

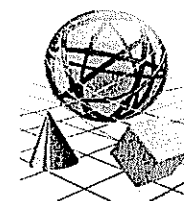
#### AUTORZY

MARCIN ANHOLCER (rozdziały 3, 12, 15.1, 15.2, 15.3)  
HELENA GASPAR-WIELOCH (rozdziały 6, 7)  
MARCIN GODLEWSKI (rozdziały 13, 14, 15.4)  
JERZY MARCINKOWSKI (rozdziały 4, 5)  
PRZEMYSŁAW STODULNY (rozdziały 10, 11)  
WOJCIECH SIKORA (wstęp, rozdziały 1, 2, 8, 9)

# BADANIA OPERACYJNE

redakcja naukowa

Wojciech Sikora



*Zmarłemu Profesorowi Edmundowi Ignasiakowi  
w podzięce za Jego inspiracje, opiekę i życzliwość  
Autorzy*

Polskie Wydawnictwo Ekonomiczne  
Warszawa 2008

BADANIA  
OPERACYJNE

Recenzent  
PROF. DR HAB. JULIUSZ SIEDLECKI

Okladkę projektował  
PAWEŁ ROSOLEK

Redaktor  
BOGUMILA FRÜBOES

Redaktor techniczny  
JOLANTA CZAPSKA



Copyright by Polskie Wydawnictwo Ekonomiczne S.A.  
Warszawa 2008

**W-149361**



009-000149361

BG SGGW  
Wyp. Stud.

ISBN 978-83-208-1743-0

Polskie Wydawnictwo Ekonomiczne S.A.  
ul. Canaletta 4, 00-099 Warszawa  
tel. (centrala) 0-22 827 80 01, faks 0-22 827 55 67  
e-mail: pwe@pwe.com.pl  
www.pwe.com.pl

Skład i łamanie: A. Cwił, tel. 0-22 644 99 55

## Spis treści

|  |    |
|--|----|
| Wstęp .....  | 9  |
| ROZDZIAŁ 1   |    |
| OPTYMALIZACJA LINIOWA .....  | 15 |
| 1.1. Modelowanie problemów decyzyjnych .....                           | 15 |
| 1.2. Przykłady liniowych zadań decyzyjnych .....                       | 17 |
| 1.3. Postacie zadań programowania liniowego .....                      | 21 |
| 1.4. Dualność .....  | 23 |
| 1.5. Metoda geometryczna .....   | 27 |
| 1.6. Metoda simpleks .....   | 29 |
| 1.7. Analiza postoptymalizacyjna .....                                 | 38 |
| ROZDZIAŁ 2   |    |
| ZAGADNIENIA TRANSPORTOWE .....   | 44 |
| 2.1. Zamknięte zagadnienie transportowe .....                          | 44 |
| 2.2. Metoda potencjałów dla zamkniętego zagadnienia transportowego ..  | 45 |
| 2.3. Zagadnienie pośrednika .....                                      | 49 |
| 2.4. Zagadnienie transportowe z ograniczoną przepustowością tras ..... | 53 |
| 2.5. Zagadnienie transportowe z kryterium czasu .....                  | 57 |
| ROZDZIAŁ 3   |    |
| PRZEPIŁY WY W SIECIACH .....   | 61 |
| 3.1. Sieci .....   | 61 |
| 3.2. Zagadnienie maksymalnego przepływu .....                          | 62 |
| 3.3. Zagadnienie najkrótszej drogi .....                               | 74 |
| 3.4. Zagadnienie przepływu o minimalnym koszcie .....                  | 78 |

|   |     |
|---|-----|
| ROZDZIAŁ 4  |     |
| <b>OPTYMALIZACJA WIELOKRYTERIALNA</b> . . . . .   |     |
| 4.1. Matematyczny model sytuacji decyzyjnej w przypadku istnienia wielu kryteriów oceny . . . . .           | 83  |
| 4.2. Optymalność w programowaniu wielokryterialnym . . . . .  | 85  |
| 4.3. Postępowanie w przypadku nieporównywalności decyzji . . . . .  | 89  |
| 4.4. Liniowe problemy wielokryterialnego podejmowania decyzji . . . . .                                     | 105 |
| ROZDZIAŁ 5  |     |
| <b>PROGRAMOWANIE NIELINIOWE</b> . . . . .   |     |
| 5.1. Wprowadzenie . . . . .   | 115 |
| 5.2. Własności problemów programowania nieliniowego . . . . .   | 118 |
| 5.3. Klasyfikacja problemów programowania nieliniowego . . . . .  | 122 |
| 5.4. Problem Kuhna-Tuckera . . . . .  | 124 |
| 5.5. Algorytmy wyznaczania ekstremum lokalnego . . . . .  | 128 |
| 5.6. Wybrane problemy programowania nieliniowego sprowadzalne do problemu programowania liniowego . . . . . | 136 |
| ROZDZIAŁ 6  |     |
| <b>PROGRAMOWANIE DYNAMICZNE</b> . . . . .   |     |
| 6.1. Istota programowania dynamicznego. Zasada optymalności Bellmana . . . . .                              | 142 |
| 6.2. Zagadnienie optymalnego rozdziału zasobu . . . . .   | 145 |
| ROZDZIAŁ 7  |     |
| <b>ANALIZA SIECIOWA PRZEDSIĘWZIĘĆ</b> . . . . .   |     |
| 7.1. Sieciowe modele przedsięwzięć . . . . .  | 157 |
| 7.2. Metoda ścieżki krytycznej . . . . .  | 162 |
| 7.3. Technika oceny i kontroli programu . . . . .   | 170 |
| 7.4. Analiza czasowo-kosztowa . . . . .   | 174 |
| 7.5. Analiza zasobowa . . . . .   | 183 |
| ROZDZIAŁ 8  |     |
| <b>OPTYMALIZACJA DYSKRETNA</b> . . . . .  |     |
| 8.1. Wprowadzenie . . . . .   | 189 |
| 8.2. Zagadnienie przydziału . . . . .   | 190 |
| 8.3. Zagadnienie lokalizacji produkcji . . . . .  | 192 |
| 8.4. Zagadnienie wyznaczania harmonogramu realizacji prac . . . . .   | 195 |
| 8.5. Zagadnienie komiwojażera . . . . .   | 197 |
| 8.6. Zagadnienie rozwózki . . . . .   | 199 |
| 8.7. Idea metody podziału i ograniczeń . . . . .  | 201 |
| 8.8. Metoda podziału i ograniczeń dla zadania PCL . . . . .   | 204 |

|   |     |
|---|-----|
| 8.9. Algorytm Little'a dla zagadnienia komiwojażera . . . . .             | 208 |
| 8.10. Algorytm Johnsona dla ustalenia kolejności obróbki detali . . . . . | 212 |

## ROZDZIAŁ 9

### PROGRAMOWANIE W WARUNKACH RYZYKA

|  |     |
|--|-----|
| 9.1. Wprowadzenie . . . . .  | 216 |
| 9.2. Zagadnienie ogrodnika — ryzyko decyzji . . . . .                    | 217 |
| 9.3. Zagadnienie gazeciarza . . . . .                                    | 221 |
| 9.4. Zagadnienie ustalania optymalnej liczby części zamiennych . . . . . | 226 |
| 9.5. Systemy masowej obsługi . . . . .                                   | 230 |

## ROZDZIAŁ 10

### DRZEWA DECYZYJNE

|   |     |
|---|-----|
| 10.1. Wprowadzenie . . . . .  | 237 |
| 10.2. Struktura drzew decyzyjnych . . . . .   | 238 |
| 10.3. Maksymalizacja oczekiwanej korzyści (minimalizacja oczekiwanego kosztu) . . . . . | 240 |
| 10.4. Decyzje sekwencyjne w warunkach niepewności . . . . .                             | 242 |
| 10.5. Problem pozyskania dodatkowej informacji . . . . .                                | 247 |
| 10.6. Maksymalizacja oczekiwanej użyteczności . . . . .                                 | 251 |

## ROZDZIAŁ 11

### PROGRAMOWANIE W WARUNKACH NIEPEWNOŚCI

|                                      |     |
|--------------------------------------|-----|
| 11.1. Wprowadzenie . . . . .         | 256 |
| 11.2. Gry z naturą . . . . .         | 257 |
| 11.3. Elementy teorii gier . . . . . | 264 |

## ROZDZIAŁ 12

### ZARZĄDZANIE ZAPASAMI

|  |     |
|--|-----|
| 12.1. Koszty zapasów . . . . .   | 278 |
| 12.2. Klasyczny model EOQ i jego rozszerzenia . . . . .                            | 279 |
| 12.3. Polityka zamawiania w przypadku zapotrzebowania zmiennego w czasie . . . . . | 288 |

## ROZDZIAŁ 13

### OPTYMALIZACJA DECYZJI INWESTYCYJNYCH

|  |     |
|--|-----|
| 13.1. Wprowadzenie . . . . .   | 293 |
| 13.2. Pomiar zysku i ryzyka inwestycji w papiery wartościowe . . . . . | 293 |
| 13.3. Portfel papierów wartościowych . . . . .                         | 298 |

|  |            |
|--|------------|
| 13.4. Kryteria wyboru optymalnego portfela .....                       | 305        |
| 13.5. Krótka sprzedaż jako przykładowe rozwinięcie analizy portfelowej | 309        |
| ROZDZIAŁ 14  |            |
| <b>SZTUCZNE SIECI NEURONOWE .....</b>                                  | <b>312</b> |
| 14.1. Wprowadzenie .....   | 312        |
| 14.2. Budowa sztucznej sieci neuronowej .....                          | 313        |
| 14.3. Uczenie sieci neuronowej .....                                   | 320        |
| 14.4. Praktyczne aspekty stosowania sieci neuronowych .....            | 328        |
| ROZDZIAŁ 15  |            |
| <b>METODY HEURYSTYCZNE .....</b>                                       | <b>332</b> |
| 15.1. Wprowadzenie .....   | 332        |
| 15.2. Heurystyki lokalnych poszukiwań .....                            | 333        |
| 15.3. Algorytmy mrówkowe .....   | 343        |
| 15.4. Algorytmy genetyczne .....                                       | 351        |
| <b>Tablica. Dystrybuanta rozkładu normalnego .....</b>                 | <b>367</b> |
| <b>Indeks .....</b>  | <b>369</b> |

## Wstęp

### Decyzje ekonomiczne a badania operacyjne

Każda działalność ludzka, w tym gospodarcza, wiąże się z planowaniem, organizowaniem, realizacją i kontrolowaniem wykonania zbioru działań. Działalność ta jest podporządkowana określonej celowi (celom) i wymaga dysponowania zasobami finansowymi i materialnymi oraz zasobem informacji. Następstwem obserwowanych obecnie zmian gospodarczych jest zwiększająca się złożoność działań. Klasyczne sposoby zarządzania, oparte na intuicji i doświadczeniu, stają się coraz mniej efektywne. W związku z czym obserwuje się wzrost zapotrzebowania na metody usprawniające ten proces. Możliwości usprawnień upatruje się w stosowaniu metod ilościowych.

**Badania operacyjne** można zdefiniować jako zespół modeli i metod poszukiwania optymalnych rozwiązań (tj. decyzji najlepiej realizujących preferencje decydenta) w danych warunkach ekonomicznych (sytuacjach decyzyjnych). Ze względu na przedmiot badań zalicza się je do nauk o zarządzaniu, natomiast stosowane metody, których celem jest kwantyfikacja i obiektywizacja procesu decyzyjnego oraz wszechstronna analiza konsekwencji podjętych decyzji, wymagają zastosowania programowania matematycznego i informatyki. Jest to interdyscyplinarna dziedzina wiedzy próbująca łączyć podejście charakterystyczne dla nauk humanistycznych z metodami typowymi dla nauk przyrodniczych. Interdyscyplinarność ta jest zarówno siłą, jak i przyczyną trudności natury metodycznej i praktycznej.

Przedmiotem zainteresowania badań operacyjnych jest proces decyzyjny. Dokonując jego analizy, posługujemy się modelami i metodami matematycznymi (analitycznymi), heurystycznymi oraz symulacją komputerową. Koniecznością staje się operowanie dużymi zbiorami informacji, której zgromadzenie, magazynowanie, przetworzenie oraz dystrybucję ułatwia technika komputerowa.

## Zakres badań operacyjnych

Początki badań operacyjnych sięgają okresu przed II wojną światową. Za ich prekursorów uważa się Kantorowicza, jednego z pierwszych laureatów Nagrody Nobla w dziedzinie ekonomii i von Neumana, twórcę teorii gier i podstaw informatyki. Nazwa **badania operacyjne** przyjęła się podczas II wojny światowej i określała zbiór metod związanych z analizowaniem i planowaniem operacji wojskowych.

Badania operacyjne bardzo ściśle są związane z programowaniem matematycznym i teorią podejmowania decyzji. Trudno wyznaczyć wyraźne granice między tymi dyscyplinami. Różnice między nimi sprowadzają się do innego rozłożenia akcentów. Programowanie matematyczne zajmuje się głównie konstrukcją i analizą właściwości algorytmów rozwiązywania określonych klas problemów optymalizacyjnych, badania operacyjne — budową modeli różnych sytuacji decyzyjnych, a teoria podejmowania decyzji — wypracowaniem odpowiednich reguł decyzyjnych na podstawie analizy właściwości konkretnych (mniej lub bardziej sformalizowanych) modeli podejmowania decyzji. Niekiedy te trzy nazwy traktujemy jako synonimy.

Badania operacyjne pozostają w bliskim związku z ekonomią matematyczną i ekonometrią — dziedzinami wiedzy, których przedmiotem zainteresowań jest ilościowy opis rzeczywistości gospodarczej. Ustalenia tych nauk w postaci zależności ilościowych (ilościowych praw ekonomicznych) wykorzystywane są do podejmowania optymalnych decyzji, głównie mikroekonomicznych. Wyróżnikiem i ambicją badań operacyjnych jest aktywne podejście do procesu decyzyjnego, jego analiza i usprawnianie.

## Metodologia postępowania

Proces modelowania możemy wyobrazić sobie jako badanie właściwości układu elementów o złożonych powiązaniach, który zachowuje się w sposób celowy. Warunkiem usprawnienia procesu decyzyjnego jest przeprowadzenie analizy i obiektywnej oceny podejmowanych decyzji. W tym celu wypracowana została metodologia postępowania charakterystyczna dla badań operacyjnych.

Sprowadza się ona do następujących etapów:

- 1) sformułowanie problemu decyzyjnego, czyli sporządzenie uproszczonego (ograniczonego do istotnych elementów) słownego opisu fragmentu interesującej nas rzeczywistości gospodarczej (sytuacji decyzyjnej);
- 2) budowa modelu matematycznego sytuacji decyzyjnej;
- 3) wybór lub opracowanie odpowiedniego algorytmu wyznaczania rozwiązania optymalnego lub przybliżonego problemu optymalizacyjnego;
- 4) pozyskanie i przetworzenie informacji niezbędnej do ustalenia (oszacowania) wartości parametrów problemu optymalizacyjnego;

- 5) implementacja i przetestowanie algorytmu oraz rozwiązania zadania optymalizacyjnego dla danych rzeczywistych;
- 6) analiza wrażliwości rozwiązania zadania optymalizacyjnego;
- 7) weryfikacja modelu;
- 8) wdrożenie rozwiązania.

Proces wspomaganie decyzji nie sprowadza się do wykonania prostej sekwencji wymienionych kroków. Zazwyczaj skonstruowany na początku model nieadekwatnie opisuje rzeczywistość, z czego wynika potrzeba (nieraz wielokrotnego) powrotu do drugiego kroku.

Najtrudniejsze jest wdrożenie rozwiązania. O końcowym sukcesie decyduje gotowość decydenta do zracjonalizowania swoich działań i przeprowadzenia — niekiedy trudnych — zmian organizacyjnych.

## Typy modeli i zadań

Decyzje możemy podejmować w warunkach deterministycznych (pewności) lub niedeterministycznych (ryzyka i niepewności). Podział ten prowadzi do wyróżnienia modeli:

- 1) deterministycznych,
- 2) niedeterministycznych
  - stochastycznych,
  - podejmowania decyzji w warunkach niepewności.

O **modelu deterministycznym** mówimy wówczas, gdy podejmujemy decyzje w warunkach pewności. Zakładamy, że parametry modelu są znane i stałe. Oznacza to, że rozwiązanie optymalne — jeżeli model realistycznie odzwierciedla proces decyzyjny — można zasadniczo utożsamiać z decyzją optymalną. Uzasadnione jest także — jeżeli nie wiąże się to z nadmiernie wysokimi kosztami — wyznaczenie rozwiązania optymalnego.

O **modelu stochastycznym** mówimy wtedy, gdy podejmujemy decyzje w warunkach ryzyka. Zakładamy, że niektóre parametry modelu są **zmiennymi losowymi o znanym rozkładzie prawdopodobieństwa**. Wynik decyzji jest wypadkową działań podjętych przez decydenta i czynników losowych. Nie można wówczas utożsamiać rozwiązania optymalnego z decyzją optymalną, gdyż w momencie podejmowania decyzji nie wiemy, jakie wartości przyjmują parametry modelu, które są realizacjami zmiennych losowych. Nie zawsze wydaje się zasadne ponoszenie zbyt dużych kosztów na poszukiwanie rozwiązania optymalnego.

O **podejmowaniu decyzji w warunkach niepewności** mówimy, gdy parametry modelu mogą przyjmować różne wartości w zależności od tego, jaki wystąpi stan otoczenia (natury). Prawdopodobieństwa wystąpienia tych stanów nie są znane.

Często podejmując decyzję, musimy brać także pod uwagę intencjonalne działania partnera (przeciwnika). Proces decyzyjny można w tym przypadku opisać za pomocą modeli **teorii gier**, której twórcami byli von Neuman, Morgenstern i Nash.

Możemy podzielić decyzje na operacyjne i strategiczne. Kryterium jest horyzont czasu. Podział ten prowadzi do wyróżnienia modeli operacyjnych oraz strategicznych.

O **modelach operacyjnych** mówimy wówczas, gdy podejmujemy decyzje o krótkim horyzoncie czasu, zazwyczaj cechujące się dużym stopniem powtarzalności.

O **modelach strategicznych** mówimy wtedy, gdy podejmujemy decyzje o dalekosiężnych następstwach. Konstrukcja tych modeli napotyka wiele trudności związanych z kwantyfikacją słabo ustrukturalizowanego procesu decyzyjnego. Kwantyfikacja pozwala na systematyczny przegląd możliwych do podjęcia działań i umożliwia — często niestety przybliżoną — analizę konsekwencji rozpatrywanych decyzji. Analiza wrażliwości pozwala na lepszy wgląd w dynamikę procesu decyzyjnego i ocenę następstw podjęcia decyzji bardziej precyzyjną niż w przypadku metod czysto intuicyjnych.

Problemy decyzyjne (optymalizacyjne) możemy podzielić ze względu na liczbę kryteriów oceny na problemy optymalizacji jednokryterialnej oraz problemy optymalizacji wielokryterialnej.

Problemy **optymalizacji jednokryterialnej** opisują sytuację, w której decydent przy podejmowaniu decyzji kieruje się jednym kryterium. Są to sytuacje typowe dla działań rutynowych. W przypadku decyzji strategicznych, słabo ustrukturalizowanych, podejmowanych przez ciała kolegialne, trzeba wykorzystać problemy **optymalizacji wielokryterialnej**.

Ze względu na typ relacji zachodzących między wielkościami, na które decydent ma wpływ (zmiennymi), wyróżniamy problemy liniowe oraz problemy nieliniowe.

Do niedawna najczęściej wykorzystywano **modele liniowe**. Rozwój teorii programowania nieliniowego oraz wzrastające możliwości obliczeniowe pozwalają na zastosowanie **modeli nieliniowych** umożliwiających bardziej realistyczne modelowanie złożonych procesów decyzyjnych.

Ze względu na typ zmiennych decyzyjnych wyróżniamy problemy ze zmiennymi ciągłymi i problemy ze zmiennymi dyskretnymi.

W przypadku zmiennych ciągłych wyznaczenie rozwiązania jest o wiele łatwiejsze aniżeli w przypadku zadań ze zmiennymi dyskretnymi.

Wybór postaci analitycznej modelu jest kompromisem między wiernością opisu przez model realiów procesu decyzyjnego a możliwościami wyznaczenia w rozsądnym czasie i przy akceptowalnych kosztach rozwiązania optymalnego. Niekiedy uzyskanie rozwiązania optymalnego jest niecelowe lub zbyt kosztowne — ograniczamy się wtedy do wyznaczenia rozwiązania przybliżonego (suboptymalnego).

Zastosowanie modelu w praktyce wymaga opracowania algorytmu pozwalają-

cego na wyznaczenie rozwiązania optymalnego lub przybliżonego. Jest to proces kosztowny i czasochłonny, co skłania — jeżeli jest to możliwe — do sprawdzenia poddawanego analizie zagadnienia do zadań, dla których takie algorytmy zostały opracowane i oprogramowane. W tej sytuacji zupełnie naturalny jest wzrost zainteresowania niealgorytmicznym podejściem do wyznaczania rozwiązań, czerpiącym inspirację z nauk biologicznych. Tym nowym kierunkom badań, tj. sztucznym sieciom neuronowym, algorytmom genetycznym i inteligencji roju, poświęcono ostatnie dwa rozdziały książki.

## Badania operacyjne i komputery

Szybki rozwój badań operacyjnych w drugiej połowie XX w. możliwy był dzięki rozwojowi informatyki. Praktyczne problemy, które cechują się dużą liczbą zmiennych decyzyjnych i parametrów, nie mogą być rozwiązywane ręcznie — konieczne jest użycie właściwego sprzętu i oprogramowania.

Odpowiednie aplikacje są często tworzone specjalnie na zamówienie klienta i w dużym stopniu dostosowane do jego dokładnie zdefiniowanych potrzeb. Do rozwiązywania najczęściej spotykanych problemów można jednak używać oprogramowania dostępnego na rynku.

Powszechnie dostępnym narzędziem jest Solver, będący składową MS Excel. Umożliwia on rozwiązywanie zadań programowania liniowego i nieliniowego o stosunkowo dużych rozmiarach. Niestety nie najlepiej radzi on sobie z programowaniem całkowitoliczbowym i binarnym. Przykładem innej aplikacji tej samej firmy jest MS Project wspomagający zarządzanie projektami. Stosunkowo dobrze rozbudowany moduł związany z badaniami operacyjnymi (*Operations Research* — OR) znaleźć można w pakiecie SAS, będącym bardzo rozbudowaną aplikacją do gromadzenia i analizy danych. Moduły OR są również obecne w zintegrowanym systemie zarządzania przedsiębiorstwem SAP.

## Układ pracy

Książka, którą oddajemy do rąk Czytelnika, jest poświęcona badaniom operacyjnym. Jest to w zamyśle autorów przystępny podręcznik dla studentów wydziałów zarządzania i ekonomii uczelni ekonomicznych, technicznych, uniwersytetów oraz licznych niepaństwowych szkół biznesu. Może także zainteresować praktyków pragnących stosować metody badań operacyjnych w procesie podejmowania decyzji. Dla zrozumienia treści książki wystarczy znajomość matematyki na poziomie wykładów w uczelniach ekonomicznych.

Podręcznik składa się z 15 rozdziałów, z których każdy jest poświęcony określonej klasie zagadnień oraz sposobom ich rozwiązywania. Po opisie zagad-

nienia jest sformułowany jego model matematyczny oraz podana metoda rozwiązania. Rozważania teoretyczne są ilustrowane przykładami liczbowymi. Każdy rozdział kończy się bibliografią oraz pytaniami kontrolnymi, pozwalającymi Czytelnikowi ocenić stopień przyswojenia materiału.

Książka powinna być czytana w całości kolejnymi rozdziałami, można jednak także studiować wybrane rozdziały z osobna.

Oddany Państwu do studiowania podręcznik jest cennym uzupełnieniem dwóch prac z zakresu badań operacyjnych wydanych w ostatnich latach przez PWE: książki T. Trzaskalika, *Wprowadzenie do badań operacyjnych z komputerem*, (2003 r.) oraz książki *Decyzje menedżerskie z Excelem*, pod redakcją T. Szapiro (2000 r.).

Autorami niniejszego podręcznika są pracownicy Katedry Badań Operacyjnych Akademii Ekonomicznej w Poznaniu. Książka nawiązuje układem i stylem do podręcznika *Badania operacyjne*, pod redakcją E. Ignasiaka, wydanego trzykrotnie przez PWE (w latach 1996, 1997, 2001).

Przedwcześnie zmarły w 2003 r. Profesor Ignasiak był nie tylko współautorem i redaktorem tego podręcznika, ale naszym nauczycielem i przelozonym. Książkę tę dedykujemy pamięci Profesora Ignasiaka.

# ROZDZIAŁ 1

## OPTIMALIZACJA LINIOWA



### 1.1. Modelowanie problemów decyzyjnych

W wielu różnych sytuacjach podejmujemy decyzje. Sytuacje te nazywamy sytuacjami decyzyjnymi, a osobę podejmującą decyzję — **decydentem**. Warunki, w jakich działa decydent, nie pozwalają na wybór dowolnej decyzji. Decyzję zgodną z warunkami ograniczającymi nazywa się **decyzją dopuszczalną**.

Nie każda decyzja dopuszczalna jest równie dobra. W świetle celów, jakie sobie stawia decydent, jedne decyzje mogą być lepsze, inne gorsze. Stąd problem wyboru decyzji najlepszej, zwanej **decyzją optymalną**. Wybór decyzji optymalnej wymaga przyjęcia określonego kryterium, według którego oceniamy decyzje jako lepsze lub gorsze. Kryterium to nazywamy **kryterium wyboru** (oceny).

#### PRZYKŁAD 1.1

Możemy podjąć jedną z trzech decyzji inwestycyjnych. Nakłady inwestycyjne oraz oczekiwany roczny zysk osiągnięty z tych inwestycji pokazano w tablicy 1.1. Która z trzech decyzji jest optymalna?

TABLICA 1.1  
Nakłady i zyski dla trzech decyzji

| Decyzje                       | A  | B  | C  |
|-------------------------------|----|----|----|
| Nakłady inwestycyjne (mln zł) | 40 | 50 | 30 |
| Zyski (mln zł)                | 8  | 4  | 6  |

Na pytanie to nie możemy odpowiedzieć, gdyż nie zostało przyjęte żadne kryterium wyboru. Jeżeli kryterium oceny będzie minimalizacja nakładów, to



najlepszą decyzją jest decyzja  $C$ . Jeśli natomiast jako kryterium wyboru przyjmujemy maksymalizację zysku, to najlepszą decyzją jest decyzja  $A$ . W przypadku maksymalizacji stopy zysku najlepszymi decyzjami są  $A$  i  $C$ .

Opis określonej sytuacji decyzyjnej nazywamy **problemem (zagadnieniem) decyzyjnym**. Dalej będziemy rozpatrywać tylko takie sytuacje, w których warunki ograniczające, kryterium wyboru i decyzje dają się opisać w języku matematycznym. Zapis problemu decyzyjnego w języku matematycznym to formułowanie **modelu matematycznego**. Model matematyczny problemu decyzyjnego nazywamy **zadaniem decyzyjnym**.

Warunki ograniczające najczęściej są opisywane za pomocą układów równań lub nierówności. W równaniach tych (lub nierównościach) występować będą pewne wielkości dane, zwane **parametrami**, oraz wielkości, które należy ustalić, zwane **zmiennymi decyzyjnymi**.

Oprócz warunków ograniczających, w zadaniu decyzyjnym mogą także występować **warunki znaku zmiennych** (np. warunek nieujemności) lub typu zmiennych (np. warunek ich ciągłości, całkowitoliczbowości lub binarności).

Decyzje dopuszczalne utożsamiać będziemy z takim układem wartości zmiennych (układem liczb), które spełniają wszystkie warunki opisujące badaną sytuację. Rolę kryterium wyboru będzie pełnił pewna funkcja zmiennych decyzyjnych mierząca cel, który chce osiągnąć decydent. Funkcję tą nazywa się **funkcją celu**.

Wybór decyzji optymalnej polega na ustaleniu takiej decyzji dopuszczalnej, przy której funkcja celu osiąga wartość najkorzystniejszą, tzn. w zależności od badanej sytuacji wartość minimalną lub maksymalną.

Jeżeli przez  $D$  oznaczymy zbiór dopuszczalnych decyzji, przez  $x$  dowolną decyzję, a przez  $f$  — funkcję celu, to zadanie decyzyjne można zapisać następująco:

znajdź taką decyzję dopuszczalną  $x^* \in D$ , że:  
 $f(x^*) = \max \{f(x) | x \in D\}$ , gdy maksymalizujemy funkcję celu,  
 $f(x^*) = \min \{f(x) | x \in D\}$ , gdy minimalizujemy funkcję celu.

Częściej (ale mniej dokładnie) zapisuje się to w postaci:

$$f(x) \rightarrow \max \quad \left| \quad f(x) \rightarrow \min \right. \\ \text{dla } x \in D \quad \left| \quad \text{dla } x \in D.$$

Opisanie sytuacji decyzyjnej w języku matematycznym ma na celu sprowadzenie problemu wyboru najlepszej decyzji do rozwiązania pewnego jednoznacznie określonego zadania matematycznego. Aby rozwiązanie takiego zadania rzeczywiście pozwoliło ustalić najlepszą decyzję, trzeba je tak sformułować, żeby dokładnie opisywało daną sytuację decyzyjną.

W tym celu należy ustalić:

- 1) jakie wielkości mają być wyznaczone i odpowiednio je oznaczyć (tzn. należy podać **zmiennie decyzyjne**);

- 2) jakie wielkości są dane (określić **parametry zadania**);
- 3) jakie ograniczenia musi spełnić dopuszczalna decyzja i sformułować je w postaci równań lub nierówności, wiążąc zmiennie decyzyjne (zapisać **warunki ograniczające**);
- 4) cel, jaki chce osiągnąć decydent oraz sformułować funkcję zmiennych decyzyjnych określającą poziom realizacji celu (podać **funkcję celu**).

Jeżeli w zadaniu decyzyjnym funkcja celu oraz warunki ograniczające są liniowe, to zadanie takie nazywamy **liniowym zadaniem decyzyjnym (LZD)**.

## 1.2. Przykłady liniowych zadań decyzyjnych

### 1.2.1. Zagadnienie diety

W pewnej stolówce studenckiej mamy do dyspozycji  $n$  różnych produktów spożywczych  $P_1, P_2, \dots, P_n$ , którymi można żywić studentów. Z produktów tych dietetyk ma sporządzić dietę, tzn. ma określić, ile każdego produktu ma dziennie spożyć jeden student. Wymaga się, aby dzienna dawka żywności dostarczała określoną ilość takich składników odżywczych jak białko, węglowodany, tłuszcze, witaminy, sole mineralne itp. Weźmy pod uwagę  $m$  składników odżywczych  $S_1, S_2, \dots, S_m$ . Załóżmy, że każdy student powinien otrzymać co najmniej  $b_1$  jednostek składnika  $S_1$ ,  $b_2$  jednostek składnika  $S_2$ ,  $b_m$  jednostek składnika  $S_m$ . Ponieważ nadmierne spożycie niektórych składników może być szkodliwe dla zdrowia, stąd żądamy, aby zjadano ich nie więcej niż  $d_1, d_2, \dots, d_m$ . Znamy zawartość poszczególnych składników w każdym produkcie. Oznaczmy przez  $a_{ij}$  ilość  $i$ -tego składnika w jednostce  $j$ -tego produktu. Informacje te przedstawiono w tablicy 1.2.

Chcemy także, aby dieta była urozmaicona, tzn. aby nie ograniczała się do kilku produktów (np. chleba, kapusty i sera). Można to uzyskać, określając minimalne i maksymalne ilości poszczególnych produktów w diecie. Znamy także cenę każdego produktu. Oznaczmy przez:  $c_j$  — cenę  $j$ -tego produktu,  $g_j$  — minimalną ilość  $j$ -tego produktu w diecie,  $h_j$  — maksymalną ilość  $j$ -tego produktu w diecie.

Należy ustalić taki plan żywienia studentów, aby koszt diety był minimalny, przy spełnieniu wszystkich postawionych ograniczeń.

Zanim wyznaczmy optymalną dietę, musimy sformułować liniowe zadanie decyzyjne opisujące powyższą sytuację decyzyjną. Zmiennymi decyzyjnymi, czyli wielkościami do ustalenia, będą ilości poszczególnych produktów w diecie. Oznaczmy przez  $x_j$  ilość  $j$ -tego produktu w diecie.

Składnika  $S_1$  w produkcie  $P_1$  dostarczymy  $a_{11}x_1$  jednostek, w produkcie  $P_2$  —  $a_{12}x_2$  jednostek itd. Ogółem we wszystkich produktach zawartych w diecie

TABLICA 1.2  
Macierz zawartości  
składników odżywczych

|       | $P_1$    | $P_2$    | ... | $P_n$    |
|-------|----------|----------|-----|----------|
| $S_1$ | $a_{11}$ | $a_{12}$ | ... | $a_{1n}$ |
| $S_2$ | $a_{21}$ | $a_{22}$ | ... | $a_{2n}$ |
| ⋮     | ⋮        | ⋮        | ... | ⋮        |
| $S_m$ | $a_{m1}$ | $a_{m2}$ | ... | $a_{mn}$ |

student otrzyma  $a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n$  jednostek składnika  $S_1$ . Aby ta ilość spełniała stawiane ograniczenia, musi zachodzić:

$$b_1 \leq \sum_{j=1}^n a_{1j}x_j \leq d_1.$$

Ponieważ podobne ograniczenia dotyczą także pozostałych składników odżywczych, można to ogólnie zapisać w następujący sposób:

$$b_i \leq \sum_{j=1}^n a_{ij}x_j \leq d_i \quad (i = 1, 2, \dots, m).$$

Ilość każdego produktu w diecie nie może być mniejsza od ilości minimalnej oraz większa od ilości maksymalnej, co zapiszemy w postaci ograniczenia:

$$g_j \leq x_j \leq h_j \quad (j = 1, 2, \dots, n).$$

Koszt wyżywienia jednego studenta wynosi:

$$\sum_{j=1}^n c_j x_j.$$

**Model matematyczny.** Problem wyznaczania optymalnej diety sprowadza się do znalezienia takich wartości zmiennych decyzyjnych  $x_1, x_2, \dots, x_n$ , aby:

$$\sum_{j=1}^n c_j x_j \rightarrow \min, \quad (1.1)$$

przy ograniczeniach

$$b_i \leq \sum_{j=1}^n a_{ij}x_j \leq d_i \quad (i = 1, \dots, m), \quad (1.2)$$

$$g_j \leq x_j \leq h_j \quad (j = 1, \dots, n). \quad (1.3)$$

Jest to liniowe zadanie decyzyjne, gdyż funkcja celu (1.1) oraz warunki (1.2)–(1.3) są liniowe. Rozwiązanie spełniające warunki (1.2)–(1.3) jest rozwiązaniem dopuszczalnym problemu, a rozwiązanie spełniające także warunek (1.1) — rozwiązaniem optymalnym. Ponieważ w zadaniu (1.1)–(1.3) wszystkie zmienne są ciągłe, stąd zadanie jest zadaniem programowania liniowego (PL).

### 1.2.2. Zagadnienie ustalania planu działania

W zakładzie można produkować  $n$  wyrobów. Do produkcji tych wyrobów potrzeba  $m$  rodzajów środków. Zakład dysponuje określonym zasobem każdego środka produkcji. Wiemy, jaka jest norma zużycia każdego środka produkcji. Niektóre środki (jeżeli jest ich zbyt dużo) można łatwo sprzedać. Także brakujące ilości tych środków można dodatkowo zakupić. Znamy cenę zakupu oraz sprzedaży każdego środka „mobilnego”.

Dla każdego „mobilnego” środka produkcji i każdego wyrobu znamy maksymalny popyt. Należy ustalić taki plan działania dla zakładu, aby osiągnięty dochód był maksymalny.

Przez dochód zakładu rozumiemy różnicę między przychodem ze sprzedaży wyprodukowanych wyrobów oraz zbędnych środków a wydatkami poniesionymi na dodatkowy zakup potrzebnych środków.

Niech  $M$  będzie zbiorem numerów środków „mobilnych”, czyli takich, które można dodatkowo sprzedać lub kupić, a  $N$  — zbiorem numerów pozostałych środków „niemobilnych”, których ilości nie można zmienić.

Parametry modelu matematycznego będą następujące:

- $a_{ij}$  — ilość  $i$ -tego środka produkcji zużyta na wykonanie jednostki  $j$ -tego wyrobu,
- $b_i$  — posiadany zasób  $i$ -tego środka produkcji,
- $c_j$  — cena sprzedaży jednostki  $j$ -tego wyrobu,
- $p_i$  — cena sprzedaży jednostki  $i$ -tego środka,
- $k_i$  — cena zakupu jednostki  $i$ -tego środka,
- $g_j$  — maksymalna ilość  $j$ -tego wyrobu, jaką można sprzedać,
- $d_i$  — maksymalna ilość  $i$ -tego środka, jaką można sprzedać.

Zakładamy ponadto, że cena sprzedaży  $i$ -tego środka musi być mniejsza od ceny jego zakupu:

$$p_i < k_i \quad (i \in M).$$

Zmiennymi decyzyjnymi w tym zagadnieniu są:

- $x_j$  — wielkość produkcji (sprzedaży)  $j$ -tego wyrobu,
- $y_i$  — wielkość sprzedaży  $i$ -tego środka „mobilnego”,
- $z_i$  — wielkość zakupu  $i$ -tego środka „mobilnego”.

Model matematyczny zagadnienia wyboru optymalnego planu działania sprowadza się do znalezienia takich wartości zmiennych  $x_j$ ,  $y_i$ ,  $z_i$ , aby:

$$\sum_{j=1}^n c_j x_j + \sum_{i \in M} p_i y_i - \sum_{i \in M} k_i z_i \rightarrow \max, \quad (1.4)$$

przy ograniczeniach:

$$\sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i \in N), \quad (1.5)$$

$$\sum_{j=1}^n a_{ij} x_j + y_i - z_i \leq b_i \quad (i \in M), \quad (1.6)$$

$$0 \leq x_j \leq g_j \quad (j = 1, \dots, n), \quad (1.7)$$

$$0 \leq y_i \leq d_i \quad (i \in M), \quad (1.8)$$

$$z_i \geq 0 \quad (i \in M). \quad (1.9)$$

Zadanie (1.4)–(1.9) jest liniowym zadaniem decyzyjnym (LZD), gdyż zarówno funkcja celu (1.4) jest liniowa, jak i warunki ograniczające (1.5)–(1.9) są nierównościami liniowymi. W tym zadaniu wszystkie zmienne są ciągłe, a więc jest to zadanie programowania liniowego (PL).

Funkcja celu (1.4) postuluje, aby dochód — będący różnicą między przychodem ze sprzedaży wyrobów oraz zbędnych środków produkcji a wydatkami poniesionymi na zakup dodatkowych środków — był maksymalny. Warunek (1.5) zapewnia, że dla każdego środka „niemobilnego” zużycie nie przekroczy posiadanego zasobu. Natomiast warunek (1.6) gwarantuje, że dla środka „mobilnego” łączne jego rozdysponowanie nie przekroczy ilości, jaką będziemy dysponowali.

Warunki (1.7) i (1.8) zapewniają, że nie wyprodukujemy więcej wyrobów oraz nie sprzedamy więcej zbędnych środków niż jest to możliwe ze względu na ograniczenia popytowe.

W opisie sytuacji przyjmuje się, że jeżeli następuje sprzedaż zewnętrznego środka produkcji, to równocześnie się go nie kupuje, i na odwrót. Przy założeniu, że cena sprzedaży zewnętrznego środka jest niższa od ceny jego zakupu, model (1.4)–(1.9) prawidłowo opisuje daną sytuację decyzyjną. Jeżeli to założenie nie będzie spełnione, w zadaniu (1.4)–(1.9) będzie opłacalny zakup środków w celu ich dalszej odsprzedaży. Aby to wykluczyć, należy dodatkowo żądać:

$$y_i z_i = 0 \quad (i \in M) \quad (1.10)$$

Zadanie zawierające nieliniowy warunek (1.10) jest już nieliniowym zadaniem decyzyjnym (NZD).

### 1.3. Postacie zadań programowania liniowego

Formułowane w poprzednim podrozdziale modele matematyczne były liniowymi zadaniami decyzyjnymi. Jeżeli w zadaniu decyzyjnym wszystkie relacje są liniowe oraz wszystkie zmienne są ciągłe, to takie zadanie nazywamy **zadaniem programowania liniowego (PL)**.

Ogólna postać zadania PL jest następująca:

$$\sum_{j=1}^n c_j x_j \rightarrow \max (\min), \quad (1.11)$$

$$\sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i = 1, 2, \dots, m), \quad (1.12)$$

$$\sum_{j=1}^n a_{ij} x_j \geq b_i \quad (i = m + 1, \dots, p), \quad (1.13)$$

$$\sum_{j=1}^n a_{ij} x_j = b_i \quad (i = p + 1, \dots, r), \quad (1.14)$$

$$x_j \geq 0 \quad (j = 1, 2, \dots, n), \quad (1.15)$$

gdzie  $n_1 \leq n$ .

Każdy wektor zmiennych decyzyjnych  $x = (x_1, x_2, \dots, x_n)$ , spełniający warunki ograniczające (1.11)–(1.15), nazywamy **rozwiązaniem dopuszczalnym** zadania PL. Rozwiązanie dopuszczalne, dla którego funkcja celu (1.26) osiąga maksimum, (minimum) nazywamy **rozwiązaniem optymalnym**.

Parametrami w tym zadaniu są  $c_j$ ,  $b_i$  oraz  $a_{ij}$  ( $i = 1, 2, \dots, r$ ;  $j = 1, 2, \dots, n$ ). Parametr  $c_j$  nazywamy  $j$ -tą wagą funkcji celu, parametr  $b_i$  —  $i$ -tym wyrazem wolnym, a parametr  $a_{ij}$  — współczynnikiem macierzy ograniczeń stojącym w  $i$ -tym wierszu i  $j$ -tej kolumnie.

Istotne znaczenie przy formułowaniu zadań PL, a także przy ich rozwiązywaniu mają dwie postacie szczególne — tzw. postać standardowa i postać kanoniczna.

Zadaniem PL o **postaci standardowej** nazywamy zadanie, w którym wszystkie ograniczenia są nierównościami typu  $\leq$  dla zadań na maksimum bądź nierównościami typu  $\geq$  dla zadań na minimum oraz wszystkie zmienne muszą być nieujemne.

Zadaniem PL o **postaci kanonicznej** nazywamy zadanie, w którym wszystkie warunki ograniczające są równaniami oraz na wszystkie zmienne są nałożone warunki nieujemności.

Zadaniem PL o postaci standardowej są więc zadania:

$$\begin{array}{l|l} \sum_{j=1}^n c_j x_j \rightarrow \max, & \sum_{j=1}^n c_j x_j \rightarrow \min, \\ \sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i = 1, 2, \dots, m), & \sum_{j=1}^n a_{ij} x_j \geq b_i \quad (i = 1, 2, \dots, m), \\ x_j \geq 0 \quad (j = 1, 2, \dots, n), & x_j \geq 0 \quad (j = 1, 2, \dots, n) \end{array}$$

Nierówność  $\leq$  dla zadania na maksimum oraz nierówność  $\geq$  dla zadania na minimum nazywamy **nierównościami typowymi**.

Podstawowym zabiegiem stosowanym przy rozwiązywaniu zadań decyzyjnych jest sprowadzanie ich do równoważnej, lecz wygodniejszej ze względów numerycznych postaci. Dla zadania PL rozwiązywanego metodą simpleks będzie to postać kanoniczna.

Aby uzyskać równoważną dla danego zadania postać kanoniczną, nierówności sprowadzamy do równości, wprowadzając tzw. **zmienną swobodną**. Z każdym warunkiem w postaci nierówności związana będzie właściwa temu warunkowi zmienna swobodna równoważąca jego stronę lewą z prawą. Zmienne te do funkcji celu wchodzi z **zerowymi wagami**.

Jeżeli w zadaniu PL jest nierówność typu  $\leq$ , to zmienną swobodną  $x_{n+i} \geq 0$  oznaczamy jako:

$$x_{n+i} = b_i - \sum_{j=1}^n a_{ij} x_j.$$

Jeżeli natomiast w zadaniu PL jest nierówność typu  $\geq$ , to zmienną swobodną  $x_{n+i} \geq 0$  definiujemy jako:

$$x_{n+i} = \sum_{j=1}^n a_{ij} x_j - b_i.$$

### PRZYKŁAD 1.2

Mamy następujące zadanie PL:

$$\begin{cases} -2x_1 + 4x_2 + 2x_3 \rightarrow \max, \\ 2x_1 + 3x_2 + x_3 \geq 3, \\ -4x_1 + 2x_2 + 3x_3 \leq 5, \\ x_1 \leq 0, x_3 \geq 0. \end{cases} \quad (1.16)$$

Aby zadanie to zapisać w równoważnej postaci kanonicznej, należy obydwie nierówności sprowadzić do równości oraz na wszystkie zmienne nałożyć warunki nieujemności. W tym celu wprowadzamy zmienne swobodne  $x_4$  i  $x_5$ , a także zmienne  $x'_1, x'_2, x''_2$  zamiast zmiennych  $x_1, x_2$ , przy czym  $x_1 = -x'_1, x_2 = x'_2 - x''_2$ .

Równoważną postać kanoniczną zapiszemy więc następująco:

$$\begin{cases} 2x'_1 + 4(x'_2 - x''_2) + 2x_3 + 0x_4 + 0x_5 \rightarrow \max, \\ -2x'_1 + 3(x'_2 - x''_2) + x_3 - x_4 + 0x_5 = 3, \\ 4x'_1 + 2(x'_2 - x''_2) + 3x_3 + 0x_4 + x_5 = 5, \\ x'_1, x'_2, x''_2, x_3, x_4, x_5 \geq 0. \end{cases}$$

Jeżeli mamy zadanie PL z funkcją celu:

$$\sum_{j=1}^n c_j x_j \rightarrow \max,$$

to równoważne mu będzie zadanie z funkcją celu:

$$\sum_{j=1}^n (-c_j) x_j \rightarrow \min.$$

## 1.4. Dualność

### 1.4.1. Reguły tworzenia zadania dualnego

Z każdym zadaniem PL sprzężone jest pewne inne zadanie PL, zwane **zadaniem dualnym**. Jeżeli zadaniem pierwotnym (ZP) jest zadanie:

$$\begin{cases} \sum_{j=1}^n c_j x_j \rightarrow \max, \\ \sum_{j=1}^n a_{ij} x_j \leq b_i \quad (i = 1, 2, \dots, m), \\ x_j \geq 0 \quad (j = 1, 2, \dots, n), \end{cases} \quad (1.17)$$

to zadaniem dualnym (ZD) będzie zadanie:

$$\begin{cases} \sum_{i=1}^m b_i y_i \rightarrow \min, \\ \sum_{i=1}^m a_{ij} y_i \geq c_j \quad (j = 1, 2, \dots, n), \\ y_i \geq 0 \quad (i = 1, 2, \dots, m). \end{cases} \quad (1.18)$$

Z relacji zachodzących między zadaniem pierwotnym a zadaniem dualnym wynika, że:

- 1) w zadaniu dualnym jest tyle zmiennych, ile nierówności w zadaniu pierwotnym (każdemu warunkowi ZP odpowiada jedna zmienna ZD);
- 2) w zadaniu dualnym jest tyle warunków, ile zmiennych w zadaniu pierwotnym;

- 3) wagi funkcji celu zadania pierwotnego są wyrazami wolnymi w zadaniu dualnym;
- 4) wyrazy wolne zadania pierwotnego są wagami funkcji celu w zadaniu dualnym;
- 5) macierz współczynników zadania dualnego jest transpozycją macierzy współczynników zadania pierwotnego;
- 6) jeżeli zadanie pierwotne jest na maksimum, to zadanie dualne jest na minimum, i odwrotnie.

Zauważmy ponadto, że jeżeli zadanie pierwotne jest w postaci standardowej, to zadanie dualne jest także w postaci standardowej. Jeżeli natomiast zadanie pierwotne nie jest w postaci standardowej, to dualne także nie będzie w postaci standardowej (ZD odbiega na tyle od postaci standardowej, na ile odbiega od niej ZP).

W przypadku ogólnym stosujemy ponadto następujące dodatkowe reguły tworzenia zadania dualnego:

- 1) jeżeli w ZP  $i$ -ty warunek jest równością, to odpowiadająca mu zmienna nie ma ograniczeń;
- 2) jeżeli w ZP  $i$ -ty warunek jest nietypową nierównością, to w ZD zmienna  $y_i \leq 0$ ;
- 3) jeżeli w ZP na zmienną  $x_j$  nie nałożono ograniczeń, to  $j$ -ty warunek ZD jest równością;
- 4) jeżeli w ZP zmienna  $x_j \leq 0$ , to w ZD  $j$ -ty warunek jest nietypową nierównością.

### PRZYKŁAD 1.3

Należy utworzyć zadanie dualne do zadania pierwotnego:

|                                 |                  |      |
|---------------------------------|------------------|------|
| zadanie pierwotne:              | zmienne dualne:  |      |
| $6x_1 + 8x_2 \rightarrow \max,$ | $y_1 \geq 0,$    | (ZP) |
| $4x_1 + 6x_2 \leq 10,$          | $y_2$ — dowolne, |      |
| $3x_1 + 2x_2 = 4,$              | $y_3 \leq 0.$    |      |
| $2x_1 + 2x_2 \geq 2,$           |                  |      |
| $x_1$ — dowolne, $x_2 \leq 0,$  |                  |      |

Zadanie dualne będzie miało trzy zmienne i dwa warunki ograniczające (poza warunkami znaku):

$$\begin{aligned} 10y_1 + 4y_2 + 2y_3 &\rightarrow \min, \\ 4y_1 + 3y_2 + 2y_3 &= 6, \\ 6y_1 + y_2 + 2y_3 &\geq 8, \\ y_1 &\geq 0, y_2 \text{ — dowolne}, y_3 \leq 0. \end{aligned}$$

## 1.4.2. Twierdzenia o dualności

Między rozwiązaniami zadania pierwotnego (1.17) i zadania dualnego (1.18) zachodzą zależności, które można sformułować w postaci następujących twierdzeń.

### O istnieniu

Jeżeli ZP i ZD mają rozwiązania dopuszczalne, to obydwa mają rozwiązania optymalne. Jeżeli natomiast chociaż jedno z nich nie ma rozwiązania dopuszczalnego, to obydwa nie mają rozwiązań optymalnych.

Jeżeli  $x_1, x_2, \dots, x_n$  jest rozwiązaniem dopuszczalnym zadania prymalnego, a  $y_1, y_2, \dots, y_m$  rozwiązaniem dopuszczalnym zadania dualnego, to między wartościami funkcji celu zachodzi nierówność:

$$\sum_{j=1}^n c_j x_j \leq \sum_{i=1}^m b_i y_i \quad (1.19)$$

Dla rozwiązań dopuszczalnych wartość funkcji celu ZP nie może być większa od wartości funkcji celu ZD.

### O optymalności

Jeżeli istnieją dwa takie rozwiązania dopuszczalne  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$  (ZP) i  $\bar{y}_1, \bar{y}_2, \dots, \bar{y}_m$  (ZD), że:

$$\sum_{j=1}^n c_j \bar{x}_j = \sum_{i=1}^m b_i \bar{y}_i, \quad (1.20)$$

to obydwa rozwiązania są rozwiązaniami optymalnymi.

### O równowadze

Jeżeli  $x_1, x_2, \dots, x_n$  jest rozwiązaniem dopuszczalnym ZP oraz  $y_1, y_2, \dots, y_m$  jest rozwiązaniem dopuszczalnym ZD, to — aby te rozwiązania były rozwiązaniami optymalnymi — wystarcza, że spełnione będą następujące warunki:

$$\sum_{j=1}^n a_{ij} x_j < b_i \Rightarrow y_i = 0, \quad (1.21)$$

$$\sum_{i=1}^m a_{ij}y_i > c_j \Rightarrow x_j = 0, \quad (1.22)$$

$$y_i > 0 \Rightarrow \sum_{j=1}^n a_{ij}x_j = b_i, \quad (1.23)$$

$$x_j > 0 \Rightarrow \sum_{i=1}^m a_{ij}y_i = c_j. \quad (1.24)$$

Z twierdzenia o równowadze korzystamy w celu sprawdzania optymalności znanego rozwiązania dopuszczalnego lub do znalezienia rozwiązania optymalnego dla przypadku szczególnego, gdy zadanie PL ma tylko dwa warunki ograniczające.

### 1.4.3. Interpretacja ekonomiczna zadania dualnego

Przyjmijmy, że zadanie pierwotne (1.17) opisuje problem maksymalizacji przychodu osiąganego z produkcji  $n$  wyrobów. Zużycie środków produkcji nie może przekroczyć zasobów, jakimi dysponujemy. Niech waga  $c_j$  oznacza cenę  $j$ -tego wyrobu, współczynnik  $a_{ij}$  — wielkość zużycia  $i$ -tego środka na produkcję jednostki  $j$ -tego wyrobu, wyraz wolny  $b_i$  — zasób  $i$ -tego środka produkcji, a zmienna  $x_j$  — wielkość produkcji  $j$ -tego wyrobu.

Aby nierówności w zadaniu (1.18) miały sens, zmienną  $y_i$  interpretujemy jako cenę  $i$ -tego środka produkcji. Załóżmy, że konkurent chce nabyć od producenta środki produkcji. Jaka cenę powinien zaoferować? Z pewnością chciałby odkupić środki produkcji najtaniej. Proponuje więc, aby suma  $\sum_{i=1}^m b_i y_i$ , czyli wartość funkcji celu zadania dualnego, była minimalna.

Konkurent musi się liczyć z faktem, że jeżeli zaoferuje producentowi zbyt niską cenę, to ten środków tych nie sprzeda. Cena za niska to taka, kiedy przychód ze sprzedaży środków produkcji byłby niższy od przychodu, jaki producent może uzyskać kierując je do produkcji. Gdyby producent sprzedał środki niezbędne do produkcji jednostki  $j$ -tego produktu po cenach  $y_i$  ( $i = 1, 2, \dots, m$ ), wówczas dostał-

by  $\sum_{i=1}^m a_{ij}y_i$ .

Oplaci się więc sprzedać środki, jeżeli:

$$\sum_{i=1}^m a_{ij}y_i \geq c_j \quad (j = 1, 2, \dots, n). \quad (1.25)$$

Warunek (1.25) to ograniczenie zadania dualnego. Zadanie dualne jest więc zadaniem, jakie powinien rozwiązać konkurent pragnący nabyć środki produkcji od

producenta, jeżeli chciałby działać racjonalnie i liczy na racjonalne zachowanie producenta.

Jak wynika z twierdzenia 1.3, optymalna wartość zmiennej  $y_i$  określa, o ile wzrośnie (zmniejszy się) przychód, jeżeli zwiększymy (zmniejszymy) zasób  $i$ -tego środka produkcji o jednostkę. Ten wniosek jest prawdziwy, gdy zmiany mieszczą się w dopuszczalnych granicach i dotyczą tylko jednego środka. Zmienna dualna określa więc zgodnie z neoklasyczną teorią ekonomii krańcową produktywność jednostki  $i$ -tego środka.

Jeżeli produktywność  $i$ -tego środka wyznaczona przez optymalne  $y_i$  równa się 10 zł, a cena  $c_i$  po jakiej nabywa producent  $i$ -ty środek wynosi 8 zł, to opłaci się tak zwiększyć zasób  $i$ -tego środka aż nastąpi zrównanie wartości  $y_i$  z wartością  $c_i$ . Jeżeli natomiast  $c_i$  wynosi 12 zł i po tej cenie można sprzedać jednostkę  $i$ -tego środka, to przy  $y_i$  równej 10 zł należy zmniejszyć zasób  $i$ -tego środka, gdyż więcej zyskamy przeznaczając jednostkę  $i$ -tego środka na sprzedaż niż do produkcji.

Dosyć oczywistą treść ekonomiczną mają w tej sytuacji warunki (1.21)–(1.24) twierdzenia o równowadze:

- 1) jeżeli zużycie  $i$ -tego środka produkcji jest mniejsze od posiadanego zasobu, to wycena (krańcowa produktywność) jednostki  $i$ -tego środka jest zerowa;
- 2) jeżeli wartość środków zużytych na wytworzenie jednostki  $j$ -tego produktu jest większa od jego ceny, to produkcja tego wyrobu jest zerowa;
- 3) jeżeli wycena  $i$ -tego środka jest dodatnia, to zużycie środka musi być równe jego zasobowi;
- 4) jeżeli produkcja  $j$ -tego wyrobu jest dodatnia, to wartość środków zużytych na jednostkę  $j$ -tego produktu jest równa jego cenie.

## 1.5. Metoda geometryczna

Zadanie PL może mieć rozwiązania dopuszczalne lub być zadaniem sprzecznym, niemającym rozwiązania dopuszczalnego. Jeżeli zadanie PL ma rozwiązania dopuszczalne, to zachodzi jedna z trzech możliwości:

- 1) jest jedno rozwiązanie optymalne;
- 2) jest wiele rozwiązań optymalnych;
- 3) brak rozwiązania optymalnego.

Znaleźć rozwiązanie optymalne lub pokazać, że zadanie go nie ma jest bardzo łatwo w przypadku dwóch zmiennych. Stosujemy wówczas **metodę geometryczną**. Ze względu na to ograniczenie nie ma ona dużego znaczenia praktycznego, ma natomiast ogromne walory dydaktyczne. Pozwala doskonale zilustrować procedurę znajdowania rozwiązania optymalnego. Istotę tej metody wyjaśnimy, rozwiązując przykład liczbowy.

## PRZYKŁAD 1.4

Rozważmy zadanie PL z dwoma zmiennymi:

$$\begin{aligned} \text{(FC)} \quad & 4x_1 + 2x_2 \rightarrow \max, \\ \text{(I)} \quad & 2x_1 - x_2 \leq 2, \\ \text{(II)} \quad & x_2 \leq 3, \\ \text{(III)} \quad & 2x_1 + x_2 \geq 2, \\ \text{(IV)} \quad & x_1, x_2 \geq 0. \end{aligned}$$

Aby rozwiązać zadanie PL metodą geometryczną, rysujemy układ współrzędnych, na osiach  $x_1, x_2$  nanosimy jednostki. Następnie rysujemy proste odpowiadające poszczególnym nierównościom zadania. Z nierównością (I) jest związana prosta o równaniu  $2x_1 - x_2 = 2$ . Znajdujemy jej punkty przecięcia z osią  $x_1$  ( $x_1 = 1$ ) oraz osią  $x_2$  ( $x_2 = -2$ ). Ponieważ nierówność (I) odpowiada półpłaszczyzna punktów leżących zarówno na tej prostej, jak i na lewo od niej, to obszar ten zaznaczamy strzałką. Z kolei nanosimy proste odpowiadające nierównościom II i III. Warunki nieujemności zmiennych zaznaczamy, umieszczając strzałki na osi  $x_1$  (dla  $x_1 \geq 0$ ) i osi  $x_2$  (dla  $x_2 \geq 0$ ). Uzyskujemy w ten sposób pięć półpłaszczyzn. Część wspólna półpłaszczyzn, czyli zbiór punktów o współrzędnych  $(x_1, x_2)$  spełniających wszystkie warunki ograniczające jest zbiorem rozwiązań dopuszczalnych. Zbiór ten to wielokąt  $A, B, C, D$ . Na rysunku 1.1 zaznaczamy go jako obszar zakreskowany.

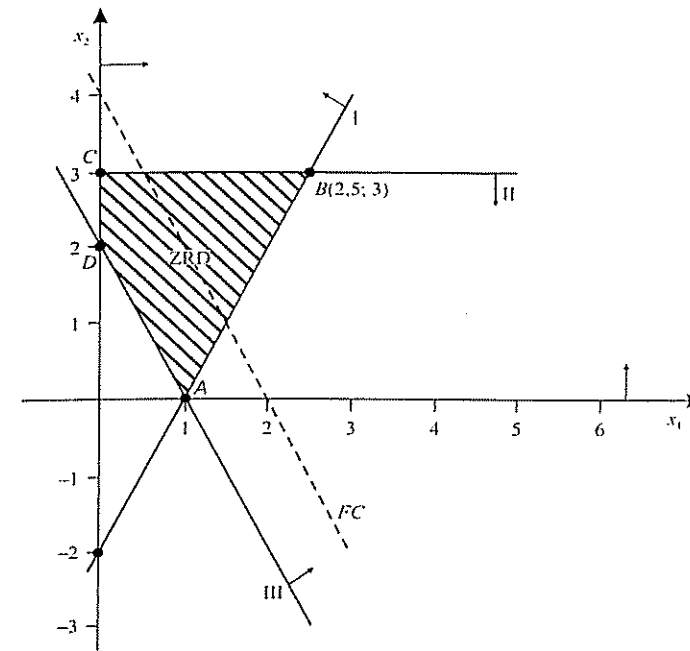
Aby znaleźć rozwiązanie optymalne zadania PL, przyjmujemy dowolną, lecz dogodną wartość funkcji celu, np. 8, i rysujemy prostą o równaniu  $4x_1 + 2x_2 = 8$ . Na rysunku 1.1 zaznaczono ją jako prostą przerywaną (FC). Jest to izokwanta, czyli linia tych samych wartości celu. Jeżeli weźmiemy większą wartość funkcji celu, to uzyskamy prostą równoległą, która będzie leżała powyżej prostej (FC)<sup>1</sup>. Jeżeli przyjmiemy mniejszą wartość, to nowa izokwanta będzie leżała poniżej prostej (FC). Jeżeli izokwantę (FC) przesuniemy równolegle w górę, to będzie ona styczna do obszaru rozwiązań w punkcie  $B(2,5; 3)$ ; w tym punkcie osiąga ona wartość największą w zbiorze rozwiązań dopuszczalnych (ZRD). Punkt  $B$  o współrzędnych  $x_1 = 2,5, x_2 = 3$  jest rozwiązaniem optymalnym zadania (FC)–(IV). Aby znaleźć te współrzędne, rozwiązujemy następujący układ równań:

$$\begin{cases} 2x_1 - x_2 = 2, \\ x_2 = 3. \end{cases}$$

Gdybyśmy szukali minimum tej samej funkcji, wówczas rozwiązaniem optymalnym byłby każdy punkt na odcinku  $[A, D]$ .

<sup>1</sup> Izokwanta o wyższej wartości funkcji celu leży poniżej tylko w przypadku, gdy obydwie wagi są ujemne.

RYSUNEK 1.1  
Metoda geometryczna



Problem znajdowania rozwiązania zadania PL metodą geometryczną sprowadza się do:

- 1) wyznaczenia półpłaszczyzn odpowiadających poszczególnym nierównościom;
- 2) znalezienia części wspólnej dla wszystkich półpłaszczyzn, czyli ZRD;
- 3) wyszukaniu w ZRD rozwiązania najlepszego dla przyjętej funkcji celu.

Jeżeli ZRD jest zbiorem pustym lub zbiorem nieograniczonym w kierunku wzrostu wartości funkcji celu dla zadania na maksimum lub spadku dla zadania na minimum, to zadanie nie ma rozwiązania optymalnego.

## 1.6. Metoda simpleks

Metoda simpleks jest podstawową metodą znajdowania optymalnych rozwiązań zadań programowania liniowego. Jest to metoda ogólna, pozwalająca rozwiązać każde zadanie PL. Polega ona na sekwencyjnym, ściśle określonym przeglądzie rozwiązań bazowych.

### 1.6.1. Rozwiązanie bazowe zadania PL

Rozwiązanie bazowe jest związane z postacią kanoniczną. Załóżmy więc, że dane jest zadanie PL o postaci kanonicznej:

$$\begin{aligned} cx &\rightarrow \max, \\ Ax &= b, \\ x &\geq 0, \end{aligned} \quad (1.26)$$

gdzie:

- A — macierz wymiarowa współczynników ( $m \cdot n$ ),
- x —  $n$ -wymiarowy wektor zmiennych,
- b —  $m$ -wymiarowy wektor wyrazów wolnych,
- c —  $n$ -wymiarowy wektor wag funkcji celu<sup>2</sup>.

Niech **B** oznacza macierz kwadratową  $m$ -tego stopnia składającą się z  $m$  linio-wo niezależnych kolumn macierzy A. Wyznacznik tej macierzy jest oczywiście różny od zera:

$$\det(\mathbf{B}) \neq 0.$$

Macierz **B** nazywać będziemy **bazą**, jej kolumny — **kolumnami bazowymi**, a pozostałe kolumny macierzy A — **kolumnami niebazowymi**. Zmienne związane z kolumnami bazowymi nazywamy **zmiennymi bazowymi**, a pozostałe zmienne — **zmiennymi niebazowymi**. Oznaczamy przez  $Z_B$  zbiór zmiennych bazowych, a przez  $Z_N$  — zbiór zmiennych niebazowych.

Z każdą bazą **B** układu  $Ax = b$  jest związane **rozwiązanie bazowe**. Jeżeli układ  $Ax = b$  jest niesprzeczny oraz  $n > m$ , to układ ten ma nieskończenie wiele rozwiązań, ale skończoną liczbę rozwiązań bazowych. Układ  $m$  równań z  $n$  zmiennymi ma co najwyżej  $\binom{n}{m} = \frac{n!}{m!(n-m)!}$  rozwiązań bazowych.

### 1.6.2. Postać bazowa zadania PL

Każdej bazie **B** odpowiada określony podział zmiennych na zmienne bazowe i niebazowe oraz związane z nią rozwiązanie bazowe. Przy danej bazie **B** wektor zmiennych x oraz macierz współczynników A można przedstawić jako:

$$x = (x_B, x_N), \quad A = (B, N).$$

<sup>2</sup> Wektory x i b są wektorami kolumnowymi, a wektor c — wektorem wierszowym.

Wówczas układ równań  $Ax = b$  zapiszemy w postaci:

$$Bx_B + Nx_N = b. \quad (1.27)$$

Mnożąc lewostronnie układ (1.27) przez  $B^{-1}$ , otrzymujemy **postać bazową**:

$$Ix_B + Wx_N = b^*, \quad (1.28)$$

gdzie:

$$I = B^{-1}B,$$

$$W = B^{-1}N,$$

$$b^* = B^{-1}b.$$

Z postaci bazowej (1.28) łatwo można odczytać rozwiązanie bazowe:

$$x_B = b^*, \quad x_N = 0.$$

Jeżeli dla danej bazy **B**:

$$x_B = B^{-1}b \geq 0,$$

to rozwiązanie bazowe jest **rozwiązaniem dopuszczalnym**.

Dwie bazy **B** i **B'** nazywamy **bazami sąsiednimi**, jeżeli się różnią tylko jedną kolumną macierzy A. Podobnie dwa rozwiązania bazowe będziemy nazywali **rozwiązaniami sąsiednimi**, jeżeli będą się różnić tylko jedną zmienną bazową.

Przechodzenie w metodzie simpleks od jednego rozwiązania bazowego do drugiego, sąsiedniego w sensie rachunkowym, polega na przekształceniu układu równań (1.28) od jednej postaci bazowej do drugiej. Oczywiście najlepiej, gdy początkowa postać kanoniczna jest także postacią bazową.

### 1.6.3. Idea metody simpleks

Zastosowanie pełnego przeglądu zbioru rozwiązań bazowych jest nieefektywne ze względu na liczbę tych rozwiązań oraz wielkość każdego układu równań, jaki należy przekształcać. Jeżeli  $n = 20$ , a  $m = 10$ , to rozwiązań bazowych może być  $\binom{20}{10} = 184\,756$ .

W metodzie simpleks stosujemy przegląd ukierunkowany zbioru rozwiązań bazowych. Przechodzimy od jednego rozwiązania bazowego dopuszczalnego do drugiego, o którym wiemy, że jest nie gorsze od poprzedniego. Pomijamy więc rozwiązania bazowe niedopuszczalne oraz te, które są gorsze od aktualnie rozpatrywanego.

Przegląd ukierunkowany zbioru rozwiązań bazowych zadania PL w metodzie simpleks sprowadza się do realizacji następujących kroków (jeżeli zadanie ma rozwiązanie optymalne):



- Krok A. Wyznaczamy rozwiązanie wyjściowe, dopuszczalne i bazowe.  
 Krok B. Sprawdzamy, czy aktualne rozwiązanie bazowe jest optymalne. Jeżeli tak to koniec obliczeń. Jeżeli nie, to idź do kroku C.  
 Krok C. Przechodzimy do sąsiedniego rozwiązania bazowego, o którym wiemy, że jest nie gorsze od poprzedniego i wracamy do kroku B.

### 1.6.4. Tablica simpleksowa

Postać bazową wraz ze wszystkimi elementami niezbędnymi do oceny, czy rozwiązanie bazowe jest optymalne, czy też nie jest, wygodnie zapisać w postaci tzw. **tablicy simpleksowej**  $[t_{ij}]$ . Elementy  $j$ -tej kolumny tablicy simpleksowej oraz kolumny wartości zmiennych bazowych definiujemy następująco:

$$\begin{bmatrix} t_{1,j} \\ t_{2,j} \\ \vdots \\ t_{m,j} \\ t_{0,j} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{A}_j \\ \vdots \\ c_j - \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}_j \end{bmatrix} \cdot \begin{bmatrix} t_{1,0} \\ t_{2,0} \\ \vdots \\ t_{m,0} \\ t_{0,0} \end{bmatrix} = \begin{bmatrix} \mathbf{B}^{-1}\mathbf{b} \\ \vdots \\ \mathbf{c}_B\mathbf{B}^{-1}\mathbf{b} \end{bmatrix},$$

gdzie  $\mathbf{A}_j$  to  $j$ -ta kolumna macierzy współczynników  $\mathbf{A}$ .

Układ równań (1.28) można więc zapisać:

$$x_{(i)} = t_{i0} - \sum_{j \in Z_N} t_{ij}x_j \quad (i = 1, 2, \dots, m),$$

gdzie  $x_{(i)}$  to zmienna bazowa związana z  $i$ -tym wierszem tablicy simpleksowej.

Wartość funkcji celu wynosi:

$$x_0 = t_{00} + \sum_{j \in Z_N} t_{0j}x_j$$

Wskaźnik kryterium optymalności, stojący przy  $j$ -tej zmiennej, obliczamy w następujący sposób:

$$t_{0j} = c_j - \mathbf{c}_B\mathbf{B}^{-1}\mathbf{A}_j = c_j - \sum_{i \in Z_B} c_i t_{ij} = c_j - z_j. \quad (1.29)$$

Element  $t_{ij} > 0$  oznacza, o ile zmniejszy się wartość zmiennej bazowej  $x_{(i)}$  jeżeli zmienna niebazowa  $x_j$  przyjmie wartość jeden. Podobnie, element  $t_{ij} < 0$

określa, o ile zwiększy się wartość zmiennej  $x_{(i)}$ , gdy zmienna niebazowa  $x_j$  przyjmie wartość jeden.

### 1.6.5. Początkowe rozwiązanie bazowe

Aby rozwiązać zadanie PL metodą simpleks, należy je sprowadzić do postaci bazowej z nieujemnym wektorem wyrazów wolnych  $\mathbf{b}$ . Przyjmując, że zmiennymi bazowymi są zmienne związane z wektorami jednostkowymi, uzyskujemy bardzo łatwo początkowe rozwiązanie bazowe:

$$\mathbf{x}_B = \mathbf{b}, \quad \mathbf{x}_N = \mathbf{0}.$$

Jeżeli  $i$ -ty wyraz wolny jest ujemny, to przemnażamy odpowiadające mu  $i$ -te równanie obustronnie przez  $(-1)$ . Gorzej jest, gdy macierz współczynników postaci kanonicznej nie zawiera macierzy jednostkowej. Można wówczas:

- 1) tak przekształcić układ równań, stosując reguły metody eliminacji Gaussa, aby zawierał on macierz jednostkową;
- 2) sztucznie utworzyć taką macierz, uzupełniając macierz współczynników  $\mathbf{A}$  o odpowiednią liczbę brakujących wektorów jednostkowych.

Będziemy stosowali ten drugi wariant procedury. Jest to **metoda sztucznej bazy**. W ślad za wprowadzonymi wektorami jednostkowymi rozszerzamy listę zmiennych o tzw. **zmienne sztuczne**.

Waga przy zmiennej sztucznej jest taka, że nieopłacalne jest pozostawienie tej zmiennej w rozwiązaniu optymalnym. Jest to bardzo duża liczba dodatnia dla zadania na minimum oraz liczba ujemna i duża co do modułu w zadaniu na maksimum.

Stosowanie zmiennych sztucznych jest chwytem rachunkowym mającym na celu szybkie uzyskanie początkowego rozwiązania bazowego. Wprowadzając je, uzyskujemy nowe, równoważne, pomocnicze zadanie PL. Optymalne rozwiązanie zadania pomocniczego wyznacza optymalne rozwiązanie zadania początkowego, jeżeli wszystkie zmienne sztuczne w rozwiązaniu optymalnym są zerowe. Jeżeli natomiast chociaż jedna zmienna sztuczna jest dodatnia, to początkowe zadanie jest sprzeczne.

### 1.6.6. Kryterium simpleksowe i reguły przekształceń postaci bazowej

Z każdą zmienną  $x_j$  związany jest wskaźnik kryterium optymalności  $t_{0j}$ . Wskaźnik ten dla zmiennych bazowych zawsze jest zerowy. Natomiast dla zmiennych niebazowych wskaźnik  $t_{0j}$  określa, jak się zmieni wartość funkcji celu, jeżeli zmien-

na  $x_j$  przyjmie wartość jeden, a wartości zmiennych bazowych zostaną odpowiednio zmienione w celu zachowania dopuszczalności rozwiązania.

Za pomocą **kryterium simpleks** można ocenić:

- 1) czy rozwiązanie bazowe jest optymalne;
- 2) którą zmienną niebazową oplaca się wprowadzić do bazy;
- 3) czy istnieje jedno czy więcej rozwiązań optymalnych.

#### 1.6.6.1. Kryterium optymalności

Dane rozwiązanie bazowe jest optymalne, jeżeli wskaźniki kryterium simpleks dla zmiennych niebazowych są:

- 1) niedodatnie w zadaniu PL na maksimum, czyli:

$$t_{0j} \leq 0 \quad (j \in Z_N);$$

- 2) nieujemne w zadaniu PL na minimum, czyli:

$$t_{0j} \geq 0 \quad (j \in Z_N).$$

#### 1.6.6.2. Reguła wyboru zmiennej wprowadzanej do bazy

Jeżeli dane rozwiązanie bazowe nie jest optymalne, to istnieje chociaż jeden wskaźnik optymalności dodatni dla zadania na maksimum lub ujemny dla zadania na minimum. Należy ustalić zmienną niebazową, której wprowadzenie do nowej bazy spowoduje największą poprawę funkcji celu.

Dla zadania na maksimum będzie to zmienna  $x_k$  o maksymalnym wskaźniku optymalności, dla której:

$$t_{0k} = \max \{t_{0j} | j \in Z_N\}. \quad (1.30)$$

Dla zadania PL na minimum będzie to zmienna  $x_k$  o minimalnej wartości wskaźnika optymalności, dla której:

$$t_{0k} = \min \{t_{0j} | j \in Z_N\}. \quad (1.31)$$

#### 1.6.6.3. Reguła wyboru zmiennej usuwanej z bazy

Mając ustaloną zmienną wprowadzaną do bazy, musimy wyznaczyć zmienną usuwaną z bazy tak, aby nowe rozwiązanie bazowe pozostało dopuszczalne. Będzie to zmienna  $x_{(r)}$ , dla której:

$$\frac{t_{r0}}{t_{rk}} = \min \left\{ \frac{t_{i0}}{t_{ik}} \mid t_{ik} > 0 \right\}. \quad (1.32)$$

Wybór innej zmiennej niż  $x_{(r)}$  powoduje uzyskanie nowego rozwiązania bazowego, które będzie niedopuszczalne. Brak w  $k$ -tej kolumnie tablicy simpleksowej elementu dodatniego to sygnał, że zadanie nie ma rozwiązania optymalnego i należy przerwać obliczenia.

Jeżeli jest więcej zmiennych spełniających albo warunek (1.30), albo warunek (1.31), albo warunek (1.32), to zawsze wybieramy zmienną o najniższym numerze.

#### 1.6.6.4. Reguły wyznaczania nowej tablicy simpleksowej

Mając ustaloną zmienną wprowadzaną i zmienną usuwaną z bazy (odpowiednią kolumnę i wiersz nazywamy **kolumną** i **wierszem centralnym**) przechodzimy do nowej bazy i związanego z nią nowego rozwiązania bazowego  $x'$ . Nowej bazie odpowiada nowa tablica simpleksowa  $T' = [t'_{ij}]$ , której elementy ustalamy na podstawie tablicy  $T$  według wzorów:

$$\begin{cases} t'_{ij} = \frac{t_{ij}}{t_{rk}} & (j = 0, 1, \dots, n), \\ t'_{ij} = t_{ij} - t_{ik}t'_{rj} & (i \neq r, j = 0, 1, \dots, n). \end{cases} \quad (1.33)$$

Przekształcając elementy starej tablicy  $T$  według podanych wzorów, otrzymujemy w nowej tablicy  $T'$  wektory jednostkowe dla nowego zestawu zmiennych bazowych.

#### 1.6.7. Degeneracja rozwiązania bazowego

**Degeneracja rozwiązania bazowego** zachodzi, gdy chociaż jedna zmienna bazowa jest zerowa. Wówczas mogą wystąpić tzw. martwe kroki, w których przechodzimy od jednego rozwiązania bazowego do drugiego związanego z tym samym punktem. Wartości zmiennych nie ulegają w tym przypadku zmianie. Podobnie jak wartość funkcji celu, która się nie poprawia. Teoretycznie możliwe jest powstanie cyklu, czyli ciągu rozwiązań, który będzie się stale powtarzał. Na szczęście z praktyki degeneracja rozwiązania bazowego albo znika, albo kolejne rozwiązanie zdegenerowane okazuje się rozwiązaniem optymalnym.

#### PRZYKŁAD 1.5

Rozwiążmy zadanie PL:

$$\begin{cases} 3x_1 + 2x_2 \rightarrow \max, \\ 2x_1 + 3x_2 \leq 12, \\ 2x_1 \leq 6, \\ x_1, x_2 \geq 0. \end{cases} \quad (1.34)$$

Wprowadzając zmienne swobodne  $x_3$  i  $x_4$ , sprowadzamy to zadanie do równoważnej postaci kanonicznej:

$$\begin{cases} 3x_1 + 2x_2 + 0x_3 + 0x_4 \rightarrow \max, \\ 2x_1 + 3x_2 + x_3 + 0x_4 = 12, \\ 2x_1 + 0x_2 + 0x_3 + x_4 = 6, \\ x_1, x_2, x_3, x_4 \geq 0. \end{cases} \quad (1.35)$$

Postać kanoniczna jest zarazem postacią bazową generującą początkowe, dopuszczalne rozwiązanie bazowe. Jeżeli zmiennymi bazowymi będą zmienne swobodne  $x_3$  i  $x_4$  z wektorami jednostkowymi:

$$A_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

to z układu równań (1.35) łatwo uzyskać początkowe rozwiązanie bazowe  $x^1$ :

$$x_1 = x_2 = 0, \quad x_3 = 12, \quad x_4 = 6.$$

Wyjściowej postaci bazowej odpowiada początkowa tablica simpleksowa  $T^1$  (tablica 1.3):

TABLICA 1.3  
Tablica simpleksowa  $T^1$

| $c_j$    |                | 3     | 2     | 0     | 0     | $b_i$            |
|----------|----------------|-------|-------|-------|-------|------------------|
| $c_i$    | zmienna bazowa | $x_1$ | $x_2$ | $x_3$ | $x_4$ |                  |
| 0        | $x_3$          | 2     | 3     | 1     | 0     | 12               |
| 0        | $x_4$          | 2     | 0     | 0     | 1     | 6                |
| $z_j$    |                | 0     | 0     | 0     | 0     | $x_0 \downarrow$ |
| $t_{0j}$ |                | 3     | 2     | 0     | 0     | 0                |

Z tablicy simpleksowej  $T^1$  odczytujemy rozwiązanie bazowe  $x^1$ . Pamiętajmy, że zmienne niebazowe są zerowe:  $x_1 = x_2 = 0$ , a wartości zmiennych bazowych otrzymujemy z kolumny  $b_i^*$ :

$$x_3 = 12, \quad x_4 = 6, \quad x_0 = 0.$$

Aby sprawdzić, czy rozwiązanie  $x^1$  jest optymalne, musimy najpierw obliczyć parametry:

$$z_j = \sum_{i \in Z_n} c_i t_{ij},$$

a następnie wskaźniki kryterium optymalności  $t_{0j} = c_j - z_j$ .

Rozwiązanie  $x^1$  nie jest optymalne, gdyż  $t_{01} = 3$ ,  $t_{02} = 2$ , czyli są dodatnie, a zadanie jest na maksimum. Zmienną wprowadzaną do bazy będzie zmienne  $x_1$ , a zmienną usuwaną z bazy będzie zmienne  $x_4$ , gdyż  $6/2 < 12/2$ . Element  $x_{21} = 2$  jest elementem centralnym przekształcenia simpleksowego. Nową postać bazową zawiera tablica 1.4.

TABLICA 1.4  
Tablica simpleksowa  $T^2$

| $c_j$    |                | 3     | 2     | 0     | 0     | $b_i$            |
|----------|----------------|-------|-------|-------|-------|------------------|
| $c_i$    | zmienna bazowa | $x_1$ | $x_2$ | $x_3$ | $x_4$ |                  |
| 0        | $x_3$          | 0     | 3     | 1     | -1    | 6                |
| 3        | $x_1$          | 1     | 0     | 0     | 1/2   | 3                |
| $z_j$    |                | 3     | 0     | 0     | 3/2   | $x_0 \downarrow$ |
| $t_{0j}$ |                | 0     | 2     | 0     | -3/2  | 9                |

Elementy tablicy simpleksowej  $T^2$  uzyskujemy na podstawie wzoru (1.33). Najpierw ustalamy elementy wiersza centralnego, a następnie elementy wierszy pozostałych. Nowe rozwiązanie bazowe  $x^2$ :  $x_1 = 3$ ,  $x_3 = 6$ ,  $x_0 = 9$ ,  $x_2 = x_4 = 0$  nie jest optymalne, gdyż  $t_{02} = 2 > 0$ . Zmienna  $x_2$  musi być wprowadzona do bazy, z której usuwamy zmienną  $x_3$ . Nową postacią bazową odpowiadającą nowej bazie przedstawiono w tablicy 1.5.

TABLICA 1.5  
Tablica simpleksowa  $T^3$

| $c_j$    |                | 3     | 2     | 0     | 0     | $b_i$            |
|----------|----------------|-------|-------|-------|-------|------------------|
| $c_i$    | zmienna bazowa | $x_1$ | $x_2$ | $x_3$ | $x_4$ |                  |
| 2        | $x_2$          | 0     | 1     | 1/3   | -1/3  | 2                |
| 3        | $x_1$          | 1     | 0     | 0     | 1/2   | 3                |
| $z_j$    |                | 3     | 2     | 2/3   | 5/6   | $x_0 \downarrow$ |
| $t_{0j}$ |                | 0     | 0     | -2/3  | -5/6  | 13               |

Rozwiązanie bazowe  $x^3$ :  $x_1 = 3$ ,  $x_2 = 2$ ,  $x_3 = x_4 = 0$ ,  $x_0 = 13$  jest rozwiązaniem optymalnym zadania (1.35), a tym samym i zadania (1.34). Z końcowej tablicy simpleksowej łatwo także odczytać rozwiązanie optymalne zadania dualnego:  $y_1 = z_3 = 2/3$ ,  $y_2 = z_4 = 5/6$ .

Z interpretacji zmiennych dualnych wynika, że zwiększenie wyrazu wolnego  $b_1$  o jednostkę spowoduje wzrost funkcji celu o  $y_1 = 2/3$ .

## 1.7. Analiza postoptymalizacyjna

### 1.7.1. Idea analizy

**Analiza postoptymalizacyjna** (zwana często **analizą wrażliwości**) jest to analiza prowadzona po uzyskaniu rozwiązania optymalnego. Chodzi w niej o ustalenie, jak zmieni się rozwiązanie optymalne i jakie będą tego skutki, jeżeli zmienią się parametry zadania. Zachodzi jeden z trzech przypadków:

- 1) rozwiązanie optymalne się nie zmienia (zmiany parametrów są nieistotne),
- 2) rozwiązanie optymalne zmienia się w sposób przewidziany (zmiana parametrów jest istotna),
- 3) rozwiązanie optymalne zmienia się w sposób nieprzewidziany (zmiana parametrów jest nadmierna), co wymaga rozwiązania zadania jeszcze raz, chociaż nie zawsze od początku.

Jeżeli zapiszemy zadanie PL w formie macierzowo-wektorowej:

$$\begin{aligned} cx &\Rightarrow \max, \\ Ax &\leq b, \\ x &\geq 0, \end{aligned} \quad (1.36)$$

to widzimy, że rozwiązanie optymalne  $x^0$  zadania (1.36) zależy od macierzy współczynników  $A$ , wektora wag funkcji celu  $c$  oraz wektora wyrazów wolnych  $b$ .

Wyróżniamy trzy przypadki:

- 1) zmieniają się parametry w co najmniej dwóch zbiorów danych:  $A$ ,  $b$  lub  $c$ .
- 2) zmienia się jeden wektor:
  - wag funkcji celu  $c$ ,
  - wyrazów wolnych  $b$ ,
  - $j$ -ty wektor macierzy współczynników  $A_j$ ;
- 3) zmienia się jeden parametr:
  - $j$ -ta waga funkcji celu  $c_j$ ,
  - $i$ -ty wyraz wolny  $b_i$ ,
  - współczynnik  $a_{ij}$  macierzy ograniczeń  $A$ .

W pierwszym przypadku zmiany parametrów są tak duże, że zazwyczaj nie potrafimy powiedzieć, jakie będzie nowe rozwiązanie optymalne bez rozwiązania zadania od początku. W drugim przypadku interesuje nas pytanie, czy zastąpienie wyjściowego wektora parametrów zmianą rozwiązania optymalne, czy też nie. Analizę prowadzoną w tym przypadku nazywa się **analizą wektorową**. W trzecim przypadku, gdy zmieniamy wartość jednego parametru, można stosować te same procedury, co w analizie wektorowej. Zazwyczaj jednak pytamy się jeszcze o to, w jakim przedziale może się zmieniać dany parametr, nie powodując zmiany

rozwiązania optymalnego lub powodując jego zmianę w sposób przewidywalny. Analizę tego typu nazywamy **analizą przedziałową**.

### 1.7.2. Analiza wektorowa

W **analizie wektorowej** zastępujemy wyjściowy wektor parametrów nowym wektorem. Interesuje nas pytanie, czy rozwiązanie optymalne nie ulegnie zmianie lub jeżeli się zmieni, to czy zmieni się także jego struktura (układ zmiennych bazowych i niebazowych). Obliczenia prowadzimy na podstawie końcowej tablicy sympleksowej.

#### 1.7.2.1. Zmiana wektora wag

Jeżeli wyjściowy wektor wag  $c$  zastąpimy nowym wektorem  $c'$ , to dla rozwiązania z końcowej tablicy sympleksowej liczymy nowe kryterium optymalności:

$$t'_{0j} = c'_j - z'_j, \quad z'_j = \sum_{i \in Z_n} t_{ij} c'_j. \quad (1.37)$$

Jeżeli spełniony jest warunek:

$$\bigwedge_{j \in Z_n} t'_{0j} \leq 0, \quad (1.38)$$

to rozwiązanie optymalne się nie zmienia. Jeżeli natomiast warunek (1.38) nie jest spełniony, to rozwiązanie optymalne się zmienia. Aby ustalić nowe rozwiązanie optymalne, musimy wykonać co najmniej jedną iterację sympleksową, przyjmując za punkt wyjścia końcową tablicę sympleksową i nowe kryterium optymalności.

#### 1.7.2.2. Zmiana wektora wyrazów wolnych

Jeżeli wektor wyrazów wolnych  $b$  zastępujemy wektorem  $b'$ , to nowe rozwiązanie bazowe generowane przez dotychczasową bazę  $B$  ustalamy zgodnie z wzorem:

$$x'_B = B^{-1}b'. \quad (1.39)$$

Jeżeli spełniony jest warunek:

$$x'_B \geq 0, \quad (1.40)$$

to nowe rozwiązanie bazowe jest dopuszczalne, a tym samym jest także optymalne. Nie zmienia się bowiem kryterium optymalności. Jeżeli natomiast warunek (1.40) nie jest spełniony, to rozwiązanie bazowe jest niedopuszczalne, czyli zmiana rozwiązania optymalnego jest nieprzewidywalna. Aby wyznaczyć nowe rozwiązanie optymalne, trzeba wykonać co najmniej jedną iterację tzw. **dualnej metody sympleks**.

### 1.7.2.3. Zmiana wektora macierzy współczynników

Jeżeli wyjściowy wektor  $A_j$ , zastąpimy wektorem  $A'_j$ , to sprawdzamy  $j$ -ty warunek zadania dualnego:

$$\sum_{i=1}^m a'_{ij} y_i^0 \geq c_j \quad (1.41)$$

Gdy warunek (1.41) jest spełniony, wówczas rozwiązanie optymalne  $x^0$  nie ulega zmianie. Jeżeli natomiast:

$$\sum_{i=1}^m a'_{ij} y_i^0 < c_j,$$

to rozwiązanie optymalne się zmienia.

#### PRZYKŁAD 1.6

Rozważmy zadanie z przykładu 1.5, nadając mu następującą interpretację ekonomiczną. Firma może produkować dwa wyroby  $W_1$  i  $W_2$ , wykorzystując przy tym dwa środki produkcji  $S_1$  i  $S_2$ . Ceny sprzedaży wyrobów wynoszą:  $c_1 = 3$  tys. zł,  $c_2 = 2$  tys. zł. Znamy normy zużycia tych środków dotyczące poszczególnych wyrobów oraz ich zasoby:

$$A_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \quad b = \begin{bmatrix} 12 \\ 6 \end{bmatrix}.$$

Należy ustalić plan produkcji maksymalizującej przychód. Rozwiązanie optymalne tego zadania zawiera tablica 1.5:  $x_1 = 3$ ,  $x_2 = 2$ ,  $x_3 = 0$ ,  $x_4 = 0$ ,  $x_0 = 13$ , czyli należy produkować trzy jednostki wyrobu  $W_1$ , dwie jednostki wyrobu  $W_2$ , zużycie środka  $S_1$  i  $S_2$  jest pełne, a przychód wynosi 13 tys. zł. Zmienne dualne  $y_1 = 2/3$  oraz  $y_2 = 5/6$  odpowiadają wycenie jednostki, odpowiednio, środka  $S_1$  i  $S_2$ .

Po pierwsze, należy ustalić, czy dla nowego wektora cen  $c' = [2, 4]$  rozwiązanie optymalne ulega zmianie. Obliczamy zgodnie z wzorem (1.37) nowe wskaźniki kryterium optymalności:

$$t'_{03} = 0 - [1/3 \cdot 4 + 0 \cdot 2] = -4/3 < 0,$$

$$t'_{04} = 0 - [(-1/3) \cdot 4 + 1/2 \cdot 2] = 1/3 > 0.$$

Ponieważ wskaźnik  $t'_{04}$  jest dodatni, to rozwiązanie optymalne ulega zmianie.

Po drugie, trzeba określić, jakie będzie rozwiązanie optymalne, jeżeli zmienia się zasoby środków i wektor  $b' = [9; 12]$ ?

Macierz  $B^1$  przedstawiono w tablicy 1.5 (s. 37) w kolumnie  $x_3$  i  $x_4$ . Nowe rozwiązanie bazowe ustalamy zgodnie z wzorem (1.39):

$$x_{B1} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1/2 \\ 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} 9 \\ 12 \end{bmatrix} = \begin{bmatrix} 6 \\ -1 \end{bmatrix}.$$

Ponieważ nowe rozwiązanie, generowane przez aktualną bazę  $B$ , jest niedopuszczalne, stąd zmiana rozwiązania jest nieprzewidywalna.

Po trzecie, należy sprawdzić, czy produkcja wyrobu  $W_5$  z ceną sprzedaży  $c_5 = 2$  oraz z wektorem norm zużycia  $A_5 = [2; 1]$ , będzie opłacalna.

Sprawdzamy zgodnie z wzorem (1.41), czy dla nowego wyrobu odpowiedni warunek zadania dualnego jest spełniony:

$$2 \cdot 2/3 + 1 \cdot 5/6 = 13/6 > 2.$$

Wartość środków produkcji przypadająca na jednostkę nowego wyrobu  $W_5$  (13/6 tys. zł) jest większa od jego ceny (2 tys. zł), a zatem produkcja  $W_5$  jest nieopłacalna. Tym samym rozwiązanie optymalne nie ulega zmianie.

### 1.7.3. Analiza przedziałowa

Celem analizy przedziałowej jest ustalenie dopuszczalnych przedziałów zmian parametru, przy których rozwiązanie optymalne zadania nie zmienia się lub zmienia się w sposób przewidywalny.

#### 1.7.3.1. Dopuszczalny przedział dla zmian wagi funkcji celu

Oznaczmy przez  $\Delta c_j = c'_j - c_j$  zmianę  $j$ -tej wagi funkcji celu. Dopuszczalny przedział dla  $\Delta c_j$  można wyznaczyć z układu nierówności:

$$\sum_{i \in Z_N} t_{ij} c'_i \geq c'_j \quad (j \in Z_N)$$

lub bezpośrednio na podstawie następujących wzorów:

1) dla wagi niebazowej:

$$-\infty < \Delta c_j \leq -t_{0j}; \quad (1.42)$$

2) dla wagi zmiennej bazowej:

$$\max \left\{ \frac{t_{0j}}{t_{k,j}} \mid j \in Z_N, t_{k,j} > 0 \right\} \leq \Delta c_{(k)} \leq \min \left\{ \frac{t_{0j}}{t_{k,j}} \mid j \in Z_N, t_{k,j} < 0 \right\}. \quad (1.43)$$

Dla zadania na minimum zamiast wzorów (1.42) i (1.43) stosujemy wzory:  
 $-t_{0j} \leq \Delta c_j < +\infty$ .

$$\max \left\{ \frac{t_{0j}}{t_{k,j}} \mid j \in Z_N, t_{k,j} < 0 \right\} \leq \Delta c_{(k)} \leq \min \left\{ \frac{t_{0j}}{t_{k,j}} \mid j \in Z_N, t_{k,j} > 0 \right\}.$$

### 1.7.3.2. Dopuszczalny przedział zmian wyrazu wolnego

Interesuje nas ustalenie takiego przedziału zmian wyrazu wolnego, przy którym rozwiązanie optymalne zmienia się w sposób przewidywalny, czyli zmieniają się wartości zmiennych bazowych, natomiast nie zmienia się skład (struktura) zmiennych bazowych i niebazowych. Dopuszczalny przedział obliczamy z nierówności:

$$B^{-1}b' \geq 0, \quad (1.44)$$

gdzie  $b'_i = b_i + \Delta b_i$ , a  $\Delta b_i$  to zmiana wartości  $i$ -tego wyrazu wolnego.

Zamiast rozwiązywać układ nierówności (1.44), wygodniej jest zastosować wzór:

$$\max \left\{ \frac{-b_k^*}{t_{k,l_i}} \mid t_{k,l_i} > 0 \right\} \leq \Delta b_i \leq \min \left\{ \frac{-b_k^*}{t_{k,l_i}} \mid t_{k,l_i} < 0 \right\},$$

gdzie  $l_i$  to numer kolumny, w której w pierwszej tablicy simpleksowej stał  $i$ -ty wektor jednostkowy.

#### PRZYKŁAD 1.7

Rozważamy ponownie zadanie z przykładów 1.5 i 1.6.

Po pierwsze, należy ustalić, o ile można zwiększyć cenę wyrobu  $W_2$ , nie powodując zmiany rozwiązania optymalnego.

Dopuszczalny przedział zmian dla ceny wyrobu  $W_2$  wynosi:

$$-2 \leq \Delta c_2 \leq 5/2,$$

czyli cena wyrobu  $W_2$  może wzrosnąć maksymalnie o 2,5 tys. zł.

Po drugie, firma zamierza dokupić 3 jednostki środka  $S_1$ , płacąc 1 tys. zł za jednostkę. Czy ten zakup jest dla firmy opłacalny?

Dopuszczalny przedział zmian dla wyrazu  $b_1$  wynosi:

$$-6 \leq \Delta b_1 \leq +\infty.$$

Ponieważ zakup środka  $S_1$  mieści się w dopuszczalnym przedziale, to na każdej zakupionej jednostce firma traci  $(5/6 - 1)$ , czyli zakup jest nieopłacalny.

Po trzecie, firma zwiększa ilość środka  $S_2$  o 8 jednostek ( $\Delta b_2 = 8$ ). Należy ustalić, o ile zwiększy się przychód i jakie będzie nowe rozwiązanie optymalne. Ponieważ dopuszczalny przedział wynosi:

$$-6 \leq \Delta b_1 \leq 6,$$

to zmiana  $\Delta b_2 = 8$  nie mieści się w tym przedziale. Zmiana rozwiązania optymalnego nie jest więc przewidywalna. Przychód zwiększy się w tym przypadku o co najmniej:

$$\Delta x_0 = y_2 \cdot \Delta b_2 = 5/6 \cdot 6 = 5 \text{ tys. zł.}$$

Dokładnej zmiany przychodu w tej sytuacji nie można podać. Można natomiast uzyskać wiele istotnych informacji, ale wymaga to wyznaczenia dopuszczalnych przedziałów zmian dla samych zmiennych, co jest już bardziej skomplikowane.

#### POLECENIA I PYTANIA KONTROLNE

1. Co to jest decyzja (rozwiązanie) optymalne?
2. Jakie dane (parametry) występują w zagadnieniu diety?
3. Kiedy model matematyczny sytuacji decyzyjnej jest zadaniem PL?
4. Jakie są zasadnicze różnice między postacią standardową a postacią kanoniczną zadania PL?
5. Kiedy graficznie w  $R^2$  poznajemy, że rozwiązanie bazowe jest zdegenerowane?
6. Jakie trudności powoduje degeneracja w metodzie simpleks?
7. Jak poznajemy w metodzie simpleks, że zadanie jest sprzeczne?
8. Jakie są zasadnicze różnice między analizą wektorową a analizą przedziałową?

#### BIBLIOGRAFIA

- Czerwiński Z. (red.), 1983, *Optymalna organizacja złożonych działań*, PWE, Warszawa.  
 Czerwiński Z., 1984, *Matematyka na usługach ekonomii*, PWN, Warszawa.  
 Gale D., 1969, *Teoria liniowych modeli ekonomicznych*, PWN, Warszawa.  
 Ignasiak E. (red.), 2001, *Badania operacyjne*, PWE, Warszawa.  
 Nykowski I., 1986, *Programowanie liniowe*, PWE, Warszawa.  
 Sysło M., Deo N., Kowalik J., 1993, *Algorytmy optymalizacji dyskretnej*, Wydawnictwo Naukowe PWN, Warszawa.  
 Trzaskalik T., 2003, *Wprowadzenie do badań operacyjnych z komputerem*, PWE, Warszawa.  
 Wagner H.M., 1980, *Badania operacyjne*, PWE, Warszawa.

## ROZDZIAŁ 2

# ZAGADNIENIA TRANSPORTOWE



### 2.1. Zamknięte zagadnienie transportowe

Mamy  $m$  dostawców i  $n$  odbiorców jednorodnego dobra. Znamy podaż każdego dostawcy, popyt każdego odbiorcy oraz jednostkowe koszty transportu od dostawców do odbiorców. Należy ustalić taki plan przewozu towaru od dostawców do odbiorców, aby łączne koszty transportu były minimalne.

Jeżeli w zadaniu transportowym łączna podaż jest równa łącznemu popytowi, to takie zadanie transportowe jest zamknięte (zbilansowane). W przeciwnym przypadku zadanie transportowe jest otwarte.

Oznaczmy przez:

- $a_i$  — podaż  $i$ -tego dostawcy ( $i = 1, \dots, m$ ),
- $b_j$  — popyt  $j$ -tego odbiorcy ( $j = 1, \dots, n$ ),
- $c_{ij}$  — jednostkowy koszt transportu od  $i$ -tego dostawcy do  $j$ -tego odbiorcy.

Zbilansowanie zadania oznacza, że łączna podaż jest równa łącznemu popytowi:

$$\sum_{i=1}^m a_i = \sum_{j=1}^n b_j.$$

Wprowadzimy zmienne decyzyjne  $x_{ij}$  — ilość dobra przesyłana od  $i$ -tego dostawcy do  $j$ -tego odbiorcy. Model matematyczny **zamkniętego zagadnienia transportowego (ZZT)** formułujemy następująco:

Znajdź takie wartości zmiennych  $x_{ij}$ , aby:

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min, \quad (2.1)$$

$$\sum_{j=1}^n x_{ij} = a_i \quad (i = 1, \dots, m), \quad (2.2)$$

$$\sum_{i=1}^m x_{ij} = b_j \quad (j = 1, \dots, n), \quad (2.3)$$

$$x_{ij} \geq 0 \quad (i = 1, \dots, m; j = 1, \dots, n). \quad (2.4)$$

Funkcja celu (2.1) minimalizuje łączne koszty transportu, warunek (2.2) zapewnia, że każdy dostawca wyśle tyle towaru, ile ma, a warunek (2.3) wymusza, aby każdy odbiorca dostał tyle towaru, ile chce.

Zadanie (2.1)–(2.4) jest liniowym zadaniem decyzyjnym, a więc można je rozwiązać metodą simpleks. Ze względu na specjalną strukturę macierzy ograniczeń, lepiej stosować specjalne algorytmy służące rozwiązaniu ZZT i zagadnień pokrewnych. W następnym podrozdziale przedstawimy jedną z bardziej popularnych metod — metodę potencjałów.

### 2.2. Metoda potencjałów dla zamkniętego zagadnienia transportowego

Metodę potencjałów sprowadza się do realizacji następujących kroków:

- Krok 1. Wyznaczamy wyjściowe rozwiązanie, dopuszczalne i bazowe.
- Krok 2. Sprawdzamy, czy aktualne rozwiązanie jest optymalne. Jeżeli tak, to koniec obliczeń. Jeżeli nie, to przechodzimy do kroku 3.
- Krok 3. Ustalamy tzw. trasę centralną i tworzymy schemat poprawy rozwiązania.
- Krok 4. Na podstawie ustalonego schematu poprawy wyznaczamy nowe rozwiązanie bazowe i wracamy do kroku 2.

Realizując podane kroki, uzyskujemy ciąg rozwiązań, z których każde jest bazowe, dopuszczalne i nie gorsze od poprzedniego, a ostatnie będzie optymalne.

Metoda potencjałów jest szczególnym przypadkiem metody simpleks. Omówmy teraz dokładnie kolejne kroki.

**Krok 1. Wyznaczenie rozwiązania wyjściowego.** Im lepsze ustalimy rozwiązanie wyjściowe, tym szybciej uzyskamy rozwiązanie optymalne. Rozwiązanie wyjściowe możemy wyznaczyć, stosując różne procedury. Jedną z nich, stosunkowo prostą i dającą dobre wyniki, jest **metoda minimalnego elementu macierzy (MEM)**. Polega ona na tym, że:

- 1) w macierzy kosztów transportu  $C = [c_{ij}]$  szukamy elementu minimalnego, czyli trasy o najniższym koszcie;
- 2) dla tej trasy ustalamy maksymalny przewóz, jaki może być zrealizowany ze względu na podaż i popyt;
- 3) następnie aktualizujemy (zmniejszamy) podaż odpowiedniego dostawcy i popyt odpowiedniego odbiorcy (o wyznaczony przewóz);

- 4) macierz  $C$  aktualizujemy, usuwając z niej wiersz bądź kolumnę, które odpowiada zerowa podaż lub zerowy popyt;
- 5) wracamy do fazy 1 i rozpoczynamy działania w „zredukowanej” macierzy kosztów.

Jeżeli podczas poszukiwania wyjściowego rozwiązania bazowego zeruje się jednocześnie podaż i popyt, to **rozwiązanie bazowe będzie zdegenerowane**. Wykreślamy wówczas wiersz z zerową podażą, pod warunkiem, że nie jest to ostatni dostępny wiersz. Gdy wybrany element jest ostatnim dopuszczalnym elementem macierzy, wówczas kończymy postępowanie (wykreślamy wiersz i kolumnę jednocześnie). Jeżeli jest więcej elementów minimalnych, to wybieramy pierwszy z nich. W ten sposób po  $n + m - 1$  etapach wyznaczamy  $n + m - 1$  **tras bazowych**. Pozostałe trasy są **niebazowe**, a więc charakteryzują się zerowymi przewozami.

**Krok 2. Sprawdzenie, czy rozwiązanie bazowe jest optymalne.** Aby sprawdzić, czy aktualne rozwiązanie bazowe jest optymalne, ustalamy na podstawie **tras bazowych** takie wielkości  $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_n$ , aby:

$$\alpha_i + \beta_j = c_{ij} \quad \langle i, j \rangle \in B, \quad (2.5)$$

gdzie  $B$  to zbiór tras bazowych dla aktualnego rozwiązania.

Ponieważ wielkości  $\alpha_i$  oraz  $\beta_j$  jest  $m + n$ , a tras bazowych  $n + m - 1$ , to umownie możemy przyjąć, że  $\alpha_1 = 0$ . Wartości pozostałych wielkości ustalamy rekurencyjnie, korzystając z wzoru (2.5).

Następnie dla **tras niebazowych** liczymy wskaźniki kryterium optymalności:

$$\Delta_{ij} = c_{ij} - \alpha_i - \beta_j \quad \langle i, j \rangle \in N, \quad (2.6)$$

gdzie  $N$  oznacza zbiór tras niebazowych. Jeżeli:

$$\Delta_{ij} \geq 0 \quad \langle i, j \rangle \in N, \quad (2.7)$$

to dane rozwiązanie bazowe jest optymalne.

$\Delta_{ij}$  mówi nam o ile zwiększy się wartość funkcji celu, jeżeli zwiększymy przewóz na trasie  $\langle i, j \rangle$  o jednostkę, zmieniając — odpowiednio — przewozy na pozostałych trasach tak, aby rozwiązanie pozostało nadal dopuszczalne.

**Krok 3. Poprawa rozwiązania.** Jeżeli rozwiązanie nie jest optymalne, to szukamy trasy niebazowej  $\langle k, l \rangle$ , dla której:

$$\Delta_{kl} = \min \{ \Delta_{ij} | \langle i, j \rangle \in N \}.$$

Trasę tę nazywamy **trasą centralną**. Generuje ona nowe rozwiązania bazowe. Trasa centralna  $\langle k, l \rangle$  wraz z elementami zbioru  $B$  tworzy cykl  $C_{kl}$ , czyli drogę zamkniętą, która ma:

- 1) parzystą liczbę tras;
- 2) dokładnie dwa elementy w każdym wierszu i kolumnie, przez które przechodzi

Trasy tworzące cykl  $C_{kl}$  cechujemy na przemian znakami „+” i „-”. Na trasach ze znakiem „+” (oznaczymy ich zbiór przez  $L^+$ ) zwiększamy przewozy, na trasach ze znakiem „-” (oznaczymy ich zbiór przez  $L^-$ ) zmniejszamy przewozy.

Następnie wybieramy spośród tras, na których zmniejszamy przewozy, trasę  $\langle p, q \rangle$  o minimalnym przewozie:

$$x_{pq} = \min \{ x_{ij} | \langle i, j \rangle \in L^- \}. \quad (2.8)$$

Nowe rozwiązanie bazowe wyznaczamy zgodnie z wzorem:

$$x_{ij} = \begin{cases} x_{ij} - x_{pq} & \text{dla } \langle i, j \rangle \in L^-, \\ x_{ij} + x_{pq} & \text{dla } \langle i, j \rangle \in L^+, \\ x_{ij} & \text{dla pozostałych tras.} \end{cases} \quad (2.9)$$

Nowy zbiór bazowy  $B'$  różni od poprzedniego zbioru  $B$  tylko jedną trasą bazową. W miejsce trasy  $\langle p, q \rangle$ , którą usuwamy z bazy, wchodzi trasa  $\langle k, l \rangle$ .

Uwaga: Koszt jednostkowy  $c_{ij}$ , przewóz  $x_{ij}$ , wskaźnik optymalności  $\Delta_{ij}$  będziemy zapisywać w polu odpowiadającym trasie  $\langle i, j \rangle$  zgodnie ze schematem:

|               |          |
|---------------|----------|
| $c_{ij}$      |          |
| $\Delta_{ij}$ | $x_{ij}$ |

Przewozy  $x_{ij}$  wpisujemy tylko dla tras bazowych, natomiast wskaźniki  $\Delta_{ij}$  tylko dla tras niebazowych.

### PRZYKŁAD 2.1

Trzech dostawców dostarcza towar trzem odbiorcom. Podaż dostawców, popyt odbiorców oraz jednostkowe koszty transportu (w zł) zawiera tablica 2.1. Należy:

- 1) ustalić taki plan przewozów, aby łączne koszty transportu były minimalne;
- 2) odpowiedzieć, czy rozwiązanie optymalne ulegnie zmianie, jeżeli na trasie  $\langle 3, 2 \rangle$  koszty transportu będą mniejsze o 2 zł.

Zauważmy, że łączna podaż równa się łącznemu popytowi, a więc mamy do czynienia z zadaniem zamkniętym. Rozwiązanie wyjściowe wyznaczamy metodą MEM. W macierzy kosztów  $C$  szukamy minimalnego elementu. Jest to element  $c_{1,3} = 2$ . Na trasie  $\langle 1, 3 \rangle$  umieszczamy więc maksymalny przewóz  $x_{1,3} = \min \{ 20, 20 \} = 20$ . Następnie aktualizujemy podaż  $a'_1 = 20 - 20 = 0$  oraz popyt  $b'_3 = 20 - 20 = 0$ . Z uwagi na to, że zarówno „zeruje się podaż”, jak i popyt, wykreślamy wiersz.

Elementem minimalnym w macierzy zredukowanej jest  $c_{2,1} = 4$ . Na trasie  $\langle 2, 1 \rangle$  umieszczamy  $x_{2,1} = 18$  i wykreślamy kolumnę „1”. Kolejnymi minimalnymi



ad 1.

TABLICA 2.1  
Dane liczbowe

|                  |    |    |    |       |
|------------------|----|----|----|-------|
| $i \backslash j$ | 1  | 2  | 3  | $a_i$ |
| 1                | 7  | 4  | 2  | 20    |
| 2                | 4  | 5  | 13 | 30    |
| 3                | 5  | 11 | 8  | 20    |
| $b_j$            | 18 | 32 | 20 |       |

TABLICA 2.2  
Rozwiązanie bazowe  $X^1$

|                      |    |    |    |       |
|----------------------|----|----|----|-------|
| $a_i \backslash b_j$ | 18 | 32 | 20 | $a_i$ |
| 20                   | 7  | 4  | 2  | 0     |
| 30                   | -3 | -1 | 20 | 0     |
| 30                   | 4  | 5  | 13 | 0     |
| 20                   | 5  | 11 | 8  | 6     |
| $\beta_j$            | 4  | 5  | 2  |       |

elementami są:  $c_{2,2} = 5$ ,  $c_{3,3} = 8$ ,  $c_{3,2} = 11$ . Wyjściowe, bazowe rozwiązanie  $X^1$  zawiera tablica 2.2. Koszt transportu  $K^1 = 2 \cdot 20 + 4 \cdot 18 + 5 \cdot 12 + 11 \cdot 20 + 8 \cdot 0 = 392$ .

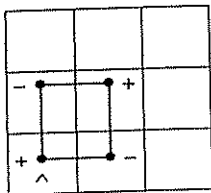
Aby sprawdzić, czy rozwiązanie to jest optymalne, wyznaczmy wartości  $\alpha_i, \beta_j$  spełniające warunek (2.5):

$$\begin{aligned} \alpha_1 &= 0, & \beta_2 &= c_{3,2} - \alpha_3 = 11 - 6 = 5, \\ \beta_3 &= c_{1,3} - \alpha_1 = 2 - 0 = 2, & \alpha_2 &= c_{2,2} - \beta_2 = 5 - 5 = 0, \\ \alpha_3 &= c_{3,3} - \beta_3 = 8 - 2 = 6, & \beta_1 &= c_{2,1} - \alpha_2 = 4 - 0 = 4. \end{aligned}$$

Zapisujemy je w wierszu „ $\beta_j$ ” oraz kolumnie „ $\alpha_i$ ” w tablicy 2.2. Następnie liczymy  $\Delta_{ij}$  zgodnie ze wzorem (2.6). Wartość wskaźników optymalności zapisujemy w lewym, dolnym rogu pól niebazowych.

Ponieważ  $\Delta_{1,2} = -1$ ,  $\Delta_{3,1} = -5$ , to rozwiązanie nie jest optymalne. Trasą centralną jest trasa  $\langle 3, 1 \rangle$ , która generuje cykl  $C_{3,1}$  podany w tablicy 2.3.

TABLICA 2.3  
Schemat zmiany



TABLICA 2.4  
Rozwiązanie bazowe  $X^2$

|                      |    |    |    |       |
|----------------------|----|----|----|-------|
| $a_i \backslash b_j$ | 18 | 32 | 20 | $a_i$ |
| 20                   | 7  | 4  | 2  | 0     |
| 30                   | 4  | 5  | 13 | 0     |
| 20                   | 5  | 11 | 8  | 6     |
| $\beta_j$            | -1 | 5  | 2  |       |

Minimalny przewóz na trasach oznaczonych znakiem „-” wynosi:

$$\min\{18, 20\} = 18,$$

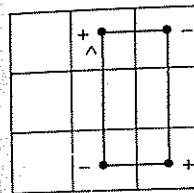
a więc na trasach tworzących cykl zmieniamy przewozy o 18 jednostek. Nowe rozwiązanie bazowe  $X^2$  zawiera tablica 2.4. Koszty zmniejszają się o  $5 \cdot 18 = 90$  zł, czyli wynoszą  $K^2 = K^1 - \Delta K^1 = 392 - 90 = 302$  zł.

Rozwiązanie  $X^2$  także nie jest optymalne, gdyż  $\Delta_{1,2} = -1$ . Cykl  $C_{1,2}$  jest podany w tablicy 2.5. Ponieważ minimalny przewóz na trasach oznaczonych symbolem „-” wynosi:

$$\min\{20, 2\} = 2,$$

to na trasach tworzących ten cykl zmieniamy przewozy o 2 jednostki. Nowe rozwiązanie  $X^3$  podane w tablicy 2.6 jest optymalne. Minimalny koszt transportu  $K^3 = 300$  zł.

TABLICA 2.5  
Schemat zmiany



TABLICA 2.6  
Rozwiązanie bazowe  $X^3$

|                      |    |    |    |       |
|----------------------|----|----|----|-------|
| $a_i \backslash b_j$ | 18 | 32 | 20 | $a_i$ |
| 20                   | 7  | 4  | 2  | 0     |
| 30                   | 4  | 5  | 13 | 1     |
| 20                   | 5  | 11 | 8  | 6     |
| $\beta_j$            | -1 | 4  | 2  |       |

ad 2. Jeżeli na trasie  $\langle 3, 2 \rangle$  obniżymy koszty o 2 zł, czyli  $c'_{3,2} = 9$ , to  $\Delta'_{3,2} = -1$  i tym samym rozwiązanie  $X^3$  nie będzie optymalne. Koszty zmniejszają się co najmniej o 2 zł, co łatwo sprawdzić, jeżeli poprawimy rozwiązanie  $X^3$  przy zmienionej macierzy kosztów  $C'$ .

### 2.3. Zagadnienie pośrednika

Pośrednik nabywa towar od  $m$  dostawców, przewozi i sprzedaje go  $n$  odbiorcom. Znamy:

- $a_i$  — maksymalną ilość towaru, jaką można kupić u  $i$ -tego dostawcy (jego podaż),
- $b_j$  — maksymalną ilość towaru, jaką można sprzedać  $j$ -temu odbiorcy (jego popyt),
- $k_i$  — cenę zakupu u  $i$ -tego dostawcy,

$p_j$  — cenę sprzedaży  $j$ -temu odbiorcy,

$c_{ij}$  — jednostkowy koszt transportu na trasie do  $i$ -tego dostawcy do  $j$ -tego odbiorcy

Należy ustalić taki plan zakupów, transportu i sprzedaży, aby dochód pośrednika był maksymalny.

$Dochód = przychód ze sprzedaży - koszt zakupu - koszt transportu.$

Zakładamy, że pośrednik nie tworzy zapasów, a więc zakupiona ilość towaru jest równa ilości sprzedanej. Oznaczmy ponadto przez:

$x_{ij}$  — wielkość przewozu na trasie  $\langle i, j \rangle$ ,

$d_{ij}$  — dochód jednostkowy z trasy  $\langle i, j \rangle$ ,  $d_{ij} = p_j - k_i - c_{ij}$ .

Model matematyczny zagadnienia pośrednika można zapisać następująco: znaleźć takie wartości zmiennych  $x_{ij}$ , aby:

$$\sum_{i=1}^m \sum_{j=1}^n d_{ij} x_{ij} \rightarrow \max, \quad (2.10)$$

$$\sum_{j=1}^n x_{ij} \leq a_i \quad (i = 1, \dots, m), \quad (2.11)$$

$$\sum_{i=1}^m x_{ij} \leq b_j \quad (j = 1, \dots, n), \quad (2.12)$$

$$x_{ij} \geq 0 \quad (i = 1, \dots, m; j = 1, \dots, n). \quad (2.13)$$

Funkcja celu (2.10) maksymalizuje dochód pośrednika. Warunek (2.11) zapewnia, że od każdego dostawcy wywozi się (i kupuje) nie więcej niż wynosi jego podaż. Natomiast warunek (2.12) gwarantuje, że każdemu odbiorcy dostarcza się (i sprzedaje) nie więcej niż wynosi jego popyt. Plan przewozu określa jednoznacznie plan zakupu i sprzedaży.

Zadanie (2.10)–(2.13) jest liniowym zadaniem decyzyjnym. Można je rozwiązać metodą potencjałów. Wymaga to następujących żądań.

Krok 1. Sprowadzamy zagadnienia do ZZT poprzez wprowadzenie:

1) fikcyjnego dostawcy z podażą  $a_{m+1} = \sum_{j=1}^n b_j$ ;

2) fikcyjnego odbiorcy z popytem  $b_{n+1} = \sum_{i=1}^m a_i$ ;

3) zerowych dochodów jednostkowych na trasach od fikcyjnego dostawcy oraz do fikcyjnego odbiorcy.

Krok 2. Wyznaczamy rozwiązanie wyjściowe, stosując metodę maksymalnego elementu macierzy, czyli rozwiązywanie zadania rozpoczynamy od tras najbardziej dochodowych.

Krok 3. Rozwiązanie jest optymalne, jeżeli:

$$\Delta_{ij} \leq 0 \quad \langle i, j \rangle \in N.$$

### PRZYKŁAD 2.2

Pośrednik kupuje towar od dwóch dostawców i sprzedaje go trzem odbiorcom. Podaż dostawców, popyt odbiorców, jednostkowe koszty transportu, ceny zakupu u dostawców oraz ceny sprzedaży odbiorcom (tys. zł) przedstawiono w tabelicy 2.7 ( $FO$  to fikcyjny odbiorca, a  $FD$  to fikcyjny dostawca):

- ustalić taki plan zakupu, przewozu i sprzedaży, aby dochód pośrednika był maksymalny,
- dla optymalnego rozwiązania ustalić przychód, koszt zakupu, koszt transportu i dochód.

ad 1.

TABLICA 2.7  
Dane liczbowe

| $a_i \backslash b_j$ | 18 | 32 | 20 | $k_i$ |
|----------------------|----|----|----|-------|
| 20                   | 5  | 3  | 8  | 6     |
| 30                   | 9  | 2  | 4  | 9     |
| $p_j$                | 15 | 14 | 16 |       |

TABLICA 2.8  
Rozwiązanie bazowe  $X^1$

| $a_i \backslash b_j$ | 15 | 12 | 18 | $FO$<br>50 | $\alpha_i$ |
|----------------------|----|----|----|------------|------------|
| 20                   | 4  | 5  | 2  | 0          | 0          |
|                      | 8  | 12 | -5 | -4         |            |
| 30                   | -3 | 3  | 3  | 0          | -4         |
|                      | -3 | 2  | 18 | 12         |            |
| $FD$                 | 0  | 0  | 0  | 0          | -4         |
| 45                   | 7  | -1 | -3 | 38         |            |
| $\beta_j$            | 4  | 5  | 7  | 4          |            |

Aby rozwiązać zadanie pośrednika sprowadzamy je do ZZT, wprowadzając fikcyjnego dostawcę z podażą  $a_3 = 45$  oraz fikcyjnego odbiorcę z popytem  $b_4 = 50$ . Dochody jednostkowe dla tras „fikcyjnych” są zerowe. Natomiast dochody jednostkowe dla tras „rzeczywistych” są następujące:

$$d_{11} = 15 - 6 - 5 = 4, \quad d_{12} = 14 - 6 - 3 = 5, \quad d_{13} = 16 - 6 - 8 = 2,$$

$$d_{21} = 15 - 9 - 9 = -3, \quad d_{22} = 14 - 9 - 2 = 3, \quad d_{23} = 16 - 9 - 4 = 3.$$

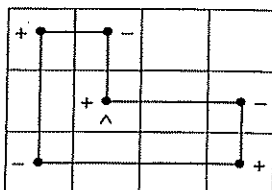
W tabelicy 2.8 dochody jednostkowe są podane w lewym, górnym rogu każdego pola. Rozwiązanie wyjściowe wyznaczamy, stosując metodę maksymalnego elementu macierzy dochodów. Przewozy zostały podane w tabelicy 2.8 w prawym, dolnym rogu każdego pola. Aby sprawdzić, czy wyznaczone rozwiązanie jest optymalne, liczymy zmienne  $\alpha_i, \beta_j$  tak samo jak dla ZZT oraz ustalamy  $\Delta_{ij}$  (w lewym, dolnym rogu każdego pola dla tras niebazowych).

Ponieważ  $\Delta_{22} = 2$ , to rozwiązanie  $X^1$  nie jest optymalne. Dochód  $D^1 = 4 \cdot 8 + 5 \cdot 12 + 3 \cdot 18 = 146$  tys. zł. Każda jednostka towaru przewożona po trasie  $\langle 2, 2 \rangle$  zwiększa dochód o 2 tys. zł. Ciekawa jest interpretacja przewozu na trasie  $\langle 3, 4 \rangle$ .  $x_{3,4} = 38$  określa, ile dobra zakupiono, przewozi się i sprzedaje.

Schemat zmiany rozwiązania jest podany w tablicy 2.9. Dla tras oznaczonych symbolem „-” minimalny przewóz wynosi  $\min\{12, 12, 7\} = 7$ .

Na trasach cyklu  $C_{2,2}$  zmieniamy przewozy o 7 jednostek. Nowe rozwiązanie jest optymalne. Zawiera je tablica (2.10).

TABLICA 2.9  
Schemat zmiany



TABLICA 2.10  
Rozwiązanie bazowe  $X^2$

| $a_i \backslash b_j$ | 15       | 12     | 18      | FO<br>50 | $\alpha_i$ |
|----------------------|----------|--------|---------|----------|------------|
| 20                   | 4<br>15  | 5<br>5 | 2<br>-3 | 0<br>-2  | 0          |
| 30                   | -3<br>-5 | 3<br>7 | 3<br>18 | 0<br>5   | -2         |
| FD                   | 0        | 0      | 0       | 0        | -2         |
| 45                   | -2       | -3     | -3      | 45       |            |
| $\beta_j$            | 4        | 5      | 5       | 2        |            |

Z rozwiązania łatwo odczytać, że od dostawcy 1 pośrednik kupuje 20 jednostek, a od dostawcy 2 — 25 jednostek. Natomiast odbiorcy 1 sprzedaje on 15 jednostek, odbiorcy 2 — 12 jednostek, a odbiorcy 3 — 18 jednostek.

ad 2.

Przychód ze sprzedaży:  $P = 15 \cdot 15 + 14 \cdot 12 + 16 \cdot 18 = 681$  tys. zł,

Koszt transportu:  $K_1 = 5 \cdot 15 + 3 \cdot 5 + 2 \cdot 7 + 4 \cdot 18 = 176$  tys. zł,

Koszt zakupu:  $K_2 = 6 \cdot 20 + 9 \cdot 25 = 345$  tys. zł,

Dochód pośrednika:  $D = P - K_1 - K_2 = 681 - 176 - 345 = 160$  tys. zł.

Dochód pośrednika możemy obliczyć także bezpośrednio, opierając się na funkcji celu (2.10):

$$D^2 = \sum_{i=1}^m \sum_{j=1}^n d_{ij} x_{ij} = 4 \cdot 15 + 5 \cdot 15 + 3 \cdot 7 + 3 \cdot 18 = 160 \text{ tys. zł.}$$

## 2.4. Zagadnienie transportowe z ograniczoną przepustowością tras

Mamy  $m$  dostawców i  $n$  odbiorców. Znamy podaż dostawców, popyt odbiorców i jednostkowe koszty transportu dla poszczególnych tras. Przepustowość pewnych lub wszystkich tras jest ograniczona. Należy ustalić plan dostaw minimalizujący koszty transportu.

Skorzystamy z tych samych oznaczeń, co w podrozdziale 2.1. Oznaczmy ponadto przez:

$H$  — zbiór tras z ograniczoną przepustowością,

$h_{ij}$  — przepustowość trasy  $\langle i, j \rangle$ .

Założmy także, że:

$$\sum_{i=1}^m a_i > \sum_{j=1}^n b_j. \quad (2.14)$$

Zadanie decyzyjne polega na wyznaczeniu  $x_{ij}$  w taki sposób, aby:

$$\sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min, \quad (2.15)$$

$$\sum_{j=1}^n x_{ij} \leq a_i \quad (i = 1, \dots, m), \quad (2.16)$$

$$\sum_{i=1}^m x_{ij} = b_j \quad (j = 1, \dots, n), \quad (2.17)$$

$$x_{ij} \leq h_{ij} \quad \langle i, j \rangle \in H, \quad (2.18)$$

$$x_{ij} \geq 0 \quad (i = 1, \dots, m; j = 1, \dots, n). \quad (2.19)$$

Zadanie (2.15)–(2.19) rozwiązujemy, stosując odpowiednio zmodyfikowaną metodą potencjałów. Modyfikacji ulega:

- 1) procedura wyznaczania rozwiązania wyjściowego;
- 2) procedura sprawdzania optymalności rozwiązania;
- 3) procedura zmiany rozwiązania.

ZZT i zagadnienie pośrednika zawsze mają rozwiązania dopuszczalne, dlatego wyjściowe rozwiązania łatwo wyznaczyć. Gdy niektóre trasy są niedopuszczalne lub mają ograniczoną przepustowość, wówczas zadanie może nie mieć rozwiązań dopuszczalnych lub mieć rozwiązanie dopuszczalne, lecz trudniej jest dla niego wyznaczyć rozwiązanie początkowe. Można w takim przypadku zastosować dwa podejścia:

- 1) procedurę zmodyfikowaną;
- 2) procedurę uogólnioną.

Procedura pierwsza polega na podziale macierzy kosztów  $C$  na kilka podmacierzy i rozpatrywanie ich w odpowiedniej kolejności. Stosujemy ją np. w dwuetapowym zagadnieniu transportowym. W naszym przypadku bezpiecznie jest zastosować procedurę drugą.

Procedura uogólniona polega na tym, że:

- 1) szukamy minimalnego elementu macierzy kosztów;
- 2) na wybranej trasie umieszczamy maksymalny przewóz, na jaki pozwalają aktualna podaż, popyt i przepustowość trasy.

Jeżeli na danej trasie  $\langle i, j \rangle \in H$  musimy umieścić przewóz większy od jego przepustowości:

$$x_{ij} > h_{ij},$$

to zmieniamy koszt tej trasy:

$$c_{ij} = M,$$

a więc nadmierny przewóz zostaje obciążony karą ( $M$  — dostatecznie duża liczba).

Jeżeli rozwiązanie bazowe  $X$  zadania (2.15)–(2.19) jest niezdegenerowane, to wszystkie trasy dzielimy na:

- 1) bazowe I rodzaju ( $B^1$ ), dla których  $0 < x_{ij} < h_{ij}$ ;
- 2) bazowe II rodzaju ( $B^2$ ), dla których  $x_{ij} = h_{ij}$ ;
- 3) niebazowe I rodzaju ( $N^1$ ), dla których  $x_{ij} = 0$ ;
- 4) niebazowe II rodzaju ( $N^2$ ), dla których  $x_{ij} = h_{ij}$ .

Wskaźniki optymalności  $\Delta_{ij}$  liczymy według wzoru:

$$\Delta_{ij} = c_{ij} - \alpha_i - \beta_j. \quad (2.21)$$

Jeżeli rozwiązanie bazowe  $X$  zadania (2.15)–(2.19) spełnia warunki:

$$B^2 = \emptyset, \quad (2.22)$$

$$\Delta_{ij} \geq 0, \quad \langle i, j \rangle \in N^1, \quad (2.23)$$

$$\Delta_{ij} \leq 0, \quad \langle i, j \rangle \in N^2, \quad (2.24)$$

to rozwiązanie  $X$  jest rozwiązaniem optymalnym.

W przypadku zadania na maksimum zamiast warunków (2.23) i (2.24) mamy:

$$\Delta_{ij} \leq 0, \quad \langle i, j \rangle \in N^1, \quad (2.25)$$

$$\Delta_{ij} \geq 0, \quad \langle i, j \rangle \in N^2. \quad (2.26)$$

### PRZYKŁAD 2.3

Trzy cukrownie:  $C_1, C_2, C_3$  dostarczają cukier do dwóch hurtowni  $H_1$  i  $H_2$ . Zdolności produkcyjne cukrowni (tys. t), popyt hurtowni (tys. t), jednostkowe koszty transportu (mln zł) dla poszczególnych tras oraz jednostkowe koszty produkcji zmienne (mln zł/tys. t) w cukrowniach zawiera tablica 2.11.

Hurtownia  $H_2$  wymaga, aby dostawy z cukrowni  $C_3$  nie przekroczyły 60% jej popytu. Ponadto zdolności produkcyjne wszystkich cukrowni muszą być wykorzystane w co najmniej 60%. Należy ustalić taki plan produkcji i dostaw, aby łączne koszty produkcji i transportu były minimalne.

TABLICA 2.11  
Dane liczbowe

|       | $H_1$ | $H_2$ | Zdolności produkcyjne | Koszty produkcji |
|-------|-------|-------|-----------------------|------------------|
| $C_1$ | 0,7   | 0,2   | 40                    | 0,7              |
| $C_2$ | 0,4   | 0,8   | 60                    | 0,8              |
| $C_3$ | 0,6   | 0,9   | 100                   | 0,6              |
| Popyt | 60    | 80    |                       |                  |

TABLICA 2.12  
Rozwiązanie bazowe  $X^1$

| $a_i \backslash b_j$ | 60   | 80  | FO        | $\alpha_i$ |
|----------------------|------|-----|-----------|------------|
| 40                   | 1,4  | 0,9 | 0 16      | 0          |
| 60                   | 9,3  | *24 | 9,1 16    | 9,1        |
| 100                  | 1,2  | 1,6 | 0 24      | 9,1        |
|                      |      | *36 | -8,4 0 24 |            |
|                      | 1,2  | 1,5 | 48 0 40   | 9,1        |
|                      | *24  | 10  | **56      | *20        |
| $\beta_j$            | -7,9 | 0,9 | -9,1      | —          |

Zbilansowane zadanie wymaga wprowadzenia fikcyjnego odbiorcy (FO) z popytem 60 i zerowymi kosztami. W każdym wierszu do kosztów transportu są dodane koszty produkcji (zob. tablica 2.12). Trasy w kolumnie FO oznaczają niewykorzystane zdolności produkcyjne. Ponieważ nie mogą być one większe od 40% zdolności, stąd trasy te mają ograniczone przepustowości zapisane w prawym, górnym rogu odpowiednich pól w tablicy 2.12. Wynoszą one kolejno: 16, 24 i 40. Także trasa  $\langle 3, 2 \rangle$  ma ograniczoną przepustowość, gdyż co najwyżej 60% dostaw, czyli 48 tys. t może pochodzić z cukrowni  $C_3$ . Trasy bazowe typu  $B^1$  zaznaczamy \*, a typu  $B^2$  — \*\*. Przewozy  $x_{13} = 16$  oraz  $x_{23} = 24$  (w tablicy 2.12) są na trasach należących do zbioru  $N^2$ .

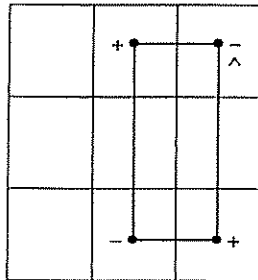
Wyznaczone w tablicy 2.12 rozwiązanie wyjściowe jest niedopuszczalne, gdyż na trasie  $\langle 3, 2 \rangle$  przewóz  $x_{32} = 56$  przekracza przepustowość tej trasy  $h_{32} = 48$ .

Dlatego trasie tej przydziela się karny koszt  $c_{32} = 10$ . Obciąża on przewozy większe od przepustowości. Dla tak uzyskanego rozwiązania zadania pomocniczego sprawdzamy, czy jest spełnione kryterium optymalności. Nie jest ono spełnione dla trasy  $\langle 1, 3 \rangle \in N^2$  oraz trasy  $\langle 2, 2 \rangle \in N^1$ ;  $\Delta_{13} = 9,1$  mówi nam, że zmniejszenie przewozu o jednostkę na trasie  $\langle 1, 3 \rangle$  zmniejszy koszt o 9,1. Ponieważ  $\max\{9, 1; |-8,4|\} = 9,1$ , to trasą centralną jest trasa  $\langle 1, 3 \rangle$ . Cykl zmiany rozwiązania generowany przez  $\langle 1, 3 \rangle$  jest podany w tabelicy 2.13.

Na trasach tworzących cykl zmieniamy przewozy o 8, gdyż  $\min\{16, 20, 8, \infty\} = 8$ . Nowe rozwiązanie  $X^2$ , już dopuszczalne, zawiera tabela 2.14. Kryterium optymalności nie jest spełnione dla trasy  $\langle 3, 2 \rangle$ .

Trasą centralną jest trasa  $\langle 3, 2 \rangle \in N^2$ . Cykl zmiany rozwiązania przedstawiony w tabelicy 2.15. Ponieważ  $\min\{48, \infty, 8, 12\} = 8$ , to na trasach cyklu zmieniamy przewozy o 8.

TABLICA 2.13  
Schemat zmiany

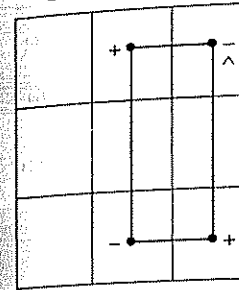


TABLICA 2.14  
Rozwiązanie bazowe  $X^2$

| $a_i \backslash b_j$ | 60         | 80         | FO       | $a_i$    |
|----------------------|------------|------------|----------|----------|
| 40                   | 1,4<br>0,2 | 0,9<br>*32 | 0<br>*8  | 16<br>8  |
| 60                   | 1,2<br>*36 | 1,6<br>0,7 | 0<br>0   | 24<br>24 |
| 100                  | 1,2<br>*24 | 1,5<br>0,6 | 48<br>48 | 0<br>*28 |
| $\beta_j$            | 1,2        | -0,9       | 0        |          |

Nowe rozwiązanie  $X^3$ , już optymalne, ukazuje tabela 2.16. Nie jest ono jedynym rozwiązaniem optymalnym, gdyż  $\Delta_{23} = 0$ . Łączne koszty produkcji i transportu wynoszą 168 mln zł. Cukrownia  $C_1$  produkuje 40 tys. t,  $C_2$  — 36 tys. t, a  $C_3$  — 64 tys. t.

TABLICA 2.15  
Schemat zmiany



TABLICA 2.16  
Rozwiązanie bazowe  $X^3$

| $a_i \backslash b_j$ | 60         | 80         | FO        | $a_i$    |
|----------------------|------------|------------|-----------|----------|
| 40                   | 1,4<br>0,8 | 0,9<br>*40 | 0<br>0,6  | 16<br>8  |
| 60                   | 1,2<br>*36 | 1,6<br>0,1 | 0<br>0    | 24<br>24 |
| 100                  | 1,2<br>*24 | 1,5<br>*40 | 48<br>*36 | 0<br>36  |
| $\beta_j$            | 0,6        | 0,9        | -0,6      |          |

## 2.5. Zagadnienie transportowe z kryterium czasu

Mamy  $m$  dostawców i  $n$  odbiorców dobra, które szybko powinno być dostarczone odbiorcom. Znamy podaż dostawców, popyt odbiorców oraz czasy dostawy od każdego dostawcy do każdego odbiorcy. Należy ustalić taki plan dostaw, aby najdłuższy czas dostawy był minimalny.

Oznaczmy przez:

$t_{ij}$  — czas dostawy dobra od  $i$ -tego dostawcy do  $j$ -tego odbiorcy,

$F^r$  — zbiór tras o dodatnich przewozach w  $r$ -tym rozwiązaniu bazowym,

$F^r = \{\langle i, j \rangle | x_{ij} > 0\}$ .

Zakładamy także, że

$$\sum_{i=1}^m a_i > \sum_{j=1}^n b_j \tag{2.27}$$

Zadanie decyzyjne polega na wyznaczeniu takich wartości zmiennych  $x_{ij}$ , aby:

$$\max\{t_{ij} | \langle i, j \rangle \in F^r\} \Rightarrow \min, \tag{2.28}$$

$$\sum_{j=1}^n x_{ij} = a_i \quad (i = 1, \dots, m), \tag{2.29}$$

$$\sum_{i=1}^m x_{ij} = b_j \quad (j = 1, \dots, n), \tag{2.30}$$

$$x_{ij} \geq 0 \quad (i = 1, \dots, m; j = 1, \dots, n). \tag{2.31}$$

Zadanie (2.28)–(2.31), ze względu na funkcję celu, nie jest zadaniem PL. Metodę potencjałów można wykorzystać tylko pośrednio.

Wyznaczamy dopuszczalne rozwiązanie bazowe  $X'$ , np. metodą MEM. Dla tego rozwiązania ustalamy maksymalny czas realizacji dostawy  $T'$ :

$$T' = \max \{t_{ij} | \langle i, j \rangle \in F'\}.$$

Następnie definiujemy macierz kosztów  $K' = [k'_{ij}]$ , gdzie:

$$k'_{ij} = \begin{cases} 0, & \text{jeżeli } t_{ij} < T', \\ 1, & \text{jeżeli } t_{ij} = T', \\ 10, & \text{jeżeli } t_{ij} > T'. \end{cases} \tag{2.32}$$

Wyznaczamy wartości zmiennych  $\alpha_i$  oraz  $\beta_j$  spełniające warunek:

$$\alpha_i + \beta_j = k'_{ij} \quad \langle i, j \rangle \in B. \tag{2.33}$$

Obliczamy wskaźniki optymalności:

$$\Delta_{ij} = k'_{ij} - \alpha_i - \beta_j. \tag{2.34}$$

Jeżeli dla ostatniego rozwiązania bazowego  $X'$ :

$$\Delta_{ij} \geq 0 \quad \langle i, j \rangle \in N, \tag{2.35}$$

to rozwiązanie jest optymalne. Maksymalny czas dostaw wynosi  $T'$ .

**PRZYKŁAD 2.4**

Trzy mleczarnie dostarczają mleko do trzech supermarketów. Znamy podaż w mleczarni, popyt supermarketów na mleko oraz czasy dostawy mleka (w godz.) od poszczególnych dostawców do odbiorców (tablica 2.17). Trzeba ustalić taki plan dostaw, aby najdłuższy czas dostawy był minimalny.

Stosując klasyczną metodę MEM, wyznaczamy rozwiązanie wyjściowe (podane w tablicy 2.18). Dla rozwiązania  $X^1$  ustalamy najdłuższy czas dostawy  $T^1 = \max \{7, 14, 3, 10, 5\} = 14$ . Zgodnie ze wzorem (2.32) definiujemy macierz kosztów  $K^1$  (tablica 2.18).

Ponieważ rozwiązanie  $X^1$  nie jest optymalne, to poprawiamy je zgodnie z schematem podanym w tablicy 2.19. Nowe rozwiązanie zawiera tablica 2.20.

TABLICA 2.17  
Dane liczbowe

|       |       |    |    |    |  |
|-------|-------|----|----|----|--|
|       | $b_j$ | 25 | 18 | 27 |  |
| $a_i$ |       |    |    |    |  |
| 30    |       | 7  | 14 | 6  |  |
| 30    |       | 8  | 4  | 3  |  |
| 30    |       | 9  | 10 | 5  |  |

TABLICA 2.19  
Schemat zmiany

|  |  |   |  |  |   |
|--|--|---|--|--|---|
|  |  | - |  |  | + |
|  |  |   |  |  | ^ |
|  |  |   |  |  |   |
|  |  | + |  |  | - |

TABLICA 2.21  
Schemat zmiany

|  |  |   |  |  |   |
|--|--|---|--|--|---|
|  |  |   |  |  |   |
|  |  | + |  |  | - |
|  |  |   |  |  | ^ |
|  |  | - |  |  | + |

TABLICA 2.18  
Rozwiązanie bazowe  $X^1$

|           |       |    |    |    |            |
|-----------|-------|----|----|----|------------|
|           | $b_j$ | 25 | 18 | 37 | $\alpha_i$ |
| $a_i$     |       |    |    |    |            |
| 30        |       | 0  | 1  | 0  | 0          |
|           |       | 25 | 5  | -1 |            |
| 20        |       | 0  | 0  | 0  | -1         |
|           |       | 1  | 0  | 20 |            |
| 30        |       | 0  | 0  | 0  | -1         |
|           |       | 1  | 13 | 17 |            |
| $\beta_j$ |       | 0  | 1  | 1  |            |

TABLICA 2.20  
Rozwiązanie bazowe  $X^2$

|           |       |    |    |    |            |
|-----------|-------|----|----|----|------------|
|           | $b_j$ | 25 | 18 | 37 | $\alpha_i$ |
| $a_i$     |       |    |    |    |            |
| 30        |       | 0  | 10 | 0  | 0          |
|           |       | 25 | 9  | 5  |            |
| 20        |       | 0  | 0  | 0  | 0          |
|           |       | 0  | -1 | 20 |            |
| 30        |       | 0  | 1  | 0  | 0          |
|           |       | 0  | 18 | 12 |            |
| $\beta_j$ |       | 0  | 1  | 0  |            |

TABLICA 2.22  
Rozwiązanie bazowe  $X^3$

|           |       |    |    |    |            |
|-----------|-------|----|----|----|------------|
|           | $b_j$ | 25 | 18 | 37 | $\alpha_i$ |
| $a_i$     |       |    |    |    |            |
| 30        |       | 1  | 10 | 0  | 0          |
|           |       | 25 | 10 | 5  |            |
| 20        |       | 10 | 0  | 0  | 0          |
|           |       | 9  | 18 | 2  |            |
| 30        |       | 10 | 10 | 0  | 0          |
|           |       | 9  | 10 | 30 |            |
| $\beta_j$ |       | 1  | 0  | 0  |            |

Dla rozwiązania  $X^2$  najdłuższy czas dostawy  $T^2 = \max \{7, 6, 3, 10, 5\} = 10$ . Definiujemy nową macierz kosztów  $K^2$  (podaną w tabelicy 2.20). Rozwiązanie  $X^2$  jest nieoptymalne. Schemat zmiany rozwiązania podano w tabelicy 2.21. Nowe rozwiązanie,  $X^3$ , przedstawiono w tabelicy 2.22. Rozwiązanie  $X^3$  spełnia już kryterium optymalności,  $T^3 = 7$  jest minimalnym czasem realizacji najdłuższej dostawy.

#### POLECENIA I PYTANIA KONTROLNE

1. Co trzeba zrobić, aby móc rozwiązać otwarte zagadnienie transportowe za pomocą metody potencjałów?
2. Kiedy w zagadnieniu pośrednika wprowadzamy fikcyjnego dostawcę lub fikcyjnego odbiorcę oraz z jaką podażą lub popytem?
3. Co to jest zmienna bazowa oraz zmienna niebazowa II rodzaju?
4. Jak uwzględnić w zagadnieniach transportowych rozwiązywanych metodą potencjałów, że pewne trasy są niedopuszczalne?
5. O czym mówi wskaźnik kryterium optymalności  $\Delta_{ij} = 2$  dla zagadnienia z pośrednikiem?
6. Kiedy zagadnienie transportowe ma zawsze rozwiązania dopuszczalne i łatwy sposób wyznaczyć rozwiązanie początkowe?

#### BIBLIOGRAFIA

- Buga J., Nykowski I., 1974, *Zagadnienia transportowe w programowaniu liniowym*, PWN, Warszawa.
- Guzik B. (red.), 2001, *Ekonometria i badania operacyjne. Zagadnienia podstawowe*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, nr 115.
- Sikora W., 1993, *Modele i metody optymalnej dystrybucji dóbr*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Zeszyt nr 127.
- Trzaskalik T., 2003, *Wprowadzenie do badań operacyjnych z komputerem*, PWE, Warszawa.

## ROZDZIAŁ 3 PRZEPIŁY WY W SIECIACH



### 3.1. Sieci

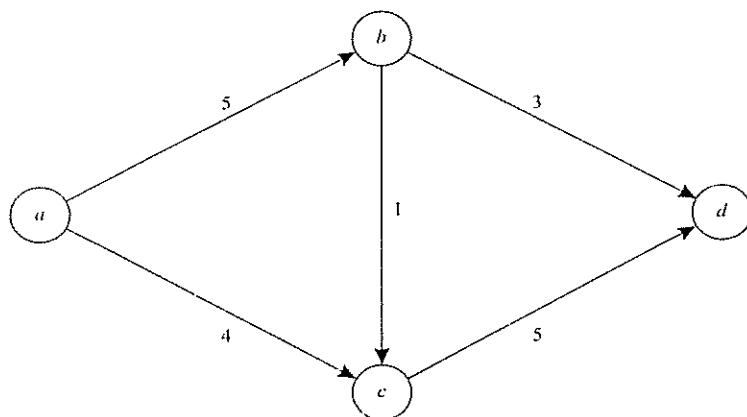
W praktyce gospodarczej znaleźć można wiele zagadnień, które dadzą się modelować za pomocą **sieci** lub, ogólniej, **grafów**. W czasie dotychczasowej lektury Czytelnik zapoznał się już ze szczególnym przykładem zagadnienia sieciowego, jakim jest zadanie transportowe, omówione w rozdziale 2. Przykładem zastosowania sieci jest również analiza czasowa przedsięwzięć, o której będzie mowa w rozdziale 7. Stanowiące szczególny typ grafów drzewa mają zastosowanie w modelowaniu decyzji sekwencyjnych (zob. rozdział 10).

Wyobraźmy sobie sieć połączeń różnych punktów geograficznych za pomocą ścieżek (dróg). Ścieżki te miejscami się przecinają, a pewne ich fragmenty są wspólne. Najkrótsze, niepodzielne fragmenty ścieżek nazywamy **lukami** (**krawędziami**) sieci, a punkty, w których się łączą — **węzłami** (**wierzchołkami**) sieci. Zakładamy, że luki są zorientowane, tzn. że dla każdego z nich możemy określić węzeł początkowy i węzeł końcowy. Węzeł, który nie jest końcem żadnego luku (a więc taki, z którego luki tylko „wychodzą”) nazywamy **źródłem**. Węzeł niebędący początkiem żadnego luku (a więc taki, do którego luki wyłącznie „wchodzą”) nazywamy **ujściem**. W ogólnym przypadku sieć może mieć wiele źródeł i wiele ujść. W dalszej części rozdziału skoncentrujemy się wyłącznie na sieciach z jednym źródłem i jednym ujściem.

W zależności od zastosowań lukom i węzłom można przypisać **etykiety** będące najczęściej liczbami (w takim przypadku nazywane są one często **wagami**).

Przykład sieci z jednym źródłem, jednym ujściem, etykietami przypisanymi węzłom i wagami przypisanymi wierzchołkom przedstawiono na rysunku 3.1. Łuki są reprezentowane przez strzałki, a węzły — przez kółka.

W tym rozdziale omówione są trzy zagadnienia związane z sieciami: problem maksymalnego przepływu, problem najkrótszej drogi i problem przepływu o minimalnym koszcie (ten ostatni jest w istocie uogólnieniem zadania transportowego).

RYSUNEK 3.1  
Przykład sieci

W przypadku każdego z omawianych zagadnień, po krótkim wprowadzeniu zostanie przedstawiony jego model matematyczny, a następnie wybrana metoda jego rozwiązywania. Całość jest zilustrowana przykładem. W podsumowaniu jest przedstawiona lista wybranych zastosowań.

Jeżeli nie będzie napisane inaczej, to sieć oznaczać będziemy przez  $N$ , zbiór węzłów przez  $V(N)$ , zbiór luków natomiast przez  $E(N)$ . Poszczególne węzły oznaczać będziemy małymi literami, w szczególności źródłu przypisana będzie litera  $s$ , a ujściu — litera  $t$ . Łuki oznaczać będziemy najczęściej jako pary węzłów (choć czasami niekiedy przypiszemy im odrębne oznaczenia). Zbiór węzłów sąsiadujących z węzłem  $u$  składa się z dwóch zbiorów:  $N^+(u)$  oznacza zbiór węzłów kończących łuki, które rozpoczynają się w  $u$  (czyli zbiór **następników węzła**  $u$ ), a  $N^-(u)$  zbiór węzłów rozpoczynających łuki kończące się w  $u$  (czyli zbiór **poprzedników węzła**  $u$ ). Odnosząc się do rysunku 3.1:  $V(N) = \{a, b, c, d\}$ ,  $E(N) = \{(a, b), (a, c), (b, c), (b, d), (c, d)\}$ ,  $s = a$ ,  $t = d$ ,  $N^+(a) = \{b, c\}$ ,  $N^+(b) = \{c, d\}$ ,  $N^+(c) = \{d\}$ ,  $N^+(d) = \emptyset$ ,  $N^-(a) = \emptyset$ ,  $N^-(b) = \{a\}$ ,  $N^-(c) = \{a, b\}$ ,  $N^-(d) = \{b, c\}$ .

## 3.2. Zagadnienie maksymalnego przepływu

Rozpatrzmy następujący problem. Należy przesłać jak największą ilość pewnego dobra ze źródła do ujścia sieci. Przesyłane dobro może zostać rozdzielone, tzn. można jego część transportować wybraną ścieżką, resztę zaś pozostałymi. Dla każdego łuku znana jest jego maksymalna przepustowość.

Wprowadźmy następujące oznaczenia:

- $y$  — całkowity przepływ,
- $x_{uv}$  — przepływ wzdłuż łuku  $(u, v)$ ,
- $h_{uv}$  — maksymalny dopuszczalny przepływ wzdłuż łuku  $(u, v)$ , czyli **przepustowość** łuku  $(u, v)$ .

Model matematyczny omawianego zagadnienia przyjmie wówczas następującą postać:

$$\begin{aligned}
 (0) \quad & y \rightarrow \max, \\
 (1) \quad & \sum_{v \in N^+(u)} x_{uv} = \sum_{v \in N^-(u)} x_{vu}, \quad u \in V(N) \setminus \{s, t\}, \\
 (2) \quad & \sum_{v \in N^+(s)} x_{sv} = y, \\
 (3) \quad & \sum_{v \in N^-(t)} x_{vt} = y, \\
 (4) \quad & 0 \leq x_{uv} \leq h_{uv}, \quad (u, v) \in E(N).
 \end{aligned} \tag{3.1}$$

Funkcja celu (0) odpowiada postulatowi maksymalizacji łącznego przepływu. Warunek (1) zapewnia, że łączny odpływ dobra z każdego węzła poza źródłem i ujściem jest równy jego przyptywowi do tego węzła. Warunki (2) i (3) zapewniają, że łączny odpływ ze źródła i łączny napływ do ujścia są równe łącznemu przepływowi w sieci. Wreszcie warunek (4) oznacza, że przepływ wzdłuż dowolnego łuku nie może przekroczyć maksymalnej dopuszczalnej wartości i nie może być ujemny.

Zadanie maksymalnego przepływu w sieci jest — jak widać — zadaniem programowania liniowego. Można je więc rozwiązać metodami ogólnymi, np. metodą sympleks. Ze względu na specyficzną postać warunków ograniczających (macierz ograniczeń jest bardzo rzadka, a jej wszystkie elementy są równe 1, 0 albo  $-1$ ) jest to nieefektywne. Dlatego też dla rozwiązywania tego typu zagadnień opracowano specjalne metody. Jedną z nich, algorytm Forda i Fulkersona, opiszemy w tym podrzdziale. Zanim jednak przejdziemy do opisu algorytmu, wprowadzimy kilka niezbędnych definicji i sformulujemy jedno twierdzenie konieczne do zrozumienia zasady jego działania.

### DEFINICJA 3.1

#### Przekrój sieci, przepustowość przekroju

Przekrój sieci to zbiór luków dzielący zbiór węzłów  $V(N)$  na dwa podzbiory  $V_1(N)$  i  $V_2(N)$  takie, że każdy łuk przekroju ma wierzchołek początkowy w  $V_1(N)$  i końcowy w  $V_2(N)$ , a ponadto  $s \in V_1(N)$  i  $t \in V_2(N)$ . Przepustowością przekroju nazywamy sumę przepustowości luków tworzących ten przekrój.



Każdy przekrój ma tę właściwość, że po usunięciu z sieci należących do niego luków dla dowolnej pary węzłów  $u$  i  $v$  takich, że  $u \in V_s(N)$  i  $v \in V_t(N)$  nie istnieje ścieżka z  $u$  do  $v$ . Przykładem przekroju w sieci na rysunku 3.1 jest zbiór luków  $\{(a, b), (c, d)\}$ . Gdyby wagi przypisane poszczególnym lukom reprezentowały przepustowości, wówczas przepustowość przekroju wynosiłaby  $h_{ab} + h_{cd} = 5 + 5 = 10$ . Odpowiednie zbiory węzłów są określone następująco:  $V_s(N) = \{a, c\}$ ,  $V_t(N) = \{b, d\}$ . Zauważmy, że istnieje ścieżka łącząca wierzchołki  $c$  i  $b$  (krawędź  $(b, c)$ ); nie jest to jednak ścieżka z  $c$  do  $b$ , lecz w przeciwnym kierunku.

### DEFINICJA 3.2 Sieć rezydualna

Sieć rezydualna to sieć  $N^*$  o tym samym zbiorze węzłów, co sieć  $N$  i zbiorze luków utworzonych w następujący sposób:

- jeżeli dla luku  $(u, v)$  sieci  $N$  jest spełniony warunek  $x_{uv} = 0$ , to w skład sieci  $N^*$  wchodzi tylko luk  $(u, v)$ , a jego przepustowość w sieci rezydualnej jest równa  $h_{uv}^* = h_{uv}$ ;
- jeżeli dla luku  $(u, v)$  sieci  $N$  jest spełniony warunek  $x_{uv} = h_{uv}$ , to w skład sieci  $N^*$  wchodzi tylko luk  $(v, u)$ , a jego przepustowość w sieci rezydualnej jest równa  $h_{vu}^* = h_{uv}$ ;
- jeżeli dla luku  $(u, v)$  sieci  $N$  jest spełniony warunek  $0 < x_{uv} < h_{uv}$ , to w skład sieci  $N^*$  wchodzi oba luki  $(u, v)$  i  $(v, u)$ , a ich przepustowości w sieci rezydualnej są równe, odpowiednio,  $h_{uv}^* = h_{uv} - x_{uv}$  i  $h_{vu}^* = x_{uv}$ .

Jak łatwo zauważyć, sieć rezydualną można zdefiniować równoważnie nieco inaczej. Każdemu lukowi w wyjściowej sieci przypisujemy luk o przeciwnym zwrocie, a przepustowości dla wszystkich definiujemy zgodnie z ostatnim wzorem. Następnie usuwamy wszystkie luki, których przepustowość jest równa 0.

Przyjmijmy, że przepływy w sieci z rysunku 3.1 są określone następująco (liczby w nawiasach na rysunku 3.2).

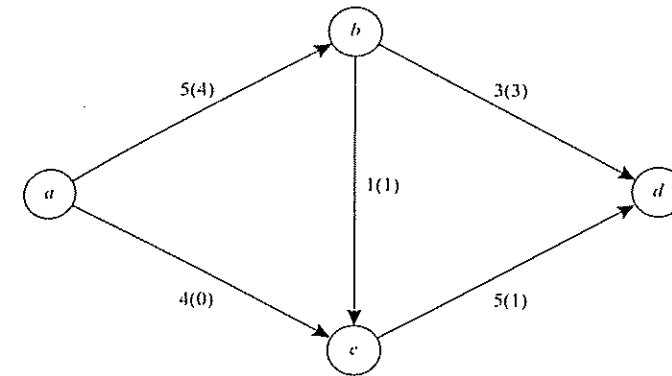
Wówczas sieć rezydualna przyjmie przedstawioną na rysunku 3.3 postać (liczby przy lukach oznaczają zmodyfikowane przepustowości).

### DEFINICJA 3.3

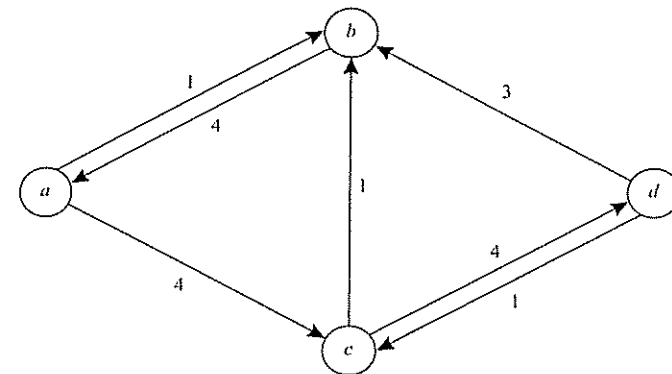
#### Ścieżka powiększająca, przepustowość ścieżki powiększającej

Ścieżka powiększająca to dowolna ścieżka w sieci rezydualnej łącząca źródło z ujściem. Przepustowość ścieżki powiększającej jest równa najmniejszej spośród przepustowości wszystkich luków wchodzących w jej skład.

RYSUNEK 3.2  
Przykładowe przepływy



RYSUNEK 3.3  
Sieć rezydualna



Ścieżkę powiększającą można zdefiniować też bez wprowadzania pojęcia sieci rezydualnej: jest to ścieżka w wyjściowej sieci niekoniecznie zachowująca zwroty luków, przy czym przepływy na lukach wchodzących w jej skład, których zwrot jest zgodny z jej kierunkiem, są mniejsze od przepustowości, a przepływy na lukach wchodzących w jej skład, których zwrot jest przeciwny do jej kierunku, są większe od 0. W takiej sytuacji przepustowość ścieżki jest definiowana jako minimum spośród przepustowości wszystkich luków o zgodnych zwrotach pomniejszonych o przepływy i przepływów na wszystkich lukach o przeciwnych zwrotach.

Jedyną ścieżką powiększającą na rysunkach 3.2 i 3.3 jest ścieżka  $(a, c, d)$ . Jej przepustowość jest równa  $\min\{h_{ac}^*, h_{cd}^*\} = \min\{4, 4\} = 4$ .

Po zaznajomieniu się z najważniejszymi pojęciami możemy sformułować twierdzenie, na którym opiera się algorytm rozwiązywania zadania maksymalnego przepływu.

### O maksymalnym przepływie i minimalnym przekroju (Ford i Fulkerson, 1955 r.<sup>1</sup>)

Załóżmy, że dana jest sieć  $N$ . Następujące trzy warunki są równoważne:

- największy przepływ w  $N$  jest równy  $y$ ;
- przepustowość najmniejszego przekroju w  $N$  jest równa  $y$ ;
- żadna sieć rezydualna  $N^*$  określona przez sieć  $N$  i przepływ  $y$  nie zawiera ścieżki powiększającej.

Idea działania algorytmu powinna być już jasna dla Czytelnika. W kolejnych krokach znajdujemy ścieżki powiększające, zwiększając przepływ wzdłuż każdej z nich o maksymalną możliwą wartość, czyli o jej przepustowość. Należy przy tym pamiętać, że powiększanie przepływu wzdłuż ścieżki powiększającej w sieci rezydualnej jest równoważne z powiększaniem przepływów w wyjściowej sieci na lukach o zwrotach zgodnych ze ścieżką i zmniejszaniu przepływów na lukach o przeciwnych zwrotach. Jeżeli w którymś kroku nie można wyznaczyć ścieżki powiększającej, oznacza to, że osiągnięty został maksymalny możliwy przepływ.

Metodę wyznaczania ścieżki powiększającej zaproponowali Edmonds i Karpa. Przed jej wprowadzeniem oryginalny algorytm Forda i Fulkersona mógł zachowywać się nieefektywnie nawet w przypadku bardzo małych zadań.

Algorytm generowania ścieżki autorstwa Edmondsa i Karpa wykorzystuje odczekanie węzłów. Każdemu węzłowi  $v$  są przy tym przypisane dwie etykiety:  $p(v)$  oznaczająca jego poprzednik i  $d(v)$  oznaczająca maksymalne możliwe zwiększenie przepływu. Przepustowość ścieżki powiększającej jest równa  $d(t)$ .

W każdym kroku rozpatrywane są trzy zbiory węzłów:

- $V_1$  — węzły odczekane i sprawdzone,
- $V_2$  — węzły odczekane i niesprawdzone,
- $V_3$  — węzły nieodczekane.

Algorytm Forda, Fulkersona, Edmondsa i Karpa można zapisać następująco:

#### Algorytm 3.1. Algorytm FFEK

Krok 1. Niech  $V_1 = \emptyset$ ,  $V_2 = \{s\}$ ,  $V_3 = V(N) \setminus \{s\}$ . Niech  $p(v) = „-”$  dla każdego  $v \in V(N)$ . Niech  $d(s) = \infty$  oraz  $d(v) = „-”$  dla każdego  $v \in V_3$ .

<sup>1</sup> W swojej pracy Ford i Fulkerson [1969] poszukiwali w istocie minimalnego przekroju, a nie maksymalnego przepływu. Miało to swoje uzasadnienie w militarnych zastosowaniach ich badań.

Krok 2. Jeżeli  $V_2 = \emptyset$ , stop. Nie ma ścieżki powiększającej. Aktualny przepływ jest największy, minimalny przekrój składa się z luków, których jeden węzeł znajduje się w zbiorze  $V_1$ , a drugi w zbiorze  $V_3$ . Jeżeli  $V_2 \neq \emptyset$  i  $t \in V_3$ , przejdź do kroku 3. Jeżeli  $t \in V_2$ , przejdź do kroku 4.

Krok 3. Niech  $v$  oznacza pierwszy węzeł ze zbioru  $V_2$ . Sprawdź go w następujący sposób. Najpierw sprawdź wszystkie poprzedniki  $u \in V_3$  węzła  $v$ . Jeżeli  $x_{uv} > 0$ , podstaw  $p(u) = „v”$ ,  $d(u) = \min\{d(v), x_{uv}\}$  i przesuń  $u$  z  $V_3$  na koniec  $V_2$ . Następnie sprawdź wszystkie następniki  $w \in V_3$  węzła  $v$ . Jeżeli  $x_{vw} < h_{vw}$ , podstaw  $p(w) = „v”$ ,  $d(w) = \min\{d(v), h_{vw} - x_{vw}\}$  i przesuń  $w$  z  $V_3$  na koniec  $V_2$ . Po sprawdzeniu wszystkich sąsiadów węzła  $v$  przesuń  $v$  z  $V_2$  na koniec  $V_1$ . Wróć do kroku 2.

Krok 4. Niech  $\delta = d(t)$ . Ścieżka powiększająca jest określona przez ciąg węzłów  $t, p(t), p(p(t)), \dots, s$  (w odwrotnej kolejności). Zmień przepływy na lukach wchodzących w skład ścieżki powiększającej zgodnie z zasadami:

- 1) jeżeli zwrot luku jest zgodny z kierunkiem ścieżki, zwiększ przepływ na tym luku o  $\delta$ ;
- 2) jeżeli zwrot luku jest przeciwny do kierunku ścieżki, zmniejsz przepływ na tym luku o  $\delta$ .

Wróć do kroku 1.

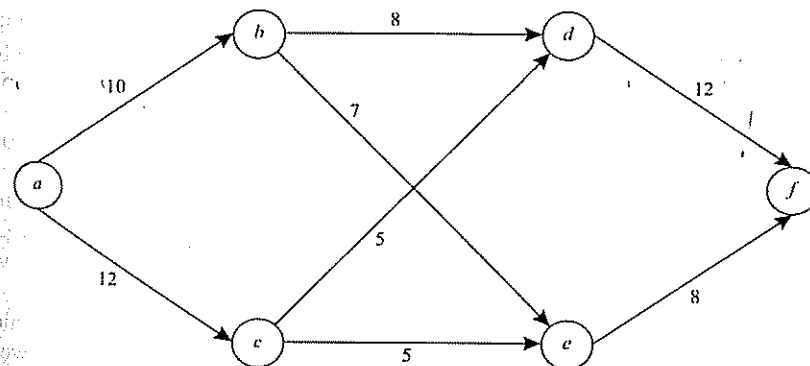
Działanie algorytmu zostanie zilustrowane przykładem liczbowym.

### PRZYKŁAD 3.1

#### Działanie algorytmu FFEK

Niech dana będzie sieć jak na rysunku 3.4. Wyznaczyć maksymalny przepływ (uwaga:  $s = a$ ,  $t = f$ ).

RYSUNEK 3.4  
Przykładowa sieć



Iteracje algorytmu zostały zapisane niżej. W każdej iteracji etykiety węzłów zostały naniesione na rysunek i przedstawione w tabelicy, przy czym na rysunkach znajdują się tylko ostateczne wartości etykiet. W części iteracji do tabelicy i rysunków dołączono również zwięzły komentarz.

Iteracja 1.

TABLICA 3.1  
Podzbiory zbioru węzłów i etykiety węzłów w trakcie iteracji 1

| Wiersz | $V_1$        | $V_2$     | $V_3$           | $a$    | $b$     | $c$     | $d$    | $e$    | $f$    |
|--------|--------------|-----------|-----------------|--------|---------|---------|--------|--------|--------|
| 1      | {-}          | {a}       | {b, c, d, e, f} | (-, ∞) | (-, -)  | (-, -)  | (-, -) | (-, -) | (-, -) |
| 2      | {a}          | {b, c}    | {d, e, f}       | (-, ∞) | (a, 10) | (a, 12) | (-, -) | (-, -) | (-, -) |
| 3      | {a, b}       | {c, d, e} | {f}             | (-, ∞) | (a, 10) | (a, 12) | (b, 8) | (b, 7) | (-, -) |
| 4      | {a, b, c}    | {d, e}    | {f}             | (-, ∞) | (a, 10) | (a, 12) | (b, 8) | (b, 7) | (-, -) |
| 5      | {a, b, c, d} | {e, f}    | {-}             | (-, ∞) | (a, 10) | (a, 12) | (b, 8) | (b, 7) | (d, 8) |
| 6      | {a, b, c, d} | {e, f}    | {-}             | (-, ∞) | (a, 10) | (a, 12) | (b, 8) | (b, 7) | (d, 8) |

Wiersz 1 (Krok 1). Ustalamy początkowe wartości etykiet.  $V_1$  jest zbiorem pustym, jedynym elementem  $V_2$  jest  $s = a$ , pozostałe węzły są elementami  $V_3$ .

Wiersz 2 (Krok 2).  $V_2$  nie jest zbiorem pustym, więc kontynuujemy. Ponieważ  $t = f \in V_3$ , przechodzimy do kroku 3. (Krok 3) pierwszym węzłem w  $V_2$  jest  $a$ . Nie ma on żadnych poprzedników w  $V_3$ . Ma dwa następniki w  $V_3$ :  $b$  i  $c$ . Skoro  $x_{ab} = 0 < h_{ab} = 10$ , ustalamy  $p(b) = a$  i  $d(b) = \min\{d(a), h_{ab} - x_{ab}\} = \min\{\infty, 10 - 0\} = 10$ . Przesuwamy  $b$  z  $V_3$  na koniec  $V_2$ . Skoro  $x_{ac} = 0 < h_{ac} = 12$ , ustalamy  $p(c) = a$  i  $d(c) = \min\{d(a), h_{ac} - x_{ac}\} = \min\{\infty, 12 - 0\} = 12$ . Przesuwamy  $c$  z  $V_3$  na koniec  $V_2$ . Przenosimy  $a$  z  $V_2$  do  $V_1$  i wracamy do kroku 2.

Wiersz 3 (Krok 2).  $V_2$  nie jest zbiorem pustym, więc kontynuujemy. Ponieważ  $t = f \in V_3$ , przechodzimy do kroku 3. (Krok 3) pierwszym węzłem w  $V_2$  jest  $b$ . Nie ma on żadnych poprzedników w  $V_3$ . Ma dwa następniki w  $V_3$ :  $d$  i  $e$ . Skoro  $x_{bd} = 0 < h_{bd} = 8$ , ustalamy  $p(d) = b$  i  $d(d) = \min\{d(b), h_{bd} - x_{bd}\} = \min\{10, 8 - 0\} = 8$ . Przesuwamy  $d$  z  $V_3$  na koniec  $V_2$ . Skoro  $x_{be} = 0 < h_{be} = 7$ , ustalamy  $p(e) = b$  i  $d(e) = \min\{d(b), h_{be} - x_{be}\} = \min\{10, 7 - 0\} = 7$ . Przesuwamy  $e$  z  $V_3$  na koniec  $V_2$ . Przenosimy  $b$  z  $V_2$  do  $V_1$  i wracamy do kroku 2.

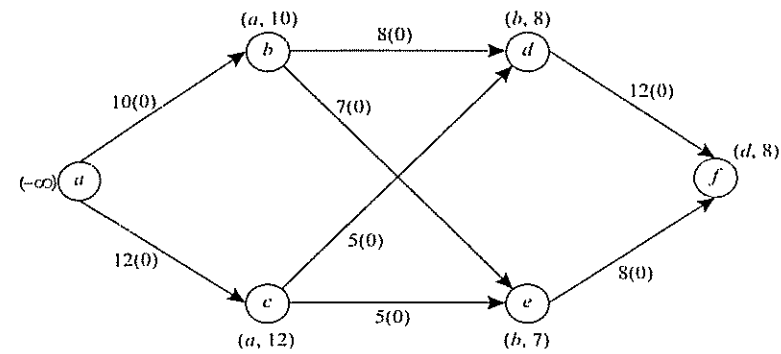
Wiersz 4 (Krok 2).  $V_2$  nie jest zbiorem pustym, więc kontynuujemy. Ponieważ  $t = f \in V_3$ , przechodzimy do kroku 3. (Krok 3) pierwszym węzłem w  $V_2$  jest  $c$ . Nie ma on żadnych poprzedników ani następników w  $V_3$ . Przenosimy  $c$  z  $V_2$  do  $V_1$  i wracamy do kroku 2.

Wiersz 5 (Krok 2).  $V_2$  nie jest zbiorem pustym, więc kontynuujemy. Ponieważ  $t = f \in V_3$ , przechodzimy do kroku 3. (Krok 3) pierwszym węzłem

w  $V_2$  jest  $d$ . Nie ma on żadnych poprzedników w  $V_3$ . Ma jeden następnik w  $V_3$ :  $f$ . Skoro  $x_{df} = 0 < h_{df} = 12$ , ustalamy  $p(f) = d$  i  $d(f) = \min\{d(d), h_{df} - x_{df}\} = \min\{8, 12 - 0\} = 8$ . Przesuwamy  $f$  z  $V_3$  na koniec  $V_2$ . Przenosimy  $d$  z  $V_2$  do  $V_1$  i wracamy do kroku 2.

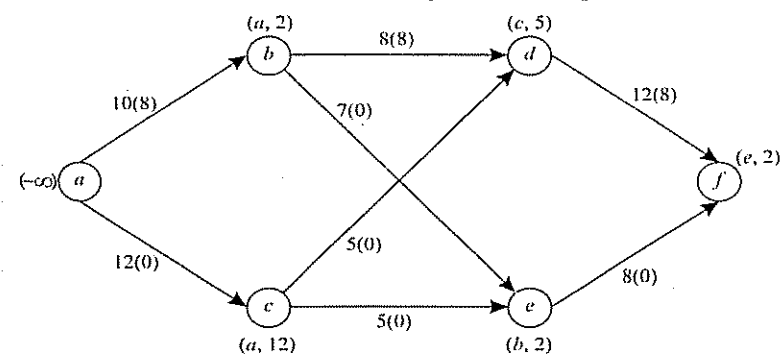
Wiersz 6 (Krok 2).  $V_2$  nie jest zbiorem pustym, więc kontynuujemy. Ponieważ  $t = f \in V_3$ , przechodzimy do kroku 4. (Krok 4)  $\delta = d(f) = 8$ . Ścieżka powiększająca jest postaci (w odwrotnym porządku):  $f, p(f) = d, p(d) = b, p(b) = a$ . Ponieważ wszystkie luki wchodzące w jej skład mają orientacje zgodną z kierunkiem ścieżki jako całości, zwiększamy przewozy na nich wszystkich o 8.

RYSUNEK 3.5  
Ostateczne wartości etykiet w iteracji 1



Iteracja 2.

RYSUNEK 3.6  
Ostateczne wartości etykiet w iteracji 2



TABLICA 3.2

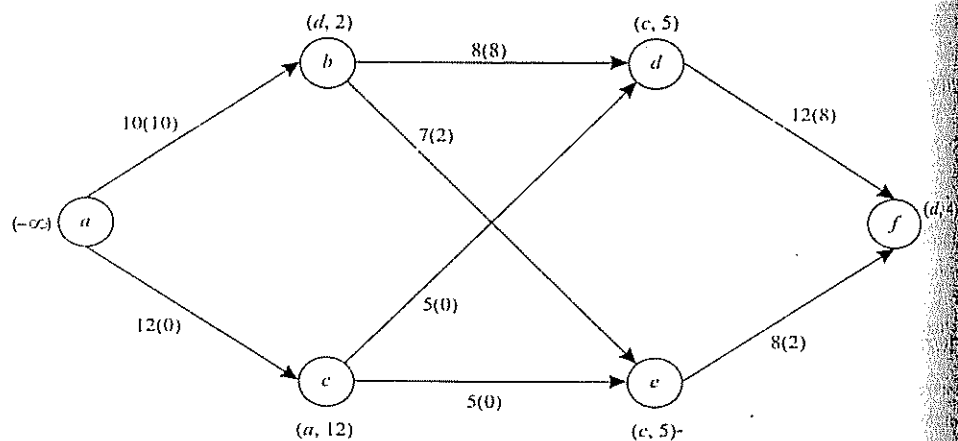
Podzbiory zbioru węzłów i etykiety węzłów w trakcie iteracji 2

| Wiersz | $V_1$        | $V_2$  | $V_3$           | $a$    | $b$    | $c$     | $d$    | $e$    | $f$    |
|--------|--------------|--------|-----------------|--------|--------|---------|--------|--------|--------|
| 1      | {-}          | {a}    | {b, c, d, e, f} | (-, ∞) | (-, -) | (-, -)  | (-, -) | (-, -) | (-, -) |
| 2      | {a}          | {b, c} | {d, e, f}       | (-, ∞) | (a, 2) | (a, 12) | (-, -) | (-, -) | (-, -) |
| 3      | {a, b}       | {c, e} | {d, f}          | (-, ∞) | (a, 2) | (a, 12) | (-, -) | (b, 2) | (-, -) |
| 4      | {a, b, c}    | {c, d} | {f}             | (-, ∞) | (a, 2) | (a, 12) | (c, 5) | (b, 2) | (-, -) |
| 5      | {a, b, c, e} | {d, f} | {-}             | (-, ∞) | (a, 2) | (a, 12) | (c, 5) | (b, 2) | (e, 2) |
| 6      | {a, b, c, e} | {d, f} | {-}             | (-, ∞) | (a, 2) | (a, 12) | (c, 5) | (b, 2) | (e, 2) |

Iteracja 3.

RYSUNEK 3.7

Ostateczne wartości etykiet w iteracji 3



TABLICA 3.3

Podzbiory zbioru węzłów i etykiety węzłów w trakcie iteracji 3

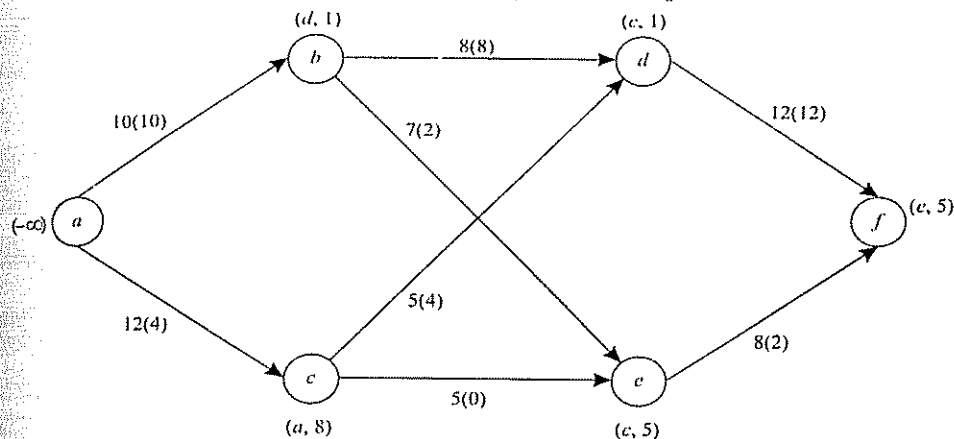
| Wiersz | $V_1$     | $V_2$     | $V_3$           | $a$    | $b$    | $c$     | $d$    | $e$    | $f$    |
|--------|-----------|-----------|-----------------|--------|--------|---------|--------|--------|--------|
| 1      | {-}       | {a}       | {b, c, d, e, f} | (-, ∞) | (-, -) | (-, -)  | (-, -) | (-, -) | (-, -) |
| 2      | {a}       | {c}       | {b, d, e, f}    | (-, ∞) | (-, -) | (a, 12) | (-, -) | (-, -) | (-, -) |
| 3      | {a, c}    | {d, e}    | {b, f}          | (-, ∞) | (-, -) | (a, 12) | (c, 5) | (c, 5) | (-, -) |
| 4      | {a, c, d} | {e, b, f} | {-}             | (-, ∞) | (d, 5) | (a, 12) | (c, 5) | (c, 5) | (d, 4) |
| 5      | {a, c, d} | {e, b, f} | {-}             | (-, ∞) | (d, 5) | (a, 12) | (c, 5) | (c, 5) | (d, 4) |

W trzeciej iteracji (wiersz 4) po raz pierwszy pojawia się cechowanie wierzchołka „wstecz”: pierwszym elementem  $V_2$  jest  $d$  mający jeden poprzednik w  $V_3$ :  $b$ . Ponieważ  $x_{bd} = 8 > 0$ , podstawiamy  $p(b) = d$  i  $d(b) = \min \{d(d), x_{bd}\} = \min \{5, 8\} = 5$ .

Iteracja 4.

RYSUNEK 3.8

Ostateczne wartości etykiet w iteracji 4



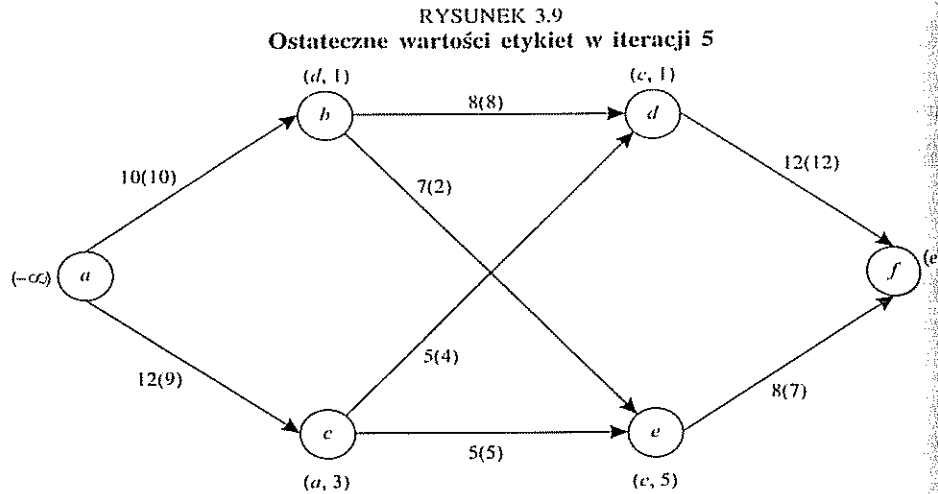
TABLICA 3.4

Podzbiory zbioru węzłów i etykiety węzłów w trakcie iteracji 4

| Wiersz | $V_1$        | $V_2$  | $V_3$           | $a$    | $b$    | $c$    | $d$    | $e$    | $f$    |
|--------|--------------|--------|-----------------|--------|--------|--------|--------|--------|--------|
| 1      | {-}          | {a}    | {b, c, d, e, f} | (-, ∞) | (-, -) | (-, -) | (-, -) | (-, -) | (-, -) |
| 2      | {a}          | {c}    | {b, d, e, f}    | (-, ∞) | (-, -) | (a, 8) | (-, -) | (-, -) | (-, -) |
| 3      | {a, c}       | {d, e} | {b, f}          | (-, ∞) | (-, -) | (a, 8) | (c, 1) | (c, 5) | (-, -) |
| 4      | {a, c, d}    | {e, b} | {f}             | (-, ∞) | (d, 1) | (a, 8) | (c, 1) | (c, 5) | (-, -) |
| 5      | {a, c, d, e} | {b, f} | {-}             | (-, ∞) | (d, 1) | (a, 8) | (c, 1) | (c, 5) | (e, 5) |
| 6      | {a, c, d, e} | {b, f} | {-}             | (-, ∞) | (d, 1) | (a, 8) | (c, 1) | (c, 5) | (e, 5) |

Iteracja 5.

Podobnie jak w dwóch poprzednich, pojawia się cechowanie węzła „wstecz” (wiersz 4). Tym razem łuk  $(b, d)$  wchodzi w skład ścieżki powiększającej, a więc pod koniec iteracji przepływ wzdłuż tego łuku zostaje zmniejszony o  $d(f) = 1$  jednostkę.

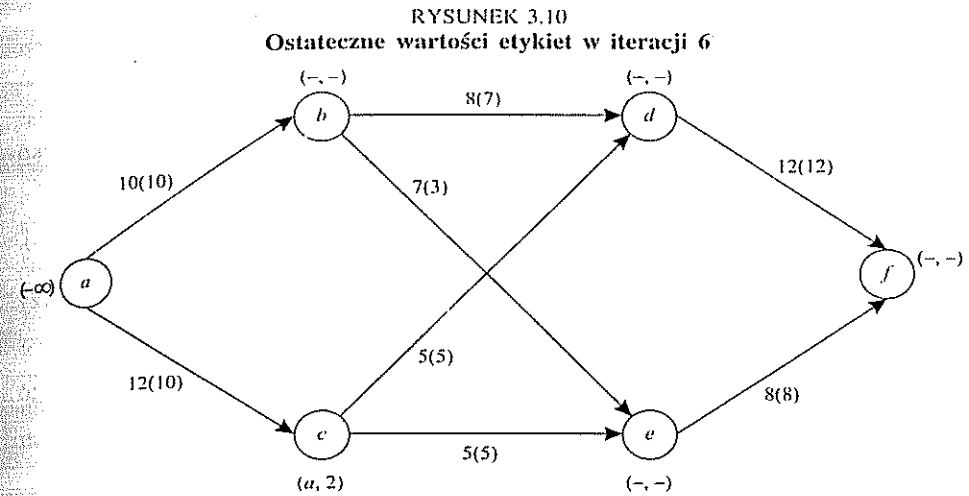


**TABLICA 3.5**  
Podzbiory zbioru węzłów i etykiety węzłów w trakcie iteracji 5

| Wiersz | $V_1$           | $V_2$ | $V_3$           | $a$    | $b$    | $c$    | $d$    | $e$    | $f$    |
|--------|-----------------|-------|-----------------|--------|--------|--------|--------|--------|--------|
| 1      | {-}             | {a}   | {b, c, d, e, f} | (-, ∞) | (-, -) | (-, -) | (-, -) | (-, -) | (-, -) |
| 2      | {a}             | {c}   | {b, d, e, f}    | (-, ∞) | (-, -) | (a, 3) | (-, -) | (-, -) | (-, -) |
| 3      | {a, c}          | {d}   | {b, e, f}       | (-, ∞) | (-, -) | (a, 3) | (c, 1) | (-, -) | (-, -) |
| 4      | {a, c, d}       | {b}   | {e, f}          | (-, ∞) | (d, 1) | (a, 3) | (c, 1) | (-, -) | (-, -) |
| 5      | {a, c, d, b}    | {e}   | {f}             | (-, ∞) | (d, 1) | (a, 3) | (c, 1) | (b, 1) | (-, -) |
| 6      | {a, c, d, b, e} | {f}   | {-}             | (-, ∞) | (d, 1) | (a, 3) | (c, 1) | (b, 1) | (e, 1) |
| 7      | {a, c, d, b, e} | {f}   | {-}             | (-, ∞) | (d, 1) | (a, 3) | (c, 1) | (b, 1) | (e, 1) |

#### Iteracja 6.

W iteracji 6 (wiersz 4) okazuje się, że niemożliwe jest wyznaczenie kolejnej ścieżki powiększającej. Zbiór  $V_2$  jest pusty, natomiast  $f \in V_3$ . Wynika stąd, że uzyskany przepływ  $y = 20$  jest największym możliwym. Przepływy wzdłuż łuków równe są liczbom w nawiasach na rysunku 3.10. Minimalny przekrój tworzą łuki o początku w  $V_1$  i końcu w  $V_3$ , a więc  $(a, b)$ ,  $(c, d)$  i  $(c, e)$ . Oczywiście algorytm FFEK wskazał tylko jeden przykładowy minimalny przekrój, znalezienie innych (o ile istnieją) pozostawiamy Czytelnikowi.



**TABLICA 3.6**  
Podzbiory zbioru węzłów i etykiety węzłów w trakcie iteracji 6

| Wiersz | $V_1$  | $V_2$ | $V_3$           | $a$    | $b$    | $c$    | $d$    | $e$    | $f$    |
|--------|--------|-------|-----------------|--------|--------|--------|--------|--------|--------|
| 1      | {-}    | {a}   | {b, c, d, e, f} | (-, ∞) | (-, -) | (-, -) | (-, -) | (-, -) | (-, -) |
| 2      | {a}    | {c}   | {b, d, e, f}    | (-, ∞) | (-, -) | (a, 2) | (-, -) | (-, -) | (-, -) |
| 3      | {a, c} | {-}   | {b, d, e, f}    | (-, ∞) | (-, -) | (a, 2) | (-, -) | (-, -) | (-, -) |
| 4      | {a, c} | {-}   | {b, d, e, f}    | (-, ∞) | (-, -) | (a, 2) | (-, -) | (-, -) | (-, -) |

Opisane wyżej zadanie maksymalnego przepływu ma wiele praktycznych zastosowań. Pierwsze z nich to oryginalny problem rozwiązany przez Forda i Fulkersona. Mieli oni za zadanie określić, w jaki sposób najskuteczniej sparaliżować ruch pociągów po wschodniej stronie Żelaznej Kurtyny (jak widać, rozwiązali raczej problem minimalnego przekroju niż maksymalnego przepływu). Idąc dalej tym tropem, można zastosować zadanie maksymalnego przepływu do wyznaczania słabych punktów tzw. wąskich gardel w sieciach komunikacyjnych i telekomunikacyjnych. Ich znajomość może pozwolić z jednej strony na zwiększenie bezpieczeństwa sieci, z drugiej zaś na efektywne zwiększenie jej przepustowości.

Zagadnienie maksymalnego przepływu może również posłużyć jako zadanie pomocnicze w problemach innego typu. Jako najważniejsze należy wymienić niektóre warianty zagadnienia szeregowania zadań, zagadnienia przydziału i wyznaczanie czynności do skrócenia w czasowo-kosztowej analizie przedsięwzięcia.

### 3.3. Zagadnienie najkrótszej drogi

Problem najkrótszej drogi jest jednym z najlepiej znanych i najdokładniej zbadanych zagadnień optymalizacji na sieciach. Jego ważność bierze się nie tylko z bezpośrednich zastosowań, lecz także z faktu, że zagadnienie to jest podproblemem w bardziej skomplikowanych zagadnieniach optymalizacyjnych.

Załóżmy, że każdemu łukowi sieci  $(u, v)$  przypisana została jego długość  $l(u, v)$ . Problem polega na znalezieniu najkrótszej drogi ze źródła albo do ujścia, albo do wszystkich (poza źródłem) węzłów w sieci. Można przy tym wyróżnić trzy przypadki:

- 1) długości wszystkich łuków są nieujemne;
- 2) w sieci występują łuki o ujemnych długościach, nie tworzą jednak żadnego cyklu o ujemnej długości;
- 3) w sieci występują cykle o ujemnej długości.

W trzecim przypadku rozwiązanie problemu najkrótszej drogi jest niemożliwe. Drugi przypadek jest uogólnieniem pierwszego. Ponadto rozwiązywanie tego typu problemów będzie niezbędne w dalszej części rozdziału. Z tego też powodu zajmiemy się tylko tym drugim przypadkiem. Interesować nas przy tym będzie odnalezienie najkrótszej drogi ze źródła do wszystkich pozostałych węzłów.

Oznaczmy przez  $d(v)$  najkrótszą odległość ze źródła do węzła  $v$ . Wówczas zadanie wyznaczenia najkrótszej drogi można zapisać następująco:

$$(0) \quad Md(t) + \sum_{v \in V(N) \setminus \{s, t\}} d(v) \rightarrow \max,$$

$$(1) \quad d(v) - d(u) \leq l(u, v), \quad (u, v) \in E(N),$$

$$(2) \quad d(s) = 0.$$
(3.2)

W modelu (3.2)  $M$  oznacza odpowiednio dużą liczbę. Warunek (2) oznacza, że odległość ze źródła do źródła wynosi 0. Warunki (1) łącznie z funkcją celu oznaczają, że dystans od źródła do dowolnego węzła musi być nie większy niż dystans mierzony dowolną drogą, z drugiej jednak strony maksymalny możliwy (czyli ostatecznie równy najmniejszemu spośród dystansów mierzonych różnymi drogami). Wysoka waga funkcji celu przy zmiennej reprezentującej odległość do ujścia zapewnia, że odległość do ujścia będzie rzeczywiście minimalna (w przypadku przypisania wszystkim zmiennym jednakowych wag zadanie mogłoby mieć nieskończenie wiele optimum, w których wartości zmiennych różniłyby się o stałą liczbę; w takiej sytuacji poszukiwane rozwiązanie byłoby jednym z nich, ale nie można by jednoznacznie określić którym).

Podane zadanie jest zadaniem programowania liniowego. Ze względu jednak na swój sieciowy charakter ma specyficzną postać macierzy ograniczeń i rozwiązywanie go ogólnymi metodami nie jest efektywne. Zauważmy też, że rozwiązanie

powyższego zadania pozwala jedynie stwierdzić, jakie są najkrótsze odległości ze źródła do pozostałych węzłów. Rozwiązując je, nie otrzymamy jednak odpowiedzi, które łuki tworzą najkrótszą drogę<sup>2</sup>. Z tych też powodów dla zagadnienia najkrótszej drogi opracowano wiele specjalnych algorytmów. W dalszej części tego podrozdziału przedstawiony zostanie jeden z nich, autorstwa Bellmana i Moore'a, z późniejszymi poprawkami d'Escopo i Papego. Algorytm ten jest często określany krótko jako algorytm Bellmana. W dalszej części będziemy się do niego odwoływać jako do algorytmu BMEP. Jego idea polega na sprawdzaniu w każdym kroku, czy istnieje możliwość skrócenia drogi do któregośkolwiek węzła poprzez zmianę jego poprzednika na dotychczasowej najkrótszej ścieżce. Informacje na temat poprzedników i dystansów są przechowywane przy użyciu dwóch etykiet przypisanych każdemu węzłowi  $v$ :  $d(v)$  reprezentującej dystans i  $p(v)$  reprezentującej poprzednik. Wierzchołki, które należy jeszcze sprawdzić, są przechowywane w kolejce  $Q$ . Algorytm BMEP można zapisać następująco.

#### Algorytm 3.2. Algorytm BMEP

- Krok 1. Dla wszystkich węzłów  $v \in V(N)$  przyjmij  $p(v) = \text{„-”}$ , dla wszystkich węzłów  $v \in V(N) \setminus \{s\}$  podstaw  $d(v) = \infty$  oraz  $d(s) = 0$ . Niech  $Q = \{s\}$ .
- Krok 2. Jeżeli  $Q = \emptyset$ , stop. Otrzymane rozwiązanie jest optymalne. Najkrótsza odległość do każdego węzła  $v$  jest reprezentowana przez  $d(v)$ , a najkrótsza ścieżka przez ciąg węzłów  $v, p(v), p(p(v)), \dots, s$  (w odwrotnej kolejności). Jeżeli  $Q \neq \emptyset$ , przejdź do kroku 3.
- Krok 3. Usuń pierwszy węzeł  $u$  z  $Q$ . Wykonaj krok 4 dla każdego następnika  $v$  węzła  $u$  i wróć do kroku 2.
- Krok 4. Jeżeli  $d(v) > l(u, v) + d(u)$ , podstaw  $d(v) = l(u, v) + d(u)$  i przyjmij  $p(v) = u$ . Ponadto, jeżeli  $v \notin Q$ :
  - 1) to w przypadku, gdy  $v$  należał do  $Q$  w którejś z poprzednich iteracji, dodaj  $v$  na początku  $Q$ ;
  - 2) to w przypadku, gdy  $v$  jeszcze nie należał do  $Q$ , dodaj  $v$  na końcu  $Q$ .

Działanie algorytmu zilustrowano przykładem liczbowym.

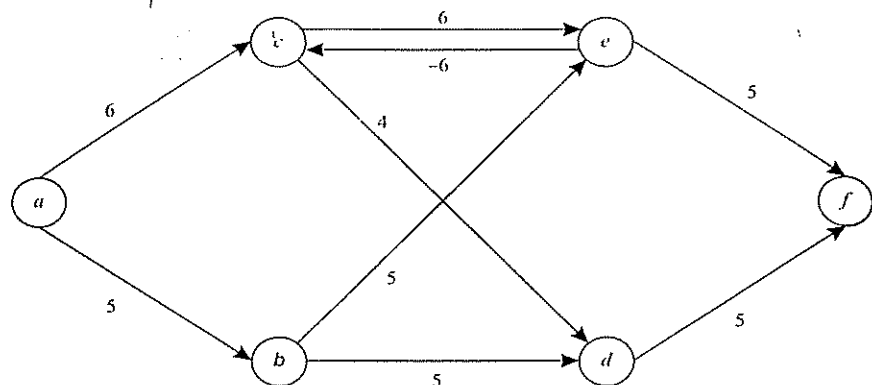
#### PRZYKŁAD 3.2

##### Działanie algorytmu BMEP

Niech dana będzie sieć jak na rysunku 3.11. Znaleźć najkrótszą drogę ze źródła do ujścia i pozostałych węzłów (uwaga:  $s = a, t = f$ ).

<sup>2</sup> Dokładniej, jesteśmy w stanie określić tylko, które łuki na pewno nie należą do najkrótszych dróg. Są to mianowicie łuki, dla których warunki z grupy (1) są niewiążące (spełnione z nierównościami).

RYСУNEK 3.11  
Przykładowa sieć



Poniżej przedstawione zostały wszystkie iteracje algorytmu w wersji tabelarycznej (dla każdej z nich podano postać kolejki  $Q$  oraz wartości etykiet węzłów). Poszczególne kroki zostały opatrzone komentarzami.

TABLICA 3.7  
Kolejka i etykiety węzłów w algorytmie BMEP

| Iteracja | $Q$           | $a$    | $b$            | $c$            | $d$            | $e$            | $f$            |
|----------|---------------|--------|----------------|----------------|----------------|----------------|----------------|
| 0        | { $a$ }       | (-, 0) | (-, $\infty$ ) | (-, $\infty$ ) | (-, $\infty$ ) | (-, $\infty$ ) | (-, $\infty$ ) |
| 1        | { $b, c$ }    | (-, 0) | ( $a, 5$ )     | ( $a, 6$ )     | (-, $\infty$ ) | (-, $\infty$ ) | (-, $\infty$ ) |
| 2        | { $c, d, e$ } | (-, 0) | ( $a, 5$ )     | ( $a, 6$ )     | ( $b, 10$ )    | ( $b, 10$ )    | (-, $\infty$ ) |
| 3        | { $d, e$ }    | (-, 0) | ( $a, 5$ )     | ( $a, 6$ )     | ( $b, 10$ )    | ( $b, 10$ )    | (-, $\infty$ ) |
| 4        | { $e, f$ }    | (-, 0) | ( $a, 5$ )     | ( $a, 6$ )     | ( $b, 10$ )    | ( $b, 10$ )    | ( $d, 15$ )    |
| 5        | { $c, f$ }    | (-, 0) | ( $a, 5$ )     | ( $e, 4$ )     | ( $b, 10$ )    | ( $b, 10$ )    | ( $d, 15$ )    |
| 6        | { $d, f$ }    | (-, 0) | ( $a, 5$ )     | ( $e, 4$ )     | ( $c, 8$ )     | ( $b, 10$ )    | ( $d, 15$ )    |
| 7        | { $f$ }       | (-, 0) | ( $a, 5$ )     | ( $e, 4$ )     | ( $c, 8$ )     | ( $b, 10$ )    | ( $d, 13$ )    |
| 8        | {-}           | (-, 0) | ( $a, 5$ )     | ( $e, 4$ )     | ( $c, 8$ )     | ( $b, 10$ )    | ( $d, 13$ )    |

Iteracja 0 (krok 1). Nadajemy początkowe wartości etykietom wierzchołków. Jedynym elementem  $Q$  jest  $a$ .

Iteracja 1 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $a$ , pierwszy (i jedyny) element  $Q$ . Węzeł  $a$  ma dwa następniki:  $b$  i  $c$ . Skoro  $d(b) = \infty > l(a, b) + d(a) = 5 + 0 = 5$ , podstawiamy  $d(b) = 5$  i  $p(b) = a$ . Po-

nieważ  $b$  nie należał jeszcze do kolejki  $Q$ , umieszczamy go na jej końcu. Skoro  $d(c) = \infty > l(a, c) + d(a) = 6 + 0 = 6$ , podstawiamy  $d(c) = 6$  i  $p(c) = a$ . Ponieważ  $c$  nie należał jeszcze do kolejki  $Q$ , umieszczamy go na jej końcu. Wracamy do kroku 2.

Iteracja 2 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $b$ , pierwszy element  $Q$ . Węzeł  $b$  ma dwa następniki:  $d$  i  $e$ . Skoro  $d(d) = \infty > l(b, d) + d(b) = 5 + 5 = 10$ , podstawiamy  $d(d) = 10$  i  $p(d) = b$ . Ponieważ  $d$  nie należał jeszcze do kolejki  $Q$ , umieszczamy go na jej końcu. Skoro  $d(e) = \infty > l(b, e) + d(b) = 5 + 5 = 10$ , podstawiamy  $d(e) = 10$  i  $p(e) = b$ . Ponieważ  $e$  nie należał jeszcze do kolejki  $Q$ , umieszczamy go na jej końcu. Wracamy do kroku 2.

Iteracja 3 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $c$ , pierwszy element  $Q$ . Węzeł  $c$  ma dwa następniki:  $d$  i  $e$ . Skoro  $d(d) = 10 = l(c, d) + d(c) = 4 + 6 = 10$ , nic nie zmieniamy. Skoro  $d(e) = 10 < l(c, e) + d(c) = 6 + 6 = 12$ , nic nie zmieniamy. Wracamy do kroku 2.

Iteracja 4 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $d$ , pierwszy element  $Q$ . Węzeł  $d$  ma jeden następnik:  $f$ . Skoro  $d(f) = \infty > l(d, f) + d(d) = 5 + 10 = 15$ , podstawiamy  $d(f) = 15$  i  $p(f) = d$ . Ponieważ  $f$  nie należał jeszcze do kolejki  $Q$ , umieszczamy go na jej końcu. Wracamy do kroku 2.

Iteracja 5 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $e$ , pierwszy element  $Q$ . Węzeł  $e$  ma dwa następniki:  $c$  i  $f$ . Skoro  $d(c) = 6 > l(e, c) + d(e) = -6 + 10 = 4$ , podstawiamy  $d(c) = 4$  i  $p(c) = e$ . Ponieważ  $c$  należał już do kolejki  $Q$ , umieszczamy go na jej początku. Skoro  $d(f) = 15 = l(e, f) + d(e) = 5 + 10 = 15$ , nic nie zmieniamy. Wracamy do kroku 2.

Iteracja 6 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $c$ , pierwszy element  $Q$ . Węzeł  $c$  ma dwa następniki:  $d$  i  $e$ . Skoro  $d(d) = 10 > l(c, d) + d(c) = 4 + 4 = 8$ , podstawiamy  $d(d) = 8$  i  $p(d) = c$ . Ponieważ  $d$  należał już do kolejki  $Q$ , umieszczamy go na jej początku. Skoro  $d(e) = 10 = l(c, e) + d(c) = 6 + 4 = 10$ , nic nie zmieniamy. Wracamy do kroku 2.

Iteracja 7 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $d$ , pierwszy element  $Q$ . Węzeł  $d$  ma jeden następnik:  $f$ . Skoro  $d(f) = 15 > l(d, f) + d(d) = 5 + 8 = 13$ , podstawiamy  $d(f) = 13$  i  $p(f) = d$ . Ponieważ  $f$  należał już do kolejki  $Q$ , nie musimy go do niej dodawać.

Iteracja 8 (kroki 2–4).  $Q$  jest niepusta, więc kontynuujemy. Usuujemy  $f$ , pierwszy element  $Q$ . Węzeł  $f$  nie ma następników, więc wracamy do kroku 2.

Iteracja 9 (krok 2; nie ujęta w tablicy).  $Q$  jest pusta, więc algorytm kończy swoje działanie. Jak widać, długość najkrótszej drogi ze źródła do ujścia wynosi  $d(t) = d(f) = 13$ , a sama najkrótsza droga jest postaci (w odwrotnym porządku):  $t = f, p(f) = d, p(d) = c, p(c) = e, p(e) = b, p(b) = a = s$ .

Odnalezienie najkrótszych dróg do pozostałych węzłów i odczytanie ich długości pozostawiamy Czytelnikowi.

Problem najkrótszej drogi ma liczne zastosowania. Najbardziej oczywiste i naturalne jest — rzecz jasna — odnajdywanie najkrótszych połączeń w sieciach komunikacyjnych i telekomunikacyjnych. Nie jest to jednak jedyny sposób jego wykorzystywania w praktyce. Zadania znalezienia najkrótszej drogi pojawiają się m.in. w systemach rozpoznawania mowy i podczas segmentacji obrazów. Mają również zastosowanie jako zadania pomocnicze w innego typu zagadnieniach optymalizacji na sieciach. Przykładem takiego zagadnienia jest problem przepływu o minimalnym koszcie, opisany w kolejnym podrozdziale.

### 3.4. Zagadnienie przepływu o minimalnym koszcie

Zagadnienie przepływu o minimalnym koszcie wiąże się w pewien sposób z zagadnieniem maksymalnego przepływu, opisany w podrozdziale 3.2. Podobnie jak tam, dana jest sieć  $N$ , a dobro transportowane jest ze źródła do ujścia. Dane są też maksymalne przepustowości  $h_{uv}$  poszczególnych luków. Różnica polega na tym, że tym razem nie poszukujemy maksymalnego przepływu, ale staramy się znaleźć przepływ o określonej z góry wielkości  $y^*$  i minimalnym koszcie. Dane są przy tym koszty jednostkowe  $c_{uv}$  odpowiadające poszczególnym lukom  $(u, v)$ . Oznaczmy przez  $x_{uv}$  przepływ wzdłuż luku  $(u, v)$ . Zadanie przepływu o minimalnym koszcie można wówczas zapisać następująco:

$$(0) \quad \sum_{(u,v) \in V(N)} c_{uv} x_{uv} \rightarrow \min,$$

$$(1) \quad \sum_{v \in N^+(u)} x_{uv} = \sum_{v \in N^-(u)} x_{vu}, \quad u \in V(N) \setminus \{s, t\},$$

$$(2) \quad \sum_{v \in N^+(s)} x_{sv} = y^*,$$

$$(3) \quad \sum_{v \in N^-(t)} x_{vt} = y^*,$$

$$(4) \quad 0 \leq x_{uv} \leq h_{uv}, \quad (u, v) \in E(N).$$

Aby powyższe zadanie miało przynajmniej jedno rozwiązanie dopuszczalne (a co za tym idzie, aby posiadało rozwiązanie optymalne), pożądaný przepływ  $y^*$  nie może przekraczać maksymalnego przepływu w sieci  $N$ .

Zaprezentowany model opisuje zadanie programowania liniowego. Zadanie to, podobnie jak w dwóch poprzednich przypadkach, ma specyficzną, sieciową strukturę

ture i jego rozwiązywanie ogólnymi metodami PL jest nieefektywne. W dalszej części tego podrozdziału jest przedstawiony jeden spośród specjalnych algorytmów rozwiązywania tego typu zagadnień, metoda Busackera-Gowena.

Idea metody jest zwiększanie przepływu wzdłuż kolejnych ścieżek powiększających o jak największą ilość (równą, oczywiście, ich przepustowości). Kolejność wyznaczania ścieżek nie jest przy tym uzależniona od porządku w zbiorze węzłów (jak to ma miejsce w przypadku algorytmu FFEK), ale od ich długości, przy czym rolę odległości pełnią jednostkowe koszty. Najkrótsze ścieżki są wyznaczane w sieci rezydualnej konstruowanej jak w przypadku algorytmu FFEK, przy czym koszty jednostkowe odpowiadające nowo powstałym lukom są wyznaczane zgodnie ze wzorem:  $c'_{vu} = -c_{uv}$ . Algorytm Busackera-Gowena (BG) można więc zapisać następująco:

#### Algorytm 3.3. Algorytm BG

Krok 1. Przyjmij aktualny przepływ równy  $y$  i aktualny skumulowany koszt  $C = 0$ . Niech  $N^* = N$ . Przejdź do kroku 2.

Krok 2. Jeżeli  $y = y^*$ , stop. Wyznaczone rozwiązanie jest optymalne. Optymalne przepływy są równe przepustowościom luków ze zbioru  $E(N^*) \setminus E(N)$ , a minimalny koszt jest równy  $C$ . Jeżeli  $y < y^*$ , przejdź do kroku 3.

Krok 3. Jeżeli w sieci  $N^*$  nie istnieje żadna ścieżka ze źródła do ujścia, stop. Zadanie nie ma rozwiązań dopuszczalnych. W przeciwnym razie znajdź w  $N^*$  najkrótszą ścieżkę  $P$ , biorąc koszty jednostkowe w miejsce odległości. Niech  $\delta = \min \{ \min \{ h_{uv} : (u, v) \in P \}; y^* - y \}$ . Zmodyfikuj sieć  $N^*$  według następujących reguł (modyfikacji podlegają wszystkie luki należące do  $P$  i luki do nich przeciwne i tylko one):

1) jeżeli dla pewnego luku  $(u, v) \in P$ ,  $(u, v) \in E(N)$  i  $(v, u) \notin E(N^*)$ , dodaj luk  $(v, u)$  do  $N^*$  i podstaw:  $h'_{uv} = h_{uv} - \delta$ ,  $h'_{vu} = \delta$ ,  $c'_{vu} = -c_{uv}$ ; jeżeli po wykonaniu wymienionych działań  $h'_{uv} = 0$ , usuń luk  $(u, v)$  z  $N^*$ ;

2) jeżeli dla pewnego luku  $(u, v) \in P$ ,  $(u, v) \in E(N)$  i  $(v, u) \in E(N^*)$ , podstaw:  $h'_{uv} = h_{uv} - \delta$ ,  $h'_{vu} = h_{vu} + \delta$ ; jeżeli po wykonaniu wymienionych działań  $h'_{uv} = 0$ , usuń luk  $(u, v)$  z  $N^*$ ;

3) jeżeli dla pewnego luku  $(u, v) \in P$ ,  $(v, u) \in E(N)$  i  $(v, u) \notin E(N^*)$ , dodaj luk  $(v, u)$  do  $N^*$  i podstaw:  $h'_{uv} = h_{uv} - \delta$ ,  $h'_{vu} = \delta$ ,  $c'_{vu} = c_{uv}$ ; jeżeli po wykonaniu wymienionych działań  $h'_{uv} = 0$ , usuń luk  $(u, v)$  z  $N^*$ ;

4) jeżeli dla pewnego luku  $(u, v) \in P$ ,  $(v, u) \in E(N)$  i  $(v, u) \in E(N^*)$ , podstaw:  $h'_{uv} = h_{uv} - \delta$ ,  $h'_{vu} = h_{vu} + \delta$ ; jeżeli po wykonaniu wymienionych działań  $h'_{uv} = 0$ , usuń luk  $(u, v)$  z  $N^*$ .

Podstaw  $y = y + \delta$ . Podstaw  $C = C + L\delta$ , gdzie  $L$  oznacza długość  $P$ . Wróć do kroku 2.

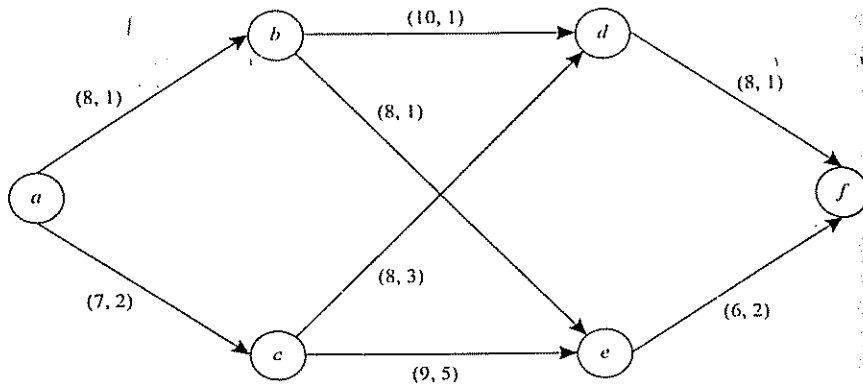
Działanie algorytmu BG ilustruje następujący przykład liczbowy.



## PRZYKŁAD 3.3

## Działanie algorytmu BG

Niech dana będzie sieć jak na rysunku 3.12. Wyznaczyć przepływ o minimalnym koszcie i wielkości  $y^* = 12$  (uwaga:  $s = a$ ,  $t = f$ , pierwsze liczby w nawiasach to przepustowości luków, drugie to jednostkowe koszty).

RYSUNEK 3.12  
Przykładowa sieć

Iteracja 1.

Sieć rezydualna ma postać jak na rysunku 3.12.

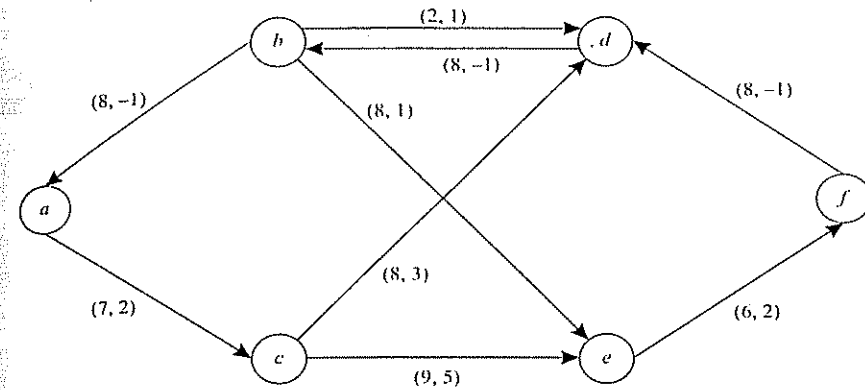
Krok 2.  $y = 0$ ,  $C = 0$ . Skoro  $y < y^*$ , przechodzimy do kroku 3.

Krok 3. Najkrótsza ścieżka:  $(a, b, d, f)$ ,  $L = 3$ .

$\delta = \min\{\min\{h_{ab}, h_{bd}, h_{df}\}, y^* - y\} = \min\{\min\{8, 10, 8\}, 12 - 0\} = 8$ . Skoro  $(a, b) \in N$  i  $(b, a) \notin N^*$ , dodajemy  $(b, a)$  do  $N^*$  i podstawiamy  $c_{ba}^* = -c_{ab}^* = -1$ ,  $h_{ba}^* = \delta = 8$ ,  $h_{ab}^* = h_{ab}^* - \delta = 0$ . Skoro  $h_{ab}^* = 0$ , usuwamy  $(a, b)$  z  $N^*$ . Skoro  $(b, d) \in N$  i  $(d, b) \notin N^*$ , dodajemy  $(d, b)$  do  $N^*$  i podstawiamy  $c_{db}^* = -c_{bd}^* = -1$ ,  $h_{db}^* = \delta = 8$ ,  $h_{bd}^* = h_{bd}^* - \delta = 2$ . Skoro  $(d, f) \in N$  i  $(f, d) \notin N^*$ , dodajemy  $(f, d)$  do  $N^*$  i podstawiamy  $c_{fd}^* = -c_{df}^* = -1$ ,  $h_{fd}^* = \delta = 8$ ,  $h_{df}^* = h_{df}^* - \delta = 0$ . Skoro  $h_{df}^* = 0$ , usuwamy  $(d, f)$  z  $N^*$ . Podstawiamy  $y = y + \delta = 0 + 8 = 8$  i  $C = C + \delta L = 0 + 8 \cdot 3 = 24$ . Wracamy do kroku 2.

Iteracja 2.

Sieć rezydualną, po pierwszej iteracji, przedstawia rysunek 3.13.

RYSUNEK 3.13  
Sieć rezydualna po pierwszej iteracji

Krok 2.  $y = 8$ ,  $C = 24$ . Skoro  $y < y^*$ , przechodzimy do kroku 3.

Krok 3. Najkrótsza ścieżka:  $(a, c, d, b, e, f)$ ,  $L = 7$ .

$\delta = \min\{\min\{h_{ac}, h_{cb}, h_{db}, h_{be}, h_{ef}\}, y^* - y\} = \min\{\min\{7, 8, 8, 8, 6\}, 12 - 8\} = 4$ . Skoro  $(a, c) \in N$  i  $(c, a) \notin N^*$ , dodajemy  $(c, a)$  do  $N^*$  i podstawiamy  $c_{ca}^* = -c_{ac}^* = -2$ ,  $h_{ca}^* = \delta = 4$ ,  $h_{ac}^* = h_{ac}^* - \delta = 3$ . Skoro  $(c, d) \in N$  i  $(d, c) \notin N^*$ , dodajemy  $(d, c)$  do  $N^*$  i podstawiamy  $c_{dc}^* = -c_{cd}^* = -3$ ,  $h_{dc}^* = \delta = 4$ ,  $h_{cd}^* = h_{cd}^* - \delta = 5$ . Skoro  $(b, d) \in N$  i  $(b, d) \in N^*$ , podstawiamy  $h_{bd}^* = h_{bd}^* + \delta = 6$ ,  $h_{db}^* = h_{db}^* - \delta = 4$ . Skoro  $(b, e) \in N$  i  $(e, b) \notin N^*$ , dodajemy  $(e, b)$  do  $N^*$  i podstawiamy  $c_{eb}^* = -c_{be}^* = -1$ ,  $h_{eb}^* = \delta = 4$ ,  $h_{be}^* = h_{be}^* - \delta = 4$ . Skoro  $(e, f) \in N$  i  $(f, e) \notin N^*$ , dodajemy  $(f, e)$  do  $N^*$  i podstawiamy  $c_{fe}^* = -c_{ef}^* = -1$ ,  $h_{fe}^* = \delta = 4$ ,  $h_{ef}^* = h_{ef}^* - \delta = 2$ . Podstawiamy  $y = y + \delta = 8 + 4 = 12$  i  $C = C + \delta L = 24 + 4 \cdot 7 = 52$ . Wracamy do kroku 2.

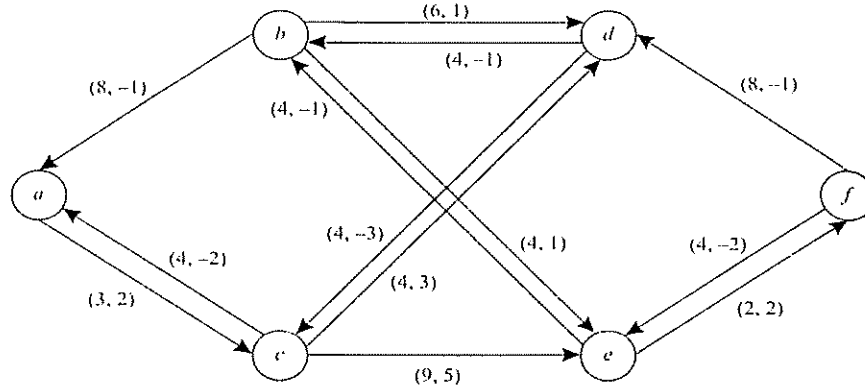
Iteracja 3.

Sieć rezydualną po drugiej iteracji przedstawia rysunek 3.14.

Krok 2.  $y = 12$ ,  $C = 52$ . Skoro  $y = y^*$ , stop. Optymalne przewozy równe są przepustowościom na dodanych lukach, a więc  $x_{ab} = 8$ ,  $x_{ac} = 4$ ,  $x_{bd} = 4$ ,  $x_{be} = 4$ ,  $x_{cd} = 4$ ,  $x_{df} = 8$ ,  $x_{ef} = 4$ .

Podstawowe zastosowanie zagadnienie przepływu o minimalnym koszcie znajduje w optymalizacji decyzji transportowych. Do mniej tradycyjnych zastosowań zaliczyć należy optymalizację czasu przesyłu danych w złożonych sieciach telekomunikacyjnych i odczytywanie obrazów w niektórych typach urządzeń do prześwietleń.

RYSUNEK 3.14  
Sieć rezydualna po drugiej iteracji



#### POLECENIA I PYTANIA KONTROLNE

1. Wymień i scharakteryzuj elementy składowe sieci.
2. Sformułuj zagadnienie maksymalnego przepływu.
3. Przedstaw algorytm FFEK.
4. Sformułuj zagadnienie najkrótszej drogi.
5. Przedstaw algorytm BMEP.
6. Sformułuj zagadnienie przepływu o minimalnym koszcie.
7. Przedstaw algorytm BG.

#### BIBLIOGRAFIA

- Ford L.R., Fulkerson D.R., 1969, *Przepływy w sieciach*, PWN, Warszawa.
- Muller Y., 1971, *Wprowadzenie do nauki organizacji i badań operacyjnych*, PWE, Warszawa.
- Sysło M.M., Deo N., Kowalik J.S., 1993, *Algorytmy optymalizacji dyskretnej*, Wydawnictwo Naukowe PWN, Warszawa.

## ROZDZIAŁ 4

# OPTIMALIZACJA WIELOKRYTERIALNA



### 4.1. Matematyczny model sytuacji decyzyjnej w przypadku istnienia wielu kryteriów oceny

Kwantyfikacja preferencji decydenta w postaci jednej funkcji celu (kryterium optymalizacji) prowadzi do konstrukcji modelu przybierającego postać problemu programowania matematycznego. W większości przypadków decydent, podejmując decyzje, pragnie w maksymalnym stopniu zrealizować jednocześnie wiele celów<sup>1</sup>. Przykład stanowi sytuacja, w której należy dokonać wyboru najlepszego wyrobu z zbioru wyrobów, z których każdy jest scharakteryzowany przez wiele cech. Często też decyzje są podejmowane przez ciała kolegialne. W tym przypadku potrzeba uwzględnienia przy podejmowaniu decyzji wielu kryteriów oceny jest oczywista.

Jeżeli przy podejmowaniu decyzji kierujemy się jednym kryterium, to konstruujemy model sytuacji decyzyjnej i podejmujemy próbę wyznaczenia rozwiązania optymalnego (lub przybliżonego) otrzymanego problemu programowania matematycznego, które może stanowić podstawę do podjęcia decyzji: tym lepsza, im model wierniej odzwierciedla realia procesu decyzyjnego. W przypadku istnienia wielu kryteriów oceny nie istnieje jednoznaczny, powszechnie akceptowany sposób wyznaczenia rozwiązania „optymalnego”. Można wskazać wiele metod wyznaczania takiego rozwiązania. Ich charakterystyczną cechą jest konieczność uzyskania od decydenta informacji na temat ważności kryteriów, która nosi charakter subiektywny, co ma istotne konsekwencje natury metodycznej i praktycznej<sup>2</sup>. Oznacza to bowiem, że żaden z zaproponowanych sposobów wyboru nie jest w jakimkolwiek sensie „lepszy” od pozostałych. W sensie czysto praktycznym

<sup>1</sup> Kierowanie się przy podejmowaniu decyzji jednym przez nikogo niekwestionowanym kryterium należy do rzadkości i jest charakterystyczne dla działalności o rutynowym charakterze.

<sup>2</sup> Można domniemywać, że jedną z przyczyn pojawienia się wielu metod wyznaczania „najlepszego” rozwiązania była próba zobiektywizowania procesu jego uzyskania.

proceedzi natomiast do stwierdzenia, że dwaj decydenci dysponujący tym samym zasobem informacji obiektywnej (bardziej precyzyjnie i realistycznie: potencjalnie ogólnie dostępnej) mogą podjąć różne decyzje i żadnemu z nich nie będzie można postawić zarzutu, że podjęta przez niego decyzja jest błędna. Nie istnieją zabiegi matematyczne mogące zmienić ten stan rzeczy, gdyż subiektywizm stanowi immanentną cechę decyzji wielokryterialnych.

W rozdziale dokonamy przeglądu wybranych, najczęściej spotykanych metod wyznaczania „najlepszych” decyzji w przypadku istnienia wielu kryteriów oceny. Przegląd ten nie jest w żadnym stopniu kompletny. W zamierzeniu ma on odzwierciedlać podstawowe idee, a nie koncentrować się na opisie metod, których ze względu na niemożność obiektywizacji procedury wyboru najlepszej decyzji pojawiło się bardzo wiele.

Będziemy zakładać, że każdemu celowi odpowiada kryterium pozwalające na pomiar stopnia jego realizacji. Kryteria te przyjęło się określać mianem **kryteriów cząstkowych**. Przyjmujemy założenie, że decydent, podejmując decyzję dopuszczalną  $x = [x_1, \dots, x_n]^T \in D$ , chce w maksymalnym stopniu zrealizować cele  $f_1, \dots, f_k$ . Zakładamy, że cele te są **mieralne**, tj. że mogą być wyrażone przez funkcje liczbowe. Przy tych założeniach ocena decyzji dana jest przez wektor ocen cząstkowych  $y^T = [f_1(x), \dots, f_k(x)]^T \in S$ , gdzie  $S$  to zbiór ocen. W zbiorze  $D$  należy wyznaczyć „najlepszą” decyzję ze względu na rozpatrywany zbiór kryteriów. Zbiór  $D$  określa się mianem **przestrzeni decyzyjnej**, a zbiór  $S$  — **przestrzeni kryterialnej**.

Zbiór decyzji może być zbiorem dyskretnym lub ciągłym. W drugim przypadku **sensownie jest założyć jego zwartość** (tj. domkniętość i ograniczoność). Jeżeli zbiór decyzji jest zbiorem dyskretnym, to jesteśmy zainteresowani problemem wyboru „najlepszej” decyzji, posortowaniem elementów zbioru bądź dokonaniem ich uporządkowania<sup>3</sup>. Nasze zainteresowania skupimy przede wszystkim na pierwszym z problemów, tj. na problemie wyboru.

W przypadku gdy zbiór decyzji nie jest zbiorem dyskretnym, proponuje się wiele strategii postępowania. W rozdziale ograniczymy się do krótkiego opisu tzw. wielokryterialnego programowania liniowego (metody ADBASE) oraz metod, których istota sprowadza się do transformacji problemu z wieloma kryteriami do problemu z jednym kryterium, tj. do problemu programowania liniowego.

<sup>3</sup> W przypadku istnienia wielu kryteriów oceny wybór „najlepszej” decyzji nie jest procesem obiektywnym. Dlatego sensowniej jest niekiedy ograniczyć się do posortowania elementów zbioru, tj. do przydziału decyzji do pewnych z góry określonych kategorii (ze wskazaniem — o ile jest to możliwe — odległości między nimi) lub jego uporządkowania, tj. do podziału decyzji na klasy decyzji równie dobrych. Pomoc w artykułowaniu preferencji, ich modelowaniu oraz realizacji analitycznej fazy wspomaganie decyzji jest udzielana przez analityka. Wybór należy do decydenta. Z tego względu procedury oferowane przez programowanie wielokryterialne powinny pojęciowo być tak proste, aby osoba niemająca przygotowania formalnego mogła je świadomie stosować — samodzielnie lub w współpracy z analitykiem.

Omawiając wybrane metody programowania wielokryterialnego rozpatrzmy cztery przypadki:

- 1) w zbiorze kryteriów decydent nie określił preferencji między celami;
- 2) preferencje zostały podane w postaci informacji o koniecznym poziomie realizacji każdego z celów;
- 3) preferencje zostały określone w postaci funkcji użyteczności;
- 4) preferencje zostały podane w postaci hierarchii celów.

Analizując każdy z przypadków, będziemy (jeżeli zachodzi taka potrzeba) dokonywać rozróżnienia między sytuacją, w której zbiór decyzji jest zbiorem dyskretnym oraz kiedy nim nie jest.

## 4.2. Optymalność w programowaniu wielokryterialnym

Załóżmy, że każde z kryteriów cząstkowych jest maksymalizowane. Powiemy, że decyzja  $x^1$  jest lepsza od  $x^2$  ze względu na kryterium  $f_i$ , jeżeli  $f_i(x^1) > f_i(x^2)$ . Każda z decyzji opisana jest przez wektor ocen, a więc ich porównywanie można sprowadzić do porównywania wektorów ocen w przestrzeni kryterialnej. W przestrzeni tej określa się **relację preporządku**, wymagając, aby była ona zwrotna i przechodnia. Definiując tę relację, korzystamy ze składowych wektorów ocen  $y = [f_1(x), \dots, f_k(x)]^T$ . Jeżeli  $y^1 \geq y^2$ , to powiemy, że  $y^1$  jest nie gorszy od  $y^2$ . Jeżeli  $y^1 = y^2$ , to wektory (a mówiąc ściśle decyzje korespondujące z nimi) są równie dobre<sup>5</sup>.

Innymi słowy, ze względu na zbiór kryteriów  $f_1, \dots, f_k$  decyzja  $x^1$  jest nie gorsza od  $x^2$ , jeżeli  $f_i(x^1) \geq f_i(x^2)$ ,  $i = 1, \dots, K$ , a decyzja  $x^1$  jest równoważna decyzji  $x^2$ , jeżeli  $f_i(x^1) = f_i(x^2)$ ,  $i = 1, \dots, K$ .

Powiemy, że wektor ocen  $y^1$  dominuje wektor  $y^2$ , jeżeli  $y^1$  jest nie gorszy od  $y^2$  oraz istnieje co najmniej jedna składowa wektora  $y^1$  większa od odpowiadającej jej składowej  $y^2$ , tj. jeżeli:

$$\begin{aligned} f_i(x^1) &\geq f_i(x^2) \quad \text{dla wszystkich } i = 1, \dots, K, \\ f_i(x^1) &> f_i(x^2) \quad \text{dla co najmniej jednego } i. \end{aligned} \quad (4.1)$$

Jeżeli w zbiorze kryteriów nie jest możliwe określenie dla decyzji  $x^1, x^2$  żadnej z tych relacji, powiemy, że są one **nieporównywalne**.

<sup>4</sup> Decyzja może polegać na określeniu wartości współrzędnych wektora (np. w przypadku problemu wyboru optymalnego asortymentu produkcji) lub też na wyborze pewnego elementu ze skończonego (na ogół) zbioru. W pierwszym przypadku do oznaczenia decyzji będziemy używać pogrubionej czcionki.

<sup>5</sup> Relacje „nie gorszy od” oraz „równoważny” są zwrotne i przechodnie.

W realnych sytuacjach decyzyjnych rzadko zdarza się, aby istniał taki wektor ocen, który dominuje wszystkie pozostałe wektory, tj. taka decyzja  $y^* \in D$ , dla której wszystkie kryteria cząstkowe osiągają jednocześnie swoje maksima (w ogólnym przypadku optima). Decyzja lepsza ze względu na jedno kryterium bywa zazwyczaj gorsza ze względu na inne kryteria. Proponuje się redefinicję terminu „optymalny”, zastępując kryterium optymalności znane z analizy jednokryterialnej, tzw. **optymalnością w sensie Pareto**.

**Niezdominowanym wektorem ocen** nazywamy taki wektor ocen, którego żaden inny wektor ocen nie dominuje. Odpowiednikiem problemu maksymalizacji (optymalizacji) jednokryterialnej jest w przypadku istnienia wielu kryteriów **problem maksymalizacji (optymalizacji) wektorowej**. Przyjmuje się, że zbiorem **rozwiązań wektorowo optymalnych** problemu maksymalizacji wektorowej jest zbiór rozwiązań dopuszczalnych, których wektory ocen są niezdominowane. Są to więc takie rozwiązania, że nie istnieją inne rozwiązania umożliwiające poprawę wartości jednego z kryteriów bez pogorszenia wartości co najmniej jednego z pozostałych kryteriów.

**Rozwiązania niezdominowane** w przestrzeni kryterialnej nazywać będziemy **rozwiązaniami Pareto-optymalnymi**, a odpowiadające im rozwiązania w przestrzeni decyzyjnej — **rozwiązaniami sprawnymi**. Rozwiązania sprawne są z definicji nieporównywalne. Zaletą rozwiązania optymalnego w sensie Pareto stanowi to, że jego wyznaczenie nie wymaga uzyskania od decydenta informacji o charakterze subiektywnym, określającej jego preferencje. Wystarczy, że decydent poda zbiór kryteriów cząstkowych<sup>6</sup>. W praktyce zbiór decyzji Pareto-optymalnych jest rzadko jednoelementowy, co wymaga jego zawężenia w celu wyboru najlepszej decyzji. Proponowane sposoby postępowania opiszemy w podrozdziale 4.3<sup>7</sup>.

<sup>6</sup> Drugą zaletą jest to, że wyznaczając decyzje Pareto-optymalne ograniczamy się do operacji porównywania, co oznacza, że pomiar nasilenia każdej cechy może być dokonywany na skali porządkowej. Jest to taka skala, w przypadku której, jeżeli ze względu na rozpatrywaną cechę decyzja A jest lepsza od B, to wartość przypisana nasileniu cechy pierwszej z nich jest większa od wartości przypisanej drugiej. Nie jest istotne, o ile większa. Przyporządkowanie wartości liczbowych cesze decyzji jest dość arbitralne — wystarczy, aby został spełniony ww. warunek. Jedyną dopuszczalną operacją jest porównywanie. Nie można powiedzieć, że decyzja A jest dwa razy lepsza od B, jeżeli wartość przypisana cesze decyzji A jest dwa razy większa od wartości przypisanej cesze decyzji B. Aby można było nie tylko porównywać nasilenie cechy, lecz także stwierdzić, ile razy nasilenie cechy jest większe, trzeba posłużyć się skalą interwałową. Wartość liczbową określającą nasilenie cechy jest proporcjonalna do jej nasilenia. Przykładem pomiaru na skali interwałowej jest pomiar zysku wyrażonego w jednostkach pieniężnych.

<sup>7</sup> Często oczekuje się, że programowanie wielokryterialne pozwoli na podanie procedur umożliwiających proces podejmowania decyzji przy wielorakości celów. Oczekiwania te są bezzasadne ze względu na konieczność posługiwania się informacją noszącą niekiedy wysoce subiektywny charakter w celu wyboru najlepszej decyzji ze zbioru decyzji nieporównywalnych. Subiektywizm ten nie jest słabością programowania wielokryterialnego — próba kwantyfikacji uoacznia problemy sprowadzając decyzji do porównywalności będące konsekwencją subiektywizmu.

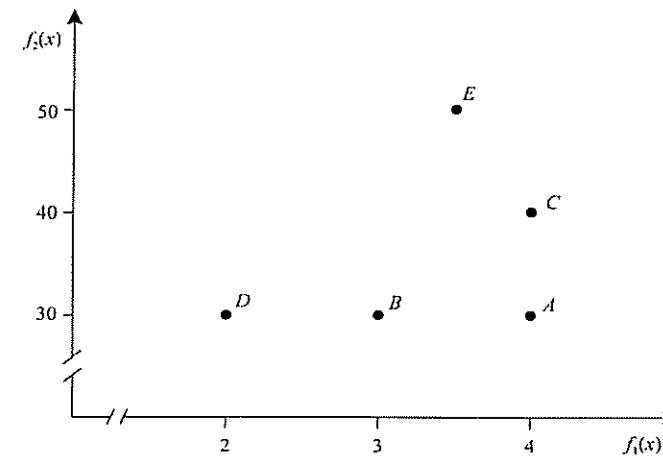
PRZYKŁAD 4.1<sup>8</sup>

W tabelicy 4.1 podano zyski i udział w rynku, jaki firma spodziewa się osiągnąć, stosując wybrane strategie marketingowe. Wyznaczmy strategie Pareto-optymalne.

TABLICA 4.1  
Ocena strategii marketingowych ze względu na zysk i udział w rynku

| Kryteria                     | Strategie marketingowe |    |    |    |     |
|------------------------------|------------------------|----|----|----|-----|
|                              | A                      | B  | C  | D  | E   |
| Zysk (mln zł) ( $f_1$ )      | 4                      | 3  | 4  | 2  | 3,5 |
| Udział w rynku (%) ( $f_2$ ) | 30                     | 30 | 40 | 30 | 50  |

RYSUNEK 4.1  
Wyznaczanie strategii Pareto-optymalnych w przypadku istnienia dwóch kryteriów oceny



U w a g a: Współrzędne każdego z punktów są dane przez wartości kryteriów cząstkowych charakteryzujących poszczególne strategie (przy sporządzaniu rysunku wykorzystano przestrzeń kryterialną). Oba kryteria są maksymalizowane. Strategiami Pareto-optymalnymi są C i E.

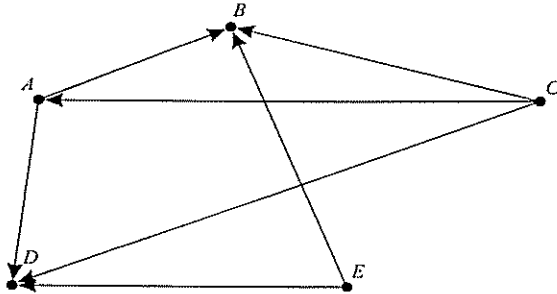
Dla strategii C i E nie można wskazać strategii lepszych ze względu na jedno kryterium, a nie gorszych ze względu na drugie. Są one Pareto-optymalne.

<sup>8</sup> Sposób wyznaczania decyzji sprawnych (przypadek ciągły) ilustruje przykład 4.5.

Pozostałe strategie to strategie zdominowane. Dla każdej ze strategii zdominowanych można wskazać lepszą od niej strategię ze zbioru strategii Pareto- optymalnych. Na przykład strategie  $C$  i  $E$  są lepsze od strategii  $B$ . Nie każda ze strategii Pareto- optymalnych musi dominować strategię zdominowaną. Strategia  $A$  jest zdominowana przez  $C$ , ale nie przez  $E$ .

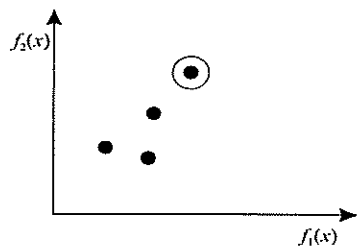
Alternatywnym sposobem prezentacji relacji preferencji jest **diagram Hassego**. Jest to graf, którego wierzchołki reprezentują decyzje. Jeżeli decyzja  $X$  dominuje decyzję  $Y$ , od wierzchołka reprezentującego pierwszą z nich kreślimy łuk prowadzący do wierzchołka reprezentującego drugą decyzję. Jeżeli decyzje są nieporównywalne, to odpowiadające im wierzchołki nie są połączone łukiem. Z decyzji Pareto- optymalnymi korespondują wierzchołki, do których nie dochodzą żadne łuki. Opisana wyżej reprezentacja relacji preferencji jest niezbędna, jeżeli wektory ocen decyzji zawierają więcej niż dwie składowe. Rysunek 4.2 przedstawia diagram Hassego dla zbioru strategii marketingowych z przykładu 4.1.

RYSUNEK 4.2  
Diagram Hassego  
dla strategii marketingowych z przykładu 4.1

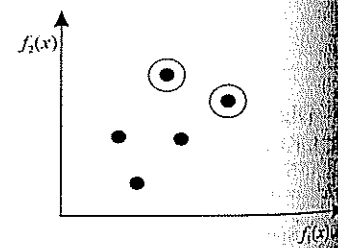


RYSUNEK 4.3  
Własności zbioru strategii  
Pareto- optymalnych

A. Jedna decyzja jest Pareto- optymalna

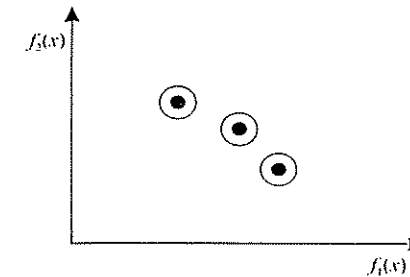


B. Dwie decyzje są Pareto- optymalne



CD. RYSUNKU 4.3

C. Wszystkie decyzje są Pareto- optymalne



Na rysunku 4.3 przedstawiono przypadek, kiedy istnieje jedna decyzja Pareto- optymalna, decyzje Pareto- optymalne są podzbiorem zbioru decyzji dopuszczalnych oraz zbiory decyzji dopuszczalnych i Pareto- optymalnych się pokrywają. Założono, że kryteria cząstkowe są maksymalizowane. Z rysunku wynika, że optymalność w sensie Pareto nie pozwala na ogół na jednoznaczny wybór „najlepszej” decyzji. Konieczne jest zatem zawężenie tego zbioru do zbioru jednoelementowego.

### 4.3. Postępowanie w przypadku nieporównywalności decyzji

Wobec nieporównywalności decyzji Pareto- optymalnych nie można powiedzieć, że jedna z nich jest lepsza od drugiej. W takiej sytuacji proponuje się wiele strategii postępowania mających na celu wybór „najlepszej” decyzji. Wprowadzając dodatkowe warunki, zawężamy zbiór decyzji Pareto- optymalnych do jednego punktu utożsamianego z tzw. **decyzją kompromisową**. Problem optymalizacji wielokryterialnej staramy się sprowadzić do problemu optymalizacji jednokryterialnej, konstruując funkcję nazywaną metakryterium (kryterium zastępczym), w której ujmujemy wszystkie kryteria cząstkowe. Możemy także wyznaczyć decyzję najbardziej zbliżoną do ideału, tj. najmniej oddaloną od tzw. punktu idealnego.

#### 4.3.1. Maksymalizacja celu głównego

Informacja o wymaganiach dotyczących realizacji celów może przybierać dwie formy:

- 1) ograniczeń z góry lub z dołu na poziomy realizacji celów,
- 2) wektora-wzorca, tj. wektora postulowanych pożądanych poziomów realizacji celów.

Rozważmy pierwszy przypadek. Jedno z kryteriów cząstkowych traktujemy jako cel główny, postulując jego maksymalizację<sup>9</sup>. Chcąc uwzględnić pozostałe kryteria cząstkowe, żądamy, aby każde z nich zostało zrealizowane w minimalnym satysfakcjonującym stopniu. Przyjmując, że  $f_1$  jest celem głównym, a  $f_2, \dots, f_K$  celami pobocznymi, otrzymujemy następujący problem optymalizacyjny:

$$f_1(x) \rightarrow \max,$$

$$x \in D,$$

$$f_i(x) \geq b_i \quad i = 2, \dots, k_1,$$

$$f_i(x) \leq b_i \quad i = k_1 + 1, \dots, K.$$

gdzie  $b_2, \dots, b_K$  są wymaganymi poziomami realizacji celów cząstkowych, cele  $f_2, \dots, f_{k_1}$  są maksymalizowane, natomiast cele  $f_{k_1+1}, \dots, f_K$  — minimalizowane<sup>10</sup>.

Nadanie  $b_i$  zbyt dużych (w przypadku maksymalizacji) lub zbyt małych (w przypadku minimalizacji) wartości — co oznacza próbę uwzględnienia w możliwie dużym stopniu kryteriów cząstkowych — prowadzi do nadmiernego ograniczenia zbioru  $D$ , czego konsekwencją jest zmiana rozwiązania optymalnego i sprzeczność problemu (4.2)–(4.5). Zawężenie zbioru  $D$  nie zawsze musi spowodować zmianę rozwiązania optymalnego problemu (4.2)–(4.5). Prowadzi ono do zmiany tego rozwiązania, jeżeli w wyniku zawężenia zbioru  $D$  otrzymane rozwiązanie optymalne przestało być dopuszczalne.

W drugim przypadku postulat wyznaczenia decyzji, dla której wartości kryteriów cząstkowych w możliwie niewielkim stopniu odbiegają od wartości pożądaných, prowadzi do sformułowania tzw. problemu programowania celowego, sprowadzającego się do zastąpienia wielu kryteriów cząstkowych jednym kryterium nazywanym metakryterium (kryterium zastępczym). Dlatego problem ten został omówiony w następnym punkcie.

Niekiedy proponuje się, aby funkcja celu minimalizowała maksymalne odchylenie wartości kryterium cząstkowego od jego wartości pożądanej w zbiorze rozwiązań kryteriów cząstkowych. Rozważmy wersję tego problemu, którą jest łatwo sprowadzić do problemu programowania liniowego:

$$\max_{i=1, \dots, K} |c^i x - c_{i0}| \rightarrow \min,$$

$$Ax = b,$$

$$x \geq 0.$$

<sup>9</sup> Kierunek optymalizacji jest nieistotny. Jeżeli chcemy, aby funkcja, która ma być minimalizowana, była maksymalizowana, to współczynniki stojące przy zmiennych wystarczy pomnożyć przez  $-1$ .

<sup>10</sup> Cele te określa się mianem poziomów aspiracji, a problem (4.2)–(4.5) — problemem wyznaczającym satysfakcjonujących poziomów kryteriów.

gdzie  $c_{i0}$  jest postulowaną wartością funkcji celu  $c^i x$  dla  $i$ -tego kryterium cząstkowego, a  $c^i$  — wektorem wag stojących przy zmiennych odpowiadających  $i$ -temu kryterium<sup>11</sup>.

W celu sprowadzenia problemu (4.6)–(4.8) do problemu liniowego wprowadzamy zmienną  $y$  spełniającą warunki:

$$c^i x - c_{i0} \leq y, \quad i = 1, \dots, K, \quad (4.9)$$

$$-c^i x + c_{i0} \leq y, \quad i = 1, \dots, K, \quad (4.10)$$

które są równoważne warunkowi  $\max_{i=1, \dots, K} |c^i x - c_{i0}| \leq y$ . Następnie wyznaczamy rozwiązanie optymalne  $(x^*, y^*)$  problemu:

$$y \rightarrow \min, \quad (4.11)$$

przy warunkach (4.7)–(4.10), który jest problemem PL. Wektor  $x^*$  jest rozwiązaniem optymalnym problemu wyjściowego, tj. problemu (4.6)–(4.8).

### 4.3.2. Metakryterium

Zakłada się, że decydent — mimo że kieruje się w swym postępowaniu wieloma kryteriami — jest zdolny do przypisania układowi wartości osiąganych przez te kryteria dla każdej decyzji dopuszczalnej jednej wartości liczbowej nazywanej użytecznością.

Innymi słowy, zakłada się, że decydent jest zdolny do wskazania funkcji<sup>12</sup>:

$$u(x) = u[f_1(x), f_2(x), \dots, f_K(x)]^T, \quad (4.12)$$

<sup>11</sup> Z numerycznego punktu widzenia nie jest istotne, czy zbiór  $D$  jest określony przez nierówności  $Ax \leq b$ , czy też przez równość  $Ax = b$ . Jeżeli problem jest w postaci standardowej, to — wprowadzając zmienne swobodne i przepisując im zerowe wagi — uzyskujemy problem w postaci kanonicznej. Zastępując warunek (4.7) warunkiem (4.3) otrzymujemy ogólniejszą wersję problemu, w przypadku którego wyznaczenie rozwiązania optymalnego może być wysoce kłopotliwe, zwłaszcza gdy funkcja  $f_1$  nie jest wypukła (zob. rozdz. 5).

<sup>12</sup> Zakładamy *implicitnie*, że funkcja ta istnieje. Niech w zbiorze  $D$  preferencje decydenta będą skwantyfikowane za pomocą relacji  $S$ . Relację tę nazywamy racjonalną, jeżeli ma ona własności zupełności (spójności i zwrotności) i przechodności. Funkcję  $u: D \rightarrow R$  nazywamy funkcją użyteczności reprezentującą relację preferencji  $S$ , jeżeli dla każdych  $x, y \in D$  zachodzi  $xSy \rightarrow u(x) \geq u(y)$ . Warunkiem koniecznym istnienia funkcji użyteczności reprezentującej relację preferencji decydenta  $S$  jest spełnienie przez nią warunków racjonalności (zob. [Trzaskalik, 2006, s. 64–66]). Jeżeli np. decyzja jest podejmowana w warunkach ryzyka i wektory ocen niezbyt się od siebie różniące nie są rozróżnialne, zdefiniowanie funkcji użyteczności nie jest możliwe. Istotnie, decyzje  $x, y, z \in D$  mogą mieć tę własność, że dla decydenta decyzje  $x$  i  $y$  oraz  $y$  i  $z$  są nierozróżnialne, ale decyzja  $x$  jest preferowana nad  $z$ . Relacja preferencji nie jest racjonalna w sensie wyżej podanej definicji, co implikuje nieistnienie funkcji użyteczności.

której maksymalizacja jest równoważna wyznaczeniu najlepszego rozwiązania wzdględu na wszystkie kryteria. Funkcję tę niekiedy nazywa się **metakryterium**.

#### 4.3.2.1. Ważona suma kryteriów cząstkowych

Funkcja metakryterium przybiera następującą postać:

$$u(\mathbf{x}) = \sum_{i=1}^K w_i f_i(\mathbf{x}), \quad (4.13)$$

gdzie  $w_i, w_i \geq 0, i = 1, \dots, K$  nazywane są **wagami**; określają one względne znaczenie przypisywane  $i$ -temu kryterium cząstkowemu, a ich wartości są w sposób subiektywny określane przez decydenta.

Ponieważ istotne są tylko ich stosunki, a nie wartości absolutne, najczęściej poddaje się je zabiegowi **normalizacji**, przyjmując, że<sup>13</sup>:

$$\sum_{i=1}^K w_i = 1. \quad (4.14)$$

Warunkiem sensowności tego podejścia jest wyrażenie wartości kryteriów cząstkowych w tych samych jednostkach pomiaru i na tej samej skali liczbowej; co niestety znacznie ogranicza zakres jego stosowalności.

#### 4.3.2.2. Ważona suma stopni realizacji kryteriów cząstkowych

Próbą rozwiązania problemu nieporównywalności skal liczbowych i jednostek jest maksymalizacja **ważonej sumy stopni realizacji celów cząstkowych**.

Załóżmy, że wszystkie kryteria cząstkowe są maksymalizowane, a zbiór  $D$  jest niepusty. Oznaczmy przez  $f_i^*$  maksymalną wartość  $i$ -tego kryterium osiąganą w zbiorze  $D$ :

$$f_i^* = \max_{\mathbf{x} \in D} \{f_i(\mathbf{x})\}, \quad i = 1, \dots, K. \quad (4.15)$$

Zakładamy, że  $f_i^* > 0, f_i(\mathbf{x}) \geq 0, i = 1, \dots, K, \mathbf{x} \in D$ .

Zdefiniujmy wielkość  $\hat{r}_i(\mathbf{x}) = \frac{f_i(\mathbf{x})}{f_i^*}$ , którą będziemy nazywać **stopniem realizacji**

**cji I rodzaju**,  $i$ -tego celu cząstkowego przez decyzję  $\mathbf{x}$ . Jest ona wielkością bezwymiarową, unormowaną na przedziale  $[0, 1]$ . Założenie, że  $f_i^* > 0$  zapewni wykonalność dzielenia przez  $f_i^*$ .

<sup>13</sup> Dalej będziemy zakładać, że wagi zostały poddane zabiegowi normalizacji.

Metakryterium przyjmuje postać:

$$m(\mathbf{x}) = \sum_{i=1}^K w_i \cdot \hat{r}_i(\mathbf{x}), \quad \mathbf{x} \in D. \quad (4.16)$$

Ponieważ  $\hat{r}_i(\mathbf{x}), i = 1, \dots, K$  nie mają miana, metakryterium można stosować w przypadku kryteriów wyrażonych w różnych jednostkach lub na różnych skalach liczbowych.

Metakryterium stanowi **syntetyczny miernik satysfakcji**, osiągając wartość równą 1 w punkcie  $\bar{\mathbf{x}}$  (jeżeli taki punkt istnieje), w którym wszystkie kryteria cząstkowe osiągają swoje maksima cząstkowe, tzn. ich stopnie realizacji są równe 1. Istotnie, w punkcie  $\bar{\mathbf{x}}, m(\bar{\mathbf{x}}) = \sum_{i=1}^K w_i = 1$  (z uwagi na unormowanie wag).

Jeżeli istnieje takie  $\bar{\mathbf{x}} \in D$ , że dla tych  $i$ , dla których  $w_i > 0, f_i(\bar{\mathbf{x}}) = 0$ , to  $m(\bar{\mathbf{x}}) = 0$ .

Założenie, że  $f_i^* > 0$  jest niekiedy krępujące. Ponadto nie jest możliwe uwzględnienie w metakryterium kryteriów, które są minimalizowane. Rozwiązaniem tego problemu jest bardziej elastyczna definicja realizacji celu cząstkowego.

Niech  $f_i^{**}$  oznacza minimalną wartość  $i$ -tego kryterium osiąganą w zbiorze  $D$ :

$$f_i^{**} = \min_{\mathbf{x} \in D} \{f_i(\mathbf{x})\}, \quad i = 1, \dots, K, \quad \mathbf{x} \in D. \quad (4.17)$$

**Stopień realizacji II rodzaju**  $i$ -tego celu cząstkowego  $\tilde{r}_i(\mathbf{x})$  przez decyzję  $\mathbf{x} \in D$  w przypadku maksymalizacji wyznacza się według wzoru:

$$\tilde{r}_i(\mathbf{x}) = \frac{f_i(\mathbf{x}) - f_i^{**}}{f_i^* - f_i^{**}}, \quad i = 1, \dots, K, \quad \mathbf{x} \in D. \quad (4.18)$$

Ze wzoru (4.18) wynika, że  $0 \leq \tilde{r}_i(\mathbf{x}) \leq 1, i = 1, \dots, K, \mathbf{x} \in D; \tilde{r}_i(\bar{\mathbf{x}}) = 0$  dla takiego  $\bar{\mathbf{x}}$ , dla którego  $f_i(\bar{\mathbf{x}}) = f_i^{**}$  oraz  $\tilde{r}_i(\bar{\mathbf{x}}) = 1$ , dla takiego  $\bar{\mathbf{x}}$ , dla którego  $f_i(\bar{\mathbf{x}}) = f_i^*$ , co oznacza, że stopień realizacji celu cząstkowego II rodzaju jest wielkością unormowaną na przedziale  $[0, 1]$ . Zauważmy, że nie wymaga się już spełnienia krępujących warunków  $f_i(\mathbf{x}) \geq 0, f_i^* > 0, i = 1, \dots, K, \mathbf{x} \in D$ .

W przypadku minimalizacji  $i$ -tego kryterium stopień realizacji II rodzaju  $\tilde{r}_i(\mathbf{x})$  proponuje się wyznaczać według wzoru<sup>14</sup>:

$$\tilde{r}_i(\mathbf{x}) = \frac{f_i^* - f_i(\mathbf{x})}{f_i^* - f_i^{**}}. \quad (4.19)$$

<sup>14</sup> Możliwa jest taka modyfikacja stopnia realizacji I rodzaju, aby był możliwy pomiar stopnia realizacji w przypadku minimalizacji. Zastępując  $\frac{f_i(\mathbf{x})}{f_i^*}$  przez  $\frac{f_i^{**}}{f_i(\mathbf{x})}, f_i^{**} > 0, i = 1, \dots, K$ , otrzymujemy unormowany w przedziale  $(0, 1]$  stopień realizacji, który ma jednak tę niepożądaną cechę, że niewystarczająco różnicuje warianty decyzyjne dla „dużych”  $f_i^{**}$ , a nadmiernie dla  $f_i^{**}$  bliskich 0.

Ze wzoru (4.19) wynika, że  $0 \leq \tilde{r}_i(x) \leq 1$ ,  $i = 1, \dots, K$ ,  $x \in D$ ;  $\tilde{r}_i(\bar{x}) = 0$  dla takiego  $\bar{x}$ , dla którego  $f_i(\bar{x}) = f_i^*$  oraz  $\tilde{r}_i(\bar{x}) = 1$  dla takiego  $\bar{x}$ , dla którego  $f_i(\bar{x}) = f_i^{\max}$ .

### PRZYKŁAD 4.2

Zagadnienie wyznaczania rankingu banków.

Każdy z banków oceniany jest ze względu na następujące kryteria:

- 1) wysokość oprocentowania depozytów rocznych o zmiennej stopie procentowej (%);
- 2) liczbę placówek;
- 3) poziom bezpieczeństwa szacowany przez ekspertów w skali od 1 do 10 (1 oznacza najniższy, a 10 — najwyższy poziom bezpieczeństwa).

Pierwsze kryterium ocenia zyskowność inwestycji, drugie — dostępność usług, a trzecie — poziom bezpieczeństwa inwestycji.

TABLICA 4.2  
Oceny banków ze względu na kryteria cząstkowe

| Kryteria | Banki |     |      |     |     |
|----------|-------|-----|------|-----|-----|
|          | A     | B   | C    | D   | E   |
| $f_1(x)$ | 4,5   | 5   | 4,75 | 5   | 5   |
| $f_2(x)$ | 1 000 | 200 | 600  | 400 | 100 |
| $f_3(x)$ | 8     | 6   | 10   | 7   | 5   |

W tabelicy 4.2 podano oceny banków ze względu na rozpatrywane kryteria. Przeprowadziwszy badania marketingowe w grupie potencjalnych klientów poszczególnym kryteriom przypisano następujące wagi: 0,6, 0,2, 0,2<sup>15</sup>.

Sporządzmy ranking banków, uwzględniając tylko te, dla których stopnie realizacji celów cząstkowych są nie mniejsze od 0,7, 0,2, 0,5<sup>16</sup>.

Uwzględnienie w rankingu tylko tych banków, w przypadku których stopnie realizacji celów cząstkowych są nie mniejsze od wymaganych, ma następującą interpretację ekonomiczną. Jeżeli stopnie realizacji celów cząstkowych są zbyt

<sup>15</sup> Istnieją zasadniczo dwie metody wyznaczania wag. Jedną z nich, metoda delficka, polega na ich wyznaczeniu przez ekspertów. Druga, często stosowana w badaniach marketingowych, sprowadza się do wyznaczenia wag na podstawie opinii klientów pytanych o ważność poszczególnych kryteriów. Zazwyczaj opinie klientów zbiera się, przeprowadzając badania ankietowe.

<sup>16</sup> Zauważmy, że nawet to sformułowanie nie jest zbyt precyzyjne, jeżeli sposób wyznaczania wartości stopnia realizacji celu nie został zdefiniowany.

niskie, świadczy to o tym, że dany bank oferuje depozyty o zbyt małej dochodowości, dostępność usług jest nieakceptowana, a stopień bezpieczeństwa niewielki<sup>17</sup>.

Cele cząstkowe są wyrażone w różnych jednostkach, dlatego — sporządzając ranking banków — korzystamy z metakryterium (4.16) wyznaczając wartość  $m(x)$  dla każdego z banków, a następnie porządkujemy je na skali preferencji od najlepszego do najgorszego.

Korzystając ze wzoru (4.15), wyznaczamy maksymalne wartości kryteriów cząstkowych  $f_i^*$ ,  $i = 1, \dots, 3$  osiągnięte w zbiorze  $D$ :  $f_1^* = 5$ ,  $f_2^* = 1000$ ,  $f_3^* = 10$ , a następnie konstruujemy macierz stopni realizacji I rodzaju celów cząstkowych (tablica 4.3).

TABLICA 4.3  
Stopnie realizacji I rodzaju celów cząstkowych dla banków z przykładu 4.2

| Stopnie realizacji | Banki |      |      |      |     | $w_i$ |
|--------------------|-------|------|------|------|-----|-------|
|                    | A     | B    | C    | D    | E   |       |
| $\tilde{r}_1(x)$   | 0,9   | 1    | 0,95 | 1    | 1   | 0,6   |
| $\tilde{r}_2(x)$   | 1     | 0,2  | 0,6  | 0,4  | 0,1 | 0,2   |
| $\tilde{r}_3(x)$   | 0,8   | 0,6  | 1    | 0,7  | 0,5 | 0,2   |
| $m(x)$             | 0,9   | 0,76 | 0,89 | 0,82 | —   | ×     |

Ze zbioru banków eliminujemy bank  $E$  z uwagi na niespełnienie warunku realizacji drugiego kryterium w wymaganym stopniu. Ranking banków przedstawia się następująco:  $A - C - D - B$ .

Stopień realizacji celu cząstkowego można wyznaczać, stosując wiele formuł. Sporządzmy ranking banków, korzystając ze wzoru (4.18).

Wyznaczamy, stosując wzór (4.17), minimalne wartości kryteriów cząstkowych  $f_i^{**}$ ,  $i = 1, \dots, 3$  osiągnięte w zbiorze  $D$  (pomijamy bank  $E$ ):  $f_1^{**} = 4,5$ ,  $f_2^{**} = 200$ ,  $f_3^{**} = 6$ ,  $f_1^* - f_1^{**} = 0,5$ ,  $f_2^* - f_2^{**} = 800$ ,  $f_3^* - f_3^{**} = 4$ .

Macierz stopni realizacji II rodzaju celów cząstkowych przybiera postać przedstawioną w tabelicy 4.4.

<sup>17</sup> Wydaje się, że nieco arbitralna eliminacja pewnych banków nie jest właściwa, gdyż zastosowanie metody rangowania powinno doprowadzić do umieszczenia ich na końcu rankingu. Pogląd ten jest niestety błędny. W przypadku posługiwania się stopniem realizacji I rodzaju nie napotkamy żadnego problemu — najslabsze banki będą umieszczone na końcu rankingu. W przypadku wykorzystywania stopnia realizacji II rodzaju niewyeliminowanie banków „odstających” prowadzi do zmniejszenia względnych różnic między bankami, dla których wartości kryteriów cząstkowych różnią się nieznacznie. Powoduje to, że waga odpowiedniego kryterium cząstkowego zostaje w niejawnym sposób zanizowana.



TABLICA 4.4  
Sposób wyznaczania stopni realizacji II rodzaju celów cząstkowych dla banków z przykładu 4.2

| Stopnie realizacji | Banki |     |     |     | $w_i$ |
|--------------------|-------|-----|-----|-----|-------|
|                    | A     | B   | C   | D   |       |
| $\bar{r}_1(x)$     | 0     | 1   | 1/2 | 1   | 0,6   |
| $\bar{r}_2(x)$     | 1     | 0   | 1/2 | 1/4 | 0,2   |
| $\bar{r}_3(x)$     | 1/2   | 0   | 1   | 1/4 | 0,2   |
| $m(x)$             | 0,3   | 0,6 | 0,5 | 0,7 | ×     |

Ranking:  $D - B - C - A$ .

Ranking może być przeprowadzony przy zastosowaniu różnych formuł, z których każda ma swoje wady i zalety. Trudno generalnie uznać, że wybrana formuła góruje nad pozostałymi, z czego wynika, że ustalanie rankingów jest obciążone dużym stopniem arbitralności.

Sposób wyznaczania stopni realizacji celów cząstkowych według formuły (4.18) faworyzuje względne różnice kosztów wartości absolutnych, co może w sposób niepożądany zniekształcać ranking w przypadku, gdy różnice względne osiągają niewielkie wartości dla kryterium, któremu przypisano dużą wagę — tymi słowami powoduje to przecenianie małych różnic.

Nader istotnym zagadnieniem jest sposób ustalania wartości wag przypisywanych kryteriom cząstkowym. Ze względu na subiektywizm związany z wyznaczaniem tych wartości pojawia się możliwość manipulowania wynikami. Jest oczywiste, że jeżeli istnieje bank, który jest najlepszy ze względu na wszystkie kryteria cząstkowe, to znajdzie się on na pierwszym miejscu w rankingu niezależnie od struktury wag. Niemniej jednak manipulując wagami, można znacząco wpłynąć na ranking.

W przykładzie bank, dla którego stopnie realizacji kryteriów cząstkowych są mniejsze od wymaganych, został wyeliminowany z rankingów i postępowanie to nie budzi zastrzeżeń<sup>18</sup>. Niekiedy jednak z rankingów trzeba także wyeliminować banki oferujące np. znacząco wyższe od rynkowych oprocentowanie depozytów, ponieważ waż może to świadczyć nie o dobrej, lecz o jego trudnej sytuacji finansowej i problemach z płynnością, wymuszających konieczność pozyskania środków finansowych.

<sup>18</sup> Czytelnik zechce sporządzić ranking banków, korzystając ze stopnia realizacji celu II rodzaju, uwzględniając wszystkie banki i porównać wyniki. Pouczające jest prześledzenie, w jaki sposób ranking wpływa na usunięcie banku, który ze względu na jedno kryterium cząstkowe odbiega od pozostałych banków. Nieusunięcie banku E prowadzi do zmniejszenia różnic między wartościami stopni realizacji II rodzaju między pierwszymi czterema bankami. Mniejsze zróżnicowanie jest równoważne z mniejszym wpływem rozpatrywanego kryterium na pozycję banku w rankingu.

wych na niekorzystnych warunkach. Włączenie takiego banku do rankingów w sposób niepożądany wpływa na wartości stopni realizacji celów cząstkowych, zniekształcając uzyskane wyniki. Łatwo sprawdzić, że jeżeli stopień realizacji celu cząstkowego jest wyznaczany ze wzoru  $\frac{f_i(x)}{f_i^*}$ , to zawyżenie wartości  $f_i^*$  prowadzi do

zaniżania zarówno wielkości  $\hat{r}_i(x) = \frac{f_i(x)}{f_i^*}$  dla banków, w przypadku których stopień realizacji celu cząstkowego nie osiąga maksimum, jak i do zaniżania różnic między nimi. Jeżeli wartość metakryterium jest wyznaczana za pomocą wzoru (4.16), to  $i$ -te kryterium cząstkowe w mniejszym stopniu wpływa na ustalenie kolejności banków w rankingu. Wynika stąd, że konstrukcja zbioru porównywanych obiektów ekonomicznych powinna być poprzedzona przeprowadzeniem szczegółowej analizy merytorycznej.

Jest rzeczą oczywistą, że wybór kryteriów cząstkowych bywa arbitralny i — co należy wyraźnie podkreślić — nie musi wcale wynikać z chęci manipulowania wynikami.

Przedstawione postępowanie znajduje zastosowanie w wylanianiu zwycięzcy przetargów<sup>19</sup>. Wskazane wyżej mankamenty wyznaczania rankingów na podstawie metodologii proponowanej na gruncie programowania wielokryterialnego są dziedziczone przez procedury przetargowe odwołujące się do metod bazujących na tym podejściu.

Przykład 4.2 ilustruje zastosowanie omawianych metakryteriów w przypadku, gdy zbiór decyzji jest dyskretny. Postępowanie w przypadku ciągłym ilustruje przykład 4.7.

Problem wyznaczenia decyzji, dla której wartości kryteriów cząstkowych w możliwie małym stopniu odbiegają od wektora-wzorca, tj. wektora postulowanych poziomów realizacji celów, prowadzi do tzw. **problemu programowania celowego**, przyjmującego postać<sup>20</sup>:

$$\sum_{i=1}^k w_i |c_i x - c_{i0}| \rightarrow \min, \quad (4.20)$$

$$Ax \leq b, \quad (4.21)$$

$$x \geq 0. \quad (4.22)$$

Wielkość  $c_{i0}$ , podobnie jak w przypadku problemu (4.6)–(4.8), jest postulowanym poziomem realizacji wartości funkcji celu dla  $i$ -tego kryterium cząstkowego.

<sup>19</sup> Wyznaczając stopnie realizacji celów cząstkowych, w zależności od kierunku optymalizacji, korzysta się ze wzoru (4.18) lub (4.19).

<sup>20</sup> Analizowany problem przypomina problem (4.6)–(4.8). Zamiast minimalizować największe, co do wartości bezwzględnej, odchylenie, minimalizujemy ważoną sumę odchylenia.

Nie musi być ona równa optymalnej wartości  $i$ -tego kryterium osiągalnej w zbiorze  $D$  określonym przez warunki (4.21)–(4.22). Istota podejścia polega na wyborze takiego wektora  $x \in D$ , dla którego odchylenia wartości cząstkowych funkcji celu  $c^i x$  od wartości postulowanych (pożądanych)  $c_{i0}$  są możliwie małe. Jeżeli istnieje taki punkt  $\bar{x} \in D$ , że  $c^i \bar{x} = c_{i0}$ ,  $i = 1, \dots, K$ , to (4.20) osiąga w punkcie optymalnym, którym jest  $\bar{x}$ , wartość równą 0.

Problem (4.20)–(4.22) jest problemem programowania nieliniowego. Po dokonaniu prostych przekształceń można go sprowadzić do problemu programowania liniowego.

Zdefiniujmy zmienne:

$$y_i = \max \{0, c^i x - c_{i0}\},$$

$$z_i = \max \{0, c_{i0} - c^i x\}.$$

Łatwo zauważyć, że:

$$|c^i x - c_{i0}| = y_i + z_i \quad (y_i + z_i \text{ jest zawsze nieujemne}),$$

$$c^i x - c_{i0} = y_i - z_i \quad (\text{jest spełnione dla dowolnych } c^i x - c_{i0}),$$

$$y_i, z_i \geq 0, \quad y_i z_i = 0.$$

Nierówności  $y_i, z_i \geq 0$ ,  $i = 1, \dots, K$  wynikają wprost z definicji zmiennej. Jeżeli  $c^i x - c_{i0} \geq 0$ , to  $y_i = c^i x - c_{i0} \geq 0$ , a  $z_i = 0$ , z czego wynika, że  $y_i z_i = 0$ . Jeżeli  $c^i x - c_{i0} < 0$ , to  $y_i = 0$ , a  $z_i = c_{i0} - c^i x > 0$ , a więc także w tym przypadku  $y_i z_i = 0$ .

Z przytoczonych zależności natychmiast wynika, że dla  $c^i x - c_{i0} \geq 0$ ,  $y_i + z_i = c^i x - c_{i0}$ , natomiast gdy  $c^i x - c_{i0} < 0$ , to  $y_i + z_i = -c^i x + c_{i0}$ , co implikuje równość  $|c^i x - c_{i0}| = y_i + z_i$ . Po elementarnych przekształceniach otrzymujemy równość  $c^i x - c_{i0} = y_i - z_i$ . Ostatecznie problem (4.20)–(4.22) przybiera postać:

$$\sum_{i=1}^K w_i (y_i + z_i) \rightarrow \min, \quad (4.23)$$

$$c^i x - y_i + z_i = c_{i0}, \quad i = 1, \dots, K, \quad (4.24)$$

$$Ax \leq b, \quad (4.25)$$

$$x \geq 0, \quad y_i, z_i \geq 0, \quad i = 1, \dots, K. \quad (4.26)$$

Warunki  $y_i z_i = 0$  zostały pominięte — w przeciwnym przypadku otrzymamy problem nie byłby problemem PL. Łatwo jednak dowiedzieć, że jeżeli  $(x^*, y^*, z^*)$  jest rozwiązaniem optymalnym problemu (4.23)–(4.26), to  $y_i^* z_i^* = 0$ ,  $i = 1, \dots, K$ .

Teoretycznym uzasadnieniem wykorzystywania pomocniczego problemu liniowego do wyznaczenia rozwiązania optymalnego problemu wyjściowego jest następujące twierdzenie.

Jeżeli  $x^*$  jest rozwiązaniem optymalnym problemu (4.20)–(4.22), to  $(x^*, y^*, z^*)$  jest rozwiązaniem optymalnym problemu (4.23)–(4.26) i *vice versa*.

Różnice  $c^i x - c_{i0}$  nie są wielkościami bezwymiarowymi, co prowadzi do niezgodności skal liczbowych lub jednostek. Odchylenia między wartością kryterium cząstkowego dla danej decyzji a wartością wzorca nie są porównywalne. Zmiana jednostek, w których wyraża się wartość kryterium cząstkowego, prowadzi w sposób niejawnny do zmiany wagi przypisywanej temu kryterium. Ponadto odchyleniom *in plus* i *in minus* od wartości pożądanej przypisywana są te same wagi<sup>21</sup>.

### PRZYKŁAD 4.3

Rozważmy wielokryterialną wersję klasycznego **problemu wyboru optymalnego asortymentu produkcji** (zob. rozdział 1). Za optymalny plan produkcji uznamy taki plan, który jest technologicznie dopuszczalny i gwarantuje możliwe najmniejsze odchylenia wartości kryteriów cząstkowych (zysku, przychodu itp.) od wartości postulowanych.

Wprowadźmy następujące oznaczenia:

$$c^1 = [2 \ 5 \ 4] \quad \text{— wektor cen zbytu (tys. zł),}$$

$$c^2 = [1 \ 1,5 \ 2] \quad \text{— wektor zysków jednostkowych (tys. zł),}$$

$$c_{10} = 100, \quad c_{20} = 50 \quad \text{— pożądana wielkość przychodu i zysku (tys. zł),}$$

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 1 \end{bmatrix} \quad \text{— macierz nakładów jednostkowych surowców niezbędnych do wytworzenia wyrobów (kg/sztukę),}$$

$$b = \begin{bmatrix} 100 \\ 150 \end{bmatrix} \quad \text{— możliwości przerobu surowców (kg),}$$

$$w_1 = 0,8, \quad w_2 = 0,2 \quad \text{— wagi przypisywane przez decydenta odchyleniom od wartości pożądanych; im większa waga, tym mniej pożądane odchylenie,}$$

$$x \quad \text{— wektor określający wielkość produkcji wyrobów (sztuki).}$$

<sup>21</sup> Z problemem tym można się jednak uporać, zastępując wyrażenie  $w_i(y_i + z_i)$  wyrażeniem  $d_i y_i + h_i z_i$ ,  $d_i \geq 0$ ,  $h_i \geq 0$ ,  $d_i + h_i > 0$ . Można dowiedzieć, że także w tym przypadku, jeżeli  $(x^*, y^*, z^*)$  jest rozwiązaniem optymalnym pomocniczego problemu liniowego z funkcją celu  $\sum_{i=1}^K (d_i y_i + h_i z_i) \rightarrow \min$ , przy warunkach (4.24)–(4.26), to  $y_i^* z_i^* = 0$ ,  $i = 1, \dots, K$ .

Problem (4.20)–(4.22) przyjmuje postać:

$$\begin{aligned} 0,8|2x_1 + 5x_2 + 4x_3 - 100| + 0,2|x_1 + 1,5x_2 + x_3 - 50| &\rightarrow \min, \\ 2x_1 + x_2 + 3x_3 &\leq 100, \\ x_1 + 2x_2 + x_3 &\leq 150, \\ x_1, x_2, x_3 &\geq 0. \end{aligned}$$

Po wprowadzeniu zmiennych swobodnych oraz zmiennych:

$$\begin{aligned} y_1 &= \max\{0, 2x_1 + 5x_2 + 4x_3 - 100\}, \\ y_2 &= \max\{0, x_1 + 1,5x_2 + x_3 - 50\}, \\ z_1 &= \max\{0, 100 - 2x_1 - 5x_2 - 4x_3\}, \\ z_2 &= \max\{0, 50 - x_1 - 1,5x_2 - x_3\}, \end{aligned}$$

otrzymujemy następujące zadanie programowania liniowego:

$$\begin{aligned} 0,8(y_1 + z_1) + 0,2(y_2 + z_2) &\rightarrow \min, \\ 2x_1 + 5x_2 + 4x_3 - y_1 + z_1 &= 100, \\ x_1 + 1,5x_2 + x_3 - y_2 + z_2 &= 50, \\ 2x_1 + x_2 + 3x_3 + x_4 &= 100, \\ x_1 + 2x_2 + x_3 + x_5 &= 150, \\ x_1, \dots, x_5, y_1, y_2, z_1, z_2 &\geq 0. \end{aligned}$$

Nawiążmy do uwagi odnoszącej się do problemu niezgodności skal liczbowych lub jednostek. Gdyby zyski zostały wyrażone nie w tysiącach zł, ale w zł, wówczas funkcja celu zadania wyjściowego przyjęłaby postać:

$$0,8|2x_1 + 5x_2 + 4x_3 - 100| + 0,2|1000x_1 + 1500x_2 + 1000x_3 - 50000| \rightarrow \min.$$

Funkcja celu pomocniczego zadania liniowego nie ulega wprawdzie zmianie, ale drugi warunek ograniczający trzeba zastąpić warunkiem:

$$1000x_1 + 1500x_2 + 1000x_3 - y_2 + z_2 = 50000,$$

co prowadzi do zmiany rozwiązania optymalnego pomocniczego zadania liniowego.

Podany przykład jest dość trywialny i nieco sztuczny. Zauważmy, że w przypadku wyrażenia kryteriów cząstkowych w różnych jednostkach następstwa właściwego ich doboru nie są tak ewidentne. Jeżeli np. jedno kryterium jest wyrażone w złotych przychodu, a drugie w „udziale w rynku”, to w ogóle trudno powiedzieć, czy dobór jednostek jest właściwy w tym sensie, że niejawnie faworyzuje jednego z kryteriów. Z przykładu jasno wynika, że problem uzgodnienia jednostek pomiaru nie jest trywialny, a „niewłaściwy” ich dobór może — niekiedy w sposób niezauważony — wpływać na uzyskiwane wyniki.

### 4.3.3. Programowanie leksykograficzne

Wymaga się uporządkowania kryteriów od najważniejszego do najmniej ważnego. Niech  $f_i$  będzie  $i$ -tym co do ważności kryterium, a  $D_i$  zbiorem decyzji (skończonym lub nieskończonym), dla których to kryterium osiąga optimum. Proponuje się następujący algorytm wyznaczania najlepszej decyzji.

#### Algorytm 4.1. Algorytm programowania leksykograficznego

Krok 1. Wybierz kryterium o aktualnie najwyższym priorytecie (na początku będzie nim  $f_1$ ) i wyznacz zbiór  $D_i$ . Jeżeli  $i$ -te kryterium jest maksymalizowane, to  $D_i = \left\{x: \max_{x \in D_{i-1}} f_i(x)\right\}$ . W przypadku minimalizacji kryterium

$$D_i = \left\{x: \min_{x \in D_{i-1}} f_i(x)\right\}. \text{ Dla } i = 1 \text{ przyjmij } D_0 = D.$$

Krok 2. Jeżeli  $i = K$  (wszystkie kryteria zostały rozpatrzone) lub  $i < K$  i  $D_i$  jest jednoelementowy, stop. Zbiór  $D_i$  uznaje się za zbiór rozwiązań optymalnych. W przeciwnym przypadku, tj. gdy  $i < K$  (nie zostały rozpatrzone wszystkie kryteria) oraz  $D_i$  jest wieloelementowy (można go ograniczyć, wykorzystując pozostałe kryteria), podstaw  $i := i + 1$  i przejdź do kroku 1.

Algorytm ulega zakończeniu, gdy zbiór  $D_i$  jest jednoelementowy i zbiór kryteriów nie został wyczerpany ( $i < K$ ),  $D_i$  jest jednoelementowy i wszystkie kryteria zostały rozpatrzone ( $i = K$ ) lub też po rozpatrzeniu wszystkich kryteriów  $D_K$  jest wieloelementowy.

Wadą przedstawionej propozycji jest to, że zbiór  $D_i$  staje się jednoelementowy po rozpatrzeniu pierwszych 2–3 kryteriów; pozostałe kryteria nie mają wpływu na otrzymane rozwiązanie. Dlatego proponuje się modyfikację polegającą na tym, aby zbioru  $D_i$  nie tworzyły te elementy  $x \in D_{i-1}$ , dla których  $f_i$  osiąga optimum, lecz te, dla których — w przypadku maksymalizacji kryterium — zachodzi nierówność:

$$f_i(x) \geq (1 - d_i)f_i^*, \quad (4.27)$$

gdzie  $d_i$  jest współczynnikiem tolerowanych odchyłeń względnych unormowanym na przedziale  $[0, 1]$ . Jeżeli  $d_i = 0$ , to żadne odchylenia nie są tolerowane.

Prawa strona (4.27) jest nieujemna, zatem korzystając z tej nierówności *implicite* zakłada się, że  $f_i(x) \geq 0$ ,  $i = 1, \dots, K$ ,  $x \in D^{22}$ . Jeżeli założenie to jest nie do

<sup>22</sup> Elementy zbioru  $x \in D_{i-1}$ , dla których  $f_i(x) < 0$ , nie mogą być wyeliminowane bez względu na wartości przyjmowane przez  $d_i$ , co pozostaje w sprzeczności z celem przyświecającym wprowadzeniu współczynników tolerowanych odchyłeń.

przyjęcia, nierówność (4.27) proponuje się zastąpić nierównościami:

$$f_i(x) \geq f_i^* - d_i(f_i^* - f_i^{**}). \quad (4.28)$$

Dla  $d_i = 0$  prawa strona nierówności (4.28) jest równa  $f_i^*$  (zerowa tolerancja), a dla  $d_i = 1$  równa się  $f_i^{**}$  (całkowita tolerancja).

Jeżeli kryterium jest minimalizowane, korzysta się z nierówności:

$$f_i(x) \leq f_i^{**} + d_i(f_i^* - f_i^{**}). \quad (4.29)$$

Dla  $d_i = 0$  prawa strona nierówności (4.29) jest równa  $f_i^{**}$ , a dla  $d_i = 1$  równa  $f_i^*$ .

Niekiedy postuluje się, aby  $d_i$  dodatkowo spełniały<sup>23</sup> nierówności  $d_i \geq d_{i+1}$ ,  $i = 2, \dots, K$ . Nierówności te zapewniają, że im mniej ważne kryterium, tym większe odchylenia od maksymalnej (minimalnej) wartości funkcji celu mogą być tolerowane. Algorytm, w którym zbiory  $D_i$  wyznacza się korzystając z nierówności (4.27)–(4.29), nosi nazwę **algorytmu programowania leksykograficznego z relaksacją hierarchizacji ważności celów**<sup>24</sup>.

#### PRZYKŁAD 4.4

W tabelicy 4.5 podano oceny sześciu sokowirówek z punktu widzenia trzech kryteriów uszeregowanych od najbardziej do najmniej ważnego: ceny ( $f_1$ ), efektywności pracy ( $f_2$ ) oraz głośności ( $f_3$ ). Efektywność pracy (ilość i jakość wyciskanego soku) i głośność były oceniane przez klientów w skali od 1 do 6, 1 — najlepsza (najcichsza), 6 — najgorsza (najgłośniejsza)<sup>25</sup>. Współczynniki tolerancji są równe  $d_1 = 0,2$ ,  $d_2 = 0,3$ ,  $d_3 = 0,4$ .

Wszystkie kryteria są minimalizowane<sup>26</sup>, eliminacji dokonujemy zatem korzystając ze wzoru (4.29). Zbiór  $D_0 = D = \{1, 2, 3, 4, 5, 6\}$ . Wyznaczamy minimalną wartość pierwszego kryterium w zbiorze  $D_0$ :  $f_1^{**} = 150$ . Zbiór  $D_1$  składa się z elementów, dla których spełniana jest nierówność  $f_1(x) \leq f_1^{**} + d_1(f_1^* - f_1^{**}) = 150 + 0,2(250 - 150) = 170$ :  $D_1 = \{1, 4\}$ . W zbiorze  $D_1 = \{1, 4\}$  minimalna wartość drugiego kryterium jest równa 3, a maksymalna 5. Wyznaczamy zbiór  $D_2$ , biorąc pod uwagę elementy należące do  $D_1$ , które spełniają nierówność  $f_2(x) \leq f_2^{**} + d_2(f_2^* - f_2^{**}) = 3 + 0,3(5 - 3) = 3,6$ . Zbiór  $D_2$  jest zbiorem jednoelementowym  $D_2 = \{1\}$ , co kończy postępowanie.

<sup>23</sup> Założenie to przyjmuje się zarówno w przypadku maksymalizacji, jak i minimalizacji.

<sup>24</sup> Zaproponowana relaksacja nie gwarantuje, że wszystkie kryteria będą rozpatrzone.

<sup>25</sup> Głośność i efektywność pracy można zmierzyć w sposób obiektywny, ale oceny te nie muszą pokrywać się z subiektywną oceną klientów. Na przykład poziom hałasu mierzony w dB może być postrzegany jako mniej lub bardziej uciążliwy w zależności od ilości składowych harmonicznych i nieharmonicznych w widmie dźwięku. Dlatego nie zawsze jest celowe — nawet jeżeli jest to możliwe — korzystanie z obiektywnych mierników cech.

<sup>26</sup> Ze sposobu zdefiniowania skali pomiarowej wynika, że kryteria drugie i trzecie są minimalizowane.

TABLICA 4.5  
Oceny sokowirówek

| Sokowirówka | $f_1$ | $f_2$ | $f_3$ |
|-------------|-------|-------|-------|
| 1           | 170   | 3     | 4     |
| 2           | 220   | 2     | 2     |
| 3           | 180   | 2     | 3     |
| 4           | 150   | 5     | 4     |
| 5           | 250   | 1     | 6     |
| 6           | 190   | 6     | 3     |

Zastosowanie algorytmu programowania leksykograficznego z relaksacją hierarchizacji ważności celów w przypadku ciągłym ilustruje przykład 4.8.

#### 4.3.4. Minimalizacja odległości od punktu idealnego

Niech  $f_1^*, \dots, f_k^*$  będą maksymalnymi wartościami funkcji  $f_1, \dots, f_k$  osiąganymi w zbiorze  $D$ . Oznacza to, że istnieją takie  $\bar{x}^1, \dots, \bar{x}^k \in D$ , że:

$$f_1(\bar{x}^1) = f_1^*, f_2(\bar{x}^2) = f_2^*, \dots, f_k(\bar{x}^k) = f_k^*. \quad (4.30)$$

Nie wynika z tego, że w zbiorze  $D$  istnieje takie  $x^0$ , dla którego:

$$f_1(x^0) = f_1^*, f_2(x^0) = f_2^*, \dots, f_k(x^0) = f_k^*. \quad (4.31)$$

tj. taka decyzja, która maksymalizuje (ogólnie optymalizuje) wszystkie kryteria cząstkowe. Może ona w ogóle nie istnieć lub też istnieć, ale znajdować się poza zbiorem  $D$ .

Można jednak postulować wybór takiej decyzji  $x^* \in D$ , która leży „najbliżej” punktu idealnego  $M$  o współrzędnych  $f_1^*, \dots, f_k^*$  i uznać ją za optymalną.

Punktem optymalnym byłby taki punkt  $x^* \in D$ , dla którego wyrażenie:

$$\left\{ \sum_{i=1}^k [f_i(x^0) - f_i(x^*)]^p \right\}^{1/p}, \quad p > 1, \quad (4.32)$$

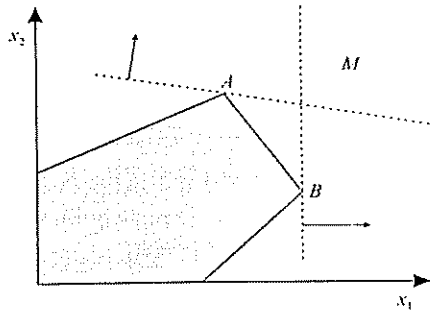
osiąga minimum<sup>27</sup>. Zazwyczaj przyjmuje się, że  $p = 2$ , co oznacza minimalizację odległości euklidesowej.

<sup>27</sup> Jeżeli  $x^0 \in D$ , to (4.32) jest równe 0, a  $x^0 = x^*$ .

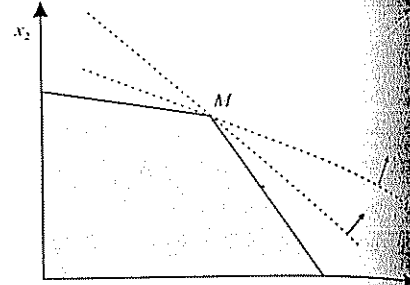
RYSUNEK 4.4

## Sposób wyznaczania punktu idealnego

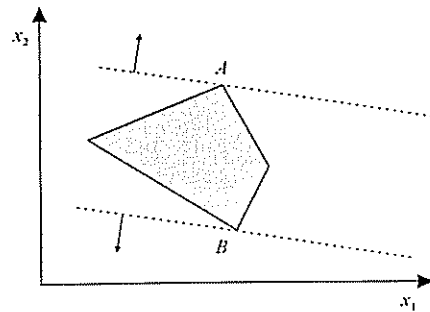
A. Maksima cząstkowe znajdują się w punktach A i B. Punkt idealny M nie jest dopuszczalny. Strzałki pokazują gradient funkcji



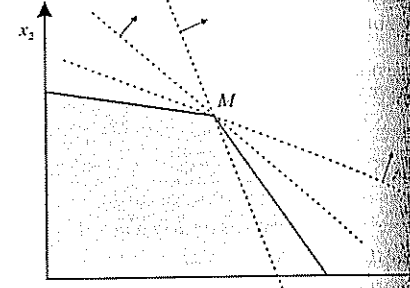
B. Maksima cząstkowe pokrywają się. Punkt idealny M jest dopuszczalny



C. Izokwanty równoległe. Punkt idealny nie istnieje



D. Nie istnieje punkt, w którym przecinają się izokwanty korespondujące z optimum cząstkowymi. Punkt idealny nie istnieje



U w a g a: Sporządzając rysunek przyjęto założenie, że zarówno warunki ograniczające, jak i funkcja celu są liniowe

Punkt  $M$  należy do  $D$ , jeżeli istnieje taki punkt, w którym wszystkie kryteria cząstkowe osiągają swoje maksima (optima). W przeciwnym razie punkt ten leży poza zbiorem  $D$ . Obie sytuacje zilustrowano na rysunku 4.4 (A i B).

Punkt idealny nie istnieje, jeżeli:

- 1) izokwanty są równoległe i się nie pokrywają (rysunek 4.4C);
- 2) istnieje więcej kryteriów cząstkowych (rysunek 4.4D).

W drugim przypadku tylko w wyjątkowej sytuacji będzie istniał punkt, w którym przecinają się wszystkie izokwanty przechodzące przez punkty korespondujące z optimum cząstkowymi.

## 4.4. Liniowe problemy wielokryterialnego podejmowania decyzji

Rozważmy problem tzw. wektorowego programowania liniowego:

$$Cx \rightarrow \max, \quad (4.33)$$

$$Ax = b, \quad (4.34)$$

$$x \geq 0. \quad (4.35)$$

Zakłada się, że wszystkie kryteria cząstkowe są funkcjami liniowymi, a  $C$  jest macierzą współczynników kryteriów cząstkowych, której wiersze korespondują z poszczególnymi kryteriami. Funkcja celu postuluje maksymalizację wektora ocen cząstkowych. Wyznaczanie rozwiązań optymalnych problemu (4.33)–(4.35) utożsamia się z wyznaczeniem wszystkich rozwiązań sprawnych. Dowodzi się, że **niebazowe rozwiązania sprawne** są kombinacjami wypukłymi<sup>28</sup> sprawnych rozwiązań bazowych, wobec czego problem poszukiwania **rozwiązań wektorowo optymalnych** problemu (4.33)–(4.35) sprowadza się do efektywnego sposobu wyznaczania bazowych rozwiązań sprawnych. W celu wyznaczenia pierwszego rozwiązania sprawnego, którym w szczególności może być pierwsze bazowe rozwiązanie dopuszczalne, można zastosować metodę Bensona.

Zbiór baz sprawnych jest skończony. Każde dwie bazy sprawne można otrzymać, generując ciąg sąsiednich baz sprawnych.

Zbiór bazowych rozwiązań sprawnych jest podzbiorem wszystkich rozwiązań bazowych (4.34)–(4.35), których liczba jest skończona. Wynika stąd skończoność zbioru bazowych rozwiązań sprawnych. Generowanie sąsiednich baz sprawnych polega na wymianie w bazie jednej zmiennej. Metodą realizującą tę ideę jest ADBASE<sup>29</sup>. Ponieważ zbiór bazowych rozwiązań sprawnych jest zazwyczaj liczny, proponuje się jego ograniczenie przez zastosowanie metod filtracji rozpinającej lub zacieśniającej<sup>30</sup>. Wykorzystując przykład 4.5, naszkicujemy idee leżące u pod-

<sup>28</sup> Kombinacją wypukłą rozwiązań  $x^1, \dots, x^n$  nazywamy rozwiązanie  $x = \alpha_1 x^1 + \dots + \alpha_n x^n$ , gdzie  $\alpha_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n \alpha_i = 1$ .

<sup>29</sup> Opis metody w literaturze polskiej można znaleźć w [Trzaskalik, 2006, s. 21–29]. Znajduje się tam także krótki opis metody Bensona.

<sup>30</sup> Procedury filtracji są heurystycznymi algorytmami wspomagającymi decydenta przy wyborze rozwiązania końcowego ze skończonego zbioru rozwiązań o dużej liczebności. Filtracja rozpinająca pozwala na wybór rozwiązań sprawnych, których wzajemne odległości są nie mniejsze od zadanych przez decydenta. Odległością może być np. odległość euklidesowa. Istota postępowania sprowadza się

staw programowania wektorowego oraz filtracji. Następnie pokażemy sposób postępowania prowadzący do maksymalizacji celu głównego oraz zastosowanie metryki kryterium w przypadku ciągłym. Przykład zakończymy opisem sposobu wyznaczania punktu idealnego i punktu dopuszczalnego najbardziej zbliżonego do idealnego.

Decyzje podejmowane w każdym przedsiębiorstwie są wypadkową gry sił różnych grup interesów. Właściciel jest zainteresowany maksymalizacją zysku, menedżer maksymalizacją przychodu (z uwagi na prowizyjny system wynagrodzeń, uzależniający je od przychodu), a pracownicy z oczywistych względów maksymalizacją liczby roboczogodzin.

Rozważmy wielokryterialną wersję problemu wyboru optymalnego asortymentu produkcji, sprowadzającego się do wyboru asortymentu produkcji dopuszczalnego ze względu na możliwości przerobu surowców i maksymalizującego następujące kryteria cząstkowe: zysk, przychód oraz liczbę roboczogodzin<sup>31</sup>.

#### PRZYKŁAD 4.5

Firma wytwarza dwa wyroby — A i B. Ceny zbytu wyrobów wynoszą 18 i 6 zł/sztukę, a zyski jednostkowe — 2 zł w przypadku każdego wyrobu. Do wytworzenia jednostki wyrobu A potrzebna jest jedna roboczogodzina, wyrobu B — pięć roboczogodzin. Do produkcji wyrobów wykorzystuje się dwa surowce — S1 i S2. W okresie planistycznym można przerobić 18 t surowca S1 oraz 12 t surowca S2. Nakłady surowców niezbędne do wytworzenia jednostki wyrobów (kg/jednostkę) podano w tablicy 4.6.

Zmiennymi decyzyjnymi są wielkości produkcji każdego z wyrobów. Dla ułatwienia obliczeń przyjmijmy, że zostały one wyrażone w tys. sztuk<sup>32</sup>.

do wyznaczenia mniej licznego zbioru rozwiązań wystarczająco zróżnicowanych. W przypadku filtracji zacieśniającej decydent podaje rozwiązanie (wektor) i oczekuje wyznaczenia zadanej liczby rozwiązań (wektorów) leżących najbliżej zadanego rozwiązania. W obu przypadkach rola analityka polega nie na zasugerowaniu konkretnego rozwiązania, ale na wskazaniu zbioru decyzji odznaczających się pożądanymi cechami (zob. [Trzaskalik, 2006, s. 30–33]).

<sup>31</sup> To, czy dane kryterium ma być maksymalizowane czy też minimalizowane, zależy od kontekstu ekonomicznego. Przyjęcie w przykładzie, że liczba roboczogodzin jest minimalizowana, byłoby równoznaczne z nieuwzględnieniem interesów pracowników.

<sup>32</sup> Przy tak zdefiniowanych zmiennych lewe strony nierówności (4.39) są wyrażone w tonach, co sprawia, że nie ma potrzeby mnożenia prawych stron przez 1000 w celu zapewnienia zgodności jednostek. Rozwiązania optymalne sformułowanych zadań optymalizacyjnych nie muszą być całkowitoliczbowe. Proponuje się wówczas obcięcie części ułamkowej. Jeżeli współczynniki macierzy A są liczbami nieujemnymi, nie prowadzi to, ze względu na zwrot nierówności, do pojawienia się rozwiązań niedopuszczalnych, aczkolwiek uzyskane w ten sposób rozwiązanie całkowitoliczbowe nie musi być optymalne. Jeżeli składowe wektora x są „dużymi” liczbami, nie powoduje to wystąpienia znaczących błędów.

TABLICA 4.6  
Jednostkowe nakłady surowców  
niezbędne do wytworzenia  
wyrobów A i B

| Surowce | Wyroby |   |
|---------|--------|---|
|         | A      | B |
| S1      | 2      | 3 |
| S2      | 2      | 1 |

Wyznamy plany produkcji maksymalizujące zysk, przychód oraz liczbę roboczogodzin. Jeżeli plany te się pokrywają, postępowanie ulega zakończeniu. Problem sprowadza się do wyznaczenia maksimum następujących funkcji:

$$f_1(x) = 2x_1 + 2x_2, \quad (4.36)$$

$$f_2(x) = 18x_1 + 6x_2, \quad (4.37)$$

$$f_3(x) = x_1 + 5x_2, \quad (4.38)$$

gdzie zbiór rozwiązań dopuszczalnych  $D$  jest dany przez ograniczenia:

$$\begin{aligned} 2x_1 + 3x_2 &\leq 18 & (0, 6) & (9, 0) \\ 2x_1 + x_2 &\leq 12 & (0, 12) & (6, 0) \\ x_1, x_2 &\geq 0. \end{aligned} \quad (4.39)$$

Ze względu na jednostki, w których wyrażono zmienne decyzyjne, pierwsze dwa kryteria są wyrażone w tys. zł, a ostatnie w tys. roboczogodzin. Łatwo sprawdzić, że maksima cząstkowe się nie pokrywają. Funkcja  $f_1$  osiąga maksimum w punkcie B o współrzędnych  $[4, 5 \ 3]^T$ ,  $f_2$  w punkcie C o współrzędnych  $[6 \ 0]^T$ , a  $f_3$  w punkcie A o współrzędnych  $[0 \ 6]^T$ , co pokazano na rysunku 4.5<sup>33</sup>. Maksima kryteriów osiągnane w zbiorze danym przez (4.39) są następujące:  $f_1^* = 15$ ,  $f_2^* = 108$  i  $f_3^* = 30$ .

Optima cząstkowe nie pokrywają się, a zatem zbiór rozwiązań sprawnych nie jest jednoelementowy<sup>34</sup>. Konieczne jest zastosowanie jednej z metod, opisanych w punktach 4.3.1–4.3.4, w celu jego zawężenia do zbioru jednoelementowego

W celu wyznaczenia rozwiązania optymalnego można zastosować algorytmy programowania całkowitoliczbowego (zob. rozdział 8), które jednak cechują się małą efektywnością i rozwiązywanie zadań o dużych rozmiarach natrafia istotne przeszkody natury obliczeniowej. Ich zastosowanie jest jednak konieczne, gdy zmienne mogą przyjmować „małe” wartości, w szczególności wartości binarne.

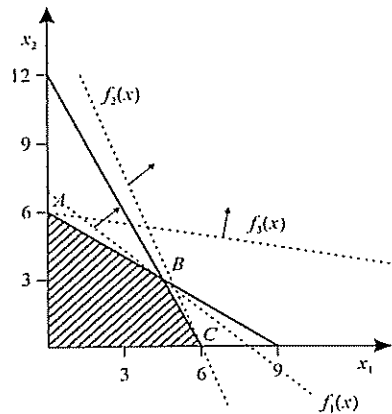
<sup>33</sup> Wszystkie rysunki ilustrujące zastosowanie metod opisanych w punktach 4.3.1–4.3.4 w przypadku ciągłym wykonano w przestrzeni decyzyjnej.

<sup>34</sup> Przypominamy, że rozwiązania nazywamy Pareto-optymalnymi, gdy korzystamy z przestrzeni kryterialnej, a sprawnymi — w przypadku przestrzeni decyzyjnej. W przypadku ciągłym wygodniej jest ilustrować postępowanie, korzystając z przestrzeni decyzyjnej.

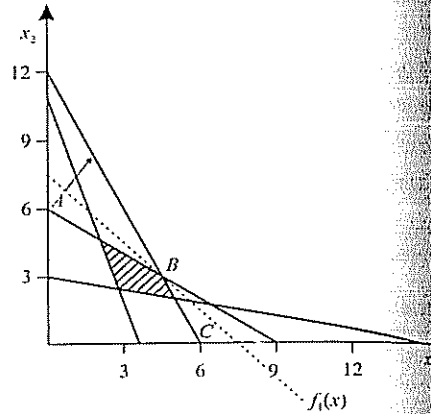
RYSUNEK 4.5

### Maksymalizacja celu głównego zapewniająca realizację celów cząstkowych w satysfakcjonującym stopniu

A. Optima cząstkowe znajdują się w punktach  $A, B, C$ . Liniami kropkowanymi narysowano izokwanty odpowiadające kryteriom cząstkowym. Strzałki określają kierunki gradientów kryteriów cząstkowych



B. Zaznaczono nowy zbiór rozwiązań dopuszczalnych powstały po dodaniu warunków (4.40) i (4.41). Punktem optymalnym jest punkt



wyznaczającego „najlepszy” plan produkcji. Niżej pokażemy, w jaki sposób zastosować przedstawione metody w przypadku, gdy zbiór rozwiązań nie jest zbiorem dyskretnym. Postępowanie rozpoczynamy od zilustrowania sposobu zastosowania metody filtracji zacieśniającej.

Łatwo sprawdzić, że optymalność w sensie Pareto nie pozwala na wybór optymalnego planu produkcji dla dowolnego, dwuelementowego podzbioru kryteriów. Rozważmy pierwsze dwa kryteria. Zbiór rozwiązań sprawnych leży na odcinku  $\overline{BC}$ . Przechodząc z punktu  $B$  do  $C$  podwyższamy wartość pierwszego kryterium i obniżamy wartość drugiego kryterium. Zbiór ten jest nieskończony, co nie pozwala na wybór „najlepszego” planu produkcji. Czytelnik zechce sprawdzić, że w przypadku pierwszego i trzeciego kryterium zbiór rozwiązań sprawnych dany jest przez odcinek  $\overline{AB}$ .

Rozwiązania sprawne mogą leżeć nie na odcinku, ale na łamanej. Z takiej sytuacji mamy do czynienia w przypadku, gdy kryteriami cząstkowymi są kryteria drugie i trzecie. Zbiorem rozwiązań sprawnych jest łamana  $ABC$ . Łatwo sprawdzić, że w przypadku rozpatrywania wszystkich kryteriów zbiorem rozwiązań sprawnych jest także łamana  $ABC$ , a sprawne rozwiązania bazowe korespondują z punktami wierzchołkowymi  $A, B, C$ . Efektywny numerycznie sposób wyznaczania wszy-

kich bazowych rozwiązań sprawnych realizuje — jak wcześniej wspomniano — metoda ADBASE.

Powiedzmy, że decydenta interesują dwa sprawne rozwiązania bazowe leżące najbliżej rozwiązania  $x_1 = 4, x_2 = 3$ . Metodą filtracji zacieśniającej wyznaczmy te rozwiązania, zakładając, że odległość między nimi jest odległością euklidesową. Odległości między bazowymi rozwiązaniami sprawnymi korespondującymi z punktami  $A, B, C$  a rozwiązaniem podanym przez decydenta są równe<sup>35</sup>:

$$\sqrt{(0-4)^2 + (6-3)^2} = 5,$$

$$\sqrt{(4,5-4)^2 + (3-3)^2} = 0,5,$$

$$\sqrt{(6-4)^2 + (0-3)^2} \approx 3,61.$$

Punktami położonymi najbliżej punktu wskazanego przez decydenta są  $B$  i  $C$ .

Rozważmy teraz zagadnienie maksymalizacji celu głównego przy założeniu, że cele poboczne muszą być zrealizowane w minimalnym, satysfakcjonującym stopniu.

#### PRZYKŁAD 4.6 (cd. 4.5)

Wyznaczmy plan produkcji maksymalizujący zysk przy założeniu, że przychód nie może być mniejszy od 60 tys. zł (zachowanie wymaganego udziału w rynku), a liczba roboczogodzin nie mniejsza od 15 tys. (wynegocjowany przez załogę minimalny poziom zatrudnienia). Jest to problem, przed którym stoi właściciel, który — realizując własny interes — musi uwzględnić interesy sił istniejących w przedsiębiorstwie. Postulat realizacji kryteriów cząstkowych w stopniu nie mniejszym od wymaganego można zapisać w postaci następujących nierówności:

$$18x_1 + 6x_2 \geq 60 \quad (0, 10) (3\frac{1}{3}, 0), \quad (4.40)$$

$$x_1 + 5x_2 \geq 15 \quad (0, 3) (15, 0) \quad (4.41)$$

W celu wyznaczenia optymalnego planu produkcji maksymalizujemy (4.36) przy warunkach (4.39) uzupełnionych o warunki (4.40)–(4.41). Rozwiązaniem optymalnym jest, jak łatwo sprawdzić, punkt  $B(4,5, 3)$ , co ilustruje rysunek 4.5B. Zauważmy, że chociaż zbiór rozwiązań dopuszczalnych uległ ograniczeniu, to rozwiązanie optymalne nie uległo zmianie, co oznacza, że uwzględnienie przez właściciela kryteriów cząstkowych nie ograniczyło możliwości maksymalizacji zysku.

Powiedzmy, że załoga wynegocjowała poziom zatrudnienia, mierzony liczbą roboczogodzin, nie mniejszy od 25 tys. Zastępując warunek (4.41) warunkiem:

$$x_1 + 5x_2 \geq 25, \quad (4.42)$$

<sup>35</sup> Wyznaczając odległości euklidesowe, bierzemy pod uwagę tylko wartości zmiennych  $x_1, x_2$ .

otrzymujemy zadanie, którego rozwiązaniem optymalnym jest  $x_1 = 2\frac{1}{7}$ ,  $x_2 = 4\frac{1}{7}$ . Zysk odpowiadający nowemu planowi produkcji wynosi  $13\frac{2}{7}$  tys. zł. W tym przypadku ograniczenie zbioru rozwiązań dopuszczalnych pociągnęło za sobą zmianę rozwiązania optymalnego, gdyż poprzednie rozwiązanie optymalne utraciło własność dopuszczalności, co zechce sprawdzić Czytelnik sporządzając odpowiedni rysunek.

Można w tym momencie sformułować uwagę ogólnej natury: próba uwzględnienia w zbyt dużym stopniu interesów uczestników procesu decyzyjnego może doprowadzić nie tylko do ograniczenia możliwości realizacji celu, który został uznany za cel główny, lecz także do sprzeczności zadania.

Prześledźmy teraz sposób sprowadzenia problemu programowania wielokryterialnego do problemu jednokryterialnego za pomocą metakryterium przybierającego postać ważonej sumy kryteriów oraz ważonej sumy stopni realizacji celów cząstkowych.

#### PRZYKŁAD 4.7 (cd. 4.5)

Zakładając, że kryteria maksymalizacji zysku i przychodu są tak samo ważne dla decydenta, wyznaczmy rozwiązanie kompromisowe maksymalizujące ważoną sumę kryteriów. Metakryterium przyjmuje postać:

$$\frac{1}{2}(2x_1 + 2x_2) + \frac{1}{2}(18x_1 + 6x_2) = 10x_1 + 4x_2. \quad (4.43)$$

Maksymalizując (4.43) przy warunkach (4.39), otrzymujemy rozwiązanie optymalne  $[6 \ 0]^T$ , któremu odpowiada wartość funkcji celu równa 60. Zauważmy, że wartości tej trudno nadać sensowną interpretację. Ponadto, zastosowanie metakryterium może budzić zastrzeżenia z powodu niezgodnych skal liczbowych. Co prawda zysk i przychód są wyrażone w tych samych jednostkach, ale pomiar dokonuje się za pomocą różnych skal liczbowych: złoty zysku ma zupełnie inne znaczenie ekonomiczne niż złoty przychodu. Dlatego zakres stosowalności tego metakryterium jest bardzo ograniczony.

Pragnąc uwolnić się od problemu niezgodności jednostek i skal liczbowych, proponuje się maksymalizację ważonej sumy stopni realizacji celów cząstkowych. Zakładając, że wagi pierwszego i trzeciego kryterium są jednakowe, wyznaczmy plan produkcji maksymalizujący tę wielkość. Posłużmy się stopniem realizacji celu cząstkowego I rodzaju. Metakryterium przyjmuje postać:

$$\frac{1}{2} \cdot \frac{2x_1 + 2x_2}{15} + \frac{1}{2} \cdot \frac{x_1 + 5x_2}{30} = \frac{5x_1 + 9x_2}{60}. \quad (4.44)$$

Maksymalizując (4.44) przy warunkach (4.39), otrzymujemy rozwiązanie optymalne  $[0 \ 6]^T$ : stopnie realizacji celów cząstkowych są odpowiednio równe  $\frac{2 \cdot 0 + 2 \cdot 6}{15} = 0,8$  i  $\frac{1 \cdot 0 + 5 \cdot 6}{30} = 1$ .

Stopień realizacji celu cząstkowego można zdefiniować na wiele sposobów. Ponieważ w przykładzie  $f_i^* = 0$ ,  $i = 1, \dots, 3$ , to w przypadku maksymalizacji zastosowanie stopni realizacji I i II rodzaju prowadzi do otrzymania identycznych wyników.

Kolejnym sposobem wyznaczenia „najlepszego” planu produkcji jest zastosowanie algorytmu programowania leksykograficznego z relaksacją hierarchizacji ważności celów.

#### PRZYKŁAD 4.8 (cd. 4.5)

Przyjmując, że najważniejszym kryterium jest maksymalizacja zysku, wyznaczmy plan produkcji maksymalizujący liczbę roboczogodzin, zapewniający, że zysk dla tego planu nie będzie odbiegać o więcej niż 20% od maksymalnego osiągalnego zysku.

Maksymalny osiągalny zysk w zbiorze danym układem nierówności (4.39) jest równy  $f_1^* = 15$ . Niech współczynnik tolerowanych odchyień  $d_1$  będzie równy 0,2. Skorzystajmy z nierówności  $f_1(x) \geq (1 - d_1)f_1^*$ . Podstawiając  $f_1(x) = 2x_1 + 2x_2$ , uzyskujemy nierówność:

$$2x_1 + 2x_2 \geq 12. \quad (4.45)$$

Zbiór  $D_1$  otrzymujemy, dołączając do układu nierówności (4.39) nierówność (4.45) i w tak określonym zbiorze poszukujemy maksimum drugiej co do ważności funkcji, tj.  $f_3$ . Łatwo sprawdzić, że jej maksimum znajduje się w punkcie o współrzędnych  $[0 \ 6]^T$ . Punkt optymalny został wyznaczony, co kończy postępowanie, które ilustruje rysunek 4.6A.

Osobliwością uzyskanego wyniku jest to, że druga co do ważności funkcja kryterium jest realizowana w uzyskanym punkcie w 100%, a pierwsza w 80%. Przy założeniu, że maksymalne odchylenie nie może przekroczyć 10%, tj. że  $d_1 = 0,1$ , nierówność ograniczająca zbiór  $D$  przyjmuje postać  $2x_1 + 2x_2 \geq 13,5$ , a funkcja  $f_3$  osiąga swoje maksimum w punkcie o współrzędnych  $[2,25 \ 4,5]^T$ .

Na zakończenie prześledźmy sposób wyznaczania punktu idealnego. Zakładając, że decydent jest zainteresowany maksymalizacją zysku i liczby roboczogodzin, wyznaczmy punkt idealny oraz punkt dopuszczalny położony najbliżej punktu idealnego.



## PRZYKŁAD 4.9 (cd. 4.5)

W celu wyznaczenia punktu idealnego  $M$  znajdujemy współrzędne punktu, w którym przecinają się izokwanty przechodzące przez optima cząstkowe korespondujące z tymi/kryteriami. W tym celu wyznaczamy rozwiązanie układu równań

$$2x_1 + 2x_2 = 15,$$

$$x_1 + 5x_2 = 30,$$

którym jest  $x_1 = 1\frac{7}{8}$ ,  $x_2 = 5\frac{5}{8}$ .

W celu wyznaczenia współrzędnych punktu  $O$  należącego do zbioru rozwiązań dopuszczalnych i położonego najbliżej punktu idealnego nie ma potrzeby wyznaczania minimum funkcji (4.32) na zbiorze danym przez (4.39). Z rysunku 4.6 wynika, że punkt  $O$  musi leżeć na odcinku  $AB$ . Rzutujemy zatem punkt  $M$  na ten odcinek, kreśląc prostopadłą do tego odcinka przechodzącą przez punkt  $M$ . Równanie prostopadłej wyraża się wzorem  $2x_1 - 3x_2 = 2 \cdot 1\frac{7}{8} - 3 \cdot 5\frac{5}{8} = -13\frac{1}{8}$ . Współrzędne poszukiwanego punktu otrzymujemy z układu równań:

$$2x_1 - 3x_2 = -13\frac{1}{8},$$

$$2x_1 + 3x_2 = 18,$$

którego rozwiązaniem jest  $x_1 = 1,2185$ ,  $x_2 = 5,1875$ .

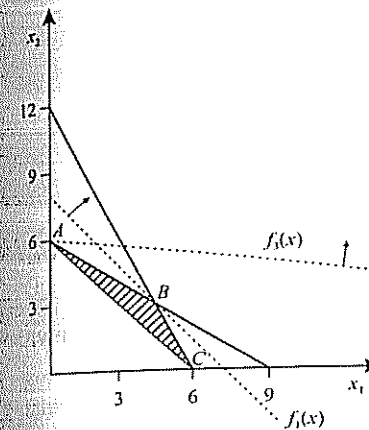
Celem tego rozdziału była prezentacja idei leżących u podstaw konstrukcji metod programowania wielokryterialnego. Czytelnika bliżej zainteresowanego tymi metodami oraz możliwościami ich zastosowania w praktyce odsyłamy do literatury.

Z uwagi na charakter podręcznika ograniczono się do opisu wybranych metod programowania wielokryterialnego, koncentrując się na zagadnieniu wyboru polegającym na wyznaczeniu najlepszego wariantu decyzyjnego. Pominięto problem sortowania polegający na przydziale wariantów decyzyjnych do określonych z góry kategorii oraz problem porządkowania sprowadzający się do podziału wariantów na klasy wariantów równie dobrych. Istotnym nieomówionym problemem jest wpływ ryzyka na proces podejmowania decyzji. Ryzyko prowadzi do nierozróżnialności zbliżonych do siebie wariantów decyzyjnych, która może być modelowana przez zastąpienie kryterium tzw. pseudokryterium. W podręczniku zdefiniowano zwykłą relację dominacji, nie uwzględniając możliwości modelowania preferencji decydenta za pomocą dominacji stochastycznych. Bardzo skrótowo potraktowano problem pomiaru i wyboru właściwej skali. Istotnym pominiętym zagadnieniem jest kwestia oceny stopnia racjonalności preferencji decydenta, zaproponowanie postępowania pozwalającego na ujawnienie nieracjonalnych preferencji oraz algorytmów prowadzących do wyboru sensownych decyzji.

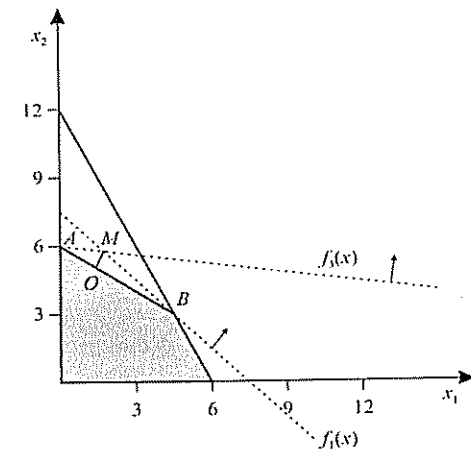
RYSUNEK 4.6

## Ilustracja działania algorytmu leksykograficznego oraz sposobu znajdowania punktu idealnego

A. Ilustracja postępowania w przypadku zastosowania algorytmu programowania leksykograficznego. Zaznaczono zmodyfikowany zbiór rozwiązań dopuszczalnych. Optimum znajduje się w punkcie  $A$



B. Punkt  $M$  jest punktem idealnym, leżącym poza zbiorem rozwiązań dopuszczalnych. Dokonując jego rzutu na odcinek  $AB$ , otrzymujemy punkt  $O$



Wiele z omawianych metod ma wersję „dialogową”, umożliwiającą decydentowi stopniowe dochodzenie do najlepszego rozwiązania (np. metoda Bipolar). Syntetyczny opis wymienionych zagadnień oraz przykłady zastosowań zainteresowany Czytelnik znajdzie w pracy T. Trzaskalika [2006].

## POLECENIA I PYTANIA KONTROLNE

1. Omówić problem subiektywizmu w programowaniu wielokryterialnym.
2. Zdefiniować optymalność w sensie Pareto. Dlaczego na ogół nie pozwala ona na wybór najlepszej decyzji?
3. Omówić problem nieporównywalności skal liczbowych i jednostek. Jakie sposoby postępowania proponuje się w przypadku wystąpienia nieporównywalności?
4. W jaki sposób można zdefiniować stopień realizacji celu w przypadku minimalizacji i maksymalizacji? Jakie ograniczenia wiążą się z wykorzystywaniem stopnia realizacji celu I rodzaju?
5. Na przykładzie wyznaczania rankingu banków wskazać na problemy doboru

obiektów do analizy. Uzasadnić konieczność wstępnej eliminacji obiektów z rankingu na podstawie analizy merytorycznej.

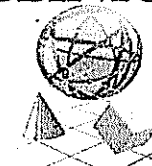
6. Jakie założenia trzeba przyjąć, konstruując funkcję użyteczności? Dlaczego jest możliwe zdefiniowanie funkcji użyteczności, gdy decydent działa w warunkach ryzyka?
7. Przedstawić ideę algorytmu programowania leksykograficznego. Na czym polega wyższość wersji algorytmu z relaksacją hierarchii ważności celów?
8. Omówić zagadnienie wyboru decyzji najbardziej zbliżonej do ideału. Kiedy i dlaczego nie istnieje punkt idealny?
9. Zdefiniować problem wektorowego programowania liniowego.
10. Przedstawić ideę filtracji rozpinającej i zacieśniającej.

#### BIBLIOGRAFIA

- Czerwiński Z., 1984, *Matematyka na usługach ekonomii*, PWN, Warszawa.
- Galas Z., Nykowski I., Żółkiewski Z., 1987, *Programowanie wielokryterialne*, PWE, Warszawa.
- Roy B., 1990, *Wielokryterialne wspomaganie decyzji*, WNT, Warszawa.
- Trzaskalik T. (red.), 2006, *Metody wielokryterialne na polskim rynku finansowym*, PWE, Warszawa.

## ROZDZIAŁ 5

# PROGRAMOWANIE NIELINIOWE



### 5.1. Wprowadzenie

Opracowanie kompletnej teorii programowania liniowego i efektywnych algorytmów wyznaczania rozwiązań optymalnych zadań PL o dużej liczbie zmiennych i warunków ograniczających oraz łatwość implementacji zaproponowanych algorytmów skłaniała do sprowadzania problemów decyzyjnych spotykanych w działalności gospodarczej do tego typu zadań. Od dawna wskazywano na mankamenty tego podejścia, argumentując, że prowadzi ono do konstrukcji modeli nadmiernie uproszczonych, o wątpliwej wartości praktycznej. Obecnie obserwuje się powolne odejście od **paradygmatu liniowości**. Przyczyny tego stanu rzeczy mają charakter zarówno techniczny, jak i merytoryczny. Rozwój teorii programowania nieliniowego, owocujący opracowaniem i implementacją akceptowalnie efektywnych algorytmów wyznaczania rozwiązań optymalnych i przybliżonych wielu klas problemów programowania nieliniowego spotykanych w praktyce oraz powszechna dostępność pakietów obliczeniowych pozwalających na wyznaczanie rozwiązań zadań o dużych rozmiarach, przyczynia się do rosnącej popularności tych metod. Dążenie do realistycznego modelowania zjawisk gospodarczych o rosnącej złożoności także skłania do odejścia od paradygmatu liniowości.

W rozdziale przedstawiono metody wyznaczania rozwiązań optymalnych i przybliżonych (suboptymalnych) klas problemów programowania nieliniowego najczęściej spotykanych w zastosowaniach praktycznych. Dokonując doboru metod wyznaczania ekstremum wybrano te, z którymi z największym prawdopodobieństwem zetknie się Czytelnik, rozwiązując problemy optymalizacji nieliniowej za pomocą dostępnych na rynku pakietów obliczeniowych. W ostatnim podrozdziale przedstawiono problemy programowania nieliniowego spotykane w praktyce, które po prostych przekształceniach można sprowadzić do problemu programowania liniowego.

Analizę własności i metod rozwiązywania problemów programowania nieliniowego poprzedzimy omówieniem zalet i ograniczeń paradygmatu liniowości.

Rozważmy liniową wersję **problemu wyboru optymalnego asortymentu produkcji**<sup>1</sup>. Polega on na wyznaczeniu asortymentu produkcji maksymalizującego przychód (zysk) producenta, dopuszczalnego ze względu na dostępność środków produkcji (ich możliwości przerobu) oraz możliwości zbytu, wymagane proporcje między wytwarzanymi wyrobami itp.

Wprowadźmy następujące oznaczenia:

- $c_j$  — cena zbytu (zysk jednostkowy) wyrobu  $j$ ,  $j = 1, \dots, n$ ,  
 $b_i$  — zapas (możliwości przerobu) środka produkcji  $i$ ,  $i = 1, \dots, m$ ,  
 $a_{ij}$  — nakład  $i$ -tego środka produkcji na wytworzenie jednostki wyrobu  $j$ ,  
 $x_j$  — wielkość produkcji wyrobu  $j$ .

Model matematyczny zagadnienia (w wersji podstawowej) przybiera następującą postać:

$$\sum_{j=1}^n c_j x_j \rightarrow \max, \quad (5.1)$$

$$\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, \dots, m, \quad (5.2)$$

$$x_j \geq 0, \quad j = 1, \dots, n. \quad (5.3)$$

Funkcja celu (5.1) postuluje wybór asortymentu produkcji  $x = [x_1, \dots, x_n]$  maksymalizujący przychód (w przypadku, gdy wagi  $c_j$  interpretujemy jako ceny zbytu wyrobów) lub zysk (w przypadku, gdy wagi określają zyski jednostkowe). Warunki (5.2) zapewniają, że do produkcji każdego z wyrobów nie zużyjemy więcej środka produkcji niż mamy do dyspozycji. Interpretacja warunków nieujemności (5.3) jest oczywista.

Jeżeli przychód ma być liniową funkcją ilości wytwarzanych wyrobów, konieczne jest przyjęcie założenia o niezależności ceny zbytu od wielkości produkcji. Jest ono akceptowalne jedynie w przypadku producenta, którego wielkość sprzedaży nie wywiera wpływu na cenę wyrobu. Liniowa funkcja celu opisuje więc przychód producenta w przypadku konkurencji doskonałej. Jeżeli cena zbytu zależy od wielkości sprzedaży, jak to ma miejsce w przypadku oligopolu lub monopolu, przychód jest nieliniową funkcją wielkości produkcji.

Niech cena zbytu będzie malejącą funkcją wielkości sprzedaży postaci:

$$c_j = k_j + \frac{a_j}{\sqrt{x_j}}, \quad a_j > 0, \quad k_j > 0, \quad j = 1, \dots, n. \quad (5.4)$$

<sup>1</sup> W rozdziale I została przedstawiona najprostsza liniowa wersja problemu. Dla zachowania ciągłości rozważań przytaczamy jego liniową wersję.

Wówczas przychód opisuje funkcja:

$$\sum_{j=1}^n \left( k_j + \frac{a_j}{\sqrt{x_j}} \right) x_j = \sum_{j=1}^n (k_j x_j + a_j \sqrt{x_j}) \quad (5.5)$$

nieliniowa względem wielkości sprzedaży.

Jeżeli zysk całkowity dany jest przez wzór (5.1), to zysk jednostkowy nie jest zależny od wielkości produkcji, co pozostaje w sprzeczności zarówno z teorią mikroekonomii, jak i wynikami badań empirycznych, gdyż implikuje niewystępowanie efektu korzyści skali, sugerując, że zysk można osiągnąć przy dowolnej skali produkcji. Tymczasem problem możliwości osiągnięcia przez przedsiębiorstwo wielkości produkcji pozwalającej na osiąganie zysku stanowi jedną z istotnych barier wejścia na rynek, a korzyści skali są ważnym czynnikiem sprzyjającym monopolizacji.

Liniowość warunków ograniczających implikuje, że nakłady środków produkcji są proporcjonalne do wielkości produkcji, co oznacza stałość nakładów jednostkowych niezależnie od skali produkcji. W wielu branżach założenie to jest wysoce nierealistyczne. Jego uchylenie prowadzi do nieliniowych ograniczeń.

Reasumując, paradygmat liniowości zakłada, że:

- 1) nakłady oraz wyniki są proporcjonalne do stopnia aktywności gospodarczej,
- 2) zmienne mogą przyjmować dowolne wartości (są ciągłe)<sup>2</sup>,
- 3) interakcje między zmiennymi są **addytywne**<sup>3</sup>.

Ostatnie z założeń oznacza niemożność modelowania efektów substytucji i komplementarności<sup>4</sup>.

<sup>2</sup> Przyjęcie założenia, że zmienne mogą przyjmować tylko wartości całkowite, prowadzi do problemu programowania liniowego w liczbach całkowitych (zob. podrozdział 8.8).

<sup>3</sup> Oznacza to, że np. w problemie (5.1)–(5.3) zmiany zapotrzebowania na środki produkcji i zmiany przychodu zależą liniowo od zmian wielkości produkcji poszczególnych wyrobów: zapotrzebowanie na środki produkcji (przychód) jest sumą wyjściowego zapotrzebowania (przychodu) i przyrostu.

<sup>4</sup> Efekt substytucji polega na możliwości wytworzenia danego produktu za pomocą różnych kombinacji środków produkcji, np. robocizny i surowca. W przypadku problemu PL jednostkowe nakłady środków produkcji są stałe, co wyklucza możliwość modelowania efektu substytucji, gdyż dany produkt jest wytwarzany za pomocą jednej kombinacji środków produkcji określonej przez nakłady jednostkowe.

Efekt komplementarności polega na tym, że jeden produkt można zastąpić drugim: istnieje wiele koszyków dóbr (wyrobów) o jednakowej użyteczności (przynoszących identyczny przychód lub zysk). W przypadku problemu (5.1)–(5.3) istnieje tylko jedna kombinacja wielkości produkcji maksymalizująca przychód (pomijamy przypadek istnienia wielu rozwiązań bazowych).

## 5.2. Własności problemów programowania nieliniowego

Dana jest przestrzeń  $R^n$ , zbiór  $D \subseteq R^n$  oraz funkcja  $f: D \rightarrow R$ . **Problem optymalizacji** polega na znalezieniu takiego punktu  $x^* \in D$  (jeżeli taki punkt istnieje), dla którego  $f(x^*) \leq f(x)$  dla każdego  $x \in D$ . **Zbiór rozwiązań dopuszczalnych**  $D$  jest zbiorem punktów spełniających układ nierówności i równań:  $g_i(x) \leq 0, i = 1, \dots, m_1, h_i(x) = 0, i = 1, \dots, m_2$ . Jeżeli warunki w postaci równań i nierówności nie występują, to problem nosi nazwę **problemu bez ograniczeń (problemu optymalizacji bezwarunkowej)**, a  $D = R^n$ . W przeciwnym przypadku określa się go mianem **problemu optymalizacji z ograniczeniami (problemu optymalizacji warunkowej)**. Jeżeli jedna z funkcji  $f, g_i, i = 1, \dots, m_1, h_i, i = 1, \dots, m_2$ , jest funkcją nieliniową, to problem optymalizacji nazywamy problemem **optymalizacji nieliniowej**. Będziemy zakładać, że funkcja  $f$  oraz funkcje  $g_i, i = 1, \dots, m_1, h_i, i = 1, \dots, m_2$ , są funkcjami ciągłymi.

Niech  $\bar{x} \in D$ . Jeżeli istnieje  $\epsilon$ -otoczenie  $U(\bar{x})$  punktu  $\bar{x}$ , takie, że dla każdego  $x \in U(\bar{x})$  zachodzi nierówność  $f(x) \geq f(\bar{x})$ , to  $\bar{x}$  nazywamy **minimum lokalnym**. Jeżeli dla każdego  $x \in U(\bar{x}), f(x) > f(\bar{x})$ , to mówimy o **właściwym minimum lokalnym**. Minimum lokalne nie może być punktem brzegowym zbioru  $D$ , gdyż w tym przypadku nie istnieje otoczenie  $U(\bar{x})$  punktu  $\bar{x} \in D$  całkowicie zawarte w zbiorze  $D$ . Jeżeli  $\bar{x}$  jest punktem brzegowym, dla którego jest spełniony warunek  $f(x) \geq f(\bar{x})$  dla każdego  $x \in D$ , to mówimy, że w punkcie  $\bar{x}$  funkcja osiąga **kres dolny**. Jeżeli dla każdego  $x \in D$  zachodzi nierówność  $f(x) \geq f(\bar{x})$ , powiemy, że w punkcie  $\bar{x}$  funkcja ma **niewłaściwe minimum globalne**. W przypadku zachodzenia ostrej nierówności  $f(x) > f(\bar{x})$ , powiemy, że funkcja ma **właściwe minimum globalne**. Gdy funkcja  $f$  ma w zbiorze  $D$  wiele minimów lokalnych, wówczas pojawia się problem wyznaczenia minimum globalnego, co nie jest zagadnieniem trywialnym.

Problem minimalizacji bezwarunkowej sprowadza się do poszukiwania minimalnej wartości funkcji o wartościach rzeczywistych, tj. do problemu<sup>5</sup>:

$$f(x) \rightarrow \min, \quad x \in R^n. \quad (5.6)$$

Zagadnienie poszukiwania wartości minimalnej lub maksymalnej danej funkcji na danym zbiorze nie zawsze jest dobrze postawione. Kwestię tę w niektórych przypadkach rozstrzyga następujące twierdzenie.

<sup>5</sup> Kierunek optymalizacji nie jest żadnym ograniczeniem, gdyż maksymalizacja funkcji  $f$  jest równoznaczna z minimalizacją funkcji  $-f$ . Dlatego na początku podrzdziału mówi się o problemie optymalizacji.

### Weierstrassa o istnieniu

Jeżeli  $f$  jest ciągła, a zbiór  $D$  zwarty (domknięty i ograniczony), to funkcja  $f$  osiąga wartość minimalną i maksymalną w pewnym punkcie tego zbioru<sup>6</sup>.

Problem programowania nieliniowego z ograniczeniami (problem optymalizacji warunkowej) przybiera postać:

$$f(x) \rightarrow \min, \quad (5.7)$$

$$g_i(x) \leq 0, \quad i = 1, \dots, m_1, \quad (5.8)$$

$$h_i(x) = 0, \quad i = 1, \dots, m_2. \quad (5.9)$$

Zagadnienie polega na wyznaczeniu minimalnej wartości funkcji  $f$  (kresu dolnego lub minimum lokalnego, w którym funkcja osiąga najmniejszą wartość) na zbiorze danym przez (5.8)–(5.9). Rozsądnie jest przyjąć, że kres ten istnieje. Istotnie, jeżeli rozpatrywany problem stanowi model zagadnienia ekonomicznego, to — jeżeli został on właściwie skonstruowany — funkcja kryterium powinna na zbiorze rozwiązań dopuszczalnych przyjmować skończone wartości. Występowanie nieostrych nierówności zapewnia wówczas, że funkcja osiąga kres dolny.

Będziemy dalej zakładać, że zbiór rozwiązań dopuszczalnych problemu (5.7)–(5.9) jest niepusty, tj. że problem jest niesprzeczny. Założymy również, że zbiór ten jest ograniczony. Jeżeli konstruując model sytuacji decyzyjnej otrzymujemy problem sprzeczny, świadczy to zazwyczaj o próbie uwzględnienia zbyt wielu ograniczeń. Nieograniczoność zbioru rozwiązań dopuszczalnych sugeruje nieuwzględnienie w modelu istotnych ograniczeń. Nieograniczoność zbioru rozwiązań dopuszczalnych może (choć nie musi) prowadzić do braku skończonego ekstremum. Sytuacja taka jest dowodem błędów popełnionych w procesie modelowania.

W procesie modelowania istotne jest dokonanie sensownego kompromisu między wiernością opisu przez model sytuacji decyzyjnej (nadmierna wierność prowadzi do konstrukcji modeli o bardzo dużej złożoności) a możliwościami efektywnego wyznaczenia rozwiązań optymalnych (co skłania do budowy modeli niezbyt skomplikowanych). Warunkiem łatwego wyznaczenia rozwiązania optymalnego jest konstrukcja problemu, dla którego opracowano efektywne algorytmy uzyskiwania rozwiązań optymalnych zaimplementowane w powszechnie do-

<sup>6</sup> Funkcja  $f$  wartość minimalną może osiągać w punkcie wewnętrznym lub brzegowym zbioru  $D$ . W pierwszym przypadku, jeżeli funkcja jest różniczkowalna, wartość pochodnych cząstkowych jest w tym punkcie równa zero. W punkcie brzegowym pochodna nie jest określona, aczkolwiek właśnie w tym punkcie funkcja może osiągać najmniejszą wartość. Przykład stanowi wypukła, malejąca funkcja jednej zmiennej określona w przedziale  $[a, b]$ . Chociaż w punkcie  $b$  pochodna jest nieokreślona, właśnie tu funkcja osiąga minimum.

stępnym pakietach obliczeniowych<sup>7</sup>. Celem rozdziału jest zapoznanie Czytelnika z ideami najczęściej wykorzystywanych algorytmów i zwrócenie uwagi na pułapki czyhające na użytkowników powszechnie dostępnych pakietów optymalizacyjnych.

W dalszych rozważaniach ograniczymy się zasadniczo do zagadnienia szukania minimum funkcji o określonych właściwościach. Niech  $D \subset R^n$  będzie zbiorem wypukłym. Funkcję o wartościach rzeczywistych nazywamy **funkcją wypukłą** na zbiorze  $D$ , jeżeli dla dowolnych  $x^1, x^2 \in D$  i dowolnego  $\alpha \in [0, 1]$  zachodzi nierówność:

$$f[\alpha \cdot x^1 + (1 - \alpha) \cdot x^2] \leq \alpha \cdot f(x^1) + (1 - \alpha) \cdot f(x^2). \quad (5.10)$$

Funkcja o wartościach rzeczywistych jest nazywana **ściśle wypukłą**<sup>8</sup> w zbiorze  $D$ , jeżeli dla dowolnych  $x^1 \neq x^2, x^1, x^2 \in D$  i dla dowolnego  $\alpha \in (0, 1)$  zachodzi nierówność<sup>9</sup>:

$$f[\alpha \cdot x^1 + (1 - \alpha) \cdot x^2] < \alpha \cdot f(x^1) + (1 - \alpha) \cdot f(x^2). \quad (5.11)$$

Na rysunku 5.1A przedstawiono warstwicę niewypukłej, a na rysunku 5.1B wypukłej funkcji dwóch zmiennych.

Funkcja  $f$  jest **funkcją wklęsłą**, jeżeli  $-f$  jest funkcją wypukłą (i odwrotnie). Funkcję  $f$  nazywamy **ściśle wklęsłą**, jeżeli  $-f$  jest funkcją ściśle wypukłą (i odwrotnie). Funkcja liniowa jest zarówno funkcją wypukłą, jak i wklęsłą. Badanie wklęsłości funkcji  $f$  sprowadza się do badania wypukłości funkcji  $-f$ .

Przy badaniu wypukłości funkcji bardzo użyteczne jest założenie o jej różniczkowalności.

<sup>7</sup> Poszukiwanie rozwiązania optymalnego nie zawsze jest celowe, zwłaszcza gdy wymaga poniesienia dużych nakładów obliczeniowych. W większości przypadków spotykanych w praktyce podejmowania decyzji wartości parametrów modelu są znane tylko z pewnym przybliżeniem, co wynika z niemożności (pracochłonności) wyznaczenia ich dokładnych wartości lub działania w warunkach ryzyka czy też niepewności (zob. rozdziały 9 i 11). Dlatego często zadawaliśmy się wyznaczeniem rozwiązania przybliżonego (suboptymalnego). Rozwiązanie takie można otrzymać, stosując metody heurystyczne (zob. rozdział 15), lub modyfikując metody dokładne. Wyznaczenie rozwiązania suboptymalnego za pomocą odpowiednio zmodyfikowanych metod dokładnych pozwala na wydajne zmniejszenie nakładów obliczeniowych w porównaniu z nakładami niezbędnymi do wyznaczenia rozwiązania optymalnego.

<sup>8</sup> Pojęcie wypukłości funkcji ma sens, jeżeli jej dziedzina  $D$  jest zbiorem wypukłym. W przeciwnym przypadku mogą istnieć takie  $x^1, x^2 \in D$  i  $\alpha \in (0, 1)$ , że istnieje punkt  $\alpha x^1 + (1 - \alpha)x^2 \notin D$ , w którym  $f[\alpha x^1 + (1 - \alpha)x^2]$  nie jest określona. Przypominamy, że zbiór  $D$  jest zbiorem wypukłym, jeżeli dla dowolnych  $x^1, x^2 \in D$   $\alpha x^1 + (1 - \alpha)x^2 \in D, 0 < \alpha < 1$ . Oznacza to, że dowolne dwa punkty należące do zbioru można połączyć odcinkiem, który w całości zawiera się w tym zbiorze.

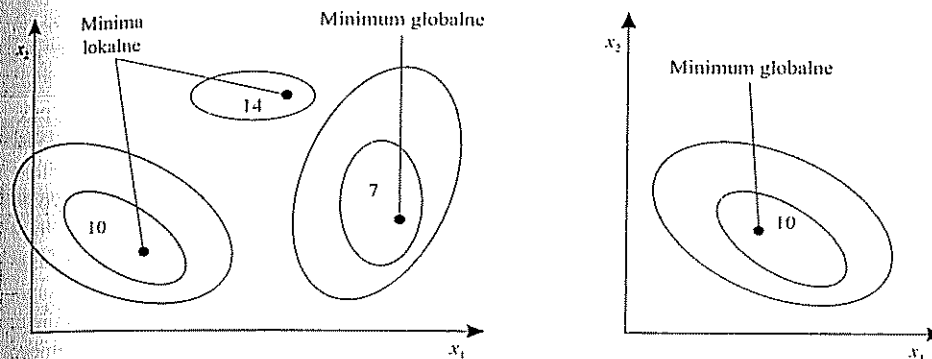
<sup>9</sup> W przypadku wypukłej funkcji jednej zmiennej wykres takiej funkcji nie jest położony powyżej odcinka łączącego  $x^1$  z  $x^2$ . W przypadku funkcji ściśle wypukłej wykres ten jest położony poniżej odcinka.

Załóżmy, że funkcja  $f$  jest różniczkowalna w  $R^n$ , natomiast wektor  $\nabla f(x^0) = \left[ \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right]^T$  jest nazywany **gradientem funkcji  $f$**  w punkcie  $x^0$ . Gradient wskazuje na kierunek, w którym przyrost wartości funkcji jest najszybszy.

RYSUNEK 5.1

Wykresy funkcji niewypukłej i wypukłej dwóch zmiennych

- A. Funkcja ma trzy minima, w tym jedno globalne. Nie jest to funkcja wypukła. Warstwicami są zbiorymi punktów, w których funkcja osiąga tę samą wartość
- B. Funkcja ma jedno minimum lokalne będące minimum globalnym. Funkcja ta jest funkcją wypukłą



Załóżmy, że funkcja  $f$  jest dwukrotnie różniczkowalna w  $R^n$ . Macierz:

$$\nabla^2 f(x^0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix}$$

nazywa się **hesjanem funkcji**.

Niech  $f$  będzie dwukrotnie różniczkowalna w  $R^n$ . Jedną z metod pozwalających na stwierdzenie wypukłości funkcji jest zbadanie określoności jej hesjanu. Funkcja jest wypukła w otwartym (tj. niezawierającym punktów brzegowych) zbiorze  $D$  wtedy i tylko wtedy, gdy dla każdego  $x \in D$  jej hesjan jest nieujemnie określony.

Jeżeli dla każdego  $x \in D$  hesjan  $\nabla^2 f(x)$  funkcji  $f$  jest dodatnio określony, to funkcja ta jest ściśle wypukła. Założenie o otwartości zbioru jest niezbędne, gdyż hesjan nie jest określony w punkcie brzegowym zbioru.

### 5.3. Klasyfikacja problemów programowania nieliniowego

Proponuje się wiele klasyfikacji problemów programowania nieliniowego. Będziemy wyróżniać dwa podstawowe typy problemów programowania nieliniowego, tj. problemy programowania wypukłego i problemy programowania niewypukłego. Szczególnym przypadkiem problemu programowania wypukłego, do którego sprowadzić można wiele zagadnień z zakresu analizy decyzyjnej, jest problem programowania kwadratowego<sup>10</sup>.

**Problem programowania wypukłego** polega na minimalizacji funkcji wypukłej (maksymalizacji funkcji wklęsłej) na zbiorze wypukłym. Niech  $f, g_i, i = 1, \dots, m_1$  oraz  $h_i, i = 1, \dots, m_2$  będą funkcjami wypukłymi. Dowodzi się, że zbiory określone przez nierówność  $g_i(x) \leq 0$  oraz równanie  $h_i(x) = 0$  są wówczas zbiorami wypukłymi. Ponieważ część wspólna zbiorów wypukłych jest zbiorem wypukłym, zbiór dany przez ograniczenia (5.8)–(5.9) jest wypukły. Założenie o wypukłości funkcji  $f$  jest bardzo użyteczne, gdyż ułatwia wyznaczenie minimum globalnego. Pomimo, że minimum lokalne wypukłej funkcji  $f$  na zbiorze wypukłym  $D$  (jeżeli istnieje) jest w tym zbiorze jej minimum globalnym, wyznaczenie minimum globalnego sprowadza się do znalezienia minimum lokalnego lub (jeżeli minimum takie nie istnieje) kresu dolnego.

Algorytmy poszukiwania minimum funkcji ciągłych (ewentualnie mających ciągle pierwsze i drugie pochodne) pozwalają jedynie na wyznaczenie minimum lokalnego. Jeżeli  $f$  jest funkcją wypukłą, to problem wyznaczenia minimum globalnego redukuje się do o wiele prostszego problemu wyznaczenia minimum lokalnego.

Wartość minimalna funkcji może być osiągana w punkcie brzegowym (w szczególności w wierzchołku zbioru rozwiązań dopuszczalnych, jeżeli ograniczenia są liniowe) lub wewnętrznym. W przypadku problemu PL wartość ta jest osiągana w punkcie wierzchołkowym. Jeżeli zbiór  $D$  dany jest przez ograniczenia liniowe, ale funkcja celu nie jest funkcją liniową, to wartość minimalna nie musi być osiągana w punkcie wierzchołkowym (w ogólności punkcie brzegowym).

Szczególnym przypadkiem problemu programowania wypukłego jest problem

<sup>10</sup> Podział ten wynika ze specyfiki metod wyznaczania minimum globalnego. Związły przedmiot metod rozwiązywania problemów programowania niewypukłego można znaleźć np. w [Grabowski 1980].

**programowania kwadratowego.** Zakłada się, że funkcja  $f$  jest wypukłą funkcją kwadratową, a funkcje  $g_i, i = 1, \dots, m_1$  są funkcjami liniowymi, a więc wypukłymi, co implikuje wypukłość zbioru  $D$ . Zagadnienie sprowadza się do poszukiwania minimum wypukłej funkcji kwadratowej na zbiorze wypukłym.

Problem programowania kwadratowego przyjmuje postać:

$$cx + \frac{1}{2}x^T Ex \rightarrow \min, \quad (5.12)$$

$$Ax \leq b, \quad (5.13)$$

$$x \geq 0, \quad (5.14)$$

gdzie  $c = [c_1, \dots, c_n]$ ,  $b = [b_1, \dots, b_m]^T$ ,  $A = [a_{ij}]_{n \times m}$ ,  $E = [e_{ij}]_{n \times n}$ ,  $x = [x_1, \dots, x_n]^T$ .  $E$  jest macierzą określoną nieujemnie<sup>11</sup>, co gwarantuje wypukłość składnika kwadratowego funkcji celu<sup>12</sup>. Ponieważ funkcja liniowa jest wypukła, a suma funkcji wypukłych jest funkcją wypukłą, to (5.12) jest funkcją wypukłą.

Problem programowania kwadratowego ma wiele ważnych zastosowań praktycznych. Wiele zagadnień decyzyjnych związanych z działalnością przedsiębiorstwa, np. zagadnienie załadunku z kwadratową funkcją kosztu, nieliniowe zagadnienie pośrednika, są zagadnieniami programowania kwadratowego. Do problemu programowania kwadratowego sprowadza się wiele zagadnień z dziedziny optymalizacji działalności finansowej, w szczególności analizy portfelowej, które zostały przedstawione w rozdziale 13.

Stwierdzenie, czy funkcja celu jest wypukła, może niekiedy być kłopotliwe<sup>13</sup>. Często jednak wypukłość funkcji gwarantują założenia przyjmowane przy konstrukcji modelu matematycznego sytuacji decyzyjnej. W wielu przypadkach zasadne jest przyjęcie założenia, że funkcje kosztu oraz zysku jednostkowego są U-kształtne<sup>14</sup>. Niekiedy wypukłość funkcji wynika z własności statystycznych

<sup>11</sup> Macierz kwadratowa  $A_{n \times n}$  jest określona dodatnio wtedy i tylko wtedy, gdy  $M_1 > 0, M_2 > 0, \dots, M_n > 0$ , a określona ujemnie wtedy i tylko wtedy, gdy  $M_1 < 0, M_2 > 0, M_3 < 0, \dots$ , gdzie:

$$M_1 = a_{11}, \quad M_r = \det \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \dots & a_{rr} \end{bmatrix}, \quad r = 2, \dots, n,$$

gdzie  $a_{ij}$  oznacza wyznacznik, a  $M_r$  — minory główne macierzy  $A$ .

<sup>12</sup> Jeżeli macierz  $E$  jest dodatnio określona, to  $x^T Ex$  jest funkcją ściśle wypukłą.

<sup>13</sup> Jedną z metod pozwalających na stwierdzenie wypukłości funkcji jest zbadanie określoności jej hesjanu.

<sup>14</sup> Funkcją U-kształtną nazywamy funkcję wypukłą jednej zmiennej określoną na pewnym przedziale, osiągającą minimum wewnątrz tego przedziału. Nazwa pochodzi od kształtu wykresu funkcji. Suma funkcji wypukłych (w szczególności U-kształtnych) jest funkcją wypukłą. Z tego faktu często korzysta się, konstruując funkcję celu wielu zagadnień optymalizacyjnych.

modelowanych procesów. Na przykład macierz kowariancji (zob. rozdział 13) nieujemnie określona, co gwarantuje wypukłość funkcji kwantyfikującej ryzyko mierzone wariancją.

## 5.4. Problem Kuhna-Tuckera<sup>15</sup>

Rozważmy następujący problem programowania nieliniowego:

$$f(x) \rightarrow \min, \quad (5.15)$$

$$g_i(x) \leq 0, \quad i = 1, \dots, m_1, \quad (5.16)$$

$$h_i(x) = 0, \quad i = 1, \dots, m_2, \quad (5.17)$$

gdzie  $f, g_i(x), i = 1, \dots, m_1, h_i(x), i = 1, \dots, m_2$  są określone w  $R^n$ .

Nie są znane metody analityczne (zob. podrozdział 5.5) wyznaczania ekstremum funkcji wielu zmiennych przy warunkach ograniczających. Zamiast rozwiązywać problem programowania nieliniowego z warunkami ograniczającymi można wyznaczyć rozwiązanie optymalne równoważnego problemu niezawierającego tych warunków. Problem pomocniczy konstruuje się w ten sposób, że jego rozwiązanie optymalne (jeżeli jest rozwiązaniem dopuszczalnym problemu wyjściowego) wyznacza rozwiązanie optymalne problemu wyjściowego. Problem pomocniczy otrzymuje się, tworząc tzw. funkcję Lagrange'a.

Rozważmy problem:

$$f(x) \rightarrow \min, \quad (5.18)$$

$$h_i(x) = 0, \quad i = 1, \dots, m_2. \quad (5.19)$$

Załóżmy, że  $f$  i  $h_i, i = 1, \dots, m_2$  są różniczkowalne w  $R^n$ . Funkcję:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m_2} \lambda_i h_i(x) \quad (5.20)$$

nazywa się funkcją Lagrange'a.

W każdym punkcie  $x \in D$ , tj. spełniającym (5.19),  $L(x, \lambda) = f(x)$ . Wyznaczenie minimum globalnego funkcji należącego do  $D$  jest równoznaczne z wyznaczeniem minimum (globalnego) funkcji  $f$ . Zagadnienie minimalizacji (5.20) nazywa się **problemem Lagrange'a**, a  $\lambda_i$  noszą nazwę **mnożników Lagrange'a**.

<sup>15</sup> W literaturze problem ten określa się niekiedy mianem problemu Karush-Kuhna-Tuckera.

Jeżeli funkcja Lagrange'a osiąga minimum w punkcie  $\bar{x}$ , to<sup>16</sup>:

$$\nabla f(\bar{x}) + \sum_{i=1}^{m_2} \bar{\lambda}_i \nabla h_i(\bar{x}) = 0. \quad (5.21)$$

Warunek (5.21) jest warunkiem koniecznym (ale nie dostatecznym) istnienia minimum lokalnego funkcji  $L$ . Jeżeli  $\bar{x}$  spełnia jednocześnie (5.19), to w tym punkcie także funkcja  $f$  może osiągać minimum lokalne. Rozwiązanie optymalne  $(x^*, \lambda^*)$ ,  $x^* \in D$  problemu (5.20) wyznacza rozwiązanie optymalne  $x^*$  problemu (5.18)–(5.19). Wyznaczywszy punkty, dla których zachodzi (5.21), należy zbadać w którym z nich funkcja  $f$  osiąga minimum.

### PRZYKŁAD 5.1

Rozważmy zadanie<sup>17</sup>:

$$f(x_1, x_2) = 2x_1^2 + 3x_2^2 \rightarrow \min,$$

$$h_i(x_1, x_2) = x_1 + x_2 - 1 = 0.$$

Funkcja Lagrange'a przyjmuje postać:

$$L(x, \lambda) = f(x) + \lambda h(x) = 2x_1^2 + 3x_2^2 + \lambda(x_1 + x_2 - 1).$$

W celu wyznaczenia punktu, w którym może znajdować się minimum, wyznaczamy pierwsze pochodne i przyrównujemy je do 0. Aby otrzymać rozwiązanie utworzonego układu równań, dodajemy równanie  $h(x_1, x_2) = x_1 + x_2 - 1 = 0$ .

$$\frac{\partial L(x, \lambda)}{\partial x_1} = 4x_1 + \lambda = 0,$$

$$\frac{\partial L(x, \lambda)}{\partial x_2} = 6x_2 + \lambda = 0,$$

$$h(x_1, x_2) = x_1 + x_2 - 1 = 0.$$

Rozwiązaniem układu równań jest  $\bar{x}_1 = 0,6, \bar{x}_2 = 0,4, \bar{\lambda} = -2,4$ .

<sup>16</sup> Przez  $\bar{x}$  oznaczamy punkt, w którym potencjalnie znajduje się minimum (spełniony jest warunek konieczny istnienia minimum), a przez  $x^*$  — punkt, w którym minimum (globalne) jest osiągane (spełniony jest warunek dostateczny).

Wątpliwość może budzić sytuacja, gdy minimum jest osiągane w punkcie brzegowym zbioru danego przez (5.19); pochodna funkcji  $f$  w tym punkcie nie istnieje. Niemniej jednak w przypadku minimalizacji (5.20) zakłada się różniczkowalność wszystkich funkcji w  $R^n$  i przy tym założeniu wyznacza się minimum  $L$ .

<sup>17</sup> Będziemy rozróżniać problem od zadania. O zadaniu mówimy, gdy parametrom zostały nadane konkretne wartości. Zadanie jest szczególnym przypadkiem (instancją) problemu.

Hesjan<sup>18</sup>:

$$\nabla^2 L(x, \lambda) = \begin{bmatrix} 4 & 0 \\ 0 & 6 \end{bmatrix}$$

jest dodatnio określony, gdyż:

$$|4| = 4 > 0, \quad \begin{vmatrix} 4 & 0 \\ 0 & 6 \end{vmatrix} = 4 \cdot 6 - 0 \cdot 0 = 24 > 0,$$

co oznacza, że w punkcie  $[0,6 \ 0,4]^T$  funkcja  $L(x, \lambda)$ , a więc i  $f(x)$ , istotnie osiąga minimum — z uwagi na wypukłość funkcji — globalne.

Wyznaczenie punktu, w którym może znajdować się minimum, wymagało rozwiązania układu równań liniowych. W ogólnym przypadku (tj. dla dowolnych funkcji  $f$  i  $h_i$ ) otrzymujemy układ równań nieliniowych, którego rozwiązanie może być bardzo trudne, co implikuje małą praktyczną przydatność wyznaczania minimum funkcji w sposób analityczny.

Rozważmy obecnie przypadek bardziej ogólny, kiedy ograniczenia mogą być także nierównościami.

Dla dowolnego  $x \in D$  zdefiniujmy zbiór aktywnych ograniczeń nierównościami  $A(x) = \{i : g_i(x) = 0\}$  oraz zbiór kierunków  $D(x) = \{d \in R^n : \nabla g_i(x)^T d \leq 0, i \in A(x), \nabla h_i(x)^T d = 0, i = 1, \dots, m_2\}$ .

Punkt  $x \in D$  nazywamy **regularnym**, jeżeli dla dowolnego kierunku  $d \in D(x)$  istnieje krzywa gładka zaczynająca się w punkcie  $x$ , styczna w tym punkcie do  $D$ , której pewna początkowa część zawiera się w  $D$ .

Niżej podajemy warunki służące do sprawdzenia regularności.

Warunki regularności:

1. **Warunek Karlina.** Jeżeli  $g_i, i = 1, \dots, m_1$  są wypukłe,  $h_i, i = 1, \dots, m_2$  są liniowe, to każdy punkt  $x \in D$  jest punktem regularnym.
2. **Warunek Slatera.** Jeżeli funkcje  $g_i, i = 1, \dots, m_1$  są wypukłe,  $h_i, i = 1, \dots, m_2$  są liniowe oraz istnieje punkt  $x$  taki, że  $g_i(x) < 0, i = 1, \dots, m_1$  i  $h_i(x) = 0, i = 1, \dots, m_2$ , to punkt  $x \in D$  jest punktem regularnym.
3. **Warunek liniowej niezależności.** Jeżeli w punkcie  $x$  gradienty  $\nabla g_i(x), i \in A(x)$  oraz  $\nabla h_i(x), i = 1, \dots, m_2$  tworzą układ wektorów liniowo niezależnych, to punkt  $x$  jest punktem regularnym.

Niech funkcje  $f, g_i, i = 1, \dots, m_1, h_i, i = 1, \dots, m_2$  będą różniczkowalne w  $R^n$ .

<sup>18</sup> Badając określoność hesjanu pomijamy zmienne  $\lambda_r$ .

### Kuhna-Tuckera

Jeżeli  $\bar{x}$  jest minimum lokalnym problemu (5.15)–(5.17) oraz punktem regularnym, to istnieją liczby  $\bar{\lambda}_i, i = 1, \dots, m_1$  i  $\bar{\mu}_i, i = 1, \dots, m_2$ , takie że:

$$\nabla f(\bar{x}) + \sum_{i=1}^{m_1} \bar{\lambda}_i \nabla g_i(\bar{x}) + \sum_{i=1}^{m_2} \bar{\mu}_i \nabla h_i(\bar{x}) = 0, \quad (5.22)$$

$$\bar{\lambda}_i g_i(\bar{x}) = 0, \quad i = 1, \dots, m_1, \quad (5.23)$$

$$\bar{\lambda}_i \geq 0, \quad i = 1, \dots, m_1. \quad (5.24)$$

Warunki nieujemności nakłada się na mnożniki odpowiadające ograniczeniom nierównościowym. Warunki (5.22)–(5.24) noszą nazwę **warunków Kuhna-Tuckera**, krótko warunków K-T. Są to warunki konieczne. Pomijając warunki  $g_i(x) \leq 0, i = 1, \dots, m_1$ , otrzymujemy problem Lagrange'a<sup>20</sup>.

Kolejne twierdzenie określa warunki wystarczające **Kuhna-Tuckera**.

Jeżeli funkcja  $f$  jest wypukła, funkcje  $g_i, i = 1, \dots, m_1$  są wypukłe, a  $h_i, i = 1, \dots, m_2$  są liniowe<sup>21</sup>, to każdy punkt, w którym są spełnione warunki konieczne K-T jest minimum globalnym.

### PRZYKŁAD 5.2

Rozważmy zadanie:

$$f(x_1, x_2) = 2x_1^2 + x_2^2 \rightarrow \min,$$

$$g_1(x_1, x_2) = x_1 + x_2 - 10 \leq 0,$$

$$g_2(x_1, x_2) = x_1^2 + x_2^2 \leq 4,$$

$$h(x_1, x_2) = 3x_1 + x_2 - 1 = 0.$$

$$\nabla f(x) = \begin{bmatrix} 4x_1 \\ 2x_2 \end{bmatrix}, \quad \nabla g_1(x) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \nabla g_2(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}, \quad \nabla h(x) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

<sup>19</sup> Warunki konieczne i wystarczające Kuhna-Tuckera formułuje się przy różnych założeniach. Przytoczyliśmy najprostszą wersję twierdzenia, pomijając warunki konieczne i wystarczające drugiego rzędu.

<sup>20</sup> Zauważmy, że w punkcie  $\bar{x}$  spełniającym warunki konieczne K-T pochodne cząstkowe  $\nabla f(\bar{x})$  funkcji  $f$  nie ulegają wyzerowaniu.

<sup>21</sup> Jest to równoznaczne ze spełnieniem warunku Karlina.



Warunki K-T przyjmują postać:

$$\nabla f(\mathbf{x}) + \sum_{i=1}^{m_1} \lambda_i \nabla g_i(\mathbf{x}) + \sum_{i=1}^{m_2} \mu_i \nabla h_i(\mathbf{x}) = \begin{bmatrix} 4x_1 \\ 2x_2 \end{bmatrix} + \lambda_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} + \mu \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$$\lambda_1 g_1(\mathbf{x}) = \lambda_1 (x_1 + x_2 - 10) = 0,$$

$$\lambda_2 g_2(\mathbf{x}) = \lambda_2 (x_1^2 + x_2^2 - 4) = 0,$$

$$\lambda_i \geq 0, \quad i = 1, 2.$$

Otrzymaliśmy układ czterech równań z pięcioma niewiadomymi. W celu wyznaczenia rozwiązania tego układu dołączamy do niego równanie  $3x_1 + x_2 - 1 = 0$ . Jednym z rozwiązań jest  $\bar{x}_1 \approx 0,27$ ,  $\bar{x}_2 \approx 0,18$ ,  $\bar{\lambda}_1 = \bar{\lambda}_2 = 0$ ,  $\bar{\mu} \approx 0,36$ . W naszym przypadku funkcja celu jest funkcją wypukłą, zatem warunek konieczny istnienia minimum jest jednocześnie warunkiem dostatecznym, a minimum lokalne jest minimum globalnym, co bezpośrednio wynika z twierdzenia 5.3. Hessian — jak łatwo sprawdzić — jest dodatnio określony w każdym punkcie dopuszczalnym<sup>22</sup>.

Podobnie jak w przypadku problemu Lagrange'a, podstawowa trudność natury numerycznej uzyskania rozwiązania optymalnego na drodze analitycznej polega na wyznaczeniu rozwiązań układu równań nieliniowych<sup>23</sup>, co w praktyce uniemożliwia bezpośrednie zastosowanie warunków K-T do poszukiwania minimum funkcji nieliniowej przy warunkach ograniczających<sup>24</sup>. Natomiast łatwo jest sprawdzić, czy w danym punkcie  $\mathbf{x}$  są spełnione warunki K-T. Układ (5.22)–(5.23) redukuje się bowiem niezależnie od postaci analitycznej funkcji  $f$ ,  $g_i$ ,  $i = 1, \dots, m_1$ ,  $h_i$ ,  $i = 1, \dots, m_2$ , do układu równań liniowych; wielkości  $\nabla f(\mathbf{x})$ ,  $\nabla g_i(\mathbf{x})$ ,  $\nabla h_i(\mathbf{x})$  i  $g_i(\mathbf{x})$  są znane, niewiadomymi są  $\lambda_i$  oraz  $\mu_i$ . Należy jedynie wyznaczyć rozwiązania układu równań (5.22)–(5.23), traktując  $\lambda_i$  i  $\mu_i$  jako niewiadome i sprawdzić, czy spełniają one (5.24).

## 5.5. Algorytmy wyznaczania ekstremum lokalnego

Ekstremum funkcji (w przypadku optymalizacji bezwarunkowej i warunkowej) możemy wyznaczać, stosując **metody analityczne** bądź **numeryczne**. Przykładem pierwszej z metod jest wyznaczenie minimum funkcji kwadratowej. W większości

<sup>22</sup> W ogólnym przypadku punkt spełniający warunki K-T nie musi być minimum lokalnym.

<sup>23</sup> Układ równań (5.22)–(5.23) jest liniowy tylko wtedy, gdy funkcja  $f$  jest funkcją kwadratową a  $g_i$ ,  $i = 1, \dots, m_1$  i  $h_i$ ,  $i = 1, \dots, m_2$  są funkcjami liniowymi.

<sup>24</sup> Sformułowanie warunków K-T dla konkretnego typu problemu prowadzi zazwyczaj do zagadnienia stanowiącego punkt wyjścia do opracowania algorytmu wyznaczającego minimum lokalne problemu wyjściowego. Opis sposobu postępowania, ze względu na jego wysoce specjalistyczny charakter, pomijamy.

przypadków nie są znane metody wyznaczania ekstremum funkcji za pomocą wzorów określających w sposób jawny zależność między współrzędnymi punktu, w którym funkcja osiąga ekstremum, a wartościami parametrów (współczynników stojących przy zmiennych). W takim przypadku (spotykanym zazwyczaj w praktyce) korzystamy z metod numerycznych pozwalających w sposób iteracyjny wyznaczyć rozwiązanie optymalne lub rozwiązanie niezbyt od niego odległe. Idea postępowania sprowadza się do wyznaczenia **rozwiązania początkowego** i poprawy tego rozwiązania do momentu uzyskania rozwiązania optymalnego lub przybliżonego (suboptymalnego). Realizacja naszkicowanego postępowania wymaga wyznaczenia dopuszczalnego rozwiązania wyjściowego (**punktu startowego**), poprawy tego rozwiązania w kolejnych krokach i określenia tzw. **kryterium stopu**, które umożliwi stwierdzenie, czy aktualne rozwiązanie jest rozwiązaniem optymalnym lub wystarczająco bliskim rozwiązaniu optymalnemu<sup>25</sup>.

Poszukując optimum funkcji (zarówno w przypadku optymalizacji bezwarunkowej, jak i warunkowej), stajemy przed następującym dylematem: czy skorzystać z algorytmu pozwalającego na wyznaczenie ekstremum lokalnego szerokiej klasy problemów, ale być może mało efektywnego w konkretnym przypadku, czy też posłużyć się algorytmem opracowanym specjalnie w celu wyznaczenia optimum problemu należącego do pewnej wąskiej klasy, ale potencjalnie bardziej efektywnym, gdyż korzystającym z informacji zawartej w szczególnej strukturze problemu. Wybór nie jest wcale oczywisty i często przesądzają o nim względy praktyczne, tj. dostępność pakietu obliczeniowego realizującego możliwy do zastosowania algorytm. Z tego powodu wybór pada często na algorytm niewyspecjalizowany, umożliwiający wyznaczenie ekstremum szerokiej klasy problemów.

Opracowano wiele **metod wyspecjalizowanych** wyznaczających minima lokalne problemów szczególnie często spotykanych w praktyce. Do tego rodzaju zagadnień można zaliczyć przede wszystkim problem programowania kwadratowego. Algorytmy te są w większości przypadków zmodyfikowanymi wersjami metody simpleks [Grabowski, 1980, rozdz. 14].

**Metody niewyspecjalizowane** wymagają ciągłości (niekiedy różniczkowalności) funkcji kryterium i nie korzystają z informacji zawartej w szczególnej strukturze problemu. Są to klasyczne metody iteracyjne, pozwalające na uzyskanie ekstremum lokalnego z zadaną dokładnością zależną od nakładu obliczeniowego (ilości kroków), jaki jesteśmy gotowi zaakceptować.

Można dokonać następującego ich podziału [Wit, 1986]:

1) **metody bezgradientowe**: korzysta się wyłącznie z wartości funkcji (nie jest wymagana jej różniczkowalność);

<sup>25</sup> Przykładem algorytmu realizującego naszkicowaną ideę jest algorytm simpleks. Punktem startowym jest wyjściowe bazowe rozwiązanie dopuszczalne. Kryteria usuwania i wprowadzania zmiennych do bazy pozwalają w kolejnym kroku przejść do bazy nie gorszej od już uzyskanej, a kryteria simpleks umożliwiają sprawdzenie, czy uzyskane rozwiązanie jest optymalne. Pełnią zatem rolę kryterium stopu.

- 2) metody gradientowe: w trakcie obliczeń konieczne jest wyznaczanie gradientu funkcji;
- 3) inne metody: korzystają z wartości pochodnych wyższego rzędu.

Wybór metody zależy od właściwości funkcji i doświadczeń numerycznych określających jej efektywność w przypadku wyznaczania optimum problemów danej klasy. Na prostych przykładach dokonamy ilustracji sposobu działania wybranych metod. Omawiane procedury obliczeniowe zostały zaimplementowane w powszechnie dostępnych pakietach obliczeniowych.

### 5.5.1. Wybrane metody gradientowe minimalizacji w problemach programowania nieliniowego bez ograniczeń

Rozważmy problem optymalizacji bezwarunkowej:

$$f(\mathbf{x}) \rightarrow \min. \quad (5.25)$$

Będziemy zakładać, że  $f$  jest funkcją ciągłą. Istnieje wiele metod wyznaczania minimum lokalnego funkcji  $f$  z dowolną dokładnością. Postępowanie sprowadza się do iteracyjnego przybliżania się do optimum lokalnego. Po dokonaniu wyboru punktu startowego, staramy się w każdym kroku przybliżyć do minimum, przysuwając się w obranym zgodnie z pewnymi regułami kierunku, nazywanym kierunkiem poprawy. Innymi słowy, konstruujemy sekwencję punktów  $\{\mathbf{x}^k\} \subset R^n$  w następujący sposób:

- 1) w punkcie  $\mathbf{x}^k$  wybieramy kierunek  $\mathbf{d}^k \in R^n$ ;
- 2) określamy wartość współczynnika  $t_k$  regulującego długość kroku; nadaje mu się taką wartość, aby spełniona była nierówność:

$$f(\mathbf{x}^k + t_k \mathbf{d}^k) < f(\mathbf{x}^k), \quad (5.26)$$

a następnie przechodzi się do kolejnego punktu:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \mathbf{d}^k. \quad (5.27)$$

Wektor  $\mathbf{d}^k$  nazywamy **kierunkiem spadku** (poprawy) funkcji  $f$  w punkcie  $\mathbf{x}^k$  jeżeli istnieje takie  $t_k > 0$ , że dla każdego  $t_k \in (0, \bar{t}_k]$   $f(\mathbf{x}^k + t_k \mathbf{d}^k) < f(\mathbf{x}^k)$ . Problem sprowadza się do wyboru takiego  $t_k$ , dla którego spadek wartości funkcji jest maksymalny. Nadanie  $t_k$  niewłaściwej wartości prowadzi do sytuacji przedstawionych na rysunku 5.2. Jeżeli wektor  $\mathbf{d}$  ma niewłaściwą długość, to z punktu  $\mathbf{x}^k$  przechodzimy do  $\mathbf{x}^2$  lub  $\mathbf{x}^3$  zamiast do  $\mathbf{x}^1$ .

Reasumując, postępowanie można ująć w postaci następującej sekwencji kroków:

- Krok 1. Wyznacz punkt startowy  $\mathbf{x}^0$  i podstaw  $k := 0$ .
- Krok 2. Sprawdź, czy w punkcie  $\mathbf{x}^k$  funkcja osiąga minimum. Jeżeli tak, stop. W przeciwnym przypadku przejdź do kroku 3.
- Krok 3. Wyznacz  $\mathbf{d}^k$  i  $t_k$ , a następnie przejdź do punktu  $\mathbf{x}^{k+1} := \mathbf{x}^k + t_k \mathbf{d}^k$ . Podstaw  $k := k + 1$  i wróć do kroku 2.

Wyznaczanie minimum funkcji wielu zmiennych redukuje się do sekwencyjnego rozwiązywania problemów z jedną zmienną:

$$f(\mathbf{x}^k + t_k \mathbf{d}^k) \rightarrow \min, \\ 0 < t_k < \bar{t}_k,$$

gdzie  $\bar{t}_k$  oznacza maksymalną wartość  $t_k$ ,  $\mathbf{x}^k$ ,  $\mathbf{d}^k$  są dane (tj. są one parametrami), a tylko  $t_k$  jest zmienną<sup>26</sup>.

Stosując pakiety obliczeniowe niezwykle rzadko udaje się dokładnie wyznaczyć wartość minimum ze względu na dokonywanie obliczeń w arytmetyce o skończonej dokładności oraz kumulację błędów<sup>27</sup>. Dlatego proponuje się różne kryteria stopu. Zakończenie postępowania może następować np. kiedy w kolejnych iteracjach poprawa wartości funkcji jest mniejsza od zadanej. Często w pakietach obliczeniowych określa się maksymalną liczbę iteracji, po której algorytm ulega zakończeniu, a którą użytkownik, w celu uzyskania rozwiązania bliższego optimum, może zmienić.

Sposób wyznaczania  $\mathbf{d}^k$  zależy od własności funkcji. Jeżeli funkcja jest różniczkowalna, to w celu wyznaczania kierunku spadku korzysta się najczęściej z gradientu. Jeżeli funkcja jest tylko ciągła, trzeba uciec się do innych metod<sup>28</sup>.

<sup>26</sup> Tylko w szczególnych przypadkach udaje się w sposób analityczny wyznaczyć minimum funkcji jednej zmiennej. Istnieje wiele prostych i uniwersalnych metod wyznaczania minimum takiej funkcji, np. przeszukiwanie dychotomiczne, metoda złotego podziału, przeszukiwanie z użyciem współczynników Fibonacciego oraz metoda bisekcji i Newtona. Wymaga się jedynie, aby w dostatecznie dużym otoczeniu minimum funkcja była wypukła — niekoniecznie różniczkowalna (tylko ostatnia z wymienionych metod wymaga różniczkowalności funkcji). Opis algorytmów wyznaczania minimum można znaleźć w każdym podręczniku metod optymalizacji nieliniowej.

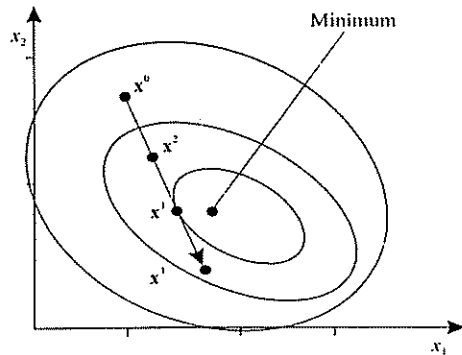
<sup>27</sup> Opis problemów numerycznych stanowi domenę metod numerycznych. Zainteresowany Czytelnik opis tych problemów znajdzie w każdym podręczniku z tej dziedziny. Stosując opisane niżej metody, możemy otrzymać rozwiązanie dowolnie bliskie rozwiązaniu optimum.

<sup>28</sup> Przykładami takich metod są metoda cyklicznego przeglądu współrzędnych oraz algorytm Hooke-Jeevesa (zob. np. [Brdyś, Ruszczyski, 1985, Wit, 1986]). Idea metod polega na systematycznym przeglądzie kierunków, w których wartość funkcji maleje i przejściu do kolejnego punktu. Przejście następuje w kierunku zapewniającym maksymalny spadek wartości funkcji.

RYSUNEK 5.2

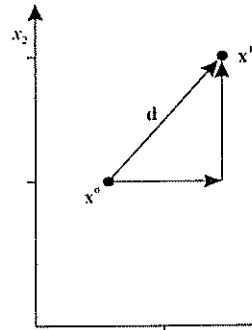
## Ilustracja sposobu wyznaczania kolejnych rozwiązań w metodach iteracyjnych

A. Dobór długości wektora  $d$ . „Przestrzelenie”: wektor przesunięcia  $d$  jest zbyt długi. Następstwem jest oddalanie się od minimum: zamiast przejść do  $x^1$ , przechodzimy do  $x^2$ . Jeżeli wektor jest zbyt krótki, następuje przejście do  $x^2$ . Nadając odpowiednią wartość  $t_k$ , eliminujemy problem niewłaściwej długości kroku



B. Sposób konstrukcji wektora przesunięcia

Długość wektora jest równa  $\sqrt{\sum_{j=1}^2 \left(\frac{\partial f(x^0)}{\partial x_j}\right)^2}$ . Dla dużych co do wartości bezwzględnej wartości pochodnych cząstkowych wektor ten jest zbyt „długi”, stąd potrzeba kontroli jego długości



## 5.5.1.1. Metoda największego spadku

Załóżmy, że funkcja  $f$  jest funkcją jednokrotnie różniczkowalną w  $R^n$  (dokładnie mającą pochodne cząstkowe pierwszego rzędu w  $R^n$ ). Startując z punktu  $x^0$ , konstruujemy ciąg  $\{x^k\}$ , korzystając ze wzoru:

$$x^{k+1} = x^k - \nabla f(x^k) \quad (5.28)$$

lub

$$x^{k+1} = x^k - t_k \nabla f(x^k), \quad (5.29)$$

gdzie  $t_k$  wybieramy w ten sposób, aby minimalizować wyrażenie  $f(x^k - t_k \nabla f(x^k))$  dla  $t_k > 0$ . Wersja metody dana przez (5.29) nosi nazwę metody z regulowaną długością kroku<sup>29</sup>.

<sup>29</sup> Wektor przesunięcia  $d^k = -\nabla f(x^k)$ . Naszym celem jest zmniejszenie wartości funkcji, przesuwanie się więc w kierunku przeciwnym do gradientu. Z tego powodu  $t_k$  może przyjmować tylko wartości dodatnie. Gdyby  $t_k$  przyjmowało wartości ujemne, przesuwalibyśmy się w kierunku, w którym wartość funkcji rośnie.

## PRZYKŁAD 5.3

Rozważmy zadanie  $f(x_1, x_2) = x_1^2 + x_2^2 \rightarrow \min$ . Minimum funkcji znajduje się w punkcie  $x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ . Wyznamy to minimum, stosując metodę największego spadku.

Niech punktem startowym będzie  $x^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ;  $td^0 = -t \nabla f(x^0) = -t \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}_{(1,1)} = \begin{bmatrix} -2t \\ -2t \end{bmatrix}$ .

$$x^0 + td^0 = x^0 - t \nabla f(x^0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -2t \\ -2t \end{bmatrix} = \begin{bmatrix} 1 - 2t \\ 1 - 2t \end{bmatrix}.$$

Tak więc  $\min_{t > 0} f(x^0 + td^0) = \min_{t > 0} [(1 - 2t)^2]$ ;  $t^* = \frac{1}{2}$ . Dokonując podstawienia

$$t^* = 1/2, \text{ otrzymujemy } x^1 = x^0 - t^* \nabla f(x^0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix}_{(1,1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Minimum zostało uzyskane po wykonaniu jednej iteracji<sup>30</sup>. Można dowiedzieć, że w przypadku funkcji kwadratowej zastosowanie metody największego spadku z regulowaną długością kroku pozwala na otrzymanie rozwiązania optymalnego w jednym kroku.

## 5.5.1.2. Metoda Newtona-Raphsona

Załóżmy, że funkcja  $f$  jest dwukrotnie różniczkowalną w  $R^n$  (dokładniej: ma pochodne cząstkowe drugiego rzędu w  $R^n$ ). W metodzie Newtona-Raphsona ciąg punktów  $\{x^k\}$  uzyskuje się, stosując wzór:

$$x^{k+1} = x^k - t_k \nabla^2 f(x^k)^{-1} \nabla f(x^k), \quad (5.30)$$

gdzie  $\nabla^2 f(x^k)^{-1}$  oznacza macierz odwrotną do hesjanu funkcji  $f$  w punkcie  $x^k$ .

## PRZYKŁAD 5.4

Dana jest funkcja  $f(x_1, x_2) = x_1^2 + x_2^2 - 2x_1x_2$  i punkt startowy  $x^0 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ;  $f(x^0) = 1$ . Funkcja osiąga minimum niewłaściwe dla  $x_1 = x_2$  i  $x_1 = -x_2$ ;  $f(x^*) = 0$ . W celu iteracyjnego wyznaczenia minimum, stosujemy wzór (5.30).

<sup>30</sup> W celu uproszczenia zapisu pominięto indeks dolny przy  $t_k$ .

Gradient funkcji jest równy  $\nabla f(x) = \begin{bmatrix} 2x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{bmatrix}$ . Wykonujemy iteracje algorytmu Newtona-Raphsona:

$$x^1 = x^0 - t \nabla f(x^0) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - t \begin{bmatrix} 2x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{bmatrix}_{(2,3)} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - t \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 + 2t \\ 3 - 2t \end{bmatrix}.$$

Wyznaczamy rozwiązanie optymalne zadania:

$$\min_{t > 0} f(x^0 + t d^0) = \min_{t > 0} [2(2 + 2t)^2 + (3 - 2t)^2 - 2(2 + 2t)(3 - 2t)]; \quad t^* = 1/4.$$

Podstawiając  $t^* = 1/4$  do formuły rekurencyjnej, otrzymujemy:

$$x^1 = x^0 - t^* \nabla f(x^0) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2^{1/2} \\ 2^{1/2} \end{bmatrix},$$

a więc punkt leżący na jednej z prostych wyznaczających rozwiązania optymalne.

Z przykładu wynika, że w celu stwierdzenia, czy wyznaczone metodami iteracyjnymi minimum jest jedyne, konieczna jest szczegółowa analiza właściwości funkcji kryterium.

W przypadku funkcji nieograniczonej z dołu iteracje byłyby kontynuowane w nieskończoność. W praktyce, próbując wyznaczyć rozwiązanie problemu z nieograniczoną funkcją celu za pomocą pakietu optymalizacyjnego (Solvera), najczęściej otrzymamy komunikat o braku zbieżności rozwiązania do optimum.

Intuicja podpowiada, że im bliżej minimum się znajdujemy, tym mniejsze wartości przyjmują pochodne cząstkowe, z czego wynika, że wektor przesunięcia ma mniejszą długość. Oznacza to, że zmiany wartości funkcji w kolejnych iteracjach stają się coraz mniejsze. Uzasadnia to korzystanie z kryterium stopu posługującego zakończenie obliczeń w przypadku, gdy zmiany wartości funkcji są mniejsze od pewnej, wystarczająco małej, dodatniej stałej. Łatwo jest także uzasadnić zakończenie obliczeń po wykonaniu pewnej, wystarczająco dużej, liczby iteracji (kryterium często wykorzystywane w pakietach optymalizacyjnych). Jeżeli przybliżanie się do minimum jest (w sensie zmian wartości funkcji celu czy też długości kroku) coraz powolniejsze, to po wykonaniu odpowiednio dużej liczby iteracji możemy oczekiwać, że zmiany wartości zmiennych (wartości funkcji) kolejnych iteracji będą już bardzo małe i bliskie wartościom przyjmowanym w rozwiązaniu optymalnym.

Stosując dowolny algorytm iteracyjny z regulowaną długością kroku, przechodzimy w kolejnym kroku do rozwiązania, w którym funkcja kryterium przyjmuje mniejszą wartość. Jeżeli funkcja ma wiele minimum lokalnych, to to, które

z nich osiągniemy stosując dowolny algorytm z regulowaną długością kroku, zależy od wyboru punktu startowego<sup>31</sup>.

Metody iteracyjne pozwalają na wyznaczenie minimum lokalnego. Wartości funkcji w minimum lokalnych mogą się od siebie znacznie różnić, w szczególności istotnie odbiegać od minimum globalnego. Nierozpoznanie struktury zadania może prowadzić do wzięcia minimum lokalnego za globalne, czego następstwem może być uzyskanie rozwiązania odległego (w sensie wartości funkcji celu lub wartości zmiennych) od minimum globalnego<sup>32</sup>.

### 5.5.2. Idea metod minimalizacji w problemach programowania nieliniowego z ograniczeniami

Obecnie poczynimy kilka uwag na temat sposobu wyznaczania optimum w przypadku problemów z warunkami ograniczającymi. Rozważmy problem (5.7)–(5.9). Wyznaczanie minimum tego problemu można sprowadzić do wyznaczania minimum kolejno rozwiązywanych pomocniczych problemów bez ograniczeń.

W  $k$ -tej iteracji wyznaczamy rozwiązanie optymalne problemu pomocniczego, które oznaczmy przez  $x^k$ . Następnie modyfikujemy parametry problemu pomocniczego w taki sposób, aby ciąg  $\{x^k\}$  rozwiązań optymalnych rozwiązywanych problemów dążył do rozwiązania optymalnego problemu z ograniczeniami  $x^*$ . Istnieje wiele metod realizujących nasz skicowaną ideę.

Jedno z podejść polega na zastąpieniu problemu z ograniczeniami problemem optymalizacji bezwarunkowej przyjmującym postać tzw. **funkcji kary**. Im bardziej rozwiązanie optymalne tego problemu jest odległe od rozwiązania dopuszczalnego problemu wyjściowego, tym funkcja ta przybiera większe wartości. Sekwencyjna modyfikacja parametrów funkcji kary zapewnia, że w kolejnych krokach otrzymujemy rozwiązania w coraz mniejszym stopniu naruszające warunki dopuszczalności i coraz bardziej zbliżone do minimum problemu z ograniczeniami.

<sup>31</sup> Jednym ze sposobów numerycznego wyznaczenia minimum globalnego w przypadku funkcji, która nie jest wypukła, jest wielokrotne powtarzanie procedury poszukiwania minimum lokalnego. Przy każdym powtórzeniu korzysta się z innego (np. losowego wygenerowanego, spełniającego warunki dopuszczalności) punktu startowego. Obecnie bada się możliwości wykorzystania szeregu algorytmów, np. algorytmów genetycznych (zob. rozdział 15) w celu sprawnego wyznaczenia minimum globalnego funkcji niebędącej funkcją wypukłą. Trudność wyznaczenia minimum dowolnej funkcji nakłada istotne praktyczne ograniczenia na proces modelowania. Można zawsze zadać pytanie, jaki sens ma konstrukcja modelu wiernie odzwierciedlającego realia procesu decyzyjnego, skoro wyznaczenie rozwiązania optymalnego będzie trudne lub niemożliwe.

<sup>32</sup> Czytelnika zainteresowanego własnościami opisywanych metod, np. problemami doboru punktu startowego, rodzajami i szybkością zbieżności — ze względu na wysoce specjalistyczny charakter tych zagadnień — odsyłamy do literatury zamieszczonej na końcu rozdziału.

Metody kierunków dopuszczalnych sprowadzają się do wyznaczenia dopuszczalnego rozwiązania wyjściowego i poprawy tego rozwiązania polegającej na przejściu do lepszego i zarazem dopuszczalnego rozwiązania wzdłuż tzw. kierunku dopuszczalnego, tj. takiego kierunku, który zapewnia, że nie opuścimy zbioru rozwiązań dopuszczalnych. Sposób wyboru kierunku oraz długości kroku zależy od konkretnej metody<sup>33</sup>. Wiele metod optymalizacji bezwarunkowej może być zmodyfikowanych w ten sposób, aby można było je wykorzystać do rozwiązywania problemów programowania nieliniowego z ograniczeniami<sup>34</sup>.

## 5.6. Wybrane problemy programowania nieliniowego sprowadzalne do problemu programowania liniowego

W wielu przypadkach udaje się sprowadzić problem programowania nieliniowego do równoważnego problemu programowania liniowego. Równoważność polega na tym, że znając rozwiązanie optymalne pomocniczego problemu liniowego, można wyznaczyć rozwiązanie optymalne wyjściowego problemu nieliniowego.

Przedstawimy dwa — często spotykane w praktyce modelowania — problemy programowania nieliniowego sprowadzalne do problemu programowania liniowego, a mianowicie problem programowania hiperbolicznego oraz problem z maksymalizacją funkcją celu. W obu przypadkach warunki ograniczające są liniowe, natomiast funkcja celu w pierwszym przypadku jest funkcją ułamkową, a w drugim w funkcji celu występuje operator maksimum lub minimum. W przypadku programowania hiperbolicznego chcemy, konstruując pewien wskaźnik, w jednej funkcji celu zawrzeć dwa kryteria. W drugim przypadku dążymy do maksymalizacji wartości kryterium cząstkowego osiągającego w zbiorze kryteriów minimalną wartość. Do wymienionych problemów można sprowadzić niektóre z zagadnień programowania wielokryterialnego, w szczególności dwukryterialnego (zob. rozdział 4).

### 5.6.1. Programowanie hiperboliczne

Wiele zagadnień ekonomicznych, w szczególności z zakresu planowania produkcji, sprowadza się do wyznaczenia planu maksymalizującego (bądź minimalizującego) wartość pewnego wskaźnika (np. rentowność brutto lub netto). Tłumaczy

<sup>33</sup> Wyznaczając jego długość musimy zadbać o to, aby nie opuścić zbioru rozwiązań dopuszczalnych.

<sup>34</sup> Modyfikacje polegają nie tylko na odmiennym niż w przypadku optymalizacji bezwarunkowej wyznaczeniu długości wektora przesunięcia, ale na zaproponowaniu postępowania, jeżeli spadek wartości funkcji jest możliwy tylko wówczas, gdy poruszamy się wzdłuż brzegu zbioru  $D$ . Czytelnik zainteresowanego sposobem postępowania odsyłamy do literatury.

dużą popularność funkcji ułamkowej. W zdecydowanej większości przypadków zasadne jest przyjęcie założenia o proporcjonalności wyników do nakładów, co uzasadnia liniowość warunków ograniczających.

Problem minimalizacji (maksymalizacji) funkcji ułamkowej przy liniowych ograniczeniach nosi nazwę problemu programowania hiperbolicznego (ilorazowego). Problem ten może być za pomocą prostych przekształceń sprowadzony do problemu programowania liniowego.

Problem programowania hiperbolicznego przyjmuje postać:

$$\frac{cx + c_0}{dx + d_0} \rightarrow \min, \quad (5.31)$$

$$Ax \leq 0, \quad (5.32)$$

$$x \geq 0. \quad (5.33)$$

Zakłada się, że  $D = \{x : Ax \leq b, x \geq 0\}$  jest zwarty (domknięty i ograniczony), a  $dx + d_0 > 0$  dla każdego  $x \in D$ .

W celu wyznaczenia rozwiązania optymalnego (5.31)–(5.33) tworzymy pomocniczy problem programowania liniowego, którego rozwiązanie optymalne (w większości przypadków) pozwala wyznaczyć rozwiązanie optymalne problemu wyjściowego.

Niech  $S = dx + d_0 > 0$ ,  $y = [y_1, \dots, y_n]^T$ ,  $y_j = \frac{x_j}{dx + d_0}$ ,  $y_0 = \frac{1}{S} = \frac{1}{dx + d_0}$ ; wynika stąd, że  $x_j = y_j(dx + d_0) = 1$  (w zapisie wektorowym  $x = y(dx + d_0)$ ), co ostatecznie implikuje równość  $y_0(dx + d_0) = 1$ . Dzielenie przez  $dx + d_0$  jest zawsze wykonalne na mocy założenia  $dx + d_0 > 0$ . Wynika stąd, że  $y_j = x_j y_0$  (w postaci wektorowej  $y = xy_0$ ).

Po wykonaniu prostych podstawień problem (5.31)–(5.33) możemy sprowadzić do następującego problemu programowania liniowego:

$$cy + c_0 y_0 \rightarrow \min, \quad (5.34)$$

$$Ay - by_0 \leq 0, \quad (5.35)$$

$$dy + d_0 y_0 = 1, \quad (5.36)$$

$$y \geq 0, \quad (5.37)$$

$$y_0 > 0. \quad (5.38)$$

W celu otrzymania (5.34) wykonujemy następujący ciąg przekształceń:

$$\frac{cx + c_0}{dx + d_0} = c \frac{x}{dx + d_0} + c_0 \frac{1}{dx + d_0} = c \frac{y(dx + d_0)}{dx + d_0} + c_0 \frac{y_0(dx + d_0)}{dx + d_0} = cy + c_0 y_0.$$

Podstawiając w (5.32)  $x = y(dx + d_0)$ , dzieląc obie strony przez  $dx + d_0$ ,

a następnie podstawiając  $\frac{1}{dx + d_0} = y_0$ , uzyskujemy (5.35). Przekształcenia przedstawiają się następująco:

$$Ax \leq b \Leftrightarrow Ay(dx + d_0) \leq b \Leftrightarrow Ay \leq b \frac{1}{dx + d_0} \Leftrightarrow Ay \leq by_0.$$

Warunek (5.36) otrzymujemy, dokonując poniższych przekształceń:

$$dy + d_0 y_0 = d \frac{x}{dx + d_0} + d_0 \frac{1}{dx + d_0} = \frac{dx + d_0}{dx + d_0} = 1.$$

Dodatność  $y_0$  wynika wprost z definicji (podobnie jak nieujemność wektora  $y$ ).

Opisany sposób postępowania określa się mianem metody **Charnesa-Coopera**.

Problemu (5.34)–(5.38) nie można rozwiązać metodą simpleks z uwagi na warunek (5.38). Wygodniej jest zastąpić go warunkiem:

$$y_0 \geq 0 \quad (5.38')$$

i sprawdzić, czy w rozwiązaniu optymalnym (5.34)–(5.38') warunek (5.38) jest spełniony. Jeżeli w otrzymanym rozwiązaniu  $y_0^* = 0$ , problem (5.31)–(5.33) nie ma rozwiązania optymalnego (zob. twierdzenie 5.5).

### PRZYKŁAD 5.5

Dane jest zadanie programowania hiperbolicznego:

$$\frac{2x_1 + 3x_2 + 3}{3x_1 + 2x_2 - 12} \rightarrow \max,$$

$$x_1 + 3x_2 \geq 10,$$

$$3x_1 - x_2 \geq 10,$$

$$2x_1 - 3x_2 \leq 11,$$

$$2x_1 + 5x_2 \leq 35,$$

$$x_1, x_2 \geq 0.$$

Sprowadzamy zadanie do postaci kanonicznej:

$$\frac{2x_1 + 3x_2 + 3}{3x_1 + 2x_2 - 12} \rightarrow \max,$$

$$x_1 + 3x_2 - x_3 = 10,$$

$$3x_1 - x_2 - x_4 = 10,$$

$$2x_1 - 3x_2 + x_5 = 11,$$

$$2x_1 + 5x_2 + x_6 = 35,$$

$$x_1, \dots, x_6 \geq 0.$$

Wprowadzamy nowe zmienne w celu sprowadzenia rozważanego zadania do zadania programowania liniowego. Dla utrzymania jednolitości zapisu zmienne  $x_3, \dots, x_6$  zastąpiono zmiennymi  $y_3, \dots, y_6$ ,  $x_j = y_j$ ,  $j = 3, \dots, 6$ .

$$2y_1 + 3y_2 + 3y_0 \rightarrow \max,$$

$$y_1 + 3y_2 - y_3 - 10y_0 = 0,$$

$$2y_1 - y_2 - y_4 - 10y_0 = 0,$$

$$2y_1 - 3y_2 + y_5 - 11y_0 = 0,$$

$$2y_1 + 5y_2 + y_6 - 35y_0 = 0,$$

$$3y_1 + 2y_2 - 12y_0 = 1,$$

$$y_1, \dots, y_6 \geq 0, \quad y_0 > 0.$$

Rozwiązaniem optymalnym zadania pomocniczego jest:

$$y_1^* = 1, y_2^* = 0,5, y_3^* = y_4^* = 0, y_5^* = 2,25, y_6^* = 4,25, y_0^* = 0,25, g(y^*) = 4,25.$$

Korzystając z zależności  $x_j = y_j/y_0$ , otrzymujemy  $x_1^* = 4, x_2^* = 2, f(x^*) = 4,25$ .

Metodę Charnesa-Coopera uzasadniają następujące twierdzenia.

Jeżeli istnieje rozwiązanie optymalne  $(y^*, y_0^*)$  problemu (5.34)–(5.38'), takie że  $y_0^* > 0$ , to wektor  $x = \frac{y^*}{y_0^*}$  jest rozwiązaniem optymalnym (5.31)–(5.33).

Jeżeli  $D \neq \emptyset$  oraz istnieje rozwiązanie optymalne (5.34)–(5.38')  $(y^*, y_0^*)$ , dla którego zachodzi  $y_0^* = 0$ , to nie istnieje rozwiązanie optymalne (5.31)–(5.33).

Jeżeli  $D \neq \emptyset$  i funkcja celu  $cy + c_0 y_0$  jest nieograniczona od dołu, to funkcja celu (5.31) jest także nieograniczona od dołu.

Jeżeli problem (5.34)–(5.38') jest sprzeczny, to problem (5.31)–(5.33) jest także sprzeczny.

Wynika stąd, że rozwiązanie optymalne problemu (5.31)–(5.33) istnieje wtedy i tylko wtedy, gdy istnieje rozwiązanie optymalne  $(y^*, y_0^*)$  problemu (5.34)–(5.38'), takie że  $y_0^* > 0$ .

W większości problemów spotykanych w praktyce jest spełniony warunek  $dx + d_0 > 0$ , co zabezpiecza przed wystąpieniem braku skończonego ekstremum. Zazwyczaj w rozwiązaniu optymalnym ( $y^*$ ,  $y_0^*$ ) problemu (5.34)–(5.38')  $y_0^* > 0$ , co umożliwia wyznaczenie rozwiązania optymalnego problemu wyjściowego.

### 5.6.2. Maksyminowa funkcja celu

Dany jest problem:

$$\min_{k=1, \dots, s} \{c^k x + c_{k0}\} \rightarrow \max, \quad (5.39)$$

$$Ax = b, \quad (5.40)$$

$$x \geq 0. \quad (5.41)$$

Rozważmy ciąg problemów:

$$x_0 \rightarrow \max, \quad x_0 = \min_{k=1, \dots, s} \{c^k x + c_{k0}\}, \quad Ax = b, \quad x \geq 0,$$

$$x_0 \rightarrow \max, \quad x_0 \leq \min_{k=1, \dots, s} \{c^k x + c_{k0}\}, \quad Ax = b, \quad x \geq 0,$$

$$x_0 \rightarrow \max, \quad x_0 \leq c^k x + c_{k0}, \quad k = 1, \dots, s, \quad Ax = b, \quad x \geq 0,$$

Pierwszy z nich jest inaczej zapisanym problemem (5.39)–(5.41). Przy przejściu od drugiego do trzeciego problemu zbiór rozwiązań dopuszczalnych ulega poszerzeniu, nie ulega natomiast zmianie zbiór rozwiązań optymalnych.

Ostatecznie problem (5.39)–(5.41) możemy zastąpić równoważnym (ze względu na zbiór rozwiązań optymalnych) problemem programowania liniowego:

$$x_0 \rightarrow \max, \quad (5.42)$$

$$x_0 - c^k x \leq c_{k0}, \quad k = 1, \dots, s \quad (5.43)$$

i warunkach (5.40)–(5.41). Na  $x_0$  nie nakłada się warunków znakowych, ponieważ  $c^k x + c_0$  jest nieograniczone co do znaku<sup>35</sup>.

Stosując elementarne przekształcenia, można w identyczny sposób sprowadzić do problemu PL problem:

$$\min_{k=1, \dots, s} \{c^k x + c_{k0}\} \rightarrow \max, \quad (5.44)$$

przy warunkach (5.40)–(5.41).

<sup>35</sup> Problemu (5.42)–(5.43) nie możemy rozwiązywać metodą simpleks, gdyż na jedną ze zmiennych nie został nałożony warunek nieujemności. Zastępując zmienną  $x_0$  zmiennymi  $x_0', x_0'' \geq 0$ ,  $x_0 = x_0'' - x_0'$ , sprowadzamy rozważany problem do problemu, w którym na wszystkie zmienne został nałożony warunek nieujemności.

### POLECENIA I PYTANIA KONTROLNE

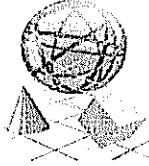
1. Omówić wady paradygmatu liniowości. Podać konkretny przykład z ekonomii.
2. Zdefiniować terminy: minimum lokalne, minimum globalne, kres dolny, funkcja wypukła, funkcja wklęsła. Podać ilustrację geometryczną dla funkcji jednej i dwóch zmiennych.
3. Podać klasyfikację problemów programowania nieliniowego. Dlaczego w zastosowaniach istotną rolę odgrywają problemy programowania wypukłego?
4. Zdefiniować problem programowania kwadratowego.
5. Omówić problem Lagrange'a i Kuhna-Tuckera. Wskazać różnice między tymi problemami.
6. Wskazać różnice między metodami analitycznymi a numerycznymi wyznaczania rozwiązań optymalnych problemów programowania nieliniowego.
7. Omówić działanie metody największego spadku i Newtona. Na czym polega wyższość metod z regulowaną długością kroku? Sporządzić odpowiednie rysunki.
8. Uzasadnić posługiwanie się omawianymi kryteriami stopu.
9. Na czym polega trudność wyznaczenia rozwiązania optymalnego (minimum globalnego) w przypadku, gdy funkcja celu nie jest funkcją wypukłą?
10. Omówić sposób wyznaczania rozwiązania optymalnego problemu programowania hiperbolicznego. Kiedy na podstawie rozwiązania problemu pomocniczego nie można wyznaczyć rozwiązania optymalnego problemu wyjściowego?

### BIBLIOGRAFIA

- Bazara M. S., Shetty C. M., 1980, *Nonlinear Programming. Theory and Algorithms*, John Wiley & Sons.
- Brassard G., Bratley P., 1969, *Fundamentals of Algorithms*, Prentice Hall.
- Brdys M., Ruszczyński A., 1985, *Metody optymalizacji w zadaniach*, WNT, Warszawa.
- Dornqvist A., 1967, *Metody numeryczne*, PWN, Warszawa.
- Grabowski W., 1980, *Programowanie matematyczne*, PWN, Warszawa.
- McCormick G.P., 1983, *Nonlinear Programming. Theory and Applications*, John Wiley & Sons.
- Reiner H., Panos M., P., 1995, *Handbook of Global Optimization*, Kluwer Academic Publishers.
- Runka H.J., 1997, *Programowanie matematyczne, część II*, Akademia Ekonomiczna w Poznaniu, Poznań.
- Shapiro J.F., 1980, *Mathematical Programming. Theory, Algorithms and Applications*, John Wiley & Sons.
- Wit R., 1986, *Metody programowania nieliniowego*, WNT, Warszawa.

## ROZDZIAŁ 6

# PROGRAMOWANIE DYNAMICZNE

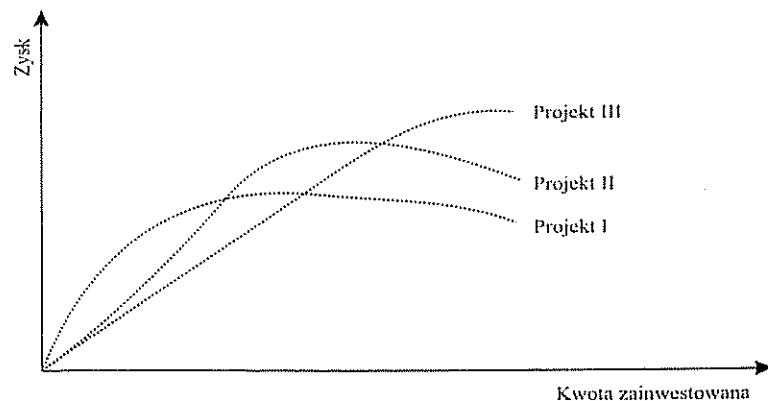


### 6.1. Istota programowania dynamicznego. Zasada optymalności Bellmana

Załóżmy, że dysponujemy pewną sumą pieniędzy, którą możemy przeznaczyć na trzy różne projekty. Przewidywane zyski z realizacji tych projektów zależą od rodzaju projektu oraz od wielkości funduszu przydzielonego na jego wykonanie (rysunek 6.1).

Zależałoby nam na takim rozdzieleniu posiadanych środków między projekty, aby łączny zysk był maksymalny. Jeżeli optymalnej strategii będziemy szukać, badając wszystkie możliwe kombinacje<sup>1</sup>, to może się okazać, że zabierze nam to

RYSUNEK 6.1  
Przewidywane zyski dla trzech projektów



<sup>1</sup> Na rysunku 6.1 krzywe zysku dla poszczególnych działalności są punktowe, a zatem zbiór możliwych strategii alokacji posiadanej kwoty pieniężnej jest skończony.

bardzo dużo czasu. Zdecydowanie szybciej i sprawniej poradzimy sobie z tym problemem, stosując metody programowania dynamicznego, któremu poświęcono ten rozdział.

Metody te znajdują właśnie zastosowanie w problemach kombinatorycznych. Są one efektywne, gdyż zamiast żmudnego badania wszystkich możliwych wariantów wykorzystywane są odpowiednie równania funkcyjne, które zostaną szczegółowo omówione w dalszej części rozdziału.

Programowanie dynamiczne wspomaga zarządzanie procesami wieloetapowymi. Charakterystyczne dla metod programowania dynamicznego jest to, że proces (lub problem) zostaje podzielony na fazy. Faza nie musi wcale oznaczać początku nowego okresu. Określa ona raczej pewien etap w naszych rozważaniach, np. nowy wariant rozwiązania (dotychczas pominięty) bądź wprowadzenie nowej zmiennej. Programowanie dynamiczne niekoniecznie dotyczy więc zadań rozwijających się w czasie. Zagadnienia, do których stosuje się programowanie dynamiczne i traktuje się czas jako zmienną podstawową, są np. związane z ustaleniem wieloletniego planu inwestycyjnego, ze sterowaniem zapasami czy też z opracowaniem cyklu remontów i wymiany urządzeń produkcyjnych [Lesz, 1968; Trzaskalik, 2003]. Wymienione zagadnienia wiążą się z procesami decyzyjnymi, które są z natury wieloetapowe. Językiem programowania dynamicznego opisywane są też zadania, w których procesu w ogóle nie ma, a zamiast niego kilka zjawisk występuje równocześnie. Poszczególne zjawiska są wprowadzane do obliczeń po kolei, co sprawia, że wspomniany „proces” staje się wyłącznie procesem myślowym. Do tej kategorii zagadnień można zaliczyć m.in. zadania przydziału zasobu<sup>2</sup>, problem najdłuższej lub najkrótszej drogi w sieci<sup>3</sup>, analizę czasowo-kosztową przedsięwzięć [Hindelang, Muth, 1979; Robinson, 1975], zagadnienie załadunku czy też zagadnienie dyliżansu [Kukuła, 1996]. Różnorodność zadań, do których są stosowane metody programowania dynamicznego, sprawia, że często nie mogą być one rozwiązywane według jednego schematu<sup>4</sup>. Nie istnieje też możliwość ich jednolitego zapisu.

Stosując programowanie dynamiczne, nie szukamy maksimum (minimum) funkcji wielu zmiennych, lecz staramy się rozwiązać ciąg podzadań polegających na wyznaczeniu ekstremum funkcji jednej zmiennej (choć nie wciąż tej samej).

Istnieją metody umożliwiające obliczenie ekstremum funkcji wielu zmiennych, np. klasyczne metody rachunku różniczkowego. Polegają one na ustaleniu wzorów

<sup>2</sup> Dotyczyć one mogą np. zasilania z jednego źródła urządzeń połączonych równolegle, alokacji kapitału w różnych środkach masowego przekazu na skuteczną reklamę nowego produktu czy też finansowania przedsięwzięć inwestycyjnych (zob. [Kukuła, 1996, s. 205–209]).

<sup>3</sup> Zob. odpowiednio podrozdziały 7.2 i 3.3. Stosując CPM w celu wyznaczenia ścieżki krytycznej (najdłuższej), określamy tzw. momenty zajścia zdarzeń. Sposób ich ustalania przypomina właśnie tok postępowania przyjęty w programowaniu dynamicznym (por. [Lesz, 1968, s. 29–33]).

<sup>4</sup> Ciekawostką jest to, że każde zadanie z zakresu programowania liniowego można rozwiązać za pomocą metod programowania dynamicznego (zob. [Lesz, 1968, s. 97–102]).



na pochodne cząstkowe tej funkcji względem wszystkich zmiennych po kolei i na przyrównaniu tych pochodnych do zera. Okazuje się jednak, że otrzymanie rozwiązania, dla którego gradient jest zerowy, stanowi warunek konieczny, lecz nie dostateczny dla istnienia ekstremum. Pochodna jest równa zero nie tylko w punktach, gdzie występują ekstrema, lecz także w punktach siodłowych. Kolejne ograniczenie rachunku różniczkowego wynika z tego, że w wielu sytuacjach poszukujemy ekstremum w pewnym obszarze ograniczonym, a tymczasem przyrównując pochodne do zera, otrzymujemy ekstrema lokalne. Nie znajdujemy natomiast zazwyczaj wartości najmniejszych lub największych położonych na krańcach interesującego nas obszaru. Wreszcie, rachunek różniczkowy traci na znaczeniu, gdy mamy do czynienia z funkcjami nieróżniczkowalnymi (np. funkcją schodkową) bądź z maksymalizacją na zbiorach nieciągłych. Chociaż rachunek różniczkowy jest bardzo pomocnym narzędziem, istnieją problemy optymalizacyjne, które z uwagi na swój charakter — powinny zostać rozwiązane metodami programowania dynamicznego. Pokonują one bowiem wymienione wyżej trudności.

**Metoda programowania dynamicznego** polega na ustaleniu ekstremów za pomocą szeregu równań rekurencyjnych, pozwalających przejść z jednego etapu do kolejnego. Przy wyprowadzaniu tychże równań korzystamy, niezależnie od sposobu rozwiązywania danego problemu, z następującego twierdzenia.

### Zasada optymalności Bellmana

Strategia optymalna ma tę własność, że niezależnie od początkowego stanu i początkowej decyzji pozostałe decyzje muszą stanowić strategię optymalną ze względu na stan wynikający z pierwszej decyzji.

Zgodnie z tą zasadą, ciąg decyzji  $x_1^*, x_2^*, \dots, x_N^*$  będzie optymalną strategią w procesie  $N$ -etapowym tylko wtedy, gdy ciąg decyzji  $x_1^*, x_2^*, \dots, x_{N-1}^*$  będzie optymalną strategią w procesie  $N-1$ -etapowym. Strategią nazywamy funkcję, która każdemu stanowi dopuszczalnemu, w jakim może znaleźć się proces, przyporządkowuje pewną decyzję dopuszczalną. Strategia optymalna natomiast przyporządkowuje temuż stanowi decyzję optymalną i pozwala wygenerować optymalną realizację procesu, czyli taki ciąg stanów i decyzji, który optymalizuje wartość funkcji celu. Kierowanie się powyższą zasadą znacząco przyspiesza dojście do optimum.

U podstaw procedury programowania dynamicznego leży jeszcze jedna własność, która umożliwia podejmowanie decyzji w procesie  $N$ -etapowym.

### Własność Markowa

Wartość funkcji celu w zadaniu  $N$ -etapowym jest sumą wartości uzyskanych w poszczególnych etapach, a wartość uzyskana w  $k$ -tym etapie zależy od stanu w etapie poprzednim oraz od decyzji podjętej w  $k$ -tym etapie. Nie zależy natomiast od tego, jaką drogą system doszedł do etapu  $k-1$ .

## 6.2. Zagadnienie optymalnego rozdziału zasobu

Zagadnienie optymalnego rozdziału zasobu dotyczy sytuacji, w której mamy do dyspozycji określoną ilość jakiegoś zasobu<sup>5</sup>. Może on być użytkowany na rozmaite sposoby. Każde możliwe zastosowanie nazywać będziemy działalnością. W wyniku skierowania całości środka bądź jego części do wybranej działalności pojawia się przychód mierzony niekoniecznie w jednostkach pieniężnych<sup>6</sup>. Jego wielkość zależy od ilości zużytego środka i od rodzaju działalności, do której został skierowany.

Omawiane zagadnienie analizowane jest przy następujących założeniach [Bellman, Dreyfus, 1967]:

- 1) rozdział zasobu dokonywany jest w jednostkach całkowitych;
- 2) przychody z poszczególnych działalności zmierzyć można tą samą jednostką miary;
- 3) przychód z dowolnej działalności jest niezależny od ilości zasobu skierowanego do innej działalności;
- 4) przychód całkowity, tj. użyteczność całego procesu alokacji, jest sumą przychodów cząstkowych rozumianych jako użyteczności pochodzące z kolejnych działalności.

Niech  $g_j(x_j)$  oznacza funkcję mierzącą zależność przychodu (zysku) z  $j$ -tej działalności od ilości przydzielonego jej środka ( $x_j$ )<sup>7</sup>. Interesuje nas taki rozdział

<sup>5</sup> Na przykład siłę roboczą, pieniądze, maszyny, pojazdy, surowiec.

<sup>6</sup> Na przykład wykorzystane maszyny mogą wpłynąć na przyrost posiadanych pieniędzy z tytułu sprzedaży wyprodukowanych przez nie wyrobów, ale mogą też wytworzyć inne urządzenia produkcyjne bądź zaowocować wzrostem wydajności przedsiębiorstwa.

<sup>7</sup> Ustalenie postaci funkcji  $g_j(x_j)$  jest dość skomplikowane, a czasem wręcz niemożliwe. Dlatego metody programowania dynamicznego są często metodami numerycznymi, czyli takimi, w których funkcje są podawane nie w postaci ogólnego wzoru analitycznego, lecz w formie wartości w tabeli. Tablicowany jest oczywiście tylko pewien określony zbiór wartości funkcji odpowiadający skończonemu zbiorowi argumentów, wyrażonemu zazwyczaj w liczbach całkowitych.

zasobu między  $N$  działalnościami, który maksymalizuje przychód (zysk) całkowity będący funkcją  $N$  zmiennych:

$$R(x_1, x_2, \dots, x_N) = \sum_{j=1}^N g_j(x_j), \quad (6.1)$$

przy warunkach

$$\sum_{j=1}^N x_j \leq a, \quad (6.2)$$

$$x_j \in d_j, \quad j = 1, 2, \dots, N, \quad (6.3)$$

Warunek (6.2) wynika z tego, że wielkość zasobu, którą mamy do dyspozycji, jest ograniczona i wynosi  $a$ . Warunek ten został celowo zapisany w postaci nierówności, gdyż maksymalizacja przychodu może, lecz nie musi oznaczać całkowitego zużycia dostępnej ilości zasobu. Zagadnienie optymalnej alokacji zasobu można również rozpatrywać przy założeniu, że posiadany zasób należy w pełni wykorzystać. Wówczas wyrażenie (6.2) zapisujemy jako równanie. W przypadku tak postawionego problemu otrzymane rozwiązanie optymalne okaże się racjonalne tylko wtedy, gdy funkcje przychodów (zysków) całkowitych dla poszczególnych działalności będą monotoniczne (rosnące).

Symbol  $d_j$  oznacza dziedzinę zmiennej  $x_j$ , czyli jej skończony zbiór stanów dopuszczalnych, taki że  $d_j \in \langle d_j^-, d_j^+ \rangle$ . Zbiór ten składa się najczęściej z nieujemnych wartości całkowitych. Jeżeli zbiory stanów dopuszczalnych ( $d_j$ ) trzech zmiennych niezależnych ( $x_j$ ) są pięcioelementowe, tj. każda zmienna może przyjąć pięć różnych wartości, to proces maksymalizacji dla  $N = 3$  zmiennych prowadzi do wyboru spośród  $5^N = 5^3 = 125$  różnych możliwości, przy założeniu, że pomijamy warunek (6.2). Lista kombinacji zawęży się, gdy z góry określamy maksymalną łączną wielkość zasobu do rozdziału.

W kontekście zagadnienia rozdziału zasobu **zasada optymalności Bellmana**, której dokładna treść przytoczona została w podrozdziale 6.1, głosi, że gdy wybraliśmy jakąś początkową wartość  $x_N$ , to nie rozpatrujemy już wszystkich strategii możliwych przy danej wartości  $x_N$ , ale tylko te strategie, które są optymalne dla procesu o  $N - 1$  etapach, przy zasobach równych  $a - x_N$  (zob. [Bellman, Dreyfus, 1967]). Przypomnijmy, że to twierdzenie pozwala na dekompozycję zadania wyjściowego na ciąg powiązanych ze sobą prostszych zadań, które należy rozwiązywać po kolei. Na podstawie optimów warunkowych dla poszczególnych etapów można ustalić optymalne rozwiązanie zadania (6.1)–(6.3).

W zagadnieniu rozdziału zasobu zakładamy, że alokacje są dokonywane pojedynczo. Na każdym etapie podejmowana jest decyzja dotycząca jednej działalności przy różnych możliwych stanach układu. Stanem układu na  $k$ -tym etapie jest ilość środka, jaką przeznaczamy do rozdziału między działalnościami rozpatrywanymi na etapie 1, 2, ...,  $k$ . Stan układu  $x$  może przyjąć wartości ze zbioru  $\{0, 1, \dots, a\}$ .

Jeżeli przyjmiemy, że  $x_k$  oznacza alokację przeprowadzoną w  $k$ -tej działalności oraz, że:

$$0 \leq x_k \leq x, \quad (6.4)$$

to pozostałą ilość zasobu ( $x - x_k$ ) warto zużyć w taki sposób, aby dała maksymalny przychód (zysk) z pozostałych  $(k - 1)$  działalności równy  $f_{k-1}(x - x_k)$ . W tej sytuacji początkowe skierowanie zasobu  $x_k$  do  $k$ -tej działalności zaowocuje przychodem całkowitym wynoszącym:

$$g_k(x_k) + f_{k-1}(x - x_k). \quad (6.5)$$

Wyrażenie (6.5) wynika z własności Markowa omówionej w podrozdziale 6.1. Należy oczywiście ustalić taką wartość  $x_k$ , która maksymalizuje funkcję (6.5). Dochodzimy zatem do podstawowego równania funkcyjnego:

$$f_k(x) = \max_{\substack{0 \leq x_k \leq x \\ x_k \in d_k}} \{g_k(x_k) + f_{k-1}(x - x_k)\}, \quad k = 2, \dots, N, \quad (6.6)$$

gdzie  $f_k(x)$  oznacza maksymalny przychód (zysk) z rozdziału co najwyżej  $x$  jednostek zasobu między  $k$  działalnościami<sup>8</sup>. Równanie funkcyjne dla etapu  $k = 1$  przyjmie postać:

$$f_1(x) = \max_{\substack{0 \leq x_1 \leq x \\ x_1 \in d_1}} \{g_1(x_1)\}. \quad (6.7)$$

### PRZYKŁAD 6.1

Samorząd województwa zamierza uruchomić maksymalnie pięć punktów informacji turystycznej w trzech miastach  $A$ ,  $B$  i  $C$ . Koszt miesięcznego utrzymania jednego punktu wynosi 20 tys. zł. W tabelicy 6.1 podano szacunkową miesięczną

TABLICA 6.1  
Liczba turystów korzystających z oferty

| Liczba punktów informacji turystycznej | Łączna liczba turystów korzystających z oferty turystycznej |     |     |
|--|---|-----|-----|
|  | A   | B   | C   |
| 0                                      | 35  | 45  | 50  |
| 1                                      | 80  | 75  | 95  |
| 2                                      | 100   | 105 | 110 |
| 3                                      | 115   | 130 | 115 |

<sup>8</sup> Przykładowo, dla funkcji  $f_2(x)$  należy porównać następujące sumy:  $g_2(0) + f_1(x)$ ,  $g_2(1) + f_1(x - 1)$ , ...,  $g_2(x) + f_1(0)$  i zapamiętać największą z nich.

liczbę turystów, którzy przy danej liczbie punktów informacyjnych skorzystają z atrakcji turystycznych przygotowanych przez samorząd. Szacuje się, że każdy turysta przynosi miesięczny dochód równy 1 tys. zł. Samorząd zastanawia się, ile punktów informacji turystycznej powinien uruchomić w poszczególnych miastach, aby całkowity zysk miesięczny był maksymalny. Koszty uruchomienia punktów informacyjnych można pominąć.

Zanim rozwiążemy problem za pomocą programowania dynamicznego, powinniśmy ustalić wartości zysków  $g_j(x_j)$  dla wybranych argumentów<sup>9</sup>.

TABLICA 6.2  
Zyski dla poszczególnych miast

| Liczba punktów informacji turystycznej | Zyski (tys. zł) |            |            |
|--|-----------------|------------|------------|
|  | A               | B          | C          |
|  | $g_1(x_1)$      | $g_2(x_2)$ | $g_3(x_3)$ |
| 0                                      | 35              | 45         | 50         |
| 1                                      | 60              | 55         | 75         |
| 2                                      | 60              | 65         | 70         |
| 3                                      | 55              | 70         | 55         |

Każde miasto potraktujemy jako odrębną działalność. Liczba działalności decyduje o liczbie etapów, stąd w rozpatrywanym zadaniu mamy do czynienia z trzema etapami.

Etap 1. Ustalamy możliwy zysk na podstawie równania (6.7), zakładając, że punkty informacyjne uruchomione zostaną jedynie w mieście A.

$$\begin{aligned}
 f_1(0) &= g_1(0) = 35, & x_1^*(0) &= 0, \\
 f_1(1) &= \max\{g_1(0), g_1(1)\} = 60, & x_1^*(1) &= 1, \\
 f_1(2) &= \max\{g_1(0), g_1(1), g_1(2)\} = 60, & x_1^*(2) &= 1, \\
 f_1(3) &= \max\{g_1(0), g_1(1), g_1(2), g_1(3)\} = 60, & x_1^*(3) &= 1, \\
 f_1(4) &= \max\{g_1(0), g_1(1), g_1(2), g_1(3)\} = 60, & x_1^*(4) &= 1, \\
 f_1(5) &= \max\{g_1(0), g_1(1), g_1(2), g_1(3)\} = 60, & x_1^*(5) &= 1.
 \end{aligned}$$

Etap 2. Korzystając ze wzoru (6.6), ustalamy zyski dla sytuacji, w której samorząd uruchamia punkty informacyjne w miastach A i B:

<sup>9</sup> Na przykład uruchomienie trzech punktów w mieście A przyniesie zysk:  $g_A(3) = 115 \cdot 1 - 3 \cdot 20 = 55$  tys. zł. W omawianym przykładzie zbiór stanów dopuszczalnych dla każdej działalności jest 4-elementowy.

$$\begin{aligned}
 f_2(0) &= g_1(0) + g_2(0) = 35 + 45 = 80, & x_2^*(0) &= 0, \\
 f_2(1) &= \max \left\{ \begin{array}{l} g_2(0) + f_1(1-0) \\ g_2(1) + f_1(1-1) \end{array} \right\} = \max \left\{ \begin{array}{l} 45 + 60 \\ 55 + 35 \end{array} \right\} = 105, & x_2^*(1) &= 0, \\
 f_2(2) &= \max \left\{ \begin{array}{l} g_2(0) + f_1(2-0) \\ g_2(1) + f_1(2-1) \\ g_2(2) + f_1(2-2) \end{array} \right\} = \max \left\{ \begin{array}{l} 45 + 60 \\ 55 + 60 \\ 65 + 35 \end{array} \right\} = 115, & x_2^*(2) &= 1, \\
 f_2(3) &= \max \left\{ \begin{array}{l} g_2(0) + f_1(3-0) \\ g_2(1) + f_1(3-1) \\ g_2(2) + f_1(3-2) \\ g_2(3) + f_1(3-3) \end{array} \right\} = \max \left\{ \begin{array}{l} 45 + 60 \\ 55 + 60 \\ 65 + 60 \\ 70 + 35 \end{array} \right\} = 125, & x_2^*(3) &= 2, \\
 f_2(4) &= \max \left\{ \begin{array}{l} g_2(0) + f_1(4-0) \\ g_2(1) + f_1(4-1) \\ g_2(2) + f_1(4-2) \\ g_2(3) + f_1(4-3) \end{array} \right\} = \max \left\{ \begin{array}{l} 45 + 60 \\ 55 + 60 \\ 65 + 60 \\ 70 + 60 \end{array} \right\} = 130, & x_2^*(4) &= 3, \\
 f_2(5) &= \max \left\{ \begin{array}{l} g_2(0) + f_1(5-0) \\ g_2(1) + f_1(5-1) \\ g_2(2) + f_1(5-2) \\ g_2(3) + f_1(5-3) \end{array} \right\} = \max \left\{ \begin{array}{l} 45 + 60 \\ 55 + 60 \\ 65 + 60 \\ 70 + 60 \end{array} \right\} = 130, & x_2^*(5) &= 3.
 \end{aligned}$$

Wartości  $x_2^*(x)$  określają, jakie środki, czyli ile punktów informacyjnych należy uruchomić w mieście B, zakładając, że w obu miastach (A i B) może pojawić się co najwyżej  $x$  punktów oraz że do końca procesu będziemy postępowali w sposób optymalny. Na razie nie wiemy, ile ostatecznie punktów zostanie uruchomionych łącznie w miastach A i B, ponieważ nie rozpatrzyliśmy jeszcze trzeciego — ostatniego etapu. Dlatego też rozwiązania optymalne wyznaczone zostały dla różnych poziomów  $x$ . Począwszy od wartości  $f_2(4)$  liczba możliwych kombinacji nie rośnie, gdyż w poszczególnych miastach działać mogą maksymalnie trzy punkty informacyjne.

Dokonane obliczenia można przedstawić w tzw. tabeli możliwych zysków (tablica 6.3).

Ustalanie zysków w tablicy 6.3 odbywa się w następujący sposób:

1. W pierwszej kolumnie zakładamy, że — niezależnie od łącznej liczby punktów informacyjnych uruchomionych w obu miastach — w mieście B nie będzie ani jednego takiego punktu. Zyski stanowią zatem sumę  $g_2(0)$  oraz  $f_1(x-0)$ .
2. W drugiej kolumnie przyjmujemy, że w mieście B działać będzie dokładnie jeden punkt informacji turystycznej. Skoro w drugim wierszu rozpatrujemy sytuację, w których łącznie w obu miastach samorząd uruchomi maksymalnie jeden punkt ( $x=1$ ), to zysk wyniesie  $g_2(1) + f_1(0) = 55 + 35 = 90$ . W trzecim wierszu są przedstawione zyski dla różnych kombinacji zakładających urucho-

TABLICA 6.3  
Tabela możliwych zysków (etap 2)

|          |       | $g_2(x_2)$ |     |     |     |
|----------|-------|------------|-----|-----|-----|
|          |       | 45         | 55  | 65  | 70  |
| $f_1(x)$ | $x_2$ | 0          | 1   | 2   | 3   |
|          | $x$   |            |     |     |     |
| 35       | 0     | 80         | —   | —   | —   |
| 60       | 1     | 105        | 90  | —   | —   |
| 60       | 2     | 105        | 115 | 100 | —   |
| 60       | 3     | 105        | 115 | 125 | 105 |
| 60       | 4     | 105        | 115 | 125 | 130 |
| 60       | 5     | 105        | 115 | 125 | 130 |

mienie co najwyżej dwóch punktów. Ponieważ wiadomo, że w drugiej kolumnie na pewno jeden punkt będzie obsługiwał turystów w mieście B, to ten drugi punkt może zostać otwarty w mieście A (zyski wyniosą  $55 + 60 = 115$ ). Jeżeli natomiast zdecydujemy się uruchomić maksymalnie trzy punkty informacyjne przy czym jeden będzie działał w mieście B, to okaże się, że tylko jeden punkt warto otworzyć w mieście A (całkowity zysk tym razem również wyniesie  $55 + 60 = 115$ ) itd.

Maksymalne wyniki otrzymane w kolejnych wierszach tablicy 6.3 odpowiadają wyznaczonym wcześniej wartościom  $f_2(x)$ . W ten sposób otrzymujemy strategię, które dla danej liczby punktów informacyjnych ( $x$ ) gwarantują maksymalny zysk.

Etap 3. Ustalamy zyski przy założeniu, że możliwe jest uruchomienie punktów informacji turystycznej we wszystkich trzech miastach. Rozpatrywane stadium procesu stanowi już ostatni etap obliczeń, zatem tym razem interesować nas będzie tylko poziom zysków dla liczby punktów podanej w treści zadania ( $x = 5$ ).

$$f_3(5) = \max \begin{cases} g_3(0) + f_2(5-0) \\ g_3(1) + f_2(5-1) \\ g_3(2) + f_2(5-2) \\ g_3(3) + f_2(5-3) \end{cases} = \max \begin{cases} 50 + 130 \\ 75 + 130 \\ 70 + 125 \\ 55 + 115 \end{cases} = 205, \quad x_3^*(5) = 1.$$

Tablica 6.4 zawiera informacje o zyskach dla różnych poziomów  $x$  i  $x_3$ . Zauważmy, że gdyby w mieście C nie został uruchomiony żaden punkt, a zysk z tego tytułu był zerowy ( $g_3(0) = 0$ ), to wartości  $f_3(x)$  w kolumnie  $x_3 = 0$  byłyby równe maksymalnym wynikom z poszczególnych wierszy tablicy 6.3. W związku z tym, że nawet przy braku punktów informacyjnych w mieście C samorząd może liczyć na zyski równe 50 tys. zł miesięcznie, wartości  $f_3(x)$  w analizowanej

TABLICA 6.4  
Tabela możliwych zysków (etap 3)

|          |       | $g_3(x_3)$ |     |     |     |
|----------|-------|------------|-----|-----|-----|
|          |       | 50         | 75  | 70  | 55  |
| $f_2(x)$ | $x_3$ | 0          | 1   | 2   | 3   |
|          | $x$   |            |     |     |     |
| 80       | 0     | 130        | —   | —   | —   |
| 105      | 1     | 155        | 155 | —   | —   |
| 115      | 2     | 165        | 180 | 150 | —   |
| 125      | 3     | 175        | 190 | 175 | 135 |
| 130      | 4     | 180        | 200 | 185 | 160 |
| 130      | 5     | 180        | 205 | 195 | 170 |

kolumnie odpowiadają maksymalnym zyskom z tablicy 6.3 powiększonym o kwotę 50 tys. zł.

Oczywiście, aby dojść do rozwiązania optymalnego, wystarczy zastosować jedno z proponowanych sposobów generowania wyników pośrednich. Można spisać maksymalne wartości zysków dla kolejnych poziomów  $x$  bądź skorzystać z zapisu tablicowego. Oba podejścia prowadzą do tej samej odpowiedzi<sup>10</sup>.

Z równań  $f_3(5) = 205$ ,  $x_3^*(5) = 1$  wynika, że maksymalny zysk z tytułu uruchomienia nie więcej niż pięciu punktów informacyjnych w miastach A, B i C wynosi 205 tys. zł, przy czym strategia optymalna zakłada, że jeden z tych punktów należy uruchomić w mieście C. Pozostałe cztery punkty warto przydzielić miastom A i B. Równania otrzymane w etapie 2:  $f_2(4) = 130$ ,  $x_2^*(4) = 3$  sugerują, że trzy punkty informacyjne powinny powstać w mieście B. Ostatni punkt przypadnie miastu A ( $f_1(1) = 60$ ,  $x_1^*(1) = 1$ ). Optymalną strategię można zapisać w następujący sposób: (1, 3, 1).

Stosując zapis tablicowy, z kolei, należy kierować się wartościami pogrubionymi. Ponieważ w sumie można uruchomić maksymalnie pięć punktów, interesować nas będzie wiersz  $x = 5$  (tablica 6.4), w którym największy zysk (205 tys. zł) pojawia się dla  $x_3 = 1$ . Aby ustalić, ile punktów należy uruchomić w miastach A i B, wystarczy spojrzeć na wyniki przedstawione w tablicy 6.3 (etap 2) i w wierszu  $x = 4$  odszukać największą wartość (130). Wartość ta znajduje się w kolumnie odpowiadającej sytuacji, w której samorząd uruchamia trzy punkty w mieście B. Dalszy sposób odczytywania rozwiązania optymalnego jest już nam znany.

<sup>10</sup> Warto również dodać, że rozpatrywanie działalności (w tym przypadku miast) w innej kolejności (np. B, C, A) nie wpływa na końcowe rozwiązanie. Takie same rezultaty otrzymamy przy każdej innej kolejności obliczeń, gdyż w zagadnieniu rozdziału zasobu optymalizacja następuje sekwencyjnie w dowolnym porządku. Właściwość ta nie jest jednak ogólna. Często charakter ograniczeń danego zadania narzuca podejście tylko w jednym lub kilku porządkach (por. [Faure, Kaufmann, 1968, s. 42]).

TABLICA 6.5  
Tabela możliwych strategii

| Liczba punktów ( $x$ ) | $k = 1$  |          | $k = 2$  |          | $k = 3$  |          |
|------------------------|----------|----------|----------|----------|----------|----------|
|                        | $x_1(x)$ | $f_1(x)$ | $x_2(x)$ | $f_2(x)$ | $x_3(x)$ | $f_3(x)$ |
| 0                      | 0        | 35       | 0        | 80       | 0        | 130      |
| 1                      | 1        | 60       | 0        | 105      | 0 ∨ 1    | 155      |
| 2                      | 1        | 60       | 1        | 115      | 1        | 180      |
| 3                      | 1        | 60       | 2        | 125      | 1        | 190      |
| 4                      | 1        | 60       | 3        | 130      | 1        | 200      |
| 5                      | 1        | 60       | 3        | 130      | 1        | 205      |

U w a g a: Symbol  $k$  oznacza numer etapu,  $x_k(x)$  — liczbę punktów uruchomionych w mieście analizowanym na  $k$ -tym etapie, a  $f_k(x)$  — maksymalny zysk z  $k$  etapów przy stanie  $x$ .

Końcowy wynik można też odczytać bezpośrednio z tabeli możliwych strategii (tablica 6.5). Stanowi ona zestawienie rozwiązań optymalnych dla różnych poziomów posiadanego zasobu ( $x$ ). Taką tabelę można uzupełnić dopiero po otrzymaniu wyników z poszczególnych etapów.

Zgodnie z zasadą optymalności Bellmana, w przypadku trzyetapowego zagadnienia maksymalizacji, odczytanie wyniku, dla danej wartości  $x$ , polega na znalezieniu odpowiadającej jej wartości  $x_2$  w kolumnie  $x_3(x)$ . Następnie należy odszukać  $x_2$  spośród wartości  $x_2(x)$ , wiedząc, że pozostała wielkość zasobu wynosi  $(x - x_2)$ . Na koniec wystarczy wyznaczyć optymalną alokację w procesie jednoetapowym przy wielkości środka wynoszącej  $(x - x_3 - x_2)$ <sup>11</sup>. Z tablicy 6.5 wynika, że gdyby samorząd mógł uruchomić tylko jeden punkt informacyjny, to w mieście C powinien działać co najwyżej jeden taki punkt, a zatem przy  $x = 1$  istnieją przynajmniej dwie strategie optymalne:  $(x_1, x_2, 0)$  oraz  $(x_1, x_2, 1)$ . Jeżeli przyjmiemy, że  $x_3 = 0$ , to  $x_2 = 0$  oraz  $x_1 = 1$ . Jeżeli natomiast  $x_3 = 1$ , to pozostałe zmienne będą równe  $x_2 = x_1 = 0$ . Miesięczny zysk będzie maksymalny, gdy samorząd uruchomi punkt informacyjny w mieście A lub C. W przypadku realizacji strategii  $(1, 0, 0)$ , zysk obliczamy w następujący sposób:

$$R(1, 0, 0) = g_1(1) + g_2(0) + g_3(0) = 60 + 45 + 50 = 155.$$

Strategia  $(0, 0, 1)$  wiąże się oczywiście z takim samym zyskiem:

$$R(0, 0, 1) = g_1(0) + g_2(0) + g_3(1) = 35 + 45 + 75 = 155.$$

Omówiony przykład dotyczył sytuacji, gdy maksymalna ilość zasobu, który należy rozdzielić, jest z góry narzucona, co wynika z równania 6.2. Może się jednak

<sup>11</sup> Wyrażenia  $(x - x_3)$  oraz  $(x - x_3 - x_2)$  opisują zależność między stanem procesu na początku następnego etapu a stanem procesu na początku etapu bieżącego i podjętą decyzją. Nazywane są one funkcjami przejścia.

okazać, że decydent będzie zainteresowany ustaleniem optymalnej strategii dla dowolnej ilości środka. Na podstawie wymiarów tablicy 6.2 wnioskujemy, że liczba punktów informacji turystycznej uruchomionych w sumie we wszystkich trzech miastach nie może przekroczyć dziewięciu, a w poszczególnych miastach mogłyby pojawić się co najwyżej trzy takie punkty. Zaprezentowana metoda programowania dynamicznego (tzw. metoda ogólna) pozwala także wyznaczyć optymalny plan alokacji, gdy wielkość zasobu nie jest z góry określona. Wystarczy tylko ustalić w etapach 1 i 2 wartości funkcji  $f_1(x)$  i  $f_2(x)$  dla  $x = 6, 7, 8, 9$ , a w etapie 3  $f_3(x)$  dla  $x = 1, 2, 3, 4, 6, 7, 8, 9$ . Strategią optymalną będzie plan spełniający warunek:

$$f_3^*(x) = \max_{x=0, 1, \dots, 9} \{f_3(x)\}.$$

Metoda ogólna nie jest skomplikowana. Jednak wraz ze wzrostem rozmiarów zadania staje się ona dość czasochłonna. Dlatego gdy jest taka możliwość, stosuje się pewne metody uproszczone, które również są procedurami dokładnymi. Okazuje się bowiem, że problemy alokacyjne, w których całkowita wielkość rozdzielonego zasobu może być dowolna, prościej jest rozwiązywać za pomocą tzw. **metody ekstremów lokalnych**. Zakłada ona wybór maksymalnej wartości funkcji  $g_j(x_j)$  oddzielnie dla każdej działalności. Wartość ta określa optymalną ilość zasobu skierowanego do poszczególnych działalności.

### PRZYKŁAD 6.2 (cd. 6.1)

Maksymalne wielkości zysków w tablicy 6.6 zostały pogrubione. W mieście A warto uruchomić jeden bądź dwa punkty informacyjne, w mieście B — trzy punkty, a w mieście C — jeden punkt. Skoro dla pierwszej działalności wyznaczono dwa ekstrema, a dla dwóch pozostałych — po jednym, to z matematycznego punktu widzenia istnieją  $2 \cdot 1 \cdot 1 = 2$  strategie optymalne:  $(1, 3, 1)$  oraz  $(2, 3, 1)$ . Pierwsza z nich zakłada uruchomienie pięciu punktów łącznie, a druga — sześciu. W obu przypadkach zyski są identyczne:  $f_3(5) = f_3(6) = 205$  tys. zł. Ponieważ szósty punkt informacyjny nie przyniesie dodatkowych zysków, można stwierdzić, że pierwszy plan jest bardziej racjonalny.

TABLICA 6.6  
Metoda ekstremów lokalnych

| Liczba punktów informacji turystycznej | Zyski (tys. zł) |           |           |
|--|-----------------|-----------|-----------|
|  | A               | B         | C         |
| 0                                      | 35              | 45        | 50        |
| 1                                      | <b>60</b>       | 55        | <b>75</b> |
| 2                                      | <b>60</b>       | 65        | 70        |
| 3                                      | 55              | <b>70</b> | 55        |

Drugą często stosowaną procedurą uproszczoną jest **metoda zysków krańcowych**<sup>12</sup>, która pozwala wyznaczyć optymalny przydział zarówno wówczas, gdy dysponujemy dowolną ilością środka, jak i wtedy, gdy zasób środka jest dany. Z tej metody można korzystać jedynie wówczas, gdy dla każdej działalności funkcja zysków krańcowych jest nierosnąca<sup>13</sup>.

### PRZYKŁAD 6.3 (cd. 6.1)

Sprawdźmy zatem, czy w rozpatrywanym zadaniu metoda ta znalazłaby zastosowanie. Dla każdej działalności obliczamy zyski krańcowe (tablica 6.7) zgodnie ze wzorem:

$$g_j(x_j) = g_j(x_j) - g_j(x_j - 1), \quad x_j \in \{d_j^- + 1, d_j^- + 2, \dots, d_j^+\}. \quad (6.8)$$

TABLICA 6.7  
Metoda zysków krańcowych

| Liczba punktów informacji turystycznej | Zyski krańcowe (tys. zł) |                 |                 |
|--|--------------------------|-----------------|-----------------|
|  | A                        | B               | C               |
| 1                                      | 25 <sup>1</sup>          | 10 <sup>1</sup> | 25 <sup>2</sup> |
| 2                                      | 0 <sup>0</sup>           | 10 <sup>1</sup> | -5              |
| 3                                      | -5                       | 5 <sup>5</sup>  | -15             |

Warunek jest spełniony. Możemy więc przystąpić do ustalenia rozwiązania optymalnego. Gdy ilość zasobu jest dowolna, należy kolejne jednostki środka przydzielać tam, gdzie zyski krańcowe są największe i tak długo, aż pojawią się ujemne wartości. Obecność zerowego zysku krańcowego przy jakiegokolwiek działości oznacza, że istnieje więcej niż jedna strategia optymalna: (1, 3, 1) oraz (2, 3, 1).

Jak już wspomniano, metoda zysków krańcowych może zastąpić metodę ogólną programowania dynamicznego również wtedy, gdy ilość środka jest ustalona. Niech  $a = 4$ . Z tablicy 6.7 wynika, że pierwszy punkt należy uruchomić w mieście A, drugi — w mieście C (lub odwrotnie), a trzeci i czwarty — w mieście B. Oto rozwiązanie optymalne:  $x_1 = 1, x_2 = 2, x_3 = 1, R(1, 2, 1) = g_1(1) + g_2(2) + g_3(1) = 60 + 65 + 75 = 200$  tys. zł.

<sup>12</sup> Gdy zadanie dotyczy maksymalizacji dochodów (minimalizacji kosztów), wówczas procedura nosi nazwę metody dochodów (kosztów) krańcowych. Metoda zysków (dochodów, kosztów) krańcowych nazywana jest również metodą zachłanną.

<sup>13</sup> W przypadku minimalizacji kosztów funkcja kosztów krańcowych powinna być niemalejąca dla każdej działalności.

Reasumując, w zagadnieniu rozdziału zasobu metodę ogólną oraz metodę zysków krańcowych stosujemy, gdy ilość środka jest z góry określona bądź dowolna, natomiast metoda ekstremów lokalnych jest przydatna tylko w tym drugim przypadku.

W przedstawionych przykładach liczbowych zadano o sprecyzowanie przedziału czasowego, którego dotyczą zarówno nakłady, jak i zyski. Gdyby jednak horyzont ekonomiczny nie został wyraźnie określony, ustalone rozwiązanie optymalne mogłoby być nie do końca jasne dla decydenta (zob. [Faure, Kaufmann, 1968]).

W rozdziale zajęliśmy się wyłącznie jednowymiarowymi procesami alokacyjnymi, czyli takimi, w których rozporządzamy tylko jednym zasobem i mamy tylko jedno ograniczenie dotyczące tego zasobu (wzór (6.2)). Zainteresowanych wielowymiarowymi procesami alokacyjnymi odsyłamy do pracy [Bellman, Dreyfus, 1967]. Ponadto należy podkreślić, że w pracy omówiono jedynie deterministyczną wersję alokacji zasobów. W literaturze znaleźć można opis modeli probabilistycznych programowania dynamicznego, mających zastosowanie m.in. w zagadnieniu rozdziału zasobu [Wagner, 1975]. Skorzystanie ze stochastycznej odmiany metody programowania dynamicznego jest konieczne wówczas, gdy zysk wynikający ze skierowania danej ilości środka do konkretnej działalności jest opisany za pomocą zmiennej losowej.

### POLECENIA I PYTANIA KONTROLNE

1. Dlaczego problemy kombinatoryczne warto rozwiązywać metodami programowania dynamicznego?
2. Kiedy mamy do czynienia z wieloetapowymi procesami decyzyjnymi? Od czego zależy liczba etapów w zagadnieniu optymalnej alokacji zasobu?
3. Jak brzmi i do czego służy zasada optymalności Bellmana?
4. Jaki cel stawiamy w zagadnieniu optymalnego rozdziału zasobu?
5. Jaka jest różnica między zbiorem stanów dopuszczalnych dla poszczególnych działalności a zbiorem stanów układu dla kolejnych etapów?
6. Jakie informacje można odczytać z tabeli możliwych strategii?
7. Wymień i porównaj metody znajdujące zastosowanie w zagadnieniu optymalnego rozdziału zasobu.

### BIBLIOGRAFIA

- Bellman R.E., Dreyfus S.E., 1958, *On the Computational Solution of Dynamic Programming Processes — I: On a Tactical Air Warfare Model of Mengel*, „Operations Research”, vol. 6, INFORMS, USA.
- Bellman R.E., Dreyfus S.E., 1967, *Programowanie dynamiczne (zastosowanie)*, PWE, Warszawa.

- Faure R., Kaufmann A., 1968, *Badania operacyjne na co dzień*, PWN, Warszawa.
- Hindelang T.J., Muth J.F., 1979, *A Dynamic Programming Algorithm for Decision CPM Networks*, „Operations Research”, vol. 27(2), INFORMS, USA.
- Kukuła K., Jędrzejczyk Z., Skrzypek J., Walkosz A., 1996, *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa.
- Lesz M., 1968, *Techniczno-ekonomiczne zastosowania metod programowania dynamicznego*, PWE, Warszawa.
- Robinson D.R., 1975, *A Dynamic Programming Solution to Cost-time Tradeoff for CPM*, „Management Science”, vol. 22, INFORMS, USA.
- Trzaskalik T., 2003, *Wprowadzenie do badań operacyjnych z komputerem*, PWE, Warszawa.
- Wagner H.M., 1975, *Principles of Operations Research with Applications to Managerial Decisions*, second edition, Prentice-Hall, Englewood Cliffs, New Jersey.

## ROZDZIAŁ 7

ANALIZA SIECIOWA  
PRZEDSIĘWZIĘĆ

## 7.1. Sieciowe modele przedsięwzięć

Przez **przedsięwzięcie** rozumie się zorganizowane działanie ludzkie, z wyróżnionym początkiem i końcem, zmierzające do osiągnięcia określonego celu i zawarte w skończonym przedziale czasu. Realizacja przedsięwzięcia (projektu) wiąże się zazwyczaj z koniecznością wykonania więcej niż jednej czynności. Niektóre z nich są wykonywane równolegle, pozostałe natomiast mogą rozpocząć się dopiero wówczas, gdy zakończone zostaną inne działania. Przykładem przedsięwzięcia jest proces rekrutacji kandydatów na studia. Wymaga on sprawnego działania uczelni oraz potencjalnych studentów. Rekrutacja powinna przebiegać w ściśle określonym czasie. Rozpoczyna się ona po przeprowadzeniu egzaminów maturalnych, a kończy na tyle wcześniej, aby kandydat przyjęty na studia mógł je rozpocząć wraz z początkiem roku akademickiego. Część prac realizowanych przez komisję rekrutacyjną może przebiegać równolegle (niezależnie), a pozostałe czynności komisja powinna wykonać w odpowiedniej, zaplanowanej kolejności, czyli sekwencyjnie (np. ogłoszenie listy rankingowej ma sens dopiero po złożeniu przez kandydatów formularza zgłoszeniowego).

Planowanie i kontrolę realizacji przedsięwzięć wspomóc mogą sieci, które stanowią graficzną reprezentację czynności wchodzących w skład projektu oraz zależności strukturalnych między nimi, czyli kolejności ich wykonania zgodnie z technologicznymi wymaganiami procesu pracy<sup>1</sup>. Sieć tworzy uporządkowany zbiór węzłów (wierzchołków) oraz luków. Dwoma najczęściej spotykanymi technikami sieciowymi są:

<sup>1</sup> Metody planowania sieciowego znalazły początkowo zastosowanie w amerykańskim przemyśle zbrojeniowym do koordynacji dostaw poszczególnych przedsiębiorstw wykonujących części składowe raket „Polaris” oraz sprzętu wojennego.

- 1) AOA (*activities-on-arcs*), gdy czynności reprezentowane są za pomocą luków, a zdarzenia za pomocą węzłów<sup>2</sup> — tak skonstruowana sieć nosi nazwę grafu czynności;
- 2) AON (*activities-on-nodes*), gdy węzły ukazują czynności, a luki relacje poprzedzania<sup>3</sup> — w tym przypadku mamy do czynienia z grafem stanów (zdarzeń).

W pracy wykorzystamy technikę AOA.

#### DEFINICJA 7.1

Graf nosi nazwę **sieci czynności**, jeżeli jest grafem:

- skierowanym (składającym się z połączeń o określonej orientacji),
- antysymetrycznym (złożonym z dróg umożliwiających przejście tylko w jednym kierunku: od węzła początkowego do końcowego),
- spójnym (łączącym wszystkie węzły i luki),
- acyklicznym (pozbawionym obwodu zamkniętego wewnątrz sieci<sup>4</sup>).

Budowa sieci typu AOA wymaga znajomości wszystkich czynności wchodzących w skład przedsięwzięcia oraz zbioru czynności bezpośrednio je poprzedzających. Przy konstrukcji sieci należy pamiętać o następujących zasadach:

1. W celu zachowania przejrzystości sieci pełne nazwy czynności (np. zamieszczenie informacji o rekrutacji na studia stacjonarne na stronie internetowej uczelni) można zastąpić kolejnymi literami alfabetu lub symbolem  $\langle i, j \rangle$ , gdzie  $i$  to numer zdarzenia rozpoczynającego daną czynność, a  $j$  numer zdarzenia kończącego tę czynność.
2. Sieć zawiera dokładnie jedno zdarzenie początkowe i jedno zdarzenie końcowe. Zdarzenie początkowe grafu nie jest poprzedzone żadną czynnością, a w zdarzeniu końcowym nie rozpoczyna się żadna czynność. Oznacza to, że wszystkie zdarzenia w sieci, z wyjątkiem początkowego i końcowego, powinny być początkiem i końcem co najmniej jednej czynności.
3. Do zdarzenia rozpoczynającego daną czynność powinny dochodzić jedynie te czynności, które bezpośrednio ją poprzedzają.
4. Dwie (lub więcej) czynności mogą rozpocząć się od tego samego zdarzenia tylko wówczas, gdy mają identyczny zbiór czynności poprzedzających.
5. Dwa dowolnie wybrane węzły może łączyć bezpośrednio co najwyżej jeden luk

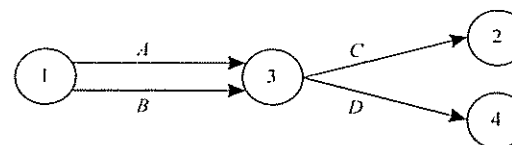
<sup>2</sup> W przypadku techniki AOA zdarzenie utożsamiane jest z wystąpieniem pewnego zdefiniowanego stanu, natomiast czynność jest zadaniem, które musi być wykonane między dwoma zdarzeniami.

<sup>3</sup> Rysowanie sieci typu AON jest prostsze niż technika AOA ze względu na brak czynności pozornych, o których będzie mowa w dalszej części rozdziału. Jednak znacznie łatwiej przeprowadzić obliczenia związane z analizą czasową przedsięwzięcia dla sieci typu AOA (podrozdział 7.2).

<sup>4</sup> Pojawienie się pętli lub cyklu oznaczałoby, że czynności nie są wykonywane jednorazowo.

6. Każdy wierzchołek otrzymuje odrębny numer, stanowiący liczbę całkowitą (1, 2, ...,  $n$ ), w taki sposób, że zdarzenie początkowe przyjmuje numer 1, końcowe —  $n$  i jeżeli  $\langle i, j \rangle$  oznacza czynność, to  $i < j$  (zdarzenie będące początkiem czynności powinno mieć numer mniejszy niż zdarzenie, które kończy tę czynność).
7. Należy w miarę możliwości unikać sytuacji, w której luki przecinają się ze sobą.

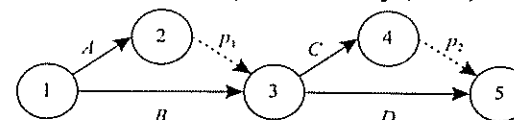
RYSUNEK 7.1  
Niepoprawny wykres sieciowy



Na rysunku 7.1 zilustrowano przykład sieci typu AOA obciążonej trzema błędami. Naruszone zostały reguły opisane w punkcie drugim, piątym oraz szóstym. Po pierwsze, sieć ma dwa zdarzenia końcowe (2 i 4). Po drugie, istnieją dwie czynności (A i B), które mają wspólny początek i koniec. Każdą z nich można opisać symbolem  $\langle 1, 3 \rangle$ . Nie są one więc rozróżniane przez komputer stosujący nazewnictwo czynności oparte na numeracji węzłów. Po trzecie, czynność  $\langle 3, 2 \rangle$  nie spełnia warunku  $i < j$ . Aby usunąć wymienione błędy, konieczne jest wprowadzenie dwóch **czynności pozornych (fikcyjnych)**, oznaczonych w sieci przerywanymi strzałkami. Czynność fikcyjna nie wiąże się z realizacją jakiegokolwiek działania. Dzięki niej jest możliwe uwzględnienie w grafie wymaganych relacji poprzedzania oraz przestrzeganie zasad drugiej, trzeciej, czwartej i piątej.

Rysunek 7.2 przedstawia poprawną wersję sieci. Wprowadzono dodatkowe zdarzenie (obecnie z numerem 2) stanowiące koniec czynności A. Zakończenie czynności B ma miejsce w zdarzeniu 3, zatem zlikwidowano obecność pary zdarzeń połączonych bezpośrednio dwiema czynnościami. Z rysunku 7.1 wynika, że czynności C i D występują dopiero po zakończeniu czynności A i B. Taką relację następstwa mogła zagwarantować jedynie czynność pozorna  $p_1$ . Narysowanie tej czynności przed, a nie po czynności A byłoby również poprawne. Dołączenie drugiej takiej czynności ( $p_2$ ) pozwoliło zachować w sieci tylko jedno zdarzenie

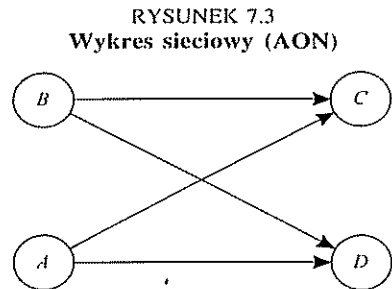
RYSUNEK 7.2  
Poprawiony wykres sieciowy (AOA)





końcowe (czyli takie, do którego strzałki jedynie dochodzą). Czynność  $p_2$  mogłaby również dobrze mieć przeciwny zwrot, ponieważ w przypadku drugiej czynności fikcyjnej chodziło wyłącznie o „doprowadzenie” zbędnego zdarzenia końcowego do innego, już istniejącego, zdarzenia końcowego. Wówczas ostatecznym wierzchołkiem końcowym byłoby zdarzenie kończące czynność  $C$ , a numery dwóch ostatnich zdarzeń musiałyby być przestawione<sup>5</sup>. Zauważmy, że luków  $C$  oraz  $D$  można było bezpośrednio (tzn. bez czynności pozornej) doprowadzić do zdarzenia końcowego, gdyż zasada opisana w punkcie piątym znów zostałaby naruszona. Trzeci wymieniony zarzut dotyczył numeracji zdarzeń, które tym razem ponumerowano zgodnie z instrukcją.

Strukturę grafu sieciowego narysowanego dla tego samego przedsięwzięcia, lecz techniką AON, prezentujemy na rysunku 7.3. Obok modeli sieciowych będących jednym ze sposobów prezentacji struktury projektu, znane są także wykresy Gantta (podrozdział 7.2) oraz **macierze powiązań (macierze incydentów)**, w których „1” oznacza połączenie między poszczególnymi wierzchołkami grafu narysowanego techniką AOA, a „0” — brak połączenia. W tabelicy 7.1 przedstawiono macierz powiązań dla przedsięwzięcia opisanego na rysunku 7.2.



TABLICA 7.1  
Macierz powiązań

|   | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 |
| 3 | 0 | 0 | 0 | 1 | 1 |
| 4 | 0 | 0 | 0 | 0 | 1 |
| 5 | 0 | 0 | 0 | 0 | 0 |

W przypadku bardziej złożonych przedsięwzięć omówiony niżej algorytm generowania sieci typu AOA może okazać się bardzo przydatny.

**Algorytm 7.1. Generowanie sieci typu AOA**

Krok 1. Wypisać zbiory czynności poprzedzających, ich niepuste części wspólne oraz zbiór wszystkich tych czynności, które nie wystąpiły w żadnym zbiorze, a następnie każdemu utworzonemu zbiorowi przypisać wierzchołek sieci.

<sup>5</sup> Czynności fikcyjne odgrywają niezwykle istotną rolę, lecz zaleca się, aby sieć zawierała jak najmniej takich czynności. Nadmierna liczba czynności pozornych przedłuża czas obliczeń (zob. podrozdział 7.2 i 7.3).

- Krok 2. Narysować luki odpowiadające czynnościom tak, aby rozpoczynały się od zbioru ich poprzedników, a kończyły w najmniejszym zbiorze zawierającym nazwę danej czynności. Czynności bez poprzedników zaczynają się od zbioru pustego.
- Krok 3. Jeżeli dany niepusty zbiór zawiera się w innym zbiorze, wprowadzić czynność pozorną zwróconą w kierunku bardziej licznego zbioru.
- Krok 4. Jeżeli dwa zbiory połączone są więcej niż jednym lukiem, dołączyć czynność fikcyjną. Ponumerować węzły tak, aby dla każdej czynności zdarzenie rozpoczynające miało numer niższy niż kończące. Liczba wszystkich zbiorów wyznacza liczbę zdarzeń.

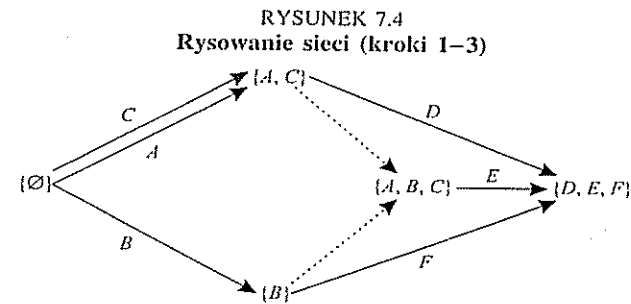
**PRZYKŁAD 7.1**

Działanie algorytmu zilustrujemy przykładem. Dane o pewnym projekcie zawiera tabela 7.2.

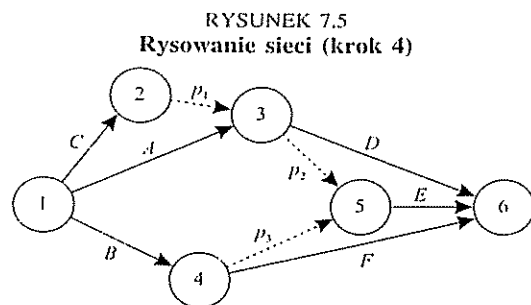
TABLICA 7.2  
Czynności wchodzące w skład przedsięwzięcia

| Czynności               | A | B | C | D    | E       | F |
|-------------------------|---|---|---|------|---------|---|
| Czynności poprzedzające | — | — | — | A, C | A, B, C | B |

- Krok 1. Oto zbiory poprzedników:  $\{A, C\}$ ,  $\{A, B, C\}$ ,  $\{B\}$ . Częścią wspólną zbiorów  $\{A, C\}$  i  $\{A, B, C\}$  jest zbiór  $\{A, C\}$ , ale ten już został wymieniony. Dla pozostałych par zbiorów część wspólna jest pusta bądź już została uwzględniona. Zbiór czynności, które nie pojawiły się w żadnym z podanych zbiorów, ma postać  $\{D, E, F\}$ .
- Krok 2 i 3. Rysunek 7.4 przedstawia luki odpowiadające czynnościom  $A-F$  oraz dołączone czynności fikcyjne.



Krok 4. Ze względu na pojawienie się dwóch luków o wspólnych zbiorach początkowym i końcowym, dodajemy kolejną czynność pozorną. Zbiory zastępujemy odpowiednio ponumerowanymi węzłami (rysunek 7.5).



## 7.2. Metoda ścieżki krytycznej (CPM)

Na podstawie modelu sieciowego obrazującego strukturę przedsięwzięcia oraz oszacowanych normalnych (nominalnych) czasów trwania czynności wchodzących w jego skład można przeprowadzić analizę czasową, która udzieli odpowiedzi na następujące pytania:

- 1) kiedy najwcześniej rozpatrywane przedsięwzięcie zostanie zrealizowane?
- 2) opóźnienie których czynności przesunie moment zakończenia inwestycji?
- 3) które działania można wykonać wolniej bądź rozpocząć później, nie opóźniając przy tym realizacji całego projektu?

W przypadku rekrutacji kandydatów na studia (zob. podrozdział 7.1), znajomość odpowiedzi na te pytania pozwoli komisji rekrutacyjnej rozpocząć prace na tyle wcześniej, aby termin wywieszenia listy osób przyjętych na uczelnię został dotrzymany.

**Metoda ścieżki krytycznej** (*critical path method*, CPM), wspomaga właśnie analizę czasową przedsięwzięcia. Metoda ta została wprowadzona w Stanach Zjednoczonych w latach 1956–1957. Stanowi ona wynik współpracy koncernu chemicznego du Pont de Nemours (M.R. Walker) oraz koncernu fabryk elektronicznych maszyn matematycznych Remington Rand Univac (J.E. Kelley). Stosowanie CPM pozwala rozpoznać „wąskie gardła” odpowiednio wcześniej, przyczyniając się tym samym do ograniczenia przestojów i do zwiększenia oszczędności. Z CPM można korzystać wówczas, gdy wszystkie czynności wchodzące w skład planowanego projektu są rzeczywiście realizowane, a ich czasy trwania mają charakter deterministyczny.

Omówienie procedury opartej na ścieżce krytycznej wymaga jednoczesnego wprowadzenia kilku istotnych pojęć.

### DEFINICJA 7.2

Ścieżką jest każdy ciąg kolejno następujących po sobie czynności.

### DEFINICJA 7.3

Ścieżka pełna to taka ścieżka, której pierwsza czynność zaczyna się w zdarzeniu początkowym, a ostatnia kończy się w zdarzeniu końcowym sieci.

### DEFINICJA 7.4

Czas trwania czynności to nakład czasu potrzebny na jej realizację<sup>6</sup>.

Znając czasy trwania poszczególnych czynności, możemy wyznaczyć **najkrótszy czas realizacji przedsięwzięcia** ( $T^*$ ). W tym celu wystarczy zastosować jedną z następujących metod:

#### Metoda I

- wypisać wszystkie ścieżki pełne,
- obliczyć czas trwania (czas przejścia) każdej ścieżki przez dodanie czasów trwania wszystkich czynności na niej leżących,
- wybrać tę ścieżkę (ścieżki), która charakteryzuje się najdłuższym czasem trwania.

Jest rzeczą oczywistą, że przedsięwzięcie można uznać za zakończone w momencie, w którym dobiegła końca realizacja wszystkich czynności wchodzących w jego skład i, właśnie dlatego, najkrótszy czas wykonania inwestycji jest zdeterminowany przez czas trwania najdłuższej, a nie najkrótszej ścieżki. Przy okazji warto dodać, że w omawianym zagadnieniu najdłuższa ścieżka wcale nie musi zawierać największej liczby luków!

<sup>6</sup> Zakładamy, że czas trwania czynności jest możliwy do ustalenia na podstawie norm pracy, cykli normatywnych, długoletnich doświadczeń itp. Czas trwania jest wyrażany w neutralnych (kolejnych) lub kalendarzowych jednostkach czasu. W pierwszym przypadku terminy początku i końca trwania czynności są określane w odniesieniu do terminu rozpoczęcia projektu, tj. w datach względnych. Ujęcie kalendarzowe oznacza podanie konkretnej (kalendarzowej) daty realizacji działania. Naturalne jednostki czasu powinny być dla całego planu jednakoowe (np. roboczogodziny, dni robocze). Ich wielkość zależy m.in. od charakteru przedsięwzięcia i przewidywanego czasu jego trwania.

Metoda II

Gdy liczba ścieżek w grafie jest dość duża, ustalenie najkrótszego czasu realizacji przedsięwzięcia na podstawie **najwcześniejszych możliwych** ( $t'_i$ ) i **najpóźniejszych dopuszczalnych** ( $t''_i$ ) momentów zajścia zdarzeń, może być mało czasochłonne:

$$t'_i = \max_k (t'_k + t_{ki}) \quad i = 2, \dots, n, \tag{7.1}$$

gdzie:

$k$  — zdarzenie początkowe czynności, które kończą się zdarzeniem  $i$ ,

$t_{ki}$  — czas trwania czynności  $\langle k, i \rangle$ .

$$t''_i = \max_k (t''_k - t_{ik}), \quad i = n - 1, \dots, 1, \tag{7.2}$$

gdzie:

$k$  — zdarzenie końcowe czynności rozpoczynających się zdarzeniem  $i$ ,

$t_{ik}$  — czas trwania czynności  $\langle i, k \rangle$ .

Moment  $t'_i$  to moment zajścia  $i$ -tego zdarzenia, w którym najwcześniej ukończone zostaną wszystkie wchodzące do niego czynności. Wartości  $t'_i$  wyznaczamy w sieci metodą „krok do przodu”, rozpoczynając od zdarzenia początkowego, dla którego  $t'_1 = 0$ , a skończywszy na zdarzeniu  $n$ . Najwcześniejsze możliwe momenty zajścia zdarzeń są podawane po nawiasach otwierających. Z kolei  $t''_i$  określa moment, w którym najpóźniej ukończone być muszą wszystkie czynności wchodzące do  $i$ -tego zdarzenia, aby nie opóźnił się moment realizacji całego przedsięwzięcia w stosunku do najkrótszego możliwego czasu jego realizacji ( $T^*$ ). Wartości  $t''_i$  wyznaczamy w sieci metodą „krok do tyłu”, rozpoczynając od zdarzenia końcowego, dla którego  $t''_n = t'_n$ , i kończąc na pierwszym zdarzeniu. Najpóźniejsze dopuszczalne momenty zajścia zdarzeń są podawane przed nawiasami zamykającymi. Wystąpienie zdarzenia końcowego jest równoznaczne z zakończeniem przedsięwzięcia, stąd:  $T^* = t'_n = t''_n$ .

Metoda III

Najkrótszy czas realizacji przedsięwzięcia można też ustalić, rozwiązując odpowiednie zadanie programowania liniowego:

$$t_n \rightarrow \min, \tag{7.3}$$

$$t_j - t_i \geq t_{ij}, \quad \langle i, j \rangle \in U, \quad i < j, \tag{7.4}$$

$$t_i \geq 0, \quad i = 1, 2, \dots, n, \tag{7.5}$$

gdzie:

$t_i$  — najwcześniejszy możliwy moment rozpoczęcia czynności  $\langle i, j \rangle$ ,

$t_j$  — najwcześniejszy możliwy moment zakończenia czynności  $\langle i, j \rangle$ ,

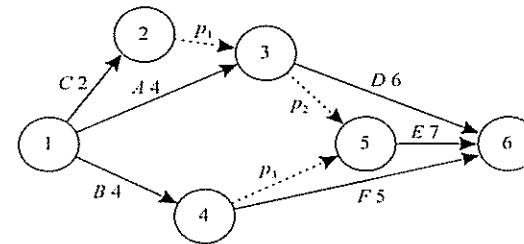
- $t_n$  — najwcześniejszy moment zajścia  $n$ -tego zdarzenia, czyli  $T^*$ ,
- $t_{ij}$  — czas trwania czynności  $\langle i, j \rangle$  (parametr zadania),
- $U$  — zbiór czynności danej sieci.

PRZYKŁAD 7.2

Tym razem wprowadźmy dodatkowo czasy trwania poszczególnych czynności (w dniach). W przypadku czynności fikcyjnych czas ten będzie równy zero (rysunek 7.6). Z tablicy 7.3 wynika, że w grafie znajdują się dwie ścieżki o najdłuższym czasie trwania:  $(A-p_2-E)$  oraz  $(B-p_3-E)$ . Najkrótszy czas realizacji przedsięwzięcia wynosi 11 dni.

RYSUNEK 7.6

Sieć czynności wraz z ich czasem trwania



TABLICA 7.3  
Ścieżki pełne sieci

| Ścieżki      | C-p <sub>1</sub> -D | C-p <sub>1</sub> -p <sub>2</sub> -E | A-D | A-p <sub>2</sub> -E | B-p <sub>3</sub> -E | B-F |
|--------------|---------------------|-------------------------------------|-----|---------------------|---------------------|-----|
| Czas trwania | 8                   | 9                                   | 10  | 11                  | 11                  | 9   |

Chociaż moment zakończenia inwestycji jest już znany, ustalenie momentów zajścia zdarzeń ( $t'_i$  i  $t''_i$ ) ma nadal istotne znaczenie, gdyż to właśnie znajomość tych wartości pozwoli przeprowadzić dogłębną analizę czasową przedsięwzięcia.

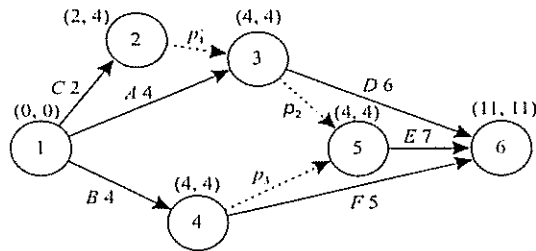
W tablicy 7.4 przedstawiono sposób i kolejność obliczania poszczególnych momentów, a na rysunku 7.7 zestawiono ostateczne wyniki. Przy wyznaczaniu  $t'_i$  ( $t''_i$ ) dla danego węzła należy wziąć pod uwagę wszystkie łuki do niego wchodzące (z niego wychodzące)<sup>7</sup>. Na podstawie najwcześniejszego możliwego momentu zajścia ostatniego zdarzenia również dochodzimy do wniosku, że realizacja przed-

<sup>7</sup> Dodajmy, że niepoprawna numeracja węzłów (zob. podrozdział 7.1) może skutkować błędnie obliczonymi momentami zdarzeń. Jeżeli  $t'_i \neq 0$  lub/i w przypadku chociażby jednego wierzchołka  $t'_i > t''_i$ , to wartości w nawiasach powinny być ponownie obliczone w celu wykrycia błędu.

TABLICA 7.4  
Wyznaczanie momentów zajścia zdarzeń

| Etap 1  | $t_1'$         | $t_2'$           | $t_3'$   | $t_4'$   | $t_5'$                                   | $t_6'$   |
|---------|----------------|------------------|--|--|--|--|
| Wzór    | $t_1' = 0$     | $t_1' + t_{12}$  | $\max\{(t_2' + t_{23})$<br>$(t_1' + t_{13})\}$   | $t_1' + t_{14}$                                  | $\max\{(t_3' + t_{35})(t_4' + t_{45})\}$ | $\max\{(t_5' + t_{56})$<br>$(t_4' + t_{46})\}$   |
| Wartość | 0              | 2                | $\max\{2, 4\} = 4$                               | 4  | $\max\{4, 4\} = 4$                       | $\max\{10, 11, 9\} = 11$                         |
| Etap 2  | $t_1''$        | $t_2''$          | $t_3''$  | $t_4''$  | $t_5''$                                  | $t_6''$  |
| Wzór    | $t_6'' = t_6'$ | $t_6'' - t_{56}$ | $\min\{(t_3'' - t_{23})$<br>$(t_6'' - t_{46})\}$ | $\min\{(t_6'' - t_{56})$<br>$(t_5'' - t_{45})\}$ | $t_5'' - t_{23}$                         | $\min\{(t_5'' - t_{45})$<br>$(t_6'' - t_{46})\}$ |
| Wartość | 11             | 4                | $\min\{4, 6\} = 4$                               | $\min\{5, 4\} = 4$                               | 4  | $\min\{2, 0, 0\} = 0$                            |

RYSUNEK 7.7  
Sieć czynności wraz z ustalonymi momentami zajścia zdarzeń



siewzięcia wymaga przynajmniej 11 dni, przy założeniu, iż wszystkie czynności zostaną wykonane zgodnie z planem<sup>8</sup>.

Obliczone momenty pozwalają także wyznaczyć różne zapasy czasu dla poszczególnych działań. **Calkowity zapas czasu**, określony za pomocą równania (7.6), oznacza, o ile czas trwania czynności  $\langle i, j \rangle$  może być wydłużony bądź moment jej rozpoczęcia opóźniony, aby najkrótszy czas realizacji przedsięwzięcia się nie zmienił:

$$z_{ij}^c = t_j'' - t_i' - t_{ij} \quad (7.6)$$

**Zapas swobodny** danej czynności (wzór (7.7)) może być wykorzystany bez wpływu na zapasy czynności tej samej ścieżki:

$$z_{ij}^s = t_j' - t_i' - t_{ij} \quad (7.7)$$

<sup>8</sup> Oczywiście w praktyce planiści momentom zajścia kolejnych zdarzeń przypisują konkretne daty kalendarzowe (np. 15 czerwca 2009 r.).

Wykorzystanie **zapasu warunkowego** danego działania (wzór (7.8)) nie zmniejsza zapasów czasu czynności poprzednich znajdujących się na tej samej ścieżce:

$$z_{ij}^w = t_j'' - t_i'' - t_{ij} \quad (7.8)$$

**Niezależny zapas czasu** (wzór (7.9)) informuje o rezerwie, której wykorzystanie nie ma wpływu na zapas jakiegokolwiek innej czynności<sup>9</sup>:

$$z_{ij}^n = \max\{0, (t_j' - t_i'' - t_{ij})\} \quad (7.9)$$

Czynność charakteryzująca się zerowym zapasem całkowitym przyjmuje miano **czynności krytycznej**, gdyż jakiegokolwiek wydłużenie jej czasu trwania opóźnia moment zakończenia inwestycji. Pozostałe zapasy czasu czynności krytycznej są także zerowe. Ciąg czynności krytycznych, rozpoczynający się zdarzeniem początkowym, a kończący w zdarzeniu końcowym, tworzy **ścieżkę krytyczną**. Skoro ścieżka ta nie ma żadnej rezerwy, oznacza to, że nie istnieje w grafie droga dłużej od niej trwająca. Czas przejścia ścieżki krytycznej jest najdłuższy i decyduje o terminie końcowym całego przedsięwzięcia. Jeżeli czas dyrektywny (o którym będzie mowa w podrozdziale 7.4) jest różny od najkrótszego czasu realizacji przedsięwzięcia ustalonego na podstawie najwcześniejszego momentu zajścia zdarzenia końcowego, to całkowity zapas czasu czynności leżących na najdłuższej ścieżce w sieci nie jest zerowy, lecz nadal mniejszy niż rezerwy pozostałych czynności. Dlatego ogólniejsza definicja ścieżki krytycznej brzmi następująco: jest to droga o najmniejszym całkowitym zapasie czasu. CPM polega właśnie na zidentyfikowaniu ścieżek krytycznych w celu:

- 1) ustalenia najkrótszego czasu wykonania projektu;
- 2) wyznaczenia rezerw dla czynności leżących na pozostałych ścieżkach;
- 3) umożliwienia kontroli realizacji tych czynności, od których bezpośrednio zależy moment zakończenia inwestycji.

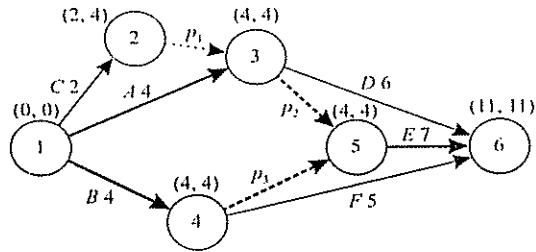
Zapasy czynności zawiera tablica 7.5. Warto zaznaczyć, że czynność pozorna, mimo zerowego czasu trwania, niekoniecznie musi być czynnością krytyczną. Jej status zależy od całkowitego zapasu czasu. Na rysunku 7.8 ścieżki krytyczne zostały pogrubione.

TABLICA 7.5  
Zapasy czasu czynności

| Czynności        | A | B | C | D | E | F | $p_1$ | $p_2$ | $p_3$ |
|------------------|---|---|---|---|---|---|-------|-------|-------|
| Zapas całkowity  | 0 | 0 | 2 | 1 | 0 | 2 | 2     | 0     | 0     |
| Zapas swobodny   | 0 | 0 | 0 | 1 | 0 | 2 | 2     | 0     | 0     |
| Zapas warunkowy  | 0 | 0 | 2 | 1 | 0 | 2 | 0     | 0     | 0     |
| Zapas niezależny | 0 | 0 | 0 | 1 | 0 | 2 | 0     | 0     | 0     |

<sup>9</sup> Zapas niezależny jest częścią składową zapasu warunkowego, a więc jeżeli  $z_{ij}^w = 0$ , to na pewno  $z_{ij}^n = 0$ .

RYSUNEK 7.8  
Sieć czynności z wyróżnionymi ścieżkami krytycznymi



W analizie czasowej przedsięwzięć największą rolę odgrywają całkowite zapasy czasu. Poniżej zamieszczono dwie istotne informacje o tej kategorii zapasów:

1. Znajomość całkowitego zapasu czasu danej czynności ma znaczenie wówczas, gdy pragniemy zbadać wpływ wydłużenia jej czasu trwania na moment zakończenia inwestycji. W przypadku przyspieszenia wybranego działania ustaleniu nowego czasu realizacji przedsięwzięcia na podstawie całkowitego zapasu skróconej czynności jest niemożliwe. Wiemy, że np. wydłużenie czasu trwania czynności C o 5 dni spowoduje opóźnienie realizacji projektu o 3 dni ( $\Delta T^* = \Delta t_{12} - z_{12}^* = 5 - 2 = 3$ ,  $T^* = 14$  dni). Nie jesteśmy natomiast w stanie ocenić skutków skrócenia czasu trwania czynności E o 3 dni, biorąc pod uwagę jej całkowity запас czasu. Można by wysunąć przypuszczenie, że skrócenie czynności E należy do obu (czyli wszystkich) istniejących, najdłuższych ścieżek w sieci, to jej przyspieszenie pociągnie za sobą skrócenie tych dróg o 3 dni, a tym samym zakończenie całego przedsięwzięcia o 3 dni wcześniej. Nie jest to jednak prawda, ponieważ najkrótszy czas realizacji inwestycji wyznaczany jest zawsze przez długość ścieżki krytycznej, a po skróceniu czynności E, w sieci pojawi się jedna zupełnie nowa ścieżka krytyczna A-D, trwająca 10 dni.
2. Wydłużenie czasu trwania czynności krytycznej o  $k$  jednostek powoduje opóźnienie realizacji inwestycji o tyle samo jednostek. Wydłużenie czasu trwania czynności niekrytycznej  $\langle i, j \rangle$  o  $k$  jednostek powoduje opóźnienie momentu zakończenia przedsięwzięcia o  $\max\{0, k - z_{ij}^*\}$  jednostek.

Oprócz zapasów ustalanych dla czynności, można również wyznaczyć luz dla zdarzeń zgodnie z formułą:

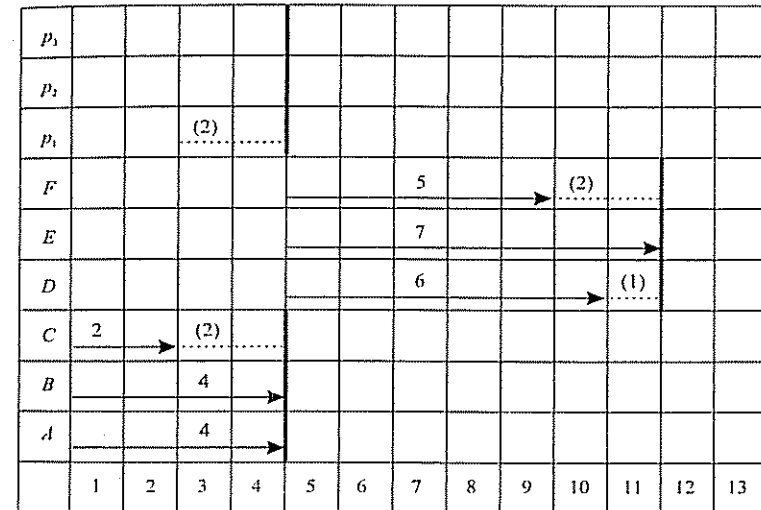
$$L_i = t_i'' - t_i' \quad (7.10)$$

Jeżeli dla  $i$ -tego zdarzenia  $L_i = 0$ , to jakiegokolwiek opóźnienie momentu jego wystąpienia powoduje wydłużenie czasu realizacji całej inwestycji. Takie zdarzenia

nosi nazwę **zdarzenia krytycznego**. Rozpatrywany model sieciowy stanowi dowód na to, że nie każdy ciąg zdarzeń krytycznych tworzy ścieżkę krytyczną (zob. [Hendrickson, Au, 1989; Idźkiewicz, 1967]). Drogi A-D i B-F, chociaż zawierają same zdarzenia krytyczne, nie są wcale najdłuższymi ścieżkami w sieci.

Na koniec tego podrozdziału warto jeszcze zapoznać się z inną formą graficznej prezentacji struktury przedsięwzięcia. Mając dane czasy trwania poszczególnych czynności, najwcześniejsze możliwe momenty zajścia zdarzeń je rozpoczynających oraz ich całkowite zapasy czasu, możliwe jest skonstruowanie **harmonogramu Gantta** (*Gantt chart*; rysunek 7.9). W odróżnieniu od modelu sieciowego, wiąże się on ze skalą czasu. Długość łuków przedstawiających kolejne czynności jest w tym przypadku zależna od ich czasów trwania. Linia przerywana dotyczy całkowitego zapasu czasu danego działania, a początek strzałki przypada na moment  $t_i'$ . Harmonogramy mają jedną istotną przewagę nad wykresami sieciowymi, gdyż ukazują, które czynności są realizowane w danym momencie<sup>10</sup>.

RYSUNEK 7.9  
Harmonogram Gantta



<sup>10</sup> Głównym mankamentem pierwotnej wersji wykresów Gantta jest brak informacji o zależnościach, jakie istnieją między czynnościami. W późniejszych wersjach zaczęto także umieszczać informacje dotyczące ścieżki krytycznej, całkowitych zapasów czasu działań, a nawet zależności między działaniami.

### 7.3. Technika oceny i kontroli programu (PERT)

Omówiona w podrozdziale 7.2 metoda ścieżki krytycznej należy do grupy metod deterministycznych, które można stosować wówczas, gdy czasy trwania poszczególnych czynności są dokładnie znane. Zdarza się jednak, że precyzyjne ustalenie czasów realizacji działań przed rozpoczęciem inwestycji jest niemożliwe<sup>11</sup>. Wówczas w celu przeprowadzenia analizy czasowej przedsięwzięcia, należy skorzystać z metod niedeterministycznych (probabilistycznych, stochastycznych), np. z techniki oceny i kontroli programu (*program evaluation and review technique* - PERT), która traktuje czasy trwania poszczególnych czynności i momenty zajścia zdarzeń jako zmienne losowe o określonej funkcji gęstości prawdopodobieństwa. Technika ta powstała w 1958 r. w związku z programem budowy rakiety Polaris, który obejmował prace perspektywiczne, badawcze, konstrukcyjne i produkcyjne. Dzięki zastosowaniu tej metody, program wykonano w ciągu 3,5 roku, a nie — jak pierwotnie zaplanowano — w ciągu 5 lat.

Powróćmy na chwilę do przywołanego już dwukrotnie przykładu rekrutacji kandydatów na studia. Chociaż uczelnie przeprowadzają rekrutację regularnie (zazwyczaj co roku), może wystąpić taka sytuacja, że na skutek decyzji ministerialnych, zmian organizacyjnych wewnątrz szkoły bądź nowej strategii innych uczelni stanowiących dla niej konkurencję proces rekrutacyjny przebiegać będzie inaczej niż w ubiegłych latach, co utrudni dokładne ustalenie czasów trwania poszczególnych czynności. W takim przypadku PERT okazuje się niezwykle cennym narzędziem. Metoda ta pozwala odpowiedzieć m.in. na następujące pytania:

- 1) jaki jest oczekiwany czas realizacji całego projektu i jego wariancja?
- 2) jakie jest prawdopodobieństwo wykonania przedsięwzięcia w zadanym czasie?

Procedura ustalania ścieżki krytycznej na podstawie PERT jest podobna do tej, którą stosuje się przy CPM, ale tym razem wszystkie obliczenia opierają się nie na normalnym, lecz na **oczekiwanym czasie realizacji czynności** ( $t_0$ ) wyznaczonym według wzoru (7.11). Formuła oczywiście nie dotyczy czynności fikcyjnych:

$$t_0 = \frac{a + 4m + b}{6}, \quad (7.11)$$

<sup>11</sup> Zwłaszcza gdy dana czynność nie była uprzednio realizowana. Realizacja przedsięwzięć zależy m.in. od pogody, sprawnego sprzętu, dyspozycyjności pracowników, a te czynniki są zmiennie w czasie.

<sup>12</sup> Warto jednak zaznaczyć, że PERT pod względem struktury logicznej modelu sieciowego zaliczana jest do metod analizy sieciowej typu DAN (*deterministic analysis network*), ponieważ w trakcie realizacji przedsięwzięcia wszystkie czynności przedstawione w sieci są realizowane. W przypadku stochastycznej struktury logicznej tylko część czynności przedstawiona w sieci, z określonym prawdopodobieństwem większym od zera, bierze udział w realizacji przedsięwzięcia.

gdzie:

$a$  — czas optymistyczny realizacji czynności w okolicznościach sprzyjających,  
 $m$  — czas najbardziej prawdopodobny realizacji czynności w warunkach normalnych,

$b$  — czas pesymistyczny realizacji czynności w warunkach niesprzyjających.

Równanie (7.11) ma sens<sup>13</sup>, gdy  $a \leq m \leq b$ . Stopień niepewności związany z przewidywanym czasem realizacji czynności  $\langle i, j \rangle$ , a wynikający z różnicy między czasem optymistycznym i pesymistycznym, można zmierzyć wariancją:

$$v_{ij} = \left( \frac{b - a}{6} \right)^2. \quad (7.12)$$

Skorzystajmy z sieci omówionej w podrozdziale 7.2. Dla przykładowych wartości  $a$ ,  $m$  i  $b$  w tabelicy 7.6 podano oczekiwane czasy trwania czynności (w dniach) oraz wariancje (w dniach do kwadratu) i prawdopodobne odchylenia (w dniach) czasu wykonania działania od czasu oczekiwanego.

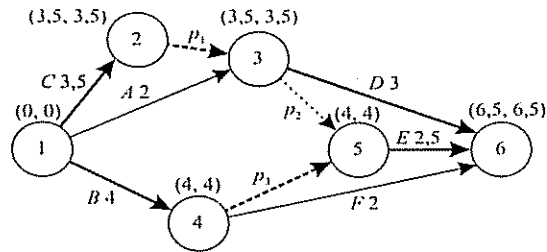
TABLICA 7.6  
Oczekiwany czas trwania oraz wariancja dla poszczególnych czynności

| Czynności | Czas optymistyczny (a) | Czas najbardziej prawdopodobny (m) | Czas pesymistyczny (b) | Czas oczekiwany ( $t_0$ ) | Wariancja (v) | Odchylenie (s) |
|-----------|------------------------|------------------------------------|------------------------|---------------------------|---------------|----------------|
| A         | 1                      | 2                                  | 3                      | 2                         | 1/9           | 1/3            |
| B         | 4                      | 4                                  | 4                      | 4                         | 0             | 0              |
| C         | 1                      | 3,25                               | 7                      | 3,5                       | 1             | 1              |
| D         | 2                      | 2,5                                | 6                      | 3                         | 4/9           | 2/3            |
| E         | 2,5                    | 2,5                                | 2,5                    | 2,5                       | 0             | 0              |
| F         | 1                      | 2                                  | 3                      | 2                         | 1/9           | 1/3            |

Ustalone na podstawie czasów  $t_0$  oczekiwane momenty zajścia zdarzeń przedstawiono na rysunku 7.10.

<sup>13</sup> Z praktyki zarządzania projektem wynika, że faktyczny czas realizacji czynności często przekracza wartość podawaną jako najbardziej prawdopodobną, gdyż ludzie mają tendencję do zbyt dużego optymizmu w swoich szacunkach. Stąd, formuła (7.11) opiera się na statystycznym rozkładzie  $\beta$ , którego cechą charakterystyczną jest odchylenie krzywej będącej graficznym obrazem funkcji gęstości w prawo. Takie odchylenie sprawia, że obliczony oczekiwany czas trwania czynności jest równy bądź większy od najbardziej prawdopodobnego, a to z kolei pozwala uzyskać czas oczekiwany na poziomie bardziej zbliżonym do rzeczywistego. Szczególnym przypadkiem asymetrycznego rozkładu  $\beta$  jest symetryczny rozkład normalny.

RYSUNEK 7.10  
Sieć czynności wraz z oczekiwanymi momentami zajścia zdarzeń



Z długości ścieżek krytycznych ( $C-p_1-D$ ,  $B-p_3-E$ ) wynika, że najwcześniejszy oczekiwany czas realizacji przedsięwzięcia wynosi  $T_0 = 6,5$  dnia. W związku z tym, że ścieżki krytyczne zostały wyróżnione, to biorąc pod uwagę oczekiwane czasy trwania czynności (obarczone odchyleniem standardowym), nie możemy mieć stuprocentowej pewności co do ich ostatecznej długości. Nie wiadomo również, czy wymienione drogi okażą się rzeczywiście najdłuższymi ścieżkami w sieci, tym bardziej, że czasy trwania poszczególnych ścieżek nie różnią się bardzo od siebie. Dlatego też powinniśmy określić możliwe odchylenie rzeczywistego momentu zakończenia inwestycji od momentu oczekiwanego. W tym celu należy najpierw obliczyć wariancję oczekiwanego czasu realizacji projektu przez zsumowanie wariancji czasów realizacji czynności wchodzących w skład ścieżki krytycznej, a następnie z otrzymanego wyniku wyciągnąć pierwiastek. Gdy graf składa się z kilku ścieżek krytycznych, wówczas prawdopodobny błąd związany z czasem wykonania przedsięwzięcia oblicza się na podstawie tej ścieżki, dla której suma wariancji jest największa:

$$s_{T_0} = \sqrt{\max\{(v_C + v_D), (v_B + v_E)\}} = \sqrt{\max\{(1 + 4/9), (0 + 0)\}} = 1,202.$$

Skoro oczekiwane czasy trwania czynności oraz całego projektu zawierają w sobie jakąś dozę niepewności, to błędem obarczone są również momenty zajścia zdarzeń. **Wariancje najwcześniejszych momentów** wyznaczamy, kumulując wariancje czasów trwania czynności począwszy od źródła grafu, dla którego  $v_1 = 0$ :

$$v_i = \max\{v_k + v_{ki}\}, \quad i = 2, \dots, n, \quad (7.13)$$

gdzie:

$k$  — zdarzenie początkowe czynności, które kończą się zdarzeniem  $i$ ,

$v_{ki}$  — wariancja czasu trwania czynności  $\langle k, i \rangle$ .

<sup>14</sup> W omawianej metodzie ustalona droga krytyczna wykazuje tylko największe prawdopodobieństwo zostania nią w stosunku do innych dróg, których szanse są mniejsze. Nie należy zatem bagatelizować ścieżek podkrytycznych (subkrytycznych), czyli ścieżek, których całkowity zapas czasu jest bliski zeru.

**Wariancje najpóźniejszych momentów** zajścia zdarzeń ustalamy, kumulując wariancje czasów trwania czynności począwszy od końca grafu, przy czym  $v_n = 0$ :

$$v_i = \max\{v_k + v_{ik}\}, \quad i = n-1, \dots, 1, \quad (7.14)$$

gdzie:

$k$  — zdarzenie końcowe czynności rozpoczynających się zdarzeniem  $i$ ,

$v_{ik}$  — wariancja czasu trwania czynności  $\langle i, k \rangle$ .

TABLICA 7.7  
Wariancje momentów zajścia zdarzeń

| Zdarzenie | 1    | 2   | 3   | 4   | 5 | 6    |
|-----------|------|-----|-----|-----|---|------|
| $v_i'$    | 0    | 1   | 1   | 0   | 1 | 13/9 |
| $v_i''$   | 13/9 | 4/9 | 4/9 | 1/9 | 0 | 0    |

Obliczone wariancje zamieszczono w tabeli 7.7. Znając czas oczekiwany realizacji projektu oraz jego odchylenie standardowe, można określić **prawdopodobieństwo zakończenia projektu w dowolnym zadanym czasie** ( $T_c$ ). Wystarczy wyznaczyć parametr  $z$ , który jest niezbędny do odnalezienia wartości tego prawdopodobieństwa w tabeli dystrybuanty rozkładu normalnego<sup>15</sup>:

$$z = \frac{T_c - T_0}{s_{T_0}}. \quad (7.15)$$

Przypuśćmy, że interesuje nas realizacja przedsięwzięcia w ciągu 8 dni:

$$z = \frac{8 - 6,5}{1,202} = 1,25.$$

Obliczonej wartości odpowiada prawdopodobieństwo na poziomie 0,8944.

Na ogół, gdy wartość prawdopodobieństwa leży poniżej 0,25, program obarczony jest znacznym ryzykiem. Wartości leżące między 0,25 a 0,6 odpowiadają ryzyku normalnemu i świadczą o dobrym bilansie zasobów. Wynik przekraczający 60% może oznaczać zaangażowanie nadmiernych środków.

Na podstawie wzoru (7.15) dochodzimy również do wniosku, że czas niezbędny do zakończenia projektu, który będzie dotrzymany z prawdopodobieństwem 0,99, wynosi 9,3 dnia. Gdy wymagany przez odbiorcę termin wykonania jest krótszy niż zezwala na to dopuszczalne ryzyko mierzone prawdopodobieństwem (np. 0,25), konieczna jest kontrola sieci i zbadanie możliwości skrócenia czasów trwania działań.

<sup>15</sup> Przyjmuje się, że szukane prawdopodobieństwo ma charakterystykę rozkładu normalnego, ponieważ faktyczne czasy realizacji odbiegają od obliczonych czasów oczekiwanych zgodnie z tym rozkładem.

Do wad analizowanej metody należy zaliczyć trudności w szacowaniu parametrów oraz probabilistyczny i subiektywny charakter otrzymanych wyników. Ponadto wybór rozkładu prawdopodobieństwa według funkcji  $\beta$  dla czasów trwania czynności został podyktowany przede wszystkim intuicją. Wreszcie, zgodnie z twierdzeniem centralnym (zob. [Bładowski, 1970]), sumowanie kolejnych wartości losowych dotyczących czynności leżących na ścieżce krytycznej w celu określenia oczekiwanego momentu zakończenia przedsięwzięcia i związanego z nim błędu wyrażonego odchyleniem standardowym jest dopuszczalne, jeżeli założymy, że czynności są niezależne od siebie. Tymczasem graf przedstawia wzajemne powiązania między działaniami, co przeczy temu upraszczającemu założeniu. Dodajmy, że rozpatrywanej technice towarzyszy ryzyko nadmiernego koncentrowania się na czynnościach krytycznych i zaniedbywania pozostałych czynności, które w rzeczywistości mogą się na tyle opóźnić, że spowodują wydłużenie czasu realizacji inwestycji. Z drugiej jednak strony PERT uwzględnia kilka wariantów czasowych, co pozwala zarządzać projektem w warunkach niepewności. PERT przypomina metodę ścieżki krytycznej (podrozdział 7.2). W obu przypadkach mamy do czynienia z siecią sformalizowaną, w której czynności są jednoznacznie zdefiniowane i niezmiennie, a zależności między nimi są stałe, co utrudnia elastyczne reagowanie na zmiany sytuacji w czasie realizacji projektu<sup>16</sup>. Stosowanie wymienionych technik zmniejsza ryzyko przekroczenia budżetu i terminów wykonania przedsięwzięcia.

## 7.4. Analiza czasowo-kosztowa

Załóżmy, że znany jest już moment zakończenia prac komisji rekrutacyjnej mającej ogłosić listę osób przyjętych na studia. Jednak z przyczyn organizacyjnych podjęta zostaje nagle decyzja o konieczności wcześniejszego zakończenia prac, co z kolei może narazić uczelnię na dodatkowe wydatki<sup>17</sup>. Uczelnia powinna skrócić czas trwania tych prac, których przyspieszenie rzeczywiście wpłynie na wcześniejsze zakończenie rekrutacji i z którymi są związane najniższe koszty skrócenia.

Okazuje się zatem, że oprócz analizy czasowej przedsięwzięcia, równie ważnym zagadnieniem jest aspekt ekonomiczny realizacji projektu i możliwości modyfikacji modelu przez kompresję sieci wynikającą ze zbyt długiego dla

<sup>16</sup> Możliwość wielowariantowego ustalenia zależności między zdarzeniami w tej samej sieci oraz swobodnego dobierania w toku realizacji projektu innych dróg postępowania niż pierwotnie ustalone dają sieci stochastyczne, np. sieć typu GERT (*graphical evaluation and review technique*).

<sup>17</sup> Wydatki te mogą pojawić się, gdy w celu szybszej realizacji prac konieczne będzie poszerzenie składu osobowego komisji rekrutacyjnej. Jeżeli dojdzie do wydłużenia czasu pracy członków komisji do późnych godzin wieczornych, uczelnia poniesie wyższe opłaty za energię elektryczną itd.

inwestora lub odbiorcy okresu wykonywania tego projektu. Potrzeba przyspieszenia realizacji przedsięwzięcia, przy jednoczesnym dążeniu do minimalizacji kosztów bezpośrednich (związanych z konkretną czynnością)<sup>18</sup>, pociąga za sobą konieczność przeprowadzenia **analizy czasowo-kosztowej**. Zakłada ona, że skróceniu czasu realizacji inwestycji towarzyszy wzrost kosztów bezpośrednich. Niezależnie od metody przyjętej przy ustalaniu momentu zakończenia inwestycji (CPM, PERT), sam proces skracania czasu realizacji przedsięwzięcia przebiega podobnie. Dlatego też w tym podrozdziale przedstawiona zostanie jedynie analiza CPM-COST. Rozpatrzmy dwa następujące przypadki:

- 1) decydent minimalizuje czas realizacji przedsięwzięcia ( $T$ ), mając na względzie dostępne środki ( $K^d$ ), które może przeznaczyć na akcelerację (*budget problem*):

$$T(\mathbf{X}) \rightarrow \min, \quad (7.16)$$

$$K(\mathbf{X}) \leq K^d. \quad (7.17)$$

gdzie  $\mathbf{X} = [x_1, x_2, \dots, x_m]^T$  to wektor czasów wykonania czynności danego przedsięwzięcia;

- 2) decydent dąży do realizacji inwestycji w czasie nie dłuższym niż zadany czas dyrektywny ( $T^d$ ), przy czym zamierza to uczynić jak najtaniej (*deadline problem*):

$$K(\mathbf{X}) \rightarrow \min, \quad (7.18)$$

$$T(\mathbf{X}) \leq T^d. \quad (7.19)$$

Przedmiotem naszych zainteresowań będą zmiany czasu trwania czynności o charakterze dyskretnym. Koszty skrócenia danego działania o każdą kolejną jednostkę mogą wzrastać z tą samą (I) bądź różną prędkością (II). W tym pierwszym przypadku koszt jednostkowy (tzw. średni gradient kosztu) ustalić można w następujący sposób:

$$s_{ij} = \frac{k_{ij}^s - k_{ij}^n}{t_{ij}^n - t_{ij}^s}, \quad (7.20)$$

gdzie:

$s_{ij}$  — wzrost kosztu związany ze skróceniem czasu czynności  $\langle i, j \rangle$  o jedną jednostkę,

$t_{ij}^n$  — normalny czas trwania czynności  $\langle i, j \rangle$ ,

$t_{ij}^s$  — graniczny czas trwania czynności  $\langle i, j \rangle$ ,

<sup>18</sup> Do kosztów bezpośrednich należą koszty robocizny, materiałów, nakłady na urządzenia produkcyjne oraz koszty skrócenia czasu czynności. Obok kosztów bezpośrednich wyróżnia się koszty pośrednie, do których zaliczyć można koszty administracyjne, podatki, kary umowne związane z niedotrzymaniem ustalonego terminu wykonania pracy. Koszty pośrednie dotyczą przedsięwzięcia jako całości.



$k_{ij}^n$  — koszt bezpośredni związany z realizacją czynności  $\langle i, j \rangle$  w czasie normalnym,  
 $k_{ij}^g$  — koszt bezpośredni związany z realizacją czynności  $\langle i, j \rangle$  w czasie granicznym.

Czas graniczny jest definiowany jako najkrótszy możliwy ze względów technicznych i technologicznych czas na wykonanie danego działania.

Modele (7.21–7.27) i (7.28–7.34) dotyczą minimalizacji czasu przy danych kosztach.

Identyczne koszty skrócenia (I):

$$t_n \rightarrow \min.$$

$$t_j \geq t_{ij}^n - y_{ij} + t_i \quad \langle i, j \rangle \in U.$$

$$t_1 = 0.$$

$$\sum_{ij=1}^m s_{ij} y_{ij} \leq K^d.$$

$$t_i, t_j \geq 0.$$

$$y_{ij} \leq P_{ij}, \quad \langle i, j \rangle \in U.$$

$$y_{ij} \in N, \quad \langle i, j \rangle \in U.$$

Różne koszty skrócenia (II):

$$t_n \rightarrow \min.$$

$$y_{ij}^1 \geq y_{ij}^2 \geq \dots \geq y_{ij}^p, \quad \langle i, j \rangle \in U.$$

$$t_j \geq t_{ij}^n - \sum_{p=1}^p y_{ij}^p + t_i, \quad \langle i, j \rangle \in U.$$

$$t_1 = 0.$$

$$\sum_{ij=1}^m \sum_{p=1}^p k_{ij}^p y_{ij}^p \leq K^d.$$

$$t_i, t_j \geq 0.$$

$$y_{ij}^p \in \{0, 1\}, \quad \langle i, j \rangle \in U.$$

W przypadku minimalizacji kosztu przy zadanym czasie należałoby rozwiązać zadanie (7.35–7.41) lub (7.42–7.48).

Identyczne koszty skrócenia (I):

$$\sum_{ij=1}^m s_{ij} y_{ij} \rightarrow \min.$$

$$t_j \geq t_{ij}^n - y_{ij} + t_i, \quad \langle i, j \rangle \in U.$$

$$t_1 = 0.$$

$$t_n \leq T^d. \quad (7.38)$$

$$t_i, t_j \geq 0. \quad (7.39)$$

$$y_{ij} \leq P_{ij}, \quad \langle i, j \rangle \in U. \quad (7.40)$$

$$y_{ij} \in N, \quad \langle i, j \rangle \in U. \quad (7.41)$$

Różne koszty skrócenia (II):

$$\sum_{ij=1}^m \sum_{p=1}^p k_{ij}^p y_{ij}^p \rightarrow \min. \quad (7.42)$$

$$y_{ij}^1 \geq y_{ij}^2 \geq \dots \geq y_{ij}^p, \quad \langle i, j \rangle \in U. \quad (7.43)$$

$$t_j \geq t_{ij}^n - \sum_{p=1}^p y_{ij}^p + t_i, \quad \langle i, j \rangle \in U. \quad (7.44)$$

$$t_1 = 0. \quad (7.45)$$

$$t_n \leq T^d. \quad (7.46)$$

$$t_i, t_j \geq 0. \quad (7.47)$$

$$y_{ij}^p \in \{0, 1\}, \quad \langle i, j \rangle \in U. \quad (7.48)$$

gdzie:

$t_i(t_j)$  — moment zajścia  $i$ -tego ( $j$ -tego) zdarzenia,

$n$  — liczba zdarzeń,

$y_{ij}$  — liczba jednostek czasu, o jaką należy skrócić czynność  $\langle i, j \rangle$ ,

$U$  — zbiór czynności danej sieci,

$m$  — liczba czynności,

$s_{ij}$  — jednostkowy koszt skrócenia czynności  $\langle i, j \rangle$ ,

$K^d$  — dostępne środki,

$P_{ij}$  — maksymalna liczba jednostek, o którą skrócenie czasu trwania czynności  $\langle i, j \rangle$  jest możliwe (różnica między czasem normalnym a czasem granicznym),

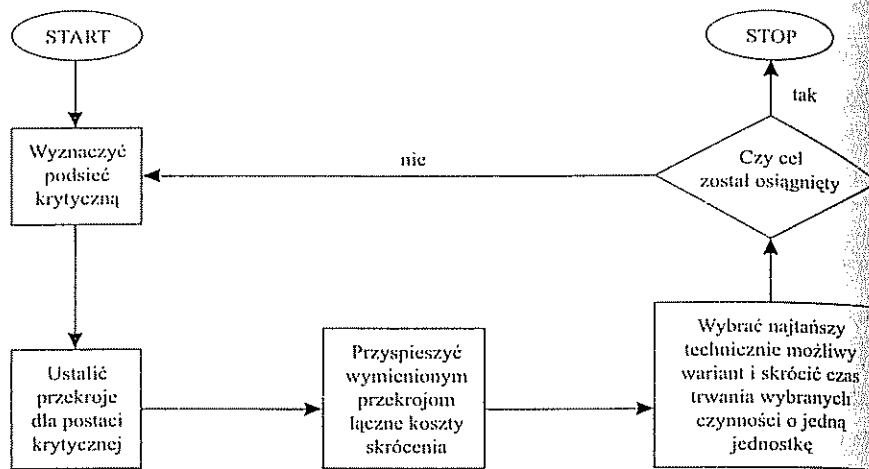
$y_{ij}^p$  — zmienna przyjmie wartość 1, gdy czynność  $\langle i, j \rangle$  zostanie skrócona po raz  $p$ -ty,

$k_{ij}^p$  — koszt skrócenia po raz  $p$ -ty czynności  $\langle i, j \rangle$ ,

$T^d$  — czas dyrektywny.

Warunki (7.22), (7.30), (7.36) oraz (7.44) dotyczą struktury sieci. Dzięki wyrażeniom (7.29) i (7.43) czas trwania czynności można skrócić po raz  $p$ -ty tylko wtedy, gdy został on już skrócony ( $p-1$ ) razy. W przypadku małych grafów rozwiązanie opisanych zadań nawet za pomocą komputera może okazać się bardziej pracochłonne niż ustalenie optymalnego wyniku na podstawie algorytmu kompresji sieci. W literaturze znaleźć można uproszczone procedury przybliżone oraz algorytmy dokładne, lecz bardziej złożone. Najprościej jest wykonać kroki zaprezentowane na rysunku 7.11.

RYSUNEK 7.11  
Przybliżony algorytm kompresji sieci



Przekrojem podsieci krytycznej jest każdy minimalny zbiór luków, przez które przechodzą wszystkie ścieżki krytyczne (por. podrozdział 3.2)<sup>19</sup>. W skład przekroju wchodzi jedynie czynności rzeczywiste, a nie fikcyjne.

Osiągnięcie celu jest równoznaczne z otrzymaniem rozwiązania zadania (7.16)–(7.17) lub (7.18)–(7.19).

Z opisu metody wynika, że:

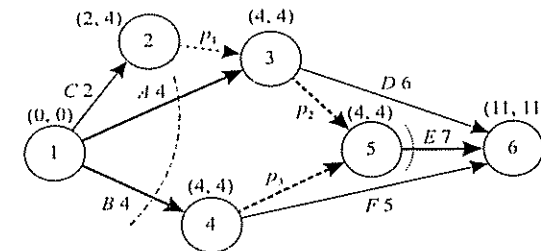
- 1) w celu skrócenia czasu realizacji całego przedsięwzięcia należy rozpatrywać jedynie ścieżki krytyczne, ponieważ to one decydują o momencie zakończenia inwestycji;
- 2) każde skrócenie momentu zakończenia przedsięwzięcia może wpłynąć na kształt podsieci krytycznej, czyli na liczbę ścieżek krytycznych;
- 3) ścieżka krytyczna może przestać być krytyczną, jeżeli nie zachowa w następnej iteracji zerowego zapasu czasu;
- 4) skracanie czasu realizacji przedsięwzięcia sukcesywnie o jedną jednostkę jest konieczne, zwłaszcza gdy w sieci występuje kilka ścieżek krytycznych; jeżeli sieć składa się tylko z jednej drogi krytycznej, i to znacznie dłuższej od pozostałych ścieżek, wybraną czynność krytyczną można w danej iteracji skrócić od razu o kilka jednostek, pamiętając o ograniczeniach wynikających z ustalonego czasu granicznego oraz z dostępnych środków przeznaczonych na akcelerację działań.

<sup>19</sup> W analizie czasowo-kosztowej przekrój konstruujemy tylko dla zbioru ścieżek krytycznych. Natomiast przy ustalaniu maksymalnego przepływu przekrój tworzymy, biorąc pod uwagę wszystkie ścieżki sieci (por. podrozdział 3.2).

PRZYKŁAD 7.3

Załóżmy, że czas trwania inwestycji opisanej na rysunku 7.12 należy skrócić do 9 dni.

RYSUNEK 7.12  
Wyjściowa sieć ( $T^* = 11$ )



Koszty realizacji przedsięwzięcia wynoszą 100 tys. zł, a koszty skrócenia (w tys. zł) poszczególnych czynności o każdy kolejny dzień podano w tabeli 7.8<sup>20</sup>.

TABLICA 7.8  
Koszty skrócenia czasu trwania działań o kolejne dni

| Czynności                                    | A | B | C | D | E | F |
|--|---|---|---|---|---|---|
| Koszt skrócenia o pierwszy dzień ( $k_i^1$ ) | 2 | 4 | 1 | 5 | 2 | 2 |
| Koszt skrócenia o drugi dzień ( $k_i^2$ )    | — | 6 | — | 5 | 3 | — |

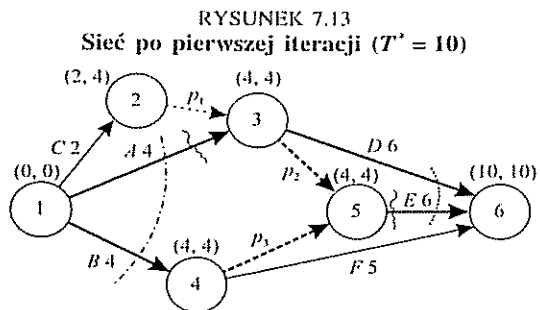
Oto koszt skrócenia odpowiadający przekrojom wyznaczonym dla podsieci krytycznej zawierającej dwie ścieżki ( $A-p_1-E$  i  $B-p_2-E$ ):

$$-k_A^1 + k_E^1 = 2 + 4 = 6$$

$$-k_B^1 = 2$$

W wyniku przyspieszenia projektu do 10 dni według tańszego wariantu (skrócenie czynności E o 1 dzień) pojawia się w sieci trzecia ścieżka krytyczna:  $A-D$  (rysunek 7.13).

<sup>20</sup> Koszt ten nie dotyczy czynności fikcyjnych, których czas trwania już wynosi zero. Znak „—” oznacza, że danej czynności nie można już skrócić.



W dalszej kolejności skrócenie czasu trwania czynności A i E okazuje się najtańsze:

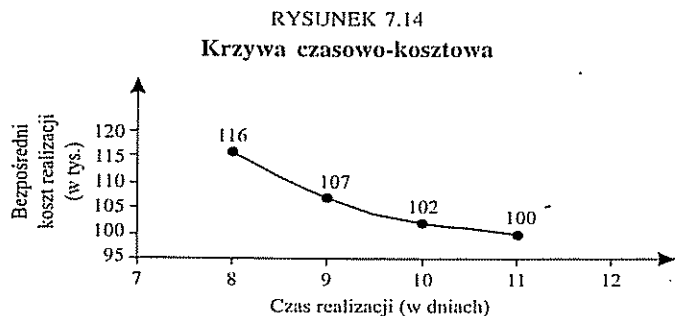
$$-k_A^1 + k_B^1 = 2 + 4 = 6$$

$$-k_A^1 + k_E^2 = 2 + 3 = 5$$

$$-k_D^1 + k_E^2 = 5 + 3 = 8$$

Łączny koszt skrócenia przedsięwzięcia z 11 do 9 dni wynosi  $2 + 5 = 7$  tys. zł.

Wybór przekroju A + E zmienia podsieć krytyczną. Skrócenie tych czynności o jeden dzień powoduje przyspieszenie ścieżki A-p<sub>2</sub>-E od razu o dwa dni. Tym samym droga ta przestaje być krytyczna!



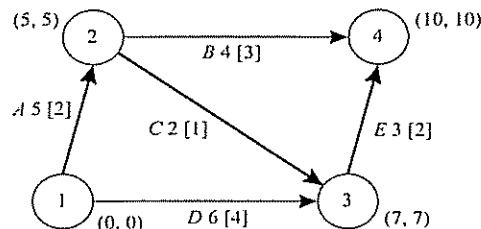
Krzywą czasowo-kosztową obrazującą łączne koszty bezpośrednie dla różnych wartości  $T$  przedstawiono na rysunku 7.14. Krzywa ta zawiera również informacje o koszcie, gdy  $T = 8$  dni. W trzeciej iteracji, gdy ścieżki krytyczne to A-D, B-p<sub>2</sub>-E i B-F, warto skorzystać z przekroju D + B, gdyż pozostałe przekroje A + B, A + E + F, D + E + F są niedopuszczalne. Z uwagi na czasy graniczne, dalsze skracanie nie jest możliwe.

W pracy są rozpatrywane zmiany czasu realizacji przedsięwzięcia o charakterze dyskretnym. Dlatego też rozwiązania leżące na poszczególnych fragmentach krzywej, z wyłączeniem wyróżnionych punktów, są jedynie przybliżeniem tych wyników, które moglibyśmy uzyskać, rozpatrując problem w wersji ciągłej. O dokładnym przebiegu krzywej czasowo-kosztowej decydują funkcje kosztów skrócenia przyspieszonych czynności.

Omówiony przykład pozwala wyciągnąć jeszcze dwa interesujące wnioski dotyczące zaprezentowanego algorytmu. Po pierwsze, w przypadku tej metody istnieje możliwość otrzymania w danej iteracji podsieci krytycznej o mniejszej liczbie ścieżek krytycznych niż w iteracji poprzedniej. Po drugie, skrócenie czasu realizacji przedsięwzięcia o jedną jednostkę nie musi oznaczać, że wszystkie najdłuższe drogi zostaną także dokładnie o tę wielkość skrócone. Wystarczy przyjąć założenie, że należy je przyspieszyć o przynajmniej jedną jednostkę<sup>21</sup>.

Opisany algorytm jest dość skutecznym narzędziem kompresji sieci. Istnieją jednak nawet proste przykłady, na podstawie których można wykazać, że procedura ta daje gorsze wyniki niż model optymalizacyjny (zob. [Gaspars, 2006]). Według przedstawionego algorytmu skrócenie czasu wykonania przedsięwzięcia opisanego

RYSUNEK 7.15  
Przykład, w którym stosowanie prezentowanego algorytmu nie prowadzi do optimum ( $T^d = 8$ )



<sup>21</sup> W literaturze można znaleźć opis procedury, zgodnie z którą ścieżki krytyczne powinny być właśnie skracane każdorazowo o dokładnie jedną jednostkę (por. [Bładowski, 1970; Grucza, Ogonek, Trocki, 2003; Idźkiewicz, 1967; Kaufmann, Desbazeille, Ventura, 1964; Kopańska-Bródka, 1998; Kukuła, 1996]). Gdybyśmy analizowany przykład próbowali rozwiązać tą metodą, okazałoby się, że lista, rozpatrywanych w drugiej iteracji wariantów skrócenia czasu realizacji przedsięwzięcia byłaby mniejsza, gdyż wybrany przekrój A + E, który wiązał się ze skróceniem czasu trwania ścieżki A-p<sub>2</sub>-E o więcej niż jedną jednostkę, nie byłby tym razem w ogóle brany pod uwagę. Ograniczanie się przy wyborze najtańszego wariantu do kombinacji skracających wszystkie drogi krytyczne o dokładnie jedną jednostkę może skutkować uzyskaniem rozwiązania, dla którego łączne koszty skrócenia są znacznie wyższe niż wynikałoby to z rozwiązania optymalnego, a czasem nawet niemożnością wyznaczenia rozwiązania dopuszczalnego, mimo że takie istnieje.

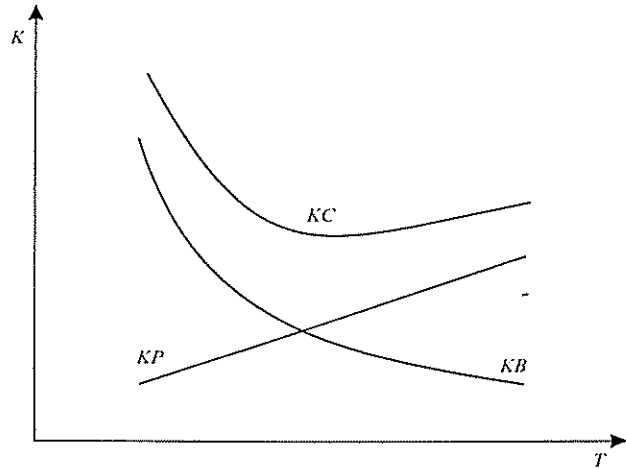
na rysunku 7.15 do 8 dni (tym razem koszty skrócenia podane w nawiasach kwadratowych są stałe, a czasy graniczne wynoszą, odpowiednio, 2, 1, 1, 3, 1) wymaga przyspieszenia najpierw czynności C, a potem czynności A i E, co w sumie kosztować będzie 5 tys. zł. Tymczasem okazuje się, że wystarczy skrócić tylko działania A i E, aby uzyskać ten sam efekt przy niższych kosztach (w sumie 4 tys. zł)!

Rozwiązanie optymalne zaprezentowanego problemu uzyskaliby na pewno m.in.:

- 1) Fulkerson i Kelley, którzy problem kompresji sieci sprowadzili za pomocą programu dualnego do zagadnienia maksymalnego przepływu w sieci<sup>22</sup>,
- 2) Siemens, dla którego punktem wyjścia nie było iteracyjne skracanie najdłuższych ścieżek na podstawie przekrojów, lecz ustalenie czasu, o który każda ścieżka powinna być przynajmniej skrócona, aby czas dyrektywny został osiągnięty<sup>23</sup>,
- 3) Phillips i Dessouky, twórcy koncepcji minimalnego cięcia w grafie (*minimum cut*)<sup>24</sup>.

Procedury tych autorów zostały opracowane przy założeniu, że krańcowe koszty skracania są stałe. Niektóre techniki zaliczyć można do ręcznych procedur

RYSUNEK 7.16  
Krzywe kosztów przedsięwzięcia



<sup>22</sup> Przyjęli oni, że koszt skrócenia czynności w zadaniu pierwotnym odpowiada przepustowości luku w zadaniu dualnym.

<sup>23</sup> Algorytm Siemens jest algorytmem przybliżonym. Goyal udoskonalił go w 1975 r.

<sup>24</sup> Cięciem grafu nazywamy taki zbiór luków, że po jego usunięciu graf rozpada się na co najmniej dwie składowe. Minimalne cięcie to cięcie o minimalnej liczbie luków. Koncepcja minimalnego cięcia jest dość prosta i zapewnia dojście do optimum. Stanowi ona uproszczoną wersję algorytmu Fulkersona

obliczeniowych (zob. [Phillips, Dessouky, 1977; Siemens, 1971]), inne — opierające się na programowaniu matematycznym — wymagają zastosowania komputera (np. [Kelley, 1961]). Nie wszystkie metody dają zawsze najlepszy wynik (zob. [Siemens, 1971]).

Wszystkie wymienione metody opracowano z myślą o małych projektach. Analizę czasowo-kosztową większych przedsięwzięć warto przeprowadzić, korzystając ze specjalnych programów komputerowych (zob. [Liu, Burns, Feng, 1995]).

W pracy uwzględniono jedynie koszty bezpośrednie (*KB*). W przypadku kosztów pośrednich (*KP*) zależność czas–koszt jest odwrotna. Krótszemu terminowi zakończenia inwestycji towarzyszą niższe koszty pośrednie. Przebieg całkowitych kosztów (*KC*) jej realizacji można opisać np. funkcją paraboliczną, której ekstremum oznacza optymalny harmonogram prac z punktu widzenia minimalizacji łącznych kosztów (rysunek 7.16).

## 7.5. Analiza zasobowa

Załóżmy, że znany jest harmonogram prac komisji rekrutującej kandydatów na uczelnię. Okazuje się jednak, że plan ten nie jest możliwy do wykonania, ponieważ liczba dostępnych komputerów, przy których członkowie komisji mieliby pracować, jest niewystarczająca. Jeżeli nie ma możliwości dostosowania wyposażenia uczelni do potrzeb komisji, to konieczna jest weryfikacja czasu i kolejności realizacji planowanych działań, która najczęściej wiąże się z opóźnieniem momentu zakończenia prac.

Przykład ten uświadomił nam, że planowanie sieciowe, przeprowadzone przy założeniu, że po pierwsze liczba osób oraz urządzeń produkcyjnych potrzebnych do realizacji projektu nie stanowi ograniczenia i, po drugie, pracowników można zawsze zwolnić z pracy lub przydzielić do innych zajęć, może być potraktowane jako zagadnienie czysto teoretyczne, pozbawione praktycznego zastosowania. Dlatego też warto przeprowadzić **analizę zasobową przedsięwzięcia**. Przez pojęcie zasobu rozumiemy będziemy zarówno sprzęt, urządzenia, jak i materiały, siłę roboczą, powierzchnię, przestrzeń produkcyjną oraz środki finansowe. Z uwagi na ostatnią wymienioną kategorię przyjęło się traktować analizę czasowo-kosztową (zob. podrozdział 7.4) jako element analizy zasobowej. W literaturze znaleźć można modele minimalizujące czas realizacji projektu przy limitowanych zasobach. Ze względu na binarny charakter i dużą liczbę zmiennych w tych modelach, zaleca się zastąpienie zadania optymalizacyjnego odpowiednim algorytmem, który pozwala wygenerować zadowalające rozwiązanie.

W celu obniżenia ilości zużywanego zasobu do poziomu jego dostępności (*d*), wystarczy czasami opóźnić moment rozpoczęcia niektórych czynności niekrytycznych w granicach ich zapasu, bez naruszenia terminu końcowego.

PRZYKŁAD 7.4

Rysunek 7.17 przedstawia dzienne zapotrzebowanie na robotników dla poszczególnych czynności (A–D) wchodzących w skład pewnego przedsięwzięcia. Dółny wiersz zawiera informację o łącznym dziennym zapotrzebowaniu (Z) przy założeniu, że wszystkie działania  $\langle i, j \rangle$  rozpoczynają się w momencie  $t'_i$ .

RYSUNEK 7.17  
Plan niedopuszczalny

|       |   |   |       |       |   |       |       |   |
|-------|---|---|-------|-------|---|-------|-------|---|
| D     |   |   | 5     | 5     | 5 | ..... | ..... |   |
| C     |   |   |       |       | 2 | 2     | 2     |   |
| B     | 3 | 3 | ..... | ..... |   |       |       |   |
| A     | 4 | 4 | 4     | 4     |   |       |       |   |
| Dzień | 1 | 2 | 3     | 4     | 5 | 6     | 7     | 8 |
| Z     | 7 | 7 | 9     | 9     | 7 | 2     | 2     |   |

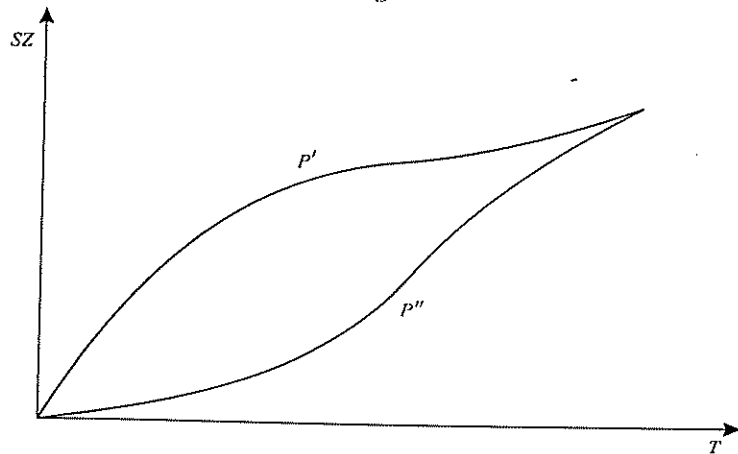
RYSUNEK 7.18  
Plan dopuszczalny

|       |   |   |       |       |   |   |   |   |
|-------|---|---|-------|-------|---|---|---|---|
| D     |   |   |       |       | 5 | 5 | 5 |   |
| C     |   |   |       |       | 2 | 2 | 2 |   |
| B     | 3 | 3 | ..... | ..... |   |   |   |   |
| A     | 4 | 4 | 4     | 4     |   |   |   |   |
| Dzień | 1 | 2 | 3     | 4     | 5 | 6 | 7 | 8 |
| Z     | 7 | 7 | 4     | 4     | 7 | 7 | 7 |   |

Przypuśćmy, że dysponujemy jedynie 7 robotnikami ( $d=7$ ), co oznacza, że w trzecim i czwartym dniu realizacji przedsięwzięcia jest ich za mało. Okazuje się, że opóźnienie momentu rozpoczęcia czynności D o dwa dni zmniejszy szczytowe zatrudnienie, nie wydłużając przy tym realizacji samej inwestycji (rysunek 7.18).

Jeżeli w celu dostosowania zapotrzebowania na środki produkcyjne do dostępnej ilości dochodzi do opóźnienia momentów rozpoczęcia niektórych czynności

RYSUNEK 7.19  
Esogram

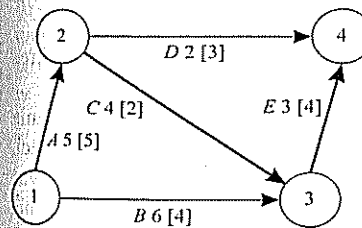


niekrytycznych w granicach ich zapasu, a struktura sieci (czyli kolejność działań) i czas realizacji przedsięwzięcia nie ulegają zmianie, to ustalonymu harmonogramowi prac odpowiada krzywa skumulowanego zapotrzebowania (SZ), leżąca w obszarze między krzywymi P' i P'' (rysunek 7.19). Krzywe dotyczą sytuacji, w której wszystkie działania zaczynają się, odpowiednio, w momencie t' i t''. Noszą one nazwę esogramu.

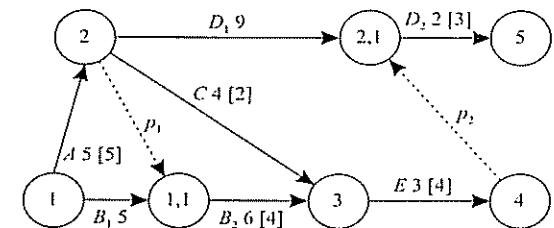
PRZYKŁAD 7.5

Teraz założmy, że realizacja przedsięwzięcia opisanego na rysunku 7.20 wymaga każdego dnia konkretnej liczby pracowników (podanej w nawiasie kwadratowym), a siła robocza, którą dysponuje wykonawca, wynosi 6 osób dziennie ( $d=6$ ). Tym razem moment zakończenia inwestycji będzie musiał zostać przesunięty przynajmniej o 4 dni (z 12 do 16 dni). Skorygowany plan działania przedstawiono na rysunku 7.21. Zawiera on dodatkowe czynności rzeczywiste, oznaczające okresy oczekiwania, oraz czynności pozorne, określające kolejność następowania po sobie czynności, do których potrzebne są te same środki produkcyjne. Wprowadzenie czynności pozornych<sup>25</sup> sprawia, że niektóre czynności dotychczas równoległe (np. A i B) stają się czynnościami szeregowymi, a to — z kolei — zmienia przebieg drogi krytycznej uzależnionej teraz nie tylko od wymagań technologicznych, lecz także od możliwości realizacji programu przy deficycie zasobów potrzebnych do jego wykonania. Czynność B<sub>1</sub> określa liczbę dni, które należy odczekać zanim pracownicy zakończą czynność A i będą mogli rozpocząć czynność B (opisaną lukiem B<sub>2</sub>).

RYSUNEK 7.20  
Plan niedopuszczalny



RYSUNEK 7.21  
Plan dopuszczalny

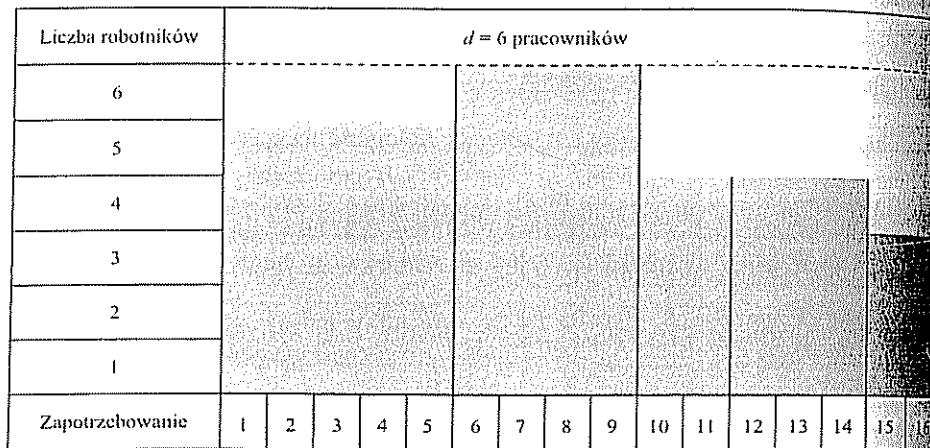


Kolejność przydziału ograniczonego środka produkcyjnego, minimalizującą czas wykonania programu, można też ustalić za pomocą następujących zasad [Bładowski, 1970]:

<sup>25</sup> Tym razem dla czynności pozornych możliwe jest naruszenie warunku  $i < j$  (zob. podrozdział 7.1).

- 1) pierwszeństwo mają czynności krytyczne, tuż za nimi — czynności niekrytyczne o najmniejszym zapasie czasu; jeżeli dwie czynności niekrytyczne mają ten sam zapas, to pierwszeństwo przyznaje się tej, która wykazuje większe zapotrzebowanie;
- 2) czynności już rozpoczęte są priorytetowe;
- 3) czynności, które w danym przedziale czasowym nie otrzymały przydziału, czekają na swoją kolejkę w następnych przedziałach czasowych; przedział czasowy kończy się w dniu, w którym wykonano którąkolwiek z czynności mających przydział.

RYСУNEK 7.22  
Dzienne zapotrzebowanie na siłę roboczą



Działanie algorytmu dla omawianego przykładu ilustruje tablica 7.9. Wykres dziennego zapotrzebowania na siłę roboczą przedstawiono na rysunku 7.22.

Zaprezentowany algorytm stanowi jedną z możliwych metod rozwiązania omówionego zadania. Jeżeli zależy nam na zachowaniu pierwotnej struktury sieci, a ograniczonym zasobem jest np. siła robocza, to można spróbować oszacować czasy trwania czynności, przyjmując, że dane działanie wykona mniej osób niż początkowo zakładano. Jeśli czynności niekrytyczne będą dysponowały dużymi zapasami czasu, to może się okazać, że przedsięwzięcie będzie mogło zostać zakończone zgodnie z planem.

W pracy skoncentrowano się jedynie na wybranych problemach dotyczących analizy zasobowej. Po pierwsze, rozpatrywano wyłącznie sytuację, w której czas trwania czynności jest dokładnie znany. Po drugie, przyjęto, że dostępność środków produkcyjnych jest stała w czasie realizacji przedsięwzięcia. Po trzecie, w ob-

TABLICA 7.9  
Procedura rozdziału środka

| Krok (przedział czasów) | Możliwe czynności | Status czynności        | Zapotrzebowanie | Kolejność | Czynności otrzymujące przydział |
|-------------------------|-------------------|-------------------------|-----------------|-----------|---------------------------------|
| I krok (od 0 do 5)      | A                 | krytyczna               | 5               | 1         | A                               |
|                         | B                 | niekrytyczna, zapas = 3 | 4               | 2         |                                 |
| II krok (od 5 do 9)     | B                 | zapas zmalał do 0       | 4               | 2         | B                               |
|                         | C                 | krytyczna               | 2               | 1         | C                               |
|                         | D                 | niekrytyczna, zapas = 5 | 3               | 3         |                                 |
| III krok (od 9 do 11)   | B                 | w toku                  | 4               | 1         | B                               |
|                         | D                 | zapas zmalał do 1       | 3               | 3         |                                 |
|                         | E                 | krytyczna               | 4               | 2         |                                 |
| IV krok (od 11 do 14)   | D                 | zapas zmalał do 0       | 3               | 2         |                                 |
|                         | E                 | krytyczna               | 4               | 1         | E                               |
| V krok (od 14 do 16)    | D                 | zapas zmalał do 0       | 3               | 1         | D                               |

zadaniach należało ustalić optymalny harmonogram prac przy założeniu, że tylko jeden zasób jest ograniczony. Sposób rozwiązywania problemów, w których mamy do czynienia z nierównomierną podażą<sup>26</sup> lub z kilkoma deficytowymi środkami produkcyjnymi, przedstawili m.in. Bładowski [1970] oraz Hendrickson i Au [1998]. Po czwarte, pominięto przypadek, w którym przedsiębiorstwo realizuje równolegle więcej niż jedno przedsięwzięcie, korzystając z tych samych zasobów (zob. [Idźkiewicz, 1970]). Po piąte, zaprezentowane przykłady dotyczyły zasobu odnawialnego (siła robocza), czyli takiego, który po wykonaniu danej czynności jest przydzielany kolejnym działaniom. Problemem alokacji zasobów nieodnawialnych zajął się również Bładowski [1970]. I po szóste, omówiono jedynie zagadnienie minimalizacji czasu trwania przedsięwzięcia przy danym limicie zasobów. Drugim równoległym celem mogłoby być równomierne rozmieszczenie zasobów w czasie realizacji projektu przy zadanym terminie zakończenia inwestycji.

#### POLECENIA I PYTANIA KONTROLNE

1. Czym różni się technika AOA od techniki AON?
2. Jakimi własnościami charakteryzuje się sieć czynności?
3. Jak przedstawiamy w sieci czynność pozorną? Jaki jest jej czas trwania? Jaka ona odgrywa rolę w analizie czasowej i zasobowej?

<sup>26</sup> Sytuacja taka może mieć miejsce, gdy środkiem produkcyjnym jest np. cement dostarczany na budowę partiami, co kilka dni.

4. W jaki sposób można wyznaczyć najkrótszy czas realizacji przedsięwzięcia? Rozpatrzyć problem dla wersji deterministycznej i indeterministycznej.
5. Czym charakteryzuje się ścieżka krytyczna?
6. Jak zmiana (wydłużenie bądź skrócenie) czasu trwania czynności może wpłynąć na moment zakończenia przedsięwzięcia? Analizę przeprowadzić dla czynności krytycznych i niekrytycznych.
7. Jakie problemy optymalizacyjne są najczęściej rozpatrywane w analizie czasowo-kosztowej oraz w analizie zasobowej przedsięwzięć?
8. Jakie informacje można odczytać z krzywej czasowo-kosztowej dla czynności i dla przedsięwzięcia?
9. Które czynności mają pierwszeństwo przy alokacji ograniczonego środka produkcyjnego?

## BIBLIOGRAFIA

- Bładowski S., 1970, *Metody sieciowe w planowaniu i organizacji pracy*, PWE, Warszawa.
- Fulkerson D.R., 1961, *A Network Flow Computation for Project Cost Curves*, „Management Science”, vol. 7(2), INFORMS, USA.
- Gaspars H., 2006, *Analiza czasowo-kosztowa (CPM-COST). Algorytm a model optymalizacyjny*, „Badania Operacyjne i Decyzje”, nr 1, Wydawnictwo Politechniki Wrocławskiej.
- Gedymin O., 1974, *Metody optymalizacji w planowaniu sieciowym*, PWN, Warszawa.
- Hendrickson C., Au T., 1989, *Project Management for Construction. Fundamental Concepts for Owners, Engineers, Architects and Builders*, Prentice Hall Englewood Cliffs, New Jersey.
- Idzkiewicz A.Z., 1967, *PERT. Metody analizy sieciowej*, PWN, Warszawa.
- Ignasiak E., Borucki W., Marcinkowski J., Sikora W., 1996, *Badania operacyjne*, PWN, Warszawa.
- Kaufmann A., Desbazeille G., Ventura E., 1964, *La méthode du chemin critique*, DUNOD, Paris.
- Kelley J.E., 1961, *Critical Path Planning and Scheduling: Mathematical Basis*, „Operations Research”, vol. 9(3), INFORMS, USA.
- Kelley J.E., Walker M.R., 1959, *Critical Path Planning and Scheduling: An Introduction*, Mauchly Associates, Ambler, Pennsylvania.
- Kopańska-Bródka D., 1998, *Wprowadzenie do badań operacyjnych*, wyd. 2 poprawione, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice.
- Kukuła K., Jędrzejczyk Z., Skrzypek J., Walkosz A., 1996, *Badania operacyjne w projektach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa.
- Liu L., Burns S.A., Feng C.-W., 1995, *Construction Time-Cost Tradeoff Analysis Using LP/IP Hybrid Method*, „Journal of Construction Engineering and Management”, vol. 121(4), ASCE, RESTON, VA, USA.
- Moder J.J., Phillips C.R., 1964, *Project Management with CPM and PERT*, Reinhold Publishing Corporation, New York.
- Phillips S.J., Dessouky M.I., 1977, *Solving the Time/Cost Tradeoff Problem Using the Minimum Cut Concept*, „Management Science”, vol. 24(4), INFORMS, USA.
- Siemens N., 1971, *A Simple CPM Time-Cost Tradeoff Algorithm*, „Management Science”, vol. 17(6), INFORMS, USA.
- Trocki M., Grucza B., Ogonek K., 2003, *Zarządzanie projektami*, PWE, Warszawa.

## ROZDZIAŁ 8

OPTIMALIZACJA  
DYSKRETNA

## 8.1. Wprowadzenie

Jest wiele sytuacji decyzyjnych, których nie można opisać używając tylko zmiennych ciągłych. Związane jest to z nieciągłością wielu procesów ekonomicznych. Samochód do miejscowości  $M$  jedzie jeden, dwa lub trzy razy. Robotnika można przydzielić do stanowiska  $S_1$ ,  $S_2$ ,  $S_3$  lub  $S_4$ . Określony projekt inwestycyjny jest realizowany lub nie. Zakład produkcyjny jest lokalizowany w określonym miejscu spośród wielu możliwych.

Żądamy więc, aby pewne wielkości, które mamy wyznaczyć (zmiennie decyzyjne), przyjmowały wartość ze zbioru dyskretnego, np. ze zbioru liczb całkowitych lub ze zbioru liczb binarnych  $\{0, 1\}$ .

Zagadnienie decyzyjne, w którym chociaż jedna zmienna decyzyjna przyjmuje wartości dyskretne, nazywamy **dyskretnym zagadnieniem decyzyjnym**. Model matematyczny opisujący tę sytuację nazywamy **dyskretnym zadaniem decyzyjnym**.

W rozdziale tym ograniczymy się do omówienia zagadnień dyskretnych, w których wszystkie relacje zachodzące między wielkościami są liniowe. Stąd sformułowane zadania są zadaniami **programowania dyskretnego liniowego (PDL)**. Wśród nich możemy wyróżnić trzy grupy zadań:

- 1) zadania **programowania całkowitoliczbowego liniowego (PCL)**, gdzie wszystkie zmienne są liczbami całkowitymi;
- 2) zadania **programowania binarnego liniowego (PBL)**, gdzie wszystkie zmienne są binarne;
- 3) zadania **programowania mieszanego liniowego (PML)**, gdzie część zmiennych to zmienne ciągłe, część — zmienne całkowite, a część — zmienne binarne.

Formułujemy model matematyczny w postaci zadania PDL, w którym nieciągłość zmiennych powoduje, że rozwiązuje się ono w ogólnym przypadku trudniej niż zadania PL. Można wyróżnić dwa zasadnicze podejścia do rozwiązywania zagadnień dyskretnych:

- 1) formułujemy model matematyczny w postaci zadania PDL, które następnie rozwiązujemy metodami ogólnymi;
- 2) zagadnienia o szczególnej strukturze rozwiązujemy bezpośrednio specjalnymi algorytmami przystosowanymi do struktury zagadnienia.

Pierwsze podejście trudno stosować, gdy model matematyczny w postaci zadania PDL prowadzi do zadania o dużych rozmiarach (kilka tysięcy zmiennych i warunków), a taka sytuacja występuje często przy optymalizacji dyskretniej.

Uniwersalną metodą rozwiązywania zagadnień dyskretnych jest **metoda podziału i ograniczeń**. Można ją stosować zarówno w jednym, jak i w drugim podejściu. Przy rozwiązywaniu zadań PDL możemy także korzystać z **metody odcięć**. W przypadku podejścia drugiego, oprócz metod dokładnych, stosujemy często algorytmy przybliżone (heurystyczne), pozwalające uzyskać rozwiązania suboptymalne. W rozdziale tym przedstawiamy kilka podstawowych zagadnień dyskretnych: problem przydziału, problem lokalizacji produkcji, problem ustalania harmonogramu realizacji prac, problem komiwojażera, problem rozwózki. Następnie omówimy ideę metody podziału i ograniczeń oraz jej zastosowanie do rozwiązywania zadania PDL i problemu komiwojażera. Na koniec opiszemy algorytm Johnsona dla ustalenia kolejności obróbki  $n$  detali na dwóch maszynach.

## 8.2. Zagadnienie przydziału

### PRZYKŁAD 8.1

Na wydziale obróbki mamy cztery obrabiarki i czterech robotników, którzy mogą je obsługiwać. Znamy wydajność każdego robotnika na poszczególnych obrabiarkach. Wydajność tę określa liczba detali, które dany robotnik może wykonać na określonej maszynie. Przedstawiamy je w tabelicy 8.1.

TABLICA 8.1  
Macierz wydajności

| $w_{ij}$ | R1 | R2 | R3 | R4 |
|----------|----|----|----|----|
| M1       | 6  | 8  | 10 | 5  |
| M2       | 11 | 7  | 9  | 6  |
| M3       | 10 | 5  | 8  | 9  |
| M4       | 13 | 12 | 7  | 10 |

Opisaną sytuację łatwo uogólnić. Mamy  $n$  stanowisk i  $n$  pracowników. Znamy macierz wydajności  $W = [w_{ij}]$ , gdzie  $w_{ij}$  jest wydajnością  $j$ -tego pracownika na  $i$ -tym stanowisku pracy. Należy ustalić taki przydział pracowników do stanowisk, aby wydajność całego zespołu była maksymalna, a każde stanowisko obsadzone przez jednego pracownika.

### 8.2.1. Liniowy model matematyczny

Wprowadzimy zmienne decyzyjne:

$$x_{ij} = \begin{cases} 1, & \text{gdy pracownik } j \text{ jest przydzielony do } i\text{-tego stanowiska,} \\ 0, & \text{w przypadku przeciwnym.} \end{cases}$$

Problem optymalnego przydziału można sformułować w postaci następującego liniowego zadania decyzyjnego: znajdź takie wartości zmiennych  $x_{ij}$ , aby:

$$\sum_{i=1}^n \sum_{j=1}^n w_{ij} x_{ij} \rightarrow \max, \quad (8.1)$$

przy warunkach:

$$\sum_{j=1}^n x_{ij} = 1 \quad (i = 1, \dots, n), \quad (8.2)$$

$$\sum_{i=1}^n x_{ij} = 1 \quad (j = 1, \dots, n), \quad (8.3)$$

$$x_{ij} \in \{0, 1\} \quad (i = 1, 2, \dots, n, j = 1, 2, \dots, n). \quad (8.4)$$

Zadanie (8.1)–(8.4) jest zadaniem programowania binarnego liniowego (PBL). Funkcja celu postuluje maksymalizację łącznej wydajności całego zespołu. Warunek (8.2) zapewnia, że na każde stanowisko będzie przydzielony tylko jeden pracownik. Natomiast warunek (8.3) wymusza, aby każdy pracownik był przydzielony tylko do jednego stanowiska. Zadanie optymalnego przydziału, mimo że jest klasycznym problemem programowania dyskretnego, może być rozwiązywane metodami programowania liniowego (metodą sympleks lub metodą potencjałów). Macierz współczynników tego zadania jest tzw. **macierzą unimodularną**, czyli taką, że każdy jej podwyznacznik stopnia  $m$  jest równy  $-1$ ,  $0$  lub  $1$ , a wektor wyrazów wolnych jest całkowitoliczbowy. Dowodzi się, że w tej sytuacji każde rozwiązanie bazowe (a więc również rozwiązanie optymalne) będzie spełniało warunek całkowitoliczbowości.



### 8.2.2. Kombinatoryczny model matematyczny

Zagadnienia przydziału nie musimy formułować w postaci zadania PBL (8.1–8.4). Można je sformułować w nieco innej postaci. Niech  $v = (v_1, v_2, \dots, v_n)$  będzie dowolną permutacją liczb  $1, 2, \dots, n$ . Przyjmijmy, że zapis ten oznacza, że stanowisko pierwsze jest obsadzone przez pracownika o numerze  $v_1$ , stanowisko drugie — przez pracownika o numerze  $v_2$ , a stanowisko  $n$ -te — przez pracownika o numerze  $v_n$ . Tak więc każda permutacja liczb naturalnych  $1, 2, \dots, n$  jest rozwiązaniem dopuszczalnym zagadnienia przydziału. Zbór wszystkich permutacji (rozwiązań) oznaczmy przez  $D$ .

Niech dalej  $f(v)$  będzie wydajnością zespołu dla rozwiązania dopuszczalnego. Zagadnienie przydziału można w tej konwencji zapisać następująco: znajdź taką obsadę stanowisk (taką permutację  $v^*$ ), aby:

$$v^* \in D,$$

$$f(v^*) = \max \{f(v) | v \in D\}.$$

Zapis (8.5)–(8.6) jest typowy dla tych problemów dyskretnych, które problemami kombinatorycznymi. Zbór rozwiązań dopuszczalnych jest skończony o ściśle ustalonej liczebności (dla zagadnienia przydziału  $n!$  elementów), a decyzja dopuszczalną można łatwo określić.

Ze względu na strukturę zagadnienia przydziału istnieje wiele efektywnych specjalnych metod jego rozwiązywania. Nie będziemy się nimi zajmować, odsyłając Czytelnika do literatury (zob. np. [Ford, Fulkerson, 1969; Gale, 1969]).

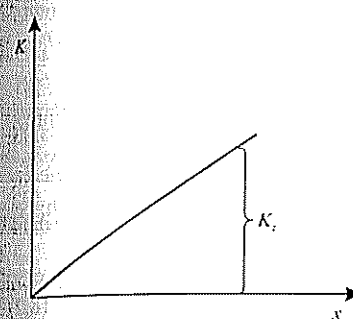
### 8.3. Zagadnienie lokalizacji produkcji

Firma produkująca jednorodny produkt (np. piwo, cukier) ma  $m$  zakładów, z których towar dostarczany jest do  $n$  odbiorców. Znamy popyt każdego odbiorcy. Łączna zdolność produkcyjna zakładów nie pozwala pokryć zapotrzebowania odbiorców. Stąd przewiduje się rozbudowę mocy produkcyjnych przez budowę nowych zakładów lub modernizację już istniejących. W starych zakładach produkcja jest prowadzona według „starej” technologii, a w zakładach nowych (lub zmodernizowanych) według „nowej” technologii. Dla każdego „istniejącego” lub „potencjalnego” punktu produkcji potrafimy określić początkową lub nową zdolność produkcyjną.

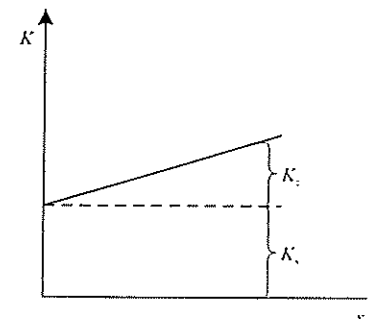
Koszt całkowity produkcji ( $K$ ) w starym zakładzie ograniczamy do kosztu zmiennego ( $K_z$ ), gdyż koszt stały produkcji ( $K_s$ ) od nas już w tym momencie

zależy. Kształtowanie się tego kosztu w zależności od wielkości produkcji przedstawiono na rysunku 8.1. Zakładamy zależność liniową między tymi wielkościami. W zakładach nowych lub modernizowanych koszt całkowity składa się z kosztu zmiennego oraz kosztu stałego. Koszt stały to głównie amortyzacja nakładów inwestycyjnych. Zależność kosztu produkcji od rozmiarów produkcji zobrazowano na rysunku 8.2.

RYSUNEK 8.1  
Koszty w zakładzie „starym”



RYSUNEK 8.2  
Koszty w zakładzie „nowym”



Znamy także wielkość nakładów, jakie musimy ponieść na inwestycje w każdym zakładzie. Wiemy, jaka jest cena sprzedaży towaru każdemu odbiorcy oraz jakie są jednostkowe koszty transportu od każdego punktu produkcji do poszczególnych odbiorców.

Należy tak ustalić plan lokalizacji inwestycji, plan produkcji oraz plan transportu towaru, aby dochód osiągany przez firmę był maksymalny.

Przez dochód rozumiemy różnicę między przychodem ze sprzedaży towarów a kosztami produkcji powiększonymi o koszty transportu.

Przyjmijmy oznaczenia:

- $M$  — zbiór punktów, w których może być prowadzona produkcja,
- $M_1$  — zbiór „istniejących” punktów produkcji (stare i zmodernizowane zakłady),
- $M_2$  — zbiór „potencjalnych” punktów produkcji, w których są lokalizowane nowe zakłady,
- $b_j$  — popyt  $j$ -tego odbiorcy,
- $a_i$  — początkowa zdolność produkcyjna (podaż)  $i$ -tego punktu ( $i \in M_1$ ),
- $d_i$  — nowa zdolność produkcyjna (podaż)  $i$ -tego punktu ( $i \in M$ ),
- $R$  — wielkość środków, jakie możemy przeznaczyć na inwestycje,
- $r_i$  — wielkość nakładów, jakie przeznaczamy na inwestycje w  $i$ -tym punkcie,
- $c_i$  — jednostkowy koszt zmienny w  $i$ -tym punkcie, gdy produkujemy „starą” technologią,

- $c_i$  — jednostkowy koszt zmienny w  $i$ -tym punkcie, gdy produkujemy „nową” technologią,  
 $s_i$  — koszt stały (amortyzacja) w  $i$ -tym punkcie, jeżeli były prowadzone inwestycje,  
 $c_{ij}$  — koszt jednostkowy transportu od  $i$ -tego punktu do  $j$ -tego dostawcy,  
 $p_j$  — cena sprzedaży jednostki towaru  $j$ -temu odbiorcy.

Wprowadźmy następujące zmienne decyzyjne:

$$y_i = \begin{cases} 1, & \text{gdy w } i\text{-tym punkcie budujemy nowy zakład lub modernizujemy stary} \\ 0, & \text{w przypadku przeciwnym.} \end{cases}$$

$\bar{x}_i$  — wielkość produkcji w  $i$ -tym punkcie, według „starej” technologii ( $i \in M_1$ ),

$x_i$  — wielkość produkcji w  $i$ -tym punkcie, według „nowej” technologii ( $i \in M$ ),

$x_{ij}$  — wielkość przewozu towaru od  $i$ -tego punktu produkcji do  $j$ -tego odbiorcy.

Model matematyczny zagadnienia rozmieszczenia produkcji sprowadza się do następującego zadania: Znajdź takie wartości zmiennych decyzyjnych  $y_i, \bar{x}_i, x_i, x_{ij}$ , aby:

$$\sum_{j=1}^n \left( p_j \sum_{i \in M} x_{ij} \right) - \sum_{i \in M} s_i y_i - \sum_{i \in M_1} \bar{c}_i \bar{x}_i - \sum_{i \in M} c_i x_i - \sum_{i \in M} \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \max, \quad (8.7)$$

przy warunkach:

$$\sum_{i \in M} r_i y_i \leq R, \quad (8.8)$$

$$x_i \leq d_i y_i \quad (i \in M), \quad (8.9)$$

$$\bar{x}_i \leq a_i (1 - y_i) \quad (i \in M_1), \quad (8.10)$$

$$\sum_{j=1}^n x_{ij} \leq x_i + \bar{x}_i \quad (i \in M_1), \quad (8.11)$$

$$\sum_{j=1}^n x_{ij} \leq x_i \quad (i \in M_2), \quad (8.12)$$

$$\sum_{i \in M} x_{ij} \leq b_j \quad (j = 1, \dots, n), \quad (8.13)$$

$$y_i \in \{0, 1\} \quad (i \in M), \quad (8.14)$$

$$x_i, \bar{x}_i, x_{ij} \geq 0 \quad (i \in M, j = 1, \dots, n). \quad (8.15)$$

Zadanie (8.7)–(8.15) jest zadaniem PML, gdyż część zmiennych to zmienne ciągłe, a część to zmienne binarne. Wszystkie warunki i funkcja celu są liniowe.

W funkcji celu (8.7) pierwszy człon określa przychód ze sprzedaży towaru odbiorcom, drugi człon — koszty stałe produkcji, które ponosimy, jeżeli są prowadzone inwestycje, trzeci człon — koszt zmienny produkcji według „starej” technologii, czwarty człon — koszt zmienny produkcji według „nowej” technologii, a ostatni człon — łączne koszty transportu.

Warunek (8.8) zapewnia, że na inwestycje nie przeznaczymy więcej środków niż posiadamy. Warunki (8.9) i (8.10) gwarantują, że produkcja według „nowej” technologii nie przekroczy nowej zdolności produkcyjnej, a produkcja według „starej” technologii nie przekroczy początkowej zdolności produkcyjnej oraz obydwie technologie nie będą realizowane jednocześnie w tym samym zakładzie.

Warunki (8.11)–(8.13) zapewniają, że z każdego punktu nie wywieziemy więcej towaru niż wynosi jego produkcja oraz każdemu odbiorcy dostarczymy nie więcej towaru niż wynosi jego popyt. Warunek (8.14) wyraża niepodzielność obiektów inwestycyjnych.

Zagadnienie lokalizacji produkcji można rozpatrywać w wersji uproszczonej, gdy nie ma żadnych starych zakładów, a za kryterium oceny przyjmiemy minimalizację łącznych kosztów produkcji i kosztów transportu.

Zagadnienie lokalizacji możemy rozwiązać, stosując podejście I lub podejście II. W przypadku podejścia I poważnym utrudnieniem są rozmiary zadania.

## 8.4. Zagadnienie wyznaczania harmonogramu realizacji prac

Rozważmy pewien wariant zagadnienia harmonogramowania prac związany z minimalizacją zatrudnionych robotników.

Firma ma wykonać  $m$  prac w okresie  $T$  dni. Dla każdej pracy znamy czas niezbędny do jej wykonania oraz liczbą potrzebnych robotników. Zakładamy, że umiejętności niezbędne do wykonania poszczególnych czynności są zbliżone, więc mogą je realizować ci sami robotnicy. Zbiór prac jest zbiorem zależnym, co oznacza, że zakończenie pewnych prac warunkuje możliwość rozpoczęcia następujących. Dana jest więc określona relacja poprzedzania  $P$ .

Należy ustalić taki harmonogram realizacji prac, aby liczba robotników, których firma musi zatrudnić, była minimalna.

Oznaczmy przez:

$j$  — numer dnia ( $j = 1, \dots, T$ ),

$i, k$  — numery prac ( $i, k = 1, \dots, m$ ),

$a_i$  — liczbę robotników niezbędnych do wykonania  $i$ -tej pracy,

$t_i$  — czas niezbędny na wykonanie  $i$ -tej pracy,

$T_i$  — zbiór dni, w jakich może być rozpoczęta  $i$ -ta praca.

Zakładamy, że zbiory  $T_i$  są tak określone, że spełniają ograniczenia wynikające z relacji poprzedzania oraz wszystkie prace mogą być zakończone w okresie  $T$  dni. Oczywiście, zbiory  $T_i$  nie mogą być puste oraz dla chociaż części prac powinny zawierać więcej niż jeden dzień. Przyjmijmy ponadto, że każda praca raz rozpoczęta jest realizowana bez przerw, przy tej samej i niezmiennej w czasie liczbie robotników. Liczba zatrudnionych robotników nie ulega zmianie w całym okresie.

Jeżeli para prac  $\langle i, k \rangle \in P$ , to zakończenie czynności  $i$ -tej umożliwia rozpoczęcie czynności  $k$ -tej.

Aby sformułować model matematyczny, musimy operować zmiennymi binarnymi.

Wprowadzimy zmienne decyzyjne:

$$y_{ij} = \begin{cases} 1, & \text{gdy } i\text{-ta praca jest wykonywana w } j\text{-tym dniu,} \\ 0, & \text{w przypadku przeciwnym} \end{cases}$$

lub

$$x_{ij} = \begin{cases} 1, & \text{gdy } i\text{-ta praca jest rozpoczynana w } j\text{-tym dniu,} \\ 0, & \text{w przypadku przeciwnym.} \end{cases}$$

Pierwszy typ zmiennych, chociaż ułatwia sformułowanie modelu, prowadzi do zadania o większych rozmiarach niż drugi typ zmiennych. Dalej będziemy korzystali z tych ostatnich. Wprowadźmy dodatkową pomocniczą zmienną decyzyjną  $y$ , która oznacza liczbę zatrudnionych robotników.

Sformułowanie modelu matematycznego dla tak na pozór prostego zagadnienia, jakie rozpatrujemy, nie jest sprawą łatwą, dlatego przedstawimy najpierw uwagi wyjaśniające kolejne zapisy.

Zauważmy, że w  $p$ -tym dniu są wykonywane nie tylko prace, które w tym dniu się rozpoczęły, lecz także wszystkie prace rozpoczęte wcześniej i niezakończone do tego dnia. Praca  $i$ -ta będzie realizowana w  $p$ -tym dniu, jeżeli:

$$\sum_{p-t_i < j \leq p} x_{ij} = 1.$$

Stąd  $\sum_{i=1}^m \left( a_i \sum_{p-t_i < j \leq p} x_{ij} \right)$  określa, ilu robotników potrzebuje firma w  $p$ -tym dniu.

Ponieważ każda praca rozpoczyna się raz, czyli:

$$\sum_{j \in T_i} x_{ij} = 1,$$

to  $\sum_{j \in T_i} j x_{ij}$  wyznacza dzień, w którym rozpoczyna się  $i$ -ta praca.

Ustalanie harmonogramu realizacji prac sprowadza się do wyznaczenia takich wartości zmiennych  $x_{ij}$  oraz  $y$ , aby:

$$y \rightarrow \min, \quad (8.16)$$

przy warunkach:

$$\sum_{j \in T_i} x_{ij} = 1 \quad (i = 1, \dots, m), \quad (8.17)$$

$$\sum_{j \in T_i} j x_{ij} + t_i \leq \sum_{j \in T_k} j x_{kj} \quad \langle i, k \rangle \in P, \quad (8.18)$$

$$y \geq \sum_{i=1}^m \left( a_i \sum_{p-t_i < j \leq p} x_{ij} \right) \quad (p = 1, \dots, T), \quad (8.19)$$

$$x_{ij} \in \{0, 1\} \quad \begin{matrix} (i = 1, \dots, m), \\ j \in T_i, \end{matrix} \quad (8.20)$$

$$y \text{ — całkowite.} \quad (8.21)$$

Zadanie (8.16)–(8.21) jest zadaniem PML. Funkcja celu (8.16) minimalizuje liczbę zatrudnionych pracowników. Warunek (8.17) zakłada, że każda praca musi rozpocząć się raz. Warunek (8.18) gwarantuje, że  $k$ -ta praca nie rozpocznie się wcześniej niż zostanie zakończona praca  $i$ -ta, jeżeli ta ostatnia poprzedza pracę  $k$ -tej. Warunek (8.19) zapewnia, że w każdym dniu zapotrzebowanie na robotników (wyrażone przez prawą część nierówności) nie przekroczy liczby zatrudnionych pracowników.

Ważnym problemem są rozmiary zadania (8.16)–(8.21). Jeżeli przyjmiemy liczebność zbioru  $P$ ,  $|P| = 2m$ , a liczebność zbioru  $T_i$  jest równa średnio  $|T_i| = 0,2T$ , to zadanie będzie miało  $L(z) = m \cdot 0,2T + 1$  zmiennych oraz  $L(w) = T + m + 2m$  warunków.

Dla  $m = 30$  oraz  $T = 30$  mamy 601 zmiennych oraz 190 warunków. Zadanie to można rozwiązać, stosując jeden z ogólnych algorytmów metody podziału i ograniczeń.

## 8.5. Zagadnienie komiwojażera

Rozważmy klasyczny problem optymalizacji dyskretniej. Komiwojażer wyrusza z pewnego miasta, ma odwiedzić kilka miejscowości i wrócić do punktu startu. Każde z miast ma być odwiedzone tylko raz w dowolnej kolejności.

Sformułujmy ten problem bardziej formalnie. Dany jest zbiór miast oraz nieujemna, kwadratowa macierz odległości (kosztu, czasu przejazdu)  $C = |c_{ij}|$ , gdzie  $c_{ij}$  określa odległość (koszt lub czas przejazdu) między miastem  $i$  a miastem  $j$ . Należy znaleźć taką drogę zamkniętą, przechodzącą przez wszystkie miejscowości, która jest minimalna.

Droga zamknięta, zwana dalej **marszrutą**, składa się z  $n$  odcinków, które nazywać będziemy trasami. Ponieważ marszruta nie może zawierać trasy  $\langle i, i \rangle$ , więc przyjmujemy, że  $c_{ii} = \infty$ , dla  $i = 1, 2, \dots, n$ .

Liczba rozwiązań (marszrut) w zagadnieniu komiwojażera jest równa  $(n-1)!$ . Dla  $n = 10$  mamy już  $9! = 362\,880$  rozwiązań. Przegląd zupełny zbioru w celu znalezienia rozwiązania optymalnego jest efektywny tylko dla małych  $n$  ( $n \leq 8$ ). Jeżeli chcemy skorzystać z metod ogólnych, musimy sformułować model w postaci zadania PDL.

Niech  $V$  będzie zbiorem wszystkich tras, a  $x_{ij}$  — zmienną binarną, która przyjmuje wartości:

$$x_{ij} = \begin{cases} 1, & \text{gdy marszruta zawiera trasę } \langle i, j \rangle, \\ 0, & \text{w przypadku przeciwnym.} \end{cases}$$

a  $z_j$  jest zmienną, która każdemu miastu przyporządkowuje liczbę całkowitą określającą kolejność odwiedzania tego miasta, przy założeniu, że punktem startu jest miejscowość o numerze „1”.

Problemem komiwojażera sprawdza się do następującego zdania: wyznaczyć takie wartości zmiennych  $x_{ij}$  oraz  $z_j$ , aby:

$$\sum_{\langle i, j \rangle \in V} c_{ij} x_{ij} \Rightarrow \min, \quad (8.22)$$

$$\sum_{i=1}^n x_{ij} = 1 \quad (j = 1, \dots, n), \quad (8.23)$$

$$\sum_{j=1}^n x_{ij} = 1 \quad (i = 1, \dots, n), \quad (8.24)$$

$$z_i - z_j + n x_{ij} \leq n - 1 \quad (i = 1, 2, \dots, n, j = 2, 3, \dots, n, i \neq j), \quad (8.25)$$

$$z_j \geq 0, z_j \in C \quad (j = 1, \dots, n), \quad (8.26)$$

$$x_{ij} \in \{0, 1\} \quad (\langle i, j \rangle \in V) \quad (8.27)$$

Funkcja celu (8.22) minimalizuje długość marszruty. Warunki (8.23) i (8.24) zapewniają, że komiwojażer przez każdy punkt przechodzi dokładnie jeden raz. Niestety nie gwarantują one, że z wybranych  $n$  tras tworzymy tylko jedną marszrutę. Niekiedy uzyskuje się kilka podcykli. Tworzenie podcykli wyklucza warunek (8.25).

Jeżeli przyjmiemy, że punktem startu jest miasto o numerze „1”, to zmienna  $z_j$  przypisuje kolejno odwiedzanym miejscowości coraz wyższe liczby całkowite.

Dla rozwiązania dopuszczalnego, np. (1, 4, 5, 3, 2, 1), zmienne  $z_j$  cechują kolejne wierzchołki, przyporządkowując im wartości:  $z_1 = 0, z_4 = 1, z_5 = 2, z_3 = 3, z_2 = 4$ . Oczywiście miasto startu nie może mieć drugiej cechy „5”, dlatego warunek (8.25) nie dotyczy tras wchodzących do miasta „startowego”, czyli „1”.

Zadanie (8.22)–(8.27) jest zadaniem PML o dużych rozmiarach. Liczba zmiennych to:

$$L(z) = n + n(n-1) = n^2,$$

a liczba warunków to:

$$L(w) = 2n + n(n-1) - (n-1) = 2n + (n-1)^2.$$

Dla  $n = 10$  mamy więc 100 zmiennych oraz 101 warunków. Zadania o tak dużych rozmiarach trudno się rozwiązuje, dlatego opracowano wiele specjalnych algorytmów rozwiązania zagadnienia komiwojażera, opartych na modelu kombinatorycznym.

Niech miasto „1” będzie umownie punktem startu, a  $v = (v_1, v_2, \dots, v_{n-1})$  dowolną permutacją liczb naturalnych  $2, 3, \dots, n$ . Permutacja  $v = (v_1, v_2, v_3, v_4) = (3, 4, 2, 5)$  oznacza, że kolejno są odwiedzane miasta: „3”, „4”, „2” i „5” i wracamy do punktu startu.

Model kombinatoryczny można sformułować następująco: znajdź taką permutację (marszrutę)  $v^*$ , aby:

$$f(v^*) = \min \{f(v) | v \in D\}, \quad (8.28)$$

gdzie:

$D$  — zbiór wszystkich permutacji,

$f(v)$  — długość marszruty dla permutacji  $v$ .

## 8.6. Zagadnienie rozwózki

Często mamy do czynienia z sytuacją, kiedy pewien jednorodny produkt musi być rozwieziony od producenta do wielu jego odbiorców. Z cukrowni rozwiezimy wyprodukowany cukier, z mleczarni masło lub mleko, z browaru — piwo itd. Niekiedy mamy sytuację odwrotną, zwłaszcza w przemyśle spożywczym zakupiony surowiec od wielu jego producentów, trzeba przewieźć do zakładu, w którym nastąpi jego przeróbka. Tego typu zagadnienie nazywamy **zagadnieniami rozwózkowo-przywozowymi**. Dla uproszczenia rozważań przyjmijmy, że dalej omawiamy tylko zagadnienie rozwózki.

Zakładamy, że jest baza będąca miejscem produkcji jednorodnego towaru oraz postój parku transportowego. Zakładamy dalej, że dana jest liczba pojazdów o jednakowej ładowności. Znamy popyt każdego odbiorcy oraz macierz odległości (kosztu przewozu, czasu przewozu) między wszystkimi punktami odbioru.

Zakładamy, że popyt każdego odbiorcy jest mniejszy od ładowności pojazdu, a łączne zapotrzebowanie wszystkich punktów odbioru jest mniejsze od ładowności

całego parku transportowego. Przyjmujemy, że towar jest dostarczony do odbiorcy w okresie planistycznym (w dniu, tygodniu) przez jeden pojazd.

Należy ustalić taki zbiór marszrut, aby:

- 1) popyt każdego odbiorcy był zrealizowany przez jeden pojazd;
- 2) ładowność każdego pojazdu nie była przekroczona;
- 3) długość wszystkich marszrut była minimalna.

Wprowadźmy następujące oznaczenia:

- $P$  — zbiór wszystkich punktów odbioru,  
 $\bar{P}$  — zbiór wszystkich punktów odbioru i dostawy,  
 $n$  — liczba odbiorców (punktów odbioru),  
 $V$  — zbiór wszystkich połączeń między punktami,  
 $m$  — liczba pojazdów,  
 $l$  — ładowność każdego pojazdu,  
 $b_j$  — popyt  $j$ -tego odbiorcy,  
 $c_{ij}$  — odległość od punktu  $i$  do punktu  $j$  (długość trasy  $\langle i, j \rangle$ ).

$$P = \{1, 2, \dots, n\}, \quad \bar{P} = \{0, 1, \dots, n\},$$

$$V = \{\langle i, j \rangle | i, j \in \bar{P} \wedge i \neq j\}.$$

Zgodnie z przyjętymi założeniami dane te muszą spełniać warunki:

$$\sum_{j=1}^n b_j \leq ml, \quad (8.29)$$

$$b_j < l \quad (j \in P). \quad (8.30)$$

Dla każdego pojazdu wyznaczamy jedną marszrutę; łącznie będzie ich  $m$ .

Przyjmijmy następujące oznaczenia:

$x_{ij}$  — ilość towaru przewożona na trasie  $\langle i, j \rangle$ ,

$$y_{ij} = \begin{cases} 1, & \text{gdy pojazd przechodzi z miasta } i \text{ do miasta } j \text{ (pokonuje trasę } \langle i, j \rangle \text{)} \\ 0, & \text{w przypadku przeciwnym.} \end{cases}$$

Zadanie decyzyjne będzie miało postać: znajdź takie wartości zmiennych  $x_{ij}$  oraz  $y_{ij}$ , aby:

$$\sum_{\langle i, j \rangle \in V} c_{ij} y_{ij} \Rightarrow \min, \quad (8.31)$$

przy warunkach:

$$\sum_{i \in \bar{P}} y_{ij} = 1 \quad (j \in P), \quad (8.32)$$

$$\sum_{i \in \bar{P}} y_{i,j} = 1 \quad (j \in P), \quad (8.33)$$

$$\sum_{i \in \bar{P}} y_{i0} = m, \quad (8.34)$$

$$\sum_{i \in P} y_{0i} = m, \quad (8.35)$$

$$\sum_{i \in P} x_{ij} = \sum_{i \in P} x_{ji} + b_j \quad (j \in P), \quad (8.36)$$

$$\sum_{i \in P} x_{0i} = \sum_{i \in P} b_i, \quad (8.37)$$

$$x_{ij} \leq l y_{ij} \quad (\langle i, j \rangle \in V), \quad (8.38)$$

$$x_{ij} \geq 0 \quad (\langle i, j \rangle \in V), \quad (8.39)$$

$$y_{ij} \in \{0, 1\} \quad (\langle i, j \rangle \in V). \quad (8.40)$$

Zadanie (8.23)–(8.40) jest zadaniem PML, gdyż część zmiennych jest ciągła, a część to zmienne binarne. Wszystkie warunki i funkcje celu są liniowe.

Funkcja celu (8.31) minimalizuje długość wszystkich marszrut (przebytej drogi pojazdów). Warunki (8.32) i (8.33) zapewniają, że dla każdego odbiorcy wjeżdża i wyjeżdża jeden pojazd. Warunki (8.34) i (8.35) wymuszają, aby z bazy wyjechało i do niej wróciło dokładnie  $m$  pojazdów. Warunek (8.36) zapewnia, że w każdym punkcie zostawimy tyle, ile wynosi jego popyt, a warunek (8.37) pozwala wywieźć z bazy tyle towaru, ile wynosi łączny popyt odbiorców. Warunek (8.38) zapewnia, że na każdej trasie przewieziemy nie więcej towaru niż wynosi ładowność pojazdu. Jeżeli danej trasy pojazd nie pokonuje, to przewóz towaru na tej trasie jest zerowy.

Ważnym problemem są rozmiary zadania (8.31)–(8.40). Będzie ono miało  $L(z) = 2n(n+1)$  zmiennych oraz  $L(w) = 3n + n(n+1) + 3$  warunków.

Dla  $n = 30$  mamy 840 zmiennych oraz 483 warunki. W praktyce każdy punkt ma tylko bezpośrednie połączenie z kilkoma najbliższymi sąsiadami. Jeżeli przyjmujemy, że średnio jest ich 5, to zadanie redukuje się do wymiarów:

$$L(z) = 2 \cdot 5 \cdot 21 = 210,$$

$$L(w) = 3 \cdot 20 + 5 \cdot 21 + 3 = 168.$$

Musimy wówczas nieco zmodyfikować warunki: po trasach wychodzących z bazy może przechodzić więcej niż jeden pojazd.

Zagadnienie rozwózki można rozwiązać metodami ogólnymi opartymi na metodzie podziału i ograniczeń bądź specjalnymi algorytmami suboptymalnymi.

## 8.7. Idea metody podziału i ograniczeń

Dla dyskretnych zadań decyzyjnych podstawową i uniwersalną metodą ich rozwiązywania jest metoda podziału i ograniczeń.

Metoda ta nie jest ściśle ustalonym algorytmem postępowania podobnym np. do

metody sympleks lub metody potencjałów. Jest to raczej określone podejście do rozwiązywania pewnej klasy (dyskretnych) zadań optymalizacyjnych.

Aby wyjaśnić ideę metody podziału i ograniczeń, rozważmy dyskretne zadanie decyzyjne o postaci: znajdź takie rozwiązanie  $x^*$ , aby:

$$x^* \in D, \quad (8.41)$$

$$f(x^*) = \max \{f(x) | x \in D\}$$

lub:

$$x^* \in D, \quad (8.42)$$

$$f(x^*) = \min \{f(x) | x \in D\}.$$

Oznaczmy zadanie (8.41) symbolem  $P$ , a zadanie (8.42) symbolem  $P'$ .

Jeżeli zbiór rozwiązań dopuszczalnych  $D$  jest niewielki (zawiera kilkadziesiąt lub kilkaset elementów), to najprościej zastosować przegląd zupełny tego zbioru czyli:

- 1) wygenerować wszystkie jego elementy (rozwiązania);
- 2) ustalić wartość funkcji celu  $f(x)$  dla każdego rozwiązania;
- 3) wybrać rozwiązanie maksymalizujące lub minimalizujące wartość funkcji celu.

Zauważmy, że np. dla zagadnienia optymalnego przydziału przy  $n = 10$  zbiór  $D$  będzie zawierał  $n! = 10! = 3\,628\,800$  elementów. W tym przypadku przegląd zupełny, nawet przy zastosowaniu komputerów, byłby bardzo uciążliwy.

Idea metody podziału i ograniczeń polega na tym, że przegląd zupełny zbioru  $D$  zastępujemy przeglądem ukierunkowanym. Pozwala on ocenić pośrednio pewne podzbiory rozwiązań i je ewentualnie odrzucić lub czasowo pominąć bez straty rozwiązania optymalnego.

Zbiór  $D$  dzielimy stopniowo na coraz mniejsze podzbiory, dla których liczymy:

- 1) w zadaniu na maksimum  $P$  **kres górny**, czyli oszacowanie z góry wartości funkcji celu dla rozwiązań należących do danego podzbioru;
- 2) w zadaniu na minimum  $P'$  **kres dolny**, czyli oszacowanie z dołu wartości funkcji celu dla rozwiązań należących do danego podzbioru.

Następnie wybieramy **podzbiór perspektywiczny o maksymalnym (minimalnym) kresie górnym (dolnym)** oraz dokonujemy jego podziału, jeżeli jest to możliwe i celowe. Podzbiór, który uległ podziałowi, nazywamy **podzbiorem biernym**. Natomiast podzbiór, który nie został jeszcze podzielony, nazywamy **podzbiorem aktywnym**. Te ostatnie podzbiory dzielimy na:

- 1) podzbiory otwarte, które mogą ulec podziałowi;
- 2) podzbiory zamknięte, które nie ulegają podziałowi.

Które z podzbiorów otwartych wybierzemy do podziału w danym kroku obliczeń, zależy od ich kresów górnych (dolnych).

Funkcję  $w$  określoną na rodzinie  $2^D$  podzbiorów zbioru  $D$  nazywamy **funkcją ograniczającą**, jeżeli spełnia następujące warunki:

dla zadania  $P$ :

$$(a) x \in D_1 \subset D \Rightarrow f(x) \leq w(D_1),$$

$$(b) D_2 \subset D_1 \subset D \Rightarrow w(D_2) \leq w(D_1),$$

$$(c) D_1 = \{x\} \subset D \Rightarrow w(D_1) = f(x),$$

dla zadania  $P'$ :

$$f(x) \geq w(D_1),$$

$$w(D_2) \geq w(D_1),$$

$$w(D_1) = f(x).$$

Wartości, jakie przyjmuje funkcja ograniczająca  $w$  dla poszczególnych podzbiorów zbioru  $D$  nazywamy ich **kresami górnymi (dolnymi)**.

Warunki (a)–(c) nie określają bezpośrednio funkcji ograniczającej  $w$ . Przy wyborze konkretnej postaci funkcji  $w$  musimy brać pod uwagę jeszcze dwa postulaty:

- 1) łatwość wyznaczania jej wartości, czyli kresów górnych (dolnych) dla poszczególnych podzbiorów;
- 2) dokładność, z jaką kresy górne (dolne) szacują wartość funkcji  $f$ .

Efektywność metody podziału i ograniczeń zależy od jakości oszacowań, które ustalamy dla poszczególnych podzbiorów. Im te oszacowania z góry (z dołu) wartości funkcji celu  $f$  są dokładniejsze, tym mniej potrzeba podziałów, a tym samym szybciej można znaleźć rozwiązanie optymalne zadania  $P(P')$ . Funkcja  $w$ , której wartości łatwo ustalić, daje na ogół gorsze oszacowania niż funkcja, której wartości są wyznaczane bardziej pracochłonnymi metodami. Konieczny jest pewien kompromis między prostotą a dokładnością.

Idealna byłaby taka funkcja ograniczająca  $w$ , aby:

$$w(D) = \max \{f(x) | x \in D\} \text{ dla } D \in G,$$

gdzie  $G$  jest zbiorem wszystkich podzbiorów zbioru  $D$ .

W tym przypadku rozwiązanie optymalne może być uzyskane bardzo szybko.

Niech  $G_r$  będzie zbiorem wszystkich podzbiorów aktywnych w  $r$ -tym kroku obliczeń. Zbiorem perspektywicznym w  $r$ -tym kroku jest taki zbiór  $D_p$ , dla którego:

$$w(D_p) = \max \{w(D_i) | D_i \in G^r\} \text{ dla zadań na maksimum}$$

lub:

$$w(D_p) = \min \{w(D_i) | D_i \in G^r\} \text{ dla zadań na minimum.}$$

Jeżeli znajdziemy taki podzbiór  $D_k$  i związane z nim takie rozwiązanie  $\bar{x}$ , że:

$$\bar{x} \in D_k \wedge w(D_k) = f(\bar{x}),$$

to zbiór  $D_k$  zamykamy, gdyż nie zawiera on lepszego rozwiązania od  $\bar{x}$ , a element  $\bar{x}$  nazywamy rozwiązaniem **lokalnie najlepszym**.

Proces podziału zbioru  $D$  kontynuujemy tak długo, aż znajdziemy podzbiór **perspektywiczny zamknięty**. Wynika to z twierdzenia 8.1.

Jeżeli podzbiorem perspektywicznym jest podzbiór zamknięty  $D_p$  z elementem najlepszym  $x^*$ , to  $x^*$  jest optymalnym rozwiązaniem zadania  $P(P')$ .

Optymalność rozwiązania  $x^*$  dla zadania  $P$  wynika z następujących zależności:

$$f(x^*) = w(D_p) = \max \{w(D_i) | D_i \in G^r\} \geq \max \{f(x) | x \in D\}.$$

## 8.8. Metoda podziału i ograniczeń dla zadań PCL

Jeżeli mamy zadanie PCL:

$$cx \rightarrow \max, \quad (8.43)$$

$$Ax \leq b, \quad (8.44)$$

$$x \geq 0, \quad (8.45)$$

$$x \text{ — wektor całkowitoliczbowy,} \quad (8.46)$$

to najprościej próbować rozwiązać to zadanie, pomijając w nim warunek całkowitoliczbowości (8.46). Uzyskane w ten sposób zadanie PL (8.43)–(8.45) jest liniową relaksacją zadania PCL. Jeżeli zadanie PL nie ma rozwiązania optymalnego, to zadanie PCL także go nie ma. Jeżeli rozwiązanie optymalne zadania PL spełnia warunek całkowitoliczbowości, to jest ono także optymalnym rozwiązaniem zadania PCL.

Jeżeli rozwiązanie optymalne zadania PL nie jest całkowitoliczbowe, to można zastosować metodę podziału i ograniczeń. Dalej omówimy dokładnie ten przypadek.

Oznaczmy przez:

- $\tilde{D}$  — zbiór rozwiązań dopuszczalnych zadania PL,
- $D$  — zbiór rozwiązań dopuszczalnych zadania PCL,
- $x_0$  — wartość funkcji celu dla rozwiązania optymalnego zadania PL,
- $N(x_0)$  — największa liczba całkowita nie większa od  $x_0$ ,
- $\tilde{D} = \{x | Ax \leq b, x \geq 0\}$ ,
- $D = \{x | x \in \tilde{D}, x \text{ — wektor całkowitoliczbowy}\}$ ,
- $D \subset \tilde{D}$ ,
- $\max \{cx | x \in D\} \leq x_0 = \max \{cx | x \in \tilde{D}\}$ .

Podział zbioru  $D$  na coraz mniejsze podzbiory można przedstawić w postaci tzw. **drzewa podziału**. Z dowolnym  $l$ -tym węzłem (wierzchołkiem) drzewa są związane: zbiór  $D_l$ , zadanie  $PL^l$ , jego rozwiązanie optymalne  $x^l$ , kres górny w  $(D_l)$ . Dla każdego podzbioru  $D_l$  kres górny ustalamy zgodnie z wzorem:

$$w(D_l) = \begin{cases} x_0^l, & \text{jeżeli wektor wag } c \text{ nie jest całkowitoliczbowy,} \\ N(x_0^l), & \text{jeżeli wektor wag } c \text{ jest całkowitoliczbowy.} \end{cases} \quad (8.47)$$

Zbiór  $D_l$  zamykamy, jeżeli:

- 1) rozwiązanie optymalne zadania  $PL^l$  jest całkowitoliczbowe; wtedy  $w(D_l) = cx^l$ ;
- 2) zbiór rozwiązań dopuszczalnych  $D_l$  zadania  $PL^l$  jest pusty, to umownie przyjmujemy, że  $w(D_l) = -\infty$ .

W wyniku  $r$  podziałów zbioru  $D$  uzyskujemy  $2r$  jego podzbiorów. Jeżeli podzbiorem perspektywicznym  $r+1$  jest podzbiór otwarty  $D_p$ , to dzielimy go na dwa podzbiory  $D_{2r+1}$  oraz  $D_{2r+2}$ . Podstawą podziału jest pierwsza zmienna w rozwiązaniu optymalnym  $x^p$  niespełniająca warunku całkowitoliczbowości. Załóżmy, że jest to zmienna  $x_k$ . Wówczas:

$$N(x_k^p) < x_k^p < N(x_k^p) + 1,$$

$$D_{2r+1} = \{x | x \in D_p, \wedge x_k \leq N(x_k^p)\},$$

$$D_{2r+2} = \{x | x \in D_p, \wedge x_k \geq N(x_k^p) + 1\}.$$

Otrzymujemy więc dwa nowe zadania:  $P^{2r+1}$  oraz  $P^{2r+2}$  związane z tymi podzbiarami. W drzewie podziału pojawią się dwa nowe odpowiadające im węzły. Optymalne rozwiązanie zadań PL wyznaczają kresy górne nowych podzbiorów. Wśród podzbiorów aktywnych wybieramy podzbiór perspektywiczny o maksymalnym kresie górnym.

Proces podziału kontynuujemy tak długo, aż znajdziemy podzbiór perspektywiczny zamknięty. Związane z nim rozwiązanie całkowitoliczbowe jest rozwiązaniem optymalnym zadania PCL.

Jeżeli związany z tym podzbiorem kres górny jest równy  $-\infty$ , to zadanie PCL nie ma rozwiązania optymalnego.

### PRZYKŁAD 8.2

Rozważmy następujące zadanie PCL:

$$7x_1 + 6x_2 \rightarrow \max, \quad (8.48)$$

$$2x_1 + 3x_2 \leq 6, \quad (8.49)$$

$$4x_1 + 2x_2 \leq 8, \quad (8.50)$$

$$x_1, x_2 \geq 0, \quad (8.51)$$

$$x_1, x_2 \text{ — całkowite.} \quad (8.52)$$

Jeżeli interesuje nas rozwiązanie przybliżone, które w najgorszym przypadku różni się wartością funkcji celu od rozwiązania optymalnego o  $\varepsilon = 1$ , to takim rozwiązaniem jest punkt  $x^2$ , o którym już po pierwszym podziale można powiedzieć, że jest rozwiązaniem dopuszczalnym zadania PCL o dokładności  $\varepsilon = 1$ , gdyż na tym etapie podziału zbiorem perspektywicznym był zbiór  $D_1$ , a:

$$w(D_1) - f(x^2) = 15 - 14 = 1 \leq \varepsilon = 1.$$

## 8.9. Algorytm Little'a dla zagadnienia komiwojażera

Do wyznaczenia najkrótszej drogi zamkniętej, czyli marszruty, możemy wykorzystać algorytm Little'a, który należy do klasy algorytmów podziału i ograniczenia. Wyznaczenie drogi najkrótszej następuje przez ustalenie tras, z których ona się składa. Dalej omówimy tylko nowe, specyficzne dla algorytmu Little'a elementy.

**Standaryzacja macierzy odległości** polega na takim jej przekształceniu, aby w każdym wierszu i każdej kolumnie uzyskać co najmniej jedno zero, a wszystkie jej elementy pozostały nieujemne. W tym celu wyznaczamy element minimalny w każdym wierszu macierzy  $C$ :

$$a_i = \min_j \{c_{ij}\} \quad (i = 1, \dots, n)$$

oraz liczymy dla każdej kolumny wartość:

$$b_j = \min_i \{c_{ij} - a_i\} \quad (j = 1, \dots, n).$$

Elementy macierzy standaryzowanej  $C' = [c'_{ij}]$  są równe:

$$c'_{ij} = c_{ij} - a_i - b_j.$$

Suma współczynników  $a_i$  oraz  $b_j$  wyznacza kres dolny zbioru wszystkich rozwiązań  $D$ :

$$w(D_0) = \sum_{i=1}^n a_i + \sum_{j=1}^n b_j.$$

W macierzy  $C'$  mamy co najmniej  $n$  elementów zerowych. Dla każdej trasy „zerowej” ustalamy koszt rezygnacji z niej. Dla trasy  $\langle g, p \rangle$  jest on równy sumie minimalnych elementów w  $g$ -tym wierszu i  $p$ -tej kolumnie, czyli:

$$d_{gp} = \min_{j \neq p} \{c'_{gj}\} + \min_{i \neq g} \{c'_{i,p}\}.$$

Oczywiście pomijamy sam element  $c'_{gp} = 0$ .

Następnie wybieramy trasę centralną o maksymalnym koszcie rezygnacji. Trasa centralna  $\langle k, l \rangle$  jest podstawą podziału zbioru  $D$  na dwa podzbiory:

$$D_1 = \{v | v \in D \wedge \langle k, l \rangle \in T(v)\},$$

$$D_2 = \{v | v \in D \wedge \langle k, l \rangle \notin T(v)\},$$

gdzie  $T(v)$  jest zbiorem tras tworzących marszrutę  $v$ .

W wyniku  $r$  podziałów otrzymujemy  $2r$  podzbiorów  $D$ . Jeżeli podzbiorem perspektywicznym  $r+1$  jest podzbiór otwarty  $D_r$ , dzielimy go na dwa podzbiory  $D_{2r+1}$  i  $D_{2r+2}$ .

Kres dolny dla zbioru  $D_r$  ustalamy zgodnie ze wzorem:

$$w(D_r) = w(D) + d_{k,l}.$$

W celu ustalenia kresu dolnego dla zbioru  $D_r$  tworzymy macierz zredukowaną  $C'_r$  przez następujące czynności:

- 1) z macierz  $C'$  wykreślamy  $k$ -ty wiersz i  $l$ -tą kolumnę;
- 2) blokujemy trasę umożliwiającą powstanie podcykli;
- 3) dokonujemy standaryzacji tak uzyskanej macierzy  $C'_r$ .

Kres dolny dla  $D_r$ :

$$w(D_r) = w(D) + s_1,$$

gdzie  $s_1$  jest sumą współczynników standaryzujących macierz  $C'_r$ .

Jeżeli w wyniku  $r+1$  podziału dzielony jest podzbiór  $D_r$  na podstawie trasy centralnej  $\langle v, u \rangle$ , to:

$$w(D_{2r+2}) = w(D_r) + d_{v,u},$$

$$w(D_{2r+1}) = w(D_r) + s_2.$$

Spośród podzbiorów aktywnych wybieramy podzbiór perspektywiczny o minimalnym kresie dolnym. Jeżeli jest on otwarty, to go dzielimy, ustalając nową trasę centralną i kresy dolne dla nowych podzbiorów.

Gdy zbiór perspektywiczny jest jednoelementowy, wtedy zamykamy go i kończymy obliczanie. Zbiór perspektywiczny jednoelementowy zawiera optymalne rozwiązanie (minimalną drogę). Zbiorowi temu odpowiada zredukowana macierz odległości o wymiarach  $[2 \times 2]$ .

### PRZYKŁAD 8.3

Handlowiec pewnej firmy ma odwiedzić klientów znajdujących się w czterech miastach i wrócić do siedziby firmy, która mieści się w pobliżu miasta „1”. Macierz odległości  $C$  podana jest w tablicy 8.2. Macierz odległości nie jest symetryczna,



a trasy niedopuszczalne  $\langle i, i \rangle$  zostały zablokowane przez przyjęcie, że  $c_{ii} = \infty$ . Kolejność odwiedzanych miast jest dowolna. Należy wyznaczyć najkrótszą drogę zamkniętą przechodzącą przez cztery miejscowości.

Najpierw standaryzujemy macierz  $C$ . Współczynniki  $a_i$  i  $b_j$  zostały podane w tablicy 8.3 wraz z macierzą  $C'$ . Suma współczynników standaryzujących macierzy  $C$  to  $s = 21$ , stąd kres dolny zbioru  $D$  wynosi  $w(D) = 21$ . Wynika więc, że nie istnieje droga zamknięta, marszruta o długości krótszej od 21.

TABLICA 8.2  
Macierz  $C$

| $i \backslash j$ | 1        | 2        | 3        | 4        |
|------------------|----------|----------|----------|----------|
| 1                | $\infty$ | 3        | 8        | 5        |
| 2                | 10       | $\infty$ | 10       | 8        |
| 3                | 11       | 5        | $\infty$ | 8        |
| 4                | 4        | 11       | 5        | $\infty$ |

TABLICA 8.3  
Macierz  $C'$

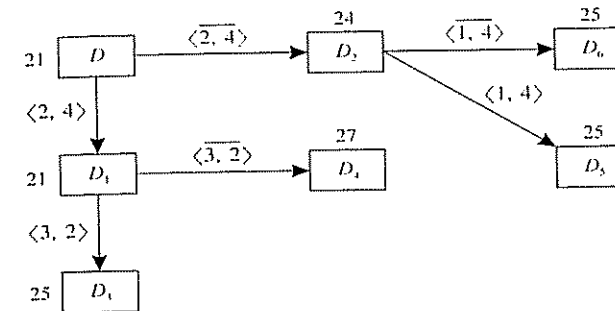
| $i \backslash j$ | 1              | 2              | 3              | 4              | $a_i$ |
|------------------|----------------|----------------|----------------|----------------|-------|
| 1                | $\infty$       | 0 <sup>2</sup> | 4              | 2              | 3     |
| 2                | 2              | $\infty$       | 1              | 0 <sup>3</sup> | 8     |
| 3                | 6              | 0 <sup>3</sup> | $\infty$       | 3              | 5     |
| 4                | 0 <sup>2</sup> | 7              | 0 <sup>4</sup> | $\infty$       | 4     |
| $b_j$            | 0              | 0              | 1              | 0              | 21    |

Dla tras zerowych w macierzy  $C'$  liczymy koszty rezygnacji. Zostały one podane w tablicy 8.3 powyżej zmodyfikowanych odległości. Maksymalny koszt rezygnacji równy 3 jest dla dwóch tras:  $\langle 2, 4 \rangle$  i  $\langle 3, 2 \rangle$ . Jako centralną wybierzemy pierwszą z nich  $\langle 2, 4 \rangle$ . Trasa  $\langle 2, 4 \rangle$  jest podstawą podziału zbioru  $D$  na dwa podzbiory  $D_1$  i  $D_2$ . Zbiór  $D_1$  składa się z dróg przechodzących przez trasę  $\langle 2, 4 \rangle$ , a zbiór  $D_2$  — z dróg, które nie przechodzą przez odcinek  $\langle 2, 4 \rangle$ . Drzewo podziału zbioru  $D$  zawiera rysunek 8.5. Na tym drzewie zapis  $\langle 2, 4 \rangle$  oznacza, że drogi tworzące zbiór  $D_2$  nie zawierają odcinka  $\langle 2, 4 \rangle$ .

Kres dolny zbioru  $D_2$   $w(D_2) = 21 + 3 = 24$ . Aby ustalić kres dolny podzbioru  $D_1$ , usuwamy z macierzy  $C'$  wiersz „2” i kolumnę „4” oraz blokujemy trasę  $\langle 4, 2 \rangle$  przyjmując, że  $c_{4,2} = \infty$ . Element blokujący możliwość tworzenia podcykli „ $\infty$ ” umieszczamy tak, aby w macierzy zredukowanej w każdym wierszu i każdej kolumnie był jeden symbol „ $\infty$ ”. Macierz  $C'_1$  jest podana w tablicy 8.4. Suma standaryzujących współczynników macierzy  $C_1$  jest zerowa, stąd:  
 $w(D_1) = 21 + 0 = 21$ .

Zauważmy, że  $s_1 = d_{4,2} = 0$ , czyli suma współczynników standaryzujących macierzy  $C_1$  jest równa kosztowi rezygnacji trasy blokowanej  $\langle 4, 2 \rangle$ . Dla tras „niezerowej” koszt rezygnacji jest zawsze równy zero (dlatego ich nie liczymy).

RYSUNEK 8.5  
Drzewo podziału zbioru rozwiązań



TABLICA 8.4  
Macierz  $C'_1$

| $i \backslash j$ | 1              | 2              | 3              | $a_i$ |
|------------------|----------------|----------------|----------------|-------|
| 1                | $\infty$       | 0 <sup>4</sup> | 4              | 0     |
| 3                | 6              | 0 <sup>6</sup> | $\infty$       | 0     |
| 4                | 0 <sup>6</sup> | $\infty$       | 0 <sup>4</sup> | 0     |
| $b_j$            | 0              | 0              | 0              | 0     |

Mamy aktualnie dwa podzbiory aktywne  $D_1$  i  $D_2$ ; jako zbiór perspektywiczny wybieramy zbiór o minimalnym kresie, czyli  $D_1$ . Odpowiada mu macierz standaryzowana  $C'_1$ .

W macierzy  $C'_1$  dla tras zerowych liczymy koszt rezygnacji oraz ustalamy koszt maksymalny, który jest równy 6 dla dwóch tras  $\langle 3, 2 \rangle$  i  $\langle 4, 1 \rangle$ . Jako trasę centralną wybieramy pierwszą z nich  $\langle 3, 2 \rangle$  i zbiór  $D_1$  dzielimy na dwa podzbiory:  $D_3$  i  $D_4$ .

Obliczmy dla nich kresy dolne:

$$w(D_3) = 21 + 4 = 25,$$

$$w(D_4) = 21 + 6 = 27,$$

gdź koszt rezygnacji z trasy  $\langle 4, 3 \rangle$ , którą musimy zablokować, wynosi  $d_{4,3} = 4$ .

Z trzech podzbiorów aktywnych  $D_3$ ,  $D_4$  i  $D_2$  jako zbiór perspektywiczny wybieramy zbiór  $D_2$ , dla którego  $w(D_2) = 24$  jest minimalne.

Nie znamy związanej z nim macierzy  $C_2$ , więc bierzemy macierz  $C'$  związaną

ze zbiorem  $D$ , który jest poprzednikiem  $D_2$  w drzewie podziału. W tej macierzy blokujemy trasę  $\langle 2, 4 \rangle$ , przyjmując  $c_{2,4} = M$  (gdzie  $M$  to duża liczba), i dokonywana jest standaryzacja. Macierz  $C'_2$  podana jest w tabelicy 8.5.

Dla tras „zerowych” ustalamy koszt rezygnacji. Dla kilku tras koszt jest maksymalny, więc wybieramy pierwszą z nich, czyli trasę  $\langle 1, 4 \rangle$  i zbiór  $D_2$  dzielimy na dwa podzbiory:  $D_5$  i  $D_6$  oraz liczymy dla nich kresy dolne:

$$w(D_5) = 24 + 1 = 25,$$

$$w(D_6) = 24 + 1 = 25.$$

TABLICA 8.5  
Macierz  $C'_2$

| $i \backslash j$ | 1        | 2        | 3        | 4        | $a_i$ |
|------------------|----------|----------|----------|----------|-------|
| 1                | $\infty$ | $0^0$    | 4        | $0^1$    | 0     |
| 2                | 1        | $\infty$ | $0^1$    | $M$      | 1     |
| 3                | 6        | $0^1$    | $\infty$ | 1        | 0     |
| 4                | $0^1$    | 7        | $0^0$    | $\infty$ | 0     |
| $b_j$            | 0        | 0        | 0        | 2        | 3     |

TABLICA 8.6  
Macierz  $C'_3$

| $i \backslash j$ | 1        | 2        | $a_i$ |
|------------------|----------|----------|-------|
| 1                | $\infty$ | 0        | 4     |
| 4                | 0        | $\infty$ | 0     |
| $b_j$            | 0        | 0        | 4     |

Aktualnie mamy trzy podzbiory z minimalnym kresem dolnym  $D_3$ ,  $D_5$  i  $D_6$ . Jako perspektywiczny zbiór wybieramy zbiór najmniej liczny, czyli  $D_3$  (jest to zbiór jednoelementowy), i go zamykamy. Uzyskaliśmy rozwiązanie optymalne. Składa się ono z tras:  $\langle 2, 4 \rangle$ ,  $\langle 3, 2 \rangle$  oraz  $\langle 1, 3 \rangle$ ,  $\langle 4, 1 \rangle$ . Te dwie ostatnie trasy odczytujemy z tabelicy 8.6, która zawiera macierz  $C'_3$ . Optymalna-droga to  $\langle 1, 3 \rangle \langle 3, 2 \rangle \langle 2, 4 \rangle$ ,  $\langle 4, 1 \rangle$  o długości 25. Handlowiec może odwiedzać miasta w kolejności: „1”, „3”, „2”, „4” lub „3”, „2”, „4”, „1”. Siedziba firmy jest bowiem w pobliżu miasta „1”.

## 8.10. Algorytm Johnsona dla ustalenia kolejności obróbki detali

Z rozważań zawartych w podrozdziale 8.4 wynika, że harmonogramowanie prac w przypadku ogólnym jest problemem trudnym do rozwiązania. Są jednak przypadki szczególne, dla których łatwo można znaleźć rozwiązanie optymalne.

Przyjmijmy, że mamy wydział obróbki, na którym  $n$  detali ma przejść proces obróbki na dwóch obrabiarkach: najpierw na  $M_1$ , a potem na  $M_2$ . Znamy czas obróbki każdego detalu na każdej maszynie. Należy ustalić taką kolejność obróbki  $n$  detali, aby czas wykonania całej partii był minimalny.

Oznaczmy przez:

- $a_j$  — czas obróbki  $j$ -tego detalu na maszynie  $M_1$ ,
- $b_j$  — czas obróbki  $j$ -tego detalu na maszynie  $M_2$ .

Niech  $v = (v_1, v_2, \dots, v_n)$  oznacza dowolną permutację liczb naturalnych 1, 2, ...,  $n$ . Jeżeli:

$$v = (v_1, v_2, v_3, v_4) = (4, 2, 1, 3),$$

oznacza to, że jako pierwszy jest wykonany detal o numerze 4, a potem detal o numerze 2, następnie o numerze 1, a na końcu o numerze 3.

Do wyznaczenia optymalnej kolejności obróbki stosujemy algorytm Johnsona, który polega na realizacji następujących kroków:

Krok 1. Przyjmij  $r = 1$ ,  $s = n$ .

Krok 2. W aktualnej macierzy czasów obróbki znajdź minimalny element. Jeżeli jest to czas  $a_k$ , to  $v_r = k$  oraz  $r := r + 1$ . Jeżeli natomiast jest to czas  $b_k$ , to  $v_s = k$  oraz  $s := s - 1$ .

Krok 3. Dokonaj redukcji macierzy czasów, wykreślając z niej  $k$ -tą kolumnę. Jeżeli zredukowana macierz nie zawiera żadnej już kolumny, przejdź do kroku 4. W przeciwnym przypadku wróć do kroku 2.

Krok 4. Koniec procedury — została ustalona optymalna kolejność.

### PRZYKŁAD 8.4

Pięć detali ma przejść proces obróbki na dwóch maszynach:  $M_1$  i  $M_2$ . Czasy obróbki zawiera tablica 8.7. Należy ustalić taką kolejność obróbki, aby łączny czas wykonania był minimalny.

TABLICA 8.7  
Macierz czasów obróbki

| Detale | 1     | 2     | 3     | 4     | 5     |
|--------|-------|-------|-------|-------|-------|
| $M_1$  | 3     | $4^3$ | 6     | 3     | $6^5$ |
| $M_2$  | $2^2$ | 5     | $4^4$ | $1^1$ | 7     |

Iteracja 1.  $r = 1$ ,  $s = 5$ . Minimalny czas  $b_4 = 1$ ,  $v_5 = 4$ ,  $s = 5 - 1 = 4$  kolejność cząstkowa  $v = (v_1, v_2, v_3, v_4, v_5) = (, , , , 4)$ . Usuwamy z macierzy czasów

kolumnę czwartą, co zaznaczamy w polu  $\langle 2, 4 \rangle$  małą cyfrą w prawym górnym rogu.

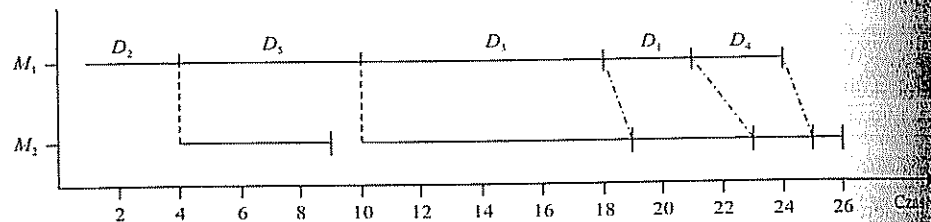
Iteracja 2. W zredukowanej macierzy czasów minimalny czas  $b_1 = 2$ ,  $v_4 = 1$ ,  $s = 4 - 1 = 3$ . Usuwamy kolumnę pierwszą, a kolejność cząstkowa  $v = (, , , 1, 4)$ .

Iteracja 3. W zredukowanej macierzy mamy dwa minimalne elementy  $a_2 = 4$  oraz  $b_3 = 4$ . Wybierzemy pierwszy z nich:  $v_1 = 2$ ,  $r = 1 + 1 = 2$ . Usuwamy kolumnę drugą, kolejność cząstkowa  $v = (2, , , 1, 4)$ .

Iteracja 4. W zredukowanej macierzy czasów minimalny element  $b_3 = 4$ ,  $c_3 = 3$ ,  $s = 3 - 1 = 2$ . Usuwamy kolumnę trzecią, a kolejność cząstkowa  $v = (2, 3, 1, 4)$ .

Iteracja 5. W zredukowanej macierzy minimalny element  $a_5 = 6$ ,  $v_2 = 5$ . Usuwamy kolumnę piątą, kolejność zupełna  $v = (2, 5, 3, 1, 4)$ . Ponieważ wykres Gantta została ostatnia kolumna, to uzyskaliśmy kompletną kolejność, która jest optymalna. Wykres Gantta dla optymalnej kolejności  $v = (2, 5, 3, 1, 4)$  zawiera rysunek 8.6. Łączny, minimalny czas obróbki wynosi 26 godzin.

RYSUNEK 8.6  
Wykres Gantta dla optymalnej kolejności obróbki



Większość algorytmów wyznaczenia harmonogramów jest albo bardzo pracochłonna, albo gwarantuje tylko rozwiązanie przybliżone.

#### POLECENIA I PTANIA KONTROLNE

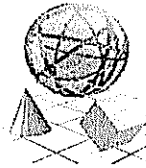
1. Jakie są dwa podstawowe podejścia do rozwiązywania problemów dyskretnych?
2. Co zasadniczo różni liniowy model zagadnienia przydziału od liniowego modelu zagadnienia komiwojażera?
3. Od czego zależy efektywność metod podziału i ograniczeń?
4. Jak można wykorzystać algorytm Little'a dla wyznaczania rozwiązań przybliżonych o zadanej dokładności?

5. Czym różni się zagadnienie rozwózki od zagadnienia zwózki?
6. Czy do ustalenia optymalnej kolejności obróbki można stosować algorytm podziału i ograniczeń? Jak wówczas dzielimy podzbiory perspektywiczne?

#### BIBLIOGRAFIA

- Czerwiński Z. (red.), 1983, *Optymalna organizacja złożonych działań*, PWN, Warszawa.  
 Ford L.R., Fullkerson D.R., 1969, *Przepływy w sieciach*, PWE, Warszawa.  
 Gale D., 1969, *Teoria liniowych modeli ekonomicznych*, PWN, Warszawa.  
 Ignasiak E. (red.), 2001, *Badania operacyjne*, PWE, Warszawa.  
 Jaworski K.M., 1999, *Metodologia projektowania realizacji budowy*, Wydawnictwo Naukowe PWN, Warszawa.  
 Sysło M., Deo N., Kowalik J., 1993, *Algorytmy optymalizacji dyskretnej*, Wydawnictwo Naukowe PWN, Warszawa.  
 Trzaskalik T., 2003, *Wprowadzenie do badań operacyjnych z komputerem*, PWE, Warszawa.  
 Wagner H.M., 1980, *Badania operacyjne*, PWE, Warszawa.

## ROZDZIAŁ 9

PROGRAMOWANIE  
W WARUNKACH RYZYKA

## 9.1. Wprowadzenie

W dotychczasowych sytuacjach decyzyjnych zakładaliśmy zazwyczaj, że podejmujemy decyzje w **warunkach pewności**, tzn. że wynik podjętej decyzji jest pewny, w postaci ściśle określonego czasu, zysku, kosztu, przychodu itd. Często w rzeczywistości wynik podjętej decyzji zależy od **stanu otoczenia**, który dalej nazywać będziemy **stanem natury**. Dla rolnictwa takim naturalnym stanem natury jest stan pogody, dla przedsiębiorcy — stan gospodarki, dla inwestora lokującego swoje pieniądze na giełdzie — sytuacja na giełdzie.

## DEFINICJA 9.1

Z podejmowaniem **decyzji w warunkach ryzyka** mamy do czynienia wówczas, gdy wynik podjętej decyzji nie jest jednoznaczny i zależy od tego, jaki wystąpi stan natury. Zakładamy, że potrafimy określić możliwe stany natury i znamy lub możemy oszacować prawdopodobieństwa ich wystąpienia.

## DEFINICJA 9.2

Z podejmowaniem **decyzji w warunkach niepewności** mamy do czynienia wówczas, gdy wynik podjętej decyzji zależy także od stanu natury, ale nie znamy prawdopodobieństwa wystąpienia stanów natury.

## DEFINICJA 9.3

Model matematyczny opisujący sytuację podejmowania decyzji w warunkach ryzyka nazywamy **modelem stochastycznym**.

## DEFINICJA 9.4

**Programowaniem w warunkach ryzyka** nazywamy procedurę wyznaczania optymalnej decyzji w warunkach ryzyka, stosując modele oraz metody formalne.

W podejmowaniu decyzji w warunkach ryzyka możemy stosować znane już modele i metody, np. zadanie PL, metodę simpleks, metodę geometryczną, a także modele oraz metody specyficzne dla tej klasy problemów.

9.2. Problem ogrodnika  
— ryzyko decyzji

## PRZYKŁAD 9.1

Ogrodnik posiadający 4 ha ziemi specjalizuje się w uprawie pomidorów ( $P$ ) i cebuli ( $C$ ). Plony zależą od stanu pogody. Zostały wyróżnione trzy stany pogody:  $S_1$ ,  $S_2$ ,  $S_3$ . Dla każdego stanu ustalono plony pomidorów i cebuli (w t/ha). Znamy prawdopodobieństwo wystąpienia stanu pogody (tablica 9.1). Znamy także cenę sprzedaży 1 kg pomidorów — 3 zł oraz 1 kg cebuli — 2 zł. Ze względów agrotechnicznych obszar uprawy pomidorów nie może być większy od 3 ha. Nakład robocizny związany ze zbiorem pomidorów wynosi 25 roboczogodzin na 1 ha, a nakład robocizny na 1 ha zbioru cebuli 50 roboczogodzin. Ogrodnik może w okresie zbioru przeznaczyć maksymalnie 150 roboczogodzin. Chciałby on zmaksymalizować osiągnięty przychód ze sprzedaży plonów.

TABLICA 9.1  
Dane liczbowe

| Stany pogody | $P(x)$ | Plony z 1 ha |     |
|--------------|--------|--------------|-----|
|              |        | $P$          | $C$ |
| $S_1$        | 0,2    | 90           | 80  |
| $S_2$        | 0,5    | 60           | 50  |
| $S_3$        | 0,3    | 30           | 40  |

Jak łatwo zauważyć, plony obu roślin są zmiennymi losowymi o znanym rozkładzie, czyli mamy do czynienia z podejmowaniem decyzji w warunkach ryzyka. Nie wiemy, jaki przychód osiągnie ogrodnik w przyszłym roku. W tej sytuacji kryterium wyboru optymalnej decyzji może być maksymalizacja oczekiwanego przychodu.

Oznaczamy przez  $E(P)$  oczekiwany plon pomidorów (w t/ha), a przez  $E(C)$  — oczekiwany plon cebuli (w t/ha):

$$E(P) = 90 \cdot 0,2 + 60 \cdot 0,5 + 30 \cdot 0,3 = 18 + 30 + 9 = 57.$$

$$E(C) = 80 \cdot 0,2 + 50 \cdot 0,5 + 40 \cdot 0,3 = 16 + 25 + 12 = 53.$$

Oznaczmy dalej przez  $x_1$  obszar uprawy pomidorów (w ha), a przez  $x_2$  — obszar uprawy cebuli (w ha). Wówczas oczekiwany przychód możemy zapisać:

$$c_1 \cdot E(P) \cdot x_1 + c_2 \cdot E(C) \cdot x_2 = 3 \cdot 57x_1 + 2 \cdot 53x_2.$$

Jeżeli  $c_1$  ( $c_2$ ), czyli cena jednej tony pomidorów (cebuli) w tys. zł, to oczekiwany przychód będzie wyrażony w tys. zł, gdyż:

$$c_1 \cdot E(P) \cdot x_1 = \text{oczekiwany przychód z uprawy pomidorów.}$$

$$\frac{\text{tys. zł}}{1 \text{ t}} \cdot \frac{\text{t}}{1 \text{ ha}} \cdot \text{ha} = \text{tys. zł.}$$

Wybór optymalnej strategii uprawy sprowadza się do rozwiązania następującego zadania:

$$f(x) = 171x_1 + 106x_2 \rightarrow \max, \quad (9.1)$$

przy warunkach:

$$x_1 + x_2 \leq 4, \quad (9.2)$$

$$25x_1 + 50x_2 \leq 150, \quad (9.3)$$

$$x_1 \leq 3, \quad (9.4)$$

$$x_1, x_2 \geq 0. \quad (9.5)$$

Zadanie (9.1)–(9.5) jest zadaniem PL z dwoma zmiennymi decyzyjnymi, które łatwo rozwiązać metodą geometryczną (rysunek 9.1).

Optymalnym jego rozwiązaniem jest  $x_1 = 3$ ,  $x_2 = 1$ ,  $f(x) = 171 \cdot 3 + 106 \cdot 1 = 513 + 106 = 619$ , czyli ogrodnik powinien uprawiać 3 ha pomidorów, 1 ha cebuli, a wówczas oczekiwany przychód wyniesie 619 tys. zł. Jaki będzie jednak rzeczywisty przychód ogrodnika zależy także od tego, który stan pogody wystąpi:

$$S1: 3 \cdot 90 \cdot 3 + 2 \cdot 80 \cdot 1 = 970,$$

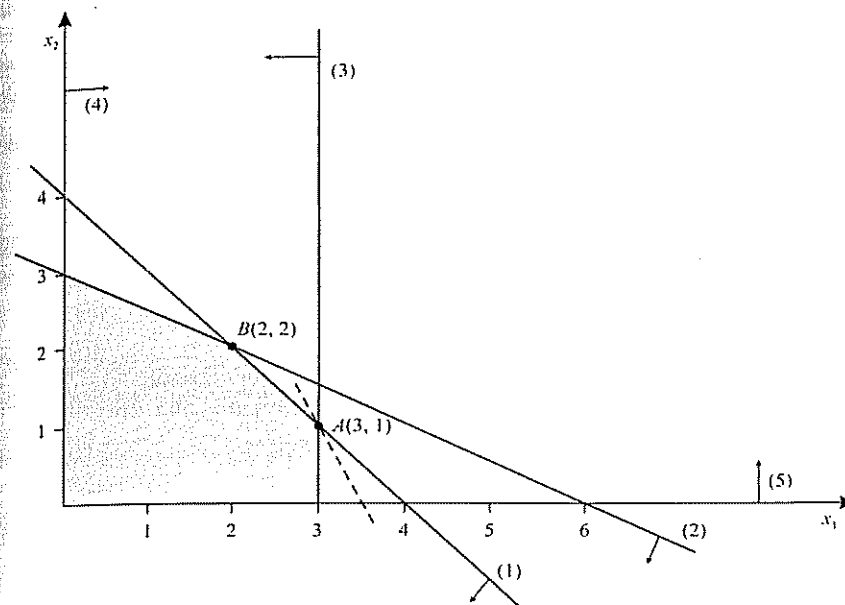
$$S2: 3 \cdot 60 \cdot 3 + 2 \cdot 50 \cdot 1 = 640,$$

$$S3: 3 \cdot 30 \cdot 3 + 2 \cdot 40 \cdot 1 = 350.$$

Najwyższy przychód byłby dla stanu pogody S1 (970 tys. zł), a najmniejszy dla stanu S3 (tylko 350 tys. zł).

Cena sprzedaży kształtuje się na rynku i zależy od wielkości podaży i popytu. O ile popyt jest raczej stabilny, o tyle podaż zależy od decyzji ogrodnika oraz od stanu pogody, który wpływa na średni plon. Nie tylko plon jest zmienną losową

RYSUNEK 9.1  
Metoda geometryczna dla zadania PL



lecz także sama cena. Załóżmy dalej, że przychód z 1 ha jest wynikiem nie tylko zmiennej wydajności, lecz także zmiennej ceny warzywa. Z reguły im niższe plony, tym wyższa cena i odwrotnie.

W tabelicy 9.2 są podane przychody z 1 ha dla poszczególnych stanów pogody i obu upraw (w tys. zł). Rozkład prawdopodobieństw nie ulega zmianie.

Oczekiwany przychód (w tys. z 1 ha) dla obydwu warzyw wynosi odpowiednio:

$$E(P) = 90 \cdot 0,2 + 180 \cdot 0,5 + 150 \cdot 0,3 = 153,$$

$$E(C) = 120 \cdot 0,2 + 100 \cdot 0,5 + 120 \cdot 0,3 = 110.$$

TABLICA 9.2  
Przychody z 1 ha dla poszczególnych stanów pogody

| Stany pogody | Rozkład prawdopodobieństw | Pomidory           | Cebula               |
|--------------|---------------------------|--------------------|----------------------|
| S1           | 0,2                       | $1 \cdot 90 = 90$  | $1,5 \cdot 80 = 120$ |
| S2           | 0,5                       | $3 \cdot 60 = 180$ | $2 \cdot 50 = 100$   |
| S3           | 0,3                       | $5 \cdot 30 = 150$ | $3 \cdot 40 = 120$   |

Oczekiwane przychody z Iha są nieco inne niż w pierwszym przypadku — dla pomidorów 153 tys. zł (wcześniej 171), a dla cebuli 110 tys. zł (wcześniej 106). Funkcja celu w zadaniu miałaby postać:

$$f(x) = 153x_1 + 110x_2 \rightarrow \max.$$

Jak łatwo sprawdzić, rozwiązanie optymalne dla funkcji celu (9.6) nie ulega zmianie. Jest nim nadal punkt  $A(3, 1)$  z oczekiwanym przychodem równym:

$$153 \cdot 3 + 110 \cdot 1 = 459 + 110 = 569.$$

Dla punktu  $B(2, 2)$  oczekiwany przychód wynosi:

$$153 \cdot 2 + 110 \cdot 2 = 306 + 220 = 526$$

i jest niższy o 43 tys. zł od maksymalnego, oczekiwanego przychodu.

Rzeczywisty przychód zależy jednak od stanu pogody i dla obu decyzji jest podany w tabelicy 9.3.

TABLICA 9.3  
Przychody dla dwóch rozwiązań

| Stany pogody         | Punkt $A(3, 1)$   | Punkt $B(2, 2)$   |
|----------------------|-------------------|-------------------|
| $S_1$                | $270 + 120 = 390$ | $180 + 240 = 420$ |
| $S_2$                | $540 + 100 = 640$ | $360 + 200 = 560$ |
| $S_3$                | $450 + 120 = 570$ | $300 + 240 = 540$ |
| $E(\text{Przychód})$ | 569               | 526               |

Zauważmy, że dla decyzji  $A$  różnica między skrajnymi przychodami wynosi 250 tys. zł, a dla decyzji  $B$  tylko 140 tys. zł, czyli ryzyko podjęcia decyzji optymalnej  $A$  jest większe niż ryzyko związane z decyzją  $B$ . Ponadto dla stanu pogody  $S_1$  lepszą decyzją jest decyzja  $B$ . Przynosi ona 420 tys. zł, gdy decyzja  $A$  tylko 390 tys. zł.

Z tego zadania wynika:

- 1) mimo że model II jest nadal uproszczonym obrazem podejmowania decyzji w warunkach ryzyka, jest to obraz znacznie bliższy rzeczywistości;
- 2) model I prowadzi do wyższego oczekiwanego przychodu, ale o bardzo dużej różnicy między skrajnymi realizacjami ( $970 - 350 = 620$ );
- 3) jeżeli przyjmiemy, że stan pogody wpływa na plony roślin, a te na ceny sprzedaży (cena rośnie, jeżeli wydajność maleje), to radykalny spadek zmienności wyniku dla modelu II jest oczywisty.

Do wyznaczenia optymalnej decyzji wykorzystany był deterministyczny aparat w postaci zadań PL i metody geometrycznej. W dalszej części rozdziału omówimy zagadnienia podejmowania decyzji w warunkach ryzyka, wymagające już innego instrumentarium.

### 9.3. Zagadnienie gazeciarza

Gazeciarz kupuje w hurtowni pewną liczbę gazet w określonej cenie. Następnie sprzedaje je nabywcom po wyższej cenie. Gazeta nie sprzedana jest zwracana do hurtowni po cenie zwrotu, niższej od ceny nabycia. Liczba gazet, jaką nabywa gazeciarz, jest dowolna, a wielkość sprzedaży zależy od popytu na gazety, który jest zmienną losową o znanym rozkładzie.

Jeżeli gazeciarz kupi zbyt mało gazet w stosunku do popytu, to traci na skutek nie wykorzystanej możliwości. Jeżeli kupi zbyt dużo gazet w stosunku do popytu, traci na gazetach, które nie zostały sprzedane. Należy ustalić, ile gazet powinien kupić gazeciarz, aby oczekiwany dochód ze sprzedaży był maksymalny.

Oznaczamy przez:

- $c_1$  — cenę zakupu gazety,
- $c_2$  — cenę sprzedaży gazety nabywcom,
- $c_3$  — cenę zwrotu gazety do hurtowni,
- $z$  — podaż gazet (liczba gazet zakupionych przez gazeciarza),
- $X$  — popyt na gazety (zmienna losowa dyskretna),
- $p(x)$  — prawdopodobieństwo, że zmienna losowa  $X$  przyjmie wartość  $x$ , czyli że popyt będzie równy  $x$ ,
- $F(x)$  — dystrybuanta rozkładu, czyli prawdopodobieństwo, że zmienna losowa  $X$  przyjmie wartość nie większą od  $x$ :

$$F(x) = P(X \leq x).$$

Zakładamy, że ceny spełniają warunek:

$$c_3 < c_1 < c_2.$$

Gdyby  $c_3 > c_1$ , wówczas gazeciarz zarabiałby na każdej gazecie, a optymalna podaż gazet musiałaby być równa maksymalnemu popytowi na nie.

Gazeta sprzedana przynosi zysk:

$$b = c_2 - c_1,$$

natomiast na gazecie, która nie znajduje nabywcy, gazeciarz ponosi stratę:

$$s = c_1 - c_3.$$

Jeżeli  $n$  to minimalny, a  $N$  to maksymalny popyt na gazety, to wówczas  $D = \{n, n+1, \dots, N\}$  jest zbiorem możliwych realizacji zmiennej losowej  $X$ .

Oznaczmy jeszcze przez:

$g(z, x)$  — dochód osiągany przez gazeciarza, jeżeli posiada  $z$  gazet, a popyt jest równy  $x$ :

$$g(z, x) = \begin{cases} bz, & \text{dla } z \leq x, \\ bx - s(z - x), & \text{dla } z > x, \end{cases} \quad (9.7)$$

$d(z)$  — oczekiwany dochód gazeciarza, jeżeli posiada  $z$  gazet:

$$d(z) = \sum_{x=n}^{z-1} [bx - s(z-x)]p(x) + \sum_{x=z}^N bzp(x). \quad (9.8)$$

Jeżeli chcemy maksymalizować oczekiwany dochód gazeciarza, to otrzymujemy następujące zadanie: znajdź takie  $z^*$ , aby:

$$d(z^*) = \max\{d(z) | z \in D\}. \quad (9.9)$$

Zadanie (9.9) można rozwiązać, stosując jedną z trzech procedur opartych na:

- 1) macierzy dochodów;
- 2) wzorze rekurencyjnym;
- 3) wzorze analitycznym.

**Procedura pierwsza — macierz dochodów.** Macierz dochodów ustalamy dla każdego wariantu podaży i popytu, określając osiągany przez gazeciarza dochód zgodnie ze wzorem (9.7). Jeżeli nie ma dodatkowych ograniczeń, zakres zmienności podaży i popytu jest taki sam.

Oczekiwany dochód dla  $i$ -tego wariantu podaży obliczamy według wzoru, który wynika z równania (9.8):

$$d_i = \sum_{j=1}^{N-n+1} d_{ij}p_j \quad (i = 1, \dots, N-n+1), \quad (9.10)$$

gdzie:

- $d_i$  — oczekiwany dochód dla  $i$ -tego wariantu podaży,
- $d_{ij}$  — dochód osiągany dla  $i$ -tego wariantu podaży i  $j$ -tego wariantu popytu,
- $p_j$  — prawdopodobieństwo zajścia  $j$ -tego wariantu popytu.

Następnie znajdujemy ten wariant podaży, dla którego oczekiwany dochód jest maksymalny.

Procedura pierwsza to kompletny przegląd możliwości — jest prosta i ogólna. Przy większej liczbie realizacji zmiennej losowej  $X$ , metoda będzie jednak bardzo uciążliwa, np. jeżeli  $X$  ma 100 realizacji, to macierz dochodów ma wymiar  $[100 \times 100]$ . Lepiej wtedy zastosować procedurę drugą lub trzecią.

**Procedura druga — wzór rekurencyjny.** Jeżeli zwiększymy podaż z  $(z-1)$  do  $z$  gazet, to dodatkowa  $z$ -ta gazeta przyniesie stratę  $s$ , gdy będzie zbędna (dla

$z \leq z-1$ ) lub zysk  $b$  w pozostałych przypadkach. Dochód osiągnięty z dodatkowej  $z$ -tej gazety jest równy:

$$-sF(z-1) + b[1 - F(z-1)] = b - (b+s)F(z-1).$$

Otrzymujemy w ten sposób wzór rekurencyjny:

$$d(z) = d(z-1) + b - (b+s)F(z-1) \quad (z = n+1, \dots, N), \quad (9.11)$$

który pozwala ustalić dochód osiągany z  $z$  gazet, jeżeli znamy dochód z  $(z-1)$  gazet.

Dla pierwszego wariantu podaży gazet:

$$d(n) = bn. \quad (9.12)$$

Procedura druga jest inną formą (bardziej oszczędną) przeglądu zbioru wszystkich możliwych rozwiązań. Nie musimy znać macierzy dochodów. Najmniej pracochłonna jest jednak procedura trzecia.

**Procedura trzecia — wzór analityczny.** Zauważmy, że  $z$ -tą gazetę warto nabyć, jeżeli oczekiwany dochód, jaki ona przynosi, jest nieujemny, czyli:

$$b - (b+s)F(z-1) \geq 0. \quad (9.13)$$

Gazety dodatkowej ( $z+1$ ) nie warto kupować, jeżeli nie przynosi ona oczekiwanego dochodu, czyli:

$$b - (b+s)F(z) \leq 0. \quad (9.14)$$

Przekształcając nierówność (9.13) i (9.14), otrzymujemy:

$$F(z-1) \leq \frac{b}{b+s} \leq F(z). \quad (9.15)$$

Wielkość  $z$  spełniająca ten warunek jest optymalną podażą gazet.

## PRZYKŁAD 9.2

Gazeciarz w ciągu ostatnich 10 tygodni badał, jaki był popyt na tygodnik „Zorza”. Wyniki obserwacji są zawarte w tablicy 9.4. Cena zakupu „Zorzy”  $c_1 = 3$  zł, cena sprzedaży  $c_2 = 5$  zł, cena zwrotu  $c_3 = 0$  zł. Należy ustalić, ile gazet powinien zakupić gazeciarz, aby oczekiwany dochód był maksymalny.

TABLICA 9.4  
Popyt na gazety w kolejnych tygodniach

| Tydzień        | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 |
|----------------|----|----|----|----|----|----|----|----|----|----|
| Popyt (sztuki) | 12 | 15 | 10 | 12 | 11 | 14 | 13 | 11 | 13 | 12 |

Najpierw ustalamy na podstawie danych z tablicy 9.4, że popyt przyjmuje wartości z przedziału  $[10, 15]$  i określamy, w ilu przypadkach popyt osiąga konkretną wartość z tego przedziału. Na podstawie tych ustaleń łatwo jest oszacować rozkład prawdopodobieństwa  $p(x)$  oraz dystrybuantę rozkładu  $F(x)$  (tablica 9.5).

TABLICA 9.5  
Rozkład prawdopodobieństwa popytu

| Popyt ( $x$ )     | 10  | 11  | 12  | 13  | 14  | 15  |
|-------------------|-----|-----|-----|-----|-----|-----|
| Liczba przypadków | 1   | 2   | 3   | 2   | 1   | 1   |
| $p(x)$            | 0,1 | 0,2 | 0,3 | 0,2 | 0,1 | 0,1 |
| $F(x)$            | 0,1 | 0,3 | 0,6 | 0,8 | 0,9 | 1,0 |

Szacunki rozkładu popytu  $p(x)$  byłyby bardziej wiarygodne, gdyby szereg obserwacji był dłuższy. Powinien on zawierać 30–50 obserwacji. Gdyby popyt przyjmował wartości z przedziału  $[1, 100]$ , wówczas powinno być 200–300 obserwacji, co często w praktyce nie jest możliwe. Wiedzy należałoby badać popyt nie dla konkretnych wartości, lecz dla przedziałów np.  $[1-10]$ ,  $[11-20]$ , ...,  $[91-100]$ .

Wyznaczamy jednostkowy dochód  $b = c_2 - c_1 = 5 - 3 = 2$  oraz jednostkową stratę  $s = c_1 - c_3 = 3 - 0 = 3$ .

Procedura pierwsza — macierz dochodów. Ustalamy dochody dla każdego wariantu podaży i popytu zgodnie ze wzorem (9.7). Dochody te są zamieszczone w tablicy 9.6. W ostatniej kolumnie zamieszczono oczekiwane dochody dla każdej wartości podaży.

TABLICA 9.6  
Macierz dochodów

| $z \backslash x$ | $p(x)$ |     |     |     |     |     | Oczekiwany dochód $d(z)$ |
|------------------|--------|-----|-----|-----|-----|-----|--------------------------|
|                  | 0,1    | 0,2 | 0,3 | 0,2 | 0,1 | 0,1 |                          |
| 10               | 20     | 20  | 20  | 20  | 20  | 20  | 20                       |
| 11               | 17     | 22  | 22  | 22  | 22  | 22  | 21,5                     |
| 12               | 14     | 19  | 24  | 24  | 24  | 24  | 22                       |
| 13               | 11     | 16  | 21  | 26  | 26  | 26  | 21                       |
| 14               | 8      | 13  | 18  | 23  | 28  | 28  | 19                       |
| 15               | 5      | 10  | 15  | 20  | 25  | 30  | 16,5                     |

Przykładowo oczekiwany dochód dla podaży równej 13 liczymy następująco:

$$d(13) = 11 \cdot 0,1 + 16 \cdot 0,2 + 21 \cdot 0,3 + 26 \cdot (0,2 + 0,1 + 0,1) = 21.$$

Ponieważ maksymalny oczekiwany dochód  $d(12) = 22$ , czyli optymalną podażą gazet jest  $z^* = 12$ . Przy tej optymalnej podaży w konkretnym dniu nie wiemy, ile wynosi dokładnie dochód gazeciarza. Można tylko podać, że wyniesie on: 14 zł z prawdopodobieństwem 0,1, 19 zł z prawdopodobieństwem 0,2 oraz 24 zł z prawdopodobieństwem 0,7.

Procedura druga — wzór rekurencyjny. W tym przypadku stosujemy wzory (9.11) i (9.12). Oczekiwany dochód dla kolejnych wariantów podaży wynosi:

$$d(10) = 2 \cdot 10 = 20,$$

$$d(11) = 20 + 2 - 5 \cdot 0,1 = 21,5,$$

$$d(12) = 21,5 + 2 - 5 \cdot 0,3 = 22,$$

$$d(13) = 20 + 2 - 5 \cdot 0,6 = 21,$$

$$d(14) = 21 + 2 - 5 \cdot 0,8 = 19,$$

$$d(15) = 19 + 2 - 5 \cdot 0,9 = 16,5.$$

Są to oczywiście te same wartości, które uzyskano na podstawie macierzy dochodów, a zatem  $z^* = 12$ .

Łatwo możemy ustalić, ile średnio traciłby gazeciarz, gdyby stosował nieoptymalną politykę zakupu, kupując stale np. 18 gazet. Dla  $z > N$  korzystamy ze wzoru:

$$d(z) = d(N) - s(z - N). \quad (9.16)$$

Każda gazeta ponad maksymalny popyt jest zbędna, czyli generuje stratę  $s$ . Mamy więc:

$$d(18) = 16,5 - 3(18 - 15) = 7,5.$$

Oczekiwana strata wyniesie w tym przypadku:

$$d(12) - d(18) = 22 - 7,5 = 14,5 \text{ zł}.$$

Procedura trzecia — wzór analityczny. Liczymy iloraz:

$$\frac{b}{b+s} = \frac{2}{2+3} = 0,4.$$

Ponieważ dystrybuanta rozkładu jest taka, że:

$$0,3 \leq 0,4 \leq 0,6 = F(12) \Rightarrow z^* = 12,$$

to optymalna podaż wynosi 12 gazet.

Jeżeli gazeciarz stale nabywa 12 gazet, popyt na gazety został dobrze oszacowany, jego rozkład nie ulega zmianom w czasie, ceny nie zmieniają się w czasie, to po kilkunastu tygodniach średni dochód gazeciarza powinien dążyć do wartości oczekiwanej. Im dłużej optymalna strategia będzie realizowana, tym różnica między obiema wartościami będzie coraz mniejsza.



Którą z trzech opisanych procedur stosujemy, zależy od rozmiarów zadania oraz informacji, jakie chcemy dodatkowo uzyskać. Jeżeli  $n = 10$ ,  $N = 50$ , to bardzo wygodnie ustalić optymalną podaż, stosując wzór analityczny. W tym przypadku jednak kłopotliwe jest obliczenie oczekiwanego maksymalnego dochodu  $d(z^*)$ . Możemy go ustalić, generując odpowiedni wiersz macierzy dochodów, a potem licząc  $d(z^*)$  lub stosując wzór rekurencyjny dla ustalenia kolejnych  $d(n)$ ,  $d(n+1)$ , ...,  $d(z^*)$ .

Jeżeli  $z^* = 15$  i zależy nam na określeniu poziomu maksymalnego oczekiwanego dochodu, lepiej zastosować do wyznaczenia  $d(z^*)$  wzór rekurencyjny niż pierwszy sposób. Gdy  $z^* = 40$ , wówczas wygodniej skorzystać z pierwszego sposobu.

Model matematyczny zagadnienia gazeciarza obejmuje wiele innych sytuacji decyzyjnych. Na przykład znamy koszt  $k$  produkcji 1 t chleba w piekarni, cenę sprzedaży  $c_s$  i cenę zwrotu  $c_z$ . Popyt na chleb jest zmienną losową. W tym przypadku problem piekarni bezpośrednio nadaje się do zastosowania opisanego modelu i możemy skorzystać z podanych procedur.

Niekiedy jednak sytuacje decyzyjne mają tylko zbliżoną postać do modelu gazeciarza i wówczas musimy zmodyfikować sam model, a także podane procedury rozwiązania.

## 9.4. Zagadnienie ustalania optymalnej liczby części zamiennych

Prowadzenie niemal każdej działalności wymaga utrzymywania określonej wielkości zapasów różnych części zamiennych. Przyczyną tworzenia zapasu są awarie maszyn. Uszkodzony element należy wymienić na nowy. Liczba awarii jest zmienną losową o znanym lub dającym się oszacować rozkładzie. Określa on popyt na części zamienne, który także jest losowy.

Oznaczmy przez:

- $X$  — liczbę awarii (popyt na części; jest to zmienna losowa, dyskretna),
- $p(x)$  — prawdopodobieństwo, że liczba awarii (popyt na części) równy jest  $x$ ,
- $z$  — wielkość zapasu (podaży) części,
- $n$  — minimalny popyt na części,
- $N$  — maksymalny popyt na części,
- $D$  — zbiór możliwych realizacji zmiennej losowej  $X$ ,
- $k_1$  — koszt nabycia jednej części zamiennej, gdy kupujemy ją razem z maszyną,
- $k_2$  — koszt nabycia części, gdy element ulega awarii,
- $s_1$  — strata jednostkowa z tytułu nadmiaru części,
- $s_2$  — strata jednostkowa z tytułu niedoboru części,
- $s(z, x)$  — strata, jeżeli zapas równy jest  $z$ , a popyt wynosi  $x$ ,

$f(z)$  — oczekiwana strata, jeżeli zapas jest równy  $z$ ;  
 $D = \{n, n+1, \dots, N-1, N\}$ ,  
 $s_1 = k_1$ ,  
 $s_2 = k_2 - k_1$ .

$$s(z, x) = \begin{cases} s_1(z-x), & \text{dla } z > x, \\ 0, & \text{dla } z = x, \\ s_2(x-z), & \text{dla } z < x, \end{cases} \quad (9.17)$$

$$f(z) = \sum_{x=n}^{z-1} s_1(z-x)p(x) + \sum_{x=z+1}^N s_2(x-z)p(x). \quad (9.18)$$

Problem decyzyjny sprowadza się do znalezienia takiego zapasu  $z^*$ , aby:

$$f(z^*) = \min \{f(z) | z \in D\}. \quad (9.19)$$

Zadanie (9.19) można rozwiązać, podobnie jak zagadnienie gazeciarza, stosując jedną z trzech procedur opartych na:

- 1) macierzy strat;
- 2) wzorze rekurencyjnym;
- 3) wzorze analitycznym.

**Procedura pierwsza — macierz strat.** Macierz strat ustalamy, obliczając stratę dla każdego wariantu podaży i popytu zgodnie ze wzorem (9.17). Następnie liczymy oczekiwaną stratę:

$$f_i = \sum_{j=1}^{N-n+1} f_{ij} p_j \quad (i = 1, \dots, N-n-1), \quad (9.20)$$

gdzie:

- $f_i$  — oczekiwana utrata dla  $i$ -tego wariantu podaży,
- $f_{ij}$  — strata dla  $i$ -tego wariantu podaży i  $j$ -tego wariantu popytu,
- $p_j$  — prawdopodobieństwo zajścia  $j$ -tego wariantu popytu.

Na koniec ustalamy  $z^*$ , dla którego oczekiwana strata jest minimalna.

**Procedura druga — wzór rekurencyjny.** Jeżeli zwiększymy zapasy części  $z(z-1)$  do  $z$ , to  $z$ -ta część przynosi stratę  $s_1$ , gdy podaż jest nadmierna. W pozostałych przypadkach zmniejsza ona stratę o  $s_2$ . Oczekiwana strata z dodatkowej  $z$ -tej części wynosi:

$$s_1 F(z-1) - s_2 [1 - F(z-1)] = -s_2 + (s_1 + s_2)F(z-1).$$

Wzór rekurencyjny na oczekiwaną stratę ma postać:

$$f(z) = f(z-1) - s_2 + (s_1 + s_2)F(z-1) \quad (z = n+1, \dots, N). \quad (9.21)$$

Dla pierwszego wariantu ( $z = n$ ):

$$f(n) = \sum_{x=n}^N s_2(x-n)p(x). \quad (9.22)$$

Mając ustalone oczekiwane straty dla wszystkich wariantów podaży, szukamy  $z^*$  minimalizujących te straty.

**Procedura trzecia — wzór analityczny.** Nabycie dodatkowej  $z$ -tej części jest nieopłacalne, jeżeli nie zwiększa to oczekiwanej straty, czyli:

$$-s_2 + (s_1 + s_2)F(z-1) \leq 0. \quad (9.23)$$

Nabycie następnej ( $z+1$ ) części jest nieopłacalne, jeżeli nie zmniejsza to straty, czyli:

$$-s_2 + (s_1 + s_2)F(z) \geq 0. \quad (9.24)$$

Z układu nierówności (9.23) i (9.24) otrzymujemy następujący wzór analityczny:

$$F(z-1) \leq \frac{s_2}{s_1 + s_2} \leq F(z). \quad (9.25)$$

Wartość  $z$  spełniająca warunek (9.25) wyznacza optymalną wielkość zapasu.

### PRZYKŁAD 9.3

Piekarnia kupuje piec do wypieku chleba. Cennym i ważnym elementem tego pieca jest sterownik. Niestety ulega on awariom. Awaria wymaga wymiany zepsutego sterownika na nowy. Koszt nabycia sterownika, gdy kupujemy go razem z piecem wynosi  $k_1 = 300$  zł. Jeżeli nabywany jest osobno, w momencie awarii, to koszt wynosi  $k_2 = 500$  zł. Producent przebadał 100 pieców tego samego typu i zaobserwował rozkład awarii, który został przytoczony w tabeli 9.7.

TABLICA 9.7  
Rozkład prawdopodobieństwa awarii sterownika

| Liczba awarii     | 0   | 1   | 2   | 3   |
|-------------------|-----|-----|-----|-----|
| Liczba przypadków | 30  | 30  | 20  | 20  |
| $p(x)$            | 0,3 | 0,3 | 0,2 | 0,2 |
| $F(x)$            | 0,3 | 0,6 | 0,8 | 1,0 |

Należy:

- 1) wyznaczyć optymalną podaż sterowników i minimalną oczekiwaną stratę;
- 2) podać optymalny zapas, jeżeli dodatkowo należy uwzględnić koszt przestoju pieca z tytułu awarii  $s = 1000$  zł (o ile wzrasta minimalna oczekiwana strata).

- 3) obliczyć, dla jakich wartości  $k_2$  ten zakup jest optymalny, jeżeli koszt  $k_2$  jest trudny do ustalenia, a piekarnia nabyła 2 sterowniki.

Rozkład prawdopodobieństwa i jego dystrybuantę przedstawiamy w drugiej części tabeli 9.7.

Ad 1. Ustalamy straty z tytułu nadwyżki i deficytu części:

$$s_1 = k_1 = 300, \quad s_2 = k_2 - k_1 = 500 - 300 = 200.$$

Wygodnie w tym przypadku zastosować procedurę pierwszą, a więc tworzymy macierz strat i liczymy oczekiwane straty dla przewidywanych wariantów podaży. Zawarte są one w tabeli 9.8.

TABLICA 9.8  
Macierz strat

| Podaż \ Popyt | 0   | 1   | 2   | 3   | Oczekiwana strata |
|---------------|-----|-----|-----|-----|-------------------|
| 0             | 0   | 200 | 400 | 600 | 260               |
| 1             | 300 | 0   | 200 | 400 | 210               |
| 2             | 600 | 300 | 0   | 200 | 310               |
| 3             | 900 | 600 | 300 | 0   | 510               |

Ponieważ:

$$\min\{f(z)\} = 210 \Rightarrow z^* = 1,$$

należy więc kupić jeden sterownik w momencie zakupu pieca, a minimalna oczekiwana strata wyniesie wówczas 210 zł.

Ad 2. Koszt przestoju pieca łatwo uwzględnić, modyfikując stratę  $s_2$ :

$$s_2 = k_2 - k_1 + s = 500 - 300 + 1000 = 1200.$$

Stosując procedurę drugą, liczymy oczekiwane straty dla kolejnych wielkości podaży:

$$f(0) = 0 \cdot 0,3 + 1200 \cdot 0,3 + 2400 \cdot 0,2 + 3600 \cdot 0,2 = 1560,$$

$$f(1) = 1560 - 1200 + 1500 \cdot 0,3 = 810,$$

$$f(2) = 810 - 1200 + 1500 \cdot 0,6 = 510,$$

$$f(3) = 510 - 1200 + 1500 \cdot 0,8 = 510.$$

W tym przypadku mamy dwa rozwiązania optymalne. Należy zakupić 2 lub 3 sterowniki. Oczekiwane straty wynoszą:

$$f(2) = f(3) = 510 \text{ zł.}$$

Minimalna oczekiwana strata wzrosła aż o 300 zł.

Ad 3. Jeżeli pominiemy koszt przestoju pieca ( $s$ ), to wzór analityczny operujący kosztami zakupu  $k_1$  i  $k_2$  możemy zapisać w postaci:

$$F(z-1) \leq \frac{k_2 - k_1}{k_2} \leq F(z). \quad (9.26)$$

Przy  $z = 2$  i  $k_1 = 300$  warunek (9.26) zapisujemy następująco:

$$F(1) = 0,6 \leq \frac{k_2 - 300}{k_2} \leq F(2) = 0,8. \quad (9.27)$$

Przekształcając warunek (9.27), uzyskujemy:

$$750 \leq k_2 \leq 1500.$$

Zakup 2 sterowników jest optymalną strategią, jeżeli koszt nabycia sterownika mieści się w przedziale [750, 1500].

## 9.5. Systemy masowej obsługi

**Kanałem obsługi** może być okienko na poczcie, w banku, kasa w sklepie samoobsługowym, dystrybutor na stacji benzynowej, stanowiska obsługi na stacji mycia samochodów.

Jeżeli klient udał się do sklepu samoobsługowego, to po wybraniu towaru trafia do kasy. Możliwe są dwa przypadki. Pierwszy — przed kasą nie ma kolejki i klient jest natychmiast obsługiwany oraz drugi — przed kasą jest kolejka, klient ustawia się na jej końcu i czeka na obsługę. Kolejki stojące przed kasami mogą mieć różną długości w zależności od liczby osób włączających się do obsługi, liczby czynnych kas, średniego czasu obsługi klienta.

Jeżeli jest wiele czynnych kas, znikają kolejki, ale właściciel ponosi wysokie koszty opłaty personelu. Jeżeli jest mało czynnych kas, pojawiają się kolejki. Właściciel ponosi wówczas straty, gdyż część klientów rezygnuje z zakupów i przenosi się do konkurencji. Należy ustalić taką liczbę kanałów obsługi, aby łączne koszty oczekiwania i obsługi były minimalne.

W każdym systemie obsługi spotykają się dwa strumienie:

- 1) strumień zgłoszeń;
- 2) strumień obsługi.

**Strumień zgłoszeń** charakteryzowany jest przez parametr  $\lambda$ , zwany **stopą przybycia**. Stopa przybycia jest to średnia (przeciętna) liczba klientów, którzy zgłosili się do systemu w jednostce czasu (w godzinie, dniu). **Strumień obsługi** charakteryzuje parametr  $\mu$ , zwany **stopą obsługi**. Stopa obsługi określa

przeciętnie klientów może być obsłużonych w systemie w jednostce czasu. Intensywność przepływu klientów w systemie obsługi charakteryzuje parametr:

$$\pi = \frac{\lambda}{\mu}. \quad (9.28)$$

Jeżeli  $\pi > 1$ , to kolejka ciągle rośnie.

Regulamin tworzenia kolejki najczęściej kieruje się zasadą — pierwszy, który przychodzi do kolejki, pierwszy z niej odchodzi. Gdy klient oczekujący w kolejce opuszcza system obsługi, mówimy wówczas o **rezygnacji z obsługi**. **Przelew**, czyli **odmowa obsługi**, występuje wtedy, gdy w kolejce brakuje miejsca i system rezygnuje z przyjęcia zgłoszenia.

Kendall podał symbolikę służącą do identyfikacji poszczególnych systemów masowej obsługi. Zgodnie z tą symboliką system masowej obsługi oznacza się jako  $X/Y/c$ , gdzie:

- $X$  — typ rozkładu na wejściu, czyli przyjęć,
- $Y$  — typ rozkładu na wyjściu, czyli rozkład czasu obsługi,
- $c$  — liczba kanałów obsługi.

Jeżeli  $X = M$ ,  $Y = M$ ,  $c = 2$ , to mamy system z dwoma kanałami, gdzie proces zgłoszeń jest rozkładem Poissona, a czasy obsługi są określane przez rozkład wykładniczy.

Ze względu na liczbę zgłoszeń systemy dzielimy na:

- 1) systemy z nieograniczoną liczbą zgłoszeń;
- 2) systemy z ograniczoną liczbą zgłoszeń.

Ze względu na liczbę kanałów obsługi rozróżnia się:

- 1) systemy jednokanałowe;
- 2) systemy wielokanałowe.

Ze względu na zachowanie się zgłoszeń, systemy masowej obsługi dzielimy na:

- 1) systemy ze stratami;
- 2) systemy bez strat;
- 3) systemy z ograniczoną kolejką;
- 4) systemy z ograniczonym czasem przebywania w kolejce.

W pierwszym przypadku (1) system nie przyjmuje zgłoszenia, jeżeli kanały są zajęte. W drugim przypadku (2) przyjęte są wszystkie zgłoszenia. W przypadku (3) zgłoszenia przyjmowane są tak długo aż kolejka nie przekroczy zadanej długości, a w przypadku (4) — do momentu, gdy czas przebywania w systemie nie przekroczy zadanego czasu pobytu.

System  $M/M/1$  bez strat składa się z jednego kanału, o wykładniczym czasie obsługi oraz nieograniczonym strumieniu zgłoszeń zgodnym z rozkładem Poissona. Jeżeli kanał jest zajęty, tworzona jest kolejka.

System ten charakteryzuje się następującymi wielkościami osobowymi:

- 1) średnia liczba osób w systemie:

$$L_s = \frac{\lambda}{\mu - \lambda};$$

- 2) średnia liczba osób w kolejce (długość kolejki):

$$L_k = \frac{\lambda^2}{\mu(\mu - \lambda)} = \frac{\lambda}{\mu} L_s;$$

- 3) średnia liczba osób w obsłudze:

$$L_0 = \frac{\lambda}{\mu}.$$

Oczywiście:

$$L_s = L_k + L_0.$$

Podobne charakterystyki dotyczą czasu:

- 1) średni czas przebywania klienta w systemie:

$$T_s = \frac{1}{\mu - \lambda} = \frac{L_s}{\lambda};$$

- 2) średni czas pobytu w kolejce:

$$T_k = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{L_k}{\lambda};$$

- 3) średni czas obsługi:

$$T_0 = \frac{1}{\mu}.$$

Oczywiście:

$$T_s = T_k + T_0.$$

Czas pobytu w systemie  $T_s$  nie obejmuje całego czasu pobytu klienta w sklepie, gdyż pomija się czas krążenia między regałami. Podane wzory dotyczą tylko systemu obsługi  $M/M/1$ . Dla innych przypadków wielkości te liczymy, stosując bardziej skomplikowane wzory<sup>1</sup>.

**Optymalna liczba kanałów.** Systemy obsługi mogą mieć kilka kanałów obsługi więc powstaje pytanie, ile ich wybudować lub uruchomić, aby cały system obsługi działał efektywniej.

<sup>1</sup> Więcej o tych wzorach i innych przypadkach systemów masowej obsługi znajduje się w pracach Szapiro [2000, rozdział 9] oraz Jaworskiego [1999, rozdział 16].

Oznaczmy przez:

$b$  – stratę, jaką ponosi system obsługi, jeżeli klient przebywa w nim przez jednostkę czasu,

$p$  – koszt utrzymania kanału obsługi w jednostce czasu.

W systemach obsługi wielokanałowej tworzy się jedną, wspólną kolejkę (tak jest często na poczcie, w bankach) lub tworzonych jest tyle kolejek, ile jest czynnych kanałów obsługi. Rozważmy przypadek drugi. Do każdego z kanałów będzie się w jednostce czasu zgłaszało  $\lambda/x$  klientów. Każdy kanał jest elementarnym systemem obsługi, powiązanych przez zgłoszenia, które są kierowane do poszczególnych kanałów zgodnie z zasadą minimalizacji każdej kolejki.

Średnia liczba osób w  $l$ -tym systemie elementarnym wynosi:

$$L_l = \frac{\lambda/x}{\mu - \lambda/x}, \quad (9.37)$$

a średnia liczba osób w całym systemie obsługi jest następująca:

$$L_s = x \frac{\lambda/x}{\mu - \lambda/x} = \frac{\lambda}{\mu - \lambda/x}. \quad (9.38)$$

Łączny koszt systemu obsługi składającego się z  $x$  kanałów to:

$$K(x) = b \frac{\lambda/x}{\mu - \lambda/x} + px. \quad (9.39)$$

Zadanie decyzyjne sprowadza się do znalezienia takiej liczby kanałów obsługi  $x^*$ , aby:

$$K(x^*) = \min \{K(x) | x \in D\}, \quad (9.40)$$

gdzie  $D$  to zbiór dopuszczalnych wariantów liczby kanałów obsługi.

Zadanie (9.40) można rozwiązać, stosując:

- 1) przegląd zupełny zbioru  $D$ ;
- 2) rachunek różniczkowy dla wyznaczania minimum funkcji  $K(x)$ .

#### PRZYKŁAD 9.4

Hipermarket posiada 6 kas. W ciągu godziny średnio zgłaszało się do kas 60 klientów, czyli  $\lambda = 60$ . Średni czas ich obsługi wynosi 2 minuty, tym samym jeden kanał (jedna kasa) może obsługiwać w ciągu godziny przeciętnie 30 klientów ( $\mu = 30$ ). Stratę, jaką ponosi hipermarket z powodu pobytu klienta w systemie przez godzinę, ocenia się na 12 zł. Koszt uruchomienia kasy wynosi 10 zł na godzinę.

Należy po pierwsze ustalić, ile kas uruchomić, aby system obsługi generował minimalny koszt.

I tak, aby kolejki stałe nie rosły, spełniony musi być warunek:

$$\pi = \frac{\lambda}{x \cdot \mu} = \frac{60}{30x} < 1,$$

czyli  $x > 2$ . Stąd 3 kasy są minimalną liczbą kanałów, jaką należy uruchomić.

Obliczmy, stosując wzór (9.39), średni koszt dla 3, 4, 5 oraz 6 kanałów obsługi.

$$K(3) = 12 \cdot \frac{60}{(30 - 60/3)} + 10 \cdot 3 = 102 \text{ zł},$$

$$K(4) = 12 \cdot \frac{60}{(30 - 60/4)} + 10 \cdot 4 = 88 \text{ zł},$$

$$K(5) = 12 \cdot \frac{60}{(30 - 60/5)} + 10 \cdot 5 = 90 \text{ zł},$$

$$K(6) = 12 \cdot \frac{60}{(30 - 60/6)} + 10 \cdot 6 = 96 \text{ zł}.$$

Ponieważ minimalny koszt wynosi 88 zł, to optymalny system obsługi składa się z czterech kanałów (kas).

Po drugie, trzeba ustalić charakterystyki osobowo-czasowe optymalnego systemu obsługi.

Dla całego systemu obsługi składającego się z czterech kanałów:

$$L_s = \frac{\lambda}{\mu - \lambda/x} = \frac{60}{30 - 60/4} = 4,$$

$$L_k = \frac{\lambda/x}{\mu} L_s = \frac{60/4}{30} \cdot 4 = 2,$$

$$L_0 = \frac{\lambda}{\mu} = \frac{60}{30} = 2,$$

czyli w każdym momencie w systemie obsługi przebywają średnio cztery osoby, w tym dwie stoją w kolejkach, a dwie są obsługiwane w kasach.

Dla systemu elementarnego:

$$L_s^i = \frac{4}{4} = 1, \quad T_s = \frac{1}{\mu - \lambda/x} = \frac{1}{30 - 60/4} = \frac{1}{15}, \quad \text{czyli 4 min},$$

$$L_k^i = \frac{2}{4} = 0,5, \quad T_k = \frac{L_k^i}{\lambda/x} = \frac{0,5}{15} = \frac{1}{30}, \quad \text{czyli 2 min},$$

$$L_0^i = \frac{2}{4} = 0,5, \quad T_0 = \frac{L_0^i}{\lambda/x} = \frac{0,5}{15} = \frac{1}{30}, \quad \text{czyli 2 min}.$$

Średni czas pobytu każdej osoby w systemie wynosi 4 minuty, w tym 2 minuty spędza ona w kolejce, a 2 minuty w kasie.

Po trzecie, ze względu na niskie place personelu,  $p = 10$  zł, udało się zatrudnić tylko 3 kasjerki. Należy wyznaczyć charakterystyki osobowo-czasowe.

$$L_s = \frac{\lambda}{\mu - \lambda/x} = \frac{60}{30 - 20} = 6, \quad L_s^i = \frac{6}{3} = 2, \quad T_s = \frac{1}{\mu - \lambda/x} = \frac{1}{30 - 60/3} = \frac{1}{10}, \quad \text{czyli 6 min},$$

$$L_k = \frac{\lambda/x}{\mu} L_s = \frac{20}{30} \cdot 6 = 4, \quad L_k^i = \frac{4}{3} = 1,33, \quad T_k = \frac{L_k^i}{\lambda/x} = \frac{4/3}{60/3} = \frac{1}{15}, \quad \text{czyli 4 min},$$

$$L_0 = \frac{\lambda}{\mu} = \frac{60}{30} = 2, \quad L_0^i = \frac{2}{3} = 0,67, \quad T_0 = \frac{L_0^i}{\lambda/x} = \frac{2/3}{60/3} = \frac{1}{30}, \quad \text{czyli 2 min}.$$

Po czwarte, ile warto dodatkowo zapłacić czwartej kasjerce? Ponieważ  $K(3) = 102$ , a  $K(4) = 88$ , więc opłaca się dodatkowo zapłacić co najwyżej 14 zł.

I po piąte, trzeba określić charakterystyki czasowo-osobowe systemu dla 6 kas.

$$L_s = \frac{60}{30 - 10} = 3, \quad L_s^i = \frac{3}{6} = \frac{1}{2}, \quad T_s = \frac{1}{30 - 10} = \frac{1}{20}, \quad \text{czyli 3 min},$$

$$L_k = \frac{10}{30} \cdot 3 = 1, \quad L_k^i = \frac{1}{6}, \quad T_k = \frac{1}{6} / \frac{60}{6} = \frac{1}{60}, \quad \text{czyli 1 min},$$

$$L_0 = \frac{60}{30} = 2, \quad L_0^i = \frac{2}{6} = \frac{1}{3}, \quad T_0 = \frac{1}{3} / \frac{60}{6} = \frac{1}{30}, \quad \text{czyli 2 min}.$$

Widzimy więc, że podwojenie liczby czynnych kas z 3 do 6 zmniejsza liczbę osób w systemie z 6 do 3, w kolejkach z 4 do 1, natomiast ich czas pobytu w systemie skraca się z 6 do 3 min, a czas pobytu w kolejce z 4 do 1 min.

#### POLECENIA I PYTANIA KONTROLNE

1. Co to jest podejmowanie decyzji w warunkach ryzyka?
2. Jaka jest zasadnicza różnica między podejmowaniem decyzji w warunkach ryzyka i w warunkach niepewności?
3. Jakie kryterium stosujemy, podejmując decyzję w warunkach ryzyka?
4. Kiedy średni dochód będzie bliski oczekiwanej wartości dochodu?
5. Jakie parametry początkowe, a jakie wtórne występują w zagadnieniu gajeciarza?

6. Co pozwala wyznaczyć wzór analityczny, a co wzór rekurencyjny w zagadnieniu gazeciarka?
7. Jaka będzie rzeczywista strata przy zakupie optymalnej liczby części zamiennych?
8. Znamy koszt zakupu części zamiennej w warunkach normalnych, rozkład awarii tej części oraz liczbę zakupionych części zamiennych. Co można ustalić i z jakiego wówczas wzoru należy skorzystać?
9. Jakie na ogół rozkłady przyjmujemy w zagadnieniu ustalania liczby kanałów obsługi?
10. Czy z podanych w książce wzorów można korzystać, jeżeli rozpatrujemy wiele kanałów obsługi, ale z jedną centralną kolejką?
11. Jakie wnioski wynikają z charakterystyk osobowo-czasowych rozważanego systemu obsługi, jeżeli zwiększymy liczbę kanałów z 4 do 6?

#### BIBLIOGRAFIA

- Czerwiński Z., 1984, *Matematyka na usługach ekonomii*, PWN, Warszawa.
- Jaworski K.M., 1999, *Metodologia projektowania realizacji budowy*, Wydawnictwo Naukowe PWN, Warszawa.
- Muller Y., 1971, *Wprowadzenie do nauki organizacji i badań operacyjnych*, część 1 i 2, PWN, Warszawa.
- Szapiro T. (red.), 2000, *Decyzje menedżerskie z Excelem*, PWE, Warszawa.
- Trzaskalik T., 2003, *Wprowadzenie do badań operacyjnych z komputerem*, PWE, Warszawa.

## ROZDZIAŁ 10 DRZEWA DECYZYJNE



### 10.1. Wprowadzenie

W tym rozdziale podręcznika są przedstawione zagadnienia związane z drzewami decyzyjnymi. Stanowią one wygodne narzędzie analizy i podejmowania decyzji w warunkach ryzyka. Ryzyko w tym kontekście oznacza, że decydent nie ma pełnej wiedzy o tym, jakie będą rezultaty jego decyzji ani też nie ma pełnej kontroli nad tymi rezultatami. Jest jednak w stanie określić możliwe stany otoczenia oraz przewidywane następstwa swoich decyzji w zależności od zaistniałych okoliczności. Zna również lub potrafi ocenić prawdopodobieństwa zaistnienia poszczególnych zdarzeń.

Aby zilustrować istotę i zastosowanie drzew decyzyjnych, posłużymy się na początek prostym przykładem problemu decyzyjnego.

#### PRZYKŁAD 10.1

Pan Nowak wybiera się w najbliższy poniedziałek w interesach do Paryża. Ma nadzieję załatwić wszystkie sprawy związane z podpisaniem pewnego kontraktu w jeden dzień i wrócić do Warszawy w środę. Może on pojechać swoim samochodem, co będzie wiązało się z kosztem 1200 zł lub polecieć samolotem linii LOT. Bilet w jedną stronę na lot z Warszawy do Paryża kosztuje 650 zł, przy czym, jeśli kupuje się bilet od razu w obie strony cena wynosi 1000 zł. Opłata związana ze zmianą terminu lotu wynosi 100 zł. W międzyczasie pan Nowak dowiedział się, że jego dobry kolega w środę będzie wracał samochodem do Polski i mógłby go zabrać. Pan Nowak chętnie skorzystałby z okazji, gdyby udało mu się załatwić wszystkie sprawy w jeden dzień tak, aby w środę móc udać się w podróż powrotną. Pan Nowak będzie zainteresowany poniesieniem jak najmniejszych kosztów związanych z podróżą.

W przedstawionej sytuacji pan Nowak na etapie planowania podróży do Paryża ma do wyboru trzy warianty decyzyjne:

- 1) jechać swoim samochodem;
- 2) odbyć podróż samolotem (kupić bilet na samolot w obie strony);
- 3) kupić bilet na samolot tylko w jedną stronę, licząc, że przy szczęśliwym zbiegu okoliczności uda mu się załatwić interesy w jeden dzień i wrócić z kolegą „okazją”.

W przypadku pierwszego wariantu pan Nowak będzie uwolniony od niepewności związanej z załatwianiem spraw w Paryżu i koszt podróży będzie znany i stały (1200 zł). W dwóch pozostałych przypadkach koszt będzie zależny od tego, czy Pan Nowak załatwi wszystkie sprawy w jeden dzień, czy też będzie zmuszony wracać później.

Opisany problem można przedstawić w postaci tabeli ponoszonych kosztów, gdzie w wierszach mamy możliwe warianty decyzji, natomiast w kolumnach zdarzenia niezależne od decydenta: przebieg negocjacji, załatwianie formalności (tablica 10.1).

TABLICA 10.1  
Macierz kosztów (zł)

| Wyszczególnienie                      | Powrót w środę | Powrót w terminie późniejszym |
|---------------------------------------|----------------|-------------------------------|
| Jechać samochodem                     | 1 200          | 1 200                         |
| Kupić bilet na samolot w jedną stronę | 650            | 1 300                         |
| Kupić bilet na samolot w obie strony  | 1 000          | 1 100                         |

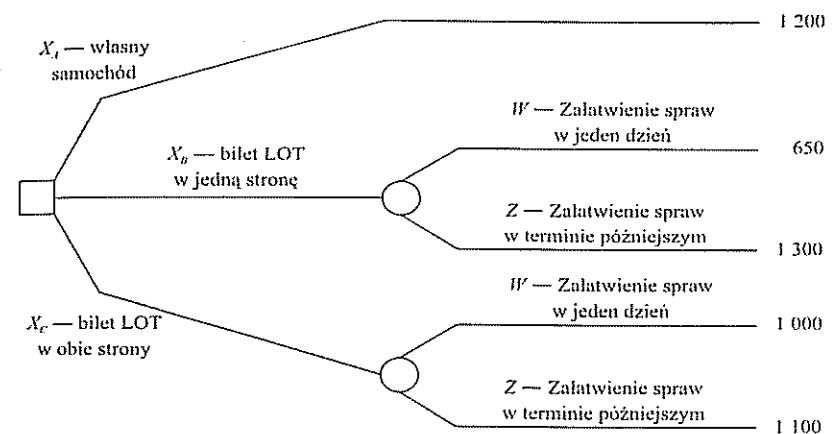
## 10.2. Struktura drzew decyzyjnych

Przedstawienie problemu decyzyjnego w postaci jednej tablicy decyzyjnej może być niejednokrotnie trudne, a czasem wręcz niemożliwe. Chodzi w szczególności o klasę decyzji sekwencyjnych (wieloetapowych) podejmowanych w sytuacji niepewności, a więc takich, gdzie decydent podejmuje wiele powiązanych ze sobą decyzji uzależnionych od nieznanego stanu otoczenia, czy też skutków wcześniej podjętych decyzji. **Drzewa decyzyjne (dendryty)** stanowią wygodną formę przedstawienia problemu decyzyjnego w warunkach ryzyka, jego analizy oraz samego rozwiązania. Mają one szczególne zastosowanie do problemów bardziej rozbudowanych, gdzie analiza dwuwymiarowa (stany-decyzje) już nie wystarcza, a zastosowanie samych formuł rachunku prawdopodobieństwa okazuje się być mało czytelne!

Drzewo decyzyjne stanowi szczególną formę grafu. Składa się z **gałęzi** oraz **węzłów**. Węzły dzielimy na **węzły decyzyjne** oraz **węzły losowe**, przy czym te pierwsze reprezentują warianty pozostające pod kontrolą decydenta (decydent ma swobodny wybór decyzji), natomiast węzły losowe reprezentują rozgałęzienia będące poza kontrolą decydenta (wyznaczane przez czynniki zewnętrzne). Gałęzie oznaczamy liniami, węzły decyzyjne kwadratami, a węzły losowe kółkami. Końcowe gałęzie drzewa mają przypisane, wyrażone ilościowo, rezultaty badanego procesu (korzyści bądź koszty). W każdym wariantcie są one zależne od sekwencji podejmowanych decyzji oraz zaistniałych okoliczności zewnętrznych.

Drzewo decyzyjne opisujące przytoczony w przykładzie 10.1 problem pana Nowaka miałoby następującą postać (rysunek 10.1).

RYSUNEK 10.1  
Drzewo decyzyjne problemu pana Nowaka



Poruszając się wzdłuż gałęzi drzewa decyzyjnego od lewej do prawej strony, możemy prześledzić całą strukturę problemu decyzyjnego. Gałęzie wychodzące z węzłów decyzyjnych są możliwymi wariantami działania, natomiast węzły losowe oznaczają niepewny i niezależny od decydenta stan otoczenia, który przekłada się na wynik decyzji.

Jeżeli zostanie wybrana decyzja  $X_a$  (podróż własnym samochodem), to — bez względu na to, ile będą trwały negocjacje i załatwianie formalności — koszt podróży wyniesie 1200 zł. Z początkowego węzła decyzyjnego wychodzi jedna gałęzia, na której końcu przypisano koszt związany z wyborem tego wariantu.

W przypadku wyboru decyzji  $X_b$  (zakup biletu LOT w jedną stronę) koszty transportu zależą od tego, czy pan Nowak załatwi wszystkie sprawy związane z kontraktem w jeden dzień. Niepewność obrazuje węzeł losowy, z którego odcho-

dzą dwie krawędzie prezentujące dwie możliwe do zaistnienia sytuacje ( $W$  oraz  $Z$ ). Jeżeli uda się załatwić wszystkie sprawy w jeden dzień, to w drodze powrotnej pan Nowak może, nie ponosząc dodatkowych kosztów, skorzystać z uprzejmości znajomego. Jeśli ten wariant się nie powiedzie, będzie musiał dokupić jeden bilet lotniczy za 650 zł.

Wreszcie w przypadku  $X_C$  (zakup biletu LOT w obie strony za 1000 zł), kiedy nie uda się załatwić wszystkiego w jeden dzień i trzeba będzie zmienić datę lotu, wiąże się z dodatkowym kosztem w wysokości 100 zł.

### 10.3. Maksymalizacja oczekiwanej korzyści (minimalizacja oczekiwanego kosztu)

W przypadku problemu podejmowania decyzji w warunkach niepewności można wykorzystać rachunek prawdopodobieństwa dla wyboru decyzji optymalnej. Niezbędna staje się wówczas znajomość rozkładu prawdopodobieństwa zaistnienia określonych stanów otoczenia lub jego założenie. Możemy mieć więc do czynienia z szacowaniem prawdopodobieństwa na podstawie rozkładu częstości zdarzeń występujących w przeszłości bądź z prawdopodobieństwem subiektywnym określonym przez grupę respondentów, zespół ekspertów lub przez samego decydenta.

Przypuśćmy, że pan Nowak ocenia, iż prawdopodobieństwo zaistnienia każdego z rozważanych wariantów sytuacji jest takie samo i wynosi 0,5 (brak informacji o wielkości prawdopodobieństwa intuicyjnie można zastąpić założeniem o rozkładzie jednostajnym)<sup>1</sup>. Jeśli odpowiednie gałęzie wychodzące z węzłów losowych oznaczymy prawdopodobieństwami (jak na rysunku 10.2), to dla każdej podjętej decyzji  $x_j$  będziemy mogli wyznaczyć oczekiwany koszt według formuły:

$$EC(x_j) = \sum_{i=1}^n R_{ij} p_i, \quad j = A, B, C, \quad (10.1)$$

gdzie:

$EC(x_j)$  — wartość oczekiwana kosztu w przypadku podjęcia decyzji  $x_j$ ,

$R_{ij}$  — koszt ponoszony w przypadku podjęcia decyzji  $x_j$  i zaistnienia  $i$ -tego wariantu sytuacji,

$p_i$  — prawdopodobieństwo zaistnienia  $i$ -tego wariantu sytuacji.

<sup>1</sup> Zgodnie z regułą Bayesa (nazywaną także regułą totalnej ignorancji, zasadą braku dostatecznej informacji lub regułą zupełnej niewiedzy) wszystkie zdarzenia losowe zawierające dokładnie jedno zdarzenie elementarne są jednakowo prawdopodobne, a więc:

$$\forall_{\omega_i \in \Omega} P(\{\omega_i\}) = \frac{1}{m}, \quad \text{gdzie } i = 1, 2, \dots, m, \quad \Omega = \{\omega_1, \omega_2, \dots, \omega_m\}.$$

Ogólna reguła analizy drzew decyzyjnych mówi, że decyzję optymalną wyznacza się, pokonując tę samą drogę, co przy budowie drzewa, tylko w przeciwnym kierunku. Poczynając od końców gałęzi drzewa dla każdego węzła losowego oblicza się wartość oczekiwaną spodziewanego kosztu (straty, korzyści), natomiast w każdym węzle decyzyjnym wybiera się minimalny koszt (minimalną stratę lub maksymalną spodziewaną korzyść) oraz gałąź, która pozwala na jej uzyskanie.

W przedstawionym przykładzie dla każdego węzła losowego mamy obliczone oczekiwane koszty transportu oraz wskazaną decyzję (gałąź), która prowadzi do osiągnięcia minimalnego kosztu (rysunek 10.2).

Wartość oczekiwana kosztu transportu jest wyznaczana według formuły:

$$EC(x_B) = 650 \cdot \frac{1}{2} + 1300 \cdot \frac{1}{2} = 975,$$

$$EC(x_C) = 1000 \cdot \frac{1}{2} + 1100 \cdot \frac{1}{2} = 1050.$$

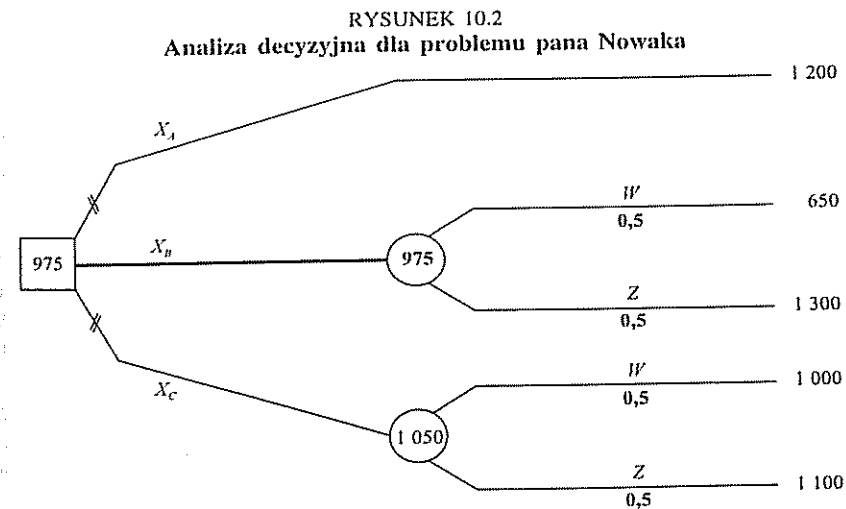
Koszt w przypadku decyzji  $X_A$  przy przyjętych założeniach jest pewny i wynosi 1200 zł. Problem wyboru optymalnej decyzji sprowadza się wówczas do znalezienia takiego dopuszczalnego wariantu decyzyjnego  $x_j^*$ , dla którego:

$$EC(x_j^*) = \min_{x_j \in D} \{EC(x_j)\}, \quad (10.2)$$

gdzie:

$EC$  — wartość oczekiwana kosztu,

$D$  — zbiór dopuszczalnych wariantów.





Z przeprowadzonej analizy decyzyjnej wynika, że pan Nowak powinien wybrać podróż samolotem i zdecydować się na kupno biletu tylko w jedną stronę (wariant  $X_n$ ). Pozostałe warianty decyzyjne charakteryzują się wyższym oczekiwanym kosztem podróży (blokujemy gałęzie decyzyjne im odpowiadające znakiem ||).

## 10.4. Decyzje sekwencyjne w warunkach niepewności

Decydent, rozważając konkretny problem, często zdaje sobie sprawę, że zadanie nie ograniczy się do podjęcia jednej decyzji i oczekiwania na jej efekty. Nierzadko mamy do czynienia z ciągiem decyzji, przy czym kolejne decyzje są podejmowane w nowej sytuacji oraz na podstawie coraz bogatszego zbioru danych. Zmiany sytuacji decyzyjnej są efektem działania zarówno czynników zewnętrznych, jak i wcześniej podjętych decyzji. Optymalizacja w zakresie wieloetapowego procesu decyzyjnego musi obejmować całą złożoną strukturę wariantów decyzyjnych (wybór określonego wariantu decyzyjnego na danym etapie może uniemożliwić wybór innego wariantu na dalszych etapach). Dla zilustrowania zastosowania drzew decyzyjnych do wieloetapowych procesów decyzyjnych prześledźmy przykład

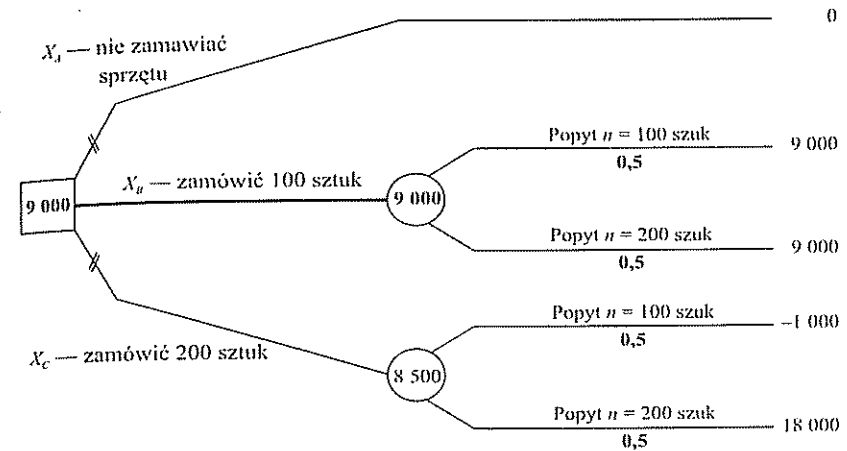
### PRZYKŁAD 10.2

Rozważmy problem decyzyjny kierownika pewnej hurtowni. We wrześniu musi on podjąć decyzję dotyczącą zamówienia u producenta nart i butów narciarskich. Popyt na sprzęt narciarski jest wielkością nieznaną i czynnikiem wprowadzającym element ryzyka do działalności hurtownika. Rozważa on zamówienie 100 lub 200 sztuk sprzętu po cenie 540 zł za komplet. Kierownik może także wcale nie zamawiać sprzętu. Zakładamy, że kierownik hurtowni ustalił cenę sprzedaży na poziomie 630 zł. Niesprzedany sprzęt narciarski może być wyprzedany po sezonie, po cenie 440 zł. Ilustrację problemu oraz jego analizę przedstawiono za pomocą drzewa decyzyjnego zamieszczonego na rysunku 10.3.

Przy założeniu jednakowych prawdopodobieństw dla wariantów popytu hurtownik powinien zamówić 100 sztuk sprzętu i może liczyć na 9000 zł dochodu.

Założmy dalej, że hurtownik może odłożyć decyzję o zamówieniu sprzętu narciarskiego do końca października. Wówczas mógłby nabyć sprzęt po cenie 570 zł za komplet. Dodatkowo kierownik hurtowni mógłby zastosować zauważoną przez siebie regułę, że w 15 na 20 przypadków po zimnym październiku (średnia temperatura poniżej 5°C) następowała mroźna i śnieżna zima, w przeciwnym natomiast sytuacji, po ciepłym październiku w 12 na 20 przypadków następowała łagodna zima. Zebrane dane zostały zamieszczone w tabelicy 10.2.

RYSUNEK 10.3  
Drzewo decyzyjne dla problemu kierownika hurtowni



TABLICA 10.2  
Dane historyczne dotyczące pogody w październiku i w sezonie zimowym

| Wyszczególnienie   | Mroźna i śnieżna zima | Łagodna zima |
|--------------------|-----------------------|--------------|
| Zimny październik  | 15                    | 5            |
| Ciepły październik | 8                     | 12           |

Kierownik hurtowni przypuszcza słusznie, że popyt na sprzęt narciarski będzie zależał od warunków pogodowych w miesiącach zimowych. Przy mroźnej i śnieżnej zimie wzrasta szansa na większy popyt, który — jak zakłada kierownik — wyniesie co najmniej 200 sztuk. Szacowana cena zbytu wynosi 690 zł. W przypadku zimy łagodnej hurtownik zakłada, że popyt wyniesie ok. 100 sztuk, a cena sprzedaży — 620 zł. Wybór optymalnej decyzji w tym momencie może się odbywać w porządku sekwencyjnym (etapowym), co pokazano na drzewie decyzyjnym (rysunek 10.4).

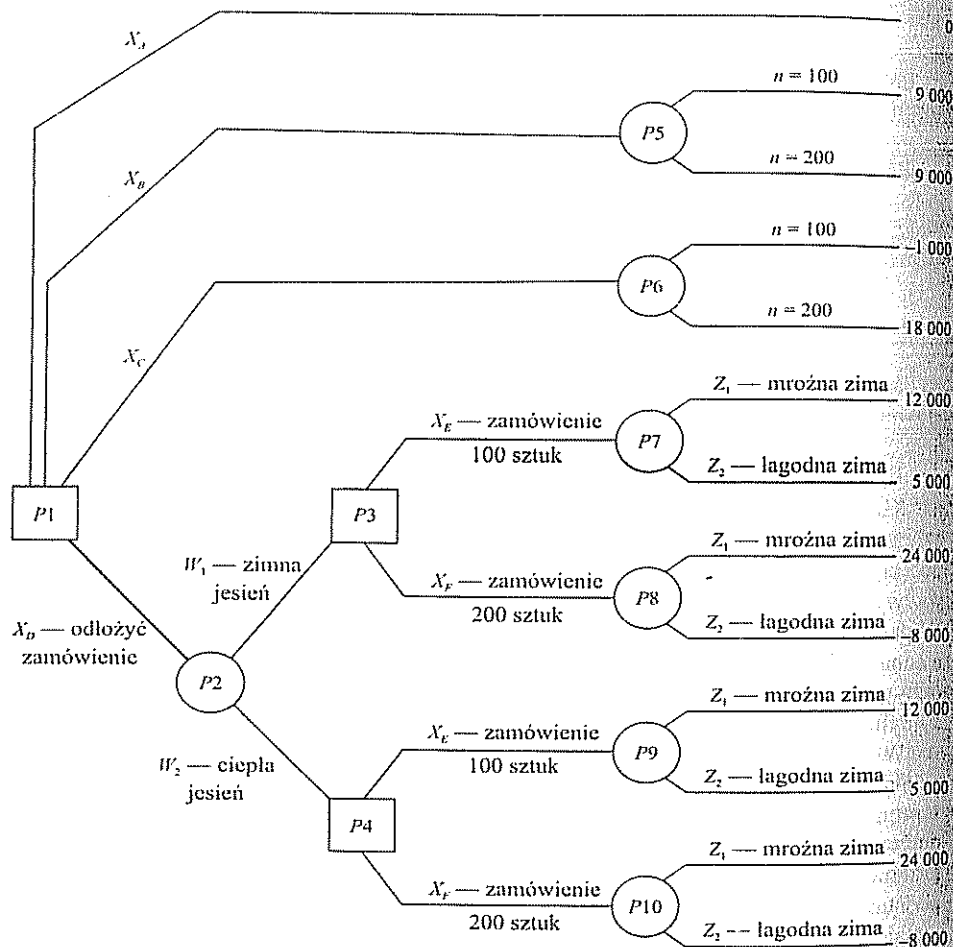
W pierwszej fazie kierownik hurtowni decyduje o tym, czy podejmować działalność, a jeśli tak, to czy złożyć zamówienie natychmiast, korzystając z promocyjnej ceny (wariant B i C), czy też wykorzystać dodatkowe informacje dotyczące pogody, które będą dostępne pod koniec października, wstrzymując się z zamówieniem (opcja D). W razie wyboru tego wariantu, w drugiej fazie, na podstawie posiadanych informacji o pogodzie w październiku (którą decydent uznaje za zapowiedź zimy) hurtownik decyduje o wielkości zamówienia. Decyzja

o odłożeniu decyzji wymaga jednak analizy dostępnych danych i prognozy. Można się bowiem okazać, że zwleknięcie ze złożeniem zamówienia jest nieopłacalne ze względu na rosnącą cenę, po której hurtownik może towar zakupić u producenta.

W drugiej fazie problemu ostateczny wynik (osiągnięty dochód) jest obciążony ryzykiem, którego źródła trzeba upatrywać w potwierdzeniu przyjętej reguły prognozowania pogody.

Dla określenia prawdopodobieństw wystąpienia poszczególnych sytuacji losowych wykorzystamy rachunek prawdopodobieństwa.

RYСУNEK 10.4  
Drzewo decyzyjne dla problemu kierownika hurtowni



Wprowadźmy następujące oznaczenia:

- $W_1$  — zdarzenie polegające na wystąpieniu zimnego października (średnia temperatura poniżej  $5^{\circ}\text{C}$ ),
- $W_2$  — zdarzenie polegające na wystąpieniu ciepłego października (średnia temperatura nie niższa niż  $5^{\circ}\text{C}$ ),
- $Z_1$  — zdarzenie polegające na wystąpieniu mroźnej i śnieżnej zimy<sup>2</sup>,
- $Z_2$  — zdarzenie polegające na wystąpieniu łagodnej zimy.

Obliczamy prawdopodobieństwo *a priori* tego, że średnia temperatura października będzie wynosiła poniżej  $5^{\circ}\text{C}$  (zimny październik) oraz tego, że średnia temperatura nie spadnie poniżej  $5^{\circ}\text{C}$  (ciepły październik). Na podstawie danych historycznych z tabeli 10.2 mamy:

$$P(W_1) = P(W_2) = \frac{20}{40} = 0,5.$$

Następnie możemy obliczyć prawdopodobieństwa warunkowe tego, że po zimnym październiku wystąpi mroźna i śnieżna zima:

$$P(Z_1/W_1) = \frac{P(Z_1 \cap W_1)}{P(W_1)} = \frac{15}{40} = 0,375.$$

I analogicznie:

$$P(Z_2/W_1) = 0,25,$$

$$P(Z_1/W_2) = 0,4,$$

$$P(Z_2/W_2) = 0,6.$$

Analizę decyzyjną tego rozbudowanego drzewa zaczynamy od końców gałęzi. I tak w węzle losowym  $P_7$  obliczamy wartość oczekiwaną dochodu według formuły:

$$ED_7 = 0,75 \cdot 12\,000 + 0,25 \cdot 5\,000 = 10\,250,$$

dla węzła losowego  $P_8$  obliczamy wartość oczekiwaną dochodu:

$$ED_8 = 0,75 \cdot 24\,000 + 0,25 \cdot (-8\,000) = 16\,000 \text{ itd.}$$

Kierując się kryterium maksymalizacji korzyści (dochodu), w każdym węzle decyzyjnym wybieramy maksymalną wartość oczekiwaną dochodu. I tak w węzłach decyzyjnych  $P_3$  i  $P_4$  mamy:

$$ED_3 = \max\{10\,250 \quad 16\,000\} = 16\,000,$$

$$ED_4 = \max\{7\,800 \quad 4\,800\} = 7\,800,$$

<sup>2</sup> Przyjęto, że mroźna i śnieżna zima oznacza co najmniej 50 dni utrzymującego się śniegu przy ujemnych temperaturach w sezonie (od 15 grudnia do 15 marca).

$$\begin{aligned}P(R_1/Z_1) &= 0,8, \\P(R_2/Z_1) &= 0,2, \\P(R_1/Z_2) &= 0,6, \\P(R_2/Z_2) &= 0,4.\end{aligned}$$

Wielkości  $P(R_1/Z_1)$  oraz  $P(R_2/Z_2)$  odzwierciedlają jakość prognoz stacji meteorologicznej. Obserwacja pogody w poprzednich latach pozwala stwierdzić, że prawdopodobieństwo wystąpienia mroźnej i śnieżnej zimy wynosi 0,575 (por. tablica 10.2), stąd:

$$\begin{aligned}P(Z_1) &= 0,575, \\P(Z_2) &= 0,425.\end{aligned}$$

Brakujące prawdopodobieństwa  $P(R_1)$  oraz  $P(R_2)$  można wyznaczyć na podstawie znanych prawdopodobieństw *a posteriori* oraz przy zastosowaniu wzoru na prawdopodobieństwo całkowite:

$$P(A) = \sum_{i=1}^n P(A/B_i) \cdot P(B_i). \quad (10.3)$$

A więc dla naszej sytuacji mamy:

$$\begin{aligned}P(R_1) &= P(R_1/Z_1) \cdot P(Z_1) + P(R_1/Z_2) \cdot P(Z_2) = 0,8 \cdot 0,575 + 0,6 \cdot 0,425 = 0,715, \\P(R_2) &= P(R_2/Z_1) \cdot P(Z_1) + P(R_2/Z_2) \cdot P(Z_2) = 0,2 \cdot 0,575 + 0,4 \cdot 0,425 = 0,285.\end{aligned}$$

Dla wyznaczenia prawdopodobieństw warunkowych *a priori* typu  $P(Z_i/R_j)$  wykorzystamy twierdzenie Bayesa.

#### WZROST DOCHODU Z ZAKUPU SPRZĘTU

Jeżeli zdarzenia  $A_i$  spełniają warunki:  $A_i \cap A_j = \emptyset$  dla  $i \neq j$  ( $i, j = 1, 2, \dots, n$ ) i  $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$  oraz  $P(B) > 0$ , to dla każdego zdarzenia  $A_j$  z pewnego zbioru zdarzeń zachodzi:

$$P(A_j/B) = \frac{P(A_j) \cdot P(B/A_j)}{P(B)} = \frac{P(A_j) \cdot P(B/A_j)}{\sum_{i=1}^n P(A_i) \cdot P(B/A_i)}. \quad (10.4)$$

W przytoczonym wzorze po jednej stronie mamy prawdopodobieństwo typu przyczyna-skutek (*a posteriori*), a po drugiej stronie prawdopodobieństwo typu symptom-diagnoza (*a priori*). Na podstawie danej formuły mamy:

$$P(Z_1/R_1) = \frac{P(Z_1) \cdot P(R_1/Z_1)}{P(R_1)} = \frac{0,575 \cdot 0,8}{0,715} \approx 0,643,$$

$$\begin{aligned}P(Z_2/R_1) &= \frac{P(Z_2) \cdot P(R_1/Z_2)}{P(R_1)} = \frac{0,425 \cdot 0,6}{0,715} \approx 0,357, \\P(Z_1/R_2) &= \frac{P(Z_1) \cdot P(R_2/Z_1)}{P(R_2)} = \frac{0,575 \cdot 0,2}{0,285} \approx 0,404, \\P(Z_2/R_2) &= \frac{P(Z_2) \cdot P(R_2/Z_2)}{P(R_2)} = \frac{0,425 \cdot 0,4}{0,285} \approx 0,596.\end{aligned}$$

Aktualizując informacje uzyskane na podstawie spodziewanych prognoz stacji meteorologicznej, kierownik hurtowni mógł narysować drzewo decyzyjne takie, jak na rysunku 10.6. Z przeprowadzonej analizy wynika, że zlecając sporządzenie prognozy stacji meteorologicznej uzyskał on dodatkowe informacje na temat niepewnej sytuacji. Dzięki temu wzrósł jego oczekiwany dochód z 11 900 zł (własne obserwacje pogody w październiku) do 16 400 zł (prognoza meteorologiczna). Przyrost ten jest częściowo spowodowany korzystniejszą ceną zakupu sprzętu w związku z możliwością wcześniejszego jego zamówienia. Nowe informacje pozwoliły z większą precyzją oszacować szukane prawdopodobieństwa. Wyraźne zwiększenie kwoty potencjalnego dochodu może mieć znaczenie w przypadku fundamentalnej decyzji, czy w ogóle rozpocząć działalność handlową w nadchodzącym sezonie. Przedsiębiorca bowiem w związku z niepewną sytuacją może stwierdzić, że kwota potencjalnego zysku jest zbyt mała, aby pokryć podejmowane ryzyko działalności (awersja do ryzyka).

Wartość informacji zdobytej dzięki zleceniu prognozy można wyrazić przyrostem oczekiwanego dochodu. Zlecenie prognozy długookresowej byłoby uzasadnione z ekonomicznego punktu widzenia, jeśli jej koszt nie wyniósłby więcej niż przyrost oczekiwanego dochodu, a więc ok. 4500 zł.

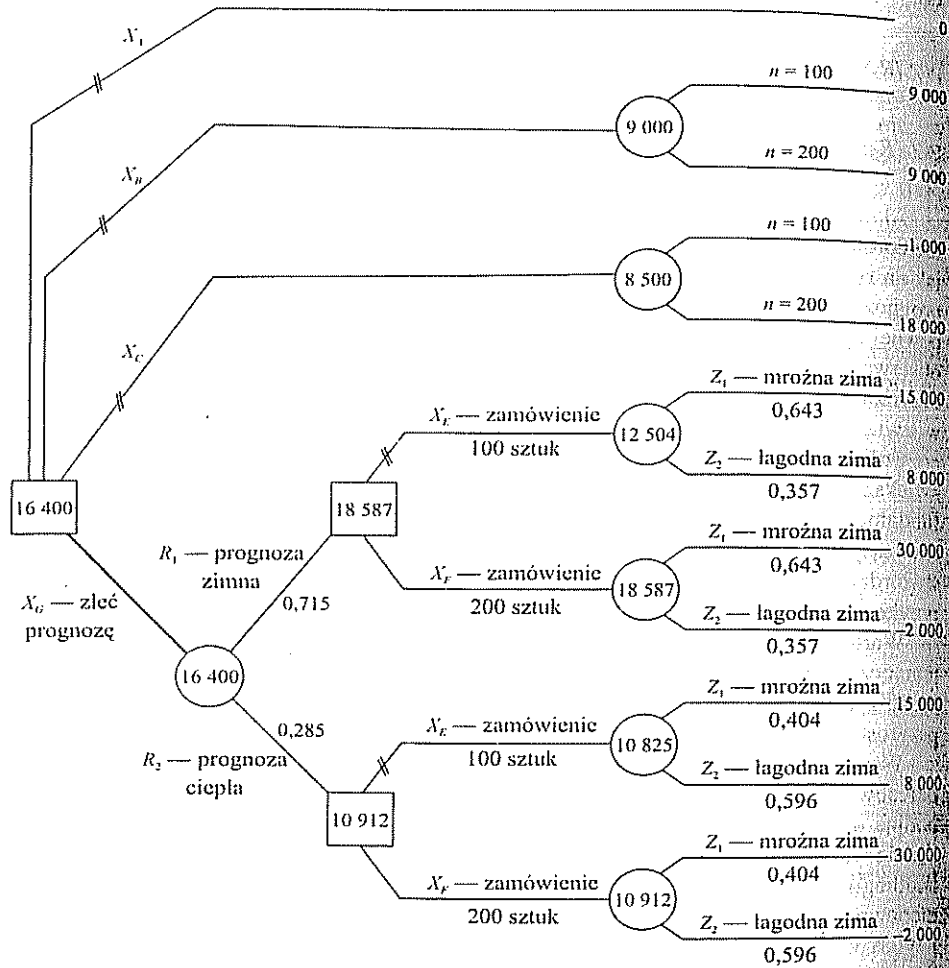
Warto jeszcze zauważyć, że bez względu na to, co pokaże prognoza długookresowa sporządzona przez stację meteorologiczną, przedsiębiorca powinien podjąć decyzję  $x_F$  (zamówienie 200 sztuk sprzętu). A zatem wynik prognozy nie będzie miał wpływu na decyzję hurtownika. Dochodzimy tutaj do drugiego istotnego czynnika warunkującego optymalne pozyskiwanie informacji w sytuacji niepewności: pozyskanie danej informacji jest korzystne i uzasadnione wtedy i tylko wtedy, gdy może ona w istotny sposób wpłynąć na rzeczywiste decyzje decydenta.

Zatem, postępując racjonalnie właściciel hurtowni powinien, bez zlecenia prognoz meteorologicznych (i ponoszenia dodatkowych kosztów), zamówić we wrześniu 200 sztuk sprzętu narciarskiego. Okazuje się jednak, że sama analiza decyzyjna i zebrane na jej potrzeby informacje zwiększyły wartość oczekiwanego zysku.

W prezentowanych rozważaniach zastosowanie znalazła jedna z najczęściej stosowanych metod podejmowania decyzji w warunkach niepewności, a mianowicie reguła maksymalizacji wartości oczekiwanej. Kryterium to pozwala oszaco-

RYSUNEK 10.6

Drzewo decyzyjne i analiza decyzyjna dla rozwiniętego problemu hurtownika



wać przyszłą wartość osiągniętego dochodu, przy czym wartość ta jest obciążona ryzykiem. W związku z tym warto zwrócić uwagę, że reguła ta jest w pełni uzasadniona w momencie, gdy proces decyzyjny jest powtarzany wiele razy w tych samych warunkach. Wówczas wielokrotny wybór danej decyzji bądź ich sekwencji przyniesie decydentowi przeciętny dochód w wysokości oszacowanej wartości oczekiwanej. Wynika to z prawa wielkich liczb, które można zapisać w postaci twierdzenia 10.2.

Jeżeli  $X_i$  są niezależnymi zmiennymi losowymi o identycznym rozkładzie z wartością oczekiwaną  $E(X_i) = \mu$  i skończoną wariancją  $V(X_i) = \sigma^2$ , to:

$$\forall \epsilon > 0 \lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \epsilon\right) = 0. \quad (10.5)$$

Poza tym reguła maksymalizacji wartości oczekiwanej będzie wskazywać optymalny ciąg decyzji w przypadku, gdy abstrahujemy od indywidualnego stosunku decydenta do podejmowanego ryzyka. Tymczasem z obserwacji zarówno rzeczywistości gospodarczej, jak i licznych badań wynika, że różne osoby mają różne nastawienie do ryzyka związanego z daną decyzją. W efekcie w tej samej sytuacji różni decydenci, mając tę samą wiedzę o rozpatrywanym zjawisku, mogą podjąć zupełnie odmienne decyzje. Aby w procesie decyzyjnym uwzględnić problem skłonności do ryzyka, można zastosować kryterium maksymalizacji oczekiwanej użyteczności.

## 10.6. Maksymalizacja oczekiwanej użyteczności

Użyteczność można traktować jako subiektywną miarę cenneści określonych wyników decyzji lub stopień zadowolenia decydenta z osiągniętych korzyści.

Funkcję użyteczności nazywamy niemalejącą funkcją  $U(x)$  określoną na zbiorze wartości  $W$  funkcji korzyści taką, że:

$$U: W \rightarrow \mathbb{R}, \quad (10.6)$$

$$U'(x) \geq 0. \quad (10.7)$$

Załóżmy, że zbiór wartości funkcji korzyści jest przedziałem ograniczonym, a więc  $w \in \langle w_0, w_1 \rangle$  oraz, że funkcja użyteczności jest unormowana w przedziale  $\langle 0, 1 \rangle$ . Wtedy:

$$U(w_0) = 0 \quad \text{oraz} \quad U(w_1) = 1.$$

Załóżmy, że spośród trzech wartości  $w_1, w_2, w_3$  decydent preferuje  $w_1$  nad  $w_2$  oraz  $w_2$  nad  $w_3$ . Jeśli przedstawimy decydentowi do wyboru jeden z dwóch wariantów: otrzymanie korzyści równej  $w_2$  lub udział w grze, w której może on wygrać  $w_1$  z prawdopodobieństwem  $p$  lub  $w_3$  z prawdopodobieństwem  $1 - p$ , to możemy wyznaczyć taką wartość  $p$ , dla której decydentowi będzie obojętne czy otrzyma pewną korzyść  $w_2$ , czy też korzyść wynikającą z udziału w grze. Chodzi więc o rozwiązanie następującego równania:

$$U(w_2) = U[pw_1 + (1 - p)w_3].$$

Przyjmując, że  $U(w_1) = 1$  oraz  $U(w_3) = 0$ , otrzymujemy, że  $U(w_2) = p$ . Dla każdej interesującej nas realizacji wartości  $w$  będzie zatem chodziło o znalezienie takiego prawdopodobieństwa  $p$ , przy którym pewna korzyść  $w_2$  będzie oceniana przez decydenta tak samo oceniana, jak oczekiwana wartość wygranej  $pw_1 + (1-p)w_3$ .

Jeżeli powrócimy do problemu hurtownika, to w rozważanej działalności mamy możliwość uzyskania różnych wartości dochodu, które będziemy traktować jako realizacje funkcji  $W$ . I tak w najgorszej z możliwych sytuacji przedsiębiorca może stracić 8000 zł (przy wykorzystaniu reguły października, zob. rysunek 10.4), natomiast w najlepszej sytuacji osiągnąć dochód równy 30 000 zł. A zatem zakładamy  $U(-8\ 000) = 0$  oraz  $U(30\ 000) = 1$ .

Rozważając problem zlecenia prognozy pogody w sezonie zimowym, w najgorszym przypadku mógłby liczyć się ze stratą w wysokości 2000 zł (nie licząc kosztów usługi i innych kosztów związanych z prowadzoną działalnością). Miałby wówczas do porównania perspektywę  $w_2 = -2\ 000$  oraz  $w_1 = 30\ 000$  osiąganego z prawdopodobieństwem  $p$  i  $w_3 = -8\ 000$  osiąganego z prawdopodobieństwem  $p - 1$ . Gdyby  $p$  przyjmowało wartość bliską 1, wówczas niewątpliwie wybrałby podjęcie działalności. W miarę jak osiągnięcie  $w_1 = 30\ 000$  stawałoby się mało prawdopodobne, przychyliłby się do opcji  $w_2 = -2\ 000$ . Załóżmy, że dla  $p = 0,03$  byłoby mu obojętne czy osiągnąć pewną stratę  $w_2 = -2\ 000$ , czy też wybrać prowadzenie działalności obarczonej dużym ryzykiem.

Użyteczność rozwiązania  $w_2 = -2\ 000$  wynosi więc:

$$U(-2\ 000) = U(30\ 000) \cdot 0,03 + U(-8\ 000) \cdot 0,97 = 0,03 + 0 = 0,03.$$

Przy realizacji strategii G-E oraz zaistnieniu ciągu stanów  $R_1 Z_2$  (niespełnienie się prognozy dotyczącej mroźnej i śnieżnej zimy) hurtownik może osiągnąć dochód równy 8000 zł. W związku z tym decydent musi sobie odpowiedzieć na pytanie, przy jakiej wartości  $p$  zachodzi<sup>4</sup>:

$$U(8\ 000) = p \cdot 30\ 000 + (1 - p)(-8\ 000).$$

Jeżeli  $p = 0,4$ , to:

$$U(8\ 000) = U(30\ 000) \cdot 0,4 + U(-8\ 000) \cdot 0,6 = 0,4.$$

TABLICA 10.3  
Oszacowana użyteczność dochodów hurtownika

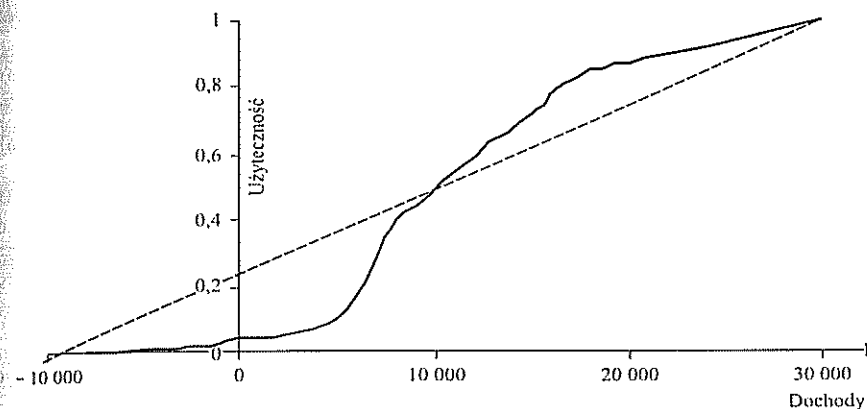
| $w$    | -8 000 | -2 000 | -1 000 | 0    | 5 000 | 8 000 | 9 000 | 12 000 | 15 000 | 16 000 | 18 000 | 24 000 | 30 000 |
|--------|--------|--------|--------|------|-------|-------|-------|--------|--------|--------|--------|--------|--------|
| $U(w)$ | 0      | 0,02   | 0,03   | 0,04 | 0,1   | 0,4   | 0,45  | 0,6    | 0,72   | 0,78   | 0,85   | 0,92   | 1      |

<sup>4</sup> Do szacowania użyteczności kolejnych wartości można też wykorzystać wcześniej ustalone.

Pozostałe oszacowane wartości funkcji użyteczności przedstawiono w tabelicy 10.3.

Jeżeli naniesiemy znalezione wartości na układ współrzędnych oraz połączymy je liniami, to otrzymamy przybliżony wykres krzywej użyteczności decydenta (rysunek 10.7). Krzywa ta opisuje postawę decydenta wobec ryzyka łączącego się z podejmowaną działalnością (przerwana prosta na schemacie przedstawia krzywą użyteczności decydenta obojętnego wobec ryzyka)<sup>5</sup>.

RYSUNEK 10.7  
Hipotetyczna krzywa użyteczności decydenta



Zastępując osiągnięte dochody użytecznościami, możemy do problemu kierownika hurtowni zastosować regułę maksymalizacji użyteczności (zob. rysunek 10.8).

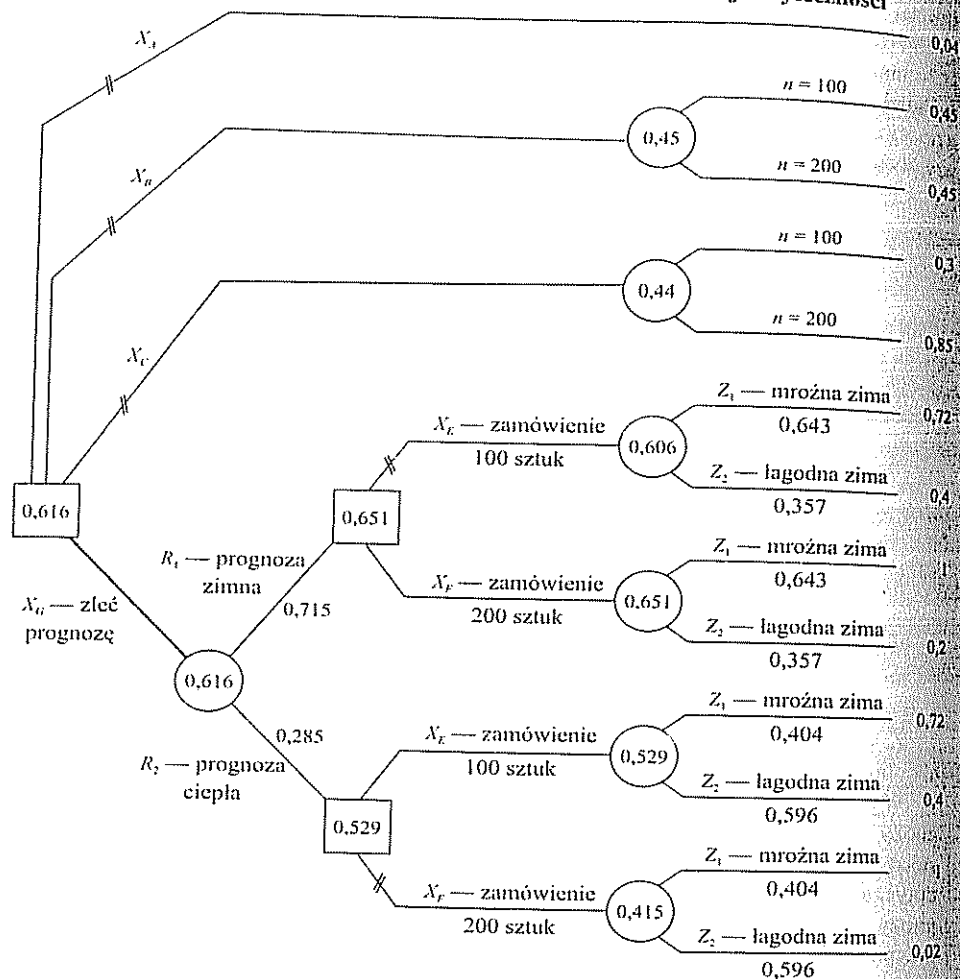
Wyciągnięty poprzednio wniosek o bezcelowości zlecenia profesjonalnej prognozy pogody trzeba by w tym przypadku zmodyfikować. W przypadku zastosowania reguły maksymalizacji oczekiwanej użyteczności hurtownika decyzja dotycząca wielkości zamówienia powinna być uzależniona od otrzymanej prognozy. W przypadku gdy prognoza będzie niekorzystna (łagodna zima), zamówienie powinno wynieść 100 sztuk (oczekiwana użyteczność w węźle decyzyjnym równa 0,529), natomiast gdy prognoza będzie korzystna, hurtownik powinien zamówić 200 sztuk sprzętu ( $U = 0,651$ ).

Zaprezentowana w powyższych rozważaniach metoda szacowania użyteczności na podstawie rozpatrywania ciągu hipotetycznych loterii czy gier jest tylko jedną z propozycji. Zastosowanie jej w praktyce może się wiązać z problemem niekonsekwentnych odpowiedzi decydenta. Może się to objawiać w nieprzechodnich

<sup>5</sup> W praktyce można spotkać różne przebiegi krzywych użyteczności (przykłady podano np. w pracy [Moore, 1975]).

RYSUNEK 10.8

Model decyzyjny hurtownika przy zastosowaniu funkcji użyteczności



preferencjach między wariantami czy niespełnieniem założenia o niemalejącej funkcji preferencji. Poza tym decydent może mieć trudności w identyfikowaniu hipotetycznej sytuacji z rzeczywistym problemem gospodarczym. Inne, bardziej zaawansowane metody wyznaczania funkcji preferencji można znaleźć m.in. w [Coombs, Dawes, Tversky, 1997; Tyszka, 1986; Varian, 2002].

W problemie hurtownika przyjęliśmy dla uproszczenia, że rozkład popytu jest dwuwartościowy. Nietrudno byłoby rozbudować drzewo decyzyjne w taki sposób,

żeby uwzględnić więcej interesujących decydenta wariantów wielkości popytu. Obliczenia związane z rozbudowanym problemem decyzyjnym można by przeprowadzić z wykorzystaniem arkusza kalkulacyjnego Excel<sup>6</sup> lub specjalistycznego oprogramowania<sup>7</sup>. Do modelowania podobnych sytuacji decyzyjnych można także wykorzystać rozkłady ciągłe<sup>8</sup>.

#### POLECENIA I PYTANIA KONTROLNE

1. Przeprowadź analizę wariantu przewidywania warunków zimowych na podstawie pogody październikowej (wariant  $x_D$ ) przy zastosowaniu reguły maksymalizacji oczekiwanej użyteczności. Czy wariant  $x_D$  byłby lepszy od  $x_G$ ?
2. W jakich przypadkach i dlaczego drzewa decyzyjne stanowią lepsze narzędzie analizy i podejmowania decyzji niż macierze wypłat (tablice decyzyjne)?
3. Wymień i scharakteryzuj elementy tworzące strukturę drzewa decyzyjnego.
4. Na czym polega i jak przebiega analiza drzew decyzyjnych?
5. Podaj podstawowe reguły optymalnego pozyskiwania informacji w analizie decyzyjnej w sytuacji niepewności.
6. Kiedy w analizie drzew decyzyjnych wykorzystujemy wzór Bayesa?
7. Jakie jest uzasadnienie wykorzystywania reguły maksymalizacji użyteczności w analizie decyzyjnej w sytuacji niepewności?

#### BIBLIOGRAFIA

- Coombs C.H., Dawes R.M., Tversky A., 1997, *Wprowadzenie do psychologii matematycznej*, Wydawnictwo Naukowe PWN, Warszawa.
- Köfler E., 1993, *Podejmowanie decyzji przy niepełnej informacji*, Real Publisher, Warszawa.
- Luce R. D., Reiffa H., 1963, *Gry i decyzje*, PWN, Warszawa.
- Moore P.G., 1975, *Ryzyko w podejmowaniu decyzji*, PWE, Warszawa.
- Samuelson W.F., Marks S.G., 1998, *Ekonomia menedżerska*, PWE, Warszawa.
- Szapiro T., 1993, *Co decyduje o decyzji*, Wydawnictwo Naukowe PWN, Warszawa.
- Szapiro T. (red.), 2000, *Decyzje menedżerskie z Excelem*, PWE, Warszawa.
- Tyszka T., 1986, *Analiza decyzyjna i psychologia decyzji*, PWN, Warszawa.
- Varian H.R., 2002, *Mikroekonomia: ujęcie nowoczesne*, Wydawnictwo Naukowe PWN, Warszawa.

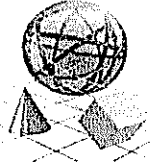
<sup>6</sup> Przykłady zastosowania arkusza kalkulacyjnego do analizy drzew decyzyjnych można znaleźć w [Szapiro, 1993].

<sup>7</sup> Przykładem takiego oprogramowania jest Arborist, który umożliwia definiowanie i analizę problemu decyzyjnego z wykorzystaniem drzew decyzyjnych w wygodnym trybie dialogowym.

<sup>8</sup> Więcej o zastosowaniu ciągłych rozkładów prawdopodobieństwa do analizy drzew decyzyjnych można znaleźć w [Luce, Reiffa, 1963].

## ROZDZIAŁ 11

# PROGRAMOWANIE W WARUNKACH NIEPEWNOŚCI



### 11.1. Wprowadzenie

Niepewność jest nieodłącznym elementem naszego życia i może dotyczyć zdarzeń przeszłych, teraźniejszych, a zwłaszcza przyszłych. Najczęściej źródeł niepewności szukamy w otoczeniu zewnętrznym, którego stany leżą poza naszym bezpośrednim wpływem. Z niepewnością zmagają się przedsiębiorca zakładający firmę i rozpoczynający działalność w określonej branży, dostawca startujący w przetargu czy inwestor kupujący jednostki uczestnictwa w funduszu inwestycyjnym. We wszystkich przedstawionych przykładach niepewność oznacza sytuację, w której decydent nie ma pełnej wiedzy ani o tym, jakie będą rezultaty jego decyzji, ani też nie ma pełnej kontroli nad tymi rezultatami. W literaturze termin **niepewność** często występuje razem z pojęciem **ryzyka**, a nierzadko jest z nim utożsamiany. Według Knighta ryzyko jest przypadkowym wydarzeniem charakteryzującym się znanym rozkładem prawdopodobieństw jego wystąpienia, podczas gdy niepewność to taki rodzaj przypadkowości, którego rozkład prawdopodobieństwa nie jest znany [Gontarek, 2001, s. 79].

W tym rozdziale podręcznika są przedstawione dwa obszary zagadnień: podejmowanie decyzji w sytuacji niepewności (tzw. gry z naturą) oraz gry dwuosobowe, które są przykładem zagadnień stanowiących przedmiot zainteresowań teorii gier.

### 11.2. Gry z naturą

Mimo nieco mylącej nazwy<sup>1</sup>, pierwszy z prezentowanych obszarów dotyczy zagadnień, w których niepewne, nieantagonistyczne otoczenie (natura) pełni bierną rolę w procesie podejmowania decyzji. Zakładamy, że decydent potrafi wskazać obszar decyzji dopuszczalnych, możliwe stany natury oraz możliwe do osiągnięcia rezultaty w każdej z hipotetycznych sytuacji. Przyjmujemy również, że decydent nie zna w danym momencie rozkładów prawdopodobieństwa zaistnienia danych stanów otoczenia w przyszłości lub nie chce korzystać z dostępnej wiedzy. Ponadto w dalszej części rozważań będziemy zajmować się tylko sytuacjami, w których liczba wariantów decyzyjnych, a także stanów natury jest skończona.

#### PRZYKŁAD 11.1

Właściciel gospodarstwa ogrodniczego posiada pole o powierzchni 100 ha. W najbliższym roku rozważa on możliwość uprawy ogórków, marchwi, fasoli i selera. Wielkości plonów, a także pośrednio ceny warzyw, są uzależnione od warunków pogodowych panujących w okresie wiosenno-letnim. Zakładamy, że pogoda może być sucha i zimna (S-Z), sucha i ciepła (S-C), wilgotna i ciepła (W-C) lub wilgotna i zimna (W-Z). Przewidywane wielkości uzyskanych przez ogrodnika dochodów z ha (tys. zł) w zależności od stanów pogody przedstawiono w tabeli 11.1.

TABLICA 11.1  
Macierz dochodów dla działalności ogrodnika

| Uprawy  | Warunki pogodowe |     |     |     |
|---------|------------------|-----|-----|-----|
|         | S-Z              | S-C | W-C | W-Z |
| Ogórki  | 6                | 7   | 16  | 10  |
| Marchew | 13               | 10  | 14  | 7   |
| Fasola  | 10               | 9   | 11  | 11  |
| Seler   | 8                | 14  | 13  | 10  |

Ogrodnik będzie zainteresowany uprawą takiego warzywa (takich warzyw), która zapewni mu maksymalny dochód (optymalna strategia ogrodnika).

<sup>1</sup> Pojęcie gry zakłada istnienie co najmniej dwóch graczy, z których każdy przez wybór określonej strategii ma możliwość wpływania na wynik gry. Chociaż natura niejednokrotnie jest źródłem niepewności w działalności ekonomicznej, trudno jej przypisać przymiot świadomego działania „na korzyść” czy „na niekorzyść” decydentów.

## DEFINICJA 11.1

**Strategią** decydenta (gracza) będziemy nazywać jeden z dostępnych sposobów postępowania w określonej sytuacji, natomiast o **decyzji** będziemy mówić w przypadku, gdy w określonej sytuacji decydent wybrał tę, a nie inną strategię.

## 11.2.1. Strategie czyste

Przypuśćmy, że ogrodnik zdecyduje się na uprawę tylko jednego warzywa, a więc będzie zainteresowany znalezieniem optymalnej strategii czystej.

## DEFINICJA 11.2

**Strategią czystą** nazywamy jednorazową decyzję decydenta o wyborze jednego z dopuszczalnych wariantów decyzyjnych.

Wśród najczęściej stosowanych metod wspomagających proces decyzyjny w warunkach niepewności należy wymienić:

- 1) regułę Walda (maximin);
- 2) regułę Bayesa (Laplace'a);
- 3) regułę Hurwicza;
- 4) regułę Savage'a (minimalnego żalu).

Wybór określonej reguły, a także zastosowanie wybranej ich sekwencji, może mieć bezpośredni wpływ na otrzymane rezultaty, o czym należy pamiętać. Na przedstawionym przykładzie omówimy właściwości tych metod oraz wskażemy na możliwości ich zastosowania.

Niech macierz  $A = [a_{ij}]$  będzie macierzą wypłat. Element  $a_{ij}$  oznacza wypłatę (korzyść) decydenta w przypadku wyboru  $i$ -tej strategii, przy założeniu, że wystąpi  $j$ -ty stan otoczenia (natury).

**Reguła Walda** jest regułą zachowawczą. Bazuje ona na założeniu, że decydent — licząc się nawet z najbardziej niekorzystnymi warunkami — chciałby uzyskać maksymalną korzyść.

Strategię  $k$  możemy uznać za optymalną w sensie kryterium Walda, jeżeli:

$$a_k = \max_i \{a_i\}, \quad (11.1)$$

gdzie  $a_i = \min_j \{a_{ij}\}$ .

Poszukiwanie najlepszej decyzji przebiega następująco:

**Krok 1.** Dla każdej strategii  $i$  wybieramy najniższy możliwy do osiągnięcia dochód  $a_i$  (minimum dla wierszy).

**Krok 2.** Ze wszystkich strategii wybieramy tę, dla której minimalna wartość dochodu  $a_i$  jest maksymalna.

W przypadku ogrodnika reguła ta prowadzi do wyboru uprawy fasoli (tablica 11.2), gdyż w najgorszym razie przewidywany dochód wyniesie 9 tys. zł/ha. Wskazuje się w związku z tym, że reguła Walda to reguła decydenta ostrożnego, zabezpieczającego się na wypadek najgorszego stanu natury<sup>2</sup>.

**Reguła Bayesa**, nazywana też regułą Laplace'a, odnosi się do zasady totalnej ignorancji, braku dostatecznej racji lub zupełnej niewiedzy, która mówi, że jeżeli nie znamy rozkładu prawdopodobieństwa, możemy przyjąć, że wszystkie zdarzenia losowe zawierające dokładnie jedno zdarzenie elementarne są jednakowo prawdopodobne.

Strategię  $k$  możemy uznać za optymalną w sensie kryterium Bayesa, jeśli:

$$a_k = \max_i \{\bar{a}_i\}, \quad (11.2)$$

$$\text{gdzie } \bar{a}_i = \frac{1}{m} \sum_{j=1}^m a_{ij}.$$

Wyznaczenie strategii optymalnej według reguły Bayesa sprowadza się do:

**Krok 1.** Ustalenia dla każdej strategii  $i$  średniego dochodu  $\bar{a}_i$ .

**Krok 2.** Wyboru jako optymalnej tej strategii, dla której średni dochód jest największy.

Zgodnie z tą regułą w przytoczonym przykładzie ogrodnik powinien wybrać uprawę selera (tablica 11.2). Powołując się na prawa wielkich liczb, korzystanie z tej reguły byłoby bardziej uzasadnione, gdyby wybraną strategię stosować wielokrotnie<sup>3</sup>.

**Reguła Hurwicza** pozwala uwzględniać indywidualną skłonność decydenta do podejmowania ryzyka przez określanie tzw. współczynnika ostrożności  $\alpha$ . Współczynnik ten przyjmuje wartości z przedziału  $\langle 0, 1 \rangle$ , gdzie wartość 1 oznacza skrajną awersję do ryzyka, a wartość 0 — całkowitą skłonność do ryzyka.

Poszukiwanie najlepszej strategii według reguły Hurwicza sprowadza się do wykonania następujących kroków:

**Krok 1.** Wyznaczenia minimalnych ( $a_i$ ) oraz maksymalnych ( $b_i$ ) wartości dochodów możliwych do uzyskania w przypadku przyjęcia określonej strategii, gdzie:

$$a_i = \min_j \{a_{ij}\}, \quad b_i = \max_j \{a_{ij}\}. \quad (11.3)$$

<sup>2</sup> W przypadku decydentów o dużej skłonności do ryzyka, nastawionych na najwyższe wygrane (korzyści), można wykorzystać regułę maximax.

<sup>3</sup> Wówczas średni dochód uzyskiwany z 1 ha powinien zbliżać się do wartości oczekiwanej.



Krok 2. Przyjęcia takiego współczynnika ostrożności  $\alpha$ , że  $\alpha \in (0, 1)$ .

Krok 3. Wyznaczenia dla każdej strategii takiego wskaźnika  $h_i$ , że:

$$h_i(\alpha) = \alpha a_i + (1 - \alpha)b_i. \quad (11.4)$$

Krok 4. W sensie reguły Hurwicza strategią optymalną jest ta strategia  $k$ , dla której:

$$h_k = \max_i h_i. \quad (11.5)$$

Jak łatwo zauważyć, w przypadku całkowitej awersji do ryzyka ( $\alpha = 1$ ) reguła Hurwicza sprowadza się do reguły maximin, a w przypadku skrajnej skłonności do ryzyka ( $\alpha = 0$ ) wyniki otrzymane na podstawie reguły Hurwicza są tożsame z wynikami uzyskanymi za pomocą metody maximax.

Jeżeli w prezentowanym przykładzie założymy znaczną skłonność decydenta do ryzyka ( $\alpha = 0,2$ ), to reguła Hurwicza jako optymalną strategię wskaże uprawę ogórków. Wystarczy jednak zmienić wartość współczynnika ostrożności, a rezultaty mogą być zupełnie inne (tablica 11.2).

TABLICA 11.2  
Wyniki zastosowania różnych reguł decyzyjnych dla problemu ogrodnika

| Uprawy  | Warunki pogodowe |     |     |     | Maximin<br>min | Bayesa<br>średnia | Hurwicza         |                  |
|---------|------------------|-----|-----|-----|----------------|-------------------|------------------|------------------|
|         | S-Z              | S-C | W-C | W-Z |                |                   | $h_{\alpha=0,2}$ | $h_{\alpha=0,7}$ |
| Ogórki  | 6                | 7   | 16  | 10  | 6              | 9,75              | 14               | 9                |
| Marchew | 13               | 10  | 14  | 7   | 7              | 11                | 12,6             | 9,1              |
| Fasola  | 10               | 9   | 11  | 11  | 9              | 10,25             | 10,6             | 9,6              |
| Seler   | 8                | 14  | 13  | 10  | 8              | 11,25             | 12,8             | 9,8              |
| Max     | 13               | 14  | 16  | 11  |                |                   |                  |                  |

**Reguła Savage'a** bazuje na postulatcie minimalizacji utraconych potencjalnych korzyści związanych z podjęciem decyzji, która w rezultacie okazała się nieoptymalna ze względu na istniejący stan natury. Wyznaczając strategię najlepszą z punktu widzenia tego kryterium, postępujemy następująco:

Krok 1. Wyznaczamy dla każdego stanu natury maksymalną wartość dochodu  $a_j$ , gdzie:

$$a_j = \max_i \{a_{ij}\}. \quad (11.6)$$

Krok 2. Tworzymy macierz względnych strat  $W$  taką, że:

$$w_{ij} = a_j - a_{ij}. \quad (11.7)$$

Elementy macierzy  $W$  oznaczają wielkość różnicy między korzyścią, jaką osiągniemy podejmując  $i$ -tą decyzję, a korzyścią, jaką osiągnęlibyśmy podejmując najkorzystniejszą decyzję w danym stanie natury.

Dla tak wyznaczonej macierzy względnych strat stosujemy regułę minimax, a więc:

Krok 3. Dla każdej decyzji wybieramy maksymalną względną stratę  $w_i$ , gdzie:

$$w_i = \max_j \{w_{ij}\}. \quad (11.8)$$

Krok 4. Jako optymalną wybieramy taką strategię  $k$ , dla której:

$$w_k = \min_i \{w_i\}. \quad (11.9)$$

Macierz strat dla problemu ogrodnika jest przedstawiona w tablicy 11.3.

TABLICA 11.3  
Macierz względnych strat dla problemu ogrodnika

| Uprawy  | Warunki pogodowe |     |     |     | Savage'a<br>max |
|---------|------------------|-----|-----|-----|-----------------|
|         | S-Z              | S-C | W-C | W-Z |                 |
| Ogórki  | 7                | 7   | 0   | 1   | 7               |
| Marchew | 0                | 4   | 2   | 4   | 4               |
| Fasola  | 3                | 5   | 5   | 0   | 5               |
| Seler   | 5                | 0   | 3   | 1   | 5               |

Metoda Savage'a sugeruje ogrodnikowi wybór uprawy marchwi. Reguła ta jest użyteczna w przypadku, gdy decydent chce minimalizować utracone korzyści związane z podjęciem decyzji, która to decyzja w zaistniałym stanie natury nie daje maksymalnych korzyści.

Podsumowując, trzeba stwierdzić, że każda z przedstawionych metod może wskazywać nie jedno, ale kilka najlepszych rozwiązań. W tym sensie może nie rozstrzygać jednoznacznie, jaką strategię czystą powinien przyjąć decydent. Ponadto każda z opisanych reguł, ze względu na swoją specyfikę i odmienne podejście, może wskazywać w danym przypadku inne najlepsze rozwiązanie, czego przykładem był przytoczony problem ogrodnika (każda z zastosowanych reguł wskazywała na wybór innej strategii). Okazuje się zatem, że wybór konkretnej reguły ma wpływ na otrzymane rezultaty. A zatem świadomy wybór reguły może poprzedzać całą analizę decyzyjną i powinien uwzględniać specyfikę rozpatrywanego problemu, skłonność decydenta do ryzyka (stopień ostrożności), a także — co się z tym wiąże — ocenę użyteczności potencjalnych korzyści. W uproszczeniu można stwierdzić, że reguła Walda (maximin) oraz reguła Savage'a należą do grupy reguł zachowawczych, reguła Hurwicza pozostawia pewien margines indywidualnego, subiektywnego wyboru decydenta, a reguła Bayesa zakłada, że wypłata będzie średnią z możliwych. Nie bezzasadne może być także zastosowanie kilku metod do konkretnego problemu oraz dokonanie arbitralnego wyboru rozwiązania najlepszego w opinii konkretnego decydenta. Takie postępowanie z pewnością daje możliwość głębszej analizy problemu.

Warto podkreślić, że trzy pierwsze reguły (Walda, Hurwicza i Savage'a) mogą być z powodzeniem stosowane do wyboru strategii czystej realizowanej jednokrotnie, natomiast w przypadku wielokrotnej realizacji określonej strategii czystej uzasadnione jest użycie reguły Bayesa.

### 11.2.2. Strategie mieszane

Dotychczas rozpatrywaliśmy jedynie problem jednorazowego wyboru jednego z dopuszczalnych wariantów decyzyjnych, czyli tzw. strategii czystej. W praktyce decydent może realizować jednocześnie kilka wariantów decyzyjnych lub działając długofalowo opracowywać strategię działania na dłuższy okres. Ma on wówczas możliwość stosowania poszczególnych wariantów decyzyjnych z określoną częstością. Mówimy wtedy o strategiach mieszanych.

#### DEFINICJA 11.3

Strategią mieszaną nazywamy liniową kombinację wypukłą strategii czystych.

W przypadku jednorazowego zastosowania strategii mieszanej<sup>4</sup> uzasadnione będzie podejście ostrożne bazujące na maksymalizacji minimalnej możliwej korzyści<sup>5</sup>.

Oznaczmy przez  $x_i$  udział  $i$ -tej strategii czystej. Przez  $v$  oznaczmy minimalną korzyść, jaką odniesie decydent bez względu na okoliczności zewnętrzne (stan natury). Wówczas problem wyboru strategii mieszanej można przedstawić następująco:

$$v \rightarrow \max, \quad (11.10)$$

$$\sum_{i=1}^n a_{ij} x_i \geq v, \quad j = 1, 2, \dots, m,$$

$$\sum_{i=1}^n x_i = 1,$$

$$x_i \geq 0 \quad i = 1, 2, \dots, n.$$

<sup>4</sup> Strategia mieszana w takim sensie oznacza jednorazowe zastosowanie kombinacji strategii czystych.

<sup>5</sup> Zgodnie z sugestią Walda, decydent powinien pamiętać, że niezależnie od własnej decyzji o wyborze określonego wariantu zawsze — wcześniej lub później — musi wystąpić niekorzystny stan otoczenia zewnętrznego, w wyniku czego otrzyma on niekorzystne rezultaty działań własnych.

Wyznaczenie optymalnej strategii mieszanej stosowanej jednorazowo dla problemu ogrodnika sprowadza się do rozwiązania następującego zadania:

$$v \rightarrow \max, \quad (11.11)$$

$$6x_1 + 13x_2 + 10x_3 + 8x_4 \geq v,$$

$$7x_1 + 10x_2 + 9x_3 + 14x_4 \geq v,$$

$$16x_1 + 14x_2 + 11x_3 + 13x_4 \geq v,$$

$$10x_1 + 7x_2 + 11x_3 + 10x_4 \geq v,$$

$$x_1 + x_2 + x_3 + x_4 = 100,$$

$$x_i \geq 0 \quad i = 1, 2, \dots, n.$$

gdzie:

$x_i$  — obszar uprawy  $i$ -tego warzywa (ha),

$v$  — minimalny zagwarantowany dochód ogrodnika (tys. zł).

Rozwiązaniem jest wektor [ $x_1 \approx 0$ ,  $x_2 \approx 17$ ,  $x_3 \approx 64$ ,  $x_4 \approx 19$ ], co oznacza, że ogrodnik powinien na polu o powierzchni 100 ha uprawiać 17 ha marchwi, 64 ha fasoli oraz 19 ha selera.

W przypadku wielokrotnego stosowania strategii mieszanej, jeśli nie ma podstaw do twierdzenia, że któryś z możliwych stanów natury będzie występował z większą częstotliwością, można przyjąć, że średnie uzyskiwane rezultaty będą się zbliżać do wartości oczekiwanej. Zasadne wydaje się zatem zastosowanie reguły Bayesa, co ogólnie można zapisać następująco:

$$\sum_{i=1}^n \bar{a}_i x_i \rightarrow \max, \quad (11.12)$$

$$\sum_{i=1}^n x_i = 1,$$

$$x_i \geq 0, \quad i = 1, 2, \dots, n,$$

gdzie:

$$\bar{a}_i = \frac{1}{m} \sum_{j=1}^m a_{ij}. \quad (11.13)$$

W przypadku braku dodatkowych ograniczeń strategia mieszana realizowana wielokrotnie będzie odpowiadała najlepszej strategii czystej lub liniowej kombinacji najlepszych strategii czystych wyznaczonych według reguły Bayesa.

## 11.3. Elementy teorii gier

Potocznie **teoria gier** może kojarzyć się z brydżem, szachami czy pokerem. O ile w grach towarzyskich można zastosować pewne elementy teorii, o tyle do takich dziedzin, jak ekonomia, socjologia, psychologia czy nauki wojskowe, teoria gier wniosła nową jakość w zakresie analizy i podejmowania decyzji. Teoria gier zajmuje się analizą sytuacji konfliktu i kooperacji, w której występuje co najmniej dwóch graczy podejmujących świadome decyzje. Każdy z graczy może wywierać wpływ na przebieg oraz wynik gry, podejmując określone decyzje (przyjmując strategię), przy czym ostateczny wynik zależy również od strategii pozostałych graczy. Jeśli założymy, że każdy z graczy jest racjonalny, to przyjmujemy, że dąży on w krótkiej lub dłuższej perspektywie do osiągnięcia najlepszego, ze swojego punktu widzenia, wyniku gry. Mimo że naturalnym stanem jest występowanie konfliktu między graczami, mogą wystąpić sytuacje, w których koordynacja działań i decyzji uczestników gry prowadzi do poprawy wyniku uzyskiwanego przez każdego z nich<sup>6</sup>. A więc istnieje przestrzeń dla kooperacji. Klasycznym przykładem są porozumienia cenowe, zawierane przez podmioty bardzo skoncentrowanych sektorów w celu utrzymania cen na wyższym poziomie i czerpania z tego tytułu dodatkowych korzyści.

### DEFINICJA 11.4

Przez **grę** będziemy rozumieli sytuację, w której:

- można wskazać co najmniej dwóch graczy (graczem może być człowiek lub określona zbiorowość, a także podmioty bezosobowe, takie jak komputer, rynek),
- każdy z graczy dysponuje zbiorem możliwych strategii określających decyzje podejmowane przez niego w toku gry,
- wynik gry jest efektem wybieranych strategii przez poszczególnych graczy
- każdemu wynikowi gry można przypisać liczbę określającą wielkość wygranej bądź przegranej poszczególnych graczy (wyplata).

Gry możemy dzielić w zależności od przyjętego kryterium. Przykłady typologii gier przedstawiono w tabelicy 11.4.

W dalszej części ograniczymy się w rozważaniach do niekooperacyjnych gier dwuosobowych ze skończoną liczbą strategii.

<sup>6</sup> Konflikt bierze się stąd, że dążenie do osiągnięcia najlepszego dla siebie wyniku przez jednego z graczy stoi w sprzeczności z interesami innych uczestników gry, którzy chcieliby poprowadzić grę w taki sposób, aby maksymalizować swoją korzyść.

TABLICA 11.4  
Typologia gier

| Kryterium klasyfikacji              | Typ gier   |
|-------------------------------------|--|
| Liczba uczestników gry              | Gry dwuosobowe<br>Gry $n$ -osobowe                                       |
| Bilans wypląt                       | Gry o sumie zero<br>Gry o sumie niezerowej                               |
| Możliwości komunikowania się graczy | Gry kooperacyjne<br>Gry niekooperacyjne                                  |
| Charakter gry (powtarzalność)       | Gry rozgrywane jednokrotnie<br>Gry rozgrywane wielokrotnie (sekwencyjne) |
| Dostępna informacja                 | Gry z pełną informacją<br>Gry z niepełną informacją                      |
| Liczba strategii                    | Gry ze skończoną liczbą strategii<br>Gry z nieskończoną liczbą strategii |

### 11.3.1. Gry dwuosobowe o sumie zero

Grę najczęściej określa się w postaci macierzowej, która w przypadku dwóch graczy sprowadza się do macierzy wypląt, gdzie w wierszach umieszczamy strategię pierwszego gracza, w kolumnach strategię drugiego gracza, a na ich przecięciu wyrażone ilościowo wygrane lub przegrane (wyplaty) graczy.

Wygodnym sposobem analizy nieskomplikowanych gier sekwencyjnych jest drzewo decyzyjne, przedstawiające możliwe przebiegi gry w zależności od posunięć graczy (gra w postaci ekstensywnej).

### PRZYKŁAD 11.2

#### Gra w cyfry

Rozważmy grę, w której udział biorą dwaj gracze  $A$  i  $B$ . Wypisują oni równocześnie na karteczkach cyfrę od 1 do 4. Gracze umówili się, że jeśli obie cyfry są takie same, to żaden z nich ani nie przegrywa, ani nie wygrywa. W pozostałych przypadkach ten gracz, który wskazał cyfrę mniejszą, wypłaca temu, który wskazał cyfrę większą kwotę równą cyfrze podanej przez przeciwnika. Wynik gry można przedstawić w tabelicy wypląt, która zawiera wielkości wygranych i przegranych graczy w zależności od wybranych przez nich strategii (tablica 11.5).

Oznaczmy przez  $a_{ij}$  wygraną pierwszego gracza, jeżeli zastosuje strategię  $i$ , a drugi gracz strategię  $j$ . Jeśli (tak jak w powyższym przypadku) wielkość wygranej jednego z graczy  $a_{ij}$  jest równa przegranej drugiego gracza  $-b_{ij}$  ( $i$  odwrotnie), to mówimy o **grze dwuosobowej o sumie zero**. Wówczas elementy macierzy wypląt

TABLICA 11.5  
Macierz wpłat dla gry w cyfry

| Gracz B \ Gracz A | 1       | 2       | 3       | 4       |
|-------------------|---------|---------|---------|---------|
| 1                 | (0, 0)  | (-2, 2) | (-3, 3) | (-4, 4) |
| 2                 | (2, -2) | (0, 0)  | (-3, 3) | (-4, 4) |
| 3                 | (3, -3) | (3, -3) | (0, 0)  | (-4, 4) |
| 4                 | (4, -4) | (4, -4) | (4, -4) | (0, 0)  |

możemy zredukować do jednej wartości  $a_{ij}$ , która oznaczać będzie wielkość wygranej pierwszego gracza i jednocześnie wielkość przegranej drugiego gracza.

Gdy gracze podejmują decyzje jednocześnie i niezależnie od siebie, wówczas mamy do czynienia z **grą statyczną**<sup>7</sup>. Analiza gry zmierza albo do znalezienia wyniku gry (stanu równowagi ogólnej lub warunkowej oraz wartości gry), albo do ustalenia przebiegu gry w postaci dynamicznej.

Postępując racjonalnie, a więc dążąc do maksymalizacji swojej wygranej, gracz A nie powinien wybrać ani strategii 1, ani 2, ani 3, gdyż dla każdej z tych strategii istnieje strategia lepsza. Strategie 1, 2 i 3 dla gracza A nazywamy **strategiami zdominowanymi**.

## DEFINICJA 11.5

Strategię  $i$  będziemy nazywać zdominowaną przez strategię  $k$ , jeśli:

$$\forall_{j=1, \dots, m} a_{ij} \leq a_{kj} \quad \text{oraz} \quad \exists_l a_{il} < a_{kl} \quad (11.14)$$

gdzie przez  $j = 1, 2, \dots, l, \dots, m$  oznaczamy strategię drugiego gracza.

Jest oczywiste, że racjonalnie postępujący gracz dokona wyboru swojego działania spośród strategii niezdominowanych. Podobna analiza prowadzona z punktu widzenia gracza B prowadzi do analogicznych wniosków, że jedyną strategią niezdominowaną w jego przypadku jest również strategia 4. Mówimy, że strategia 4 jest dla obu graczy **strategią dominującą**.

<sup>7</sup> W odróżnieniu od gry w postaci statycznej, gra dynamiczna jest rozgrywana sekwencyjnie przez graczy w ustalonej kolejności.

## DEFINICJA 11.6

Strategię  $k$  będziemy nazywać dominującą, jeśli:

$$a_{kj} = \max_i \{a_{ij}\}, \quad j = 1, \dots, m. \quad (11.15)$$

Strategia dominująca to najlepsza możliwa reakcja na każdą strategię zastosowaną przez konkurenta. Odstąpienie od strategii dominującej nieuchronnie prowadzi do pogorszenia wyniku w przypadku gry niekooperacyjnej.

Szukając rozwiązania gry, zamiast redukować macierz o strategię zdominowaną, możemy rozważyć istnienie strategii dominującej (tablica 11.6). Z punktu widzenia gracza A, który chce maksymalizować wielkość swojej wypłaty  $a_{ij}$ , strategią dominującą będzie strategia 4. Gdyby założyć, że gracz B wybierze strategię 1, wówczas racjonalną odpowiedzią gracza A jest wybór strategii 4 (najwyższa wygrana), co zaznaczamy  $\bigcirc$ . Podobnie gdyby założyć, że gracz B wybierze strategię 2, 3 lub 4, wówczas racjonalną odpowiedzią gracza A jest również wybór strategii 4. Bez względu więc na to, co zrobi gracz B, wybór strategii 4 przez gracza A zapewnia mu większą wypłatę niż pozostałe strategie.

TABLICA 11.6  
Rozwiązanie gry w cyfry

| Gracz B \ Gracz A | 1            | 2            | 3            | 4            |
|-------------------|--------------|--------------|--------------|--------------|
| 1                 | 0            | -2           | -3           | $\square$ -4 |
| 2                 | 2            | 0            | -3           | $\square$ -4 |
| 3                 | 3            | 3            | 0            | $\square$ -4 |
| 4                 | $\bigcirc$ 4 | $\bigcirc$ 4 | $\bigcirc$ 4 | $\bigcirc$ 0 |

Gracz B będzie zainteresowany wyborem takiej strategii, która zapewni mu maksymalną wypłatę, co oznacza minimalną wygraną gracza A. Z tego punktu widzenia wybór przez gracza B strategii 4 jest rozwiązaniem najlepszym. Jeśli założyć, że gracz A wybierze strategię 1, to najlepszą decyzją gracza B jest wybór strategii 4 (najniższa wygrana gracza A, co oznacza najwyższą wygraną gracza B). Wybór ten oznaczamy przez  $\square$ . Analogicznie, jeśli założyć, że gracz A wybierze strategię 2, 3 lub 4, to najlepszą decyzją gracza B jest także wybór strategii 4. Strategia 4 jest strategią dominującą gracza B.

Jeżeli w grach dwuosobowych o sumie zero obaj gracze mają strategie

dominujące, to gra zawiera **stan równowagi** określony przez parę tych strategii (element macierzy wypłat, w którym  $\circ$  pokryło się z  $\square$ ). W prezentowanym przykładzie będzie to para strategii: 4, 4, niedająca żadnemu z graczy żadnej wygranej (wypłaty równe 0). Stan taki jest stabilny, gdyż żadnemu z graczy nie opłaca się odstąpić od strategii dominującej, niezależnie od tego, co zrobi konkurent. Ogólnie stan równowagi w grze dwuosobowej istnieje, gdy chociaż jeden z graczy posiada strategię dominującą.

Gdyby gra była rozgrywana wielokrotnie i, z jakichś powodów, jeden z graczy odstąpiłby od swojej strategii dominującej, wówczas kolejne racjonalne decyzje graczy sprawią, że gra powróci do punktu równowagi: 4, 4, co widać na diagramie przesunięć (tablica 11.7).

TABLICA 11.7  
Diagram przesunięć dla gry w cyfry

|         |  |   |   |   |   |
|---------|--|---|---|---|---|
| Gracz B |  | 1 | 2 | 3 | 4 |
| Gracz A |  |   |   |   |   |
| 1       |  |   |   |   |   |
| 2       |  |   |   |   |   |
| 3       |  |   |   |   |   |
| 4       |  |   |   |   |   |

Stan równowagi określany jest także jako **punkt siodłowy**. Dla jego wyznaczenia stosujemy regułę maximin dla jednego gracza oraz minimax dla drugiego gracza. Zakładamy, że obaj gracze chcą maksymalizować (minimalizować) swoją minimalną wygraną (maksymalną przegraną), a więc przyjmują postępowanie zachowawcze. Jeśli zastosowanie reguły maximin prowadzi do wygranej pierwszego gracza równej  $w_1^*$ , a zastosowanie reguły minimax prowadzi do wygranej drugiego gracza na poziomie  $w_2^*$  oraz spełniony jest warunek:

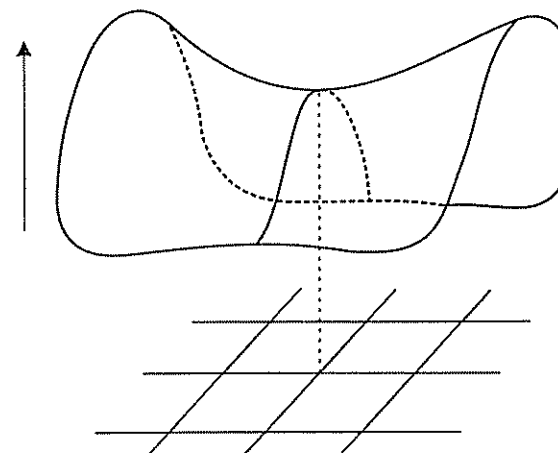
$$w_1^* = w_2^* \quad (11.16)$$

czyli:

$$\max_i (\min_j \{a_{ij}\}) = \min_j (\max_i \{a_{ij}\}),$$

to gra zawiera **punkt siodłowy**, a obaj gracze powinni realizować zawierające go strategie. Punkt siodłowy stanowi rozwiązanie gry. Oznacza on, że wygrane obu graczy ( $w_1^*$ ,  $w_2^*$ ) przyjmują najmniejszą wartość w swoim wierszu i jednocześnie największą wartość w swojej kolumnie (rysunek 11.1).

RYSUNEK 11.1  
Punkt siodłowy



Źródło: [Straffin, 2001, s. 9].

### PRZYKŁAD 11.3

Kandydaci na prezydenta prowadzą kampanię przedwyborczą. Załóżmy, że każdy z nich może przyjąć jedną z trzech strategii: odbywać spotkania z wyborcami głównie w dużych miastach (A), odbywać spotkania głównie w małych miastach i na wsiach (B) lub poświęcić się przede wszystkim poprawie wizerunku w mediach (C). Załóżmy, że w sondażach każdy z kandydatów otrzymuje podobny stopień poparcia. W zależności od zastosowanej strategii stopień poparcia dla kandydatów może ulegać zmianie, co przedstawiono w tablicy 11.8.

TABLICA 11.8  
Zmiany poparcia (w %) kandydatów  
pod wpływem zastosowanej  
kampanii przedwyborczej

| Kandydat I | Kandydat II  |              |           | min |
|------------|--------------|--------------|-----------|-----|
|            | A            | B            | C         |     |
| A          | $\square -2$ | -4           | -1        | -4  |
| B          | $\circ 4$    | $\square -6$ | -1        | -6  |
| C          | 0            | $\circ 2$    | $\circ 0$ | 0   |
| max        | 4            | 2            | 0         |     |

Przeprowadzając analizę macierzy wypłat, stwierdzamy, że żaden z graczy nie ma ani strategii dominującej, ani strategii zdominowanych. Stosując regułę maximin dla kandydata I wyznaczamy dla każdego wiersza minimalne wygrane (zmiany poparcia dla kandydata), a spośród strategii wybieramy tę, dla której minimalna wartość jest największa (strategia C). W przypadku kandydata II stosujemy regułę minimax z uwagi na fakt, że wygrane w macierzy wypłat oznaczają dla gracza II przegrane. Dla każdej kolumny wybieramy maksymalną wypłatę i spośród strategii wybieramy tę, dla której maksymalna wygrana jest minimalna (strategia C). Punktem siodłowym będzie więc para strategii (C, C), gdyż  $w_I = w_{II}$ . Obaj kandydaci stosując strategię C, maksymalizują swój „poziom bezpieczeństwo”, a więc gwarantują sobie, że bez względu na to, co zrobi przeciwnik polityczny, zmiana poparcia nie będzie mniejsza niż 0 (utrzymanie co najmniej *status quo*).

Istnieją przykłady gier, w których mamy kilka punktów siodłowych lub punkt siodłowy w ogóle nie istnieje w zbiorze strategii czystych.

### 11.3.2. Gry dwuosobowe o sumie niezerowej

Dla gier o sumie niezerowej wygrane jednego z graczy nie muszą oznaczać przegranych drugiego gracza<sup>8</sup>. W odróżnieniu od gier o sumie zero, gdzie przyjmowaliśmy ich ściśle antagonistyczny charakter, w grach o sumie niezerowej interesy obu graczy nie muszą być ściśle konkurencyjne. W pewnych sytuacjach możliwość kooperacji daje szansę poprawy wyniku gry obu graczy.

#### PRZYKŁAD 11.4

##### Dylemat więźnia

Dwie osoby podejrzewane o wspólne popełnienie przestępstwa trafiły do aresztu i są osobno przesłuchiwane. Prokurator przedstawił każdemu z nich możliwe wyniki postępowania. Jeżeli jeden z nich przyzna się do popełnionego czynu, a drugi nie, to przyznający się zostanie nagrodzony (wypłata 1), a ten, który się nie przyznał, dostanie surowy wyrok (-2). Jeśli obaj się przyznają, dostaną łagodny wyrok (-1), jeśli natomiast obaj się nie przyznają, zostaną uwolnieni (0). Macierz wypłat dla powyższego zagadnienia przedstawiona została w tablicy 11.9.

Każdy z graczy posiada strategię dominującą i w obu przypadkach jest to opcja „przyznać się”, co każdemu z oskarżonych daje perspektywę łagodnej kary. Stan równowagi wyznacza więc para strategii (P, P). Jak łatwo zauważyć, nie jest to

TABLICA 11.9  
Macierz wypłat dla zagadnienia dylematu więźnia

|         |                      | Gracz II             |                  |
|---------|----------------------|----------------------|------------------|
|         |                      | Nie przyznać się (N) | Przyznać się (P) |
| Gracz I | Nie przyznać się (N) | (0, 0)               | (-2, 1)          |
|         | Przyznać się (P)     | (1, -2)              | (-1, -1)         |

rozwiązanie optymalne w sensie Pareto, gdyż solidarne przyjęcie przez obu graczy strategii N daje poprawę ich sytuacji (uwolnienie). Taki stan, przy założeniu racjonalnie postępujących graczy, mógłby zaistnieć w przypadku, gdyby gracze mieli możliwość dogadania się (kooperacji) oraz zachowywali się lojalnie wobec siebie. Rozwiązanie (N, N) jest bowiem niestabilne, gdyż każdy z graczy miałby pokusę zmiany strategii. Zagadnienie dylematu więźnia ilustruje występujący konflikt między racjonalnością indywidualną (reprezentowaną przez kryterium dominacji) a racjonalnością (korzyścią) grupową, reprezentowaną przez kryterium optymalności Pareto.

W grach o sumie niezerowej strategii równowagi nie zawsze istnieją. Ten problem można rozwiązać przez zastosowanie słabszej koncepcji **równowagi Nasha**. Niech będą dane wypłaty poszczególnych graczy osiągane przy wyborze wybranych strategii, gdzie przez  $a_{ij}$  oznaczmy wygraną pierwszego gracza, a przez  $b_{ij}$  wygraną drugiego gracza.

#### DEFINICJA 11.7

Strategie  $k$  oraz  $l$  pozostają w równowadze Nasha, jeśli spełniony jest warunek:

$$\begin{aligned} \forall_{i \neq k} a_{ki} \geq a_{ii} \quad \text{oraz} \quad \exists_{i \neq k} a_{ki} > a_{ii}, \\ \forall_{j \neq l} b_{kl} \geq b_{kj} \quad \text{oraz} \quad \exists_{j \neq l} b_{kl} > b_{kj}. \end{aligned} \quad (11.17)$$

Równowaga w sensie Nasha składa się ze strategii poszczególnych graczy będących najlepszymi odpowiedziami na siebie nawzajem. Strategie graczy tworzą równowagę Nasha, jeżeli maksymalizują wypłatę gracza przy danym wyborze drugiego gracza. Taki stan określa się często **równowagą warunkową**, gdyż danemu graczowi nie oplaca się zmieniać strategii, jeżeli nie zrobi tego drugi gracz. Równowaga Nasha może ukształtować się w położeniu nieoptymalnym dla obu graczy (obaj gracze mogliby poprawić swoje wypłaty).

<sup>8</sup> W tym sensie gry o sumie zerowej są szczególnym przypadkiem gier o sumie niezerowej

Nash, 1951 r.

Każda dwuosobowa gra niekooperacyjna o sumie niezerowej ze skończonym zbiorem strategii czystych zawiera co najmniej jedną równowagę w zbiorze strategii czystych bądź mieszanych.

### PRZYKŁAD 11.5

#### Duopol Cournota

Dwa zakłady *A* i *B* całkowicie opanowały rynek pewnego produktu. Obaj producenci mogą podjąć decyzję o stosowaniu określonej strategii cenowej. Przyjmijmy ogólnie, że będzie to strategia wysokiej ceny (*W*), strategia średniej ceny (*S*) i strategia niskiej ceny (*N*). Osiągane przez producentów przychody ze sprzedaży będą zależne od wielkości sprzedaży tego produktu oraz oferowanej ceny. Macierz dochodów została przedstawiona w tablicy 11.10.

TABLICA 11.10  
Macierz wypłat dla problemu gry konkurencyjnej duopolistów

|         |   | Firma B  |          |          |
|---------|---|----------|----------|----------|
|         |   | W        | S        | N        |
| Firma A | W | (16, 19) | (15, 18) | (19, 21) |
|         | S | (21, 17) | (19, 19) | (17, 18) |
|         | N | (15, 20) | (17, 22) | (18, 19) |

Rozważając istnienie strategii dominujących dla obu graczy, stwierdzamy, że żadna z firm nie posiada takiej strategii (punkt równowagi ogólnej nie istnieje). Poszukując punktów równowagi w sensie Nasha zakładamy, że:

1. Firma *B* przyjmuje strategię wysokich cen (*W*). Wówczas najlepszą odpowiedzią firmy *A* byłoby przyjęcie strategii średnich cen (*S*) i zapewnienie sobie dochodu na poziomie 21. Ta para strategii (*S*, *W*) nie jest w równowadze z uwagi na to, że firma *B* będzie miała interes w zmianie strategii na *S*, co poprawi jej przychody (wzrost z 17 do 19).
2. Firma *B* przyjmuje strategię średnich cen (*S*). Wówczas najlepszą odpowiedzią firmy *A* byłoby przyjęcie strategii średnich cen (*S*) i zapewnienie sobie dochodu na poziomie 19. Żadna z firm nie będzie zainteresowana zmianą strategii, gdyż wiązałoby się to z pogorszeniem wyników. Para strategii (*S*, *S*) jest w równowadze w sensie Nasha.

3. Firma *B* przyjmuje strategię niskich cen (*N*). Wówczas najlepszą odpowiedzią firmy *A* byłoby przyjęcie strategii wysokich cen (*W*) i zapewnienie sobie dochodu na poziomie 19. Żadna z firm nie będzie zainteresowana zmianą strategii, gdyż wiązałoby się to z pogorszeniem wyników. Para strategii (*W*, *N*) jest w równowadze w sensie Nasha.

W przedstawionym przykładzie mamy więc dwie równowagi w sensie Nasha. Charakteryzują się one tym, że żadnemu z graczy nie opłaca się zmienić strategii, chyba że zrobi to drugi gracz.

W praktyce możemy spotkać się z sytuacją braku rozwiązania gry w sensie Nasha. Dla następującej macierzy wypłat (tablica 11.11) nie istnieje ani równowaga ogólna, ani równowaga w sensie Nasha w zbiorze strategii czystych.

TABLICA 11.11  
Przykład gry o sumie niezerowej

| Gracz I | Gracz II |         |         |
|---------|----------|---------|---------|
|         | A        | B       | C       |
| A       | (2, -2)  | (-1, 0) | (1, -2) |
| B       | (0, 0)   | (0, -1) | (2, -1) |
| C       | (1, -1)  | (1, 1)  | (0, 2)  |

### 11.3.3. Strategie mieszane w grach dwuosobowych

Jeżeli gra nie zawiera stanu równowagi w zbiorze strategii czystych, możemy określić dla obu graczy najlepsze strategie mieszane, które w tym przypadku będą oznaczać częstość (prawdopodobieństwo) wyboru określonej strategii czystej na kolejnym etapie<sup>9</sup>. Niech wektor  $\mathbf{X} = [x_1, x_2, \dots, x_n]$  oznacza prawdopodobieństwa, z jakimi gracz *A* wybiera poszczególne strategie czyste  $A_1, A_2, \dots, A_m$ , natomiast wektor  $\mathbf{Y} = [y_1, y_2, \dots, y_m]$  oznacza prawdopodobieństwa, z jakimi gracz *B* wybiera swoje strategie czyste  $B_1, B_2, \dots, B_m$ . Każdy z graczy będzie dążył do osiągnięcia maksymalnej wartości oczekiwanej swojej wygranej. Jeśli gracz *B* zastosuje strategię  $B_j$ , a gracz *A* strategię mieszaną daną przez wektor  $\mathbf{X} = [x_1, x_2, \dots, x_n]$ , to dla dowolnej macierzy wypłat  $W = [a_{ij}, b_{ij}]$  gracz *A* uzyska wygraną równą:

$$a_{1j}x_1 + a_{2j}x_2 + \dots + a_{nj}x_n \quad (11.18)$$

Przez  $v_A$  oznaczmy oczekiwaną wypłatę gracza *A* uzyskiwaną przy realizacji strategii  $\mathbf{X}$ .

<sup>9</sup> W przypadku strategii mieszanych zakładamy, że gry są powtarzalne.

Gracz  $A$  będzie zatem dążył do maksymalizacji swojej wygranej bez względu na to, co zrobi gracz  $B$ . Problem poszukiwania optymalnej strategii mieszanej dla gracza  $A$  można zapisać następująco:

$$\begin{aligned} v_A &\rightarrow \max, \\ \sum_{i=1}^n a_{ij}x_i &\geq v_A, \quad j = 1, 2, \dots, m, \\ \sum_{i=1}^n x_i &= 1, \\ x_i &\geq 0 \quad i = 1, 2, \dots, n. \end{aligned} \quad (11.19)$$

Podobne rozumowanie można przeprowadzić z punktu widzenia drugiego gracza. Jeśli gracz  $A$  przyjmie określoną strategię  $A_i$ , natomiast gracz  $B$  zastosuje strategię mieszaną daną przez wektor  $\bar{Y} = [y_1, y_2, \dots, y_m]$ , to wygraną gracza  $B$  można ogólnie zapisać w postaci:

$$b_{i1}y_1 + b_{i2}y_2 + \dots + b_{im}y_m. \quad (11.20)$$

Problem drugiego gracza definiujemy następująco:

$$\begin{aligned} v_B &\rightarrow \max, \\ \sum_{j=1}^m b_{ij}y_j &\geq v_B, \quad i = 1, 2, \dots, n, \\ \sum_{j=1}^m y_j &= 1, \\ y_j &\geq 0, \quad j = 1, 2, \dots, m. \end{aligned} \quad (11.21)$$

W szczególnym przypadku gier o sumie zero (jeśli  $a_{ij} = -b_{ij}$ ) powyższe zadania są problemami dualnymi względem siebie, a więc ich rozwiązania spełniają warunek:

$$v_A = v_B. \quad (11.22)$$

### PRZYKŁAD 11.6

#### Walka płci

Rozpatrzmy problem dwóch graczy: męża i żony. Mąż bardzo lubi chodzić na mecze piłkarskie ( $P$ ), trochę mniej do kina ( $K$ ) i wcale nie lubi chodzić do filharmonii ( $F$ ). Żona z kolei bardzo lubi chodzić do filharmonii ( $F$ ), trochę mniej do kina, a mecze piłkarskie „omija z daleka”. Mąż i żona określili użyteczność różnych wariantów spędzania czasu w pojedynkę lub wspólnie. Macierz wypłat (użyteczności) obu graczy przedstawiono w tabelicy 11.12.

Jak łatwo sprawdzić, żaden z graczy nie ma strategii dominującej, nie istnieje także stabilny stan równowagi w sensie Nasha. Szukając równowagi w zbiorze strategii mieszanych, rozwiązujemy dwa zadania programowania liniowego.

TABLICA 11.12  
Użyteczność poszczególnych strategii dla żony i męża

| Mąż | Żona    |        |         |
|-----|---------|--------|---------|
|     | A       | B      | C       |
| P   | (10, 3) | (5, 7) | (8, 4)  |
| K   | (3, 4)  | (6, 5) | (3, 6)  |
| F   | (0, 0)  | (0, 2) | (7, 14) |

Z punktu widzenia męża problem przyjmuje następującą postać:

$$\begin{aligned} v_A &\rightarrow \max, \\ 10x_P + 3x_K + 0x_F &\geq v_A, \\ 5x_P + 6x_K + 0x_F &\geq v_A, \\ 8x_P + 3x_K + 7x_F &\geq v_A, \\ x_P + x_K + x_F &= 1, \\ x_P, x_K, x_F &\geq 0. \end{aligned} \quad (11.23)$$

Rozwiązaniem jest wektor  $X = [0,5, 0,5, 0]$ , co oznacza, że aby maksymalizować swoją użyteczność, mąż przy każdej okazji powinien wybierać z jednakowym prawdopodobieństwem pójście na mecz lub do kina.

Rozwiązując podobny problem poszukiwania strategii mieszanej dla żony, rozwiązujemy poniższe zadanie:

$$\begin{aligned} v_B &\rightarrow \max, \\ 3y_P + 7y_K + 4y_F &\geq v_B, \\ 4y_P + 5y_K + 6y_F &\geq v_B, \\ 0y_P + 2y_K + 14y_F &\geq v_B, \\ y_P + y_K + y_F &= 1, \\ y_P, y_K, y_F &\geq 0. \end{aligned} \quad (11.24)$$

Rozwiązaniem jest wektor  $Y = [0, 0,5, 0,5]$ , co oznacza, że aby maksymalizować swoją użyteczność, żona przy każdej okazji powinna wybierać z jednakową częstotliwością pójście do kina lub do filharmonii.



Rozwiązanie gry oznacza, że gracze (mąż i żona), stosując optymalne dla siebie strategie mieszane nigdy nie spotkają się ani na meczu piłkarskim, ani w filharmonii, a więc ani żona, ani mąż nie osiągną swoich maksymalnych użyteczności. Mąż, gdyby żona towarzyszyła mu na meczu piłkarskim (10), żona gdyby udało się wyjść z mężem do filharmonii (14). Oczekiwana użyteczność dla obu graczy wynosi w tym przypadku tylko 5,5, podczas gdy stosując pewien kompromis (pójście raz na mecz piłkarski, raz do filharmonii) oboje mogliby osiągnąć oczekiwaną użyteczność na poziomie 8,5.

Jak starano się pokazać w prezentowanych przykładach, teoria gier ma swoje odniesienie do życia i praktyki gospodarczej. Zajmuje się nie tylko szukaniem rozwiązań w pewnych sytuacjach, w których występuje co najmniej dwóch niekoniecznie antagonistycznie nastawionych graczy, lecz także zmierza do opisu i wyjaśniania motywów i zasad ludzkiego działania. Trzeba więc wspomnieć o związkach teorii gier z psychologią behawioralną czy badaniami zachowań społecznych.

Warto na koniec zaznaczyć, że rozważania przedstawione w tym rozdziale prezentowały jedynie wybrane metody podejmowania decyzji w warunkach niepewności. Teoria gier została przedstawiona na przykładzie pewnej klasy gier statycznych, w których udział bierze dwóch graczy, postępujących racjonalnie, posiadających skończoną liczbę strategii, mających całkowitą wiedzę o strategiach i macierzy wypłat, zasadniczo niepodejmujących współpracy. Czytelników zainteresowanych rozwinięciem tego tematu odsyłamy do dostępnej literatury.

#### POLECENIA I PYTANIA KONTROLNE

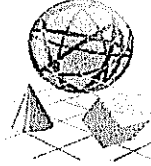
1. Wyjaśnij pojęcia strategii czystej i strategii mieszanej.
2. Scharakteryzuj podstawowe reguły decyzyjne w warunkach niepewności. Określ podobieństwa i różnice między nimi.
3. Uzasadnij, dlaczego w przypadku wyboru strategii czystej realizowanej wielokrotnie korzystamy z reguły Bayesa, a w przypadku wyboru strategii mieszanej realizowanej jednorazowo z reguły Walda.
4. Wyjaśnij pojęcia strategii zdominowanej i strategii dominującej.
5. Podaj warunek istnienia równowagi ogólnej w grach o sumie zero.
6. Na dowolnym przykładzie gry o sumie niezerowej wyjaśnij pojęcie równowagi Nasha. Czym różni się koncepcja Nasha od koncepcji równowagi ogólnej?
7. Wyjaśnij na dowolnym przykładzie, na czym polega konflikt między racjonalnością indywidualną a grupową.

#### BIBLIOGRAFIA

- Gontarek D., 2001, *Nowoczesne metody zarządzania ryzykiem finansowym*, WIG-Press, Warszawa.
- Kofler E., 1993, *Podejmowanie decyzji przy niepełnej informacji*, Real Publisher, Warszawa.
- Luce R.D., Reiffa H., 1963, *Gry i decyzje*, PWN, Warszawa.
- Małowski M., Wieczorek A., Sosnowska H., 2004, *Konkurencja i kooperacja. Teoria gier w ekonomii i naukach społecznych*, Wydawnictwo Naukowe PWN, Warszawa.
- Straffin P.D., 2001, *Teoria gier*, Schölar, Warszawa.
- Trzaskalik T., 2003, *Wprowadzenie do badań operacyjnych z komputerem*, PWE, Warszawa.
- Watson J., 2005, *Strategia: wprowadzenie do teorii gier*, Wydawnictwo Naukowo-Techniczne, Warszawa.

## ROZDZIAŁ 12

# ZARZĄDZANIE ZAPASAMI



### 12.1. Koszty zapasów

W procesie produkcji, na różnych jego etapach, niezbędne są materiały i surowce. Jednak ich pozyskiwanie i przechowywanie jest związane z ponoszeniem pewnych kosztów. Można wśród nich wyróżnić w szczególności koszty zakupu, koszty magazynowania, koszty transportu, koszty realizacji dostaw czy też koszty niedoboru zapasów (rozumiane najczęściej jako koszty utraconych korzyści związane z utratą potencjalnych klientów). Suma tych wszystkich (a niekiedy nie tylko tych) kosztów to całkowity **koszt funkcjonowania systemu zapasów**.

Minimalizacja kosztów utrzymywania zapasów jest przeważnie głównym celem. Nie należy jednak zapominać również o innych celach, jak zapewnienie rytmicznych dostaw i minimalizacja czasu przechowywania. Ze względu na losowy niekiedy charakter zapotrzebowania, należy mieć na uwadze utrzymanie pewnego zapasu bezpieczeństwa. Nie bez znaczenia są różnice w jakości usług świadczonych przez poszczególnych dostawców. Trzeba wreszcie pamiętać o przesunięciu czasowym między złożeniem zamówienia a jego realizacją.

Spełnieniu wszystkich wymienionych postulatów służy właściwa **polityka zarządzania zapasami**. Składają na nią: wybór dostawcy (w dalszych rozważaniach całkowicie pominiemy ten aspekt), określenie terminów i wielkości zamówień. Często określa się, przy jakim poziomie zapasów w magazynie należy złożyć zamówienie. Poziom ten nazywa się **poziomem odnowienia**; obliczany jest on na podstawie znajomości tempa zużycia zasobu, wielkości dostawy i czasu jej realizacji.

W dalszej części tego rozdziału Czytelnik zapozna się z klasycznym modelem zarządzania zapasami opartym na idei EOQ i jego rozszerzeniami. Następnie przedstawionych zostanie kilka modeli stosowanych w sytuacji nierównomiernego zapotrzebowania na magazynowany surowiec.

## 12.2. Klasyczny model EOQ i jego rozszerzenia

### 12.2.1. Klasyczny model EOQ

EOQ to skrót od *economic order quantity* (**ekonomiczna wielkość zamówienia**). Model został po raz pierwszy wprowadzony przez Harrisa, zasługę tę przypisuje się jednak Wilsonowi. Często model jest wręcz nazywany **modelem Wilsona**.

W klasycznym modelu EOQ zakłada się, że:

- 1) zużycie zasobu w czasie jest równomierne, a zapotrzebowanie  $D$  na ten zasób jest znane;
- 2) każde zamówienie opiewa na tę samą ilość zasobu i w związku z tym jednakowe dostawy są realizowane w stałych odstępach czasu;
- 3) zapotrzebowanie w każdym momencie musi być zaspokojone;
- 4) znane i stałe są: jednostkowa cena zakupu  $C_u$ , jednostkowy koszt magazynowania  $C_h$  i jednostkowy koszt realizacji zamówienia  $C_r$  (koszt stały zamówienia, niezależny od jego wielkości);
- 5) znany i stały jest czas realizacji dostawy  $L$ .

Zmienność poziomu zasobu w magazynie w czasie została przedstawiona na rysunku 12.1.  $Q_0$  oznacza ilość zasobu w magazynie,  $t$  — czas,  $Q'$  — wielkość dostawy, a  $T_0$  — czas upływający między jedną dostawą a drugą. Oczywiście czas ten jest równy:

$$T_0 = \frac{Q}{D} T, \quad (12.1)$$

gdzie:

$T$  — długość całego rozpatrywanego okresu,

$Q$  — wielkość jednej dostawy.

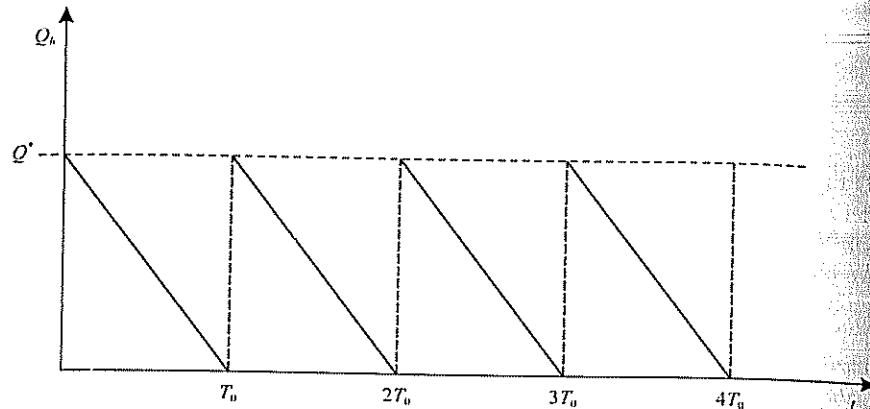
Łączny zakup w rozpatrywanym okresie nie zależy od wielkości jednej dostawy, ale od łącznego zapotrzebowania. Stąd całkowity koszt zakupu wyniesie:

$$K_u = C_u D. \quad (12.2)$$

Ekonomicznie uzasadnione jest takie planowanie dostaw, aby kolejna dostawa była realizowana w chwili, gdy skończą się zapasy z poprzedniej. Można więc założyć, że w każdym momencie magazynowana jest średnio ilość zasobu równa połowie wielkości dostawy. Stąd całkowity koszt magazynowania jest równy:

$$K_h = C_h \frac{Q}{2}. \quad (12.3)$$

RYSUNEK 12.1  
Zmienność poziomu zasobu w magazynie w czasie



Liczba zamówień to iloraz zapotrzebowania na zasób w całym okresie przez wielkość dostawy. Stąd koszt realizacji zamówień jest równy:

$$K_r = C_r \frac{D}{Q}. \quad (12.4)$$

Jak widać, jedynie koszt zakupu jest stały. Dwa pozostałe koszty zmieniają się wraz z wielkością zamówienia. Przy tym wraz ze wzrostem partii koszty magazynowania rosną (im większa partia, tym więcej towaru na raz musi być magazynowane), a koszty realizacji zamówień maleją (im większa partia, tym mniej zamówień trzeba zrealizować). Zadanie polega na minimalizacji łącznego kosztu, a więc na znalezieniu pewnego kompromisu między rosnącym kosztem magazynowania i malejącym kosztem realizacji zamówienia. Ta optymalna wielkość dostawy to właśnie EOQ. Zależności te przedstawione zostały na rysunku 12.2.  $KC$  oznacza koszt całkowity,  $KC_{\min}$  jego wartość minimalną, a  $Q^*$  optymalną wielkość dostawy (czyli EOQ).

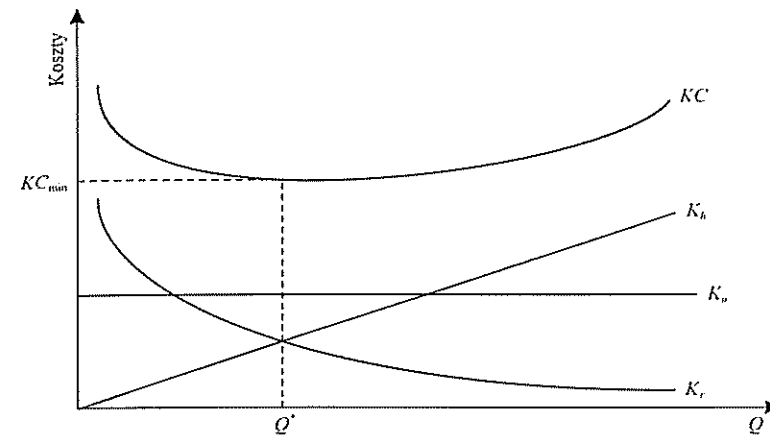
Całkowity koszt magazynowania więc jest równy:

$$KC(Q) = K_u + K_h + K_r = C_u D + C_h \frac{Q}{2} + C_r \frac{D}{Q}. \quad (12.5)$$

Aby wyznaczyć wielkość dostawy minimalizującą koszt, musimy przyrównać do 0 pierwszą pochodną powyższego wyrażenia:

$$\frac{dKC}{dQ} = 0. \quad (12.6)$$

RYSUNEK 12.2  
Koszty funkcjonowania systemu zapasów



Jest to równoważne równaniu:

$$\frac{C_h}{2} - C_r \frac{D}{Q^2} = 0. \quad (12.7)$$

Ostatecznym rozwiązaniem jest:

$$Q^* = \sqrt{\frac{2DC_r}{C_h}}. \quad (12.8)$$

### PRZYKŁAD 12.1

#### Klasyczny model EOQ

Łączne zapotrzebowanie na surowiec w ciągu dziewięciu tygodni wynosi 270 kg. Jednostkowy koszt magazynowania to 2 zł za 1 kg za tydzień, koszt realizacji zamówienia — 30 zł, jednostkowa cena zakupu — 10 zł, czas realizacji dostawy — 2 tygodnie. Określić optymalną wielkość dostawy i politykę zamawiania. Wyznaczyć wartość minimalnego kosztu funkcjonowania systemu zapasów.

Z warunków zadania wynika, że:  $D = 270$ ,  $C_h = 2$ ,  $C_r = 30$ ,  $C_u = 10$ ,  $L = 2$ . Stąd:

$$Q^* = \sqrt{\frac{2DC_r}{C_h}} = \sqrt{\frac{2 \cdot 270 \cdot 30}{2}} = 90.$$

Optymalna wielkość zamówienia wynosi więc 90 kg. Czas upływający między dostawami to  $T_0 = 3$  tygodnie, a więc polityka zamawiania powinna być taka, żeby

co trzy tygodnie przyjeżdżała nowa dostawa. Należy pamiętać o dwutygodniowym czasie realizacji dostawy. Między złożeniem zamówienia a jego realizacją zasób jest cały czas zużywany. Zamówienie musi więc zostać złożone w momencie, gdy poziom zapasów spadnie do poziomu dwutygodniowej rezerwy (ewentualnie pomniejszonej o łączną wielkość dostaw będących już w drodze). Skoro średnio na tydzień zużywa się 30 kg zasobu, to jego dwutygodniowa rezerwa musi być równa 60 kg. Ostatecznie więc polityka zamawiania przyjmuje następującą postać: zamówić 90 kg zasobu na 2 tygodnie przed rozpoczęciem rozpatrywanego okresu, a następnie zamawiać 90 kg za każdym razem, gdy poziom zasobu osiągnie 60 kg, oprócz ostatniego okresu.

Harmonogram składania i realizacji zamówień został ujęty w tabelicy 12.1.

TABLICA 12.1  
Harmonogram składania i realizacji zamówień

| Tydzień    | -1 | 0 | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | Razem |
|------------|----|---|----|----|----|----|----|----|----|----|----|-------|
| Zużycie    |    |   | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 270   |
| Dostawa    |    |   | 90 |    |    |    |    |    |    | 90 |    | 270   |
| Zamówienie | 90 |   |    | 90 |    |    | 90 |    |    |    |    | 270   |

Całkowity koszt funkcjonowania systemu zapasów przy takiej polityce wynosi:

$$KC(Q) = 10 \cdot 270 + 2 \cdot \frac{90}{2} + 30 \cdot \frac{270}{90} = 2\,880 \text{ zł.}$$

### 12.2.2. Model EOQ z uwzględnieniem zapasu w drodze

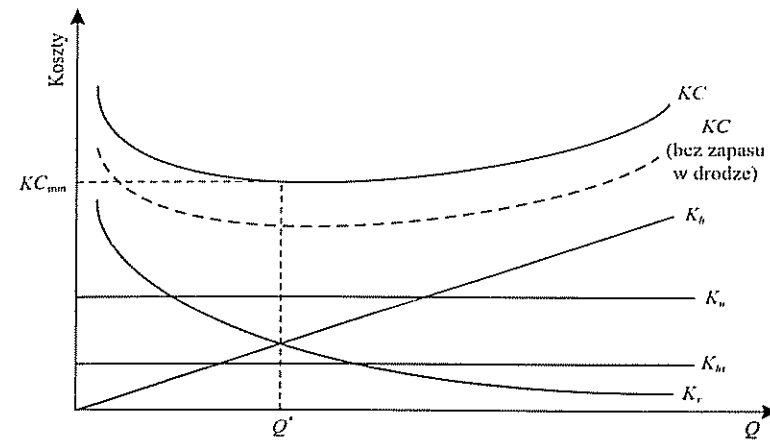
W modelu EOQ z uwzględnieniem zapasu w drodze<sup>1</sup> zakłada się, że przez pewien czas ( $t_m$ ) zapas jest utrzymywany w drodze, przy czym koszt utrzymania tego zapasu jest stały i wynosi  $C_{hr}$ . Przy obliczaniu całkowitego kosztu należy w takiej sytuacji uwzględnić również całkowity koszt utrzymania zapasu w drodze, czyli:

$$K_{hr} = \frac{t_m}{T} C_{hr} D. \quad (12.9)$$

Jak widać, koszt ten nie zależy od wielkości dostawy, zatem nie będzie miał wpływu na wielkość ekonomicznej partii dostawy (a więc również całej polityki

<sup>1</sup> W literaturze można spotkać również określenie zapas *in transit*.

RYSUNEK 12.3  
Koszty funkcjonowania systemu zapasów z uwzględnieniem zapasu w drodze



zamawiania), a jedynie na wartość całkowitego kosztu, który w tym przypadku będzie równy:

$$KC(Q) = K_u + K_h + K_r + K_{hr} = C_u D + C_h \frac{Q}{2} + C_r \frac{D}{Q} + \frac{t_m}{T} C_{hr} D. \quad (12.10)$$

Zależność ta pokazana jest na rysunku 12.3. Funkcja kosztu całkowitego, odpowiadającego sytuacji bez uwzględniania kosztów utrzymania zapasu w drodze, została przedstawiona za pomocą linii przerywanej.

#### PRZYKŁAD 12.2 (cd. 12.1)

Jednostkowy koszt utrzymywania zapasu w drodze wynosi 4 zł, a czas jego utrzymywania 2 tygodnie. Jak wpłynie to na politykę zamawiania i całkowity koszt funkcjonowania systemu zapasów?

Koszt utrzymania zapasu w drodze to:

$$K_{hr} = \frac{t_m}{T} C_{hr} D = \frac{2}{9} \cdot 4 \cdot 270 = 240 \text{ zł,}$$

stad całkowity koszt funkcjonowania systemu zapasów wzrośnie do:

$$K_{hr} = 2\,880 + 240 = 3\,120 \text{ zł.}$$

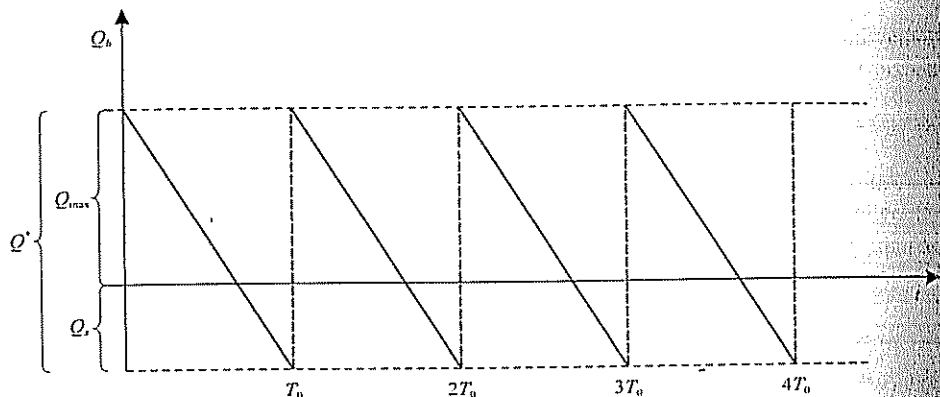
Polityka zamawiania nie zmieni się.

### 12.2.3. Model EOQ z uwzględnieniem planowych niedoborów

Niekiedy pełne zaspokajanie potrzeb klientów czy też linii produkcyjnej jest zbyt kosztowne, a potencjalne straty wynikające z niewielkiego niedoboru są nieproporcjonalnie małe. W takiej sytuacji rozwiązaniem może być dopuszczenie pewnego, z góry założonego, niedoboru przez pewien okres. Sytuację tę obrazuje rysunek 12.4, gdzie  $Q_{\max}$  oznacza maksymalny poziom zasobu w magazynie, a  $Q_s$  maksymalny dopuszczalny niedobór.

Niezależnie od dopuszczonego poziomu niedoboru, koszt zakupu się nie zmieni, gdyż nie jest on zależny od wielkości dostawy. Nie zmieni się również zależność wielkości kosztu realizacji zamówień od wielkości dostawy  $Q$ . Nie można już tego jednak powiedzieć o pozostałych składowych kosztu całkowitego.

RYSUNEK 12.4  
Zmienność poziomu zasobu w magazynie  
w czasie w sytuacji planowego niedoboru



Średnia ilość zasobu w magazynie będzie mniejsza niż w przypadku braku niedoboru proporcjonalnie do wielkości niedoboru. Podobnie rzecz się ma z czasem magazynowania zasobu. Wynika stąd, że koszt magazynowania jest w tym przypadku równy:

$$K_h = C_h \frac{Q}{2} \left( \frac{Q - Q_s}{Q} \right)^2 = C_h \frac{(Q - Q_s)^2}{2Q} \quad (12.11)$$

Analogicznie, koszt niedoboru jest równy ( $C_s$  oznacza jednostkowy koszt niedoboru):

$$K_r = C_s \frac{Q}{2} \left( \frac{Q_s}{Q} \right)^2 = C_s \frac{Q_s^2}{2Q} \quad (12.12)$$

Ostatecznie koszt całkowity systemu zapasów wynosi:

$$KC(Q, Q_s) = K_u + K_h + K_r + K_s = C_u D + C_h \frac{(Q - Q_s)^2}{2Q} + C_r \frac{D}{Q} + C_s \frac{Q_s^2}{2Q} \quad (12.13)$$

Jak widać, tym razem koszt całkowity jest funkcją dwóch zmiennych: wielkości zamówienia i wielkości niedoboru. Aby wyznaczyć ich optymalne wartości, trzeba przyrównać do 0 obie pochodne cząstkowe i rozwiązać odpowiedni układ równań:

$$\begin{cases} \frac{\partial KC}{\partial Q} = 0, \\ \frac{\partial KC}{\partial Q_s} = 0, \end{cases} \quad (12.14)$$

czyli:

$$\begin{cases} \frac{C_h}{2} \frac{2Q(Q - Q_s) - (Q - Q_s)^2}{Q^2} - C_r \frac{D}{Q^2} - C_s \frac{Q_s^2}{2Q^2} = 0, \\ C_h \frac{2(Q_s - Q)}{2Q} + C_s \frac{2Q_s}{2Q} = 0. \end{cases} \quad (12.15)$$

Układ ten upraszcza się do postaci:

$$\begin{cases} C_h Q^2 - (C_h + C_s) Q_s^2 - 2C_r D = 0, \\ C_h Q - (C_h + C_s) Q_s = 0, \end{cases} \quad (12.16)$$

stąd ostatecznie otrzymujemy rozwiązanie:

$$\begin{cases} Q^* = \sqrt{\frac{2DC_r}{C_h}} \sqrt{\frac{C_s + C_h}{C_s}}, \\ Q_s^* = \sqrt{\frac{2DC_r}{C_s}} \sqrt{\frac{C_h}{C_s + C_h}}. \end{cases} \quad (12.17)$$

Maksymalny poziom zapasu w magazynie to:

$$Q_{\max} = Q^* - Q_s^* = \sqrt{\frac{2DC_r}{C_h}} \sqrt{\frac{C_s}{C_s + C_h}} \quad (12.18)$$

## PRZYKŁAD 12.3 (cd. 12.1)

Załóżmy, że w omawianej sytuacji dopuszczalny jest niedobór. Jednostkowy koszt niedoboru wynosi przy tym 1,6 zł. Jak wpłynie to na politykę zamawiania?

Skoro  $C_s = 1,6$ , to — korzystając z odpowiednich wzorów — otrzymujemy

$$\begin{cases} Q^* = \sqrt{\frac{2DC_r}{C_h}} \sqrt{\frac{C_s + C_h}{C_s}} = \sqrt{\frac{2 \cdot 270 \cdot 30}{2}} \sqrt{\frac{1,6 + 2}{1,6}} = 135, \\ Q_s^* = \sqrt{\frac{2DC_r}{C_s}} \sqrt{\frac{C_h}{C_s + C_h}} = \sqrt{\frac{2 \cdot 270 \cdot 30}{1,6}} \sqrt{\frac{2}{1,6 + 2}} = 75. \end{cases}$$

Harmonogram składania i realizacji zamówień został ujęty w tabeli 12.2.

TABLICA 12.2  
Harmonogram składania i realizacji zamówień

| Tydzień    | 1   | 2  | 3   | 4  | 5   | 6  | 7   | 8  | 9  | Razem |
|------------|-----|----|-----|----|-----|----|-----|----|----|-------|
| Zużycie    | 30  | 30 | 30  | 30 | 30  | 30 | 30  | 30 | 30 | 270   |
| Dostawa    |     |    | 135 |    |     |    | 135 |    |    | 270   |
| Zamówienie | 135 |    |     |    | 135 |    |     |    |    | 270   |

## 12.2.4. Stochastyczny model EOQ

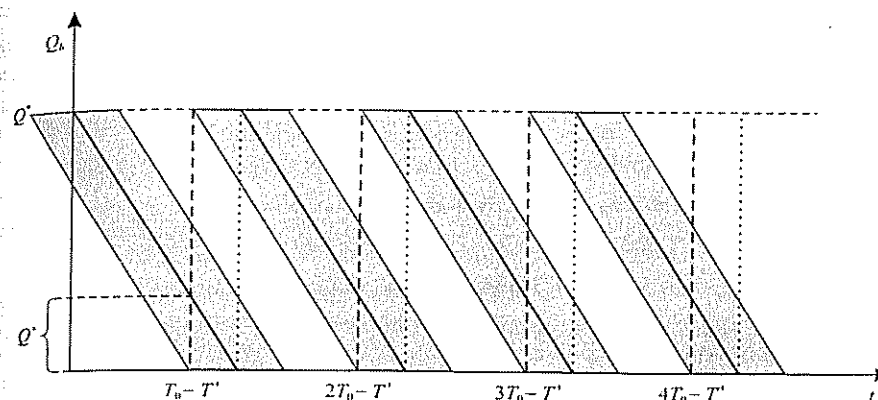
Założenie o znajomości przyszłej wielkości popytu bywa nierealistyczne. Popyt, jak wiele innych wielkości, podlega losowym fluktuacjom. Z tego też względu niekiedy jest sensowne zastosowanie stochastycznego podejścia przy planowaniu polityki zarządzania zapasami.

W celu zapewnienia odpowiedniego poziomu obsługi-utrzymywany jest tzw. zapas bezpieczeństwa. Z góry zakładamy pewien zadowalający poziom obsługi  $1 - p$ , a następnie dobieramy termin dostawy w ten sposób, żeby prawdopodobieństwo wystąpienia braku było nie większe niż  $p$  (czyli prawdopodobieństwo zaspokojenia potrzeb było równe  $1 - p$ ). Sytuacja ta jest przedstawiona na rysunku 12.5, gdzie  $Q^*$  oznacza zapas bezpieczeństwa, a  $T^*$  średni czas zużycia ilości zasobu odpowiadającej zapasowi bezpieczeństwa (czyli, równoważnie, część czasu dostawy proporcjonalną do zapasu bezpieczeństwa). Na rysunku 12.5 zacięziono obszar, na którym — z prawdopodobieństwem  $1 - p$  — będzie się mieścić rzeczywista ilość zasobu w magazynie.

Zapasz bezpieczeństwa jest wyznaczany zgodnie ze wzorem:

$$Q^* = \sqrt{Lz}\sigma, \quad (12.19)$$

RYSUNEK 12.5  
Zmienność poziomu zasobu w magazynie  
w czasie w sytuacji losowej zmienności popytu



gdzie:

$Q^*$  — zapas bezpieczeństwa,

$L$  — czas dostawy,

$\sigma$  — odchylenie standardowe popytu,

$z$  — wielkość, dla której dystrybuanta rozkładu popytu osiąga wartość  $1 - p$ .

Najczęściej przyjmujemy, że popyt ma rozkład normalny.

## PRZYKŁAD 12.4 (cd. 12.1)

Załóżmy, że odchylenie standardowe tygodniowego popytu jest równe 9, a planowany poziom obsługi to 99% (przyjmujemy, że popyt cechuje się rozkładem normalnym). Jaka powinna być polityka zamawiania?

$Q^* = 90$  (jak w przypadku deterministycznego EOQ),

$z = 2,326$ ,

$Q_s^* = \sqrt{2} \cdot 2,326 \cdot 9 = 29,61 \approx 30$ .

Ponieważ średni tygodniowy popyt równy jest właśnie 30, to każda partia towaru powinna być zamówiona (i dostarczona) o tydzień wcześniej niż w przypadku deterministycznego popytu. Politykę zamawiania w tym przypadku przedstawia tabela 12.3. Pierwsza dostawa, jako jedyna, jest realizowana według takiego samego harmonogramu, jak w przypadku popytu deterministycznego (wcześniej na pewno nie wystąpi zapotrzebowanie). Kolejne są realizowane już na podstawie przyspieszonego harmonogramu.

TABLICA 12.3  
Harmonogram składania i realizacji zamówień

| Tydzień    | -1 | 0 | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | Razem |
|------------|----|---|----|----|----|----|----|----|----|----|----|-------|
| Zużycie    |    |   | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 270   |
| Dostawa    |    |   | 90 |    | 90 |    |    | 90 |    |    |    | 270   |
| Zamówienie | 90 |   | 90 |    |    | 90 |    |    |    |    |    | 270   |

## 12.3. Polityka zamawiania w przypadku zapotrzebowania zmiennego w czasie

We wszystkich omówionych niżej przypadkach zakładamy, że zapotrzebowanie na zasób jest zmienne w czasie, ale znane. Skoncentrujemy się przy tym na polityce zamawiania, zakładając w razie potrzeby, że wielkość EOQ jest nam już znana (została wyznaczona wcześniej).

### 12.3.1. Zamówienie zgodne z zapotrzebowaniem

Zamówienie zgodne z zapotrzebowaniem to metoda zamawiania zgodna w pełni z filozofią *just-in-time* (JIT). W każdym okresie chcemy zrealizować dostawę zasobu w ilości równej dokładnie zapotrzebowaniu. Metoda ta pozwala całkowicie zrezygnować z magazynowania zasobu, ale znacznie podwyższa koszty realizacji zamówienia. Jest szczególnie dobra w przypadku towarów szybko psujących się, ale jej zastosowanie wymaga wysokiego poziomu niezawodności dostawców.

#### PRZYKŁAD 12.5

W tabeli 12.4. pokazano, w jaki sposób powinny być realizowane (i zamawiane) dostawy w przypadku przykładowego popytu.

TABLICA 12.4  
Harmonogram składania i realizacji zamówień

| Tydzień    | -1 | 0  | 1  | 2  | 3 | 4  | 5  | 6  | 7  | 8  | 9  | Razem |
|------------|----|----|----|----|---|----|----|----|----|----|----|-------|
| Zużycie    |    |    | 70 | 20 |   | 80 |    | 40 | 10 | 20 | 60 | 300   |
| Dostawa    |    |    | 70 | 20 |   | 80 |    | 40 | 10 | 20 | 60 | 300   |
| Zamówienie | 70 | 20 |    | 80 |   | 40 | 10 | 20 | 60 |    |    | 300   |

### 12.3.2. Stała wielkość zamówienia

W przypadku tej metody, najpierw jest ustalana pewna stała wielkość zamówienia. Najczęściej wyznacza się ją arbitralnie, czasem jednak używa się do tego deterministycznego modelu EOQ. Następnie dostawy są realizowane w taki sposób, żeby zamówiony towar dotarł w najpóźniejszym możliwym momencie umożliwiającym pełne pokrycie zapotrzebowania.

#### PRZYKŁAD 12.6 (cd. 12.5)

W tabeli 12.5 pokazano, w jaki sposób powinny być realizowane (i zamawiane) dostawy przykładowego popytu. Zakładamy przy tym, że wyznaczona uprzednio wartość EOQ wynosi 100 (pozostałe założenia się nie zmieniają).

TABLICA 12.5  
Harmonogram składania i realizacji zamówień

| Tydzień    | -1  | 0 | 1   | 2   | 3 | 4   | 5 | 6   | 7  | 8  | 9  | Razem |
|------------|-----|---|-----|-----|---|-----|---|-----|----|----|----|-------|
| Zużycie    |     |   | 70  | 20  |   | 80  |   | 40  | 10 | 20 | 60 | 300   |
| Dostawa    |     |   | 100 |     |   | 100 |   | 100 |    |    |    | 300   |
| Zamówienie | 100 |   |     | 100 |   | 100 |   |     |    |    |    | 300   |

### 12.3.3. Stały okres zamówienia

W tym przypadku punktem wyjścia jest ustalenie długości okresu, który wpływa między kolejnymi zamówieniami (lub, równoważnie, dostawami). Okres ten jest ustalany albo arbitralnie, albo na podstawie deterministycznego modelu EOQ (po wyznaczeniu wielkości dostawy, z odpowiedniej proporcji względem całościowego popytu i całkowitego okresu, na jaki planujemy, wyznaczamy okres zamówienia). Wielkość dostaw jest ustalana na najniższym możliwym poziomie umożliwiającym pełne pokrycie zapotrzebowania. Jeżeli przy tym dostawa miałaby przypaść na okres, w którym zapotrzebowanie jest zerowe, następuje jej przesunięcie na kolejny okres.

#### PRZYKŁAD 12.7 (cd. 12.5)

W tabeli 12.6 pokazano, w jaki sposób powinny być realizowane (i zamawiane) dostawy przykładowego popytu. Zakładamy przy tym, że wyznaczony stały okres dostawy wynosi 2 tygodnie.

TABLICA 12.6  
Harmonogram składania i realizacji zamówień

| Tydzień    | -1 | 0 | 1  | 2  | 3 | 4  | 5 | 6  | 7  | 8  | 9  | Razem |
|------------|----|---|----|----|---|----|---|----|----|----|----|-------|
| Zużycie    |    |   | 70 | 20 |   | 80 |   | 40 | 10 | 20 | 60 | 300   |
| Dostawa    |    |   | 90 |    |   | 80 |   | 50 |    | 80 |    | 300   |
| Zamówienie | 90 |   |    | 80 |   | 50 |   | 80 |    |    |    | 300   |

Jak widać, druga dostawa, która powinna być zrealizowana w trzecim tygodniu została przesunięta na tydzień czwarty (gdyż w trzecim tygodniu nie przewiduje się żadnego zapotrzebowania na zasób). Konsekwentnie, wszystkie kolejne dostawy również zostały przesunięte o jeden tydzień.

### 12.3.4. Najniższy koszt jednostkowy

W przypadku tej metody wielkość dostawy jest wyznaczana iteracyjnie tak, aby zminimalizować łączny koszt zamówienia i utrzymania zapasu przypadający na jednostkę zasobu. Wielkość dostawy jest ustalana na najniższym możliwym poziomie umożliwiającym pełne pokrycie zapotrzebowania.

Funkcja kosztu jednostkowego jest wypukła, w związku z czym okres zamówienia zwiększamy do momentu, kiedy kolejne jego wydłużenie spowodowałoby wzrost kosztu (ze względu na wypukłość koszt jednostkowy ma tylko jedno minimum).

#### PRZYKŁAD 12.8 (cd. 12.5)

Załóżmy, że koszt realizacji zamówienia wynosi 550 zł, a koszt utrzymania 1 kg zapasu przez tydzień 1 zł. Należy wyznaczyć momenty złożenia i wielkości dostaw minimalizujące jednostkowy koszt.

W tabelicy 12.7 zostały zamieszczone obliczenia pozwalające wyznaczyć moment realizacji i wielkość pierwszej dostawy. Przykładowo, koszt dla czwartego tygodnia został obliczony następująco. Realizacja w pierwszym tygodniu dostawy zaspokajającej zapotrzebowanie na cztery pierwsze tygodnie wymusiłaby przechowywanie 70 kg z pierwszego tygodnia przez 0 tygodni (zostałaby od razu zużyta), 20 kg z drugiego tygodnia przez 1 tydzień i 80 kg z czwartego tygodnia przez 3 tygodnie. Stąd łączny koszt magazynowania wynosi:

$$T_h = 1 \cdot (70 \cdot 0 + 20 \cdot 1 + 80 \cdot 3) = 260 \text{ zł.}$$

Łączną wielkość zamówienia obliczono jako sumę zapotrzebowania z czterech tygodni łącznie:

$$Q = 70 + 20 + 0 + 80 = 170.$$

TABLICA 12.7  
Wyznaczanie kosztu jednostkowego

| Tydzień  | Zapotrzebowanie | Czas utrzymania | Wielkość partii | Całkowity koszt utrzymania | Jednostkowy koszt utrzymania | Jednostkowy koszt zamówienia | Jednostkowy koszt ogółem |
|----------|-----------------|-----------------|-----------------|----------------------------|------------------------------|------------------------------|--------------------------|
| Pierwszy | 70              | 0               | 70              | 0                          | 0,00                         | 7,86                         | 7,86                     |
| Drugi    | 20              | 1               | 90              | 20                         | 0,22                         | 6,11                         | 6,33                     |
| Trzeci   |                 | 2               | 90              | 20                         | 0,22                         | 6,11                         | 6,33                     |
| Czwarty  | 80              | 3               | 170             | 260                        | 1,53                         | 3,24                         | 4,77                     |
| Piąty    |                 | 4               | 170             | 260                        | 1,53                         | 3,24                         | 4,77                     |
| Szósty   | 40              | 5               | 210             | 460                        | 2,19                         | 2,62                         | 4,81                     |

Jednostkowy koszt magazynowania to iloraz kosztu całkowitego przez łączną wielkość zamówienia:

$$U_h = T_h / Q = 260 / 170 = 1,53 \text{ zł.}$$

Podobnie, jednostkowy koszt zamówienia to iloraz kosztu zamówienia przez łączną wielkość zamówienia:

$$U_r = T_r / Q = 550 / 170 = 3,24 \text{ zł.}$$

Łączny koszt jednostkowy jest oczywiście sumą obu obliczonych kosztów jednostkowych.

W podobny sposób wyznaczono koszt jednostkowy dla wszystkich tygodni. Jak widać, w szóstym tygodniu koszt jednostkowy zaczyna wzrastać. Oznacza to, że okres minimalizujący wielkość jednostkowego kosztu wynosi 5 tygodni (a w rzeczywistości nawet 4, gdyż w piątym tygodniu nie przewiduje się żadnego zapotrzebowania).

Dalszy ciąg obliczeń wykonujemy, rozpoczynając od szóstego tygodnia. Wszystkie obliczenia zawarto w tabelicy 12.8.

TABLICA 12.8  
Wyznaczanie kosztu jednostkowego (cd.)

| Tydzień   | Zapotrzebowanie | Czas utrzymania | Wielkość partii | Całkowity koszt utrzymania | Jednostkowy koszt utrzymania | Jednostkowy koszt zamówienia | Jednostkowy koszt ogółem |
|-----------|-----------------|-----------------|-----------------|----------------------------|------------------------------|------------------------------|--------------------------|
| Szósty    | 40              | 0               | 40              | 0                          | 0,00                         | 13,75                        | 13,75                    |
| Siódmy    | 10              | 1               | 50              | 10                         | 0,20                         | 11,00                        | 11,20                    |
| Ósmy      | 20              | 2               | 70              | 50                         | 0,71                         | 7,86                         | 8,57                     |
| Dziewiąty | 60              | 3               | 130             | 230                        | 1,77                         | 4,23                         | 6,00                     |



TABLICA 12.9  
Harmonogram składania i realizacji zamówień

| Tydzień    | -1  | 0 | 1   | 2  | 3 | 4   | 5 | 6   | 7  | 8  | 9  | Razem |
|------------|-----|---|-----|----|---|-----|---|-----|----|----|----|-------|
| Zużycie    |     |   | 70  | 20 |   | 80  |   | 40  | 10 | 20 | 60 | 300   |
| Dostawa    |     |   | 170 |    |   |     |   | 130 |    |    |    | 300   |
| Zamówienie | 170 |   |     |    |   | 130 |   |     |    |    |    | 300   |

Jak widać, tym razem optymalna dostawa powinna obejmować wszystkie 4 pozostałe tygodnie, do dziewiątego włącznie. Ostateczny harmonogram dostaw został przedstawiony w tablicy 12.9.

#### POLECENIA I PYTANIA KONTROLNE

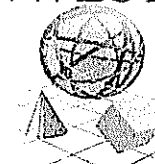
1. Scharakteryzuj koszty występujące w systemie zapasów.
2. Omów klasyczny model EOQ i jego deterministyczne rozszerzenia.
3. Omów stochastyczny model EOQ.
4. Scharakteryzuj metody zarządzania zapasami, jeżeli zapotrzebowanie jest zmienne w czasie.

#### BIBLIOGRAFIA

- Krawczyk S., 1996, *Badania operacyjne dla menedżerów*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
- Sarjusz-Wolski Z., 1998, *Strategia zarządzania zaopatrzeniem*, Agencja Wydawnicza Placet, Warszawa.
- Golemska E. (red.), 1999, *Kompendium wiedzy o logistyce*, Wydawnictwo Naukowe PWN, Warszawa-Poznań.

## ROZDZIAŁ 13

# OPTIMALIZACJA DECYZJI INWESTYCYJNYCH



### 13.1. Wprowadzenie

Metody ilościowe — w tym techniki optymalizacyjne wywodzące się z badań operacyjnych — są powszechnie stosowane do wspomagania decyzji inwestycyjnych. W niniejszym podręczniku przedmiotem rozważań będą inwestycje kapitałowe, ze szczególnym uwzględnieniem inwestycji w papiery wartościowe o zmiennej stopie dochodu, których typowym przykładem są akcje. Należy jednocześnie podkreślić, że tak zdefiniowany obszar nie pokrywa ani całości tematyki analizy inwestycji, ani wkładu nauk ilościowych w tę dziedzinę ekonomii. W tym rozdziale są przedstawione wybrane modele opisujące funkcjonowanie rynków finansowych, a także metody pozwalające podejmować racjonalne oraz — w świetle odpowiednio zdefiniowanych kryteriów — optymalne decyzje.

### 13.2. Pomiar zysku i ryzyka inwestycji w papiery wartościowe

Celem inwestowania w papiery wartościowe jest pomnażanie ułokowanego kapitału. Aby móc analizować i porównywać różne warianty inwestycyjne, należy zdefiniować miary, które pozwolą ocenić, w jakim stopniu dana inwestycja spełnia tak postawiony cel.

Podstawową charakterystyką dochodowości papieru wartościowego jest *stopa zwrotu*. Dla wybranego okresu  $t$  jest ona obliczana następująco:

$$R_t = \frac{P_t - P_{t-1} + C_t}{P_{t-1}}, \quad (13.1)$$

gdzie:

- $R_t$  — stopa zwrotu z papieru wartościowego w okresie  $t$ ,  
 $P_t$  — cena papieru wartościowego na koniec okresu  $t$ ,  
 $P_{t-1}$  — cena papieru wartościowego na początku okresu  $t$  (równoważna cenie tego waloru na koniec okresu  $t-1$ ),  
 $C_t$  — dodatkowe przepływy pieniężne związane z posiadaniem papieru wartościowego w okresie  $t$ , zarówno dodatnie (np. wypłacona posiadaczowi akcji dywidenda), jak i ujemne (np. poniesione koszty obsługi rachunku inwestycyjnego w domu maklerskim).

Wielkość ta informuje o zysku z posiadanej akcji osiągniętym w wyniku zmiany jej ceny, z uwzględnieniem ewentualnego poniesienia wydatków lub uzyskania przychodu związanego z posiadaniem waloru. Zysk ten jest odniesiony do nakładu na zakup akcji, w związku z czym uzyskana wartość stopy zwrotu jest wielkością względną. Możliwe jest obliczenie wartości stopy zwrotu dla przeszłych okresów na podstawie zaobserwowanych historycznych wartości ceny waloru i wartości dodatkowych przepływów pieniężnych z nim związanych, jednak jeśli rozpatrywany okres  $t$  ma się odnosić do przyszłości (co jest konieczne, jeśli analiza ma służyć podejmowaniu decyzji), to — wśród wielkości występujących we wzorze — jedynie początkowa cena akcji jest wartością znaną. Przyjmijmy zatem, że wielkości nieznane będziemy traktować jak **zmiennie losowe**, których poszczególnym realizacjom (np. cenom akcji w przyszłości) przyporządkowane są odpowiednie prawdopodobieństwa wystąpienia. Automatycznie stopa zwrotu, która jest funkcją zmiennych losowych, również staje się zmienną losową. Jeżeli przyjmiemy, że **rozkład prawdopodobieństwa stopy zwrotu jest dyskretny** i znane są możliwe warianty wartości przyszłej stopy zwrotu z danej akcji wraz z odpowiadającymi im prawdopodobieństwami, to możemy wyznaczyć **wartość oczekiwaną stopy zwrotu**:

$$E(R) = \sum_{i=1}^m r_i p_i \quad (13.2)$$

$$\text{przy } \sum_{i=1}^m p_i = 1,$$

gdzie:

- $E(R)$  — wartość oczekiwana stopy zwrotu,  
 $r_i$  — stopa zwrotu z akcji w wariantcie  $i$ ,  
 $p_i$  — prawdopodobieństwo wystąpienia wariantu  $i$ ,  
 $m$  — liczba rozpatrywanych wariantów wartości stopy zwrotu.

Rozkład prawdopodobieństwa zmiennej losowej stopy zwrotu można szacować np. na podstawie opinii ekspertów czy analityków giełdowych, ale prostszym

i częściej stosowanym podejściem jest wykorzystanie danych historycznych i przyjęcie założenia, że w rozpatrywanym okresie przyszłości notowanie danego waloru będzie zachowywało się tak samo, jak w pewnym okresie przeszłości, tj. powtórzy się któraś z zaobserwowanych już wcześniej stóp zwrotu. Jeśli założymy, że prawdopodobieństwo powtórzenia się jest takie samo dla każdej z  $T$  stóp zwrotu obserwowanych w przeszłości, to oczekiwana stopa zwrotu z akcji będzie wyrażona średnią arytmetyczną z historycznych stóp zwrotu:

$$E(R) = \bar{R} = \sum_{i=1}^T \left( r_i \frac{1}{T} \right) = \frac{r_1 + r_2 + \dots + r_T}{T} \quad (13.3)$$

### PRZYKŁAD 13.1

Przeanalizujmy kształtowanie się w okresie jednego roku (między 1.11.2005 a 31.10.2006) miesięcznych stóp zwrotu akcji dwóch notowanych na warszawskiej Giełdzie Papierów Wartościowych spółek — KGHM Polska Miedź SA oraz Netia Holdings SA. W rozpatrywanym okresie obie spółki wypłacały dywidendę, która powinna być uwzględniona w obliczeniach historycznych stóp zwrotu. Całość obliczeń przedstawia tablica 13.1.

TABLICA 13.1  
Wartości notowań, dywidendy oraz stopy zwrotu rozpatrywanych spółek

| Data       | Notowanie |       | Dywidenda |       | Stopa zwrotu |         |
|------------|-----------|-------|-----------|-------|--------------|---------|
|            | KGHM      | Netia | KGHM      | Netia | KGHM         | Netia   |
| 2005.10.31 | 46,20     | 4,65  |           |       | —            | —       |
| 2005.11.30 | 50,40     | 5,50  |           |       | 0,0909       | 0,1828  |
| 2005.12.31 | 62,50     | 5,70  |           |       | 0,2401       | 0,0364  |
| 2006.01.31 | 83,50     | 5,70  |           |       | 0,3360       | 0,0000  |
| 2006.02.28 | 74,00     | 5,45  |           |       | -0,1138      | -0,0439 |
| 2006.03.31 | 83,00     | 4,83  |           |       | 0,1216       | -0,1138 |
| 2006.04.30 | 112,50    | 5,65  |           | 0,13  | 0,3554       | 0,1967  |
| 2006.05.31 | 105,00    | 4,73  |           |       | -0,0667      | -0,1628 |
| 2006.06.30 | 108,00    | 4,47  |           |       | 0,0286       | -0,0550 |
| 2006.07.31 | 120,00    | 4,25  | 10,00     |       | 0,2037       | -0,0492 |
| 2006.08.31 | 105,50    | 4,20  |           |       | -0,1208      | -0,0118 |
| 2006.09.30 | 100,70    | 4,68  |           |       | -0,0455      | 0,1143  |
| 2006.10.31 | 112,50    | 5,13  |           |       | 0,1172       | 0,0962  |
| Średnia    | ×         | ×     | ×         | ×     | 0,0956       | 0,0158  |

Jeżeli przyjmiemy, że historyczne wartości miesięcznych stóp zwrotu odzwierciedlają z równym prawdopodobieństwem nieznane przyszłe wartości stóp zwrotu, to zgodnie ze wzorem (13.3) oczekiwana miesięczna stopa zwrotu z akcji spółki KGHM wyniesie 0,0956 (9,56%), a z akcji Netia — 0,0158 (1,58%).

Analiza przykładu 13.1 prowadzi do wniosku, że oczekiwana zyskowność inwestycji w akcje KGHM jest wyższa niż w przypadku akcji spółki Netia. Sama wartość oczekiwana stopy zwrotu nie jest jednak wystarczającym kryterium wyboru przedmiotu inwestowania, gdyż jest ona wyłącznie pewną prognozą nieznaną, przyszłej wartości stopy zwrotu, nie dając przy tym żadnej informacji o jakości (dokładności) tej prognozy. Z tego powodu, inwestor jest zmuszony uwzględnić w swojej analizie dodatkowe, pomocnicze kryterium — ryzyko.

Najpopularniejszą miarą ryzyka jest **wariancja**:

$$V(R) = E[R - E(R)]^2. \quad (13.4)$$

Jeżeli dany jest dyskretny rozkład zmiennej losowej stopy zwrotu, to wariancję obliczamy według wzoru:

$$V(R) = \sum_{i=1}^m p_i (r_i - E(R))^2. \quad (13.5)$$

TABLICA 13.2  
Obliczenia na potrzeby wyznaczenia  
wariancji stopy zwrotu  
dla każdej ze spółek

| Data       | $(r_i - E(R))^2$ |        |
|------------|------------------|--------|
|            | KGHM             | Netia  |
| 2005.11.30 | 0,0000           | 0,0279 |
| 2005.12.31 | 0,0209           | 0,0004 |
| 2006.01.31 | 0,0578           | 0,0003 |
| 2006.02.28 | 0,0438           | 0,0036 |
| 2006.03.31 | 0,0007           | 0,0168 |
| 2006.04.30 | 0,0675           | 0,0327 |
| 2006.05.31 | 0,0263           | 0,0319 |
| 2006.06.30 | 0,0045           | 0,0050 |
| 2006.07.31 | 0,0117           | 0,0042 |
| 2006.08.31 | 0,0468           | 0,0008 |
| 2006.09.30 | 0,0199           | 0,0097 |
| 2006.10.31 | 0,0005           | 0,0065 |
| Suma       | 0,3004           | 0,1397 |

Gdy szacujemy wariancję na podstawie rozkładu z próby historycznej, wtedy posługujemy się wzorem:

$$V(R) = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{R})^2. \quad (13.6)$$

Wariancja jest **miarą rozproszenia zmiennej losowej wokół wartości oczekiwanej**, co w naszym przypadku jest zbliżone z ryzykiem nieosiągnięcia zamierzonej (prognozowanej) wartości stopy zwrotu. Często, ze względu na łatwiejszą interpretację, ryzyko wyraża się **odchyleniem standardowym**, tj. pierwiastkiem z wariancji:

$$D(R) = \sqrt{V(R)}. \quad (13.7)$$

Wartość odchylenia standardowego odzwierciedla wyrażoną w punktach procentowych przeciętną różnicę między realizacją zmiennej losowej (tu: stopy zwrotu) a jej wartością oczekiwaną.

### PRZYKŁAD 13.2 (cd. 13.1)

Wyznamy wariancję stopy zwrotu akcji KGHM i akcji Netia. Tablica 13.2 zawiera potrzebne obliczenia.

Korzystając ze wzoru (13.6) i podstawiając  $T=12$ , obliczamy<sup>1</sup>:

$$V(R_A) = \frac{1}{12-1} 0,3004 = 0,02731,$$

$$V(R_B) = \frac{1}{12-1} 0,1397 = 0,01270.$$

Okazuje się, że wariancja stopy zwrotu akcji KGHM jest znacznie wyższa niż wariancja stopy zwrotu akcji Netia, co wskazuje na większą stabilność (mniejsze ryzyko) stopy zwrotu drugiej z wymienionych spółek.

Przedstawione charakterystyki stopy zwrotu z inwestycji jako zmiennej losowej są podstawowymi narzędziami analizy prognozowanej rentowności inwestycji w akcje. Zaprezentowane miary pozwalają nie tylko ocenić przewidywaną zyskowność inwestycji w walor danego rodzaju, lecz także oszacować poziom ryzyka inwestycji.

<sup>1</sup> Dla uproszczenia notacji przyjmijmy, że charakterystyki odnoszące się do akcji KGHM będą zapisywane z indeksem A, a te odnoszące się do akcji Netii — z indeksem B.

### 13.3. Portfel papierów wartościowych

Kurs papieru wartościowego na rynku giełdowym jest wypadkową bardzo wielu czynników. Na wzrost i spadek notowań spółki mają wpływ zarówno czynniki wewnętrzne związane z daną firmą, jak i czynniki zewnętrzne. Brak możliwości przewidzenia przyszłego stanu makro- i mikrootoczenia rodzi ryzyko inwestycji w akcje. Jednym ze sposobów redukcji tego ryzyka jest konstruowanie odpowiednio **zdywersyfikowanego portfela akcji**, czyli rozdzielenie całości inwestycji między pewną liczbę możliwie zróżnicowanych spółek.

#### PRZYKŁAD 13.3

Jednym z czynników kreujących ryzyko stopy zwrotu z akcji spółki giełdowej są wahania kursów walut. Jeśli spółka  $X$  eksportuje znaczną część swoich produktów bądź usług, to nagle umocnienie się rodzimej waluty (wzrost jej kursu wobec kursu walut zagranicznych) jest dla firmy niekorzystną wiadomością i może przełożyć się na spadek notowań spółki  $X$  na giełdzie papierów wartościowych (i, w konsekwencji, osiągnięcie stopy zwrotu niższej niż oczekiwana). Załóżmy, że istnieje również spółka  $Y$ , której domeną jest z kolei import produktów, przez co to samo zdarzenie, które wpływało niekorzystnie na spółkę  $X$  (umocnienie krajowej waluty), wpłynie pozytywnie na wyniki spółki  $Y$  i prawdopodobnie przełoży się również na wzrost jej notowań (i, w konsekwencji, osiągnięcie stopy zwrotu wyższej niż oczekiwana). Inwestor mógłby spróbować uniezależnić się od ryzyka związanego ze zmianą kursów walut przez skonstruowanie portfela papierów wartościowych, w którym znajdą się jednocześnie akcje spółek  $X$  i  $Y$  — te dwa komponenty portfela będą w przeciwny sposób reagować na nieprzewidziane zmiany na rynku walutowym, przez co powstałe odchylenia będą się nawzajem redukować, prowadząc do zwiększonej stabilności całego portfela (względem tego wyróżnionego źródła ryzyka).

Miarą pozwalającą wyodrębnić spółki, których notowania reagują w odmienny sposób na zdarzenia zachodzące w otoczeniu, jest **kowariancja stóp zwrotu**. Jest ona dana wzorem:

$$\text{cov}(R_A, R_B) = E[(R_A - E(R_A))(R_B - E(R_B))]. \quad (13.8)$$

Analogicznie do wartości oczekiwanej i odchylenia standardowego (wariancji) oszacowania wartości kowariancji będziemy dokonywać na podstawie przyjętego rozkładu prawdopodobieństwa wystąpienia poszczególnych wariantów stóp zwrotu (wzór (13.9)) lub na podstawie próbki obserwowanych w przeszłości wartości stóp zwrotu poszczególnych walorów (wzór (13.10)):

$$\text{cov}(R_A, R_B) = \sum_{i=1}^m p_i (r_{iA} - E(R_A))(r_{iB} - E(R_B)), \quad (13.9)$$

gdzie:  
 $r_{iA}, r_{iB}$  — wartości stopy zwrotu z walorów (odpowiednio  $A$  i  $B$ ) w  $i$ -tym wariantcie otoczenia,  
 $E(R_A), E(R_B)$  — wartości oczekiwane stóp zwrotu z walorów  $A$  i  $B$ ,  
 $p_i$  — prawdopodobieństwo wystąpienia  $i$ -tego wariantu otoczenia,  
 $m$  — liczba rozpatrywanych wariantów par stóp zwrotu.

$$\text{cov}(R_A, R_B) = \frac{1}{T-1} \sum_{t=1}^T (r_{tA} - \bar{R}_A)(r_{tB} - \bar{R}_B), \quad (13.10)$$

gdzie:  
 $r_{tA}, r_{tB}$  — realizacje stopy zwrotu z walorów (odpowiednio  $A$  i  $B$ ) w  $t$ -tej obserwacji historycznej,  
 $\bar{R}_A, \bar{R}_B$  — średnie arytmetyczne (wartości oczekiwane) stopy zwrotu z walorów  $A$  i  $B$ ,  
 $T$  — liczba obserwacji stóp zwrotu w przeszłości.

Oszacowanie kowariancji pozwala analizować, w jaki sposób stopy zwrotu poszczególnych walorów odchylają się **jednocześnie** od swoich wartości oczekiwanych. Mniejsza niż zero wartość kowariancji sugeruje, że gdy jeden z walorów uzyskuje realizację stopy zwrotu słabszą (niższą) niż wartość oczekiwana, wtedy realizacja stopy zwrotu z drugiego waloru lokuje się powyżej swojej wartości oczekiwanej. Zestawienie ze sobą w portfelu tak zachowujących się akcji powodowałoby — na skutek wzajemnej redukcji odchyień od wartości oczekiwanych — zmniejszenie stopnia niepewności co do końcowego rezultatu inwestycji. Prognoza stopy zwrotu z portfela oparta na oczekiwanych stopach zwrotu z poszczególnych akcji zyskałaby na wiarygodności.

Kowariancja daje informację o wzajemnych kierunkach odchylenia się realizacji stóp zwrotu od wartości oczekiwanych, niewiele jednak mówi o sile tej zależności, gdyż jest **wielkością niestandardyzowaną**. Aby ocenić nie tylko kierunek, lecz także i siłę współzależności odchyień stóp zwrotu, przeprowadza się standaryzację kowariancji, otrzymując **współczynnik korelacji**:

$$\rho_{AB} = \frac{\text{cov}(R_A, R_B)}{\sqrt{V(R_A)V(R_B)}}. \quad (13.11)$$

Znak współczynnika korelacji jest taki sam jak znak odpowiadającej mu kowariancji, ale wartość tego współczynnika jest unormowana w przedziale  $[-1, 1]$ . Wartość  $-1$  oznacza, że każde odchylenie realizacji jednej ze stóp zwrotu od wartości oczekiwanej spotyka się z idealnie proporcjonalnym co do wielkości,

ale przeciwnym co do znaku, odchyleniem realizacji stopy zwrotu drugiego waloru. Współczynnik korelacji równy 1 charakteryzuje natomiast zmienne losowe, których realizacje odchylają się od wartości oczekiwanej dokładnie tak samo, zarówno pod względem kierunku, jak i siły. Zerowa wartość współczynnika korelacji sugeruje brak zależności między wahaniami stóp zwrotu danej pary akcji.

Przyjmijmy, że inwestor może posiadać akcje  $n$  różnych spółek. Kapitał zaangażowany w akcje każdej ze spółek oznaczamy będziemy przez  $w_k$  ( $k = 1, 2, \dots, n$ ). Stąd, łączna kwota zainwestowanych środków to:

$$w = \sum_{k=1}^n w_k. \quad (13.12)$$

Zakładamy<sup>2</sup>, że  $w_k \geq 0$  i  $w > 0$ .

Zbiór inwestycji to **portfel papierów wartościowych**. Dla każdego z papierów możemy wyznaczyć jego **udział wartościowy w portfelu**:

$$x_k = \frac{w_k}{w}, \quad k = 1, 2, \dots, n. \quad (13.13)$$

Wtedy, oczywiście, udziały poszczególnych walorów sumują się do jedności (do 100%):

$$\sum_{k=1}^n x_k = 1. \quad (13.14)$$

**Oczekiwana stopę zwrotu z portfela papierów wartościowych** będziemy obliczać według następującej formuły:

$$E(R_p) = \sum_{k=1}^n x_k E(R_k) = \sum_{k=1}^n x_k \bar{R}_k, \quad (13.15)$$

gdzie  $E(R_k)$  jest oczekiwaną stopą zwrotu z  $k$ -tego waloru.

W szczególnym przypadku portfela dwuskładnikowego, czyli kiedy  $n = 2$ , wzór redukuje się do postaci:

$$E(R_p) = x_1 E(R_1) + x_2 E(R_2) = x_1 \bar{R}_1 + x_2 \bar{R}_2 \quad (13.16)$$

Oczekiwana stopa zwrotu z portfela jest więc średnią ważoną oczekiwanych stóp zwrotu wchodzących w jego skład papierów wartościowych, przy czym wagi odpowiadają udziałom wartościowym poszczególnych walorów w portfelu. Istotnym wnioskiem wynikającym z postaci wzoru (13.15) oraz z założenia o nieujemności udziałów składników jest to, że oczekiwana stopa zwrotu z portfela  $E(R_p)$

<sup>2</sup> Założenie o nieujemności udziałów każdego waloru jest w istocie założeniem o braku tzw. krótkiej sprzedaży. Założenie to zostanie uchylone w dalszej części rozdziału.

zawiera się między najniższą a najwyższą oczekiwaną stopą zwrotu spośród walorów wchodzących w jego skład:

$$\min_k E(R_k) \leq E(R_p) \leq \max_k E(R_k). \quad (13.17)$$

**Ryzyko stopy zwrotu portfela** może być — analogicznie do przypadku pojedynczej akcji — mierzone wariancją:

$$V(R_p) = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \text{cov}(R_i, R_j). \quad (13.18)$$

Dla portfela dwuskładnikowego wzór redukuje się do:

$$V(R_p) = x_1^2 V(R_1) + x_2^2 V(R_2) + 2x_1 x_2 \text{cov}(R_1, R_2). \quad (13.19)$$

Naturalnie, wariancja stopy zwrotu portfela jest kwadratem odchylenia standardowego.

#### PRZYKŁAD 13.4

Rozpatrzmy portfel akcji składający się w 20% z akcji spółki KGHM oraz w 80% z akcji spółki Netia. Charakterystyki stopy zwrotu i ryzyka dla tych dwóch walorów zostały już wyznaczone w przykładach 13.1 i 13.2, tj.:

- 1) wartość oczekiwana stopy zwrotu z akcji KGHM wynosi  $E(R_A) = 0,0956$  (9,56%);
- 2) wartość oczekiwana stopy zwrotu z akcji NETIA wynosi  $E(R_B) = 0,0158$  (1,58%);
- 3) wariancja stopy zwrotu KGHM wynosi  $V(R_A) = 0,0273$ ;
- 4) wariancja stopy zwrotu NETIA wynosi  $V(R_B) = 0,0127$ .

Dodatkowo, na podstawie wzoru (13.10) obliczamy wartość kowariancji stóp zwrotu tych dwóch spółek. Otrzymujemy  $\text{cov}(R_A, R_B) = 0,006251$ .

Udziały poszczególnych spółek wynoszą  $x_A = 0,2$  oraz  $x_B = 0,8$ . Zgodnie z wzorem (13.16) wartość oczekiwana stopy zwrotu z portfela to:

$$E(R_p) = x_A E(R_A) + x_B E(R_B) = 0,2 \cdot 0,0956 + 0,8 \cdot 0,0158 = 0,03177 \quad (3,18\%).$$

Ryzyko portfela mierzone wariancją zgodnie z wzorem (13.19) wynosi:

$$\begin{aligned} V(R_p) &= x_A^2 V(R_A) + x_B^2 V(R_B) + 2x_A x_B \text{cov}(R_A, R_B) = \\ &= 0,2^2 \cdot 0,0273 + 0,8^2 \cdot 0,0127 + 2 \cdot 0,2 \cdot 0,8 \cdot 0,006251 = 0,0114. \end{aligned}$$

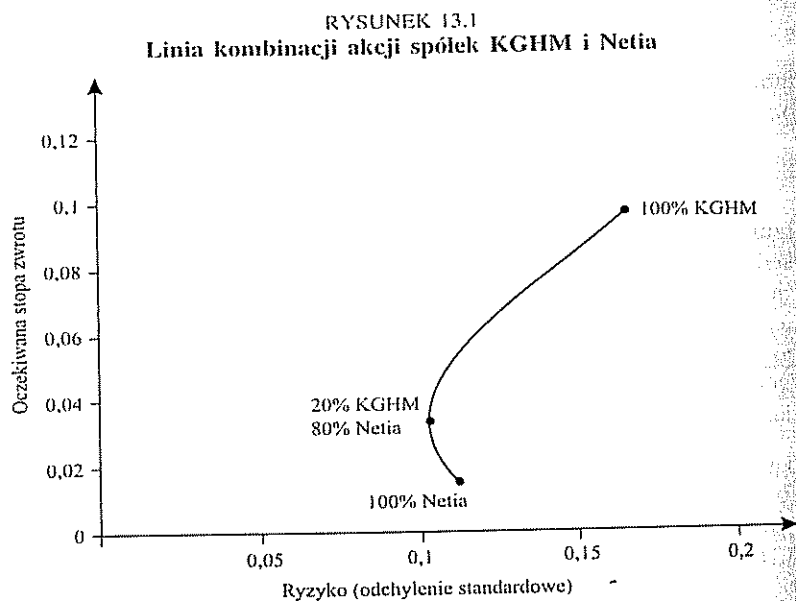
Warto zauważyć, że mierzone wariancją stopy zwrotu ryzyko otrzymanego portfela jest mniejsze niż ryzyko każdego ze składników osobno.

Korzystając z zależności podanych wzorami (13.15)–(13.16) oraz (13.18)–(13.19), możemy dokonać graficznej ilustracji wpływu dywersyfikacji portfela na jego zyskowność i ryzyko. W tym celu nanosimy na wykres punkty odpowiadające

wszystkim możliwym udziałom wybranych dwóch walorów w portfelu, odkładając na osi odciętych odchylenie standardowe stopy zwrotu z portfela (jako miarę ryzyka), a na osi rzędnych — oczekiwaną stopę zwrotu portfela. Powstałe w ten sposób krzywe noszą miano **linii kombinacji**.

### PRZYKŁAD 13.5 (cd. 13.1)

Linia kombinacji utworzona na podstawie portfeli dwuskładnikowych zawierających akcje KGHM oraz Netia ma postać przedstawioną na rysunku 13.1.



Kształt (a dokładnie siła zakrzywienia linii kombinacji) jest uzależniony od kierunku i siły współzależności między stopami zwrotu rozpatrywanych akcji. Współczynnik korelacji stóp zwrotu akcji KGHM oraz Netia obliczony zgodnie z wzorem (13.11) wynosi:

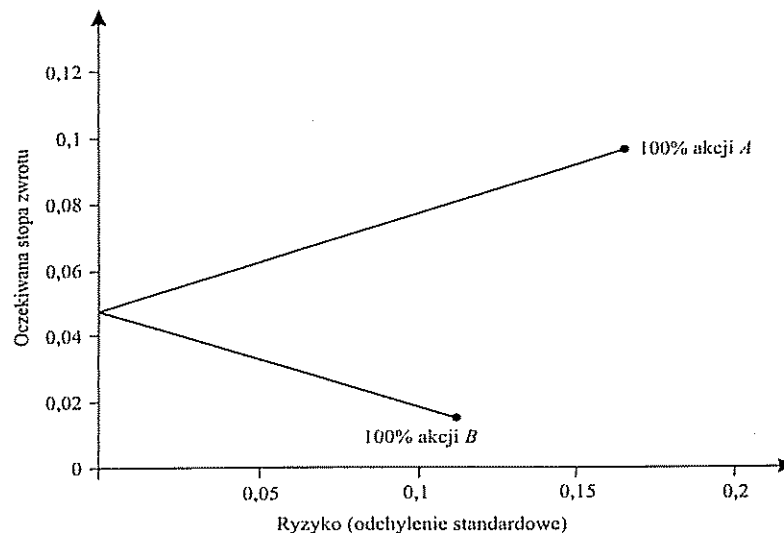
$$\rho_{AB} = \frac{\text{cov}(R_A, R_B)}{\sqrt{V(R_A)} \sqrt{V(R_B)}} = \frac{0,006251}{\sqrt{0,0273} \cdot 0,0127} = 0,33565.$$

W rozpatrywanym przypadku stopy zwrotu spółek są słabo dodatnio skorelowane. Im współczynnik korelacji jest mniejszy, tym większa jest potencjalna korzyść z dywersyfikacji portfela dwuskładnikowego.

Rozpatrzmy dwa skrajne przypadki wartości współczynnika korelacji stóp zwrotu:

- Jeżeli mamy do czynienia z doskonałą korelacją ujemną ( $\rho_{AB} = -1$ ), to linia kombinacji przyjmuje kształt zaprezentowany na rysunku 13.2. Co szczególnie warto odnotować, w przypadku doskonałej ujemnej korelacji istnieje portfel, którego teoretyczne ryzyko mierzone wariancją jest równe 0.

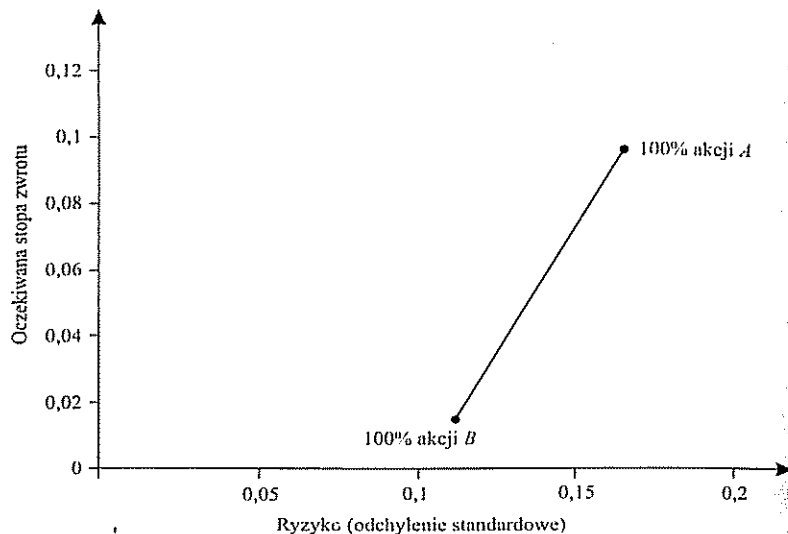
RYSUNEK 13.2  
Hipotetyczna linia kombinacji spółek o korelacji stóp zwrotu wynoszącej  $-1$



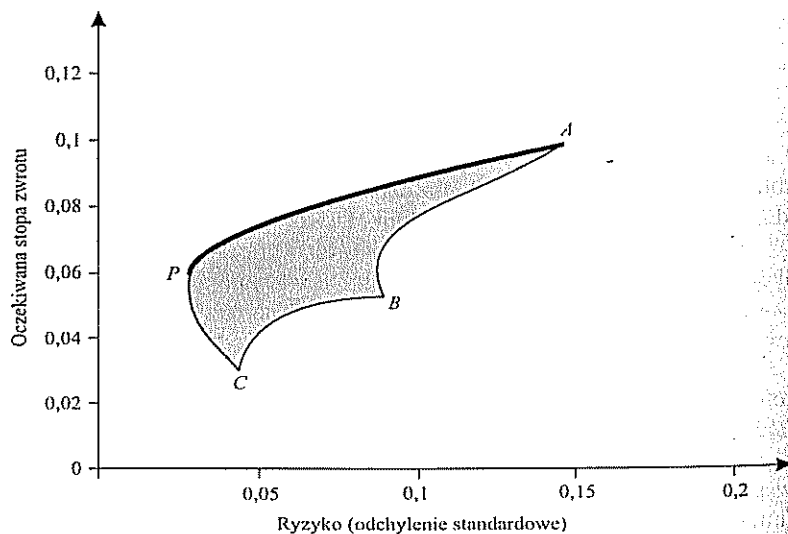
- Jeśli mamy do czynienia z doskonałą korelacją dodatnią ( $\rho_{AB} = 1$ ), to linia kombinacji jest prostym odcinkiem łączącym punkty reprezentujące portfele składające się w 100% z jednej z akcji. Utworzenie portfela z akcji, których stopy zwrotu są doskonale dodatnio skorelowane nie przynosi dodatkowej redukcji ryzyka, tzn. wartość oczekiwana i wariancja stopy zwrotu portfela jest zwykłą średnią ważoną odpowiednich parametrów poszczególnych składników.

Uogólnieniem idei linii kombinacji na przypadek portfela o większej niż dwa liczbie składników, jest **zbiór portfeli dopuszczalnych**, nazywany zwykle od nazwiska twórcy klasycznej analizy portfelowej **pociskiem Markowitza**. Graficznie powstaje on w wyniku naniesienia w układzie współrzędnych punktów reprezentujących wszystkie możliwe portfele  $n$ -składnikowe.

RYSUNEK 13.3  
Hipotetyczna linia kombinacji spółek  
o korelacji stóp zwrotu równej 1



RYSUNEK 13.4  
Zbiór portfeli dopuszczalnych oraz linia portfeli efektywnych



Na rysunku 13.4 literami *A*, *B*, *C* oznaczono punkty reprezentujące portfele składające się w 100% z akcji odpowiednio każdej z tych trzech spółek. Pozostałe punkty leżące wewnątrz zacienionego obszaru lub na jego krawędziach przedstawiają portfele złożone z możliwych kombinacji dwóch lub trzech spośród spółek *A*, *B*, *C*.

Punkt *P* reprezentuje **portfel o minimalnym ryzyku** (mierzonym wariancją lub odchyleniem standardowym) spośród wszystkich portfeli dopuszczalnych.

Górna krawędź zbioru portfeli dopuszczalnych, powyżej portfela minimalnego ryzyka, stanowi **zbiór portfeli efektywnych**. Portfel nazywamy efektywnym, jeżeli nie istnieje ani żaden inny portfel charakteryzujący się mniejszym ryzykiem przy tej samej oczekiwanej stopie zwrotu, ani portfel o wyższej stopie zwrotu przy tym samym poziomie ryzyka. Inaczej ujmując, jeśli wybór portfela potraktujemy jako dwukryteriowe zadanie maksymalizacji wartości oczekiwanej stopy zwrotu i minimalizacji jej wariancji, to zbiór portfeli efektywnych jest odpowiednikiem zbioru rozwiązań Pareto-optymalnych. Racjonalny inwestor powinien ograniczyć swój wybór do takiego zbioru wariantów inwestycyjnych.

## 13.4. Kryteria wyboru optymalnego portfela

W poprzednim podrozdziale ustaliliśmy, że inwestor dokonuje wyboru spośród portfeli efektywnych. Stwierdzenie to nie wyjaśnia jednak, w jaki sposób decydent ma porównać między sobą poszczególne portfele efektywne i dokonać ostatecznego wyboru. Można rozpatrywać wiele różnych podejść do przedstawionego problemu.

### 13.4.1. Maksymalizacja oczekiwanej stopy zwrotu przy założonym poziomie akceptowanego ryzyka

Wybór portfela sprowadza się w tym przypadku do znalezienia rozwiązania optymalnego poniższego zadania:

$$\begin{aligned}
 E(R_p) &\rightarrow \max, \\
 V(R_p) &\leq V^*, \\
 \sum_{i=1}^n x_i &= 1, \\
 x_i &\geq 0, \quad i = 1, 2, \dots, n,
 \end{aligned}
 \tag{13.20}$$

gdzie:

- $x_i$  — zmienna decyzyjna określająca udział wartościowy  $i$ -tego waloru w portfelu,  
 $n$  — liczba rozpatrywanych walorów,  
 $V^*$  — określony przez inwestora maksymalny akceptowany poziom wariacji stopy zwrotu portfela,  
 $E(R_p), V(R_p)$  — wartość oczekiwana stopy zwrotu z portfela oraz oszacowanie jej wariacji obliczane zgodnie ze wzorami podanymi w poprzednich podrozdziałach.

### 13.4.2. Minimalizacja ryzyka przy założonym poziomie akceptowanej stopy zwrotu

Struktura portfela jest wyznaczana przez rozwiązanie optymalne zadania:

$$\begin{aligned} V(R_p) &\rightarrow \min, \\ E(R_p) &\geq R^*, \\ \sum_{i=1}^n x_i &= 1, \\ x_i &\geq 0, \quad i = 1, 2, \dots, n, \end{aligned} \quad (13.21)$$

gdzie  $R^*$  jest określoną przez inwestora minimalną akceptowaną stopą zwrotu z inwestycji, natomiast pozostałe parametry są analogiczne jak w zadaniu określonym postacią (13.20).

### 13.4.3. Maksymalizacja funkcji użyteczności inwestora

Jest to bardziej wyrafinowane podejście do zagadnienia wyznaczania optymalnego portfela papierów wartościowych. Opiera się ono na założeniu, że inwestorzy różnią się między sobą **skłonnością do ryzyka**. Ideą jest skonstruowanie funkcji użyteczności, która będzie stanowiła o ocenie rozpatrywanych portfeli w świetle ich oczekiwanych stóp zwrotu, poziomu wariacji (czy odchylenia standardowego) oraz chęci inwestora do podejmowania ryzyka. Mimo że odpowiednia konstrukcja portfela pozwala zmniejszać niepewność co do przyszłej wartości inwestycji, to nadal chęć osiągania wyższych zysków wiąże się z koniecznością podejmowania dodatkowego ryzyka. Od inwestora zależy, jaki wzrost oczekiwanego dochodu jest dla niego akceptowalny jako rekompensata za dodatkową jednostkę ryzyka.

Można przyjąć następującą funkcję użyteczności:

$$U = E(R_p) - \frac{1}{2} \lambda V(R_p). \quad (13.22)$$

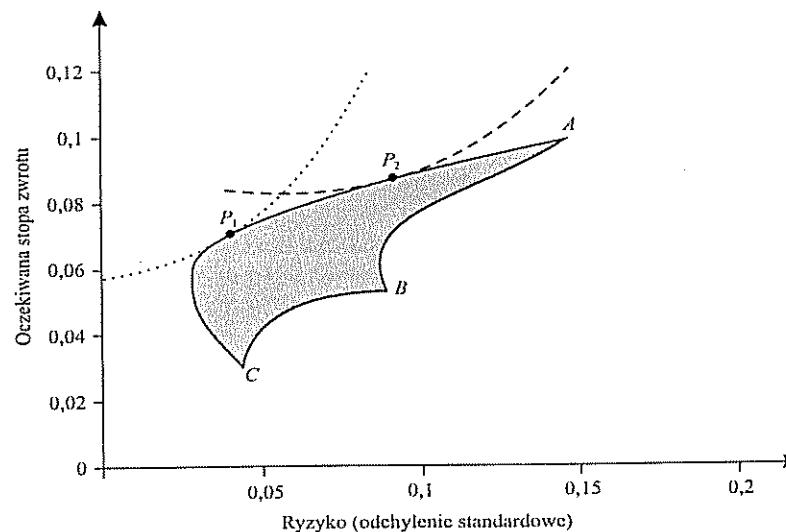
Parametr  $\lambda$  określa stopień ostrożności inwestora. Połowa wartości tego parametru może być interpretowana jako wynagrodzenie (wyrażone, w punktach procentowych, jako wzrost oczekiwanej stopy zwrotu) uznawane przez inwestora za odpowiednie za podjęcie dodatkowego ryzyka (wzrost wariacji stopy zwrotu o jedną jednostkę).

Formalnie, wybór portfela polega na wyznaczeniu rozwiązania optymalnego zadania:

$$\begin{aligned} E(R_p) - \frac{1}{2} \lambda V(R_p) &\rightarrow \max, \\ \sum_{i=1}^n x_i &= 1, \\ x_i &\geq 0, \quad i = 1, 2, \dots, n. \end{aligned} \quad (13.23)$$

Ideę wyboru portfela według zadania (13.23) można przedstawić w formie graficznej, stosując **izokwanty**, czyli zbiory punktów o tej samej wartości funkcji użyteczności. Im wyższa jest wartość parametru  $\lambda$ , tym bardziej stromy jest

RYSUNEK 13.5  
Wybór optymalnego portfela uwzględniający indywidualną dla inwestora skłonność do ryzyka





przebieg izokwant. Optymalny z punktu widzenia konkretnego inwestora portfel jest wyznaczony przez punkt styczności krzywej portfeli efektywnych do możliwie najwyżej położonej izokwenty. Rysunek 13.5 przedstawia sytuację wyboru portfela z punktu widzenia dwóch inwestorów, z których jeden charakteryzuje się względną ostrożnością (wysoka wartość  $\lambda$  — kropkowana izokwanta), drugi zaś wykazuje pewną skłonność do ryzyka (mała wartość  $\lambda$  — kreskowana izokwanta). Zgodnie z oczekiwaniami portfel wybrany przez pierwszego inwestora cechuje się mniejszą oczekiwaną stopą zwrotu oraz niższym poziomem ryzyka mierzonym odchyleniem standardowym.

### 13.4.4. Maksymalizacja premii za podejmowane ryzyko

Inwestor podejmuje zwiększone ryzyko, jeśli w jego ocenie działanie to zostanie wynagrodzone osiągnięciem zysku wyższego niż w przypadku działania obciążonego mniejszym ryzykiem. Inwestor musi określić, jaki poziom dodatkowego ryzyka warto zaakceptować w zamian za zwiększenie przewidywanej stopy zwrotu o jednostkę (np. o jeden punkt procentowy).

Punktem wyjścia w rozumowaniu inwestora może być inwestycja pozbawiona ryzyka (np. ulokowanie kapitału w bonach skarbowych). Stopa zwrotu z takiej inwestycji jest nazywana **stopą wolną od ryzyka**. Inwestor jest skłonny podejmować jakiegokolwiek ryzyko jedynie w przypadku, gdy wartość oczekiwanej stopy zwrotu przewyższy wartość stopy zwrotu wolnej od ryzyka. W szczególności inwestor może być zainteresowany wyborem takiej inwestycji, która wymaga jak najmniejszego wzrostu ryzyka w zamian za każdą jednostkę stopy zwrotu ponad stopę wolną od ryzyka. Ujmując ten sam problem od przeciwnej strony, inwestor dąży do zmaksymalizowania nadwyżki stopy zwrotu ponad stopę wolną od ryzyka (**premii za ryzyko**) przypadającej na każdą jednostkę podejmowanego ryzyka. Reguła ta jest wyrażona przez zadanie programowania nieliniowego:

$$\frac{E(R_p) - r_f}{S(R_p)} \rightarrow \max,$$

$$\sum_{i=1}^n x_i = 1,$$

$$x_i \geq 0, \quad i = 1, 2, \dots, n.$$

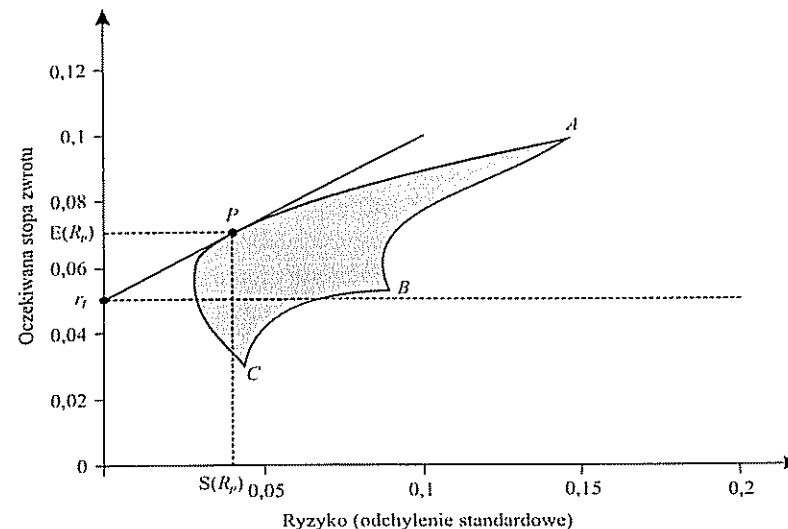
(13.24)

gdzie  $r_f$  oznacza wartość stopy zwrotu wolnej od ryzyka.

Poszukiwanie optymalnego rozwiązania powyższego zadania można zilustrować graficznie jako poszukiwanie punktu styczności półprostej o początku w punkcie  $(0, r_f)$  do zbioru portfeli efektywnych.

RYSUNEK 13.6

Ilustracja geometryczna wyboru portfela maksymalizującego premię za ryzyko



### 13.5. Krótka sprzedaż jako przykładowe rozwinięcie analizy portfelowej

**Krótka sprzedaż** to metoda inwestowania, której celem jest osiągnięcie zysku na spadku notowań akcji w czasie. Istotą tej strategii jest pożyczanie akcji od podmiotu, który takie akcje posiada (np. biura maklerskiego), sprzedaż ich na rynku po bieżącej cenie, a następnie — po upływie pewnego okresu — odkupienie akcji po nowej cenie i ich zwrot do pożyczkodawcy. Inwestor osiąga zysk, jeśli cena, po której sprzedał pożyczone akcje była wyższa, niż cena, po której je później odkupił<sup>3</sup>. Wejście w transakcję krótkiej sprzedaży jest określane mianem zajęcia **krótkiej pozycji** na danym papierze wartościowym, w przeciwieństwie do operacji zwykłego kupna akcji, które jest nazywane zajęciem **pozycji długiej**.

W przypadku inwestowania w portfel akcji, a nie tylko w pojedynczy walor, korzystanie z krótkiej sprzedaży pozwala osiągnąć dodatkowe zyski, gdyż tymczasowo posiadane środki, uzyskane ze sprzedaży pożyczonych papierów, można wykorzystać w celu zajęcia długiej pozycji na akcjach innej spółki.

<sup>3</sup> W praktyce uwzględnić należy również wynagrodzenie dla pożyczkodawcy; w niniejszych rozważaniach ten aspekt zostanie jednak pominięty.

## PRZYKŁAD 13.6

Inwestor posiada kapitał w wysokości 1000 USD. Na rynku są notowane akcje spółki *A* z bieżącą ceną 10 USD oraz akcje spółki *B* z bieżącą ceną 16 USD. Inwestor jest przekonany, że cena akcji spółki *A* będzie rosła, natomiast cena akcji spółki *B* będzie malała. W konsekwencji dokonał następujących transakcji:

- 1) zajął krótką pozycję na 40 sztukach akcji *B* (tzn. pożyczył je od Domu Maklerskiego i sprzedał po bieżącej cenie uzyskując w ten sposób 640 USD);
- 2) przeznaczył posiadane 1000 USD oraz 640 USD uzyskane ze sprzedaży akcji *B* na zakup 164 sztuk akcji spółki *A*.

Po miesiącu kurs akcji *A* wyniósł 12 USD, a kurs akcji *B* — 14 USD. Inwestor sprzedał posiadane 164 sztuk akcji *A* za łączną kwotę 1968 USD, a 40 sztuk akcji spółki *B* odkupił za kwotę 560 USD i zwrócił pożyczkodawcy. Saldo kapitału inwestora po przeprowadzonych transakcjach wynosi  $1968 - 560 = 1408$  USD, co daje miesięczną stopę zwrotu inwestycji na poziomie 40,8%.

Krótką sprzedaż można uwzględnić w zadaniu optymalizacji portfela papierów wartościowych przez dopuszczenie ujemnych udziałów poszczególnych spółek w portfelu. Wtedy udział większy od zera oznacza przeznaczenie odpowiedniej części kapitału na zajęcie długiej pozycji, a udział mniejszy od zera, odpowiednio, zajęcie pozycji krótkiej.

W przykładzie 13.5 wartość pożyczonych akcji spółki *B* wynosiła 640 USD i stanowiła 64% posiadanego kapitału, stąd  $x_B = -0,64$ , a akcje spółki *A* kupiono za kwotę 1640 USD, co stanowi 164% wyjściowego kapitału inwestora, dlatego przyjmujemy  $x_A = 1,64$ .

Wiemy, że stopy zwrotu poszczególnych spółek wyniosły, odpowiednio,  $R_A = 0,2$  oraz  $R_B = -0,125$ , co pozwala obliczyć — zgodnie ze znanym już wzorem (13.16) — wartość stopy zwrotu całego portfela, tj.  $R_p = 1,64 \cdot 0,2 + (-0,64) \cdot (-0,125) = 0,408$  (40,8%). Otrzymany wynik jest zgodny z wcześniejszymi obliczeniami. Analogicznie można wyznaczyć wariancję portfela, w którym jeden lub kilka składników podlega transakcji krótkiej sprzedaży; w tym celu wystarczy zastosować poznane już wzory (13.18)–(13.19).

Należy zauważyć, że w przypadku stosowania krótkiej sprzedaży nie jest spełniony układ nierówności (13.17), tzn. inwestując w portfel akcji z użyciem krótkiej sprzedaży, można osiągnąć stopę zwrotu niższą niż stopa zwrotu najmniej zyskownego spośród walorów wchodzących w jego skład, lub wyższą, niż stopa zwrotu najbardziej zyskownego spośród nich. Jest to efekt działania tzw. **dźwigni finansowej**<sup>4</sup>, który może być dla inwestora korzystny (ponieważ inwestując kapitał

<sup>4</sup> Więcej o dźwigni finansowej można znaleźć w podręcznikach z zakresu finansów przedsiębiorstw.

z pożyczonych akcji, można znacząco podnieść rentowność ulokowanych środków własnych), ale rodzi też znacznie wyższe ryzyko (ponieważ ryzykiem straty jest obciążony zarówno kapitał własny, jak i pożyczony). Ponoszone ryzyko jest tym większe, im większe jest zaangażowanie kapitałowe w krótką sprzedaż. Poniższe zadanie optymalizacji portfela papierów wartościowych uwzględnia dodatkowy warunek ograniczający skalę stosowania krótkiej sprzedaży:

$$\begin{aligned} E(R_p) &\rightarrow \max, \\ V(R_p) &\leq V^*, \\ \sum_{i=1}^n x_i &= 1, \\ \sum_{i=1}^n |x_i| &\leq 1 + 2H, \\ x_i &\geq 0, \quad i = 1, 2, \dots, n, \end{aligned} \tag{13.25}$$

gdzie  $H$  oznacza maksymalny dopuszczalny stosunek wartości krótkiej sprzedaży do wartości inwestowanego kapitału własnego.

## POLECENIA I PYTANIA KONTROLNE

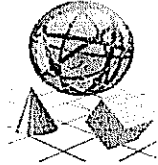
1. Dlaczego stopa zwrotu nie jest wystarczającym kryterium podejmowania decyzji inwestycyjnych?
2. Objasnij pojęcie dywersyfikacji portfela i uzasadnij stosowanie dywersyfikacji jako narzędzia redukcji ryzyka inwestycji.
3. W jaki sposób kształt linii kombinacji zależy od współczynnika korelacji stóp zwrotu spółek?
4. Czym różni się linia kombinacji od zbioru portfeli dopuszczalnych?
5. Dlaczego racjonalny inwestor powinien ograniczyć się do portfeli leżących na linii portfeli efektywnych?
6. Jakie przykładowe podejścia może zastosować inwestor, aby wybrać optymalny dla siebie portfel papierów wartościowych?

## BIBLIOGRAFIA

- R. Haugen, 1996, *Teoria nowoczesnego inwestowania. Obszerny podręcznik analizy portfelowej*, WIG Press, Warszawa.
- K. Jajuga, T. Jajuga, 1993, *Jak inwestować w papiery wartościowe*, Wydawnictwo Naukowe PWN, Warszawa.
- W. Jurek, 2001, *Konstrukcja i analiza portfela papierów wartościowych o zmiennym dochodzie*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.
- W. Tarczyński, 1997, *Rynki kapitałowe. Analiza portfelowa. Analiza banków*, Placet, Warszawa.

## ROZDZIAŁ 14

# SZTUCZNE SIECI NEURONOWE



### 14.1. Wprowadzenie

**Sztuczne sieci neuronowe** są to modelowane przy użyciu technik komputerowych struktury, których celem jest przekształcanie informacji w sposób wzorowany na działaniu mózgu, czy też ogólniej — systemu nerwowego żywych istot. Sieć neuronowa składa się z dużej liczby wykonujących proste operacje elementów przetwarzających (neuronów) oraz licznych połączeń między nimi. Elementy te przyjmują sygnały wejściowe (w przypadku biologicznego neuronu takim sygnałem jest impuls elektryczny), przetwarzają je i przesyłają do innych elementów.

W biologicznym odpowiedniku sieci neuronowej typowa droga, którą przebywa informacja, prowadzi od nerwów sensorycznych, poprzez skomplikowaną sieć wzajemnie połączonych komórek nerwowych przetwarzających sygnały, aż do nerwów motorycznych. Podobnie w sztucznej sieci, poszczególne neurony mogą pełnić jedną z trzech ról:

- 1) neuronu wejściowego, który pobiera informacje wejściowe z zewnętrznego środowiska;
- 2) neuronu ukrytego (wewnętrznego), który stanowi pośrednie ogniwo na drodze kolejnych przekształceń przenoszonych sygnałów;
- 3) neuronu wyjściowego, który stanowi wyprowadzenie sygnału odzwierciedlającego rezultat przetwarzania dokonanego przez sieć.

Mimo że zasada działania pojedynczego neuronu jest bardzo prosta, to — dzięki występowaniu ich w dużej liczbie i we wzajemnych połączeniach — zorganizowana sieć neuronów jest w stanie odzwierciedlać pewną wiedzę i generować zadowalające rozwiązania postawionych zadań. W odróżnieniu od klasycznego sposobu przetwarzania informacji, stosowanego we współczesnych komputerach, sieć neuronowa nie wymaga zaprogramowania algorytmu, który opisywałby krok po kroku kolejne czynności do wykonania, aby przekształcić dane wejściowe

w oczekiwany rezultat — „wiedza” sieci o sposobie rozwiązywania zadania jest ukryta w jej strukturze, tj. w charakterystyce neuronów i połączeń, z których sieć się składa, a charakterystyka ta powstaje nie poprzez zaprogramowanie, a przez uczenie. Zadaniem użytkownika jest wstępnie zdefiniować sieć (określić jej typ, liczbę neuronów, podział na warstwy itp.<sup>1</sup>) oraz dostarczyć przykładowe dane, na podstawie których sieć nauczy się generować poprawne wyniki. Zaletą takiego podejścia jest to, że nawet sam użytkownik nie musi wiedzieć, w jaki sposób dane wejściowe zadania powinny być przekształcane w poprawną odpowiedź. Inne zalety sieci neuronowych, jakie są najczęściej wskazywane w literaturze, to fakt, że poszczególne ich elementy przetwarzają informacje równolegle (w przeciwieństwie do przetwarzania sekwencyjnego, charakterystycznego dla sposobu działania klasycznych komputerów), co może prowadzić do szybszego otrzymania wyniku, oraz to, że sieci są odporne na uszkodzenia, tzn. nawet w przypadku, gdy część neuronów lub połączeń nie będzie funkcjonować, cała sieć nadal działa, generując co najwyżej gorsze rozwiązanie.

Sieci neuronowe mogą być wykorzystywane do rozwiązywania różnych typów problemów, z których najczęściej wyróżnia się poniższe kategorie:

- 1) **prognozowanie**, czyli przewidywanie natężenia pewnej cechy bądź zjawiska, na podstawie znanych informacji wejściowych, np. prognozowanie kursów akcji na najbliższej sesji giełdowej, prognozowanie zapotrzebowania na energię dostarczaną z danej elektrowni;
- 2) **klasyfikacja i rozpoznawanie**, czyli wskazywanie grupy, do której dany obiekt z największym prawdopodobieństwem należy, biorąc pod uwagę jego cechy, np. klasyfikacja kredytobiorców jako wiarygodnych lub niewiarygodnych na podstawie informacji podanych we wniosku kredytowym oraz historii kredytowej, rozpoznawanie znaków drogowych na podstawie obrazu uzyskanego z zainstalowanej w samochodzie kamery obserwującej pobocze;
- 3) **optymalizacja**, czyli poszukiwanie rozwiązań danego problemu spełniających w jak najwyższym stopniu postawione kryteria, np. wyznaczenie najkrótszej trasy rozwózki towaru, przebiegającej przez wszystkie punkty na liście (problem komiwojażera i jego pochodne).

### 14.2. Budowa sztucznej sieci neuronowej

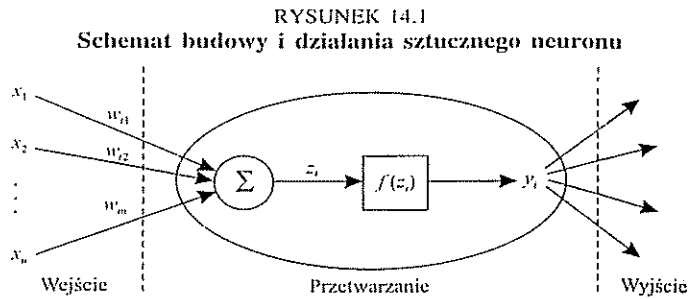
Na wejściu do neuronu występuje  $n$  połączeń prowadzących od innych neuronów, których bieżące stany (wartości ich sygnałów) są opisane jako  $x_1, x_2, \dots, x_n$ . Parametr  $w_{ij}$  określa wagę przypisaną połączeniu<sup>2</sup> prowadzącemu z neuronu  $j$  do

<sup>1</sup> Aspekty te będą szczegółowo omawiane w dalszej części rozdziału.

<sup>2</sup> Proszę zwrócić uwagę na nietypową kolejność indeksów w parametrze  $w_{ij}$ .

neuronu  $i$ . Sygnały na wejściu do neuronu są sumowane zgodnie z ich wagami, tj. zagregowany sygnał wchodzący do neuronu  $i$  jest dany wzorem:

$$z_i = \sum_j x_j w_{ij} \quad (14.1)$$



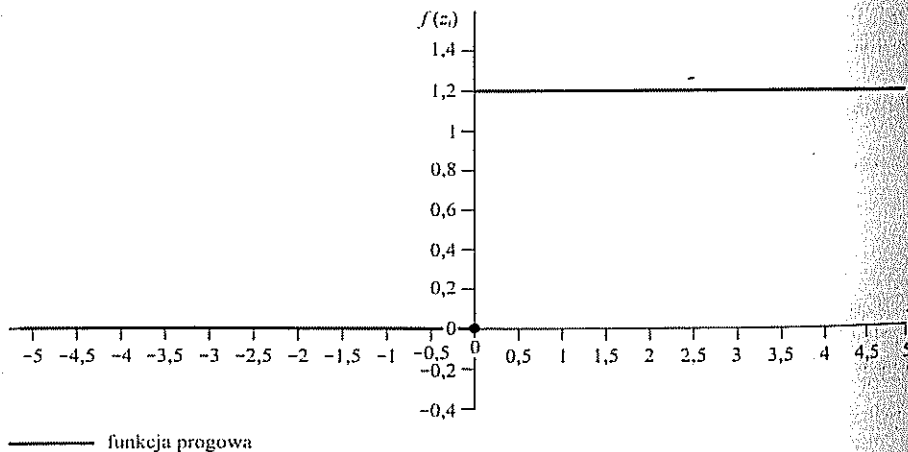
Zagregowany sygnał jest przekształcany zgodnie z przyjętą funkcją aktywacji  $f$ . Otrzymana wielkość staje się wartością sygnału rozpatrywanego neuronu i zostaje przesłana wzdłuż połączeń wychodzących do kolejnych neuronów w sieci.

Funkcja aktywacji może przyjmować różną postać. Najpopularniejsze są:

1) **funkcja progowa** o postaci:

$$y_i = f(z_i) = \begin{cases} 1, & \text{gdy } z_i > 0, \\ 0, & \text{gdy } z_i \leq 0, \end{cases}$$

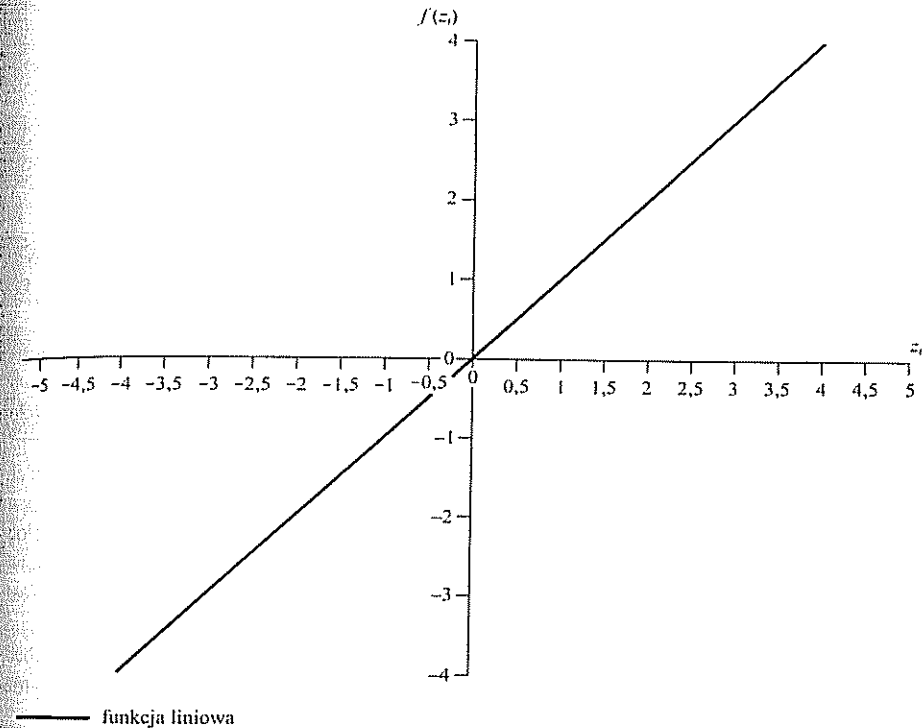
RYSUNEK 14.2  
Wykres progowej funkcji aktywacji



2) **funkcja liniowa** o postaci:

$$y_i = f(z_i) = z_i$$

RYSUNEK 14.3  
Wykres liniowej funkcji aktywacji



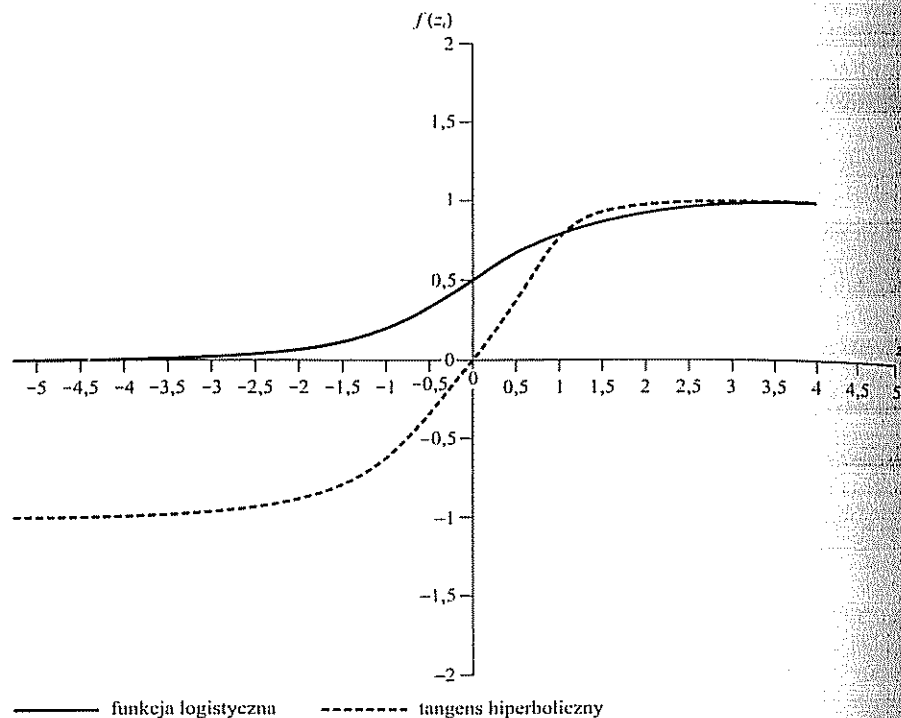
3) **funkcja nieliniowa S-kształtna (sigmoidalna)**, np. logistyczna:

$$y_i = f(z_i) = \frac{1}{1 + e^{-z_i}}$$

lub **tangens hiperboliczny**:

$$y_i = f(z_i) = \frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}}$$

RYSUNEK 14.4  
Wykresy S-kształtnych funkcji aktywacji



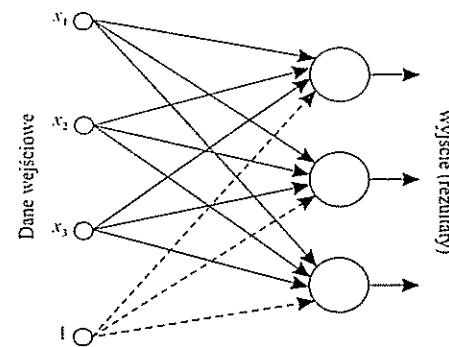
Ze względu na architekturę połączeń występujących między neuronami wyróżnia się następujące typy sieci neuronowych:

- 1) jednowarstwowe;
- 2) wielowarstwowe;
- 3) rekurencyjne.

Podział na warstwy polega na pogrupowaniu neuronów w uporządkowane klasy w taki sposób, że do neuronów w danej klasie prowadzą połączenia wyłącznie od neuronów z wcześniejszych klas (warstw). Takie pogrupowanie jest zawsze możliwe, jeżeli połączenia między neuronami nie mają charakteru cyklicznego. Podział neuronów na warstwy ułatwia komputerową implementację modelu sieci neuronowej.

**Sieć jednowarstwowa**, niemająca połączeń rekurencyjnych, jest najprostszym wariantem architektury sieci neuronowej. Przykładową strukturę takiej sieci przedstawia rysunek 14.5.

RYSUNEK 14.5  
Schemat budowy przykładowej sieci jednowarstwowej



Neurony, w których następuje przetwarzanie sygnału, są na rysunku oznaczone dużymi okręgami i tworzą warstwę neuronów. Komórki oznaczone mniejszymi okręgami również formalnie są neuronami, ale ich rola jest pasywna, gdyż jedynie wprowadzają one sygnały wejściowe do sieci i nie przetwarzają ich (dlatego też nie są traktowane jako warstwa sieci). Szczególnym przypadkiem jest neuron wejściowy oznaczony 1 — jest to tzw. **bias**, neuron pomocniczy, wysyłający stale do połączonych z nim neuronów sygnał o sile 1. Dzięki jego obecności sieć ma większe możliwości dopasowywania swojego działania do oczekiwanych rezultatów. W większości przypadków stosowania sieci neuronowych bias jest uwzględniany w strukturze sieci, często — ze względu na swój pomocniczy charakter — nie będąc nawet uwidacznianym na schemacie (dlatego połączenia z neuronu pomocniczego na rysunku 14.5 oznaczono przerywanymi liniami).

Powyższa sieć neuronowa ma trzy wejścia (oraz bias) oraz trzy neurony wyjściowe. W przedstawionej sieci neurony wejściowe są połączone z neuronami wyjściowymi na zasadzie każdy z każdym — jest to podejście najczęściej stosowane, gdyż trudno z wyprzedzeniem przewidzieć, które połączenia w sieci będą potrzebne dla poprawnego jej działania, a które okażą się zbędne.

Każde połączenie posiada wagę, której wartość jest ustalana w procesie uczenia sieci w taki sposób, aby kolejne przetworzenia sygnałów wejściowych dały w neuronach wyjściowych rezultaty możliwie jak najbliższe pożądanym wynikom. Łatwo zauważyć, że gdybyśmy rozpatrzyli sieć jednowarstwową z neuronami o liniowej funkcji aktywacji, to sygnał na *i*-tym neuronie wyjściowym będzie kombinacją liniową sygnałów wejściowych (z wyrazem wolnym), tj.:

$$y_i = \left( \sum_{j=1}^{n-1} x_j w_{ij} \right) + w_{im} \quad (14.2)$$

gdzie:

$n$  — liczba neuronów wejściowych (z których  $n-1$  to faktyczne sygnały wejściowe, a neuron  $n$  to bias),

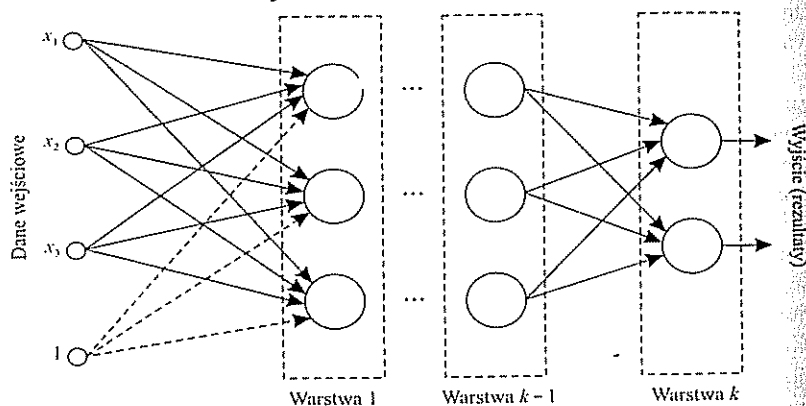
$x_j$  — wartość sygnału na wejściu  $j$ ,

$w_{ij}$  — waga połączenia między neuronem  $j$  a neuronem  $i$ .

Poszukiwanie takich wag  $w_{ij}$ , aby otrzymywane w wyniku działania sieci neuronowych wartości  $y_i$  były jak najbliższe pożądanym rezultatom, przywodzi na myśl analogię do poszukiwania parametrów liniowej funkcji regresji. Naturalnie, wartości parametrów funkcji regresji można skuteczniej wyznaczać za pomocą odpowiednich istniejących metod rachunkowych, niemniej jednak powyższy przykład stanowi dobry punkt wyjścia do analizy potencjalnych możliwości wynikających z budowy sieci bardziej skomplikowanych, w których neurony będą posługiwały się funkcją aktywacji inną niż liniowa oraz będą zorganizowane w więcej niż jedną warstwę.

RYSUNEK 14.6

Schemat budowy sieci wielowarstwowej (o  $k$  warstwach)



Rozpatrzmy sieć  $k$ -warstwową, której ogólną strukturę przedstawiono na rysunku 14.6. W schemacie, oprócz neuronów wejściowych oraz warstwy neuronów wyjściowych, występuje również co najmniej jedna warstwa neuronów ukrytych. Zastosowanie warstwy lub warstw ukrytych oraz nieliniowej (np. S-kształtnej) funkcji aktywacji pozwala sieci lepiej prognozować wartości sygnałów na wyjściu. W szczególności przyjmuje się, że sieć z dwiema warstwami ukrytymi, składającymi się z neuronów o logistycznej funkcji aktywacji, jest w stanie dokonywać regresji nieliniowej odpowiadającej praktycznie dowolnej

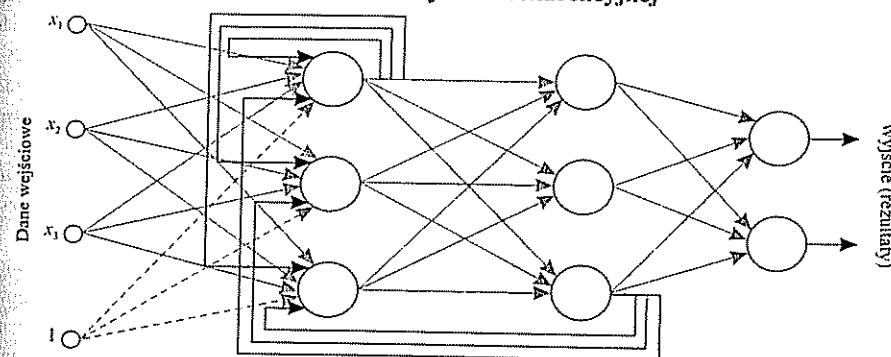
postaci funkcyjnej. O ile w przypadku regresji liniowej używanie sieci neuronowej zamiast metod analitycznych nie miało uzasadnienia, o tyle w przypadku regresji nieliniowej jest inaczej. Klasyczne metody regresji nieliniowej pozwalają wyznaczać optymalne wartości parametrów, ale względem już ustalonej ogólnej postaci funkcji (np. funkcji potęgowej, logarytmicznej). Zastosowanie sieci neuronowej daje przewagę w tym zakresie, że nie jest konieczne określanie z góry postaci ogólnej funkcji regresji.

Wszystkie rozpatrywane powyżej warianty opisywały sieci **jednokierunkowe** (*feedforward*), tzn. takie, gdzie neurony można było uszeregować w kolejności od „wcześniejszego” do „późniejszego”, zgodnie z kierunkiem przebiegu sygnałów w sieci. Inaczej mówiąc, graf, którego wierzchołkami są występujące w sieci neurony, a łukami połączenia między nimi, miał właściwość **acykliczności**. Sieci neuronowe, opisane grafem pozbawionym tej właściwości, są nazywane **sieciami rekurencyjnymi**.

Ideę sieci rekurencyjnej ilustruje rysunek 14.7. Dla większej przejrzystości połączenia niemające rekurencyjnego charakteru zaznaczono kolorem szarym.

RYSUNEK 14.7

Schemat budowy sieci rekurencyjnej



Sieci rekurencyjne stanowią kolejny krok w kierunku budowania architektur coraz wierniej odpowiadających prawdziwemu systemowi nerwowemu, pozwalających rozwiązywać coraz bardziej skomplikowane problemy. Jednocześnie występowanie cyklicznych połączeń między neuronami wprowadza dodatkowe trudności w opisie i analizie sieci. Szczegółowe rozważania w zakresie sieci rekurencyjnych wykraczają poza ramy tego podręcznika i są pominięte. W szczególności podrozdział dotyczący uczenia sieci neuronowych będzie dotyczył jedynie tematyki sieci jednokierunkowych.

## 14.3. Uczenie sieci neuronowej

Tworzenie sieci neuronowej, która ma rozwiązywać zadany problem, sprowadza się do realizacji dwóch głównych etapów, tj.:

- 1) określenia struktury sieci, tzn. liczby neuronów, ewentualnego ich podziału na warstwy, wyboru typu funkcji aktywacji dla poszczególnych neuronów;
- 2) nauczenia sieci działania odpowiedniego dla rozwiązywanego problemu.

Pierwszy z tych etapów wymaga aktywnego i koncepcyjnego udziału użytkownika, a drugi — po określeniu stosownych założeń — jest realizowany automatycznie.

Rozróżnia się dwa typy uczenia sieci: z nauczycielem (*supervised learning*) oraz bez nauczyciela (*unsupervised learning*).

### 14.3.1. Uczenie z nauczycielem

Wagi przyporządkowane połączeniom między poszczególnymi neuronami muszą zostać ustalone w taki sposób, aby zminimalizować błąd na wyjściu sieci (różnicę między sygnałami generowanymi przez sieć a pożądanymi wartościami tych sygnałów). W tym celu użytkownik musi zgromadzić możliwie duży zbiór przykładowych danych wejściowych dla zagadnienia, które sieć ma rozwiązywać, wraz z pożądanymi wartościami wyników (sygnałów na wyjściu sieci). Właściwe wartości wag są następnie wyznaczane przy użyciu odpowiedniego algorytmu na bazie analizy zachowania sieci w poszczególnych przykładowych zadaniach ze zbioru uczącego. Podstawowym algorytmem poszukiwania właściwych wag jest **metoda spadku gradientu**.

Rozpatrzmy najprostszy przypadek, jakim jest sieć jednokierunkowa, jednowarstwowa. Niech  $K$  oznacza liczbę neuronów wejściowych, a  $L$  liczbę neuronów wyjściowych. Każdy neuron wyjściowy jest połączony z każdym neuronem wejściowym, w związku z czym występuje  $K \times L$  połączeń, z których każde ma swoją wagę. Zakładamy, że neurony wyjściowe posługują się liniową funkcją aktywacji, co oznacza, że sygnały na wyjściu są liniową kombinacją sygnałów wejściowych. Dla każdego przypadku ze zbioru uczącego są dane:

$\mathbf{x} = [x_i]_{K \times 1}$  — wektor sygnałów wejściowych,

$\mathbf{d} = [d_i]_{L \times 1}$  — wektor poświadczonych wartości sygnałów wyjściowych.

Oznaczmy macierz wag połączeń przez  $\mathbf{W} = [w_{ij}]_{L \times K}$ . Dla ustalonych wartości wag sygnały na wyjściu z sieci mogą być obliczone według wzoru:

$$y_i = \sum_{j=1}^K x_j w_{ij}, \quad \text{dla każdego } i = 1, 2, \dots, L. \quad (14.3)$$

Alternatywnie, można zdefiniować wektor sygnałów wyjściowych  $\mathbf{y} = [y_i]_{L \times 1}$ . Wtedy w zapisie wektorowym:

$$\mathbf{y} = \mathbf{W}\mathbf{x}. \quad (14.4)$$

Niech  $E$  oznacza funkcję błędu sygnałów wyjściowych, której wartość jest obliczana jako połowa<sup>3</sup> sumy kwadratów różnic między pożądanymi a otrzymanymi wartościami sygnałów wyjściowych:

$$E = \frac{1}{2} \sum_{i=1}^L (d_i - y_i)^2. \quad (14.5)$$

Mając zdefiniowaną funkcję błędu sieci neuronowej, można — przy użyciu pochodnych — zbadać wpływ wag poszczególnych połączeń na wartość błędu całej sieci. Łatwo wyprowadzić, że:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}} = -(d_i - y_i)x_j, \quad \text{dla } i = 1, 2, \dots, L \text{ oraz } j = 1, 2, \dots, K. \quad (14.6)$$

Wartość obliczonej pochodnej stanowi oszacowanie wzrostu funkcji błędu w odpowiedzi na jednostkowy wzrost wartości wagi  $w_{ij}$ . Aby poprawić wartość błędu sieci neuronowej, należałoby do obecnej wartości  $w_{ij}$  dodać współczynnik korekty dany wzorem:

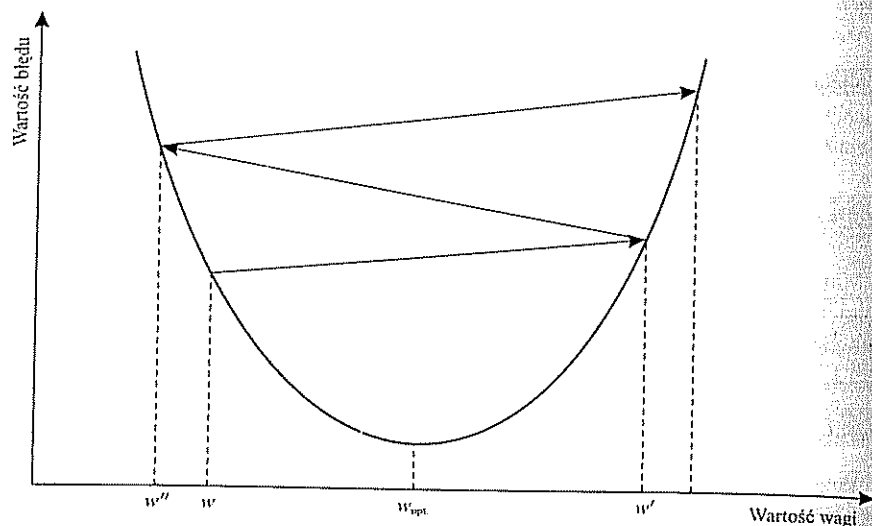
$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta (d_i - y_i)x_j, \quad (14.7)$$

gdzie  $\eta$  jest współczynnikiem uczenia określającym tempo modyfikowania wag połączeń.

Dochodzenie do optymalnej wartości wagi połączenia w metodzie spadku gradientu polega na krokowym zwiększaniu lub zmniejszaniu wagi w kierunku wskazywanym przez nachylenie krzywej błędu w danym punkcie (a więc wartość pochodnej funkcji błędu  $E$  względem wagi połączenia). **Współczynnik uczenia** decyduje o długości kroku w kierunku oczekiwanej poprawy i jest istotnym czynnikiem wpływającym na skuteczność uczenia sieci. Zbyt mała wartość współczynnika uczenia będzie skutkować bardzo wolnym dochodzeniem do optymalnych wartości wag, natomiast zbyt duża jego wartość może powodować brak zbieżności wag do optymalnych wartości. Ideę drugiego ze wspomnianych problemów przedstawiono na rysunku 14.8.

<sup>3</sup> Współczynnik  $1/2$  jest wprowadzany, aby uzyskać prostszą postać pochodnej funkcji błędu, co będzie pomocne w dalszych krokach.

RYSUNEK 14.8  
Ilustracja konsekwencji zbyt wysokiej wartości współczynnika uczenia



W przykładowej sytuacji przedstawionej na rysunku 14.8 początkowa wartość wagi wynosiła  $w$  i — zgodnie z nachyleniem krzywej błędu — musiała zostać zwiększona, aby dotrzeć do minimum funkcji błędu. Z powodu zbyt dużej długości kroku nowa wartość wagi wyniosła  $w'$  i musiała zostać w dalszej kolejności zmniejszona, aby uzyskać spadek błędu. Zbyt długi krok spowodował, że waga została jednak zmniejszona do poziomu  $w''$ , konsekwentnie oddalając się od optymalnego poziomu, zamiast się do niego zbliżać. Dobór takiej wartości współczynnika uczenia, która nie będzie ani zbyt mała, ani zbyt duża, jest uzależniony od konkretnego problemu i jest najczęściej dokonywany drogą eksperymentalną.

Całościowy schemat uczenia jednokierunkowej i jednowarstwowej sieci neuronowej można opisać za pomocą algorytmu.

#### Algorytm 14.1

Krok 1. Przyjmij losowe wartości początkowe wag połączeń.

Krok 2. Wyzeruj współczynniki korekty wag dla wszystkich połączeń:

$$\Delta w_{ij} := 0, \text{ dla } i = 1, 2, \dots, L \text{ oraz } j = 1, 2, \dots, K.$$

Krok 3. Dla każdego zestawu danych wejściowych (dla każdej pary  $x, d$ ):

- używając bieżących wag połączeń w sieci, oblicz wartości sygnałów wyjściowych  $y_1, y_2, \dots, y_L$  otrzymane po wprowadzeniu wartości wektora  $x$  jako sygnałów wejściowych do sieci neuronowej.
- dla każdego połączenia  $\langle i, j \rangle$  powiększ wartość  $\Delta w_{ij}$  o wielkość korekty wynikającej z błędu sieci neuronowej w rozpatrywanym zestawie danych wejściowych:

$$\Delta w_{ij} := \Delta w_{ij} + \eta(d_i - y_i)x_j.$$

Krok 4. Dla każdego połączenia  $\langle i, j \rangle$  wprowadź korektę wagi:

$$w_{ij} := w_{ij} + \Delta w_{ij}.$$

Krok 5. Wróć do kroku 2.

Pojedyncze wykonanie kroków 2–4 jest nazywane **epoką** (*epoch*). Proces uczenia trwa wiele epok i jest przerywany, gdy błędy popełniane przez sieć przestają ulegać zauważalnej poprawie lub gdy występują oznaki przeuczenia sieci<sup>4</sup>.

#### PRZYKŁAD 14.1

Przeanalizujemy przebieg uczenia prostej sieci jednowarstwowej przyjmującej dwa sygnały wejściowe (oraz bias) i mającej jedno wyjście. Dane są cztery przykładowe zestawy wartości sygnałów wejściowych wraz z odpowiadającymi im pożądanymi wartościami sygnału wyjściowego (tablica 14.1).

TABLICA 14.1  
Dane dla procesu uczenia  
sieci neuronowej

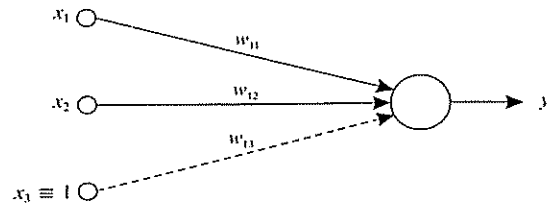
| Zestaw | Wejście |       | Wyjście |
|--------|---------|-------|---------|
|        | $x_1$   | $x_2$ | $d$     |
| 1      | 2       | 2     | 7       |
| 2      | 1       | 4     | 7       |
| 3      | -4      | 5     | -11     |
| 4      | 0       | 2     | -1      |

Zgodnie z algorytmem 14.1 wyznaczamy losowo początkowe wartości wag. Przyjmijmy, że otrzymane wartości to:  $w_{11} = 0,399$ ,  $w_{12} = 0,922$ ,  $w_{13} = 0,013$ .

<sup>4</sup> O tym zjawisku będzie mowa w dalszej części rozdziału.



RYSUNEK 14.9  
Schemat sieci neuronowej do przykładu 14.1



Dla przyjętych wartości wag połączeń obliczony — zgodnie ze wzorem (14.3) — sygnał na wyjściu sieci ( $y$ ) oraz obliczona — zgodnie ze wzorem (14.5) — wielkość błędu neuronu wyjściowego ( $E$ ) są zawarte w tabelicy 14.2.

TABLICA 14.2  
Rezultaty generowane przez sieć neuronową przy początkowym układzie wag

| Zestaw danych wejściowych | Sygnał wyjściowy $y$ | Błąd neuronu wyjściowego $E$ |
|---------------------------|----------------------|------------------------------|
| 1                         | 2,655                | 9,436                        |
| 2                         | 4,100                | 4,205                        |
| 3                         | 3,027                | 98,378                       |
| 4                         | 1,857                | 4,081                        |
| Przeciętna wartość błędu  |                      | 29,026                       |

Wielkości korekt, obliczone zgodnie ze wzorem (14.7) dla każdej wagi oraz z uwzględnieniem każdego z zestawów danych uczących, są zapisane w tabelicy 14.3. Na potrzeby obliczeń przyjęto wartość współczynnika uczenia na poziomie  $\eta = 0,01$ .

TABLICA 14.3  
Obliczenia korekt wag sieci neuronowej

| Waga     | Wartość początkowa | Wielkość korekty |          |          |          |          | Wartość skorygowana |
|----------|--------------------|------------------|----------|----------|----------|----------|---------------------|
|          |                    | zestaw 1         | zestaw 2 | zestaw 3 | zestaw 4 | łącznie  |                     |
| $w_{11}$ | 0,399              | 0,0869           | 0,029    | 0,56108  | 0        | 0,67698  | 1,07598             |
| $w_{12}$ | 0,922              | 0,0869           | 0,116    | -0,70135 | -0,05714 | -0,55559 | 0,36641             |
| $w_{13}$ | 0,013              | 0,04345          | 0,029    | -0,14027 | -0,02857 | -0,09639 | -0,08339            |

Tabela 14.4 zawiera obliczenia nowych wartości sygnałów wyjściowych sieci dla poszczególnych przypadków zestawów uczących, uwzględniając nowe wartości wag połączeń. Można zauważyć, że przeciętna wielkość błędu neuronu wyjściowego uległa znaczącemu zmniejszeniu w porównaniu z przedstawioną w tabelicy 14.2.

TABLICA 14.4  
Rezultaty generowane przez sieć neuronową po pierwszej korekcie wag

| Zestaw danych wejściowych | Sygnał wyjściowy $y$ | Błąd neuronu wyjściowego $E$ |
|---------------------------|----------------------|------------------------------|
| 1                         | 2,801                | 8,815801                     |
| 2                         | 2,457                | 10,31942                     |
| 3                         | -2,557               | 35,64212                     |
| 4                         | 0,649                | 1,359601                     |
| Przeciętna wartość błędu  |                      | 14,03424                     |

W rezultacie wielokrotnego powtarzania kroków 2–5 algorytmu wagi połączeń będą stopniowo korygowane w kierunku pozwalającym na redukcję błędu otrzymywanych sygnałów wyjściowych.

Algorytm 14.1 pozwala korygować wagi jedynie połączeń prowadzących do neuronów wyjściowych sieci, gdyż tylko dla nich są znane pożądane stany sygnałów, przez co możliwe jest porównanie bieżącego działania połączeń ze stanem, jaki należy uzyskać. W związku z tym metoda ta jest wystarczająca jedynie dla sieci jednowarstwowych. Uogólnieniem przedstawionego algorytmu na przypadku sieci, które mogą zawierać również ukryte warstwy neuronów, jest **metoda wstecznej propagacji błędów** (*backpropagation*).

Przyjmijmy, że neurony w sieci są ponumerowane liczbami naturalnymi. Wprowadźmy oznaczenia:

- $P^-(i)$  — zbiór neuronów poprzedzających neuron  $i$ , tzn. takich, z których sygnał jest przesyłany do neuronu  $i$ ,
- $P^+(i)$  — zbiór neuronów następujących po neuronie  $i$ , tzn. takich, do których jest wysyłany sygnał z neuronu  $i$ ,
- $z_i$  — zagregowany sygnał wchodzący do neuronu  $i$ ,
- $y_i$  — sygnał wychodzący z neuronu  $i$ ,
- $f(\cdot)$  — funkcja aktywacji neuronu  $i$ .

Dla neuronów wejściowych, czyli takich neuronów  $i$ , że  $P^-(i) = \emptyset$ , sygnał wchodzący do neuronu jest sygnałem wejściowym sieci. W pozostałych przypad-

kach, czyli dla neuronów należących do warstwy ukrytej lub wyjściowej, wartość  $z_i$  jest obliczana jako ważona suma sygnałów od neuronów je poprzedzających:

$$z_i = \sum_{j \in P^*(i)} y_j w_{ij}. \quad (14.8)$$

Sygnał wychodzący z neuronu stanowi przekształcenie sygnału wchodzącego dokonane za pomocą funkcji aktywacji, tj.:

$$y_i = f_i(z_i). \quad (14.9)$$

Dla neuronów wejściowych funkcja aktywacji jest zawsze liniowa ( $f_i(x) = x$ ), a dla pozostałych neuronów może mieć dowolną postać, np. logistyczną.

Rozpatrzmy teraz sposób uczenia sieci wielowarstwowej poprzez korygowanie wag w sposób analogiczny do przedstawionego wcześniej dla sieci jednowarstwowej.

Podobnie jak poprzednio, chcemy dla każdego połączenia w sieci wyznaczyć wartość pochodnej funkcji błędu sieci neuronowej względem wagi połączenia. Dla połączenia prowadzącego od neuronu  $j$  do neuronu  $i$  pochodna ma postać:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial z_i} \frac{\partial z_i}{\partial w_{ij}} = \frac{\partial E}{\partial z_i} y_j. \quad (14.10)$$

Łatwo zauważyć, że czynnik  $\frac{\partial E}{\partial z_i}$  jest wspólny dla pochodnych względem wag każdego z połączeń prowadzących do  $i$ -go neuronu. Czynnik ten jest nazywany **sygnałem błędu neuronu  $i$** . Jest on oznaczany symbolem  $\delta_i$ :

$$\delta_i = \frac{\partial E}{\partial z_i} = \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial z_i} = \frac{\partial E}{\partial y_i} f'_i(z_i). \quad (14.11)$$

Wartość  $\delta_i$  łatwo wyznaczyć, jeśli  $i$  jest neuronem wyjściowym w sieci. Wtedy:

$$\delta_i = \frac{\partial E}{\partial z_i} = \frac{\partial E}{\partial y_i} f'_i(z_i) = -(d_i - y_i) f'_i(z_i). \quad (14.12)$$

Jeżeli  $i$  należy do warstwy ukrytej, to  $\delta_i$  wyznacza się na podstawie wartości sygnałów błędu wszystkich następników neuronu  $i$ :

$$\delta_i = \frac{\partial E}{\partial z_i} = \sum_{k \in P^*(i)} \delta_k w_{ki} f'_i(z_i). \quad (14.13)$$

Wzór (14.13) objaśnia genezę nazwy metody wstecznej propagacji błędu — wartości sygnału błędu poszczególnych neuronów są przekazywane od warstwy wyjściowej do kolejnych warstw poprzedzających.

Dla stosowania wzorów (14.11)–(14.13) konieczna jest znajomość pochodnej funkcji aktywacji. W szczególności dla funkcji progowej pochodna nie istnieje w punkcie  $z_i = 0$ . Problem ten nie występuje w przypadku pozostałych wprowadzanych we wstępie typach funkcji, a funkcja progowa może być z powodzeniem aproksymowana przy użyciu funkcji  $S$ -kształtnych. Pochodne najczęściej stosowanych funkcji zawiera poniższe zestawienie:

1) dla funkcji liniowej:

$$f'(z_k) = 1; \quad (14.14)$$

2) dla funkcji logistycznej:

$$f'(z_k) = f(z_k)(1 - f(z_k)) = y_k(1 - y_k); \quad (14.15)$$

3) dla funkcji tangensa hiperbolicznego:

$$f'(z_k) = 1 - f(z_k)^2 = 1 - y_k^2. \quad (14.16)$$

Uogólnioną procedurę poprawy wag połączeń można sformułować w postaci algorytmu 14.2.

#### Algorytm 14.2

Krok 1. Przyjmij losowe wartości początkowe wag połączeń.

Krok 2. Wyzeruj współczynniki korekty wag dla wszystkich połączeń:

$$\forall_{i,j: j \in P^*(i)} \Delta w_{ij} := 0.$$

Krok 3. Dla każdego zestawu danych wejściowych (dla każdej pary  $x, d$ ):

- używając bieżących wag połączeń w sieci, oblicz wartości sygnałów wyjściowych otrzymane po wprowadzeniu wartości wektora  $x$  jako sygnałów wejściowych do sieci neuronowej,
- dla każdego neuronu wyjściowego oblicz wartość sygnału błędu ( $\delta$ ) za pomocą wzoru (14.12), następnie dla neuronów w warstwach ukrytych, w kolejności od ostatniej do pierwszej, wyznacz wartości sygnału błędu, korzystając z wzoru (14.13),
- dla każdego połączenia  $\langle i, j \rangle$  powiększ wartość  $\Delta w_{ij}$  o wielkość korekty wynikającej z błędu sieci neuronowej w rozpatrywanym zestawie danych wejściowych:

$$\Delta w_{ij} := \Delta w_{ij} - \eta \delta_i y_j.$$

Krok 4. Dla każdego połączenia  $\langle i, j \rangle$  dokonaj korekty wagi:

$$w_{ij} := w_{ij} + \Delta w_{ij}.$$

Krok 5. Wróć do kroku 2.

### 14.3.2. Uczenie bez nauczyciela

W odróżnieniu od omawianej powyżej sytuacji, gdzie sieć neuronowa była uczona właściwego zachowania poprzez analizę przykładowych danych wejściowych do problemu w powiązaniu z odpowiadającymi im oczekiwanymi rezultatami, uczenie bez nauczyciela opiera się jedynie na przetwarzaniu przykładowych danych wejściowych bez wskazywania, jaka powinna być prawidłowa odpowiedź w danej sytuacji. Celem działania tak budowanej sieci jest rozpoznanie struktury danych, w szczególności dokonanie segmentacji obiektów podawanych na wejściu do sieci. Problem ten różni się od klasycznego zadania klasyfikacji, gdzie celem sieci neuronowej było nauczyć się określania z możliwie wysoką skutecznością, do której ze znanych kategorii dany obiekt — uwzględniając poszczególne jego cechy — należy, przy czym zbiór uczący zawierał informacje o poprawnej kategorii obiektu. W zagadnieniu segmentacji sieć musi sama wykryć, na jakie kategorie i w jaki sposób podzielić obiekty podawane na wejściu. Samouczące sieci są również wykorzystywane np. w zagadnieniach kompresji obrazów.

Najbardziej znanym wariantem sieci uczonej bez nauczyciela są **samoorganizujące mapy cech** (*self organising feature map*, SOFM), nazywane też od nazwiska twórcy **sieciami Kohonena**.

Poprzestajemy na zaprezentowaniu jedynie rysu problematyki uczenia bez nauczyciela, gdyż szczegółowe rozważanie tego tematu wykracza poza ramy niniejszego podręcznika.

## 14.4. Praktyczne aspekty stosowania sieci neuronowych

### 14.4.1. Dobór odpowiedniej architektury sieci

Budowa sieci, która będzie sprawnie rozwiązywała postawiony problem, nie jest zadaniem łatwym i niejednokrotnie wymaga podejścia eksperymentalnego, wspartego rozpoznaniem sugestii zawartych w literaturze omawiającej szczegółowo problematykę sieci neuronowych w kontekście poszczególnych zagadnień, np. prognozowania kursów akcji czy rozpoznawania obrazów. Dodatkowym problemem, nawet przy ustalonej już ogólnej architekturze sieci, będzie dobór właściwej liczby neuronów w poszczególnych warstwach. Zbyt mała liczba neuronów w stosunku do stopnia komplikacji rozpatrywanego zagadnienia może okazać się niewystarczająca do właściwego nauczenia sieci poprawnych zachowań. Z kolei, zbyt dużo neuronów może doprowadzić do wystąpienia zjawiska **przeuczenia sieci neuronowej**, tj. sytuacji, gdzie sieć, zamiast wykryć i nauczyć się ogólnych reguł

łączących dane wejściowe z poprawnymi rezultatami, „zapamiętuje” wartości oczekiwanych rezultatów zadań ze zbioru uczącego, dzięki czemu radzi sobie doskonale z przykładowymi zestawami danych wejściowych, ale może okazać się zupełnie nieskuteczna w rozwiązywaniu nowego zadania spoza poznanego zestawu. Można wyróżnić dwa typowe podejścia do problemu ustalania liczby neuronów:

- 1) rozpocząć od sieci ze stosunkowo małą liczbą neuronów i stopniowo tę liczbę zwiększać tak długo, jak długo powoduje to poprawę dokładności wyników podawanych przez sieć neuronową;
- 2) rozpocząć od sieci z dużą liczbą neuronów i stopniowo je usuwać aż zacznie to powodować wyraźne pogarszanie się jakości wyników podawanych przez sieć.

Aby poprawnie ocenić jakość rezultatów generowanych przez sieć, w szczególności w kontekście zjawiska przeuczenia sieci, należy stosować podział zestawów przykładowych danych wejściowych na zestaw uczący i zestaw walidacyjny (sprawdzający).

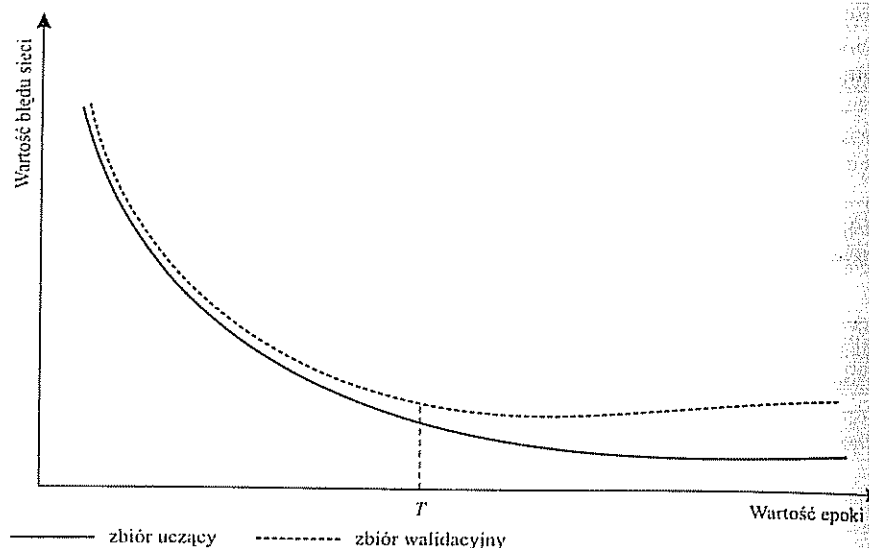
### 14.4.2. Podział przykładowych zestawów danych na zestaw uczący i zestaw walidacyjny

Sieć neuronowa uczy się, analizując odpowiednią liczbę przykładowych zestawów, składających się z opisu zadania w postaci wartości sygnałów wejściowych oraz określenia oczekiwanego wyniku w postaci poprawnych wartości sygnałów na wyjściu. Sieć neuronowa w procesie uczenia dostosowuje swoją budowę tak, aby na danym zbiorze przykładowych zadań przetwarzać sygnały wejściowe w rezultaty jak najbliższe pożądanym wartościom sygnałów wyjściowych. Zakładamy, że sieć, która nauczy się rozwiązywać zadania z przykładowych zestawów, będzie potrafiła uogólnić zauważone zależności i w konsekwencji rozwiązywać problemy tego samego typu, nie uwzględnione w zbiorze uczącym. W celu zapewnienia odpowiedniego poziomu ogólności działania sieci neuronowej wskazany jest podział dostępnych zestawów przykładowych zadań na zbiór uczący i zbiór walidacyjny. Sieć neuronowa jest uczona wyłącznie na bazie zestawów danych ze zbioru uczącego, ale na bieżąco oceniana jest także skuteczność działania sieci w zadaniach ze zbioru walidacyjnego. Jeżeli zachowania wyuczone przez sieć mają charakter ogólny dla danego problemu, czyli nie występuje zjawisko przeuczenia sieci, to wielkości błędów popełnianych w zadaniach z obu zbiorów powinny być zbliżone. Rysunek 14.10 ilustruje ideę kontrolowania jakości wyników w zbiorze uczącym i zbiorze walidacyjnym po kolejnych epokach procesu uczenia sieci.

Analizując przebieg krzywych na rysunku 14.10, można wysnuć przypuszczenie, że w początkowych  $T$  epokach sieć uogólniała zjawiska widoczne w zbiorze uczącym, przez co jakość wyników generowanych w zbiorze walidacyjnym (który

RYSUNEK 14.10

Ilustracja porównania skuteczności sieci neuronowej w zbiorze uczącym i w zbiorze walidacyjnym



nie brał udziału w uczeniu) również się polepszała i była zbliżona do jakości wyników ze zbioru uczącego. W kolejnych epokach sieć zaczęła jednak coraz bardziej dopasowywać się do konkretnych przypadków zawartych w zbiorze uczącym, przez co coraz bardziej traciła umiejętność poprawnego uogólniania wiedzy. Proces uczenia powinien być zatem przerwany po  $T$  epokach.

### 14.4.3. Odpowiedni dobór struktury zbioru uczącego

Zbiór uczący powinien zawierać przykłady zadań reprezentatywne dla całego spektrum przypadków, jakie mogą wystąpić w rzeczywistości. Oznacza to w szczególności, że:

1. Wzorcowe dane powinny zawierać przypadki pokrywające wszystkie możliwe typy sytuacji potencjalnie występujących w rozpatrywanym problemie. Przykładowo, jeśli rozpatrywanym zagadnieniem jest prognozowanie upadłości przedsiębiorstw, to zbiór uczący powinien zawierać przypadki firm:
  - różnej wielkości (małych, średnich i dużych),
  - o różnym charakterze własnościowym (państwowych i prywatnych),

- o różnej formie prawnej (spółki akcyjne, spółki z ograniczoną odpowiedzialnością, spółki cywilne itp.),
  - z różnych branż,
  - itp.
2. Rozkład wariantów przypadków w zbiorze uczącym powinien być zbliżony do rozkładu wariantów w zbiorze, na którym sieć będzie używana. Załóżmy, że w całej gospodarce przeciętnie 10% przedsiębiorstw jest zagrożonych upadłością w ciągu najbliższego roku. Przyjmijmy, że zbiór uczący skonstruowano w taki sposób, że połowę jego składu stanowią przypadki zagrożone upadłością, a połowę niezagrożone. Sieć neuronowa nauczona na takim zbiorze będzie miała tendencję do zbyt łatwego klasyfikowania przedsiębiorstw jako zagrożonych. Przykładowo, gdyby sieć mylnie klasyfikowała co dziesiąte dobre przedsiębiorstwo jako zagrożone, wtedy w zbiorze uczącym błąd dotyczyłby jedynie 5% przypadków (co dziesiąte przedsiębiorstwo spośród niezagrożonych). Tymczasem w pełnej populacji przedsiębiorstw, gdzie firmy niezagrożone stanowią 90%, sklasyfikowanie co dziesiątej z nich jako zagrożonej upadłością powoduje błędne wskazanie w 9% wszystkich przypadków.

### POLECENIA I PYTANIA KONTROLNE

1. Omów budowę sztucznego neuronu.
2. W jaki sposób należy dobierać liczbę warstw sieci oraz liczbę neuronów w poszczególnych warstwach?
3. Objasnij wpływ wartości współczynnika uczenia na efektywność uczenia sieci neuronowej.
4. Dlaczego metoda wstecznej propagacji błędów nie może być zastosowana w przypadku sieci rekurencyjnych?
5. Na czym polega zjawisko przeuczenia sieci i jak można się przed nim bronić?
6. Dlaczego dane wejściowe i wyjściowe ze zbioru walidacyjnego nie powinny brać udziału w uczeniu sieci?

### BIBLIOGRAFIA

- Rutkowska, D., Piliński M., Rutkowski L., 1999, *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*, Wydawnictwo Naukowe PWN, Warszawa, Łódź.
- StatSoft, 2007, *Elektroniczny Podręcznik Statystyki PL*, Kraków, <http://www.statsoft.pl/text-book/stathome.html> (stan na 01.01.2007).
- Tadeusiewicz, R., 1993, *Sieci neuronowe*, Akademicka Oficyna Wydawnicza, Warszawa.
- Witkowska, D., 2002, *Sztuczne sieci neuronowe i metody statystyczne: wybrane zagadnienia finansowe*, C.H. Beck, Warszawa.

## ROZDZIAŁ 15 METODY HEURYSTYCZNE



### 15.1. Wprowadzenie

Większość omówionych dotychczas algorytmów rozwiązywania zadań z jednym kryterium (a więc metody omówione w rozdziałach 1–3 i 5–13), chociaż dotyczy różnych zagadnień, ma jedną wspólną cechę. Prawie każda z opisanych metod<sup>1</sup> pozwala znaleźć w skończonej liczbie kroków rozwiązanie optymalne odpowiedniego zagadnienia.

W praktyce często nie jest konieczne wyznaczanie optimum. Korzyści z wyznaczenia dokładnego rozwiązania są czasem mniejsze niż straty związane z czasem jego otrzymania. Problem ten uwidacznia się szczególnie w przypadku zadań optymalizacji dyskretnej, takich jak omawiane już wcześniej zagadnienie komiwojażera. Rozwiązanie nawet stosunkowo niewielkiego zadania tego typu na szybkim komputerze może zająć kilka godzin, a nawet dni.

Z tego względu niekiedy pożądane jest użycie metody pozwalającej wyznaczyć jedynie pewne przybliżenie rozwiązania optymalnego<sup>2</sup>, za to w krótkim czasie. Takie przybliżone algorytmy są nazywane **metodami heurystycznymi** lub po prostu **heurystykami**.

W tym rozdziale, ze względu na złożoność tematu, przedstawione są jedynie wybrane metody heurystyczne. Aby zapewnić przejrzystość i ułatwić Czytelnikowi uchwycenie najistotniejszych różnic między nimi, ich działanie zostało omówione na przykładzie jednego problemu — zagadnienia komiwojażera. Przypomnijmy, że problem ten polega na znalezieniu najkrótszego cyklu łączącego wszystkie podane

<sup>1</sup> Wyjątek stanowią np. metody programowania nieliniowego niewyznaczające przeważnie dokładnego optimum. Jednak także w ich przypadku możliwe jest otrzymanie rozwiązania przybliżonego, cechującego się z góry określoną dokładnością.

<sup>2</sup> Przykładem takiej metody jest algorytm skracania sieci czynności opisany w podrozdziale 7.4 (rysunek 7.11).

punkty (np. miejscowości), przy czym są znane odległości między poszczególnymi parami punktów<sup>3</sup>.

W kolejnych podrozdziałach są przedstawione pokrótce: heurystyki lokalnych poszukiwań, algorytmy mrówkowe i algorytmy genetyczne.

### 15.2. Heurystyki lokalnych poszukiwań

**Heurystyki lokalnych poszukiwań** to metody, które nie wykorzystują globalnej struktury problemu, bazując jedynie na lokalnych właściwościach zbioru rozwiązań dopuszczalnych.

W tym rozdziale są omówione dwie heurystyki lokalnych poszukiwań. Pierwsza z nich polega na stopniowym powiększaniu cyklu przez dołączanie do niego kolejnych wierzchołków. Ideą drugiej jest modyfikowanie kompletnych cykli.

#### 15.2.1. Heurystyka włączania

Idea działania **heurystyki włączania** jest następująca. Mając dane częściowe rozwiązanie (częściowy cykl), dołącza się kolejny wierzchołek. Dołączanie należy przy tym podzielić na dwa etapy. Najpierw następuje wybór kolejnego wierzchołka, następnie zaś ustalenie miejsca, w którym ma on być dołączony do cyklu. Obie czynności są powtarzane aż do otrzymania kompletnego cyklu. Heurystykę włączania można więc zapisać tak, jak pokazuje algorytm 15.1.

##### Algorytm 15.1. Heurystyka włączania

- Krok 1. Wybierz pierwszy wierzchołek.
- Krok 2. Jeżeli wszystkie wierzchołki zostały już włączone — stop. W przeciwnym razie przejdź do kroku 3.
- Krok 3. Wybierz wierzchołek spośród jeszcze nie włączonych, stosując wybraną metodę.
- Krok 4. Wybierz miejsce włączenia nowego wierzchołka w taki sposób, aby nowo powstały cykl był jak najkrótszy.

Istnieją różne sposoby wyboru wierzchołka włączanego do cyklu. Trzy z nich to wybór najbliższego sąsiada, wybór najdalszego sąsiada i wybór losowego sąsiada<sup>4</sup>.

<sup>3</sup> Dokładniej zagadnienie to zostało opisane w rozdziale 8 (podrozdziały 8.5 i 8.8).

<sup>4</sup> Niekiedy metodę włączania nazywa się metodą najbliższego sąsiada albo metodą najdalszego sąsiada, w zależności od zastosowanego sposobu wyboru wierzchołka włączanego do cyklu.

W przypadku reguły najbliższego sąsiada jest wybierany wierzchołek położony najbliżej cyklu. Odległość od cyklu jest liczona przy tym jako minimum z odległości między nowym wierzchołkiem a wierzchołkami już należącymi do cyklu. W przypadku wyboru reguły najdalszego sąsiada jest wybierany wierzchołek położony najdalej od cyklu, a w przypadku reguły losowego sąsiada nowy wierzchołek jest losowany. Prawdopodobieństwa wyboru są przy tym ustalane najczęściej albo jako jednakowe dla wszystkich wierzchołków, albo jako odwrotnie proporcjonalne do ich odległości od cyklu. Doświadczenia pokazały, że najskuteczniejszy jest wybór najdalszego sąsiada. Na przykładzie 15.1 pokazano działanie tej właśnie reguły.

### PRZYKŁAD 15.1

Rozwiązać zadanie komiwojażera opisane macierzą odległości przedstawioną w tablicy 15.1, używając metody włączania z wyborem najdalszego sąsiada.

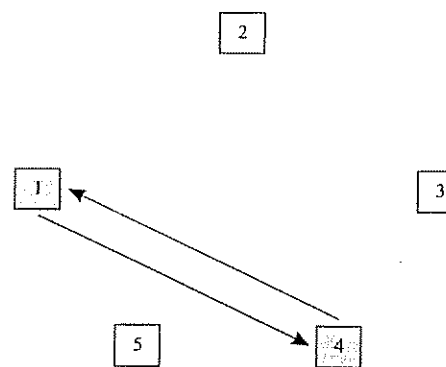
TABLICA 15.1  
Macierz odległości

|   | 1  | 2  | 3  | 4  | 5  |
|---|----|----|----|----|----|
| 1 | —  | 31 | 26 | 37 | 3  |
| 2 | 33 | —  | 6  | 36 | 28 |
| 3 | 22 | 11 | —  | 5  | 28 |
| 4 | 27 | 8  | 5  | —  | 35 |
| 5 | 6  | 27 | 37 | 17 | —  |

Konstrukcję rozwiązania rozpoczniemy od wierzchołka 1. Jak widać, odległości z tego wierzchołka do pozostałych czterech wynoszą, odpowiednio, 31, 26, 37 i 3. Wynika z tego, że najdalszym sąsiadem jest wierzchołek 4, a więc to on zostanie włączony do cyklu. Jego dołączenie jest przy tym jednoznaczne, a więc po pierwszej iteracji cykl częściowy wygląda jak na rysunku 15.1, a jego długość wynosi:  $L = c_{14} + c_{41} = 37 + 27 = 64$ .

W kolejnej iteracji rozpatrujemy odległość jeszcze nie dołączonych wierzchołków od wierzchołków 1 i 4 (w zredukowanej macierzy w tablicy 15.2 odległości te zostały wytłuszczone).

RYSUNEK 15.1  
Postać cyklu po pierwszej iteracji



Jak widać, odległość od cyklu równa jest  $\min\{31, 8\} = 8$  w przypadku wierzchołka 2,  $\min\{26, 5\} = 5$  w przypadku wierzchołka 3 i  $\min\{3, 35\} = 3$  w przypadku wierzchołka 5. Wynika stąd, że do cyklu należy włączyć wierzchołek 2.

TABLICA 15.2  
Zredukowana macierz odległości

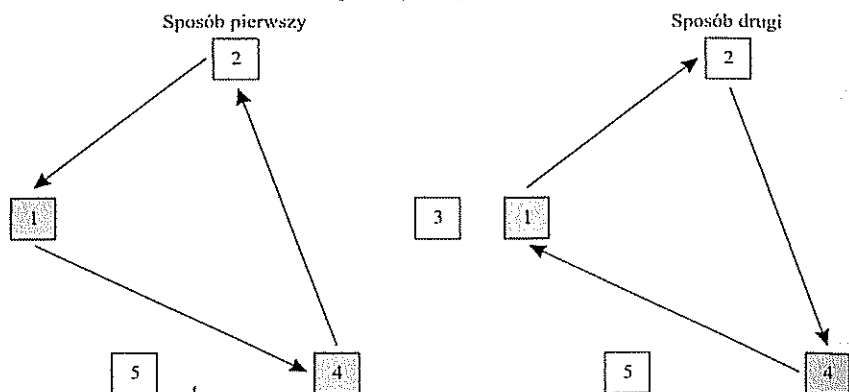
|   | 1 | 2  | 3  | 4 | 5  |
|---|---|----|----|---|----|
| 1 | — | 31 | 26 | — | 3  |
| 2 | — | —  | 6  | — | 28 |
| 3 | — | 11 | —  | — | 28 |
| 4 | — | 8  | 5  | — | 35 |
| 5 | — | 27 | 37 | — | —  |

Wierzchołek ten może być włączony w jeden z dwóch sposobów pokazanych na rysunku 15.2.

Jeżeli wierzchołek zostanie włączony sposobem pierwszym, długość tak utworzonego cyklu wyniesie:  $L = c_{14} + c_{42} + c_{21} = 37 + 8 + 33 = 78$ . W drugim przypadku długość nowego cyklu będzie równa:  $L = c_{12} + c_{24} + c_{41} = 31 + 36 + 27 = 94$ . Należy wybrać pierwszy sposób.

RYSUNEK 15.2

## Dwa sposoby włączenia wierzchołka 2



W kolejnej iteracji ponownie posługujemy się zredukowaną macierzą odległości, jak w tabelicy 15.3. Jak widać, odległości wierzchołków od cyklu są równe  $\min(26, 6, 5) = 5$  w przypadku wierzchołka 3 i  $\min(3, 28, 35) = 3$  w przypadku wierzchołka 5. Wynika stąd, że do cyklu należy włączyć wierzchołek 3.

TABLICA 15.3  
Zredukowana macierz odległości

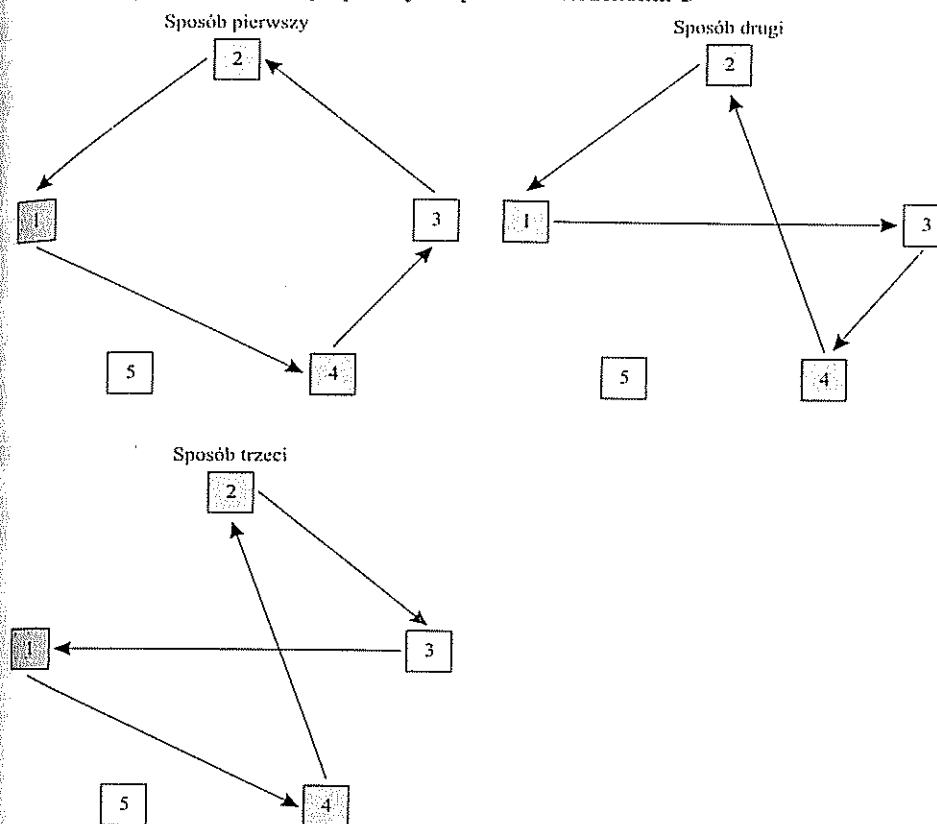
|   | 1 | 2 | 3  | 4 | 5  |
|---|---|---|----|---|----|
| 1 | — | — | 26 | — | 3  |
| 2 | — | — | 6  | — | 28 |
| 3 | — | — | —  | — | 28 |
| 4 | — | — | 5  | — | 35 |
| 5 | — | — | 37 | — | —  |

Wierzchołek ten może zostać włączony na jeden z trzech sposobów pokazanych na rysunku 15.3.

Jeżeli cykl zostanie utworzony pierwszym sposobem, jego długość wyniesie:  $L = c_{14} + c_{43} + c_{32} + c_{21} = 37 + 5 + 11 + 33 = 86$ . Podobnie w przypadku drugiego

RYSUNEK 15.3

## Trzy sposoby włączenia wierzchołka 3

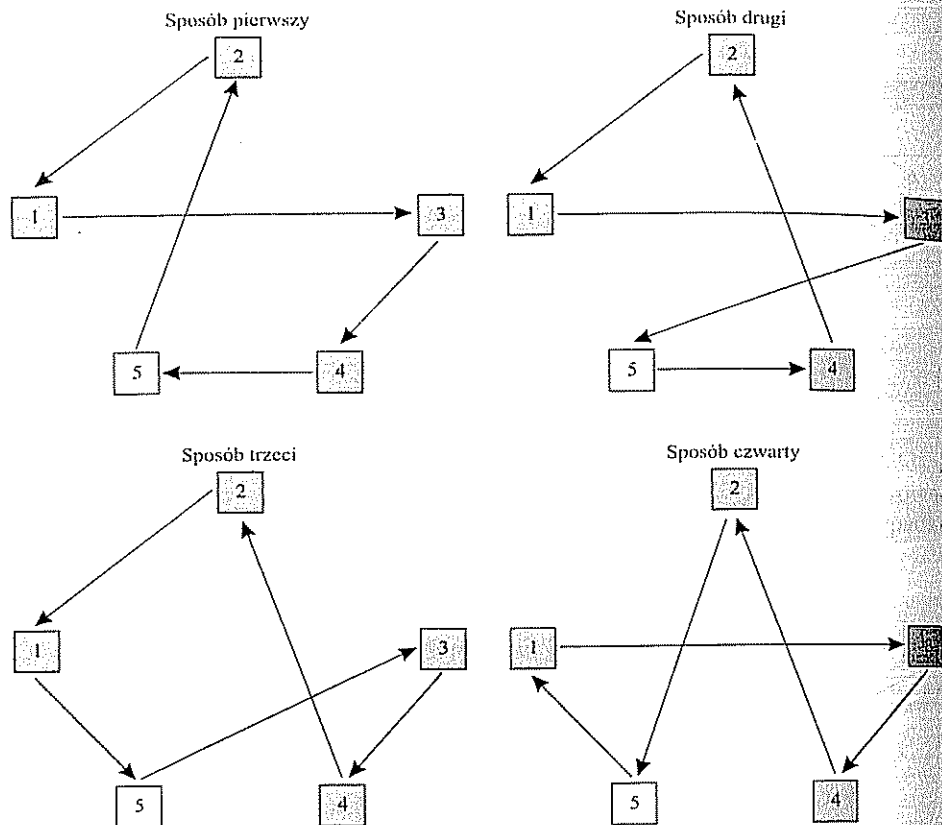


sposobu będzie równa:  $L = c_{13} + c_{34} + c_{42} + c_{21} = 26 + 5 + 8 + 33 = 72$ , a w przypadku trzeciego:  $L = c_{14} + c_{42} + c_{23} + c_{31} = 37 + 8 + 6 + 22 = 73$ . Należy wybrać wariant drugi. Ostatnim wierzchołkiem jest wierzchołek 5. Pozostaje ustalić miejsce jego włączenia. Możliwe kombinacje są przedstawione na rysunku 15.4.

W zależności od wyboru sposobu długość cyklu wyniesie:  $L = c_{13} + c_{34} + c_{45} + c_{52} + c_{21} = 26 + 5 + 35 + 27 + 33 = 126$  (sposób pierwszy),  $L = c_{13} + c_{35} + c_{54} + c_{42} + c_{21} = 26 + 28 + 17 + 8 + 33 = 122$  (sposób drugi),  $L = c_{15} + c_{53} + c_{34} + c_{42} + c_{21} = 3 + 37 + 5 + 8 + 33 = 86$  (sposób trzeci) albo  $L = c_{13} + c_{34} + c_{42} + c_{25} + c_{51} = 26 + 5 + 8 + 28 + 6 = 73$  (sposób czwarty). Jak widać, najlepszy jest sposób czwarty. Stąd ostatecznie najlepszy cykl o długości 73 ma postać jak na rysunku 15.5.

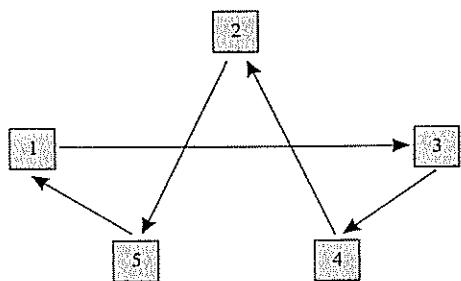
RYSUNEK 15.4

## Cztery sposoby włączenia wierzchołka 5



RYSUNEK 15.5

## Najlepszy cykl wyznaczony metodą włączania

15.2.2. Heurystyki  $k$ -optymalne

Idea działania kolejnej metody (czy też grupy metod), czyli **heurystyki  $k$ -optymalnej**, polega na poprawianiu już istniejącego kompletnego rozwiązania (a nie częściowego, jak to miało miejsce w przypadku algorytmu włączania). Dokładniej, sprawdza się, czy wymiana  $k$  krawędzi na inne, niewystępujące w danym rozwiązaniu, przyniesie poprawę rozwiązania. Jeżeli tak, to wybierana jest zamiana powodująca największe zmniejszenie odległości. Jeżeli nie, oznacza to, że zostało znalezione najlepsze rozwiązanie.

Za  $k$  podstawia się najczęściej 2 albo 3 — przy większych wartościach poziom komplikacji poszczególnych kroków (i wynikający z niego wzrost czasu działania metody) nie jest już w zasadzie rekompensowany zwiększeniem dokładności rozwiązania. Stąd czasem mówi się konkretnie o heurystyce 2-optymalnej i 3-optymalnej.

Heurystykę  $k$ -optymalną można zapisać tak, jak przedstawiono w algorytmie 15.2.

Algorytm 15.2. Heurystyka  $k$ -optymalna

- Krok 1. Wyznacz dowolną metodą pierwsze rozwiązanie.
- Krok 2. Dla każdych  $k$  krawędzi, z których żadne dwie nie sąsiadują ze sobą, sprawdź czy ich wymiana na inne  $k$  krawędzi, przy zachowaniu spójności cyklu, może spowodować jego skrócenie. Jeżeli nie — stop. Znalezione rozwiązanie uznajemy za ostateczne. Jeżeli tak, przejdź do kroku 3.
- Krok 3. Wybierz wymianę gwarantującą największe zmniejszenie długości cyklu i wróć do kroku 2.

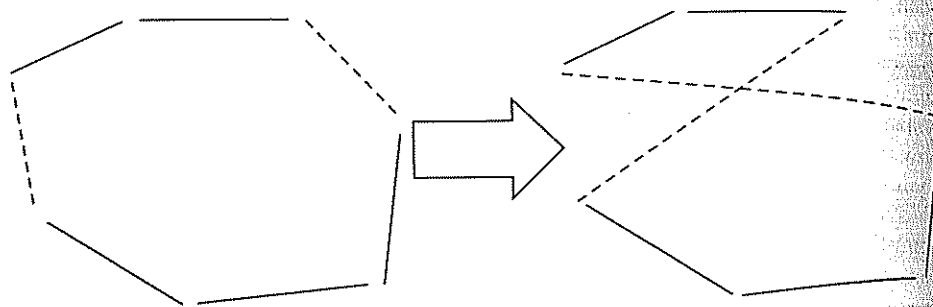
Tak jak to jest wspomniane w kroku 2, nie każda wymiana jest dopuszczalna. W szczególności w przypadku heurystyki 2-optymalnej jest możliwa tylko jedna wymiana dla zadania komiwojagera z symetryczną macierzą odległości (kierunek przemieszczania się z miasta do miasta nie ma wówczas znaczenia) i dwie możliwe wymiany dla zadania z niesymetryczną macierzą odległości. Sytuacje te są przedstawione na rysunkach 15.6 i 15.7. Linie przerywane oznaczają przykładowe krawędzie wymieniane i te, które mogą się pojawić w ich miejsce. W przypadku niesymetrycznej macierzy odległości należy zwrócić uwagę na zmianę orientacji części starego cyklu.

W przypadku heurystyki 3-optymalnej pojawia się już 7 możliwości dla zadania z symetryczną macierzą odległości (i dwukrotnie więcej dla zadania z macierzą niesymetryczną). Wraz ze wzrostem  $k$  liczba potencjalnych wymian bardzo szybko wzrasta.

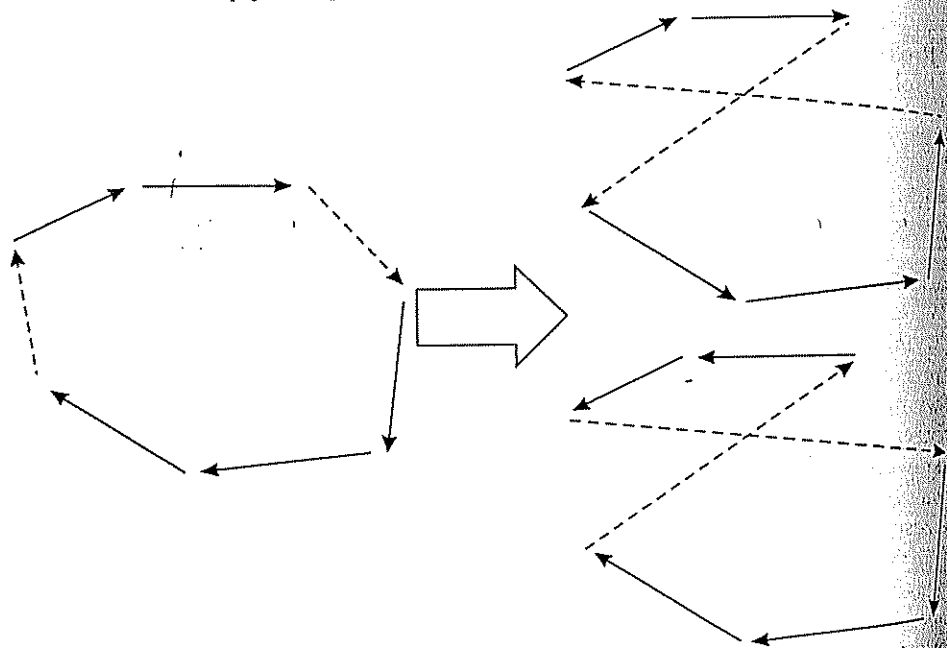
Stosowanie heurystyki  $k$ -optymalnej (dla dowolnego  $k$ ) ma sens jedynie wtedy, gdy w problemie występuje co najmniej  $2k$  wierzchołków (w przeciwnym razie nie jest możliwe znalezienie  $k$  rozłącznych luków do wymiany).



RYSUNEK 15.6  
Zamiana krawędzi w heurystyce  
2-optimalnej — symetryczna macierz odległości



RYSUNEK 15.7  
Zamiana krawędzi w heurystyce  
2-optimalnej — niesymetryczna macierz odległości



Na zakończenie tego podrozdziału zilustrujemy działanie heurystyki 2-optimalnej przykładem obliczeniowym.

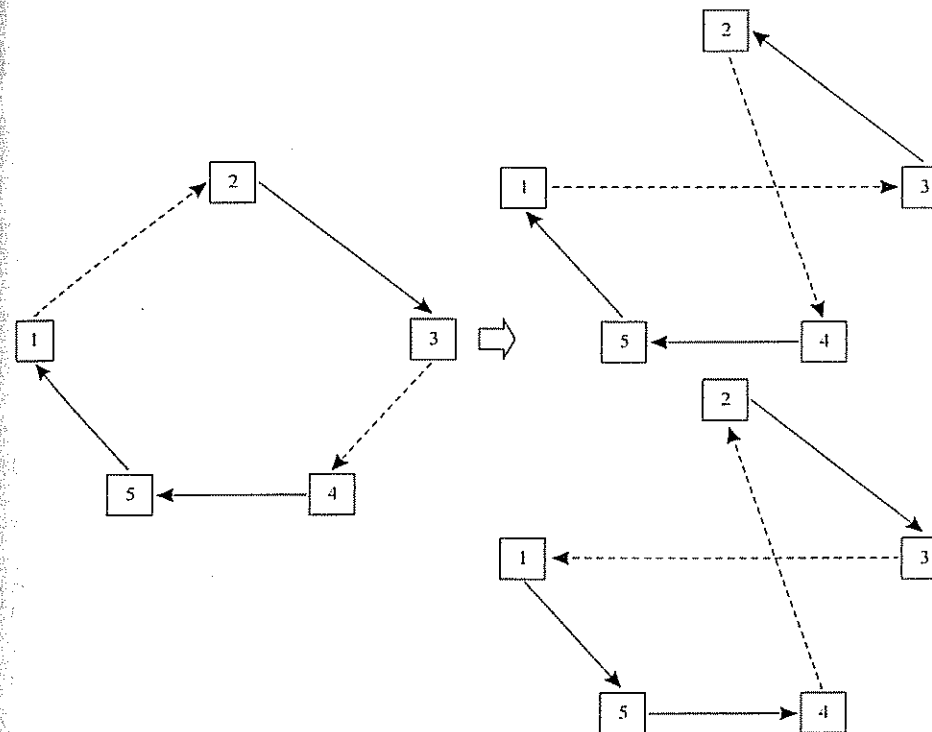
PRZYKŁAD 15.2 (cd. 15.1)

Rozwiązać rozpatrywany problem za pomocą heurystyki 2-optimalnej.

Załóżmy, że wyjściowy cykl ma postać (1, 2, 3, 4, 5). Jego długość wynosi:  $L = c_{12} + c_{23} + c_{34} + c_{45} + c_{51} = 31 + 6 + 5 + 35 + 6 = 83$ . Wymianie może ulec pięć par luków: (1, 2) i (3, 4), (1, 2) i (4, 5), (2, 3) i (4, 5), (2, 3) i (5, 1) albo (3, 4) i (5, 1).

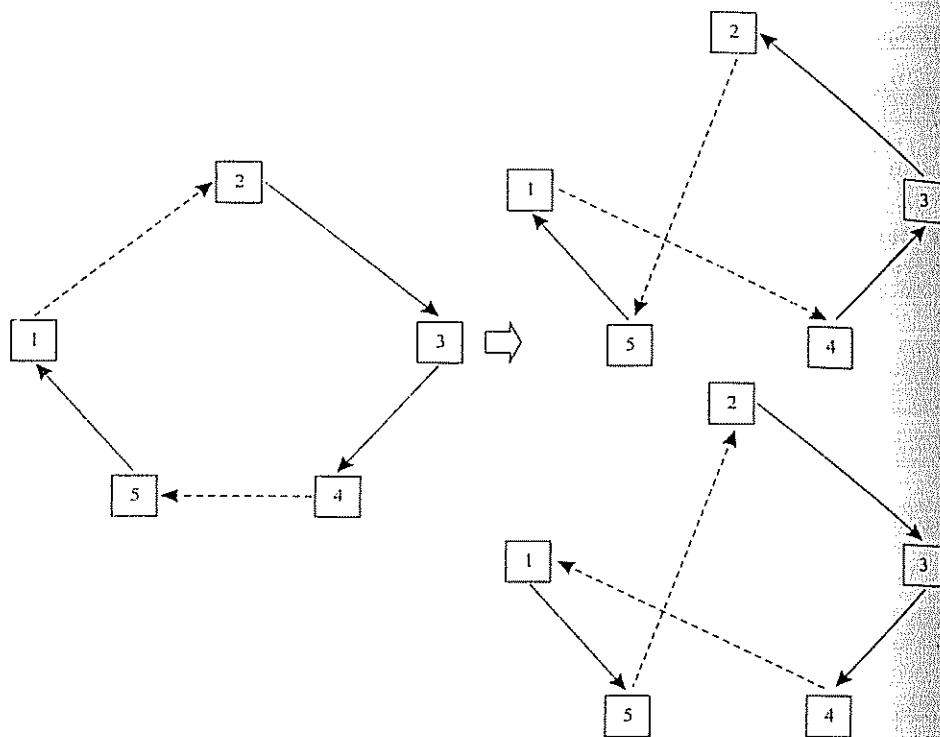
Rozpatrzmy najpierw wymianę pary luków (1, 2) i (3, 4). W jej wyniku powstają dwa cykle: (1, 3, 2, 4, 5) i (1, 5, 4, 2, 3). Sytuację tę ilustruje rysunek 15.8. Cykl (1, 3, 2, 4, 5) ma długość:  $L = c_{13} + c_{32} + c_{24} + c_{45} + c_{51} = 26 + 11 + 36 + 35 + 6 = 114$ . Z kolei długość przeciwnie do niego zorientowanego cyklu (1, 5, 4, 2, 3) jest równa:  $L = c_{15} + c_{54} + c_{42} + c_{23} + c_{31} = 3 + 17 + 8 + 6 + 22 = 56$ .

RYSUNEK 15.8  
Wymiana luków (1, 2) i (3, 4)



Przejdźmy teraz do wymiany pary luków (1, 2) i (4, 5). W jej wyniku powstają dwa cykle: (1, 4, 3, 2, 5) i (1, 5, 2, 3, 4). Sytuację tę ilustruje rysunek 15.9.

RYSUNEK 15.9  
Wymiana luków (1, 2) i (4, 5)



Cykl (1, 4, 3, 2, 5) ma długość:  $L = c_{14} + c_{43} + c_{32} + c_{25} + c_{51} = 37 + 5 + 11 + 28 + 6 = 87$ . Z kolei długość przeciwnie do niego zorientowanego cyklu (1, 5, 2, 3, 4) jest równa:  $L = c_{15} + c_{52} + c_{23} + c_{34} + c_{41} = 3 + 27 + 6 + 5 + 27 = 68$ .

W podobny sposób przeanalizować należy wymianę trzech pozostałych par luków i powstałe w ich wyniku sześć cykli.

Po sprawdzeniu wszystkich możliwych wariantów okazuje się, że wymiana luków może spowodować skrócenie cyklu. Ponadto, najkrótszym cyklem spośród wszystkich dziesięciu mogących potencjalnie powstać jest utworzony z wymiany luków (1, 2) i (3, 4) cykl (1, 5, 4, 2, 3). On więc będzie wyjściowym cyklem w kolejnej iteracji.

Ponowne sprawdzenie wszystkich możliwych wariantów wymiany luków pozwala stwierdzić, że dalsze skrócenie cyklu nie jest możliwe. Tak więc cykl (1, 5, 4, 2, 3) o długości 56 jest poszukiwanym rozwiązaniem.

## 15.3. Algorytmy mrówkowe

### 15.3.1. Ogólnie o algorytmach mrówkowych

Algorytmy mrówkowe należą do metod określaných wspólnym mianem **metod sztucznej inteligencji**. Bierze się to z faktu, że naśladują one działanie istot żywych.

W tym konkretnie przypadku mamy do czynienia z tzw. **inteligencją roju**. Jest to pojęcie określające zachowanie roju owadów społecznych (takich jak właśnie mrówki). Choć owad jako jednostka w żaden sposób nie może być nazwany istotą inteligentną, to rój takich stworzeń potrafi wykonywać z pozytywnym skutkiem złożone przedsięwzięcia (zbieranie pożywienia, budowa kopców i mrowisk itp.).

W przypadku mrówek badacz zainspirował sposób komunikacji: z jednej strony prymitywny (a więc prosty do zaprogramowania), z drugiej zaś na tyle skuteczny, że umożliwia kolonii mrówek szybko optymalizować szlaki komunikacyjne między mrowiskiem a źródłem pożywienia.

Ów mechanizm komunikacji opiera się na tzw. **szlaku feromonowym**. Każda mrówka idąc pozostawia za sobą substancję zwaną feromonem, na której zapach reagują inne mrówki. Ponieważ feromon stopniowo paruje, jest tym lepiej wyczuwalny, im później został pozostawiony. Jednak z drugiej strony, jeżeli jakaś mrówka przeszła swoją drogę więcej razy (bo zajęło jej to mniej czasu niż pozostałym owadom), pozostawiła feromon kilkakrotnie, przez co jest on lepiej wyczuwalny. Najświeższe i najliczniej powtórzone szlaki feromonowe pozostawiają mrówki, którym udało się znaleźć najkrótszą drogę. Ponieważ owady starają się wybierać trasy z największym stężeniem owej substancji, to ostatecznie cały rój zmierza po pożywienie jedną, najkrótszą (lub raczej prawie najkrótszą) drogą.

Sztuczne mrówki to proste obiekty działające na podobnej zasadzie jak ich naturalne odpowiedniki. Zostały jednak nieco udoskonalone, żeby nie powielać wad występujących w naturze.

Po pierwsze, sztuczne mrówki działają najczęściej w turach, tzn. że niezależnie od czasu przebycia trasy „czekają” na swoje towarzyszkę i dopiero na początku kolejnej tury wyruszają dalej.

Po drugie, nie kierują się jedynie ilością feromonu na trasie, lecz także jej lokalnymi właściwościami. Posługując się luźną analogią do natury, można to sobie wyobrazić w ten sposób, że mrówka nie patrzy wyłącznie na stężenie substancji zapachowej, ale również na ukształtowanie terenu. Jeżeli czeka ją strome podejście, to — nawet gdy jest suto skropione feromonem — jest duża szansa, że zdecyduje się obejść wzniesienie dookoła.

Po trzecie wreszcie, sztuczne mrówki nie wybierają „najlepszej” trasy (a więc tej o największym stężeniu feromonu), lecz losują dalszy kierunek marszu spośród

wszystkich możliwych. Niedopuszczalne są tylko ruchy, które powodowałyby cofanie się. Są one zapisywane na tzw. **liście tabu**. Prawdopodobieństwa wyboru każdego kierunku są proporcjonalne do jego oceny (która, przypomnijmy, zależy nie tylko od stężenia feromonu).

Ogólną postać algorytmu mrówkowego przedstawia algorytm 15.3.

### Algorytm 15.3. Algorytm mrówkowy

- Krok 1. Ustal maksymalną liczbę iteracji  $t_{\max}$  i liczbę mrówek  $k_{\max}$ . Wykonaj dwa kolejne kroki dla  $t$  równego od 1 do  $t_{\max}$ .
- Krok 2. Dla każdej mrówki zbuduj rozwiązanie, wybierając jego kolejną składową zgodnie z odpowiednio obliczonym prawdopodobieństwem.
- Krok 3. Uaktualnij wartości feromonu dla poszczególnych składowych zgodnie z uzyskanymi w poprzednim kroku rozwiązaniami i przyjętym schematem parowania feromonu.

Ten podstawowy schemat doczekał się wielu modyfikacji i rozszerzeń dostosowanych do różnych typów zagadnień, na których opisanie nie ma tutaj niestety miejsca. Warto jednak wspomnieć o jednej z nich. Polega ona na wprowadzeniu tzw. **mrówek elitarnych**. W tej wersji najlepsze rozwiązanie uzyskane w danej iteracji (albo we wszystkich iteracjach do obecnej włącznie) jest dodatkowo wzmacniane (czyli na odpowiadającym mu szlaku pozostawiana jest większa ilość feromonu). Możliwe jest również wprowadzenie **mrówek elitarnych z rangami**, dzięki czemu można premiować nie jedno, lecz kilka najlepszych rozwiązań.

Dobór wszystkich parametrów algorytmu odbywa się zwykle metodą prób i błędów na pewnym zbiorze testowych zadań, dla których są znane rozwiązania dokładne. Proces optymalizacji wartości parametrów jest nazywany **kalibracją algorytmu**.

Algorytmy mrówkowe nadają się najlepiej do rozwiązywania zagadnień optymalizacji kombinatorycznej. Klasycznym przykładem zastosowania algorytmu mrówkowego jest problem komiwojażera. Dlatego dalszy ciąg tego podrozdziału poświęcony będzie bardziej szczegółowemu omówieniu algorytmu mrówkowego dla tego właśnie zagadnienia.

### 15.3.2. Algorytm mrówkowy dla zagadnienia komiwojażera

Załóżmy, że dany jest problem komiwojażera z  $n$  wierzchołkami. Dana jest macierz odległości, przy czym dla każdego łuku  $(i, j)$  długość łuku (odległość)  $c_{ij}$  jest dodatnia.

Każda mrówka rozpoczyna tworzenie nowego rozwiązania od pewnego, ustalonego miasta, a następnie wybiera losowo  $n$  łuków. Dołączenie kolejnego łuku

powoduje skrócenie listy możliwych wyborów o łuki kończące się w tym samym wierzchołku (tworzy się w ten sposób listę tabu). W praktyce, ze względu na ilość danych do zapamiętania, lista tabu nie jest zbiorem łuków (odpowiadających kolejnym ruchom mrówki), lecz wierzchołków. Użycie takiej postaci listy tabu w niczym nie przeszkadza, gdyż wyeliminowanie łuków kończących się w wybranych wierzchołku jest równoważne z wyeliminowaniem tego wierzchołka.

Niech  $t$  oznacza numer aktualnie wykonywanej iteracji. Prawdopodobieństwo wyboru przez mrówkę kolejnego łuku jest wyznaczane zgodnie ze wzorem:

$$p_{ij}^k(t) = \frac{a_{ij}(t)}{\sum_{l \in N_i^k} a_{il}(t)}, \quad i = 1, \dots, n, \quad j = 1, \dots, n, \quad i \neq j, \quad (15.1)$$

przy czym:

$$a_{ij}(t) = \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_i^k} [\tau_{il}(t)]^\alpha [\eta_{il}]^\beta}, \quad i = 1, \dots, n, \quad j = 1, \dots, n, \quad i \neq j, \quad (15.2)$$

są elementami tzw. **tablicy decyzyjnej mrówek**, odnawianej w każdej iteracji, gdzie  $k$  oznacza numer mrówki,  $\tau_{ij}(t)$  — ilość śladu feromonowego na łuku  $(i, j)$ ,  $\eta_{ij}$  — wartość lokalnej funkcji kryterium (najczęściej przyjmuje się  $\eta_{ij} = 1/c_{ij}$ ),  $N_i^k$  — zbiór sąsiadów wierzchołka  $i$ , do których  $k$ -ta mrówka może jeszcze dojść (odpowiednie krawędzie nie znajdują się na jej liście tabu), a  $\alpha$  i  $\beta$  to parametry kontrolujące działanie algorytmu. Odpowiednie dobranie wartości tych ostatnich jest kluczem do sukcesu: zbyt duża wartość  $\alpha$  będzie powodować wpadanie w lokalne optima, a zbyt duża wartość  $\beta$  utratę danych o rozwiązaniach uzyskiwanych w poprzednich iteracjach i w konsekwencji zachowanie mrówek zbliżone do błędzenia losowego.

Początkową ilość feromonu na poszczególnych trasach określa się arbitralnie. Jej przyrost w iteracji  $t$  oblicza się zgodnie ze wzorem:

$$\Delta_{ij}^k(t) = \begin{cases} \frac{q}{L^k(t)}, & (i, j) \in T^k(t), \\ 0, & (i, j) \notin T^k(t), \end{cases} \quad i = 1, \dots, n, \quad j = 1, \dots, n, \quad i \neq j, \quad (15.3)$$

gdzie:

$q$  — parametr algorytmu,

$T^k(t)$  — cykl znaleziony przez mrówkę  $k$  w iteracji  $t$ ,

$L^k(t)$  — długość iteracji  $t$ .

Przyrost feromonu dla trasy oblicza się zgodnie ze wzorem:

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \sum_{k=1}^m \Delta_{ij}^k(t), \quad (15.4)$$

gdzie  $\rho \in (0, 1)$  jest współczynnikiem parowania feromonu.

Dobór dwóch ostatnich parametrów ma również istotny wpływ na działanie algorytmu. Zbyt duża wartość  $\rho$  wywołuje podobne skutki jak w przypadku parametru  $\beta$ , czyli utratę informacji o poprzednich rozwiązaniach i w konsekwencji zachowanie mrówek zbliżone do błędzenia losowego. Z kolei zbyt duża wartość  $q$  zwiększa prawdopodobieństwo wpadnięcia w lokalne optimum.

Za ostateczne rozwiązanie uznaje się najlepsze spośród uzyskanych przez wszystkie mrówki we wszystkich iteracjach.

Działanie poszczególnych fragmentów algorytmu zostało zilustrowane na przykładzie 15.3.

### PRZYKŁAD 15.3 (cd. 15.1)

Załóżmy, że do rozwiązania zadania użyto dwóch mrówek. Niech  $\alpha = 0,7$ ,  $\beta = 0,3$ ,  $\rho = 0,1$ ,  $q = 20$ . Przyjmijmy ponadto, że po kilku iteracjach ilości feromonu  $\tau_{ij}$  przypisane poszczególnym lukom mają takie wartości, jak w tabelicy 15.4.

TABLICA 15.4  
Przykładowe ilości feromonu

|   | 1    | 2    | 3    | 4    | 5    |
|---|------|------|------|------|------|
| 1 | —    | 0,47 | 0,07 | 0,05 | 0,44 |
| 2 | 0,46 | —    | 2,71 | 0,33 | 0,07 |
| 3 | 0,76 | 1,57 | —    | 1,13 | 0,30 |
| 4 | 0,18 | 2,28 | 2,84 | —    | 0,06 |
| 5 | 0,66 | 0,28 | 0,31 | 0,76 | —    |

Załóżmy wreszcie, że najkrótszy dotychczas znaleziony cykl to (1, 5, 2, 3, 4) o długości 68. Zostanie teraz wykonana kolejna iteracja algorytmu.

Po wyznaczeniu odwrotności odległości wartości lokalnej oceny poszczególnych luków  $\eta_{ij}$  są równe liczbom z tabelicy 15.5. Przykładowo, dla łuku (1, 2) mamy:  $\eta_{12} = 1/31 = 0,03$ , a dla łuku (4, 2):  $\eta_{42} = 1/8 = 0,13$ .

Wartości poszczególnych elementów tabelicy decyzyjnej mrówek obliczamy zgodnie ze wzorem (15.2), przykładowo:

$$a_{12} = \frac{(0,47)^{0,7}(0,03)^{0,3}}{(0,47)^{0,7}(0,03)^{0,3} + (0,07)^{0,7}(0,04)^{0,3} + (0,05)^{0,7}(0,03)^{0,3} + (0,44)^{0,7}(0,33)^{0,3}} = 0,29.$$

TABLICA 15.5  
Wartości lokalnej oceny luków

|   | 1    | 2    | 3    | 4    | 5    |
|---|------|------|------|------|------|
| 1 | —    | 0,03 | 0,04 | 0,03 | 0,33 |
| 2 | 0,03 | —    | 0,17 | 0,03 | 0,04 |
| 3 | 0,05 | 0,09 | —    | 0,2  | 0,04 |
| 4 | 0,04 | 0,13 | 0,20 | —    | 0,03 |
| 5 | 0,17 | 0,04 | 0,03 | 0,06 | —    |

Ostateczna postać tabelicy decyzyjnej została przedstawiona w tabelicy 15.6.

TABLICA 15.6  
Tabela decyzyjna mrówek

|   | 1    | 2    | 3    | 4    | 5    |
|---|------|------|------|------|------|
| 1 | —    | 0,29 | 0,08 | 0,06 | 0,57 |
| 2 | 0,13 | —    | 0,74 | 0,1  | 0,04 |
| 3 | 0,18 | 0,36 | —    | 0,37 | 0,09 |
| 4 | 0,05 | 0,40 | 0,53 | —    | 0,02 |
| 5 | 0,40 | 0,14 | 0,14 | 0,32 | —    |

Załóżmy, że obie mrówki rozpoczynają budowę cyklu od wierzchołka 1. Prawdopodobieństwa wyboru kolejnych wierzchołków będziemy obliczać, korzystając ze wzoru (15.1).

Zacznijmy od pierwszej mrówki. Jej lista tabu składa się tylko z wierzchołka 1. Prawdopodobieństwa wylosowania kolejnych wierzchołków równe są:

$$p_{12} = \frac{0,29}{0,29 + 0,08 + 0,06 + 0,57} = 0,29,$$

$$p_{13} = \frac{0,08}{0,29 + 0,08 + 0,06 + 0,57} = 0,08,$$

$$p_{14} = \frac{0,06}{0,29 + 0,08 + 0,06 + 0,57} = 0,06,$$

$$p_{15} = \frac{0,57}{0,29 + 0,08 + 0,06 + 0,57} = 0,57.$$

Załóżmy, że w wyniku losowania został wybrany wierzchołek 2. Tak więc pierwszym wybranym łukiem jest łuk (1, 2), a lista tabu powiększa się o wierzchołek 2. W kolejnym kroku należy wyznaczyć prawdopodobieństwa wyboru wierzchołków 3, 4 i 5 (tym razem wierzchołkiem początkowym jest wierzchołek 2):

$$p_{23} = \frac{0,74}{0,74 + 0,1 + 0,04} = 0,84,$$

$$p_{24} = \frac{0,1}{0,74 + 0,1 + 0,04} = 0,11,$$

$$p_{25} = \frac{0,04}{0,74 + 0,1 + 0,04} = 0,05.$$

Załóżmy, że w wyniku losowania wybrany został wierzchołek 3. Kolejnym wybranym łukiem jest więc łuk (2, 3), a lista tabu powiększa się o wierzchołek 3. W kolejnym kroku należy wyznaczyć prawdopodobieństwa wyboru wierzchołków 4 i 5 (wierzchołkiem początkowym jest wierzchołek 3):

$$p_{34} = \frac{0,37}{0,37 + 0,09} = 0,80,$$

$$p_{35} = \frac{0,09}{0,37 + 0,09} = 0,20.$$

Przyjmijmy, że tym razem wylosowany został wierzchołek 4. Kolejnym łukiem w cyklu jest więc łuk (3, 4), a lista tabu powiększa się o wierzchołek 4. Ponieważ poza listą tabu pozostał jeszcze tylko wierzchołek 5, zostanie on na pewno dołączony do wierzchołka 4, a co za tym idzie cykl powiększy się o dwa ostatnie łuki: (4, 5) i (5, 1). Cykl wyznaczony przez pierwszą mrówkę przyjmie więc postać (1, 2, 3, 4, 5), a jego długość równa jest 83. Najlepszym cyklem pozostaje wyznaczony uprzednio cykl o długości 68.

Przejdźmy teraz do drugiej mrówki. Jej lista tabu składa się tylko z wierzchołka 1. Prawdopodobieństwa wylosowania kolejnych wierzchołków równe są:

$$p_{12} = \frac{0,29}{0,29 + 0,08 + 0,06 + 0,57} = 0,29,$$

$$p_{13} = \frac{0,08}{0,29 + 0,08 + 0,06 + 0,57} = 0,08,$$

$$p_{14} = \frac{0,06}{0,29 + 0,08 + 0,06 + 0,57} = 0,06,$$

$$p_{15} = \frac{0,57}{0,29 + 0,08 + 0,06 + 0,57} = 0,57.$$

Załóżmy, że w wyniku losowania został wybrany wierzchołek 5. Tak więc pierwszym wybranym łukiem jest łuk (1, 5), a lista tabu powiększa się o wierzchołek 5. W kolejnym kroku należy wyznaczyć prawdopodobieństwa wyboru wierzchołków 2, 3 i 4 (tym razem wierzchołkiem początkowym jest wierzchołek 5):

$$p_{52} = \frac{0,14}{0,14 + 0,14 + 0,32} = 0,23,$$

$$p_{53} = \frac{0,14}{0,14 + 0,14 + 0,32} = 0,23,$$

$$p_{54} = \frac{0,32}{0,14 + 0,14 + 0,32} = 0,54.$$

Załóżmy, że w wyniku losowania wybrany został wierzchołek 4. Kolejnym wybranym łukiem jest więc łuk (5, 4), a lista tabu powiększa się o wierzchołek 4. W kolejnym kroku należy wyznaczyć prawdopodobieństwa wyboru wierzchołków 2 i 3 (wierzchołkiem początkowym jest wierzchołek 4):

$$p_{42} = \frac{0,40}{0,40 + 0,53} = 0,43,$$

$$p_{43} = \frac{0,53}{0,40 + 0,53} = 0,57.$$

Przyjmijmy, że tym razem wylosowany został wierzchołek 2. Kolejnym łukiem w cyklu jest więc łuk (4, 2), a lista tabu powiększa się o wierzchołek 2. Ponieważ poza listą tabu pozostał jeszcze tylko wierzchołek 3, zostanie on na pewno dołączony do wierzchołka 2, a co za tym idzie cykl powiększy się o dwa ostatnie łuki: (2, 3) i (3, 1). Cykl wyznaczony przez drugą mrówkę przyjmie więc postać (1, 5, 4, 2, 3), a jego długość równa jest 56. Nowy cykl jest lepszy od dotychczasowego, zostaje więc zapamiętany jako najlepszy.

Po wyznaczeniu rozwiązań przez wszystkie mrówki następuje modyfikacja ilości feromonu na poszczególnych łukach na podstawie wzorów (15.3) i (15.4). Dla wszystkich łuków należących do cyklu wyznaczonego przez mrówkę pierwszą przyrost ilości feromonu jest równy:

$$\Delta_{ij}^1 = \frac{20}{83} = 0,24.$$

Dla każdego łuku należącego do cyklu wyznaczonego przez mrówkę drugą wynosi natomiast:

$$\Delta_{ij}^2 = \frac{20}{56} = 0,36.$$

Łuk (1, 3) nie należy do żadnego z cykli, w związku z tym przypisana mu nowa wartość feromonu zmniejsza się w wyniku odparowania:

$$\tau_{13} = (1 - 0,1) \cdot 0,07 = 0,06.$$

Łuk (1, 2) należy wyłącznie do cyklu zbudowanego przez mrówkę pierwszą, więc przypisana mu ilość feromonu zmniejsza się w wyniku odparowania, ale też zwiększy o 0,24:

$$\tau_{12} = (1 - 0,1) \cdot 0,47 + 0,24 = 0,66.$$

Podobnie, na zmianę ilości feromonu przypisanego łukowi (1, 5) wpływ będzie miał tylko przyrost przypisany mrówce drugiej:

$$\tau_{15} = (1 - 0,1) \cdot 0,44 + 0,36 = 0,75.$$

Łuk (2, 3) występuje w obu cyklach, więc przypisana mu ilość feromonu zwiększy się w dwójnasób:

$$\tau_{23} = (1 - 0,1) \cdot 2,71 + 0,24 + 0,36 = 3,04.$$

Wszystkie wyznaczone nowe ilości feromonu zostały umieszczone w tabeli 15.7.

TABLICA 15.7  
Nowe ilości feromonu

|   | 1    | 2    | 3    | 4    | 5    |
|---|------|------|------|------|------|
| 1 | —    | 0,66 | 0,06 | 0,05 | 0,75 |
| 2 | 0,41 | —    | 3,04 | 0,30 | 0,06 |
| 3 | 1,04 | 1,41 | —    | 1,26 | 0,27 |
| 4 | 0,16 | 2,41 | 2,56 | —    | 0,29 |
| 5 | 0,83 | 0,25 | 0,28 | 1,04 | —    |

## 15.4. Algorytmy genetyczne

### 15.4.1. Wprowadzenie

Algorytmy genetyczne stanowią próbę wykorzystania mechanizmów znanych z teorii ewolucji żywych organizmów oraz z nauki o genetyce do poszukiwania optymalnych rozwiązań sztucznych, stworzonych przez człowieka problemów. Tak jak wśród dwóch przedstawicieli tego samego gatunku żywych organizmów jeden jest silniejszy, a drugi słabszy, tak samo możemy wskazać lepsze i gorsze rozwiązania postawionego problemu decyzyjnego. Jeżeli więc rozpatrzemy pewien zbiór rozwiązań (niekoniecznie wszystkich możliwych), to możemy go porównać do populacji organizmów. Przypisując poszczególnym rozwiązaniom-osobnikom ich indywidualne oceny przystosowania do postawionych warunków (wartości **funkcji przystosowania**), możemy symulować procesy ewolucyjne, reprodukcję w kolejnych „pokoleniach” lepsze spośród rozwiązań, eliminując natomiast te, które słabo spełniają zadane kryteria.

W naturze o ogólnym przystosowaniu danego organizmu decyduje zespół cech, które mogą poszczególne osobniki do siebie upodabniać, lub różnić. Analogicznie, w teorii **algorytmów genetycznych** zakłada się, że rozwiązanie problemu decyzyjnego nie jest nierozdzieloną całością, lecz sumą pewnych elementów składowych (idei cząstkowych). Pozwala to na wzbogacenie ewolucyjnej reguły „silniejszy wygrywa” o analogię do procesu reprodukcji kodu genetycznego, a konkretnie o możliwość zestawiania ze sobą (na wzór krzyżowania) idei cząstkowych pochodzących z różnych rozwiązań i otrzymywania w ten sposób innowacyjnych rozwiązań rozpatrywanego problemu. Od strony technicznej wymaga to przyjęcia sposobu przekształcania konkretnego rozwiązania w jednoznacznie je identyfikujący ciąg kodowy (odpowiednik kodu genetycznego). Tak zdefiniowane **ciągi kodowe** można poddawać przetwarzaniu, np. krzyżować parami czy poddawać mutacji, otrzymując tą drogą ciągi kodowe reprezentujące nowe rozwiązania.

Formalnie, ciąg składający się z  $n$  znaków, z których każdy może występować w jednej z  $m$  odmian, można przedstawić:

$$a_1, a_2, a_3, \dots, a_{n-1}, a_n \quad \text{gdzie } a_i \in \{w_1, w_2, \dots, w_n\}.$$

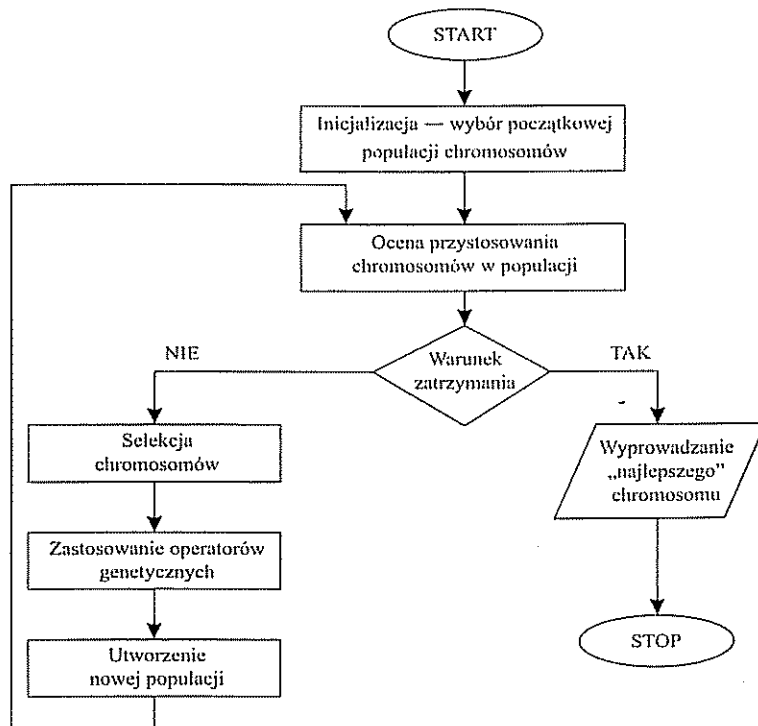
Jeżeli przyjmiemy, że każda kombinacja znaków (każdy ciąg) jednoznacznie definiuje nieco inne rozwiązanie postawionego problemu, to  $n$ -znakowy ciąg opisany na  $m$ -elementowym alfabecie daje możliwość wyróżnienia  $m^n$  różnych rozwiązań. Stosując analogię genetyczną, ciąg kodowy reprezentujący pewne rozwiązanie bywa nazywany **genotypem** lub **chromosomem** (w typowych implementacjach algorytmu genotyp składa się z jednego chromosomu, a więc określenia te mogą być stosowane zamiennie).

Schemat działania typowego algorytmu genetycznego jest następujący:

### Algorytm 15.4. Algorytm genetyczny

- Krok 1. Wyznacz (losowo)  $N$ -elementową populację rozwiązań rozpatrywanego problemu.
- Krok 2. Wyznacz dla każdego z rozwiązań w populacji wartość funkcji przystosowania.
- Krok 3. Czy spełniony jest warunek zatrzymania algorytmu? Jeśli tak, idź do kroku 7, jeśli nie, idź do kroku 4.
- Krok 4. Wylosuj skład  $\frac{N}{2}$  par rozwiązań (tzw. pulę rodzicielską), przyjmując, że prawdopodobieństwo znalezienia się w puli jest rosnące względem wartości funkcji przystosowania. Wielokrotne występowanie tego samego rozwiązania w puli (a nawet w jednej parze) jest dopuszczalne.

RYSUNEK 15.10  
Schemat blokowy algorytmu genetycznego



Źródło: [D. Rutkowska, M. Piliński, L. Rutkowski, 1999, s. 137].

Krok 5. Utwórz nowe  $N$ -elementowe pokolenie rozwiązań, przetwarzając pulę rodzicielską, wykorzystując w tym celu **operatory genetyczne (krzyżowanie, mutacja)**.

Krok 6. Idź do kroku 2.

Krok 7. Wybierz najlepsze rozwiązanie z bieżącej populacji.

Krok 8. Koniec.

### 15.4.2. Techniczne aspekty funkcjonowania algorytmu genetycznego

W większości implementacji algorytmów genetycznych kod, w którym zapisywane są rozwiązania rozpatrywanego problemu optymalizacyjnego, ma postać binarną. Oznacza to, że każdy znak ciągu może przyjmować jedną z dwóch wartości: 0 lub 1. Pozwala to m.in. na proste odwzorowywanie wartości liczbowych, gdyż można wykorzystać w tym celu zapis dwójkowy.

Przykładowo, ciąg binarny o postaci:

0101110

może reprezentować liczbę o wartości 46 ( $0 \cdot 2^6 + 1 \cdot 2^5 + 0 \cdot 2^4 + 1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^1 + 0 \cdot 2^0$ ).

Ogólnie, za pomocą  $n$ -znakowego ciągu binarnego można zapisać liczby całkowite z przedziału  $[0, 2^n - 1]$ .

Całkowitoliczbowość wynikająca z zapisu dwójkowego nie musi bezpośrednio przekładać się na całkowitoliczbowość wartości zmiennych decyzyjnych zadania. Bez trudu można zastosować przekształcenie osłabiające warunek całkowitoliczbowości do warunku dyskretyzacji zmiennych. Jeżeli szukamy wartości zmiennej  $x$  w przedziale  $[a, b]$ , to wystarczy przyjąć, że ciąg składający się z samych zer odpowiada wartości  $a$ , a ciąg samych jedynek odpowiada wartości  $b$ . Dokładność odwzorowania wartości z danego przedziału zależy od liczby znaków w ciągu.

Przykładowo, rozpatrzmy możliwość odwzorowania wartości  $x \in [-2, 5]$ . Żądamy dokładności opisu zmiennej  $x$  do trzeciego miejsca po przecinku (w zapisie dziesiętnym). Oznacza to konieczność przyjęcia zapisu kodowego, który uwzględniłby co najmniej  $(5 - (-2)) \cdot 10^3 + 1 = 7001$  różnych stanów. Zapis dwójkowy składający się z  $n$  znaków pozwala wyróżnić  $2^n$  stanów:

$$2^{12} = 4\ 096,$$

$$2^{13} = 8\ 192.$$

Oznacza to, że aby przedstawić liczbę z zakresu  $[-2, 5]$  z dokładnością do  $1/1000$  potrzebne jest przyjęcie zapisu kodowego o długości nie mniejszej niż 13 znaków. Uogólniając, liczba znaków zapisu dwójkowego potrzebna do odzwierciedlenia

wartości zmiennej należącej do przedziału  $[a, b]$ , z dokładnością do  $k$  miejsc po przecinku, jest równa najmniejszej liczbie  $n$  spełniającej warunek:

$$2^n - 1 \geq (b - a) \cdot 10^k. \quad (15.5)$$

Znając  $n$ -znakowy ciąg  $Z$ , wartość odpowiadającą mu zmiennej  $x$  otrzymujemy, korzystając z następującego wzoru:

$$x = a + \frac{\text{dec}(Z)(b - a)}{(2^n - 1)}, \quad (15.6)$$

gdzie  $\text{dec}(Z)$  jest wartością dziesiętną liczby opisanej podanym ciągiem dwójkowym  $Z$ .

W powyższym przypadku ciąg 0110100111010 reprezentuje liczbę:

$$x = -2 + \frac{3386 \cdot (5 - (-2))}{(2^{13} - 1)} \approx 0,894.$$

Jeżeli do opisanego rozwiązania zadania potrzeba więcej niż jednej zmiennej, to ciąg kodowy będzie składał się z podciągów określających wartości poszczególnych zmiennych. Pierwsze  $n_1$  znaków w ciągu będzie odpowiadało zmiennej  $x_1$ , kolejne  $n_2$  znaków zmiennej  $x_2$  itd.

**Kryterium oceny rozwiązań** jest w algorytmie genetycznym wyrażane przez funkcję przystosowania. Rozpatrzmy funkcję  $f$ , która:

- 1) jest nieujemna dla każdego rozwiązania dopuszczalnego,
- 2) jest rosnąca w miarę poprawy rozwiązań.

Niech  $f_i$  oznacza wartość funkcji przystosowania  $i$ -tego rozwiązania, gdzie  $i = 1, 2, \dots, N$ , przy czym  $N$  oznacza liczebność populacji rozwiązań rozpatrywanych łącznie w pojedynczej iteracji algorytmu. W najbardziej typowej implementacji algorytmu genetycznego skład puli rodzicielskiej jest generowany przy użyciu tzw. **metody ruletki**. Obsadzenie  $N$  miejsc w puli rodzicielskiej można zilustrować jako  $N$ -krotny wybór ciągu kodowego za pomocą koła ruletki, na którym każdemu ciągowi z populacji jest przyporządkowany wycinek o wielkości proporcjonalnej do wartości funkcji przystosowania danego rozwiązania. Jeżeli prawdopodobieństwo zajęcia pojedynczego miejsca w puli rodzicielskiej przez dany ciąg jest proporcjonalne do wartości funkcji przystosowania oraz każde miejsce w puli musi zostać obsadzone, to prawdopodobieństwo, że  $i$ -ty ciąg kodowy zostanie w pojedynczym losowaniu wytypowany do zajęcia miejsca w puli rodzicielskiej jest równe:

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j}. \quad (15.7)$$

Wtedy wartość oczekiwana liczby kopii  $i$ -tego rozwiązania w puli rodzicielskiej wynosi:

$$\mu_i = p_i N = \frac{f_i}{\sum_{j=1}^N f_j} = \frac{f_i}{\bar{f}}, \quad (15.8)$$

gdzie  $\bar{f}$  jest średnią arytmetyczną wartości funkcji przystosowania w bieżącej populacji.

Łatwo zauważyć, że ciągi kodowe reprezentujące rozwiązania lepsze niż średnia będą (przeciętnie) zwiększały swój udział w populacji (wartość oczekiwana liczby ich kopii będzie większa niż 1) kosztem ciągów kodowych odpowiadających rozwiązaniom słabszym niż przeciętna w populacji. Jest to zależność korzystna z punktu widzenia procesu ewoluowania rozwiązań w kolejnych pokoleniach.

Typowymi operacjami, które są wykonywane na populacji rodzicielskiej w celu utworzenia nowego pokolenia, są krzyżowanie oraz mutacja.

**Krzyżowanie** polega na zamianie między parą rozwiązań fragmentów ich ciągów kodowych. Podstawowym typem krzyżowania jest **krzyżowanie jednopunktowe**. Polega ono na „rozcieciu” obu krzyżowanych ciągów kodowych w losowo wybranym miejscu (tym samym dla obu ciągów) i zamianie odciętych fragmentów między ciągami.

Przykładowo, dwa ciągi:

Rodzic 1 1 1 0 0 1 0 1  
 Rodzic 2 0 1 1 1 0 1 1

rozciete między trzecim a czwartym znakiem i skrzyżowane dadzą następujące „potomstwo”:

Dziecko 1 1 1 0 1 0 1 1  
 Dziecko 2 0 1 1 0 1 0 1

**Mutacja** jest operacją losowej zmiany wartości wybranego znaku w ciągu kodowym. W przypadku kodu binarnego oznacza to zmianę na losowo wybranej pozycji z 0 na 1 lub odwrotnie.

W większości implementacji algorytmów genetycznych przyjmuje się, że zarówno krzyżowanie, jak i mutacja nie występuje w 100% przypadków. Określa się prawdopodobieństwo krzyżowania  $p_c$  oraz prawdopodobieństwo mutacji  $p_m$ . O ile jednak prawdopodobieństwo krzyżowania zazwyczaj przyjmowane jest na dość wysokim poziomie ( $p_c > 0,5$ ), o tyle sugeruje się, aby mutacja była procesem zachodzącym rzadko ( $p_m < 0,1$ ).



Utworzone ciągi potomne stają się **nową populacją**. Proces oceny i reprodukcji populacji jest powtarzany. Cały cykl jest realizowany przez szereg **pokoleń** aż do spełnienia **kryterium zatrzymania algorytmu**. Rozpatruje się kilka możliwych typów kryteriów:

- 1) upływanie założonej liczby pokoleń;
- 2) zbieżność wartości funkcji przystosowania, tj. brak znaczącej poprawy średniej wartości funkcji przystosowania w kilku następujących po sobie pokoleniach;
- 3) zbieżność składu populacji, tj. osiągnięcie przez ciągi kodowe w populacji określonego stopnia wzajemnego podobieństwa.

#### PRZYKŁAD 15.4

Rozpatrzmy poniższe zadanie optymalizacyjne:

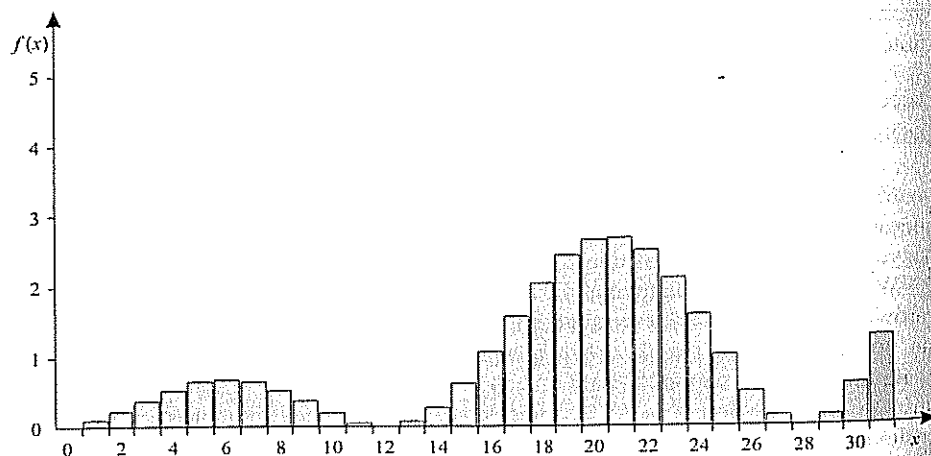
$$f(x) = \left( \sin\left(\frac{\pi x}{8}\right) + 1 \right) \frac{x}{15} \rightarrow \max.$$

$$0 \leq x \leq 31$$

$$x \in C$$

Wskazanie rozwiązania optymalnego powyższego zadania nie nastęca trudności, gdyż przy tak niewielkim, skończonym zbiorze rozwiązań dopuszczalnych zadowalająca byłaby nawet metoda przeglądu zupełnego rozwiązań. W celu zilustrowania działania algorytmu genetycznego przeprowadzimy jednak szczegółową analizę problemu.

RYSUNEK 15.11  
Wykres maksymalizowanej funkcji



Zadanie ma 32 rozwiązania, z których każde jest jednoznacznie opisane za pomocą odpowiedniej wartości zmiennej  $x$ . Możemy przyjąć, że ciąg kodowy reprezentujący dane rozwiązanie będzie zapisem wartości  $x$  w systemie dwójkowym, czyli będzie ciągiem pięciu zer lub jedynek.

Przyjmijmy, że w skład populacji w pojedynczym pokoleniu będzie wchodziło sześć rozwiązań. Pierwsze pokolenie zostanie utworzone w sposób losowy, czyli ciągi kodowe zostaną utworzone na podstawie wartości podawanych przez generator liczb pseudolosowych, np. w arkuszu Microsoft Excel.

TABLICA 15.8  
Skład populacji początkowej  
(uzyskany losowo)

| Lp. | Ciąg kodowy | Odkodowana wartość $x$ |
|-----|-------------|------------------------|
| 1   | 11000       | 24                     |
| 2   | 11010       | 26                     |
| 3   | 11110       | 30                     |
| 4   | 10011       | 19                     |
| 5   | 11101       | 29                     |
| 6   | 01010       | 10                     |

Każde z rozwiązań wchodzących w skład populacji zostaje ocenione przy użyciu wartości funkcji przystosowania  $f$ . Na podstawie wyników są określone prawdopodobieństwa, z jakimi za pomocą selekcji ruletkowej ciągi kodowe będą wybierane do udziału w puli rodzicielskiej.

Skumulowane wartości prawdopodobieństw wylosowania poszczególnych ciągów są wykorzystywane przy realizacji selekcji ruletkowej. Ostatnia kolumna w tabeli 15.9 przedstawia oczekiwaną liczbę reprezentantów danego ciągu w puli rodzicielskiej.

Sześć miejsc w puli rodzicielskiej jest obsadzane według następującej procedury.

Generowana jest liczba pseudolosowa  $r \in [0, 1)$ . Miejsce w puli jest obsadzane ciągiem o numerze  $i$ , dla którego zachodzi:

$$\text{cum}(i-1) \leq r < \text{cum}(i). \quad (15.9)$$

Zakładamy, że  $\text{cum}(0) = 0$ .

W opisany powyżej sposób został wygenerowany skład puli rodzicielskiej.

Zakładamy prawdopodobieństwo krzyżowania  $p_c = 1$ , czyli kolejne pary puli rodzicielskiej będą wzajemnie krzyżowane. Stosując ponownie generator liczb pseudolosowych, wyznaczono dla każdej pary ciągów miejsce krzyżowania

TABLICA 15.9  
Obliczenia do operacji reprodukcji

| $i$ | Ciąg kodowy | $x$     | $f_i$ | $p_i = \frac{f_i}{\sum_{j=1}^6 f_j}$ | $\text{cum}(i) = \sum_{j=1}^i p_j$ | $\mu_i = \frac{f_i}{f}$ |
|-----|-------------|---------|-------|--------------------------------------|------------------------------------|-------------------------|
| 1   | 11000       | 24      | 1,600 | 0,292                                | 0,292                              | 1,754                   |
| 2   | 11010       | 26      | 0,508 | 0,093                                | 0,385                              | 0,557                   |
| 3   | 11110       | 30      | 0,586 | 0,107                                | 0,492                              | 0,642                   |
| 4   | 10011       | 19      | 2,437 | 0,445                                | 0,937                              | 2,672                   |
| 5   | 11101       | 29      | 0,147 | 0,027                                | 0,964                              | 0,161                   |
| 6   | 01010       | 10      | 0,195 | 0,036                                | 1,000                              | 0,214                   |
|     |             | Suma    | 5,473 | 1                                    |                                    | 6                       |
|     |             | Średnia | 0,912 |                                      |                                    |                         |

(w tablicy 15.10 oznaczone pionową kreską). Przypomnijmy, że chociaż w prezentowanym przykładzie taka sytuacja nie wystąpiła (por. tablica 15.10), generalnie dopuszczalne jest krzyżowanie ciągu kodowego z inną kopią jego samego w puli rodzicielskiej.

Każdy gen (znak) w puli rodzicielskiej może ulec mutacji z prawdopodobieństwem  $p_m$ . Przyjęto  $p_m = 0,05$ , co daje oczekiwaną liczbę mutacji w pokoleniu:  $0,01 \cdot 5 \cdot 6 = 1,5$ . Przyjmijmy, że w wyniku losowania okazało się, że mutacji ulegnie piąty znak w drugim ciągu kodowym puli rodzicielskiej (podkreślony w tablicy).

TABLICA 15.10  
Pula rodzicielska pierwszego pokolenia

| Numer pary | Numer ciągu kodowego | Ciąg kodowy | Ciąg kodowy po mutacji i krzyżowaniu |
|------------|----------------------|-------------|--------------------------------------|
| 1          | 1                    | 1100 0      | 1100  <u>1</u>                       |
|            | 2                    | 1101 0      | 1101 0                               |
| 2          | 3                    | 111 10      | 111 11                               |
|            | 4                    | 100 11      | 100 10                               |
| 3          | 1                    | 1100 0      | 1100 1                               |
|            | 4                    | 1001 1      | 1001 0                               |

Odpowiednie pary ciągów kodowych w puli rodzicielskiej są teraz łączone „na krzyż” zgodnie z ustalonym miejscem przecięcia, dając nowe pokolenie (tablica 15.11).

TABLICA 15.11  
Nowe pokolenie rozwiązań

| Lp. | Ciąg kodowy | $x$     | $f(x)$ |
|-----|-------------|---------|--------|
| 1   | 11001       | 25      | 1,029  |
| 2   | 11010       | 26      | 0,508  |
| 3   | 11111       | 31      | 1,276  |
| 4   | 10010       | 18      | 2,049  |
| 5   | 11001       | 25      | 1,029  |
| 6   | 10010       | 18      | 2,049  |
|     |             | Suma    | 7,938  |
|     |             | Średnia | 1,323  |

Uzyskano nowe pokolenie, którego członkowie lepiej niż ich „rodzice” spełniają założone kryterium — osiągają wyższe wartości funkcji przystosowania (średnio 1,323 w porównaniu do 0,912 w pokoleniu pierwszym).

Otrzymana populacja ponownie podlegałaby ocenie i w analogicznym do przeprowadzonego procesie dałaby kolejne, trzecie pokolenie. Powtarzanie procedury przez wiele pokoleń powinno prowadzić do uzyskiwania coraz lepszych rozwiązań, aż do otrzymania rozwiązania satysfakcjonującego.

### 15.4.3. Adaptacja algorytmu genetycznego na potrzeby dowolnego zadania

W powyższych rozważaniach omawiany był problem optymalizacyjny o szczególnej postaci, gdyż przy przyjętym sposobie kodowania każdy możliwy ciąg kodowy reprezentował pewne rozwiązanie dopuszczalne. Pozytywną konsekwencją tego faktu było również to, że dowolne przekształcenia na ciągach kodowych (krzyżowanie i mutacja) prowadziły do otrzymywania rozwiązań, które również były dopuszczalne.

Należy odrzucić założenie, że dla każdego potencjalnego problemu optymalizacyjnego uda się skonstruować taki sposób kodowania rozwiązania, który spełniałby warunki wskazane w poprzednim akapicie.

Rozpatrzmy problem komiwojażera. Przyjmijmy, że do odwiedzenia jest  $n$  miast  $M = \{m_1, m_2, \dots, m_n\}$ , z których każde musi być odwiedzone dokładnie jeden raz. Dowolne rozwiązanie (trasę komiwojażera) można przedstawić jako permutację zbioru  $M$ :

$$a_1 a_2 a_3 \dots a_n \dots a_n \dots a_1$$

gdzie:

$$a_i \in M, \quad i = 1, 2, \dots, n,$$

$$\forall_{\substack{i \neq j \\ i=1,2,\dots,n \\ j=1,2,\dots,n}} a_i \neq a_j,$$

$$i=1,2,\dots,n$$

Rozwiązanie może więc zostać określone przez ciąg  $n$  znaków z  $n$ -elementowego alfabetu. Taki sposób zapisu oznacza, że nie każdy możliwy ciąg kodowy reprezentuje rozwiązanie dopuszczalne zadania. Istnieje  $n!$  wszystkich możliwych do utworzenia ciągów kodowych, podczas gdy tylko  $n!$  spośród nich daje rozwiązania dopuszczalne, tzn. takie, gdzie każdy znak (miasto) pojawia się w ciągu (na trasie) dokładnie jeden raz. O ile zmodyfikowanie procedury losowania wyjściowej populacji rozwiązań w taki sposób, żeby generowała jedynie dopuszczalne rozwiązania, nie stanowi problemu, o tyle wykonywanie na populacji typowych operatorów genetycznych krzyżowania i mutacji będzie już z pewnością prowadziło do powstawania rozwiązań niedopuszczalnych.

Teoria algorytmów genetycznych proponuje trzy podstawowe podejścia w opisywanej sytuacji:

- 1) **zastosowanie funkcji kary**, która będzie obciążała wartości funkcji przystosowania rozwiązań niedopuszczalnych, stymulując algorytm do samoczynnego kierowania się w stronę rozwiązań zgodnych z postawionymi warunkami dopuszczalności;
- 2) **zastosowanie procedur naprawczych**, czyli pomocniczych algorytmów dokonujących w możliwie efektywny sposób przekształcenia napotkanego rozwiązania niedopuszczalnego na jak najbliższe mu ideowo rozwiązanie dopuszczalne zadania;
- 3) **skonstruowanie specyficznych operatorów** o działaniu analogicznym do krzyżowania i mutacji, które dawałyby jednak zawsze jako wynik dopuszczalne rozwiązanie problemu, pod warunkiem, że ciągi kodowe „rodziców” reprezentują również rozwiązania dopuszczalne (operatory takie bazują na wiedzy specyficznej dla zadania).

W odniesieniu do zadania komiwojażera literatura algorytmów genetycznych proponuje kilka wariantów specyficznych operatorów.

#### 15.4.3.1. Operator krzyżowania PMX (partially mapped crossover)

Ciągi kodowe rodziców są rozcinane w dwóch losowo wybranych miejscach, po czym fragment „ścieżki” między punktami rozcięcia jest przenoszony „na krzyż” do ciągów potomnych. Pozostałe znaki ciągów potomnych są kopiowane wprost z odpowiednich ciągów rodzicielskich, jeżeli nie powoduje to niedopuszczalności ścieżki (pojawienia się dwukrotnie tego samego miasta na trasie). Brakujące po tym

etapie znaki w ciągu uzupełnia się na podstawie odwzorowania miast we fragmentach ścieżki znajdujących się między miejscami cięcia.

Rozpatrzmy krzyżowanie dwóch ciągów rodzicielskich reprezentujących przykładowe rozwiązania zadania komiwojażera z dziewięcioma miastami na trasie. Fragmenty ciągów kodowych leżące między wylosowanymi punktami rozcięcia zostały zacieniowane:

|       |   |   |   |   |   |   |   |   |   |
|-------|---|---|---|---|---|---|---|---|---|
| $R_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $R_2$ | 5 | 3 | 6 | 8 | 7 | 9 | 2 | 1 | 4 |

Etap 1. Wylosowane fragmenty tras przenosimy krzyżowo do ciągów potomnych:

|       |  |  |   |   |   |   |  |  |  |
|-------|--|--|---|---|---|---|--|--|--|
| $D_1$ |  |  | 6 | 8 | 7 | 9 |  |  |  |
| $D_2$ |  |  | 3 | 4 | 5 | 6 |  |  |  |

Etap 2. Przepisujemy między  $R_1$  a  $D_1$  i między  $R_2$  i  $D_2$  te znaki z ciągów kodowych, które nie powodują niedopuszczalności rozwiązania (powstania duplikatów):

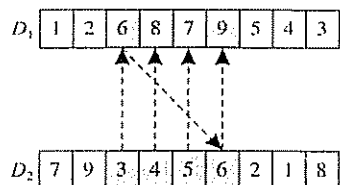
|       |   |   |   |   |   |   |   |   |  |
|-------|---|---|---|---|---|---|---|---|--|
| $D_1$ | 1 | 2 | 6 | 8 | 7 | 9 |   |   |  |
| $D_2$ |   |   | 3 | 4 | 5 | 6 | 2 | 1 |  |

Przykładowo, nie mogliśmy wpisać 7 na siódmym miejscu w ciągu  $D_1$ , gdyż znak 7 już w tym ciągu występuje, itd.

Etap 3. Uzupełniamy brakujące pola w ciągu  $D_1$ . Na pozycji siódmej powinniśmy wpisać 7 (z  $R_1$ ), ale 7 już występuje w  $D_1$ . Wobec tego zauważamy, że obecnie istniejąca 7 w  $D_1$  jest odwzorowana w 5 w ciągu  $D_2$  i właśnie taki znak wpisujemy w puste miejsce (odwzorowania są na zamieszczonym poniżej schemacie oznaczone przerywanymi lukami). Analogicznie, na ósmej pozycji nie możemy wpisać 8. Zamiast tego wpisujemy kod, w którym 8 jest odwzorowane, czyli 4. Na dziewiątej pozycji powinniśmy wpisać 9. Zauważmy, że 9 jest odwzorowane w 6, którego też nie możemy wpisać do  $D_1$ , gdyż taki znak już tam występuje. W takiej sytuacji szukamy kolejnego odwzorowania, tym razem dla kodu 6 i stwierdzamy, że jest nim 3 i ten właśnie kod wpisujemy w dziewiąte pole kodu  $D_1$ :

|       |   |   |   |   |   |   |   |   |   |
|-------|---|---|---|---|---|---|---|---|---|
| $D_1$ | 1 | 2 | 6 | 8 | 7 | 9 | 5 | 4 | 3 |
| $D_2$ |   |   | 3 | 4 | 5 | 6 | 2 | 1 |   |

Analogicznie, uzupełniamy pola w ciągu kodowym  $D_2$ . Na pierwszej pozycji powinniśmy wpisać kod 5 (z pierwszej pozycji  $R_2$ ), ale zamiast tego wpisujemy jego odwzorowanie, czyli 7. Na pozycji drugiej zamiast 3 wpisujemy 9 (po dwukrotnym odwzorowaniu), a na pozycji dziewiątej, zamiast 4 wpisujemy 8.



### 15.4.3.2. Operator krzyżowania CX (cycle crossover)

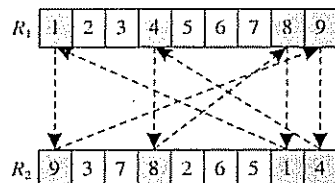
Działanie operatora można opisać następującymi krokami:

- Krok 1. Wylosuj pozycję w ciągu i przyjmij ją jako bieżącą.
- Krok 2. Przepisz znak z bieżącej pozycji z kodu  $R_1$  do  $D_1$  i z kodu  $R_2$  do  $D_2$ .
- Krok 3. Znajdź pozycję w kodzie  $R_1$ , na której znajduje się taki sam znak, jak na bieżącej pozycji w kodzie  $R_2$ .
- Krok 4. Czy znaleziona pozycja jest tą samą, od której rozpoczęto działanie w kroku 1? Nie: przyjmij znalezioną pozycję jako bieżącą i idź do kroku 2. Tak: idź do kroku 5.
- Krok 5. Uzupełnij brakujące znaki kodu  $D_1$ , przepisując je wprost z kodu  $R_2$ . Uzupełnij brakujące znaki kodu  $D_2$  przepisując je wprost z kodu  $R_1$ .

Rozpatrzmy dwa rozwiązania, które chcemy skrzyżować:

|       |   |   |   |   |   |   |   |   |   |
|-------|---|---|---|---|---|---|---|---|---|
| $R_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| $R_2$ | 9 | 3 | 7 | 8 | 2 | 6 | 5 | 1 | 4 |

Bez utraty ogólności możemy przyjąć, że w kroku 1 powyższej procedury wylosowano 1 jako wyjściową pozycję.



Na powyższym schemacie zaciemnione są te pozycje w ciągu, które pełniły rolę pozycji bieżących w kolejnych iteracjach procedury. Strzałki wskazują kolejność przechodzenia między pozycjami (począwszy od pozycji 1).

Po wykonaniu kroków 1–4 otrzymujemy poniższe potomne ciągi kodowe:

|       |   |  |  |   |  |  |  |   |   |
|-------|---|--|--|---|--|--|--|---|---|
| $D_1$ | 1 |  |  | 4 |  |  |  | 8 | 9 |
| $D_2$ | 9 |  |  | 8 |  |  |  | 1 | 4 |

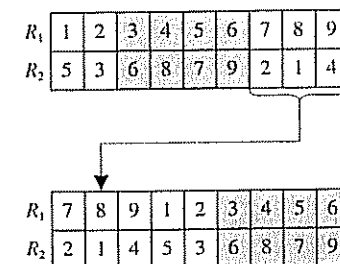
Brakujące pola są uzupełniane przez przepisanie krzyżowo wartości z ciągów rodzicielskich:

|       |   |   |   |   |   |   |   |   |   |
|-------|---|---|---|---|---|---|---|---|---|
| $D_1$ | 1 | 3 | 7 | 4 | 2 | 6 | 5 | 8 | 9 |
| $D_2$ | 9 | 2 | 3 | 8 | 5 | 6 | 7 | 1 | 4 |

### 15.4.3.3. Operator krzyżowania OX (order crossover)

Podobnie jak w przypadku operatora PMX wybieramy losowo fragment ciągu kodowego (losując dwa punkty cięcia).

Zauważmy, że w rozwiązaniu problemu komiwojażera istotna jest jedynie kolejność miast na trasie, a nie to, od którego miasta trasa się rozpoczyna, wobec czego bez zmiany rozwiązania możemy zmienić układ pozycji tak, aby wylosowany fragment pokrywał ostatnie pozycje ciągu:



Przepisujemy bez zmian wylosowany fragment z ciągu z  $R_1$  do  $D_1$  i z  $R_2$  do  $D_2$ .

|       |  |  |  |  |   |   |   |   |  |
|-------|--|--|--|--|---|---|---|---|--|
| $D_1$ |  |  |  |  | 3 | 4 | 5 | 6 |  |
| $D_2$ |  |  |  |  | 6 | 8 | 7 | 9 |  |

Brakujące pola w ciągu  $D_1$  uzupełniamy, przepisując kolejne kody z ciągu  $R_2$  i pomijając te, które już w  $D_1$  występują. Analogicznie, uzupełniamy ciąg kodowy  $D_2$ , wpisując kody z ciągu  $R_1$ .

|       |   |   |   |   |   |   |   |   |   |
|-------|---|---|---|---|---|---|---|---|---|
| $D_1$ | 2 | 1 | 8 | 7 | 9 | 3 | 4 | 5 | 6 |
| $D_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 7 | 9 |

### PRZYKŁAD 15.5 (cd. 15.1)

Przyjmijmy, że populacja w pojedynczym pokoleniu w algorytmie genetycznym będzie składała się z sześciu rozwiązań, których początkowa postać jest losowa. Tablica 15.12 zawiera przykładowo wylosowany zbiór sześciu rozwiązań wraz z odpowiadającymi tym rozwiązaniom długościami cyklu. Funkcja celu w zadaniu komiwojażera powinna być minimalizowana, dlatego — aby uzyskać funkcję przystosowania, która będzie rosnąca w miarę poprawy rozwiązania — zastąpmy długość cyklu jej odwrotnością. W tabelicy podano również obliczone wartości prawdopodobieństw wylosowania każdego z ciągów w pojedynczym losowaniu metodą ruletki, pomocnicze skumulowane prawdopodobieństwa oraz wartości oczekiwane liczby kopii poszczególnych rozwiązań w puli rodzicielskiej.

TABLICA 15.12  
Skład i charakterystyki rozwiązań w populacji początkowej

| Numer rozwiązania $i$ | Kolejność miast | Długość, $l_i$ | Funkcja przystosowania $f_i = \frac{1}{\text{długość}_i}$ | $p_i = \frac{f_i}{\sum_{j=1}^6 f_j}$ | $\text{cum}(i) = \sum_{j=1}^i p_j$ | $\mu_i = \frac{f_i}{f}$ |
|-----------------------|-----------------|----------------|---|--------------------------------------|------------------------------------|-------------------------|
| 1                     | (3, 4, 2, 1, 5) | 86             | 0,011628  | 0,175247                             | 0,175247                           | 1,051484                |
| 2                     | (4, 3, 5, 2, 1) | 130            | 0,007692  | 0,115933                             | 0,29118                            | 0,695597                |
| 3                     | (2, 3, 4, 1, 5) | 68             | 0,014706  | 0,221636                             | 0,512816                           | 1,329817                |
| 4                     | (3, 1, 4, 5, 2) | 127            | 0,007874  | 0,118671                             | 0,631488                           | 0,712028                |
| 5                     | (5, 1, 3, 4, 2) | 73             | 0,013699  | 0,206456                             | 0,837943                           | 1,238734                |
| 6                     | (5, 2, 4, 3, 1) | 93             | 0,010753  | 0,162057                             | 1                                  | 0,97234                 |

Przyjmijmy, że w wyniku sześciokrotnego wygenerowania liczby pseudolosowej z przedziału  $[0, 1)$  oraz zastosowania warunku o postaci (15.9) otrzymano poniższy zbiór rozwiązań stanowiących pulę rodzicielską (tablica 15.13).

Niech ramki wokół fragmentów ciągów kodowych rozwiązań oznaczają wylosowane fragmenty ciągów, które będą brały udział w krzyżowaniu według operato-

TABLICA 15.13  
Skład uzyskanej losowo puli rodzicielskiej

| Lp. | Kolejność miast | Koszt |
|-----|-----------------|-------|
| 1   | (2, 3, 4, 1, 5) | 68    |
| 2   | (5, 1, 3, 4, 2) | 73    |
| 3   | (5, 2, 4, 3, 1) | 93    |
| 4   | (3, 4, 2, 1, 5) | 86    |
| 5   | (3, 4, 2, 1, 5) | 86    |
| 6   | (2, 3, 4, 1, 5) | 68    |

TABLICA 15.14  
Nowa populacja uzyskana w wyniku krzyżowania PMX w poszczególnych parach puli rodzicielskiej

| Lp. | Kolejność miast | Koszt |
|-----|-----------------|-------|
| 1   | (5, 3, 1, 4, 2) | 132   |
| 2   | (2, 4, 3, 1, 5) | 93    |
| 3   | (5, 4, 2, 3, 1) | 56    |
| 4   | (3, 2, 4, 1, 5) | 114   |
| 5   | (2, 3, 4, 1, 5) | 68    |
| 6   | (3, 4, 2, 1, 5) | 86    |

ra PMX. Wtedy rozwiązania uzyskane z krzyżowania w parach (1, 2), (3, 4) oraz (5, 6) mają postać podaną w tabelicy 15.14.

Analiza tabelicy 15.14 pokazuje, że losowość zakorzeniona w algorytmie genetycznym może doprowadzić do uzyskania w nowym pokoleniu rozwiązania, które będzie gorsze niż którekolwiek z rozpatrywanych wcześniej (rozwiązanie pierwsze w tabelicy 15.14), ale też daje szansę na znalezienie rozwiązań lepszych niż znane dotychczas — w powyższym przypadku takim rozwiązaniem jest trasa prowadząca komiwojażera przez miasta (5, 4, 2, 3, 1) o łącznym koszcie wynoszącym 56. Naturalnie, zaprezentowane obliczenia stanowią ilustrację działania algorytmu genetycznego na przestrzeni zaledwie jednego pokolenia rozwiązań. Można oczekiwać, że satysfakcjonujące rezultaty „ewolucji” rozwiązań będą widoczne po wielokrotnym powtórzeniu opisanej procedury.

## POLECENIA I PYTANIA KONTROLNE

1. Co odróżnia metody heurystyczne od metod dokładnych?
2. Wymień i scharakteryzuj metody przyłączania wierzchołków w algorytmie włączania.
3. Opisz działanie heurystyki 2-optimalnej.
4. Wymień najważniejsze różnice między sztucznymi a naturalnymi mrówkami.
5. Podaj ogólny schemat algorytmu mrówkowego.
6. Opisz poszczególne kroki algorytmu genetycznego.
7. W jaki sposób losowanie puli rodzicielskiej metodą ruletki faworyzuje rozwiązania, które są lepsze niż inne?
8. Na podstawie jakich czynników może być podejmowana decyzja o zatrzymaniu procesu symulowania kolejnych pokoleń w algorytmie genetycznym?
9. W jaki sposób algorytm genetyczny może radzić sobie z problemem niedopuszczalności niektórych rozwiązań?

## BIBLIOGRAFIA

- Dorigo M., Stuetzle T., 2004, *Ant Colony Optimization*, MIT Press, Cambridge–Massachusetts–London.
- Goldberg D., 1998, *Algorytmy genetyczne i ich zastosowania*, WNT, Warszawa.
- Michalewicz Z., 1996, *Algorytmy genetyczne + struktury danych = programy ewolucyjne*, WNT, Warszawa.
- Rutkowska D., Piliński M., Rutkowski L., *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*, Wydawnictwo Naukowe PWN, Warszawa.
- Syso M.M., Deo N., Kowalik J.S., 1993, *Algorytmy optymalizacji dyskretnej*, Wydawnictwo Naukowe PWN, Warszawa.

## TABLICA

Dystrybuanta rozkładu normalnego  $N(0, 1)$ 

|     | 0      | 0,01   | 0,02   | 0,03   | 0,04   | 0,05   | 0,06   | 0,07   | 0,08   | 0,09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0   | 0,5000 | 0,5040 | 0,5080 | 0,5120 | 0,5160 | 0,5199 | 0,5239 | 0,5279 | 0,5319 | 0,5359 |
| 0,1 | 0,5398 | 0,5438 | 0,5478 | 0,5517 | 0,5557 | 0,5596 | 0,5636 | 0,5675 | 0,5714 | 0,5753 |
| 0,2 | 0,5793 | 0,5832 | 0,5871 | 0,5910 | 0,5948 | 0,5987 | 0,6026 | 0,6064 | 0,6103 | 0,6141 |
| 0,3 | 0,6179 | 0,6217 | 0,6255 | 0,6293 | 0,6331 | 0,6368 | 0,6406 | 0,6443 | 0,6480 | 0,6517 |
| 0,4 | 0,6554 | 0,6591 | 0,6628 | 0,6664 | 0,6700 | 0,6736 | 0,6772 | 0,6808 | 0,6844 | 0,6879 |
| 0,5 | 0,6915 | 0,6950 | 0,6985 | 0,7019 | 0,7054 | 0,7088 | 0,7123 | 0,7157 | 0,7190 | 0,7224 |
| 0,6 | 0,7257 | 0,7291 | 0,7324 | 0,7357 | 0,7389 | 0,7422 | 0,7454 | 0,7486 | 0,7517 | 0,7549 |
| 0,7 | 0,7580 | 0,7611 | 0,7642 | 0,7673 | 0,7704 | 0,7734 | 0,7764 | 0,7794 | 0,7823 | 0,7852 |
| 0,8 | 0,7881 | 0,7910 | 0,7939 | 0,7967 | 0,7995 | 0,8023 | 0,8051 | 0,8078 | 0,8106 | 0,8133 |
| 0,9 | 0,8159 | 0,8186 | 0,8212 | 0,8238 | 0,8264 | 0,8289 | 0,8315 | 0,8340 | 0,8365 | 0,8389 |
| 1   | 0,8413 | 0,8438 | 0,8461 | 0,8485 | 0,8508 | 0,8531 | 0,8554 | 0,8577 | 0,8599 | 0,8621 |
| 1,1 | 0,8643 | 0,8665 | 0,8686 | 0,8708 | 0,8729 | 0,8749 | 0,8770 | 0,8790 | 0,8810 | 0,8830 |
| 1,2 | 0,8849 | 0,8869 | 0,8888 | 0,8907 | 0,8925 | 0,8944 | 0,8962 | 0,8980 | 0,8997 | 0,9015 |
| 1,3 | 0,9032 | 0,9049 | 0,9066 | 0,9082 | 0,9099 | 0,9115 | 0,9131 | 0,9147 | 0,9162 | 0,9177 |
| 1,4 | 0,9192 | 0,9207 | 0,9222 | 0,9236 | 0,9251 | 0,9265 | 0,9279 | 0,9292 | 0,9306 | 0,9319 |
| 1,5 | 0,9332 | 0,9345 | 0,9357 | 0,9370 | 0,9382 | 0,9394 | 0,9406 | 0,9418 | 0,9429 | 0,9441 |
| 1,6 | 0,9452 | 0,9463 | 0,9474 | 0,9484 | 0,9495 | 0,9505 | 0,9515 | 0,9525 | 0,9535 | 0,9545 |
| 1,7 | 0,9554 | 0,9564 | 0,9573 | 0,9582 | 0,9591 | 0,9599 | 0,9608 | 0,9616 | 0,9625 | 0,9633 |
| 1,8 | 0,9641 | 0,9649 | 0,9656 | 0,9664 | 0,9671 | 0,9678 | 0,9686 | 0,9693 | 0,9699 | 0,9706 |
| 1,9 | 0,9713 | 0,9719 | 0,9726 | 0,9732 | 0,9738 | 0,9744 | 0,9750 | 0,9756 | 0,9761 | 0,9767 |
| 2   | 0,9772 | 0,9778 | 0,9783 | 0,9788 | 0,9793 | 0,9798 | 0,9803 | 0,9808 | 0,9812 | 0,9817 |
| 2,1 | 0,9821 | 0,9826 | 0,9830 | 0,9834 | 0,9838 | 0,9842 | 0,9846 | 0,9850 | 0,9854 | 0,9857 |
| 2,2 | 0,9861 | 0,9864 | 0,9868 | 0,9871 | 0,9875 | 0,9878 | 0,9881 | 0,9884 | 0,9887 | 0,9890 |
| 2,3 | 0,9893 | 0,9896 | 0,9898 | 0,9901 | 0,9904 | 0,9906 | 0,9909 | 0,9911 | 0,9913 | 0,9916 |
| 2,4 | 0,9918 | 0,9920 | 0,9922 | 0,9925 | 0,9927 | 0,9929 | 0,9931 | 0,9932 | 0,9934 | 0,9936 |
| 2,5 | 0,9938 | 0,9940 | 0,9941 | 0,9943 | 0,9945 | 0,9946 | 0,9948 | 0,9949 | 0,9951 | 0,9952 |
| 2,6 | 0,9953 | 0,9955 | 0,9956 | 0,9957 | 0,9959 | 0,9960 | 0,9961 | 0,9962 | 0,9963 | 0,9964 |
| 2,7 | 0,9965 | 0,9966 | 0,9967 | 0,9968 | 0,9969 | 0,9970 | 0,9971 | 0,9972 | 0,9973 | 0,9974 |
| 2,8 | 0,9974 | 0,9975 | 0,9976 | 0,9977 | 0,9977 | 0,9978 | 0,9979 | 0,9979 | 0,9980 | 0,9981 |
| 2,9 | 0,9981 | 0,9982 | 0,9982 | 0,9983 | 0,9984 | 0,9984 | 0,9985 | 0,9985 | 0,9986 | 0,9986 |
| 3   | 0,9987 | 0,9987 | 0,9987 | 0,9988 | 0,9988 | 0,9989 | 0,9989 | 0,9989 | 0,9990 | 0,9990 |

## Indeks

- Algorytm
  - Bellmana 75
  - BG 79
  - BMEP 75
  - Busackera-Gowena 79
  - FFEK 66
  - Forda-Fulkersona 66
  - Johnsona 212
  - Little'a 208
  - mrówkowy 343
  - programowania leksykograficznego 101
  - z relaksacją hierarchizacji ważności celów 102–103
- Analiza
  - czasowa przedsięwzięcia 162
  - czasowo-kosztowa przedsięwzięcia 175
  - postoptymalizacyjna 38
  - przedziałowa 39
  - wektorowa 38
  - wrażliwości 38
  - zasobowa przedsięwzięcia 183
- AOA (*Activities-on-arcs*) 158
- AON (*Activities-on-nodes*) 159
- Baza 30
- Bias 317
- Całkowity zapas czasu czynności 166
- Cel
  - główny 90
  - mierzalny 84
  - poboczny 90
- Ciąg kodowy 353
- CPM (*Critical Path Method*) 162
- Czas trwania czynności 163
- Czynność
  - fikcyjna 159
  - krytyczna 167
  - pozorna 159
- Decydent 15
- Decyzja
  - dopuszczalna 15
  - optymalna 15
- Decyzje nieporównywalne 85
- Degeneracja rozwiązania bazowego 35
- Diagram Hassego 87
- Drzewo
  - decyzyjne (budowa) 239
  - podziału 205
- Dywersyfikacja inwestycji w papiery wartościowe 298
- Efekt
  - komplementarności 117
  - substytucji 117
- Ekonomiczna wielkość zamówienia 279
- Ekstremum 119
- EOQ 279
- Etykieta 61

- Filtracja
  - rozpinająca 105
  - zacieśniająca 105
- Funkcja
  - aktywacji neuronu 314
  - kary 135
  - Lagrange'a 124
  - ograniczająca 203
  - przystosowania 351
  - ściśle
    - wklęsła 120
    - wypukła 120
  - U-kształtna 123
  - użyteczności 91
    - inwestora 206
  - wklęsła 120
  - wypukła 120
- Gra 264
  - dwuosobowa o sumie zero 265
- Gradient funkcji 121
- Graf 61
- Graniczny czas trwania czynności 176
- Harmonogram Gantta 169
- Hesjan 121
- Heurystyka 332
  - k- optymalna 339
  - lokalnych poszukiwań 333
  - włączania 333
- Inteligencja roju 343
- Jednokierunkowa sieć neuronowa 319
- Kalibracja algorytmu 344
- Kanał obsługi 230
- Kierunek
  - dopuszczalny 136
  - spadku (poprawy) 130
- Kombinacja wypukła 105
- Kompresja sieci 174
- Koszt funkcjonowania systemu zapasów 278
- Kowariancja stóp zwrotu 298
- Krawędź sieci 61
- Kres
  - dolny 202
  - górnym 202
- Krok (długość) 130
- Krótką sprzedaż 309
- Kryteria cząstkowe 84
- Kryterium
  - wyboru 15
  - zatrzymania algorytmu 356
- Krzywa
  - użyteczności decydenta 253
  - czasowo-kosztowa przedsięwzięcia 180
- Krzyżowanie 355
- Linia kombinacji 302
- Liniowa funkcja aktywacji neuronu 315
- Liniowe zadanie decyzyjne (LZD) 17
- Lista tabu 344
- Luz dla zdarzeń 168
- Luk sieci 61
- Macierz
  - określona
    - dodatnio 123
    - ujemnie 123
  - dochodów 222
  - incydencji 160
  - powiązań 160
  - strat 227
  - unimodularna 191
- Maksymalizacja
  - premier za ryzyko 308
  - użyteczności inwestora 306
- Marszruta 197
- Metakryterium 92
- Metoda
  - Charnesa-Coopera 138
  - ekstremów lokalnych 153
  - geometryczna 27
  - heurystyczna 332
  - Newtona-Raphsona 133
  - odcięć 190
  - podziału i ograniczeń 190

- potencjałów 45
- programowania dynamicznego 144
- ruletki 354
- simpleks 29
- spadku gradientu 320
- sztucznej inteligencji 343
- ścieżki krytycznej 162
- wstecznej propagacji błędów 325
- z regulowaną długością kroku 133
- zysków (dochodów, kosztów) krańcowych 154
- Metody
  - analityczne 128
  - bezgradientowe 129
  - gradientowe 130
  - iteracyjne 129
  - kierunków dopuszczalnych 136
  - numeryczne 128
  - wyspecjalizowane 129
- Minimum
  - globalne
    - niewłaściwe 118
    - właściwe 119
  - lokalne
    - niewłaściwe 118
    - właściwe 118
- Mnożniki Lagrange'a 124
- Model
  - matematyczny 16
  - stochastyczny 216
  - Wilsona 279
- Mrówki
  - elitarnie 344
  - z rangami 344
- Mutacja 355
- Najkrótszy czas realizacji przedsięwzięcia 163
- Najpóźniejszy dopuszczalny moment zajścia zdarzenia 164
- Najwcześniejszy możliwy moment zajścia zdarzenia 164
- Następnik węzła 62
- Neuron
  - ukryty 312
  - wejściowy 312
  - wyjściowy 312
- Niezależny zapas czasu czynności 167
- Normalizacja 92
- Oczekiwany czas realizacji czynności 170
- Odchylenie
  - rzeczywistego momentu zakończenia inwestycji od momentu oczekiwanego 172
  - standardowe stopy zwrotu 297
- Odległość euklidesowa 103
- Ograniczenia aktywne 126
- Operatory genetyczne 355
  - specyficzne dla zadania 360
- Optymalność w sensie Pareto 86
- Paradygmat liniowości 115
- Parametr 16
- PERT (*Program Evaluation and Review Technique*) 170
- Pocisk Markowitza 303
- Podjmowanie decyzji w warunkach ryzyka 216
- Podzbiór
  - aktywny 202
  - bierny 202
  - perspektywiczny 202
- Polityka zarządzania zapasami 278
- Poprzednik węzła 62
- Populacja rozwiązań 351
- Portfel
  - efektywny 305
  - papierów wartościowych 300
- Postać
  - bazowa 31
  - kanoniczna 21
  - standardowa 21
- Postulowana wartość funkcji celu 91
- Poziom
  - aspiracji 90
  - odnowienia 278
- Prawdopodobieństwo zakończenia projektu w dowolnym zadanym czasie 173



- Premia za ryzyko 308  
 Problem  
   decyzyjny 15  
   komivożażera 359  
   Lagrange'a 124  
   niesprzeczny 119  
   optymalizacji 118  
     bezwarunkowej (bez ograniczeń) 118  
     nieliniowej 118  
     warunkowej (z ograniczeniami) 118  
     wektorowej 86  
   programowania celowego 97  
     hiperbolicznego 137  
     kwadratowego 123  
     wypukłego 122  
   wektorowego programowania liniowego 105  
   wyboru optymalnego asortymentu produkcji  
     z nieliniową funkcją celu 116  
     w przypadku wielu kryteriów oceny 99  
     satisfakcjonujących poziomów kryteriów 90  
     z maksymalną funkcją celu 140  
 Procedury naprawcze 360  
 Progowa funkcja aktywacji neuronu 314  
 Przedsięwzięcie 157  
 Przekrój  
   podsieci krytycznej 178  
   sieci 63  
 Przepustowość  
   przekroju 63  
   ścieżki powiększającej 64  
 Przestrzeń  
   decyzyjna 84  
   kryterialna 84  
 Przeuczenie sieci neuronowej 328  
 Puła rodzicielska 352  
 Punkt  
   idealny 103  
   regularny 126  
   siodłowy 268  
   startowy 129  
 Reguła  
   Bayesa 240  
   Hurwicza 259  
   losowego sąsiada 334  
   maksymalizacji oczekiwanej użyteczności 251  
   maksymalizacji wartości oczekiwanej 249  
   najbliższego sąsiada 334  
   najdalszego sąsiada 334  
   racjonalnego pozyskiwania informacji 247  
   Savage'a 260  
   Walda 258  
 Rekurencyjna sieć neuronowa 319  
 Relacja  
   preporządku 85  
   racjonalna 91  
 Rozkład prawdopodobieństwa stopy zwrotu 294  
 Rozwiązanie  
   bazowe 30  
   dopuszczalne 21  
   lokalne najlepsze 203  
   niezdominowane 86  
   optymalne 21  
   Pareto — optymalne 86  
   początkowe 129  
   sprawne 86  
   niebazowe 105  
   wektorowo optymalne 86  
 Równowaga Nasha 271  
 Sieć 61  
   czynności 158  
   rezydualna 64  
 Sigmoidalna funkcja aktywacji neuronu 315  
 Skala  
   interwałowa 86  
   porządkowa 86  
 Skłonność do ryzyka 306  
 Standaryzacja macierzy odległości 208  
 Stopa  
   obsługi 230  
   przybycia 230  
   zwrotu 293  
   wolna od ryzyka 308

- Stopień realizacji celu  
   I rodzaju 92  
   II rodzaju 93  
 Strategia  
   czysta 258  
   dominująca 267  
   mieszana 262  
   zdominowana 266  
 Strumień  
   obsługi 230  
   zgłoszeń 230  
 Swobodny zapas czasu czynności 166  
 Sygnał błędu neuronu 326  
 Syntetyczny miernik satysfakcji 93  
 Szlak feromonowy 343  
 Ścieżka 163  
   krytyczna 167  
   pełna 163  
   powiększająca 64  
 Tablica  
   decyzyjna mrówek 345  
   simpleksowa 32  
 Technika oceny i kontroli programu 170  
 Twierdzenie  
   Bayesa 248  
   Kuhna-Tuckera 127  
   Weierstrassa o istnieniu 119  
 Uczenie  
   bez nauczyciela 328  
   z nauczycielem 320  
 Udział wartościowy papieru wartościowego w portfelu 300  
 Ujście 61  
 Użyteczność 251  
 Waga 61  
 Wariancja  
   najpóźniejszego momentu zajścia zdarzenia 173  
   najwcześniejszego momentu zajścia zdarzenia 172  
   stopy zwrotu 296  
 Warstwa neuronów 316  
 Wartość oczekiwana stopy zwrotu 294  
 Warunki Kuhna-Tuckera  
   dostateczne 127  
   konieczne 127  
 Warunek regularności 126  
   Karlina 126  
   liniowej niezależności 126  
   Slatera 126  
 Warunkowy zapas czasu czynności 167  
 Ważona suma stopni realizacji celów 92  
 Wektor  
   niezdominowany 86  
   ocen 85  
   wag 91  
   wzorzec 97  
 Wektorowe programowanie liniowe 10  
 Węzeł sieci 61  
 Wielowarstwowa sieć neuronowa 318  
 Wierzchołek sieci 61  
 Własność Markowa 145  
 Współczynnik  
   korelacji stóp zwrotu 299  
   uczenia sieci neuronowej 321  
 Wykres Gantta 169  
 Wzór  
   analityczny 223  
   rekurencyjny 222  
 Zadanie  
   decyzyjne 16  
   dualne (ZD) 23  
   programowania  
     całkowitoliczbowego liniowego (PCL) 189  
     liniowego (PL) 21  
     mieszanego liniowego (PML) 189  
     binarnego liniowego (PBL) 189  
 Zagadnienie  
   gazeciarza 221  
   komivożażera 197  
   pośrednika 50  
   przydziału 190  
   rozwózkowo-przewozowe 197  
   transportowe z kryterium czasu 57  
     z ograniczoną przepustowością tras 53

- Zmienna swobodna 22  
Zamknięte zagadnienie transportowe 44  
Zapis binarny 353  
Zasada optymalności Bellmana 144  
Zbiór  
  aktywnych ograniczeń 126  
  ocen 84  
  portfeli  
    opuszczalnych 303  
    efektywnych 305  
  zbior rozwiązań dopuszczalnych 118  
  wypukły 120  
  zwarty 84
- Zdarzenie krytyczne 169  
Zestaw  
  uczący 329  
  walidacyjny 329  
Zmienna  
  bazowa 30  
  decyzyjna 16  
  losowa 294  
  niebazowa 30  
  swobodna 22  
  sztuczna 33
- Źródło 61

Szkoła Główna Gospodarstwa Wiejskiego  
w Warszawie  
BIBLIOTEKA GŁÓWNA IM. WŁADYSŁAWA GRABSKIEGO  
STUDENCKA WYPOŻYCZALNIA PODRĘCZNIKÓW  
02-787 Warszawa, ul. Nowoursynowska 161  
tel. 022 59-35723 (5-9)