Na mocy twierdzenia 2, część wspólna wszystkich podpółgrup zawierających zbiór A (uwzględniając, naturalnie, sam zbiór S), jest podpółgrupą zawierającą zbiór A. Musi to być najmniejsza podpółgrupa zawierająca A, czyli jest to A^+. Obserwacja ta pokazuje metodę zdefiniowania półgrupy A^+ bez potrzeby opisywania jej elementów za pomocą zbioru A.

PRZYKŁAD 10 W przypadku półgrup teza twierdzenia Lagrange'a może być w spektakularny sposób fałszywa. Weźmy dowolny niepusty zbiór S i określmy działanie $\square$ w S wzorem $s \square t = t$ dla wszystkich $s, t \in S$; tzn. iloczynem dwóch elementów jest po prostu zawsze drugi z nich. Wtedy $(s \square t) \square u = u = s \square u = s \square (t \square u)$, a więc działanie $\square$ jest łączne i $(S, \square)$ jest półgrupą. Każdy niepusty podzbiór zbioru S jest podpółgrupą, ponieważ jest zamknięty ze względu na działanie $\square$. ∎

Wśród funkcji przekształcających półgrupy w półgrupy szczególnie ważne są, podobnie jak dla grup, te, które zachowują strukturę półgrupy. Niech $(S, \bullet)$ i $(T, \square)$ będą półgrupami. Funkcję $\varphi\colon S \to T$ nazywamy **homomorfizmem półgrup**, jeśli

$$\varphi(s_1 \bullet s_2) = \varphi(s_1) \square \varphi(s_2) \quad \text{dla wszystkich} \quad s_1, s_2 \in S.$$

Jeśli φ jest zarazem przekształceniem wzajemnie jednoznacznym, to homomorfizm φ nazywamy **izomorfizmem półgrup**. Jeśli istnieje izomorfizm półgrupy S na półgrupę T, to mówimy, że półgrupy S i T są **izomorficzne** i piszemy $S \simeq T$ (lub $(S, \bullet) \simeq (T, \square)$, jeśli chcemy zwrócić uwagę na odpowiednie działania).

PRZYKŁAD 11 (a) Niech $(S, \bullet)$ oznacza półgrupę $(\mathbb{P}, +)$, a $(T, \square)$ — półgrupę $(\mathbb{P}, \cdot)$. Niech $\varphi(m) = 2^m$ dla $m \in \mathbb{P}$. Ponieważ

$$\varphi(m + n) = 2^{m+n} = 2^m \cdot 2^n = \varphi(m) \cdot \varphi(n),$$

to φ jest homomorfizmem $(\mathbb{P}, +)$ w $(\mathbb{P}, \cdot)$.

(b) W przykładzie 2 z § 12.6 zauważyliśmy, że MOD p jest homomorfizmem grupy $(\mathbb{Z}, +)$ na grupę $(\mathbb{Z}_p, +_p)$, ale że równocześnie MOD p nie jest homomorfizmem grupy $(\mathbb{Z}, \cdot)$ na $(\mathbb{Z}_p, *_p)$ z tej prostej przyczyny, że $(\mathbb{Z}_p, *_p)$ nie jest grupą. Ale $(\mathbb{Z}_p, *_p)$ jest półgrupą i, jak zauważyliśmy w przykładzie 2(b) z § 12.6, MOD p jest homomorfizmem półgrup. ∎

PRZYKŁAD 12 Niech U będzie dowolnym zbiorem. Definiujemy „funkcję dopełnienia" φ z $\mathcal{P}(U)$ w $\mathcal{P}(U)$ wzorem $\varphi(A) = A^c$ dla $A \in \mathcal{P}(U)$.

Na mocy prawa De Morgana z tablicy 1.1 w § 1.2,

$$\varphi(A \cup B) = (A \cup B)^c = A^c \cap B^c = \varphi(A) \cap \varphi(B),$$

a więc φ jest homomorfizmem półgrupy $(\mathcal{P}(U), \cup)$ w półgrupę $(\mathcal{P}(U), \cap)$. Funkcja φ jest przekształceniem wzajemnie jednoznacznym zbioru $\mathcal{P}(U)$ na zbiór $\mathcal{P}(U)$, jest zatem izomorfizmem $(\mathcal{P}(U), \cup)$ na $(\mathcal{P}(U), \cap)$. Jednocześnie funkcja φ jest odwrotna do samej siebie (dlaczego?), więc φ jest także izomorfizmem $(\mathcal{P}(U), \cap)$ na $(\mathcal{P}(U), \cup)$. Zauważmy, że funkcja φ przeprowadza jedynkę $\emptyset$ półgrupy $(\mathcal{P}(U), \cup)$ na jedynkę U półgrupy $(\mathcal{P}(U), \cap)$. ∎

Istnieją dwa różne sposoby wykorzystania izomorfizmów, o czym przekonaliśmy się, badając izomorfizmy grafów. Czasami chcemy zwrócić uwagę na fakt, że dwie na pierwszy rzut oka różne półgrupy mają w istocie identyczną strukturę. Innym razem identyczność struktur jest oczywista, ale chcemy badać różne możliwe izomorfizmy, na przykład półgrupy S na S, aby uzyskać informację o symetriach badanego obiektu.

ĆWICZENIA DO § 12.7

1. Struktura $(\mathbb{N}, \cdot)$ jest półgrupą.
   (a) Czy jest to półgrupa przemienna?
   (b) Czy w tej półgrupie istnieje jedynka?
   (c) Jeśli tak, to czy istnieją elementy odwrotne? Jeśli tak, określ je.
   (d) Czy jest to monoid bądź grupa?
2. Zbiór $\mathcal{P}(U)$ jest półgrupą z różnicą symetryczną $\div$ jako działaniem. Powtórz ćwiczenie 1 w odniesieniu do tej półgrupy.
3. Powtórz ćwiczenie 1 w odniesieniu do półgrupy $\text{FUN}(\mathbb{R}, \mathbb{R})$ wszystkich funkcji o wartościach rzeczywistych określonych na zbiorze $\mathbb{R}$ wraz z działaniem dodawania.
4. Powtórz ćwiczenie 1 w odniesieniu do półgrupy $\text{FUN}(\mathbb{R}, \mathbb{R})$ wraz z działaniem mnożenia.
5. Zbiór $\mathfrak{M}_{2,2}$ wszystkich macierzy wymiaru 2×2 wraz z mnożeniem macierzy jako działaniem jest półgrupą.
   (a) Czy półgrupa ta jest przemienna?
   (b) Czy w półgrupie tej istnieje jedynka?
   (c) Jeśli tak, to czy istnieją elementy odwrotne?
   (d) Czy półgrupa ta jest monoidem bądź grupą?
6. Niech $\Sigma = \{a, b, c, d\}$. Weźmy słowa: $w_1 = bad$, $w_2 = cab$ i $w_3 = abcd$.
   (a) Znajdź słowa w_1w_2, w_2w_1, $w_2w_3w_2w_1$ i $w_3w_2w_3$.
   (b) Znajdź słowa w_1^2, w_2^3 i λ^4.

7. Niech Σ będzie zwyczajnym alfabetem angielskim. Weźmy angielskie słowa: $w_1 = break$, $w_2 = fast$, $w_3 = lunch$ i $w_4 = food$.
(a) Znajdź słowa λw_1, $w_2\lambda$, w_2w_4, w_3w_1 i $w_4\lambda w_4$.
(b) Porównaj ze sobą słowa w_1w_2 i w_2w_1.
(c) Znajdź słowa w_2^2, w_4^2, $w_2^2w_4w_1^2$ i λ^{73}.

8. Wykaż, że zbiór $\mathbb{R}^+ = \{x \in \mathbb{R}\colon x > 0\}$ nie jest półgrupą ze względu na dwuargumentowe działanie $(x, y) \to x/y$.

9. (a) Przekonaj się, że zbiór $\mathbb{N}$ jest półgrupą ze względu na dwuargumentowe działanie $(m, n) \to \min\{m, n\}$ a także ze względu na działanie $(m, n) \to \max\{m, n\}$.
(b) Czy półgrupy z części (a) są monoidami?

10. (a) Wykaż, że zbiór $\mathbb{P}$ jest półgrupą ze względu na działanie $(m, n) \to \mathrm{NWD}(m, n)$, gdzie $\mathrm{NWD}(m, n)$ oznacza największy wspólny dzielnik liczb m i n.
(b) Wykaż, że zbiór $\mathbb{P}$ jest półgrupą ze względu na działanie $(m, n) \to \mathrm{NWW}(m, n)$, gdzie $\mathrm{NWW}(m, n)$ oznacza najmniejszą wspólną wielokrotność liczb m i n.
(c) Czy półgrupy z części (a) i (b) są monoidami?

11. Opisz elementy każdej z następujących podpółgrup półgrupy $(\mathbb{Z}, +)$:
(a) $\{1\}^+$, (b) $\{0\}^+$, (c) $\{-1, 2\}^+$,
(d) $\mathbb{P}^+$, (e) $\mathbb{Z}^+$, (f) $\{2, 3\}^+$,
(g) $\{6\}^+ \cap \{9\}^+$.

12. Opisz elementy każdej z następujących podpółgrup półgrupy $(\mathbb{Z}, \cdot)$:
(a) $\{1\}^+$, (b) $\{0\}^+$, (c) $\{-1, 2\}^+$,
(d) $\mathbb{P}^+$, (e) $\mathbb{Z}^+$, (f) $\{2, 3\}^+$.

13. Które z półgrup z ćwiczenia 11 są półgrupami cyklicznymi? Uzasadnij swoje odpowiedzi.

14. Które z półgrup z ćwiczenia 12 są półgrupami cyklicznymi? Uzasadnij swoje odpowiedzi.

15. Jeśli A jest podzbiorem monoidu $(M, \Box)$, w którym jedynką jest e, to **podmonoid** generowany przez A definiujemy jako zbiór $A^+ \cup \{e\}$. Znajdź następujące monoidy.
(a) Podmonoid monoidu $(\mathbb{Z}, +)$ generowany przez $\{2\}$.
(b) Podmonoid monoidu $(\mathbb{Z}, +)$ generowany przez $\{1, -1\}$.
(c) Podmonoid monoidu $(\mathbb{Z}, +)$ generowany przez $\{0\}$.
(d) Podmonoid monoidu $(\mathbb{Z}, \cdot)$ generowany przez $\{1\}$.
(e) Podmonoid monoidu Σ^*, wraz z konkatenacją jako działaniem, generowany przez Σ.

16. (a) Podaj przykład półgrupy cyklicznej wraz z jej podpółgrupą, która nie jest cykliczna.
(b) Podaj przykład grupy cyklicznej, która nie jest półgrupą cykliczną.
(c) Podaj przykład grupy cyklicznej, która jest półgrupą cykliczną.

17. Wypisz elementy każdej z następujących skończonych podpółgrup półgrupy $(\mathfrak{M}_{3,3}, \cdot)$:

(a) $\left\{ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}^{+}$,

(b) $\left\{ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \right\}^{+}$,

(c) $\left\{ \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix} \right\}^{+}$.

18. (a) Które z podpółgrup z ćwiczenia 17 są grupami?
(b) Które są przemienne?
(c) Które są cykliczne?

19. (a) Znajdź część wspólną podpółgrup $2\mathbb{P}$ i $3\mathbb{P}$ półgrupy $(\mathbb{P}, +)$.
(b) Czy część wspólna, o której mowa w części (a) jest półgrupą cykliczną?
(c) Powtórz część (b) dla półgrupy $(\mathbb{P}, \cdot)$.

20. (a) Znajdź część wspólną podpółgrup półgrupy $(\mathbb{P}, \cdot)$ generowanych, odpowiednio, przez 2 i przez 3.
(b) Czy część wspólna, o której mowa w części (a) jest półgrupą cykliczną? Odpowiedź uzasadnij.

21. (a) Znajdź część wspólną trzech następujących podpółgrup półgrupy $(\mathbb{P}, \cdot)$: $4\mathbb{P}$, $6\mathbb{P}$, $10\mathbb{P}$.
(b) Czy część wspólna z części (a) jest półgrupą cykliczną? Odpowiedź uzasadnij.
(c) Powtórz ćwiczenie (b) dla półgrupy $(\mathbb{P}, +)$.

22. Które z następujących funkcji φ są homomorfizmami półgrupy $(\mathbb{P}, +)$ w półgrupę $(\mathbb{P}, \cdot)$?
(a) $\varphi(n) = 2^n$, (b) $\varphi(n) = n$, (c) $\varphi(n) = (-1)^n$,
(d) $\varphi(n) = 2n$, (e) $\varphi(n) = 2^{n+1}$.

23. Które z homomorfizmów z ćwiczenia 22 są izomorfizmami? Odpowiedź krótko uzasadnij.

24. Niech Σ będzie alfabetem angielskim. Na zbiorze Σ^* zdefiniujmy funkcję φ wzorem $\varphi(w) =$ długość słowa w. Uzasadnij, dlaczego φ jest homomorfizmem półgrupy Σ^* (wraz ze zwykłym działaniem) w półgrupę $(\mathbb{N}, +)$.

25. Wykaż, że jeśli S jest półgrupą, zbiór A generuje S i φ jest homomorfizmem określonym na S, to zbiór $\varphi(A)$ generuje półgrupę $\varphi(S)$.

26. (a) Wykaż, że jeśli φ jest homomorfizmem półgrupy $(S, \bullet)$ w półgrupę $(T, \square)$ i ψ jest homomorfizmem półgrupy $(T, \square)$ w półgrupę $(U, \triangle)$, to $\psi \circ \varphi$ jest homomorfizmem półgrup.

(b) Wykaż, że jeśli φ jest izomorfizmem półgrupy $(S, \bullet)$ na półgrupę $(T, \square)$, to φ^{-1} jest izomorfizmem półgrupy $(T, \square)$ na półgrupę $(S, \bullet)$.

(c) Wykorzystaj wyniki części (a) i (b) do wykazania, że relacja $\simeq$ jest zwrotna, symetryczna i przechodnia.

27. Element z półgrupy $(S, \bullet)$ nazywamy **elementem zerowym** lub **zerem** półgrupy S, jeśli

$$z \bullet s = s \bullet z = z \quad \text{dla wszystkich} \quad s \quad \text{z} \quad S.$$

(a) Wykaż, że półgrupa nie może mieć więcej niż jednego elementu zerowego.

(b) Podaj przykład nieskończonej półgrupy mającej element zerowy.

(c) Podaj przykład skończonej półgrupy o co najmniej dwóch elementach, która ma element zerowy.

28. Niech z będzie elementem zerowym półgrupy $(S, \bullet)$ i niech φ będzie homomorfizmem półgrupy $(S, \bullet)$ w półgrupę $(T, \square)$.

(a) Wykaż, że $\varphi(z)$ jest elementem zerowym półgrupy $\varphi(S)$.

(b) Czy $\varphi(z)$ musi być elementem zerowym półgrupy $(T, \square)$? Uzasadnij swoją odpowiedź.

29. Przypuśćmy, że $(S, \bullet)$ jest monoidem z jedynką e i niech φ będzie homomorfizmem półgrupy $(S, \bullet)$ w półgrupę $(T, \square)$.

(a) Wykaż, że półgrupa $(\varphi(S), \square)$ jest monoidem z jedynką $\varphi(e)$.

(b) Czy element $\varphi(e)$ musi być jedynką półgrupy $(T, \square)$? Uzasadnij swoją odpowiedź.

§ 12.8. Inne systemy algebraiczne

Dotychczas w tym rozdziale zajmowaliśmy się zbiorami z jednym tylko dwuargumentowym działaniem. W wielu ważnych i dobrze znanych strukturach algebraicznych są dwa dwuargumentowe działania, oznaczane najczęściej przez $+$ i $\cdot$. „Dodawanie" $+$ ma na ogół bardzo dobre własności, podczas gdy własności „mnożenia" $\cdot$ są przeważnie nieco gorsze; są też zwykle spełnione prawa rozdzielności, wiążące te dwa działania ze sobą. W tym paragrafie przedstawimy pokrótce pierścienie i ciała, czyli dwa najważniejsze rodzaje struktur algebraicznych z dwoma działaniami, podamy szereg przykładów i omówimy podstawowe fakty dotyczące homomorfizmów takich systemów.

PRZYKŁAD 1 (a) Każdy ze zbiorów $\mathbb{Z}$, $\mathbb{Q}$ i $\mathbb{R}$ jest zamknięty ze względu na zwykłe dodawanie i mnożenie. Zarówno $+$, jak i $\cdot$ są w każdym

z wymienionych zbiorów działaniami przemiennymi i łącznymi. Ponadto spełniają one prawa rozdzielności:

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c) \quad \text{oraz} \quad (a + b) \cdot c = (a \cdot c) + (b \cdot c).$$

(b) Zbiór $\mathfrak{M}_{n,n}$ wszystkich macierzy wymiaru $n \times n$ o wyrazach rzeczywistych jest zamknięty zarówno ze względu na dodawanie, jak i mnożenie macierzy. Oba te działania są łączne, dodawanie jest przemienne, ale mnożenie przemienne nie jest. Na przykład,

$$\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\neq \begin{bmatrix} 5 & 5 \\ -5 & -5 \end{bmatrix} = \begin{bmatrix} 3 & -1 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}.$$

Prawa rozdzielności

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = (\mathbf{AB}) + (\mathbf{AC}) \quad \text{oraz} \quad (\mathbf{A} + \mathbf{B})\mathbf{C} = (\mathbf{AC}) + (\mathbf{BC})$$

są spełnione.

(c) Dla $p \geq 2$, działania $+_p$ i $*_p$ w zbiorze $\mathbb{Z}_p$ są przemienne i łączne, a prawa rozdzielności są spełnione na mocy twierdzenia 4 z § 3.6. ∎

Struktury takie jak w przykładzie 1 występują dostatecznie często, by zasługiwać na odrębną nazwę. **Pierścień** $(R, +, \cdot)$ jest to zbiór R zamknięty ze względu na dwa dwuargumentowe działania, oznaczane na ogół przez $+$ oraz $\cdot$, takie że

(a) $(R, +)$ jest grupą przemienną,
(b) $(R, \cdot)$ jest półgrupą,
(c) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ i $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$ dla wszystkich elementów $a, b, c \in R$.

Jeśli półgrupa $(R, \cdot)$ jest również przemienna, to mówimy, że pierścień $(R, +, \cdot)$ jest **przemienny**. Pierścień $(\mathfrak{M}_{n,n}, +, \cdot)$ z przykładu 1(b) nie jest przemienny, o ile $n > 1$. Pozostałe pierścienie z przykładu 1 są przemienne. Pierścień ma zawsze element neutralny względem dodawania, który oznaczany jest przez 0. Jeśli ma on również różny od 0 element neutralny względem mnożenia, to element ten oznaczamy zwykle przez 1 i mówimy, że pierścień ten jest **pierścieniem z jedynką**. Każdy z pierścieni z przykładu 1 jest pierścieniem z jedynką. Pierścień $(2\mathbb{Z}, +, \cdot)$ złożony z parzystych liczb całkowitych ze zwykłymi działaniami sumy i iloczynu jest przykładem pierścienia przemiennego bez jedynki.

PRZYKŁAD 2 Zbiór POLY $(\mathbb{R})$ wszystkich wielomianów zmiennej x o współczynnikach rzeczywistych jest zamknięty ze względu na $+$ oraz $\cdot$.

Dodawanie wielomianów jest łatwe. Przykładowo,

$$(2x^2 + 3x + 1) + (-5x^4 + 4x - 2) = -5x^4 + 2x^2 + 7x - 1.$$

Jeśli mamy dane wielomiany $p(x)$ i $q(x)$, to dodajemy po prostu do siebie współczynniki przy x^k wielomianów $p(x)$ i $q(x)$, otrzymując współczynnik przy x^k wielomianu $p(x) + q(x)$. Dodawanie jest oczywiście łączne i przemienne; $(\text{POLY}(\mathbb{R}), +)$ jest istotnie grupą przemienną.

Mnożenie jest trochę bardziej skomplikowane. Na przykład,

$$\begin{aligned}(2x^2 - 3x + 1) \cdot (4x^3 - 2x^2) &= 2x^2 \cdot (4x^3 - 2x^2) - 3x \cdot (4x^3 - 2x^2) + 1 \cdot (4x^3 - 2x^2) \\ &= 8x^5 - 4x^4 - 12x^4 + 6x^3 + 4x^3 - 2x^2 \\ &= 8x^5 - 16x^4 + 10x^3 - 2x^2.\end{aligned}$$

Używamy wielokrotnie praw rozdzielności, a następnie zbieramy razem wyrazy, w których x występuje w tej samej potędze.

Współczynnik przy x^k wielomianu będącego iloczynem

$$(a_m x^m + \ldots + a_1 x + a_0) \cdot (b_n x^n + \ldots + b_1 x + b_0),$$

jak się okazuje, jest równy

$$(*) \qquad a_0 \cdot b_k + a_1 \cdot b_{k-1} + \ldots + a_i \cdot b_{k-i} + \ldots + a_k \cdot b_0,$$

gdzie umawiamy się, że $a_i = 0$ dla $i > m$ i $b_j = 0$ dla $j > n$. Można sprawdzić, że tak określone mnożenie jest łączne i przemienne oraz rozdzielne względem dodawania.

Rozdzielność wydaje się być tu sprawą oczywistą. Czyż nie używaliśmy jej definiując iloczyn? Nie, rozdzielności użyliśmy jedynie, aby uzasadnić wprowadzoną definicję iloczynu. W rzeczywistości, gdybyśmy po prostu zaczęli od zdefiniowania działania $\cdot$ za pomocą wyrażenia $(*)$, to nudny dowód pokazałby, że wszystkie inne dobre własności mnożenia już z tej definicji wynikają.

Wraz z tymi działaniami, $(\text{POLY}(\mathbb{R}), +, \cdot)$ jest pierścieniem przemiennym. To samo dotyczy struktur $(\text{POLY}(\mathbb{Z}), +, \cdot)$, $(\text{POLY}(\mathbb{Q}), +, \cdot)$ i $(\text{POLY}(\mathbb{Z}_p), +, \cdot)$, które otrzymamy rozpatrując wielomiany o współczynnikach, odpowiednio, w zbiorze $\mathbb{Z}$, $\mathbb{Q}$ i $\mathbb{Z}_p$. (Gdy współczynniki należą do $\mathbb{Z}_p$, to na nich używamy oczywiście działań $+_p$ i $*_p$). Jedynkami tych pierścieni są wielomiany stałe $e(x) = 1$. ∎

Prawa rozdzielności sprawiają, że obliczenia na elementach pierścienia w znacznym stopniu przypominają dobrze nam znaną arytmetykę w zbiorze $\mathbb{Z}$, z tym oczywistym zastrzeżeniem, że musimy uważać na elementy, które nie są ze sobą przemienne. Na przykład, mamy

$$\begin{aligned}(a+b)^2 &= (a+b)\cdot(a+b) = [a\cdot(a+b)] + [b\cdot(a+b)]\\ &= (a\cdot a) + (a\cdot b) + (b\cdot a) + (b\cdot b)\\ &= a^2 + a\cdot b + b\cdot a + b^2,\end{aligned}$$

ale ostatnia suma nie jest równa $a^2 + 2(a\cdot b) + b^2$, chyba, że $a\cdot b = b\cdot a$. Jako inny przykład, mamy

$$(0\cdot a) + (0\cdot a) = (0+0)\cdot a = 0\cdot a,$$

a więc

$$0\cdot a = (0\cdot a) + (0\cdot a) - (0\cdot a) = (0\cdot a) - (0\cdot a) = 0.$$

Analogicznie, $a\cdot 0 = 0$ dla wszystkich elementów a danego pierścienia. Można też pokazać, że $(-a)\cdot b = -(a\cdot b) = a\cdot(-b)$ (ćwiczenie 8).

Nie możemy liczyć na to. że w jakimś pierścieniu będziemy mogli dzielić przez 0, ale w pierścieniach $(\mathbb{Q},+,\cdot)$ i $(\mathbb{R},+,\cdot)$ możemy dzielić przez dowolny element różny od 0. Z drugiej strony, w pierścieniu $(\mathbb{Z},+,\cdot)$ uda nam się podzielić 6 przez 3, ale nie możemy podzielić 6 przez 5 tak, by wynik pozostał w $\mathbb{Z}$. **Ciało** jest to taki pierścień przemienny $(R,+,\cdot)$, którego niezerowe elementy tworzą grupę ze względu na mnożenie. Jeśli a jest różnym od 0 elementem ciała, to element doń odwrotny oznaczany jest zwykle przez a^{-1} lub $\frac{1}{a}$ i ma on tę własność, że $a^{-1}\cdot a = a\cdot a^{-1} = 1$. Piszemy też b/a zamiast $b\cdot a^{-1}$, co uzasadniają równości $(b/a)\cdot a = b\cdot a^{-1}\cdot a = b$.

Ponieważ zbiór wszystkich niezerowych elementów ciała jest zamknięty ze względu na mnożenie, to ciała mają następującą własność:

(DC) jeśli $a\cdot b = 0$, to $a = 0$ lub $b = 0$.

Własność ta może być spełniona nawet, jeśli nie dla wszystkich elementów istnieją elementy odwrotne. Przykładem jest pierścień $(\mathbb{Z},+,\cdot)$. Pierścień przemienny mający własność (DC) nazywany jest **dziedziną całkowitości**. Pierścienie takie tworzą ważną klasę pośrednią między ciałami, a ogólnymi pierścieniami przemiennymi z jedynką. Są to dokładnie te pierścienie, które speł-

niają następujące **prawo skracania**

$$\text{jeśli} \quad a \cdot c = a \cdot d \quad \text{oraz} \quad a \neq 0, \quad \text{to} \quad c = d,$$

ponieważ z warunków $a \cdot (c - d) = 0$ i $a \neq 0$ wynika, że $c - d = 0$.

Można wykazać, że każda skończona dziedzina całkowitości jest ciałem (ćwiczenie 14(c)).

PRZYKŁAD 3 (a) Grupy są zawsze niepuste, więc w dowolnym ciele element neutralny ze względu na mnożenie jest różny od 0. Najmniejszym możliwym ciałem jest $(\mathbb{Z}_2, +_2, *_2)$, które ma jedynie dwa elementy 0 i 1, a działania na nich opisują poniższe tablice.

$+_2$	0	1
0	0	1
1	1	0

$*_2$	0	1
0	0	0
1	0	1

(b) Zbiór $\text{FUN}(\mathbb{R}, \mathbb{R})$ jest pierścieniem, w którym działania $f + g$ i $f \cdot g$ definiujemy wzorami

$$(f + g)(x) = f(x) + g(x) \quad \text{oraz} \quad (f \cdot g)(x) = f(x) \cdot g(x)$$

dla wszystkich $x \in \mathbb{R}$. Zerem jest funkcja stała 0, dana wzorem $0(x) = 0$ dla wszystkich $x \in \mathbb{R}$. (Zob. ćwiczenie 5 z § 12.6, gdzie omówione jest dodawanie funkcji). Pierścień ten jest przemienny, ale nie jest dziedziną całkowitości, mimo że jego mnożenie zdefiniowane zostało za pomocą mnożenia w ciele $\mathbb{R}$. Na przykład, niech $f(x) = 0$ dla $x \leq 0$ i $f(x) = 1$ dla $x > 0$ oraz niech $g(x) = 1$ dla $x \leq 0$ i $g(x) = 0$ dla $x > 0$. Wówczas $(f \cdot g)(x) = 0$ dla każdego x, a więc $f \cdot g = 0$, ale $f \neq 0$ i $g \neq 0$.

(c) Pierścień przemienny $(\mathbb{Z}_4, +_4, *_4)$ nie jest dziedziną całkowitości; $2 *_4 2 = 0$, ale $2 \neq 0$.

(d) Pierścienie wielomianów $\text{POLY}(\mathbb{R})$, $\text{POLY}(\mathbb{Q})$ i $\text{POLY}(\mathbb{Z})$ są dziedzinami całkowitości. Jeśli $a(x) \neq 0$ i $b(x) \neq 0$, to $a(x) \cdot b(x) \neq 0$. (Wielomian jest równy 0 jedynie wówczas, gdy wszystkie jego współczynniki są równe 0). Nie jest trudno przekonać się, że

$$\begin{aligned}(a_m x^m + \ldots + a_1 x + a_0) \cdot (b_n x^n + \ldots + b_1 x + b_0) = \\ = (a_m \cdot b_n) x^{m+n} + \text{ wyrazy, w których } x \text{ występuje} \\ \text{w niższej potędze.}\end{aligned}$$

Jeśli $a_m \neq 0$ i $b_n \neq 0$, to $a_m \cdot b_n \neq 0$; tzn. jeśli $a(x) \neq 0$ i $b(x) \neq 0$, to $a(x) \cdot b(x) \neq 0$.

(e) Pierścień wielomianów $\text{POLY}(\mathbb{Z}_4)$ nie jest dziedziną całkowitości. Skoro bowiem sam pierścień $(\mathbb{Z}_4, +_4, *_4)$ nie jest dziedziną całkowitości, to istnieją niezerowe wielomiany stałe, których iloczyn jest wielomianem zerowym.

(f) Okazuje, że jeśli p jest liczbą pierwszą, to $\mathbb{Z}_p$ jest ciałem (zob. ćwiczenie16). Rozumowanie z punktu (d) pokazuje, że w tym przypadku pierścień $\text{POLY}(\mathbb{Z}_p)$ jest dziedziną całkowitości. ∎

Podpierścień pierścienia R jest to po prostu podzbiór zbioru R, który sam jest pierścieniem ze względu na działania z pierścienia R. **Podciało ciała** F jest to podpierścień ciała F, który sam jest ciałem; w szczególności, należy do niego jedynka 1 oraz jest on zamknięty na branie elementów odwrotnych.

PRZYKŁAD 4 (a) Podpierścienie pierścienia $(\mathbb{Z}, +, \cdot)$ to między innymi $2\mathbb{Z}$, $73\mathbb{Z}$ i $\{0\}$. W rzeczywistości, na mocy twierdzenia 3 z § 12.5, podpierścienie pierścienia $\mathbb{Z}$ to dokładnie pierścienie postaci $n\mathbb{Z}$, gdzie n jest liczbą całkowitą. Pierścień $(\mathbb{Z}_p, +_p, *_p)$ nie jest podpierścieniem pierścienia $\mathbb{Z}$; pierścienie te mają w rzeczywistości zupełnie różną strukturę. Na przykład, dla każdego elementu a pierścienia $\mathbb{Z}_p$ mamy $a +_p a +_p \ldots +_p a = 0$, jeśli suma ta składa się z p wyrazów, podczas gdy w pierścieniu $\mathbb{Z}$ mamy $a + a + \ldots + a = pa$.

(b) Każdy podpierścień danego ciała, zawierający jedynkę, jest oczywiście dziedziną całkowitości. W szczególności podpierścień $(\mathbb{Z}, +, \cdot)$ ciała $(\mathbb{R}, +, \cdot)$ jest dziedziną całkowitości. Nie jest on ciałem, gdyż jedynie elementy 1 i -1 mają w $\mathbb{Z}$ elementy odwrotne ze względu na mnożenie. Ciało $(\mathbb{Q}, +, \cdot)$ jest podciałem ciała $(\mathbb{R}, +, \cdot)$.

(c) Pierścień wielomianów $(\text{POLY}(\mathbb{Z}), +, \cdot)$ jest podpierścieniem pierścienia $(\text{POLY}(\mathbb{Q}), +, \cdot)$, który sam jest podpierścieniem pierścienia $(\text{POLY}(\mathbb{R}), +, \cdot)$. ∎

Właściwymi przekształceniami z punktu widzenia ich wykorzystania do badania pierścieni są te, które mają związek zarówno ze strukturą dodawania, jak i mnożenia. **Homomorfizmem pierścieni**, pierścienia $(R, +, \cdot)$ w pierścień $(S, +, \cdot)$, nazywamy funkcję $\varphi\colon R \to S$ taką, że

$$\varphi(a + b) = \varphi(a) + \varphi(b) \quad \text{oraz} \quad \varphi(a \cdot b) = \varphi(a) \cdot \varphi(b)$$

dla wszystkich elementów $a, b \in R$. Działania występujące w powyższych równościach po lewej stronie są działaniami pierścienia R, po prawej — pierścienia S. Zatem homomorfizm pierścieni jest to po prostu funkcja będąca zarówno homomorfizmem grup, $(R, +)$ w $(S, +)$, jak i homomorfizmem półgrup, $(R, \cdot)$ w $(S, \cdot)$.

PRZYKŁAD 5 (a) Funkcja φ z $\mathbb{Z}$ w $\mathbb{Z}_p$, zdefiniowana wzorem $\varphi(m) = m \bmod p$, jest homomorfizmem pierścieni, jak to zostało odnotowane we wniosku z twierdzenia 2 z § 3.6.

(b) Funkcja φ z $\mathbb{Z}$ w $\mathbb{Z}$, zdefiniowana wzorem $\varphi(m) = 3m$ jest homomorfizmem grupy $\mathbb{Z}$ w grupę $\mathbb{Z}$, ale nie jest homomorfizmem pierścienia $\mathbb{Z}$ w pierścień $\mathbb{Z}$, ponieważ $\varphi(m \cdot n) = 3mn$, podczas gdy $\varphi(m) \cdot \varphi(n) = 3m \cdot 3n = 9mn$.

(c) Przekształcenie φ zbioru $\text{FUN}(\mathbb{R}, \mathbb{R})$ w $\mathbb{R}$ dane wzorem $\varphi(f) = f(10)$ jest homomorfizmem pierścieni, gdyż

$$\varphi(f+g) = (f+g)(10) = f(10) + g(10) = \varphi(f) + \varphi(g)$$

oraz

$$\varphi(f \cdot g) = (f \cdot g)(10) = f(10) \cdot g(10) = \varphi(f) \cdot \varphi(g).$$

Ogólniej, dla dowolnego zbioru S i dowolnego pierścienia R, w zbiorze $\text{FUN}(S, R)$ można wprowadzić strukturę pierścienia dokładnie tak, jak zrobiliśmy to w przypadku zbioru $\text{FUN}(\mathbb{R}, \mathbb{R})$. Każdemu elementowi $s \in S$ odpowiada homomorfizm φ pierścienia $\text{FUN}(S, R)$ w pierścień R, zwany **homomorfizmem ewaluacji**, dany wzorem $\varphi(f) = f(s)$ dla wszystkich $f \in \text{FUN}(S, R)$.

(d) Pierścień $\text{POLY}(\mathbb{R})$ możemy uważać za podpierścień pierścienia $\text{FUN}(\mathbb{R}, \mathbb{R})$, gdyż każdy wielomian jednoznacznie definiuje funkcję i można pokazać, że funkcje odpowiadające dwóm różnym wielomianom muszą być różne (tzn. mają różne wykresy). Homomorfizm ewaluacji taki, jak $f \to f(73)$, pierścienia $\text{FUN}(\mathbb{R}, \mathbb{R})$ w $\mathbb{R}$ wyznacza homomorfizm pierścienia $\text{POLY}(\mathbb{R})$ w $\mathbb{R}$, zdefiniowany w tym przypadku wzorem $p \to p(73)$. Homomorfizm ewaluacji $p \to p(0)$ przypisuje każdemu wielomianowi jego wyraz wolny a_0.

Ewaluacja jest ważna w projektowaniu algorytmów szybkiego mnożenia bardzo dużych liczb całkowitych przez siebie. Aby zobaczyć ideę tego pomysłu, zaobserwuj, że ewaluacja w punkcie 10 określa odpowiedniości

$$3x^4 + 2x^3 + x + 1 \to 32011$$

oraz

$$2x^2 + 7x + 8 \to 278.$$

Aby pomnożyć 32011 przez 278 moglibyśmy pomnożyć przez siebie dwa powyższe wielomiany: $(3x^4+2x^3+x+1)\cdot(2x^2+7x+8) = p(x)$, a następnie wziąć ewaluację $p(10)$. (Inny przykład znajduje się w ćwiczeniu 4). Wygląda to na bardzo trudny sposób wykonywania mnożenia. Kluczową rolę odgrywa tu jednakże fakt,

że współczynniki wielomianów są małe, a szybka transformata Fouriera daje bardzo szybką metodę mnożenia przez siebie wielomianów wysokich stopni. ∎

Ponieważ homomorfizmy pierścieni są jednocześnie homomorfizmami ich grup addytywnych, więc mają one jądra. Niech φ będzie homomorfizmem pierścienia $(R, +, \cdot)$ w pierścień $(S, +, \cdot)$. Jądrem homomorfizmu φ jest zbiór $K = \{a \in R\colon \varphi(a) = \varphi(0)\}$ i dla dowolnych elementów $a, b \in R$ mamy

$$\varphi(a) = \varphi(b) \Leftrightarrow a - b \in K \Leftrightarrow a + K = b + K.$$

Wszystko tutaj zapisane jest w notacji addytywnej, a więc $a + K$ jest zbiorem $\{a+k\colon k \in K\}$ będącym warstwą podgrupy K w grupie $(R, +)$. Tak jak poprzednio, homomorfizm φ jest różnowartościowy wtedy i tylko wtedy, gdy $K = \{0\}$. Jądro homomorfizmu φ jest szczególnego rodzaju podpierścieniem pierścienia R. Oczywiście jest on podgrupą addytywnej grupy pierścienia, ale jest także zamknięty ze względu na mnożenie i to nie tylko przez swoje własne elementy, ale nawet przez inne elementy pierścienia R:

$$a \in K \text{ i } r \in R \text{ pociąga za sobą, że } a \cdot r,\ r \cdot a \in K.$$

Powód leży w tym, że $\varphi(a) = \varphi(0)$ i jeśli $r \in R$, to

$$\varphi(a \cdot r) = \varphi(a) \cdot \varphi(r) = \varphi(0) \cdot \varphi(r) = \varphi(0 \cdot r) = \varphi(0)$$

i analogicznie $\varphi(r \cdot a) = \varphi(0)$.

Podgrupa I addytywnej grupy pierścienia $(R, +, \cdot)$ nazywana jest **ideałem pierścienia** R, jeśli $r \cdot a \in I$ i $a \cdot r \in I$ dla wszystkich elementów $a \in I$ i $r \in R$. Jak zauważyliśmy w poprzednim akapicie, jądra homomorfizmów pierścieni są ideałami i można też wykazać (ćwiczenie 10), że każdy ideał jest jądrem pewnego homomorfizmu.

PRZYKŁAD 6 (a) Jeśli pierścień R jest przemienny oraz $a \in R$, to zbiór $R \cdot a = \{r \cdot a\colon r \in R\}$ jest ideałem pierścienia R. Aby to sprawdzić, zauważmy, że dla dowolnych elementów $r, s \in R$, mamy $r \cdot a + s \cdot a = (r+s) \cdot a \in R \cdot a$ oraz $-(r \cdot a) = (-r) \cdot a \in R \cdot a$, a więc zbiór $R \cdot a$ jest podgrupą grupy $(R, +)$. Ponieważ $s \cdot (r \cdot a) = (s \cdot r) \cdot a \in R \cdot a$ dla $r, s \in R$, to zbiór $R \cdot a$ jest zamknięty ze względu na mnożenie przez elementy pierścienia R. Ideał postaci $R \cdot a$ nazywany jest **ideałem głównym**.

Zgodnie z twierdzeniem 3 z § 12.5, wszystkie podgrupy grupy $(\mathbb{Z}, +)$ są postaci $n\mathbb{Z}$, a więc wszystkie ideały pierścienia $(\mathbb{Z}, +, \cdot)$ są główne. Okazuje się, że to samo dotyczy ideałów pierścienia

POLY $(\mathbb{R})$, ale sytuacja taka jest bardzo wyjątkowa. Przykładowo, w przemiennym pierścieniu POLY $(\mathbb{Z})$, złożonym z wielomianów o współczynnikach całkowitych, zbiór wszystkich wielomianów $a_n x^n + \ldots + a_1 x + a_0$, w których współczynnik a_0 jest liczbą parzystą, jest ideałem, który nie jest główny (ćwiczenie 17).

(b) Ideały ciał nie są ciekawe. Przypuśćmy, że I jest różnym od zbioru $\{0\}$ ideałem ciała F. Niech $0 \neq a \in I$. Dla każdego elementu $b \in F$ mamy $b = (b \cdot a^{-1}) \cdot a \in F \cdot a \subseteq I$, a więc $I = F$. Znaczy to, że jedyne ideały ciała F to $\{0\}$ i F. ∎

Jeśli I jest ideałem pierścienia R, to grupie R/I, złożonej z addytywnych warstw $r + I = \{r + i\colon i \in I\}$, można w naturalny sposób nadać strukturę pierścienia. Definiujemy

$$(r + I) + (s + I) = (r + s) + I$$

oraz

$$(r + I) \cdot (s + I) = r \cdot s + I.$$

Z twierdzenia 2(b) z § 12.6 wiemy już, że dodawanie w zbiorze R/I jest poprawnie określone; sprawdzimy teraz poprawność definicji mnożenia. Jeśli $r + I = r' + I$ i $s + I = s' + I$, to elementy $r - r'$ i $s - s'$ należą do ideału I, a stąd

$$\begin{aligned} r \cdot s - r' \cdot s' &= r \cdot s - r \cdot s' + r \cdot s' - r' \cdot s' \\ &= r \cdot (s - s') + (r - r') \cdot s' \in r \cdot I \ + I \cdot s' \subseteq I. \end{aligned}$$

Tak więc $r \cdot s + I = r' \cdot s' + I$ i nasza definicja iloczynu nie zależy od wyboru reprezentantów warstw $r + I$ i $s + I$. Inne części definicji pierścienia są łatwe do sprawdzenia.

Twierdzenie o izomorfizmie dla grup prowadzi do analogicznego rezultatu dla pierścieni. Rozważmy homomorfizm φ pierścienia R w pierścień S, o jądrze I. Wówczas funkcja φ jest homomorfizmem grupy $(R, +)$ w grupę $(S, +)$, a więc na mocy twierdzenia o izomorfizmie z § 12.6 wiemy już, że przekształcenie φ^* zbioru R/I w zbiór $\varphi(R)$, dane wzorem $\varphi^*(r + I) = \varphi(r)$ dla $r \in R$, jest izomorfizmem grupy R/I na grupę $\varphi(R)$. Ponieważ

$$\begin{aligned} \varphi^*((r + I) \cdot (r' + I)) &= \varphi^*((r \cdot r') + I) = \varphi(r \cdot r') \\ &= \varphi(r) \cdot \varphi(r') = \varphi^*(r + I) \cdot \varphi^*(r' + I), \end{aligned}$$

to przekształcenie φ^* jest w rzeczywistości **homomorfizmem pierścieni**. Zatem φ^* jest izomorfizmem pierścienia R/I na pierścień $\varphi(R)$, tzn. takim homomorfizmem tych pierścieni, który jest różnowartościowy i „na". Wykazaliśmy co następuje:

Twierdzenie 1

Niech φ będzie homomorfizmem pierścienia R w pierścień S, o jądrze I. Wówczas przekształcenie $r+I \longrightarrow \varphi(r)$ jest izomorfizmem pierścienia R/I na pierścień $\varphi(R)$.

PRZYKŁAD 7

(a) Homomorfizm pierścienia $\mathbb{Z}$ w pierścień $\mathbb{Z}_2$, dany wzorem $n \longrightarrow n \bmod 2$, ma jądro $2\mathbb{Z} =$ PARZYSTE. Pierścień $\mathbb{Z}/2\mathbb{Z}$ ma tylko dwa elementy, PARZYSTE oraz $2\mathbb{Z}+1 =$ NIEPARZYSTE, z działaniami pokazanymi w tablicach poniżej. Izomorfizm pierścienia $\mathbb{Z}/2\mathbb{Z}$ na pierścień $\mathbb{Z}_2$ to: PARZYSTE $\longrightarrow 0$, NIEPARZYSTE $\longrightarrow 1$.

$+$	PARZYSTE	NIEPARZYSTE
PARZYSTE	PARZYSTE	NIEPARZYSTE
NIEPARZYSTE	NIEPARZYSTE	PARZYSTE

$*_2$	PARZYSTE	NIEPARZYSTE
PARZYSTE	PARZYSTE	PARZYSTE
NIEPARZYSTE	PARZYSTE	NIEPARZYSTE

(b) Homomorfizm ewaluacji $p(x) \longrightarrow p(0)$ przekształca POLY$(\mathbb{R})$ na $\mathbb{R}$. Jego jądrem jest zbiór wszystkich wielomianów o wyrazie stałym równym 0, tzn. ideał główny $x \cdot$ POLY$(\mathbb{R})$, złożony ze wszystkich wielokrotności wielomianu x. Warstwy tego ideału są postaci $r + x \cdot$ POLY$(\mathbb{R})$ dla stałych $r \in \mathbb{R}$. Izomorfizmem pierścienia POLY$(\mathbb{R})/(x \cdot$ POLY$(\mathbb{R}))$ na $\mathbb{R}$ jest po prostu funkcja

$$r + x \cdot \text{POLY}(\mathbb{R}) \longrightarrow r.$$

(c) Ogólniej, jądrem ewaluacji $p(x) \longrightarrow p(a)$ jest ideał główny $(x-a) \cdot$ POLY$(\mathbb{R})$ złożony ze wszystkich wielokrotności wielomianu $x-a$ (ćwiczenie 5). Warstwy tego ideału są postaci $r+(x-a)\cdot$ POLY$(\mathbb{R})$, a izomorfizmem pierścienia POLY$(\mathbb{R})/(x-a)\cdot$ POLY$(\mathbb{R}))$ na $\mathbb{R}$ jest funkcja

$$r + (x-a) \cdot \text{POLY}(\mathbb{R}) \longrightarrow r. \qquad \blacksquare$$

PRZYKŁAD 8

W przykładzie 8(b) z § 12.6 możemy dodać jeszcze jedno działanie i nadać zbiorowi $\mathbb{Z}_2 \times \mathbb{Z}_3$ strukturę pierścienia przyjmując, że

$$(m,n) + (j,k) = (m +_2 j, n +_3 k)$$

oraz

$$(m,n) \cdot (j,k) = (m *_2 j, n *_3 k).$$

Przekształcenie φ: $m \longrightarrow (m \bmod 2, m \bmod 3)$ jest homomorfizmem pierścienia $\mathbb{Z}$ na pierścień $\mathbb{Z}_2 \times \mathbb{Z}_3$. Jego jądrem jest zbiór

$$\{m \in \mathbb{Z}\colon m \equiv 0 \pmod 2 \text{ oraz } m \equiv 0 \pmod 3\}$$
$$= \{m \in \mathbb{Z}\colon m \equiv 0 \pmod 6\} = 6\mathbb{Z}.$$

Zatem na mocy twierdzenia 1, pierścień $\varphi(\mathbb{Z})$ jest izomorficzny z pierścieniem $\mathbb{Z}/6\mathbb{Z}$, tzn. jest izomorficzny z pierścieniem $\mathbb{Z}_6$. Znaczy to, że

pierścień $\mathbb{Z}_2 \times \mathbb{Z}_3$ jest izomorficzny z pierścieniem $\mathbb{Z}_6$. ∎

Przykład 8 można uogólnić. Jeśli $R_1, \ldots, R_n$ jest ciągiem pierścieni, niekoniecznie różnych, to iloczynowi kartezjańskiemu $R_1 \times \ldots \times R_n$ można nadać strukturę pierścienia definiując

$$(r_1, \ldots, r_n) + (s_1, \ldots, s_n) = (r_1 + s_1, \ldots, r_n + s_n)$$

oraz

$$(r_1, \ldots, r_n) \cdot (s_1, \ldots, s_n) = (r_1 \cdot s_1, \ldots, r_n \cdot s_n),$$

gdzie działania na k-tej współrzędnej są działaniami odpowiedniego pierścienia R_k. Jeśli $\varphi_1, \ldots, \varphi_n$ są homomorfizmami pewnego pierścienia R w, odpowiednio, $R_1, \ldots, R_n$, to można sprawdzić (ćwiczenie 12), że przekształcenie φ z R w $R_1 \times \ldots \times R_n$, zdefiniowane wzorem $\varphi(r) = (\varphi_1(r), \ldots, \varphi_n(r))$, jest homomorfizmem. Jego jądrem jest zbiór

$$\{r \in R\colon \varphi_k(r) = 0 \text{ dla wszystkich } k = 1, \ldots, n\},$$

tzn. jest to część wspólna jąder homomorfizmów $\varphi_1, \ldots, \varphi_n$.

Przypuśćmy teraz, że $I_1, \ldots, I_n$ są ideałami pierścienia R i niech dla $k = 1, \ldots, n$, φ_k będzie naturalnym homomorfizmem pierścienia R na pierścień R/I_k, danym wzorem $\varphi_k(r) = r + I_k$. Wówczas homomorfizm φ, opisany w poprzednim akapicie, zdefiniowany jest wzorem $\varphi(r) = (r + I_1, \ldots, r + I_n)$ dla $r \in R$. Ponieważ ideał I_k jest jądrem homomorfizmu φ_k, to otrzymujemy co następuje.

Twierdzenie 2

Niech $I_1, \ldots, I_n$ będą ideałami pierścienia R. Wówczas $I_1 \cap \ldots \cap I_n$ jest ideałem pierścienia R oraz pierścień $R/(I_1 \cap \ldots \cap I_n)$ jest izomorficzny z pewnym podpierścieniem pierścienia $(R/I_1) \times \ldots \times (R/I_n)$.

W przykładzie 8 $I_1 = 2\mathbb{Z}$ i $I_2 = 3\mathbb{Z}$ i pierścień $\mathbb{Z}/(2\mathbb{Z} \cap 3\mathbb{Z})$ był izomorficzny z całym pierścieniem $(\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/3\mathbb{Z})$, ale w ogólnej sytuacji pierścień $\varphi(R)$ jest jedynie podpierścieniem pierścienia $R_1 \times \ldots \times R_n$. Na przykład, w pierścieniu $\mathbb{Z}$ mamy $6\mathbb{Z} \cap 10\mathbb{Z} \cap 15\mathbb{Z} = 30\mathbb{Z}$ (sprawdź to). Zatem pierścień postaci

$\mathbb{Z}/(6\mathbb{Z} \cap 10\mathbb{Z} \cap 15\mathbb{Z}) = \mathbb{Z}/30\mathbb{Z}$ ma 30 elementów, podczas gdy pierścień $(\mathbb{Z}/6\mathbb{Z}) \times (\mathbb{Z}/10\mathbb{Z}) \times (\mathbb{Z}/15\mathbb{Z})$ ma $6 \cdot 10 \cdot 15 = 900$ elementów. Inny przykład podany jest w ćwiczeniu 9.

Wniosek

Jeśli $p_1, \ldots, p_n$ są różnymi liczbami pierwszymi, to pierścień $\mathbb{Z}_{p_1 \cdot \ldots \cdot p_n}$ jest izomorficzny z pierścieniem $\mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_n}$.

Dowód. Weźmy ideały $I_1 = p_1\mathbb{Z}, \ldots, I_n = p_n\mathbb{Z}$. Ich część wspólna $I_1 \cap \ldots \cap I_n$ jest równa $p_1 \cdot \ldots \cdot p_n\mathbb{Z}$, ponieważ dowolna liczba, która jest podzielna przez każdą z liczb p_k, musi też być podzielna przez ich iloczyn. Mamy teraz $\mathbb{Z}/I_k = \mathbb{Z}/p_k\mathbb{Z} \simeq \mathbb{Z}_{p_k}$, a więc pierścień $\mathbb{Z}/I_1 \times \ldots \times \mathbb{Z}/I_n$ jest izomorficzny z pierścieniem $\mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_n}$, który ma $p_1 \cdot \ldots \cdot p_n$ elementów, czyli tyle samo, ile pierścień $\mathbb{Z}/(p_1 \cdot \ldots \cdot p_n\mathbb{Z})$. Wobec twierdzenia 2, pierścień $\mathbb{Z}/(p_1 \cdot \ldots \cdot p_n\mathbb{Z})$ jest izomorficzny z podpierścieniem pierścienia $\mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_n}$. Ponieważ oba te pierścienie mają tyle samo elementów oraz $Z_{(p_1 \cdot \ldots \cdot p_n)} \simeq \mathbb{Z}/(p_1 \cdot \ldots \cdot p_n\mathbb{Z})$, więc otrzymujemy stąd dowodzony wniosek. ∎

PRZYKŁAD 9

Powyższy wniosek stanowi szczególny przypadek **chińskiego twierdzenia o resztach**, które mogło być używane w starożytnych Chinach do badania liczebności wojsk. Weźmy niewielką armię, taką, o której wiemy, że ma mniej niż 1000 żołnierzy. Zauważmy, że $7 \cdot 11 \cdot 13 = 1001$ i że nasz wniosek mówi, iż liczby $0, 1, \ldots, 1000$ ze zbioru $\mathbb{Z}_{1001}$ są we wzajemnie jednoznacznej odpowiedniości z trójkami liczb, należącymi do zbioru $\mathbb{Z}_7 \times \mathbb{Z}_{11} \times \mathbb{Z}_{13}$. Odpowiedniość ta wygląda następująco:

$$m \leftrightarrow (m \bmod 7, m \bmod 11, m \bmod 13),$$

a więc, jeśli liczebność armii wynosi N żołnierzy, to znajomość liczb $N \bmod 7$, $N \bmod 11$ i $N \bmod 13$ pozwala nam w zasadzie określić wartość N.

Polećmy żołnierzom zebrać się w grupach po 7 i powiedzieć nam, ilu żołnierzy zostało. Powiedzmy, że został 1. Następnie zgrupujmy ich po 11 i stwierdźmy, że zostało ich, powiedzmy, 2. W końcu zgrupujmy ich po 13 i niech wówczas zostanie ich 8. Informacje te wystarczają do określenia liczby N. Jest ona równa 827.

Do obliczania wartości N potrzebny jest jakiś algorytm lub zespół rachmistrzów. W dzisiejszych czasach zastosowalibyśmy szybką metodę opartą na algorytmie Euklidesa i pomysłach z § 4.6. Moglibyśmy też użyć liczb pierwszych znacznie większych

niż 7, 11 i 13. Liczba pierwsza p, mająca 9 cyfr w układzie dziesiętnym, może być zapisana w komputerze jako słowo 32-bitowe, co oznacza, że obliczeń modulo p można dokonywać szybko. Chińskie twierdzenie o resztach pozwala nam użyć trzech takich dużych liczb pierwszych p, q, r do wyznaczania liczb aż do liczby $p \cdot q \cdot r - 1$, tzn. aż do ok. 10^{30}. Teoria pierścieni pozwala nam przeprowadzać obliczenia całkowitoliczbowe modulo p, q i r, a na koniec zebrać razem ich wyniki. Sytuacja przypomina trochę wykonywanie zdjęć jakiegoś obiektu z trzech różnych stron dla nabrania wyobrażenia o tym, jak wygląda on w całości. ∎

PRZYKŁAD 10 Na interpolację wielomianową można spojrzeć w kontekście twierdzenia 2. Oto prosta ilustracja. Przypuśćmy, że chcemy znaleźć wielomian $p(x)$ należący do zbioru $\text{POLY}(\mathbb{R})$ taki, że $p(1) = 5$, $p(4) = 8$ i $p(6) = 7$ i przypuśćmy, że chcemy również, by miał on stopień co najwyżej 2, to znaczy by $p(x) = ax^2 + bx + c$ dla pewnych nieznanych współczynników a, b i c.

Ewaluacja w punkcie 1 jest homomorfizmem, oznaczmy go przez φ_1, pierścienia $\text{POLY}(\mathbb{R})$ na $\mathbb{R}$. Wiemy, że $\varphi_1(p(x)) = p(1) = 5$. Ewaluacje w punktach 4 i 6 dają kolejne dwa homomorfizmy, φ_4 i φ_6, przy czym $\varphi_4(p(x)) = 8$ i $\varphi_6(p(x)) = 7$. Jak zauważyliśmy w przykładzie 7(c), jądrem homomorfizmu φ_1 jest ideał $I_1 = (x-1) \cdot \text{POLY}(\mathbb{R})$, podczas gdy homomorfizmy φ_4 i φ_6 mają jądra $I_4 = (x-4) \cdot \text{POLY}(\mathbb{R})$ i $I_6 = (x-6) \cdot \text{POLY}(\mathbb{R})$.

Na mocy twierdzenia 2 przekształcenie

$$p(x) + I_1 \cap I_4 \cap I_6 \longrightarrow (p(x) + I_1, p(x) + I_4, p(x) + I_6)$$

jest izomorfizmem pierścienia $\text{POLY}(\mathbb{R})/(I_1 \cap I_4 \cap I_6)$ w iloczyn pierścieni $\text{POLY}(\mathbb{R})/I_1 \times \text{POLY}(\mathbb{R})/I_4 \times \text{POLY}(\mathbb{R})/I_6$, który sam jest izomorficzny z $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Złożeniem tych izomorfizmów jest izomorfizm pierścienia $\text{POLY}(\mathbb{R})/(I_1 \cap I_4 \cap I_6)$ w $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ dany wzorem

$$p(x) + I_1 \cap I_4 \cap I_6 \longrightarrow (p(1), p(4), p(6)).$$

Nasze zagadnienie interpolacyjne polega na znalezieniu takiego wielomianu $p(x)$, który jest przeprowadzony na odpowiednią trójkę $(p(1), p(4), p(6))$. W naszym przykładzie, ciąg (5, 8, 7) odpowiada wielomianowi $p(x) = -0,3x^2 + 2,5x + 2,8$.

Szybkie algorytmy, służące do rozwiązywania zadań dotyczących chińskiego problemu reszt, można dostosować do rozwiązywania zagadnienia interpolacji wielomianowej. Z kolei interpolacja daje następującą metodę mnożenia wielomianów. Aby znaleźć wielomian $a(x) \cdot b(x)$, możemy obliczyć wartości $a(r)$ i $b(r)$ dla wielu argumentów r, w każdym przypadku wykonać mnożenie

$a(r) \cdot b(r)$, a następnie użyć metody interpolacji do znalezienia takiego wielomianu $p(x)$, dla którego $p(r) = a(r) \cdot b(r)$ dla każdego r. Metody takie jak ta są istotą szybkich algorytmów służących do wykonywania działań na dużych liczbach całkowitych w arytmetyce „nieskończonej precyzji".

O pierścieniach i ideałach można powiedzieć znacznie więcej. Wprowadziliśmy jedynie najbardziej podstawowe pojęcia i fakty oraz pokazaliśmy kilka przykładów. Mamy jednak nadzieję, że daliśmy Czytelnikowi jakieś pojęcie o rodzaju pytań, które odnośnie tych systemów rozsądnie jest stawiać oraz o rodzaju odpowiedzi, które można otrzymać. Badanie grup, pierścieni i ciał stanowi znaczną część dziedziny matematyki zwanej algebrą abstrakcyjną. W tym momencie jesteś Czytelniku dobrze przygotowany do przeczytania jakiejś książki wprowadzającej w tę dyscyplinę.

ĆWICZENIA DO § 12.8

W tych ćwiczeniach słowa „homomorfizm" i „izomorfizm" znaczą „homomorfizm pierścieni" i „izomorfizm pierścieni".

1. Które z następujących zbiorów są podpierścieniami pierścienia $(\mathbb{R}, +, \cdot)$?

(a) $2\mathbb{Z}$, (b) $2\mathbb{R}$,
(c) $\mathbb{N}$, (d) $\{m + n\sqrt{2}\colon m, n \in \mathbb{Z}\}$,
(e) $\{m/2\colon m \in \mathbb{Z}\}$, (f) $\{m/2^a\colon m \in \mathbb{Z}, a \in \mathbb{P}\}$.

2. (a) Dla każdego z tych podzbiorów z ćwiczenia 1, które są podpierścieniami pierścienia $\mathbb{R}$, sprawdź, że jest on zamknięty ze względu na dodawanie i mnożenie.

(b) Dla każdego z tych podzbiorów z ćwiczenia 1, które nie są podpierścieniami pierścienia $\mathbb{R}$, wskaż własność przysługującą podpierścieniowi, której zbiór ten nie spełnia.

3. Które z następujących funkcji są homomorfizmami? W każdym z przypadków uzasadnij swoją odpowiedź.

(a) $\varphi\colon \text{FUN}(\mathbb{R}, \mathbb{R}) \to \mathbb{R}$ dana wzorem $\varphi(f) = f(0)$.
(b) $\varphi\colon \mathbb{R} \to \mathbb{R}$ dana wzorem $\varphi(r) = r^2$.
(c) $\varphi\colon \mathbb{R} \to \text{FUN}(\mathbb{R}, \mathbb{R})$ dana wzorem $(\varphi(r))(x) = r$. To znaczy, że $\varphi(r)$ jest funkcją stałą na zbiorze $\mathbb{R}$, przyjmującą dla każdego argumentu x wartość r.
(d) $\varphi\colon \mathbb{Z} \to \mathbb{R}$ dana wzorem $\varphi(n) = n$.
(e) $\varphi\colon \mathbb{Z}/3\mathbb{Z} \to \mathbb{Z}/6\mathbb{Z}$ dana wzorem $\varphi(n + 3\mathbb{Z}) = 2n + 6\mathbb{Z}$.

4. Niech $p(x) = 8x^4 + 3x^3 + x + 7$ i $q(x) = x^4 + x^2 + x + 1$.

(a) Znajdź $p(10)$ i $q(10)$.

(b) Oblicz iloczyn 83017×10111 w trudny sposób, tzn. znajdując wielomian $p(x) \cdot q(x) = r(x)$, a następnie jego wartość $r(10)$. Zwróć uwagę na problem „przeniesień".

5. (a) Wykaż, że jeśli wielomian $p(x) = q(x) \cdot (x - a) + b$ należy do zbioru $\text{POLY}(\mathbb{R})$ i liczby a oraz b należą do $\mathbb{R}$, to $b = p(a)$.

(b) Wykaż, że jeśli

$$q_k(x) = \sum_{i=1}^{k} a^{i-1} x^{k-i} = x^{k-1} + ax^{k-2} + \ldots + a^{k-2}x + a^{k-1},$$

to $x^k - a^k = q_k(x) \cdot (x - a)$.

(c) Wykorzystaj część (b) do wykazania, że jeśli

$$p(x) = \sum_{k=0}^{n} c_k x^k \in \text{POLY}(\mathbb{R}) \text{ oraz } a \in \mathbb{R},$$

to $p(x) = q(x) \cdot (x - a) + p(a)$ dla pewnego wielomianu $q(x) \in \text{POLY}(\mathbb{R})$. To znaczy, udowodnij znane z algebry twierdzenie o reszcie.

(d) Wykaż, że jądrem homomorfizmu ewaluacji $p(x) \longrightarrow p(a)$ pierścienia $\text{POLY}(\mathbb{R})$ w $\mathbb{R}$ jest ideał $(x - a) \cdot \text{POLY}(\mathbb{R})$ składający się z wielokrotności wielomianu $x - a$.

6. Znajdź jądro homomorfizmu φ pierścienia $\text{FUN}(\mathbb{R}, \mathbb{R})$ w $\mathbb{R}$ z przykładu 5(c).

7. Weźmy pierścień $\mathbb{Z}$. Zapisz każdy z następujących zbiorów w postaci $n\mathbb{Z}$, gdzie $n \in \mathbb{N}$:

(a) $6\mathbb{Z} \cap 8\mathbb{Z}$, (b) $6\mathbb{Z} + 8\mathbb{Z}$,

(c) $3\mathbb{Z} + 2\mathbb{Z}$, (d) $6\mathbb{Z} + 10\mathbb{Z} + 15\mathbb{Z}$.

8. Wykaż, że w dowolnym pierścieniu zachodzi $(-a) \cdot b = -(a \cdot b) = a \cdot (-b)$, dla wszystkich elementów a i b.

9. (a) Sprawdź, że przekształcenie φ pierścienia $\mathbb{Z}_{12}$ w pierścień $\mathbb{Z}_4 \times \mathbb{Z}_6$ dane wzorem $\varphi(m) = (m \text{ MOD } 4, m \text{ MOD } 6)$ jest poprawnie zdefiniowanym homomorfizmem.

(b) Znajdź jądro homomorfizmu φ.

(c) Znajdź element zbioru $\mathbb{Z}_4 \times \mathbb{Z}_6$, który nie należy do przeciwdziedziny funkcji φ.

(d) Którym elementom zbioru $\mathbb{Z}_{12}$ przekształcenie φ przyporządkowuje parę (1,3)?

10. Niech I będzie ideałem pierścienia R.

(a) Wykaż, że odwzorowanie $r \longrightarrow r + I$ jest homomorfizmem pierścienia R na pierścień R/I.

(b) Znajdź jądro tego homomorfizmu.

11. (a) Wykaż, że pierścień $(\mathbb{Z}_6, +_6, *_6)$ nie jest ciałem.

(b) Wykaż, że pierścień $(\mathbb{Z}_5, +_5, *_5)$ jest ciałem.

(c) Wykaż, że jeśli pierścienie F i K są ciałami, to pierścień $F \times K$ nie jest ciałem.

12. Jeśli $R_1, \ldots, R_n$ są pierścieniami oraz $\varphi_1, \ldots, \varphi_n$ są homomorfizmami pierścienia R w, odpowiednio, pierścień $R_1, \ldots, R_n$, to przekształcenie φ dane wzorem $\varphi(r) = (\varphi_1(r), \ldots, \varphi_n(r))$ jest homomorfizmem pierścienia R w pierścień $R_1 \times \ldots \times R_n$. Sprawdź ten fakt dla $n = 2$.

13. (a) Wykaż, że jeśli φ jest homomorfizmem ciała F w pierścień R, to albo $\varphi(a) = 0$ dla wszystkich elementów $a \in F$, albo przekształcenie φ jest różnowartościowe.

(b) Wykaż, że jeśli I jest ideałem pierścienia R oraz θ jest homomorfizmem pierścienia R w pierścień S, to $\theta(I)$ jest ideałem pierścienia S.

(c) Wykaż, że jeśli w części (b) pierścień S jest ciałem, to albo $\theta(I) = S$, albo $\theta(I) = \{0\}$.

14. Niech R będzie pierścieniem przemiennym z jedynką.

(a) Wykaż, że R jest dziedziną całkowitości wtedy i tylko wtedy, gdy dla każdego różnego od 0 elementu $a \in R$ przekształcenie R w R dane wzorem $r \to a \cdot r$ jest różnowartościowe.

(b) Wykaż, że R jest ciałem wtedy i tylko wtedy, gdy dla każdego różnego od 0 elementu $a \in R$, odwzorowanie to jest wzajemnie jednoznacznym przekształceniem zbioru R na siebie.

(c) Wykaż, że każda skończona dziedzina całkowitości jest ciałem.

15. (a) Znajdź taki ideał I pierścienia $\mathbb{Z}$, dla którego pierścień $\mathbb{Z}/I$ jest izomorficzny z pierścieniem $\mathbb{Z}_3 \times \mathbb{Z}_5$.

(b) Wskaż izomorfizm między pierścieniami $\mathbb{Z}_{12}$ i $\mathbb{Z}_3 \times \mathbb{Z}_4$.

16. Niech p będzie liczbą pierwszą. Ćwiczenie to pokazuje, że pierścień przemienny $\mathbb{Z}_p$ jest ciałem.

(a) Wykaż, że jeśli $0 < k < p$, to istnieją liczby całkowite s i t takie, że $k \cdot s + p \cdot t = 1$.

(b) Przy oznaczeniach z części (a) wykaż, że w pierścieniu $\mathbb{Z}_p$ element $s \bmod p$ jest elementem odwrotnym ze względu na mnożenie do elementu k. (Istnieje szybki algorytm obliczania s i t, a więc rachunki w $\mathbb{Z}_p$ możemy wykonywać sprawnie nawet wtedy, gdy liczba p jest bardzo duża).

17. Rozważmy pierścień $R = \text{POLY}(\mathbb{Z})$ wielomianów zmiennej x o współczynnikach całkowitych.

(a) Opisz elementy ideałów $R \cdot 2$, $R \cdot x$ i $R \cdot 2 + R \cdot x$.

(b) Wykaż, że w pierścieniu R nie istnieje wielomian p taki, że $R \cdot p = R \cdot 2 + R \cdot x$.

18. Zbiorowi $\mathbb{B} \times \mathbb{B}$ można nadać strukturę pierścienia w inny jeszcze, od podanego w tym paragrafie, sposób. Zdefiniujmy działania $+$ oraz $\bullet$ za pomocą następujących tablic:

$+$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(0,0)$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(1,0)$	$(1,0)$	$(0,0)$	$(1,1)$	$(0,1)$
$(0,1)$	$(0,1)$	$(1,1)$	$(0,0)$	$(1,0)$
$(1,1)$	$(1,1)$	$(0,1)$	$(1,0)$	$(0,0)$

$\bullet$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$	$(0,0)$
$(1,0)$	$(0,0)$	$(1,0)$	$(0,1)$	$(1,1)$
$(0,1)$	$(0,0)$	$(0,1)$	$(1,1)$	$(1,0)$
$(1,1)$	$(0,0)$	$(1,1)$	$(1,0)$	$(0,1)$

Sprawdź, że zbiór $\mathbb{B} \times \mathbb{B}$ jest grupą ze względu na dodawanie oraz że jego elementy różne od 0 tworzą grupę ze względu na mnożenie. *Sugestia*: Oszczędź sobie pracy przez wskazanie grup, które są izomorficzne ze strukturami, o których chcesz dowieść, że są grupami. (Ciała skończone, takie jak to, są ważne przy tworzeniu kodów algebraicznych, minimalizujących efekty zakłóceń podczas transmisji danych.)

To, co jest najważniejsze w tym rozdziale

Jak zwykle: Co to znaczy? Dlaczego tutaj się znalazło? Jak mogę to zastosować? Myśl o przykładach. W rozdziale tym jest wiele pojęć i pomysłów. Główne wątki to podsystemy, homomorfizmy i działania grup. Podczas powtarzania materiału staraj się dostrzec, jak poszczególne zagadnienia wiążą się z tymi tematami.

Pojęcia i oznaczenia

permutacja, grupa permutacji, grupa symetryczna S_n
- cykl, cykle rozłączne, cykl długości m
- grupa działająca na zbiorze X, na zbiorze $\text{FUN}(X, K)$
- orbita
- punkt stały, $\text{FIX}_G(x)$, $\text{FIX}_X(g)$
- obcięcie

zastosowania grup permutacji
- automorfizm grafu lub grafu skierowanego, $\text{AUT}(D)$, $\text{AUT}(H)$
- kolorowanie, kolorowania równoważne

system algebraiczny
- półgrupa, monoid, grupa
  - element neutralny, operacja brania elementu odwrotnego, g^{-1}, $-g$
- pierścień, dziedzina całkowitości, ciało
  - zero, jedynka
- podgrupa (podpierścień itd.), właściwa, trywialna
  - podgrupa $\langle A \rangle$ generowana przez zbiór A
  - podpółgrupa A^+ generowana przez zbiór A
- grupa cykliczna, rząd elementu

homomorfizm, izomorfizm, $\simeq$
- grup, półgrup, pierścieni

jądro
dzielnik normalny, ideał
ideał główny
warstwa, gH, Hg, $g+H$
naturalne działanie w zbiorze G/K, naturalne działania w zbiorze R/I
naturalny homomorfizm $\nu\colon G \to G/K$ lub $R \to R/I$

Fakty dotyczące działań grup na zbiorach

Każda permutacja z grupy S_n jest iloczynem rozłącznych cykli; jej rząd jest równy najmniejszej wspólnej wielokrotności długości tych cykli.

Orbity działania grupy G tworzą podział zbioru, na którym grupa G działa.

Zbiory wyrazów rozłącznych cykli, których iloczynem jest permutacja g, tworzą rodzinę wszystkich $\langle g\rangle$-orbit.

$|G| = |Gx| \cdot |\operatorname{FIX}_G(x)|$ dla każdego elementu x zbioru X, na którym działa grupa G.

Liczba G-orbit w zbiorze X jest równa średniej liczbie punktów stałych elementu grupy G.

Liczba orbit działania grupy G na zbiorze $\operatorname{FUN}(X,K)$ wynosi

$$\frac{1}{|G|}\sum_{g\in G}|K|^{m(g)},$$

gdzie $m(g)$ jest liczbą orbit działania grupy $\langle g\rangle$ na zbiorze X.

Liczba klas złożonych z G-równoważnych kolorowań zbioru X za pomocą k-kolorów wynosi

$$C(k) = \frac{1}{|G|}\sum_{g\in G}k^{m(g)},$$

Ogólne fakty algebraiczne

Skracanie w grupie jest dozwolone. W dziedzinie całkowitości dozwolone jest dzielenie obu stron równości przez elementy różne od 0.

Skończony podzbiór grupy, który jest niepusty i zamknięty ze względu na działanie grupowe, jest podgrupą tej grupy.

Części wspólne podgrup (podpółgrup, podpierścieni, dzielników normalnych itd.) są podgrupami (podpółgrupami itd.).

Podpółgrupa (podgrupa) generowana przez zbiór A składa się z iloczynów elementów zbioru A (i ich odwrotności).

Podgrupami grupy $(\mathbb{Z},+)$ są cykliczne grupy $n\mathbb{Z}$. Są to zarazem wszystkie ideały pierścienia $(\mathbb{Z},+,\cdot)$.

Warstwy podgrupy tworzą podział grupy na zbiory jednakowej liczności (z których jeden jest daną podgrupą).

Twierdzenia Lagrange'a: $|G| = |G/H| \cdot |H|$.

Rząd elementu skończonej grupy G dzieli $|G|$.

Dzielniki normalne są to jądra homomorfizmów grup; ideały są to jądra homomorfizmów pierścieni.

Homomorfizmy przeprowadzają elementy neutralne na elementy neutralne i elementy odwrotne na elementy odwrotne.

Twierdzenie o izomorfizmie: Jeśli K jest jądrem homomorfizmu φ, określonego na grupie G, to $G/K \simeq \varphi(G)$. Analogiczne zdanie jest prawdziwe dla pierścieni.

Homomorfizm grup bądź pierścieni jest różnowartościowy wtedy i tylko wtedy, gdy jego jądro składa się z jednego tylko elementu.

Każda grupa cykliczna jest izomorficzna z $(\mathbb{Z}, +)$ lub $(\mathbb{Z}_p, +_p)$.

Jeśli $I_1, \ldots, I_n$ są ideałami pierścienia R, to $R/(I_1 \cap \ldots \cap I_n)$ jest izomorficzny z pewnym podpierścieniem pierścienia $(R/I_1) \times \ldots \times (R/I_n)$.

Jeśli $p_1, \ldots, p_n$ są różnymi liczbami pierwszymi, to

$$\mathbb{Z}_{p_1 \cdot \ldots \cdot p_n} \simeq \mathbb{Z}_{p_1} \times \ldots \times \mathbb{Z}_{p_n}$$

(jest to szczególny przypadek chińskiego twierdzenia o resztach).

Teoria pierścieni znajduje zastosowanie do interpolacji wielomianowej.

13. RACHUNEK PREDYKATÓW I ZBIORY NIESKOŃCZONE

W tym rozdziale omawiamy zagadnienia, które równie dobrze mogłyby się znaleźć w rozdziałach 2 i 5, ale odłożyliśmy je na później, aby nie zakłócać głównego nurtu myśli przedstawianych w tej książce. Nieformalne wprowadzenie przez nas kwantyfikatorów $\forall$ i $\exists$ w rozdziale 2 było odpowiednie w przypadku zwykłego ich użycia, gdyż nawet stosunkowo niepełne wprowadzenie do rachunku predykatów byłoby znacznie bardziej skomplikowane niż stopień zaawansowania tej książki w tamtym rozdziale. I tak, wyjaśnienia w następnych dwóch paragrafach ledwo sugerują pytania, które mogłyby się pojawić w formalnym wykładzie tego tematu.

Powodem wyłączenia zbiorów nieskończonych z rozdziału, w którym omawiamy zliczanie, jest po prostu to, że idee i metody potrzebne do badania zbiorów nieskończonych są zupełnie inne niż odpowiednie metody dla zbiorów skończonych. Intuicja dotycząca zbiorów skończonych i intuicja dotycząca zbiorów nieskończonych powinny być zupełnie inne.

§ 13.1. Kwantyfikatory

Rachunek zdań opisany w rozdziale 2 jest przyjemną, zupełną i niesprzeczną teorią logiki, ale jest on zupełnie nieprzydatny w większości zagadnień matematycznych. Problem polega na tym, że rachunek zdań nie pozwala na użycie nieskończonej liczby zdań, a sposób notacji jest niewygodny nawet wtedy, kiedy trzeba zapisać duży skończony zbiór zdań. Na przykład, często spotykamy nieskończony ciąg zdań $p(n)$ dla $n \in \mathbb{N}$. Nieformalne

stwierdzenie „zdanie $p(n)$ jest prawdziwe dla wszystkich n" oznacza „zdanie $p(0)$ jest prawdziwe", „zdanie $p(1)$ jest prawdziwe", „zdanie $p(2)$ jest prawdziwe" itd. Jedynym zapisem dostępnym w rachunku zdań jest coś w rodzaju $p(0) \wedge p(1) \wedge p(2) \wedge \ldots$, a tego nie możemy przyjąć. Podobnie, stwierdzenie „zdanie $p(n)$ jest prawdziwe dla pewnego n" odpowiadałoby nieakceptowalnemu wyrażeniu $p(0) \vee p(1) \vee p(2) \vee \ldots$. Aby poradzić sobie z tym problemem, będziemy używać **kwantyfikatorów** $\forall$ i $\exists$, które wprowadziliśmy w § 2.1. Musimy rozwinąć reguły używania tych nowych symboli i łączenia ich ze starymi. Ten rozszerzony system symboli i reguł jest nazywany **rachunkiem predykatów**.

Przypomnijmy, że kwantyfikatory stosuje się do rodzin zdań $\{p(x)\colon\ x \in U\}$; niepusty zbiór U nazywamy **uniwersum**, **zbiorem uniwersalnym**, **przestrzenią** lub **dziedziną**. Zdaniu złożonemu $\forall x\ p(x)$ przypisujemy wartości logiczne w następujący sposób:

> Zdanie $\forall x\, p(x)$ jest prawdziwe, jeśli zdanie $p(x)$ jest prawdziwe dla każdego $x \in U$; w przeciwnym przypadku zdanie $\forall x\, p(x)$ jest fałszywe.

Zdanie złożone $\exists x\, p(x)$ ma następujące wartości logiczne:

> Zdanie $\exists x\, p(x)$ jest prawdziwe, jeśli zdanie $p(x)$ jest prawdziwe dla co najmniej jednego $x \in U$;
> zdanie $\exists x\, p(x)$ jest fałszywe, jeśli zdanie $p(x)$ jest fałszywe dla każdego $x \in U$.

PRZYKŁAD 1 Zdania złożone nie są w pełni zdefiniowane, chyba że określona jest ich dziedzina. Zatem wyrażenie $\forall x(x^2 \geq x)$ jest niejednoznaczne, dopóki nie określimy jego dziedziny, którą może być zbiór $\mathbb{R}$, przedział $[0, \infty)$ czy jakiś inny podzbiór zbioru $\mathbb{R}$. Ogólnie, wartość logiczna zależy od dziedziny. Na przykład, powyższe zdanie jest prawdziwe w dziedzinie $[1, \infty)$, a fałszywe dla $[0, \infty)$ czy dla $\mathbb{R}$. ∎

PRZYKŁAD 2 Czasami spotykamy w algebrze funkcję stałą, taką jak na przykład funkcja $f(x) = 2$ dla $x \in \mathbb{R}$, gdzie zmienna x nie występuje po prawej stronie definicji f. Chociaż wartość $f(x)$ nie zależy od wyboru x, jednakże rozpatrujemy f jako funkcję zmiennej x. Podobnie w logice często spotykamy zdania $p(x)$, których wartość logiczna nie zależy od wyboru $x \in U$. Weźmy dwa sztuczne przykłady: $p(n)$ =„2 jest liczbą pierwszą" oraz $q(n)$ =„16 jest liczbą pierwszą", gdzie dziedziną jest zbiór $\mathbb{N}$. Ponieważ wszystkie zdania $p(n)$ są prawdziwe, to oba zdania $\exists n\, p(n)$ i $\forall n\, p(n)$ są

prawdziwe. Ponieważ wszystkie zdania $q(n)$ są fałszywe, to oba zdania $\exists n\, q(n)$ i $\forall n\, q(n)$ są fałszywe. Takie zdania $p(x)$, których wartość logiczna nie zależy od x, to dokładnie te zdania, którymi zajmowaliśmy się wcześniej w rachunku zdań. W pewnym sensie rachunek zdań zajmuje takie samo miejsce w rachunku predykatów, jak funkcje stałe w badaniu wszystkich funkcji. ∎

Zbadajmy dokładniej zdanie $\forall x\, p(x)$. Wyrażenie $p(x)$ nazywamy **predykatem**. W gramatyce predykat (grupa orzeczenia, predykatyw) jest częścią zdania, która mówi coś o podmiocie zdania. Na przykład, „ ... poleciał na Księżyc" i „ ... jest większa od pojemnika na chleb" są predykatami. Aby utworzyć zdanie, dodajemy podmiot. Na przykład, predykat „ ... jest większa od pojemnika na chleb" stanie się zdaniem „ta książka jest większa od pojemnika na chleb", jeśli podamy podmiot „ta książka". Jeśli oznaczymy predykat symbolem p, to zdanie mogłoby być oznaczone symbolem p(ta książka). Każdy podmiot daje zdanie $p(x)$.

Predykatem w naszym ujęciu symbolicznym jest funkcja, która daje zdanie zawsze wtedy, gdy dostarczamy jej element dziedziny, to znaczy jest funkcją przyjmującą wartości będące zdaniami (funkcją zdaniową), o dziedzinie U. Jak zwykle, oznaczymy taką funkcję przez $p(x)$. Zmienną x w wyrażeniu $p(x)$ nazywamy **zmienną wolną** tego predykatu. Kiedy x przebiega zbiór U, wartość logiczna $p(x)$ może się zmieniać. Dla kontrastu, zdanie $\forall x\, p(x)$ ma dobrze określoną wartość logiczną, która nie zależy od x. Zmienną x w zdaniu $\forall x\, p(x)$ nazywamy **zmienną związaną**; jest ona związana kwantyfikatorem $\forall$. Ponieważ zdanie $\forall x\, p(x)$ ma ustalone znaczenie i wartość logiczną, byłoby niecelowe i nienaturalne kwantyfikowanie go ponownie. To znaczy, że byłoby niepotrzebne wprowadzanie kwantyfikatorów $\forall x[\forall x\, p(x)]$ oraz $\exists x[\forall x\, p(x)]$, ponieważ wartości logiczne tych zdań są takie same jak zdania $\forall x\, p(x)$.

Możemy też rozpatrywać predykaty, które są funkcjami więcej niż jednej zmiennej, być może pochodzących z więcej niż jednej dziedziny i w takich przypadkach jest naturalne użycie wielu kwantyfikatorów.

PRZYKŁAD 3 (a) Niech $\mathbb{N}$ będzie ustaloną dziedziną i dla każdej z liczb m i n ze zbioru $\mathbb{N}$ niech $p(m, n)$ będzie zdaniem „$m < n$". Możemy uważać, że te zdania są indeksowane elementami zbioru $\mathbb{N}\times\mathbb{N}$ i myśleć o $\mathbb{N}\times\mathbb{N}$ jako o dziedzinie, ale na razie wolimy zmienne m i n traktować oddzielnie. Obie zmienne m i n są wolne, w tym sensie, że

znaczenie i wartość logiczna zdania $p(m,n)$ zależy zarówno od m, jak i od n. W wyrażeniu $\exists m\, p(m,n)$ zmienna m jest związana, ale zmienna n jest wolna. Zdanie $\exists m\, p(m,n)$ czytamy: „istnieje liczba $m \in \mathbb{N}$ taka, że $m < n$", zatem zdanie $\exists m\, p(m,0)$ jest fałszywe, zdanie $\exists m\, p(m,1)$ jest prawdziwe, zdanie $\exists m\, p(m,2)$ jest prawdziwe itd. Dla każdego wybranego n zdanie $\exists m\, p(m,n)$ jest albo prawdziwe, albo fałszywe; jego wartość logiczna nie zależy od m, a zależy tylko od samego n. Znaczące jest skwantyfikowanie zdania $\exists m\, p(m,n)$ ze względu na zmienną wolną n, aby otrzymać zdania $\forall n[\exists m\, p(m,n)]$ i $\exists n[\exists m\, p(m,n)]$. Zdanie $\forall n[\exists m\, p(m,n)]$ jest fałszywe, ponieważ zdanie $\exists m\, p(m,0)$ jest fałszywe, a zdanie $\exists n[\exists m\, p(m,n)]$ jest prawdziwe, ponieważ na przykład $\exists m\, p(m,1)$ jest zdaniem prawdziwym. Odtąd będziemy zazwyczaj opuszczać nawiasy kwadratowe [ ] i pisać $\forall n \exists m\, p(m,n)$ i $\exists n \exists m\, p(m,n)$.

Jest osiem sposobów zastosowania dwóch kwantyfikatorów do dwóch zmiennych: $\forall m\, \forall n$, $\forall n\, \forall m$, $\exists m\, \exists n$, $\exists n\, \exists m$, $\forall m\, \exists n$, $\exists n\, \forall m$, $\forall n\, \exists m$, $\exists m\, \forall n$. Pierwsze dwa okazują się być logicznie równoważne, w sensie, który dokładnie zdefiniujemy w § 13.2, ponieważ mają one to samo znaczenie co $\forall (m,n)\, p(m,n)$, gdzie (m,n) przebiega nową dziedzinę $\mathbb{N} \times \mathbb{N}$. Podobnie, zdania $\exists m\, \exists n\, p(m,n)$ oraz $\exists n \exists m\, p(m,n)$ są logicznie równoważne. Do pozostałych czterech należy podchodzić ostrożnie. W naszym ostatnim przykładzie zaobserwowaliśmy już, że $\forall n\, \exists m\, p(m,n)$ jest zdaniem fałszywym. Nieważne, jakie jest m, zdanie $p(m,0)$ jest fałszywe, więc zdanie $\forall n\, p(m,n)$ jest fałszywe, a zatem $\exists m\, \forall n\, p(m,n)$ jest także zdaniem fałszywym. Analizując zdanie $\forall m\, \exists n\, p(m,n)$, zauważ, że dla każdego m zdanie $\exists n\, p(m,n)$ jest prawdziwe, ponieważ zdanie $p(m,m+1)$ jest prawdziwe. Zatem zdanie $\forall m\, \exists n\, p(m,n)$ jest również prawdziwe. Analizując zdanie $\exists n\, \forall m\, p(m,n)$, zauważ, że dla każdego n zdanie $\forall m\, p(m,n)$ jest fałszywe, ponieważ na przykład zdanie $p(n,n)$ jest fałszywe. Zatem zdanie $\exists n\, \forall m\, p(m,n)$ jest również fałszywe. Powtórzmy:

> w tym przykładzie $\forall m\, \exists n\, p(m,n)$ jest zdaniem prawdziwym, podczas gdy $\exists n\, \forall m\, p(m,n)$ jest zdaniem fałszywym.

To pierwsze zdanie mówi, zgodnie z prawdą, że dla każdej liczby m istnieje liczba większa n. Drugie zdanie mówi, że istnieje liczba n większa niż wszystkie liczby m, co jest nieprawdą.

(b) Oto mniej matematyczny przykład ilustrujący, jak ważna jest kolejność w przypadku dwóch kwantyfikatorów $\forall$ i $\exists$. Niech dziedziną będzie zbiór wszystkich ludzi i niech

$$p(x,y) = \text{„}y \text{ jest matką } x\text{"}.$$

Wtedy zdanie $\forall x \exists y\, p(x,y)$ mówi, że każdy ma matkę, co jest prawdą. Z drugiej strony, zdanie $\exists y \forall x\, p(x,y)$ mówi, że jedna osoba jest matką każdego człowieka, co jest nieprawdą.

Zdanie $\forall y \exists x\, p(x,y)$ mówi, że każdy jest matką, a zdanie $\exists x \forall y\, p(x,y)$ mówi, że dla kogoś każdy jest jego lub jej matką. Oczywiście oba te zdania są fałszywe. ∎

Mając w pamięci te przykłady, przechodzimy teraz do bardziej formalnych wyjaśnień. Niech $U_1, U_2, \ldots, U_n$ będą zbiorami niepustymi. **Predykatem n-argumentowym** nad $U_1 \times U_2 \times \ldots \times U_n$ jest funkcja $p(x_1, x_2, \ldots, x_n)$ o dziedzinie $U_1 \times U_2 \times \ldots \times U_n$, której wartościami są zdania. Wszystkie zmienne $x_1, x_2, \ldots, x_n$ w $p(x_1, x_2, \ldots, x_n)$ są **zmiennymi wolnymi** tego predykatu i każda zmienna x_j przebiega odpowiednią dziedzinę U_j. Słowo „wolna" znaczy, że za zmienną x_j można podstawiać daną szczególną wartość ze zbioru U_j w każdym miejscu, w którym występuje x_j.

Jeśli podstawimy jakąś wartość zamiast x_j — powiedzmy, dla ustalenia uwagi, podstawimy a zamiast x_1 w $p(x_1, x_2, \ldots, x_n)$, to otrzymamy predykat $p(a, x_2, \ldots, x_n)$, w którym jest wolnych pozostałych $n-1$ zmiennych $x_2, \ldots, x_n$, ale nie jest już wolna zmienna x_1. Zastosowanie kwantyfikatora $\forall x_j$ lub $\exists x_j$ do predykatu $p(x_1, x_2, \ldots, x_n)$ daje predykat $\forall x_j\, p(x_1, x_2, \ldots, x_n)$ lub $\exists x_j\, p(x_1, x_2, \ldots, x_n)$, którego wartość zależy tylko od wartości pozostałych $n-1$ zmiennych, innych niż x_j. Mówimy, że kwantyfikator wiąże zmienną x_j, czyniąc x_j zmienną związaną dla danego predykatu. Dodanie n kwantyfikatorów, po jednym dla każdej zmiennej, sprawia, że wszystkie zmienne są związane dla tego predykatu i daje zdanie, którego wartość logiczną można określić, jeśli zastosujemy do dziedzin $U_1, U_2, \ldots, U_n$ reguły dotyczące kwantyfikatorów $\forall x$ i $\exists x$ omówione przed przykładem 1.

PRZYKŁAD 4 (a) Weźmy zdanie

$$\forall m \exists n[n > 2^m]; \tag{1}$$

$p(m,n) =$ „$n > 2^m$" jest tutaj predykatem dwuargumentowym nad $\mathbb{N} \times \mathbb{N}$. To znaczy, m i n mogą przebiegać zbiór $\mathbb{N}$. Przypomnijmy naszą umowę dotyczącą opuszczania nawiasów — wyrażenie (1) oznacza

$$\forall m[\exists n[n > 2^m]].$$

Obie zmienne m i n są związane. Aby określić wartość logiczną zdania (1), weźmy wewnętrzne wyrażenie $\exists n[n > 2^m]$, w którym n jest zmienną związaną, a m jest zmienną wolną. Ustalmy

w myśli zmienną wolną m i zauważmy, że zdanie „$n > 2^m$" jest prawdziwe dla pewnych $n \in \mathbb{N}$, na przykład dla $n = 2^m + 1$. Wynika stąd, że $\exists n[n > 2^m]$ jest zdaniem prawdziwym. Ten proces myślowy sprawdza się dla każdego $m \in \mathbb{N}$, więc wnioskujemy, że $\exists n[n > 2^m]$ jest prawdziwe dla wszystkich m. To znaczy, że (1) jest zdaniem prawdziwym.

Jeśli przestawimy kwantyfikatory w wyrażeniu (1), otrzymamy

$$\exists n \forall m[n > 2^m]. \tag{2}$$

Jest to zdanie fałszywe, ponieważ zdanie $\forall m[n > 2^m]$ jest fałszywe dla każdego n, gdyż zdanie „$n > 2^m$" jest fałszywe dla $m = n$.

(b) Weźmy zdania

$$\forall x \exists y[x + y = 0], \tag{3}$$

$$\exists y \forall x[x + y = 0], \tag{4}$$

$$\forall x \exists y[xy = 0], \tag{5}$$

$$\exists y \forall x[xy = 0], \tag{6}$$

których dziedziną jest w każdym przypadku zbiór $\mathbb{R}$.

Aby zbadać (3), weźmy ustaloną liczbę x. Wtedy zdanie $\exists y[x+y = 0]$ jest prawdziwe, ponieważ wybierając $y = -x$, otrzymujemy zdanie prawdziwe „$x+y = 0$". To znaczy, że $\exists y[x+y = 0]$ jest zdaniem prawdziwym dla wszystkich x, a więc (3) jest zdaniem prawdziwym.

Aby zbadać (4), weźmy ustaloną wartość y. Wtedy zdanie $\forall x[x+y = 0]$ nie jest zdaniem prawdziwym, ponieważ wybierając $x = 1 - y$, stwierdzamy, że zdanie „$x + y = 0$" jest fałszywe. To znaczy, że dla każdego y zdanie $\forall x[x + y = 0]$ jest zdaniem fałszywym, a więc (4) jest zdaniem fałszywym.

Zdanie (5) jest prawdziwe, ponieważ zdanie $\exists y[xy = 0]$ jest prawdziwe dla wszystkich x. Rzeczywiście, wybierając $y = 0$, otrzymujemy zdanie prawdziwe „$xy = 0$".

Aby uporać się z (6), zbadamy zdanie $\forall x[xy = 0]$. Jeśli $y = 0$, zdanie to jest oczywiście prawdziwe. Ponieważ zdanie $\forall x[xy = 0]$ jest prawdziwe dla pewnego y, mianowicie dla $y = 0$, zdanie (6) jest zdaniem prawdziwym.

W następnym paragrafie zobaczymy, że z prawdziwości zdania (6) wynika na podstawie czysto logicznych rozważań prawdziwość zdania (5); to znaczy, że zdanie

$$\exists y \forall x p(x, y) \rightarrow \forall x \exists y p(x, y)$$

jest zawsze prawdziwe. ∎

Zauważyliśmy już, że n-argumentowy predykat staje się predykatem $(n-1)$-argumentowym, kiedy związemy jedną ze zmiennych kwantyfikatorem. Jego wartość logiczna zależy od pozostałych $n-1$ zmiennych wolnych i w szczególności nie zależy od tego, jak nazwiemy zmienną związaną. Zatem, jeśli $p(x)$ jest jednoargumentowym predykatem o dziedzinie U, to wszystkie zdania $\forall x\, p(x)$, $\forall y\, p(y)$ i $\forall t\, p(t)$ mają tę samą wartość logiczną, mianowicie wartość prawdy, jeśli $p(u)$ jest zdaniem prawdziwym dla każdego u i wartość fałszu w przeciwnym przypadku. Podobnie, jeśli $q(x,y)$ jest 2-argumentowym predykatem o dziedzinach U i V, to wszystkie zdania $\exists y\, q(x,y)$, $\exists t\, q(x,t)$, $\exists s\, q(x,s)$ opisują ten sam jednoargumentowy predykat, mianowicie predykat, który ma wartość logiczną prawdy dla danego $x \in U$ wtedy i tylko wtedy, gdy $q(x,v)$ jest zdaniem prawdziwym dla jakiegoś v z dziedziny V, do której należy druga zmienna. Z drugiej strony, predykat $\exists x\, q(x,x)$ nie jest taki sam jak ostatnie trzy. Różnica polega na tym, że kwantyfikator w tym przypadku wiąże obie zmienne wolne.

PRZYKŁAD 5 Niech U i V będą równe $\mathbb{N}$ i niech $q(x,y) =$„$x > y$”. Wtedy $\exists x\, q(x,y)$ jest jednoargumentowym predykatem „pewien element zbioru $\mathbb{N}$ jest większy od y”, tak samo jak $\exists t\, q(t,y)$. Predykat $\exists y\, q(x,y)$ jest jednoargumentowym predykatem „istnieje element zbioru $\mathbb{N}$ mniejszy od x”, który jest tym samym predykatem co $\exists s\, q(x,s)$, i ma on wartość prawdy dla $x > 0$ i fałszu dla $x = 0$. Ale $\exists x\, q(x,x)$ jest zupełnie innym zdaniem „$x > x$ dla pewnego x” i jego wartością logiczną jest fałsz. ∎

ĆWICZENIA DO § 13.1

Tak jak w rozdziale 2, wartości logiczne „prawda” i „fałsz” można zapisywać, odpowiednio, jako 1 i 0.

1. Podaj wartości logiczne następujących wyrażeń, gdzie dziedziną jest zbiór $\mathbb{N}$:

(a) $\forall m\, \exists n[2n = m]$,
(b) $\exists n\, \forall m[2m = n]$,
(c) $\forall m\, \exists n[2m = n]$,
(d) $\exists n\, \forall m[2n = m]$,
(e) $\forall m\, \forall n[\neg\{2n = m\}]$.

2. Podaj wartości logiczne następujących wyrażeń, gdzie dziedziną jest zbiór $\mathbb{R}$:

(a) $\forall x\, \exists y[xy = 1]$,
(b) $\exists y\, \forall y[xy = 1]$,
(c) $\exists x\, \exists y[xy = 1]$,
(d) $\forall x\, \forall y[(x+y)^2 = x^2+y^2]$,
(e) $\forall x\, \exists y[(x+y)^2 = x^2+y^2]$,
(f) $\exists y\, \forall x[(x+y)^2 = x^2+y^2]$.
(g) $\exists x\, \exists y[(x+2y=4) \wedge (2x-y=2)]$,
(h) $\exists x\, \exists y[x^2+y^2+1 = 2xy]$.

3. Zapisz następujące zdania za pomocą symboliki logicznej. Sprawdź, czy związałeś wszystkie zmienne. Kiedy używasz kwantyfikatorów, określ dziedziny; weź zbiór $\mathbb{R}$, jeśli dziedzina nie jest zaznaczona.
(a) Jeśli $x < y$ i $y < z$, to $x < z$.
(b) Dla każdej liczby $x > 0$ istnieje liczba $n \in \mathbb{N}$ taka, że $n > x$ oraz $x > 1/n$.
(c) Dla każdych liczb $m, n \in \mathbb{N}$ istnieje liczba $p \in \mathbb{N}$ taka, że $m < p$ oraz $p < n$.
(d) Istnieje liczba $u \in \mathbb{N}$ taka, że $un = n$ dla wszystkich $n \in \mathbb{N}$.
(e) Dla każdej liczby $n \in \mathbb{N}$ istnieje liczba $m \in \mathbb{N}$ taka, że $m < n$.
(f) Dla każdej liczby $n \in \mathbb{N}$ istnieje liczba $m \in \mathbb{N}$ taka, że $2^m \leq n$ oraz $n < 2^{m+1}$.

4. Podaj wartość logiczną zdań z ćwiczenia 3.

5. Zapisz następujące zdania za pomocą symboliki logicznej; dziedziną jest zbiór Σ^* słów utworzonych z liter skończonego alfabetu Σ.
(a) Jeśli $w_1 w_2 = w_1 w_3$, to $w_2 = w_3$.
(b) Jeśli długość$(w) = 1$, to $w \in \Sigma$.
(c) $w_1 w_2 = w_2 w_1$ dla wszystkich $w_1, w_2 \in \Sigma^*$.

6. Podaj wartość logiczną zdań z ćwiczenia 5.

7. Określ, które zmienne w następujących wyrażeniach są wolne, a które związane:
(a) $\forall x \, \exists z [\sin(x + y) = \cos(z - y)]$,
(b) $\exists x [xy = xz \rightarrow y = z]$,
(c) $\exists x \, \exists z [x^2 + z^2 = y]$.

8. Weźmy wyrażenie $x + y = y + x$.
(a) Określ zmienne wolne i związane w tym wyrażeniu.
(b) Dodaj kwantyfikatory ogólne dla dziedziny $\mathbb{R}$, aby otrzymać zdanie. Czy to zdanie jest prawdziwe?
(c) Dodaj kwantyfikatory szczegółowe dla dziedziny $\mathbb{R}$, aby otrzymać zdanie. Czy to zdanie jest prawdziwe?

9. Powtórz ćwiczenie 8 dla wyrażenia $(x - y)^2 = x^2 - y^2$.

10. Weźmy zdanie $\forall m \exists n [m + n = 7]$.
(a) Czy to zdanie jest prawdziwe, jeśli każdą z dziedzin jest zbiór $\mathbb{N}$?
(b) Czy to zdanie jest prawdziwe, jeśli każdą z dziedzin jest zbiór $\mathbb{Z}$?

11. Powtórz ćwiczenie 10 dla zdania $\forall n \, \exists m [m + 1 = n]$.

12. Weźmy zdanie $\forall x \, \exists y [(x^2 + 1)y = 1]$.
(a) Czy to zdanie jest prawdziwe, jeśli każdą z dziedzin jest zbiór $\mathbb{N}$?
(b) Czy to zdanie jest prawdziwe, jeśli każdą z dziedzin jest zbiór $\mathbb{Q}$?
(c) Czy to zdanie jest prawdziwe, jeśli każdą z dziedzin jest zbiór $\mathbb{R}$?

13. Innym przydatnym kwantyfikatorem jest kwantyfikator $\exists!$, gdzie napis $\exists! x\, p(x)$ czytamy „istnieje dokładnie jeden taki x, że $p(x)$". To zdanie złożone ma wartość logiczną prawdy, jeśli $p(x)$ jest prawdziwe dla dokładnie jednej wartości x w dziedzinie; w przeciwnym przypadku jest ono fałszywe. Zapisz następujące zdania za pomocą symboliki logicznej.

(a) Istnieje dokładnie jedna liczba $x \in \mathbb{R}$ taka, że $x+y = y$ dla wszystkich $y \in \mathbb{R}$.

(b) Równanie $x^2 = x$ ma dokładnie jedno rozwiązanie.

(c) Dokładnie jeden zbiór jest podzbiorem wszystkich zbiorów z $\mathcal{P}(\mathbb{N})$.

(d) Jeśli $f\colon\ A \to B$, to dla każdego $a \in A$ istnieje dokładnie jedno $b \in B$ takie, że $f(a) = b$.

(e) Jeśli $f\colon\ A \to B$ jest funkcją różnowartościową, to dla każdego $b \in B$ istnieje dokładnie jedno $a \in A$ takie, że $f(a) = b$.

14. Podaj wartość logiczną zdań z ćwiczenia 13.

15. W tym zadaniu $A = \{0, 2, 4, 6, 8, 10\}$, a dziedziną jest zbiór $\mathbb{N}$. Czy następujące zdania są prawdziwe czy fałszywe?

(a) A jest zbiorem liczb parzystych w $\mathbb{N}$ mniejszych od 12.

(b) $A = \{0, 2, 4, 6, \ldots\}$.

(c) $A = \{n \in \mathbb{N}\colon\ 2n < 24\}$.

(d) $A = \{n \in \mathbb{N}\colon\ \forall m[2m = n \ \to\ m < 6]\}$.

(e) $A = \{n \in \mathbb{N}\colon\ \forall m[2m = n \wedge m < 6]\}$.

(f) $A = \{n \in \mathbb{N}\colon\ \exists m[2m = n \ \to\ m < 6]\}$.

(g) $A = \{n \in \mathbb{N}\colon\ \exists m[2m = n \wedge m < 6]\}$.

(h) $A = \{n \in \mathbb{N}\colon\ \exists! m[2m = n \wedge m < 6]\}$.

(i) $A = \{n \in \mathbb{N}\colon\ n$ jest liczbą parzystą i $n^2 \leq 100\}$.

(j) $\forall n[n \in A \ \to\ n \leq 10]$.

(k) $3 \in A \ \to\ 3 < 10$.

(l) $12 \in A \ \to\ 12 < 10$.

(m) $8 \in A \ \to\ 8 < 10$.

16. Niech dziedziną będzie zbiór $\mathbb{N}$ i niech $p(n) =$„n jest liczbą pierwszą" oraz $e(n) =$„n jest liczbą parzystą". Zapisz następujące wyrażenia w języku polskim:

(a) $\exists m\, \forall n[e(n) \wedge p(m+n)]$,

(b) $\forall n \exists m[\neg e(n) \ \to\ e(m+n)]$.

Zapisz następujące zdania za pomocą symboliki logicznej, używając p i e.

(c) Istnieją dwie liczby pierwsze, których suma jest liczbą parzystą.

(d) Jeśli suma dwóch liczb pierwszych jest parzysta, to żadna z tych liczb nie jest równa 2.

(e) Suma dwóch liczb pierwszych jest nieparzysta.

17. Podaj wartość logiczną zdań z ćwiczenia 16.

§ 13.2. Elementarny rachunek predykatów

Jednym z celów w rozdziale 2 było nauczenie się, jak rozpoznać, czy dwa zdania złożone są logicznie równoważne lub czy jedno zdanie logicznie implikuje inne zdanie. Takie pytania są nawet jeszcze ważniejsze, a także trochę trudniejsze, kiedy dopuszczamy użycie kwantyfikatorów $\forall$ i $\exists$ przy tworzeniu wyrażeń złożonych. Kiedy mieliśmy do czynienia tylko ze spójnikami $\neg$, $\wedge$, $\vee$, $\rightarrow$ oraz $\leftrightarrow$, stosowaliśmy tabele wartości logicznych, aby podzielić nasze postępowanie na mniejsze kroki. Aby móc zrobić podobny podział w rachunku predykatów, musimy najpierw dokładnie określić, jakimi typami wyrażeń musimy się posługiwać, a następnie poznać, jakie są związki między kwantyfikatorami $\forall$ i $\exists$ oraz między nimi a spójnikami logicznymi $\neg$, $\wedge$, $\vee$, $\rightarrow$ i $\leftrightarrow$.

Pojęcia „dowód" i „reguła wnioskowania", które omówiliśmy w § 2.4 dla rachunku zdań, mogą również być rozszerzone na rachunek predykatów. Nie powinno nas dziwić, że więcej możliwych wyrażeń powoduje więcej komplikacji. Nawet niezbyt dokładne omówienie tego tematu mogłoby stanowić istotną część nowej książki. W tym paragrafie omówimy po prostu kilka najbardziej podstawowych i użytecznych zależności między kwantyfikatorami i spójnikami logicznymi.

PRZYKŁAD 1 Kiedy słyszymy w reklamie telewizyjnej, że „samochody nie są takie same", wiemy, że naprawdę oznacza to, iż „nie wszystkie samochody są takie same". Różnica byłaby nawet bardziej widoczna, gdybyśmy porównali zdania „wszystkie samochody nie są żółte" i „nie wszystkie samochody są żółte". Pierwsze z tych stwierdzeń jest fałszywe — są jakieś żółte samochody — a drugie jest prawdziwe. Zdania te mają różne wartości logiczne. Dla danej dziedziny złożonej z samochodów, jeśli $y(c)$ jest predykatem „c jest żółty", to pierwsze zdanie ma postać $\forall c(\neg y(c))$, podczas gdy drugie $\neg(\forall c y(c))$.

To są proste przykłady predykatów, w tym przypadku zdań, zbudowanych przy użyciu kwantyfikatora $\forall$ i spójnika logicznego $\neg$. Bardziej skomplikowane predykaty pojawiają się często przy weryfikacji programów komputerowych zawierających instrukcje wyboru, aby można było zobaczyć, jak będą zachowywały się te programy dla różnych danych wejściowych.

Nasze proste predykaty nie mają takiej samej wartości logicznej dla danego zbioru samochodów i dla tej szczególnej interpretacji $y(c)$ — „c jest żółty". Czy mogłyby być one logicznie równoważne, gdybyśmy mieli inną dziedzinę lub inne znaczenie $y(c)$?

Jest to jedno z ogólnych pytań, na które planujemy odpowiedzieć. Zob. przykład 4(b) poniżej, gdzie kontynuujemy tę dyskusję. ■

Na początku musimy opisać wyrażenia logiczne, z którymi będziemy mieli do czynienia. W rozdziale 2 stosowaliśmy termin „zdanie złożone" w sposób nieformalny, aby opisać zdania zbudowane z prostszych zdań. W ćwiczeniu 15 w § 7.2 podaliśmy rekurencyjną definicję wyrażeń poprawnie zbudowanych w rachunku zdań; są to zdania złożone z rozdziału 2. W ten sam sposób użyjemy definicji rekurencyjnej, aby dokładnie zdefiniować „zdania złożone" w rachunku predykatów. Przyjrzymy się też bliżej temu, co zrobiliśmy w rachunku zdań.

W rozdziale 2 używaliśmy spójników logicznych do budowania zdań z symboli p, q, r itd., które uważaliśmy za nazwy innych zdań. Te symbole można traktować jako zmienne, a zdanie złożone takie jak $p \wedge (\neg q)$ można traktować jako funkcję dwóch zmiennych p i q; jego wartość logiczna to prawda lub fałsz, w zależności od wartości logicznych p i q oraz w zależności od tego, jaką postać ma to zdanie złożone. Tablicą wartości logicznych dla zdania złożonego jest po prostu lista wartości funkcji, gdy zmienne zmieniają się niezależnie w zbiorze wartości {prawda, fałsz}. W rachunku zdań potrzebowaliśmy tylko takich zmiennych, ale w rachunku predykatów musimy też rozważać zmienne przebiegające dziedziny nieskończone, takie jak $\mathbb{R}$, $\mathbb{Z}$ i $\mathfrak{M}_{m,n}$.

Przypuśćmy, że dany jest zbiór niepustych dziedzin, które przebiegają zmienne wolne wszystkich rozważanych predykatów. Definiujemy **zbiór predykatów** złożonych w następujący sposób:

(P_1) Zmienne logiczne są predykatami złożonymi.

(P_2) Predykaty n-argumentowe są predykatami złożonymi dla $n \geq 1$.

(R_1) Jeśli $p(x)$ i $q(x)$ są predykatami złożonymi ze zmienną wolną x, to również

$$\neg p(x),\ (p(x) \vee q(x)),\ (p(x) \wedge q(x)),$$
$$(p(x) \to q(x)) \text{ oraz } (p(x) \leftrightarrow q(x))$$

są predykatami złożonymi ze zmienną wolną x.

(R_2) Jeśli $p(x)$ jest predykatem złożonym ze zmienną wolną x, to

$$(\forall x\, p(x)) \text{ oraz } (\exists x\, p(x))$$

są predykatami złożonymi, w których zmienna x nie jest zmienną wolną.

Kiedy piszemy tutaj „$p(x)$", chcemy zaznaczyć, że w danym predykacie złożonym x jest jedną z jego zmiennych wolnych. Dopuszczamy możliwość, że wartość logiczna $p(x)$ jest w istocie niezależna od wyboru x. W szczególności możemy uważać, że zdanie to ma w trywialny sposób zmienną wolną x, tak więc $p(x)$ i $q(x)$ w (R_1) mogą być po prostu zdaniami p i q. Jeśli usuniemy (P_2) i (R_2) oraz odwołania do zmiennych wolnych w (R_1), otrzymamy rekurencyjną definicję wyrażeń poprawnie zbudowanych w rachunku zdań, znajdującą się w ćwiczeniu 15 w § 7.2. Przy naszych obecnych ustaleniach istnieją predykaty złożone, poza predykatami utworzonymi w (P_1) i (R_1), które są zdaniami. Jeśli wszystkie zmienne w predykacie złożonym są związane, to predykat jest zdaniem. Rozszerzymy naszą definicję i nazwiemy predykat złożony bez zmiennych wolnych **zdaniem złożonym**. Na przykład

$$((\exists x(\exists z\, p(x,z))) \rightarrow (\forall y(\neg r(y))))$$

jest zdaniem złożonym bez zmiennych wolnych. Z drugiej strony,

$$(p(x) \vee (\neg \forall y\, q(x,y)))$$

oraz

$$((\exists z\, p(x,z)) \rightarrow (\forall y(\neg r(y))))$$

są predykatami złożonymi ze zmienną wolną x.

Liczba i miejsce nawiasów w predykacie złożonym są dokładnie określone przez naszą definicję rekurencyjną. W praktyce dla przejrzystości możemy dodać lub usunąć niektóre nawiasy. Na przykład, możemy zapisać wyrażenie $((\forall x\, p(x)) \rightarrow (\exists x\, p(x)))$ jako $\forall x\, p(x) \rightarrow \exists x\, p(x)$ oraz $(\exists x \neg p(x))$ jako $\exists x(\neg p(x))$. Czasami używamy też nawiasów kwadratowych lub klamrowych zamiast zwykłych nawiasów.

Wartość logiczna zdania złożonego zazwyczaj zależy od wyboru dziedzin, do których należą zmienne związane, ale istnieją ważne przykłady, w których wartość logiczna nie tylko nie zależy od wyboru dziedziny, ale tak naprawdę jest też niezależna od wartości zmiennych logicznych. Zdanie złożone, które jest prawdziwe dla wszystkich dziedzin i wszystkich wartości swoich zmiennych logicznych nazywamy **tautologią**. W tej definicji rozszerzamy pojęcie tautologii z rozdziału 2, gdzie nie było dziedziny, o którą trzeba byłoby się troszczyć.

PRZYKŁAD 2 (a) Ważną klasę tautologii stanowią uogólnione prawa De Morgana; por. je z prawami 8a-8d w tablicy 2.1 w § 2.2. Oto one:

$$(1) \qquad \neg \forall x\, p(x) \leftrightarrow \exists x[\neg p(x)],$$

$$(2) \qquad \neg \exists x\, p(x) \leftrightarrow \forall x[\neg p(x)],$$

$$\forall x\, p(x) \leftrightarrow \neg\exists x[\neg p(x)], \tag{3}$$

$$\exists x\, p(x) \leftrightarrow \neg\forall x[\neg p(x)]. \tag{4}$$

Aby stwierdzić, że prawo (1) jest tautologią, zauważmy, że zdanie $\neg\forall x\, p(x)$ ma wartość logiczną prawdy dokładnie wtedy, gdy zdanie $\forall x\, p(x)$ ma wartość logiczną fałszu, a to zachodzi zawsze wtedy, gdy istnieje x w danej dziedzinie taki, że $p(x)$ jest zdaniem fałszywym, tzn. taki, że $\neg p(x)$ jest zdaniem prawdziwym. Zatem zdanie $\neg\forall x\, p(x)$ jest prawdziwe dokładnie wtedy, gdy zdanie $\exists x(\neg p(x))$ jest prawdziwe. Rozumowanie to nie zależy od wyboru dziedziny, a więc prawo (1) jest tautologią.

Prawo De Morgana (2) można przeanalizować w podobny sposób. Inaczej, możemy wyprowadzić prawo (2) z prawa (1), podstawiając predykat jednoargumentowy $\neg p(x)$ zamiast $p(x)$, aby otrzymać

$$\neg\forall x[\neg p(x)] \leftrightarrow \exists x[\neg\neg p(x)].$$

Reguły podstawiania podane w § 2.4 nadal obowiązują, a więc możemy podstawić $p(x)$ zamiast $\neg\neg p(x)$, aby otrzymać równoważne wyrażenie

$$\neg\forall x[\neg p(x)] \leftrightarrow \exists x[p(x)].$$

Jest to prawo De Morgana (4) i jeśli weźmiemy zaprzeczenie obu stron, otrzymamy prawo (2). Stosując prawo (2) do $\neg p(x)$ otrzymujemy prawo (3).

(b) Weźmy znowu predykat $y(c) =$„c jest żółty", gdzie c przebiega zbiór samochodów. Prawo De Morgana (1) mówi nam, że

$$\neg(\forall c\, y(c)) \leftrightarrow \exists c(\neg y(c))$$

jest tautologią. Wnioskujemy stąd, że $\neg(\forall c\, y(c))$ i $\exists c(\neg y(c))$ muszą mieć tę samą wartość logiczną z czysto logicznych powodów; nie musimy odwoływać się do kontekstu samochodów. W przykładzie 1 stwierdziliśmy, że $\neg(\forall c\, y(c))$, tzn. „nie wszystkie samochody są żółte" jest zdaniem prawdziwym, a więc $\exists c(\neg y(c))$ też musi być prawdziwe. Oczywiście jest prawdą, że „istnieje samochód, który nie jest żółty". ■

PRZYKŁAD 3 (a) Następujący predykat złożony jest prawdziwy dla każdego dwuargumentowego predykatu $p(x, y)$:

$$\exists x\, \forall y\, p(x, y) \rightarrow \forall y\, \exists x\, p(x, y). \tag{$*$}$$

Innymi słowy, $(*)$ jest tautologią. Aby to stwierdzić, przypuśćmy, że lewa strona implikacji $(*)$ $\exists x\, \forall y\, p(x, y)$ jest prawdziwa. Wtedy

istnieje x_0 w danej dziedzinie takie, że zdanie $\forall y\, p(x_0, y)$ jest prawdziwe, a więc zdanie $p(x_0, y)$ jest prawdziwe dla wszystkich y. Zatem dla każdego y zdanie $\exists x\, p(x, y)$ jest prawdziwe; tak naprawdę to samo x_0 jest dobre dla wszystkich y. Ponieważ zdanie $\exists x\, p(x, y)$ jest prawdziwe dla wszystkich y, prawa strona implikacji $(*)$ jest prawdziwa. Ponieważ prawa strona $(*)$ jest prawdziwa zawsze wtedy, gdy prawdziwa jest lewa strona, to $(*)$ jest tautologią.

(b) Implikacja odwrotna do $(*)$, mianowicie

$$\forall y\, \exists x\, p(x, y) \to \exists x\, \forall y\, p(x, y),$$

nie jest tautologią, jak zauważyliśmy to w przykładzie 3 w § 13.1. Oto inny bardzo prosty przykład. Niech $p(x, y)$ będzie dwuargumentowym predykatem „$x = y$" w dwuelementowej dziedzinie $U = \{a, b\}$. Zauważmy, że zdanie $\exists x\, p(x, a)$ jest prawdziwe, ponieważ $p(x, a)$ jest prawdziwe dla $x = a$. Podobnie, zdanie $\exists x\, p(x, b)$ jest prawdziwe, więc zdanie $\forall y\, \exists x\, p(x, y)$ jest prawdziwe.

Z drugiej strony, jak zauważyliśmy w dowodzie $(*)$, zdanie $\exists x\, \forall y\, p(x, y$ jest prawdziwe tylko wtedy, gdy $\forall y\, p(x_0, y)$ jest prawdziwe dla pewnego x_0. Ponieważ x_0 musi być równe a lub b, to albo $\forall y\, p(a, y)$, albo $\forall y\, p(b, y)$ byłoby zdaniem prawdziwym. Ale zdanie $\forall y\, p(a, y)$ jest fałszywe, ponieważ $p(a, y)$ jest fałszywe dla $y = b$ i podobnie $\forall y\, p(b, y)$ jest fałszywe. Zatem w tej sytuacji zdanie $\forall y\, \exists x\, p(x, y) \to \exists x\, \forall y\, p(x, y)$ jest fałszywe. A więc to zdanie złożone nie jest tautologią. ∎

Tak jak w rachunku zdań, mówimy, że dwa zdania złożone P i Q są **logicznie równoważne**, co zapisujemy $P \Leftrightarrow Q$, jeśli zdanie $P \leftrightarrow Q$ jest tautologią. Również P **logicznie implikuje** Q, jeśli zdanie $P \to Q$ jest tautologią, co zapisujemy $P \Rightarrow Q$. W tablicy 13.1 pokazane są niektóre przydatne równoważności i implikacje logiczne. Zaczynamy numerowanie reguł od 35, ponieważ rozdział 2 zawiera prawa od 1 do 34.

W przykładach 2 i 3 omawialiśmy tautologie odpowiadające prawom 37 i regule 36. Pozostałe reguły są łatwe do sprawdzenia.

PRZYKŁAD 4 (a) Aby dowieść prawdziwości reguły 35b, to znaczy sprawdzić, że

$$\exists x\, \exists y\, p(x, y) \leftrightarrow \exists y\, \exists x\, p(x, y)$$

jest tautologią, musimy sprawdzić, czy to zdanie ma wartość prawdy we wszystkich możliwych dziedzinach. Na podstawie definicji $\leftrightarrow$ musimy tylko sprawdzić, że zdanie $\exists x\, \exists y\, p(x, y)$ ma wartość prawdy w danej dziedzinie wtedy i tylko wtedy, gdy zdanie

Tablica 13.1. Zależności logiczne w rachunku predykatów

35a.	$\forall x\,\forall y\,p(x,y) \Leftrightarrow \forall y\,\forall x\,p(x,y)$	
b.	$\exists x\,\exists y\,p(x,y) \Leftrightarrow \exists y\,\exists x\,p(x,y)$	
36.	$\exists x\,\forall y\,p(x,y) \Leftrightarrow \forall y\,\exists x\,p(x,y)$	
37a.	$\neg\forall x\,p(x) \Leftrightarrow \exists x[\neg p(x)]$	prawa De Morgana
b.	$\neg\exists x\,p(x) \Leftrightarrow \forall x[\neg p(x)]$	
c.	$\forall x\,p(x) \Leftrightarrow \neg\exists x[\neg p(x)]$	
d.	$\exists x\,p(x) \Leftrightarrow \neg\forall x[\neg p(x)]$	
38.	$\forall x\,p(x) \Rightarrow \exists x\,p(x)$	

$\exists y\,\exists x\,p(x,y)$ ma wartość prawdy w tej dziedzinie. Przypuśćmy, że zdanie $\exists x\,\exists y\,p(x,y)$ jest prawdziwe. Wtedy zdanie $\exists y\,p(x_0,y)$ jest prawdziwe dla pewnego x_0 w tej dziedzinie, więc $p(x_0,y_0)$ jest prawdziwe dla pewnego y_0 z tej dziedziny. Zatem zdanie $\exists x\,p(x,y_0)$ jest prawdziwe, a więc zdanie $\exists y\,\exists x\,p(x,y)$ jest prawdziwe. Implikacji w drugą stronę dowodzi się podobnie. Ponadto, zdania $\exists x\,\exists y\,p(x,y)$ i $\exists y\,\exists x\,p(x,y)$ są logicznie równoważne zdaniu $\exists(x,y)p(x,y)$, gdzie (x,y) przebiega zbiór $U_1 \times U_2$, a U_1 i U_2 są dziedzinami zmiennych x i y.

(b) Reguła 38, zastosowana do predykatu $\neg p(x)$ zamiast do $p(x)$, daje

$$\forall x\neg p(x) \Rightarrow \exists x\neg p(x).$$

Wtedy stosując prawo De Morgana 37a do zdania $\exists x\neg p(x)$, otrzymujemy

$$\forall x\neg p(x) \Rightarrow \neg\forall x\,p(x).$$

Implikacja odwrotna jest oczywiście fałszywa, jak widzieliśmy w przykładzie 1. ∎

Prawa De Morgana od 37a do 37d można stosować wielokrotnie, aby zaprzeczać dowolne zdania z kwantyfikatorami. Na przykład

$$\neg\exists w\,\forall x\,\exists y\exists z\,p(w,x,y,z)$$

jest kolejno równoważne logicznie zdaniom

$$\forall w[\neg\forall x\,\exists y\,\exists z\,p(w,x,y,z)],$$
$$\forall w\,\exists x[\neg\exists y\,\exists z\,p(w,x,y,z)],$$
$$\forall w\,\exists x\,\forall y[\neg\exists z\,p(w,x,y,z)],$$
$$\forall w\,\exists x\,\forall y\,\forall z[\neg p(w,x,y,z)].$$

Przykład ten ilustruje ogólną zasadę: zaprzeczenie predykatu poprzedzonego kwantyfikatorami jest logicznie równoważne zdaniu otrzymanemu przez zastąpienie każdego kwantyfikatora $\forall$ kwantyfikatorem $\exists$ i każdego kwantyfikatora $\exists$ kwantyfikatorem $\forall$ oraz zastąpienie samego predykatu jego zaprzeczeniem.

PRZYKŁAD 5 (a) Zdanie

$$\forall x\, \forall y\, \exists z[x < z < y] \tag{1}$$

mówi, że dla każdych liczb x i y istnieje liczba z zawarta między nimi. Jego wartość logiczna, która tak naprawdę nas w tej chwili nie obchodzi, zależy od wyboru dziedziny dla x, y i z (zob. ćwiczenie 14). Zaprzeczeniem zdania (1) jest zdanie

$$\exists x\, \exists y\, \forall z\{\neg[x < z < y]\}.$$

Ponieważ „$x < z < y$" oznacza „$(x < z) \wedge (z < y)$", możemy zastosować podstawowe prawo De Morgana z rozdziału 2, aby otrzymać

$$\neg[x < z < y] \Leftrightarrow \neg(x < z) \vee \neg(z < y) \Leftrightarrow (x \geq z) \vee (z \geq y).$$

Zatem zaprzeczenie zdania (1) jest logicznie równoważne zdaniu

$$\exists x\, \exists y\, \forall z[(z \leq x) \vee (z \geq y)].$$

Zdanie to mówi, że dla pewnych x_0 i y_0 każda liczba z jest albo mniejsza lub równa x_0 albo większa lub równa y_0.

(b) Weźmy dziedziny U_1, U_2 i U_3 złożone odpowiednio z fabryk, materiałów do produkcji i urządzeń. Niech $p(x, y)$ będzie predykatem „x produkuje y" i niech $q(y, z)$ będzie „y jest częścią składową z". Predykat

$$p(x, y) \wedge q(y, z)$$

oznacza „x produkuje y, który jest częścią składową z". Zdanie

$$\forall x\, \forall z\, \exists y[p(x, y) \wedge q(y, z)]$$

oznacza, że każda fabryka produkuje pewną część składową każdego urządzenia. Jego zaprzeczeniem jest zdanie

$$\neg\forall x\, \forall z\, \exists y[p(x, y) \wedge q(y, z)],$$

które jest logicznie równoważne na podstawie prawa De Morgana zdaniu

$$\exists x \exists z \forall y[\neg p(x, y) \vee \neg q(y, z)].$$

Zaprzeczenie to interpretujemy tak, że istnieje fabryka x_0 i urządzenie z_0 takie, że dla każdej części składowej albo x_0 nie produkuje jej, albo nie jest ona częścią składową z_0. Równoważną postacią tej negacji jest zdanie

$$\exists x\, \exists z \neg \exists y [p(x,y) \wedge q(y,z)],$$

które interpretujemy tak, że istnieje fabryka x_0 i urządzenie z_0 takie, że żadna część składowa nie jest produkowana przez x_0 i nie jest częścią składową z_0.

Porównaj ten przykład z przykładem (a). W przypadku (a) zaprzeczenie również było równoważne zdaniu

$$\exists x\, \exists y \neg \exists z [x < z < y],$$

tzn. istnieją x_0 i y_0 i nie istnieje z leżący między nimi.

(c) Zaprzeczeniem zdania

$$\forall x\, \forall y [x < y \to x^2 < y^2] \tag{2}$$

jest zdanie

$$\exists x\, \exists y \{\neg [x < y \to x^2 < y^2]\}.$$

Na podstawie reguły 10a z tablicy 2.1 z § 2.2 i prawa De Morgana otrzymujemy $\neg(p \to q) \Leftrightarrow \neg(\neg p \vee q) \Leftrightarrow p \wedge \neg q$. Tak więc $\neg[x < y \to x^2 < y^2] \Leftrightarrow (x < y) \wedge (x^2 \geq y^2)$. Zatem zaprzeczenie zdania (2) jest logicznie równoważne zdaniu

$$\exists x\, \exists y [(x < y) \wedge (x^2 \geq y^2)].$$

Zdanie to ma postać prawdy, jeśli dziedziną x i y jest zbiór $\mathbb{Z}$ i postać fałszu, jeśli dziedziną jest zbiór $\mathbb{N}$, oczywiście odwrotnie niż w zdaniu (2). ∎

PRZYKŁAD 6 Niech dziedzina U składa się z dwóch elementów a i b. Prawo De Morgana 37a ma wtedy postać

$$\neg[p(a) \wedge p(b)] \Leftrightarrow [\neg p(a)] \vee [\neg p(b)].$$

Jest to prawo De Morgana 8b z tabl. 2.1 w § 2.2, z tym tylko, że zamiast p i q mamy $p(a)$ i $p(b)$. ∎

Zdanie ogólne często ma postać $\forall x\, p(x)$, gdzie x przebiega pewną dziedzinę. Zdanie to jest fałszywe wtedy i tylko wtedy, gdy zdanie $\exists x[\neg p(x)]$ jest prawdziwe, na podstawie prawa De Morgana 37a. Zatem zdanie $\forall x\, p(x)$ jest fałszywe, jeśli można wskazać jakiś x_0, dla którego $p(x_0)$ jest fałszywe. Jak zauważyliśmy to w § 2.1 (po przykładzie 11), taki x_0 nazywamy **kontrprzykładem** dla zdania $\forall x\, p(x)$. W przykładzie 12 w tamtym paragrafie podanych jest kilka przykładów. Tutaj podajemy jeszcze dwa.

PRZYKŁAD 7 (a) Macierze

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{oraz} \quad \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$$

są kontrprzykładem na (fałszywe) stwierdzenie „jeśli macierze $\mathbf{A}$ i $\mathbf{B}$ o wymiarach 2×2 spełniają zależność $\mathbf{AB} = \mathbf{0}$, to $\mathbf{A} = \mathbf{0}$ lub $\mathbf{B} = \mathbf{0}$". To ogólne stwierdzenie można zapisać w postaci

$$\forall \mathbf{A} \forall \mathbf{B}[\mathbf{AB} = \mathbf{0} \rightarrow (\mathbf{A} = \mathbf{0} \vee \mathbf{B} = \mathbf{0})].$$

(b) Zdanie „każdy graf spójny ma cykl Eulera" jest zdaniem fałszywym. Na podstawie twierdzenia Eulera z § 6.2 każdy graf spójny mający wierzchołek stopnia nieparzystego może służyć jako kontrprzykład na to stwierdzenie. Najprostszym kontrprzykładem jest graf o dwóch wierzchołkach i z jedną krawędzią łączącą je. ∎

Warto podkreślić, że implikacja 38 w tabl. 13.1 i jej równoważne wersje z przykładu 4(b) są prawdziwe, ponieważ ograniczyliśmy naszą uwagę do dziedzin niepustych. Gdybyśmy dopuścili dziedzinę pustą, to implikacja 38 byłaby fałszywa. W tym przypadku zdanie $\forall x\, p(x)$ byłoby prawdziwe „w próżni", podczas gdy zdanie $\exists x\, p(x)$ byłoby fałszywe. Jest prawdą, że każdy człowiek o trzech głowach jest bogaty. (Nie zgadzasz się? Podaj kontrprzykład.) Ale nie jest prawdą, że istnieje bogaty człowiek, który ma trzy głowy. Dziedzina tutaj składa się ze wszystkich ludzi o trzech głowach, a $p(x)$ oznacza „x jest bogaty".

Druga implikacja w przykładzie 4(b) również nie jest spełniona w dziedzinie pustej, ale jest nieco trudniej ją przeanalizować. Jest prawdą, że każdy człowiek o trzech głowach nie jest bogaty, ale jest fałszem, że nie każdy człowiek o trzech głowach jest bogaty. Ludzie o trzech głowach mają zadziwiające cechy.

ĆWICZENIA DO § 13.2

1. Weźmy dziedzinę U_1 składającą się z członków klubu i dziedzinę U_2 składającą się z linii lotniczych. Niech $p(x, y)$ będzie predykatem „x był pasażerem linii y" lub równoważnie „y miał x jako pasażera". Napisz, co oznaczają następujące zdania
   (a) reguła 35a, (b) reguła 35b, (c) reguła 36.
2. Weźmy zbiór U wszystkich profesorów uniwersytetu. Niech $p(x)$ będzie predykatem „x lubi muzykę trash metal".
   (a) Zapisz zdanie „nie wszyscy profesorowie uniwersytetu lubią muzykę trash metal" za pomocą symboli rachunku predykatów.

(b) Zrób to samo dla zdania „żaden profesor uniwersytetu nie lubi muzyki trash metal".
(c) Czy którekolwiek ze zdań z ćwiczeń (a) i (b) implikuje drugie z nich? Wyjaśnij to.
(d) Napisz, co oznacza prawo 37b dla tej dziedziny U i tego $p(x)$.
(e) Zrób to samo z prawem 37d.

3. Zinterpretuj prawa De Morgana 37b, 37c i 37d dla predykatu $y(c) =$„c jest żółty", gdzie c przebiega dziedzinę wszystkich samochodów.

4. Niech $p(x, y)$ będzie predykatem „$x \neq y$" w dziedzinie będącej zbiorem $\mathbb{N}$. Podaj wartość logiczną każdego z następujących zdań:
(a) $\forall y\, \exists x\, p(x, y) \to \exists x\, \forall y\, p(x, y)$,
(b) $\neg\forall x\, \exists y\, p(x, y) \to \neg\exists y\, \forall x\, p(x, y)$,
(c) $\forall y\, p(x, y)$,
(d) $\forall x[\neg\forall y\, p(x, y) \vee \exists y\, p(x, y)]$.

5. Pokaż, że następujące prawa z tabl. 13.1 sprowadzają się do praw z tabl. 2.1 z § 2.2, gdy dziedzina U ma dwa elementy, a i b:
(a) prawo 37d, (b) prawo 37b.

6. (a) Pokaż, że implikacja logiczna

$$[\exists x\, p(x)] \wedge [\exists x\, q(x)] \Rightarrow \exists x[p(x) \wedge q(x)]$$

jest fałszywa. Możesz to zrobić, definiując predykaty $p(x)$ i $q(x)$, dla których ta implikacja nie zachodzi.
(b) Zrób to samo dla implikacji logicznej

$$\exists x\, \forall y\, p(x, y) \Rightarrow \forall x\, \exists y\, p(x, y).$$

(Por. to z prawdziwą implikacją z reguły 36).

7. Dla dziedziny $\mathbb{N}$ napisz zaprzeczenie wyrażenia $\forall n[p(n) \to p(n + 1)]$ bez użycia kwantyfikatora $\forall$.

8. Napisz zaprzeczenie wyrażenia $\exists x\, \forall y\, \exists z[z > y \to z < x^2]$ bez użycia spójnika $\neg$.

9. (a) Napisz zaprzeczenie wyrażenia

$$P = \forall x\, \forall y[x < y \to \exists z\{x < z < y\}]$$

bez użycia spójnika $\neg$.
(b) Wyznacz wartość logiczną P, jeśli dziedziną x, y i z jest zbiór $\mathbb{R}$ lub $\mathbb{Q}$.
(c) Wyznacz wartość logiczną P, jeśli dziedziną jest zbiór $\mathbb{N}$ lub $\mathbb{Z}$.

10. Podaj kontrprzykład dla każdego z następujących stwierdzeń.
(a) Każda liczba całkowita parzysta jest iloczynem dwóch liczb całkowitych parzystych.
(b) $|S \cup T| = |S| + |T|$ dla każdych dwóch skończonych zbiorów S i T.
(c) Każda liczba całkowita dodatnia postaci $6k-1$ jest liczbą pierwszą.
(d) Każdy graf ma parzystą liczbę krawędzi.
(e) Wszystkie wykłady z matematyki są zabawne.

11. Nasza definicja predykatu złożonego nie dopuszcza wyrażeń takich jak $\exists x\, p(x,x)$, gdzie $p(x,y)$ jest predykatem dwuargumentowym. Opisz predykat $q(x,y)$ taki, że zdanie

$$\exists x\, \exists y[p(x,y) \wedge q(x,y)]$$

jest prawdziwe wtedy i tylko wtedy, gdy $p(x,x)$ jest prawdziwe dla pewnego x.

12. Gdy dziedzina jest zbiorem pustym, zdanie $\forall x\, p(x)$ jest prawdziwe „w próżni", bez względu na to, czy $p(x)$ oraz $\exists x\, p(x)$ są fałszywe. Opisz tę sytuację dla dziedziny z dokładnie jednym elementem.

13. Stwierdzenie „istnieją dowolnie duże liczby całkowite n takie, że $p(n)$ jest prawdziwe" można przetłumaczyć na zdanie

$$\forall N\, \exists n[(n \geq N) \wedge p(n)]$$

o dziedzinie $\mathbb{P}$. Napisz zaprzeczenie tego zdania używając spójnika $\rightarrow$ i nie używając spójnika $\neg$. Twoja odpowiedź powinna się tłumaczyć na stwierdzenie, z którego wynika, że $p(n)$ jest prawdziwe tylko dla skończonego zbioru n.

14. (a) Wybierz dziedziny dla x, y i z tak, by zdanie (1) w przykładzie 5 było prawdziwe.
(b) Wybierz dziedziny tak, by zdanie (1) w przykładzie 5 było fałszywe.

§ 13.3. Zbiory nieskończone

W rozdziale 5 skoncentrowaliśmy się na zliczaniu zbiorów skończonych. Zliczanie zbiorów nieskończonych to zupełnie inna sprawa. Być może myślisz, że wszystkie one są tej samej wielkości i że może istnieć tylko jedna nieskończoność. Przeczytaj dalej, by zobaczyć, co myślą na ten temat matematycy. Odłóż na bok swoją intuicję; dużo ci ona tutaj nie pomoże. Temat ten jest fascynujący, ale w tym paragrafie prześlizgniemy się tylko po jego powierzchni.

Matematycy znają sposób klasyfikowania zbiorów nieskończonych zgodnie z ich „wielkością". Po pierwsze uogólniają oni pojęcie „dwa zbiory są tej samej wielkości". Kluczem do tego powszechnie przyjętego właściwego podejścia jest następująca elementarna obserwacja: dwa zbiory skończone są tej samej wielkości[1] wtedy i tylko wtedy, gdy istnieje **przekształcenie**

[1] Zbiory tej samej wielkości (ang. of the same size) nazywamy zwykle zbiorami równolicznymi (ang. equipollent, equinumerous, equivalent).

wzajemnie jednoznaczne jednego zbioru na drugi. Idąc tym tropem, określamy, że dwa zbiory S i T, skończone lub nieskończone, są **tej samej wielkości**, jeśli istnieje przekształcenie wzajemnie jednoznaczne jednego zbioru na drugi. W tej książce nie będziemy badać szczegółowo tej metody klasyfikacji zbiorów, ale będziemy rozróżniać między sobą dwa typy zbiorów nieskończonych.

Każdy zbiór, który jest tej samej wielkości co zbiór $\mathbb{P}$ liczb całkowitych dodatnich będziemy nazywać zbiorem **przeliczalnym**. Tak więc zbiór S jest przeliczalny wtedy i tylko wtedy, gdy istnieje **przekształcenie wzajemnie jednoznaczne** zbioru $\mathbb{P}$ na zbiór S. Zbiór jest **co najwyżej przeliczalny**, jeśli jest skończony lub przeliczalny. Można policzyć lub wypisać elementy takiego niepustego zbioru numerując je liczbami ze zbioru $\{1, 2, \ldots, n\}$ dla pewnego $n \in \mathbb{P}$ lub liczbami z całego zbioru $\mathbb{P}$. W przypadku zbioru nieskończonego ta lista nie zakończy się. Jak można się spodziewać, zbiór jest **nieprzeliczalny**, jeśli nie jest co najwyżej przeliczalny.

PRZYKŁAD 1

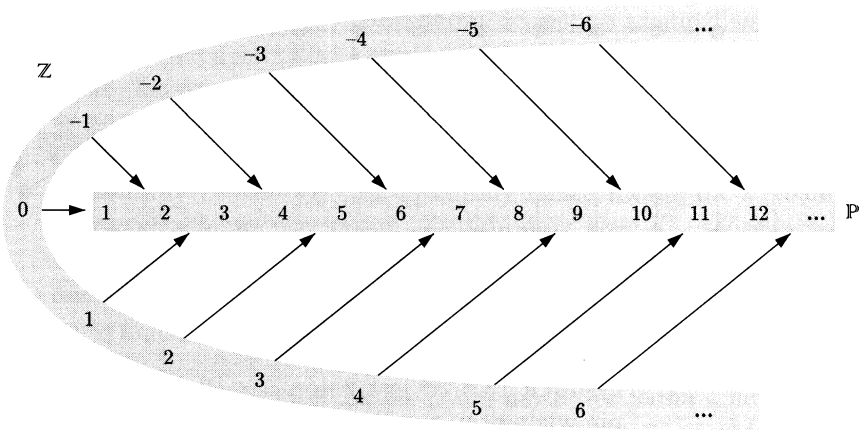
(a) Zbiór $\mathbb{N}$ jest przeliczalny, ponieważ funkcja $f(n) = n - 1$ określa przekształcenie wzajemnie jednoznaczne zbioru $\mathbb{P}$ na zbiór $\mathbb{N}$. Przekształcenie do niego odwrotne f^{-1} jest przekształceniem wzajemnie jednoznacznym zbioru $\mathbb{N}$ na zbiór $\mathbb{P}$; zauważmy, że $f^{-1}(n) = n + 1$ dla $n \in \mathbb{N}$. Nawet pomimo tego, że $\mathbb{P}$ jest podzbiorem właściwym zbioru $\mathbb{N}$, z naszej definicji wynika, że P ma tę samą wielkość co zbiór $\mathbb{N}$. To może być zaskakujące, ponieważ podobna sytuacja nie zdarza się w przypadku zbiorów skończonych. Ale w końcu zbiór $\mathbb{N}$ ma tylko jeden element, który nie należy do $\mathbb{P}$.

(b) Zbiór $\mathbb{Z}$ wszystkich liczb całkowitych jest też przeliczalny. Funkcja wzajemnie jednoznaczna f przekształcająca zbiór $\mathbb{Z}$ na zbiór $\mathbb{P}$ jest pokazana na rys. 13.1, gdzie wygodnie było narysować zbiór $\mathbb{Z}$ w taki wygięty sposób. Funkcja ta jest dana wzorem

$$f(n) = \begin{cases} 2n + 1, & \text{jeśli } n \geq 0, \\ -2n, & \text{jeśli } n < 0. \end{cases}$$

Chociaż $\mathbb{Z}$ wydaje się być dwa razy większy niż $\mathbb{P}$, zbiory te są tej samej wielkości. Bądź ostrożny! Na twojej intuicji dotyczącej zbiorów skończonych nie można polegać lub — patrząc na sytuację bardziej pozytywnie może powinniśmy udoskonalić naszą intuicję, gdy mamy do czynienia ze zbiorami nieskończonymi.

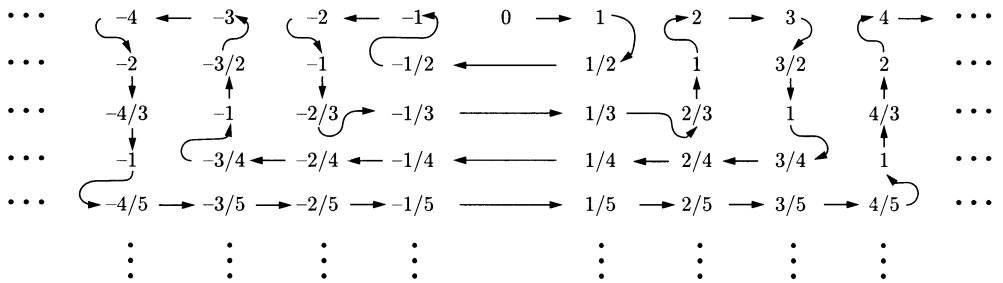
(c) Nawet zbiór $\mathbb{Q}$ wszystkich liczb wymiernych jest przeliczalny. Ten fakt jest zaskakujący, ponieważ zbiór liczb wymier-



przekształcenie wzajemnie jednoznaczne zbioru $\mathbb{Z}$ na zbiór $\mathbb{P}$

Rysunek 13.1

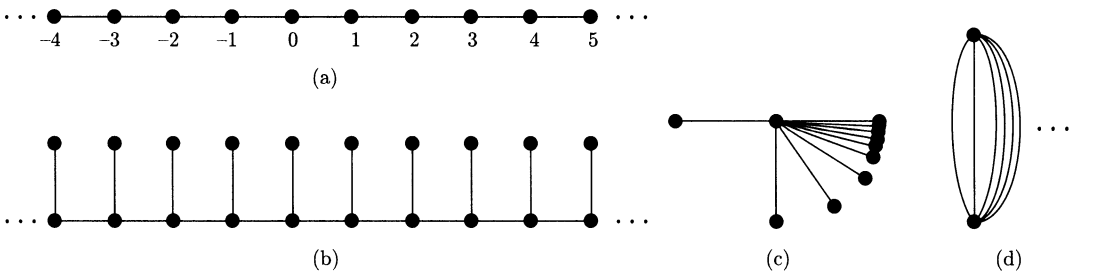
nych rozkłada się równomiernie w całym zbiorze $\mathbb{R}$. Chcąc podać przekształcenie wzajemnie jednoznaczne zbioru $\mathbb{P}$ na zbiór $\mathbb{Q}$, stwierdzamy, że rysunek jest wart tyle co tysiąc wzorów; zob. rys. 13.2. Funkcję f otrzymujemy idąc za strzałkami i omijając to, co się powtarza. Zatem $f(1) = 0$, $f(2) = 1$, $f(3) = \frac{1}{2}$, $f(4) = -\frac{1}{2}$, $f(5) = -1$, $f(6) = -2$, $f(7) = -\frac{2}{3}$ itd. ∎



ustawienie liczb wymiernych w ciąg

Rysunek 13.2

PRZYKŁAD 2 Prawie wszystkie nasze przykłady grafów i drzew miały skończenie wiele wierzchołków i krawędzi. Jednakże nie ma takich ograniczeń w ogólnych definicjach. Na rysunku 13.3 pokazane są przykłady kilku grafów nieskończonych. Zbiorem wierzchołków na rysunku 13.3(a) jest zbiór $\mathbb{Z}$ i tylko sąsiednie liczby całkowite są połączone krawędzią. Zauważmy, że to drzewo nieskończone nie ma liści, podczas gdy każde drzewo skończone, mające więcej niż jeden wierzchołek, ma liście. Zbiorem wierzchołków na



Rysunek 13.3

rysunku 13.3(b) jest zbiór $\mathbb{Z} \times \{0,1\}$; to drzewo ma nieskończenie wiele liści. Środkowy wierzchołek drzewa przedstawionego na rysunku 13.3(c) ma stopień nieskończony; wszystkie inne wierzchołki są liśćmi. Są tylko dwa wierzchołki w grafie przedstawionym na rysunku 13.3(d), ale połączone są one nieskończenie wieloma krawędziami.

We wszystkich tych przykładach zbiory wierzchołków i krawędzi są przeliczalne. Grafy nie muszą być przeliczalne, ale trudno jest narysować lub inaczej przedstawić grafy nieprzeliczalne. ∎

W następnym przykładzie pokażemy metodę, która pochodzi od Georga Cantora, ojca teorii mnogości i która jest nazywana metodą przekątniową Cantora. Wyniki przykładu (b) mogą wydać ci się bardziej interesujące, ale szczegóły przykładu (a) są łatwiejsze do zrozumienia.

PRZYKŁAD 3 (a) Twierdzimy, że zbiór $\text{FUN}(\mathbb{P}, \{0,1\})$ wszystkich funkcji ze zbioru $\mathbb{P}$ w zbiór $\{0,1\}$ jest nieprzeliczalny. Równoważnie, zbiór wszystkich nieskończonych ciągów zer i jedynek jest nieprzeliczalny. Oczywiście zbiór $\text{FUN}(\mathbb{P}, \{0,1\})$ jest nieskończony, więc gdyby był on przeliczalny, to istniałby nieskończony ciąg $\{f_1, f_2, \ldots\}$ wszystkich funkcji z tego zbioru. Definiujemy na zbiorze $\mathbb{P}$ funkcję f^* w następujący sposób:

$$f^*(n) = \begin{cases} 0, & \text{jeśli } f_n(n) = 1, \\ 1, & \text{jeśli } f_n(n) = 0. \end{cases}$$

Dla każdego $n \in \mathbb{P}$ z konstrukcji wynika, że $f^*(n) \neq f_n(n)$, a więc funkcja f^* musi być różna od każdej funkcji f_n. Zatem $\{f_1, f_2, \ldots\}$ nie jest ciągiem wszystkich funkcji ze zbioru $\text{FUN}(\mathbb{P}, \{0,1\})$. Sprzeczność ta pokazuje, że $\text{FUN}(\mathbb{P}, \{0,1\})$ jest zbiorem nieprzeliczalnym.

(b) Przedział $[0,1)$ jest nieprzeliczalny. Gdyby był on przeliczalny, istniałaby funkcja wzajemnie jednoznaczna f przekształ-

cająca $\mathbb{P}$ na zbiór $[0,1)$. Pokażemy, że jest to niemożliwe. Każda liczba z przedziału $[0,1)$ ma rozwinięcie dziesiętne $0,d_1d_2d_3\ldots$, gdzie każde d_j jest cyfrą ze zbioru $\{0,1,2,3,4,5,6,7,8,9\}$. W szczególności, każda liczba $f(k)$ ma postać $0,d_{1k}d_{2k}d_{3k}\ldots$; symbol d_{nk} oznacza tutaj n-tą cyfrę w rozwinięciu liczby $f(k)$. Popatrzmy na tablicę 13.2 i skupmy naszą uwagę na wyróżnionych cyfrach na przekątnej. Definiujemy ciąg d^*, którego n-ty wyraz d_n^* jest zbudowany w następujący sposób: jeśli $d_{nn}\neq 1$, niech $d_n^*=1$ oraz jeśli $d_{nn}=1$, niech $d_n^*=2$. Cała rzecz polega na tym, że $d_n^*\neq d_{nn}$ dla wszystkich $n\in\mathbb{P}$. Teraz $0,d_1^*d_2^*d_3^*\ldots$ oznacza liczbę $x\in[0,1)$, która dla każdego $n\in\mathbb{P}$ różni się od liczby $f(n)$ n-tą cyfrą. Zatem x nie może być jedną z liczb $f(n)$; tzn. $x\notin \mathrm{Im}(f)$, więc f nie przekształca zbioru $\mathbb{P}$ na zbiór $[0,1)$. A więc przedział $[0,1)$ jest nieprzeliczalny.

Tablica 13.2

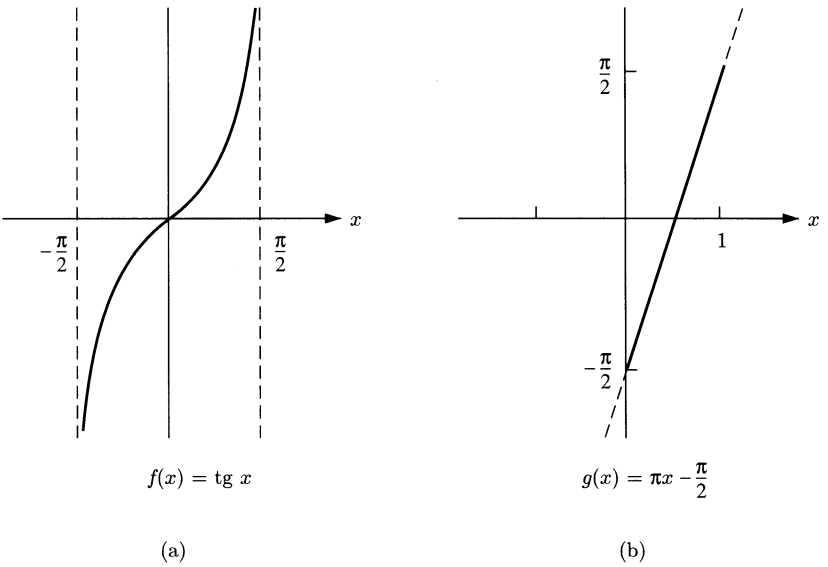
$$
\begin{aligned}
f(1) &= 0,\mathbf{d_{11}}d_{21}d_{31}d_{41}\ldots\\
f(2) &= 0,d_{12}\mathbf{d_{22}}d_{32}d_{42}\ldots\\
f(3) &= 0,d_{13}d_{23}\mathbf{d_{33}}d_{43}\ldots\\
f(4) &= 0,d_{14}d_{24}d_{34}\mathbf{d_{44}}\ldots\\
&\cdot\\
&\cdot\\
&\cdot
\end{aligned}
$$

Zauważmy, że zrobiliśmy to tak, iż wszystkie cyfry liczby x były jedynkami lub dwójkami. Taki wybór był całkiem dowolny, poza tym, że starannie unikaliśmy zer i dziewiątek, ponieważ istnieją pewne liczby, których rozwinięcia zawierają 0 i 9 i liczby te mają dwa rozwinięcia dziesiętne. Na przykład liczby $0,250000\ldots$ oraz $0,249999\ldots$ oznaczają tę samą liczbę z przedziału $[0,1)$. ∎

Dowód przedstawiony w przykładzie 3(b) może być zmodyfikowany tak, by można było dowieść, że zbiory $\mathbb{R}$ i $(0,1)$ są nieprzeliczalne; w rzeczywistości, wszystkie przedziały $[a,b]$, $[a,b)$, $(a,b]$ i (a,b) są nieprzeliczalne, jeśli $a<b$. Na podstawie ćwiczenia 9 można inaczej pokazać, że te zbiory są nieprzeliczalne. Polega to na udowodnieniu, że istnieją odwzorowania wzajemnie jednoznaczne między nimi. Tak naprawdę, są też odwzorowania wzajemnie jednoznaczne między nimi i przedziałami nieograniczonymi. Pokazanie, że istnieją takie odwzorowania wzajemnie jednoznaczne, może być wyzwaniem. Podamy kilka trikowych ro-

zumowań w następnym przykładzie i poprosimy o podanie kilku łatwiejszych w ćwiczeniu 3.

PRZYKŁAD 4 (a) Pokażemy, że istnieje odwzorowanie wzajemnie jednoznaczne zbioru $\mathbb{R}$ na zbiór $(0,1)$, a więc zbiory te są tej samej wielkości. Chociaż nie są tu koniecznie potrzebne funkcje trygonometryczne (por. ćwiczenie 5), użyjemy gotowej funkcji trygonometrycznej, mianowicie funkcji tangens; zob. rys. 13.4(a). Funkcja f dana wzorem $f(x) = \operatorname{tg} x$ jest różnowartościowa w przedziale $(-\pi/2, \pi/2)$ i przekształca ten przedział na zbiór $\mathbb{R}$. Łatwo znaleźć funkcję liniową g przekształcającą przedział $(0,1)$ na przedział $(-\pi/2, \pi/2)$, jest nią funkcja $g(x) = \pi x - \pi/2$; zob. rys. 13.4(b). Funkcja złożona $f \circ g$ jest przekształceniem wzajemnie jednoznacznym zbioru $(0,1)$ na zbiór $\mathbb{R}$.



Rysunek 13.4

(b) Pokażemy, że przedziały $[0,1)$ i $(0,1)$ są tej samej wielkości. Nie ma żadnego prostego wzoru dającego przekształcenie różnowartościowe jednego z tych zbiorów na drugi. Pomysł polega na wyizolowaniu pewnego nieskończonego ciągu w przedziale $(0,1)$, powiedzmy ciągu $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$, a następnie przekształceniu go na ciąg $0, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$, pozostawiając resztę nie naruszoną. To znaczy, niech

$$C = (0,1) \setminus \left\{ \frac{1}{n} : \; n = 2, 3, 4, \ldots \right\}$$

i definiujemy funkcję

$$f(x)=\begin{cases} 0, & \text{jeśli } x = 1/2, \\ \frac{1}{n-1}, & \text{jeśli } x = 1/n \text{ dla pewnej liczby całkowitej } n \geq 3, \\ x, & \text{jeśli } x \in C. \end{cases}$$

Zobacz tabl. 13.3. ∎

Tablica 13.3

$$(0,1) = C \cup \left\{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots\right\}$$
$$\downarrow \quad \downarrow \downarrow \downarrow \downarrow$$
$$[0,1) = C \cup \left\{0, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right\}$$
$$f\colon (0,1) \to [0,1)$$

Udowodnimy teraz dwa podstawowe fakty o zbiorach co najwyżej przeliczalnych.

Twierdzenie

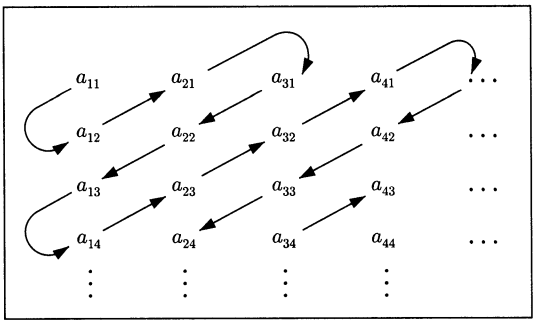
(a) Podzbiory zbiorów co najwyżej przeliczalnych są co najwyżej przeliczalne.
(b) Suma co najwyżej przeliczalnie wielu zbiorów co najwyżej przeliczalnych jest zbiorem co najwyżej przeliczalnym.

Dowód. (a) Wystarczy pokazać, że podzbiory zbioru $\mathbb{P}$ są co najwyżej przeliczalne. Weźmy podzbiór A zbioru $\mathbb{P}$. Oczywiście A jest co najwyżej przeliczalny, jeśli jest skończony. Przypuśćmy, że A jest zbiorem nieskończonym. Zastosujemy zasadę dobrego uporządkowania. Definiujemy $f(1)$ jako najmniejszy element zbioru A. Następnie definiujemy $f(2)$ jako najmniejszy element zbioru $A\backslash\{f(1)\}$, $f(3)$ jako najmniejszy element zbioru $A\backslash\{f(1), f(2)\}$ itd. Postępujemy dalej w ten sposób tak, by $f(n+1)$ był najmniejszym elementem niepustego zbioru $A\backslash\{f(k)\colon\ 1 \leq k \leq n\}$ dla każdego $n \in \mathbb{P}$. Łatwo sprawdzić, że ta definicja rekurencyjna określa funkcję różnowartościową f przekształcającą $\mathbb{P}$ na zbiór A (ćwiczenie 10), a więc A jest zbiorem przeliczalnym.

(b) Stwierdzenie (b) oznacza, że jeśli I jest zbiorem co najwyżej przeliczalnym i jeśli $\{A_i\colon\ i \in I\}$ jest rodziną zbiorów co najwyżej przeliczalnych, to suma $\bigcup_{i\in I} A_i$ jest przeliczalna. Możemy przyjąć, że każdy zbiór A_i jest niepusty i że suma $\bigcup_{i\in I} A_i$

jest nieskończona oraz możemy założyć, że $I = \mathbb{P}$ lub że zbiór I ma postać $\{1, 2, \ldots, n\}$. Jeśli $I = \{1, 2, \ldots, n\}$, możemy zdefiniować $A_i = A_n$ dla $i > n$ i otrzymać rodzinę zbiorów $\{A_i\colon\ i \in \mathbb{P}\}$, mającą tę samą sumę. Zatem możemy założyć, że $I = \mathbb{P}$. Każdy zbiór A_i jest skończony lub przeliczalny. Powtarzając elementy, jeśli zbiór A_i jest skończony, możemy zapisać każdy A_i w następujący sposób:

$$A_i = \{a_{1i}, a_{2i}, a_{3i}, a_{4i}, \ldots\}.$$



Rysunek 13.5

Elementy sumy $\bigcup_{i \in I} A_i$ można wypisać w tablicy, tak jak pokazano na rys. 13.5. Strzałki na tym rysunku pokazują, jak można elementy $\bigcup_{i \in I} A_i$ ustawić w ciąg:

$$(*) \qquad a_{11}, a_{12}, a_{21}, a_{31}, a_{22}, a_{13}, a_{14}, a_{23}, a_{32}, a_{41}, \ldots.$$

Niektóre elementy mogą się powtarzać, ale ciąg ten zawiera nieskończenie wiele różnych elementów, ponieważ zbiór $\bigcup_{i \in I} A_i$ jest nieskończony. Teraz funkcję różnowartościową f przekształcającą zbiór $\mathbb{P}$ na zbiór $\bigcup_{i \in I} A_i$ otrzymujemy w następujący sposób: $f(1) = a_{11}$, $f(2)$ jest następnym elementem wypisanym w ciągu $(*)$ i różnym od $f(1)$, $f(3)$ jest następnym elementem ciągu $(*)$ różnym od $f(1)$ i od $f(2)$ itd. ∎

PRZYKŁAD 5 Rozumowanie z przykładu 1(c) pokazujące, że zbiór $\mathbb{Q}$ jest przeliczalny, jest podobne do dowodu punktu (b) twierdzenia. W rzeczywistości możemy zastosować twierdzenie, aby podać inny dowód tego, że $\mathbb{Q}$ jest zbiorem przeliczalnym. Dla każdej liczby $n \in \mathbb{P}$ niech

$$A_n = \left\{\frac{m}{n}\colon\ m \in \mathbb{Z}\right\}.$$

Zatem A_n składa się z wszystkich całkowitych wielokrotności $1/n$. Oczywiście zbiór $\mathbb{Z}$ można przekształcić w sposób wzajemnie jednoznaczny na każdy zbiór A_n (przekształcając m/n na m), więc każdy zbiór A_n jest przeliczalny. Na podstawie punktu (b) twierdzenia suma

$$\bigcup_{n\in\mathbb{P}} A_n = \mathbb{Q}$$

jest również zbiorem przeliczalnym. ∎

PRZYKŁAD 6 (a) Jeśli Σ jest alfabetem skończonym, to zbiór Σ^* wszystkich słów utworzonych z liter alfabetu Σ jest przeliczalny. Zauważmy, że zbiór Σ jest niepusty z definicji. Wiemy już, że Σ^* jest zbiorem nieskończonym. Przypomnijmy, że

$$\Sigma^* = \bigcup_{k=0}^{\infty} \Sigma^k,$$

gdzie każdy zbiór Σ^k jest skończony. Zatem Σ^* jest przeliczalną sumą zbiorów przeliczalnych, a więc Σ^* sam jest zbiorem przeliczalnym na podstawie punktu (b) twierdzenia.

(b) Wyobraźmy sobie, jeśli się nam uda, przeliczalny alfabet Σ i niech Σ^* składa się z wszystkich słów zbudowanych z liter alfabetu Σ, tzn. wszystkich skończonych ciągów liter ze zbioru Σ. Dla każdej liczby $k \in \mathbb{P}$ zbiór Σ^k wszystkich słów długości k można przekształcić w sposób wzajemnie jednoznaczny na zbiór $\Sigma^k = \Sigma \times \Sigma \times \ldots \times \Sigma$ (k razy). Tak naprawdę, w tym przekształceniu każde słowo $a_1 a_2 \ldots a_k$ przechodzi na ciąg k-elementowy $(a_1, a_2, \ldots, a_k)$. Tak więc każdy zbiór Σ^k jest przeliczalny na podstawie ćwiczenia 15. Jednoelementowy zbiór $\Sigma^0 = \{\lambda\}$ jest również co najwyżej przeliczalny. Zatem $\Sigma^* = \bigcup_{k=0}^{\infty} \Sigma^k$ jest zbiorem przeliczalnym na podstawie punktu (b) twierdzenia. ∎

PRZYKŁAD 7 Weźmy graf o zbiorze wierzchołków V i zbiorze krawędzi E. Nawet jeśli V lub E jest zbiorem nieskończonym, to z definicji każda droga ma skończoną długość. Niech $\mathcal{P}$ będzie zbiorem wszystkich dróg w tym grafie.

(a) Jeśli E jest zbiorem niepustym, to $\mathcal{P}$ jest zbiorem nieskończonym. Ponieważ, jeśli e jest jakąś krawędzią, to e, ee, eee itd. są drogami w tym grafie.

(b) Jeśli E jest zbiorem skończonym, to $\mathcal{P}$ jest zbiorem co najwyżej przeliczalnym. Na potrzeby obliczeń potraktujmy zbiór E jako alfabet. Ponieważ każda droga jest ciągiem krawędzi, odpowiada jej słowo w zbiorze E^*. Oczywiście nie wszystkie słowa

w E^* odpowiadają drogom, ponieważ końce sąsiednich krawędzi muszą się łączyć. Ale istnieje przekształcenie wzajemnie jednoznaczne zbioru $\mathcal{P}$ na pewien podzbiór zbioru E^*. Zbiór E^* jest przeliczalny na podstawie przykładu 6(a), a więc $\mathcal{P}$ jest zbiorem co najwyżej przeliczalnym na podstawie punktu (a) twierdzenia.

(c) Jeśli E jest zbiorem przeliczalnym, to $\mathcal{P}$ nadal jest zbiorem przeliczalnym. Po prostu wykorzystaj przykład 6(b) zamiast przykładu 6(a) w dyskusji przeprowadzonej w punkcie (b). ∎

ĆWICZENIA DO § 13.3

1. Niech A i B będą zbiorami skończonymi takimi, że $|A| < |B|$. Czy następujące zdania są prawdziwe czy fałszywe?
(a) Istnieje przekształcenie różnowartościowe zbioru A w zbiór B.
(b) Istnieje przekształcenie różnowartościowe zbioru A na zbiór B.
(c) Istnieje przekształcenie różnowartościowe zbioru B w zbiór A.
(d) Istnieje funkcja przekształcająca zbiór A na zbiór B.
(e) Istnieje funkcja przekształcająca zbiór B na zbiór A.

2. Czy następujące zdania są prawdziwe czy fałszywe?
(a) Zbiór liczb wymiernych dodatnich jest przeliczalny.
(b) Zbiór wszystkich liczb wymiernych jest przeliczalny.
(c) Zbiór liczb rzeczywistych dodatnich jest przeliczalny.
(d) Część wspólna dwóch zbiorów przeliczalnych jest zbiorem przeliczalnym.
(e) Istnieje przekształcenie wzajemnie jednoznaczne zbioru wszystkich liczb całkowitych parzystych na zbiór $\mathbb{N}$ liczb naturalnych.

3. Dla każdej z następujących par zbiorów podaj przekształcenie wzajemnie jednoznaczne jednego zbioru na drugi:
(a) $(0,1)$ i $(-1,1)$, (b) $[0,1)$ i $(0,1]$,
(c) $[0,1]$ i $[-5,8]$, (d) $(0,1)$ i $(1,\infty)$,
(e) $(0,1)$ i $(0,\infty)$, (f) $\mathbb{R}$ i $(0,\infty)$.

4. Niech $E = \{n \in \mathbb{N}: \ n \text{ jest liczbą parzystą}\}$. Pokaż, że zbiory E i $\mathbb{N}\setminus E$ są przeliczalne, podając funkcje wzajemnie jednoznaczne $f\colon \mathbb{P} \to E$ i $g\colon \mathbb{P} \to \mathbb{N}\setminus E$.

5. Oto jeszcze jedna funkcja wzajemnie jednoznaczna przekształcająca przedział $(0,1)$ na zbiór $\mathbb{R}$:

$$f(x) = \frac{2x-1}{x(1-x)}.$$

(a) Naszkicuj wykres funkcji f.
(b) Jeśli znasz trochę rachunek różniczkowy, udowodnij, że funkcja f jest różnowartościowa, pokazując, że jej pochodna jest dodatnia w przedziale $(0,1)$.

6. Które z następujących zbiorów są co najwyżej przeliczalne, a które są przeliczalne?
(a) $\{0, 1, 2, 3, 4\}$, (b) $\{n \in \mathbb{N}\colon\ n \leq 73\}$,
(c) $\{n \in \mathbb{Z}\colon\ n \leq 73\}$, (d) $\{n \in \mathbb{Z}\colon\ |n| \leq 73\}$,
(e) $\{5, 10, 15, 20, 25, \ldots\}$, (f) $\mathbb{N} \times \mathbb{N}$,
(g) $[\frac{1}{4}, \frac{1}{3}]$.

7. Niech Σ będzie alfabetem $\{a, b, c\}$. Które z następujących zbiorów są przeliczalne?
(a) Σ^{73}, (b) Σ^*,
(c) $\bigcup_{k=0}^{\infty} \Sigma^{2k} = \{w \in \Sigma^*\colon$ długość(w) jest liczbą parzystą$\}$,
(d) $\bigcup_{k=0}^{3} \Sigma^k$, (e) $\bigcup_{k=0}^{3} \Sigma^{2k}$.

8. Zbiór A ma m elementów, a zbiór B ma n elementów. Ile jest funkcji różnowartościowych ze zbioru A w zbiór B? *Wskazówka*: rozważ oddzielnie przypadki $m \leq n$ i $m > n$.

9. (a) Pokaż, że jeśli istnieje przekształcenie wzajemnie jednoznaczne zbioru S na pewien zbiór co najwyżej przeliczalny, to sam zbiór S jest co najwyżej przeliczalny.
(b) Pokaż, że jeśli istnieje przekształcenie wzajemnnie jednoznaczne zbioru S na pewien zbiór nieprzeliczalny, to sam zbiór S jest nieprzeliczalny.

10. Uzupełnij dowód punktu (a) twierdzenia, pokazując, że f jest funkcją różnowartościową przekształcającą $\mathbb{P}$ na A.

11. (a) Udowodnij, że jeśli zbiory S i T są co najwyżej przeliczalne, to zbiór $S \times T$ jest co najwyżej przeliczalny.
(b) Udowodnij, że jeśli funkcja f przekształca zbiór S na zbiór T i S jest zbiorem co najwyżej przeliczalnym, to T jest zbiorem co najwyżej przeliczalnym.
(c) Wykorzystaj ćwiczenia (a) i (b) i podaj inny dowód tego, że $\mathbb{Q}$ jest zbiorem co najwyżej przeliczalnym. *Sugestia*: dla $(m, n) \in \mathbb{Z} \times \mathbb{P}$ zdefiniuj $f(m, n) = m/n$.

12. Pokaż, że jeśli zbiory S i T są tej samej wielkości, to również zbiory $\mathcal{P}(S)$ i $\mathcal{P}(T)$ są tej samej wielkości.

13. (a) Pokaż, że zbiór FUN $(\mathbb{P}, \{0, 1\})$ można przekształcić w sposób wzajemnie jednoznaczny na zbiór $\mathcal{P}(\mathbb{P})$ wszystkich podzbiorów zbioru $\mathbb{P}$.
(b) Pokaż, że zbiór $\mathcal{P}(\mathbb{P})$ jest nieprzeliczalny.

14. Pokaż, że każda rozłączna rodzina niepustych podzbiorów zbioru co najwyżej przeliczalnego jest co najwyżej przeliczalna.

15. Pokaż, że jeśli S jest zbiorem co najwyżej przeliczalnym, to $S^n = S \times S \times \ldots \times S$ (n razy) jest zbiorem co najwyżej przeliczalnym dla każdego n. *Wskazówka*: wykorzystaj ćwiczenie 11(a) i indukcję.

16. Oto elegancka przejrzysta funkcja wzajemnie jednoznaczna przekształcająca zbiór $\mathbb{Q}^+$ liczb wymiernych dodatnich na zbiór $\mathbb{P}$. Dla danej liczby $m/n \in \mathbb{Q}^+$, gdzie m i n są liczbami względnie pierwszymi, zapisz $m = p_1^{m_1} \dots p_k^{m_k}$ oraz $n = q_1^{n_1} \dots q_l^{n_l}$ jako iloczyn liczb pierwszych. Definiujemy

$$f\left(\frac{m}{n}\right) = p_1^{2m_1} \dots p_k^{2m_k} \cdot q_1^{2n_1-1} \dots q_l^{2n_l-1}.$$

W szczególności, $f(m) = p_1^{2m_1} \dots p_k^{2m_k} = m^2$ dla liczb całkowitych dodatnich m.

(a) Oblicz $f(\frac{1}{8})$, $f(\frac{1}{9})$, $f(\frac{1}{10})$, $f(\frac{1}{100})$ i $f(\frac{21}{20})$.

(b) Jaki ułamek jest przekształcony na liczbę 23? a na 24? na 25? na 26? na 27? na 28?

(c) Wyjaśnij, dlaczego f jest funkcją różnowartościową oraz dlaczego f przekształca $\mathbb{Q}^+$ na $\mathbb{P}$.

Przykład ten został podany przez Yorama Saghera w listopadowym numerze z 1989 roku miesięcznika *The American Mathematical Monthly* na stronie 823.

To, co jest najważniejsze w tym rozdziale

Jak zwykle: Co to znaczy? Dlaczego znajduje się to tutaj? Jak można tego użyć? Zastanów się nad przykładami.

Pojęcia i oznaczenia

kwantyfikatory, $\forall$, $\exists$
zbiór uniwersalny, dziedzina
predykat = funkcja o wartościach będących zdaniami (funkcja zdaniowa)
predykat n-argumentowy
predykat złożony
    zmienna wolna, zmienna związana
    zdanie złożone
tautologia
logiczna równoważność, implikacja logiczna, $\Leftrightarrow$, $\Rightarrow$
kontrprzykład
zbiory nieskończone
    zbiory tej samej wielkości (równoliczne)
    zbiory co najwyżej przeliczalne
    przeliczalne
    nieprzeliczalne

Fakty

Kwantyfikatory $\forall$ i $\exists$ nie są przemienne:

$\exists x\, \forall y\, p(x,y) \rightarrow \forall y\, \exists x\, p(x,y)$ jest tautologią, a
$\forall x\, \exists y\, p(x,y) \rightarrow \exists y\, \forall x\, p(x,y)$ nie jest.

Zbiory $\mathbb{N}$, $\mathbb{Z}$ i $\mathbb{Q}$ są przeliczalne.
Zbiory $[0,1)$ i $\text{FUN}(\mathbb{P}, \{0,1\})$ są nieprzeliczalne.
Zbiory $\mathbb{R}$, $[0,1)$ i $(0,1)$ są tej samej wielkości.
Podzbiory i co najwyżej przeliczalne sumy zbiorów co najwyżej przeliczalnych są przeliczalne.

Metody

Użycie uogólnionych praw De Morgana do zaprzeczania predykatów z kwantyfikatorami.
Metoda przekątniowa Cantora do dowodzenia, że pewne zbiory są nieprzeliczalne.

ODPOWIEDZI I WSKAZÓWKI

Rozsądni studenci będą zaglądać do tych odpowiedzi i wskazówek po uprzedniej wytężonej pracy nad ćwiczeniami. Często podane są tylko wskazówki; powinniście uzupełnić wszystkie szczegóły rozwiązania po to, by sprawdzić, czy dobrze je rozumiecie, a także po to, by nabrać wprawy w przekazywaniu pomysłów matematycznych.

Paragraf 1.1

1. (a) Na przykład 0, 5, 10, 15, 20.
(c) Na przykład $\emptyset$, $\{1\}$, $\{2,3\}$, $\{3,4\}$, $\{5\}$.
(e) Na przykład 1, 1/2, 1/3, 1/4, 1/73.
(g) Na przykład 1, 2, 4, 16, 18.

3. (a) Na przykład λ, a, ab, cab, ba.
(c) $aaaa$, $aaab$, $aabb$ itd.

5. (a) 0. (c) 138. (e) 73. (g) 0.
(i) ∞. (k) ∞. (m) ∞. (o) ∞.

7. $A \subseteq A$, $B \subseteq B$, zbiór C jest podzbiorem zbioru A i zbiory C, D są podzbiorami A, B i D.

9. (a) Słowo aba należy do wszystkich trzech zbiorów i w każdym ma długość 3.
(c) Słowo cba należy do Σ_1^* i długość$(cba) = 3$.
(e) Słowo $caab$ należy do zbioru Σ_1^* i ma w nim długość 4 oraz należy do Σ_2^* i ma w nim długość 3.

11. (a) Tak.
(c) Usuwaj pierwsze litery ciągu tak długo, jak długo jest to możliwe. Jeśli otrzymasz λ, to wyjściowy ciąg był w Σ^*. W przeciwnym przypadku nie był.

Paragraf 1.2

1. (a) $\{1,2,3,5,7,9,11\}$. (c) $\{1,5,7,9,11\}$.
(e) $\{3,6,12\}$. (g) 16.

3. (a) $[2,3]$. (c) $[0,2)$. (e) $(-\infty,0)\cup(3,\infty)$.

5. (a) $\emptyset$. (c) $\emptyset$. (e) $\{\lambda, ab, ba\}$.
(g) Zbiory $B^c \cap C^c$ i $(B \cup C)^c$ są równe na mocy prawa De Morgana (można to też łatwo sprawdzić bezpośrednio), podobnie jak $(B \cap C)^c$ i $B^c \cup C^c$.

7. $A \oplus A = \emptyset$ i $A \oplus \emptyset = A$.

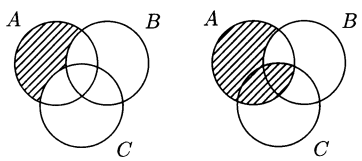
9. $(A \cap B \cap C)^c = (A \cap B)^c \cup C^c$ i $(A \cap B)^c = A^c \cup B^c$ na mocy prawa De Morgana 9b. Teraz wystarczy podstawić.

11. (a) (a,a), (a,b), (a,c), (b,a) itp. Jest dziewięć takich par.
(c) (a,a), (b,b).

13. (a) Na przykład $(0,0)$, $(1,1)$, $(2,2)$, ..., $(6,6)$.
(c) Na przykład $(6,1)$, $(6,2)$, $(6,3)$, ..., $(6,7)$.
(e) Ten zbiór ma 5 elementów. Wypisz je.

15. (a) Fałszywe. Spróbuj $A = \emptyset$.
(c) Prawdziwe. Pokaż, że $x \in B$ implikuje $x \in C$, rozpatrując dwa przypadki: $x \in A$ i $x \notin A$. Podobnie, $x \in C$ implikuje $x \in B$.
(e) Również możesz skorzystać ze wskazówki do ćwiczenia (c).

17. (a) Każdy przykład, w którym $A \neq B$, jest kontrprzykładem.
(c) Odpowiednimi diagramami Venna są:



$(A\backslash B)\backslash C$ $A\backslash(B\backslash C)$

Paragraf 1.3

1. (a) $f(3) = 27, f(1/3) = 1/3, f(-1/3) = 1/27, f(-3) = 27$.
(c) $\mathrm{Im}\,(f) = [0,\infty)$.

3. (a) Nie; zbiór S jest większy niż T.
(c) Tak. Na przykład, niech $f(1) = a$, $f(2) = b$, $f(3) = c$, $f(4) = f(5) = d$.
(e) Nie. Wynika to z ćwiczenia (a) lub z ćwiczenia (d).

5. (a) $f(2,1) = 2^2 3^1 = 12$, $f(1,2) = 2^1 3^2 = 18$ itd.
(c) Na przykład weź 5.

7. (a) Wybierz b_0 ze zbioru B. Dla każdego $a \in A$, $\text{RZUT}(a, b_0) = a$, a więc każdy element a ze zbioru A należy do przeciwdziedziny funkcji RZUT.

9. $\{n \in \mathbb{Z}$: liczba n jest parzysta$\}$.

11. (a) $f \circ g \circ h(x) = (x^8 + 1)^{-3} - 4(x^8 + 1)^{-1}$.
(c) $h \circ g \circ f(x) = ((x^3 - 4x)^2 + 1)^{-4}$.
(e) $g \circ g(x) = (x^2 + 1)^2/(1 + (x^2 + 1)^2)$.
(g) $g \circ h(x) = (x^8 + 1)^{-1}$.

13. Ponieważ $g \circ f\colon S \to U$ i $h\colon U \to V$, więc złożenie funkcji $h \circ (g \circ f)$ jest określone i przekształca zbiór S w zbiór V. Podobna uwaga dotyczy funkcji $(h \circ g) \circ f$. Pokaż, że wartości tych funkcji są takie same dla każdego $x \in S$.

15. (a) $1, 0, -1$ i 0.
(c) $g \circ f$ jest funkcją charakterystyczną zbioru $\mathbb{Z} \setminus E$. $f \circ f(n) = n - 2$ dla wszystkich $n \in \mathbb{Z}$.

Paragraf 1.4

1. (a) $f^{-1}(y) = (y - 3)/2$. (c) $h^{-1}(y) = 2 + \sqrt[3]{y}$.

3. (a) Wszystkie; sprawdź to.
(c) $\text{SUMA}^{\leftarrow}(4)$ ma 5 elementów, $\text{ILOCZYN}^{\leftarrow}(4)$ ma 3 elementy, $\text{MAX}^{\leftarrow}(4)$ ma 9 elementów, $\text{MIN}^{\leftarrow}(4)$ jest zbiorem nieskończonym.

5. (a) $f(0) = 1$, $f(1) = 2$, $f(2) = 3$, $f(3) = 4$, $f(4) = 5$, $f(73) = 74$.
(c) Funkcja f jest różnowartościowa, gdyż jeśli $f(n) = f(n')$, to mamy $n = f(n) - 1 = f(n') - 1 = n'$. Funkcja f nie jest „na", ponieważ $0 \notin \text{Im}(f)$.
(e) $g(f(n)) = \max\{0, (n + 1) - 1\} = n$, ale $f(g(0)) = f(0) = 1$.

7. (a) $(f \circ f)(x) = 1/(1/x) = x$.
(b)-(d) Sprawdź w podobny sposób.

9. Ponieważ funkcje f i g są odwracalne, więc istnieją funkcje odwrotne $f^{-1}\colon T \to S$ i $g^{-1}\colon U \to T$ oraz $f^{-1} \circ g^{-1} : U \to S$. Wystarczy zatem pokazać, że $(g \circ f) \circ (f^{-1} \circ g^{-1}) = 1_U$ i $(f^{-1} \circ g^{-1}) \circ (g \circ f) = 1_S$.

11. (a) Udowodnij: jeśli $s_1, s_2 \in S$ i $f(s_1) = f(s_2)$, to $s_1 = s_2$. Dowód będzie bardzo krótki.

13. (a) Przypuśćmy, że $t \in f(f^{\leftarrow}(B))$. Wtedy $t = f(s)$ dla pewnego $s \in f^{\leftarrow}(B)$. Ale $s \in f^{\leftarrow}(B)$ oznacza, że $f(s) \in B$. Zatem $t \in B$. Jest tak dla dowolnego t, więc $f(f^{\leftarrow}(B)) \subseteq B$.
(c) Po pierwsze, $f(f^{\leftarrow}(B_1) \cap f^{\leftarrow}(B_2)) \subseteq f(f^{\leftarrow}(B_1)) \cap f(f^{\leftarrow}(B_2)) = B_1 \cap B_2$, a więc $f^{\leftarrow}(B_1) \cap f^{\leftarrow}(B_2) \subseteq f^{\leftarrow}(B_1 \cap B_2)$. W przeciwną stronę: $f(f^{\leftarrow}(B_1 \cap B_2)) = B_1 \cap B_2 \subseteq B_1$, a więc $f^{\leftarrow}(B_1 \cap B_2) \subseteq f^{\leftarrow}(B_1)$ i podobnie $f^{\leftarrow}(B_1 \cap B_2) \subseteq f^{\leftarrow}(B_2)$. Zatem $f^{\leftarrow}(B_1 \cap B_2) \subseteq f^{\leftarrow}(B_1) \cap f^{\leftarrow}(B_2)$.

15. (a) Pierwsze zdanie pokazuje, że funkcja f w tym przykładzie nie może być różnowartościowa. Z drugiej strony, jeśli f jest funkcją stałą, to $f \circ g = f \circ h$ dla dowolnych funkcji g i h. Podaj konkretny przykład.
(c) Jeśli funkcje g i h mają tę samą dziedzinę i jeśli przeciwdziedziną funkcji f jest ta dziedzina, to z równości $g \circ f = h \circ f$ wynika $g = h$.

Paragraf 1.5

1. (a) 42. (c) 1. (e) 154.

3. (a) 3, 12, 39 i 120. (c) 3, 9 i 45.

5. (a) -2, 2, 0, 0 i 0.

7. (a) $0, 1/3, 1/2, 3/5, 2/3, 5/7$.
(c) Zauważ, że $a_{n+1} = \dfrac{(n+1)-1}{(n+1)+1} = \dfrac{n}{n+2}$ dla $n \in \mathbb{P}$.

9. (a) $0, 0, 2, 6, 12, 20, 30$.
(c) Po prostu podstaw odpowiednie wartości do obu stron.

11. (a) $0, 0, 1, 1, 2, 2, 3, 3, \ldots$.
(c) $(0,0), (0,1), (1,1), (1,2), (2,2), (2,3), (3,3), \ldots$.

13. (a)

n	n^4	4^n	n^{20}	20^n	$n!$
5	625	1024	$9{,}54 \cdot 10^{13}$	$3{,}2 \cdot 10^6$	120
10	10^4	$1{,}05 \cdot 10^6$	10^{20}	$1{,}02 \cdot 10^{13}$	$3{,}63 \cdot 10^6$
25	$3{,}91 \cdot 10^5$	$1{,}13 \cdot 10^{15}$	$9{,}09 \cdot 10^{27}$	$3{,}36 \cdot 10^{32}$	$1{,}55 \cdot 10^{25}$
50	$6{,}25 \cdot 10^6$	$1{,}27 \cdot 10^{30}$	$9{,}54 \cdot 10^{33}$	$1{,}13 \cdot 10^{65}$	$3{,}04 \cdot 10^{64}$

15. $2^n = 10^{n \cdot \log_{10} 2}$. Dlaczego? Wartości w tabeli wyglądają nieco inaczej, ponieważ wykładniki $n \cdot \log_{10} 2$ nie są liczbami całkowitymi.

Paragraf 1.6

1. (a) $k = 2$. (c) $k = 12$.

3. (a) $n!$. Zauważ, że $3^n \neq O(2^n)$, ale $3^n = O(n!)$; zob. przykład 3(b) lub ćwiczenie 8.
(c) $\log_2 n$.

5. (a) Prawdziwe. Weź $C \geq 2$.
(c) Fałszywe. $2^{2n} \leq C \cdot 2^n$ tylko wtedy, gdy $2^n \leq C$, tzn. tylko dla $n \leq \log_2 C$.

7. (a) Fałszywe. Jeśli $40^n \leq C \cdot 2^n$ dla dostatecznie dużych n, to $20^n \leq C$ dla dostatecznie dużych n.
(c) Fałszywe. Jeśli $(2n)! \leq C \cdot n!$ dla dużych n, to $(n+1)! \leq C \cdot n!$ dla dostatecznie dużych n, a więc $n + 1 \leq C$ dla dużych n, co jest niemożliwe.

9. (a) Nierówność $\frac{1}{k^2} \leq \left(\frac{1}{k-1} - \frac{1}{k}\right)$ jest równoważna z nierównością $(k-1)k \leq k^2$. Aby wykazać równość podaną we wskazówce, wypisz kilka początkowych wyrazów sumy.

11. (a) Z założenia $f(n) = 3n^4 + a(n)$ i $g(n) = 2n^3 + b(n)$, gdzie $a(n) = O(n)$ i $b(n) = O(n)$. Tak więc $f(n)+g(n) = 3n^4+(a(n)+2n^3+b(n))$ oraz $a(n) + 2n^3 + b(n) = O(n^3)$, ponieważ ciągi $a(n)$, $2n^3$ i $b(n)$ są $O(n^3)$. Zostało tu dwukrotnie użyte twierdzenie 2(b) dla $g(n) = n^3$.

13. (a) Przyjmijmy, że $c \neq 0$, ponieważ dla $c = 0$ teza jest oczywista. Istnieje liczba $C > 0$ taka, że $|f(n)| \leq C \cdot |g(n)|$ dla dużych n. Zatem $|c \cdot f(n)| \leq C \cdot |c| \cdot |g(n)|$ dla dużych n.

15. (a) $a(n)/b(n) = n^4$ nie jest $O(n^3)$ na podstawie przykładu 8(e).

17. (a) Niech $\text{CYFR}(n) = m$. Wtedy liczbę $10^{\text{CYFR}(n)}$ zapisujemy jako jedynkę z m zerami. A więc ta liczba jest dłuższa niż jakakolwiek liczba m-cyfrowa, taka jak n. Liczbę 10^{m-1} zapisujemy jako jedynkę z $m-1$ zerami, a więc jest to najmniejsza liczba m-cyfrowa.

(c) Wykorzystaj ćwiczenie (a). Pokaż, że

$$\text{CYFR}(n) \leq 1 + \log_{10} n = O(\log_{10} n).$$

Paragraf 2.1

1. (a) $p \wedge q$. (c) $\neg p \rightarrow (\neg q \wedge r)$. (e) $\neg r \rightarrow q$.

3. (a) Zdania (b) i (c) są prawdziwe. Pozostałe trzy są fałszywe.

5. Zdanie to jest prawdziwe dla wszystkich $x, y \in [0, \infty)$.

7. (a) $\neg r \rightarrow \neg q$.

(c) Jeśli fałszywe jest to, że $x = 0$ lub $x = 1$, to $x^2 \neq x$.

9. (a) Zdanie $3^3 < 3^3$ jest fałszywe.

11. (a) $(-1+1)^2 = 0 < 1 = (-1)^2$.

(c) Nie. Jeśli $x \geq 0$, to $(x+1)^2 = x^2 + 2x + 1 > x^2$.

13. (a) $(0, -1)$.

15. (a) $p \rightarrow q$. (c) $\neg r \rightarrow p$. (e) $r \rightarrow q$.

17. (a) Prawdopodobnie intencją jest: „jeśli dotkniesz tych ciastek, to sprawię ci lanie". Łatwiej sobie wyobrazić, że zdanie p w implikacji $p \rightarrow q$ jest prawdziwe w takim znaczeniu, niż w znaczeniu „jeśli chcesz dostać lanie, to dotknij tych ciastek".

(c) Jeśli nie odejdziesz, to poszczuję cię psem.

(e) Jeśli nie przestaniesz, to odejdę.

Paragraf 2.2

1. (a) Odwrotne: $(q \wedge r) \to p$. Przeciwstawne: $\neg(q \wedge r) \to \neg p$.
(c) Odwrotne: jeśli $3 + 3 = 8$, to $2 + 2 = 4$.
Przeciwstawne: jeśli $3 + 3 \neq 8$, to $2 + 2 \neq 4$.

3. (a) $q \to p$. (c) $p \to q$, $\neg q \to \neg p$, $\neg p \vee q$.

5. (a) 0. (c) 1.
Uwaga: w matrycach logicznych podane są tylko ostateczne kolumny.

7. (a)

p	q	$\neg(p \wedge q)$
0	0	1
0	1	1
1	0	1
1	1	0

(c)

p	q	$\neg p \wedge \neg q$
0	0	1
0	1	0
1	0	0
1	1	0

9.

p	q	r	ostateczne kolumny
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0

11.

p	q	r	część (a)	część (b)
0	0	0	0	0
0	0	1	1	0
0	1	0	1	1
0	1	1	1	0
1	0	0	1	1
1	0	1	1	0
1	1	0	1	1
1	1	1	1	0

13. (b)

p	$p \oplus p$
0	0
1	0

(d)

p	$(p \oplus p) \oplus p$
0	0
1	1

15. (a) Zabronione jest łowienie ryb i zabronione jest polowanie. Szkoła nie będzie otwarta w lipcu i nie będzie otwarta w sierpniu.

19. (a) Wystarczy sprawdzić tylko te wiersze, w których zdanie $[(p \wedge r) \to (q \wedge r)]$ jest fałszywe, tzn. w których zdanie $(p \wedge r)$ jest prawdziwe i zdanie $(q \wedge r)$ jest fałszywe. Zatem pozostaje do sprawdzenia jeden wiersz:

p	q	r
1	0	1

(c) Wystarczy sprawdzić tylko te wiersze, w których zdanie $[(p \wedge r) \to (q \wedge s)]$ jest fałszywe, tzn. w których zdanie $(p \wedge r)$ jest prawdziwe i zdanie $(q \wedge s)$ jest fałszywe. Zatem pozostają do sprawdzenia trzy wiersze:

p	q	r	s
1	0	1	0
1	0	1	1
1	1	1	0

21. Niech p = „dokończył kolację" i q = „został wysłany do łóżka". Wtedy zdanie p jest prawdziwe i zdanie q jest prawdziwe, a więc zdanie wypowiedziane przez matkę $\neg p \to \neg q$ ma wartość logiczną prawdy. Z logicznego punktu widzenia miała ona rację, choć nie było to miłe.

23. (a) Popatrzmy na matryce logiczne. Zdanie B ma wartość 1 w każdym wierszu, w którym zdanie A ma wartość 1, zdanie C ma wartość 1 w każdym wierszu, w którym zdanie B ma wartość 1, a więc zdanie C ma wartość 1 w każdym wierszu, w którym zdanie A ma wartość 1.

(c) Wiemy, że $P \Rightarrow Q$. Ponieważ $Q \Rightarrow R$ i $R \Rightarrow P$, więc z ćwiczenia (a) mamy $Q \Rightarrow P$. Zatem $P \Leftrightarrow Q$.

Paragraf 2.3

1. Podaj dowód bezpośredni korzystając z następującego faktu. Jeśli m i n są liczbami parzystymi, to istnieją liczby $j, k \in \mathbb{Z}$ takie, że $m = 2j$ i $n = 2k$.

3. Można tego dowieść rozpatrując cztery przypadki: por. przykład 5.

5. Można tego dowieść rozpatrując trzy przypadki: (1) $n = 3k$ dla pewnej liczby $k \in \mathbb{N}$; (2) $n = 3k + 1$ dla pewnej liczby $k \in \mathbb{N}$; (3) $n = 3k + 2$ dla pewnej liczby $k \in \mathbb{N}$.

7. (a) Podaj dowód bezpośredni, tak jak w ćwiczeniu 1.

(c) Fałszywe. Na przykład $2 + 3$, $2 + 5$ lub $2 + 11$.

(e) Fałszywe.

9. (a) Prawdziwe trywialnie.

(c) Prawdziwe „w próżni".

11. W przykładzie 2 widzieliśmy, że zbiór liczb pierwszych jest nieskończony. Rozumowanie podobne do rozumowania z przykładu 10 pokazuje, że pewne dwie liczby pierwsze mają takich samych sześć ostatnich cyfr. Ten dowód jest niekonstruktywny.

13. (a) Żadna z liczb w zbiorze

$$\{k \in \mathbb{N}: \ (n+1)! + 2 \leq k \leq (n+1)! + (n+1)\}$$

nie jest pierwsza, gdyż jeśli $2 \leq m \leq n + 1$, to liczba m jest dzielnikiem $(n + 1)!$, a więc jest też dzielnikiem $(n + 1)! + m$.

(b) Tak. Ponieważ $7! = 5040$, więc ten dowód pokazuje, że liczby od 5042 do 5047 są złożone.

(c) Po prostu dołącz 5048 do listy z ćwiczenia (b). Inny ciąg siedmiu kolejnych liczb złożonych zaczyna się od liczby 90.

15. (a) $14 = 2 \cdot 7$ i 7 jest liczbą nieparzystą. Zatem $14 = 2^1 \cdot 7$.

(c) $96 = 2 \cdot 48 = 2 \cdot 2 \cdot 24 = 2 \cdot 2 \cdot 2 \cdot 12 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 6 = 2 \cdot 2 \cdot 2 \cdot 2 \cdot 2 \cdot 3$, zatem $96 = 2^5 \cdot 3$.

Paragraf 2.4

1. (a) $\neg(p \to q) \to ((p \to q) \to p)$. (c) $p \vee \neg p$.

3. (a) Prawo 14, w którym zdanie q zastąpiono zdaniem $q \wedge r$.
(c) Prawo 8a, w którym zdanie p zastąpiono zdaniem $\neg p \wedge r$ i zdanie q zastąpiono zdaniem $q \to r$.

5. (a) Prawo 2a i reguła podstawiania (b).
(c) Prawa 10a i 1 i reguła podstawiania (b).

7. (a) Prawo 10a (gdzie zdanie q zastąpiono zdaniem s, korzystając z reguły podstawiania (a)) i prawo 11a (gdzie zdanie p zastąpiono zdaniem s i zdanie q zdaniem t, korzystając z reguły podstawiania (a)) oraz reguła podstawiania (b).
(c) Prawo 3a (gdzie zdanie p zastąpiono zdaniem $\neg p$, zdanie q zdaniem s i zdanie r zdaniem s, korzystając z reguły podstawiania (a)) oraz reguła podstawiania (b).
(e) Jeszcze raz prawo 3a i reguła podstawiania (a).

9. Kolejne równoważności otrzymujemy korzystając ze wskazanych praw za pomocą odpowiednich podstawień.

(a) $[(p \vee r) \wedge (q \to r)]$

$[(p \vee r) \wedge (\neg q \vee r)]$	prawo 10a
$[(r \vee p) \wedge (r \vee \neg q)]$	dwukrotnie prawo 2a
$r \vee (p \wedge \neg q)$	prawo 4a
$(p \wedge \neg q) \vee r$	prawo 2a

(c) $p \vee (\neg p \wedge \neg q)$

$(p \vee \neg p) \wedge (p \vee \neg q)$	prawo 4a
$t \wedge (p \vee \neg q)$	prawo 7a
$p \vee \neg q$	prawa 2b i 6d

11. Rozpatrz przypadki

p	q	r
1	0	1

lub

p	q	r
1	1	0

oraz

p	q	r
0	0	1

lub

p	q	r
0	1	0

13. 1, 2, 3: założenia
4: prawo 16, tautologia
5: 4, 1 i reguła sylogizmu hipotetycznego 33
6: 5, 3 i reguła 33
7: 6, 2 i reguła modus tollens 31

15. (a) Weź dowód oryginalny i zmień uzasadnienie dla A z „założenie" na "tautologia", tzn. sam dowód nie musi być zmieniony.
(b) Niech A będzie tautologią z przykładu 8 i skorzystaj z ćwiczenia (a).

17. (a) Skorzystaj z praw 11a i 11b.

(c) Nie. Każde zdanie, w którym występują tylko p, q, $\wedge$ i $\vee$, ma wartość logiczną 0, gdy oba zdania p i q mają wartość 0. Por. ćwiczenie 17 w § 7.2.

19. Użyj matryc logicznych.

Paragraf 2.5

1. To rozumowanie nie jest poprawne, gdyż założenia są prawdziwe, jeśli zdanie C jest prawdziwe i zdanie A jest fałszywe. Błąd polega na potraktowaniu zdań $A \to C$ i $\neg A \to \neg C$ jak zdań równoważnych.

3. (a) i (b). Zobacz ćwiczenie (c).

(c) Ten argument nie jest mocniejszy. Jeśli zdanie C oraz wszystkie zdania A_i są fałszywe, to każde założenie $A_i \to C$ jest prawdziwe, niezależnie od tego, czy zdanie C jest prawdziwe.

5. (a) Przy naturalnych oznaczeniach mamy założenia $\neg t \to \neg u$, $u \to e$ i $\neg e$. Możemy wyprowadzić zdanie $\neg u$, korzystając z prawa kontrapozycji. Nie możemy wyprowadzić ani zdania t, ani zdania $\neg t$. Oczywiście możemy wyprowadzić bardziej skomplikowane zdania, takie jak $\neg e \vee u$ czy $(u \wedge t) \to e$.

(c) Założeniami są $(p \vee d) \to r$, $n \to r$ i $\neg n$. Nie można wyprowadzić interesujących wniosków, takich jak p czy $\neg r$.

7. (a) Prawdziwe. Zdanie $A \to B$ jest założeniem. Pokazaliśmy, że zdanie $\neg A \to \neg B$ wynika z tego założenia.

(c) Prawdziwe. Z przyjętego założenia $(B \vee \neg Y) \to A$ wynika $\neg Y \to A$, a więc równoważnie $\neg\neg Y \vee A$.

9. (a) Przyjmijmy: o = „moje obliczenia zgadzają się”, r = „zapłacę rachunek za elektryczność”, z = „zabraknie mi pieniędzy” oraz p = „prądu mi nie wyłączą”. Twierdzenie mówi więc, że jeśli $(o \wedge r) \to z$ i $\neg r \to \neg p$, to $(\neg z \wedge p) \to \neg o$. Oto jeden z możliwych dowodów; uzupełnij brakujące wyjaśnienia.

1. $(o \wedge r) \to z$
2. $\neg r \to \neg p$
3. $\neg z \to \neg(o \wedge r)$ — 1; prawo kontrapozycji 9
4. $\neg z \to (\neg o \vee \neg r)$
5. $p \to r$
6. $(\neg z \wedge p) \to [(\neg o \vee \neg r) \wedge r]$ — 4,5; reguła wnioskowania odpowiadająca prawu 26b
7. $(\neg z \wedge p) \to [r \wedge (\neg o \vee \neg r)]$
8. $(\neg z \wedge p) \to [(r \wedge \neg o) \vee (r \wedge \neg r)]$
9. $(\neg z \wedge p) \to [(r \wedge \neg o) \vee$ sprzeczność$]$
10. $(\neg z \wedge p) \to (r \wedge \neg o)$ — 9; prawo identyczności 6a

11. $(\neg z \wedge p) \rightarrow (\neg o \wedge r)$
12. $(\neg o \wedge r) \rightarrow \neg o$ opuszczanie koniunkcji 17
13. $(\neg z \wedge p) \rightarrow \neg o$

(c) Przyjmijmy: p =„dostanę pracę”, c =„będę ciężko pracować”, a =„będę awansować”, z =„będę zadowolony”. Twierdzenie mówi wtedy, że: jeśli $(p \wedge c) \rightarrow a$, $a \rightarrow z$ i $\neg z$, to $\neg p \vee \neg c$. Oto jeden z możliwych dowodów; uzupełnij brakujące wyjaśnienia.

1. $(p \wedge c) \rightarrow a$
2. $a \rightarrow z$
3. $\neg z$
4. $\neg a$
5. $\neg(p \wedge c)$
6. $\neg p \vee \neg c$

11. Wykaż jak najszybciej s. Wtedy regułę sylogizmu hipotetycznego z przykładu 3 można zastąpić regułą modus ponens.

13. (a) Oto jeden z możliwych dowodów; uzupełnij brakujące wyjaśnienia.

1. $A \wedge \neg B$
2. A
3. $A \rightarrow P$
4. P
5. $\neg B$
6. $P \wedge \neg B$

Paragraf 3.1

1. (a) Relacja R_1 spełnia własność (PZ) i (S).
(c) Relacja R_3 spełnia własności (Z), (AS) i (P).
(e) Relacja R_5 spełnia tylko własność (S).

3. Relacje w (a) i (c) są zwrotne. Relacje w (c), (d), (f), (g) i (h) są symetryczne.

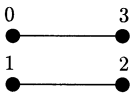
5. Relacja R_1 spełnia własności (PZ) i (S). Relacje R_2 i R_3 spełniają tylko własność (S).

7. (a) Relacja pusta spełnia (PZ), (S), (AS) i (P). Ostatnie trzy własności spełnione są „w próżni”.

9. (a) Jeśli $E \subseteq R_1$ i $E \subseteq R_2$, to $E \subseteq R_1 \cap R_2$.
(c) Przypuśćmy, że R_1 i R_2 są przechodnie. Jeśli $(x,y),(y,z) \in R_1 \cap R_2$, to $(x,y),(y,z) \in R_1$, a więc $(x,z) \in R_1$. Podobnie $(x,z) \in R_2$.

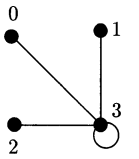
11. (a) Załóżmy, że relacja R jest symetryczna. Jeśli $(x,y) \in R$, to $(y,x) \in R$ z własności symetrii, a więc $(x,y) \in R^{\leftarrow}$. Podobnie, z $(x,y) \in R^{\leftarrow}$, wynika, że $(x,y) \in R$ (sprawdź), a więc $R = R^{\leftarrow}$. Na odwrót, przyjmij, że $R = R^{\leftarrow}$ i pokaż, że relacja R jest symetryczna.

13. (a)



(c) Zob. rysunek 3.1(a).

(e)



Paragraf 3.2

1. (a)

e	a	b	c	d	e	f
$\gamma(e)$	(x,v)	(v,x)	(v,w)	(w,y)	(w,y)	(y,x)

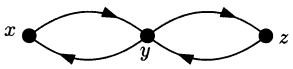
(c)

e	a	b	c	d
$\gamma(e)$	(x,w)	(w,x)	(y,z)	(z,y)

3. (a) Tak.
(c) Nie. Nie ma krawędzi od wierzchołka t do wierzchołka x.
(e) Tak.

5. (a) xwy lub $xwvzy$.
(c) vxw lub vzw lub $vzxw$ lub $vzyxw$.
(e) $zyxwv$ lub zwv lub $zxwv$ lub zyv.

7. Oto jeden przykład:



9. (a) (v,w), (v,y), (v,z). Zauważ, że (v,z) jest w relacji osiągalności, ale nie jest w relacji sąsiedztwa.
(c) Każda z nich. Relacja osiągalności jest relacją uniwersalną.

11. (a) 2.
(c) 3.
(e) To nie jest ciąg wierzchołków żadnej drogi.
(g) 3.

13. (a) Krawędzie e, f i g są wielokrotne.

15. (a) $A = \{(w,w),(w,x),(x,w),(y,y),(w,y),(y,w),(x,z),(z,x)\}$, podczas gdy relacja R zawiera wszystkie 16 par uporządkowanych wierzchołków.

17. (a) cad lub cbd. Obie mają ciąg wierzchołków równy $yvww$.
(c) Są cztery takie drogi, każda z ciągiem wierzchołków $vwwvy$.

Paragraf 3.3

1. (a) 1.
(c) 2.

3. (a) $\begin{bmatrix} -1 & 1 & 4 \\ 0 & 3 & 2 \\ 2 & -2 & 3 \end{bmatrix}$. (c) $\begin{bmatrix} 5 & 8 & 7 \\ 5 & 1 & 5 \\ 7 & 3 & 5 \end{bmatrix}$. (e) $\begin{bmatrix} 5 & 5 & 7 \\ 8 & 1 & 3 \\ 7 & 5 & 5 \end{bmatrix}$.

(g) $\begin{bmatrix} 12 & 12 & 8 \\ 12 & -4 & 8 \\ 8 & 8 & 4 \end{bmatrix}$. (i) $\begin{bmatrix} 4 & 8 & 9 \\ 6 & 4 & 3 \\ 11 & 5 & 8 \end{bmatrix}$.

5. (a) $\begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$. (c) Nie jest określona.

(e) $\begin{bmatrix} 3 & 2 & 5 & 4 \\ 2 & 5 & 4 & 7 \\ 5 & 4 & 7 & 6 \end{bmatrix}$.

7. (a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$

$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

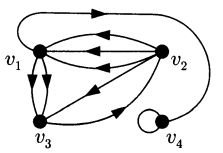
9. (a) $\begin{bmatrix} 1 & 0 \\ n & 1 \end{bmatrix}$. (c) $\{n \in \mathbb{N}: \ n \text{ jest nieparzysta}\}$.

11. (a) Wyrazem o indeksach (i, j) macierzy $a\mathbf{A}$ jest $a\mathbf{A}[i, j]$. Podobnie dla $b\mathbf{B}$, a więc wyrazem o indeksach (i, j) macierzy $a\mathbf{A} + b\mathbf{B}$ jest $a\mathbf{A}[i, j] + b\mathbf{B}[i, j]$. Zatem wyrazem o indeksach (i, j) macierzy $c(a\mathbf{A} + b\mathbf{B})$ jest $ca\mathbf{A}[i, j] + cb\mathbf{B}[i, j]$. W podobny sposób pokazujemy, że jest to wyraz o indeksach (i, j) macierzy $(ca)\mathbf{A} + (cb)\mathbf{B}$. Ponieważ wszystkie wyrazy tych macierzy są równe, to macierze $c(a\mathbf{A} + b\mathbf{B})$ i $(ca)\mathbf{A} + (cb)\mathbf{B}$ są równe.

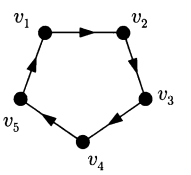
(c) Porównaj wyrazy o indeksach (j, i) macierzy $(a\mathbf{A})^T$ i $a\mathbf{A}^T$.

13. (a) $\begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$. (c) $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

15. (a)



(c)



17. (a) $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$. (c) Zob. przykład 5(a). (e) $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$.

19. (a) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. (c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. (e) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

(g) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$. (i) $\begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Paragraf 3.4

1. (a) $\begin{bmatrix} -8 & 13 \\ 2 & 9 \end{bmatrix}$. (c) $\begin{bmatrix} 31 & -16 & -6 \\ 29 & 4 & 26 \end{bmatrix}$. (e) $\begin{bmatrix} -1 & 7 \\ 8 & 0 \\ 16 & 0 \end{bmatrix}$.

3. Iloczyny zapisane w (a), (c) i (e) nie istnieją.

5. (a) $\begin{bmatrix} 1 & 10 \\ 11 & 19 \end{bmatrix}$.

7. (a) $\begin{bmatrix} 7 & 14 \\ 8 & 11 \\ 2 & -6 \end{bmatrix}$.

9. (a) 2. (c) 2.

11. (a) $\mathbf{M}^3 = \begin{bmatrix} 3 & 20 & 2 & 3 \\ 0 & 8 & 0 & 0 \\ 2 & 9 & 1 & 2 \\ 0 & 4 & 0 & 0 \end{bmatrix}$.

(c) fab, fac, fbd, fbe, fcd, fce, fhj, kjd, kje.

13. (a) Po prostu usuń strzałki z rys. 3.18.
(c) $dd, ee, de, ed, bb, cc, bc, cb, jj$.

15. (a) $\mathbf{I}^{-1} = \mathbf{I}$. (c) Nie jest odwracalna. (e) $\mathbf{D}^{-1} = \mathbf{D}$.

17. (b) Właściwa odpowiedź brzmi: $\mathbf{A}^n = \begin{bmatrix} 1 & 0 \\ n & 1 \end{bmatrix}$.

19. Dla $1 \leq k \leq p$ i $1 \leq i \leq m$

$$(\mathbf{B}^T\mathbf{A}^T)[k,i] = \sum_{j=1}^{n} \mathbf{B}^T[k,j]\mathbf{A}^T[j,i] = \sum_{j=1}^{n} \mathbf{B}[j,k]\mathbf{A}[i,j].$$

Porównaj to z wyrazem o indeksach (k,i) macierzy $(\mathbf{AB})^T$.

21. (a) W rzeczywistości $\mathbf{AB} = \mathbf{BA} = a\mathbf{B}$ dla wszystkich $\mathbf{B}$ ze zbioru $\mathfrak{M}_{2,2}$.

(b) Z równości $\mathbf{AB} = \mathbf{BA}$ dla $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ wynika, że $\begin{bmatrix} a & 0 \\ c & 0 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$, a więc $b = c = 0$. Zatem $\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$. Teraz spróbuj $\mathbf{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

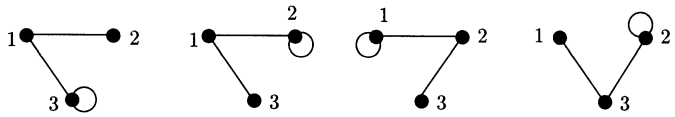
23. (a) Weź $1 \leq i \leq m$ i $1 \leq k \leq p$ i porównaj wyrazy o indeksach (i,k) macierzy $(\mathbf{A}+\mathbf{B})\mathbf{C}$ oraz $\mathbf{AC}+\mathbf{BC}$.

Paragraf 3.5

1. (a) Jest relacją równoważności.
(c) Również jest relacją równoważności, chyba że rozpatruje się również te nieliczne osoby, które mają mieszkania w więcej niż jednym stanie.
(e) Relacja $\approx$ nie jest relacją równoważności, ponieważ nie jest ona przechodnia.

3. Jak najbardziej.

5. (a) Są następujące możliwości:



7. (a) Sprawdź bezpośrednio lub zastosuj twierdzenie 2(a) do funkcji określonej wzorem $f(m) = m^2$ dla $m \in \mathbb{Z}$.

9. (a) Jest nieskończenie wiele takich klas: $\{0\}$ oraz klasy $\{n, -n\}$ dla $n \in \mathbb{P}$.

11. Zastosuj twierdzenie 2, używając funkcji długości. Klasami równoważności są zbiory Σ^k, dla $k \in \mathbb{N}$.

13. (a) Użyj brutalnej siły lub zastosuj twierdzenie 2(a) i ćwiczenie (b).

15. (a) Nie jest dobrze określona: zależy ona od wyboru reprezentantów. Na przykład $[3] = [-3]$ i $-3 \leq 2$. Gdyby definicja miała sens, otrzymalibyśmy $[3] = [-3] \leq [2]$ i stąd $3 \leq 2$.
(c) Nie dzieje się nic złego. Jeśli $[m] = [n]$, to $m^4+m^2+1 = n^4+n^2+1$.

17. (a) Ponieważ relacja $\sim$ jest zwrotna i symetryczna, więc również relacja $\approx$ ma te własności. Dla danych ciągów funkcji $f = f_1 \sim f_2 \sim \ldots \sim f_n = g$ oraz $g = g_1 \sim g_2 \sim \ldots \sim g_m = h$ mamy ciąg

$$f = f_1 \sim f_2 \sim \ldots \sim f_n \sim g_2 \sim \ldots \sim g_m = h,$$

więc $f \approx h$. Zatem relacja $\approx$ jest przechodnia.

Paragraf 3.6

1. (a) $q = 6, r = 2$. (c) $q = -7, r = 1$. (e) $q = 5711, r = 31$.

3. (a) $-4, 0, 4$. (c) $-2, 2, 6$. (e) $-4, 0, 4$.

5. (a) 1. (c) 1. (e) 0.

7. (a) 3 oraz 2.
(c) $m *_{10} k$ jest ostatnią (dziesiętną) cyfrą liczby $m \cdot k$.

9.

$+_4$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$*_4$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

11. Rozwiązaniami są odpowiednio 1, 3, 2 i 4.

13. (a) $m \equiv n \pmod 1$ dla wszystkich $m, n \in \mathbb{Z}$. Jest tylko jedna klasa równoważności.
(c) $0 = 0 +_1 0$ oraz $0 = 0 *_1 0$.

15. (a) $n = 1000a + 100b + 10c + d = a + b + c + d + 9 \cdot (111a + 11b + c)$
$\equiv a + b + c + d \pmod 9$
lub zastosuj twierdzenie 2 do kongruencji

$$1000 \equiv 100 \equiv 10 \equiv 1 \pmod 9.$$

17. Tak jak ćwiczenie 15. Zauważ, że $1000 = 91 \cdot 11 - 1$, $100 = 9 \cdot 11 + 1$, $10 = 1 \cdot 11 - 1$, więc $1000a + 100b + 10c + d \equiv -a + b - c + d \equiv 0 \pmod{11}$ wtedy i tylko wtedy, gdy $a - b + c - d \equiv 0 \pmod{11}$.

19. Mamy $q \cdot p - q' \cdot p = r' - r$, tak więc $r' - r$ jest wielokrotnością p. Ale $-p < -r \leq r' - r \leq r' < p$ i 0 jest jedyną wielokrotnością p zawartą między liczbami $-p$ i p. Zatem $r' = r$, czyli $0 = (q - q') \cdot p$ i $q = q'$.

21. (a) Z twierdzenia 3(a)

$$\begin{aligned}(m \bmod p) +_p (n \bmod p) &= (m + n) \bmod p = (n + m) \bmod p \\ &= (n \bmod p) +_p (m \bmod p).\end{aligned}$$

Ponieważ $m, n \in \mathbb{Z}_p$, więc $m \bmod p = m$ oraz $n \bmod p = n$.

Paragraf 4.1

1. (a) 0, 3, 9, 21, 45. (c) 1, 1, 1, 1, 1.

3. (a)

	m	n
na początku	0	0
po pierwszym przebiegu	1	1
po drugim przebiegu	4	2
po trzecim przebiegu	9	3
po czwartym przebiegu	16	4

5. (a) 4, 16, 36, 64.

7. (a) Jeśli $m + n$ jest liczbą parzystą, to również $(m + 1) + (n + 1) = (m + n) + 2$ jest liczbą parzystą. Oczywiście, aby się o tym przekonać, nie potrzebowaliśmy warunku dozoru pętli $1 \leq n$.

9. (a) Tak. Jeśli $i < j^2$ oraz $j \geq 1$, to $i+2 < j^2+2 < j^2+2j+1 = (j+1)^2$.
(c) Nie. Rozważ przypadek $i = j = 0$.

11. (a) $b \geq 2$. (c) $b \in \mathbb{N}$.

13. Nie. Ciąg poleceń „k: $= k^2$, wypisz k" zmienia wartość k. Algorytm A wypisuje 1, 4 i zatrzymuje się. Algorytm B wypisuje 1, 4, 9, 16, ponieważ instrukcja „dla" za każdym razem nadaje k wartość początkową.

15. (a) Tak; nowe $r < 73$ z definicji funkcji MOD.
(c) Jest to niezmiennik „w próżni", gdyż oba zdania $r \leq 0$ i $r > 0$ nie mogą być jednocześnie prawdziwe na początku pętli.

17. Zbiory w przykładach (a), (c) i (f) mają element najmniejszy; pozostałe nie mają.

19. (a) To jest niezmiennik pętli: jeśli $r > 0$ oraz a, b i r są wielokrotnościami 5, to nowe wartości b, r i $b \operatorname{MOD} r$ są wielokrotnościami 5. Zauważ, że $b \operatorname{MOD} r = b - (b \operatorname{DIV} r) \cdot r$.
(c) To jest niezmiennik pętli.

21. (a) Jeśli m jest liczbą parzystą, to $2^{(\text{nowe } k)} \cdot (\text{nowe } m) = 2^{k+1} \cdot m/2 = 2^k \cdot m = n$.

23. (a)

$a = 2$, $n = 11$	p	q	i
na początku	1	2	11
po pierwszym przebiegu	2	4	5
po drugim przebiegu	8	16	2
po trzecim przebiegu	8	256	1
po czwartym przebiegu	$256 \cdot 8$	256^2	0

(b) Pokaż, że jeśli $q^i \cdot p = a^n$, to $(\text{nowe } q)^{(\text{nowe } i)}(\text{nowe } p) = a^n$; rozważ przypadki, kiedy i jest liczbą nieparzystą i kiedy i jest liczbą parzystą. Ponieważ i przebiega malejący ciąg liczb całkowitych nieujemnych, więc ostatecznie $i = 0$ i algorytm wychodzi z pętli. W tym momencie $p = a^n$.

Paragraf 4.2

Dowody indukcyjne powinny być napisane starannie i dokładnie. Te odpowiedzi służą tylko jako wskazówki, a nie jako wzorce.

1. Sprawdź warunek początkowy. W kroku indukcyjnym przyjmij, że równość jest prawdziwa dla k. Wtedy

$$\sum_{i=1}^{k+1} i^2 = \sum_{i=1}^{k} i^2 + (k+1)^2 = \frac{k(k+1)(2k+1)}{6} + (k+1)^2.$$

Po przekształceniach algebraicznych otrzymujemy po prawej stronie

$$\frac{(k+1)(k+2)(2k+3)}{6},$$

a więc równość jest prawdziwa dla $k+1$, jeśli tylko była prawdziwa dla k.

3. (a) Weź $n = 37^{20}$ w przykładzie 1.
(c) Na podstawie punktów (a), (b) i przykładu 1 liczba

$$(37^{500} - 37^{100}) + (37^{100} - 37^{20}) + (37^{20} - 37^4)$$

jest wielokrotnością 10.
(e) Na podstawie (c) i (d), tak jak w (c).

5. Warunkiem początkowym jest „$s_0 = 2^0 a + (2^0 - 1)b$". Załóżmy indukcyjnie, że $s_k = 2^k a + (2^k - 1)b$ dla jakiegoś $k \in \mathbb{N}$. Przekształcenia algebraiczne w kroku indukcyjnym są następujące:

$$2 \cdot [2^k a + (2^k - 1)b] + b = 2^{k+1} a + 2^{k+1} b - 2b + b.$$

7. Pokaż, że $11^{k+1} - 4^{k+1} = 11 \cdot (11^k - 4^k) + 7 \cdot 4^k$. Powtórz rozumowanie z przykładu 2(d).

9. (a) Przypuśćmy, że $\sum_{i=0}^{k} 2^i = 2^{k+1} - 1$ i $0 \leq k$. Wtedy

$$\sum_{i=0}^{k+1} 2^i = \left(\sum_{i=0}^{k} 2^i\right) + 2^{k+1} = 2^{k+1} - 1 + 2^{k+1} = 2^{k+2} - 1,$$

a więc równość nadal jest prawdziwa dla nowej wartości k.
(c) Tak. $\sum_{i=0}^{0} 2^i = 1 = 2^1 - 1$ na początku, a więc pętla nigdy się nie zakończy i niezmiennik jest spełniony dla każdej wartości $k \in \mathbb{N}$.

11. (a) $1 + 3 + \ldots + (2n - 1) = n^2$.
(b) W kroku indukcyjnym

$$k^2 + [2(k + 1) - 1] = k^2 + 2k + 1 = (k + 1)^2.$$

13. (a) Załóż, że zdanie $p(k)$ jest prawdziwe. Wtedy $(k+1)^2+5(k+1)+1 = (k^2 + 5k + 1) + (2k + 6)$. Ponieważ liczba $k^2 + 5k + 1$ jest parzysta z założenia i liczba $2k + 6$ jest też oczywiście parzysta, to zdanie $p(k + 1)$ jest prawdziwe.
(b) Wszystkie zdania $p(n)$ są fałszywe. *Morał*: warunek początkowy jest niezbędny w indukcji matematycznej. Zobacz też ćwiczenia 8 i 9.

15. *Wskazówka*: $5^{k+1} - 4(k + 1) - 1 = 5(5^k - 4k - 1) + 16k$.

17. *Wskazówki*:

$$\frac{1}{n+2} + \ldots + \frac{1}{2n+2} = \left(\frac{1}{n+1} + \ldots + \frac{1}{2n}\right) + \left(\frac{1}{2n+1} + \frac{1}{2n+2} - \frac{1}{n+1}\right)$$

oraz

$$\frac{1}{2n+1} + \frac{1}{2n+2} - \frac{1}{n+1} = \frac{1}{2n+1} - \frac{1}{2n+2}.$$

Inaczej, aby uniknąć indukcji, niech $f(n) = \sum_{i=1}^{n} \frac{1}{i}$ i pokaż, że obie strony są równe $f(2n) - f(n)$.

19. *Wskazówki*: $5^{k+2} + 2 \cdot 3^{k+1} + 1 = 5(5^{k+1} + 2 \cdot 3^k + 1) - 4(3^k + 1)$. Pokaż, że liczba $3^n + 1$ jest zawsze parzysta.

21. Niech $p(n)$ będzie tutaj zdaniem „$|\sin nx| \leq n|\sin x|$ dla wszystkich $x \in \mathbb{R}$". Oczywiście zdanie $p(1)$ jest prawdziwe. Po wykonaniu przekształceń algebraicznych i trygonometrycznych, otrzymamy

$$\begin{aligned}|\sin(k+1)x| &= |\sin(kx+x)| = |\sin kx \cos x + \cos kx \sin x| \\ &\leq |\sin kx| \cdot |\cos x| + |\cos kx| \cdot |\sin x| \leq |\sin kx| + |\sin x|.\end{aligned}$$

Załóż teraz, że zdanie $p(k)$ jest prawdziwe i pokaż, że zdanie $p(k+1)$ jest prawdziwe.

Paragraf 4.3

1. (a) $1, 2, 1, 2, 1, 2, 1, 2, \ldots$.

3. (a) $\text{SEQ}(n) = 3^n$.

5. Nie. Wszystko jest w porządku aż do $\text{SEQ}(100)$, ale wartość $\text{SEQ}(101)$ nie jest określona, ponieważ nie możemy dzielić przez zero. Gdybyśmy warunek (R) ograniczyli do liczb n mniejszych lub równych 100, to otrzymalibyśmy rekurencyjnie określony ciąg skończony.

7. (a) $1, 3, 8$. (c) $s_3 = 22$, $s_4 = 60$.

9. (a) $1, 1, 2, 4$.
(c) Nasz wzór nie jest prawdziwy, ponieważ $t_0 \neq 1/2$.

11. (a) $a_6 = a_5 + 2a_4 = a_4 + 2a_3 + 2a_4 = 3(a_3 + 2a_2) + 2a_3 = 5(a_2 + 2a_1) + 6a_2 = 11(a_1 + 2a_0) + 10a_1 = 11 \cdot 3 + 10 = 43$. W tym obliczeniu używa się tylko dwóch adresów wartości pośrednich w danym momencie. Możliwe są inne obliczenia rekurencyjne, w których używa się więcej adresów.
(b) Zastosuj indukcję.

13. Zastosuj zasadę dobrego uporządkowania do zbioru S i skorzystaj z definicji rekurencyjnej ciągu FIB z przykładu 3(a).

15. $\text{SEQ}(n) = 2^{n-1}$ dla $n \geq 1$.

17. (a) $A(1) = 1$. $A(n) = n \cdot A(n-1)$. (c) Tak.

19. (a) $\{1, 110, 1200\}$.

21. (a) (P) $\text{SUMA}(1) = A_1$;
(R) $\text{SUMA}(n) = A_n \cup \text{SUMA}(n-1)$ dla $n \geq 2$.
(b) „Sumą pustą" jest $\emptyset$.

Paragraf 4.4

1. $s_n = 3 \cdot (-2)^n$ dla $n \in \mathbb{N}$.

3. Dowodzimy tego przez indukcję. Równość $s_n = a^n \cdot s_0$ zachodzi dla $n = 0$, gdyż $a^0 = 1$. Jeśli zachodzi ona dla pewnej liczby n, to $s_{n+1} = as_n = a(a^n \cdot s_0) = a^{n+1} \cdot s_0$ i ta równość zachodzi dla $n+1$.

5. $s_0 = 3^0 - 2 \cdot 0 \cdot 3^0 = 1$, $s_1 = 3^1 - 2 \cdot 1 \cdot 3^1 = -3$. Dla $n \geq 2$ mamy

$$\begin{aligned} 6s_{n-1} - 9s_{n-2} &= 6[3^{n-1} - 2(n-1) \cdot 3^{n-1}] - 9[3^{n-2} - 2(n-2) \cdot 3^{n-2}] \\ &= 3^n(1-2n) = s_n. \end{aligned}$$

7. Tym razem $c_1 = 3$ i $c_2 = 0$, a więc $s_n = 3 \cdot 2^n$ dla $n \in \mathbb{N}$.

9. Rozwiązując układ równań $1 = c_1 + c_2$ i $2 = c_1 r_1 + c_2 r_2$ z niewiadomymi c_1 i c_2, otrzymamy $c_1 = (1 + r_1)/\sqrt{5}$ i $c_2 = -(1 + r_2)/\sqrt{5}$. Stąd

$$s_n = \frac{1}{\sqrt{5}}(r_1^n + r_1^{n+1} - r_2^n - r_2^{n+1}) \quad \text{dla wszystkich} \quad n,$$

gdzie r_1, r_2 są takie jak w przykładzie 3.

11. (a) $r_1 = -3, r_2 = 2, c_1 = c_2 = 1$; stąd $s_n = (-3)^n + 2^n$ dla $n \in \mathbb{N}$.
(c) Równanie charakterystyczne ma tu jedno rozwiązanie $r = 2$; $c_1 = 1$ i $c_2 = 3$, a więc $s_n = 2^n + 3n \cdot 2^n$ dla $n \in \mathbb{N}$.
(e) $s_{2n} = 1, s_{2n+1} = 4$ dla wszystkich $n \in \mathbb{N}$.
(g) $s_n = (-3)^n$ dla $n \in \mathbb{N}$.

13. (a) $s_{2^m} = 2^m + 3 \cdot (2^m - 1) = 2^{m+2} - 3$.
(c) $s_{2^m} = \frac{5}{2} \cdot 2^m \cdot m$.
(e) $s_{2^m} = 7 - 6 \cdot 2^m$.
(g) $s_{2^m} = (6 - m)2^{m-1}$.

15. $s_{2^m} = 2^m[s_1 + \frac{1}{2}(2^m - 1)]$. Sprawdź, że ten wzór spełnia warunki $s_{2^0} = s_1$ i $s_{2^{m+1}} = 2s_{2^m} + (2^m)^2$.

17. (a) $t_{2^m} = b^m t_1 + b^{m-1} \cdot \sum_{i=0}^{m-1} \frac{f(2^i)}{b^i}$.

Paragraf 4.5

1. Pierwsza zasada indukcji wystarczy. W kroku indukcyjnym skorzystaj z równości

$$4n^2 - n + 8(n+1) - 5 = 4n^2 + 7n + 3 = 4(n+1)^2 - (n+1).$$

3. Pokaż, że liczba $n^5 - n$ jest zawsze parzysta. Skorzystaj następnie z równości

$$(n+1)^5 = n^5 + 5n^4 + 10n^3 + 10n^2 + 5n + 1$$

(wniosek ze wzoru dwumianowego Newtona).

5. Tak. Nieparzystość liczby a_n zależy wyłącznie od nieparzystości liczby a_{n-1}, gdyż liczba $2a_{n-2}$ jest parzysta, niezależnie od tego, czy liczba a_{n-2} jest parzysta, czy nie.

7. (b) $a_n = 1$ dla wszystkich $n \in \mathbb{N}$.
(c) Trzeba sprawdzić warunek początkowy dla $n = 0$ i $n = 1$. W kroku indukcyjnym weź $n \geq 2$ oraz załóż, że $a_k = 1$ dla $0 \leq k < n$. Wtedy

$$a_n = \frac{a_{n-1}^2 + a_{n-2}}{a_{n-1} + a_{n-2}} = \frac{1^2 + 1}{1 + 1} = 1.$$

To dowodzi kroku indukcyjnego, więc $a_n = 1$ dla wszystkich $n \in \mathbb{N}$ na podstawie drugiej zasady indukcji matematycznej.

9. (b) $a_n = n^2$ dla wszystkich $n \in \mathbb{N}$.

(c) Trzeba sprawdzić warunek początkowy dla $n = 0$ i $n = 1$. W kroku indukcyjnym weź $n \geq 2$ oraz załóż, że $a_k = k^2$ dla $0 \leq k < n$. Aby dowieść kroku indukcyjnego, wykaż, że $a_n = n^2$.

11. (b) Trzeba sprawdzić warunek początkowy dla $n = 0, 1$ i 2. W kroku indukcyjnym weź $n \geq 3$ oraz załóż, że liczba a_k jest nieparzysta dla $0 \leq k < n$. Wykaż, że liczba a_n też jest nieparzysta.

(c) Ponieważ nierówność ma zachodzić dla $n \geq 1$ i ponieważ w kroku indukcyjnym będziesz chciał skorzystać z równości $a_n = a_{n-1} + a_{n-2} + a_{n-3}$, więc w kroku indukcyjnym będziesz potrzebował nierówności $n - 3 \geq 1$. Zatem musisz sprawdzić warunek początkowy dla $n = 1, 2$ i 3. W kroku indukcyjnym weź $n \geq 4$ i załóż, że $a_k \leq 2^{k-1}$ dla $1 \leq k < n$. Aby dowieść kroku indukcyjnego, wykaż, że $a_n \leq 2^{n-1}$.

13. (a) 2, 3, 4, 6.

(b) Trzeba sprawdzić nierówność dla $n = 3, 4$ i 5, zanim zastosuje się drugą zasadę indukcji matematycznej do równości

$$b_n = b_{n-1} + b_{n-3}.$$

(c) Trzeba sprawdzić nierówność dla $n = 2, 3$ i 4. Następnie skorzystaj z drugiej zasady indukcji matematycznej i ćwiczenia (b). W kroku indukcyjnym weź $n \geq 5$ i załóż, że $b_k \geq (\sqrt{2})^{k-2}$ dla $2 \leq k < n$. Wtedy

$$\begin{aligned} b_n &= b_{n-1} + b_{n-3} \geq 2b_{n-3} + b_{n-3} = 3b_{n-3} \\ &\geq 3(\sqrt{2})^{n-5} > (\sqrt{2})^3(\sqrt{2})^{n-5} = (\sqrt{2})^{n-2}. \end{aligned}$$

Można też tego dowieść nie korzystając z ćwiczenia (b).

15. Sprawdź $n = 0$ i 1, zanim przejdziesz do kroku indukcyjnego. Może będzie łatwiej dowieść zdania „$\text{SEQ}(n) \leq 1$ dla wszystkich n" niezależnie od zdania „$\text{SEQ}(n) \geq 0$ dla wszystkich n". Na przykład, załóż, że $n \geq 2$ i $\text{SEQ}(k) \leq 1$ dla $0 \leq k < n$. Wtedy

$$\begin{aligned} \text{SEQ}(n) &= (1/n) \cdot \text{SEQ}(n-1) + ((n-1)/n) \cdot \text{SEQ}(n-2) \\ &\leq (1/n) + ((n-1)/n) = 1. \end{aligned}$$

Dowód tego, że $\text{SEQ}(n) \geq 0$ dla $n \geq 0$ jest prawie taki sam.

17. Pierwsza zasada indukcji wystarczy. Skorzystaj z warunku (R), by sprawdzić równość dla $n = 2$. W kroku indukcyjnym od n do $n + 1$ mamy

$$\begin{aligned} \text{FIB}(n+1) &= \text{FIB}(n) + \text{FIB}(n-1) \\ &= 1 + \sum_{k=0}^{n-2} \text{FIB}(k) + \text{FIB}(n-1) = 1 + \sum_{k=0}^{n-1} \text{FIB}(k). \end{aligned}$$

19. Dla $n > 0$ niech $L(n)$ będzie największą liczbą całkowitą postaci 2^k taką, że $2^k \leq n$. Wykaż, że $L(n) = T(n)$ dla wszystkich n, pokazując najpierw, że $L(\lfloor n/2 \rfloor) = L(n/2)$ dla $n \geq 2$, a następnie korzystając z drugiej zasady indukcji.

21. Za pomocą drugiej zasady indukcji wykaż, że $S(n) \leq n$ dla każdego n.

Paragraf 4.6

1. (a) 20. (c) 1. (e) 4. (g) 6.

3. (a) (20,14), (14,6), (6,2), (2,0); NWD = 2.
(c) (20,30), (30,20), (20,10), (10,0); NWD = 10.

5. (a) NWD$(20, 14) = 2, s = -2, t = 3$.

a	q	s	t
20		1	0
14	1	0	1
6	2	1	−1
2	3	−2	3
0			

(c) NWD$(20, 30) = 10, s = -1, t = 1$.

a	q	s	t
20		1	0
30	0	0	1
20	1	1	0
10	2	−1	1
0			

7. (a) $x = 21[\equiv -5 \pmod{26}]$.
(c) Nie istnieje rozwiązanie, gdyż liczby 4 i 26 nie są względnie pierwsze.
(e) $x = 23[\equiv -3 \pmod{26}]$.

9. (a) $x \equiv 5 \pmod{13}$.
(c) To samo rozwiązanie co w ćwiczeniu (a), gdyż $99 \equiv 8 \pmod{13}$.

11. (a) $x = 99$. A oto szczegóły: $x = 99y$, a więc $8y \equiv 99y \equiv 8 \pmod{13}$. Skróć czynnik 8 po obu stronach, by otrzymać $y \equiv 1 \pmod{13}$. Przyjmując $y = 1$, otrzymamy $x = 99$.
(b) $x = 65$.
(c) $x = 164$. To nie jest przypadek, że sumą rozwiązań ćwiczeń (a) i (b) jest rozwiązanie ćwiczenia (c). Jeśli x jest rozwiązaniem (a) i x' jest rozwiązaniem (b), to $x + x'$ jest rozwiązaniem (c).

13. Załóż, że $a = s \cdot m + t \cdot n$ i $a' = s' \cdot m + t' \cdot n$ na początku pętli. Z równości $a' = s' \cdot m + t' \cdot n$ otrzymamy na końcu równość $a_{\text{nast}} = s_{\text{nast}} \cdot m + t_{\text{nast}} \cdot n$

oraz ostatecznie

$$a'_{\text{nast}} = a' - q \cdot a = s' \cdot m + t' \cdot n - q \cdot s \cdot m - q \cdot t \cdot n$$
$$= (s' - q \cdot s) \cdot m + (t' - q \cdot t) \cdot n = s'_{\text{nast}} \cdot m + t'_{\text{nast}} \cdot n.$$

15. (a) $1 = s \cdot (m/d) + t \cdot (n/d)$ dla pewnych liczb całkowitych s i t. Zastosuj ćwiczenie 14 do liczb m/d i n/d zamiast m i n. Można przeprowadzić dłuższy dowód, korzystając z rozkładu na czynniki pierwsze.
(c) Niech $x = s \cdot a/d$.

17. (a) Sprawdź dla $l = 1$. W kroku indukcyjnym wystarczy pokazać, że jeśli $a = m = \text{FIB}\,(l+2)$ i $b = n = \text{FIB}\,(l+1)$, przy czym $l \geq 1$, to po pierwszym przebiegu pętli „dopóki" otrzymamy $a = \text{FIB}\,(l+1)$ i $b = \text{FIB}\,(l)$. Mianowicie z założenia indukcyjnego wyniknie wtedy, że po dokładnie l następnych przebiegach pętli algorytm zatrzyma się. Z definicji (a, b) w pętli „dopóki" wystarczy pokazać, że

$$\text{FIB}\,(l+2)\ \text{MOD}\ \text{FIB}\,(l+1) = \text{FIB}\,(l).$$

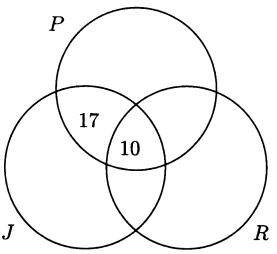
(c) Z ćwiczenia (a) wynika, że algorytm NWD wykonuje l przebiegów pętli. Z ćwiczenia (b) otrzymujemy

$$\log_2 (m+n) = \log_2 \text{FIB}\,(l+2) \leq l \quad \text{dla} \quad l \geq 3.$$

Paragraf 5.1

1. (a) 56. (c) 56. (e) 1.

3. Narysuj diagram Venna i znajduj liczby elementów kolejnych zbiorów zaczynając od środka. Liczbę 10 otrzymamy jako $|P \cap R \cap J|$; liczbę 17 otrzymamy wiedząc, że $27 = |J \cap P|$ itd. Odpowiedź: 15.



5. (a) 126. (b) 105.

7. (a) 0. (b) 840. (c) 2401.

9. (a) Liczba ta jest taka sama jak liczba sposobów wylosowania kolejno dziesięciu kart tak, by pierwsza z nich się nie powtórzyła. Stąd wynosi ona $52 \cdot (51)^9$.
(b) $52^{10} - 52(51)^9$.

11. (a) $13 \cdot \binom{4}{4} \cdot \binom{48}{1} = 624$. (b) 5108.

(c) $13 \cdot \binom{4}{3} \cdot \binom{12}{2} \cdot 4 \cdot 4 = 54912$. (d) 1 098 240.

13. (a) Jest to macierz wymiaru $n \times n$, w której na głównej przekątnej występują same zera, a na wszystkich pozostałych miejscach są jedynki.

15. (a) $n \cdot (n-1)^3$. (b) $n(n-1)(n-2)(n-3)$. (c) $n(n-1)(n-2)(n-2)$.

Paragraf 5.2

1. (a) $8/25 = 0,32$. (b) 0,20. (c) $9/25 = 0,36$.

3. (a) $\dfrac{5 \cdot 4 \cdot 3 \cdot 2}{5^4} = 0,192$. (b) $\dfrac{3^4}{5^4} = 0,1296$. (c) $2/5 = 0,40$.

5. Zauważ, że $\binom{7}{3} = 35$ jest liczbą sposobów wyboru trzech kul z naszej urny.
(a) 1/35. (b) 4/35. (c) 18/35. (d) 12/35.

7. (a) 0,2. (b) 0,9. (c) 0,6.

9. Tutaj $N = 2\ 598\ 960$.
(a) $624/N \approx 0,000240$. (b) $54\ 912/N \approx 0,0211$.
(c) $10\ 200/N \approx 0,00392$. (d) $123\ 552/N \approx 0,0475$.
(e) $1\ 098\ 240/N \approx 0,423$.

11. (a) 1/2. (b) 15/36. (c) 1/12.

13. Mamy

$$P(E_1 \cup E_2 \cup E_3) = P(E_1 \cup E_2) + P(E_3) - P((E_1 \cup E_2) \cap E_3).$$

Ale

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$$

oraz

$$\begin{aligned} P((E_1 \cup E_2) \cap E_3) &= P((E_1 \cap E_3) \cup (E_2 \cap E_3)) \\ &= P(E_1 \cap E_3) + P(E_2 \cap E_3) - P(E_1 \cap E_3 \cap E_2 \cap E_3). \end{aligned}$$

Podstaw uzyskane wyrażenia. Inny sposób rozwiązania polega na zastosowaniu równości

$$E_1 \cup E_2 \cup E_3 = (E_1 \backslash E_2) \cup (E_2 \backslash E_3) \cup (E_3 \backslash E_1) \cup (E_1 \cap E_2 \cap E_3).$$

15. (a) 1/64. (b) 6/64. (c) 15/64. (d) 20/64. (e) 22/64.

17. Niech Ω_n oznacza zbiór wszystkich ciągów długości n, złożonych z liter O i R, i niech E_n będzie zbiorem tych ciągów z Ω_n, w których O występuje parzystą liczbę razy. Wystarczy dowieść, że $|E_n| = 2^{n-1}$, co można zrobić przez indukcję.

19. (a) Jeśli Ω jest zbiorem wszystkich czteroelementowych podzbiorów zbioru S, to wszystkie wyniki są jednakowo prawdopodobne. Jeśli E_2 jest zdarzeniem „dokładnie dwie są parzyste", to

$$|E_2| = \binom{4}{2} \cdot \binom{4}{2} = 36.$$

Ponieważ

$$|\Omega| = \binom{8}{4} = 70, \text{ więc } P(E_2) = \frac{36}{70} \approx 0,514.$$

(b) 1/70. (c) 16/70. (d) 16/70. (e) 1/70.

21. (a) Przestrzeń zdarzeń elementarnych Ω składa się ze wszystkich uporządkowanych trójek (k,l,m), gdzie $k,l,m \in \{1,2,3\}$. Tak więc $|\Omega| = 3^3$. Trójki, dla których $k,l,m \in \{2,3\}$ odpowiadają sytuacji, gdy liczba 1 nie została wybrana. Zatem $P(\text{liczba 1 nie została wybrana}) = \frac{2^3}{3^3}$ i odpowiedź brzmi $1 - \frac{8}{27} \approx 0,704$.

(b) $1 - \dfrac{3^4}{4^4} \approx 0,684.$

(c) $1 - \left(\dfrac{n-1}{n}\right)^n.$

(d) $1 - 0,999999^{1000000} \approx 0,632120.$

Paragraf 5.3

1. 125.

3. W naszym zbiorze jest 466 takich liczb, zatem szukane prawdopodobieństwo wynosi 0,466. Pamiętaj, że $D_4 \cap D_6 = D_{12}$, a nie D_{24}. Zatem: $|D_4|+|D_5|+|D_6|-|D_4\cap D_5|-|D_4\cap D_6|-|D_5\cap D_6|+|D_4\cap D_5\cap D_6| = 250+200+166-50-83-33+16 = 466.$

5. (a) 0,142. (b) 0,09. (c) 0,78. (d) 0,208.

7. (a) $\binom{12+4-1}{4-1} = 455.$ (b) $\binom{4+4-1}{4-1} = 35.$

9. (a) $x^4 + 8x^3y + 24x^2y^2 + 32xy^3 + 16y^4.$
(c) $81x^4 + 108x^3 + 54x^2 + 12x + 1.$

11. $\binom{n}{r} = \dfrac{n!}{(n-r)!r!} = \dfrac{n!}{r!(n-r)!} = \dfrac{n!}{(n-(n-r))!(n-r)!} = \binom{n}{n-r}.$

13. (b) Dla każdego r, jest $\binom{n}{r}$ podzbiorów o liczności r, a zatem wszystkich podzbiorów jest $\sum_{r=0}^{n}\binom{n}{r}$.

(c) Jeśli tożsamość jest prawdziwa dla danego n, to

$$\sum_{r=0}^{n+1}\binom{n+1}{r} = 1 + \sum_{r=1}^{n}\binom{n+1}{r} + 1 = 1 + \sum_{r=1}^{n}\binom{n}{r-1} + \sum_{r=1}^{n}\binom{n}{r} + 1$$

$$= \sum_{r=1}^{n+1} \binom{n}{r-1} + \sum_{r=0}^{n} \binom{n}{r} = 2\sum_{r=0}^{n} \binom{n}{r} = 2 \cdot 2^n = 2^{n+1}.$$

15. (a) Obie strony są równe 15.

(b) Dla $n \geq m$ niech $p(n)$ oznacza zdanie „$\sum_{k=m}^{n} \binom{k}{m} = \binom{n+1}{m+1}$". Sprawdź, że zachodzi $p(m)$. Załóż, że dla pewnego $n \geq m$, $p(n)$ jest prawdą. Wtedy

$$\sum_{k=m}^{n+1} \binom{k}{m} = \left[\sum_{k=m}^{n} \binom{k}{m}\right] + \binom{n+1}{m}$$

$$= \binom{n+1}{m+1} + \binom{n+1}{m} \quad \text{(na mocy założenia indukcyjnego)}$$

$$= \binom{n+2}{m+1} \quad \text{(podaj powód)}.$$

(c) Zbiór $\mathcal{A}$ wszystkich $(m+1)$-elementowych podzbiorów zbioru $\{1, 2, \ldots, n+1\}$ jest rozłączną sumą $\bigcup_{k=m}^{n} \mathcal{A}_k$, gdzie $\mathcal{A}_k$ jest rodziną tych podzbiorów, których największym elementem jest $k+1$. Dowolny zbiór należący do $\mathcal{A}_k$ jest postaci: m-elementowy podzbiór zbioru $\{1, 2, \ldots, k\}$ z dołączoną doń liczbą $k+1$. Stąd $|\mathcal{A}| = \sum_{k=m}^{n} \binom{k}{m}$.

17. (a) Włóż 8 przedmiotów do jednego z 3 pudełek, następnie rozmieść pozostałych 6 w trzech pudełkach. Odpowiedź brzmi $3 \cdot \binom{8}{2} = 84$ sposoby.

(b) 36.

(c) Ponieważ $1 + 9 + 9 < 20$, to każda cyfra jest równa co najmniej 2. Wówczas

$$(d_1 - 2) + (d_2 - 2) + (d_3 - 2) = 20 - 6 = 14,$$

gdzie $2 \leq d_i \leq 9$ dla $i = 1, 2, 3$. Na mocy części (b) jest 36 takich liczb. Inny sposób otrzymania tej wartości polega na wykryciu prawidłowości rządzącej ciągiem:

$$299; \quad 398, 389; \quad 497, 488, 479; \quad 596, 587, 578, 569; \quad \ldots$$

Otrzymamy wtedy $1 + 2 + 3 + \ldots + 8 = \binom{9}{2} = 36$.

Paragraf 5.4

1. (a) $\dfrac{15!}{3!4!5!3!}$. (b) $\binom{15}{3}\binom{15}{4}\binom{15}{5}$.

3. (a) Tak jak w przykładzie 4, zliczaj podziały uporządkowane (A, B, C, D), gdzie $|A| = 5$, $|B| = 3$, $|C| = 2$ i $|D| = 3$. Odpowiedź: $\dfrac{13!}{5! \cdot 3! \cdot 2! \cdot 3!} = 720720$.

(b) $\frac{1}{2} \cdot \frac{13!}{4! \cdot 3! \cdot 3! \cdot 3!} = 600\,600.$

(c) Zliczaj podziały uporządkowane, gdzie $|A| = |B| = |C| = 3$ i $|D| = 4$, ale zauważ, że zmiana kolejności w ciągu A, B, C prowadzi do równoważnego zbioru komisji.

Odpowiedź: $\frac{1}{6} \cdot \frac{13!}{3! \cdot 3! \cdot 3! \cdot 4!} = 200200.$

5. (a) $3^{10} = 59049.$ (b) 252. (c) $\binom{10}{3} = 120.$

(d) 15360. (e) $\frac{10!}{3! \cdot 4! \cdot 3!} = 4200.$

(f) 55980. Należy zastosować zasadę włączeń i wyłączeń do zbiorów złożonych z ciągów, które nie zawierają, odpowiednio, zer, jedynek bądź dwójek.

7. (a) 625. (b) 505. (c) 250. (d) 303.

9. Jest $\frac{1}{2}\binom{2n}{n}$ takich podziałów nieuporządkowanych i $\binom{2n}{n}$ podziałów uporządkowanych.

11. (a) $10 \cdot \binom{8}{2} \cdot \binom{5}{2} \cdot \binom{2}{2} = 2800.$ Wybierz po prostu trzeciego zawodnika ich drużyny, a następnie skompletuj 3 drużyny z pozostałych 9 zawodników.

(b) $2/11 \approx 0,18.$

13. (a) Pomyśl o rozmieszczeniu 9 przedmiotów w 4 pudełkach, przy czym należy zacząć od włożenia jednego przedmiotu do pierwszego pudełka. $\binom{8+4-1}{4-1} = 165.$

(b) 56.

15. Piętnaście. Po prostu zliczaj. Nie znamy żadnej sprytnej sztuczki poza rozbiciem ich na typy: 4, 3-1, 2-2, 2-1-1 i 1-1-1-1. Liczby podziałów danych typów wynoszą, odpowiednio, 1, 4, 3, 6 i 1. To rozwiązuje problem, ponieważ, na mocy twierdzenia 1 z § 3.5, istnieje wzajemnie jednoznaczna odpowiedniość pomiędzy relacjami równoważności a podziałami.

Paragraf 5.5

1. (a) Zastosuj zasadę szufladkową do podziału $\{A_0, A_1, A_2\}$ zbioru S złożonego z czterech danych liczb całkowitych, gdzie $A_i = \{n \in S\colon n \equiv i (\mathrm{mod}\, 3)\}$. Można też zastosować drugą wersję zasady szufladkowej do funkcji $\mathrm{MOD}\, 3\colon S \to \mathbb{Z}_3$.

(b) Zastosuj zasadę szufladkową do funkcji $f\colon \{1, 2, \ldots, p+1\} \to \mathbb{Z}_p$ zdefiniowanej wzorem $f(m) = a_m \,\mathrm{MOD}\, p$.

3. (a) Mamy tutaj $|S| = 73$ i $73/8 > 9$, więc jedno z pudełek zawiera więcej niż 9 kulek.

5. Dla każdego czteroelementowego podzbioru B zbioru A niech $f(B)$ będzie sumą wszystkich liczb należących do B. Uzasadnij, dlaczego funkcja f przekształca zbiór wszystkich czteroelementowych podzbiorów zbioru A w zbiór $\{10, 11, 12, \ldots, 194\}$. Zauważ, że A ma $\binom{10}{4} = 210$ czteroelementowych podzbiorów. Zastosuj do funkcji f drugą wersję zasady szufladkowej.

7. Dla każdego dwuelementowego podzbioru T zbioru A niech $f(T)$ będzie sumą jego dwóch elementów. Wówczas funkcja f przekształca trzystuelementowy zbiór wszystkich dwuelementowych podzbiorów zbioru A w zbiór $\{3, 4, 5, \ldots, 299\}$.

9. Powtarzające się bloki są różnymi permutacjami wyrazów ciągu 142857.

11. (a) Przyjrzyj się sześciu następującym ciągom:

$$(n_1, n_2, n_3, n_4),\ (n_5, n_6, n_7, n_8), \ldots, (n_{21}, n_{22}, n_{23}, n_{24}).$$

(b) Wykorzystaj przykład 2(b) z § 4.2.

(c) Przyjrzyj się ośmiu następującym ciągom:

$$(n_1, n_2, n_3),\ (n_4, n_5, n_6), \ldots, (n_{22}, n_{23}, n_{24}).$$

(d) Przyjrzyj się pięciu następującym ciągom:
$(n_1, \ldots, n_5)$, $(n_6, \ldots, n_{10})$, $(n_{11}, \ldots, n_{15})$, $(n_{16}, \ldots, n_{20})$, $(n_{21}, \ldots, n_{24})$.

13. Jeśli $0 \in \mathrm{Im}(f)$, to któraś z liczb n_1, $n_1 + n_2$ lub $n_1 + n_2 + n_3$ jest podzielna przez 3. W przeciwnym przypadku funkcja f nie jest różnowartościowa i są trzy możliwości:

$$n_1 \equiv n_1 + n_2 (\mathrm{mod}\, 3); \quad n_1 \equiv n_1 + n_2 + n_3 (\mathrm{mod}\, 3);$$
$$n_1 + n_2 \equiv n_1 + n_2 + n_3 (\mathrm{mod}\, 3).$$

15. (a) 262 144. (b) 73 502. (c) 20 160. (d) 60.

17. (a) Wykaż, że w zbiorze S muszą się znaleźć oba elementy pewnej pary $(2k - 1, 2k)$.
(b) Każdą liczbę $m \in S$ przedstaw w postaci $m = 2^k \cdot n$, gdzie n jest liczbą nieparzystą; niech $f(m) = n$. Zobacz ćwiczenie 21 z § 4.1. Wówczas $f\colon S \to \{1, 3, 5, \ldots, 2n - 1\}$. Zastosuj drugą wersję zasady szufladkowej.
(c) Rozważ zbiór $S = \{2, 4, 6, \ldots, 2n\}$.

19. (a) Na mocy uwagi zamieszczonej na końcu dowodu uogólnionej zasady szufladkowej, średnia liczebność wynosi $2 \cdot 21/7 = 6$.

Paragraf 6.1

1. (a) *stv* lub *suv*; długość 2. (c) *uvwy*; długość 3.

3. (a) Prawdziwe.

5. (a) Prawdziwe.

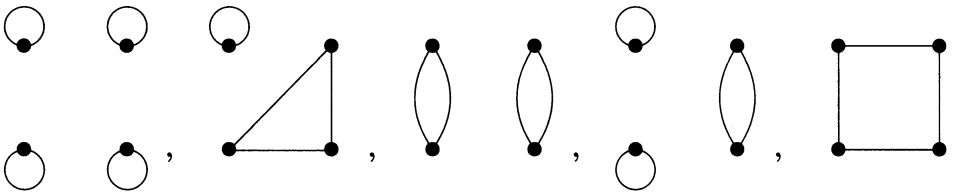
7. Zobacz rysunek 6.1(c).

9. (a)

e	a	b	c	d	e	f	g	h	k
$\gamma(e)$	$\{w,x\}$	$\{x,u\}$	$\{t,u\}$	$\{t,v\}$	$\{u,v\}$	$\{u,y\}$	$\{v,z\}$	$\{x,y\}$	$\{y,z\}$

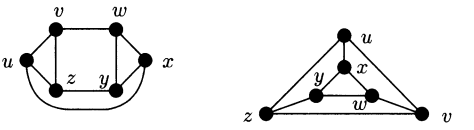
11. (a) Cykl $ebhkg$ i jego odwrócenie $gkhbe$.
(c) Cykle ecd, bhf i ich odwrócenia.

13. (a)



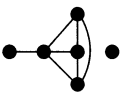
(c) Nie ma takich grafów, na podstawie twierdzenia 3, ponieważ $5 \cdot 3$ nie jest liczbą parzystą.

15. Grafy (a), (c) i (d) są regularne, a graf (b) nie jest. Grafy (a) i (c) mają cykle długości 3, a graf (d) nie ma. Możesz też policzyć krawędzie. Grafy (a) i (c) są izomorficzne; etykiety pokazują izomorfizm między grafami (a) i (c).

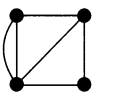


17. (a) $\binom{8}{5} = 56$. (b) 37. (c) $8 \cdot 7 + 8 \cdot 7 \cdot 6 + 8 \cdot 7 \cdot 6 \cdot 6 = 2408$.

19. (a)



(e)



(c) Zastosuj twierdzenie 3 (weź pod uwagę ćwiczenie 6(b)).
(g) K_4.

21. Załóż, że graf nie ma pętli i krawędzi wielokrotnych. Rozważ najdłuższą drogę $v_1 \ldots v_m$ o różnych wierzchołkach. Istnieje inna krawędź wychodząca z wierzchołka v_m. Dołącz ją do tej drogi, aby otrzymać drogę zamkniętą i skorzystaj ze stwierdzenia 1.

23. Wykorzystaj równość $|V(G)| = D_0(G) + D_1(G) + D_2(G) + \ldots$ i twierdzenie 3.

Paragraf 6.2

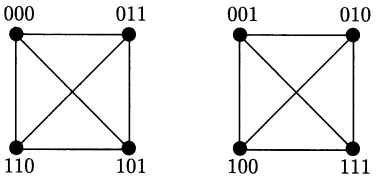
1. Tylko graf na rysunku 6.16(b) ma cykl Eulera. Aby go znaleźć, zrób ćwiczenie 2.

3. Nie załamie się aż do momentu drugiego przejścia przez inny wierzchołek stopnia 3, mianowicie t.

5. (a) Jest nią droga $v_3v_1v_2v_3v_6v_2v_4v_6v_5v_1v_4v_5v_3v_4$.

7. Nie. Krawędzie i naroża tworzą graf o ośmiu wierzchołkach, z których każdy ma stopień 3. Nie ma więc drogi Eulera. Por. rysunek 6.38(c) w § 6.5.

9. Zbiór $\{0,1\}^3$ składa się z trójek zer i jedynek, które możemy traktować jako ciągi zerojedynkowe długości 3. Tym grafem jest wtedy



(a) 2. (b) Wszystkie wierzchołki mają stopień 3. (c) Nie.

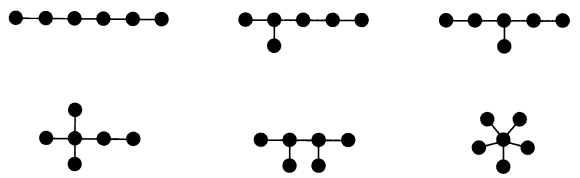
11. (a) Połącz wierzchołki stopnia nieparzystego parami za pomocą k nowych krawędzi. Nowy graf ma cykl Eulera, na podstawie twierdzenia 2. Nowe krawędzie nie pojawiają się obok siebie w tym cyklu i dzielą one cykl na k dróg prostych grafu G.

(c) Powtórz ten sam dowód. To znaczy, dodaj na przykład krawędzie $\{v_2, v_3\}$ oraz $\{v_5, v_6\}$, utwórz cykl Eulera, a następnie usuń te dwie nowe krawędzie.

13. (a) Nie jest możliwe takie przejście. Utwórz graf w następujący sposób. Umieść wierzchołek w każdym pokoju i jeden wierzchołek na zewnątrz domu. Przez każde drzwi narysuj krawędź łączącą wierzchołki znajdujące się w obszarach po obu stronach drzwi. Otrzymany graf ma dwa wierzchołki stopnia 4, trzy wierzchołki stopnia 5 i jeden wierzchołek stopnia 9. Zastosuj wniosek z twierdzenia 1.

Paragraf 6.3

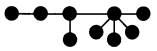
1.



3. (a) 4. (c) $4 + 2 \cdot 2 = 8$. Narysuj je.

(e) $8 \cdot 8 = 64$. Każde drzewo spinające można traktować jako parę drzew spinających, jedno dla górnej połowy i drugie dla dolnej połowy tego grafu.

5. (a) Ponieważ $2n - 2$ jest sumą stopni wierzchołków, musimy mieć $2n - 2 = 4 + 4 + 3 + 2 + 1 \cdot (n - 4)$. Rozwiąż to równanie z niewiadomą n.

7. (a) 

(b) Udowodnij przez indukcję względem n. Przypadki $n = 1, 2, 3$ są łatwe. Załóżmy, że wynik jest prawdziwy dla pewnego $n \geq 4$. Przypuśćmy, że

$$d_1 + \ldots + d_{n+1} = 2(n+1) - 2.$$

Co najmniej jedna z liczb d_k jest równa 1, powiedzmy d_{n+1}. Co najmniej jedna z liczb d_k przekracza 1, powiedzmy d_1. Definiujemy $d_1^* = d_1 - 1$ oraz $d_k^* = d_k$ dla $2 \leq k \leq n$. Wtedy

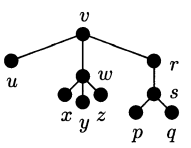
$$d_1^* + \ldots + d_n^* = 2n - 2$$

i z założenia indukcyjnego istnieje drzewo o n wierzchołkach, którego wierzchołki mają stopnie $d_1^*, \ldots, d_n^*$. Dołącz liść do wierzchołka stopnia d_1^*, aby otrzymać drzewo o $n + 1$ wierzchołkach. Nowy wierzchołek ma stopień $1 = d_{n+1}$, a wierzchołek stopnia d_1^* ma teraz stopień $d_1^* + 1 = d_1$.

9. (a) Przypuśćmy, że te składowe mają $n_1, n_2, \ldots, n_m$ wierzchołków, a więc w sumie $n_1 + n_2 + \ldots + n_m = n$. Zastosuj twierdzenie 3 do każdej składowej. Ile jest w sumie krawędzi?

11. Na podstawie lematu 1 do twierdzenia 3 musi ono być nieskończone.

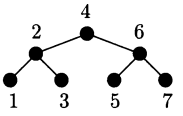
Paragraf 6.4

1. (a) 

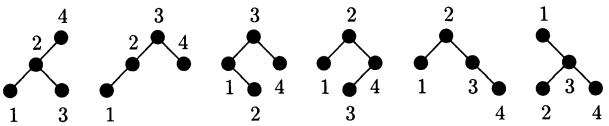
(c) 3.

3. (a) Drzewa z wyróżnionym korzeniem z rysunków 6.30(b) i 6.30(c) mają wysokość 2; drzewo z rysunku 6.30(d) ma wysokość 3.

5. (a) Jest ich siedem.
(c) Jest dokładnie jedno pełne drzewo binarne.
(d) 21.

7. (a) 

(c) Są następujące możliwości:



9. (a)



11. Z przykładu 5 mamy $(m-1)p = m^h - 1$ oraz $m^h = t$.

13. Istnieje 2^k słów długości k.

15. (a) Przesuń albo rekord Lyona, albo rekord Rossa do wierzchołka Rose'a i usuń utworzony pusty liść.

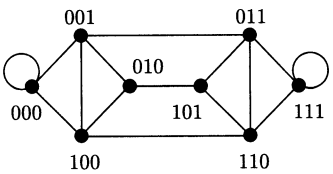
Paragraf 6.5

1. (a) W ciągu wierzchołków dla cyklu Hamiltona (jeśli taki istnieje), wierzchołek w musiałby zarówno poprzedzać wierzchołek v, jak i następować po nim. To znaczy, że wierzchołek w byłby przechodzony dwukrotnie.

3. (a) Tak. Spróbuj na przykład $v_1v_2v_6v_5v_4v_3v_1$.
(c) Nie.

5. (a) $2(n!)^2$. Zauważ, że początkowy wierzchołek może być albo w zbiorze V_1, albo w zbiorze V_2.
(c) Liczby m i n są parzyste lub m jest liczbą nieparzystą i $n = 2$ lub $m = n = 1$.

7. Oto ten graf



Jeden z możliwych cykli Hamiltona ma ciąg wierzchołków

$$000, 001, 011, 111, 110, 101, 010, 100, 000$$

odpowiadający ustawieniu cyklicznemu 00011101. Chociaż istnieją cztery istotnie różne cykle Hamiltona w zbiorze $\{0,1\}^3$, to istnieją tylko dwa różne ustawienia cykliczne, 00011101 oraz 00010111, które są odwrotne do siebie.

9. Nie ma drogi Hamiltona, ponieważ graf nie jest spójny. Graf ten jest narysowany w odpowiedzi do ćwiczenia 9 w § 6.2.

11. (a) Graf K_n^+ ma n wierzchołków i o jedną krawędź więcej niż graf K_{n-1}, a więc ma on dokładnie $\frac{1}{2}(n-1)(n-2)+1$ krawędzi.

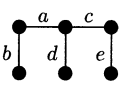
13. (a) 

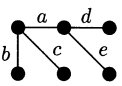
(c) Wybierz dwa wierzchołki u i v w grafie G. Jeśli nie są one połączone krawędzią w grafie G, to są połączone krawędzią w dopełnieniu tego grafu. Jeśli są połączone krawędzią w grafie G, to są w tej samej składowej G. Wybierz wierzchołek w w jakiejś innej składowej. Wtedy uwv jest drogą w dopełnieniu grafu. W każdym przypadku u i v są połączone drogą w dopełnieniu grafu.

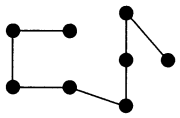
(e) Nie. Weź pod uwagę 

15. Dla danego grafu G_{n+1} weźmy podgraf H_0, gdzie $V(H_0)$ składa się ze wszystkich zerojedynkowych ciągów $(n+1)$-elementowych, w których na $(n+1)$-szym miejscu stoi 0 oraz $E(H_0)$ jest zbiorem wszystkich krawędzi grafu G_{n+1} łączących wierzchołki w $V(H_0)$. Zdefiniuj podobnie H_1. Pokaż, że H_0 i H_1 są izomorficzne z G_n, a zatem mają cykle Hamiltona. Użyj ich do zbudowania cyklu Hamiltona dla G_{n+1}. Dla $n=2$ zob. rysunek 6.31.

Paragraf 6.6

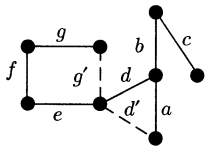
1. (a) 

(c) 

3. (a) 

Waga 1330.

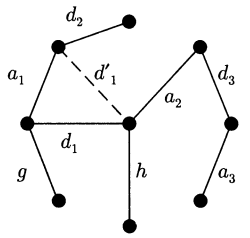
5. (a) i (c):



Można wybrać albo d, albo d' oraz albo g, albo g', a więc są cztery możliwe odpowiedzi w (a).

7. (a) $e_1, e_2, e_3, e_5, e_6, e_7, e_9$. (b) $e_7, e_5, e_2, e_1, e_3, e_6, e_9$.

9. (a)



Krawędzie a_1, a_2, a_3 mogą być wybrane w jakiejkolwiek kolejności. Tak samo krawędzie d_1, d_2, d_3. Krawędź d'_1 można wybrać zamiast krawędzi d_1. Waga jest równa 16.

11. 1687 mil.

13. Oto jeden z możliwych algorytmów:

Niech $E{:}\ = \emptyset$.

Wybierz w ze zbioru $V(G)$ i niech $V{:}\ = \{w\}$.

Dopóki $V \neq V(G)$, wykonuj

jeśli istnieje krawędź $\{u, v\} \in E(G)$ taka, że $u \in V$ oraz $v \in V(G)\backslash V$, to wybierz taką krawędź o najmniejszej wadze

dołącz krawędź $\{u, v\}$ do zbioru E i wierzchołek v do zbioru V

w przeciwnym przypadku wybierz $v \in V(G)\backslash V$ i dołącz wierzchołek v do zbioru V.

15. *Szkic*: Skorzystaj ze wskazówki. Popatrz na minimalne drzewa spinające S i T, gdzie $e \in T\backslash S$. Wtedy $S \cup \{e\}$ ma cykl zawierający krawędź $f \in S\backslash T$. Graf $(S \cup \{e\})\backslash\{f\}$ jest drzewem spinającym. $W(e) > W(f)$, co jest sprzeczne z wyborem e. Wyjaśnij dlaczego.

Paragraf 7.1

1. (a) Zastosuj indukcję względem m. Oczywiście, $2^0 \in S$ na podstawie warunku (P). Jeśli $2^m \in S$, to $2 \cdot 2^m = 2^{m+1} \in S$ na podstawie warunku (R).

(b) Zastosuj uogólnioną zasadę indukcji, gdzie $p(n) =$„n ma postać 2^m dla pewnej liczby $m \in \mathbb{N}$”.

3. (a) $S = \{(m, n){:}\ m \leq n\}$.

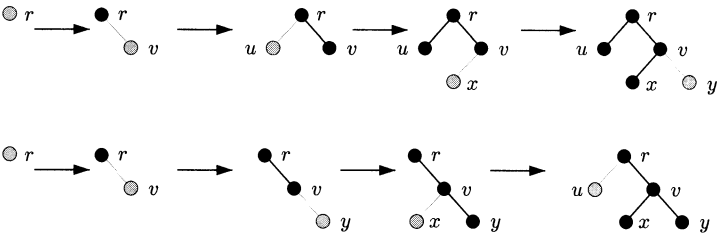
(c) Tak. Para $(0, n)$ należy do zbioru S tylko na podstawie warunku (P) albo dlatego, że para $(n-1, n-1) \in S$, a parę (m, n), gdzie $0 < m \leq n$ można otrzymać tylko z pary $(m-1, n)$.

5. (a) (P) λ należy do zbioru S;
(R) jeśli $w \in S$, to $aw \in S$ i $wb \in S$.

(c) $\lambda \in S$ na podstawie warunku (P), a więc a, aa i aab należą do zbioru S na podstawie warunku (R).

(d) Nasza definicja nie jest jednoznaczna.

7. Oto dwie możliwe konstrukcje:



9. Zastosuj uogólnioną zasadę indukcji do zdania $p(n)$ = „$n \equiv 0 \pmod 3$ lub $n \equiv 1 \pmod 3$". Oczywiście $p(1)$ jest zdaniem prawdziwym. Wystarczy pokazać, że $p(n) \Rightarrow p(3n)$ i $p(2n+1) \Rightarrow p(n)$. Pierwsza implikacja jest trywialna, ponieważ $3n \equiv 0 \pmod 3$ dla wszystkich n. Aby dowieść drugiej implikacji, pokaż, że prawdziwa jest kontrapozycja, to znaczy, że ze zdania $n \equiv 2 \pmod 3$ wynika zdanie $(2n+1) \equiv 2 \pmod 3$.

11. (a) Wykorzystaj warunki (P) i (R) do pokazania, że wyrazy ciągu 1, 2, 4, 8, 16, 5, 10, 3, 6 należą do zbioru S.

13. Wykorzystaj wskazówkę. Niech $p(w)$ = „$l'(w)$ jest liczbą liter w słowie w". Na podstawie (1) z przykładu 10(b) zdanie $p(w)$ jest prawdziwe, jeśli $w \in X = \{\lambda\} \cup \Sigma$. Przypuśćmy, że $w = uv$, gdzie $p(u)$ i $p(v)$ są zdaniami prawdziwymi i pokaż, że wtedy $p(w)$ jest zdaniem prawdziwym. Zatem $p(w)$ jest zdaniem prawdziwym dla każdego $w \in \Sigma^*$ na podstawie uogólnionej zasady indukcji.

15. (a) $(2,3)$, $(4,6)$ itd.

(b) Warunek (P) jest oczywisty, ponieważ 5 jest dzielnikiem liczby $0+0$. Jeśli chodzi o warunek (R), musisz sprawdzić, czy spełniona jest implikacja

„jeśli 5 dzieli $m+z$, to 5 dzieli $(m+2)+(n+3)$".

Inaczej, udowodnij, że każdy element zbioru S ma postać $(2k, 3k)$ dla $k \in \mathbb{N}$.

(c) Nie. Dlaczego?

17. (a) Oczywiście $A \subseteq \mathbb{N} \times \mathbb{N}$. Aby pokazać, że $\mathbb{N} \times \mathbb{N} \subseteq A$, zastosuj zwykłą zasadę indukcji matematycznej do zdań postaci

$p(k)$ = „jeśli $(m,n) \in \mathbb{N} \times \mathbb{N}$ i $m+n=k$, to $(m,n) \in A$".

(b) Niech $p(m,n)$ będzie funkcją zdaniową określoną na zbiorze $\mathbb{N} \times \mathbb{N}$. Aby pokazać, że $p(m,n)$ jest zdaniem prawdziwym dla wszystkich $(m,n) \in \mathbb{N} \times \mathbb{N}$, wystarczy pokazać:

(P) $p(0,0)$ jest zdaniem prawdziwym oraz

(I) jeśli $p(m,n)$ jest zdaniem prawdziwym, to $p(m+1,n)$ oraz $p(m,n+1)$ są zdaniami prawdziwymi.

19. (a) Dla $w \in \Sigma^*$ niech

$$p(w) = \text{„długość}(\overleftarrow{w}) = \text{długość}(w)\text{”}.$$

Zastosuj uogólnioną zasadę indukcji. Ponieważ $\overleftarrow{\lambda} = \lambda$, więc $p(\lambda)$ jest oczywiście zdaniem prawdziwym. Musisz pokazać, że jeśli $p(w)$ jest zdaniem prawdziwym, to również $p(wx)$ jest zdaniem prawdziwym:

$$\text{długość}(\overleftarrow{w}) = \text{długość}(w) \text{ implikuje długość}(\overleftarrow{wx}) = \text{długość}(wx).$$

Paragraf 7.2

1. (a) TEST(20;)
TEST(10;)
TEST(5;)
b: =fałsz; m: $= -\infty$
b: =fałsz; m: $= -\infty + 1 = -\infty$
b: =fałsz; m: $= -\infty + 1 = -\infty$

3. (a) $((x + y) + z)$ lub $(x + (y + z))$.
(c) $((xy)z)$ lub $(x(yz))$.

5. (a) Na podstawie warunku (P) x, y i 2 są wyrażeniami poprawnie zbudowanymi. Na podstawie warunku (R) wnioskujemy, że (x^2) i (y^2) są wyrażeniami poprawnie zbudowanymi. Zatem znów na podstawie warunku (R) $((x^2) + (y^2))$ jest wyrażeniem poprawnie zbudowanym.
(c) Na podstawie warunku (P) X i Y są wyrażeniami poprawnie zbudowanymi. Na podstawie warunku (R) $(X + Y)$ jest wyrażeniem poprawnie zbudowanym. Znów na podstawie warunku (R) $(X - Y)$ jest wyrażeniem poprawnie zbudowanym. W końcu $((X + Y) * (X - Y))$ jest wyrażeniem poprawnie zbudowanym.

7. Puste miejsca poniżej oznaczają, że algorytm nie kończy działania dla danej liczby n.

n	FOO	GOO	BOO	MOO	TOO	ZOO
8	8	40 320		4	4	8
9	9	362 880			4	8

9. Niech $p(k)$ będzie stwierdzeniem „algorytm ZOO daje w wyniku 2^k, jeśli tylko $2^k \leq n < 2^{k+1}$”. Dla $k = 0$ oznacza to, że „algorytm ZOO daje w wyniku 1, jeśli $n = 1$”, co jest oczywiste. Przyjmijmy, że $p(k)$ jest zdaniem prawdziwym i weźmy liczbę n taką że $2^{k+1} \leq n < 2^{k+2}$. Wtedy mamy $2^k \leq n \text{ DIV } 2 < 2^{k+1}$, więc algorytm ZOO($n$ DIV 2;) daje w wyniku $s = 2^k$. Instrukcja po słowach „w przeciwnym przypadku” w algorytmie ZOO(n;) daje $r = 2 * s = 2 * 2^k = 2^{k+1}$. Zatem algorytm ZOO daje w wyniku 2^{k+1}, jeśli tylko $2^{k+1} \leq n < 2^{k+2}$, a więc

$p(k+1)$ jest zdaniem prawdziwym. Tak więc z zasady indukcji wynika, że wszystkie zdania $p(k)$ są prawdziwe.

11. EUKLIDES$^+(108, 30;\ )$
EUKLIDES$^+(30, 18;\ )$
EUKLIDES$^+(18, 12;\ )$
EUKLIDES$^+(12, 6;\ )$
EUKLIDES$^+(6, 0;\ )$
$d:=6; s:=1; t:=0$
$d:=6; s:=0; t:=1-0\cdot(12 \text{ DIV } 6)=1$
$d:=6; s:=1; t:=0-1\cdot(18 \text{ DIV } 12)=-1$
$d:=6; s:=-1; t:=1-(-1)\cdot(30 \text{ DIV } 18)=2$
$d:=6; s:=2; t:=-1-2\cdot(108 \text{ DIV } 30)=-7.$

Oczywiście $108\cdot 2+30\cdot(-7)=6$.

13. (a) Ponieważ $\text{NWD}(m,n)=m$, gdy $n=0$, możemy przyjąć, że $n\neq 0$. Musimy sprawdzić, czy algorytm zatrzymuje się i czy jeśli liczba d' otrzymana w wyniku działania algorytmu EUKLIDES$(n, m \text{ MOD } n;\ )$ jest właściwa, to czy również liczba $d=d'$ otrzymana w wyniku działania algorytmu EUKLIDES$(m, n;\ )$ jest właściwa. Ostatnie stwierdzenie wynika z udowodnionej w § 4.6 równości $\text{NWD}(m,n)=\text{NWD}(n, m \text{ MOD } n)$.
Aby udowodnić, że algorytm kończy działanie, musimy sprawdzić, że druga zmienna n w algorytmie EUKLIDES$(\ , n;\ )$ jest w końcu równa 0. Jest to oczywiste, ponieważ w każdym kroku nowa liczba $n'=m \text{ MOD } n$ jest mniejsza od n, a więc w końcu wartości te muszą osiągnąć 0.

15. (a) p i q są wyrażeniami poprawnie zbudowanymi na podstawie warunku (P). Wyrażenie $p\vee q$ jest poprawnie zbudowane na podstawie warunku (R). Wyrażenie $\neg(p\vee q)$ jest poprawnie zbudowane na podstawie warunku (R).
(c) p, q i r są wyrażeniami poprawnie zbudowanymi na podstawie warunku (P). Wyrażenia $p\leftrightarrow q$ i $r\rightarrow p$ są poprawnie zbudowane na podstawie warunku (R). Wyrażenie $((r\rightarrow p)\vee q)$ jest poprawnie zbudowane na podstawie warunku (R), a więc również na podstawie (R) wyrażenie $((p\leftrightarrow q)\rightarrow((r\rightarrow p)\vee q))$ jest poprawnie zbudowane.

17. (a) $p, q\in\mathcal{F}$ na podstawie warunku (P). $p\vee q\in\mathcal{F}$ na podstawie warunku (R), gdzie $P=p$, $Q=q$. Zatem $(p\wedge(p\vee q))\in\mathcal{F}$ na podstawie (R), gdzie $P=p$, $Q=(p\vee q)$.
(b) Ponieważ wszystkie te p i P mogą się mylić, więc oznacz na przykład przez $r(P)$ funkcję zdaniową zmiennej $P\in\mathcal{F}$; następnie zastosuj uogólnioną zasadę indukcji. Tak więc musisz udowodnić, że wszystkie $r(P)$ są prawdziwe, gdzie

$$r(P)=\text{„jeśli } p, q \text{ są fałszywe, to } P \text{ jest fałszywe”}.$$

Aby udowodnić zdanie $r(P)$ dla wszystkich $P \in \mathcal{F}$, wystarczy pokazać, że:

(P) zdania $r(p)$ i $r(q)$ są prawdziwe;

(I) jeśli zdania $r(P)$ i $r(Q)$ są prawdziwe, to również prawdziwe są zdania $r((P \wedge Q))$ i $r((P \vee Q))$.

(c) Jeśli zdania p i q są fałszywe, to zdanie $(p \to q)$ jest prawdziwe, a więc zdanie $r((p \to q))$ jest fałszywe. Zatem zdanie $(p \to q)$ nie może być logicznie równoważne ze zdaniem należącym do zbioru $\mathcal{F}$ na podstawie ćwiczenia (b).

Ćwiczenie to daje negatywną odpowiedź na pytanie postawione w ćwiczeniu 17(c) w § 2.4: implikacja $p \to q$ nie może być zapisana w żaden sposób przy użyciu tylko p, q, $\wedge$ i $\vee$.

Paragraf 7.3

1. W porządku prefiksowym: $rxwvyzsutpq$.
W porządku postfiksowym: $vywzxtpuqsr$.

3. W porządku prefiksowym: $rtxvyzwupqs$.
W porządku postfiksowym: $vyzxwtpqsur$.

5. Porządek przechodzenia drzewa jest następujący:

$$r, w, v, w, x, y, x, z, x, w, r, u, t, u, s, p, s, q, s, u, \ r.$$

(a) W porządku prefiksowym: $rwvxyzutspq$. $L(w) = wvxyz$. $L(u) = utspq$.

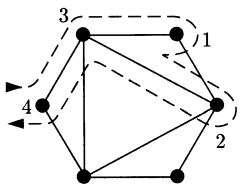
7. Porządek przechodzenia drzewa jest następujący:

$$u, x, w, v, w, y, w, x, r, z, r, t, r, x, u, s, p, s, q, s, u.$$

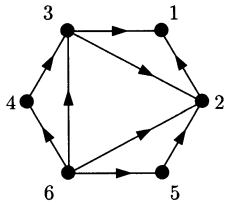
(a) W porządku postfiksowym: $vywztrxpqsu$.

(c) W porządku infiksowym: $vwyxzrtupsq$.

9. (a)

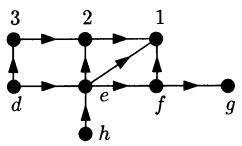


(c)

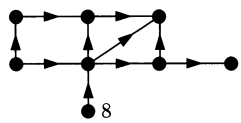


(b) Zacznij od a, wybierz jego następnik b, wybierz jego następnik c i oznacz c przez 1.
Wróć do b, wybierz jego następnik d i oznacz d przez 2.
Wróć do b i oznacz b przez 3.
Wróć do a i oznacz a przez 4.

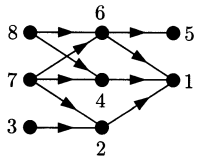
11. (a)



(c)



13. (a)



15. (a) Przypadki podstawowe to te, w których drzewo składa się tylko z korzenia r.

(c) Liczba potomków v jest dobrą miarą. Każde dziecko w wierzchołka v ma mniej potomków niż sam wierzchołek v. Miarą przypadku podstawowego jest 0.

17. Ponieważ ten graf skierowany jest acykliczny, może być co najwyżej jedna krawędź łącząca każdą z $n(n-1)/2$ par różnych wierzchołków.

Paragraf 7.4

1. Odwrotna notacja polska: $x42^\wedge - y * 23/+$.
Notacja polska: $+ * -x^\wedge 42y/23$.

3. (a) Notacja polska: $- * +ab - ab -^\wedge a2^\wedge b2$.
Notacja infiksowa: $(a+b)*(a-b)-((a^\wedge 2)-(b^\wedge 2))$.

5. (a) 20.

7. (a) $3x * 4 - 2^\wedge$.

(c) Odpowiedź zależy od tego, jak są zgrupowane wyrazy. Dla $(x-x^2)+(x^3-x^4)$ otrzymamy $xx2^\wedge - x3^\wedge x4^\wedge - +$.

9. (a) $abc**$ oraz $ab*c*$.

(c) Prawo łączności ma następującą postać: $abc** = ab*c*$. Prawo rozdzielności: $abc+* = ab*ac*+$.

11. (a) $p \to (q \vee (\neg p))$.

13. (a) Notacja infiksowa: $(p \wedge (p \to q)) \to q$.

15. (a) Obie dają $a/b+c$.

17. (a) (P) Stałe liczbowe i zmienne są wyrażeniami poprawnie zbudowanymi.

(R) Jeśli f i g są wyrażeniami poprawnie zbudowanymi, to również $+fg$, $-fg$, $*fg$, $/fg$ i $^\wedge fg$ są wyrażeniami poprawnie zbudowanymi.

19. (a) (P) Zmienne takie jak p, q, r są wyrażeniami poprawnie zbudowanymi.

(R) Jeśli P i Q są wyrażeniami poprawnie zbudowanymi, to również $PQ\vee$, $PQ\wedge$, $PQ\to$, $PQ\leftrightarrow$ i $P\neg$ są wyrażeniami poprawnie zbudowanymi.

(b) Uzasadnij po kolei, że $q\neg$, $pq\neg\wedge$ oraz $pq\neg\wedge\neg$ są wyrażeniami poprawnie zbudowanymi. Podobnie, $pq\neg\rightarrow$ jest wyrażeniem poprawnie zbudowanym. Zatem $pq\neg\wedge\neg pq\neg\rightarrow\vee$ jest wyrażeniem poprawnie zbudowanym.

(c) (P) Zmienne takie jak p, q, r są wyrażeniami poprawnie zbudowanymi.

(R) Jeśli P i Q są wyrażeniami poprawnie zbudowanymi, to również $\vee PQ$, $\wedge PQ$, $\rightarrow PQ$, $\leftrightarrow PQ$ i $\neg P$ są wyrażeniami poprawnie zbudowanymi.

Paragraf 7.5

1. (a) 35, 56, 70, 82.

3. Polecamy procedurę z przykładów 4 i 5.
(a) Waga = 84. (c) Waga = 244.

5. Wszystkie oprócz (b) są kodami prefiksowymi. W (b) 01 składa się z pierwszych dwóch cyfr 0111.

7. (b) 269.

9. (a) Wierzchołek oznaczony 0 ma tylko jedno dziecko, 00.
(b) Weź dowolny ciąg zaczynający się od 01.

11. (a) 484. (c) 373. (e) Całkowita liczba porównań wynosi 354.

13. (b) 221.

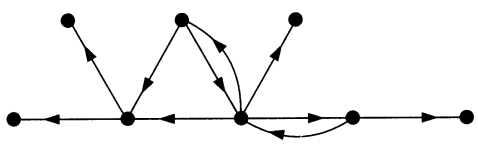
15. Na przykład w ćwiczeniu 1(a) pierwsze drzewo miało wagę równą 35. Liść o wadze **21** $= 12 + 9$ został zastąpiony poddrzewem o wagach 12 i 9 i waga całego drzewa wzrosła do 56, tzn. do $35 +$ **21**.

Paragraf 8.1

1. Ujściami są t i z. Jedynym źródłem jest u.

3. (a) $R(s) = \{s,t,u,w,x,y,z\} = R(t)$, $R(u) = \{w,x,y,z\}$, $R(w) = \{z\}$, $R(x) = \emptyset$, $R(y) = \{x,z\}$, $R(z) = \emptyset$.
(c) Droga sts jest cyklem, więc graf G nie jest acykliczny.

5. (a) Dodaj etykiety.



(c) s, u, x, z.

7. Użyj algorytmu NUMEROWANIE WIERZCHOŁKÓW. Jeden przykład etykietowania: $t = 1$, $z = 2$, $y = 3$, $w = 4$, $v = 5$, $x = 6$, $u = 7$.

9. (a) Jednym z przykładów jest: $rswtvwvuxyzvruyvsr$.

11. (a) Jeden taki graf skierowany jest przedstawiony na rysunku 8.3(b), gdzie $w = 00$, $x = 01$, $z = 11$ i $y = 10$.

(b) Jednym z możliwych ciągów jest ciąg 11101000 rozmieszczony na okręgu.

13. Pokaż, że graf $\widehat{G}$ jest też acykliczny. Zastosuj twierdzenie 2 do grafu $\widehat{G}$. Ujście dla grafu $\widehat{G}$ jest źródłem dla grafu G.

15. (a) Zobacz drugi dowód twierdzenia 2.

(b) Jeśli skończony acykliczny graf skierowany ma tylko jedno źródło, to istnieje droga do każdego wierzchołka z tego źródła.

17. (a) W dowodzie twierdzenia 1 (podanym w § 6.1) wybierz najkrótszą drogę składającą się z krawędzi danej drogi.

(b) Dla $u \neq v$ zastosuj twierdzenie 1 oraz wniosek 2. Dla $u = v$ zastosuj wniosek 1.

Paragraf 8.2

1.

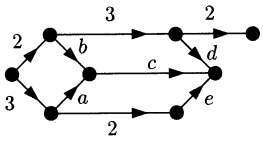
W^*	A	B	C	D
A	∞	1,0	1,4	1,2
B	0,4	∞	0,4	0,2
C	0,7	0,3	∞	0,5
D	0,8	0,5	0,2	∞

3.

W	m	q	r	s	w	x	y	z
m	∞	6	∞	2	∞	4	∞	∞
q	∞	∞	4	∞	4	∞	∞	∞
r	∞	∞	∞	∞	∞	∞	∞	3
s	∞	3	∞	∞	5	1	∞	∞
w	∞	∞	2	∞	∞	∞	2	5
x	∞	∞	∞	∞	3	∞	6	∞
y	∞	∞	∞	∞	∞	∞	∞	1
z	∞	∞	∞	∞	∞	∞	∞	∞

W^*	m	q	r	s	w	x	y	z
m	∞	5	8	2	6	3	8	9
q	∞	∞	4	∞	4	∞	6	7
r	∞	∞	∞	∞	∞	∞	∞	3
s	∞	3	6	∞	4	1	6	7
w	∞	∞	2	∞	∞	∞	2	3
x	∞	∞	5	∞	3	∞	5	6
y	∞	∞	∞	∞	∞	∞	∞	1
z	∞	∞	∞	∞	∞	∞	∞	∞

5. (a) Jeśli graf skierowany ma wszystkie strzałki prowadzące w prawo, jego wagi muszą być następujące:



gdzie $\min\{a+1, b\} = 4$ oraz $\min\{c+1, d, e\} = 3$.

7. (a) Droga $svxf$ jest jedną taką drogą. Jest też inna.

9. (a)

	s	u	v	w	x	y	f
A	0	2	7	5	5	11	15
L	0	2	9	7	5	11	15

(b) $S(v) = S(w) = 2$. $S(t) = 0$ dla wszystkich innych wierzchołków t.
(c) Droga $suxyf$ jest jedyną drogą krytyczną.
(d) Krawędzie na drodze krytycznej mają rezerwę czasową 0. Również $F(x, f) = 6$, $F(u, w) = F(v, f) = 3$ oraz $F(s, w) = F(w, y) = F(s, v) = F(v, y) = 2$.

11. (a)

	m	s	q	x	w	r	y	z
A	0	2	6	4	10	12	12	15
L	0	3	6	7	10	12	14	15

(c) Są dwie drogi krytyczne. Znajdź je.

13. (a) Są dwie drogi krytyczne: $suwxyf$ oraz $stwxyf$.
(b) 2.
(c) Są dwie krawędzie z rezerwą czasową 2.

15. Ściągnij krawędzie o wadze 0, aby ich końce pokryły się.

17. (a)

W	u	v	w	x	y
u	∞	1	∞	∞	∞
v	∞	∞	3	-2	∞
w	∞	∞	∞	∞	∞
x	4	∞	∞	∞	∞
y	∞	∞	∞	5	∞

W^*	u	v	w	x	y
u	3	1	4	-1	∞
v	2	3	3	-2	∞
w	∞	∞	∞	∞	∞
x	4	5	8	3	∞
y	9	10	13	5	∞

(c) Nie istniałyby w ogóle wagi minimalne dla dróg zawierających krawędzie u, v lub x, ponieważ wielokrotne chodzenie wzdłuż cyklu $uvxu$ za każdym razem zmniejszałoby wagę o 1.

19. (a) $FF(u, v) = A(v) - A(u) - W(u, v)$.
(c) Jest to rezerwa czasowa w v.

21. (a) $A(u) = M(s, u)$ = waga drogi maksymalnej z s do u. Jeśli istnieje krawędź (w, u), to droga maksymalna z s do w przedłużona o tę krawędź ma całkowitą wagę co najwyżej równą $M(s, u)$, tzn. że $A(w) + W(w, u) \leq A(u)$. Jeśli (w, u) jest krawędzią w drodze maksymalnej z s do u, to $A(w) + W(w, u) = A(u)$.

Paragraf 8.3

1. (a) $\mathbf{W}^* = \begin{bmatrix} \infty & 1 & 2 & 3 & 4 & 5 & 7 \\ \infty & \infty & \infty & 4 & 3 & 4 & 6 \\ \infty & \infty & \infty & 1 & 4 & 5 & 7 \\ \infty & \infty & \infty & \infty & 3 & 4 & 6 \\ \infty & \infty & \infty & \infty & \infty & 1 & 3 \\ \infty & \infty & \infty & \infty & \infty & \infty & 3 \\ \infty & \infty & \infty & \infty & \infty & \infty & \infty \end{bmatrix}$

3. (a)

L	$D(2)$	$D(3)$	$D(4)$	$D(5)$	$D(6)$	$D(7)$
$\emptyset$	1	2	∞	∞	∞	∞
$\{2\}$	1	2	5	4	∞	∞
$\{2,3\}$	1	2	3	4	9	∞
$\{2,3,4\}$	1	2	3	4	9	∞
$\{2,3,4,5\}$	1	2	3	4	5	7
$\{2,3,4,5,6\}$	1	2	3	4	5	7

Teraz już się nie zmienia

5. (a) $\mathbf{W}_2 = \begin{bmatrix} \infty & \infty & \infty & \infty & 1 & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 1 \\ \infty & \infty & \infty & 1 & \infty & 1 & \infty \\ \infty & \infty & 1 & \infty & 1 & \infty & \infty \\ 1 & \infty & \infty & 1 & 2 & \infty & \infty \\ \infty & \infty & 1 & \infty & \infty & \infty & 1 \\ \infty & 1 & \infty & \infty & \infty & 1 & 2 \end{bmatrix}$

$\mathbf{W}_4 = \begin{bmatrix} \infty & \infty & \infty & \infty & 1 & \infty & \infty \\ \infty & \infty & \infty & \infty & \infty & \infty & 1 \\ \infty & \infty & 2 & 1 & 2 & 1 & \infty \\ \infty & \infty & 1 & 2 & 1 & 2 & \infty \\ 1 & \infty & 2 & 1 & 2 & 3 & \infty \\ \infty & \infty & 1 & 2 & 3 & 2 & 1 \\ \infty & 1 & \infty & \infty & \infty & 1 & 2 \end{bmatrix}$, $\mathbf{W}_7 = \begin{bmatrix} 2 & 6 & 3 & 2 & 1 & 4 & 5 \\ 6 & 2 & 3 & 4 & 5 & 2 & 1 \\ 3 & 3 & 2 & 1 & 2 & 1 & 2 \\ 2 & 4 & 1 & 2 & 1 & 2 & 3 \\ 1 & 5 & 2 & 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 2 & 3 & 2 & 1 \\ 5 & 1 & 2 & 3 & 4 & 1 & 2 \end{bmatrix}$.

7. (a) $\mathbf{W}^* = \begin{bmatrix} \infty & 8 & 7 & 5 & 2 \\ \infty & \infty & \infty & \infty & \infty \\ \infty & 1 & \infty & \infty & \infty \\ \infty & 3 & 2 & \infty & \infty \\ \infty & 6 & 5 & 3 & \infty \end{bmatrix}$.

9. (a) $\mathbf{D}_0 = \mathbf{D}_1 = [-\infty \quad 1 \quad 2 \quad -\infty \quad -\infty \quad -\infty \quad -\infty]$,

$\mathbf{D}_2 = [-\infty \quad 1 \quad 2 \quad 5 \quad 4 \quad -\infty \quad -\infty]$,

$\mathbf{D}_3 = [-\infty \quad 1 \quad 2 \quad 5 \quad 7 \quad 9 \quad -\infty]$,

$\mathbf{D}_4 = [-\infty \quad 1 \quad 2 \quad 5 \quad 8 \quad 13 \quad -\infty]$,

$\mathbf{D}_5 = [-\infty \quad 1 \quad 2 \quad 5 \quad 8 \quad 13 \quad 11]$,

$\mathbf{D}_6 = [-\infty \quad 1 \quad 2 \quad 5 \quad 8 \quad 13 \quad 16]$.

11. (a) Algorytm mógłby dawać

L	$D(2)$	$D(3)$	$D(4)$
$\emptyset$	5	6	$-\infty$
$\{3\}$	5	6	9

Dalej bez zmian

gdzie $M(1,3) = 9$ i $M(1,4) = 12$.

(c) Oba algorytmy nadal nie dawałyby poprawnych wartości $M(1,4)$.

Paragraf 8.4

1. (a)

$$\mathbf{P}_0 = \begin{bmatrix} 0 & 2 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 5 & 6 & 0 \\ 0 & 0 & 0 & 0 & 5 & 6 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{P}_{\text{końcowa}} = \begin{bmatrix} 0 & 2 & 3 & 3 & 2 & 2 & 2 \\ 0 & 0 & 0 & 4 & 5 & 5 & 5 \\ 0 & 0 & 0 & 4 & 4 & 4 & 4 \\ 0 & 0 & 0 & 0 & 5 & 5 & 5 \\ 0 & 0 & 0 & 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

3. (a)

k	$D(2)$	$D(3)$	$D(4)$	$D(5)$	$P(2)$	$P(3)$	$P(4)$	$P(5)$
1	2	1*	7	∞	1	1	1	0
3	2*	1	7	4	1	1	1	3
2	2	1	6	4*	1	1	2	3
5	2	1	5*	4	1	1	5	3

Znak $*$ w kolumnach $D(k)$ oznacza moment, w którym wierzchołek k zostaje wybrany do zbioru L i wartość $D(k)$ zostaje ustalona.

5. (a)

$$\mathbf{W}_0 = \mathbf{W}_1 = \begin{bmatrix} \infty & 1 & \infty & 7 & \infty \\ \infty & \infty & 4 & 2 & \infty \\ \infty & \infty & \infty & \infty & 3 \\ \infty & \infty & 1 & \infty & 5 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{P}_0 = \mathbf{P}_1 = \begin{bmatrix} 0 & 2 & 0 & 4 & 0 \\ 0 & 0 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 3 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{W}_2 = \begin{bmatrix} \infty & 1 & 5 & 3 & \infty \\ \infty & \infty & 4 & 2 & \infty \\ \infty & \infty & \infty & \infty & 3 \\ \infty & \infty & 1 & \infty & 5 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 0 & 2 & 2 & 2 & 0 \\ 0 & 0 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 3 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

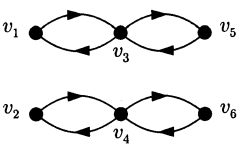
$$\mathbf{W}_3 = \begin{bmatrix} \infty & 1 & 5 & 3 & 8 \\ \infty & \infty & 4 & 2 & 7 \\ \infty & \infty & \infty & \infty & 3 \\ \infty & \infty & 1 & \infty & 4 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{P}_3 = \begin{bmatrix} 0 & 2 & 2 & 2 & 2 \\ 0 & 0 & 3 & 4 & 3 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 3 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{W}_4 = \begin{bmatrix} \infty & 1 & 4 & 3 & 7 \\ \infty & \infty & 3 & 2 & 6 \\ \infty & \infty & \infty & \infty & 3 \\ \infty & \infty & 1 & \infty & 4 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{P}_4 = \begin{bmatrix} 0 & 2 & 2 & 2 & 2 \\ 0 & 0 & 4 & 4 & 4 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 3 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

A także $\mathbf{W}_5 = \mathbf{W}^* = \mathbf{W}_4$ oraz $\mathbf{P}_5 = \mathbf{P}^* = \mathbf{P}_4$.

7. (a) Utwórz macierz wierszową $\mathbf{P}$, gdzie na początku $\mathbf{P}[j] = 1$, jeśli istnieje krawędź od wierzchołka v_1 do wierzchołka v_j oraz $\mathbf{P}[j] = 0$ w przeciwnym przypadku. Dodaj linię
zastąp $\mathbf{P}[j]$ przez k.
(b) Część tego ćwiczenia jest rozwiązana w przykładzie 3 w § 8.3.

9. (a)



(c) $\mathbf{M}_R = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$.

11. (a) Zapoczątkowanie macierzy $\mathbf{P}$ zajmuje czas rzędu $O(n)$, a krok, w którym dokonujemy podstawienia, wymaga stałego czasu podczas każdego przebiegu pętli.
(b) $O(n^3)$. Dlaczego?

Paragraf 9.1

1. (a) $\dfrac{\binom{3}{2}}{\binom{11}{2}} = \dfrac{3}{55}$. (b) $\dfrac{28}{55}$. (c) $\dfrac{3 \cdot 8}{\binom{11}{2}} = \dfrac{24}{55}$.

3. (a) $P(B_0) = P(R_0) = P(E) = \dfrac{1}{2}$ i $P(B_0 \cap R_0) = P(B_0 \cap E) = P(R_0 \cap E) = \dfrac{1}{4}$.

5. Zdarzenia $\{S, L\}$ są zależne, gdyż $P(S \cap L) = \dfrac{6}{36}$, podczas gdy $P(S) \cdot P(L) = \dfrac{15}{36} \cdot \dfrac{15}{36}$.
Zdarzenia $\{S, E\}$ są zależne, gdyż $P(S \cap E) = \dfrac{3}{36}$, podczas gdy $P(S) \cdot P(E) = \dfrac{15}{36} \cdot \dfrac{6}{36}$.

Zdarzenia $\{L, E\}$ są zależne, gdyż $P(L|E) = 0 \neq P(L)$. Analogicznie dla pary $\{L, G\}$.

7. Nie. $P(B|A) = \frac{1}{2}$, podczas gdy $P(B) = \frac{\binom{4}{2}}{2^4} = \frac{3}{8}$.

9. (a) 0,25. (b) 0,7. (c) Nie, $P(A|B) = 0,25 \neq P(A)$.

11. (a) $\frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \approx 0,00018$.
(b) $\approx 0,00048$.
(c) $1 - \frac{48}{52} \cdot \frac{47}{51} \cdot \frac{46}{50} \approx 0,217$.

13. (a) $P(B) = \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{1}{2} = \frac{47}{90}$.
(b) $P(U_1|B) = \frac{20}{47}$ $P(U_2|B) = \frac{12}{47}$, $P(U_3|B) = \frac{15}{47}$.
(c) $P(B \cap U_1) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$.

15. (a) $\frac{5}{9}$.

17. (a) $P(Z) = P(H) \cdot P(Z|H) + P(F) \cdot P(Z|F) + P(G) \cdot P(Z|G) = 0,043$.
(b) $P(H|Z) = \frac{10}{43} \approx 0,23$ itd.

19. (a) $P(D) = P(N^c \cap D) + P(N \cap D) = 0,0041$, a więc

$$P(N^c|D) = \frac{P(N^c \cap D)}{P(D)} = \frac{0,004}{0,0041} \approx 0,9756.$$

(b) $P((N^c \cap D) \cup (N \cap D^c)) = 0,9599$. Tyle wynosi prawdopodobieństwo, że wynik testu potwierdza stan badanego.
(c) $P(D|N^c) = \frac{P(D \cap N^c)}{P(N^c)} = \frac{0,004}{0,044} \approx 0,091$. Zatem prawdopodobieństwo, że ma się daną chorobę pod warunkiem, że wynik testu jest dodatni, jest mniejsze niż 0,10. Następująca tabelka może dopomóc w wyjaśnieniu sytuacji:

	D [choroba]	D^c [brak choroby]
N^c [wynik dodatni]	0,004	0,04
N^c [wynik ujemny]	0,0001	0,9559

21. (a) $1 - (1 - q)^n$. Zakładamy, że części psują się niezależnie od siebie.
(b) $1 - (0,99)^{100} \approx 0,634$.
(c) $1 - (0,999)^{100} \approx 0,952$.

23. (a) $\frac{\binom{18}{5}}{\binom{20}{5}} = \frac{15 \cdot 14}{20 \cdot 19} = \binom{21}{38} \approx 0,55$. (b) $\frac{9}{38}$.

25. Nie. Na przykład, rzuć trzy razy symetryczną monetą i niech A_k =„w k–tym rzucie wypadł orzeł”. Wtedy zdarzenia A_1, A_2, A_3 są wzajemnie

niezależne, ale zdarzenia $A_1 \cap A_2$ i $A_1 \cap A_3$ nie są niezależne. Aby to zobaczyć, wykaż, że $P(A_1 \cap A_3 | A_1 \cap A_2) \neq P(A_1 \cap A_3)$.

27. (a) Nieprawda. Jeśli byłaby to prawda, to z faktu, że zdarzenia A i B są niezależne oraz zdarzenia B i A są niezależne, wynikałoby, że A jest niezależne od A, co jest ogólnie nieprawdą na mocy części (b).
(b) Nieprawda.
(c) Kompletna nieprawda, chyba, że $P(A) = 0$ lub $P(B) = 0$.

29. (a) $$\frac{P^*(E)}{P^*(F)} = \frac{P(E|S)}{P(F|S)} = \frac{\dfrac{P(E \cap S)}{P(S)}}{\dfrac{P(F \cap S)}{P(S)}} = \frac{P(E \cap S)}{P(F \cap S)}.$$

Paragraf 9.2

1. (a) $\{3, 4, 5, 6, \ldots, 18\}$.

3. (a) $\{0, 1, 2, 3, 4, 5\}$ i $\{1, 2, 3, 4, 5, 6\}$.
(c) $P(D \leq 1) = 4/9$, $P(M \leq 3) = 1/4$, $P(D \leq 1 \text{ i } M \leq 3) = 7/36$.

5. (a) $\{1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 15, 16, 18, 20, 24, 25, 30, 36\}$.
(c) $1/9$.

7. (a) $\{0, 1, 2, 3, 4\}$. (b) $5/16$.
(c)

k	0	1	2	3	4
$P(X+Y=k)$	1/16	1/8	5/16	1/4	1/4

9. (a) Przykładowe obliczenie: $$P(X=2) = \frac{\binom{5}{2} \cdot \binom{5}{2}}{\binom{10}{4}} = \frac{10 \cdot 10}{210} = \frac{20}{42}.$$

k	0	1	2	3	4
$P(X=k)$	1/42	10/42	20/42	10/42	1/42

(b) Przykładowe obliczenie: $$P(X=2) = \frac{\binom{5}{2} \cdot \binom{5}{5}}{\binom{10}{7}} = \frac{10}{120} = \frac{1}{12}.$$

11. (a) Nie. Suma wartości funkcji $f(x)$ musi być równa 1.

13. (a) $0 \leq P(X \leq y) \leq 1$, ponieważ $0 \leq P(E) \leq 1$ dla wszystkich zdarzeń E.

15. $f(k) = P(W = k) = \left(\frac{5}{6}\right)^{k-1} \cdot \frac{1}{6}$ dla $k = 1, 2, \ldots$ i $f(x) = 0$ dla wszystkich innych wartości x.

17. (a) $P\left(\left[\frac{1}{6}, \frac{5}{6}\right]\right) = \frac{2}{3}$. (b) $P\left(\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right)\right) = \frac{2}{3}$.

19. Niech $x_1, x_2, \ldots, x_m$ oraz $y_1, y_2, \ldots, y_n$ będą zbiorami wartości zmiennych X i Y.

$(I_2) \Rightarrow (I_3)$. Aby wykazać warunek (I_3), możemy założyć, że $x = x_i$ i $y = y_j$ dla pewnych indeksów i oraz j. Niech I będzie pewnym przedziałem, do którego należy x_i, ale nie należy żaden inny element zbioru wartości zmiennej X; niech J będzie pewnym przedziałem, do którego należy y_j, ale nie należy żaden inny element zbioru wartości zmiennej Y. Wówczas

$$\begin{aligned} P(X = x_i \text{ i } Y = y_j) &= P(X \in I \text{ i } Y \in J) \\ &= P(X \in I) \cdot P(Y \in J) = P(X = x_i) \cdot P(Y = y_j). \end{aligned}$$

$(I_2) \Rightarrow (I_3)$. Dla danych przedziałów I i J niech $A = \{i\colon x_i \in I\}$ oraz $B = \{j\colon y_j \in J\}$. Wówczas zdarzenie $\{X \in I \text{ i } Y \in J\}$ można przedstawić w postaci rozłącznej sumy zdarzeń w następujący sposób:

$$\bigcup_{i \in A} \bigcup_{j \in B} \{X = x_i \text{ i } Y = y_j\};$$

wykaż teraz, że $P(X \in I \text{ i } Y \in J) = P(X \in I) \cdot P(Y \in J)$.

Paragraf 9.3

1. (a) Dość oczywista odpowiedź brzmi: 2. Dla jej potwierdzenia wykorzystaj rozkład prawdopodobieństwa podany w odpowiedzi do ćwiczenia 9 z § 9.2 i wykonaj obliczenie $\sum_{k=0}^{4} k \cdot P(X = k) = 2$.

(b) $\sigma = \sqrt{\frac{2}{3}} \approx 0,82$.

3. (a) Średnie odchylenie wynosi $\sum_{k=1}^{6} |k - 3,5| \cdot \frac{1}{6} = 1,5 < 1,71 \approx \sigma$.

(b) Średnie odchylenie wynosi $\frac{1}{2} = \sigma$.

5. $\mu_X = \mu_Y = 5/4$ i $\mu_{X+Y} = 5/2$. Wykaż, że $E(X^2) = 9/4$; wykorzystaj ten rezultat do wykazania, że $V(X) = 11/16$, a stąd $\sigma_X = \frac{1}{4}\sqrt{11} \approx 0,83$. Te same wyniki uzyskuje się dla zmiennej Y. W końcu, na mocy twierdzenia 7, $V(X + Y) = V(X) + V(Y) = 11/8$, a więc $\sigma_{X+Y} = \sqrt{\frac{11}{8}} \approx 1,17$.

7. (a) 3/5. (b) 7/5. (c) 13/5. (d) 19/5.

9. Wykaż, że $E(X^4) = 49/5$; wykorzystaj ten rezultat do wykazania, że $V(X^2) = 76/25$. Wynika stąd, że odchylenie standardowe zmiennej X^2 wynosi $\frac{1}{5}\sqrt{76} \approx 1,74$.

11. $E(X) = 5/13$. Zauważ, że zmienna losowa X jest równa sumie $X_1 + X_2 + X_3 + X_4 + X_5$, gdzie $X_i = 1$, jeśli i–ta karta jest asem i $X_i = 0$ w przeciwnym przypadku. Ile wynosi $E(X_i)$?

13. (a) Niech W będzie zmienną losową czasu oczekiwania. Jeśli wyobrazimy sobie, że jednocześnie wyciągamy z urny wszystkie kulki, to staje się jasne, że to, czy niebieska kulka będzie pierwszą z wyciągniętych kulek jest jednakowo prawdopodobne jak to, że będzie ona wyciągnięta jako druga, itd. Znaczy to, że $P(W = k) = 1/5$ dla $k = 1, 2, 3, 4, 5$. (Te równości łatwo jest także otrzymać bezpośrednio. I tak przykład $P(W = 3) = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5}$). Zatem

$$E(W) = \frac{1}{5}(1 + 2 + 3 + 4 + 5) = 3.$$

(b) Odpowiedź brzmi: 5. Można ją otrzymać za pomocą „nonsensownego" argumentu, że w każdym losowaniu spodziewamy się otrzymać $1/5$ niebieskiej kulki.

15. (a) Zmienna Y także przyjmuje każdą z wartości $1, 2, \ldots, n$ z prawdopodobieństwem $1/n$.

(c) Na mocy twierdzenia 2

$$E(X) = \frac{1}{n} + \frac{2}{n} + \ldots + \frac{n}{n}.$$

Zatem $1 + 2 + \ldots + n = n \cdot E(X) = n \cdot \frac{1}{2}(n + 1)$.

17. Jeśli $\mu = E(X)$, to $E(X + c) = \mu + c$. Stąd

$$V(X+c) = \sum_x (x-\mu-c)^2 \cdot P(X+c = x) = \sum_x (x-c-\mu)^2 \cdot P(X = x-c).$$

Zastępując wszędzie $x - c$ przez y dostajemy $\sum_y (y-\mu)^2 \cdot P(X = y) = V(X)$. Rezultat, który właśnie wykazaliśmy, jest intuicyjnie oczywisty: zmienna $X + c$ przesuwa wszystkie wartości zmiennej X o stałą c, ale nie zmienia ich wzajemnego położenia względem siebie.

Ponieważ równość $V(cX) = c^2 \cdot V(X)$ jest oczywista dla $c = 0$, więc załóżmy, że $c \neq 0$. Ponieważ $E(cX) = c \cdot \mu$, to

$$\begin{aligned} V(cX) &= \sum_x (x - c\cdot\mu)^2 \cdot P(cX = x) = c^2 \sum_x \left(\frac{x}{c} - \mu\right)^2 \cdot P\left(X = \frac{x}{c}\right) \\ &= c^2 \sum_y (y - \mu)^2 \cdot P(X = y) = c^2 \cdot V(X). \end{aligned}$$

19. (a) $E(S) = n \cdot \mu$ i $\sigma_S = \sqrt{n} \cdot \sigma$.

(b) $E\left(\frac{1}{n}S\right) = \mu$, a odchylenie standardowe zmiennej $\frac{1}{n}S$ jest równe $\frac{1}{\sqrt{n}} \cdot \sigma$.

21. (a) Ponieważ wszystkie te zmienne losowe mają skończone zbiory wartości, to wystarczy wykazać, że

$$(1) \qquad P(X_1 + X_2 = x \text{ i } X_i = x_i \text{ dla } i = 3, \ldots, n) = \\ = P(X_1 + X_2 = x) \cdot \prod_{i=3}^{n} P(X_i = x_i)$$

dla dowolnych liczb rzeczywistych $x, x_3, \ldots, x_n$. Niech A będzie zbiorem wszystkich par (u, v) liczb rzeczywistych takich, że u należy do zbioru wartości zmiennej X_1, v należy do zbioru wartości zmiennej X_2 oraz $u + v = x$. Wtedy zdarzenie $\{X_1 + X_2 = x\}$ jest równe rozłącznej sumie $\bigcup_{(u,v)\in A}\{X_1 = u \text{ i } X_2 = v\}$; zatem

$$(2) \qquad P(X_1 + X_2 = x) = \sum_{(u,v)\in A} P(X_1 = u \text{ i } X_2 = v).$$

Analogicznie,

$$P(X_1 + X_2 = x \text{ i } X_i = x_i \text{ dla } i = 3, \ldots, n) = \\ = \sum_{(u,v)\in A} P(X_1 = u, X_2 = v \text{ i } X_i = x_i \text{ dla } i = 3, \ldots, n).$$

Aby wykazać równość (1) wykorzystaj niezależność zmiennych $X_1, X_2, \ldots, X_n$, a następnie równość (2).

(b) Zastosuj indukcję względem n oraz część (a).

Paragraf 9.4

1. Gdyby te wyniki nie były wzajemnie niezależne, to musielibyśmy użyć prawdopodobieństw warunkowych, aby otrzymać następujący wzór na prawdopodobieństwo ciągu (S,S,P,S,P):
P(na pierwszym miejscu S) ·
P(na drugim miejscu S | na pierwszym miejscu S) ·
P(na trzecim miejscu P | na pierwszym i drugim miejscu S) · itd.

3. (a) Szukana wartość oczekiwana wynosi $np = 10/3$, ponieważ $n = 10$ oraz $p = 1/3$.
(c) $1 - P$(co najwyżej 2 skuteczne uderzenia) $= 1 - F(2) \approx 1 - 0,299 = 0,701$ na podstawie tablicy 9.6.

5. (a) $(0,9)^{10} \approx 0,349$ lub $F(0) \approx 0,349$ z tablicy 9.6.

7. (a) $1 - \Phi(1) \approx 1 - 0,8413 = 0,1587$. (b) $\approx 0,0227$.

9. (a) $\mu = 600$. (b) $\sigma = 20$.
(c) Tak jak w przykładzie 9, jednym z takich przedziałów jest

$$(\mu - 2\sigma, \mu + 2\sigma] = (560, 640].$$

11. (a) $\mu = 500$ i $\sigma \approx 15,81$. Zatem $10 \approx 0,632 \cdot \sigma$, a więc

$$P(490 < X \leq 510) \approx \Phi(0,63) - \Phi(-0,63) \approx 0,47.$$

(b) $\approx 0,95$.

(c) $\approx \Phi(20) - \Phi(-20)$. Ta liczba jest bardzo, ale to bardzo bliska 1, gdyż równa się $0,9999\ldots$, gdzie pierwszych 88 cyfr stanowią dziewiątki. We wszystkich praktycznych zastosowaniach można przyjąć, że zajście zdarzenia

$$\{490\ 000 < X \leq 510\ 000\}$$

jest pewne.

13. (a) 13/12. $X = X_1 + X_2 + X_3$, gdzie $X_i = 1$, jeśli i-ty eksperyment kończy się sukcesem, a $X_i = 0$ w przeciwnym przypadku. Zatem

$$E(X) = E(X_1) + E(X_2) + E(X_3).$$

(b) Na mocy twierdzenia 7 z § 9.3, $V(X) = V(X_1) + V(X_2) + V(X_3) = 95/144$. Zatem $\sigma_X = \frac{1}{12}\sqrt{95} \approx 0,81$.

(c) Nie. Dlaczego?

15. (a) Ponieważ bądź y, bądź $-y$ jest liczbą nieujemną, to możemy założyć, że nieujemna jest, powiedzmy, liczba y. Wówczas

$$\begin{aligned} 1 &= \text{pole obszaru pod wykresem krzywej } \varphi \\ &= \text{pole obszaru pod wykresem } \varphi \text{ na lewo od } y \\ &\quad + \text{pole obszaru pod wykresem } \varphi \text{ na prawo od } y \\ &= \Phi(y) + \text{pole obszaru pod wykresem } \varphi \text{ na prawo od } y. \end{aligned}$$

Ponieważ wykres funkcji φ jest symetryczny względem osi y, to

$$\begin{aligned} &\text{pole obszaru pod wykresem } \varphi \text{ na prawo od } y \\ &\quad = \text{pole obszaru pod wykresem } \varphi \text{ na lewo od } -y \\ &\quad = \Phi(-y). \end{aligned}$$

Paragraf 10.1

1. (a) Ponieważ działania $\vee$ i $\wedge$ traktują 0 i 1 tak, jakby to były wartości logiczne, więc sprawdzenie praw od 1Ba do 5Bb sprowadza się we wszystkich przypadkach do sprawdzenia odpowiednich matryc logicznych. Sprawdź tyle, ile potrzebujesz, by stało się to jasne dla ciebie.

3. Jedno rozwiązanie polega na wybraniu zbioru $S = \{1,2,3,4,5\}$ i zdefiniowaniu izomorfizmu wzorem

$$\varphi(x_1, x_2, x_3, x_4, x_5) = \{i \in S\colon\ x_i = 1\}.$$

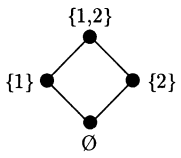
5. (a) Atomy są określone za pomocą czterech kolumn po prawej stronie następującej tablicy:

x	y	a	b	c	d
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	1

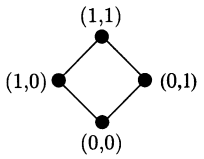
(c) Stosując oznaczenia z punktu (a), mamy $h = a \vee b \vee d$.

7. (a) Nie. Skończona algebra Boole'a ma 2^n elementów dla pewnego n.

9. (a)



(c)



11. (a) Jeśli $a \leq x$ lub $a \leq y$, to oczywiście $a \leq x \vee y$ na podstawie lematów 3(a) i 2(a). Przypuśćmy więc, że $a \leq x \vee y$. Wtedy $a = a \wedge (x \vee y) = (a \wedge x) \vee (a \wedge y)$. Jeden z elementów $a \wedge x$ i $a \wedge y$, na przykład $a \wedge x$, musi być różny od zera. Ale $a < a \wedge x \leq a$, więc $a \wedge x = a$, czyli $a \leq x$.

(c) $a \leq 1 = x \vee x'$, więc $a \leq x$ lub $a \leq x'$ na podstawie ćwiczenia (a). Z obu nierówności $a \leq x$ i $a \leq x'$ razem wynikałoby $a \leq x \wedge x' = 0$ na podstawie ćwiczenia (b), co prowadzi do sprzeczności.

13. $x \leq y \Leftrightarrow x \vee y = y \Leftrightarrow \varphi(x \vee y) = \varphi(y) \Leftrightarrow \varphi(x) \vee \varphi(y) = \varphi(y) \Leftrightarrow \varphi(x) \leq \varphi(y)$. Uzasadnij poszczególne kroki.

Paragraf 10.2

1. $x'y'z' \vee x'y'z \vee xyz'$.

3.

x	y	z	xy	$xy \vee z'$
0	0	0	0	1
0	0	1	0	0
0	1	0	0	1
0	1	1	0	0
1	0	0	0	1
1	0	1	0	0
1	1	0	1	1
1	1	1	1	1

(a) $xyz' \vee xyz$.
(c) $x'y'z' \vee x'yz' \vee xy'z' \vee xyz' \vee xyz$.

5. (a) $x_1x_2x_3'x_4 \vee x_1x_2x_3'x_4' \vee x_1'x_2x_3x_4'$.

7. (a) $xz \vee y'$.

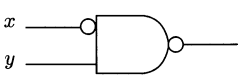
9. $y' \vee z$.

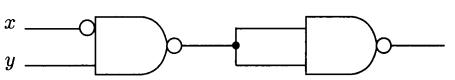
11. (a) Znajdź postać kanoniczną wyrażenia E'. Następnie znajdź $E = (E')'$ korzystając z praw De Morgana najpierw dla sum, a następnie dla iloczynów.
(b) $(x' \vee y')(x \vee y)$.

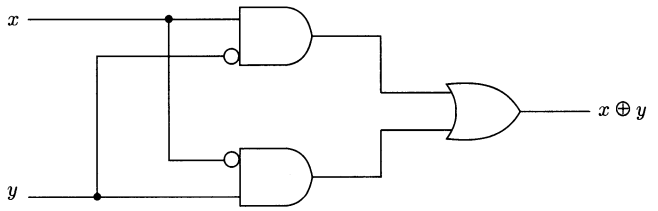
13. (a) Funkcja booleowska określona za pomocą wyrażenia $x'z \vee y'z$ przyjmuje wartość 1 dla trzech argumentów z $\mathbb{B}^3$. Funkcje booleowskie wyznaczone przez iloczyny symboli atomowych przyjmują wartość 1 dla 1, 2 lub 4 argumentów z $\mathbb{B}^3$ na podstawie ćwiczenia 12.

Paragraf 10.3

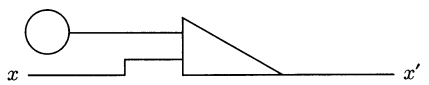
1. (a) xyz.

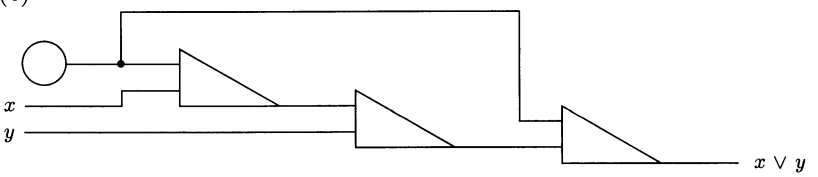
3. (a) 

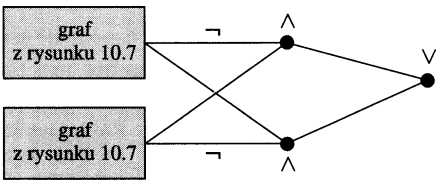
(b) 

5. (a) 

7. (a) $S = 1, C_O = 0$. (c) $S = 0, C_O = 1$.

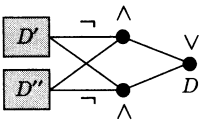
9. (a) 

(c) 

11. 

13. Wygodnie jest przyjąć, że ten fakt jest prawdziwy dla $n = 1$. Zastosujemy drugą zasadę indukcji matematycznej. Rzut oka na rysunek 10.7 pokazuje, że ten fakt jest również prawdziwy dla $n = 2$. Załóżmy więc, że jest on prawdziwy dla wszystkich j takich, że $1 \leq j < n$. Weźmy

liczbę k taką, że $2^{k-1} < n \leq 2^k$ i niech $n' = 2^{k-1}$ oraz $n'' = n - 2^{k-1}$. Z założenia indukcyjnego wynika, że istnieją grafy skierowane D' i D'' obliczające odpowiednio $x_1 \oplus x_2 \oplus \ldots \oplus x_{n'}$ i $x_{n'+1} \oplus \ldots \oplus x_n$. Graf D' ma $3(n'-1)$ wierzchołków $\wedge$ i $\vee$, a graf D'' ma $3(n''-1)$ wierzchołków $\wedge$ i $\vee$. Ponadto każda droga w grafach D' i D'' ma długość co najwyżej $2(k-1)$. Teraz tworzymy graf D tak jak na rysunku:

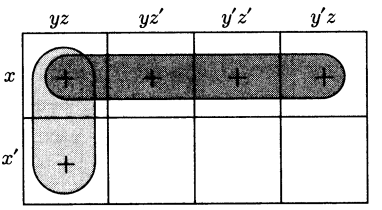


Paragraf 10.4

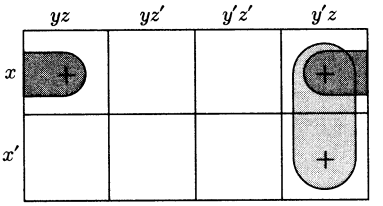
1. $xyz \vee xyz' \vee xy'z' \vee xy'z \vee x'yz \vee x'y'z = x \vee z$.

3. $xyz \vee xyz' \vee xy'z \vee x'y'z' \vee x'y'z = xz \vee xy \vee x'y' = xy \vee y'z \vee x'y'$.

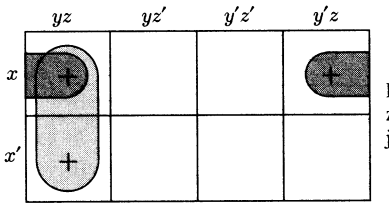
5. (a)



(c)

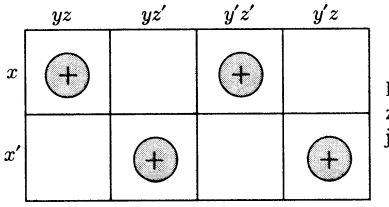


7. (a)



Każdy blok składający się z dwóch kwadratów jest istotny.

(c)

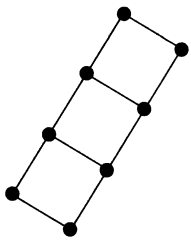


Każdy blok składający się z jednego kwadratu jest istotny.

9. (a) $z' \vee xy \vee x'y' \vee w'y$ lub $z' \vee xy \vee x'y' \vee w'x'$.
(c) $w'x'z' \vee w'xy' \vee wxy \vee wx'z \vee y'z'$.

Paragraf 11.1

1. (a) Oznacz punkty.



3. (a) h, o, p, q, r, z. (c) B i C. (e) f, z, p, nie istnieje.

5. (a) $a \vee b = \sup(a,b) = \max\{a,b\}$, $a \wedge b = \inf(a,b) = \min\{a,b\}$.
(c) 73. (e) $\sqrt{73}$.

7. (a) Przypuśćmy, że $\preceq$ jest częściowym porządkiem w zbiorze S i relacja $\succeq$ w S jest zdefiniowana w następujący sposób: $x \succeq y$ wtedy i tylko wtedy, gdy $y \preceq x$. Wtedy $x \preceq x$, a więc $x \succeq x$. Jeśli $x \succeq y$ i $y \succeq x$, to $y \preceq x$ i $x \preceq y$, a więc $x = y$. Jeśli $x \succeq y$ i $y \succeq z$, to $y \preceq x$ i $z \preceq y$, stąd $z \preceq x$, a więc $x \succeq z$. Zatem relacja $\succeq$ spełnia własności (Z), (AS) i (P).

(b) Oczywiście, $x \preceq x$, a więc relacja $\preceq$ ma własność (Z). Jeśli $x \preceq y$ i $y \preceq z$, to możliwe są cztery przypadki:

$$x = y = z, \quad x = y \prec z, \quad x \prec y = z \quad \text{oraz} \quad x \prec y \prec z.$$

Wykaż, że w każdym z tych przypadków $x \preceq z$, a więc spełniona jest własność (P). Jeśli $x \preceq y$ i $y \preceq z$, to potencjalnie mamy następujące przypadki:

$$x = y = x, \quad x = y \prec x, \quad x \prec y = x \quad \text{oraz} \quad x \prec y \prec x.$$

Tylko pierwszy z nich jest możliwy. Pozostałe trzy przeczą własności (AS) relacji $\prec$. Zatem relacja $\preceq$ ma własność (AS).

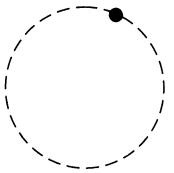
9. Ponieważ $w = w\lambda$, to $w \preceq w$ i spełniona jest własność (Z). Jeśli $w_1 \preceq w_2$ i $w_2 \preceq w_3$, to istnieją słowa u i v takie, że $w_2 = w_1 u$ i $w_3 = w_2 v$. Wtedy $w_3 = w_1 uv$, gdzie $uv \in \Sigma^*$, a więc $w_1 \preceq w_3$. Zatem spełniona jest własność (P).

11. Nie jest, jeśli tylko Σ ma więcej niż jeden element. Pokaż, że ta relacja nie jest antysymetryczna.

13. (a) Nie. Każdy skończony podzbiór zbioru $\mathbb{N}$ jest podzbiorem jakiegoś większego skończonego podzbioru $\mathbb{N}$.
(c) $\sup\{A,B\} = A \cup B$. Zauważ, że $A \cup B \in \mathcal{F}(\mathbb{N})$ dla wszystkich zbiorów $A, B \in \mathcal{F}(\mathbb{N})$.
(e) Tak.

15. (a) Jedynie relacja $\leq$. Relacja $<$ nie jest zwrotna, a relacja $\preceq$ nie jest antysymetryczna.

(c)



17. Zobacz rysunki 11.2 i 11.7 lub ćwiczenie 16, żeby poznać dwa rodzaje powodów, dla których może tak nie być.

19. (a) Wykaż, że element b spełnia definicję $\sup\{x, y, z\}$, tzn. $x \preceq b$, $y \preceq b$, $z \preceq b$ i, jeśli $x \preceq c$, $y \preceq c$, $z \preceq c$, to $b \preceq c$.
(b) Wykaż przez indukcję po n, że każdy n-elementowy podzbiór kraty ma kres górny.
(c) Wykorzystaj część (a) i przemienność działania $\vee$.

Paragraf 11.2

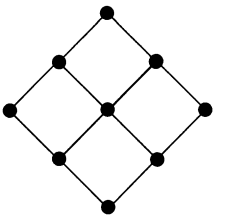
1. (a) Jednym z nich jest $\{\emptyset, \{1\}, \{1,4\}, \{1,4,3\}, \{1,4,3,5\}, \{1,4,3,5,2\}\}$.

3. (a) $501, 502, \ldots, 1000$.
(c) Tak. Pomyśl o liczbach pierwszych lub zob. ćwiczenie 17.

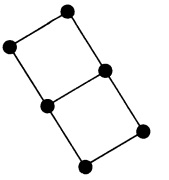
5. Tak. Jeśli $a \preceq b$, to $\sup(a,b) = b$ i $\inf(a,b) = a$.

7. (a) Przykładowo wykażemy przechodniość. Jeśli $f \preceq g$ i $g \preceq h$, to $f(t) \preceq g(t)$ i $g(t) \preceq h(t)$ w sensie relacji $\preceq$ w zbiorze S, dla wszystkich t ze zbioru T. Ponieważ relacja $\preceq$ w zbiorze S jest przechodnia, to $f(t) \preceq h(t)$ dla wszystkich t, a więc $f \preceq h$ w zbiorze $\text{FUN}(T, S)$.
(c) Mamy $f(t) \preceq f(t) \vee g(t) = h(t)$ dla wszystkich elementów t zbioru T. Stąd $f \preceq h$. Analogicznie, $g \preceq h$, a więc h jest ograniczeniem górnym zbioru $\{f, g\}$. Pokaż, że, jeśli $f \preceq k$ i $g \preceq k$, to $h \preceq k$. Stąd wyniknie, że h jest kresem górnym zbioru $\{f, g\}$.

9. (a) Oznacz punkty.



(c)



11. (a) (0,0), (0,1), (0,2), (1,0), (1,1), (1,2), (2,0), (2,1), (2,2).
(c) (3,0), (3,1), (3,2), (4,0), (4,1), (4,2).

13. (a) 000, 0010, 010, 10, 1000, 101, 11.

15. (a) w z to się słowa składa wypisz zdanie których porządku rosnącym standardowym .

17. Niech $a_1 \prec a_2 \prec \ldots \prec a_n$ będzie łańcuchem maksymalnym w zbiorze S. W S nie ma łańcucha postaci $b \prec a_1 \prec a_2 \prec \ldots \prec a_n$, a więc nie istnieje element b taki, że $b \prec a_1$. Znaczy to, że element a_1 jest minimalny. (Założenie o skończoności zbioru S jest istotne. W zbiorze liniowo uporządkowanym $(\mathbb{Z}, \leq)$ sam zbiór $\mathbb{Z}$ jest łańcuchem maksymalnym.

19. Antysymetria jest oczywista. Dla dowodu przechodniości rozważ szereg przypadków. Przypuśćmy, że

$$(s_1, \ldots, s_n) \prec (t_1, \ldots, t_n) \quad \text{oraz} \quad (t_1, \ldots, t_n) \prec (u_1, \ldots, u_n).$$

Jeśli $s_1 \prec_1 t_1$, to $s_1 \prec_1 t_1 \preceq_1 u_1$, a więc $(s_1, \ldots, s_n) \prec (u_1, \ldots, u_n)$. Jeśli $s_1 = t_1, \ldots, s_{r-1} = t_{r-1}$, $s_r \prec_r t_r$ i $t_1 = u_1, \ldots, t_{p-1} = u_{p-1}$, $t_p \prec_p u_p$ i $r < p$, to $s_1 = u_1, \ldots, s_{r-1} = u_{r-1}$ i $s_r \prec_r t_r = u_r$, a więc znowu $(s_1, \ldots, s_n) \prec (u_1, \ldots, u_n)$. Pozostałe przypadki analizuje się podobnie.

Paragraf 11.3

(a) $\mathbf{A} * \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$. Ponieważ nie jest prawdą, że $\mathbf{A} * \mathbf{A} \leq \mathbf{A}$, więc relacja R nie jest przechodnia.

(c) Nie jest przechodnia. Zauważ, że $\mathbf{A} * \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

3. (a) Macierzami relacji R i R^2 są

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{oraz} \quad \mathbf{A} * \mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

(c) Nie; porównaj macierze $\mathbf{A}$ i $\mathbf{A} * \mathbf{A}$ i zauważ, że nie jest spełniona nierówność $\mathbf{A} * \mathbf{A} \leq \mathbf{A}$.

(e) Tak.

5. (a) Macierzą relacji R^0 jest macierz jednostkowa. Macierzą relacji R jest oczywiście $\mathbf{A}$. Macierzą każdej z relacji R^n, dla $n \geq 2$, jest $\mathbf{A} * \mathbf{A}$. Należy tego dowieść przez indukcję.

(b) Relacja R jest zwrotna, ale nie jest ani symetryczna ani przechodnia.

7. (a) $\mathbf{A}_f = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ oraz $\mathbf{A}_g = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$.

(b) Będą one różne, ponieważ macierzą booleowską relacji $R_f R_g$ jest

$$\mathbf{A}_f * \mathbf{A}_g = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix};$$

to jest macierz booleowska relacji $R_{g \circ f}$, ale nie relacji $R_{f \circ g}$.

(c) Jedna tak, jedna nie.

9. (a) Relacja R_1 spełnia (PZ) i (S).

(c) Relacja R_3 spełnia (Z), (AS) i (P).

(e) Relacja R_5 spełnia tylko (S).

11. (a) Prawda. Dla każdego x, $(x,x) \in R_1 \cap R_2$, a więc $(x,x) \in R_1 R_2$.

(c) Fałsz. Weź relacje równoważności R_1 i R_2 w zbiorze $\{1,2,3\}$, o macierzach booleowskich

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{oraz} \quad \mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

13. Nie korzystaj z macierzy booleowskich; zbiory S, T, U mogą być nieskończone.

(a) Na mocy przykładu 2(a), $R_1 R_3 \cup R_1 R_4 \subseteq R_1(R_3 \cup R_4)$. Aby dowieść inkluzji odwrotnej, weź $(s,u) \in R_1(R_3 \cup R_4)$ i wykaż, że para (s,u) należy do $R_1 R_3$ lub do $R_1 R_4$.

(c) Wykaż, że $R_1(R_3 \cap R_4) \subseteq R_1 R_3 \cap R_1 R_4$. Równość nie musi zachodzić. Na przykład, rozważ relacje R_1, R_3, R_4 o macierzach booleowskich

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{A}_4 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

15. (Z) Relacja R jest zwrotna wtedy i tylko wtedy, gdy $(x,x) \in R$ dla każdego x, wtedy i tylko wtedy, gdy $\mathbf{A}[x,x] = 1$ dla każdego x.
(PZ) Analogicznie do rozumowania dla (Z); wstaw $(x,x) \notin R$ i $\mathbf{A}[x,x] = 0$.
(S) Wynika stąd, że $\mathbf{A}^T[x,y] = \mathbf{A}[y,x]$ dla każdych x, y.
(AS) Relacja R jest antysymetryczna wtedy i tylko wtedy, gdy $x = y$ ilekroć $(x,y) \in R$ i $(y,x) \in R$, tzn. ilekroć $\mathbf{A}[x,y] = \mathbf{A}^T[x,y] = 1$. Zatem relacja R jest antysymetryczna wtedy i tylko wtedy, gdy wszystkie wyrazy spoza głównej przekątnej macierzy $\mathbf{A} \wedge \mathbf{A}^T$ są równe 0.
(P) To wynika z twierdzenia 3 i punktu (a) tego zestawienia.

17. Dane macierze booleowskie $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ o wymiarach, odpowiednio, $m \times n$, $n \times p$ i $p \times q$ odpowiadają relacjom R_1, R_2, R_3, gdzie R_1 jest relacją na zbiorze $\{1,2,\ldots,m\} \times \{1,2,\ldots,n\}$ itd. Macierzami relacji

$(R_1R_2)R_3$ oraz $R_1(R_2R_3)$ są macierze $(\mathbf{A}_1 * \mathbf{A}_2) * \mathbf{A}_3$ i $\mathbf{A}_1 * (\mathbf{A}_2 * \mathbf{A}_3)$, na mocy czterokrotnego zastosowania twierdzenia 1.

19. (a) By wykazać, że relacja $R \cup E$ jest częściowym porządkiem, pokaż, że

(Z) $(x, x) \in R \cup E$ dla wszystkich $x \in S$,

(AS) $(x, y) \in R \cup E$ i $(y, x) \in R \cup E$ implikują, że $x = y$,

(P) $(x, y) \in R \cup E$ i $(y, z) \in R \cup E$ implikują, że $(x, z) \in R \cup E$.

Aby sprawdzić, że zachodzi własność (P), rozważ następujące przypadki:

$$(x, y) \in R \quad \text{i} \quad (y, z) \in R,$$
$$(x, y) \in R \quad \text{i} \quad (y, z) \in E \text{ (a więc } y = z),$$
$$(x, y) \in E \quad \text{i} \quad (y, z) \in R,$$
$$(x, y) \in E \quad \text{i} \quad (y, z) \in E.$$

Ostatnie dwa przypadki można zebrać razem, gdyż jeśli $(x, y) \in E$ oraz $(y, z) \in R \cup E$, to $(x, z) = (y, z) \in R \cup E$.

Paragraf 11.4

1. (a) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, (c) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, (e) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

3. $\{1, 2\}$, $\{3\}$.

5. (a) $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, (c) $\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$, (e) $\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$.

7. (a) $z(R)$ to zwykły porządek $\leq$.

(c) $zs(R)$ jest uniwersalną relacją w $\mathbb{P}$.

(e) Sama relacja R jest przechodnia.

9. $(p_1, p_2) \in sp(R)$, jeśli $p_1 = p_2$ lub jeśli jeden z pustelników p_1, p_2 jest Najwyższym Pustelnikiem. Z drugiej strony, $ps(R)$ jest relacją uniwersalną w zbiorze Z.W.S.P.

11. (a) Wykaż, że $p(R) \cup E \subseteq pz(R)$. Aby pokazać inkluzję odwrotną, $pz(R) \subseteq zp(R)$, wystarczy dowieść, że $z(R) \subseteq zp(R)$ oraz, że relacja $zp(R)$ jest przechodnia, ponieważ wówczas $zp(R)$ zawiera przechodnie domknięcie relacji $z(R)$.

(b) Porównaj ze sobą relacje $(R \cup E) \cup (R \cup E)^{\leftarrow}$ i $(R \cup R^{\leftarrow}) \cup E$; zob. ćwiczenie 12 z § 3.1.

13. (a) Na mocy ćwiczeń 12(a) i (b), $sz(R_1 \cup R_2) = sz(R_1) \cup sz(R_2) = R_1 \cup R_2$. Zatem $psz(R_1 \cup R_2) = p(R_1 \cup R_2)$. Zastosuj twierdzenie 3.

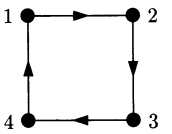
(c) Zobacz ćwiczenie 9 w § 3.1.

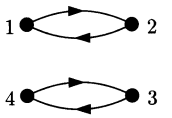
15. Do każdej relacji zawierającej relację R należeć będzie para $(1,1)$, więc relacja taka nie będzie przeciwzwrotna.

17. (a) Część wspólna wszystkich relacji zawierających relację R i mających własność p jest najmniejszą taką relacją.
(c) Nie jest spełniony warunek (i). Relacja $S \times S$ nie jest przeciwzwrotna.
(e) Nie jest spełniony warunek (ii).

Paragraf 12.1

1. (a) $(1\,5\,4\,2)$. (c) $(25)(34)$.

3. (a) 

(c) 

5. (a) $(1\,4\,2)$. (c) $(1\,3\,4\,2)$. (e) $(1\,3)(2\,4)$.

7. (a) sam cykl $(1\,2)$. (c) $(1\,6\,3)$.

9. (b) $(1\,4)(1\,3)(1\,5)(1\,2)(1\,7)$.
(c) $(k_1k_2\ldots k_m) = (k_1k_m)(k_1k_{m-1})(k_1k_{m-2})\ldots(k_1k_3)(k_1k_2)$.

11. Odpowiedzi to, odpowiednio, 1, 1, 2 i 2. Określ, które elementy mają dane rzędy.

13. (a) $\{e, (2\,5)\}$. (c) $\{e,\ (1\,6)(2\,4\,3),\ (2\,3\,4),\ (1\,6),\ (2\,4\,3),\ (1\,6)(2\,3\,4)\}$.

15. (a) e, $(1\,2)$, $(1\,2\,3)$, $(1\,2\,3\,4)$, $(1\,2\,3\,4\,5)$ i, powiedzmy, $(1\,2)(3\,4\,5)$. Permutacja $(1\,2\,3\,4\,5\,6)$ ma też rząd 6, ale jest cyklem.
(c) Cykle mają oczywiście rzędy 1, 2, 3, 4, 5 lub 6. Weźmy więc permutację przedstawioną w postaci iloczynu dwóch lub większej liczby nietrywialnych cykli; cykle długości 1 uważamy tu za trywialne. Na mocy twierdzenia 2 problem sprowadza się do przeanalizowania możliwych sposobów zapisania liczby 6 w postaci sumy, w której co najmniej dwa składniki są większe od 1.

17. Tak. $g^i \circ g^j = g^{i+j} = g^j \circ g^i$ dla wszystkich potęg elementu g.

19. (a) Jeśli j jest wielokrotnością m, powiedzmy $j = qm$, to $g^j = (g^m)^q = e^q = e$. Na odwrót, załóżmy, że $g^j = e$. Zastosuj algorytm dzielenia, by napisać, że $j = qm + r$, gdzie $0 \leq r < m$. Wówczas, jak wyżej, $g^{qm} = e$, więc $g^r = g^{qm+r} = g^j = e$. Skoro m jest najmniejszym dodatnim wykładnikiem, dla którego $g^m = e$, to liczba r musi być równa 0. Zatem j jest wielokrotnością m.
(c) Użyj oznaczeń, które zostały wprowadzone przed sformułowaniem twierdzenia 2. Załóż, że $g^j = e$ i $j > 0$. Wówczas $c_1^j c_2^j \cdots c_k^j = e$. Na mocy części (b), $c_i^j = e$ dla każdego i. Zastosuj k–krotnie część (a) i w wyniku tego stwierdź, że j jest wielokrotnością każdej z liczb

m_i. A więc j jest wspólną wielokrotnością liczb $m_1, m_2, \ldots, m_k$ i stąd $j \geq \text{NWW}(m_1, m_2, \ldots, m_k)$.

Paragraf 12.2

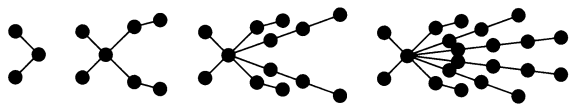
1. (a) Niech f i g będą elementami zbioru $\text{AUT}(D)$. Wówczas (x, y) jest krawędzią grafu wtedy i tylko wtedy, gdy $(g(x), g(y))$ jest jego krawędzią. Analogicznie, $(g(x), g(y))$ jest krawędzią wtedy i tylko wtedy, gdy $(f(g(x)), f(g(y)))$ jest krawędzią. Zatem (x, y) jest krawędzią wtedy i tylko wtedy, gdy $(f(g(x)), f(g(y)))$ też nią jest. Tak więc $f \circ g$ należy do $\text{AUT}(D)$. Jak zauważyliśmy w § 12.1, nie ma potrzeby sprawdzać, że jeśli $g \in \text{AUT}(D)$, to $g^{-1} \in \text{AUT}(D)$, chociaż łatwo można to zrobić bezpośrednio.

3. $|\text{AUT}(D)| = 4$, $|\text{AUT}(D)p| = |\{p, r\}| = 2$, $|\text{FIX}(p)| = |\{e, f\}| = 2$ i $2 \cdot 2 = 4$; w przypadku punktów q, r i s sprawdza się podobnie.

5. (b) i (c) Zob. przykład 2 w § 12.4.

7. Twierdzenie udowodnione w tym paragrafie pokazuje, że każda z liczb $|Gx_j|$ musi być dzielnikiem liczby $|G| = 2^k$. A więc każda z liczb $|Gx_j|$ musi być równa 1, 2, 4, ... lub 2^k. Ponieważ $|X|$ jest liczbą nieparzystą, to dla pewnego j musimy mieć $|Gx_j| = 1$. Zatem, ponownie na mocy wspomnianego twierdzenia, $|G| = |\text{FIX}(x_j)|$, a więc $G = \text{FIX}(x_j)$. Wynika stąd, że $g(x_j) = x_j$ dla wszystkich $g \in G$.

9. (a) Oto jeden z możliwych ciągów grafów.



itd.

11. $\text{FIX}(\{w, y\}) = \{e, f\}$, ponieważ zarówno punkt $f(w)$, jak i punkt $f(y)$ należą do zbioru $\{w, y\}$. Jednakże, $\text{FIX}(w) \cap \text{FIX}(y) = \{e\}$.

13. (a) $\text{AUT}(H)u = \{u, v\}$. Ponieważ $u = e(u)$ i $v = f(u)$, dobrym wyborem jest $g_1 = e$, $g_2 = f$.

(c) $\text{AUT}(H)x = \{w, x, y, z\}$. Ponieważ $w = g(x)$, $x = e(x)$, $y = f(x)$, $z = fg(x)$, dobrym wyborem jest $g_1 = e$, $g_2 = f$, $g_3 = g$, $g_4 = fg$.

15. (a) Niech, powiedzmy, $X = Gx_0$. Dla dowolnego elementu x ze zbioru X mamy oczywiście $Gx \subseteq X$. Weźmy $y \in X$. Wtedy $y = g_1(x_0)$ dla pewnego $g_1 \in G$. Mamy również $x = g_2(x_0)$ dla pewnego $g_2 \in G$. A więc $y = g_1(x_0) = g_1 \circ g_2^{-1}(x) \in Gx$.

(b) Ponieważ grupa G jest skończona, to skończony jest też zbiór $Gx = X$. Zastosuj teraz twierdzenie z tego paragafu i część (a).

17. (a) Ponieważ $e \in G$ oraz $e(x) = x$, więc para (x, x) należy do R i relacja R jest zwrotna. Ponieważ $g(x) = y \Rightarrow y = g^{-1}(x)$, więc

$(x,y) \in R \Rightarrow (y,x) \in R$, a więc relacja R jest symetryczna. Dla dowodu przechodniości zauważ, że równości $g(x) = y$ i $g'(y) = z$ implikują, że $g' \circ g(x) = z$.

Paragraf 12.3

1. (a) Na mocy przykładu 1, szukane liczby wynoszą 6, 4, 4, 2, 0, 0, 0, 0, a sumą ich jest 16.

(b) Zob. przykład 6(c), § 12.2.

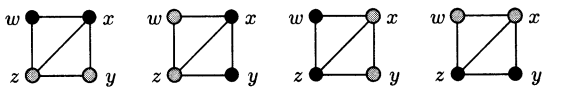
(c) Sumy te muszą być równe; wynikają one z dwóch sposobów obliczenia $|S|$, danych wzorami (1) i (2) z dowodu twierdzenia 1.

3. Automorfizmy tego grafu mogą zamieniać ze sobą w i y oraz x i z.

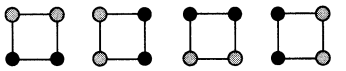
(b) $\text{FIX}(e) = \{w,x,y,z\}$, $\text{FIX}(g) = \{x,z\}$, $\text{FIX}(h) = \{w,y\}$ oraz $\text{FIX}(gh) = \emptyset$. A więc na mocy twierdzenia 1, liczba orbit grupy G wynosi $\frac{1}{4}(4+2+2+0)$. Wskaż te orbity.

5. Automorfizmy g i gh ruszają każdy z elementów orbity $\{w,y\}$. Automorfizmy h i gh ruszają każdy z elementów orbity $\{x,z\}$.

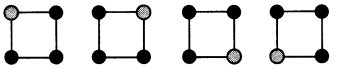
7. (a)



9. (a)



(c)



(d) Z części (a), (b) i (c) otrzymujemy trzy istotnie różne kolorowania. Dla trzech szarych wierzchołków sytuacja jest analogiczna do części (c). To razem daje nam $4+2+4+4 = 14$ kolorowań. Wszystkich kolorowań jest $2^4 = 16$. Dwa pozostałe kolorowania wykorzystują wyłącznie kolor szary lub wyłącznie czarny. Jest więc sześć istotnie różnych kolorowań: (a), (b), (c), (c) z trzema wierzchołkami szarymi, ze wszystkimi wierzchołkami szarymi bądź wszystkimi czarnymi. W następnym paragrafie poznamy o wiele lepszą metodę rozwiązania tego oraz znacznie trudniejszych problemów. Zob. przykład 2, § 12.4.

11. Permutacja h przeprowadza każdy z elementów zbioru $\{w,x,u,v\}$ na samego siebie, a więc permutacja h^* przeprowadza na samego siebie każdy z dwuelementowych podzbiorów tego zbioru. Jest $\binom{4}{2} = 6$ takich podzbiorów. Jedynym innym dwuelementowym podzbiorem, przeprowadzanym przez h^* na samego siebie jest $\{y,z\}$, a więc zbiór $\text{FIX}_T(h^*)$ ma 7 elementów.

13. (a) $|G| = 2^2$ oraz $|E| = 5$ jest liczbą nieparzystą, a więc ćwiczenie 7 z § 12.2 pokazuje, że pewna krawędź grafu musi być punktem stałym dla wszystkich elementów grupy G.

15. (a) $G = \{e, g\}$, gdzie e jest identycznością na zbiorze $V = \{u, v\}$ i $g(u) = v$, $g(v) = u$. Zarówno e^*, jak i g^* równe są identyczności na jednoelementowym zbiorze krawędzi E.

(b) Ponieważ równość $g^* = h^*$ pociąga za sobą, że $(h^{-1} \circ g)^* = e^*$, więc wystarczy pokazać, że jeśli $g^* = e^*$, to $g = e$. (Bo wtedy, jeśli $g^* = h^*$, to $(h^{-1} \circ g)^* = e^*$, co pociąga za sobą $h^{-1} \circ g = e$, a z tego wynika, że $g = h$.) Załóżmy więc, że $g^* = e^*$. Ponieważ $g^*(\{u, v\}) = \{g(u), g(v)\} = \{u, v\}$, to wszystko, co g^* może zrobić z dowolną krawędzią, to albo zostawić ją w spokoju, albo zamienić miejscami jej końce. Jeśli $g \neq e$, to funkcja g^* musi zamienić miejscami końce pewnej krawędzi $\{u, v\}$. Wtedy musi ona każdą inną krawędź $\{u, w\}$ wychodzącą z wierzchołka u doczepić do wierzchołka v, jednocześnie nie robiąc z nią nic, bądź zamieniając miejscami jej końce. Przemyśl to. Dla pewnego wierzchołka t mielibyśmy $\{u, w\} \to \{v, t\}$, ale również $\{u, w\} \to \{u, w\}$, a więc $v = w$, $t = u$ i $\{u, w\} = \{u, v\}$. Zatem krawędź $\{u, w\}$ w rzeczywistości nie mogła być różna od krawędzi $\{u, v\}$. A więc nie ma żadnych innych krawędzi wychodzących z wierzchołka u oraz, na mocy symetrii zagadnienia, z wierzchołka v. Ale wówczas, ponieważ graf H jest spójny, to $\{u, v\}$ jest jego jedyną krawędzią i H jest grafem z części (a).

Paragraf 12.4

1. (a) 4.

(c) Zastosuj twierdzenie 2, zauważając, że $m(e) = 5$, $m(f) = 4$, $|G| = 2$. Odpowiedzią jest $(k^5 + k^4)/2$.

3. (a)

	w	x	y	z
e	w	x	y	z
a	z	y	x	w

Mamy tu $m(e) = 4$, $m(a) = 2$, $|G| = 2$. Twierdzenie 2 daje wzór $C(k) = (k^4 + k^2)/2$.

(b) Tak jak w przykładzie 4, odpowiedź brzmi

$$C(4) - 4 \cdot C(3) + 6 \cdot C(2) - 4 \cdot C(1) = 136 - 4 \cdot 45 + 6 \cdot 10 - 4 \cdot 1 = 12.$$

Inne rozwiązanie polega na zauważeniu, że wszystkich permutacji danych czterech etykiet jest 24, a dwie permutacje odpowiadają równoważnym sposobom przypisania etykiet jedynie wówczas, gdy jedna jest lustrzanym odbiciem drugiej. Przykładowo, permutacja $wxyz$ jest lustrzanym odbiciem permutacji $zyxw$.

5. (a) Zastosuj twierdzenie 2, zauważając, że $m(e)=4$, $m(r)=m(r^2)=2$, $m(f)=m(g)=m(h)=3$. Odpowiedź ma postać $C(k)=(k^4+2k^2+3k^3)/6$.

(b) Zastosuj twierdzenie 2, by otrzymać wzór $C(k)=(k^3+2k+3k^2)/6$.

7. (a) Wykorzystując rysunek 12.20(a) otrzymujemy

$$C(4)-4\cdot C(3)+6\cdot C(2)-4\cdot C(1)=55-4\cdot 21+6\cdot 6-4\cdot 1=3.$$

9. (a) $C(3)-3\cdot C(2)+3\cdot C(1)=21-3\cdot 6+3\cdot 1=6$.

(b) 4.

(c) $C(5)-5\cdot C(4)+10\cdot C(3)-10\cdot C(2)+5\cdot C(1)=120-5\cdot 55+10\cdot 21-10\cdot 6+5=0$, co jest oczywiste.

11. (a) Z rysunku 12.18(b) z § 12.3 możemy odczytać, że: orbitami działania grupy $\langle e\rangle$ są zbiory $\{w\},\{x\},\{y\},\{z\}$; orbitami działania grupy $\langle g\rangle$ są zbiory $\{w,y\},\{x\},\{z\}$; orbitami działania grupy $\langle h\rangle$ są zbiory $\{w\},\{x,z\},\{y\}$; orbitami działania grupy $\langle gh\rangle$ są zbiory $\{w,y\},\{x,z\}$. A więc $m(e)=4$, $m(g)=m(h)=3$ i $m(gh)=2$.

(b) $C(k)=\frac{1}{4}(k^4+2k^3+k^2)$ na mocy twierdzenia 2 i części (a).

(c) $C(2)=\frac{1}{4}(16+16+4)=9$.

13. Stosując się do zamieszczonej sugestii otrzymujemy tablicę:

Typ	Liczba obrotów danego typu	$m(g)$, gdy grupa działa na zbiorze wierzchołków	$m(g)$, gdy grupa działa na zbiorze krawędzi
a	6	4	7
b	6	2	3
c	3	4	6
d	8	4	4
e	1	8	12

(a) $C(k)=(k^8+17k^4+6k^2)/24$. Przykładowo, jest $6+3+8=17$ obrotów, dla których $m(g)=4$.

(b) $C(k)=(k^{12}+6k^7+3k^6+8k^4+6k^3)/24$.

15. (a) W tym przypadku funkcje oznaczone na rysunku 12.24 liczbami n i $15-n$ uważane są za takie same. A więc jest 8 klas równoważności, którym odpowiada 8 istotnie różnych układów logicznych o 2 argumentach na wejściu.

(b) Tym razem, jak to zostało zauważone podczas omawiania rysunku 12.24, funkcje o numerach 2 i 4, jak również 3 i 5, są takie same. Jest teraz 6 różnych klas. Wypisz je.

Paragraf 12.5

1. (a) $\mathbb{Z}$. (c) $\mathbb{Z}$. (e) $\mathbb{Z}$.

3. (a) $3\mathbb{Z}$.

(c) Ten podzbiór nie jest podgrupą; liczby 2 i 4 do niego należą, ale liczba $2+4$ nie.
(e) $\mathbb{N} \cup (-\mathbb{N}) = \mathbb{Z} = 1 \cdot \mathbb{Z}$.

5. $h \cdot g = k \cdot g$ implikuje, że $h = h \cdot e = h \cdot (g \cdot g^{-1}) = (h \cdot g) \cdot g^{-1} = (k \cdot g) \cdot g^{-1} = k \cdot (g \cdot g^{-1}) = k \cdot e = k$.

7. (a) Wiemy, że liczba 1 jest generatorem. Pokaż, że generatorami są również liczby 2, 3 i 4.
(b) W tym przypadku jedynymi generatorami są liczby 1 i 5. Wykaż, że żadna z grup $\langle 2 \rangle$, $\langle 3 \rangle$ lub $\langle 4 \rangle$ nie jest równa $\mathbb{Z}_6$.

9. (a) $\varphi(n) = n + 3$.

11. Wykaż, że do zbiorów z części (b) i (c) należy permutacja (1 2 3), ale do żadnego z nich nie należy permutacja (1 2 3)(1 2 3). Podzbiory z części (a) i (d) to podgrupy $\text{FIX}_G(4)$ i $\text{FIX}_G(\{1,2\})$; zob. ćwiczenie 10 z § 12.2.

13. (a) Zastosuj dwukrotnie twierdzenie 1(b).
(b) Przez indukcję łatwo można pokazać, że $(g_1 \cdot g_2 \cdot \ldots \cdot g_n)^{-1} = g_n^{-1} \cdot \ldots \cdot g_2^{-1} \cdot g_1^{-1}$. W kroku indukcyjnym wykorzystuje się równość $(g_1 \cdot g_2 \cdot \ldots \cdot g_n \cdot g_{n+1})^{-1} = g_{n+1}^{-1} \cdot (g_1 \cdot g_2 \cdot \ldots \cdot g_n)^{-1}$.

15. (a) Jeśli elementy g i h należą do rozważanej części wspólnej, to oba należą do każdej z podgrup z danej rodziny, a więc to samo dotyczy ich iloczynu. Ponieważ element neutralny grupy należy do każdej z podgrup, to należy on też do ich przecięcia. Dowolny element przecięcia należy do każdej z danych podgrup, a więc to samo dotyczy też elementu doń odwrotnego. A więc ten element odwrotny również należy do rozważanego przecięcia. Wynika stąd, że rozważane przecięcie podgrup jest zamknięte ze względu na mnożenie i branie elementów odwrotnych oraz należy do niego element neutralny grupy.
(b) Dobry będzie dowolny przykład, w którym żadna z podgrup H bądź K nie zawiera drugiej. Na przykład, jeśli $G = (\mathbb{Z}, +)$, to niech $H = 2\mathbb{Z}$ i $K = 3\mathbb{Z}$.

17. (a) $6 = 3!$.
(b) $\text{FIX}_G(1) = \{e, (2\,3)\}$.
(c) $\text{FIX}_G(1) \circ (1\,2\,3) = \{(1\,2\,3), (1\,3)\}$.
(d) $(1\,2\,3) \circ \text{FIX}_G(1) = \{(1\,2\,3), (1\,2)\} \neq \{(1\,2\,3), (1\,3)\}$.
(e) Należy do niego permutacja (1 2 3), a więc jedyną warstwą lewostronną, którą zbiór ten mógłby być jest $(1\,2\,3) \circ \text{FIX}_G(1)$. Zastosuj część (d).
(f) 3.

19. (a) $\langle a \rangle = \{e, a, b\}$.
(b) $\langle a \rangle \bullet c = \{c, d, f\}$, podczas gdy $c \bullet \langle a \rangle = \{c, f, d\}$.
(c) $\langle c \rangle, \langle d \rangle, \langle f \rangle$.

(d) $|G|/|\langle d\rangle| = 6/2 = 3$. Zob. też część (e).
(e) $\{e, d\}, \{a, c\}, \{b, f\}$.

21. Do zbioru $g \cdot H$ należy element $g \cdot e = g$. Stąd, jeśli $g \cdot H = H$, to $g \in H$. Jeśli $g \in H$, to element g należy do warstwy H i do warstwy $g \cdot H$. Warstwy te nie są więc rozłączne, zatem na mocy twierdzenia 4 mamy $g \cdot H = H$.

23. (a) Dla $h \in H$ mamy $(g \cdot h)^{-1} = h^{-1} \cdot g^{-1} \in H \cdot g^{-1}$, a więc

$$\{f^{-1}\colon f \in g \cdot H\} \subseteq H \cdot g^{-1}.$$

Ponadto dla $h \in H$ element $h \cdot g^{-1} = (g \cdot h^{-1})^{-1}$ należy do zbioru $\{f^{-1}\colon f \in g \cdot H\}$, a więc

$$H \cdot g^{-1} \subseteq \{f^{-1}\colon f \in g \cdot H\}.$$

(b) $g \cdot H \to H \cdot g^{-1}$. Wykaż, że funkcja ta jest różnowartościowa.

25. (a) (Z) Zauważ, że $g^{-1} \cdot g = e \in H$.
(S) Zauważ, że $g_1^{-1} \cdot g_2 = (g_2^{-1} \cdot g_1)^{-1}$.
(P) Zauważ, że $g_3^{-1} \cdot g_1 = (g_3^{-1} \cdot g_2) \cdot (g_2^{-1} \cdot g_1)$.
(b) Wykaż, że $g_1 \cdot H = \{g \in G\colon g \sim g_1\}$.

Paragraf 12.6

1. (a), (c), (e).

3. (a) Nie jest izomorfizmem, gdyż nie przekształca $\mathbb{Z}$ na $\mathbb{Z}$.
(c) Jest izomorfizmem jako homomorfizm różnowartościowy i „na".
(e) Nie jest izomorfizmem, ponieważ funkcja $\varphi(n) = 3n$ nie przekształca $\mathbb{Z}$ na $\mathbb{Z}$.

5. (a) Wykaż, że $[f + (g + h)](x) = [(f + g) + h](x)$ dla wszystkich $x \in \mathbb{R}$; to pokaże, że działanie $+$ jest łączne w zbiorze F. Funkcja $\mathbf{0}$, gdzie $\mathbf{0}(x) = 0$ dla wszystkich $x \in \mathbb{R}$ jest elementem neutralnym w F. Elementem przeciwnym do danej funkcji jest po prostu funkcja do niej przeciwna.
(b) Tak. Uzasadnij to.
(c) $\varphi(f + g) = (f + g)(73) = f(73) + g(73) = \varphi(f) + \varphi(g)$.

7. (a) $\{0\}$. (c) $5\mathbb{Z} = \{5n\colon n \in \mathbb{Z}\}$.

9. (a) 4. (c) 3.

11. (a) Elementem neutralnym jest para (e_G, e_H), gdzie e_G i e_H są elementami neutralnymi w grupie G i H. Ponadto, $(g, h)^{-1} = (g^{-1}, h^{-1})$.
(c) $\{(e_G, h)\colon h \in H\}$.
(d) $\{(e_G, h)\colon h \in H\}$. Część (c) pokazuje, że podgrupa ta jest dzielnikiem normalnym.

13. (b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot H = \left\{ \begin{bmatrix} 0 & 1 \\ 1 & x \end{bmatrix} : x \in \mathbb{R} \right\}$, podczas gdy $H \cdot \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \left\{ \begin{bmatrix} x & 1 \\ 1 & 0 \end{bmatrix} : x \in \mathbb{R} \right\}$.

(c) $\begin{bmatrix} y & z \\ 0 & 1/y \end{bmatrix} \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/y & -z \\ 0 & y \end{bmatrix} = \begin{bmatrix} 1 & xy^2 \\ 0 & 1 \end{bmatrix}$ należy do H.

(d) Jądrem homomorfizmu φ jest grupa H.

(e) Wykorzystaj odpowiedź z części (d) i twierdzenie o izomorfizmie.

15. Przeciwobrazem zbioru złożonego z $\varphi(g)$ jest $g \cdot K$, gdzie K jest jądrem homomorfizmu φ. Na mocy założenia, $|g \cdot K| = 1$. A więc $|K| = 1$ i homomorfizm φ jest różnowartościowy, jak wynika z wniosku z twierdzenia 1.

17. (a) Wykorzystaj tożsamość $(g \cdot h) \cdot H \cdot (g \cdot h)^{-1} = g \cdot (h \cdot H \cdot h^{-1}) \cdot g^{-1}$. Ponadto, jeśli $H = g \cdot H \cdot g^{-1}$, to $g^{-1} \cdot H \cdot g = g^{-1} \cdot (g \cdot H \cdot g^{-1}) \cdot g = H$.

(b) Zbiór $\{g \in G\colon g \cdot H \cdot g^{-1} = H\}$ jest podgrupą (część (a)) grupy G zawierającą A, a G jest najmniejszą podgrupą zawierającą A. A więc mamy $\{g \in G\colon g \cdot H \cdot g^{-1} = H\} = G$.

19. (a) $e \cdot K \cdot (1\,3) \cdot K = e \cdot \{e, (1\,2)\} \cdot (1\,3) \cdot \{e, (1\,2)\} = \{(1\,3), (1\,3\,2), (1\,2\,3), (2\,3)\}$, ale warstwy mają tylko po dwa elementy.

(b) Nie. Uzasadnij.

Paragraf 12.7

1. (a) Tak.

(c) Nie. Tylko element 1 ma element odwrotny.

3. (a) Tak. (c) Tak; por. przykład 7(b).

5. (a) Nie. Podaj kontrprzykład.

(c) Nie, na przykład macierz zerowa nie ma elementu odwrotnego.

7. (a) *break*, *fast*, *fastfood*, *lunchbreak*, *foodfood*.

(c) *fastfast*, *foodfood*, *fastfastfoodbreakbreak*, λ.

9. (b) Półgrupa $(\mathbb{N}, \max)$ jest monoidem, ponieważ 0 jest elementem neutralnym. W półgrupie $(\mathbb{N}, \min)$ nie istnieje element neutralny, więc nie jest ona monoidem. Wykaż te stwierdzenia.

11. (a) $\mathbb{P}$. (c) $\mathbb{Z}$. (e) $\mathbb{Z}$. (g) $18\mathbb{P} = \{18k\colon k \in \mathbb{P}\}$.

13. $\mathbb{P} = \{1\}^+$, $\{0\} = \{0\}^+$, $18\mathbb{P} = \{18\}^+$.

15. (a) $2\mathbb{N} = \{2k\colon k \in \mathbb{N}\}$. (c) $\{0\}$. (e) Σ^*.

17. (a) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ oraz $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

(b) Składa się ona z następujących sześciu „macierzy permutacyjnych":

$$\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}, \begin{bmatrix}0&1&0\\1&0&0\\0&0&1\end{bmatrix}, \begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}, \begin{bmatrix}0&0&1\\1&0&0\\0&1&0\end{bmatrix}, \begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}.$$

Półgrupa ta jest izomorficzna z grupą S_3 wszystkich permutacji zbioru trójelementowego.

(c) Składa się ona z trzech macierzy. Znajdź je.

19. (a) $6\mathbb{P} = \{6k\colon k \in \mathbb{P}\}$.

(c) Nie. Na przykład liczby 6 i 12 nie są potęgami tej samej liczby ze zbioru $6\mathbb{P}$, a więc nie mogą należeć do tej samej pólgrupy cyklicznej.

21. (a) $60\mathbb{P}$. (c) Liczba 60 generuje addytywną półgrupę $60\mathbb{P}$.

23. Jedynie funkcja φ z części (a) jest homomorfizmem. Chociaż jest ona różnowartościowa, nie jest izomorfizmem, gdyż nie przekształca zbioru $\mathbb{P}$ na zbiór $\mathbb{P}$.

25. Ogólniej, jeśli $A \subseteq S$, to

$$\begin{aligned}\varphi(A^+) &= \varphi(\{s\colon s \text{ jest iloczynem } a_1 \cdot \ldots \cdot a_n \text{ elementów zbioru } A\})\\ &= \{\varphi(s)\colon s = a_1 \cdot \ldots \cdot a_n \text{ i } a_1, \ldots, a_n \in A\}\\ &= \{t\colon t = \varphi(a_1) \cdot \ldots \cdot \varphi(a_n) \text{ i } a_1, \ldots, a_n \in A\} = \varphi(A^+).\end{aligned}$$

27. (a) Jeśli z' jest także zerem, to $z' = z \bullet z' = z$.

(c) $(\{0,1\}, \cdot)$. Czy potrafisz znaleźć inny przykład?

29. (a) Zbiór $\varphi(S)$ jest zamknięty ze względu na iloczyny, gdyż

$$\varphi(s) \square \varphi(s') = \varphi(s \bullet s') \in \varphi(S).$$

Ponadto element $\varphi(e)$ jest elementem neutralnym w zbiorze $\varphi(S)$, ponieważ

$$\varphi(s) \square \varphi(e) = \varphi(s \bullet e) = \varphi(s) = \varphi(e \bullet s) = \varphi(e) \square \varphi(s)$$

dla wszystkich elementów $\varphi(s) \in \varphi(S)$.

(b) Nie. Weź, na przykład, dowolny monoid $(S, \bullet)$ i zdefiniuj homomorfizm $\varphi\colon (S, \bullet) \to (\mathbb{Z}, \cdot)$ wzorem $\varphi(s) = 0$ dla wszystkich elementów $s \in S$.

Paragraf 12.8

1. (a), (b), (d), (f).

3. Wszystkie z wyjątkiem (b) i (e). Funkcja (b) nie zachowuje dodawania, a funkcja (e) nie zachowuje mnożenia.

5. (a) Do obu stron równości podstaw $x = a$.

(c) Na mocy (b) mamy

$$p(x) = \sum_{k=0}^{n} c_k x^k = \sum_{k=0}^{n} c_k \{q_k(x) \cdot (x-a) + a^k\}$$
$$= \Big(\sum_{k=0}^{n} c_k q_k(x)\Big) \cdot (x-a) + \sum_{k=0}^{n} c_k a^k.$$

Zdefiniuj wielomian $q(x)$ wzorem $q(x) = \sum_{k=0}^{n} c_k q_k(x)$.

(d) Wielomian $p(x)$ należy do szukanego jądra wtedy i tylko wtedy, gdy $p(a) = 0$. Skorzystaj z części (c).

7. (a) $24\mathbb{Z}$.
(b) $2\mathbb{Z}$.
(c) $3\mathbb{Z} + 2\mathbb{Z} = 1\mathbb{Z} = \mathbb{Z}$, gdyż $1 = 3 \cdot 1 + 2 \cdot (-1) \in 3\mathbb{Z} + 2\mathbb{Z}$. (d) $\mathbb{Z}$.

9. (a) Sprawdź poprawność definicji bezpośrednio lub zastosuj twierdzenie 1 do homomorfizmu $m \longrightarrow (m \text{ MOD } 4, m \text{ MOD } 6)$ pierścienia $\mathbb{Z}$ w pierścień $\mathbb{Z}_4 \times \mathbb{Z}_6$, tak jak w przykładzie 8. Zgodnie z faktem odnotowanym w ćwiczeniu 12 wystarczy sprawdzić, że funkcje $m \longrightarrow m \text{ MOD } 4$ i $m \longrightarrow m \text{ MOD } 6$ są homomorfizmami określonymi na $\mathbb{Z}_{12}$.
(c) Para $(1, 4)$ jest jednym z 12 takich elementów; znajdź jeszcze jeden.
(d) Jedynie liczbę 9.

11. (a) Ponieważ $2 *_6 3 = 0$, to nie istnieje element odwrotny do liczby 2.
(b) Dla każdego niezerowego elementu wskaż element odwrotny. Elementy odwrotne do niezerowych elementów pierścienia $\mathbb{Z}_5$ mogą być odczytane z rysunku 3.24 z § 3.6. (Zob. ćwiczenie 16, w którym przedstawiony jest ogólny argument).
(c) Pierścień $F \times K$ nie jest nawet dziedziną całkowitości, gdyż $(1, 0) \cdot (0, 1) = (0, 0)$.

13. (a) Wobec przykładu 6(b), jądrem homomorfizmu φ jest bądź zbiór F, bądź $\{0\}$.
(c) Skorzystaj z części (b) i przykładu 6(b).

15. (a) $I = 15\mathbb{Z}$.
(b) Zobacz przykład 8. Zdefiniuj izomorfizm φ za pomocą wzoru $\varphi(m) = (m \text{ MOD } 3, m \text{ MOD } 4)$ dla $m \in \mathbb{Z}_{12}$.

17. (a) $R{\cdot}2 = \{a_0 + a_1 x + \ldots + a_n x^n {\in} R$: każda z liczb a_i jest parzysta$\}$,
$R{\cdot}x = \{a_0 + a_1 x + \ldots + a_n x^n \in R$: $a_0 = 0\}$,
$R{\cdot}2 + R \cdot x = \{a_0 + a_1 x + \ldots + a_n x^n {\in} R$: liczba a_0 jest parzysta$\}$.
(b) Przypuśćmy, że dla pewnego wielomianu $p \in R$ mamy $R \cdot p = R{\cdot}2 + R{\cdot}x$. Ponieważ $2 \in R{\cdot}p$, to p musi być wielomianem stałym; ponadto $x \in R{\cdot}p$, więc wielomian p jest równy 1 lub -1. Ale wtedy $R \cdot p = R$ jest sprzecznością.

Paragraf 13.1

1. (a) 0. (b) 0. (c) 1. (d) 0. (e) 0.

3. (a) $\forall x \forall y \forall z[((x < y) \wedge (y < z)) \rightarrow (x < z)]$; dziedziną jest zbiór $\mathbb{R}$.
(c) $\forall m \forall n \exists p[(m < p) \wedge (p < n)]$; dziedziną jest zbiór $\mathbb{N}$.
(e) $\forall n \exists m[m < n]$; dziedziną jest zbiór $\mathbb{N}$.

5. (a) $\forall w_1 \forall w_2 \forall w_3[(w_1 w_2 = w_1 w_3) \rightarrow (w_2 = w_3)]$.
(c) $\forall w_1 \forall w_2[w_1 w_2 = w_2 w_1]$.

7. (a) Zmienne x i z są związane; y jest zmienną wolną.
(c) Taka sama odpowiedź jak w punkcie (a).

9. (a) Zmienne x i y są wolne; nie ma zmiennych związanych.
(b) $\forall x \forall y[(x-y)^2 = x^2 - y^2]$ jest zdaniem fałszywym.
(c) $\exists x \exists y[(x-y)^2 = x^2 - y^2]$ jest zdaniem prawdziwym.

11. (a) Nie. Zdanie $\exists m[m+1 = n]$ jest fałszywe dla $n = 0$.
(b) Tak.

13. (a) $\exists! x \forall y[x + y = y]$.
(c) $\exists! A \forall B[A \subseteq B]$. A i B należą tutaj do dziedziny $\mathcal{P}(\mathbb{N})$. Zauważ, że $\forall B[A \subseteq B]$ jest zdaniem prawdziwym wtedy i tylko wtedy, gdy $A = \emptyset$.
(e) „$f\colon A \rightarrow B$ jest funkcją różnowartościową” $\rightarrow \forall b\ \exists! a[f(a) = b]$. Zmienna a należy tutaj do zbioru A, a b należy do zbioru B. Dobrze to widać, jeśli napisze się $\forall b \in B \exists! a \in A[f(a) = b]$.

15. (a) Prawdziwe.
(c) Fałszywe; np. 3 znajduje się w zbiorze po prawej stronie.
(e) Fałszywe; zbiór po prawej stronie jest pusty.
(g) Prawdziwe.
(i) Prawdziwe.
(k) Prawdziwe.
(m)Prawdziwe.

17. (a) 0. (c) 1. (e) 0.

Paragraf 13.2

1. (a) Każdy członek klubu był pasażerem każdej linii lotniczej wtedy i tylko wtedy, gdy każda linia lotnicza miała każdego członka klubu jako pasażera.
(c) Jeśli istnieje członek klubu, który był pasażerem każdej linii lotniczej, to każda linia lotnicza miała członka klubu jako pasażera.

3. Reguła 37b mówi, że zdanie „nie istnieje żółty samochód” jest logicznie równoważne zdaniu "żaden samochód nie jest żółty". W rzeczywistości oba te zdania są fałszywe. Reguła 37d mówi, że zdanie „istnieje żółty samochód” jest logicznie równoważne zdaniu „nie każdy samochód nie jest żółty". Oba one są prawdziwe.

5. (a) Zobacz regułę 8c.

7. $\exists n[\neg\{p(n) \to p(n+1)\}]$ lub $\exists n[p(n) \wedge \neg p(n+1)]$.

9. (a) $\exists x\, \exists y[(x < y) \wedge \forall z\{(z \leq x) \vee (y \leq z)\}]$.
(c) 0; na przykład $[x < y \to \exists z\{x < z < y\}]$ jest zdaniem fałszywym dla $x = 3$ i $y = 4$.

11. Można przyjąć, że $q(x, y)$ jest predykatem „$x = y$". Inny sposób na to, by poradzić sobie z $\exists x\, p(x, x)$, jest taki, że $r(x)$ jest jednoargumentowym predykatem $p(x, x)$. Wtedy $\exists x\, r(x)$ jest predykatem złożonym.

13. $\exists N\, \forall n[p(n) \to (n < N)]$.

Paragraf 13.3

1. (a) Prawdziwe.
(c) Fałszywe.
(e) Prawdziwe. Por. ćwiczenie 3 z § 1.3.

3. (a) Funkcja postaci $f(x) = ax + b$ będzie dobra, jeśli dobierzesz a i b tak, by $f(0) = -1$ i $f(1) = 1$. Naszkicuj rozwiązanie, aby zobaczyć, że jest dobre.
(b) Użyj g, gdzie $g(x) = 1 - x$.
(c) Zmodyfikuj sugestię podaną w (a).
(d) Użyj funkcji $x \to 1/x$.
(e) Przekształć przedział $(1, \infty)$ na $(0, \infty)$, używając funkcji $h(x) = x-1$ i złóż ją z funkcją, którą otrzymałeś jako rozwiązanie ćwiczenia (d).
(f) Powiedzmy, $f(x) = 2^x$. Naszkicuj f, aby zobaczyć, że jest ona dobra.

5. (a) Użyj danych:

x	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
$f(x)$	−8,89	−3,75	−1,90	−0,83	0	0,83	1,90	3,75	8,89

7. Tylko zbiory w (b) i (c) są przeliczalne.

9. (a) Możemy przyjąć, że S jest zbiorem nieskończonym. Niech $f\colon S \to T$ będzie przekształceniem wzajemnie jednoznacznym, gdzie T jest zbiorem co najwyżej przeliczalnym. Istnieje przekształcenie wzajemnie jednoznaczne $g\colon T \to \mathbb{P}$ (dlaczego?). Wtedy $g \circ f$ jest przekształceniem wzajemnie jednoznacznym zbioru S na zbiór $\mathbb{P}$.

11. (a) Zastosuj punkt (b) twierdzenia do zbioru $S \times T = \bigcup_{t \in T}(S \times \{t\})$. Dlaczego każdy zbiór $S \times \{t\}$ jest co najwyżej przeliczalny?
(b) Dla każdego $t \in T$ niech $g(t)$ będzie takim elementem zbioru S, że $f(g(t)) = t$. Pokaż, że g jest funkcją różnowartościową i zastosuj punkt (a) twierdzenia.

13. (a) Dla każdej funkcji $f \in \text{FUN}(\mathbb{P}, \{0,1\})$ definiujemy $\phi(f)$ jako zbiór $\{n \in \mathbb{P}\colon\ f(n) = 1\}$. Pokaż, że ϕ jest funkcją różnowartościową ze zbioru $\text{FUN}(\mathbb{P}, \{0,1\})$ na zbiór $\mathcal{P}(\mathbb{P})$.

(b) Wykorzystaj przykład 2(a) i ćwiczenie 9.

15. W kroku indukcyjnym wykorzystaj równość $S^n = S^{n-1} \times S$.

SKOROWIDZ

Alfabet 20
algebra Boole'a 588, 589
algebry Boole'a izomorficzne 597
algorytm Dijkstry 504
– – ze wskaźnikami 519
– DRZEWO 382
– dzielenia 187, 188, 201
– ETYKIETOWANIE DRZEWA 435
– EUKLIDES 424
– EUKLIDES$^+$ 424
– Euklidesa 200, 257, 264
– Fleury'ego 346
– HUFFMAN 463
– INORDER 431
– Kruskala 386
– LAS 382
– NUMEROWANIE WIERZCHOŁKÓW 481
– POSTORDER 428
– PREORDER 427
– Prima 389
– REKUR 419
– rekurencyjny 413, 414
– SORTOWANIE DRZEWA 438
– SORTOWANIE KRZAKA 448
– TEST 413
– TEST DRZEWA 415
– – WPZ 417
– UJŚCIE 480
– WAGI MAKSYMALNE 513
– Warshalla 508, 509
– – ze wskaźnikami 520
– zachłanny 389
algorytmy „dziel i rządź" 244
– poszukiwania z nawrotami 434
– przeszukiwania w głąb 427, 434
alternatywa nie wykluczająca 25, 90
alternatywa wykluczająca 25, 90, 99
atom 594
automorfizm grafu 703
– – skierowanego 701

Bramka 612
– AND 613, 615
– AND-OR-INVERT 623
– NAND 613, 615
– NOR 613, 615
– OR 613, 615
brydż 309

Ciało 774
ciąg 57
– de Bruijna rzędu n 486
– Fibonacciego 231, 241, 407
– liczb wierzchołków kolejnych stopni 333
– Lucasa 256
– skończony 62
CIĄG(n) 63, 232
CYFR(n) 74
CYFR2(n) 74
cykl 146, 328, 689
– Eulera 340
– Hamiltona 372
cykle rozłączne 689
czas działania algorytmu Fleury'ego 350

część całkowita liczby 188, 231
– ułamkowa liczby 188

Definicja jednoznaczna 406
– rekurencyjna 398, 400
– – ciągu 200, 228
deg(v) 333
diagram Hassego 600, 637
– Venna 25
długość drogi 148
długość słowa 22, 41
dołączenie liścia 401
domknięcie relacji 678
– przechodnie 678
– symetryczne 678
– zwrotne 678
dopełnienie zbioru 27
dowód formalny zdania 115
– indukcyjny 398
– konstruktywny 105
– niekonstruktywny 105
– niewprost 101
– poprawności algorytmu Fleury'ego 349
– poprawny 122
– trywialny 105
– wprost 101
– „w próżni" 104
droga 145, 148
– acykliczna 146, 328
– długości n 145, 150, 327
– Eulera 340
– Hamiltona 372
– krytyczna 496
– maksymalna 496
– między wierzchołkami 150
– minimalna 492
– prosta 328
– zamknięta 145, 150, 327
drzewo 352
– binarne 366
– o m rozgałęzieniach 366
– pełne o m rozgałęzieniach 366
– poszukiwań binarnych 360
– regularne o m rozgałęzieniach 366
– spinające 354
– uporządkowane z wyróżnionym korzeniem 367
– z m rozgałęzieniami 409
– z wagami 459
– z wyróżnionym korzeniem 360
DWA(n) 60

dystrybuanta rozkładu dwumianowego 573
– – jednostajnego 550
– – normalnego (Gaussa) 579
– zmiennej losowej 548
działania w algebrze Boole'a 589
działanie booleowskie 667
działanie grupy na zbiorze 699, 715
– przechodnie 708
dziecko 365
dziedzina 791
dziedzina całkowitości 774
dzielnik 257
– normalny 752

Element neutralny grupy 734
– – półgrupy 760
– odwrotny do elementu 734
– przeciwny do elementu 735
– rzędu nieskończonego 737
element zbioru 15
– – częściowo uporządkowanego 641
– – – –, maksymalny 641
– – – –, minimalny 642
– – – –, najmniejszy 643
– – – –, największy 643
elementy izomorficzne 176
– podobne 176
– przystające 176
etykietowanie uporządkowane 434, 436, 480

Fałsz 78
funkcja 37
– booleowska 598
– charakterystyczna 42
– dobrze określona 184
– γ 143
– głębokości 408
– identycznościowa 42
– jako relacja 136
– logarytmiczna 50
– odwracalna 49
– odwrotna 48
– określona na zbiorze 37
– prawdopodobieństwa 283
– różnowartościowa 40
– stała 42
– zdaniowa 83

Głębokość wyrażenia 426
graf 142, 327

graf acykliczny 146, 328
– dwudzielny 378
– hamiltonowski 372
– pełny 334
– regularny 334
– skierowany 142, 477
– – z wagami 490
– spójny 342
grafy izomorficzne 332
grupa 733
– cykliczna 737
– generowana przez element 696
– permutacji 687
– symetryczna 697
grupy izomorficzne 734
G-orbita 700

Hierarchia ciągów 69
hipoteza Goldbacha 78, 82, 85
homomorfizm ewaluacji 777
– grup 748
– –, naturalny (kanoniczny) 753
– pierścieni 776
– –, naturalny 779
– półgrup 767

Ideał pierścienia 778
– główny 778
iloczyn booleowski 668
– kartezjański 32
– macierzy 167
– – przez skalar 158
– minimalny 605
iloczyn prosty grup 758
iloraz 188
implikacja logiczna 95, 96, 803
– spełniona „w próżni" 104
– trywialna 105
indeg(v) 483
indukcja matematyczna 200, 217
infimum 643
iteracja 203
izomorfizm 332
– algebr Boole'a 597
– grup 735
– pierścieni 779
– półgrup 767

Jądro homomorfizmu grup 750
jedynka pierścienia 772
jedynka półgrupy 760
język 20

Klasa abstrakcji 179
– reszt modulo p 191
– równoważności 179
kod Graya 377
– Huffmana 472
– prefiksowy 466
kolorowania równoważne 717
kombinacje 277
koniec krawędzi 143
konkatenacja słów 761
kontrapozycja 81
kontrprzykład 85, 806
korzeń 360
– drzewa 402
kostka symetryczna 286
krata 644
krawędź grafu 143
– krytyczna 496
– wielokąta 144, 147
kres dolny 643
– górny 643
kreska Sheffera 121, 615
krok indukcyjny 220
– początkowy definicji 229
krzywa dzwonowa 580
kwantyfikator 83
– egzystencjalny 83
– $\exists!$ 798
– ogólny 83, 791
– szczegółowy 83, 791

Las 358
L-droga 506
– minimalna 506
lemat o zliczaniu 305
liczba elementów zbioru 33
liczby pierwsze bliźniacze 108
– względnie pierwsze 262
liść 357, 365
litera alfabetu 20
logarytm naturalny 50
losowanie
– bez zwracania 274
– ze zwracaniem 274
LUC(n) 256
luz pełny 498
– swobodny 503

Łańcuch 651
– maksymalny 652
łączność złożenia funkcji 45

Macierz 154
– booleowska 668
– jednostkowa 170
– kwadratowa 155
– odwracalna 172
– odwrotna 172
– przeciwna 157
– sąsiedztwa 158
– symetryczna 161
– transponowana 155
– wymiaru $m \times m$ 150
macierze równe 155
maksimum 643
matryca logiczna 89, 91
– – dla implikacji 90
metoda PERT 500
– przekątniowa Cantora 812
– tablic Karnaugha 610
metody dowodzenia 101
minimalne drzewo spinające 382, 385
minimum 643
mnożenie macierzy 165
modus ponendo ponens 96
– ponendo tollens 96
– tollendo tollens 96
moneta niesymetryczna 571
– symetryczna 288
monoid 761
multigraf 147
– skierowany 144

Największy wspólny dzielnik 200, 257
NAND 121
n DIV p 189
niezależność zdarzeń 532
– zmiennych losowych 545
niezmiennik izomorfizmu 332
– pętli 201, 205, 207
n MOD p 189
notacja infiksowa 451
– O 64, 67, 68, 72
– odwrotna polska 451
– polska 449, 451
– postfiksowa 451
– prefiksowa 451
– zbioru 16

n-ty wyraz ciągu 59
numer poziomu 366
NWD 257
NWD$^+$ 263

Obcięcie funkcji 52, 714
obliczenia iteracyjne 230
obraz zbioru 52
odchylenie standardowe 565
odcinek początkowy słowa 640
ODWR 56, 209
odwrócenie słowa 411
odwzorowanie 37
ograniczenie dolne 643
– górne 643
określenie implikacji za pomocą alternatywy lub koniunkcji 95
– równoważności 95
operator domknięcia 678
optymalne drzewo binarne 463
opuszczanie koniunkcji 96
orbita działania grupy 700
outdeg(v) 483
oznaczenie fałszu 89
– prawdy 89

Para uporządkowana 32
pełny graf dwudzielny 378
– sumator 620
permutacja 276, 687
– identycznościowa 688
pętla 147, 327
– „dopóki" 200, 201, 202
pierścień 772
– przemienny 772
– z jedynką 772
początek krawędzi 143
podciało 776
podgrupa 693, 734
– generowana przez zbiór 738
– niewłaściwa 694
– normalna 752
– trywialna 694
– właściwa 694
podgrupy sprzężone 757
podmonoid 769
podpierścień 776
podpółgrupa 761
podzbiór 17
– właściwy 18

podzbiór zbioru częściowo uporządkowanego 642
podział uporządkowany 307
podział zbioru 179
poker 279
porażka 570
porządek częściowy 636
– dobry 653
– leksykograficzny 656, 658
– liniowy 367, 651
– postfiksowy 428
– prefiksowy 427
– produktowy 655
– standardowy 657
postać kanoniczna wyrażenia 606
– normalna alternatywno-koniunkcyjna 606
POT(n) 235
potomek 365
półgrupa 760
– cykliczna 766
– generowana przez zbiór 764
– przemienna 760
półgrupy izomorficzne 767
półsumator 619
prawa De Morgana 94, 95, 591, 801, 804
– dylematu konstrukcyjnego 96
– – destrukcyjnego 96
– działań w algebrze Boole'a 589, 590
– idempotentności 95
– identyczności 95
– łączności 95
– łączności i skracania dla działań w zbiorze $\mathbb{R}$ 84
– pochłaniania 122
– przemienności 95
– rozdzielności 95
prawda 78
prawo eksportacji 95
– iloczynu 273
– łączności dla grup 733, 735
– – – macierzy 171
– – – relacji 665
– kontrapozycji 94, 95
– podwójnego przeczenia 95
– przemienności dla grup 735
– skracania równości (w grupie) 736
– – w dziedzinie całkowitości 775
– sumy 272
prawdopodobieństwo 283
– warunkowe 529
predykat 792
predykat n-argumentowy 794
problem mostów królewieckich 340
procedura Quine'a–McCluskeya 610
PROD(n) 236
produkt zbiorów 34
przebieg pętli 203
przechodniość implikacji 96
– równoważności 96
przecięcie zbiorów 25
przeciwdziedzina funkcji 37
przeciwobraz elementu 54
– zbioru 53
przedział 18
– domknięty 18
– otwarty 18
przekształcenie 37
– kanoniczne 182
– naturalne 182
– wzajemne jednoznaczne 40
przestrzeń 27, 791
– zdarzeń elementarnych 283
przystawanie modulo 137, 190
punkt stały permutacji 689

Quasi-porządek 637

Rachunek predykatów 83, 791
– zdań 78, 89, 109
reductio ad absurdum 95
reguła modus ponendo ponens 117
– – – tollens 117
– – tollendo tollens 117
– opuszczania koniunkcji 117
– podstawiania 109, 112
– sylogizmu hipotetycznego 117
– wprowadzania alternatywy 117
– – koniunkcji 117
reguły wnioskowania 115, 117
relacja 135
– antysymetryczna 138
– częściowego porządku 593
– dwuargumentowa 135
– kongruencji 138, 190
– należenia 15
– odwrotna 138
– osiągalności 150, 177
– przechodnia 138
– przeciwzwrotna 138
– równości 686
– równoważności 175, 176

relacja sąsiedztwa 146, 150, 177
– symetryczna 138
– uniwersalna 150
– zwrotna 138
reszta modulo p 189
– z dzielenia 188
rezerwa czasowa krawędzi 498
– – wierzchołka 496
rodzic 365
rozkład dwumianowy 547, 572
– Gaussa 579
– geometryczny 577
– jednostajny 550
– normalny 579
– prawdopodobieństwa zmiennej losowej 547
rozmieszczanie przedmiotów w pudełkach 300
rozumowanie błędne 122
rozwiązywanie kongruencji 200
równanie charakterystyczne 240
– rekurencyjne 228
równość zbiorów 17
równoważność 92
różnica zbiorów 25
– –, symetryczna 25
rysunek grafu 148
– – skierowanego 143
rysunek relacji 670
rząd elementu grupy 696, 737
RZUT 47

SEQ 229
sieć logiczna 612
– zdarzeń 495
silnia 58
SILNIA 208
SILNIA(n) 59, 229
skalar 158
składowa grafu 342
słowo 20
– puste 21
spójnik logiczny 80, 90-92
sprowadzenie do sprzeczności 96, 101
stopień wejściowy 483
– wierzchołka 333, 483
– wyjściowy 483
sukces 570
SUMAKW(n) 63
SUMA(n) 229
suma macierzy 156
– maksymalna zmiennych 611
suma zbiorów 25
supremum 643
symbol atomowy 605
SYM(n) 239

Tablica wartości logicznych 89
tautologia 93, 801
teza 115
TRANS(A) 171
transpozycja 698
treść pętli 202
twierdzenie 115
– chińskie o resztach 782
– Eulera 344
– – dla grafów skierowanych 484
– Karnaugha 624
– Lagrange'a 742
– o izomorfizmie 754

Ujście 478
układ kart w brydżu 309
– w pokerze 279
uniwersum 27, 791
uogólniona zasada indukcji 404
uzupełnienie zbioru 27

Waga 373, 385, 490
– drogi 490
– drzewa 459
– liścia 459
– maksymalna 496
– minimalna 492
wariacja r-wyrazowa 276
wariancja zmiennej losowej 565
warstwa podgrupy 741, 752
– lewostronna 740
– prawostronna 740
wartość bezwzględna 103
– elementu przy funkcji 37
– logiczna 83
– – zdania złożonego 89, 91
wartość oczekiwana zmiennej losowej 556
wartość średnia zmiennej losowej 556
warunek dozoru pętli 202
– konieczny 82
– początkowy 220
– rekurencyjny 400
– wystarczający 82
warunki poprawności dla algorytmów rekurencyjnych 424

wektor 157
– kolumnowy 155
– wierszowy 155
węzeł gałęzi 365
– końcowy 365
– wewnętrzny
wierzchołek grafu 142
– osiągalny 483
– sąsiedni 146
własności dodawania macierzy 157
wniosek 115
wnioskowanie poprawne 122
wprowadzenie alternatywy 96
wskaźnik 518
wspólny dzielnik 257
współczynnik dwumianowy 277
wzór Bayesa 537
– dwumianowy Newtona 298
– na prawdopodobieństwo całkowite 536
wyjście z pętli 203
wykres funkcji 38
wyrażenia booleowskie 601
– – n zmiennych 602
– – równoważne 604
– poprawnie zbudowane 416
– – – dla notacji polskiej 454
– – – – odwrotnej notacji polskiej 454
wysokość drzewa 366
– elementu 409
wzór jawny 200
– rekurencyjny 228

Zależność rekurencyjna 228, 239
założenie 115
zasada dobrego uporządkowania 204
– dualności 589
– indukcji 200
– – matematycznej 220
– rozmieszczania przedmiotów w pudełkach 300
– skończonej indukcji matematycznej 220, 249, 250, 251
– szufladkowa Dirichleta 315
– – uogólniona 321
– włączeń i wyłączeń 295
zbiory tej samej wielkości 810
zbiór 15
– bezpośrednich następników 479
zbiór co najwyżej przeliczalny 810
– częściowo uporządkowany 636
– dobrze uporządkowany 653
– liczb całkowitych 16
– – – dodatnich 16
– – naturalnych 16
– – rzeczywistych 16
– – wymiernych 16
– liniowo uporządkowany 651
– nieprzeliczalny 810
– nieskończony 809
– predykatów złożonych 800
– przeliczalny 810
– pusty 19
– potęgowy 19
– uniwersalny 27, 791
– zaetykietowany 438
– $\mathbb{Z}_p$ 187, 189
zdania logicznie równoważne 83, 94
– silniejsze 98
– słabsze 98
zdanie 78
– fałszywe 85
– odwrotne 81
– przeciwstawne 81
– sprzeczne 93
– warunkowe 90
zdarzenia niezależne 531
– parami niezależne 532
zdarzenie 283
– elementarne 283
zliczanie permutacji 306
– podziałów uporządkowanych 308
złożenie funkcji 43
– relacji 663, 664
zmienna losowa 542
– –, dyskretna 547
– –, dystrybuanta 548
– –, unormowana 578
– –, wariancja 565
– –, wartość oczekiwana 556
– –, zbiór wartości 543
zmienna wolna 792, 794
– zdaniowa 91
zmienne losowe niezależne 545

Źródło 478

Do nabycia w księgarniach:

J. Bartoszewicz

Wykłady ze statystyki matematycznej

R. L. Graham, D. E. Knuth, O. Patashnik

Matematyka konkretna

K. Jänich

Topologia

H.-O. Peitgen, H. Jürgens, D. Saupe

Granice chaosu. Fraktale. Cz. 2

W. Rudin

Podstawy analizy matematycznej

I. Stewart

Czy Bóg gra w kości?

M. M. Sysło, N. Deo, J. S. Kowalik

Algorytmy optymalizacji dyskretnej

z programami w języku Pascal

Książki PWN są do nabycia w księgarniach własnych PWN: **Warszawa**, ul. Miodowa 10; **Katowice**, ul. 3 Maja 12; **Kraków**, ul. Św. Tomasza 30; **Łódź**, ul. Więckowskiego 13; **Poznań**, ul. Wodna 8/9; **Wrocław**, ul. Kuźnicza 56.

Księgarnia Wysyłkowa PWN

ul. Miodowa 10, 00-251 **Warszawa**, tel. 695 41 74 lub 635 09 77, fax 695 41 52, przyjmuje zamówienia telefoniczne i pisemne.

Alfabet grecki

A	α	alfa
B	β	beta
Γ	γ	gamma
Δ	δ	delta
E	ε	epsilon
Z	ζ	dzeta
H	η	eta
Θ	θ ϑ	theta
I	ι	jota
K	κ	kappa
Λ	λ	lambda
M	μ	mi
N	ν	ni
Ξ	ξ	ksi
O	o	omikron
Π	π	pi
P	ρ	ro
Σ	σ	sigma
T	τ	tau
Υ	υ	ypsilon
Φ	φ	fi
X	χ	chi
Ψ	ψ	psi
Ω	ω	omega

Algorytmy

Algorytm Dijkstry **504**
Algorytm Dijkstry ze wskaźnikami **519**
Algorytm DRZEWO **382**
Algorytm dzielenia **201**
Algorytm ETYKIETOWANIE **443**
Algorytm ETYKIETOWANIE DRZEWA **435**
Algorytm EUKLIDES **424**
Algorytm EUKLIDES$^+$ **424**
Algorytm Euklidesa **264**
Algorytm Fleury'ego **346**
Algorytm HUFFMAN **463**
Algorytm INORDER **431**
Algorytm Kruskala **386**
Algorytm LAS **382**
Algorytm NUMEROWANIE WIERZCHOŁKÓW **481**
Algorytm NWD **259**
Algorytm NWD$^+$ **263**
Algorytm POSTORDER **438**
Algorytm PREORDER **427**
Algorytm Prima **389**
Algorytm SORTOWANIE DRZEWA **438**
Algorytm TEST **413**
Algorytm TEST DRZEWA **415**
Algorytm TEST WPZ **417**
Algorytm UJŚCIE **480**
Algorytm WAGI MAKSYMALNE **513**
Algorytm Warshalla **509**
Algorytm Warshalla ze wskaźnikami **520**