О. В. Ульянченко

Дослідження в економіці

Олександр Вікторович Ульянченко – проректор, завідувач кафедри економічної кібернетики Харківського національного аграрного університету ім. В. В. Докучаєва, фахівець в галузі використання математичних методів в економіці.

Народився у 1954 р. у селищі Покотилівка на Харківщині. У 1977 р. закінчив економічний факультет Харківського сільськогосподарського інституту. Працював науковим співробітником лабораторії економічних досліджень. У 1980 - 1983 рр. навчався в очній аспірантурі Всесоюзного науково-дослідного інституту економіки сільського господарства під керівництвом відомого вченого, професора І.Г. Ушачова. Після успішного захисту кандидатської дисертації повернувся до рідного вузу на кафедру економічної кібернетики, де працює вже майже 20 років, пройшовши шлях від асистента до професора, завідувача кафедри. Математичну підготовку одержав на вечірньому відділенні механіко-математичного факультету Харківського національного університету ім. В. Н. Каразіна.

О. В. Ульянченко є автором численних наукових праць з проблем управління ефективним використанням агроресурсів, а також навчальних посібників: "Сучасні моделі дослідження операцій в економіці", "Методи оптимізації в економіці", "Інформаційні системи та технології в аграрному менеджменті" (у співавторстві з П. А. Щербаковим), "Інформатика та комп'ютерна техніка: програмне забезпечення ЕОМ" (у співавторстві з П. А. Щербаковим, Н. М. Мартьяновою, Т. А. Бутенко), "Математичне програмування" (у співавторстві з М. Т. Лебедем, Г. Г. Хлівняком, В. О. Бабенко). Одним із перших учених економістів-аграрників обґрунтував доцільність відродження в економічній аграрній освіті курсу "Дослідження операцій в економіці". Разом з фахівцями кафедри економічної кібернетики НАУ під керівництвом професора В. А. Кадієвського розробив типову програму цього курсу та підготував підручник, у якому розглянуто не тільки традиційні теми дослідження операцій, а й нові, пов'язані з розвитком ринкових відносин в економіці.

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

О. В. Ульянченко

ДОСЛІДЖЕННЯ ОПЕРАЦІЙ В ЕКОНОМІЦІ

Підручник для студентів вузів

Харків
«Гриф»
2002

ББК 65вбя73
У51
УДК 519.8(075.8)

Рецензенти : *Путятін В. П.*, академік Української академії інформатики, завідувач кафедри економічної кібернетики ХДТУСГ;
Кадієвський В. А., д-р екон. наук, завідувач кафедри економічної кібернетики НАУ;
Сизоненко В. Л., докт. фіз.-мат. наук, професор, завідувач кафедри вищої математики та фізики ХНАУ.

Рекомендовано Міністерством освіти і науки України як підручник для студентів вищих навчальних закладів (протокол № 14/18.2-889 від 3 червня 2002 р.)

Ульянченко О. В.
У51 Дослідження операцій в економіці: Підручник для студентів вузів/ Харк. нац. аграр. ун-т ім. В. В.Докучаєва. – Харків: Гриф, 2002. - 580 с.

ISBN 966-7165-58-2.

У підручнику викладено теоретичні основи та принципи побудови моделей дослідження операцій в економіці, розглянуто їх структуру й основні компоненти. Значну увагу приділено основним поняттям, принципам і методиці дослідження операцій, розбудові виробничих функцій, прогнозуванню економічних показників, методам лінійного та нелінійного програмування, планування й управління мережами, математичному моделюванню процесів масового обслуговування, управління запасами, імітаційному моделюванню, експертним оцінкам у менеджменті, використанню пакетів прикладних програм у дослідженні операцій.

Для студентів економічних спеціальностей вищих закладів освіти III-IV рівнів акредитації.

ББК 65вбя73

© Харківський національний аграрний університет, 2002
© О. В. Ульянченко, 2002
© ВКФ "Гриф", 2002

ISBN 966-7165-58-2

ВІД АВТОРА

Видані раніше навчальні посібники "Сучасні моделі дослідження операцій в економіці" (2000 р.) та "Методи оптимізації в економіці" (2001 р.) пройшли апробацію у ряді вищих навчальних закладів України. Пропонований підручник суттєво доповнює ці видання. Автор узагальнив власний досвід читання лекцій і проведення практичних занять з курсу "Дослідження операцій в економіці" для студентів економічного факультету Харківського національного аграрного університету ім. В.В. Докучаєва.

Планування багатьох техніко-економічних процесів ґрунтується на різноманітних нормативах, базах знань, використання яких дозволяє оцінювати перебіг цих процесів і визначати оптимальні їх напрямки. Тому з урахуванням вимог програми курсу розглянуто окремі методи оптимізації, які широко й ефективно використовуються в плануванні виробництва та в економіці. У підручнику приділено увагу використанню лінійного, нелінійного та цілочислового програмування. Необхідні відомості з курсу вищої математики стисло наводяться при викладанні відповідних тем. Розглядаються також методи розбудови виробничих функцій та їх значення в моделюванні виробничих процесів при їх економічній оцінці, модулі управління складними організаційними системами та питання оптимізації розміщення коштів у виробництво, прогнозування економічних показників з використанням часових рядів, математичне моделювання процесів масового обслуговування, імітаційне моделювання операційних систем, дослідження інвестиційних проектів, експертні оцінки в менеджменті, діагностика банкрутства, використання пакетів прикладних програм тощо.

Автор впевнений, що з випуском підручника коло проблем зазначеного курсу розшириться. У ринкових умовах ефективне господарювання й управління виробничою діяльністю на будь-якому рівні в будь-якій галузі не можливі без використання сучасних методів планування й організації, основою яких є відповідний математичний апарат. Поєднання математичних методів та економічного аналізу відкриває нові можливості для економічної науки та практики.

Все це безумовно сприятиме високій фаховій підготовці спеціалістів і гарантуватиме їм подальше практичне використання знань і навичок у професійній діяльності по закінченні вузу.

Автор завдячує академіку, доктору технічних наук, професору Путятіну В.П., доктору економічних наук, професору Кадієвському В.А., доктору фізико-математичних наук, професору Сизоненку В.Л., які висловили цінні пропозиції при рецензуванні підручника, колегам з кафедри економічної кібернетики ХНАУ, редактору підручника Рамакаєвій С.О., кандидатам технічних наук доценту Хлівняку Г.Г., Бабенко В.О. та викладачу Бутенко Т.А. за надання цінних порад і практичної допомоги при підготовці рукопису до набору.

Автор буде вдячний читачам, які надішлють свої відгуки, зауваження та пропозиції на адресу Харківського національного аграрного університету ім. В.В. Докучаєва.

Сучасний рівень вимог до економічної теорії та практики зобов'язує фахівців цього профілю постійно знайомитися з передовими ідеями модельної структуризації й аналізу. В останні роки велике значення в економічних дослідженнях набули математичні методи. При розв'язанні багатьох практичних задач використовуються методи лінійного та нелінійного програмування, велику роль в економічному моделюванні відіграють методи планування й управління мережами з використанням засобів лінійної алгебри, аналіз функцій однієї та багатьох змінних і деякі інші розділи математики. Підвищується значимість динамічного програмування, теорії гри в розв'язанні господарських задач. Значне місце в економічних дослідженнях займає імітаційне моделювання, експертні оцінки тощо. З цих питань є багато спеціальної літератури, вивчення якої вимагає поглибленої підготовки.

Дослідження операцій є комплексною науковою дисципліною, яка має важливе методологічне значення у системі підготовки сучасного фахівця економічного профілю. Ця дисципліна найбільш чітко реалізує одну з основних ідей прикладної математики – ідею математичного моделювання економічних процесів. У даному підручнику автор намагається висвітлити основні розділи курсу, які використовуються при аналізі економічних систем. Викладення матеріалу характеризується достатньою простотою, що дозволяє залучити до математичних методів широке коло економістів.

Автор підручника – відомий фахівець у галузі економіко-математичних методів і моделей. Він одним із перших обґрунтував необхідність відновлення курсу "Дослідження операцій" для економічних спеціальностей у вузах аграрного профілю. Його навчальні посібники "Сучасні моделі дослідження операцій" та "Методи оптимізації в економіці" стали базовими при викладанні цієї дисципліни.

Підручник містить чимало доповнень, які стосуються поглиблення зазначеної навчальної дисципліни. Так, вперше розглянуто питання методики дослідження операцій, прогнозування економічних показників з використанням часових рядів, математичних моделей управління запасами, математичного моделювання процесів масового обслуговування, імітаційного моделювання операційних систем, а також зовсім нові питання: дослідження інвестиційних проектів, експертні оцінки в менеджменті, діагностика банкрутства, використання пакетів прикладних програм у дослідженні операцій. Це виводить підручник за межі програми зазначеного курсу. Завдяки цьому він може бути використаний в процесі економічної підготовки фахівців різних галузей народного господарства.

Пропонований підручник – це перша спроба визначити в повному обсязі тематику курсу та поширити його впровадження в навчальні плани з економічних спеціальностей вищих навчальних закладів освіти.

О. М. Онищенко,
академік НАН України,
доктор економічних наук,
професор

ВСТУП

Розвиток сучасного суспільства досяг того рівня, коли виникає нагальна потреба в розробці ефективних методів управління організаційними системами різного призначення та різних рівнів. Прикладами таких систем є окремі виробництва, галузі господарства, структури управління (військові, державні), господарські комплекси і т. ін.

Метою створення організаційних структур або систем та їх функціонування є задоволення певних запитів суспільства, тому природно виникає питання про якісні та кількісні оцінки як самих систем, так і впливу суспільства на реалізацію їх цільового призначення.

Виконання безпосередніх емпіричних досліджень відповідних процесів та явищ практично не можливе з багатьох причин: вартості досліду, певної його унікальності (неповторності), обмежень у часі, неможливості визначити результати впливу тих чи інших чинників і т. ін. З урахуванням цих обставин емпіричні дослідження заміняються дослідженнями відповідних математичних моделей.

Об'єктом наукової дисципліни "Дослідження операцій в економіці" є аналіз функціонування виробничо-господарських систем і розробка методів оптимального управління ними з використанням відповідних математичних моделей. Вирішення цих проблем досягається системним, всебічним вивченням процесів у досліджуваних системах, синтезом якісних досліджень і певного математичного апарату в поєднанні з широкими можливостями сучасних ЕОМ.

Під терміном "операція" розуміють будь-яку сукупність заходів, спрямованих на дослідження певної мети. Дослідження функціонування структури або системи можна виконувати різними способами. Наша мета – описати основні засади дослідження систем з використанням їх математичних моделей і розбудови таких моделей.

Основні поняття теорії математичного моделювання. Слово "модель" (від фр. modele – зразок) в літературі має досить широкий діапазон тлумачення. Ми розглядаємо основні моделі досліджуваних явищ і процесів в системах – математичні моделі. У літературі наводяться різні визначення поняття математичної моделі (ММ). У кожному з них наголошується на певних особливостях і цільових призначеннях ММ. Реальні соціально-економічні процеси досить складні, їх поточний стан і перебіг в часі визначаються сукупністю багатьох чинників (факторів). Прагнучи розбудувати ММ, ми з необхідністю мусимо розглянути кількісні виміри діючих чинників та описувати їх взаємозв'язки аналітичними залежностями. Чинники, вибрані для розбудови ММ за умови використання їх кількісної міри, називаються *параметрами* моделі.

Залежно від підходу до класифікації параметрів виділяють параметри *стану* та *управління*, *екзогенні* та *ендогенні*. Параметри, які використовуються для

відображення фізичного втілення досліджуваного процесу та зовнішніх умов його перебігу, називаються *екзогенними*. Їх величини визначаються поза розбудованою ММ, їх значення дозволяють виділяти споріднені системи. Змінні, величини яких обчислюються з використанням ММ, називаються *ендогенними*. Наприклад, при розбудові ММ функціонування окремого підприємства вартості використовуваних ресурсів будуть екзогенними параметрами, а їх ефективність – ендогенними. Виробничі потужності підприємства в ММ характеризуються параметрами стану, а їх завантаження в процесі виробництва визначається параметрами управління.

У подальшому будемо дотримуватися такого використання термінів: якщо маємо в ММ величини, які виділяють її з класу однорідних за призначенням, то за такими величинами залишаємо термін "параметри"; якщо величини використовуються для дослідження можливих станів системи, то будемо називати їх змінними.

Під ММ будемо розуміти сукупність математичних залежностей, за допомогою яких описується взаємозв'язок між параметрами та змінними, вибраними для дослідження властивостей об'єкта або процесу. Процес розбудови ММ називається *моделюванням*. При математичному моделюванні будемо користуватися аналітичним апаратом теорії функції кількох змінних, теорії ймовірностей та математичної статистики.

Переважна кількість проблем і задач практичної діяльності пов'язана з керованими процесами, тобто процесами, перебіг яких певною мірою обумовлений прийнятими рішеннями та заходами по їх втіленню. Обґрунтуванню рішень передує аналіз умов функціонування досліджуваного об'єкта або процесу, в якому можна виділити чотири етапи.

Перший етап – якісні дослідження взаємозв'язків з метою пошуку чинників, які обумовлюють закономірність перебігу процесу, вибір множини параметрів стану, управління, екзогенних та ендогенних. На цьому етапі необхідна співпраця фахівців відповідної галузі та математиків.

Другий етап – розбудова математичної моделі досліджуваної проблеми: в математичних поняттях необхідно описати стан об'єкта або процесу, вимоги, яким мають задовольняти величини використаних параметрів, та сформулювати так звану функцію цілі.

Витрачаючи ресурси та втілюючи відповідні заходи при використанні структури або системи певного цільового призначення, тобто реалізуючи операцію по досягненню певної мети, природно дати оцінку операції. А оскільки розбудовуємо ММ операцій, то оцінка повинна мати кількісну міру. Оцінки операцій необхідні для визначення ефективності операцій, тобто оцінки міри їх пристосування для втілення мети функціонування відповідної структури або системи. Кількісний вимір якості оцінки операцій досягається розбудовою та

використанням так званої *цільової функції*, яка є втіленням *критерію ефективності* операції. Використання цільової функції дає можливість оцінити якість операції за певним критерієм та одержати рекомендації щодо управління процесом. Критерієм ефективності організації виробництва може бути цільова функція – прибуток після реалізації продукції.

У теорії дослідження операцій в економіці розроблено ефективні методи планування та виконання відповідних розрахунків за умови відомих цілей управління та критеріїв оцінки якості можливих планів досягнення цих цілей. Труднощі в конкретних практичних ситуаціях обумовлені тим, що не лише критерій, а й мету управління часто досить складно визначити.

Практично не можна одним критерієм оцінити функціонування підприємства або галузі. Як правило, критеріїв існує чимало, і вони суперечать один одному за змістом, наприклад, собівартість і рівень якості виробів, обсяги за валом і розмаїття номенклатури. Наведемо відомий в літературі приклад. У 1960-1961 рр. ОЦ АН СРСР виконував роботи по впровадженню економіко-математичних методів у планування й управління роботою будівельних організацій Москви. Було розроблено програму організації транспортування будівельних матеріалів. Щосуботи результати планування подавалися у формі щоденних накладних для кожної машини обслуговуючих автопідприємств. По завершенні кварталу керівна інстанція оцінила роботу автопідприємств як *незадовільну*. Чому? План перевезень був розроблений за критерієм мінімальних витрат з *обов'язковим виконанням усіх замовлень* будівельних організацій. Але керівна інстанція оцінювала роботу автопідприємств за іншими критеріями: так званими тонно-кілометрами, економією бензину на цих тонно-кілометрах, тільки не за основним показником – виконанням замовлень з мінімальними витратами.

Доцільно наголосити, що правильний вибір критерію якості плану управління можна зробити лише з урахуванням вимог, згідно з метою функціонування підприємства, галузі тощо. Вибір критерію (або їх сукупності) оцінки ефективності операцій часто визначає практичну доцільність досліджень.

Третій етап – розв'язання сформульованої екстремальної задачі. Обов'язковими компонентами ММ дослідження будь-якої операції є цільова функція та обмеження на величини змінних управління з урахуванням значень екзогенних параметрів. У загальних рисах задачу дослідження операцій в економіці можна сформулювати так. Для розбудованої ММ знайти такі значення змінних управління, які мають задовольняти наявним обмеженням на їх величини та за яких цільова функція досягає екстремального значення, обумовленого сенсом операції. Необхідно наголосити, що практичні задачі, як правило, вимагають розбудови функціональних залежностей зі значною кількістю змінних за умови обмежень на їх можливі значення. Обсяги обчислень настільки значні, що їх не можна реалізувати без використання ЕОМ.

Четвертий етап – експертна оцінка результатів розрахунків стосовно досліджуваних операцій управління об'єктом або процесом. Наголошуємо на двох можливих випадках. Якщо результати незадовільні, то уточнюють вхідну інформацію та розбудовану математичну модель і повторно виконують розв'язок задачі. Якщо результати співставлення розрахунків і практичних вимог мають сенс, то вирішують питання практичного використання як розбудованої ММ для розв'язання подібних задач, так і результатів розв'язку конкретної задачі.

На завершення доцільно наголосити, що реальні економічні та виробничі процеси досить складні, але недоцільно розбудовувати громіздкі ММ для врахування "усіх" можливих чинників. Практика розбудови та використання ММ свідчить, що необхідно враховувати лише основні, найважливіші чинники. Майстерність розбудови ММ визначається вмінням відбирати саме такі чинники, що вимагає високої фахової підготовки. До цього необхідно додати, що в математиці розбудовано методи, які дозволяють відбирати такі чинники та відповідні параметри ММ, оцінюючи міру їх взаємних зв'язків.

Використовувані моделі дослідження операцій в економіці можна класифікувати в залежності від призначення та змісту модельованої операції, сенсу розв'язуваної проблеми, використовуваного математичного апарату.

Математичне програмування. Доцільно виділити клас практично значимих *оптимізаційних* моделей з використанням аналітичного апарату функцій кількох змінних при розбудові ММ відповідної операції. Методи розв'язання відповідних задач об'єднані в розділ математики, який називається *математичним програмуванням.*

Моделювання операцій планування й управління, в яких доцільно користуватися математичним апаратом функцій багатьох змінних, незважаючи на їх розмаїття, має спільний стрижень. По-перше, має бути побудована цільова функція (функція цілі) – оцінка за певним критерієм якості кожного можливого варіанта реалізації плану управління перебігом досліджуваного процесу. По-друге, необхідно враховувати, що можливості впливу на перебіг будь-якого процесу або явища завжди обмежені обсягом наявних ресурсів та іншими чинниками. Математично такі обмеження формулюються як вимоги виконання системи певних рівнянь і нерівностей для можливих значень змінних, використовуваних у ММ. Загальна задача математичного програмування формулюється так: знайти такі значення n змінних, вибраних при розбудові ММ, які задовольняють певній системі рівнянь і нерівностей:

$$\varphi_i(x_1, x_2, \ldots, x_n)\,(\leq = \geq)\, b_i \qquad (i = \overline{1, m}), \qquad (*)$$

а функція цілі досягає екстремального значення:

$$z = f(x_1, x_2, \ldots, x_n). \qquad (**)$$

Функції φ_i та f мають бути заданими. За допомогою цих множин моделюють зовнішні умови функціонування системи. У кожній з умов (*) має бути лише одне з указаних відношень $\leq$, = або $\geq$. Величини n та m ніяк не пов'язані між собою, m може бути більшим, меншим або рівним n. Зокрема m може дорівнювати нулю, що засвідчує відсутність будь-яких обмежень на можливі значення змінних ММ. До деяких змінних або до всіх може бути поставлена вимога їх невід'ємності ($x_j \geq 0,\ j = \overline{1, n}$).

Набір чисел $(x_1, x_2, \ldots, x_n)$, які задовольняють умові (*), називається *планом* (втіленням програми управління системою) задачі математичного програмування. Не кожна система (*) має розв'язок, тому й не кожна задача має план (розв'язок).

План $(x_1^0, x_2^0, \ldots, x_n^0)$, за якого функція цілі (**) досягає екстремуму, називається *оптимальним*. Задача математичного програмування може мати не один оптимальний план.

Сформульована задача математичного програмування (*) і (**) має загальний вигляд, тому є занадто складною з точки зору розбудови методів її розв'язання.

У математичному програмуванні розрізняють два напрямки. До першого належать детерміновані моделі, в яких вся вихідна інформація приймається повністю і одночасно є визначеною в межах розбудованої ММ, до другого, так званого стохастичного математичного програмування, – моделі, в яких параметри мають елементи невизначеності, деякі з них уявляються випадковими величинами з відомими законами розподілу. Наприклад, ринкові ціни на продукцію, обсяги її реалізації та темп інфляції є випадковими величинами, що необхідно враховувати при розбудові ММ планування виробництва на значний проміжок часу.

За традицією з урахуванням використовуваного математичного апарату в математичному програмуванні виділяють такі основні розділи.

Лінійне програмування. Математична модель називається лінійною, якщо вона подається лише лінійними функціональними залежностями. Нагадаємо, що функція багатьох змінних називається лінійною, якщо вона є раціональною, тобто для її запису використовуються лише арифметичні дії, і змінні вжиті у першому степені при відсутності добутків однієї змінної на іншу. У лінійному програмуванні досліджуються різні класи задач з лінійними ММ.

Нелінійне програмування. При розбудові ММ використовуються нелінійні функціональні залежності. Оскільки методи розв'язання задач нелінійного програмування вимагають великого обсягу обчислень, то важливе значення мають спеціальні типи функціональних залежностей, використовуваних при розбудові ММ:

- *опукле програмування,* коли функція цілі та обмеження на величини

змінних у ММ подаються так званими *опуклими* функціями; специфічний характер функціональних залежностей істотно спрощує розбудову числових алгоритмів розв'язання відповідних задач;

- *квадратичне програмування,* коли функція цілі є квадратичною функцією змінних, а обмеження представлені лінійними рівняннями та нерівностями;
- *цілочислове та дискретне програмування,* коли розв'язок задачі має сенс лише за умови цілочислових або дискретних значень змінних.

Як зразок задачі стохастичного математичного програмування стисло охарактеризуємо вимоги до розбудови ММ задачі комплектування верстатного парку ремонтного підприємства. Протягом певного часу на підприємство надходять замовлення на виконання ремонтних робіт. Наперед не відомі ні час надходження замовлення, ні його обсяги та відповідна вартість. Але очевидним є те, що при недостатній кількості верстатів і майстрів для реалізації замовлень замовник може звернутися за послугами до інших підприємств, а ці матимуть збитки через втрату замовлень. З іншого боку, якщо парк обладнання значно розширити, то можливі значні простої обладнання, і підприємство, витративши кошти на комплектування, продовжуватиме нести збитки, обумовлені утриманням невикористовуваного обладнання.

За описаної ситуації очікуваний прибуток підприємства є випадковою величиною, а тому не можна говорити про її найбільше значення. За таких умов на практиці за цільову функцію вибирають або математичне сподівання величини прибутку, обчислюване з використанням відомих законів розподілу надходження замовлень (такий підхід дає можливість скористатися методом лінійного програмування), або ймовірність того, що очікуваний обсяг прибутку буде не менше заданої величини.

Моделі дослідження операцій при необхідності врахування часу. Розмаїття функціонування складних виробничих і соціально-економічних систем робить неможливим використання лише моделей математичного програмування. За своїм змістом, методикою розбудови та використовуваним математичним апаратом типові моделі дослідження відповідних операцій в економіці доцільно розподілити на такі класи.

Математичні моделі управління операціями, які доцільно представити як сукупність узгоджених у часі етапів з можливістю оцінок якості управління на кожному етапі, утворюють клас задач так званого динамічного програмування. Моделі цього класу поєднані як особливостями їх розбудови, так і методами їх розв'язання.

Суттєво важливими у функціонуванні економічних систем є *процеси управління запасами.* У літературі описані різні за своєю природою та призначенням моделі розв'язання задачі управління запасами. Як наукова проблема ця задача

виникла з практичних потреб організації постачання англійської армії під час другої світової війни.

Під терміном "запаси" розуміють не лише матеріальні цінності, а й виробничі потужності, трудові ресурси і т. ін. У задачі управління запасами досліджується співвідношення вартості витрат на постачання та зберігання запасів (ресурсів) зі збитками через їх відсутність при необхідності. Наслідком розв'язання відповідної задачі є рекомендації щодо обсягів і періодів поставок, обсягів страхових запасів і т. ін. Для розв'язання задачі управління запасами в залежності від її специфіки використовують математичний апарат лінійного, нелінійного, динамічного та стохастичного програмування.

Другий важливий клас задач дослідження операцій в економіці утворюють *процеси розподілу*, обумовлені завжди наявною певною обмеженістю ресурсів і засобів, необхідних для функціонування певної системи. Зміст задачі полягає в пошуку такого плану розподілу ресурсів, який би забезпечував результат, оптимальний за певним критерієм якості. Задачі розподілу розв'язуються засобами математичного програмування та методами планування мережі. До цього ж класу належать задачі *пошуку оптимального шляху* за певними критеріями: найменших витрат часу та коштів на переміщення, найкоротшого шляху за наявності певних обмежень тощо.

Важливими виробничими задачами є задачі ремонту, експлуатації та заміни обладнання. Зміст таких задач визначається розробкою рекомендацій стосовно термінів ремонту та заміни обладнання з урахуванням витрат на виконання відповідних операцій.

Математично подібними до названих є задачі заміни кадрів і робочої сили, але розв'язання таких задач вимагає широких соціологічних досліджень, обумовлених психолого-демографічними факторами.

При вирішенні проблем пошуку оптимальних розв'язків суттєво важливими є задачі *моделювання конфліктних і конкурентних процесів.* Принципово значимою властивістю таких задач є моделювання поняття оптимальності, яке в багатьох випадках визначає і використовуваний математичний апарат і практичне втілення розв'язку. При розв'язанні таких задач широко й ефективно використовується апарат теорії ігор у поєднанні зі статистичними методами з метою пошуку такого режиму поведінки конфліктуючих сторін, який би забезпечував максимальний виграш або мінімальний програш.

Важливим елементом задач указаних класів є розробка функцій, з допомогою яких за певних вхідних даних визначається стан об'єкта або процесу. Такі функції називаються виробничими. Методи їх аналітичного зображення визначаються цільовим призначенням цих функцій і досить ґрунтовно описані в літературі та в пропонованому підручнику.

Сучасна людина постійно користується так званими *системами масового*

обслуговування: телефонними станціями, касами, медичними закладами, підприємствами торгівлі і т. ін. Моделі систем масового обслуговування розбудовуються для аналізу основ проектування відповідних систем, дослідження ефективності їх функціонування та спроможності забезпечувати цільове призначення, наприклад, визначення терміну виконання замовлень необхідної кількості каналів обслуговування, появи черги, пропускної спроможності системи і т. ін.

Створення, функціонування та розвиток будь-якої виробничої системи вимагає інвестування. Оцінки доцільності та ефективності інвестування особливо значимі, якщо необхідне довгострокове інвестування за умов інфляції. У підручнику розглядаються деякі моделі оцінок ефективності інвестування.

За ринкових відносин, розбудовуваних в Україні, в організації економічної діяльності суб'єктів господарювання вирішальне значення мають їх фінансовий стан і фінансово-юридичні особливості стану банкрутства. Ці питання теж стисло розглядаються в підручнику. Знайомство з юридично-правовими нормами необхідно не лише банкруту, а й іншим суб'єктам господарювання, взаємопов'язаним у своїй фінансово-економічній діяльності.

Деякі методи контролю якості математичного моделювання при дослідженні операцій в економіці. Математичні моделі дослідження операцій в економіці, як і будь-яке моделювання, не можуть адекватно в повному обсязі віддзеркалювати реальні процеси – вони завжди мають до певної міри *формальний зміст*, тобто представлений математичними залежностями. Для визначення доцільності й ефективності рекомендацій, одержаних при розв'язанні відповідної ММ управління, перед їх втіленням необхідно врахувати думку досвідчених фахівців певної галузі, тобто виконати експертизу – неформальний аналіз результатів математичного моделювання. У підручнику описані методи проведення експертних оцінок і поєднання формальних та неформальних методів дослідження операцій в економіці. Імітаційне моделювання при використанні математичних моделей та експертних оцінок досліджуваних операцій дає можливість з допомогою комп'ютерів швидко, якісно, з високим рівнем надійності проаналізувати ймовірні ситуації в реалізації процесу управління та оцінити можливі наслідки.

Будь-які соціально-економічні процеси відбуваються через практичну діяльність людей, об'єднаних у певні колективи. При дослідженні суспільних процесів уже на найнижчому рівні виникають труднощі моделювання, обумовлені тим, що кожний член колективу має можливість в певних межах самостійно сприймати й оцінювати інформацію, приймати та реалізувати рішення.

За таких умов поведінка колективу визначається не лише його станом у певний момент та системою зовнішніх і внутрішніх взаємозв'язків, а й наслідками індивідуального прийняття рішень. Суттєвим є і той незаперечний факт, що на

особисті рішення індивідууму впливає не лише часто неповна або навіть недостовірна інформація, а й рівень розуміння ситуації, в якій приймається рішення. Якщо Японія демонструє високий рівень виробничих технологій, то не слід забувати, що понад триста років вона є країною суцільної грамотності, а сьогодні йдеться вже про суцільну вищу освіту. Щодо колишнього СРСР чимало соціологів і політиків вважають, що радянський варіант соціалізму був прикладом державного рабовласницького устрою на новій виробничій основі.

Труднощі, обумовлені моделюванням поведінки індивідуума, у багато разів підсилюються при спробах моделювання поведінки колективу. Неможливість розробити досить вірогідну модель такої поведінки навіть при наявності певної якості моделей поведінки індивідуумів є головною складністю моделювання соціальних систем, в яких завжди виявляється нетотожність інтересів осіб, груп, рівнів ієрархії суспільства. Проблеми моделювання поведінки певних колективів як одні з найважливіших, але майже недосліджених широко вивчаються представниками різних наукових напрямків.

Тому при дослідженні операцій наголос робиться в основному на економічних показниках ефективності функціонування системи, з посилкою, що успішний економічний стан є гарантією гарного соціального стану.

Історична довідка. Функціонування будь-якого колективу вимагає певних правил управління ним. Ця проблема, можна стверджувати без заперечень, була однозначно усвідомлена з виникненням такого суспільного об'єднання, як держава. Найстародавніші закони державного устрою зафіксовані клинописом біля п'яти тисяч років тому шумерським царем Ур Намі. Усі монотеїстичні релігії прагнули упорядкувати управління суспільством при наявності значних колективів з протилежними інтересами та складною ієрархією стосунків між ними.

Розвиток суспільно-виробничих відносин обумовив появу такої науки, як політична економія (майстерність управляти економікою держави). Засновником класичної політичної економії був У. Петті (1623-1687).

Перша у світі модель господарства в державі була створена французьким ученим Ф. Кене (1694-1774). Його знаменита "Економічна таблиця" стала основою розробки багатьох математичних моделей суспільного виробництва.

Важливу роль у розвитку економічної науки XIX століття відіграла так звана *математична школа в політичній економії*. Її найвизначніші представники – О. Курно, Г. Госсен, Я. Вальрас, У. Джеванс, Ф. Еджворт, В. Парето зробили значний вклад у розробку проблем споживання, попиту і пропозицій, збалансованості (рівноваги) розвитку економіки. У дореволюційній Росії наприкінці XIX століття були виконані оригінальні економіко-математичні дослідження В. К. Дмитрієва, В. І. Борткевича, Р. М. Орженецького та ін. В. К. Дмитрієв запропонував модель повних господарських витрат праці та збалансованих цін. Вагомий вклад у розробку економіко-математичного

моделювання зробив Є. Є. Слуцький (1880-1948). У 1930-1950 рр. в СРСР не було прогресу в розвитку економіко-математичних досліджень через ідеологічні утиски тоталітарного режиму. Як зразок ідеологічних збочень наведемо цитату з підручника, виданого в 1968 р.: "Возникновение эконометрии как особого направления буржуазной экономической науки вызвано ... стремлением капиталистических монополий к совершенствованию методов эксплуатации трудящихся. Буржуазные ученые придают ей огромное значение... На самом же деле теоретические основы эконометрии являются совершенно несостоятельными... Большинство ее математических формул и уравнений являются математическим выражением антинаучных и апологетических вульгарных политэкономических категорий"*. Економетрія розглядається як поєднання економічної теорії, статистики та математики. Термін "економетрія" був введений у науковий вжиток норвезьким ученим Р. Фрішем (1895-1973). Доцільно наголосити, що Нобелівські премії з економіки присуджувалися за дослідження у галузі економіко-математичного моделювання, переважно економетрії. Лауреатами Нобелівської премії свого часу стали Р. Фріш, Я. Тінберген, П. Самуельсон, Д. Хіке, К. Ерроу, В. Леонтьєв, Т. Купманс, радянський математик Л. В. Канторович.

Розвиток і дослідження виробничих зв'язків та соціальних відносин, практика господарювання вимагають розв'язання відповідних задач управління системами. У 1930-1934 рр. талановитий російський учений М. Д. Кондратьєв (1892-1938), перебуваючи в політізоляторі, розробив динамічну модель макроекономіки, використавши апарат диференціальних рівнянь. У 1938-1939 рр. ленінградський математик Л. В. Канторович (1912-1986), досліджуючи деякі проблеми організації та планування виробництва, сформулював новий клас умовно-екстремальних задач. Відповідна галузь прикладної математики згодом була названа "лінійне програмування"; під цим терміном розуміють обґрунтування плану відповідних економічних заходів.

У 1960-1980 рр. дослідження в галузі використання економіко-математичних методів відродилися та розширилися. Вийшли з друку фундаментальні праці: "Экономический расчет наилучшего использования ресурсов" Л. В. Канторовича (1959), "Измерение затрат и их результатов в социалистическом хозяйстве" В. В. Новожилова (1959), "Экономико-математические методы и модели" В. С. Немчинова (1962).

У 70-ті роки розпочалися роботи по впровадженню досягнень економіко-математичного моделювання та теорії дослідження операцій в розробку автоматизованих систем планування й управління. Основна увага дослідників у країнах СНД зараз зосереджена на розв'язанні проблем перехідного періоду. Однією з важливих проблем західноєвропейських та американських учених є дослідження

* Браславец М. Е. Экономико-математические методы в организации планирования сельского хозяйства. – К.: Урожай, 1968.

соціально-економічних питань розбудови ринкових відносин на планетарному рівні.

Вивчення основних положень теорії дослідження операцій в економіці сприятиме формуванню конструктивного світогляду у кожного майбутнього працівника будь-якого рівня соціальної ієрархії: втілення будь-якого заходу розпочинається з плану його реалізації, вимагає часу та коштів; необхідно не лише працювати, а й оцінювати ефективність своєї праці, не лише приймати рішення, а й усвідомлювати значимість упереджуючої їх оцінки та контролю втілення.

ГЛАВА 1

ОСНОВНІ ПОНЯТТЯ ТА ПРИНЦИПИ ДОСЛІДЖЕННЯ ОПЕРАЦІЙ

1.1. Об'єкт дослідження

Аграрно-промисловий комплекс – це складна, багатоступенева галузь господарства країни. Він має широкі і глибокі зв'язки з іншими галузями виробництва. Ефективність роботи АПК визначає соціально-економічний стан всієї країни. Аграрна наука та виробничі структури можуть забезпечувати успішне функціонування АПК лише за умови постійного удосконалення методів організації й управління з використанням сучасних досягнень науки і практики дослідження операцій. Але кожний фахівець мусить чітко розуміти, що будь-які обґрунтовані і змістовні розрахунки та рекомендації мають сенс лише за умови їх практичного втілення. Завжди має бути раціональним поєднання теоретичних розрахунків і практичної діяльності.

Запити практики обумовили розробки спеціальних наукових методів розбудови й організації функціонування складних виробничо-організаційних систем. Такі методи об'єднані назвою "дослідження операцій". Термін "операція" означає виконання заходів для досягнення певної мети. Будь-яка операція реалізується через матеріальне втілення цієї мети, тобто, досліджуючи операцію, ми досліджуємо її матеріальний прояв, який може поєднувати і якісні і кількісні ознаки. Вважаємо, що операція як сукупність заходів завжди керований процес, але наслідки операції при одному й тому ж управлінні, як правило, не завжди тотожні, бо перебіг процесу може бути за різних умов.

Практика засвідчує, що виробничі й економічні процеси теж доцільно характеризувати якісними та кількісними показниками, з урахуванням закономірностей їх формування. Сумлінно виконаний якісний аналіз у багатьох випадках буває міцним підґрунтям для визначення закономірностей стосовно кількісних показників. Ці процеси вивчаються за універсальною схемою пізнання: спостереження, досвід, накопичення необхідної інформації, її аналіз, виявлення найбільш суттєвого узагальнення закономірностей, їх перевірка на практиці.

Узагальнення характерних особливостей і закономірностей, властивих досліджуваному об'єкту або процесу, виконане певними засобами зображення, називається *моделлю*. Існує чимало визначень поняття моделі, що пов'язано з їх використанням в різних галузях науки. Автори всіх визначень єдині в тому, що будь-яка модель лише наближено описує відповідне реальне явище з віддзеркаленням тих його особливостей, які були враховані при розбудові моделі.

У дослідженні операцій широко використовуються *математичні моделі*, які

ґрунтуються на формально-логічному апараті математики. Але не можна ототожнювати моделі дослідження операцій з математичними моделями, апарат яких при цьому використовується. Модель дослідження операцій завжди є більш об'ємною стосовно мети та об'єкта дослідження, ніж використовуваний математичний апарат, з допомогою якого оцінюються властивості об'єкта. Математичний апарат – це чудовий музичний інструмент з величезними можливостями, дослідження операцій – це об'єднання виконавців з їх інструментами в оркестр.

Результати лише кількісного аналізу, одержані при використанні відповідної математичної моделі, не можуть бути єдиною переконливою підставою для прийняття певного стратегічного рішення. Навіть за умов, коли кількісний аналіз відіграє визначальну роль у прийнятті управлінських рішень, не можна досягти розбудови таких моделей управління, щоб одержати інформацію в обсязі, достатньому для обґрунтування майбутнього способу дій. Управлінські рішення завжди приймає людина; розрахунки, якими б якісними вони не були, є лише підґрунтям.

Необхідно наголосити на різниці між об'єктами дослідження операцій і так званого *системного аналізу*. Фахівці системного аналізу досліджують функціонування та взаємозв'язки між окремими частинами складних *фізичних* систем: телекомунікацій, електропередач, водогонів, транспортної мережі і т. ін. *Об'єктом дослідження операцій є функціонування систем, що поєднують цілеспрямовану діяльність людей, які використовують певні засоби виробництва* (система "людина – машина"). Такі системи називають *операційними*, наголошуючи на їх цільовому призначенні: забезпеченні виконання певних дій – операцій. Будь-яка операційна система може ефективно функціонувати лише за умови врахування поведінки людей, їх психології та дій в системі "людина – машина", тому, досліджуючи такі системи, необхідно проводити так званий *біхевіористичний аналіз* – оцінювати реакцію та поведінку людей при використанні ними технічних засобів виробництва як стосовно засобів, так і поміж собою. Процес розбудови та дослідження таких систем орієнтований не лише на логічні операції із символами та на числові методи, а й на здоровий глузд. Тобто запровадження операційної системи та її ефективне функціонування не можливе поза факторами організаційної діяльності виконавців. Проектування операційної системи може стосуватися як уже діючих виробничих та управлінських структур (тоді на меті підвищення якості їх функціонування), так і запланованих до організації. В обох випадках має бути чітко окреслена мета виконання запланованого проекту дослідження операцій: сукупність заходів з підвищення ефективності функціонування, оперативного управління та оцінки можливостей надійного прогнозування.

1.2. Мета та цілі дослідження

Мета в загальних рисах визначається тим, що бажають одержати, виконуючи певні заходи. Мета і є сукупністю наслідків реалізації цих заходів. У багатьох випадках є сенс розуміти мету як сукупність цілей. Наприклад, маємо мету організувати виробництво продукції певного найменування. Для втілення цієї мети треба досягти таких цілей: підготувати відповідні кадри, забезпечити виробництво матеріально-технічними засобами, запобігати можливих шкідливих екологічних наслідків, дослідити потенційні можливості ринку збуту, уникнути можливої негативної протидії конкурентів і т. ін. Тобто мета, як правило, багатопланова та багатогранна, цілі більш конкретні.

Надалі під терміном *"операційна система"* (ОС) будемо розуміти наявну або плановану структуру (організацію) певного призначення. *Модель функціонування* такої системи постає як сукупність певних математичних моделей, правил послідовності їх використання та правил дій і обов'язків людей, які працюють у структурі.

Постановка задачі дослідження операцій вважається здійсненою, якщо визначені:

- мета та цілі функціонування системи, значимість кожної з них;
- можливі засоби досягнення необхідних цілей та критерії їх ефективності;
- оцінки можливостей реалізації засобів досягнення намічених цілей.

Розв'язання задачі – це аналіз усіх можливих альтернативних способів досягнення мети та цілей і вибір найбільш ефективних з них за певними критеріями. Природно, що значимість кожної з цілей створення операційної системи або її частин різна й обумовлена багатьма факторами, які визначають її функціонування. Наприклад, що є більш вагомим: собівартість продукції чи її якість? Як досягнути раціонального узгодження між цими властивостями? Як свідчать літературні джерела, в Японії при визначенні доцільності *вдосконалення* або *заміни технології* випуску продукції перевагу віддають підвищенню якості, навіть якщо собівартість зростає. У Європі та Америці за таких умов перевагу віддають собівартості, звичайно, забезпечуючи певний рівень якості. Взагалі визначення значимості ОС – складне та відповідальне завдання, яке, як правило, виконують фахівці та експерти відповідної галузі. Можливі схеми розбудови відповідних алгоритмів наведені у главах 2 та 19.

За будь-яких обставин для розбудови ОС завжди необхідно мати перелік цілей, упорядкованих за значимістю. Маючи певну ціль, необхідно проаналізувати можливості та засоби її досягнення, тобто *стратегію*. Якщо є можливість вибору стратегій, то з'являється задача їх порівняння між собою та вибору. Для розв'язання таких задач завжди необхідно ввести певні критерії оцінки кожної стратегії. Для

того, щоб вибір був упорядкований, критерій оцінки стратегії повинен мати властивість транзитивності: якщо за обраного критерію стратегія A краще, ніж B, а стратегія B краща, ніж C, то стратегія A має бути кращою за C. Стисло розглянемо ситуації, коли ефективність певної стратегії стосовно досягнення цілі можна характеризувати *лише одним числом,* знайденим за певною методикою або визначеним за експертними оцінками.

У випадку, якщо можна користуватися детермінованою схемою, таке число є результатом розрахунків певної функції, наприклад, собівартість або зиск після реалізації продукції. Розбудови та використання таких функцій розглянуті при аналізі відповідних математичних моделей. У випадку, коли необхідно скористатися стохастичними моделями оцінок можливих стратегій, таке число P_{ij} є ймовірністю досягнення i-ої цілі за умови втілення j-ої стратегії. Якби всі цілі були рівнозначними, то вибір оптимальних стратегій виконувався б за умовою: найкращою є така множина стратегій, для якої виконується умова $\sum_{j=1}^{n} \max_{1\le i\le m} P_{ij}$ (n – кількість допустимих стратегій; m – кількість цілей розбудови досліджуваної системи).

Але значимості цілей різні, і вплив їх досягнення на загальну ефективність системи має бути визначений з урахуванням значимості V_i кожної цілі. Найкращою буде множина стратегій, для якої виконується умова $\sum_{j=1}^{n} \max_{1\le i\le m} V_i P_{ij}$; величини V_i значимостей цілей мають бути пронормовані так, щоб виконувалась умова $\sum_{i=1}^{m} V_i = 1$.

Природно, що втілення різних стратегій може вимагати різних затрат і часу. Більш конкретно оцінки ефективності стратегій буде розглянуто стосовно відповідних математичних моделей.

1.3. Моделювання в теорії дослідження операцій

Як наголошувалося раніше, модель має віддзеркалювати найбільш суттєві властивості та взаємозв'язки досліджуваного об'єкта або процесу. При розробці моделі (моделюванні) враховуються як визначальні характеристики об'єкта, так і мета дослідження. Дослідження операцій як наука має ту принципову особливість, що її об'єктами є відносно швидко змінювані організаційно-управлінські системи, які функціонують у нестабільних зовнішніх умовах і *поєднують взаємодії людей та засобів праці*. Такі системи, як і умови їх функціонування, практично не можна створити за бажанням дослідника, тому лише моделювання є єдиним та ефективним методом дослідження.

До того ж не можна розтягувати дослідження операцій у часі, бо несвоєчасно

одержані результати можуть стосуватися минулого стану системи і будуть марними. Результати мають бути завчасними, упереджувати розвиток системи та зміни як зовнішні, так і внутрішні, в середовищі її функціонування. Прогнозування розвитку економічної або системи іншого призначення є одним з головних завдань дослідження операцій. Перелічимо основні етапи дослідження операційних систем.

Перший етап – попереднє вивчення системи та її якісних показників. Результатами мають бути загальні уявлення про систему, про її цільове призначення, структуру та умови функціонування.

Другий етап – загальна постановка задачі оцінки ефективності управління або економічного функціонування, тобто визначення мети. Формулюється мета дослідження як множина певних цілей, які необхідно досягти, визначаються критерії оптимальності стосовно цільового призначення системи, визначаються в загальних рисах засоби досягнення певних цілей та оцінка їх ефективності.

Третій етап – розробка математичної моделі досліджуваної системи або її частин. Інформація, одержана на перших двох етапах, узагальнюється, упорядковується, гіпотетичні припущення та закономірності подаються мовою математичного апарату відображення залежностей, складні системи, як правило, моделюються множиною математичних моделей, які стосуються окремих частин або різних аспектів функціонування системи. Питання математичного моделювання більш детально будуть розглянуті у главах, в яких розповідається про відповідне цільове призначення математичних моделей.

Четвертий етап – розв'язок математичної моделі, який дуже рідко вдається виконати аналітично, тому широко використовуються числові методи. Результати розв'язку мають бути представлені в такому вигляді, щоб їх можна було використати на практиці або в подальших розрахунках.

П'ятий етап – перевірка узгодженості розв'язку задачі та відповідних практичних рекомендацій з реальними даними. Модель вважається узгодженою з дійсністю, якщо вона з практично достатньою точністю визначає основні показники функціонування системи. Але при цьому необхідно враховувати можливі похибки числової реалізації математичної моделі. Доцільно наголосити, що точність розрахунків має бути адекватною точності вхідних експериментальних даних. Це є однією з причин недоцільності розробки математичних моделей з великою кількістю змінних.

Шостий етап – визначення процедури коригування моделі. Зі зміною досліджуваної системи в часі та під впливом зовнішніх умов слід коригувати модель системи та вхідні дані для числового розв'язання відповідної задачі. Необхідно мати систему правил зміни моделі та вхідних даних з відповідною корекцією управління системою. Відсутність таких правил може викликати небажані, а іноді й катастрофічні наслідки. Сумний приклад – аварія на Чорнобильській АЕС.

Сьомий етап – упровадження рішення в практику. Метою дослідження системи є практичне використання результатів дослідження. За цих умов має бути найтіснішим зв'язок між тими, хто одержує розв'язок, і тими, хто впроваджує його в практику.

1.4. Математичне моделювання

Моделі, які віддзеркалюють на якісному рівні причинні взаємозв'язки між складовими частинами досліджуваної ОС, будуються переважно за допомогою логічних символів і слів у їх геометричному зображенні.

Кількісні моделі будуються з використанням певного математичного апарату, вони більш динамічні, добре пристосовані до сучасних комп'ютерів. При моделюванні ОС поєднуються якісні та кількісні моделі. Розглянемо деякі загальні питання розробки математичних моделей.

Використання математичного апарату дає можливість:

- виділити й описати найбільш значимі зв'язки в системі в кількісних співвідношеннях;
- використовуючи моделі взаємозв'язків, методами логічного аналізу одержати інформацію, адекватну досліджуваним властивостям об'єкта;
- розв'язавши відповідну математичну модель, одержати вірогідну інформацію про стан системи та можливість прогнозувати і забезпечити її майбутній стан.

При будь-яких економічних або інших дослідженнях поєднуються теоретичні розробки, представлені математичними моделями (ММ), та практичний досвід, віддзеркалений у статистичних даних. Процес побудови, дослідження та практичного використання моделі називається *моделюванням.*

Побудова моделі ґрунтується на певній інформації про досліджуваний об'єкт. Можливості дослідження об'єкта з допомогою моделі пов'язані з тим, що ММ віддзеркалює найбільш суттєві з *погляду дослідника* властивості та взаємозв'язки досліджуваного об'єкта. Питання необхідної, достатньої й ефективної міри узгодженості оригіналу та його ММ вимагає конкретного аналізу і в багатьох випадках може бути позитивно розв'язуваним лише при проведенні об'ємних і змістовних досліджень. З формальної позиції будь-які явища, процеси та системи – це об'єднання певних елементів та існування відповідних зв'язків між ними, що забезпечує їх цілісність та стабільність стосовно зовнішніх і внутрішніх змін. Будуючи ММ, ми виділяємо значимі для нас елементи та зв'язки між ними, тобто досліджуємо певні властивості об'єкта або процесу, нехтуючи іншими, не суттєвими для мети дослідження властивостями та зв'язками. Таким чином, *моделювання завжди має цілеспрямований характер* і дає можливість досліджувати лише певні властивості об'єкта або процесу. Отже, для одного й того ж об'єкта в залежності від

мети дослідження можуть бути використані різні ММ і навпаки: ММ одного й того ж аналітичного зображення можуть віддзеркалювати зовсім різні за природою явища та процеси.

Побудувавши певну ММ, *одержуємо самостійний об'єкт дослідження*. При таких дослідженнях цілеспрямовано і свідомо в межах, обумовлених властивостями ММ, змінюються умови функціонування моделі та обчислюються відповідні дані.

Накопичивши певний обсяг знань про ММ, аналізуємо можливість ототожнення одержаних знань з відповідними властивостями об'єкта моделювання. Природно, що таке ототожнення стосується лише тих властивостей, які були віддзеркалені у ММ. Одержана з допомогою ММ інформація про реальний об'єкт або процес використовується на практиці та для розробки узагальнюючої теорії процесу, його управління, перетворення та прогнозування. Оскільки аналітичні співвідношення, використовувані в математиці, є абстрактними віддзеркаленнями об'єктивно існуючих якісних і кількісних зв'язків у явищах та процесах, то математичне моделювання в широкому розумінні є методом дослідження, який ґрунтується на математичній аналогії, явищ і процесів різних за природою. У сучасних умовах математичне моделювання невіддільне від використання широких можливостей комп'ютерів.

Змінні, які використовуються при розробці ММ за критерієм визначеності, поділяються на дві групи. Змінні, які відображають фізичне втілення досліджуваного об'єкта та зовнішні умови, називаються *екзогенними* (їх величини визнаються поза розробленою ММ), змінні, величини яких обчислюються за ММ, – *ендогенними*.

Припустимо, що необхідно визначити ефективну спеціалізацію – головну галузь багатогалузевого сільськогосподарського підприємства. Назвемо можливі *екзогенні змінні* (параметри): площі орних земель, сіножатей, пасовищ, саду, наявність засобів механізації, трудові ресурси, коефіцієнти ефективності праці в кожній галузі виробництва, транспортна мережа. *Ендогенними змінними* будуть обсяги можливих прибутків кожної галузі виробництва, питання працевлаштування; саме для їх обчислення і необхідно будувати відповідну ММ.

За критерієм цільового призначення ММ використовувані для її розробки змінні доцільно розділити так: *параметри стану* та *параметри управління*. До параметрів стану належать змінні, які виділяють досліджуваний об'єкт серед інших того ж призначення і дають можливість обчислити величини тих характеристик, для визначення яких і розробляється ММ. До параметрів управління належать характеристики, змінюючи величини яких в певних межах можна впливати на досліджуваний стан і з великої кількості можливих станів системи за певним критерієм вибрати оптимальний. Наприклад, у задачі оптимального розподілу інвестицій у багатогалузевому виробництві параметрами управління будуть обсяги коштів, а параметрами стану – розміщення коштів в кожній галузі в залежності від забезпеченості матеріально-технічними та трудовими ресурсами й ефективності функціонування галузі.

Використовувані математичні моделі можна класифікувати за різними ознаками стосовно як об'єкта, який моделюється, так і математичного апарату: макро- та мікроекономічні, теоретичні та прикладні, оптимізаційні та рівноваги, статичні та динамічні.

Макроекономічні моделі використовуються при дослідженні економіки в цілому на рівні господарства, галузі, держави та стосуються узагальнених матеріальних і фінансових показників: ВНП, обсягів споживання та інвестицій, зайнятості населення, обсягів грошової маси і т. ін. Яскравим прикладом таких моделей є моделі міжгалузевих зв'язків.

Мікроекономічні моделі стосуються дослідження взаємодії та стану структурних і функціональних складових економічної системи: фірми, підприємства. Мікроекономічне моделювання посідає значне місце в теорії дослідження операцій окремих виробництв, що обумовлено різноманіттям форм і зв'язків між об'єктами господарювання. Їх роль особливо зростає в умовах ринкової економіки, при дослідженні економічних стратегій фірм та інших виробничо-торгових утворень.

Моделі рівноваги використовуються при моделюванні ринкових відносин. Вони описують такий стан відносин між суб'єктами ринкової економіки, за якого результуюча за певними критеріями усіх стосунків між ними дорівнює нулю. В умовах неринкової економіки порушення рівноваги між одними параметрами як віддзеркалення реально існуючого стану (наприклад, незадоволення попиту споживачів) компенсується впливом інших факторів (черги, чорний ринок).

Доцільно наголосити, що проблема *оптимізації* за умов ринкових відносин має різні аспекти стосовно макро- та мікроекономіки. На мікрорівні це проблеми пошуку умов найбільшого зиску фірми, на макрорівні – вибір фірмою такої стратегії виробничо-економічної діяльності, яка б забезпечувала стабільну рівновагу на ринку при врахуванні багатьох факторів, наприклад, купівельної спроможності, виробництва супутніх товарів і т. ін.

За характером віддзеркалення взаємозв'язків між змінними, використаними при розробці моделі, розрізняються моделі *детерміновані* та *стохастичні* (*ймовірні*). У детермінованих моделях приймають, що між змінними існує однозначний функціональний зв'язок. При розробці стохастичних моделей враховуються випадкові результати в прояві зв'язків між змінними. У таких моделях широко використовується математичний апарат теорії ймовірностей та математичної статистики. Значимість обчислюваних параметрів оцінюється певною величиною їх достовірності.

За способами врахування фактора часу економіко-математичні моделі поділяються на *статичні* та *динамічні*. У статичних моделях усі досліджувані залежності співставляються або з певним проміжком часу, або з конкретним моментом. Динамічні моделі дають можливість досліджувати взаємозв'язки змінних з часом, віддзеркалюючи перебіг процесу. Наприклад, динаміка інвестування визначає динаміку

основного капіталу, що обумовлює зміну обсягів виробництва в часі. У динамічних моделях використовуються апарат диференціальних і скінченно-різницевих рівнянь та варіаційного числення, часові ряди.

Особливе місце за можливостями та способом використання посідають так звані імітаційні моделі, які, як правило, синтезують методи розбудови різних раніше названих ММ. Імітаційною називається модель, побудована для моделювання перебігу досліджуваного економіко-виробничого процесу або процесу управління в часі за умови активного втручання людини в хід процесу та принципової зміни нею подальшого перебігу процесу. Практична значимість імітаційної моделі визначається рівнем тотожності моделі відповідному процесу. Маючи імітаційну модель, користувач в залежності від стану зовнішніх умов та досліджуваного процесу може приймати відповідні рішення та за допомогою моделі визначати можливі їх наслідки. Наступне "поточне" рішення визначається користувачем при врахуванні досягнутого раніше стану та ін. Багаторазове повторення таких процедур за різних вихідних станів дає можливість оцінити наслідки перебігу процесу та вибрати план найкращого управління ним. Розбудова імітаційних моделей можлива лише за наявності широкого сервісного забезпечення, яке дозволяє користувачеві працювати з моделлю в діалоговому режимі.

Імітаційні моделі є основою так званих "ділових ігор", які широко використовуються для підготовки керівних кадрів. Значення імітаційних моделей зростає в умовах перехідної економіки, коли використання стохастичних моделей не може бути достатньо обґрунтованим. Це питання найбільш вдало розроблене білоруськими вченими*. В Республіці Бєларусь ефективно використовується імітаційна балансова середньотермінова макроекономічна модель, яка дозволяє одержувати практично значимі розрахунки навіть за нестабільних економічних умов.

Стосовно використання математичного моделювання в дослідженні операцій можна вказати, хоч і дещо умовно, на чотири аспекти.

1. *Удосконалення й упорядкування інформації про досліджувану систему.* Необхідність чіткого формулювання ММ обумовлює вимоги щодо обсягів упорядкування інформації, виявлення недоліків та визначення нової інформації. Прогрес в інформаційному забезпеченні планування й управління обумовлений стрімким зростанням можливостей комп'ютерів, їх спроможністю швидко обробляти великі обсяги різнопланової інформації.

2. *Інтенсифікація та підвищення надійності розрахунків.* Можливості ММ та комп'ютерів значно прискорюють виконання типових, стандартних розрахунків, підвищують точність, скорочують обсяги трудових витрат і вартість, дозволяють виконувати пошуки оптимальних варіантів розв'язання складних проблем управління та господарювання.

* Экономико-математические методы и модели: Учеб. пособие / Н. И. Холод, А. В. Кузнецов, Я. Н. Жихар и др.; Под общ. ред. А. В. Кузнецова. – 2-е изд. – Минск: БГЭУ, 2000.

3. *Підвищення якості аналізу функціонування систем.* Використання методів ММ значно розширює можливості кількісного аналізу: дослідження взаємозв'язків багатьох факторів, які впливають на функціонування системи, оцінка наслідків зміни зовнішніх умов та власного стану системи.

4. *Розв'язання принципово нових проблем управління системами.* Лише при використанні методів математичного моделювання можна розв'язувати багато проблем, які іншими засобами розв'язати не вдається, наприклад, дослідження фінансового та економічного стану держави, спільне функціонування різних галузей господарства і т. ін.

Сфера практичного використання методів дослідження операцій у значній мірі визначається можливостями й ефективністю ММ, інформаційним і технічним забезпеченням використовуваних моделей, але вона не тотожна математичному моделюванню. При розробці та прийнятті соціально-господарських рішень слід поєднувати формальні методи, узгоджені з існуючим законодавством, і методи економіко-математичного моделювання. Але розробка будь-яких якісних ММ може бути доцільною та ефективною лише при відповідному рівні системи управління, забезпеченості кадрами, технічному та організаційно-правому забезпеченні.

1.5. Операційні системи, призначені для людей

Розглянемо деякі загальні важливі питання розбудови та впровадження ОС.

По-перше, виконання рекомендацій, заходів і вимог стосовно функціонування будь-яких структур не може бути забезпеченим лише за рахунок розбудови множини математичних моделей незалежно від їх якості.

Визначаючи важливість використання ММ, треба чітко розуміти їх місце як у практичному управлінні, так і в прогнозуванні: вони можуть допомогти (і допомагають) у прийнятті управлінських рішень, а самі рішення приймають і відповідають за їх наслідки керівники. Розробка ОС та її впровадження – це значно більше, ніж створення та використання будь-яких ММ. Організаційна структура, яка використовує певну ОС, має функціонувати за умов впливу різних факторів як техніко-економічних, так і діяльності виконавців, саме вони забезпечують життєздатність структур. У підручнику основну увагу приділено розбудові та використанню ММ, але читач мусить розуміти не лише їх величезну значимість, а й обмеженість сфери використання без урахування діяльності людей, працюючих у відповідній структурі та використовуючих певні засоби виробництва.

По-друге, в ОС завжди необхідно приділяти увагу джерелам інформації, її якості та повноті, що в значній мірі залежить від фахової підготовки відповідальних за накопичення й аналіз інформації. Також необхідно враховувати і вимоги до обсягів та якості тієї інформації, яка передається керівникові, відповідальному за прийн-

яття управлінських рішень.

По-третє, необхідно враховувати, що часто науково розроблені засади ОС не співпадають зі звичками працюючих і керівників, їх уявленнями та розумінням перебігу подій. Вивченням таких ситуацій та розбудовою відповідних науково обґрунтованих рекомендацій інтенсивно займаються фахівці біхевіористичного аналізу, об'єктом дослідження якого є поведінка та стосунки людей в колективах при виконанні ними певних обов'язків і використанні знарядь праці. Оцінки одного й того ж явища можуть бути у різних людей діаметрально протилежними. Наведемо приклад із практики. В одному з господарств випробовували американські зернозбиральні комбайни "Джон Дір", на одному з яких погодився певний час попрацювати головний інженер. По завершенні терміну роботи головний інженер висловив думку, що комбайн чудова, висококомфортна машина, може регулювати темпи і якість збирання врожаю в залежності від вологості та засміченості. "Досвідчений" же комбайнер заявив, що "начеплені" пристрої – то зайве, без них краще, єдине, що добре, так це наявність кондиціонера.

Розбудова проектів ОС та їх впровадження, як правило, виконуються для конкретних замовників, загальні методи та вимоги до ОС є об'єктами ретельних досліджень фахівців-операційників. Упровадження та використання ОС в значній мірі обумовлені майстерністю як операційників, так і керівників відповідних структур та їх ланок. Сподіваємося, що підручник допоможе одержати загальне уявлення про зміст та ефективність ОС і сприятиме формуванню у фахівців вміння оцінювати можливості науково обґрунтованих методів організації, управління та прогнозування.

Питання для самоконтролю

1. Що вивчає наукова дисципліна "Дослідження операцій"?
2. Як ви розумієте поняття "математична модель" та її призначення?
3. Що ви розумієте під терміном "біхевіористичний аналіз"? Розв'язання яких питань забезпечують його методи?
4. Що ви розумієте під терміном "виробничо-економічна структура"?
5. З якою метою виконується дослідження операцій в певній структурі?
6. Особливості розбудови математичних моделей в теорії дослідження операцій.
7. Принципи відмінності між екзогенними та ендогенними змінними математичної моделі.
8. Як ви розумієте термін "параметри управління"?

9. Різниця між поняттями "мікроекономіка" та "макроекономіка".
10. Принципова різниця між детермінованими та стохастичними математичними моделями.
11. Який математичний апарат використовують за розбудови стохастичних ММ?

ГЛАВА 2

МЕТОДИКА ПРОВЕДЕННЯ ДОСЛІДЖЕННЯ ОПЕРАЦІЙ

Дослідження кожної функціонуючої або проектованої системи має свої особливості, тому вимагає від дослідника не лише спеціальних знань в досліджуваній сфері виробництва та відповідного математичного апарату, але й уміння логічно мислити, прояву інтуїції та ініціативи. Фахівець з дослідження операцій мусить чітко розуміти, що будь-які обґрунтовані та змістовні розрахунки і рекомендації мають сенс лише за умови їх практичного втілення та ефективності. Сподіваємося, що матеріал пропонованої глави допоможе в загальних рисах з'ясувати основні етапи дослідження операцій (ДО) та їх зміст. Зрозуміло, що для кожної досліджуваної системи етапи дослідження будуть мати своє, конкретне втілення. Доцільність чіткого уявлення про методику ДО обумовлена кількома вагомими обставинами.

Використання правильної методики суттєво підвищує ймовірність *уникнення невірно сформульованої задачі* дослідження, а також *можливості невірного розв'язання правильно поставленої задачі.* Якщо дослідник (або група) керується вірною методикою, то більше гарантій, що розбудована модель досліджуваної системи адекватна меті дослідження, а при впровадженні результатів забезпечить очікувану ефективність.

Етапи ДО в загальних рисах майже тотожні тим, які властиві організації будь-яких наукових досліджень, але мають і свої особливості, обумовлені специфікою об'єктів дослідження. Даючи перелік етапів, наголошуємо на певних конструктивних, хоч і загальних рекомендаціях щодо розв'язання задач методами дослідження операцій.

2.1. Визначення мети та значимості цілей

Поява чергової функціонуючої системи або проектування нової обумовлені певною необхідністю. Тому, прагнучи дослідити систему, доцільно знати мету і якомога повніша перелік цілей створення системи та її поточного функціонування. Наголосимо, що цілі групуються за двома ознаками: 1) цілі або результати, які необхідно досягти в процесі роботи системи; 2) цілі (результати), які необхідно зберегти (наприклад, зменшення собівартості продукції за умови збереження її якості). Визначення кількості цілей, їх значення та взаємозв'язку є суттєвим етапом дослідження системи. Метою виконання проекту ДО є оцінка та поліпшення роботи виробничої або організаційної системи в залежності від її функціонального призначення. Накреслюючи мету та цілі дослідження, необхідно потурбуватися про їх раціональне визначення: вони не повинні бути всеосяжними, щоб не перевищити межі

наявних можливостей, або другорядними, щоб не "загубити" основне призначення системи. Маючи реєстр цілей, необхідно зробити аналіз рівня їх взаємозалежності та відкинути ті, які не є визначальними або свідчать про досягнення інших цілей. Наприклад, маємо дві привабливі цілі: збільшити прибуток підприємства та зменшити виробничі витрати. Очевидно, що ціль зменшити виробничі витрати буде досягнута за умови досягнення першої цілі – збільшення прибутку, тобто вона обумовлена першою ціллю і не є незалежною. Але треба враховувати, що подібна ситуація може бути обумовлена певними засобами досягнення відповідних цілей, за інших засобів може змінитися співвідношення між цілями. Досягти чіткого розуміння мети та цілей до початку дослідження абсолютно необхідно, але, на жаль, недостатньо.

Цілі та засоби їх досягнення взаємопов'язані: їх взаємне співвідношення може змінюватися в часі та при зміні умов функціонування системи. Тому, маючи відповідні реєстри, необхідно виконати аналіз можливостей досягнення цілей на множині допустимих засобів їх досягнення (стратегій). Цілі, які за наявних стратегій не можна досягти, слід відхилити. Маючи такий наближений аналіз узгодження цілей і стратегій, перевіряють вказані питання стосовно всіх складових досліджуваної системи для завершального визначення реєстру цілей як мети розбудови проекту.

Особливої шкоди розбудові проекту ДО завдають ситуації, коли цілі повністю або частково змінюються замовником в процесі виконання проекту. У зв'язку з цим доречно дати виконавцеві проекту дослідження операцій певні рекомендації.

1. Виконавець мусить не лише слухати, а й сприймати й однозначно розуміти завдання.
2. Щоб з'ясувати мету дослідження, виконавець зобов'язаний сприймати розв'язувану задачу з позицій керівництва відповідної структури, враховуючи обмеження та вимоги іноді протилежного напрямку, серед яких виконавець має працювати та шукати компромісу.
3. Виконавець мусить знати структуру та ієрархію організації, для якої виконується проект.
4. Виконавець мусить утримуватися від власних думок про зміст досліджуваної проблеми та від затискання її в межі своїх, раніше сформованих уявлень про аналогічні задачі, не нав'язувати свого підходу до її розв'язання, поки повністю не з'ясує мету керівництва, яке замовляє проект ДО.
5. Надзвичайно важливо, щоб жодне твердження та зауваження керівництва не залишилося неперевіреним та неоднозначно сприйнятим. Пасивне, через увічливість вислуховування керівництва ніколи не призведе до з'ясування суті проблеми. Виконавець мусить питати та перевіряти однозначність розуміння кожного питання та змісту відповіді на нього.

Оскільки розглядається проблема дослідження операцій з використанням математичних моделей, в яких є кількісні показники, то виникає питання узгодження розмірностей величин, в тому числі при визначенні цілей. Питання це досить склад-

не, і немає вичерпних рекомендацій щодо його вирішення. Цілі реєстру мають, як правило, різні розмірності, а іноді навіть лише якісні характеристики. Щоб користуватися математичним апаратом, необхідно представити цілі у кількісних вимірах і ввести єдину розмірність. У кожній конкретній ситуації це питання вирішується по-різному. Одна із загальних рекомендацій така: цілі упорядковуються за ступенем їх значимості, потім до розмірності головної цілі зводяться розмірності інших. Головна ціль оцінюється переважно у грошових одиницях як найбільш широко використовуваних одиницях виміру в економічних розрахунках. Стосовно визначення еквівалентностей в розмірностях інших цілей, то це питання розв'язується або вивченням минулого досвіду з урахуванням відповідних коректив, або на підставі експертних оцінок, або поєднанням цих підходів. Наприклад, якщо необхідно оцінити вартість простою агрегату за одну годину, то це можна зробити або на підставі попередніх досліджень, якщо такі проводились, або за відповідною оцінкою експерта.

Розглянемо методику оцінки ступенів значимості цілей, вважаючи, що для якісних цілей також маємо певні кількісні виміри. Якщо немає можливості запропонувати такі кількісні еквіваленти, то слід визначити сенс часткового досягнення такої цілі в кожній із стратегій.

У теорії експертних оцінок пропонується досить універсальний *метод оцінки відносної значимості цілей реєстру* – шляхом упорядкованого уточнення порівняльних експертних оцінок для кожної цілі.

1. Цілі, занесені до реєстру, послідовно нумеруються за початковими *якісними* експертними оцінками, тобто експерти визначають ієрархію значимості цілей.

2. Кожна ціль в одержаній послідовності співставляється з деяким числом V_i ($i = 1, 2, \ldots, n$; n – кількість цілей), *меншим за одиницю*, яке характеризує значимість цілі стосовно головної, якій ставиться у відповідність число $V_1 = 1{,}00$. Цей вихідний ряд значень V_i, сформований за початковими експертними оцінками, потім буде уточнений.

3. Оцінюються значимість першої цілі стосовно решти цілей (логічної суми). Якщо експерти вважають першу ціль більш суттєвою за значимістю, ніж логічна сума решти, то при необхідності величина V_1 змінюється так, щоб виконувалася нерівність $V_1 > V_2 + V_3 + \ldots + V_n$ (випадок *А*). Якщо ж експерти вважають, що перша ціль та логічна сума решти цілей рівнозначні за значимістю, то V_1 коригується так, щоб виконувалася рівність $V_1 = V_2 + V_3 + \ldots + V_n$ (випадок *Б*).

Якщо експерти вважають, що перша ціль має менше значення, ніж решта, то V_1 коригується так, щоб виконувалася нерівність $V_1 < V_2 + V_3 + \ldots + V_n$ (випадок *В*). Відкориговане значення V_1 порівнюється із сумою $V_2 + V_3 + \ldots + V_{n-1}$ (без урахування останньої цілі в реєстрі). За умови, що знову маємо випадок *В* (тобто $V_1 < V_2 + V_3 + \ldots + V_{n-1}$) V_1 порівнюють із сумою, зменшеною на V_{n-1}. Процес порівняння виконується до тих пір, поки значимість першої цілі V_1 стане більшою або рівною зменшеній сумі решти. По завершенні цього етапу визначається величина V_1, яка характеризує порівняльну значимість першої цілі.

4. У послідовності реєстру цілей відкидається перша ціль і визначається величина V_2, як це обумовлено п. 3. Реєстр цілей відповідно зменшується, і обчислюється величина V_3 за п. 3. Процес завершується по уточненні всіх значень V_i. На останньому кроці порівнюються значимості останньої та передостанньої цілей.

5. Значення величин V_i, одержані за описаним алгоритмом, нормуються так, щоб пронормовані величини V_i^* задовольняли умові:

$$V_i^* = \frac{V_i}{\sum_{i=1}^{n} V_i}, \text{ тобто } \sum_{i=1}^{n} V_i^* = 1.$$

Розглянутий алгоритм дозволяє упорядкувати за значимістю експертні оцінки цілей та оцінити їх числовими величинами. Одержані за алгоритмом оцінки є відносними і можуть змінюватися при зміні переліку та значимості цілей, а також у часі.

Не можна починати розбудову проекту дослідження операцій за відсутності чітко визначеної мети та ієрархії цілей реалізації проекту.

2.2. Дослідження стратегій

Задача дослідження засобів досягнення цілей (стратегій) може з'явитися лише при наявності альтернатив їх досягнення. Природно очікувати, що втілення будь-якої зі стратегій може забезпечити різні ступені досягнення мети – реєстру цілей, але жодна зі стратегій не може забезпечити повне досягнення усіх цілей. Тому необхідно визначити ефективність кожної стратегії стосовно кожної цілі, тобто визначити ступінь досягнення будь-якої цілі по втіленню кожної із відповідних стратегій. Цю ефективність теж необхідно характеризувати числовою мірою або як наслідок розв'язання відповідної математичної моделі, або шляхом експертних оцінок ефективності на зразок описаного алгоритму з визначення значимості цілей. Для реалізації процесу визначення ефективності стратегій необхідно мати відомості про всі можливі стратегії. Цього можна досягти або переліком їх елементів, або значенням критеріїв, за якими визначаються стратегії. За умови відомої математичної моделі можливих стратегій такі критерії будуть представлені сукупністю лінійних або нелінійних нерівностей або рівнянь стосовно набору змінних відповідної моделі. Наголосимо, що в багатьох випадках ефективність будь-якої стратегії змінюється в часі, а також суттєво залежить від зовнішніх умов.

Можливі випадки, коли альтернативні стратегії та їх ефективність відомі з попереднього досвіду або в порівнянні з аналогічними ситуаціями. Але більш вагоми-

ми є випадки, коли жодна з відомих стратегій не може гарантувати досягнення цілі. За таких умов виникає необхідність розробки нових стратегій (засобів досягнення цілі). Тобто процедури пошуку стратегій поділяються на два класи: 1) аналіз доцільності використання відомих стратегій у розв'язуваній проблемі; 2) розбудова нових можливих стратегій. На практиці в багатьох випадках доцільно об'єднувати ці класи пошуку можливих стратегій.

Необхідно наголосити, що в багатьох випадках можлива активна або пасивна протидія процесові втілення певної стратегії, обумовлена або самою системою, або зовнішніми факторами. Наявність можливих контрстратегій суттєво знижує ефективність стратегій. Таку ситуацію слід враховувати при дослідженні системи.

2.3. Планування етапів розбудови проекту

Дослідження операцій завжди до певної міри являє собою творчий процес, обумовлений майстерністю виконавця або виконавців. Тому природним є запитання: чи можна планувати таку роботу і як її планувати? Під планом треба розуміти визначення термінів виконання певних етапів робіт та узгодженість їх послідовності в часі. Такий підхід обумовлений необхідністю контролювати процес розробки проекту. Визначення термінів виконання проміжних етапів стимулює творчу активність виконавців і є необхідною умовою своєчасного завершення проекту. Виконання будь-якої роботи поза терміном її завершення породжує пасивність і безвідповідальність. До того ж порушення графіка виконання робіт викликає додаткові труднощі: змінюється ставлення замовника до виконуваного проекту, зменшується ефективність очікуваних результатів дослідження і т. ін.

План є по суті календарним графіком виконання узгоджених між собою етапів, тому його доцільно мати у вигляді структурної схеми, на зразок наведеної на рис. 2.1, з урахуванням координат часу.

Етапи проекту, як правило, деталізуються до рівня окремих завдань, узгоджених між собою за вхідною та вихідною інформацією. Чим детальніше розроблено план, тим достовірніше та ефективніше можна обґрунтувати оцінки термінів і ресурсів, необхідних для виконання проекту. Розподіл обсягу досліджень на окремі, узгоджені в часі етапи є найкращим способом достовірної оцінки необхідних фінансових, фахових і часових ресурсів ще до початку роботи над проектом, що суттєво підвищує ймовірність своєчасного виконання проекту з дотриманням обсягу асигнувань.

При складанні плану важливе значення має розподіл робіт між виконавцями та визначення термінів їх завершення. Доцільно наголосити на такому суттєвому питанні, як *повний аналіз даних*, зібраних і накопичених в процесі досліджень. Значимість цієї вимоги обумовлена такими обставинами:

Рис. 2.1.

1. Накопичуючи інформацію та аналізуючи її лише під кутом зору виконання поставленої задачі, фахівець іноді не помічає, що відносно вихідної мети є можливість одержати певні додаткові вагомі результати, а іноді, навпаки, зрозуміти неможливість втілення результатів планованих досліджень. Наведемо приклад. На одному з крупних машинобудівних заводів колишнього СРСР керівництво ковальського цеху вирішило модернізувати роботу ковалів важких деталей, кування яких виконується за такою формальною схемою. Коваль кліщами, тримаючи їх у двох руках, бере з горнила заготівку з початковою температурою біля 1200 °С і розміщує її на ковадлі пневматичного молота. Підручний керує силою та частотою ударів молота по деталі, яку коваль переміщає під молотом. Роботою підручного коваль керує лише з допомогою міміки: спілкуватися голосом неможливо через сильний шум, будь-яка автоматизація виключена через надто різний перебіг процесу кування кожної деталі навіть одного й того ж типу, що обумовлено різною початковою темпера-

турою та зміною як деформацій, так і температури в процесі обробки. Робота вимагає високої майстерності та взаєморозуміння між працюючою парою, тому відсутність конкретного підручного унеможливлює роботу коваля. Керівництво цеху вирішило, що коваль може управляти роботою молота, тримаючи в роті спеціальний мундштук і стискаючи його з певною силою та частотою. Було витрачено значні кошти на розробку відповідних датчиків та узгодженої системи управління молотом, розроблено інструкцію, проведено експериментальні позитивні дослідження, але розробки не були впроваджені через відмову ковалів тримати мундштук у роті. Яскравий приклад того, що вивчення проблеми було лише під вузьким кутом: автоматизувати процес без урахування всіх умов його реалізації.

2. Працюючи над проектом, виконавець зобов'язаний інформувати замовника про хід роботи та важливі для практики проміжні результати. У багатьох випадках чіткий аналіз даних стосовно певного питання, їх узагальнення в схемах та діаграмах не лише має позитивний вплив на стосунки між замовником і виконавцем, але й допомагає краще організувати поточну практичну роботу.

2.4. Визначення проблеми дослідження

Визначення проблеми дослідження – надзвичайно суттєвий етап, значимість якого важко переоцінити, в багатьох випадках саме він визначає успіх та ефективність дослідження або невдачу. Уже на цьому етапі необхідні тісні ділові контакти із замовником та особами, які будуть реалізувати результати та рекомендації планового дослідження. Необхідним є і накопичення інформації стосовно чіткого розуміння таких питань: у чому суть проблеми? якими є особливості функціонування системи діючої або проектованої? що та як відбувалося в минулому? чого можна чекати в майбутньому? якими є зв'язки між різними підрозділами структури та працюючими в них? які фактори є визначальними і як вони взаємопов'язані? На основі одержаних даних і відповідного аналізу визначається суть досліджуваної проблеми та загальна схема розбудови операційної моделі і зміст подальшої роботи. Значимість даного етапу обумовлює необхідність його всебічного обговорення з усіма зацікавленими сторонами.

Перше питання, яке доцільно з'ясувати на стадії визначення проблеми: чи можна всю проблему представити сукупністю окремих, менш об'ємних, часткових субпроблем з метою їх незалежного паралельного або послідовного дослідження. Це особливо важливо в тих випадках, коли досліджувана проблема стосується всебічної діяльності організації, для якої виконується проект дослідження. Процес логічного розподілу вихідної всеохоплюючої проблеми на складові може привести до субоптимального розв'язку, але відхилення від оптимуму, як правило, будуть щонайменшими, якщо субпроблеми не перетинаються. Для пояснення сенсу розв'язання цього питання скористаємося відповідним прикладом.

Необхідно було скласти план виробництва та транспортування продукції на черговий чотиритижневий період, користуючись даними за минулий тиждень. На стадії визначення проблеми було з'ясовано ряд суттєвих питань:

1) яким має бути рівень запасів на підприємстві по завершенні чотиритижневого планового періоду при *врахуванні сезонних коливань попиту*?
2) яку доцільно мати оптимальну величину поточних запасів на складі?
3) якими мають бути обсяги партій продукції, які відвантажують на склад за умови найефективнішого використання транспортних тарифів?

Ці питання тісно пов'язані з вибором методу розв'язання, який можна практично реалізувати, розділивши проблему на частини. На рис. 2.2 наведено схему виділення складових проблеми.

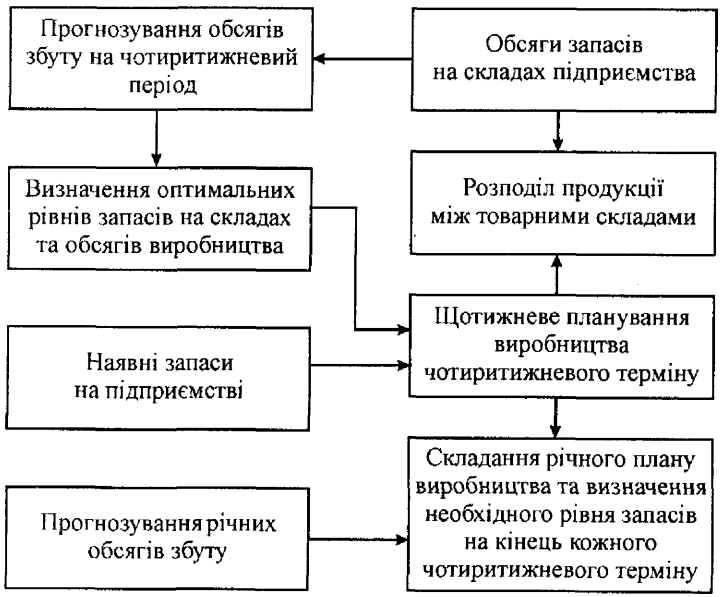


Рис. 2.2.

Було прийнято рішення про розробку таких основних моделей.

1. **Модель прогнозу попиту** на чотиритижневий термін за тижневими даними збуту кожного виду продукції.
2. **Модель визначення оптимальних обсягів складських запасів** по за-

вершенні чотиритижневого терміну за даними прогнозу попиту. За даними про поточний рівень запасів можна визначити оптимальні обсяги виробництва на черговий чотиритижневий термін.

3. **Модель планування на рік.** Мета розробки цієї моделі полягала в тому, щоб за даними річного прогнозу обсягів збуту продукції визначити в загальних рисах обсяги виробництва та рівень запасів на кожний чотиритижневий термін протягом року. Дані про рівень запасів повинні бути використані при визначенні планових завдань для системи управління запасами та в моделі планування виробництва на чотиритижневі терміни.

4. **Модель поточного планування виробництва** була розроблена для того, щоб, використовуючи величини обсягів запасів, визначених річним планом, та враховуючи вимоги стосовно складських обсягів запасів, визначити обсяги виробництва для кожного чотиритижневого циклу.

5. **Модель розподілу продукції між складами** при врахуванні результатів прогнозу обсягів збуту, даних про складські запаси та планів обсягів виробництва.

При розробці загального підходу до аналізу великомасштабних проблем досліднику доцільно дотримуватися такої принципової схеми.

1. Розпочати з основної задачі дослідження. У вищевказаному названому прикладі це задача планування чотиритижневого циклу виробництва.
2. З'ясувати, які вхідні та вихідні дані будуть потрібні при розбудові моделі основної задачі. Розв'язання цього питання дає можливість визначити зміст допоміжних моделей та необхідну для них інформацію.
3. З'ясувати наявність засобів формування необхідної вхідної інформації або використання вихідної інформації основної моделі. При їх відсутності необхідно побудувати додаткові моделі, які б забезпечили надходження відповідної інформації.
4. Чітко визначити, що має бути розроблене стосовно кожної додаткової моделі, якими є методи розрахунків та засоби використання одержаної інформації.

Якщо дотримуватися описаної схеми виконання, то є велика ймовірність ефективного розв'язання ряду важливих питань: визначення проблем, якими можна управляти, більш раціональна співпраця з персоналом замовника, ефективне впровадження результатів досліджень. Виконавець проекту ДО мусить чітко й однозначно уявляти собі призначення моделей, їх використання для розв'язання задачі управління, яку виконавець досліджує.

Доцільно уже на початку дослідження приділяти увагу створенню умов, необхідних для експлуатації майбутньої операційної системи. Ця вимога важлива тому, щоб успіх дослідження, яке завершується впровадженням нових систем і методів управління, не став тимчасовим. Якщо вивчення названих питань відкладається до етапу впровадження, то на завершення може з'ясуватися, що можливостей для успішної експлуатації розробленої системи недостатньо, а це, як правило, істотно зменшує й ефективність, і термін її існування. Наведемо реєстр контрольних питань, позитивне розв'язання яких допоможе в підвищенні ймовірності успішного функціонування впроваджуваного проекту ДО.

1. Чи передбачено необхідні кошти для неперервного коригування вхідної інформації при функціонуванні спроектованої операційної системи (ОС)?
2. В якій мірі розроблено документацію, достатню для можливих інших заходів щодо поліпшення системи?
3. Які заходи передбачено для забезпечення функціонування операційної системи за умови використання попередніх, традиційних методів управління? Це питання обумовлене тим, що досить часто спостерігається опір впровадженню нових методів, незважаючи на їх безумовну користь.
4. У якій мірі операційна система здатна адекватно й ефективно реагувати на можливі зміни умов діяльності організації чи підприємства?
5. Чи виконано необхідний аналіз можливих способів використання ОС за умови одержання інформації про її недоліки?

Друге питання, яке необхідно з'ясувати на стадії визначення проблеми, пов'язане з рівнем деталізації проекту. Основні фактори, які впливають на розв'язання названого питання – рівень фаху виконавців, обсяги наявних коштів, термін виконання проекту, мета та цілі його впровадження. Розв'язуючи питання деталізації, необхідно основну увагу звернути на міру адекватності, щоб уникнути розбудови так званої "зміщеної моделі", тобто такої, в якій не враховано певні суттєві та постійно діючі фактори. Результати ретельного дослідження кожної із складових проблеми по завершенні їх вивчення належним чином синтезуються, що дає можливість глибше та всебічно з'ясувати вихідну проблему. Природно, що за будь-якої деталізації ОС будуть похибки оцінки впливу певних факторів, але якщо вона адекватно віддзеркалює процес функціонування, використання моделі має забезпечувати практично значиму точність передбачення.

Після визначення проблеми та рівня її деталізації переходять до розбудови відповідних математичних моделей та їх узгодженості. На завершення наведемо приклад, за яких умов може з'явитися "зміщена модель".

Припустимо, виробнича діяльність характеризується такими показниками: виробничі потужності – 100 виробів за тиждень; попит – 400 виробів за чотири тижні.

Приймемо, що в математичній моделі планування виробничої потужності обрано чотиритижневий термін і на основі порівняння додаткових капітальних затрат з очікуваною економією від зменшення дефіциту та рівня товарно-матеріальних запасів аналізується доцільність капіталовкладення у розширення виробництва. За такої моделі може бути неврахованим надто суттєвий момент, який виявляється лише при більш детальному аналізі умов і показників виробничої діяльності. Припустимо, для планового терміну в *чотири тижні* маємо такі показники:

обсяг виробництва	400 виробів
попит на продукцію	400 виробів
вартість запасів	0
втрати за дефіциту	0
необхідні додаткові капіталовкладення	0

У межах чотиритижневого планового періоду коливання попиту не відчутні, що породжує ілюзію відсутності збитків, обумовлених витратами на зберігання продукції або її дефіцитом, а виробничі потужності узгоджені за обсягом запиту. Така ситуація означає, що чекати прибутків від додаткових капіталовкладень даремно. Але при аналізі тижневих проміжків стан може суттєво змінитися, наприклад, за умов, наведених у табл. 2.1.

Таблиця 2.1.

Показники	За тижневі проміжки				В цілому
	1	2	3	4	
Обсяги попиту	80	130	90	110	470
Виробництво	100	100	100	100	400
Запаси	20	0	10	0	30
Виконання запиту	80	120	90	110	400
Втрачений збут	0	10	0	0	10

Аналіз даних табл. 2.1 показує, що коливання попиту іноді перевищують наявні виробничі можливості. Таким чином, враховуючи екстремальні обсяги попиту за певні проміжки, доцільно мати необхідні запаси продукції. Відсутність таких запасів викликає збитки. Тобто є підстави для доцільності капіталовкладень в розширення виробничих потужностей.

2.5. Розбудова математичної моделі

Визначивши в загальних рисах проблему дослідження та можливості її поділу на складові частини, виконують розбудову її математичної моделі (або їх сукупності). Враховуючи розмаїття можливих ММ, яке було коротко описане у главі 1, маємо змогу лише в загальних рисах охарактеризувати цей етап, незважаючи на його надзвичайно важливе значення в розбудові проекту ДО та у функціонуванні ОС. Основну увагу буде приділено економіко-математичним моделям; поза увагою

залишаться специфічні моделі ОС, які стосуються взаємовідносин між працюючими та працюючими і використовуваними засобами праці.

Розбудова ММ допомагає об'єднати складні, хоч іноді й не чітко визначені фактори, пов'язані з проблемою прийняття рішень, у логічно чітку схему, яку можна детально проаналізувати. Саме такий аналіз дозволяє одержати й оцінити альтернативні можливості функціонування досліджуваної системи (ДС) та передбачити наслідки управлінських рішень. Модель має давати чітку та адекватну уяву про досліджувану структуру. Визначення вірного балансу між рівнем адекватності ММ тій реальній системі, яку вона має описувати, та можливостями одержати за використання моделі практично цінні розв'язки є складним і важливим завданням, яке вимагає як високого фаху, так і змістовного аналізу відповідної системи. Не можна розбудувати універсальну методику вибору та створення ММ за будь-якого конкретного випадку. Найвагомішою допомогою у розв'язанні питання математичного моделювання є знання про ММ аналогічних систем із літературних джерел і досвід.

Ретельний якісний аналіз завжди мусить передувати кількісному аналізу, щоб з'ясувати ті фактори, які є визначальними для досліджуваної проблеми або задачі. Необхідно мати попередні уявлення про основне призначення ДС та управлінських рішень, про критерій оцінки ефективності можливих альтернатив. По завершенні якісного аналізу треба мати змогу визначити: а) параметри стану, тобто ті, які ідентифікують ДС; б) сенс параметрів управління, тобто тих величин, які визначають *можливий стан* ДС; в) екзогенні змінні – величини, значення яких задані поза моделлю; г) ендогенні змінні – величини, значення яких обчислюються при використанні моделі.

Розподіл змінних і параметрів ММ на вказані класи умовний; класи частково можуть перетинатися.

Приклад. Необхідно спланувати обсяги випуску продукції декількох найменувань за умови відомих обсягів необхідних ресурсів і витрат кожного з використовуваних видів ресурсів на одиницю кожного з найменувань продукції. Задані також ринкові ціни продукції одиниці кожного найменування. Зиск після реалізації має бути максимальним.

Позначимо *загальну кількість найменувань* продукції літерою n, а окреме найменування - індексом j, тоді *обсяги продукції* кожного найменування зручно позначити x_j $(j=\overline{1,n})$, а відповідні ціни c_j. За використаних позначень величина зиску f після реалізації визначається залежністю

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

Наведена функціональна залежність показує, що чим більші обсяги випуску продукції, тим більший зиск, бо величини x_j та c_j за своїм змістом не можуть бути

від'ємними. Але слід врахувати, що обсяги ресурсів задані, тобто обмежені, а за такої умови й обсяги випуску продукції кожного найменування будуть обмеженими. Ці вимоги необхідно описати аналітичними залежностями. Позначимо *загальну кількість* ресурсів літерою m, а кожний ресурс - індексом i, тоді *обсяги кожного з ресурсів* доцільно позначити b_i $(i=\overline{1,\ m})$, а відповідні витрати на одиницю продукції кожного найменування – a_{ij} $(i=\overline{1,\ m};\ \ j=\overline{1,\ n})$. Узгодженість обсягів ресурсів та їх витрат на можливі обсяги продукції x_j аналітично буде представлена так:

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i=\overline{1,\ m}).$$

Задача пошуку оптимального функціонування системи "виробництво – реалізація продукції" за вище названих умов в термінах математичного апарату формулюється так: знайти такий план обсягів x_j виробництва продукції кожного найменування, щоб величина зиску

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

досягала найбільшого значення за умови виконання відношень:

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i=\overline{1,\ m});$$

$$x_j \ge 0 \quad (j=1,\ n).$$

За вище названих умов параметрами стану системи будуть a_{ij}, b_{ij}, c_j $(i=\overline{1,\ m};\ j=\overline{1,\ n})$, параметрами управління x_j $(j=\overline{1,\ n})$. Параметри a_{ij}, b_{ij}, c_j утворюють клас екзогенних величин, їх значення мусять бути заданими для розбудови пропонованої ММ; змінні x_j $(j=\overline{1,\ n})$ – клас ендогенних величин, їх значення обчислюють при використанні ММ. Якщо необхідно змоделювати процес вибору можливих технологій виробництва, то параметри a_{ij} доцільно віднести до класу ендогенних, або управління.

Математична модель є відображенням співвідношень складових елементів досліджуваної системи у вигляді аналітичних залежностей і логічних зв'язків. Досить часто за розбудови ММ необхідно моделювати певні технологічні процеси, моделі яких є частинами загальної моделі ДО. У загальному випадку модель віддзеркалює взаємозв'язки між параметрами управління, параметрами стану, технологічними параметрами та показниками ефективності в кількісній мірі. Якісно розбудована модель є запорукою успіху розробки та впровадження ОС; вона може значно поліпшити ефективність функціонування ДС або її економічні показники. Починаючи розбудову моделі, необхідно перш за все з'ясувати питання можливості використання в ній певних показників і відношень. Якщо в більшості задач виробництва порівняно легко оцінити залежність між приростом обсягів виробництва та відповідними результуючими показниками, то оцінки впливу певних заходів на поведінку людей зробити набагато складніше.

Можна, хоч і складно, емпірично оцінити, як пов'язані обсяги складських запасів і величина ймовірного дефіциту, але набагато важче оцінити вплив реклами на обсяги збуту, хоч немає жодних підстав заперечувати такий зв'язок.

При розбудові *операційних* моделей доцільно розглянути три взаємопов'язані напрями. **Перший** стосується питань оптимізації, тобто поліпшення організаційного управління за одним або кількома критеріями з урахуванням певних обмежень на можливі засоби таких поліпшень. Для розв'язання задачі оптимізації виробничо-господарських процесів розбудована значна кількість математичних моделей. Стосовно управлінських структур це питання набагато складніше, а можливостей значно менше.

Другий напрям обумовлений визначенням чутливості ММ до зміни величин параметрів моделі та вхідних даних. Якщо при проектуванні технічних систем прагнуть досягти їх найбільшої стабільності, тобто зменшити відхилення від стаціонарних умов експлуатації, то вартість операційної системи в значній мірі визначається рівнем реагування на зміни зовнішніх умов, модельованих вхідною інформацією. *Нечутлива система не має права на існування.* Як приклад можна навести моделі управління запасами. При розбудові таких моделей намагаються визначити стратегії в залежності від наявних запасів і прогнозу попиту, тобто дати чіткі рекомендації, що необхідно робити в залежності від рівнів запасів і попиту.

Третій напрям обумовлений визначенням взаємозв'язків окремих складових тієї організаційної структури, в якій буде працювати ОС. Рекомендації операційного аналізу мають бути узгодженими з інформаційною, керівною та виконавчою структурами, поза межами їх функціонування вони не мають сенсу. Вимоги та умови узгодженості ОС із середовищем її функціонування є найвагомішою частиною проекту ДО, розбудова якої завжди має надто специфічні особливості, вимагає високого фаху, тому важко піддається формалізації, що обумовлюється відсутністю достатньої фахової літератури.

При розбудові *моделі операційного дослідження* необхідно ретельно з'ясувати найменші дрібниці, визначити зміст і призначення вхідної інформації та мати хоча б загальну уяву про сенс вихідної інформації й її використання. Доцільно розмежувати статичні та динамічні структурні елементи й описати в математичній формі взаємозв'язки між ними. Необхідно також визначити так званий "горизонт терміну",

тобто той проміжок часу, в межах якого можуть проявлятися наслідки організаційно-управлінських рішень. Це дозволить оцінити критерії ефективності різних можливих варіантів розв'язання відповідних питань. Визначення горизонту терміну впливає на зміст обмежень, яким мусять задовольняти параметри моделі, бо при значному терміні можна, як правило, нехтувати обмеженнями короткотермінової значимості.

Наведемо короткий загальний аналіз різних типів визначення взаємозв'язків між змінними та параметрами моделі.

Відношення за означеннями. Такі аналітичні залежності визначаються або загальновизнаними правилами економічно-господарської діяльності, або певними фізичними законами. Наприклад, наявність обсягів товару на момент t визначається наявністю обсягів запасів на момент попереднього їх заміру, доданих до обсягів виробництва та зменшених на величину обсягів збуту за контрольований проміжок. Такий взаємозв'язок легко й однозначно при необхідності можна записати у вигляді формули.

Емпіричні залежності, тобто залежності за даними експериментів, визначають на основі вивчення даних за попередній проміжок, аналізу технічних і технологічних аспектів, експериментальних даних, правових вимог, а також певних припущень (гіпотез). Прикладом таких залежностей може бути схема нарахування величини податку, яка будується за правилом: величина податку P дорівнює $\lambda\%$, якщо величина річного прибутку A не більше S, а для частини прибутку, більшої S, величина P зростає і дорівнює $\beta\%$ $(\beta > \lambda)$.

При такій схемі оподаткування величина P визначається залежністю:

$$P = \begin{cases} \lambda \cdot 0{,}01A, \text{ якщо } A \le S, \\ \lambda \cdot 0{,}01S + \beta \cdot 0{,}01(A - S), \text{ якщо } A > S. \end{cases}$$

Емпіричні залежності можуть бути *детермінованими*, тобто однозначно визначеними для кожного конкретного набору змінних, або *стохастичними*, числові величини яких обчислюються з використанням статистичних даних.

Визначення емпіричних залежностей часто ускладнюється такими причинами. По-перше, дані аналізу минулого можуть бути непрезентативними для досліджуваної системи через зміни у зовнішніх умовах. Нові умови функціонування можуть вимагати зміни і кількісних вимірів, і змісту залежностей. По-друге, емпіричні залежності стосувалися умов функціонування, які не відповідали можливому рівню ефективності.

Нормативні залежності визначають, які змінні та як мають бути пов'язані між собою, і віддзеркалюють ті директивні вимоги, яким мусить бути підпорядкована досліджувана система. Наприклад, планування обсягів випуску певного виду

продукції може бути обмежене прогнозованою величиною збуту або вимогою використання певного наявного ресурсу виробництва в повному обсязі.

На завершальному етапі розбудови математичної моделі треба мати точні аналітичні залежності. Це має бути обов'язковою передумовою вибору обчислювального методу. Питання розбудови ММ більш детально описані у відповідних главах.

2.6. Інформаційне забезпечення та вибір числових методів

Інформаційне забезпечення ефективного функціонування ОС будується за принципом "банк даних", вся необхідна інформація зберігається на технічних носіях: дисках, магнітних стрічках і т. ін. Пошук і виклик інформації виконується за допомогою спеціальних програм. Для розв'язання кожної поставленої задачі формуються так звані робочі інформаційні масиви, саме їх дані є вхідною інформацією для розрахунків коефіцієнтів відповідної ММ і результатів цільового призначення ОС.

Має бути розроблена методика аналізу та використання інформації, одержаної після розв'язання кожної задачі, визначення, яку інформацію і в якому вигляді необхідно друкувати, яка інформація являє собою проміжні результати і має бути записана в "банк даних".

Оновлення та коригування інформації обумовлені можливою заміною технічного оснащення виробництва, технологічних процесів, структурно-організаційних форм, директивних документів і т. ін. У залежності від значимості інформації визначають терміни їх зберігання.

Для надійного й економного інформаційного забезпечення використовують уніфіковані системи планових документів і кодифікатори інформації, що й складає в сукупності інформаційне забезпечення розрахунків за ОС. Характер інформації визначається змістом і призначенням ОС та математичними методами, які забезпечують їх функціонування.

Завершивши розбудову ММ і визначивши зміст вхідної та вихідної інформації, яка має бути використана при прийнятті управлінських рішень, необхідно вибрати на комп'ютері числовий метод реалізації ММ або їх сукупності. Існує значна бібліотека числових методів розв'язання відповідних математичних задач. Вибираючи числовий метод або їх сукупність, слід враховувати такі аспекти:

а) доцільність використання імітаційного моделювання чи певних методів оптимізації;

б) доцільність використання статистичних даних для обчислення деяких величин чи прийняття їх як детермінованих;

в) допустимість лінеаризації при наявності в ММ нелінійних залежностей;

г) здатність наявної бібліотеки числових методів забезпечити необхідні обсяги розрахунків.

Тобто слід провести певний аналіз адекватності числових методів і розбудованої ММ. Особливо треба простежити, щоб спрощення, які приймаються при виборі або розбудові числового методу, не спотворили зміст розбудованої ММ. Вибір числових методів вимагає високого фаху виконавців та ретельної перевірки цих методів на тестових задачах, тобто таких, розв'язок яких відомий або може бути легко та безпомилково знайдений. При розв'язанні питання вибору числового методу часто виникає дилема:

а) знайти точний розв'язок спрощеної ММ;

б) знайти наближений розв'язок сформульованої в повному обсязі ММ.

Практичний досвід однозначно свідчить, що перевагу слід надавати другому варіанту. Розрахунки за моделлю, в якій враховані в більшому обсязі взаємозв'язки важливих факторів, завжди практично більш корисні, ніж точне рішення спрощеної моделі.

2.7. Розробка технічного завдання, програмування та відладка

Поза сумнівом, без використання комп'ютерів неможливо забезпечити функціонування практично значимої операційної системи. Складання програми є важливою частиною реалізації проекту ДО. Хоч воно виконується, як правило, фахівцями відповідного профілю, але відповідає за якість експлуатаційних властивостей програми розробник ОС. Розробка технічного завдання має бути виконана максимально ретельно, з чітко окресленими вимогами до форми та змісту вхідних і вихідних даних, етапів контролю, постійної та змінної інформації. Детальна, акуратна розробка технічного завдання значно допоможе якісно оформити програму та інструкцію по її використанню, що сприятиме більш швидкому впровадженню ОС і зручності в її експлуатації. Особливу увагу необхідно приділити змісту та формі представлення (документам) вхідної та вихідної інформації, їх погодженню з керівництвом і працюючими відповідними структурами. Вихідні дані мають бути однозначно зрозумілими, добре та зручно розташованими, максимально скорочувати пошук необхідної інформації та максимально забезпечувати визначення ефективності різнопланового призначення функціонуючої структури.

Орієнтація на те, що вхідну інформацію для функціонування ОС буде готувати користувач, означає, що її підготовка має бути максимально зручною, давати можливість щонайшвидшої підготовки та оновлення або заміни її частини; функції її сортування, перевірки за формою, а по можливості й за змістом має виконувати комп'ютер.

Складання нових програм або вибірка та узгодженість уже наявних програм за

рухом інформації – то справа відповідних фахівців. Але розробник проекту ДО завжди мусить розуміти свою відповідальність за ОС і ретельно та постійно контролювати як етапи програмування, так і здатність завершеної програми, особливо можливість її адаптації, до певної зміни вхідної інформації.

2.8. Накопичення даних

На етапі накопичення даних основною метою є підготовка й аналіз інформації, необхідної для перевірки правильності моделі та можливостей практичного використання результатів впровадження проекту ДО. Збирання та накопичення даних на попередніх етапах виконувалися з метою визначення проблеми дослідження та розбудови моделі ОС. Суттєвим питанням накопичення інформації є її точність та подання. Розв'язуючи його, треба з'ясувати, якої точності вимагають від вихідної інформації, використовуваної для прийняття управлінських рішень, наскільки ММ чутлива до рівня точності вхідних даних і наскільки важлива точність вхідної інформації для мети ОС. Наприклад, при дослідженні планування обсягів певного виробництва з урахуванням обсягів прогнозованого попиту слід враховувати випадковість величини обсягів попиту: іноді вона може бути визначена з точністю до 50%. За названих причин, накопичуючи дані про досліджуване виробництво в минулому: продуктивність виробництва, простої через технічні неполадки, питома вартість одиниці продукції і т. ін., не варто прагнути одержати їх з точністю до 1%. Межі їх точності ±20% не будуть визначальними для дієздатності моделі. Доцільно також наголосити, що досить часто рекомендації ОС стосуються прийняття таких рішень, які вже були в минулому за подібних чи близьких обставин, тому треба знати, які дані та з якими значеннями були використані. Можлива ситуація, коли рішення стосовно використання певних даних приймалися без належного обгрунтування – навмання. За такої ситуації треба очікувати, що врахування відповідної інформації поліпшить ефективність структури, а її керівництво треба переконати в необхідності та доцільності використання такої інформації.

2.9. Перевірка дієздатності моделі

Основними складовими цього заходу є спосіб перевірки дієздатності моделі та її реалізація. Спочатку вибирають аналітичні та експериментальні методи перевірки *узгодженості* з дійсністю результатів розрахунків за моделлю, *чутливості та дієздатності моделі.* Для перевірки моделі використовують дані, накопичені на раніше названому етапі. Наслідки перевірки засвідчать або її дієздатність і доцільність використання за умови доопрацювання, або необхідність перебудови моделей і розробки нової програми.

Узгодженість. Переконавшись у достовірності та необхідному рівні вхідної

інформації, по завершенні розрахунків перевіряють їх практичний сенс, відповідність здоровому глузду та логіці, якщо параметри вхідної інформації змінювати в певних межах, а також оцінку результатів розрахунків керівниками та фахівцями відповідної галузі. Якщо можна, то одержані результати порівнюють з результатами розв'язання аналогічних задач, можливо, за дещо зміненими моделями, іноді з'являється можливість зробити певні граничні переходи в моделі та одержати уже добре відомі й перевірені практикою результати.

Перевірка моделі на чутливість обов'язково виконується при використанні числових методів в розрахунках. Мета заходу – перевірити характер зміни певної вихідної інформації при зміні вхідної інформації в певних межах з відомим кроком. Позитивним результатом перевірки є додаткове свідчення вірогідності адекватного віддзеркалення в моделі реально існуючих закономірностей в роботі досліджуваної структури.

Реальність. Мета запровадження будь-якої ОС – поліпшити поточне функціонування структури та мати можливість оцінити наслідки управлінських рішень за певних умов у близькому або більш віддаленому майбутньому. Як переконатися в практичній доцільності рекомендацій за ОС? Одним з ефективних засобів такої перевірки є дослідження за допомогою ОС ситуацій, які були в минулому, та порівняння прийнятих управлінських рішень з рекомендаціями, запропонованими ОС за відповідних умов у минулому.

Дієздатність ОС визначає, наскільки оперативно та з якими затратами коштів і праці можна одержати інформацію, потрібну для оперативної діяльності або прогнозування майбутнього. Особлива увага має бути приділена здатності моделі давати реалістичні рекомендації при врахуванні зовнішніх умов для оцінки одноразових рішень за важливістю їх наслідків, наприклад, оцінки доцільності оновлення старого або організації нового виробництва. Помилки у висновках і рекомендаціях щодо поточних рішень набагато менш небезпечні, ніж при одноразових рішеннях. Найбільш вагомим моментом перевірки моделей прийняття одноразових рішень є аналіз і критична оцінка одержаних результатів по завершенні розрахунків. Тобто, розв'язавши задачу з використанням моделі, фахівець ДО ще багато чого мусить уточнити та перевірити, перш ніж подасть керівництву альтернативні варіанти, їх оцінки, обґрунтування відповідних висновків і порад. Виконання робіт такого змісту вимагає часу та кропіткої праці, але, безумовно, допомагає більш глибокому й осяжному розумінню досліджуваної проблеми та істотно збільшує ймовірність уникнення помилок.

Проблема впровадження та використання. На шляху впровадження проекту ОС завжди виникає чимало труднощів, обумовлених як можливими недоліками та вадами самого проекту, так і необхідністю перебороти звичний комформізм тих, хто буде працювати з використанням ОС. Ефективність впровадження ОС суттєво залежить від того, в якій мірі вдалося організувати співпрацю керівників різних ланок структури з виконавцями проекту. Перешкоди в багатьох випадках обумовлені не вагомими причинами, а лише відсутністю розуміння, певною амбіціозністю і небажанням будь-яких змін. Такі перепони можна успішно усунути лише певною наполегливістю та аргументованими доказами майбутньої користі від впровадження ОС.

Для успішного впровадження проекту реалізації ОС необхідно підготувати матеріали для ознайомлення керівників відповідних ланок з принципами функціонування впроваджуваної ОС, для проведення семінарів і переконливої демонстрації результатів, які можна досягти впровадженням проекту. Такі заходи сприяють більш ефективній співпраці, взаєморозумінню та взаємодопомозі в процесі впровадження ОС. Доцільно наголосити також на безумовній необхідності враховувати наявність неминучих збурень, обумовлених вирішенням нагальних питань впровадження ОС, бо таких ситуацій практично не можна уникнути. Фахівці з дослідження операцій радять: завжди, якщо є така можливість, порівнювати результати, одержані при використанні ОС і традиційних для структури методів. Для успішного проведення таких заходів фахівець, який упроваджує ОС, повинен мати:

1) ретельно розроблений план виконання випробувань;
2) порівняння даних, одержаних при використанні ОС і фактичних, завчасно перевірених методом оцінки;
3) засоби та можливості накопичення тих даних, порівняння й оцінка яких передбачені планом;
4) ресурси часу та фінансові, необхідні для внесення відповідних коректив у ОС;
5) принципову згоду керівництва відповідної структури на виконання аналогічних випробувань у майбутньому.

Представники сторін, які співпрацюють по впровадженню проекту ОС, повинні чітко розуміти необхідність певних зусиль для усунення можливих недоліків, які в будь-якому прояві завжди властиві новому. Світовий досвід переконливо засвідчує безумовну доцільність та ефективність використання наукових досягнень в організації функціонування структур різнопланового призначення.

Питання для самоконтролю

1. Мета та цілі розбудови проекту дослідження операцій в певній структурі.
2. Що таке стратегія досягнення цілі?
3. Основні етапи розбудови проекту дослідження операцій.

4. Різниця між поняттями математичної моделі та моделі дослідження операцій.
5. Які фактори визначають проблеми дослідження операцій?
6. За яких умов доцільно користуватися детермінованими та стохастичними математичними моделями?
7. Як ви розумієте вислів "інформаційне забезпечення ОС"?
8. Чи є відмінність між інформацією, використовуваною при розбудові проекту ДО, та інформацією, яка забезпечує функціонування ОС?
9. Що таке технічне завдання на програмування? Що одержують по завершенні етапу програмування?
10. Як можна перевірити дієздатність та ефективність ОС?
11. Як ви уявляєте собі місце ОС в управлінні структурою?
12. Які загальні вимоги необхідно виконувати при накопиченні вхідної інформації для поточного функціонування ОС?
13. Які загальні вимоги необхідно виконувати стосовно подання вихідної інформації?

ГЛАВА 3

ВИРОБНИЧІ ФУНКЦІЇ

3.1. Поняття виробничої функції

Зростаючі запити суспільства вимагають розширення номенклатури товарів і послуг, що може бути забезпечене впровадженням високоефективних технологій і виробничих процесів. У розмаїтті товарів і послуг доцільно виділити досить стабільні групи за їх функціональним призначенням. Доцільно об'єднати у групи і використовувані ресурси та технології. Такий підхід до аналізу виробничо-економічної діяльності дозволяє розробити теорії розбудови спеціальних економіко-виробничих математичних моделей дослідження виробничих процесів – так званих *виробничих функцій* (ВФ). *Мета розбудови виробничих функцій* – проаналізувати і визначити аналітичну залежність у кількісних вимірах між використовуваними ресурсами та обсягами продукції.

При розбудові виробничих функцій розглядаються тільки найбільш суттєві фактори виробництва. Як факторні ознаки використовуються: вартість основних виробничих фондів, дози внесення органічних і мінеральних добрив, забезпеченість трудовими ресурсами тощо. Результати виробництва відображаються в показниках валової та товарної продукції, прибутку, продуктивності праці тощо.

Виробничі функції застосовуються для *аналізу, планування та прогнозування* на різних рівнях управління: в сільськогосподарських підприємствах, в районах та областях за даними сукупності підприємств, у галузі в цілому, в усьому народному господарстві. При розбудові виробничих функцій на галузевому рівні і в цілому по народному господарству використовуються як результативні показники валовий суспільний продукт, національний прибуток тощо.

Виробнича функція, як і відповідне поняття математики, може бути представлена різними способами. Широко використовується табличний спосіб представлення ВФ з одним, двома або трьома входами в залежності від кількості змінних, використовуваних для побудови функції. Наприклад, якщо ВФ є функцією однієї змінної, то табличний спосіб її визначення має вигляд:

Значення аргументу	x_1	x_2	...	x_n
Значення функції	$f(x_1)$	$f(x_2)$	...	$f(x_n)$

ВФ двох аргументів таблично описують за такою схемою побудови таблиці: в крайньому лівому стовпці та у верхньому рядку записують значення аргументів, а на перетині відповідних рядків і стовпців – значення функції, як це показано в табл. 3. 1, в якій аргументи названі відповідно x та y.

Таблиця 3.1.

$x \backslash y$	y_1	y_2	…	y_n
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	…	$f(x_1, y_n)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	…	$f(x_2, y_n)$
⋮	⋮	⋮	⋮ ⋮ ⋮	⋮
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	…	$f(x_m, y_n)$

При необхідності задати функцію трьох змінних таблично роблять так: для кожного фіксованого значення одного з аргументів записують таблицю значень функції для двох інших аргументів, як це показано в табл. 3.1. Тоді ВФ трьох аргументів буде представлена множиною таблиць типу 3.1 для кожного значення третього аргументу. Схема побудови таких таблиць показана в табл. 3.2, якщо впорядкування виконати за аргументом x.

Таблиця 3.2.

$x = x_1$					$x = x_2$				
$y \backslash z$	z_1	z_2	…	z_n	$y \backslash z$	z_1	z_2	…	z_n
y_1	$f(x_1,y_1,z_1)$	$f(x_1,y_1,z_2)$	…	$f(x_1,y_1,z_n)$	y_1	$f(x_2,y_1,z_1)$	$f(x_2,y_1,z_2)$	…	$f(x_2,y_1,z_n)$
y_2	$f(x_1,y_2,z_1)$	$f(x_1,y_2,z_2)$	…	$f(x_1,y_2,z_n)$	y_2	$f(x_2,y_2,z_1)$	$f(x_2,y_2,z_2)$	…	$f(x_2,y_1,z_n)$
⋮	⋮	⋮	…	⋮	⋮	⋮	⋮	…	⋮
y_m	$f(x_1,y_m,z_1)$	$f(x_1,y_m,z_2)$	…	$f(x_1,y_m,z_n)$	y_m	$f(x_2,y_m,z_1)$	$f(x_2,y_m,z_2)$	…	$f(x_2,y_m,z_n)$
$x = x_3$					$x = x_4$				
$y \backslash z$	z_1	z_2	…		$y \backslash z$	z_1	z_2	…	
y_1	$f(x_3,y_1,z_1)$	$f(x_3,y_1,z_2)$	…		y_1	$f(x_4,y_1,z_1)$	$f(x_4,y_1,z_2)$	…	
⋮	⋮	⋮	…		⋮	⋮	⋮	…	

Табличний спосіб представлення ВФ доцільно використовувати при необхідності запису наслідків безпосередніх спостережень, тобто накопичення даних для подальшого використання.

Використовуючи термін "функція" в розбудові математичних моделей, доцільно наголосити на відмінностях його тлумачення в деяких природничих науках та в економіці. У багатьох випадках при дослідженні зв'язків між різними взаємопов'язаними факторами в природничих науках досить чітко та однозначно можна простежити *функціональну* залежність, коли *певному значенню однієї змінної* відповідає *конкретне значення іншої змінної* (функції). Наприклад, швидкість падіння тіла у вакуумі в залежності від часу, число обертів певного двигуна внутрішнього згорання – від кількості пального, вага певної рідини – від її об'єму при фіксованій

температурі і т. ін. В економіці, а також в інших науках у більшості випадків між змінними величинами спостерігаються залежності, коли певному значенню однієї змінної може відповідати не одне фіксоване значення іншої змінної, а деяка *допустима множина* значень іншої змінної. Така допустима множина, як правило, підлягає певним законам упорядкування. За таких умов залежність між досліджуваними величинами називається **статистичною**, або **стохастичною** (ймовірною).

Необхідність використання моделі статистичної залежності обумовлена тим, що відповідні ММ розбудовуються за умов, коли *значення залежної змінної* об'єктивно визначаються не лише *значенням незалежної змінної*, а й багатьма випадковими, неконтрольованими факторами, кожний з яких окремо не впливає та й не може істотно вплинути на величину значення залежної змінної, але сукупність таких факторів обумовлює певні її зміни у деяких межах. Прикладами статистичної залежності є залежність урожайності від кількості внесених добрив, щоденний попит на певний вид продукції харчування у певному районі тощо.

Неможливість установлення однозначності між порівнюваними величинами досліджуваних взаємопов'язаних факторів обумовлює необхідність пошуку закономірностей зміни їх усереднених величин. При такому підході до дослідження взаємозв'язків між факторами кожному *можливому значенню незалежної змінної ставиться у відповідність математичне сподівання величини залежної змінної.*

Кореляційною залежністю між двома змінними (або більшим числом змінних) називається функціональна залежність між значеннями однієї з них і математичним сподіванням іншої. Користуючись звичайними математичними позначеннями множини значень незалежної змінної X та залежної змінної Y, а їх елементів – відповідно x та y, кореляційну залежність у загальному тлумаченні можна записати так:

$$M_x(Y) = \varphi(x) \qquad (1)$$

($M_x(Y)$ – математичне сподівання величини Y за умови, що величина X отримає конкретне значення x (якщо досліджується кореляційна залежність від кількох змінних, x доцільно розглядати як точку в n-мірному просторі), φ – символ відповідної функціональної залежності).

Рівняння (1) називається рівнянням регресії Y по X, а його графічне зображення – лінією регресії. Теоретично для розбудови рівняння регресії необхідно знайти закон розподілу двомірної випадкової величини (X; Y). На практиці маємо лише скінченну вибірку пар значень (x_i; y_i). За таких умов можна дати лише наближену оцінку рівняння регресії (1), використовуючи вибіркові дані. Методику розбудови таких оцінок буде наведено при розгляді загальних властивостей ВФ.

Аналітичний спосіб зображення ВФ, тобто з допомогою формули, яка визначає послідовність певних дій з аргументами для одержання відповідного значення функції, є найбільш ефективним і широко використовуваним у теоретичних і прак-

тичних дослідженнях. Для *конкретного* аналітичного представлення ВФ завжди використовуються дані спостережень.

Статистичні закономірності, використовувані для розбудови ВФ досліджуються методами *кореляційного* та *регресійного аналізу*. *Основне завдання кореляційного аналізу* – дослідження взаємозв'язку між статистичними даними (спостереженнями) та оцінка його величини. *Основне завдання регресійного аналізу* – розбудова форми аналітичної залежності між досліджуваними факторами та оцінка її узгодженості з даними спостережень.

Розглянемо методи побудови аналітичного зображення ВФ за умови наявності відповідних даних спостережень.

3.2. Виробничі функції однієї змінної

З точки зору економічного використання, ВФ є функцією, незалежна змінна якої набуває значення обсягів використовуваного ресурсу (фактора виробництва), а залежна змінна, тобто функція, – значення обсягів виробництва продукції або інших показників цілеспрямованої діяльності. Запис $y = f(x)$ означає: якщо ресурс буде використаний в обсязі x певних одиниць виміру, то продукції буде одержано в обсязі $y = f(x)$ одиниць відповідного виміру. Символ f (знак функції) є математичною моделлю виробничої системи, яка перетворює ресурс у продукцію. У мікроекономіці прийнято вважати, що y – це детермінована можлива величина обсягу продукції, хоч значення її одержано, як правило, після усереднення даних спостереження, якщо ресурс витрачається в обсязі x одиниць. У макроекономіці ВФ – це *стабільно статистичний зв'язок* між затратами та випуском. В обох випадках області визначення та значення ВФ є невід'ємні числові множини $(x \geq 0;\ y \geq 0)$.

Побудова виробничих функцій. Якщо ВФ необхідно використовувати при виконанні певних досліджень, то вона має бути представлена аналітичною залежністю. У кожному конкретному випадку така залежність повинна забезпечувати найбільш точне моделювання досліджуваного економічного процесу або його складових. Зміст виразу "найбільш точно моделювати" визначається тлумаченням міри точності.

Побудові аналітичного зображення ВФ (математичної моделі) має передувати ретельний якісний та певний кількісний аналіз досліджуваного економічного процесу. Такий аналіз може бути виконаний лише за умови достовірності експериментальних даних стосовно досліджуваного процесу. Наслідком попереднього аналізу експериментальних даних має бути висновок про загальний характер зміни залежної та незалежної величин. Рекомендується зображати експериментальні дані у відповідній системі координат. Якщо маємо статистичну залежність, то мусимо розглядати

усереднені значення. Подальший процес розбудови ВФ доцільно розділити на такі етапи:

– вибір виду формули можливої аналітичної залежності шуканої ВФ від зміни незалежної величини; таких видів формул може бути декілька;

– вибір міри точності узгодження аналітичного зображення ВФ з експериментальними даними;

– на множині орієнтовно вибраних видів формул аналітичної залежності за критерієм міри точності узгодження експериментальних даних з аналітичним зображенням ВФ вибирається певна конкретна формула.

Опишемо більш детально вказані етапи, вважаючи, що маємо *достовірні* експериментальні дані. Спочатку розглянемо приклад. Припустимо, що маємо експериментальні дані, представлені в табл. 3.3 і графічно зображені у прямокутній системі координат на рис. 3.1.

Таблиця 3.3.

x	1	2	3	4	5	6	7
y	2,15	2,8	3,1	3,3	3,5	3,8	4

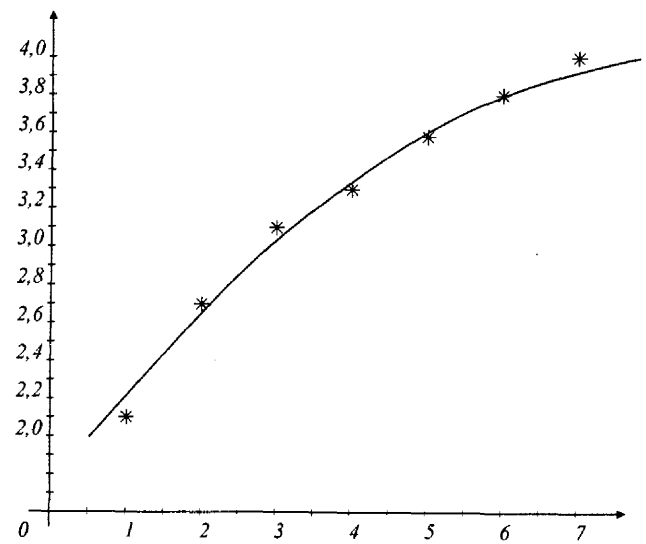


Рис. 3.1.

Аналізуючи графік, доходимо висновку, що аналітичне зображення відповідної функції доцільно шукати у вигляді або $y_1 = \sqrt{ax+b}$, або $y_2 = a\ln x + b$. Названі функції за відповідного вибору параметрів a та b мають такий же вигляд у графічному зображенні залежності від свого аргументу, як і наявні експериментальні дані. Щоб обчислити необхідні значення величин параметрів a та b, треба вибрати критерій оцінки якості наближення експериментальних даних значеннями функції у вигляді аналітичної залежності для відповідних значень незалежної змінної.

Один із найбільш поширених на практиці критеріїв такої оцінки побудований за таким принципом: наближення експериментальних даних значеннями аналітичної функції вибраного виду вважається найкращим, якщо сума квадратів різниць величин експериментальних даних та відповідних значень функції аналітичного представлення є найменшою. Використання цього критерію дає можливість вибрати найкраще наближення характеру зміни даних спостереження певною функцією серед усіх функцій даного виду, тобто функцій, тотожних в аналітичному представленні, але з різними значеннями параметрів. Метод пошуку параметрів вибраного виду функцій за вказаним критерієм називається *методом найменших квадратів* (МНК). Наголосимо, що МНК дає *апарат обчислення параметрів* вибраного виду функцій, а не *апарат вибору виду функцій*. Для нашого прикладу за названим критерієм вибираємо функцію серед функцій двох видів $y_1 = \sqrt{ax+b}$ та $y_2 = a\ln x + b$, обчисливши параметри a та b, які для функції кожного виду будуть мати свої значення.

Знайшовши найкращу за вибраним критерієм якості наближення функцію серед *функцій певного виду*, необхідно зробити подальший вибір. Для такого вибору можна скористатися тим же критерієм: найкращою є функція того виду, для якої сума квадратів різниць між експериментальними та аналітичними даними є найменшою. Іноді доцільно використати інші критерії. Наприклад, вибирають ту функцію, для якої модуль найбільшої різниці між експериментальними даними та відповідними аналітичними значеннями має менше значення.

Для запропонованого прикладу зображення конкуруючих функцій буде таким: $y_1 = \sqrt{2x+3}$, $y_2 = 0{,}85\ln x + \sqrt{5}$. Величини різниць між експериментальними даними та відповідною аналітичною апроксимацією даними наведені в табл. 3.4.

Таблиця 3.4.

x	1	2	3	4	5	6	7	max Δy
y (експер.)	2.15	2.8	3.1	3.3	3.5	3.8	4.0	
y_1	2.236	2.646	3	3.317	3.606	3.873	4.123	
$\lvert y_1 - y \rvert$	0.086	0.154	0.1	0.017	0.106	0.073	0.123	0.154
y_2	2.236	2.825	3.17	3.414	3.604	3.759	3.890	
$\lvert y_2 - y \rvert$	0.086	0.025	0.07	0.114	0.104	0.041	0.11	0.114

Аналіз результатів показує, що серед двох конкуруючих функцій з вибраним критерієм оцінки якості за кращу треба взяти функцію $y_2 = 0{,}85\ln x + \sqrt{5}$.

За будь-яких умов ВФ може бути побудова лише з використанням експериментальних даних. Вибираючи залежну змінну (найменування функцій), необхідно керуватися метою планованих досліджень. За незалежну змінну, як правило, вибирають такий показник, який найбільш істотно впливає на досліджувану функцію, на певну характеристику виробничого процесу. Так, за міру ефективності виробництва вибирають собівартість одиниці продукції, або величину зиску на одиницю вкладених коштів, або обсяги виробництва в залежності від основних фондів.

Для відібраних для дослідження показників необхідно мати відповідні кількісні дані спостережень. Щоб одержати такі дані, треба зібрати й упорядкувати необхідну інформацію, виконавши спеціальні експерименти й обробивши відповідні статистичні матеріали. Розбудова певної виробничої функції доцільна лише за умови достовірності використаних експериментальних даних.

Одним із важливих етапів розбудови ВФ є *вибір форми аналітичної залежності* між незалежною та залежною змінними. Форма залежності має якомога точніше віддзеркалювати основну тенденцію перебігу досліджуваного процесу, бути зручною в подальшому використанні, не бути занадто складною за представленням. Незважаючи на те, що стосовно вибору форми аналітичного зображення ВФ немає однозначних рекомендацій, є досить ефективні методи аналізу експериментальних даних, які дозволяють проводити цілеспрямований пошук форми аналітичного представлення. Але слід зважати на те, що математична інтерпретація економічних процесів занадто складна і в кожному конкретному випадку можна з рівнозначними доказами одержати різні форми аналітичної залежності. Для вибору конкретної форми необхідно в таких випадках виконати додаткові розрахунки. При виборі форми аналітичної залежності досить часто використовують характер зміни перших, других та різниць більш високого порядку, а також графічне зображення експериментальних даних у певній системі координат. Доцільно також враховувати досвід дослідження аналогічних процесів.

Приклад. З практичного досвіду відомо, що обсяг продукції f та величина витраченого виробничого ресурсу (наприклад, робочого часу) добре узгоджуються за аналітичної залежності ax^b $(a > 0;\ 0 < b \le 1)$, тобто ВФ у цьому випадку доцільно представити вказаною залежністю. Величини a, b – параметри ВФ f в кожному конкретному випадку необхідно обчислювати за експериментальними даними.

Загальний вигляд графіка функції $f(x) = ax^b$ $(a > 0;\ 0 < b \le 1)$ наведено на рис. 3.2. На графіку видно, що при збільшенні величини контрольованого ресурсу обсяг виробництва збільшується, але це збільшення має специфіку: кожна додаткова одиниця ресурсу дає все менший приріст продукції (див. рис. 3. 2). Ця особливість

ВФ віддзеркалює одне з фундаментальних положень економічної теорії, підтверджене широкою практикою, – *закон спадаючої ефективності*. Функція вказаного виду є типовою однофакторною ВФ і широко використовується в практиці.

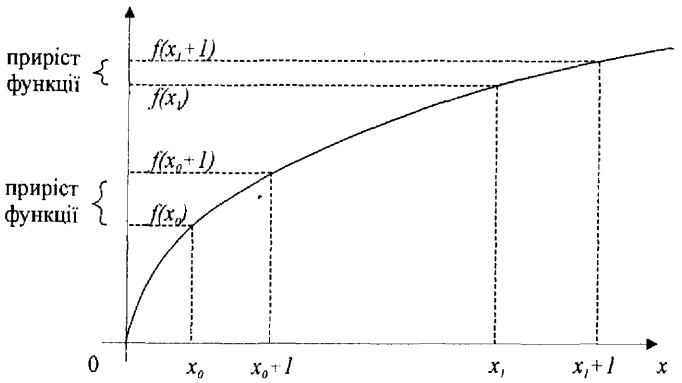


Рис. 3.2.

Дослідження багатьох економічних процесів виконуються за принципом "затрати – випуск" на мікро- та на макроекономічному рівнях. На мікроекономічному рівні ВФ $y = ax^b$ використовується в дослідженнях взаємозв'язку між величинами затраченого певного ресурсу x протягом контрольованого проміжку (місяць, квартал, рік) на окремому підприємстві та обсягом продукції y того ж підприємства. У таких випадках ВФ використовується для розв'язання *задач аналізу та планування*, а також *прогнозування*.

Вказана ВФ використовується і для дослідження взаємозв'язку між затратами певного виду та обсягами продукції або прибутків в межах регіону або країни в цілому. У такому випадку роль виробничої системи виконують регіон або країна. На макроекономічному рівні виробничі функції будують для дослідження питань аналізу, планування та прогнозування господарської діяльності.

Зміст понять затраченого ресурсу та обсягів виробництва продукції, вибір відповідних одиниць виміру визначаються метою виконуваних досліджень та наявністю вихідних експериментальних даних. На мікроекономічному рівні затрати та випуск вимірюються як у відповідних натуральних величинах, так і в певних одиницях вартості. На макроекономічному рівні затрати та їх результативність вимірюються, як правило, у вартісних показниках.

Побудова однофакторних ВФ (ВФ однієї змінної) детально розроблена з урахуванням як аналітичних аспектів, так і економічного сенсу. У табл. 3. 5 вказані найбільш широко використовувані на практиці види однофакторних функцій та ознаки,

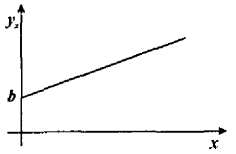
які дозволяють вибрати певну форму аналітичної залежності. Ознаки вибору ґрунтуються на дослідженні зміни перших і других різниць експериментальних даних:

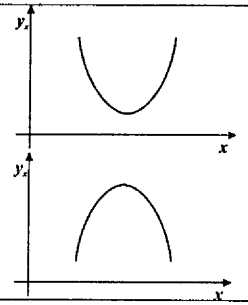
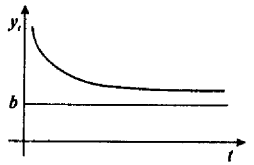
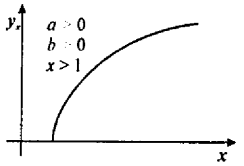
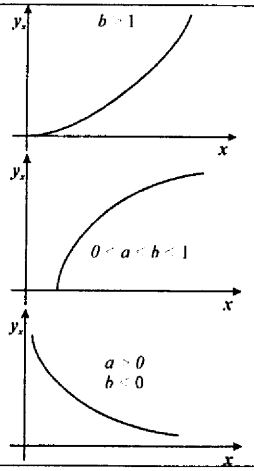
$$\Delta y_i = f(x_i + \Delta x) - f(x_i),$$

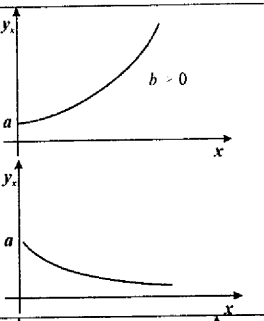
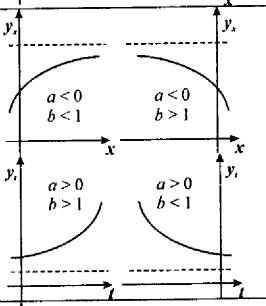
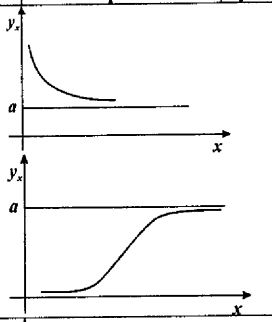
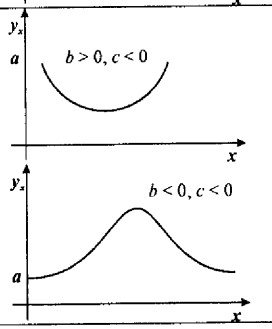
$$\Delta^2 y_i = \Delta(\Delta y_i) = (f(x_i + 2\Delta x_i) - f(x_i + \Delta x)) - (f(x_i + \Delta x) - f(x_i)) =$$
$$= f(x_i + 2\Delta x_i) - 2f(x_i + \Delta x) + f(x_i).$$

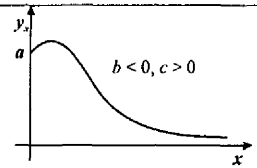
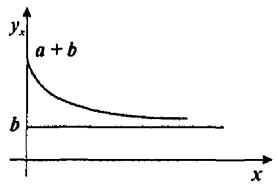
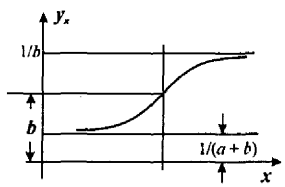
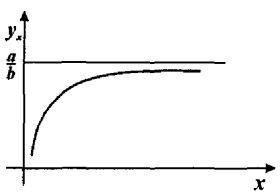
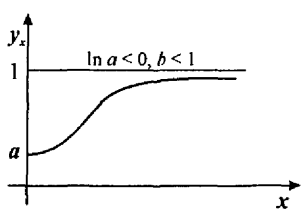
У табл. 3.5 вказані види аналітичних залежностей, їх графічне зображення та критерії вибору відповідної залежності. Таким чином, вибору аналітичної залежності передує певний додатковий аналіз експериментальних даних, який допомагає зосередити увагу подальших досліджень на формулах, краще узгоджених з характером зміни досліджуваних величин. На практиці вибір формули (форми кривої) ґрунтується на порівнянні характеру зміни приростів досліджуваного показника пояснюючого фактора для даних спостереження та аналітичної залежності. Відповідні вимоги стосовно приростів наведено у четвертому стовпці табл. 3.5. Вибір аналітичної залежності відповідає умові, за якої ці характеристики найбільш близькі. Іноді такий вибір не можна зробити однозначно й необхідні додаткові дослідження. Фахівці з питань аналізу даних спостережень та прогнозування вважають, що за усіх інших умов (кількості та похибок вихідних даних і т. ін.) помилка у виборі форми аналітичної залежності найбільш значима за своїми негативними наслідками, особливо при прогнозуванні. Розбудову аналітичного представлення ВФ за даними спостережень описано далі.

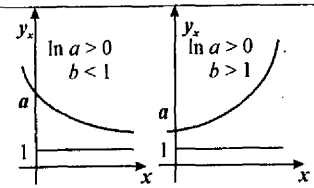
Таблиця 3.5.

№ п/п	Рівняння лінії аналітичної залежності	Графік	Відношення приростів
1	2	3	4
1.	$y = ax + b$		$\frac{\Delta y}{\Delta x} \approx const$

2.	$y = ax_2 + bx + c$		$\frac{\Delta^2 y}{(\Delta x)^2} \approx const$
3.	$y = \frac{a}{x} + b$ $\tau = \frac{1}{x}$ $y = a\tau + b$		$\frac{\Delta y}{\Delta\left(\frac{1}{x}\right)} \approx const$
4.	$y = a \ln x + b$ $\tau = \ln x$ $y = a\tau + b$		$\frac{\Delta y}{\Delta \ln \Delta x} = const$
5.	$y = axb$ $\ln y = b \ln x + \ln a$ $z = \ln y$ $\tau = \ln x$ $z = b\tau + \ln a$		$\frac{\Delta \ln y}{\Delta \ln \Delta x} = const$

6.	$y = aebx$ $\ln y = bx + \ln a$ $z = \ln y$ $z = bx + \ln a$		$\frac{\Delta \ln y}{\Delta x} = const$
6а.	Модифікована експонента $y = k + abx$ $z = \ln(y - k)$ $z = x \ln b + \ln a$		$\frac{\Delta \ln y}{\Delta x} \approx const$
7.	Експонента з насиченням $y = ae^{\frac{b}{x}}$ $\ln y = \frac{b}{x} + \ln a$ $z = \ln y;\ \tau = \frac{1}{x}$ $z = b\tau + \ln a$		$\frac{\Delta \ln y}{\Delta\left(\frac{1}{x}\right)} \approx const$
8.	Логарифмічна парабола $y = ae^{-bx^2+cx}$ $\ln y = bx_2 + cx + \ln a$ $z = \ln y$ $z = bx_2 + cx + \ln a$		$\frac{\Delta^2 \ln y}{(\Delta x)^2} \approx const$

			
9.	$y = ae{-}x + b$ $\tau = e{-}x$ $y = a\tau + b$		$\frac{\Delta y}{\Delta e^{-x}} = const$
10.	Логістична крива, або крива Перла-Рида $y = \frac{1}{ae^{-x} + b}$ $z = \frac{1}{y};\ \tau = e{-}x$ $z = a\tau + b$		$\frac{\Delta\left(\frac{1}{y}\right)}{\Delta e^{-x}} = const$
11.	$y = \frac{a}{b + \frac{1}{x}}$ $z = \frac{1}{y};\ \tau = \frac{1}{x}$ $z = \tilde{b} + \tilde{a}_\tau$ $\tilde{b} = \frac{a}{b};\ \tilde{a} = \frac{1}{a}$		$\frac{\Delta\left(\frac{1}{y}\right)}{\Delta\left(\frac{1}{x}\right)} \approx const$
12.	Крива Гомперца $y = a^{b^x}$ $\ln\ln y = x\ln b +$ $+ \ln\ln a$ $z = \ln\ln y$ $\tilde{a} = \ln b$ $\tilde{b} = \ln\ln a$		$\frac{\Delta\ln(\ln y)}{\Delta\ln x} \approx const$

	$z = \tilde{a}x + \tilde{b}$		

3.3. Багатофакторні виробничі функції

Якщо необхідно визначити вплив кількох факторів на обсяги випуску продукції за певні проміжки часу, то ВФ розбудовують як функцію кількох незалежних змінних і називають її багатофакторною:

$$Y = f(x_1, x_2, ..., x_n, a_1, a_2, ..., a_m) \qquad (2)$$

(Y – скалярна величина; $x_1, x_2, ..., x_n$ – незалежні змінні, величини яких виступають як компоненти n-мірного вектора або як координати точки у n-мірному просторі; $a_1, a_2, ..., a_m$ – параметри певної форми функціональної залежності, числові значення яких обчислюються за експериментальними даними).

За економічним змістом $x_1 \geq 0, x_2 \geq 0, ..., x_n \geq 0$; отже, областю визначення багатофакторної ВФ є множина n-мірних векторів з невід'ємними компонентами.

У мікроекономіці багатофакторні ВФ характеризують обсяг виробництва в натуральному або вартісному вимірі, обумовлений затратами робочого часу, виробничої сировини, комплектуючих, енергії, основного капіталу і т. ін. ВФ такого змісту характеризують з економічної точки зору *технологію* виробництва.

На макроекономічному рівні (дослідження стосовно регіону або країни) обсяг випуску продукції Y характеризують у *стабільних* цінах (якщо спостерігається інфляція, то необхідно зробити відповідні перерахунки); за ресурси (фактори впливу) приймають *основний капітал* K (обсяг *використаного* капіталу на контрольованому проміжку) та *живий продукт* L (обсяг живої праці, затраченої на контрольованому проміжку), представлений його вартістю. Такі змінні використовуються для розбудови двофакторної ВФ. Якщо необхідно врахувати інші аспекти виробничої діяльності, то використовується три- або чотирифакторна ВФ, наприклад, при врахуванні природних ресурсів, досягнень науково-технічного прогресу, витрат на підтримання екології на певному рівні тощо.

Виробнича функція називається *статичною*, якщо її параметри не залежать від часу. Але часто доводиться враховувати залежність і факторів, і ВФ від часу.

Приклад. Для дослідження економічного процесу в межах регіону або країни в цілому широко використовують ВФ у формі $y = a_1\, x_1^{a_2}\, x_2^{a_3}$ (a_1, a_2, a_3 – параметри ВФ, які приймають лише додатні значення; іноді виконується додаткова умова $a_2 + a_3 = 1$). ВФ обраної структури називається ВФ Кобба-Дугласа (ВФКД) – на прізвища двох американських економістів, які в 1929 р. запропонували її використання. ВФКД широко використовується в практичних дослідженнях. Завдяки своїй простоті, вона належить до класу так званих *мультиплікативних* функцій (представлена добутком). У літературі використовуються специфіковані позначення змінних цієї функції:

$$Y = a_0\, K^{a_1}\, L^{a_2} \qquad (3)$$

(K – обсяг використаного основного капіталу (основні фонди); L – обсяг затраченої живої праці).

Розглянемо графік ВФ $y = a_1\, x_1^{a_2}\, x_2^{a_3}$ за умови $a_2 + a_3 = 1$ у прямокутній тримірній системі координат. Якщо $x_1 = 0$ або $x_2 = 0$, то $y = 0$; якщо x_1 та x_2 пов'язані залежністю $x_2 = \lambda x_1$, то $y = a_1\, x_1^{a_2} (\lambda\, x_1)^{a_3} = a_1 \lambda^{a_3} x_1$ за умови $a_2 + a_3 = 1$, тобто для точок (x_1, x_2) будь-якого променя, що виходить з початку координат, величина y змінюється за лінійним законом. Отже, за умови $a_2 + a_3 = 1$ маємо конічну поверхню, ескіз якої представлено на рис. 3.3 а). Перерізи конічної поверхні площинами $x_1 = x_1^0$ та $x_2 = x_2^0$ зображені на рис. 3.3 а) та у відповідних площинах на рис. 3.3 б) та 3.3 в).

Аналітичне зображення вказаних перерізів таке: $y/_{x_1 = x_1^0} = a_1 (x_1^0)^{a_2}\, x_2^{a_3}$ та $y/_{x_2 = x_2^0} = a_1 x_1^{a_2} (x_2^0)^{a_3}$. Оскільки виконуються умови $0 < a_2 < 1$ і $0 < a_3 < 1$, то лінії в перерізах будуть відрізнятися лише кривизною. Такий характер зміни ВФ при фіксованій змінній віддзеркалює ту реальну економічну особливість, що при збільшенні величини будь-якого ресурсу обсяг виробництва збільшується, але це збільшення уповільнюється, тобто кожна додаткова одиниця певного ресурсу викликає все менший приріст обсягу виробництва (закон спадаючої економічної ефективності).

Якщо $a_2 + a_3 < 1$, то поверхня графіка відповідної ВФ уже не буде конічною, лінія перерізу такої поверхні площиною $x_2 = \lambda\, x_1$ не буде прямою, вона деформується.

На практиці однією з широко використовуваних форм аналітичної залежності ВФ є лінійна функція досліджуваних факторів $y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_n x_n$

(ЛВФ). За такої форми зміна величини певного фактора визначає зміну фактора ЛВФ як відповідний доданок. Функції такого класу називаються *адитивними*.

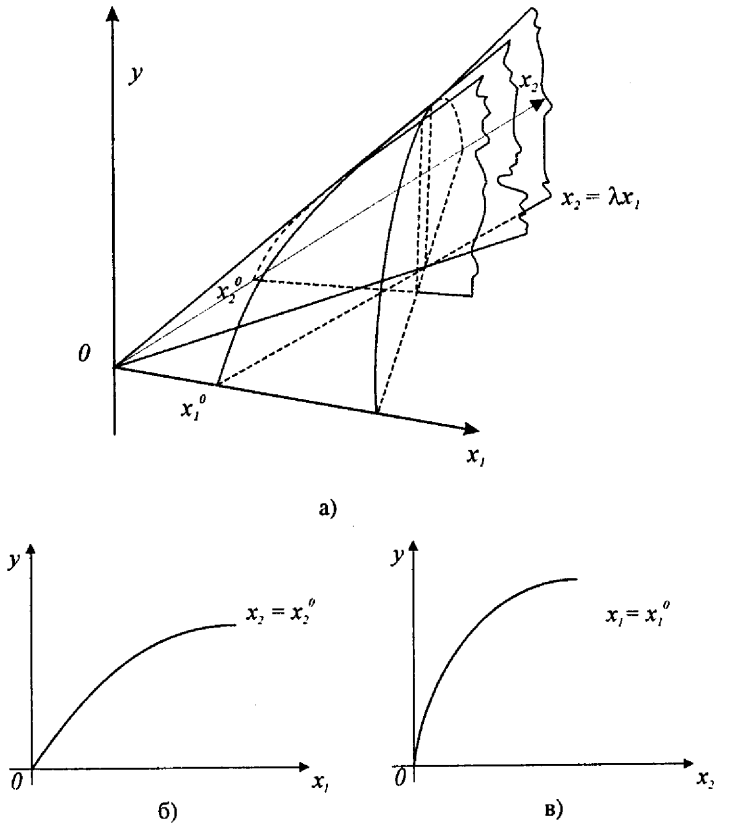


Рис. 3.3.

Наприклад, перехід від двофакторної мультиплікативної ВФ (3) до адитивної виконується логарифмуванням: $\ln Y = \ln a_0 + a_1 \ln K + a_2 \ln L$. Якщо ввести нові змінні (змінити систему відліку) $\ln Y = W$; $\ln K = x_1$; $\ln L = x_2$, то одержимо ВФ в адитивній формі: $W = \ln a_0 + a_1 x_1 + a_2 x_2$. Перехід від двофакторної адитивної функції до мультиплікативної виконується потенціюванням. Якщо маємо $y = a_0 + a_1 x_1 + a_2 x_2$, то після заміни змінних $x_1 = \ln u$, $x_2 = \ln v$ та потенціювання

запишемо $e^y = e^{a_0} u^{a_1} v^{a_2}$. Скориставшися позначеннями $Z = e^y$ та $b = e^{a_0}$, можемо записати в мультиплікативній формі $Z = bu^{a_1} v^{a_2}$.

Враховуючи практичну значимість ВФ Кобба-Дугласа (3), зупинимося на деяких її особливостях. Якщо сума показників степеня ВФ $Y = a_0 K^{a_1} L^{a_2}$ дорівнює одиниці $(a_1 + a_2 = 1)$, то доцільно записати її так:

$$\frac{Y}{L} = a_0 \left(\frac{K}{L}\right)^{a_1}. \qquad (4)$$

Відношення $Z = \frac{Y}{L}$ та $k = \frac{K}{L}$ називаються відповідно *продуктивністю праці* та *капіталозабезпеченістю праці.* За вказаних умов можемо записати однофакторну ВФ $Z = a_0 k^{a_1}$. Якщо виконується умова $0 < a_1 < 1$, то функціональна залежність віддзеркалює той об'єктивно існуючий факт, що продуктивність праці Z збільшується повільніше, ніж її капіталозабезпеченість k. Але цей висновок, як підтверджує практика господарської діяльності, має сенс лише за умови незмінних технологій та використовуваних ресурсів. При використанні досягнень науково-технічного прогресу та відповідних передових технологій економічно-виробнича ситуація змінюється, і ВФ буде зображена іншою аналітичною залежністю.

Оцінку ефективності використання коштів у виробництві виконують з допомогою відношення Y/K, яке називається *продуктивністю капіталу.* Величини K/Y та L/Y – відповідно *капіталомісткість* та *працемісткість* виробництва.

3.4. Загальні властивості виробничих функцій

Аналіз практичного використання ВФ дає можливість охарактеризувати їх певні *властивості поза формою аналітичного зображення.*

Для досліджень ВФ за будь-якої форми її зображення необхідно використовувати лише ту частину області визначення, яка задовольняє вимогам, що величини факторів невід'ємні, тобто $x_i \geq 0$ $(i = 1, 2, ..., n)$. Цільове призначення ВФ обумовлює такі властивості, які для зручності записів будуть названі лише для випадку двофакторних функцій.

1. $f(0,0) = 0$. 1а. $f(0, x_2) = f(x_1, 0) = 0$.

2. Якщо збільшується величина хоча б одного з факторів, то значення ВФ теж збільшуються, що означає умову: коли $x_1^* > x_1^0$ або $x_2^* > x_2^0$, то виконується відношення $f(x_1^*, x_2^*) > f(x_1^0, x_2^0)$. Ця властивість обумовлює, що $\frac{\partial f}{\partial x_1} > 0$ і $\frac{\partial f}{\partial x_2} > 0$ в області визначення ВФ практичної значимості.

3. Завжди $\frac{\partial^2 f}{\partial x_1^2} \leq 0$ і $\frac{\partial^2 f}{\partial x_2^2} \leq 0$.

4. Виробнича функція має бути однорідною функцією певного порядку p своїх аргументів, тобто $f(\lambda x_1, \lambda x_2) = \lambda^p f(x_1, x_2)$.

Властивості 1 та 1а свідчать про те, що без ресурсів не може бути виробництва.

Властивість 2 означає, що при збільшенні величини будь-якого фактора обсяг виробництва теж збільшується. Такий висновок справедливий в певних межах зміни факторів, які визначають величину ВФ. Наприклад, внесення добрив сприяє збільшенню врожайності.

Властивість 3 означає, що збільшення величини певного ресурсу за умови фіксованості величини решти ресурсів хоч і викликає збільшення обсягу виробництва, але характер завжди такий, що збільшення величини ресурсу на кожну додаткову одиницю обумовлює *спад приросту* обсягу виробництва *(закон спадаючої ефективності).* У прямокутній системі координат ВФ буде зображена поверхнею, розташованою в першому квадранті та випуклою догори, але крутість такої поверхні спадає зі збільшенням величин незалежних змінних.

Властивість 4 означає, що ВФ завжди має бути однорідною функцією своїх змінних. Це віддзеркалює той факт, що зміна масштабів (мірила) одиниць вимірів виробничих факторів обумовлює зміну масштабу виміру одиниць обсягів виробництва. В економічних дослідженнях величину λ розуміють як збільшення обсягів незалежних факторів за одних і тих же одиниць їх виміру. За таких умов при $p > 1$ зі збільшенням величин обсягів незалежних факторів виробництва в λ разів обсяг виробництва зросте в λ^p разів, якщо $\lambda > 1$, тобто маємо підвищення ефективності виробництва. Якщо $p < 1$ $(\lambda > 1)$, маємо спад ефективності виробництва за умови збільшення величин виробничих факторів. При $p = 1$ маємо сталу ефективність виробництва.

Для ВФ вигляду (3) властивості 1–4 виконуються за умови $a_1 + a_2 = 1$, $p = 1$, тому маємо *модель сталої ефективності виробництва.* Стосовно лінійної форми ВФ необхідно наголосити, що у формулі $y = a_0 + a_1 x_1 + a_2 x_2$ вплив величин виробничих факторів на обсяг виробництва обумовлений лише останніми двома доданками; величина a_0 визначає лише початок відліку обсягу виробництва. За таких умов властивості 1–4 теж будуть виконуватися, якщо ввести функцію $Z = Y - a_0$, значення якої дорівнює нулю, за умови $x_1 = 0; x_2 = 0$. Названу ситуацію можна тлумачити, наприклад, так: якщо в задачі дослідження ефективності збільшення врожайності в

залежності від обсягів використовуваних добрив ВФ буде представлена лінійною формою, величина a_0 визначатиме урожайність без використання добрив, тобто фактичний ефект визначатиметься лише доданками $a_1x_1 + a_2x_2$.

Множина значень (x_1, x_2), для яких виконується умова $f(x_1, x_2) = C$ (C – стала величина), називається *лінією рівня або лінією байдужості*. Для ВФ характерно, що завжди $C > 0$. За таких умов використовується термін "ізокванта". Різні сукупності значень (x_1, x_2), якщо їх величини задовольняють аналітичній лінії рівня (ізокванті), забезпечують один і той же обсяг виробництва (ефективність використання ресурсів).

Для будь-якої двофакторної ВФ за умови виконання властивостей 1–4 ізокванта (лінія рівня) випукла вниз і має вигляд, як представленого на рис. 3.4 а). Для лінійної ВФ ізокванти зображені на рис. 3.4 б).

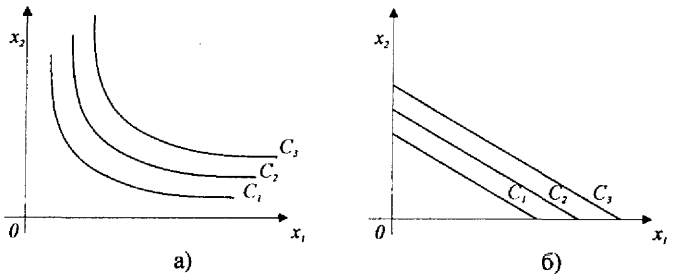


Рис. 3.4.

Для обох частин рис. 3.4 значення ізоквант збільшуються при віддаленні від початку координат.

3.5. Визначення параметрів виробничих функцій

Мета розбудови ВФ – дати математичну модель певного економічного процесу. Ефективне моделювання може бути виконане лише за умови, що йому передує ретельний аналіз досліджуваного процесу, вибір визначальних його факторів та певний аналіз їх змін і впливу на досліджуваний економічний показник. Виділення визначальних факторів (видів ресурсів) та вибір форми аналітичного зображення функції називається *специфікацією ВФ*.

Визначення величин параметрів ВФ за статистичними експериментальними даними з використанням математичного апарату кореляційного та регресійного аналізу називається *параметризацією ВФ*.

Перевірка узгодженості (адекватності) ВФ з відповідними експериментальними даними називається *верифікацією ВФ*.

Специфікація ВФ (вибір факторів та аналітичної залежності між ними) обумовлена як теоретичними міркуваннями, які враховують особливості зв'язків між конкретними ресурсами (задачі мікроекономіки) або економічних закономірностей (задачі макроекономіки), так і практичним досвідом розв'язку подібних задач і характером зміни експериментальних даних. На специфікацію та параметризацію при розбудові ВФ впливають і результати верифікації. Обчислення значень параметрів ВФ виконується в основному з допомогою метода найменших квадратів (МНК).

Прийнявши певну гіпотезу про аналітичну форму взаємозв'язків між змінними ВФ і використання параметрів в ній, необхідно вибрати критерій, який би дозволяв визначити кращий набір числових значень вибраних параметрів форми. Природно, що такий критерій має враховувати співвідношення експериментальних даних та аналітичних розрахунків за одних і тих же значень незалежних змінних (набору визначальних факторів). Різні методи розрахунків величини параметрів ґрунтуються на різних критеріях, які враховують ступінь наближення експериментальних та аналітичних даних, і дають різні значення величини параметрів ВФ за однієї й тієї ж аналітичної форми; ці параметри можна обчислювати за критерієм: абсолютна величина найбільшого відхилення аналітичних значень від відповідних даних спостережень досягає найменшого значення. До того ж виявляється, що різні значення величин параметрів, одержані за різними методами, мають різні статистичні властивості. Найбільш ґрунтовно досліджені статистичні властивості величин параметрів аналітичної форми ВФ, які одержують методом найменших квадратів (МНК). Опишемо основні етапи МНК за умови двофакторної ВФ. Позначимо: $y = f(x_1, x_2, \bar{a})$ (f – форма аналітичної залежності ВФ від факторів і вектора параметрів $\bar{a} = (a_1, a_2, ..., a_m)$; y_i^* – відомі експериментальні дані досліджуваної закономірності за певного набору значень величин визначальних факторів $x_{1,i}^*, x_{2,i}^*$). Відхиленням (похибкою) аналітичних даних від експериментальних назвемо різницю

$$y_i - y_i^* \equiv f(x_{1,i}^*, x_{2,i}^*, a_1, a_2, ..., a_m) - y_i^*. \qquad (5)$$

У МНК найкращим вважається такий набір параметрів, коли сума квадратів відхилень для всіх експериментальних точок, тобто для всіх сукупностей пар $(y_i^*, x_{1,i}^*, x_{2,i}^*)$, має найменше значення. Аналітично ця умова записується так: знайти найменше значення функції Q:

$$Q=\sum_{i=1}^{n}(f(x_{1,i}^{*};x_{2,i}^{*};a_1;a_2;...;a_m)-y_i^{*})^2. \qquad (6)$$

За відомого набору експериментальних даних $(x_{1,i}^{*}, x_{2,i}^{*})$ та y_i^{*} функція Q є функцією змінних $a_1, a_2, ..., a_m$. Якщо ВФ є диференційована функція (умова, яка завжди виконується для $x_1 > 0; x_2 > 0$), то Q теж буде диференційованою. Система розв'язувальних рівнянь для визначення параметрів будується за умови виконання вимог, що частинні похідні функції Q дорівнюють нулю, тобто:

$$\frac{\partial Q}{\partial a_k}=2\sum_{i=1}^{n}(f(x_{1,i}^{*};x_{2,i}^{*};a_1;...;a_m)-y_i^{*})\frac{\partial f(x_{1,i}^{*};x_{2,i}^{*};a_1;...;a_m)}{\partial a_k}=0 \qquad (7)$$

($k=1,2,...,m$; m – кількість параметрів в аналітичній формі зображення ВФ).

Конкретне представлення розв'язувальних рівнянь (7), які в літературі часто називаються системою нормальних рівнянь, залежить від аналітичної форми представлення ВФ. Якщо маємо двофакторну ВФ в лінійній трипараметричній формі $y=a_0+a_1x_1+a_2x_2$, то система розв'язувальних рівнянь має вигляд:

$$\begin{cases} na_0+a_1\sum x_{1,i}^{*}+a_2\sum x_{2,i}^{*}=\sum y_i^{*} \\ a_0\sum x_{1,i}^{*}+a_1\sum (x_{1,i}^{*})^2+a_2\sum x_{1,i}^{*}x_{2,i}^{*}=\sum y_i^{*}x_{1,i}^{*} \\ a_0\sum x_{2,i}^{*}+a_1\sum x_{1,i}^{*}x_{2,i}^{*}+a_2\sum (x_{2,i}^{*})^2=\sum y_i^{*}x_{2,i}^{*}. \end{cases} \qquad (8)$$

У розбудові (8) додавання виконується за даними усіх спостережень, тобто від $i=1$ до $i=n$.

Якщо ВФ нелінійна стосовно параметрів $a_1, a_2, ..., a_m$, то система рівнянь (7) теж буде нелінійною. Розв'язання таких систем виконується методами, побудованими з використанням поняття градієнта (напрямку найбільш швидкого зростання). Використовуються також методи знаходження величин параметрів шляхом безпосереднього пошуку найменшого значення функції Q (6).

На практиці широко використовуються такі форми аналітичної залежності, які після відповідних перетворень можна привести в нових координатах відліку до лінійної. Відповідні аналітичні перетворення для однофакторних функцій наведені в табл. 3.5.

Приклад. ВФ аналітичної форми $y=a_0x_1^{a_1}x_2^{a_2}$ після виконання операції логарифмування та введення нових змінних буде приведена до лінійної залежності:

$$Z=b_0+a_1u_1+a_2u_2 \quad (Z=\ln y;\ b_0=\ln a_0;\ u_1=\ln x_1;\ u_2=\ln x_2).$$

3.5.1. Елементи теорії обробки даних спостереження

Як наголошувалося раніше, дані спостережень економічних і виробничих процесів завжди мають статистичний характер, тобто певним конкретним значенням незалежних змінних відповідає можливий статистичний розподіл залежної змінної. За таких умов дані спостережень доцільно записувати у вигляді так званих кореляційних таблиць. Зразок такої таблиці наведено як результат дослідження величини зміни врожайності y (ц/га) в залежності від кількості внесених добрив x (ц/га) на площі 100 га (табл. 3.6).

Таблиця 3.6.

x \ y	10	12	14	16	18	20	Всього
10	9	4	1	–	–	–	14
30	1	10	9	3	–	–	23
50	–	2	6	14	6	–	28
70	–	–	1	10	18	6	35
Всього	10	16	17	27	24	6	100

За умови експерименту змінній x було надано чотири значення – 10; 30; 50; 70 ц/га і всю досліджувану площу розділено на чотири частини (групи) – відповідно 14; 23; 28; 35 га. На кожний гектар було внесено одну і ту ж кількість добрив в кожній групі, а урожайність виявилася різною. Вибравши прямокутну систему координат (рис. 3.5), дані табл. 3.6 представимо множиною відповідних точок. Таке зображення статистичних даних називається *полем кореляції*.

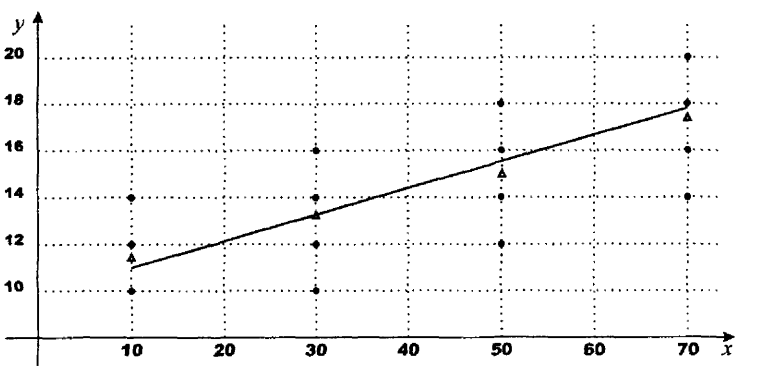


Рис. 3.5.

Природно очікувати, що збільшення кількості добрив до певної межі сприятиме підвищенню врожайності. Ця закономірність чітко спостерігається, якщо обчислити середні величини урожайності при кожній кількості добрив на гектар (для кожної групи за критерієм кількості добрив на гектар):

$\bar{y}(10) = (10 \cdot 9 + 12 \cdot 4 + 14 \cdot 1)/14 \approx 10{,}86$ ц;

$\bar{y}(30) = (10 \cdot 1 + 12 \cdot 10 + 14 \cdot 9 + 16 \cdot 3)/23 \approx 13{,}22$ ц;

$\bar{y}(50) = (12 \cdot 2 + 14 \cdot 6 + 16 \cdot 14 + 18 \cdot 6)/28 \approx 15{,}71$ ц;

$\bar{y}(70) = (14 \cdot 1 + 16 \cdot 10 + 18 \cdot 18 + 20 \cdot 6)/35 \approx 17{,}66$ ц.

Величини середніх урожайностей $\bar{y}_i$ ($i = 1; 2; 3; 4$) на рис. 3. 5 відмічені трикутниками; їх розташування в системі координат та обчислення перших різниць середніх ($\Delta\bar{y}_i = \bar{y}_{i+1} - \bar{y}_i$) свідчать про доцільність вибору лінійної залежності величини врожайності від кількості добрив. Наведемо перші різниці: $\Delta y_1 = 2{,}36$; $\Delta y_2 = 2{,}49$; $\Delta y_3 = 1{,}95$.

У загальному випадку дані спостережень записують у вигляді кореляційної табл. 3.7.

Таблиця 3.7.

x_i \ y_j	y_1	y_2	…	y_j	…	y_m	Всього n_{x_i}
x_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1m}	n_{x_1}
x_2	n_{21}	n_{22}	…	n_{2j}	…	n_{2m}	n_{x_2}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_i	n_{i1}	n_{i2}	…	n_{ij}	…	n_{im}	n_{x_i}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_l	n_{l1}	n_{l2}	…	n_{lj}	…	n_{lm}	n_{x_l}
Всього n_{y_j}	n_{y_1}	n_{y_2}	…	n_{y_j}	…	n_{y_m}	n

У табл. 3.7 n_{ij} ($i = \overline{1, l}$; $j = \overline{1, m}$) – частоти пар значень (x_i; y_j). Сума їх дає об'єм сукупності спостереження n. Змінна x приймала l різних значень: $x_1, x_2, \ldots, x_l$ з відповідними частотами $n_{x_1}, n_{x_2}, \ldots, n_{x_l}$, наведеними в останньому стовпці. Змінна y приймала m різних значень: $y_1, y_2, \ldots, y_m$; їх частоти $n_{y_1}, n_{y_2}, \ldots, n_{y_m}$ наведені в останньому рядку. Аналіз табл. 3.7 свідчить, що можливі три різних представлення сукупності n: сума частот n_{ij}, сума частот n_{x_i} та сума частот n_{y_j}:

$$n = \sum_{i=1}^{l}\sum_{j=1}^{m} n_{ij} = \sum_{i=1}^{l} n_{x_i} = \sum_{j=1}^{m} n_{y_j}. \qquad (9)$$

Будь-яка з частот n_{ij} в частинних випадках може дорівнювати нулю, але завжди можна вважати, що жодна з частот n_{x_i} та n_{y_j} не дорівнює нулю.

Всю сукупність даних спостереження можна розділити на групи двома різними способами: в залежності від значень змінної x і відповідної їй множини значень y або навпаки.

Одержавши дані спостережень у вигляді кореляційної таблиці, доцільно оцінити щільність зв'язку між ними.

Розглянемо найбільш широко використовуваний на практиці та в теоретичних дослідженнях коефіцієнт r лінійної кореляції (від лат. correlation – взаємозв'язок) між даними спостереження (скористаємося однією з модифікацій його аналітичного представлення):

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x s_y}; \qquad (10)$$

$$\overline{xy} = \left(\sum_{i=1}^{l}\sum_{j=1}^{m} x_i y_j n_{ij}\right)/n; \quad \bar{x} = \left(\sum_{i=1}^{l} x_i n_{x_i}\right)/n; \quad \bar{y} = \left(\sum_{j=1}^{m} y_j n_{y_j}\right)/n;$$

$$s_x = \sqrt{\left(\sum_{i=1}^{l} x_i^2 n_{x_i}\right)/n - \overline{x^2}}; \quad s_y = \sqrt{\left(\sum_{j=1}^{m} y_j^2 n_{y_j}\right)/n - \overline{y^2}}.$$

Вкажемо основні властивості коефіцієнта кореляції r.

1. Коефіцієнт кореляції може приймати значення лише на відрізку [–1; 1]. Чим ближче значення $|r|$ до 1, тим щільніше зв'язок між досліджуваними факторами.

2. Якщо всі значення змінних збільшити (зменшити) у кілька разів або на кілька одиниць, то величина коефіцієнта кореляції не зміниться.

3. За умови $r = \pm 1$ наявною є *лінійна функціональна залежність* між досліджуваними змінними.

4. За умови $r = 0$ не можлива лінійна кореляційна залежність, але можлива нелінійна.

Середні значення однієї із змінних в кореляційній табл. 3. 7 за умови, що величина значення іншої змінної фіксована, називаються *груповими середніми*, тобто:

$$\bar{y}_i = \left(\sum_{j=1}^{m} y_j n_{ij}\right)/n_{x_i}; \quad \bar{x}_j = \left(\sum_{i=1}^{l} x_i n_{ij}\right)/n_{y_j}. \qquad (11)$$

Розглянемо методи розбудови аналітичних залежностей між однією змінною та відповідними груповими середніми іншої змінної. Рівняння, яке пов'язує значення групової середньої та значення іншої змінної за відомої кореляційної таблиці, називається *рівнянням регресії*. Почнемо з розбудови лінійних рівнянь регресії.

3.5.2. Лінійна парна регресія

Припустимо, що маємо множину пар (x_i, $\bar{y}_i$) результатів спостереження. Є підстава вважати, що значення групових середніх величин $\bar{y}$ залежить лінійно від x, тобто:

$$\bar{y} = \alpha + \beta x \qquad (12)$$

(α та β – сталі (параметри), величини яких треба підібрати так, щоб за допомогою залежності (12) найкращим чином узгодити усереднені дані спостереження величини y та дані, розраховані за (12) при зміні x). Окремі пари даних спостереження завжди в тій чи іншій мірі відхиляються від лінійної залежності під впливом різних випадкових об'єктивних факторів, вплив яких практично не можна оцінити, а також випадкових суб'єктивних причин, наприклад, точності виконання та реєстрації замірів і т. ін. Бажаючи врахувати можливі випадкові відхилення від представлення основної закономірності, лінійну залежність між двома величинами (парну регресію) будемо розглядати у вигляді:

$$\bar{y} = \alpha + \beta x + \varepsilon(x) \qquad (13)$$

(ε – випадкова величина, яка дає можливість оцінити ймовірні випадкові відхилення даних спостереження від лінійної залежності між наявними $\bar{y}_i$ та x_i, яка за припущенням відповідає природній закономірності).

Стосовно ймовірного відхилення $\varepsilon(x)$ приймається припущення:

1) відхилення є випадковою величиною;
2) математичне сподівання $\varepsilon(x)$ дорівнює нулю;
3) дисперсія відхилень – стала величина;
4) відхилення для різних значень x не залежні між собою.

При розбудові рівняння регресії вважаємо, що результати кожного i-го спостереження, тобто спостереження, коли x дорівнює x_i, пов'язані залежністю:

$$\bar{y}_i = \alpha + \beta x_i + \varepsilon(x_i). \qquad (14)$$

Таким чином, при наявній множині спостережень зміни середніх досліджуваного показника y необхідно знайти числові значення параметрів α та β. Істинних значень параметрів одержати неможливо, оскільки наявні спостереження містять в собі випадкові похибки. Тому обчислені будь-яким методом величини параметрів α та β завжди будуть їх *статистичними оцінками*. Доцільно позначити ці статистичні оцінки a та b. За таких умов рівняння парної регресії $\bar{y} = a + bx$ необхідно розуміти як статистичну оцінку лінійної залежності $\bar{y} = \alpha + \beta x$.

Якщо в прямокутній системі координат (x; y) за вибраним мірилом по вісям координат нанести результати спостережень (x_i; $\bar{y}_i$), то точки з такими координатами не будуть розташовані на прямій лінії через наявність впливу випадкових причин на перебіг лінійного процесу (див. рис. 3.6 а). Тому однозначно вибрати значення параметрів a та b неможливо: можна побудувати нескінченну множину прямих ліній, стосовно кожної з яких точки (x_i; $\bar{y}_i$) будуть певним чином розташовані, але необхідно мати додатковий критерій, який дав би можливість однозначного вибору з множини прямих ліній.

Найбільш широко використовуваним і статистично ефективним є метод найменших квадратів (МНК). Згідно з цим методом a і b обчислюють за умови, що сума квадратів відхилень даних спостереження (x_i; $\bar{y}_i$) від значень, обчислених для відповідних x_i, для досліджуваної аналітичної залежності $y = a + bx$ досягатиме найменшої величини.

Статистична значимість МНК обумовлена тим, що за умови виконання раніше прийнятих припущень стосовно статистичних властивостей $\varepsilon(x)$ оцінки a та b мають суттєві для подальшого практичного використання властивості:

1) величини a та b, обчислені за МНК, є незміщеними оцінками α та β, тобто математичне сподівання значень a та b дорівнює істинним, але невідомим значенням α та β;
2) оцінки спроможні, тобто дисперсія оцінки кожного параметра сходить до нуля за умови зростання кількості спостережень;
3) оцінки є ефективними, тобто мають найменшу дисперсію при порівнянні їх з будь-якими оцінками, обчисленими з використанням інших методів.

Ураховуючи, що a та b є статистичними оцінками α та β, питання їх якості надзвичайно важливе для практичного використання, особливо враховуючи, що за їх допомогою може будуватися прогноз.

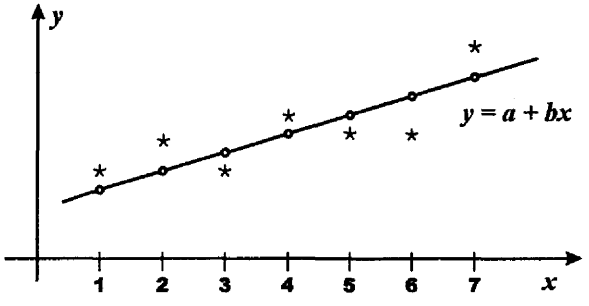
Вимога незміщеності є визначальною, бо зміщені оцінки можуть дати невірні кількісні зв'язки між досліджуваними показниками. Властивість спроможності означає, що при збільшенні кількості спостережень оцінки параметрів стають більш надійними, тобто все щільніше розташовуються навколо істинних, але невідомих значень параметрів α та β. За розбудови прогнозу властивість ефективності надзвичайно суттєва, бо вона визначає рівень можливої його похибки.

3.5.3 Оцінки параметрів лінійної парної регресії

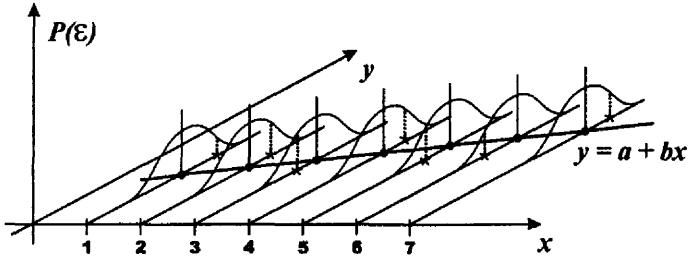
Розглянемо схематичний графік (рис. 3.6 а), на якому представлені результати спостережень (x_i; $\bar{y}_i$) та графік їх аналітичної апроксимації $y = a + bx$. Ймовірності можливих відхилень середніх величин експериментальних даних від лінійної залежності зображено на рис. 3.6 б. В околі розташування точок з координатами (x_i; $\bar{y}_i$) будемо шукати розташування деякої прямої $y = a + bx$ за таким критерієм. Для кожної заданої абсциси x_i розглянемо відхилення e_i (похибку):

$$e_i = \bar{y}_i - y_i = \bar{y}_i - (a + bx_i) \tag{15}$$

($\bar{y}_i$ – середні значення даних відповідного спостереження; y_i – величини, обумовлені вибраною аналітичною залежністю при відповідних значеннях x_i). У подальшому для зручності будемо вважати, що маємо n пар (x_i; $\bar{y}_i$).



а)



б)

Рис. 3.6.

Залежність (15) свідчить, що за відомих x_i та $\bar{y}_i$ величина похибки e_i є функцією від a та b. Згідно з МНК величини параметрів a та b необхідно вибрати так, щоб сума квадратів відхилень e_i (15), тобто величина

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (\bar{y}_i - y_i)^2 = \sum_{i=1}^{n} (\bar{y}_i - (a + bx_i))^2$$

для всіх даних спостереження була найменшою.

Оскільки функція Q за своїм аналітичним представленням є неперервною та диференційованою стосовно a та b, то мінімум може бути лише за таких їх значень, за яких виконуються умови:

$$\begin{aligned} \frac{\partial Q}{\partial a} &= -2\sum_{i=1}^{n} (\bar{y}_i - (a + bx_i)) = 0; \\ \frac{\partial Q}{\partial b} &= -2\sum_{i=1}^{n} (\bar{y}_i - (a + bx_i))x_i = 0. \end{aligned} \tag{16}$$

Виконавши алгебраїчні перетворення системи (16), одержуємо її так звану *нормальну форму* для обчислення значень a та b:

$$\begin{aligned} na + b\sum_{i=1}^{n} x_i &= \sum_{i=1}^{n} \bar{y}_i; \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{n} x_i \bar{y}_i. \end{aligned} \tag{17}$$

Розділивши кожне з одержаних рівнянь на n і замінивши послідовність їх запису, маємо:

$$\begin{aligned} \bar{y} &= a + b\bar{x}; \\ \overline{xy} &= a\bar{x} + b\overline{x^2}. \end{aligned} \tag{18}$$

Рискою над символом позначено відповідні середні арифметичні величини для всіх величин, залежних від наявних даних спостереження:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i;\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;\ \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i \bar{y}_i;\ \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2.$$

Доцільно ще раз наголосити, що $\bar{y}_i$ – групова середня (середня за фіксованого значення x), $\bar{y}$ – середня всього обсягу спостереження.

Перше рівняння системи (18) свідчить: МНК дає такі значення a та b, що лінійна аналітична залежність виконується для середніх значень множин $\{x_i\}$ та $\{y_i\}$. Графічно це означає, що пряма, рівняння якої знаходимо, проходить через точку з координатами $\left(\frac{1}{n}\sum_{i=1}^{n} x_i;\ \frac{1}{n}\sum_{i=1}^{n} y_i\right)$. Таким чином, рівняння регресії (14) можна записати так:

$$y - \bar{y} = b(x - \bar{x}). \tag{19}$$

Коефіцієнт b у рівнянні (19) називається *коефіцієнтом регресії y на x*; в літературі його позначають символом $\rho_{y/x}$:

$$b = \rho_{y/x} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - (\bar{x})^2}. \tag{20}$$

При паралельному переносі системи координат у точку $(\bar{x};\bar{y})$, якщо позначити $\tilde{x} = x - \bar{x}$, $\tilde{y} = y - \bar{y}$, рівняння регресії буде мати вигляд:

$$\tilde{y} = b\tilde{x},$$

а коефіцієнт b за нових координат обчислюється за формулою:

$$b = \frac{\sum_{i=1}^{n} \tilde{x}_i\ \tilde{y}_i}{\sum_{i=1}^{n} (\tilde{x}_i)^2}.$$

Оцінку щільності залежності між змінними кореляційної таблиці за умови лінійної залежності між ними визначають так званим коефіцієнтом кореляції r:

$$r = b\cdot s_x / s_y = (\overline{xy} - \bar{x}\cdot\bar{y}) / (s_x s_y) \tag{21}$$

(s_x та s_y – вибіркові середні квадратичні відхилення).

Величина так визначеного коефіцієнта кореляції (21) не залежить від вибору одиниць вимірів змінних. Якщо $r>0$, то кореляційний зв'язок називається прямим: зі збільшенням незалежної змінної збільшується залежна змінна. Якщо $r<0$, кореляційний зв'язок називається зворотним: збільшення незалежної змінної обумовлює зменшення залежної змінної.

Перелік основних властивостей коефіцієнта кореляції (21) наведено в коментарях до формули (10).

Рівняння лінійної регресії досить широко використовується на практиці завдяки зручності розрахунків, але в багатьох випадках воно не може забезпечити необхідну точність досліджень, тому використовують нелінійні форми зв'язків між незалежною та залежною змінними. Доцільно наголосити, що формули нелінійних залежностей часто вибирають так, щоб після певних алгебраїчних перетворень мати лінійну залежність.

3.5.4. Нелінійна парна регресія

Практика математичного моделювання соціально-економічних явищ і процесів свідчить, що не завжди можна користуватися лінійними моделями, оскільки можуть виникати необґрунтовано значні похибки моделювання. У таких випадках використовують рівняння регресії, не лінійне по незалежній змінній. У табл. 3.5 наведено набір певних апробованих практикою аналітичних залежностей нелінійних рівнянь регресії та їх графічне зображення.

Даний перелік був розбудований за вимог щонайменшої кількості параметрів у рівнянні регресії та можливості відповідними алгебраїчними перетвореннями, наведеними для кожної формули, одержати лінійну залежність за нових змінних τ та z. Окрему групу утворюють формули, в яких наявними є нелінійна залежність від аргументу (визначальної змінної) та лінійна залежність стосовно параметрів (наприклад, многочлен).

Для вибору певного аналітичного представлення рівняння регресії необхідно виконати додатковий аналіз усереднених групових наявної кореляційної табл. 3.7 для пояснювального фактора, а також врахувати досвід подібних досліджень, відомий з літературних джерел або з практики. Такий аналіз допомагає зосередити увагу при подальших розрахунках на тих видах аналітичного представлення, які краще узгоджені з характером зміни досліджуваних факторів. На практиці вибір форми рівняння регресії (згідно з графіком) ґрунтується на порівнянні характеру зміни приростів незалежної та залежної змінних даних спостереження і відповідних значень, одержуваних при використанні формули, яку бажають вибрати.

Визначальні співвідношення між приростами для кожної аналітичної залежності наведені в четвертому стовпці табл. 3.5. Розв'язувальних рівнянь не наводимо: як наголошувалося раніше, формули в табл. 3.5 підібрані так, що після відповідних перетворень в нових координатах маємо лінійну залежність.

Вибір форми нелінійної аналітичної залежності виконується на певній множині альтернатив, аналіз співвідношень між приростами змінних дозволяє зменшити

множину альтернатив і вибрати таку модель залежності, яка краще відтворює природу досліджуваного явища або процесу.

Приклад. Залежність *середньої* урожайності y (ц/га) від глибини зрошення наведено в табл. 3.8. Знайти аналітичну форму залежності урожайності від глибини зрошення (см).

Таблиця 3.8.

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Глибина зрошення x (см)	25	27	30	35	36	38	39	41	42	45	46	47	50	52	53
Урожайність y (ц/га)	23	24	27	27	32	31	33	35	34	32	29	28	25	24	25
Δx	1	3	5	1	2	1	2	1	3	1	1	3	2	1	
Δy	1	3	0	5	–1	2	2	–1	–2	–3	–1	–3	–1	–1	
$\Delta^2 y$	2	–3	5	–6	3	0	–3	–1	–1	2	–2	2	2		
$\Delta^2 y/\Delta x^2$	2	–1/3	0,2	–6	0,75	0	–0,75	–1	–1/9	2	–2	2/9	0,5		

Побудувавши в прямокутній системі координат (рис. 3.7) залежність між даними табл. 3.8 та обчисливши величини відносних приростів другого порядку, значення яких наведено в останньому рядку табл. 3.8, доходимо висновку про доцільність вибору рівняння регресії у вигляді:

$$y = a_0 + a_1 x + a_2 x^2.$$

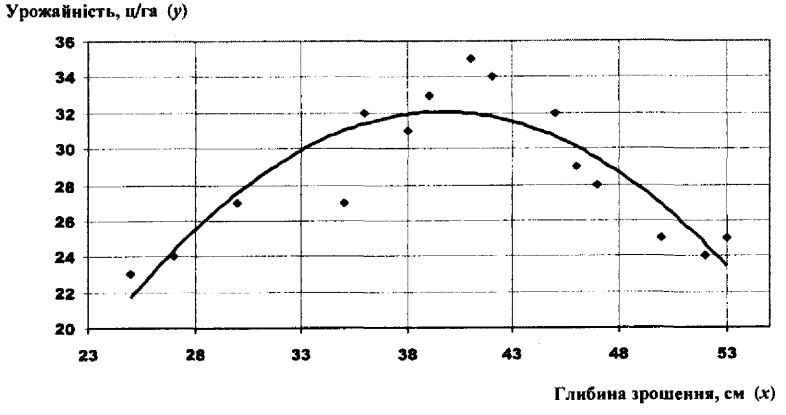


Рис. 3.7.

Вибір такої аналітичної форми рівняння регресії ґрунтується на графічному зображенні даних спостереження та на факті, що відносні прирости другого порядку коливаються біля нуля. Параметри рівняння регресії $y = a_0 + a_1 x + a_2 x^2$ знаходимо за МНК:

$$Q = \sum_{i=1}^{15} \left(\bar{y}_i - (a_0 + a_1 x + a_2 x^2)\right)^2 \to \min. \qquad (22)$$

Прирівнявши нулю частинні похідні функції Q (22), одержуємо систему розв'язувальних рівнянь, записану для загального випадку з використанням рівняння регресії у формі многочлена другого степеня:

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i x_i \\ a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} y_i x_i^2. \end{cases} \qquad (23)$$

Для розв'язуваного прикладу маємо $n = 15$. Схему розрахунків коефіцієнтів СЛАР (23) наведено в табл. 3.9.

Таблиця 3.9.

i	x_i	y_i	x_i^2	x_i^3	x_i^4	$y_i x_i$	$y_i x_i^2$	y_i^2	$\bar{y}_{x_i}$	$(\bar{y}_{x_i} - y_i)^2$
1	25	23	625	15625	390625	575	14375	529	21,7	1,69
2	27	24	729	19683	531441	648	17496	576	24,3	0,11
…	…	…	…	…	…	…	…	…	…	…
14	52	24	2704	140608	7311616	1248	64896	576	24,7	0,46
15	53	25	2809	148877	7890481	1325	70225	625	23,4	2,44
Σ	606	429	25548	1115808	50158200	17371	730123	12493	–	45,94

Після виконання необхідних обчислень система рівнянь (23) має вигляд:

$$\begin{cases} 15a_0 + 606a_1 + 25548a_2 = 429 \\ 606a_0 + 25548a_1 + 1115808a_2 = 17371 \\ 25548a_0 + 1115808a_1 + 50158200a_2 = 730123. \end{cases}$$

Розв'язавши наведену систему, маємо рівняння регресії:

$$y = -43{,}93 + 3{,}834x - 0{,}0483x^2.$$

Найбільшого значення 32,06 y досягне при $x = 39{,}64$. Графік рівняння регресії наведено на рис. 3. 7. Найбільше відхилення даних спостереження від аналітичних буде при $x = 35$ і дорівнюватиме 4,02 (12,8% від аналітичного значення). Решта відхилень суттєво менші. Практичний висновок: найкраща глибина зрошення для досліджуваної культури – біля 40 см.

3.5.5. Перевірка значимості рівняння регресії

Одержавши рівняння регресії, необхідно оцінити, наскільки якісно воно характеризує взаємозв'язок між пояснювальними та пояснюючими факторами, тобто зміну залежної величини за умови зміни незалежної. Аналізуємо кореляційну табл. 3. 7.

Загальною дисперсією S_y^2 величини пояснювального фактору (змінної y) називають середню суми квадратів різниць між даними спостереження та їх середньою величиною:

$$S_y^2 = \frac{1}{n}\sum_{j=1}^{m}(y_j - \bar{y})^2 n_{y_j}. \qquad (24)$$

Груповою дисперсією s_{iy}^2 називається дисперсія розподілу даних спостереження, що складають групу, яка відповідає значенню x_i:

$$s_{iy}^2 = \frac{1}{n_{x_i}}\sum_{j=1}^{m}(y_j - \bar{y}_i)^2 n_{ij}. \qquad (25)$$

Міжгруповою дисперсією δ_{iy}^2 називається середня арифметична квадратів різниць групових середніх $\bar{y}_i$ груп, на які розмежована сукупність даних спостереження, за умови, що вони не мають спільних елементів, та загальної середньої $\bar{y}$:

$$\delta_{iy}^2 = \frac{1}{n}\sum_{i=1}^{l}(\bar{y}_i - \bar{y})^2 n_{x_i}. \qquad (26)$$

Середня арифметична групових дисперсій:

$$\overline{s_{iy}^2} = \frac{1}{n}\sum_{i=1}^{l} s^2{}_{iy} n_{x_i}.$$

Якщо множина даних спостереження розмежована на групи, які не мають спільних елементів, то виконується рівність: загальна дисперсія S_y^2 дорівнює сумі середніх арифметичних групових дисперсій $\overline{s_{iy}^2}$ та міжгрупових дисперсій δ_{iy}^2:

$$S_y^2 = \overline{s_{iy}^2} + \delta_{iy}^2. \qquad (27)$$

Виконаємо аналіз вищеназваних різнопланових дисперсій. При наявності аналітичної залежності середніх величин $\bar{y}_i$ від x, узгодженої із закономірностями перебігу процесу, групова дисперсія (25) визначає ту частину відхилення даних спостереження y при фіксованому значенні x, яка обумовлена впливом випадкових факторів. За допомогою міжгрупової дисперсії δ_{iy}^2 (26) оцінюють ту частину зміни значень y (*пояснювального* фактора), яка визначається зміною x (*пояснюючого* фактора).

Величина

$$\eta = \sqrt{\frac{\delta_{iy}^2}{s_{iy}^2}} \qquad (28)$$

характеризує вплив зміни x на зміну y у порівнянні з впливом випадкових факторів і називається *емпіричним* (одержаним за спостереженнями) *кореляційним відношенням* y *на* x. Величина η^2, названа *емпіричним коефіцієнтом детермінації*, визначає, яка частина зміни y обумовлена зміною x при наявному рівнянні регресії.

Назвемо основні властивості кореляційного відношення (28).

1. Величина кореляційного відношення η належить проміжку [0; 1].
2. Якщо η=0, то кореляційний зв'язок між величинами відсутній.
3. Якщо η=1, то наявною є функціональна залежність між досліджуваними величинами.

За допомогою емпіричного кореляційного відношення η (28) оцінюється *міра розсіювання (відхилення) кореляційного поля* (табл. 3.7) відносно емпіричної лінії регресії, представленої ламаною, яка з'єднує у вибраній системі координат точки $(x_i, \bar{y}_i)$.

Розсіювання кореляційного поля відносно лінії регресії $y = f(x, \bar{a})$ оцінюється за допомогою *теоретичного кореляційного відношення* R:

$$R = \sqrt{\frac{\delta_{iy}^2}{S_y^2}} \qquad (29)$$

(дисперсія δ_{iy}^2 обчислюється за формулою (26), групові середні $\bar{y}_i$ замінено на $y = f(x, \bar{a})$, значення яких обчислюються за рівнянням регресії).

Показники η та R можна обчислити *за будь-якої аналітичної форми рівняння регресії.* Названі показники пов'язані з коефіцієнтом кореляції r (10):

$$0 \le |r| \le R \le \eta \le 1. \qquad (30)$$

При наявності лінійної регресії $R = |r|$. Розбіжність між величинами η^2 та R^2 (або r^2) використовується для перевірки наявності лінійної кореляційної залежності.

Прагнення використати характер зміни приростів даних спостереження для обґрунтування вибору формули вимагає домовленостей стосовно оцінок величин коливань приростів зі зміною пояснювального фактора. Під "близькими" розуміють такі величини приростів, коли зміни відносних похибок у послідовності контрольованих величин не перевищують певного значення (0,05; 0,1; 0,3). За умови, що при вибраному рівні відносної похибки контрольованих відношень критерію вибору формули відповідає кілька аналітичних залежностей, доцільно зменшити рівень відповідної похибки. Якщо знову залишиться більше одного претендента, то для завершення вибору необхідно або ввести допоміжний критерій відбору (наприклад, залишити ту форму, для якої найбільше абсолютне відхилення даних спостереження від аналітичних значень досягає найменшого значення), або виконати в повному обсязі необхідні обчислення та порівняти результати методом МНК. За форму аналітичного представлення вибирають ту формулу, для якої закономірності порівнювальних характеристик найбільш близькі. Властивості таких характеристик наведені в четвертому стовпці табл. 3.5.

Після вибору форми кривої, тобто аналітичної залежності, необхідно обчислити відповідні параметри.

Перевірка значимості рівняння регресії виконується за результатами теоретичних висновків дисперсійного аналізу після розрахунків певних характеристик.

Використовувану інформацію наведено в табл. 3.10.

У теорії дисперсійного аналізу доведено, що між названими сумами квадратів виконується залежність:

$$Q = Q_R + Q_e.$$

Таблиця 3.10.

Компоненти дисперсії	Сума квадратів	Число степенів волі	Середні квадрати
Регресія	$Q_R = \sum_{i=1}^{n} (y(x_i, \bar{a}) - \bar{y})^2$	$m-1$	$s_R^2 = \frac{Q_R}{m-1}$
Залишкова	$Q_e = \sum_{i=1}^{n} (\bar{y}_i - y(x_i, \bar{a}))^2$	$n-m$	$s_e^2 = \frac{Q_e}{(n-m)}$
Загальна	$Q = \sum_{i=1}^{n} (\bar{y}_i - y)^2$	$n-1$	–
m – число параметрів рівняння регресії n – число різних значень пояснювального фактора даних спостережень			

Рівняння регресії є значущим на рівні α, якщо величина F (критерію Фішера-Снедекора) задовольняє умові:

$$F = \frac{Q_R(n-m)}{Q_e(m-1)} > F_{\alpha, k_1, k_2}$$

(F_{α, k_1, k_2} – таблична величина названого критерію, визначена рівнем значимості α за $k_1 = m-1$; $k_2 = n-m$ степеней волі;

m – число параметрів рівняння регресії;

n – кількість груп, усереднених за пояснювальним фактором даних спостереження).

У випадку лінійної парної регресії m=2, а рівняння регресії буде значущим на рівні α, якщо

$$F = \frac{Q_R(n-2)}{Q_e} > F_{\alpha, 1, n-2}.$$

Користуючись даними останнього рядка табл. 3.9, оцінимо значення одержаного рівняння регресії $y = -43{,}93 + 3{,}8342x - 0{,}048361x^2$:

$$Q_e = \sum (\bar{y}_i(x_i) - y)^2 = 45{,}94;$$

$$Q = \sum y_i^2 - \frac{1}{n}(\sum y_i)^2 = 12493 - \frac{429^2}{15} = 223{,}6.$$

Отже, $Q_R = Q - Q_e = 223{,}6 - 45{,}94 = 177{,}66$; $F = \dfrac{177{,}66(15-3)}{45{,}94(3-1)} = 23{,}2$.

При $\alpha = 0{,}05$ та наявних степенів волі $k_1 = 2$, $k_2 = 12$ табличне значення $F_{0{,}05;\,2;\,12} = 3{,}88$. Оскільки як $F > F_{0{,}05;\,2;\,12}$, то рівняння регресії є значущим.

3.5.6. Поняття про багатовимірні кореляцію та регресійний аналіз

Економічно-господарські процеси визначаються багатьма факторами, які одночасно і сукупно діють, тому для адекватного їх моделювання необхідно користуватися функціями кількох змінних. Стисло окреслимо суть відповідного математичного апарату.

Припустимо, що маємо сукупність кількох змінних $(x_1, x_2, \ldots, x_l)$, кожна з яких може приймати випадкові значення. Схема реєстрації даних спостереження для випадку трьох змінних була наведена на початку цієї глави; при більшому числі змінних схема залишається в такій же формі: для кожного набору з трьох змінних записується таблиця, а решта змінних для цієї таблиці – сталі величини. Якщо маємо три величини, із яких одна (z) статистично залежить від двох інших величин x та y, то результати спостережень можна записати за схемою, наведеною в табл. 3.11. До формування цієї таблиці дані експерименту були упорядковані по групах за кількістю різних трійок чисел (x; y; z), всього l груп.

Таблиця 3.11.

x	y	z	Частоти
x_1	y_1	z_1	n_1
x_2	y_2	z_2	n_2
⋮	⋮	⋮	⋮
x_l	y_l	z_l	n_l
Всього			n

У загальному випадку при дослідженні певного економічного показника z залежно від декількох пояснюючих факторів, який моделюється як пошук функціональної залежності від декількох змінних, тобто $z = f(\bar{x})$ ($\bar{x}$ тлумачиться як m-мірний вектор або як точка в m-мірному просторі), задача оцінки статистичних взаємозв'язків між змінними формулюється аналогічно випадку дослідження парної регресії. Вибираємо певну функціональну залежність $z = f(\bar{\alpha}\,;\,\bar{x}) + \varepsilon$ ($\bar{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_l)$ – вектор параметрів, $\bar{x} = (x_1, x_2, \ldots, x_m)$ – вектор пояснюючих факторів, ε – випадкова похибка). Припустивши, що ця функція визначає залежність досліджуваного показника z від пояснюючих факторів $(x_1, x_2, \ldots, x_m)$ для даного економічного процесу або явища, стосовно похибки ε, як при дослідженні парної регресії, приймаємо, що при кожному замірі величина ε є випадковою в m-мірному просторі з нульовим математичним сподіванням і сталою дисперсією. Похибки кожного із замірів статистично незалежні. Для обґрунтування статистичної значимості оцінок компонент $\bar{\alpha}$ приймають, що похибки ε мають нормальний закон розподілу. За прийнятих припущень формулюється задача: за даними спостережень обсягом n необхідно оцінити значення параметрів $(\alpha_1, \alpha_2, \ldots, \alpha_l)$, тобто виконати *специфікацію* вибраної функціональної залежності $z = f(\bar{\alpha}\,;\,\bar{x}) + \varepsilon$.

Розглянемо лінійну залежність z від $\bar{x}$, тобто лінійне рівняння регресії в m-мірному просторі:

$$z = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m \qquad (31)$$

($\bar{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_l)$ – вектор невідомих параметрів розмірності $l = m + 1$).

Припустимо, що маємо n спостережень вектора $\bar{x}$ і залежної змінної z з урахуванням можливих повторів у числових значеннях змінних. Для того, щоб мати можливість розв'язати задачу обчислення "найкращих" компонент вектора $\bar{\alpha}$ (параметрів лінійної регресії), необхідно мати число спостережень $n \geq m + 1$. Якщо така умова не виконується, то можна знайти незліченну множину значень $\bar{\alpha}$, за яких умова буде виконуватись абсолютно точно. Наприклад, якщо $n = m + 1$, то величини $(\alpha_0, \alpha_1, \ldots, \alpha_m)$ визначаються однозначно і обчислюються як розв'язок відповідної системи лінійних алгебраїчних рівнянь. Якщо число спостережень n більше $m + 1$, то неможливо так вибрати величини коефіцієнтів у формулі (31), щоб були узгоджені всі результати спостережень. З'являється задача пошуку таких значень $(\alpha_0, \alpha_1, \ldots, \alpha_m)$, щоб вибрана формула в певному розумінні була найкращою для узгодження з даними спостережень, тобто прагнемо оцінити компоненти вектора $\bar{\alpha}$ за допомогою компонент вектора $\bar{a}$, значення яких обчислюємо за даними спостережень. Невід'ємна різниця $n - (m + 1)$ за такої умови називається *числом степенів волі*. Якщо число степенів волі невелике в порівнянні з одиницею, то статистична надійність розбудованої формули (31) буде невисокою, тобто додатковий експеримент для досліджуваного економічного процесу може дати результати надто далекі від тих, які визначаються за формулою для відповідних значень незалежних змінних. Наведемо геометричне тлумачення вказаних положень. Якщо маємо лише два пояснюючих фактори, то формула (31) матиме вигляд: $z = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$, що дає підстави тлумачити її як площину в прямокутній системі координат (x, y, z). За умови наявності лише трьох сукупностей спостережень (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3) величини α_0, α_1, α_2 будуть обчислені як розв'язок відповідної системи лінійних алгебраїчних рівнянь ($n = m + 1 = 3$) і числові величини даних спостережень будуть точно узгоджені між собою формулою (31). Але будь-яке наступне спостереження майже достовірно матиме результати, не узгоджені з одержаною залежністю. Практика використання рівнянь регресії свідчить, що для забезпечення достатньо високого рівня статистичної надійності необхідно, щоб *число спостережень не менше ніж у три рази перевищувало число шуканих параметрів*, тобто величин $\{\bar{a}_i\}$.

Задача розбудови рівняння багатовимірної лінійної регресії визначається пошуком $(m+1)$-мірного вектора $\bar{a}$, компоненти якого є статистичними оцінками відповідних компонент вектора $\bar{\alpha}$. Такі оцінки виконуються методом найменших квадратів з використанням даних спостережень. Обчислені за формулою (31) значення величини z повинні давати такі величини похибок стосовно даних спостережень, щоб сума квадратів цих похибок для всіх наявних спостережень була найменшою.

За умови виконання припущень стосовно статистичних властивостей відхилень ε від лінійної залежності оцінки параметрів формули (31) за МНК є незміщеними, спроможними та ефективними. Як і у випадку парної регресії, відхилення залежної змінної z в j-му спостереженні від теоретичної лінійної залежності (31) запишемо так: $e_j = z_j - \alpha_0 - \alpha_1 x_1 - \alpha_2 x_2 - \ldots - \alpha_m x_m$. Позначивши через Q суму квадратів названих відхилень по всіх даних спостережень, маємо згідно з МНК:

$$Q = \sum_{j=1}^{n} e_j^2 = \sum_{j=1}^{n}\left(z_j - \left(\alpha_0 + \sum_{i=1}^{m}\alpha_i x_{ij}\right)\right)^2 \to \min. \qquad (32)$$

Функція Q є квадратичною відносно $\{\alpha_i\}$, тому найменшого значення може досягти лише за таких величин α_i, за яких всі її частинні похідні по α_i дорівнюють нулю. Запишемо частинні похідні та прирівняємо їх нулю:

$$\begin{cases} \dfrac{\partial Q}{\partial \alpha_0} = \sum\limits_{j=1}^{n} 2\left(z_j - \left(\alpha_0 + \sum\limits_{i=1}^{m}\alpha_i x_{ij}\right)\right) = 0 \quad (i = \overline{1,m}); \\ \dfrac{\partial Q}{\partial \alpha_i} = \sum\limits_{j=1}^{n} 2\left(z_j - \left(\alpha_0 + \sum\limits_{i=1}^{m}\alpha_i x_{ij}\right)\right) x_{ij} = 0 \quad (i = \overline{1,m}). \end{cases} \qquad (33)$$

Система (33) є системою лінійних алгебраїчних рівнянь (СЛАР) стосовно $m+1$ невідомих $\alpha_0, \alpha_1, \ldots, \alpha_m$. Дані реальних статистичних спостережень свідчать, що СЛАР (33) має єдиний розв'язок. Наведена СЛАР називається *нормальною*; для її конкретного представлення та розв'язку використовується матрично-векторний запис.

При дослідженні рівняння регресії у випадку наявності лише двох пояснюючих факторів (для скорочення запису позначимо їх x та y і замінимо позначення коефіцієнтів) маємо:

$$z = ax + by + c. \qquad (34)$$

При наявності можливих частот n_j повторення даних спостереження одержимо:

$$Q = \sum\left(z_j - (ax_j + by_j + c)\right)^2 n_j. \qquad (35)$$

При записі функції Q операція додавання виконується за всіма різними результатами спостережень, кількість яких позначимо m; можлива тотожність наслідків урахована за допомогою множника частоти n_j.

Записавши систему (33) для функції (35) та упорядкувавши записи коефіцієнтів при a, b та c, одержимо СЛАР:

$$\begin{cases} c\dfrac{\sum\limits_{i=1}^{m} n_i}{n} + a\dfrac{\sum\limits_{i=1}^{m} x_i n_i}{n} + b\dfrac{\sum\limits_{i=1}^{m} y_i n_i}{n} = \dfrac{\sum\limits_{i=1}^{m} z_i n_i}{n}; \\ c\dfrac{\sum\limits_{i=1}^{m} y_i n_i}{n} + a\dfrac{\sum\limits_{i=1}^{m} x_i y_i n_i}{n} + b\dfrac{\sum\limits_{i=1}^{m} y_i^2 n_i}{n} = \dfrac{\sum\limits_{i=1}^{m} y_i z_i n_i}{n}; \\ c\dfrac{\sum\limits_{i=1}^{m} x_i n_i}{n} + a\dfrac{\sum\limits_{i=1}^{m} x_i^2 n_i}{n} + b\dfrac{\sum\limits_{i=1}^{m} x_i y_i n_i}{n} = \dfrac{\sum\limits_{i=1}^{m} x_i z_i n_i}{n}. \end{cases} \qquad (36)$$

Урахувавши, що коефіцієнти СЛАР (36) є середніми арифметичними певних величин, обумовлених даними спостережень (позначимо їх рискою зверху), одержимо систему (36) у такому записі:

$$\begin{cases} c + a\bar{x} + b\bar{y} = \bar{z}; \\ c\bar{y} + a\overline{xy} + b\overline{y^2} = \overline{yz}; \\ c\bar{x} + a\overline{x^2} + b\overline{xy} = \overline{xz}. \end{cases} \qquad (37)$$

Обчисливши з першого рівняння $c = \bar{z} - a\bar{x} - b\bar{y}$ та підставивши його значення у (34), одержимо рівняння $z = \bar{z} + a(x - \bar{x}) + b(\bar{y} - y)$.

Для знаходження величин a та b вираз для $c = \bar{z} - a\bar{x} - b\bar{y}$ підставимо у друге та третє рівняння системи (37). Після об'єднання коефіцієнтів при a та b і певних алгебраїчних перетворень маємо:

$$\begin{cases} a\left(\overline{xy} - \bar{x}\cdot\bar{y}\right) + b\left(\overline{y^2} - \bar{y}^2\right) = \overline{yz} - \bar{y}\cdot\bar{z}; \\ a\left(\overline{x^2} - \bar{x}^2\right) + b\left(\overline{xy} - \bar{x}\cdot\bar{y}\right) = \overline{xz} - \bar{x}\cdot\bar{z}. \end{cases}$$

Скориставшись тим, що дисперсія $\sigma_\xi^2 = \overline{\xi^2} - (\overline{\xi})^2$ (ξ – будь-яка випадкова величина), одержуємо:

$$\begin{cases} a(\overline{xy} - \overline{x} \cdot \overline{y}) + bs_y^2 = \overline{yz} - \overline{y} \cdot \overline{z}; \\ as_x^2 + b(\overline{xy} - \overline{x} \cdot \overline{y}) = \overline{xz} - \overline{x} \cdot \overline{z} \end{cases} \tag{38}$$

(s – вибіркова дисперсія).

Для розв'язання системи (38) введемо так звані *парні коефіцієнти кореляції*:

$$r_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{s_x \ s_y}; \ r_{xz} = \frac{\overline{xz} - \overline{x} \cdot \overline{z}}{s_x \ s_z}; \ r_{yz} = \frac{\overline{yz} - \overline{y} \cdot \overline{z}}{s_y \ s_z}, \tag{39}$$

що дає можливість записати відношення:

$$\overline{xy} - \overline{x} \cdot \overline{y} = r_{xy} s_x s_y; \ \overline{xz} - \overline{x} \cdot \overline{z} = r_{xz} s_x s_z; \ \overline{yz} - \overline{y} \cdot \overline{z} = r_{yz} s_y s_z.$$

Скориставшися наведеними відношеннями та системою (38) маємо:

$$\begin{cases} ar_{xy} s_x + bs_y = r_{yz} s_z; \\ as_x + br_{xy} s_y = r_{xz} s_z. \end{cases}$$

Розв'язавши наведену систему рівнянь, знаходимо:

$$a = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{1 - r_{xy}^2} \cdot \frac{s_z}{s_x}; \ b = \frac{r_{yz} - r_{xz} \cdot r_{xy}}{1 - r_{xy}^2} \cdot \frac{s_z}{s_y}. \tag{40}$$

Отже, одержали статистичні оцінки для коефіцієнтів лінійної регресії (31) за наявності лише трьох факторів: одного пояснювального та двох пояснюючих.

Задача кореляційного аналізу – визначення щільності лінійних кореляційних зв'язків між досліджуваними факторами (змінними x, y, z) розв'язується за допомогою сукупного коефіцієнта кореляції R, який через парні коефіцієнти кореляції (39) обчислюється за формулою:

$$R = \sqrt{\frac{r_{xz}^2 + r_{yz}^2 - 2r_{xz} \cdot r_{yz} \cdot r_{xy}}{1 - r_{xy}^2}}. \tag{41}$$

Властивості R такі:

– приймає значення лише з проміжку [0; 1].

– якщо $R = 0$, то неможлива лінійна регресія (31), але можлива нелінійна регресія або навіть функціональна залежність;

– якщо $R = 1$, то маємо *функціональну*, а не кореляційну залежність;

– чим ближче значення R до 1, тим щільніше пов'язані значення z з величинами x та y.

Отже, *сукупний коефіцієнт кореляції R (41) є мірою лінійного кореляційного зв'язку z з x та z з y*. Поряд із дослідженнями зв'язку пояснювального фактора z з пояснюючими факторами x та y часто буває важливо оцінити нарізно щільність зв'язків між z та x, а також між z та y. Такі зв'язки оцінюються за допомогою так званих частинних коефіцієнтів кореляції.

Частинний коефіцієнт кореляції $r_{xz(y)}$ між z та x визначає величину лінійного кореляційного зв'язку між z та x за умови, що y є сталою величиною:

$$r_{xz(y)} = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}}. \tag{42}$$

Величина лінійного кореляційного зв'язку між z та y за умови сталого значення x визначається частинним коефіцієнтом кореляції $r_{yz(x)}$:

$$r_{yz(x)} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}. \tag{43}$$

Частинні коефіцієнти кореляції $r_{xz(y)}$ та $r_{yz(x)}$ мають ті ж властивості, що і парний коефіцієнт кореляції (10). Якщо, наприклад, $r_{yz(x)} = 0$, то це робить неможливим лінійну кореляційну залежність між x та z, але за такої умови можлива нелінійна кореляційна залежність або функціональна.

На завершення необхідно додати, що існує широкий спектр математичних засобів моделювання економіко-господарських процесів з використанням пакетів відповідного програмного забезпечення для сучасних комп'ютерів.

3.6. Граничні та середні значення виробничих функцій

Оскільки в мікро- та макроекономіці виробничі функції мають різний зміст, то при висвітленні матеріалу цього параграфа постійно будемо на цьому наголошува-

ти. Для спрощення записів розглянемо двофакторні ВФ.

Якщо $y = f(x_1, x_2)$ є певною ВФ, побудованою для дослідження заданого економічного процесу, то дріб

$$\frac{f(x_1, x_2)}{x_i} \qquad (i = 1,\ 2) \tag{44}$$

називається *середньою продуктивністю i-го ресурсу* (фактора виробництва), або *середнім випуском за i-им ресурсом* (фактором виробництва). Для двофакторної ВФКД $Y = a_0 K^{a_1} L^{a_2}$ стосовно середніх продуктивностей Y/K та Y/L основного капіталу та праці були запроваджені терміни "капіталовіддача" та "продуктивність праці". Ці терміни застосовуються до будь-яких двофакторних ВФ, для яких досліджується вплив обсягів капіталу та праці на обсяги виробництва.

Величини перших частинних похідних $\frac{\partial f}{\partial x_1}$ та $\frac{\partial f}{\partial x_2}$ називаються *граничними продуктивностями* відповідного ресурсу (фактора виробництва). Щоб з'ясувати зміст цього терміна, необхідно згадати зміст поняття першої частинної похідної – швидкість зміни функції за умови, що швидкість зміни відповідного аргументу дорівнює одиниці. За визначенням першої частинної похідної маємо:

$$\frac{\partial f}{\partial x_i} = \lim_{\Delta x_i \to 0} \frac{f(x_i + \Delta x_i, x_k) - f(x_i, x_k)}{\Delta x_i} \Rightarrow \frac{\partial f}{\partial x_i} \approx \frac{\Delta_i f}{\Delta x_i}. \tag{45}$$

Тобто відношення частинного приросту ВФ до приросту відповідного аргумента показує, на скільки збільшиться обсяг виробництва за умови, що обсяг витрат i-го ресурсу збільшиться на одиницю, а обсяг витрат іншого ресурсу залишиться незмінним.

Відношення граничної продуктивності i-го ресурсу (фактора виробництва) до його середньої продуктивності називається *частинною еластичністю* i-го ресурсу (фактора виробництва): $E_i = \frac{\partial f}{\partial x_i} \Big/ \frac{f(x_1, x_2)}{x_i}$.

За малих приростів Δx_i маємо наближену рівність

$$E_i = \frac{\partial f}{\partial x_i} \Big/ \frac{f(x_1, x_2)}{x_i} \approx \frac{\Delta_i f}{f(x_1, x_2)} \Big/ \frac{\Delta x_i}{x_i}. \tag{46}$$

Остання залежність є відношенням двох відносних величин $\Delta_i f / f$ та $\Delta x_i / x_i$, тобто E_i показує зміну відносного обсягу випуску продукції при зміні відносних

витрат i-го ресурсу. Якщо заміри виконувати у відсотках, то рівень (46) показує, на скільки відсотків зміниться обсяг випуску, якщо відносні витрати i-го ресурсу зміняться на один відсоток за умови, що витрати іншого ресурсу залишаться незмінними.

Сума $E = E_1 + E_2$ називається *еластичністю виробництва.*

Приклад. Для ВФ $y = a_1 x_1 + a_2 x_2$ обчислити величини E_1, E_2, E.

$$E_1 = \frac{x_1}{f(x_1, x_2)} \frac{\partial f}{\partial x_1} = \frac{x_1 a_1}{a_1 x_1 + a_2 x_2}; \qquad E_2 = \frac{x_2}{f(x_1, x_2)} \frac{\partial f}{\partial x_2} = \frac{x_2 a_2}{a_1 x_1 + a_2 x_2};$$

$$E = E_1 + E_2 = 1.$$

В економічних дослідженнях досить часто виникає питання про можливість заміни одного ресурсу іншим за умови збереження обсягу виробництва. Граничною нормою заміни (заміщення) i-го ресурсу (фактора виробництва) j-им за певних їх величин називається відношення R_{ij}:

$$R_{ij} = -\frac{dx_j}{dx_i} \qquad (i, j = 1, 2) \tag{47}$$

за умови сталого значення y – величини ВФ.

Наголосимо, що i – номер того ресурсу, який замінюють; j – номер ресурсу, яким замінюють; нескінченно малі прирости dx_i та dx_j незалежні.

Наведемо правило обчислення відношення (47) за умови відомої ВФ. Припустимо, що обсяги ресурсів (факторів виробництва), маючи певні величини, змінюються так, що величина y обсягу виробництва має стале значення, тобто відносна зміна величин x_1 та x_2 відбувається так, що їх значення утворюють одну ізокванту (лінію рівня). За таких умов повний диференціал $y = f(x_1, x_2)$ дорівнюватиме нулю:

$$dy = \frac{\partial f(x_1, x_2)}{\partial x_1} dx_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} dx_2 = 0. \tag{48}$$

Розв'язуючи рівняння (48) стосовно бажаного відношення диференціалів залежних змінних (вибравши, який ресурс яким замінено), маємо:

$$\frac{dx_j}{dx_i} = -\frac{\partial f(x_1, x_2)}{\partial x_i} \Big/ \frac{\partial f(x_1, x_2)}{\partial x_j} \qquad (i, j = 1, 2;\quad i \neq j).$$

Порівнявши одержаний результат з (47), маємо:

$$R_{ij} = -\frac{dx_j}{dx_i} = \frac{\partial f(x_1, x_2)}{\partial x_i} \Big/ \frac{\partial f(x_1, x_2)}{\partial x_j} \qquad (i, j = 1, 2; \quad i \neq j). \tag{49}$$

Скориставшися рівністю (46), переконуємося, що для двофакторної ВФ виконується співвідношення:

$$R_{1,2} = \frac{E_1}{E_2} \frac{x_2}{x_1}.$$

Тобто гранична норма заміни першого ресурсу другим дорівнює відношенню еластичності випуску продукції по першому і другому ресурсах, помноженому на відношення обсягу другого до обсягу першого ресурсу. Якщо названа умова не буде виконана, то, заміняючи обсяги ресурсів, не можна забезпечити сталість обсягів виробництва.

Для ВФКД, якщо $x_1 = K$, $x_2 = L$, відношення $\frac{x_1}{x_2} = \frac{K}{L}$ називається *капіталозабезпеченістю праці*. У цьому випадку гранична норма заміни основного капіталу працею дорівнює відношенню еластичності випуску за основним капіталом і працею, поділеному на капіталозабезпеченість праці.

Для двофакторної ВФ за умови сталого обсягу виробництва та малих приростів Δx_1 та Δx_2 маємо наближену рівність:

$$R_{1,2} = -\frac{dx_2}{dx_1} \approx -\frac{\Delta x_2}{\Delta x_1}. \tag{50}$$

Тобто гранична норма заміни ресурсів $R_{1,2}$ наближено визначає, у скільки разів необхідно збільшити витрати другого ресурсу при зменшенні витрат першого ресурсу на одиницю, щоб забезпечити сталість виробництва.

Приклад. Для ВФКД $y = a_0 x_1^{a_1} x_2^{a_2}$ обчислити величини $R_{1,2}$ та $R_{2,1}$.

$$R_{1,2} = \frac{\partial f}{\partial x_1} \Big/ \frac{\partial f}{\partial x_2} = \frac{a_1 x_2}{a_2 x_1}; \quad R_{2,1} = \frac{\partial f}{\partial x_2} \Big/ \frac{\partial f}{\partial x_1} = \frac{a_2 x_1}{a_1 x_2}.$$

3.7. Врахування часу при розбудові виробничих функцій

Виробнича функція називається *динамічною*, якщо:

1) час t ураховується як самостійна змінна, від якої залежить обсяг виробництва;
2) параметри ВФ залежать від часу t.

Динамічні ВФ використовуються в основному для дослідження зміни контрольованого показника з часом і для розбудови прогнозів його значень у майбутньому.

Якщо параметри ВФ визначалися за даними часових рядів (обсяги використовуваних ресурсів і продукції) на базовому проміжку в N періодів (місяців, кварталів, років), то горизонт прогнозу за таких умов з використанням відповідної ВФ недоцільно брати більш ніж на $N/3$ періодів. Використання часових рядів і ВФ буде детально розглянуто в наступній главі.

При розбудові ВФ науково-технічний прогрес (НТП) іноді враховують шляхом використання співмножника НТП, вибраного як $e^{\rho t}$, де параметр ρ характеризує *темп приросту обсягів виробництва* за рахунок НТП:

$$y(t) = e^{\rho t} f(x_1(t), x_2(t)).$$

Така структура ВФ є найпростішим зразком динамічної ВФ, коли вплив НТП оцінюється безпосередньо, а не опосередковано через виробничі фактори. У більш складних ситуаціях НТП впливає безпосередньо на виробничі фактори, в тому числі на продуктивність праці та віддачу капіталовкладень.

На підставі даних про розвиток економіки СРСР у 1960-1985 рр. (динаміка національного прибутку, чисельність зайнятих у матеріальному виробництві, обсяги основних фондів) ВФКД без урахування НТП була представлена так: $Y = 1{,}022 K^{0{,}5382} L^{0{,}4618}$. При підстановці фактичних даних K і L за 1986 р. похибка прогнозу була 3%.

З урахуванням НТП ВФКД стосовно того ж проміжку часу має вигляд $Y = 1.038 e^{-0{,}0294t} K^{0{,}9749} L^{0{,}2399}$ *.

* Гранберг А. Г. Моделирование социалистической экономики. – М.: Экономика, 1988.

Питання для самоконтролю

1. Мета розбудови ВФ.
2. Закон спадаючої ефективності. Як закон віддзеркалюється на геометричному зображенні ВФ?
3. Призначення ВФ Кобба-Дугласа. Як визначаються середня продуктивність та капіталозабезпеченість праці?
4. Основні властивості ВФ, їх економічний сенс.
5. Ізокванта та її економічний сенс.
6. Основні етапи розбудови ВФ.
7. За якими умовами виконується параметризація ВФ?
8. Верифікація ВФ. Як вона виконується?
9. Дати визначення частинної еластичності випуску за i-им ресурсом (фактором виробництва).
10. Якою є різниця між поняттями середньої та граничної продуктивності ресурсу (фактора виробництва)?
11. Як змінюється гранична норма заміни одного ресурсу іншим при рухові по ізокванті?
12. Основні засади та сфера використання методу найменших квадратів (МНК).
13. Статистичні властивості параметрів ВФ, обчислюваних за МНК.

ГЛАВА 4

ПРОГНОЗУВАННЯ ЕКОНОМІЧНИХ ПОКАЗНИКІВ З ВИКОРИСТАННЯМ ЧАСОВИХ РЯДІВ

4.1. Загальні поняття

Прогнозування (від грец. prognosis – передбачення) має дати відповідь на запитання: що найімовірніше може статися в майбутньому та за яких умов можна досягти очікуваного результату. У залежності від вимог і ступеня конкретизації досліджуваних процесів та явищ розрізняють три форми передбачення: гіпотеза, прогноз, план.

Дослідження будь-якого процесу розпочинається з *гіпотези*, тобто припущення стосовно факторів і зв'язків між ними, механізму функціонування та перебігу процесу дослідження в часі з урахуванням експериментальних і наукових відомостей та інтуїції. Користуючись гіпотезою, будують деяку загальну схему математичної моделі (ММ) перебігу досліджуваного процесу в часі. Експериментальні дані, одержані при дослідженні процесу в минулому, дозволяють надати кількісні значення певним параметрам ММ, за допомогою яких можна оцінити значення тих параметрів, які до певної міри визначають майбутній стан процесу.

Прогноз у порівнянні з гіпотезою має бути більш визначеним, обумовленим певними перевірками гіпотези або гіпотез, використаних при розбудові ММ прогнозу. Майбутній стан будь-якого процесу в певній мірі обумовлений не лише зміною закономірностей, які його визначають, а й багатьма випадковими факторами, взаємозв'язок та виміри впливу яких в повному обсязі врахувати практично неможливо. За названої причини *прогнози завжди мають ймовірний характер*.

План являє собою систему взаємопов'язаних та взаємоузгоджених заходів, зорієнтованих на досягнення певної мети в майбутньому, з урахуванням матеріально-технічних, фінансових та інших факторів.

Мета економіко-виробничих прогнозів може бути у загальних контурах сформульована так: визначення перспектив майбутнього розвитку досліджуваного процесу з урахуванням дійсного стану в минулому, розбудова оптимальних перспективних планів з урахуванням результатів прогнозу та оцінок планових заходів по досягненню бажаних результатів.

Економіко-виробничі прогнози за масштабністю досліджуваних процесів з певною мірою умовності доцільно розділити так: 1) мікроекономічні – стосуються планування діяльності окремих фірм, підприємств, виробництв і запитів на певні види продукції; 2) галузеві – присвячені дослідженням у певній галузі промисловості; 3) прогнози закономірностей розвитку господарських комплексів і

галузей, поєднаних спільними інтересами; 4) регіональні; 5) макроекономічні, коли досліджуються найбільш загальні об'єднані показники економічного розвитку країни.

За функціональним призначенням прогнози доцільно розділити на *пошукові* та *нормативні*. При пошукових прогнозах досліджуються можливі напрямки та кількісні виміри параметрів ймовірного розвитку процесу. Мета нормативного прогнозу – дослідження можливих шляхів, фінансових, матеріально-технічних та інших засобів і термінів досягнення бажаного стану певного процесу в майбутньому.

Вихідні положення, використовувані засоби та методи оцінок якості прогнозування суттєво залежать від того, на який проміжок часу майбутнього прагнуть зробити прогноз. За величиною проміжку упередження прогнози поділяються на оперативні (як правило, на проміжок до одного місяця), короткотермінові (від кількох місяців до року), середньотермінові (від одного до п'яти років), довгострокові (понад п'ять років).

Методи прогнозування дозволяють за результатами аналізу ретроспективних зовнішніх та внутрішніх зв'язків і факторів, які визначають перебіг процесу, а також за їх кількісною мірою зробити в межах розбудованої ММ висновки стосовно очікуваного майбутнього розвитку.

У розбудові методів прогнозування використовують *два принципово різних підходи. Перший шлях* – вивчення причинно-наслідкових механізмів: вивчивши сукупність факторів, які визначають перебіг процесу, моделюють їх взаємозв'язки та наслідки за умов їх певної кількісної міри. Використовуючи професійні знання, певні теоретичні положення, практичний досвід вивчення процесів аналогічних досліджуваному, приймають певну концепцію аналітичної залежності між параметрами, вибраними для розбудови ММ.

Оскільки значення параметрів у відповідних ММ не можуть вважатися детермінованими, тобто завжди є випадковими величинами, які визначаються за статистичними даними, то ММ такого змісту та призначення називаються *економетричними*. Економетричні моделі аналітично представляються системами рівнянь та нерівностей певного змісту, які пов'язують параметри, використані при розбудові ММ. Основні засади такого підходу були стисло описані при розгляді виробничих функцій.

Другий шлях – розбудова певних аналітичних залежностей, які дають можливість скласти прогноз за наслідками аналізу часового (динамічного) ряду зміни досліджуваного показника в минулому, *не вивчаючи причин та умов його зміни*. Саме методи використання часових (динамічних) рядів для прогнозування будуть розглянуті в цій главі.

4.2. Часові ряди

4.2.1. Різновиди часових рядів

Доцільно виділяти два різновиди прогнозованих характеристик процесу, залежних від часу: *змінні стану* та *змінні інтенсивності*.

Характерною особливістю змінної стану є те, що її значення хоч і залежить від часу, але не залежить від проміжків чергових її замірів від моменту початку спостережень. Кількісна міра змінної інтенсивності істотно залежить від величини проміжку між сусідніми замірами. Наприклад, температура ґрунту біля поверхні протягом доби залежить від часу, але на її значення не впливають проміжки чергових замірів.

Інша справа – кількість опадів. Ця характеристика залежить від проміжків поточних замірів. Змінна стану характеризує кількісну міру якогось фактора, а змінна інтенсивності – швидкість зміни фактора в часі (доречно нагадати, що величини можуть змінюватися і в просторі).

Методи прогнозування змінних стану та інтенсивності відрізняються між собою такими особливостями:

- якщо заміри досліджуваного показника виконувати через різні проміжки, то величину проміжку між сусідніми замірами необхідно враховувати при дослідженні змінних інтенсивності, а для оцінки перебігу змінних стану величина проміжку не є визначальною;
- оскільки прогнози, як правило, виконуються для кількох послідовних проміжків в межах деякого часу, то величини прогнозованого показника змінної стану для кінця проміжку упередження не залежать від його проміжних значень на цьому проміжку, а для змінної інтенсивності прогноз має являти собою суму значень у проміжних точках;
- функція розподілу ймовірності похибок прогнозу в часі для змінної стану має відповідати функції розподілу ймовірностей похибок вихідних даних, а для змінної інтенсивності закон розподілу ймовірностей похибок прогнозу в часі наближується до нормального при будь-якому законі розподілу ймовірностей похибок вихідних даних, оскільки ці похибки є сумою похибок прогнозу в проміжних точках замірів.

4.2.2. Тренд

Абсолютна більшість методів прогнозування з використанням часових рядів ґрунтується на ідеї екстраполяції, тобто перенесенні на майбутнє тенденції в зміні досліджуваного показника, яка спостерігалася в минулому до усього часу включно.

Під тенденцією розуміють деякий загальний характер змін досліджуваного показника, обумовлений внутрішніми взаємозв'язками факторів, які впливають на перебіг процесу. Тенденцію зміни показника в кількісному вимірі в координатах "час – досліджуваний показник" зображують плавною лінією (траєкторією).

Досліджуючи перебіг певного показника в часі (траєкторію), виділяють умовно такі його складові частини:

- *тренд,* який віддзеркалює вплив причинно-наслідкових закономірностей, властивих досліджуваному процесу і обумовлених довгодіючими факторами його природи (наприклад, збільшення чисельності населення, вплив науково-технічного прогресу і т. ін.);
- *сезонна компонента,* обумовлена можливими повтореннями впливу деяких тимчасових факторів протягом відносно короткого терміну – місяця, тижня, певного дня тижня (наприклад, сезонний попит на товари, сезонна інтенсивність експлуатації певної техніки і т. ін.);
- *циклічна компонента,* з використанням якої прагнуть врахувати можливі періодично повторювані умови змін в перебігу досліджуваного процесу (наприклад, вплив циклів сонячної активності на чисельність певної популяції комах);
- *випадкова компонента,* за допомогою якої враховують вплив випадкових, непередбачуваних факторів.

Через неможливість або істотні труднощі в одночасному дослідженні всіх чотирьох компонент перебігу процесу в часі виділяють дві складові: невипадкову (закономірну) та випадкову. Невипадкова складова процесу об'єднує три раніше названі компоненти: тренд, сезонну та циклічну, але в літературі часто за цією складовою залишають назву "тренд". Такого тлумачення тренда і будемо дотримуватися надалі. Тренд в певному розумінні характеризує природну закономірність зміни досліджуваного показника в часі, звільнену в основному від впливу випадкових факторів. Для практичного використання поняття тренда доцільно й досить ефективно прийняти його як певну детерміновану складову динаміки розвитку процесу, яка визначається постійнодіючими факторами та взаємозв'язками між ними, що обумовлює перебіг процесу. Необхідно наголосити, що форма аналітичного представлення тренда для одного й того ж часового ряду може бути різною, але кількісні значення показника для одних і тих же точок спостереження мають бути близькими (відносні відхилення значень менші одиниці).

Однією з причин широкого використання (при прогнозуванні) методів екстраполяції тенденції зміни часових рівнів є відсутність іншої інформації, окрім дискретних значень досліджуваного показника.

При такому підході до прогнозування прагнемо врахувати лише постійнодіючі фактори, які обумовлюють перебіг процесу, а впливом множини випадкових факторів, складно оцінюваних і малозначущих у кількісному відношенні, нехтуємо.

Тому використання різних методів екстраполяції ґрунтується на таких припущеннях: 1) тенденція зміни в часі кількісної міри досліджуваного показника може бути представлена певною аналітичної формулою, а у вибраних координатах – плавною траєкторією-трендом; 2) умови, які визначали тенденцію зміни в минулому, несуттєво змінюються в недалекому майбутньому.

При дослідженні часових рядів економічних процесів визначальними є пошук і статистична оцінка тенденції перебігу процесу та можливих відхилень.

Горизонтом прогнозу, або *терміном упередження*, називається проміжок від моменту, для якого зафіксовано останнє експериментальне значення показника, до моменту, з яким співставляють прогноз.

Необхідно чітко розуміти, що екстраполяція дає можливість обчислити точкові значення прогнозованого показника. Але значення досліджуваного показника для будь-якого моменту, в тому числі для прогнозованого, є *неперервною випадковою величиною з певного інтервалу*, бо на її формування впливають не тільки детерміновані, а й чимало випадкових факторів. Отже, ймовірність того, що по досягненні в часі моменту, для якого виконано прогноз, дійсне значення показника буде рівним прогнозованому, дорівнює нулю. Є сенс говорити лише про межі, в яких буде знаходитися прогнозоване значення. Одним із шляхів одержання таких меж є *обчислення величини інтервалу довіри значень прогнозованого показника.*

Методи прогнозування істотно залежать від горизонту прогнозу: короткотермінового, середньотермінового, довготермінового.

При короткотерміновому прогнозуванні числові значення досліджуваного показника беруться за проміжки (день, тиждень, місяць, квартал) в минулому, а прогноз обчислюється на один-два вибрані проміжки в майбутньому.

При середньо- та довготерміновому прогнозуванні значення показника, як правило, щорічні, а прогноз треба будувати на кілька років наперед. Використовувані методи такого прогнозування ідентичні, тому надалі не будемо виділяти їх окремо.

4.2.3. Проміжки переоцінки та уточнення прогнозу

Величина проміжків між замірами показника з метою перевірки та уточнення раніше виконаного прогнозу залежить переважно від терміну упередження та найбільшої частоти циклічних змін у процесі при їх наявності. Тому проміжки переоцінки прогнозу змінюються в дуже широких межах: від кількох годин до років, в залежності від природи досліджуваного процесу. Так, переоцінка та уточнення обсягів лісових насаджень з періодом поновлення 80 років може виконуватись раз в десять років, а заміри споживання електроенергії щогодинно. Раціональний вибір

величини проміжків спостереження для переоцінки прогнозу має істотне значення для ефективного планування функціонування відповідних систем.

При дослідженні операцій важливими є методи пошуку нових, ефективних ММ та методів уточнення прогнозів за нових поточних спостережень.

4.2.4. Вихідні дані

Вихідні дані являють собою наслідки вибіркових спостережень або змінної стану, або змінної інтенсивності. Управлінські рішення, які приймаються з використанням результатів прогнозу, в сьогочасний момент обумовлюють певні наслідки за проміжок упередження, тому до якості вихідних даних ставляться високі вимоги. Практика свідчить, що результати спостережень досить часто реєструються зі значними похибками, обумовленими якістю самих замірів та їх записами. Крім того, досліджуваний процес за своєю природою має стохастичний (ймовірний) характер, наприклад, ефективність мінеральних та органічних добрив стосовно урожайності. Результати спостережень можуть мати аномальні ефекти, наприклад, спалах епідемії та її вплив на чисельність певного виду тварин або комах в регіоні. За названих причин не кожну сукупність даних доцільно використовувати для розбудови часового ряду, при дослідженні якого виконується прогноз. До розбудови моделі прогнозу з використанням експериментальних даних з них треба вилучити "викиди" (аномальні відхилення). Один із таких способів вилучення аномальних даних показаний на рис. 4.1, де у вибраній системі координат зображено залежність величин деякого показника від часу.

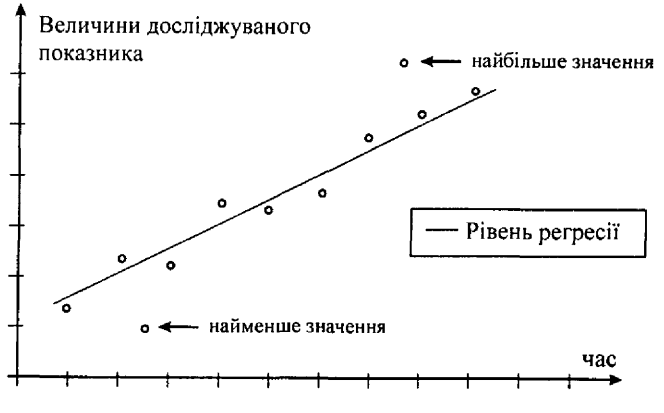


Рис. 4.1.

Оскільки найбільше та найменше значення результатів спостереження у зв'язку з їх різкими відхиленнями від решти величин можуть бути ймовірними викидами, то вилучаємо їх з наявної вибірки даних. Аналіз графічного зображення решти даних свідчить про доцільність використання лінійної залежності величини досліджуваного показника від часу.

Величини значень досліджуваного показника в часі називаються в літературі *рівнями часового ряду.*

4.3. Аналіз тенденцій зміни рівнів часового ряду

При аналізі часових рядів досліджуються тенденції трьох характеристик: тенденція зміни середніх значень рівнів часового ряду; тенденція дисперсії; тенденція автокореляції.

4.3.1. Дослідження тенденції зміни середнього рівня

До побудови тренда (траєкторії змін значень досліджуваного показника в часі) доцільно перевірити гіпотезу, чи дійсно рівні ряду змінюються в часі чи ж наявні значення різниць їх величин обумовлені лише дією випадкових факторів.

Наведемо два способи перевірки гіпотези про наявність тренда. Кількість членів ряду, як правило, не велика (10-20 значень), тому скористаємося методом перевірки для малих за обсягом вибірок за нормального закону розподілу досліджуваної ознаки. Наявний ряд значень $y_1, y_2, \ldots, y_n$ розділимо на дві групи, зберігаючи часову послідовність. Елементи кожної групи розглядаємо як незалежні вибірки з кількістю елементів n_1 та n_2 ($n_1 + n_2 = n$). Позначимо $\bar{y}_1$ та $\bar{y}_2$ відповідні значення середніх величин у кожній групі. Необхідно перевірити гіпотезу про існування різниці ($\bar{y}_1 - \bar{y}_2$) за певного рівня значимості α, тобто перевірити припущення: наявне значення величини ($\bar{y}_1 - \bar{y}_1$) є закономірним чи випадковим. По кожній групі обчислюємо виправлені вибіркові дисперсії s_1^2 та s_2^2. Нагадаємо, що виправлена вибіркова дисперсія s^2 за наявності n елементів вибірки $\{x_i\}$ обчислюється за формулою:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

($\bar{x}$ – середнє значення величин у сукупності $\{x_i\}$).

Перевірка гіпотези про рівність дисперсії за вибраного рівня значимості α

виконується за допомогою F-критерія Фішера-Снедекора, побудованого на результатах порівняня розрахункового відношення

$$F = s_1^2 / s_2^2 \text{ (індекси 1 та 2 вибираються за умови } s_1 > s_2) \quad (1)$$

з табличним $F(\alpha;\ n_1-1;\ n_2-1)$. Таблиці значень $F(\alpha;\ k_1;\ k_2)$ – критерію Фішера-Снедекора наведено в літературі з математичної статистики. Якщо обчислене значення F менше табличного за відомих α, n_1-1, n_2-1, то приймається гіпотеза про тотожність дисперсій, тобто наявна розбіжність є випадковою.

Перевірка основної гіпотези $\bar{y}_1=\bar{y}_2$ виконується з використанням T-статистики Стьюдента, для чого обчислюється величина T-критерію:

$$T = (\bar{y}_1 - \bar{y}_2)/s \quad (2)$$

(s – так звана середня виважена величина дисперсій s_1 та s_2, яка обчислюється за формулою:

$$s = \sqrt{\frac{(n_1-1)^2 s_1^2 + (n_2-1)^2 s_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}). \quad (3)$$

Якщо $T \geq T_\alpha$ (величина T_α визначається в таблиці розподілу за Стьюдентом з вибраним рівнем значимості α), то гіпотеза про відсутність тренда середніх відхиляється з ймовірністю α, в протилежному випадку приймається.

Більш надійним, але й більш об'ємним за обчисленням є метод Ф. Фостера та А. Стюарда. Згідно з цим методом, за даними досліджуваного ряду обчислюються допоміжні змінні v_t та l_t шляхом послідовного порівняння рівнів ряду. Якщо взятий рівень ряду y_t за своєю величиною *перевищує усі* попередні рівні, то змінній v_t дається значення 1, а в протилежному випадку – 0:

$$v_t = \begin{cases} 1, \text{ якщо } y_t > y_{t-1},\ y_{t-2},\ y_{t-3}, \ldots,\ y_1, \\ 0 \text{ за іншої умови.} \end{cases} \quad (4)$$

Значення змінної l_t обчислюється так:

$$l_t = \begin{cases} 1, \text{ якщо } y_t < y_{t+1},\ y_{t+2},\ y_{t+3}, \ldots,\ y_n, \\ 0 \text{ за іншої умови.} \end{cases} \quad (5)$$

Після обчислення v_t та l_t визначають ще дві величини – s_t та d_t: ($s_t = v_t + l_t$, $d_t = v_t - l_t$). Отримавши множину значень s_t і d_t, обчислюють характеристики, обумовлені всією сукупністю значень досліджуваного показника:

$$s = \sum_{i=1}^{n}(v_t + l_t); \quad d = \sum_{i=1}^{n}(v_t - l_t), \quad (6)$$

$s_t = 0$, якщо y_t не є ні найбільшим, ні найменшим значенням серед усіх попередніх значень, в протилежному випадку $s_t = 1$. Таким чином, завжди буде виконуватися умова: $0 < s \leq n-1$. Якщо всі значення досліджуваного показника тотожні (дисперсія нульова), то $s = 0$; якщо вони змінюються монотонно (спадають або зростають) або ж коливання їх чергуються, послідовно збільшуючись або зменшуючись, то $s_t = n-1$.

Можливими значеннями змінної d_t є 0; 1; –1, тому можлива нижня межа d дорівнює $(n-1)$, а верхня $(n+1)$. Ці межі досягаються за монотонних часових рядів.

Величини s і d випадкові й незалежні, як доведено в теорії, їх розподіл асимптотично збігається до нормального за умови $n \to \infty$. Вони істотно залежать від розташування величин значень показника стосовно один одного.

Для величин s і d, залежних від випадкового розподілу рівнів та кількості членів часового ряду, в табл. 4.1 наведені значення: μ – математичного сподівання та середніх квадратичних відхилень σ_1 і σ_2 відповідно для s і d.

Таблиця 4.1.

n	μ	σ_1	σ_2	n	μ	σ_1	σ_2
10	3,858	1,288	1,954	60	7,360	2,201	2,713
15	4,636	1,521	2,153	65	7,519	2,236	2,742
20	5,195	1,677	2,279	70	7,666	2,268	2,769
25	5,632	1,791	2,373	75	7,803	2,297	2,793
30	5,990	1,882	2,447	80	7,931	2,324	2,816
35	6,294	1,956	2,509	85	8,051	2,349	2,837
40	6,557	2,019	2,561	90	8,165	2,373	2,857
45	6,790	2,072	2,606	95	8,273	2,395	2,876
50	6,998	2,121	2,645	100	8,375	2,416	2,894
55	7,187	2,163	2,681				

Значення величини s, обчислене для конкретного ряду, використовується для виявлення тенденції зміни дисперсії, а значення d – тенденції зміни середньої. Для цього перевіряють гіпотези: чи можна вважати випадковими різниці $s - \mu$ та $d - 0$. Ці гіпотези можна перевірити, використовуючи T-критерій Стьюдента (2):

$$T_1 = (s - \mu)/\sigma_1; \quad T_2 = (d - 0)/\sigma_2. \qquad (7)$$

Випадкові величини T_1 і T_2 мають розподіл Стьюдента з $k = n - 1$ степенями волі. Значення T_1 і T_2, обчислені за формулою (7), порівнюються з табличними відповідно до значень критичних точок розподілу Стьюдента з $k = n - 1$ степенями волі за вибраного рівня значимості α.

Якщо $T_1 > T_{кр}(\alpha_1, k)$, то наявна тенденція зміни в дисперсіях, і вона може бути описана певним трендом, в протилежному випадку тенденція в зміні дисперсії відсутня.

Якщо $T_2 > T_{кр}(\alpha_2, k)$, то гіпотеза про відсутність тенденції зміни середнього значення відхиляється, тобто тренд існує. За умови $T_2 < T_{кр}(\alpha_2, k)$ тренд не існує.

Описаний метод не вимагає складних обчислень і легко реалізується на практиці. За наявності тенденції переходять до її аналітичного представлення засобами, які будуть описані далі.

Розглянемо приклад пошуку наявності тенденції. У табл. 4.2 наведені величини рівнів часового ряду (обсяги попиту на певний товар) та значення величин v_t і l_t.

Таблиця 4.2.

t	Рівень	v_t	l_t	t	Рівень	v_t	l_t
1	103	–	–	9	171	0	0
2	143	1	0	10	77	0	0
3	78	0	1	11	153	0	0
4	158	1	0	12	163	0	0
5	144	0	0	13	199	0	0
6	167	1	0	14	144	0	0
7	153	0	0	15	187	0	0
8	202	1	0	16	207	1	0

Скориставшись одержаними значеннями v_t і l_t, маємо $s = 6$, $d = 4$. Величини μ, σ_1 та σ_2 знаходимо за табл. 4. 1. А оскільки не наведено значення для $n = 16$ (найближчі табличні значення маємо для $n = 15$ та $n = 20$), то необхідні величини обчислюємо, скориставшися лінійною інтерполяцією. Маємо $\mu = 4{,}749$, $\sigma_1 = 1{,}552$, $\sigma_2 = 2{,}178$; $T_1 = (6 - 4{,}749)/1{,}552 = 0{,}81$; $T_2 = 4/2{,}178 = 1{,}84$. З рівнем значимості $\alpha = 0{,}1$ (це значення вибираємо самі) найближче табличне значення розподілу Стьюдента більше 1,7. Таким чином, гіпотеза про відсутність тенденції в зміні дисперсії приймається, а гіпотеза про відсутність тенденції в зміні середнього відхиляється, тобто тренд існує.

4.3.2. Елементарні засоби статистичного аналізу динаміки часових рядів

Запропоновані засоби статистичного аналізу динаміки часових рядів дозволяють зробити перший крок у дослідженнях тенденції їх розвитку та обчислити деякі узагальнені характеристики оцінки змін у часі, що дає можливість виявити характер тенденції змін у минулому та зробити до певної міри більш обґрунтований прогноз.

Результати практичних спостережень рівнів часового ряду, як наголошувалося раніше, розглядаються як сума чотирьох компонент, кожна з яких не може бути виміряна безпосередньо. Це *тренд*, обумовлений внутрішніми природними закономірностями досліджуваного процесу; *циклічна складова*, обумовлена періодичністю коливання деяких факторів, які впливають на перебіг процесу; *сезонна складова*; *випадкова величина*, яка віддзеркалює результати впливу множини випадкових факторів, кожний з яких, взятий окремо, є малозначущим. Такий покомпонентний розподіл умовний, але він ефективно допомагає в розв'язанні практичних задач і тому широко використовується при дослідженнях закономірностей зміни рівнів часових рядів.

По тенденціях зміни досліджуваний показник може бути представлений *аналітичною залежністю від часу*, що за геометричного тлумачення в координатах "час – показник" відповідає певній лінії, яку називають *траєкторією*.

Аналітична залежність досліджуваного показника від часу та його траєкторія показують наслідки впливу основних закономірностей на перебіг процесу, які до певної міри позбавлені впливу випадкових факторів. Приймають, що відхилення експериментальних даних від відповідних аналітичних величин обумовлене впливом випадкових причин. Таким чином, рівні $y(t)$ часового ряду доцільно представити у вигляді залежності

$$y(t) = f(t) + \varepsilon(t) \qquad (8)$$

($f(t)$ – аналітичне представлення тренда з урахуванням можливих циклічних і сезонних складових, $\varepsilon(t)$ – міра відхилення наявних експериментальних даних від відповідних аналітичних величин, обумовлена дією випадкових факторів, другорядних і малозначущих стосовно постійно діючих. Доцільно наголосити, що аналітичне зображення $f(t)$ для одного й того ж часового ряду може бути різним, але значення $y(t)$ за різних аналітичних представлень для одно й того ж моменту будуть близькі за величиною. Таким чином, розщеплення величини рівня ряду на детерміновану (закономірну) та випадкову складові (8) має до певної міри умовний характер).

Виділення детермінованої складової в значенні рівня ряду та побудова тренда

є основою простих статистичних методів прогнозування. Оскільки спостереження дають лише одну реалізацію часового ряду, то не можна користатися апаратом теорії ймовірностей і слід шукати інші підходи. Перейдемо до розгляду методів дослідження закономірностей зміни рівнів часових рядів.

4.3.3. Плинні середні

Плинною середньою називають середню величину рівнів ряду для кількох послідовних проміжків вихідного ряду. Використання плинних середніх дозволяє до певної міри уникнути наслідків впливу випадкових факторів і чіткіше виявити вплив основних причинних факторів, які визначають перебіг процесу.

Метод плинної середньої ґрунтується на ідеї обчислення середньої величини для проміжку, який є сумою кількох послідовних проміжків із заданими на їх кінцях рівнями досліджуваного часового ряду. Новий ряд, рівні якого є певним чином усередненими значеннями рівнів вихідного ряду, має менші відхилення від закономірних значень досліджуваного показника. Дійсно, якщо величина випадкового відхилення для певної точки y_t визначається дисперсією σ^2, то випадкове відхилення для рівнів усередненого ряду буде характеризуватися дисперсією σ^2/m (m – кількість членів ряду, використаних за обчислення середньої). Обчислена середня по об'єднаних проміжках ставиться, як правило, у відповідність *середині об'єднаного проміжку*. А оскільки об'єднаний проміжок усереднення може мати парну або непарну кількість вихідних проміжків, то методика обчислення середньої дещо змінюється.

За умови усереднення по непарній кількості $2p+1$ вихідних проміжків середня величина ставиться у відповідність центральній точці. Якщо точки проміжків, які використовуються для обчислення середньої, умовно пронумерувати від 1 до $2p+1$, то формула обчислення середньої $\bar{y}_{p+1}$ буде мати вигляд:

$$\bar{y}_{p+1} = \frac{1}{2p+1}\sum_{i=1}^{2p+1} y_i \,. \qquad (9)$$

Часовий ряд плинних середніх $\bar{y}_k$ буде мати рівні, які обчислюються за формулою:

$$\bar{y}_k = \bar{y}_{k-1} + \frac{1}{2p+1}(y_{k+p} - y_{k-p-1}),\ (k = p+2, \ldots, n-p) \qquad (10)$$

(n – кількість точок вихідного ряду; k – поточний індекс членів ряду плинних середніх).

За названих умов по завершенні операції усереднення одержаний часовий ряд буде мати на $2p$ елементів менше, ніж вихідний ряд.

Якщо для усереднення використовується парна кількість вихідних проміжків, правим кінцям яких надано номери $i+1$, $i+2$, …, $i+2p$, то величина середньої ставиться у відповідність моменту $(i+p+\frac{1}{2})\Delta t$ (Δt – величина проміжку між сусідніми послідовними моментами спостереження рівнів вихідного ряду). Так одержану середню величину не можна порівнювати зі значеннями рівнів вихідного ряду, оскільки величини належать до різних моментів часу. Тому використовують так звану *центровану плинну середню*, яка є середньою величиною двох сусідніх плинних середніх і порівнюється з центром проміжку усереднення. Якщо точки, які використовуються за обчислення плинної середньої за вибору парної кількості проміжків, умовно пронумерувати від 1 до $2p$ (нумерація правих кінців проміжків), то формула обчислення центрованої плинної середньої буде мати вигляд:

$$\bar{y}_p = \frac{0{,}5y_1 + \sum_{i=2}^{2p} y_i + 0{,}5y_{2p+1}}{2p}. \qquad (11)$$

З використанням певної кількості проміжків при усередненні кількість членів *часового ряду плинних середніх* зменшується на число членів у проміжку усереднення.

Одним із суттєвих питань використання плинних середніх при дослідженні тенденції перебігу процесу є вибір величини проміжку усереднення. Вирішенню цього питання має передувати економічний аналіз конкретного процесу поряд з побудовою діаграми в осях "час – досліджуваний показник". Діаграма дає можливість виявити в загальних рисах характер тенденції перебігу процесу, наявність можливих сезонних і циклічних коливань. Методи з використанням плинних середніх детально досліджені*.

При прогнозуванні плинну середню порівнюють із завершальною точкою проміжку усереднення і називають *адаптивною (пристосованою) плинною середньою*, яку позначають u_k. За такої умови замість формули (10) маємо:

$$u_k = \frac{1}{2p+1}\sum_{i=k-2p}^{k} y_{k-2p} = u_{k-1} + \frac{1}{2p+1}(y_k - y_{k-2p});\quad (k = 2p+1, \ldots, n). \qquad (12)$$

* Ланге О. Введение в эконометрию / Пер. с польск. – М., 1964.

Вирівнювання з допомогою простих плинних середніх в деяких випадках настільки суттєво змінює тенденцію перебігу, що значущі для виконання аналізу особливості спотворюються або зникають, тому використовуються дещо інші методи оцінок тенденції розвитку за допомогою наявних даних спостереження.

4.3.4. Зважені плинні середні

При використанні раніше описаних плинних середніх рівні ряду приймалися рівнозначимими стосовно впливу на середнє значення незалежно від їх віддаленості в часі від поточного моменту. Природно зробити припущення: те, що відбувалося в більш віддаленому часі, має менший вплив на стан процесу, ніж те, що відбувалося ближче до *досліджуваного моменту*. Аналітично таке природне припущення реалізується в такий спосіб: при обчисленні так званих *зважених плинних середніх* кожен рівень в межах проміжку усереднення домножується на певний числовий коефіцієнт – "вагу" в залежності від розташування цього рівня стосовно середини проміжка усереднення. Система таких вагових коефіцієнтів визначається за таких міркувань: приймаємо, що зміну рівнів часового ряду для *симетрично* розташованих моментів відносно моменту, для якого обчислюється зважена плинна середня, можна описати за допомогою залежного від часу многочлена певного степеня:

$$y(t) = a + bt + ct^2 + dt^3 + \ldots \qquad (13)$$

Відлік часу зручно умовно виконувати в межах проміжку усереднення від його середини. Приймаючи, що сусідні рівні ряду співставлені рівновіддаленим моментам з кроком Δt, а величина часу t буде залежною від Δt та номера точки i ($i = 0$; ±1; ±2; ...), з відомим розташуванням середини проміжку усереднення, формулу (13) запишемо так:

$$\bar{y}_i = a + b(i\Delta t) + c(i\Delta t)^2 + d(i\Delta t)^3 + \ldots \qquad (14)$$

Значення параметрів $a, b, c, d, \ldots$ обчислюються за вимоги, щоб сума квадратів відхилень наявних рівнів ряду від знайдених згідно з (14) досягала найменшого значення на проміжку усереднення. Формули розрахунків названих параметрів з використанням многочленів різних степенів відомі*. З використанням *многочлена другого степеня* за певної кількості точок на проміжку усереднення ($m = 2p + 1$) формули розрахунків зважених плинних середніх мають вигляд:

* Четыркин Е. М. Статистические методы прогнозирования. – М.: Статистика, 1975.

за $m = 5$, $\bar{y}_k = \frac{1}{35}(-3y_{k-2} + 12y_{k-1} + 17y_k + 12y_{k+1} - 3y_{k+2})$;

за $m = 7$, $\bar{y}_k = \frac{1}{21}(-2y_{k-3} + 3y_{k-2} + 6y_{k-1} + 7y_k + 6y_{k+1} + 3y_{k+2} - 2y_{k+3})$;

за $m = 9$, $\bar{y}_k = \frac{1}{231}(-21y_{k-4} + 14y_{k-3} + 39y_{k-2} + 54y_{k-1} + 59y_k + 39y_{k+1} + 54y_{k+2} + 14y_{k+3} - 21y_{k+4})$,

$k = (p+1) \div (n-p)$.

Використання зважених плинних середніх до певної міри дає можливість врахувати наявну змінну кривизни тренда.

Як свідчать наведені формули, вагові коефіцієнти, на які домножуються рівні ряду, симетричні відносно центральної точки y_k, і їх сума дорівнює одиниці (з урахуванням множника, розташованого за дужками), наприклад, при $m = 7$ маємо: $\frac{1}{21}(-2 + 3 + 6 + 7 + 6 + 3 - 2) = 1$. Оскільки зважена плинна середня дорівнює середній точці кожного вибору послідовних $2p+1$ точок, то для p початкових та p завершальних рівнів вона не може бути обчислена.

Використання плинних середніх і зважених плинних середніх доцільно розглядати як один із засобів дослідження тенденції перебігу процесу в сукупності з іншими.

4.3.5. Середні прирости

Дослідження середньої швидкості середніх приростів зміни рівнів ряду дає вагому інформацію для аналітичного моделювання тренда та прогнозування. При обчисленні середніх приростів необхідно користуватися експериментальними даними рівнів ряду. Розглянемо один із методів. Припустимо, що зміну рівнів у часі для перших $m = 2p + 1$ точок можна характеризувати лінійною залежністю:

$$y(t) = a + bt, \qquad (15)$$

де значення параметрів a та b обчислені за методом найменших квадратів (МНК), стисло описаним у главі 3.

Якщо відлік часу умовно виконати від середини проміжку розташування вибраних точок, то величина a дорівнює середній арифметичній сумі m рівнів ряду. Величини a і b за умови, що відлік виконуємо від центральної точки на кожному проміжку усереднення, обчислимо за МНК, позначивши експериментальні рівні ряду верхнім індексом "e".

Сума квадратів відхилень Q значень $y(t)$, обчислених за (15), від фактичних рівнів ряду дорівнює:

$$Q(a,b) = \sum_{i=-p}^{p}(a + bi\Delta t - y_{k+i}^{e})^2 .$$

При вибраному розташуванні центру k, величині проміжку усереднення $2p\Delta t$ і відомих величинах рівнів ряду y^e і Δt функція $Q(a, b)$ буде функцією лише двох змінних – a і b, а оскільки вона може мати лише невід'ємні значення, то її найменша величина буде в точці (a, b), в якій її частинні похідні дорівнюють нулю. Запишемо їх та прирівняємо нулю:

$$\frac{\partial Q(a,b)}{\partial a} = 2\sum_{i=-p}^{p}(a + bi\Delta t - y_{k+i}^{e}) = 2\left[(2p+1)a - \sum_{i=-p}^{p} y_{k+i}^{e}\right] = 0;$$

$$\frac{\partial Q(a,b)}{\partial b} = 2\sum_{i=-p}^{p}(a + bi\Delta t - y_{k+i}^{e})i\Delta t = 2\Delta t\left[a\sum_{i=-p}^{p} i + b\Delta t\sum_{i=-p}^{p} i^2 - \sum_{i=-p}^{p} iy_{k+i}^{e}\right] = 0.$$

Розв'язавши записані рівняння відносно a і b, одержуємо:

$$b = \frac{1}{\Delta t\sum_{i=-p}^{p} i^2}\sum_{i=-p}^{p} iy_{k+i}^{e}, \qquad a = \frac{1}{2p+1}\sum_{i=-p}^{p} y_{k+i}^{e} .$$

Скориставшися знайденими величинами a, b і залежністю (15), обчислимо $\bar{d}_k$ – середні прирости на проміжку усереднення. Наприклад, якщо $p = 2$, тобто проміжок усереднення $m = 5$ точок, одержимо, що величина середнього приросту $\bar{d}_k$ (5) дорівнює:

$$\bar{d}_k(5) = \frac{1}{4}((y_{-1} - y_{-2}) + (y_0 - y_{-1}) + (y_1 - y_0) + (y_2 - y_1)) = \frac{1}{4}(y_2 - y_{-2}).$$

Але, згідно з (15), за відомого значення b маємо:

$$\bar{d}_k(5) = \frac{1}{10}(2y_{k+2}^{e} + y_{k+1}^{e} - y_{k-1}^{e} - 2y_{k-2}^{e}).$$

За умови, що проміжок усереднення має сім точок ($p = 3$), величина середнього приросту $\bar{d}_k(7)$ визначається за формулою:

$$\bar{d}_k(7) = \frac{1}{28}(3y_{k+3}^{e} + 2y_{k+2}^{e} + y_{k+1}^{e} - y_{k-1}^{e} - 2y_{k-2}^{e} - 3y_{k-3}^{e}).$$

У загальному випадку, обчисливши приріст на першому проміжку усереднення за $2p+1$ точок, виконуємо розрахунки середніх приростів для решти членів ряду. Наведемо кілька використовуваних на практиці формул:

$$p = 1;\ m = 3;\quad \bar{d}_k(3) = \bar{d}_{k-1}(3) + \frac{1}{2}(2y_{k+1}^{e} + y_{k-2}^{e} - 3y_k^{e}),$$

$$p = 2;\ m = 5;\quad \bar{d}_k(5) = \bar{d}_{k-1}(5) + \frac{1}{10}(3y_{k+2}^{e} + 2y_{k-3}^{e} - 5y_k^{e}),$$

$$p = 3;\ m = 7;\quad \bar{d}_k(7) = \bar{d}_{k-1}(7) + \frac{1}{28}(4y_{k+3}^{e} + 3y_{k-4}^{e} - 7y_k^{e}).$$

Наявність величин середніх приростів та їх графічне зображення дають змістовну інформацію про вибір аналітичної залежності тренда.

4.3.6. Темпи зростання та середні темпи приросту

Ланцюговий темп зростання в певний момент часу визначається відношенням наступного рівня до спостережуваного і вимірюється або часткою одиниці, або відсотками. Якщо y_1, y_2, y_3, …, y_n є рівні часового ряду, то ланцюгові темпи зростання відповідно дорівнюватимуть:

$$\tau_2 = \frac{y_2}{y_1};\ \tau_3 = \frac{y_3}{y_2};\tau_4 = \frac{y_4}{y_3};\ldots;\tau_{n-1} = \frac{y_{n-1}}{y_{n-2}};\tau_n = \frac{y_n}{y_{n-1}} . \qquad (16)$$

Ланцюгові темпи приростів $\tau_{пр.\,i}$ визначаються залежністю:

$$\tau_{пр.\,i+1} = \frac{y_{i+1} - y_i}{y_i} = \tau_{i+1} - 1. \qquad (17)$$

Середній темп зростання за відомого проміжку визначається залежністю:

$$\bar{\tau} = \sqrt[n-1]{\tau_2 \cdot \tau_3 \cdots \tau_n} = \sqrt[n-1]{y_n / y_1} . \qquad (18)$$

Середній темп зростання використовується при співставленні показників динаміки за різні проміжки для певного часового ряду або при порівнянні темпів зростання кількох часових рядів.

Описані засоби обробки інформації, наявної в часовому ряді, дають можливість зробити перший цілеспрямований крок у вивченні тенденції перебігу досліджуваного процесу.

Приклад. Розглянемо результати обробки статистичних даних про обсяги реалізації (в абсолютних і відносних вимірах) певних товарів, умовно названих A і B, по м. Харкову (табл. 4.3).

Таблиця 4.3.

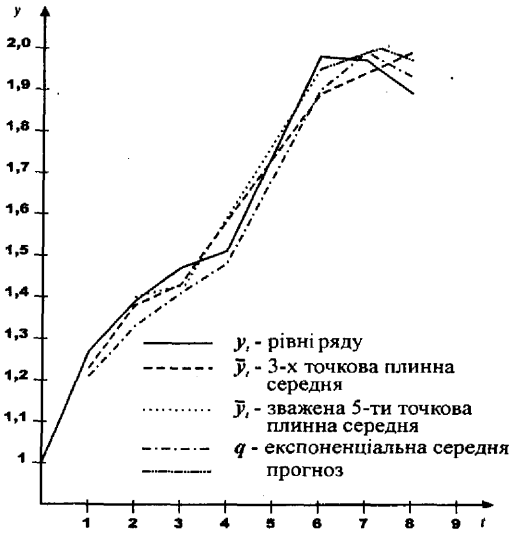
Рік	1975	1976	1977	1978	1979	1980	1981	1982	1983
t	0	1	2	3	4	5	6	7	8
Обсяги реалізації товару A									
y_t	12311	14609	17101	18213	18846	22074	24470	24874	23795
y_t	1	1,268	1,389	1,479	1,531	1,793	1,988	2,02	1,927
τ_t	–	1,268	1,096	1,065	1,035	1,171	1,109	1,016	0,95
$\tau_{пр.t}$	–	0,268	0,096	0,065	0,035	0,171	0,109	0,016	0,05
Обсяги реалізації товару B									
y_t	2057	2366	3060	3168	4007	3374	2606	2606	2377
y_t	1	1,15	1,488	1,54	1,948	11,64	11,267	1,267	1,156
τ_t		1,15	1,294	1,035	1,265	0,842	0,773		0,912
$\tau_{пр.t}$		0,15	0,294	0,035	0,265	-0,158	-0,227	0	0,088

Деякі результати статистичної обробки наведених у табл. 4.3 даних, які характеризують відносні величини обсягів реалізації товарів A і B, представлені на рис. 4.2 (відповідно позиції а і б).

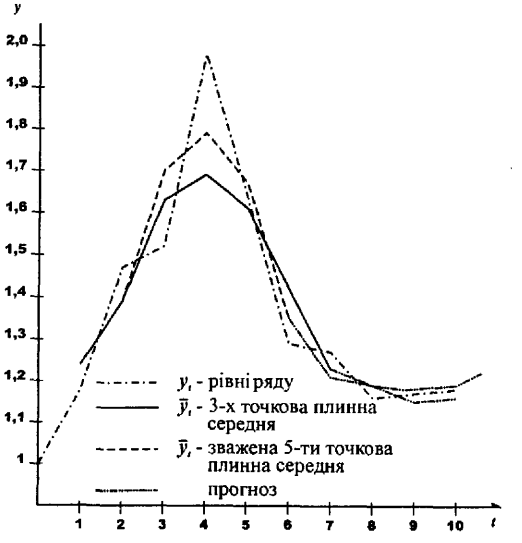
4.4. Короткотермінове прогнозування

Методи короткотермінового прогнозування з використанням часових рядів доцільно застосовувати в таких випадках: 1) при наявності даних за проміжок часу, який не перевищує одного року (за день, тиждень, місяць, квартал); 2) при виконанні прогнозу по конкретному показнику окремо для кожного наступного моменту часу.

До розбудови математичних методів прогнозування необхідно чітко уявляти собі таке. Оскільки будь-який прогноз ґрунтується на інформації про минулий стан, то він завжди буде мати похибки, а тому моделі прогнозування часто будуються з використанням ідеї мінімізації такої похибки. Прогноз дає можливість обчислити



а)



б)

Рис. 4.2.

величину досліджуваного показника лише орієнтовно, з відхиленнями від його дійсного значення, яке буде досягнуте в майбутньому. Як правило, приймають, що такі очікувані відхилення підлягають нормальному закону розподілу. Основою для такої гіпотези є те, що відхилення від природних властивих досліджуваному процесу закономірностей обумовлені дією факторів різної значимості та результативності. Приймаючи гіпотезу про нормальний закон розподілу похибок прогнозу, за міру їх відхилення від прогнозованого значення приймають середнє квадратичне відхилення, або дисперсію.

Таким чином, *зробити прогноз означає, що необхідно обчислити ймовірне очікуване значення показника на майбутній момент можливе середнє квадратичне відхилення, або дисперсію.*

4.4.1. База та горизонт прогнозу

При прогнозуванні використовують усереднені значення досліджуваного показника на деякому проміжку (день, тиждень тощо). Природно чекати, що, збільшивши кількість спостережень, тобто обсяг вибірки, можна зменшити вплив окремих випадкових факторів і як наслідок одержати більш точний прогноз. Але збільшення кількості спостережень зменшує швидкість реагування моделей прогнозу на фактичні коливання значень досліджуваного показника. Компроміс між цими двома вимогами досягається належною вибіркою кількості проміжків спостереження. Використовувана при прогнозуванні кількість проміжків спостереження називається *базою прогнозу,* кількість проміжків, на які робиться прогноз, – *горизонтом прогнозу.*

При наявності тенденції зміни досліджуваного показника в часі (методи її перевірки були описані раніше) необхідно будувати тренд. Якщо тенденція зміни не спостерігається, то такий часовий ряд називається *стаціонарним.* Типову ситуацію, за якої часовий ряд буде стаціонарним, графічно зображено на рис. 4.3.

Значення показника для моментів спостереження коливаються вверх і вниз відносно один одного під впливом випадкових факторів, але середнє значення за місяць на протязі року близьке до 93 одиниць. Як буде показано далі, використання критерію Фостера-Стьюдента свідчить про *відсутність тенденції зміни середньої* наведеного часового ряду, тобто це стаціонарний ряд.

На практиці стаціонарні ряди трапляються рідко, але поняття та методи їх дослідження широко використовуються за інших ситуацій, тому стаціонарним рядам буде приділена певна увага.

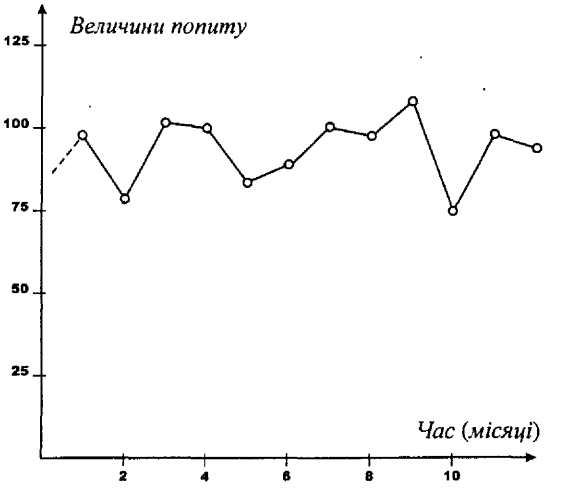


Рис. 4.3.

4.4.2. Прогнозування стаціонарних показників

Стаціонарним називається такий показник, для якого зміни в часі представлені стаціонарним рядом (рис. 4.3).

Для спрощення розрахунків відлік часу зручно упорядкувати так: за початок підрахунку взяти сьогочасний момент, а всі інші точки зіставляти з обраною. Тобто якщо точку відліку (сьогочасний момент) позначити t, то минулі проміжки (моменти) заміру показника будуть $t-1, t-2, \ldots$, а наступні $t+1, t+2, \ldots$ і т. д.

"*Примітивний*" прогноз визначається так: величина досліджуваного показника f_{t+1} в наступний проміжок дорівнює наявному значенню d_t:

$$f_{t+1} = d_t.$$

Замінивши правило відліку часу, варто замінити й позначення відповідних величин.

Таким "примітивним" прогнозом доцільно користуватися лише при проведенні певних досліджень.

Більш ефективним прогнозованим значенням показника є так звана *адаптивна*

плинна середня $m_t = \frac{1}{n}\sum_{i=t}^{t-n+1} d_i$ $(i = t;\ t-1,\ldots,\ t-n+1)$, яка є результатом усереднення його значень за n минулих проміжків і *співставляється правому кінцю* проміжка (t–1; t). За такої умови маємо:

$$f_{t+1} = m_t = m_{t-1} + \frac{1}{n}(d_t - d_{t-n}). \qquad (19)$$

Формула (19) показує, що сьогочасне значення адаптивної плинної середньої m_t відрізняється від її минулого значення m_{t-1} на $1/n$ величини різниці сьогочасного значення d_t показника та значення d_{t-n} на момент, зміщений на n проміжків в минуле. При збільшенні кількості проміжків усереднення n величина поправки $(d_t - d_{t-n})/n$ за будь-яких інших умов буде зменшуватися.

Перший доданок у формулі (19) характеризує "інерцію перебігу", а другий – вплив сьогочасного значення показника. Такий вплив визначається коефіцієнтом $1/n$.

При наявності стаціонарного ряду за прогнозовану величину f_{t+1} показника беруть значення m_t, обчислене згідно з (19) не лише на момент $t + 1$, але, при необхідності, на всі наступні моменти.

Наголосимо на особливостях визначення величини прогнозу при використанні адаптованої плинної середньої згідно з (19).

1. Величину прогнозу не можна обчислити раніше, ніж буде одержано n величин рівня ряду (кількість n вибираємо довільно, орієнтуючись на певні попередні дослідження змін контрольованого показника).

2. Рівням ряду, які відповідають вибраному проміжку усереднення, надається ваговий коефіцієнт $1/n$, а решті рівнів – нульовий. При такому підході більш близьким за часом рівням надається такий же ваговий коефіцієнт, як і більш віддаленим, хоч природно очікувати їх різну значимість для якості прогнозу.

3. Відгук значення плинної середньої на величину рівня ряду (чутливість) визначається величиною $1/n$, обернено пропорційною кількості точок на проміжку усереднення. З такої причини без зміни n чутливість виміряти неможливо.

Розглянемо середню, яка більш ефективно дозволяє зробити прогноз.

4.4.2.1. Експоненціальна зважена середня

Для більш точного обчислення величини прогнозу доцільно використовувати так звані *зважені плинні середні*, за яких певним чином враховуються часові відстані рівнів ряду від моменту прогнозування: більш віддалені мають менший вплив на

величину прогнозу*. Розглянемо розбудову так званої експоненціальної зваженої середньої. Якщо розглянути ряд $\alpha+\alpha(1-\alpha)+\alpha(1-\alpha)^2+\ldots+\alpha(1-\alpha)^n+\ldots$, то за умови $0<\alpha<1$ при $n \to \infty$ сума ряду збігається до 1, бо маємо нескінченну геометричну прогресію зі знаменником $(1-\alpha)$ меншим за 1. З використанням членів наведеного нескінченного ряду експоненціальну зважену середню u_t визначають залежністю:

$$u_i = \alpha d_t + \alpha(1-\alpha)d_{t-1} + \alpha(\alpha-1)^2 d_{t-2} + \ldots + \alpha(\alpha-1)^n d_{t-n} + \ldots$$

Якщо в послідовності доданків, починаючи з другою, винести за дужки $(1-\alpha)$, то одержимо:

$$u_i = \alpha d_t + (1-\alpha)u_{t-1}. \qquad (20)$$

З використанням експоненціальної зваженої середньої прогноз f_{t+1} на наступний проміжок визначається за формулою:

$$f_{t+1} = u_t = u_{t-1} + \alpha(d_t - u_{t-1}), \qquad (21)$$

де різниця визначає похибку прогнозу, виконаного пропонованим методом, якщо момент прогнозу співставлено точці $t-1$.

Автор методу прогнозування з використанням експоненціальної зваженої англійський вчений Г. Г. Браун радить величину α визначати залежністю: $\alpha = 2/(n+1)$ (n – кількість рівнів ряду, використаних для усереднення). У літературі описано й інші підходи до визначення величини α. Використання експоненціальної зваженої дає певні переваги над плинною середньою:

1) для прогнозування необхідно задати лише його вихідну оцінку u_{t-1}, подальше прогнозування можливе відразу по одержанні наступних даних;
2) вагові коефіцієнти спадають у міру віддалення від моменту прогнозу, тому немає точок з нульовими ваговими коефіцієнтами;
3) для визначення наступного значення прогнозу необхідні лише дві величини: попереднє значення прогнозу $f_t = u_{t-1}$ та досягнуте значення досліджуваного показника d_t;
4) чутливість експоненціальної зваженої середньої може бути змінена шляхом зміни величини α; чим більше значення α, тим більша чутливість.

На практиці рекомендується брати α з інтервалу (0,05; 0,3). Якщо кращі результати одержують за більших значень α, то це свідчить про те, що відповідний часовий ряд не є стаціонарним і необхідно замінити метод прогнозування.

* Льюис К. Д. Методы прогнозирования экономических показателей. – М.: Финансы и статистика, 1986.

4.4.3. Прогнозування нестаціонарних показників

Описані раніше способи дослідження тенденції зміни при її наявності дають можливість характеризувати її у вигляді тренда середніх. Тренди з метою їх використання для прогнозування доцільно класифікувати за характером і типом. Зміна середньої в часі визначає характер тренда, а закономірність відхилення значень досліджуваного показника від середньої величини визначає його тип.

4.4.3.1. Характери трендів

Лінійний тренд. Якщо закономірність зміни середньої в часі має лінійну залежність, то тренд називається лінійним.

Сезонний тренд. Тренд називається сезонним, якщо в зміні середньої наявні певні часові цикли. Наприклад, обсяги споживання пального та запчастин сільськогосподарськими підприємствами мають сезонні коливання протилежного характеру.

Мішаний сезонно-лінійний тренд. Такий тренд являє собою об'єднання раніше вказаних трендів. Окремі економічні показники сільськогосподарського виробництва, яке з кожним роком розширює свої обсяги, будуть мати тренди такого характеру. Динаміку зміни показника при сезонно-лінійному тренді показано на рис. 4.4.

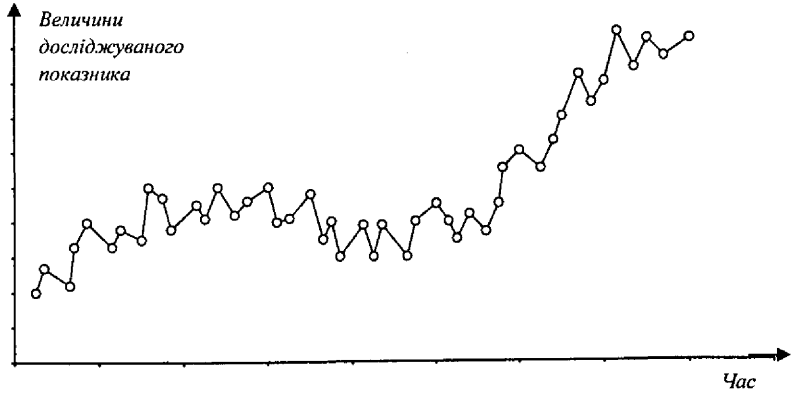


Рис. 4.4.

4.4.3.2. Типи трендів

Адитивний тренд. При адитивному тренді наявні величини досліджуваного показника відхиляються від середніх в той чи інший бік приблизно на однакову величину.

Тренд відношень або мультиплікативний. При такому тренді відношення кожних двох сусідніх відхилень від наявних значень показника стосовно середньої, яка визначає тренд, наближено рівні. Наприклад, попит на деякий товар протягом року характеризується лінійним трендом, але зростає щомісячно на 2%.

Як приклад розглянемо лінійний адитивний тренд. Показник перебігу процесу при такому тренді має середні значення, які змінюються приблизно на однакову величину з кожним наступним проміжком заміру (рис. 4.5).

Рис. 4.5.

4.4.4. Лінійно-адитивна модель прогнозування

Лінійно-адитивна модель прогнозування ґрунтується на припущенні, що середня прогнозованого показника d_t на проміжку $[t-1;\ t]$ є лінійною функцією часу:

$$d_t = \mu + \alpha t + \varepsilon(t) \qquad (22)$$

(μ – середня перебігу процесу, α – швидкість його зміни в часі, $\varepsilon(t)$ – випадкова величина з нульовим математичним сподіванням).

Для обчислення параметрів (22) пропонуються різні методи. Розглянемо деякі з них.

Метод Холта. Згідно з методом Холта* параметр α оцінюється коефіцієнтом b_t ($\alpha = b_t$) який обчислюється як експоненціально зважена середня різниці між сьогочасними експоненціально зваженими середніми u_t процесу та їх попередніми значеннями u_{t-1}. При обчисленні сьогочасного значення експоненціально зваженої середньої u_t використовується раніше обчислене значення показника зростання b_{t-1} і в такий спосіб досягається адаптація до попереднього значення тренда.

Розрахункові формули методу Холта мають вигляд:

$$u_t = Ad_t + (1 - A)(u_{t-1} + b_{t-1}), \tag{23}$$

$$b_t = B(u_t - u_{t-1}) + (1 - B)b_{t-1}. \tag{24}$$

Параметри A, B розташовані в межах від 0 до 1. За відомого значення b_t прогноз $f_{t+\tau}$ на момент $t+\tau$ обчислюється за формулою:

$$f_{t+\tau} = u_t + b_t\tau. \tag{25}$$

Величини A, B рекомендується вибирати рівними відповідно 0,1 і 0,01. Умови, за яких необхідно переглядати значення величин A, B, описано далі при обчисленні точності прогнозу. Використання формул (23) та (24) можна розпочати при наявності не менше трьох замірів досліджуваного показника, відповідно $b_{t-1} = d_{t-1} - d_{t-2}, u_{t-1} = d_{t-1}$.

Метод адаптивного згладжування Брауна ґрунтується на припущенні, що менш віддалені в часі спостереження мають більший вплив на величину прогнозу, їх значимість корегується за допомогою так званого коефіцієнта дисконтування γ. Прогноз $f_{t+\tau}$ визначається з використанням залежності:

$$f_{t+\tau} = a_0 + a_1\tau + a_2\tau^2 + \varepsilon(t)$$

$\varepsilon(t)$ – величина випадкового відхилення з математичним сподіванням, рівним нулю. Вибравши *за бажанням* величину коефіцієнта дисконтування γ, коефіцієнти a_0, a_1, a_2 вибирають так, щоб сума зважених квадратів відхилень даних спостережень d_{k-1} в

* Льюис К. Д. Методы прогнозирования экономических показателей. – М.: Финансы и статистика, 1986.

минулому та величин, обчислюваних за пропонованою формулою, досягала найменшого значення:

$$\sum_{i=0}^{n} \gamma^i (d_{t-i} - f_{t-i})^2 \to \min.$$

При наявності лінійного тренда оцінка прогнозу за методом зважених найменших квадратів обчислюється так:

$$f_{t+\tau} = u_t + b_t\tau$$

($u_t = u_{t-1} + b_{t-1} + (1 - \gamma)^2 e_t$; $e_t = d_t - f_t$; $b_t = b_{t-1}(1 - \gamma)^2 e_t$.)

Рекомендується вибирати $\gamma = 0{,}8$*.

4.4.5. Визначення точності короткотермінового прогнозу

Наведемо деякі оцінки якості прогнозу. Як правило, для оцінки якості вибраної моделі прогнозу використовують дані *за вже минулі моменти спостереження показника,* тобто якість вибраної моделі прогнозування перевіряють за вже відомими даними. Точність моделі прогнозу оцінюють величиною середнього квадратичного відхилення σ даних спостереження від величин, обчислених згідно з моделлю для відповідних моментів часу в минулому. Якщо дані спостереження позначимо d_i, а прогнозовані за моделлю прогнозу – f_i для тих же моментів часу, то середнє квадратичне відхилення σ_t обчислюватиметься за формулою:

$$\sigma_t = \sqrt{\frac{1}{n}\sum (d_t - f_t)^2}, \tag{26}$$

де додавання виконується по всіх проміжках, використаних для оцінки прогнозу. Такий підхід дозволяє перевірити вибрану модель прогнозу за даними відомого минулого.

Іноді за міру точності прогнозу доцільно використати так зване середнє абсолютне відхилення MAD_t, яке обчислюють як експоненціально зважену середню абсолютних похибок прогнозу $e_t = |d_t - f_t|$:

* Льюис К. Д. Методы прогнозирования экономических показателей. – М.: Финансы и статистика, 1986.

$$MAD_t = \alpha e_t + (1 - \alpha) MAD_{t-1}. \quad (27)$$

Доведено, що для значного класу статистичних розподілів величина середнього квадратичного відхилення σ дещо більша середнього абсолютного відхилення MAD і пропорційна йому. Стала пропорційності для різних розподілів змінюється в межах від 1,2 до 1,3. При нормальному розподілі її значення дорівнює $\sqrt{\pi/2} = 1{,}2533\ldots$

Часто для оцінки точності прогнозу використовують так звану середню похибку прогнозу $MAPE$, яку обчислюють за формулою:

$$MAPE = \frac{1}{n} \sum_{i=t}^{t-n+1} e_i, \quad (28)$$

та середню відсоткову похибку прогнозу $MAPE_p$:

$$MAPE_p = \frac{1}{n} \sum_{i=t}^{t-n+1} \frac{e_i}{d_i} 100. \quad (29)$$

Якщо $MAPE_p$ перевищує 5%, доцільно переглянути модель прогнозування*.

При порівнянні точності прогнозу у процесах різної природи використовують $MAPE_p$, оскільки така оцінка характеризує відносну точність прогнозу.

При виборі оптимальної моделі прогнозування серед кількох пропонованих використовують оцінку середнього квадрата похибки Δ_s:

$$\Delta_s = \frac{1}{n} \sum_{i=t}^{t-n+1} e_i^2. \quad (30)$$

Доцільність використання певної оцінки похибок прогнозу залежить від конкретної задачі.

4.4.6. Контроль при короткотерміновому прогнозуванні

Основною перешкодою розбудови якісних прогнозів є раптові виплески (викиди) величини досліджуваного показника. Тому за будь-якої моделі прогнозування має бути деякий механізм контролю, який дав би можливість оцінити, як спрогнозована раніше величина відхиляється від фактичної по

* Льюис К. Д. Методы прогнозирования экономических показателей. – М.: Финансы и статистика, 1986.

досягненні моменту прогнозу, та визначити, коли модель прогнозування стає неефективною. Основна мета контролю короткотермінового прогнозування – виявити випадкові виплески в значеннях рівнів показника та з'ясувати причини їх походження.

Після виявлення втрати бажаної точності прогнозу необхідно розв'язати ряд питань: яка причина виплеску, яка ймовірність появи у майбутньому виплесків; якщо велика ймовірність їх появи, то оцінити, по можливості, їх періодичність. Відповіді на ці питання знаходяться поза межами моделі прогнозу, і для їх пошуку необхідні експертний аналіз та оцінки.

До використання певних методів прогнозування необхідно з'ясувати, як треба розуміти й оцінювати виплески значень показника. Статистична значимість появи деякої події залежить, як відомо, від статистичного розподілу появи досліджуваної випадкової події. Практично побудувати відповідний розподіл не завжди можливо та легко. Тому обмежимося тлумаченням результатів відповідного аналізу використанням поняття статистичної значущості. При прогнозуванні й управлінні рівень значущості вибирається, як правило, в межах [0,05; 0,1].

Наведемо один із засобів контролю*. Пропонується так званий "контрольний сигнал спостереження" T_t, який визначається як відношення експоненціально зваженої похибки (а не її модуля) $\bar{e}_t = \alpha e_t + (1-\alpha)\bar{e}_{t-1}$ до середнього квадратичного відхилення σ_t (26):

$$T_t = \bar{e}_t / \sigma_t. \quad (31)$$

Так визначена величина T_t за деякого рівня статистичної довіри показує на міру неузгодженості множини наявних даних і прогнозу, зокрема на низьку якість прогнозу. Межами можливих значень контрольного сигналу T_t є (–1; 1). При відповідності вибраному рівню довіри значення T_t мають знаходитися в певних межах (табл. 4.4). Якщо величина T_t за модулем перевищує наведені в табл. 4.4 межі, необхідно уточнити або замінити модель прогнозування.

Таблиця 4.4.

Рівень довіри, %	Значення контрольного сигналу T_t	
	$\alpha = 0{,}1$	$\alpha = 0{,}2$
80	0,36	0,54
90	0,45	0,66
95	0,51	0,74
98	0,60	0,81
100	1,00	1,00

При обчисленні контрольного сигналу необхідно користуватися тією ж величиною α, що і для експоненціальної зваженої похибки $\bar{e}_t$ та середнього абсолютного відхилення (27). Дані табл. 4.4 свідчать: якщо, наприклад, при $\alpha = 0{,}1$

* Льюис К. Д. Методы прогнозирования экономических показателей. – М.: Финансы и статистика, 1986.

величина T_t перевищила значення 0,51, то це означає, що з 95% рівнем довіри використовувана модель прогнозу перестала бути адекватною наявним даним.

Отже, використання контрольного сигналу дозволяє ефективно стежити за відповідністю математичної моделі перебігу досліджуваного процесу. Розглянемо методи, які дозволяють при необхідності корегувати модель прогнозу.

4.4.7. Поняття про адаптивне прогнозування

Методи, які забезпечують при необхідності можливість перебудови моделі прогнозу у відповідності до фактичних поточних даних, називаються *адаптивними*. Основна властивість таких методів: з одержанням нових даних значення прогнозу, якщо воно стосувалося горизонту з більш ніж одним проміжком спостереження, змінюється у відповідності до одержаної інформації. Вимога високої якості прогнозування з необхідністю вимагає забезпечення чутливості прогнозу до зміни базових даних. Більш чутлива модель прогнозування є гарантією поліпшеної його якості. У методах прогнозування, які використовують експоненціальну зважену, більшій чутливості відповідає більше значення α і навпаки ($0<\alpha<1$). У всіх моделей прогнозування завжди певним чином використовується поточна інформація. У момент заміру, який настає за моментом виплеску даних, буде якось відображена величина виплеску. У низькочутливої моделі прогнозу величини показника за дуже короткий проміжок досягнуть попереднього середнього рівня, а для високочутливої моделі для стабілізації потрібний значно більший проміжок. Наведемо деякі адаптивні методи.

Метод Чоу. За пропозицією Чоу* метод адаптивного прогнозування будується з використанням *трьох одночасно виконуваних прогнозів* за схемою експоненціально зваженої середньої при різних значеннях величини α. Фактичний прогноз виконується з використанням експоненціально зваженої середньої за "основного" значення α. Окрім такого прогнозу, виконуються прогнози за значенням α на 0,05 більше та менше основного.

Величина α, довільно вибрана на початку прогнозування, з часом автоматично змінюється за таким правилом: якщо за критерієм оцінки якості прогнозу один із "крайніх" прогнозів стає кращим "середнього", то для наступних моментів за "основний" прогноз вибирають кращий із "крайніх", тобто відповідно замінюють величину α, потім будують наступні "крайні" прогнози при зміненому "основному" значенні α.

Метод Тригга і Ліча. Згідно з цим методом для адаптації швидкості реакції моделі прогнозу до виплесків величин показника використовують описаний раніше

* Льюис К. Д. Методы прогнозирования экономических показателей. – М.: Финансы и статистика, 1986.

контрольний сигнал T_t (31)*. За величину α експоненціально зваженої середньої вибирають абсолютне значення T_t. Такий вибір α відповідає вимогам адаптації до виплесків значень досліджуваного показника: вона збільшується при збільшенні величин коливань в даних спостереження і зменшується при їх стабілізації. При користуванні названим методом величина прогнозу $f_{t+\tau}$ обчислюється за формулою:

$$f_{t+\tau} = |T_t| d_t + (1 - |T_t|) u_{t-1},$$

а величини u_{t-1}, T_t – за формулами (20) та (31).

Якщо при значних різких змінах величини досліджуваного показника наявна стала α не забезпечує потрібної точності прогнозу, величина контрольного сигналу T_t зростає, що обумовлює зростання вагових коефіцієнтів для найближчих спостережень, – то прогноз виконується при новому рівні середньої. З перебудовою моделі прогнозування наступне значення T_t "автоматично" зменшується і модель стає менш чутливою до коливань в даних. Наведений метод значно швидше реагує на коливання фактичних величин показника, ніж метод експоненціального вирівнювання (рис. 4.6.), і використовується при наявності сезонних коливань.

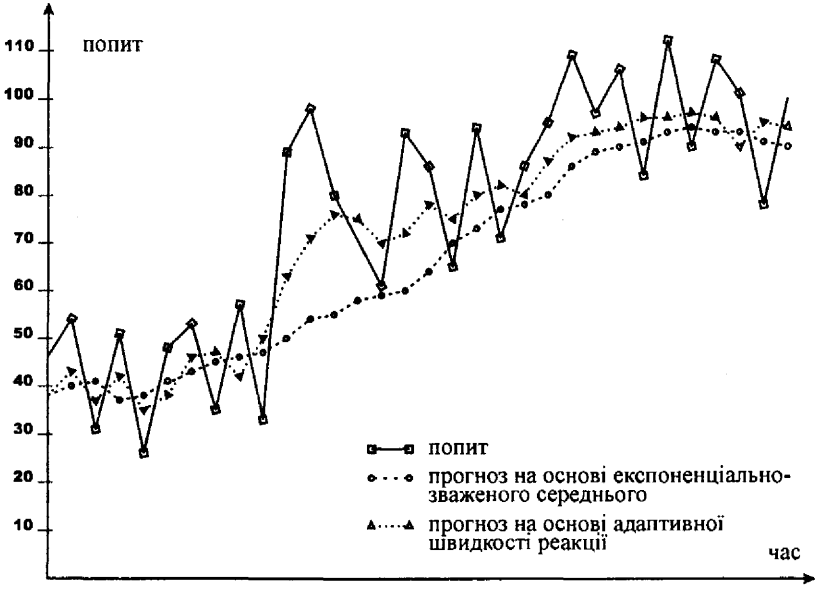


Рис. 4.6.

На завершення доцільно наголосити, що методи короткотермінового прогнозування використовуються при оперативному плануванні та управлінні, ефективність яких може бути досягнута лише з відповідним програмним забезпеченням.

4.5. Методи середньо- та довготермінового прогнозування

4.5.1. Вихідні положення

Аналітичні методи, якими користуються при середньо- та довготерміновому прогнозуванні, ідентичні, тому розглянемо їх разом і будемо користуватися терміном "середньотермінове програмування".

Середньо- та довготермінове прогнозування використовується, як правило, за таких обставин: 1) необхідно мати одноразовий прогноз, який в більшості ситуацій не треба корегувати при появі нових даних; 2) наявні щорічні дані за минулі роки; 3) прогноз треба виконати при відносно малій кількості вихідних даних; 4) прогнозується динаміка досить складного процесу, наприклад, ефективність інвестування або перспективи розвитку певної галузі.

До моделей середньотермінового прогнозування з їх високою економічною значимістю ставлять більш високі вимоги, ніж до моделей короткотермінового прогнозування, тому такі моделі більш складні, при їх розбудові широко використовуються лінійна та нелінійна регресії, велика увага приділяється оцінкам точності прогнозу.

Моделі прогнозування, розбудовані з використанням регресій, мають певні характерні особливості: 1) ряд вихідних даних мають забезпечувати певну точність прогнозування, а для цього необхідно мати відповідну кількість даних; 2) адаптація не можлива: з появою нових даних і при необхідності уточнення прогнозу виконуються нові розрахунки; 3) методи не дозволяють виконувати сезонне прогнозування; 4) у зв'язку з високою економічною та виробничою значимістю прогнозування модель має бути завжди перевірена на її відповідність прогнозованому процесу стосовно досліджуваного показника.

Можливість одержати прогноз розвитку досліджуваних соціально-економічних і виробничих процесів ґрунтується на гіпотезі, що визначальні фактори та тенденції минулого мають певну інерційність, тому вивчення та оцінка їх впливу в минулому можуть бути перенесені в майбутнє. Інерційність названих процесів проявляється двояко: по-перше, як збереження в основних проявах взаємозв'язків досліджуваного процесу з іншими, тобто збереження залежності значень прогнозованого показника від факторів, урахованих при розбудові математичної моделі; по-друге, як збереження в часі загальної тенденції перебігу досліджуваного

процесу – темпу, якості приростів, стабільності стосовно впливу множини випадкових факторів і т. ін.

Гіпотеза інерційності зв'язків між факторами, визначальними для перебігу процесу, використовується при розбудові їх аналітичного представлення, пошук якого в загальних рисах можна описати так. Значення величин досліджуваного фактора y (залежної змінної – функції) згідно з прийнятою математичною моделлю аналітично пов'язують з величинами факторів, визначальними для перебігу процесу (x_1, x_2, …, x_m), які називаються в математиці незалежними змінними – аргументами. Оскільки ми домовилися абстрагуватися від розгляду природи причинно-наслідкових зв'язків і враховувати лише числові дані стосовно як визначальних, так і досліджуваного показників, то загальна форма залежності досліджуваного показника y від часу t та факторів впливу, опосередкованих часом, буде мати вигляд:

$$y = f(y_i, t, a_j), \tag{32}$$

(y_i – рівні ряду, використані як база прогнозу при розбудові відповідної моделі; a_j – параметри тренду факторів впливу). Прикладами таких моделей можуть бути виробничі функції за умови їх залежності від кількох факторів, опосередкованих часом.

Наголосимо на зручності використання інших позначень, ніж при дослідженні короткотермінового прогнозування.

Розбудувати однозначний універсальний алгоритм визначення взаємозв'язків між базою, горизонтом і точністю прогнозу неможливо, тому слід керуватися правилом з досвіду: термін упередження не може перевищувати третини проміжку бази прогнозу.

Для визначення аналітичної залежності (32) розглянемо використання рівняння регресії досліджуваного показника в часі. Як відомо, регресійний аналіз ґрунтується на розв'язанні двох проблем: вибору незалежних змінних, які визначають характер зміни залежної величини, та представлення форм аналітичної залежності між ними (етап так званої специфікації), кількісних оцінок вибраних параметрів. За пояснюючий фактор (незалежну змінну) вибираємо час, а задача кількісних оцінок параметрів завжди розв'язується з використанням відповідного статистичного методу обробки даних спостереження. Методику розбудови рівнянь парних лінійної та нелінійної регресій розглянуто в главі 3. Наголосимо на деяких особливостях використання рівняння парної регресії при дослідженні та прогнозуванні рівнів часових рядів.

Точність прогнозу визначається не лише точністю рівняння регресії, але й тим, наскільки надійно та точно будуть передбачені *умови* перебігу процесу в майбутньому. Прогноз завжди має точковий характер, тому дійсні значення

досліджуваного показника по досягненні моменту, стосовно якого було зроблено прогноз, будуть відхилятися від прогнозованого. Надійність прогнозу визначається розбудовою так званих інтервалів довіри, в межах яких очікується прогнозована величина.

При розбудові рівняння регресії оцінюють перебіг досліджуваного економічного процесу *в минулому* за наявними даними спостережень. При прогнозуванні користуються екстраполяцією тенденції перебігу процесу в минулому, тобто продовженням у *майбутнє* умов і причинно-наслідкових взаємозв'язків. Екстраполяція наявних тенденцій, оцінених при розбудові рівняння регресії часових рядів, досить широко використовується на практиці завдяки її простоті в реалізації та можливостей скористатися відносно невеликим обсягом інформації в поєднанні з певною чіткістю вихідних положень.

Аналіз особливостей зміни приростів рівнів ряду до певної міри віддзеркалює природу динаміки досліджуваного показника, тому має передувати вибору форми кривої. Цей аналіз має три послідовних етапи: 1) усереднення рівнів ряду з використанням певної плинної середньої; 2) обчислення середніх приростів усередненого ряду; 3) обчислення певних характеристик приростів.

Аналітична залежність та відповідне графічне зображення використовуваних на практиці ліній регресії наведені в другому та третьому стовпцях табл. 3. 5 (глава 3), а алгебраїчні перетворення змінних, які дозволяють у перебудованих координатах мати лінійну залежність, – у другому стовпці.

Стисло опишемо основні труднощі, які виникають при виборі форми кривої регресії досліджуваного показника у часі. Немає чітко визначеного критерію вибору терміну усереднення. Чим більший проміжок усереднення, тим більше взаємогасяться можливі коливання значень рівнів ряду. На практиці широко використовуються проміжки усереднення від трьох до семи точок. Якщо ряд має коливання при чітко спостереженому проміжку циклу, то коливання повністю будуть усунені, якщо проміжок усереднення буде рівний або кратний періоду циклу. За умови, що використання певної плинної середньої не усуває значних коливань, то застосовують зважену плинну середню. Необхідно наголосити, що *операція усереднення може істотно змінити природу динаміки ряду.* Якщо до усереднення можна було з великою ймовірністю вважати незалежними між собою значення рівнів ряду, то після усереднення одержані рівні будуть в деякій мірі залежні між собою. Значимість цієї ситуації буде розглянуто далі при дослідженні автокореляції.

Як наголошувалося в главі 3, вибір аналітичної залежності в значній мірі визначається характером зміни наявних та обчислюваних приростів, з використанням ймовірної функціональної залежності.

За деяких умов рекомендовані контрольні відношення не можна обчислити (5, 6, 7, 8, ..., 12 табл. 3.5), якщо величини приростів не будуть додатними. У таких ситуаціях доцільно скористатися одним із заходів: 1) збільшити інтервал

усереднення при обчисленні плинної середньої; 2) замінити "аномальні" дані розрахунковими величинами, які являються певним усередненням даних, наприклад, середніми із рівнів, які передують точці "аномалії" та наступним точкам.

При необхідності аналізувати використання приростів ряду, який має спадаючу тенденцію, тобто його природні прирости не додатні, стає неможливим використання логарифмічних характеристик приростів. У такій ситуації рекомендується при необхідності досліджувати наявний ряд у зворотному напрямі відносно часу, скориставшись модулем приросту часу.

Вибравши формулу, доцільно провести також певний допоміжний аналіз, а саме побудувати *графік перших різниць* усереднених рівнів: 1) якщо перші різниці зменшуються зі сталим темпом, то необхідно вибрати модифіковану експоненту (ба табл. 3. 5); 2) якщо графік має одну вершину, то доцільно вибрати криву Гомперца (12 табл. 3.5); 3) якщо графіка за формою близький до графіка функції $y = e^{-(x-a)^2}$ при $x < a$, то доцільно вибрати логістичну криву (10 табл. 3.5).

Ускладнення аналітичної залежності підвищує точність представлення тренда, але зростаюча кількість параметрів збільшує інтервали довіри, тобто зменшує точність прогнозу, що необхідно враховувати, вибираючи формули аналітичної залежності.

На завершення доцільно наголосити, що вибір форми аналітичної залежності – наслідок змістовного аналізу конкретного процесу. Такий аналіз має не лише дати відомості про хід і кількісні характеристики процесу в минулому, але й оцінити його можливе майбутнє. Вибір форми залежності виконується на певній множині альтернатив, тому досить часто необхідно виконувати дослідження визначальних суттєвих закономірностей перебігу процесу, а не обмежуватися множиною даних спостереження. Але за будь-яких обставин описані статистичні засоби аналізу даних спостереження будуть корисними і до певної міри ефективними, бо дозволяють істотно обмежити множину альтернативних варіантів математичного моделювання перебігу процесу в часі та прогнозування його майбутнього стану.

Розв'язавши питання вибору форми аналітичної залежності між даними спостереження, необхідно розглянути методи обчислення величин параметрів обраної формули за методикою, описаною в главі 3.

4.5.2. Автокореляція

При розбудові теорії аналізу динаміки часових рядів було використано припущення про незалежність та нормальний закон розподілу випадкових факторів, які спричиняють відхилення рівнів ряду від величин, обумовлених природною суттю досліджуваного процесу та модельованих трендом з циклічною і сезонною

компонентами. Вибір аналітичного представлення закономірностей зміни рівнів ряду в часі ґрунтується на наслідках аналізу над усередненими (вирівняними) рівнями ряду. При виконанні операції усереднення одержані рівні ряду до певної міри стають залежними між собою. Якщо одержали помилкове рівняння регресії: невірно вибрали форму аналітичної залежності або з помилками обчислили коефіцієнти, то величини різниць між значеннями наявних рівнів ряду та аналітичної залежності для відповідних моментів часу не будуть незалежними між собою. Як відомо, рівень залежності між двома множинами величин визначається за допомогою *коефіцієнта кореляції*. Залежність між величинами рівнів вихідного або вирівняного рядів визначається *коефіцієнтом автокореляції (взаємозв'язку)*. Визначення величини рівня залежності між даними часового ряду досягається шляхом співставлення рівнів наявного ряду з рівнями цього ж ряду, але переміщеними на якесь число проміжків назад. Кількість проміжків, на яку відрізняються порівнювані послідовні рівні ряду, називається *лагом*.

Наприклад, якщо дані за 1995 р., 1996 р., 1997 р. і т. д. порівнюються з даними 1990 р., 1991 р. і т. д., то лаг дорівнює п'яти. Коефіцієнт автокореляції r_L з лагом L обчислюється за формулою:

$$r_L = \sum_{i=1}^{p}(d_t - \bar{d}_t)(d_{t-L} - \bar{d}_{t-L}) \Big/ \sqrt{\sum_{i=1}^{p}(d_t - \bar{d}_t)^2 \sum_{i=1}^{p}(d_{t-L} - \bar{d}_{t-L})^2 / (p-1)^2} \qquad (33)$$

(d_t – усереднені рівні наявного ряду; d_{t-L} – усереднені рівні наявного ряду, позиції яких визначені лагом L; p – кількість використовуваних пар; $\bar{d}_t$, $\bar{d}_{t-L}$ – середні величини в кожній групі даних).

Коефіцієнт автокореляції приймає значення в межах [-1; 1]. Абсолютна величина коефіцієнта автокореляції, близька до одиниці, свідчить про істотну залежність між порівнюваними групами послідовних рівнів ряду. Якщо лаг L дорівнює одиниці, тобто обчислюється автокореляція між сусідніми елементами ряду, то маємо коефіцієнт автокореляції першого порядку.

Для перевірки якості вибраної аналітичної залежності використовують *коефіцієнт автокореляції першого порядку* для ряду похибок між наявними даними ряду та розрахованими за вибраною аналітичною залежністю. Якщо названий коефіцієнт має абсолютну величину близьку до одиниці, то рівняння регресії (аналітичної залежності) є малодостовірним і його необхідно переглянути. Існує ряд методів, за допомогою яких визначають якість побудованого рівняння регресії. Розглянемо один із них – метод Дарбіна-Уотсона, використовуваний для малих за обсягом вибірок, тобто наявних рівнів ряду. У відповідності до методу використовується критерій DW, пов'язаний з гіпотезою про наявність автокореляції між сусідніми похибками $e_t = y_t - y(t)$ – значення, яке відповідає моменту t у

вибраному аналітичному тренді:

$$DW = \sum_{t=2}^{n}(e_t - e_{t-1})^2 / \sum_{t=1}^{n} e_t^2 . \qquad (34)$$

При відсутності значної автокореляції величина DW дорівнює приблизно 2, а при значній автокореляції вона дорівнює приблизно 0 або 4. Для DW-статистики розраховано критичні межі, які дозволяють прийняти або відхилити гіпотезу про відсутність автокореляції. Авторами критерію знайдені нижня d_e та верхня d_u межі при рівнях значущості 1%, 2,5% , 5%, розташування величини DW відносно них обґрунтовує ймовірність прийняття певної гіпотези.

Дані DW-статистики (значення критерію Дарбіна-Уотсона) наведено в табл. 4.8, величина m дорівнює кількості незалежних змінних при розбудові рівняння регресії (для часових рядів $m = 1$).

Таблиця 4.8.

n	Значення d_e і d_u при 5% значущості								Значення d_e і d_u при 1% значущості							
	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_e	d_u	d_e	d_u	d_e	d_u	d_e	d_u	d_e	d_u	d_e	d_u	d_e	d_u	d_e	d_u
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,95	1,15	0,86	1,27	0,77	1,41	0,69	1,57
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,32	1,40	1,28	1,45	1,24	1,49	0,20	1,54
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63

Критерій DW (34) використовується таким чином:

а) якщо обчислена величина DW опиняється в інтервалі від d_u до $(4 - d_u)$, то гіпотеза про відсутність автокореляції приймається;

б) якщо величина DW опиняється в інтервалі (d_e, d_u) або $(4 - d_e, 4 - d_u)$, то немає статистичних підстав ні прийняти, ні відхилити гіпотезу про відсутність автокореляції (ситуація невизначеності);

в) якщо $DW < d_e$ або $DW > (4 - d_e)$, то маємо підстави прийняти гіпотезу про наявність автокореляції.

Умови використання критерію зображено на рис. 4.7.

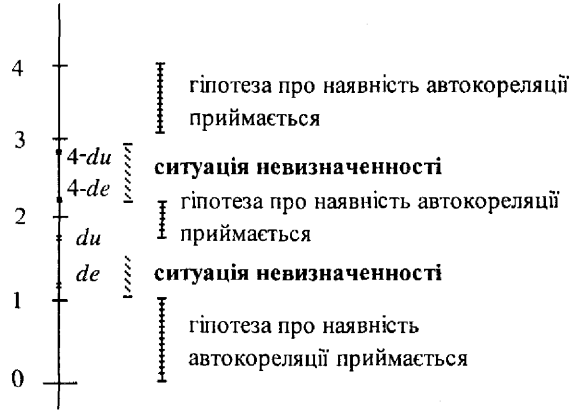


Рис. 4.7.

За умови $DW < d_e$ наявною є додатна автокореляція, за умови $DW > (4 - d_e)$ – від'ємна.

По завершенні визначення аналітичної залежності рівняння регресії необхідно обчислити DW-критерій (34); якщо буде виявлено наявність автокореляції, то необхідно переглянути побудову аналітичної залежності.

Приклад. У другому стовпці табл. 4.9 вказано обсяги споживання молочних продуктів по Харківській області за 1975-1983 рр. Аналіз графіку відносних експериментальних даних на рис. 4.8 і подальші розрахунки, частково наведені в табл. 4.9, дали підстави вибрати форму рівняння регресії у вигляді модифікованої експоненти $y = k + ab^t$ (ситуація 6а табл. 3.5).

Вилучивши теоретичні міркування щодо обґрунтування розрахунків параметрів модифікованої експоненти*, наведемо відповідні формули:

$$b = \frac{(n-1)\sum_{t=1}^{n-1} y_t y_{t+1} - \sum_{t=1}^{n-1} y_t \sum_{t=1}^{n-1} y_{t+1}}{(n-1)\sum_{t=1}^{n-1} y_t^2 - (\sum_{t=1}^{n-1} y_t)^2}, \quad a = \frac{n\sum_{t=1}^{n} b^t y_t - \sum_{t=1}^{n} b^t \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} b^{2t} - (\sum_{t=1}^{n} b^t)^2}, \quad k = \frac{\sum_{t=1}^{n} y_t - a\sum_{t=1}^{n} b^t}{n}.$$

* Четыркин Е. М. Статистические методы прогнозирования. – М.: Статистика, 1975.

Таблиця 4.9.

Рік	Дані спостереження		За рівнянням регресії		Похибка		Відхилення аналітичних і даних експерименту, %
	абсолютні (m)	відносні	абсолютні	відносні	абсолютна	відносна	
1	14928	1	14900,0	0,998	28,0	0,0018	1,878
2	16257	1,0890	15817,9	1,059	439,1	0,029	2,700
3	15767	1,0562	16272,1	1,090	-505,1	-0,0338	-3,203
4	15451	1,0350	16496,9	1,105	-1045,9	-0,0700	-6,763
5	16441	1,1013	16608,1	1,112	-167,1	-0,0111	-1,015
6	16837	1,1279	16663,1	1,116	173,9	0,0116	1,033
7	17154	1,1491	16690,4	1,118	463,6	0,0310	2,699
8	16977	1,1372	16703,9	1,118	273,1	0,018	1,608
9	17051	1,1422	16710,5	1,119	340,5	0,0228	1,997

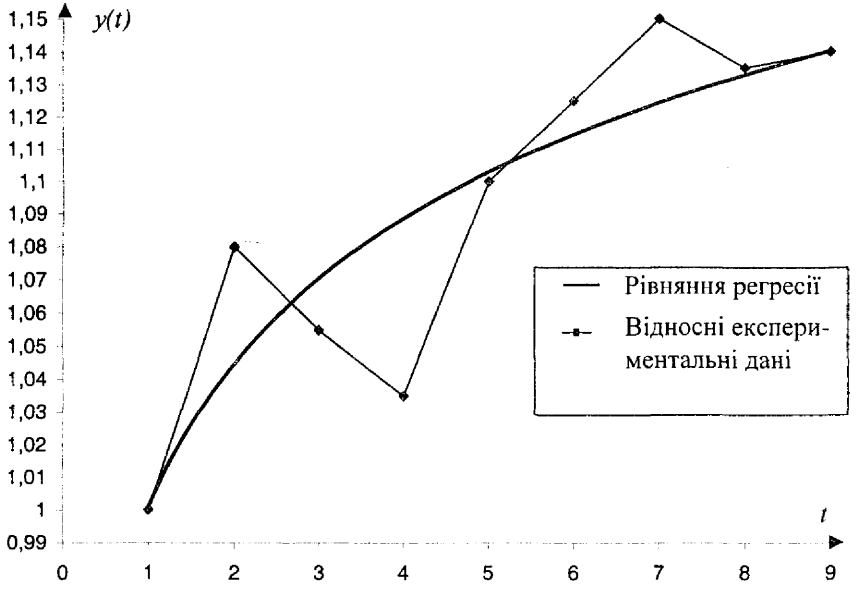


Рис. 4.8.

По виконанні необхідних обчислень маємо $k = 16717{,}05$, $a = -3672{,}07$, $b = 0{,}49483$. Величина $DW = 1{,}179$ при $m = 1$, бо враховується лише один фактор – час. Досліджуваний ряд має $n = 9$ членів, найближче табличне значення DW наведено при $n = 15$, тому для визначення необхідної величини DW використовуємо лінійну екстраполяцію. Виберемо рівень значущості 1% (0,01), враховуючи характер зміни меж d_e і d_u при $m = 1$, для лінійної екстраполяції виберемо необхідні дані при $n = 15$ і $n = 20$. Тоді величина d_u при $n = 9$ наближено дорівнюватиме:

$$d_u(9) = d_u(15) + \frac{d_u(20) - d_u(15)}{20 - 15}(9 - 15) \approx 0{,}96.$$

Так обчислена величина $DW = 1{,}179$ опиняється в інтервалі (d_u; $4 - d_u$), тобто, на рівні значущості 1% (з імовірностю помилки 0,01) можна стверджувати про відсутність автокореляції похибок. Отже рівняння регресії

$$y = 16717{,}05 - 3672{,}07\,(0{,}4948)^t \qquad (35)$$

з високою мірою точності описує тренд – основні закономірності зміни досліджуваного показника в часі. Найбільше відхилення величин, обчислених за (35), від експериментальних даних не перевищує 7% (табл. 4.9.)

4.5.3. Екстраполяція тренда для розрахунків прогнозу

Якщо рівняння регресії побудовано, тобто визначена закономірність зміни тренда від часу в минулому, то величина точкового значення прогнозу досліджуваного показника обчислюється підставлянням рівняння регресії величини відповідного моменту часу. Для досліджуваного прикладу з використанням вихідних даних за дев'ять років в одержане рівняння тренда (35) необхідно поставити величину $t = 10$. Тоді точкове значення прогнозу $\tilde{y}_9(10)$ на наступний, 10-й рік з використанням значень за дев'ять років дорівнюватиме:

$$\tilde{y}_9(10) = 16717{,}05 - 3672{,}07 \cdot (0{,}4948)^{10} = 16713{,}82.$$

Щоб переконатися в доцільності вибору форми тренда (аналітичної залежності) у вигляді модифікованої експоненти (6а табл. 3. 5) та оцінити точність прогнозу, побудуємо рівняння регресії з використанням даних за вісім початкових років. Вилучивши відповідні розрахунки, наведемо аналітичну залежність для цього випадку:

$$\tilde{y}_8(t) = 16609{,}63 - 3733{,}26 \cdot (0{,}45128)^t.$$

Підставивши у цю формулу $t = 9$ і $t = 10$, одержимо відповідно $\tilde{y}_8(9) = 16606{,}73$, $\tilde{y}_8(10) = 16608{,}33$. Перший із наведених результатів прогнозу порівняємо з величиною фактичного значення показника 17051 (табл. 4. 9) при $t = 9$. Відносна похибка прогнозованої величини складає 2,61% фактичного значення, що свідчить про високу точність прогнозу. Порівнюючи величини прогнозу для t=10 при використанні бази у вісім і дев'ять років (16608,33 та 16713,82), доходимо висновку, що вони близькі між собою. Таким чином, для наведеного прикладу за рівняння тренду доцільно вибрати модифіковану експоненту, яка аналітично описує перебіг процесу в часі відповідно до фактичних спостережених даних.

Використання рівняння регресії для обчислення величини прогнозу має сенс, якщо є підстави вважати, що зберігаються умови, які визначали перебіг процесу в минулому. З використанням екстраполяції маємо лише точкові значення прогнозу. *Збіг прогнозу та фактичних даних* по досягненні відповідного моменту – *явище практично не можливе,* тому що на фактичне значення показника впливає велика кількість випадкових факторів, а трендом можна описати лише вплив факторів, обумовлених природними закономірностями досліджуваного процесу. За названих обставин виникає необхідність пошуку інтервальних оцінок прогнозу. Такі оцінки виконуються за допомогою так званих *інтервалів довіри.* Інтервали довіри з високою мірою надійності (ймовірністю) визначають ті межі, в яких можна очікувати величину досліджуваного показника в майбутньому.

Перейдемо до розгляду питань оцінки точності та надійності прогнозу.

4.5.4. Оцінки точності та надійності прогнозу

У теорії економічного та технологічного прогнозування широко використовуються терміни "точність" та "надійність" з їх технічним тлумаченням. Наголосимо на деяких особливостях названих термінів стосовно прогнозування.

Точність прогнозу визначається як величина похибки, яка дорівнює різниці між прогнозованим і фактично досягнутим значенням досліджуваного показника. Величина відносної похибки прогнозу обчислюється за відношенням похибки до наявної досягнутої величини показника (досить часто у процентах). Такі оцінки можна дати лише у двох випадках: 1) по завершенні терміну упередження, коли одержимо фактичне значення досліджуваного показника; 2) при оцінці ефективності пропонованої моделі прогнозування за результатами перебігу процесу *в минулому* для тих моментів, для яких можна використати наявні заміри показника. У другому випадку наявну інформацію про уже відомий перебіг процесу розділяють на дві

групи. Сукупність більш ранніх даних використовують для розбудови математичної моделі, яку планують використати для прогнозування (вибір відповідної аналітичної залежності та розрахунки використаних параметрів). Після розбудови ММ другу групу даних спостереження, які відповідають більш пізнім моментам, використовують для визначення оцінок точності моделі прогнозування. При одержаних оцінках істотно підвищується рівень надійності прогнозу та довіри до результатів прогнозування, а також є певні підстави для оцінок різних ММ при їх наявності. Але величини похибок прогнозу "минулого" не варто використовувати як безсумнівний доказ ефективності апробованої ММ, бо це лише певна аргументація доцільності її використання. Величина досліджуваного показника буде визначатися станом реальних умов на проміжку прогнозування.

Точність прогнозу значною мірою залежить від точності представлення даних, які використовуються при прогнозуванні. Природно, що точність прогнозу за розрахунків її оцінки за вже досягнутих даних (прогноз expost) буде значно більшою, ніж за очікуваних величин даних (прогноз exante). В останньому випадку необхідно враховувати ймовірність похибок величин майбутніх значень параметрів.

При наявності даних про реалізацію в часі досліджуваного процесу в минулому якість прогнозування іноді оцінюють за формулою:

$$\eta = \frac{p}{p+q} \tag{36}$$

(p - кількість випадків, коли величини прогнозу опинялись в очікуваному інтервалі для прогнозу; q – кількість випадків протилежної ситуації).

Якщо *для всіх контрольованих моментів прогнози потрапляють у передбачувані інтервали,* то $q = 0$, а $\eta = 1$. Якщо прогноз для кожного контрольованого моменту не потрапив у передбачуваний інтервал, то $p = 0$ та $\eta = 0$. Такий підхід до оцінки якості прогнозу можливий лише за умови перевірки пропонованої моделі прогнозу *з використанням даних за минулі моменти.*

При наявності точкових величин прогнозу, одержаних з використанням певної ММ, для оцінки якості прогнозування можна користуватися статистичними характеристиками експериментальних даних: середньою абсолютною та середньою квадратичною похибками прогнозу, залишковою дисперсією тренда, кореляційними коефіцієнтами та ін. З використанням точкових оцінок можна обчислювати інтервали довіри для прогнозованих величин.

За пропозицією Г. Тейла* за міру якості прогнозу використовується так званий *коефіцієнт розбіжності*:

* Тейл Г. Экономические прогнозы и принятие решений. - М.: Статистика, 1971.

$$V = \frac{\sqrt{\sum (P_t - A_t)^2}}{\sqrt{\sum A_t^2}} \tag{37}$$

(P_t – величина приросту досліджуваного показника, обчислена з використанням ММ прогнозування для відповідного приросту часу; A_t – наявна величина досліджуваного показника).

Якщо для кожного контрольованого моменту прогнозу наявні величини досліджуваного показника та значення, обчислені з використанням ММ прогнозування, співпадають, то V дорівнює нулю.

Якщо при перевірці пропонованої ММ прогнозування одержують величину середньої квадратичної похибки, яку можна досягти з припущенням "стабільних" приростів досліджуваного показника, тобто $\rho_i = 0$, то V=1. За умови, що ММ прогнозування дає гірші результати, ніж з гіпотезою про відсутність приростів, буде те, що й було: $V > 1$.

Коефіцієнт розбіжності (37) може бути використаний при співставленні якості прогнозування за різними моделями.

Описані оцінки якості моделі прогнозування ґрунтуються на використанні інформації про дійсні значення досліджуваного показника, тобто стосуються минулого. *Але на практиці проблему точності прогнозу необхідно розв'язувати тоді, коли термін упередження не минув і дійсне значення прогнозованої величини досліджуваного показника не відоме.* За таких умов використовують статистичні методи оцінок точності прогнозування, пов'язані з поняттям інтервалу довіри. Модель прогнозування, яка дає менший інтервал довіри за одного й того ж значення ймовірності довіри, є більш точною.

Надійність прогнозу визначається ймовірністю реалізації відповідної оцінки прогнозу. Ймовірність реалізації оцінюється або суб'єктивно (експертні оцінки) або шляхом урахування інтервалу довіри прогнозу.

4.5.4.1. Оцінка дисперсії прогнозу

Як було показано раніше (підрозділ 3.5.2), за лінійної регресії значення y_t доцільно у багатьох випадках представляти залежністю:

$$y_t = \rho_y \cdot t + a + \varepsilon(x).$$

Схематичне розташування величин, використовуваних при розбудові рівняння регресії, наведено на рис. 4.9.

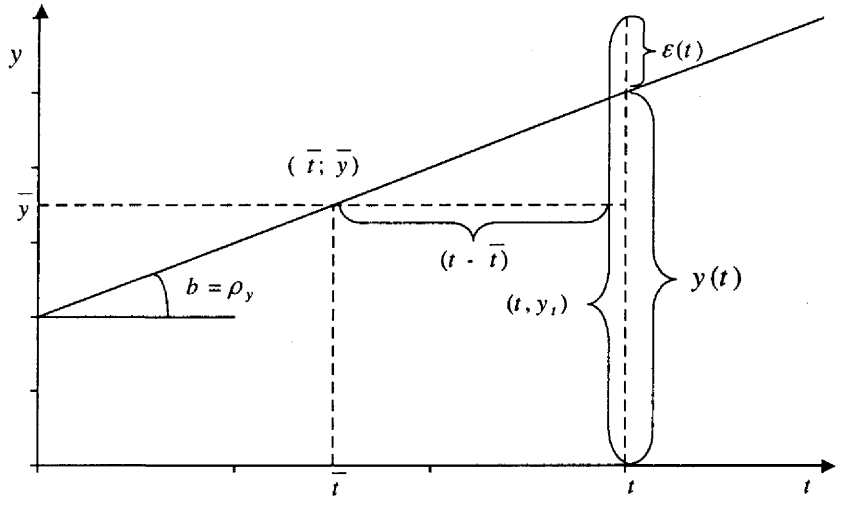


Рис. 4.9.

Величини випадкових відхилень похибок ε, обумовлених дією випадкових факторів, спостерігати безпосередньо неможливо хоча б тому, що невідомою є закономірність впливу постійно діючих факторів. Через це неможливо безпосередньо побудувати закон розподілу, тому використовуються припущення, наведені у підрозділі 3.5.2. Після перевірки значущості рівняння регресії одним з головних завдань використанням рівняння тренда (регресії) при прогнозуванні є обчислення інтервалів довіри. Величина інтервалу довіри обумовлена рівнями відхилень наявних даних спостереження від відповідних значень, обчислених з використанням рівняння тренда.

Розглянемо обчислення інтервалу довіри для рівняння регресії тренда, тобто для умовного математичного сподівання $M_t(y_t)$ при фіксованому значенні величини t.

Використавши залежність (19) глави 3, рівняння регресії запишемо так:

$$y(t) = \bar{y} + b(t - \bar{t}). \tag{38}$$

У наведеній залежності лише величина $(t - \bar{t})$ є детермінованою, а решта – випадкові величини значення яких обумовлені результатами замірів випадкових значень рівнів часового ряду. Прийнявши припущення про сталу величину дисперсії δ^2 відхилень рівнів ряду від величини відповідного математичного

сподівання та використавши властивості дисперсії двох незалежних випадкових величин, запишемо:

$$\delta_y^2 = \delta_{\bar{y}}^2 + \delta_b^2 (t - \bar{t})^2. \tag{39}$$

Якщо δ^2 є величиною дисперсії значень рівнів ряду при певному t, то величина $\delta_{\bar{y}}^2$ обчислюється так:

$$\delta_{\bar{y}}^2 = \delta^2 / n. \tag{40}$$

Для обчислення величини дисперсії δ_b^2 необхідно врахувати, що залежність (20) глави 3:

$$b = \frac{\sum_{i=1}^{n} (t_i - \bar{t})(y_t - \bar{y})}{\sum_{i=1}^{n} (t_i - \bar{t})^2} = \frac{\overline{ty} - \bar{t} \cdot \bar{y}}{\overline{t^2} - (\bar{t})^2}. \tag{41}$$

Тоді

$$\delta_b^2 = \frac{\sum_{i=1}^{n} (t_i - \bar{t})^2 \delta^2}{(\sum_{i=1}^{n} (t_i - \bar{t})^2)^2} = \frac{\delta^2}{\sum_{i=1}^{n} (t_i - \bar{t})^2}. \tag{42}$$

Замінивши невідоме значення δ^2 його вибірковою оцінкою s^2 з урахуванням (40) та (42), маємо:

$$s_y^2 = s^2 \left(\frac{1}{n} + \frac{(t - \bar{t})^2}{\sum_{i=1}^{n} (t_i - \bar{t})^2} \right). \tag{43}$$

У теорії математичної статистики доведено, що випадкова величина $(y(t) - M_t(y_t)) / s_y$ має розподіл Стьюдента з $k = n - 2$ степенями волі, а тому інтервал довіри до умовного математичного сподівання $M_t(y_t)$ при вибраній величині ймовірності довіри $\gamma = 1 - \lambda$ обчислюється так:

$$y(t) - t_{\gamma;k} s_y \le M_t(y_t) \le y(t) + t_{\gamma;k} s_y \qquad (44)$$

($t_{y;k}$ – величина функції розподілу за Стьюдентом при відомих значеннях γ $k = n - 2$; $s_y = \sqrt{s_y^2}$ – величина, обчислена згідно з (43)).

Аналізуючи залежності (43) та (44), доходимо висновку, що величина інтервалу довіри до очікуваної величини математичного сподівання залежить від розташування моменту заміру стосовно середнього значення проміжку замірів, що умовно зображено на рис. 4.10. При віддаленні від середньої точки ($\bar{t}$; $\bar{y}$) величина інтервалу довіри зростає.

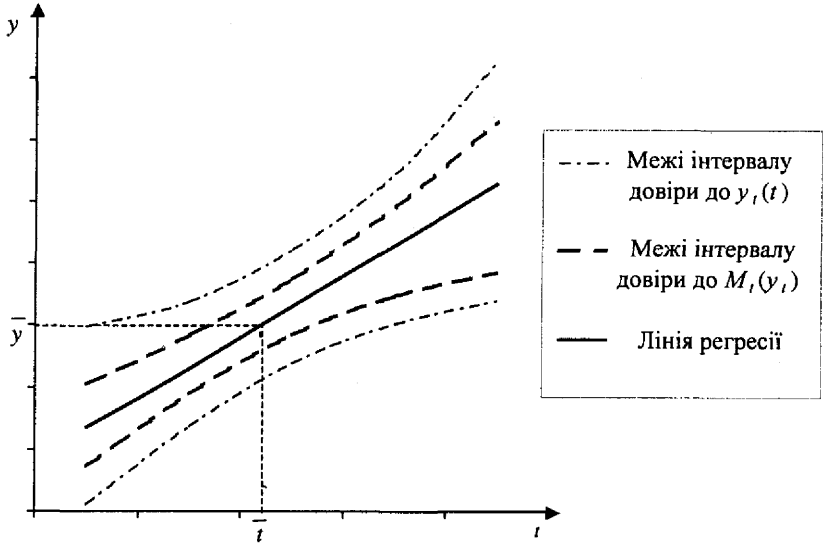


Рис. 4.10.

Необхідно наголосити, що розбудована величина інтервалу довіри (44) стосується умовного математичного сподівання, тобто розташування лінії регресії, але при окремих можливих значеннях рівнів часового ряду, які відхиляються від умовного математичного сподівання. З цієї причини при обчисленні інтервалу довіри до очікуваного значення рівня часового ряду необхідно враховувати ще одне джерело відхилень – розсіювання навколо умовного математичного сподівання, тобто величину $\delta^2 \approx s^2$. Таким чином, оцінка дисперсії очікуваного значення величини y^* при $t = t^*$ визначається формулою:

$$s_{y*}^2 = s^2 \left(1 + \frac{1}{n} + \frac{(t^* - \bar{t})^2}{\sum_{i=1}^{n} (t_i - \bar{t})^2}\right). \qquad (45)$$

Відповідний інтервал довіри буде обчислюватися за формулою:

$$y(t) - t_{\gamma;k} s_{y*}^2 \le y^* \le y(t) + t_{\gamma,k} s_{y*}^2. \qquad (46)$$

Використана у формулах (43) – (46) залишкова дисперсія вибірки s^2 при наявних рівнях часового ряду обчислюється за формулою:

$$s^2 = \left(\sum_{i=1}^{n} y_i - y(t_i))^2\right) / (n - 2) \qquad (47)$$

(y_i – наявні рівні часового ряду; $y(t_i)$ – величини, обчислені з розбудованим рівнянням тренда (регресії) при відповідних значеннях t).

При можливості екстраполяції перебігу досліджуваного показника на майбутнє в наявному рівнянні регресії для обчислення величини використовують значення $t = t_n + \tau$ (t_n – момент прогнозування, τ – горизонт прогнозування). Інтервал довіри буде обчислюватися за формулою (46) з $t^* = t_n + \tau$:

$$y(t_n + \tau) \pm t_{\gamma,k} s_p. \qquad (48)$$

(s_p – оцінка середньоквадратичної похибки прогнозу, визначення якої розглянемо далі).

Наведені формули обчислення величини прогнозу та інтервалу довіри до його значення є втіленням гіпотези про можливість перенесення на майбутнє умов, які відповідали перебігу досліджуваного показника в минулому.

4.5.4.2. Інтервал довіри прогнозу для лінійного тренда

При використанні лінійного тренда для прогнозування величина s_p обчислюється так:

$$s_p = s\sqrt{\frac{n+1}{n} + (t_n + \tau - \bar{t})^2 / (\sum_{i=1}^{n} (t_i - \bar{t})^2} \qquad (49)$$

(де s – обчислюється за формулою (47); n – кількість даних спостереження (додавання виконуються по всім спостереженням); $t_n + \tau$ – час, на який робиться

прогноз; t_i – момент прогнозування; τ – горизонт прогнозу; $\bar{t}$ – середнє значення проміжку, використаного при розбудові рівняння тренда).

При врахуванні того, що проміжки часу пронумеровані натуральними числами, корінь квадратний у (49), позначений літерою K, обчислюється так:

$$K = \sqrt{\frac{n+1}{n} + 3\frac{(n+2\tau-1)^2}{n(n^2-1)}}. \qquad (50)$$

Отже, K залежить лише від кількості точок часового ряду n та проміжку упередження (горизонту прогнозу) τ. Тому на практиці використовуються спеціальні таблиці для $K^* = tK$. З використанням K^* інтервал довіри обчислюється так: $y(t_n + \tau) \pm sK^*$.

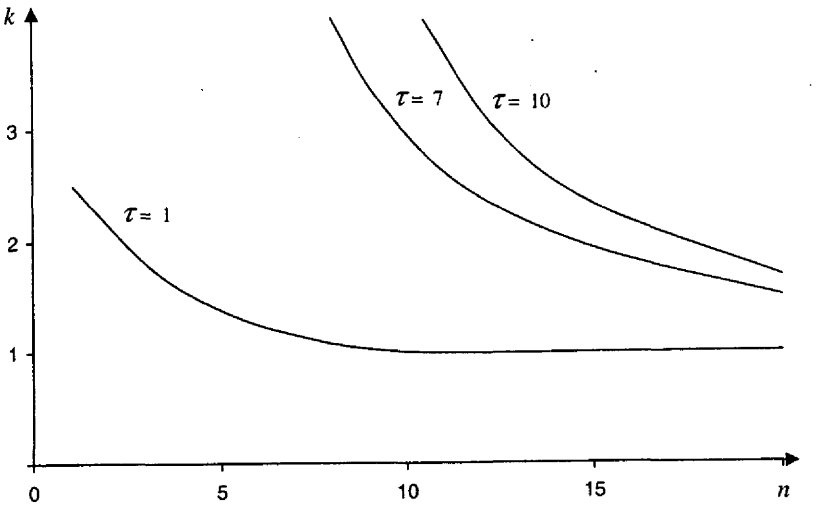
При зростанні величини n (кількість рівнів ряду, використаних для прогнозування) значення K^* та K зменшуються, бо природно очікувати збільшення точності прогнозу. При зростанні величини горизонту τ названі коефіцієнти зростають, бо чим більше τ, тим менш достовірним стає прогноз. На рис. 4.11 а показано закономірності зміни K від n при фіксованих значеннях τ, на рис. 4.11 б залежність K від τ.

Доцільно наголосити, що кількість n членів ряду впливає на величину інтервалу довіри також і через значення t_γ статистики Стьюдента, розподіл величин в якій залежить від числа степенів волі, а воно залежить від n.

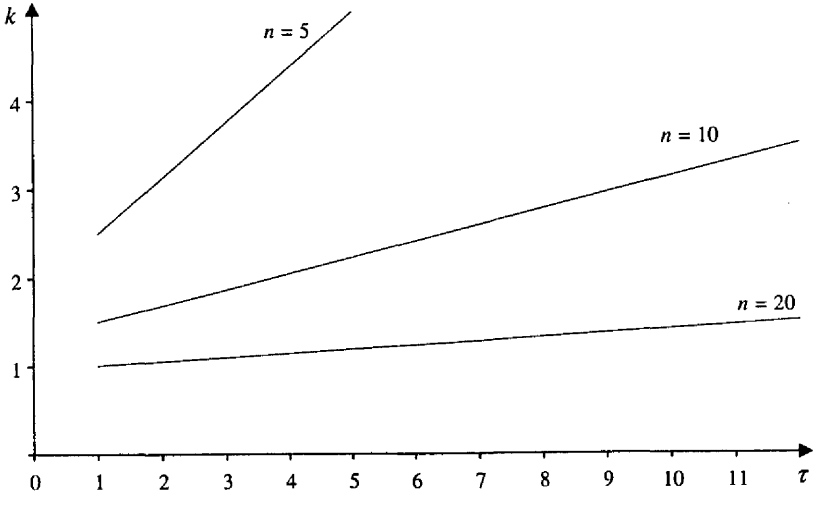
Приклад. Виконаємо розрахунки інтервалу довіри прогнозування величини попиту на наступний рік з відомим часовим рядом розподілу попиту за попередні роки згідно з табл. 4.9 (у першому та другому стовпцях вказано час і відповідний обсяг попиту).

Таблиця 4 9.

t_i	y_i	t^2	y_i^2	$t_i y_i$	$y(t_i)$	$y_i - y(t_i)$	$(y_i - y(t_i))^2$
1	213	1	45369	213	206,999	6	36
2	171	4	29241	342	232,678	-61,678	3804,176
3	291	9	84681	973	258,357	32,643	1065,565
4	309	16	95481	1236	284,036	24,964	623,201
5	317	25	100489	1585	309,715	7,285	53,071
6	362	36	131044	2172	335,394	26,606	707,879
7	351	49	123201	2451	361,073	10,073	101,465
8	361	81	130321	2888	386,752	25,752	663,165
$\sum 36$	2375	204	739827	11766			7059,2



а)



б)

Рис. 4.11.

Використавши рівняння регресії у вигляді (38) та розрахункові формули (19) і (20) глави 3, одержимо рівняння регресії (тренда):

$$y(t) = 181{,}32 + 25{,}679t.$$

Результати обчислення величин, необхідних для розбудови тренда, наведено в табл. 4.9. На рис. 4.12 зображено величини попиту (експериментальні дані) та графік одержаного рівняння регресії.

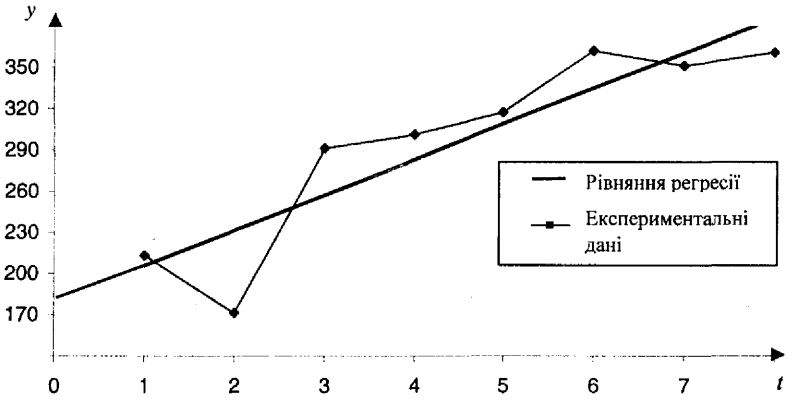


Рис. 4.12.

Перевірку значущості одержаного рівняння тренда виконаємо за F-критерієм на 5% рівні. Обчислимо необхідні суми квадратів:

а) обумовлену регресією:

$$Q_R = \sum_{i=1}^{8}(y(t_i) - \bar{y})^2 = \sum_{i=1}^{8} b^2(t_i - \bar{t})^2 25{,}679^2(204 - \frac{36^2}{8}) = 27695{,}3;$$

б) загальну:

$$Q = \sum_{i=1}^{8}(y_i - \bar{y})^2 = \sum_{i=1}^{8} y_i^2 - (\sum_{i=1}^{8} y_i)^2 / n = 739827 - \frac{2375^2}{8} = 34748{,}9;$$

в) залишкову:

$$Q_e = Q - Q_R = 34748{,}9 - 27695{,}3 = 7053{,}6.$$

Обчисливши величину $F = \frac{Q_R(n-2)}{Q_e} = \frac{27695{,}3 \cdot (8-2)}{7053{,}6} = 23{,}56$ та вибравши з таблиць F-розподілу значення $F_{0{,}05;1;6} = 5{,}99$, маємо $F > F_{0{,}05;1;6}$. Таким чином, рівнянню тренда можна довіряти з імовірністю 0,95 (95%).

Рівняння тренда $y(t) = 181{,}32 + 25{,}679t$ свідчить, що попит на товар щорічно збільшувався в середньому на 25,7 у. о. Обчислимо точкову оцінку та інтервал довіри до очікуваної величини попиту на дев'ятий рік. Точкова оцінка визначається з використанням рівняння тренда $y(9) = 412{,}4$ у. о.

Залишкову дисперсію вибірки s^2 обчислимо за формулою (47). З використанням результатів, наведених у табл. 4.9, маємо $s^2 = 7059{,}2/(8-2) = 1176{,}5$. З таблиці розподілу за Стьюдентом за $\gamma = 0{,}95$ маємо $t_{0{,}95;6} = 2{,}45$. Дисперсія s_y^2, згідно з формулою (43), дорівнює:

$$s_y^2 = 1176{,}5(\frac{1}{8} + \frac{(9-45)^2}{42}) = 714{,}3;\ s_y = \sqrt{714{,}3} = 26{,}73.$$

Таким чином, інтервальна оцінка прогнозу середнього значення попиту така: $412{,}4 + 2{,}45 \cdot 26{,}73 \le y(9) \le 412{,}4 - 2{,}45 \cdot 26{,}73$. Тобто $346{,}9 \le y(9) \le 477{,}9$.

Для обчислення інтервальної оцінки очікуваного значення $y^*(9)$ знайдемо величину відповідної дисперсії згідно з формулою (45):

$$s_{y^*}^2 = 1176{,}5(1 + \frac{1}{8} + \frac{(9-45)^2}{42}) = 1890{,}8;\ s_{y^*} = 43{,}48.$$

Отже, з використанням формули (46) інтервальна оцінка для $y^*(9)$ буде такою:

$$305{,}9 \le y^*(9) \le 518{,}9.$$

Таким чином, з імовірністю 0,95 середнє значення попиту на товар на дев'ятий рік буде в межах від 346,9 до 477,9, а його конкретне значення – від 305,9 до 518,9.

4.5.4.3. Інтервали довіри для трендів, аналітичні форми яких можна привести до лінійних

У табл. 3.5 вказано умови доцільності використання нелінійних аналітичних залежностей, які після заміни координат представлення досліджуваних факторів

можна описати як лінійні. Лінійна залежність в нових координатах дає можливість оцінити інтервали довіри в них, а потім, виконавши зворотний перехід у координатах відліку, обчислити інтервали довіри для вибраної нелінійної ММ прогнозування. У літературі описано і більш складні методи використання часових рядів при прогнозуванні соціально-економічних показників.

На завершення стисло вкажемо основні етапи розбудови середньо- та довгострокового прогнозу з використанням часових рядів.

1. Переконавшись в достатньому рівні достовірності наявних рівнів часового ряду, виконати вирівнювання його значень, використавши відповідну плинну.

2. Виконати перевірку на наявність автокореляції вирівняних рівнів ряду за критерієм *DW* (Дарбіна – Уотсона). При наявності значущої автокореляції ще раз перевірити вихідні дані та результати вирівнювання. При відсутності значущої автокореляції перейти до розбудови тренда (рівняння регресії).

3. При розбудові тренда (рівняння регресії) доцільно використовувати рекомендовані в табл. 3.5 аналітичні залежності.

4. Розбудувавши рівняння тренда, перевірити рівень його значущості, обчислити точкове значення очікуваної величини прогнозу, інтервали довіри до середнього очікуваного значення та можливого конкретного значення величини прогнозу.

4.6. Використання індикаторів та індексів

При прогнозуванні перебігу складних соціально-економічних процесів досить часто використовуються поняття індикаторів та індексів. Ці терміни дуже схожі за змістом, але віддзеркалюють дещо різні характеристики й явища.

Індикаторами (від лат. indicator – показник) в економіці називаються статистичні показники, динаміка яких характеризується стабільністю й однозначністю у зміні протягом певного економічного циклу. Приклади індикаторів: вартість обмінного курсу національної валюти, рівень безробіття, ВВП у порівняльних цінах.

При виконанні економічного аналізу відносно часу перебігу досліджуваного процесу розрізняють три *групи індикаторів*: упереджуючі, співпадаючі та запізнілі. До групи *упереджуючих* індикаторів відносять показники, зміна яких у часі передує зміні економічної кон'юнктури. Індикатори, обчислення величин яких можна виконати в моменти, співпадаючі з моментами оцінок спостережень досліджуваного процесу, відносять до групи *співпадаючих*. Величини *запізнілих* індикаторів можуть бути обчислені лише по завершенні відповідних моментів спостереження. Доцільно наголосити, що наведена класифікація індикаторів має відносно умовний характер і

суттєво залежить від особливостей розбудови економіки та накопичення відповідних даних.

В умовах стабільної ринкової економіки індикатори капіталовкладень в її розвиток об'єктивно необхідно віднести до групи запізнілих, в той же час такі капіталовкладення в перехідній економіці будуть упереджуючими. За класифікацією, прийнятою в США, до групи співпадаючих відносять ВВП, показники промислового виробництва, показники реальних прибутків населення. А індекси цін, курс акцій, курс валюти, характеристики зміни товарно-матеріальних запасів утворюють групи упереджуючих індикаторів. Але класифікація індикаторів не має чітких, однозначних критеріїв.

Характеристики поточного стану соціально-економічних процесів, представлені числами, які є результатами відповідних статистичних розрахунків, називаються *індексами* (від лат. index – показник). Термін "індикатор" є більш загальним, ніж термін "індекс", і використовується в основному при дослідженні динаміки процесу. Використання в математиці терміну "індекс" не суперечить змісту поняття index в латинській мові і є фактично показником, за допомогою якого розрізняють елементи певної множини.

Характерною та суттєвою особливістю методів, в яких використовуються індикатори та індекси, є можливість прогнозування термінів поворотних точок соціально-економічних циклів. Для цієї мети досліджується динаміка множини упереджуючих індикаторів. Якщо всі індикатори цієї множини змінюються тотожно, то прогнозування поворотної точки циклу не викликає ускладнень. Наприклад, якщо інтенсивність інфляційних процесів стабільно знижується, реальний обмінний курс національної валюти не падає, обсяги товарно-матеріальних запасів зменшуються, то можна очікувати появу поворотної точки в соціально-економічному розвитку, а саме: зростання ВВП, зменшення рівня безробіття, ріст реальних доходів населення. Тобто, досліджуючи динаміку упереджуючих індикаторів, можна визначити динаміку співпадаючих і запізнілих індикаторів.

При дослідженні змін індикаторів часто спостерігається поява фальшивих сигналів, тобто така динаміка індикаторів, яка не узгоджується з дійсним перебігом соціально-економічного процесу. Проблема розпізнавання *фальшивих сигналів* розв'язується використанням найбільш узагальнених індексів.

Узагальнений упереджуючий (співпадаючий, запізнілий) індекс визначається як середньозважена оцінка групи відповідних індикаторів. За величини вагових коефіцієнтів використовують показник значимості відповідних індикаторів, які, як правило, визначаються експертами. Якщо визначення таких вагових коефіцієнтів є складним або виникають сумніви стосовно їх величин, то всі вагові коефіцієнти значимості індикаторів даної групи вибираються рівновеликими, а узагальнений індекс обчислюється як середня оцінка по групі відповідних індикаторів. *Оцінки*

вагових коефіцієнтів визначаються так: якщо динаміка індикатора характеризує тенденцію поліпшення стану досліджуваного процесу, то індикатору надається оцінка рівна 1; якщо змін індикатора не спостерігається, то надається оцінка 0,5; за умови, що динаміка індикатора засвідчує спад його значень, оцінка індикатора дорівнює 0. Якщо обчислений таким чином узагальнений індекс буде на проміжку $[0; 0{,}5]$, то це буде свідчити про перевагу негативних тенденцій в досліджуваному процесі. Виконання умови, що значення узагальненого індексу потрапляє в проміжок $[0{,}5; 1]$, свідчить про об'єктивні тенденції покращання стану відповідного процесу.

Прогноз поворотної точки у перебігу процесу ґрунтується на співставленні оцінок упереджуючих узагальнених індексів двох послідовних проміжків спостереження: варто очікувати поворотну точку, якщо оцінки узагальнених індексів двох послідовних проміжків розташовані в різних інтервалах $[0; 0{,}5]$ та $[0{,}5; 1]$.

Природно очікувати, що за такою методикою розбудови прогнозу поворотної точки узагальнений індекс більш достовірно характеризуватиме перебіг процесу, якщо буде використано більше індикаторів.

Приклад. Динаміку упереджуючих індикаторів за вказані проміжки спостереження вказано в табл. 4.10. Чи варто очікувати поліпшення економічної ситуації?

Таблиця 4.10.

Показники	Січень–березень	Квітень–червень
Індекс цін на промислову продукцію	2,3	1,5
Індекс цін на сільськогосподарську продукцію	1,25	1,25
Індекс реального обмінного курсу	1,2	1,1
Індекс змін обсягу товарно-матеріальних запасів	1,1	0,9

Розв'язок. Обчислимо оцінки упереджуючих індикаторів, використовуючи їх динаміку, та запишемо їх у табл. 4.11.

Таблиця 4.11.

Показники	Січень-березень	Квітень-червень
Індекс цін на промислову продукцію	0	1
Індекс цін на сільськогосподарську продукцію	0,5	0,5
Індекс реального обмінного курсу	1	1
Індекс змін обсягу товарно-матеріальних запасів	0	0

Оскільки індекс цін на промислову продукцію спочатку зменшився, а потім збільшився, то у перший рядок табл. 4.11 записуємо відповідно 0 та 1. Індекс цін на сільськогосподарську продукцію стабільний, тому у другому рядку в обох стовпцях записано 0,5. Зменшення індексу реального обмінного курсу свідчить про поліпшення стану економіки, тому у третьому рядку в обох стовпцях записано 1. Індекс зміни обсягів товарно-матеріальних запасів зменшується, тому у четвертому рядку в обох стовпцях записано 0.

Припустивши тотожну значимість індикаторів, обчислимо узагальнений упереджуючий індекс як середнє арифметичне одержаних оцінок.

Узагальнений упереджуючий індекс I_1 за січень–березень дорівнює:

$$I_1 = \frac{0 + 0{,}5 + 1 + 0}{4} = 0{,}375,$$

I_2 за квітень–червень:

$$I_2 = \frac{1 + 0{,}5 + 1 + 0}{4} = 0{,}625.$$

Оскільки величини узагальнених упереджуючих індексів за два послідовних строки потрапили у два різних проміжки, а саме: $I_1 = 0{,}375 \in [0; 0{,}5]$, $I_2 = 0{,}625 \in [0{,}5; 1]$, то маємо підстави стверджувати про настання зворотного циклу в економіці, досліджувані стани якої характеризувалися згідно з табл. 4. 10.

Розглянувши ілюстративний приклад, використаємо описану методику для оцінки перебігу економічного процесу в Україні; скористаємося наведеними в табл. 4. 12 даними*.

Аналіз оцінок упереджуючих індексів, наведених в нижній частині табл. 4.12, свідчить, що 1994 рік став переломним у перебігу економіки України.

Маючи певну методику оцінки напрямку динаміки соціально-економічних процесів, бажано мати критерій оцінки швидкості їх перебігу. Для цієї мети користуються поняттям *індексу амплітуди*. Цей індекс використовується для відносного заліку швидкості перебігу досліджуваного процесу на певному проміжку у порівнянні з відомою (нормативною чи прогнозованою) величиною. Індекс амплітуди обчислюється як для окремого індикатора, так і для їх групи за такою схемою: а) обчислюється темп приросту індикатора (групи); б) розраховані темпи приросту порівнюються із заданими прогнозованими чи нормативними.

Індекс амплітуди групи індикаторів обчислюється як середньозважена величина індексів амплітуди групи індикаторів. За вагові коефіцієнти можна використати коефіцієнти значимості відповідних індикаторів.

* Мельник О. М. Інфляція: теорія і практика регулювання. – К., 1999.

Таблиця 4.12.

Показники	1991 р.	1992 р.	1993 р.	1994 р.	1995 р.
Динаміка упереджуючих індикаторів					
Середньомісячний темп інфляції	10,7	33,5	47,1	14,4	9
Індекс цін на промислову продукцію	2,6	42,3	97,7	8,7	2,7
Індекс цін на сільськогосподарську продукцію	1,8	19,3	40,6	6	42
Оцінки упереджуючих індикаторів					
Середньомісячний темп інфляції		0	0	1	1
Індекс цін на промислову продукцію		0	0	1	1
Індекс цін на сільськогосподарську продукцію		0	0	1	1
Приведений упереджуючий індекс		$\frac{0+0+0}{3}=0$	$\frac{0+0+0}{3}=0$	$\frac{1+1+1}{3}=1$	$\frac{1+1+1}{3}=1$

Якщо індекс амплітуди потрапляє у проміжок $[0;1]$, то це означає, що швидкість перебігу процесу менше швидкості, з якою порівнюють (прогнозованою чи нормативною). Якщо індекс амплітуди більше за одиницю, то це свідчить, що швидкість перебігу процесу більше швидкості, з якою виконуємо порівняння. Тобто індекс амплітуди характеризує інтенсивність перебігу процесу.

Приклад. Оцінити інтенсивність перебігу змін в економіці України в роки, вказані в табл. 4.13, де наведено середньорічні темпи приростів співпадаючих індикаторів.

Таблиця 4.13.

Показники	1991 р.	1992 р.	1993 р.	1994 р.	1995 р.	1996 р.
ВВП, % до попереднього року	– 8,7	–9,9	–14,2	–23	–12,2	–10
Реальна зарплата (1990 р.=100%)	–	81,3	36,8	33,5	36,9	35
Безробітні, % до робочої сили	–	0,3	0,4	0,4	0,6	1,6

Результати розрахунків індексів амплітуди кожного із названих співпадаючих індикаторів та порівняння їх динаміки з попереднім періодом вказані в табл. 4.14.

Таблиця 4.14.

Показники	1993 р.	1994 р.	1995 р.
Індекс амплітуди ВВП	$\frac{-14,2}{-9,9}=1,43$	$\frac{-23}{-14,2}=1,62$	$\frac{-12,2}{-23}=0,53$
Індекс реальної заробітної плати (1990 р.= =100%)	$\frac{36,8}{81,3}=0,45$	$\frac{33,5}{36,8}=0,91$	$\frac{36,9}{33,5}=1,10$
Кількість безробітних, %	$\frac{0,4}{0,3}=1,33$	$\frac{0,4}{0,4}=1$	$\frac{0,6}{0,4}=1,25$
Індекс амплітуди загалом по групі	$\frac{1,43+0,45+1,33}{3}=1,07$	$\frac{1,62+0,91+1}{3}=1,176$	$\frac{0,53+1,10+1,25}{3}=0,96$

Для 1993 та 1994 рр. індекс амплітуди загалом по групі був більшим за одиницю. Це свідчить про те, що криза характеризувалася підвищенням інтенсивності, яке почало спадати в 1995 р. Саме такий перебіг економічного процесу спостерігався в Україні в указані роки.

Характерною особливістю методу економічних індексів та індикаторів є те, що його процедура не досить чітко окреслена і в значній мірі залежить від майстерності виконавця при реалізації комплексного аналізу динаміки значної кількості індексів та індикаторів. Але для економіки, яка функціонує в більш-менш стабільних умовах, цей метод є одним з ефективних стосовно передбачення появи поворотних точок циклів соціально-економічних процесів.

Питання для самоконтролю

1. Що таке часовий ряд?
2. У чому різниця між стаціонарними та нестаціонарними часовими рядами?
3. Що таке тренд?
4. Як визначаються поняття бази та горизонту прогнозу?
5. Що таке тенденція зміни часового ряду?
6. Поняття плинних середніх та їх використання при дослідженні часових рядів.
7. Поняття середнього приросту та темпу приросту.
8. З якою метою використовують короткотермінове прогнозування?
9. Поняття адаптивного прогнозування та контролю прогнозу.
10. Мета середньо- та довготермінового прогнозування.
11. З якою метою розбудовують рівняння регресії?
12. Яким аналітичним способом записують рівняння лінійної регресії?
13. Що означають терміни "екстраполяція" та "інтерполяція"?
14. Який сенс мають поняття: точкова оцінка, інтервал довіри?
15. Який економічний сенс мають поняття: індекс та індикатор?
16. Класифікація індексів та індикаторів.
17. Які особливості розвитку соціально-економічних процесів досліджуються з використанням індексів та індикаторів?

ГЛАВА 5

ЗАГАЛЬНА ПОСТАНОВКА ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ (ЗЛП)

5.1. Приклади задач лінійного програмування

Для моделювання складних реальних процесів планування й управління необхідно враховувати чималу кількість факторів. Розглянемо приклади математичного моделювання, коли відповідну модель можна побудувати, використовуючи лінійні залежності між змінними, обраними для моделювання.

Задача оптимального планування виробництва. Припустимо, що підприємство може виробляти n найменувань продукції, використовуючи m видів ресурсів (сировина, енергоресурси, робоча сила і т. ін.), обсяги яких обмежені величинами b_i $(i=\overline{1,m})$. Відомими є витрати a_{ij} i-го ресурсу на одиницю продукції j-го найменування $(j=\overline{1,n})$ та ринкові ціни c_j одиниці відповідної продукції.

По обсягах продукції j-го найменування існують обмеження, обумовлені виробничими можливостями, що мають нижню d_j та верхню D_j межі. Необхідно скласти план виробництва продукції кожного найменування, враховуючи обсяги наявних ресурсів та обмеження на виробничі можливості, щоб одержати максимальний зиск від реалізації продукції. Так формулюючи задачу, ми приймаємо гіпотези, що вартість одиниці продукції є сталою величиною, обсяги продукції всіх найменувань будуть повністю реалізовані; витрати на зберігання ресурсів і продукції та транспортування при цьому не враховуються і т. ін. Незважаючи на такі припущення, сформульована ММ досить широко використовується на практиці при короткотерміновому плануванні.

Позначимо x_j кількість одиниць продукції j-го найменування, яку планується виробити. Величину зиску $z(x_1, x_2, ..., x_n)$ від реалізації продукції можна обчислити за формулою:

$$z(x_1, x_2, ..., x_n) = c_1x_1 + c_2x_2 + ... + c_nx_n = \sum_{j=1}^{n} c_jx_j . \qquad (1)$$

За планованих обсягів виробництва витрати i-го ресурсу будуть визначатися залежністю $a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j$ і не повинні перевищувати обсяг наявного запасу b_i: $\sum_{j=1}^{n} a_{ij}x_j \le b_i$.

Вимоги необхідності виконання обмежень по обсягах усіх ресурсів будуть представлені системою лінійних нерівностей:

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}). \tag{2}$$

Обмеження щодо можливих обсягів виробництва будуть представлені системою нерівностей:

$$d_j \le x_j \le D_j \quad (j = \overline{1, n}). \tag{3}$$

Математична модель задачі формулюється так: знайти такі величини x_j $(j = \overline{1, n})$, які б задовольняли системам нерівностей (2) і (3) та забезпечували найбільше значення функції (1).

Задача оптимізації харчового раціону для тварин. У зоотехнічній практиці давно приділяють велику увагу питанням харчового раціону для тварин. Основним питанням є збалансованість раціону за харчовими одиницями, перетравним протеїном, кальцієм, фосфором і каротином. Використання ЕОМ та відповідних ММ дозволяє спланувати раціон з урахуванням не лише енергетичного, білкового та макроелементного складу, а й амінокислотного та мікроелементного з метою повного забезпечення потреб тварин у харчових інгредієнтах за умови найменших витрат. Накопичений досвід показує, що використання методів лінійного програмування дає значний економічний ефект. Така ситуація обумовлена тим, що між кількістю споживаних тваринами харчових речовин та обсягами одержуваної від них продукції існує залежність, яку в певних межах можна описати лінійними співвідношеннями.

У загальних рисах мета задачі може бути сформульована так: використовуючи певні види кормів і добавок, що мають відомий за кількістю поживних речовин склад на одиницю виміру, так спланувати раціон харчування, щоб повністю забезпечити потреби тварин у необхідних харчових речовинах, при мінімальних витратах на корми. Звичайно, плануючи раціон харчування, слід враховувати і біологічні особливості окремих видів тварин, і сезонні умови, і термін планування раціону і т. ін.

Пропонуємо загальну схему розв'язання сформульованої задачі. Розбудовуючи ММ задачі, за змінні величини приймемо кількості кормів і кормових добавок у відповідних одиницях виміру. Умови щодо збалансованості харчових речовин у раціоні описуються певними обмеженнями на величини змінних. Коефіцієнти ММ характеризують вміст харчових речовин стосовно певних одиниць виміру в кормах і кормових добавках.

ММ сформульованої задачі. Припустимо, що маємо n видів кормів і кормових добавок (сіно, силос, концентрати і т. ін.), у кожному з яких міститься в різних

кількостях m видів харчових речовин (білки, кальцій, фосфор, вітаміни тощо). Відомо, що одиниця корму j-го виду $(j = \overline{1, n})$ має a_{ij} одиниць i-ої харчової речовини $(i = \overline{1, m})$. Якщо якась речовина в певному виді кормів відсутня, то відповідна величина a_{ij} дорівнює нулю. Необхідна потреба i-ої харчової речовини на добу має бути не менше b_i, вартість одиниці j-го виду корму або кормової добавки дорівнює c_j. Необхідно спланувати раціон – набір і кількість кормів так, щоб кількість кожної харчової речовини була не менше необхідної, а вартість раціону за такої умови – мінімальною.

Припустимо, що x_j $(j = \overline{1, n})$ – кількість одиниць корму j-го виду, який входить у добовий раціон $(x_1, x_2, ..., x_n)$. За такої умови загальна кількість i-ої харчової речовини в добовому раціоні буде визначатися сумою:

$$a_{i1}x_1 + a_{i2}x_2 + ... + a_{in}x_n = \sum_{j=1}^{n} a_{ij} x_j .$$

Ця загальна кількість має бути не менше величини відповідної харчової речовини, яку вимагають зоотехнічні норми. Стосовно переліку усіх видів необхідних харчових речовин вказана вимога буде представлена системою лінійних нерівностей:

$$\sum_{j=1}^{n} a_{ij} x_j \ge b_i \quad (i = \overline{1, m}), \tag{4}$$

до яких необхідно приєднати очевидні вимоги $x_j \ge 0$ $(j = \overline{1, n})$.

Вартість добового раціону $z(x_1, x_2, ..., x_n)$ обчислюється за формулою:

$$z(x_1, x_2, ..., x_n) = c_1 x_1 + c_2 x_2 + ... + c_n x_n = \sum_{j=1}^{n} c_j x_j . \tag{5}$$

Таким чином, задача планування добового раціону приведена за названих умов до такої ММ: знайти такі величини $(x_1, x_2, ..., x_n)$, за яких виконується система нерівностей (4), а функція $z(x_1, x_2, ..., x_n)$ досягає найменшого значення. Якщо за зоотехнічними вимогами необхідно додержуватися певних співвідношень між обсягами соковитих і грубих кормів, то треба додати відповідні залежності до системи обмежень на допустимі значення величин x_j. У ММ можна врахувати й інші зоотехнічні вимоги, наприклад, наявність деяких поживних речовин в кількостях, не менших і не більших за потрібні. Тоді відповідні нерівності системи (4) будуть записані так:

$$b_i \le \sum_{j=1}^{n} a_{ij} x_j \le B_i \quad (i = \overline{1, m}).$$

Наведена задача оптимізації раціону за умови мінімальної його вартості є типовою для класу так званих задач на оптимізацію сумішей. Такі задачі виникають тоді, коли необхідно утворити суміші різних компонентів з метою одержання сумішей з певними властивостями або замінити одні компоненти на інші (наприклад, створення найбільш економної суміші пального для двигуна внутрішнього згоряння, суміш добрив, суміш вугілля різних марок для парових котлів, склад шихти для плавок сталі і т. ін.).

5.2. Різні форми запису задачі лінійного програмування

Аналізуючи приклади наведених ЗЛП, доцільно наголосити на деяких спільних особливостях їх математичних моделей. Якість плану організації виробництва або раціону годівлі худоби оцінюється функцією n-змінних, з допомогою яких розбудовуються відповідні ММ. Функцію, якою оцінюють якість, називають *функцією цілі.*

За умовами задач необхідно знайти екстремальні значення відповідних цільових функцій. В обох задачах цільові функції були лінійними, і пошук їх екстремальних значень, якби не було вимоги обмежень, не викликає ніяких труднощів. А саме: якщо необхідно знайти найменше значення лінійної функції $\sum_{j=1}^{n} c_j x_j$ за умови, що всі $c_j \ge 0$ $(j = \overline{1, n})$, то необхідно вибрати всі $x_j = 0$ $(j = \overline{1, n})$, якщо функція визначена для $x_j \ge 0$. Якщо за названих умов для c_j та x_j $(j = \overline{1, n})$ необхідно знайти найбільше значення цільової функції, то чим більшими є значення величин x_j, тим більшим є значення функції. Але таке рішення не має практичного сенсу. Шукані величини змінних мають задовольняти певним умовам, математично представленим системами нерівностей певного сенсу.

Ураховуючи зазначені особливості задач наведених прикладів, загальну постановку ЗЛП формулюють так: знайти такі значення n-змінних $(x_1, x_2, ..., x_n)$, які задовольняють системі m лінійних нерівностей певного сенсу або рівностей і надають екстремальне значення (min або max) цільовій функції.

Математична модель ЗЛП в загальному випадку записується так: знайти екстремум функції

$$z = \sum_{j=1}^{n} c_j x_j \tag{6}$$

за умови, що x_j $(j = \overline{1, n})$ задовольняють системі лінійних нерівностей або рівностей, які називаються обмеженнями:

$$\sum_{j=1}^{n} a_{ij} x_j \le (\ge)\ b_i \quad (i = \overline{1, m}). \tag{7}$$

Приймаємо, що c_j $(j = \overline{1, n})$, b_i $(i = \overline{1, m})$ та a_{ij} $(j = \overline{1, n}; i = \overline{1, m})$ – відомі сталі, а в кожному з обмежень наявний лише один знак ≤, = або ≥. Величини n та m кількісно між собою не пов'язані: m може бути більше, менше або дорівнювати n. Якщо обмеження відсутні, то m дорівнює нулю. У багатьох задачах змінні x_j мають задовольняти вимозі невід'ємності, тобто $x_j \ge 0$ $(j = \overline{1, n})$. Набір числових значень величин $(x_1, x_2, ..., x_n)$ називається *планом* задачі. Доцільно наголосити, що не в кожній ЗЛП має бути план, бо не кожна система обмежень (7) має розв'язок.

Математична модель ЗЛП, записана у вигляді (6) та (7), називається *загальною задачею лінійного програмування.* Визначальним фактором є те, що цільова функція (6) та аналітичне представлення лівих частин обмежень (7) є лінійними стосовно змінних задачі. Усі практичні задачі, ММ яких приводяться до задач лінійного програмування, мають представлення, які є частинними стосовно (6) та (7).

Математичні методи розв'язання ЗЛП розбудовані за умов, що форма їх запису певним чином упорядкована. Розглянемо деякі форми запису ЗЛП та правила переходу від однієї форми до іншої.

Задачею лінійного програмування в *симетричній формі* запису називають задачу, математична модель якої формулюється так: знайти *максимум* функції

$$z = \sum_{j=1}^{n} c_j x_j \tag{8}$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}) \tag{9}$$

$$x_j \ge 0 \quad (j = \overline{1, n}). \tag{10}$$

Симетричну форму запису ЗЛП іноді називають *стандартною.*

Для розбудови методів розв'язання ЗЛП широко використовують запис у *канонічній формі* (приведеній): знайти *максимум* функції

$$z = \sum_{j=1}^{n} c_j x_j \tag{11}$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad (i = \overline{1, m}) \tag{12}$$

$$x_j \geq 0 \ (j = \overline{1, n}). \tag{13}$$

Канонічну форму запису ЗЛП іноді називають *основною*.

Математичні моделі реальних задач можуть мати форми, які відрізняються від вищеназваних. Наприклад: практична задача вимагає пошуку найменшого значення, моделювання обмежень приводяться до системи нерівностей різного сенсу і т. ін. Тому доводиться переходити від одних форм математичних моделей до інших, їм еквівалентних за результатами розв'язку.

Такий перехід виконується шляхом відповідних математичних перетворень. Якщо певна нерівність має сенс ≥, то домножують обидві частини нерівності на "–1". Якщо деякі змінні за змістом реальної задачі *не підлягають обмеженням невід'ємності*, то кожну з таких змінних замінюють *різницею двох нових змінних, які обов'язково мають задовольняти умові невід'ємності.* Тобто якщо деяка змінна x_r за змістом реальної задачі не зобов'язана бути обмеженою умовою невід'ємності ($x_r \geq 0$), то при необхідності її можна представити як різницю $x_r = x_l - x_k$, поставивши вимоги $x_l \geq 0$ і $x_k \geq 0$.

При необхідності задачу пошуку найменшого значення можна замінити задачею пошуку найбільшого значення. Для пояснення відповідних перетворень розглянемо дві функції: функцію $z(x_1, x_2, ..., x_n)$ та функцію з протилежним знаком "$-z(x_1, x_2, ..., x_n)$". Якщо за певних значень $(x_1^0, x_2^0, ..., x_n^0)$ функція z досягає найменшого значення, то функція "$-z(x_1, x_2, ..., x_n)$" досягає свого найбільшого значення за умови $(x_1^0, x_2^0, ..., x_n^0)$ і навпаки. Величини екстремальних значень обох функцій пов'язані залежністю: $\min z(x_1^0, x_2^0, ..., x_n^0) = -\max(-z(x_1^0, x_2^0, ..., x_n^0))$. Для функції однієї змінної наведене твердження має геометричне тлумачення, зображене на рис. 5. 1. Таким чином, якщо задачу пошуку найменшого значення функції за певних обмежень необхідно замінити задачею пошуку найбільшого значення за тих же обмежень, то треба замість вихідної функції $z(x_1, x_2, ..., x_n)$ взяти за цільову функцію $-z(x_1, x_2, ..., x_n)$, знайти її найбільше значення та змінити потім його знак на протилежний. Значення змінних, за яких досягаються екстремуми, будуть тотожними.

Еквівалентні перетворення нерівностей в рівняння або рівнянь в нерівності ґрунтуються на теоремі 5. 1 (яку даємо без доведення).

Теорема 5. 1. У будь-якому розв'язку $(x_1^*, x_2^*, ..., x_n^*)$ нерівності

$$a_1 x_1 + a_2 x_2 + ... + a_n x_n \leq b \tag{14}$$

відповідає розв'язок $(x_1^*, x_2^*, ..., x_n^*, x_{n+1}^*)$ рівняння

$$a_1 x_1 + a_2 x_2 + ... + a_n x_n + x_{n+1} = b, \tag{15}$$

в якому

$$x_{n+1}^* \geq 0, \tag{16}$$

і навпаки: кожному розв'язку $(x_1^*, x_2^*, ..., x_n^*, x_{n+1}^*)$ рівняння (15) та нерівності (16) відповідає рішення $(x_1^*, x_2^*, ..., x_n^*)$ нерівності (14).

Змінну x_{n+1} називають *додатковою.* Іноді вона має певний економічний сенс, а іноді – лише для формальних перетворень.

Приклад. Привести до канонічної форми дану ЗЛП. Знайти найменше значення

$$z = 2x_1 + 3x_2 - x_3$$

за умови

$$\begin{cases} x_1 - 2x_2 + 4x_3 \geq 5 \\ 2x_1 + x_2 \leq 9 \\ x_1 + 3x_2 - 2x_3 = 2 \\ x_1 \geq 0; x_3 \geq 0. \end{cases}$$

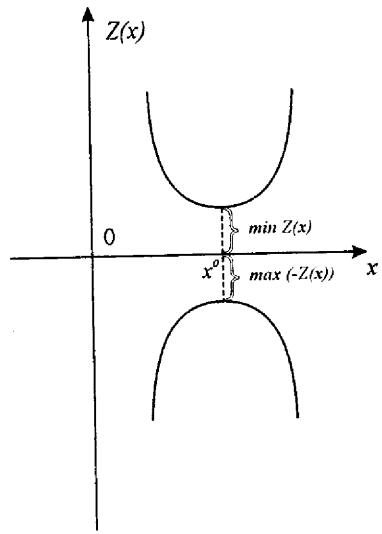


Рис. 5.1.

Замість даної функції будемо досліджувати функцію – $z = z_н = -2x_1 - 3x_2 + x_3$. Від лівої частини першої нерівності віднімемо додаткову невід'ємну змінну x_4, а до лівої частини другої нерівності додамо додаткову невід'ємну змінну x_5. Оскільки для x_2 у вихідній задачі не поставлена вимога невід'ємності, то цю змінну замінюємо різницею $x_6 - x_7$ двох невід'ємних нових змінних. Після наведених перетворень маємо канонічну форму сформульованої ЗЛП.

Знайти найбільше значення $z_н = -2x_1 - 3(x_6 - x_7) + x_3$ за умови:

$$\begin{cases} x_1 - 2(x_6 - x_7) + 4x_3 - x_4 = 5 \\ 2x_1 + (x_6 - x_7) + x_5 = 9 \\ x_1 + 3(x_6 - x_7) - 2x_3 = 2 \\ x_1 \ge 0; x_3 \ge 0; x_4 \ge 0; x_5 \ge 0; x_6 \ge 0; x_7 \ge 0. \end{cases}$$

На підставі теореми 5.1 систему m нерівностей з n змінними

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}), \tag{17}$$

використовуючи m додаткових невід'ємних змінних x_{n+i} $(i = \overline{1, m})$, можна привести до системи m лінійних рівнянь з $n + m$ змінними:

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+i} = b_i \quad (i = \overline{1, m}), \tag{18}$$

яка буде еквівалентною вихідній системі нерівностей в тому розумінні, що будь-якому розв'язку $(x_1^*, x_2^*, ..., x_n^*)$ системи (17) відповідає певний розв'язок $(x_1^*, x_2^*, ..., x_n^*, x_{n+1}^*, ..., x_{n+m}^*)$ системи (18), в якому додаткові змінні x_{n+i}^* $(i = \overline{1, m})$ невід'ємні. При необхідності найменування змінних можна замінити, щоб індекс імен змінювався неперервно.

Перехід від канонічної форми запису ЗЛП (11-13) до симетричної (8-10) виконується шляхом еквівалентних перетворень системи лінійних рівнянь (12) до системи нерівностей (9). Оскільки для подальших викладок буде використовуватися лише канонічна форма запису ЗЛП, то вищезгадані перетворення опускаємо.

5.3. Геометричне тлумачення задачі лінійного програмування

Геометричне тлумачення ЗЛП дозволяє розбудувати графічний метод її розв'язання. Такий метод досить очевидний і дозволяє з'ясувати основні особливості ЗЛП, але має надто обмежену область використання, а саме: якщо в симетричній формі запису кількість невідомих не більше ніж 3. Будемо розглядати ЗЛП у симетричній формі за умови, що кількість невідомих дорівнює 2.

Нагадаємо, що графік кожної функції $f(x, y) = 0$ в системі координат (x, y) розділяє площину змінних x та y на дві частини: для координат точок, в одній з яких виконується нерівність $f(x, y) > 0$, в іншій – $f(x, y) < 0$, а на лінії їх розмежування (графік функції $f(x, y) = 0$) рівність $f(x, y) = 0$. Щоб з'ясувати, в якій частині виконується певна нерівність, необхідно і достатньо взяти координати будь-якої точки (пробної), розташування якої відомо відносно графіка лінії $f(x, y) = 0$, і підставити у формулу $f(x, y)$. У результаті підстановки одержимо певну нерівність. Для координат усіх точок, розташованих відносно графіка $f(x, y) = 0$, як і вибраної точки, буде виконуватися нерівність того ж сенсу. Для координат точок, розташованих відносно лінії по-іншому, ніж вибрана точка, виконуватиметься нерівність протилежного сенсу. За домовленістю *область, в якій виконується нерівність певного сенсу*, по межі ($f(x, y) = 0$) *відмічається штриховкою.*

Приклади. 1. Дати графічний розв'язок нерівності

$$y^2 - x^3 \le 0.$$

Вибравши прямокутну систему координат, будуємо графік $y^2 - x^3 = 0$ $(y = \pm\sqrt{x^3})$, зображений на рис. 5.2. Вибираємо пробну точку з координатами (1; 0). Вибрані координати підставляємо в залежність $(y^2 - x^3)$ і одержуємо $(y^2 - x^3) < 0$. Отже, як для пробної точки, так і для всіх координат точок площини (x, y), розташованих справа відносно розбудваного графіка, виконується нерівність $y^2 - x^3 < 0$. На рис. 5.2 відповідна область по межі відмічена штрихами. На межі маємо $y^2 - x^3 = 0$.

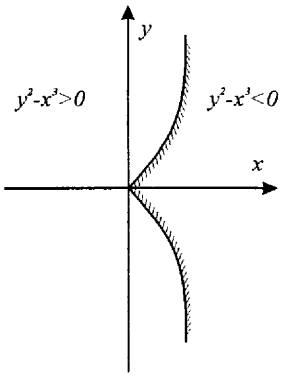


Рис. 5.2.

2. Дати графічний розв'язок нерівності

$$\frac{x^2}{4} + \frac{y^2}{9} - 1 \ge 0.$$

У прямокутній системі координат будуємо еліпс $\frac{x^2}{4} + \frac{y^2}{9} - 1 = 0$ (рис. 5.3). Вибираємо пробну точку (0; 0), координати якої підставляємо в ліву частину досліджуваної нерівності й одержуємо $\frac{x^2}{4} + \frac{y^2}{9} - 1 < 0$. Отже, задана нерівність виконується для координат точок, розташованих на еліпсі та зовні графічного зображення еліпса.

Графічний метод розв'язання ЗЛП, як правило, використовується при симетричній (стандартній) формі запису за умови наявності не більше трьох змінних.

Розглянемо розбудову методу за наявності двох змінних.

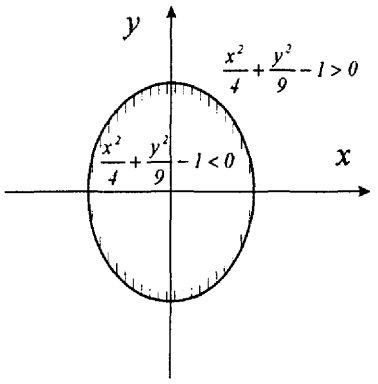


Рис. 5.3.

Припустимо, що маємо задачу: знайти найбільше значення

$$z = c_1x_1 + c_2x_2 \qquad (19)$$

за умови

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \le b_1 \\ a_{21}x_1 + a_{22}x_2 \le b_2 \\ \dots\dots\dots\dots \\ \dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 \le b_m \\ x_1 \ge 0; x_2 \ge 0. \end{cases} \qquad (20)$$

1-й етап. Будуємо графічне зображення області допустимих розв'язків згідно з умовою (20) у вибраній системі координат (x_1, x_2). Для цього спочатку будуємо, як було описано раніше, графічне зображення розв'язку кожної з нерівностей із системи (20), потім знаходимо їх спільну частину, якщо вона є.

Приклад. Побудувати графічно спільний розв'язок системи нерівностей

$$\begin{cases} x_1+3x_2 \le 15 & (1) \\ 2x_1 + 2x_2 \le 16 & (2) \\ x_2 \le 5 & (3) \\ x_1 \le 6 & (4) \\ x_1 \ge 0 & (5) \\ x_2 \ge 0. & (6) \end{cases} \qquad (21)$$

Вибираємо прямокутну систему координат (x_1, x_2) – рис. 5.4. Для графічного зображення кожної з наведених нерівностей спочатку будуємо відповідну пряму лінію, яка відповідає рівності, а потім знаходимо ту частину площини, в якій виконується задана нерівність. Виконавши побудову для кожної з нерівностей, знаходимо частину площини з точками, координати яких задовольняють наведеній системі. Область сумісних розв'язків системи заштрихована по межі всередину.

2-й етап. Будуємо лінії рівня цільової функції та визначаємо напрям її зростання. Нагадаємо, що лінією рівня функції $z = f(x, y)$ називають множину точок площини з координатами (x, y), для яких функція z має стале значення, тобто рівняння лінії рівня визначається умовою $f(x, y) = const$.

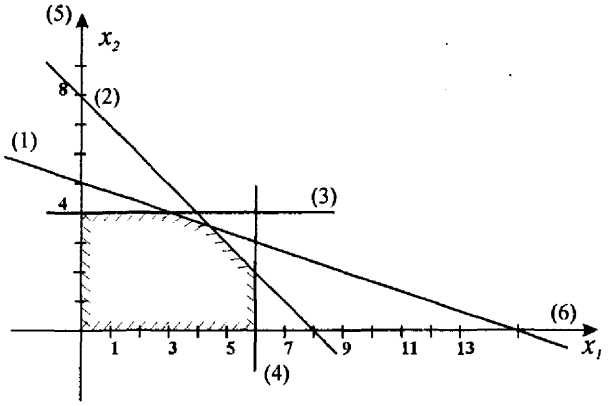


Рис. 5.4.

Поняття лінії (поверхні) рівня широко використовується в практиці, наприклад, ізотерми (лінії однакової температури) та ізобари (лінії однакового тиску) в метеорології, широти в географії, лінії рівня на топографічних картах. З курсу математики відомо, що напрям найбільшого зростання функції в кожній точці області її визначення вказує так званий градієнт, який обчислюється за формулою $grad\ z = \frac{\partial f}{\partial x}\bar{i} + \frac{\partial f}{\partial y}\bar{j}$. Градієнт завжди перпендикулярний до лінії рівня, яка проходить через точку обчислення градієнта.

У ЗЛП цільова функція завжди лінійна $z = c_1x_1 + c_2x_2$, тому *лінії рівня* такої функції *є прямими*, а $grad\ z = c_1\bar{i} + c_2\bar{j}$ не залежить від x_1, x_2, тобто має *сталу орієнтацію* за певної цільової функції.

Якщо в системі координат x_1, x_2 побудувати хоча б одну лінію рівня і $grad\ z$, то водночас буде виявлено, як треба змінювати x_1, x_2, щоб цільова функція зростала або спадала, бо напрям найбільшого спадання функції визначає антиградієнт $-\frac{\partial f}{\partial x_1}\bar{i} - \frac{\partial f}{\partial x_2}\bar{j}$.

Приклад. Припустимо, що маємо $z = 3x_1 + 5x_2$, тоді:

$$grad\ z = 3\bar{i} + 5\bar{j}.$$

Лінія рівня $z = 15$ та $grad\ z$ побудовані на рис. 5.5.

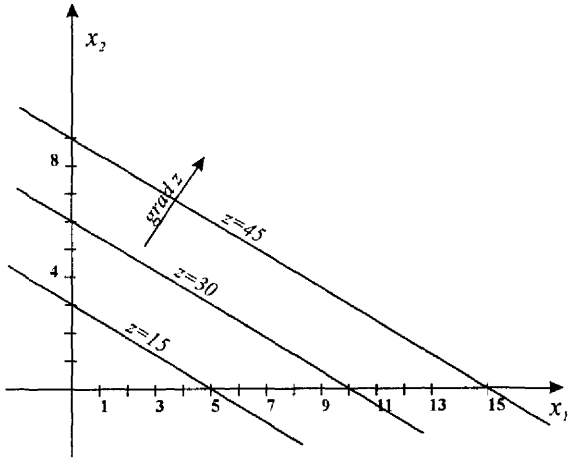


Рис. 5.5.

Оскільки z лінійна функція, то *grad z* має сталий напрям і його можна співставляти будь-якій точці площини (x_1, x_2). Чим більше лінія рівня віддалена від початку координат в напрямі градієнта, тим більшим є значення функції. На рис. 5. 5 показано лінії рівня $z = 15$, $z = 30$ та $z = 45$.

3-й етап. Знаходимо розв'язок відповідної ЗЛП з використанням графічних даних першого та другого етапів. Припустимо, що маємо графічне зображення області допустимих розв'язків ЗЛП, графік хоча б однієї лінії рівня даної цільової функції та її градієнт.

Маючи вказану інформацію, необхідно знайти лінію рівня, яка відповідає найбільшому значенню цільової функції і проходить хоча б через одну точку області допустимих розв'язків. Якщо така точка лише одна, то її координати будуть розв'язком ЗЛП. У залежності від математичної моделі практичної задачі ЗЛП може мати лише один розв'язок, нескінченну множину розв'язків або не мати жодного. Ці ситуації будуть обговорені згодом .

Приклад. Знайти найбільше значення функції $z = 4x_1 + 3x_2$ за умови:

$$\begin{cases} 2x_1 + 7x_2 \le 92 & (1) \\ -2x_1 + x_2 \le 4 & (2) \\ x_1 - 1{,}5x_2 \le 6 & (3) \\ x_1 \ge 0 & (4) \\ x_2 \ge 0 & (5) \end{cases} \qquad (22)$$

Вибравши систему координат (x_1, x_2), будуємо область допустимих розв'язків (рис. 5.6): спочатку – прямі лінії, які відповідають рівнянням, а потім знаходимо ті півплощини, для яких виконуються вказані нерівності. Областю допустимих розв'язків буде частина площини, спільна для всіх знайдених півплощин. На рис. 5.6 вона помічена штриховкою по межі із середини.

Будуємо графік однієї з ліній рівня цільової функції, наприклад, $4x_1 + 3x_2 = 24$ та $grad\ z = 4\bar{i} + 3\bar{j}$. Маючи графічне зображення області допустимих розв'язків, характер розташування ліній рівня та напрям зростання значення цільової функції, доходимо висновку, що найбільше значення цільової функції буде відповідати тій лінії рівня, яка проходить через точку $A(18; 8)$ перетину ліній $2x_1 + 7x_2 = 92$ та $x_1 - 1{,}5x_2 = 6$. Найбільше значення $z = 4{\cdot}18 + 3{\cdot}8 = 96$. Таким чином, знайдено такі значення $x_1 = 18$ та $x_2 = 8$, величини яких задовольняють приведеній системі обмежень, а цільова функція досягає найбільшого значення. Читач може безпосередньо переконатися, що при будь-якій зміні величин $x_1 = 18$ та $x_2 = 8$ або зменшиться значення цільової функції, або не будуть виконуватися вимоги щодо обмежень (22).

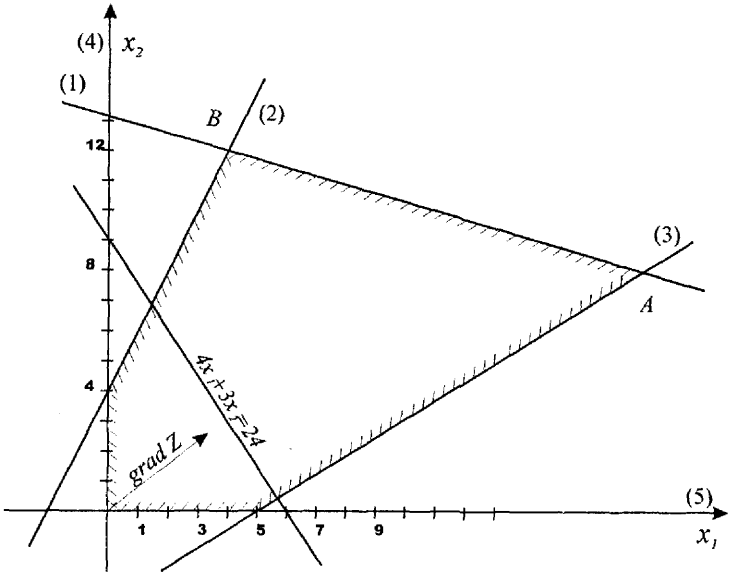


Рис. 5.6.

Проаналізувавши метод графічного розв'язання ЗЛП, доцільно зробити деякі загальні висновки стосовно властивостей її розв'язків, які зручно тлумачити геометрично.

1. Якщо існує розв'язок ЗЛП, то точка або точки, які йому відповідають, обов'язково розташовані на межі області допустимих розв'язків. Якби точки були розташовані в середині цієї області, то розташування лінії рівня можна було б змінити так, щоб значення цільової функції збільшилось (зменшилось).

2. Задача лінійного програмування буде мати нескінченну множину розв'язків, якщо лінія рівня цільової функції, на якій досягається найбільше значення, паралельна одній із сторін багатокутника, яким зображається область допустимих розв'язків.

Це так звана *ситуація альтернативного оптимуму*. Якби у щойно розглянутому прикладі (22) цільова функція була представлена формулою

$$z = 2x_1 + 7x_2, \tag{23}$$

то за такої умови відповідна ЗЛП мала б нескінченну множину розв'язків з геометричним зображенням їх на відрізку BA, розташованому на прямій $2x_1 + 7x_2 = 92$.

Розглянемо координати точок, розташованих на відрізку BA. Координати їх пов'язані залежністю $x_2 = \frac{92 - 2x_1}{7}$ за умови, що $x_1 \in [4; 18]$. Значення лівої межі відрізку одержане як результат обчислення координат точки B, яка є перетином прямих (1) та (2) умови (22) – рис. 5.6.

Переконаємося, що при $x_1 \in [4; 18]$ та $x_2 = \frac{92 - 2x_1}{7}$ будуть виконуватися умови (22), а цільова функція (23) має одне й те ж значення для всіх точок відрізку BA. Підставляємо x_1 та x_2 почергово в умови (22):

(1) $2x_1 + 7x_2 = \left| x_2 = \frac{92 - 2x_1}{7} \right| = 92 = 92;$

(2) $-2x_1 + x_2 = \left| x_2 = \frac{92 - 2x_1}{7} \right| = \frac{92 - 16x_1}{7} \le 4 \Rightarrow x_1 \ge 4;$

(3) $x_1 - 1{,}5x_2 = \left| x_2 = \frac{92 - 2x_1}{7} \right| = \frac{10x_1 - 138}{7} \le 6 \Rightarrow x_1 \le 18;$

(4) $x_1 > 0;$

(5) за умови, що $x_2 = \frac{92 - 2x_1}{7}$ при $x_1 \in [4; 18]$, $x_2 \in [8; 12] \Rightarrow x_2 > 0$.

Отже, умови (22) виконуються. Обчислимо значення цільової функції (23):

$$z = 2x_1 + 7\left(\frac{92 - 2x_1}{7}\right) = 92.$$

У тому, що значення $z = 92$ є найбільшим за умов (22) та (23), легко переконатися: відрізок BA є частиною найбільш віддаленої від початку координат лінії рівня функції (23) і має нескінченну множину точок.

3. Задача лінійного програмування може не мати жодного розв'язку. Така ситуація буде за таких умов:

а) область допустимих розв'язків пуста множина, тобто обмеження на допустимі значення змінних величин сформульовані так, що немає такої множини значень змінних, які б задовольняли всім обмеженням;

б) цільова функція не обмежена в області допустимих розв'язків.

Наприклад, знайти найбільше значення цільової функції

$$z = 4x_1 + 5x_2$$

за умови

$$\begin{cases} 5x_1 - 3x_2 \le 15 & (1) \\ -2x_1 + x_2 \le 2 & (2) \\ x_1 \ge 0 & (3) \\ x_2 \ge 0. & (4) \end{cases}$$

Область допустимих розв'язків, розташування однієї з ліній рівня цільової функції та *grad z* зображені на рис. 5.7.

Аналіз графіка показує, що із зростанням x_1 та x_2 значення цільової функції зростають, а оскільки, за умовою задачі, значення x_1 та x_2 необмежені, то і значення цільової функції цього прикладу будуть необмежено зростати.

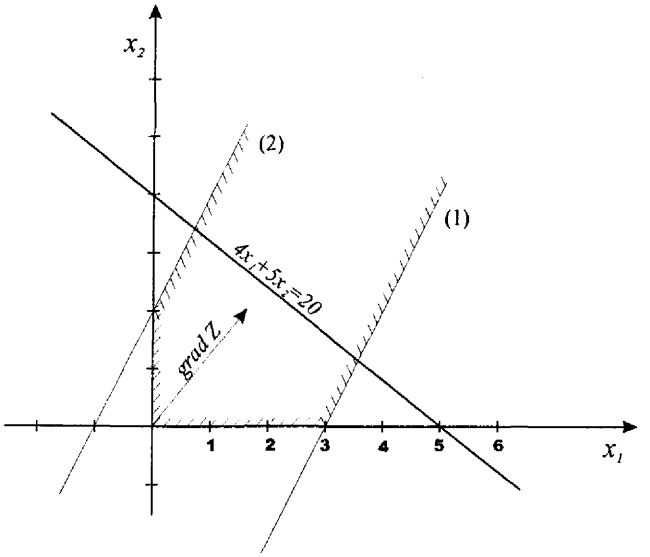


Рис. 5.7.

Зауваження. Графічним методом можна розв'язувати ЗЛП у симетричній формі за умови наявності трьох змінних, але відповідні побудови у тримірному просторі досить складні та громіздкі.

Графічний метод розв'язання ЗЛП є простим і наочним, але дуже обмеженим за своїми можливостями і не дозволяє з'ясовувати економічний зміст практичних задач. Тому розроблено аналітичні методи, основи яких будуть описані в наступних главах.

Питання для самоконтролю

1. Приклади задач лінійного програмування з практичною орієнтацією.
2. Характерні особливості ЗЛП.
3. Форми запису ЗЛП.
4. Відмінності у запису ЗЛП у стандартній та канонічній формах.
5. Що таке лінія рівня функції двох змінних та *grad*?
6. Характерні особливості лінії рівня лінійної функції та її градієнта.
7. Умови графічного розв'язання ЗЛП.
8. Основні етапи графічного розв'язання ЗЛП.
9. Графічне тлумачення можливих ситуацій у розв'язанні ЗЛП.
10. Альтернативний оптимум та його графічне тлумачення.

ГЛАВА 6

НЕОБХІДНІ ВІДОМОСТІ З ЛІНІЙНОЇ АЛГЕБРИ

6.1. Числа, вектори, *n*-мірний простір Евкліда

Поняття "число" виникло в незапам'ятні часи як один із засобів орієнтації в реальному світі і було обумовлене необхідністю підраховувати кількість людей, тварин, плодів, виробів тощо.

Стародавні сумери (шумери) Месопотамії вже шість тисячоліть тому мали високу культуру виконання розрахунків. Користувалися вони *позиційною* шістдесятирічною системою числення, ввели тижневий цикл у сім днів (за кількістю кольорів видимого спектра сонячного світла), добу розділили на 24 години, годину на 60 хвилин, хвилину на 60 секунд, коло на 360 градусів і т. ін.

Число є не лише показником кількісної міри за вибраної одиниці виміру, а й засобом порівняння певних характеристик. Саме з такою метою введено операції відношення: ≤, =, ≥ (не більше, дорівнює, не менше). Таку двоїстість сприйняття числа відмічав у своїй "Політичній арифметиці" один із засновників класичної політичної економії У. Петті (1623-1687): замість того, щоб використовувати слова в порівняльному та вищому ступені та спиратися на уявні аргументи, він став на шлях відтворення своїх думок мовою чисел, ваги та мір.

Поняття "вектор", без заперечень, виникло у фізиці і було обумовлене необхідністю відображати такі величини, для яких важливе значення мала не тільки певна кількісна міра, а й їх просторова орієнтація: переміщення, сила, швидкість і т. ін. У різних науках це поняття визначається в залежності від його цільового призначення. Для розбудови математичних моделей в економіці доцільно ввести поняття вектора як упорядкованої сукупності певної (n) кількості чисел ($a_1, a_2, ..., a_n$), стосовно якої визначаються певні правила дій. Числа ($a_1, a_2, ..., a_n$) називаються *компонентами* вектора; якщо йдеться про його геометричне тлумачення, то вони називаються *проекціями*. Упорядкованість сукупності означає, що значимість числа залежить не лише від його величини, а й від місця цього числа в сукупності. Наприклад, зиск a_i від реалізації одиниці кожного виду сільськогосподарської продукції доцільно представити так упорядкованим записом (вектором):

a_1	a_2	a_3	a_4	...	a_{n-1}	a_n
пшениця	жито	просо	горох	...	кукурудза	капуста

(n – контрольована кількість найменувань).

Якщо вектор має одну, дві або три компоненти, то одним із можливих його

тлумачень буде направлений відрізок. Зображення відповідного відрізку виконується так: вибирають відповідну систему координат (на прямій, площині або просторі), як в аналітичній геометрії; в цій системі будують точку з координатами, рівними компонентам вектора, і початок координат з'єднують відрізком з побудованою точкою, приймаючи, що відрізок орієнтований завжди від початку координат.

Якщо компонент вектора більше трьох, то наочне геометричне тлумачення таких векторів відсутнє, але зберігається геометрична термінологія, яка була розбудована за умови, що кількість компонент вектора не перевищує трьох.

Вектори вважаються *рівними* (тотожними) в тому і лише в тому випадку, якщо їх компоненти, розташовані на одних і тих же місцях, дорівнюють одна одній, тобто $\bar{a} = \bar{b}$, якщо $a_i = b_i$ для всіх $i = \overline{1, n}$. Множина векторів з n-компонентами, за умови, що кожна з компонент може прийняти будь-яке числове значення, називається *n-мірним простором*, а кількість компонент, використаних при записі вектора, – *розмірністю вектора* або *простору*.

Абстрагуючись, як це прийнято в математиці, від конкретного змісту та фізичного втілення понять, які вводяться для використання, залишають лише найбільш суттєві їх властивості, а саме правила виконання операцій над ними, щоб мати можливість використовувати їх у математичних викладках. Ці правила розбудовані так, щоб віддзеркалити певні реальні взаємозв'язки між тими об'єктами, яким відповідають вектори. Наведемо допустимі дії над векторами, даючи їм геометричне й економічне тлумачення.

1. *Добутком вектора на число* називається вектор, компоненти якого дорівнюють компонентам вихідного вектора, помноженим на дане число.

Якщо $\bar{a} = (a_1, a_2, ..., a_n)$, то $\lambda\bar{a} = (\lambda a_1, \lambda a_2, ..., \lambda a_n)$. Геометричне тлумачення: довжина вихідного вектора змінюється в λ разів, просторова орієнтація результативного вектора зберігається, якщо $\lambda > 0$, та змінюється на протилежну за умови $\lambda < 0$. Якщо компонентами вектора будуть вартості одиниць продукції певних найменувань, то множення такого вектора на число можна тлумачити як зміну відповідних вартостей за умови зміни вартості використання грошової одиниці (девальвація або ревальвація).

Операція множення вектора на число має такі властивості:

а) $\alpha(\beta\bar{a}) = (\alpha\beta)\bar{a}$ – асоціативний закон стосовно числових множників;

б) $\alpha(\bar{a} + \bar{b}) = \alpha\bar{a} + \alpha\bar{b}$; $(\alpha + \beta)\bar{a} = \alpha\bar{a} + \beta\bar{a}$ – дистрибутивний закон.

2. *Сумою двох векторів* $\bar{a} = (a_1, a_2, ..., a_n)$ та $\bar{b} = (b_1, b_2, ..., b_n)$ називається вектор $\bar{c}$, компоненти якого дорівнюють сумі відповідних компонент векторів-доданків, тобто $\bar{c} = \bar{a} + \bar{b} = (a_1 + b_1, a_2 + b_2, ..., a_n + b_n)$.

Для операції додавання векторів виконуються властивості:

а) $\bar{a} + \bar{b} = \bar{b} + \bar{a}$ – комутативний закон;

б) $\bar{a} + \bar{b} + \bar{c} = (\bar{a} + \bar{b}) + \bar{c} = \bar{a} + (\bar{b} + \bar{c})$ – асоціативний закон;

в) існує єдиний так званий нульовий вектор $\bar{0} = (0_1, 0_2, ..., 0_n)$, для якого неза-

лежно від вектора $\bar{a}$ виконується умова $\bar{a} + \bar{0} = \bar{a}$;

г) кожному вектору $\bar{a}$ n-мірного простору відповідає *протилежний* вектор $-\bar{a}$, для якого виконується умова $\bar{a} + (-\bar{a}) = \bar{0} = (0, 0, ..., 0)$.

Якщо вектори тлумачити як орієнтовані відрізки у просторі, то операції додавання векторів відповідає так зване *правило паралелограма*, згідно з яким у просторі необхідно побудувати неперервно розташовані вектори-доданки, тоді сума буде зображена орієнтованим відрізком, початок якого співпадає з початком першого вектора-доданка, а кінець – з кінцем другого вектора-доданка.

За умови, що вектори-доданки мають економічне тлумачення (наприклад, наявний машинно-тракторний парк двох різних сільськогосподарських підприємств), то вектор суми буде визначати машинно-тракторний парк при їх об'єднанні з урахуванням типів машин, кількість яких представлена відповідними компонентами.

3. *Скалярний добуток векторів.* Кожній парі векторів $\bar{a} = (a_1, a_2, ..., a_n)$ та $\bar{b} = (b_1, b_2, ..., b_n)$ ставиться у відповідність число, яке називається їх скалярним добутком $\bar{a}\,\bar{b}$ і обчислюється через їх компоненти так: $a_1b_1 + a_2b_2 + ... + a_nb_n$, тобто $\bar{a}\,\bar{b} = \sum_{i=1}^{n} a_i b_i$. Таким чином, скалярний добуток двох векторів дорівнює сумі добутків їх однойменних компонент. Якщо перший множник тлумачити як силу, що діє на матеріальну точку, а другий – як її переміщення у просторі під дією сили, то скалярний добуток визначає величину роботи, виконаної даною силою на даному переміщенні. За умови, що компоненти першого вектора визначають обсяги продажу продукції певних найменувань, а компоненти другого множника – вартості одиниць продукції тих же найменувань, то скалярний добуток визначає зиск від реалізації за наведених умов.

Для операції скалярного добутку двох векторів виконується властивість:

а) $\bar{a}\,\bar{b} = \bar{b}\,\bar{a}$ – комутативний закон;

б) $(\alpha\bar{a} + \beta\bar{b})\bar{c} = \alpha\bar{a}\,\bar{c} + \beta\bar{b}\,\bar{c}$ для будь-яких чисел α і β та будь-яких векторів з n-мірного простору;

в) $\bar{a}\,\bar{a} \geq 0$ і $\bar{a}\,\bar{a} = 0$ лише за умови, що $\bar{a} = \bar{0} = (0_1, 0_2, ..., 0_n)$. Векторний простір, в якому визначено скалярний добуток за вищеназваними умовами, називають евклідовим векторним простором.

В аналітичній геометрії, фізиці використовуються й інші операції над векторами, але для наших викладок вони не потрібні.

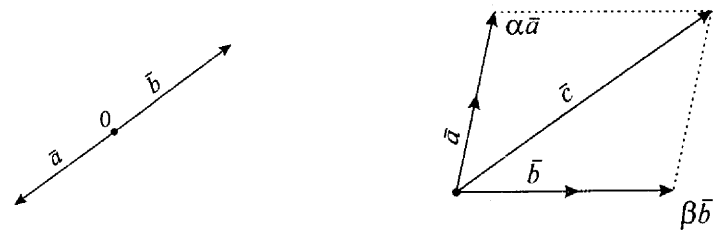
Для розуміння сенсу багатьох дій та результатів в лінійній алгебрі суттєвим є питання *лінійної залежності векторів*. Вектор $\bar{b}$ називається лінійною комбінацією векторів $\bar{a}_1, \bar{a}_2, ..., \bar{a}_m$, якщо існують такі числа $\lambda_1, \lambda_2, ..., \lambda_m$, що $\bar{b} = \lambda_1\bar{a}_1 + \lambda_2\bar{a}_2 + ... + \lambda_m\bar{a}_m$. У цьому випадку будь-яка i-та компонента вектора $\bar{b}$ дорівнює сумі добутків відповідних компонент векторів $\bar{a}_1, \bar{a}_2, ..., \bar{a}_m$ на числа $\lambda_1, \lambda_2, ..., \lambda_m$, тобто $b_i = \lambda_1 a_{1i} + \lambda_2 a_{2i} + ... + \lambda_m a_{mi} = \sum_{k=1}^{m} \lambda_k a_{ki}$ $(i = \overline{1, m})$.

Система векторів $\bar{a}_1, \bar{a}_2, ..., \bar{a}_m$ називається *лінійно залежною*, якщо існують такі числа $\lambda_1, \lambda_2, ..., \lambda_m$, серед яких хоча б одне відмінне від нуля, тобто $\sum_{i=1}^{m} \lambda_i^2 \neq 0$, для яких виконується рівність $\lambda_1\bar{a}_1 + \lambda_2\bar{a}_2 + ... + \lambda_m\bar{a}_m = 0$.

Якщо наведена рівність можлива лише за умови $\sum_{i=1}^{m} \lambda_i^2 = 0$ (тобто немає ненульових значень λ_i), то вектори $\bar{a}_1, \bar{a}_2, ..., \bar{a}_m$ є лінійно незалежними. Якщо розглядати двомірні вектори і дати їм геометричне тлумачення, то легко пояснити поняття лінійної та нелінійної залежностей. Якщо два вектори $\bar{a}$ та $\bar{b}$ колініарні, то вони завжди лінійно залежні (рис. 6. 1, а); дійсно, завжди в такій ситуації $\bar{b} = \lambda\bar{a}$, і, розглядаючи рівність $\alpha\bar{a} + \beta\bar{b} = (\alpha + \lambda\beta)\bar{a}$, досить вибрати відношення $\frac{\alpha}{\beta} = -\lambda$, щоб одержати нульовий вектор. Три вектори на площині будуть також завжди лінійно залежними, а два неколініарних вектори, кожний з яких відмінний від нуля, будуть лінійно незалежними. Візьмемо три будь-які вектори $\bar{a}$, $\bar{b}$, $\bar{c}$ на площині. Один з них, наприклад, $\bar{c}$ можна представити лінійною комбінацією двох інших (рис. 6.1 б): $\bar{c} = \alpha\,\bar{a} + \beta\,\bar{b}$; α і β визначаються через вектори $\bar{a}$, $\bar{b}$, $\bar{c}$, а лінійна комбінація

$$\lambda\bar{a} + \delta\bar{b} + \gamma\bar{c} = \lambda\bar{a} + \delta\bar{b} + \alpha\gamma\bar{a} + \beta\gamma\bar{b} = (\lambda + \gamma\alpha)\bar{a} + (\delta + \beta\gamma)\bar{b}$$

буде дорівнювати нульовому вектору за умови $\lambda = -\gamma\alpha$, $\delta = -\beta\gamma$. Отже, два неколініарних вектори на площині будуть завжди лінійно незалежними, а три – завжди лінійно залежними.



а) б)

Рис. 6.1.

Базис *n*-мірного векторного простору. Для розбудови алгоритмів розв'язання задач часто доцільно використати поняття базису векторного простору.

Базисом n-мірного векторного простору називається система n лінійно незалежних n-мірних векторів. У лінійній алгебрі доведено, що *в n-мірному векторному просторі можна завжди вибрати базис, до того ж множина базисів нескінченна.* Але будь-яка система n-мірних векторів у кількості $n+1$ буде завжди лінійно залежною. Наведемо приклади базисів (системи лінійно незалежних векторів):

$$\text{а)}\begin{cases}\bar{e}_1=(1;0;0;...;0),\\ \bar{e}_2=(0;1;0;...;0),\\ \cdot\\ \cdot\\ \cdot\\ \bar{e}_n=(0;0;0;...;1).\end{cases}\qquad \text{б)}\begin{cases}\bar{a}_1=(1;0;0;...;0),\\ \bar{a}_2=(1;1;0;...;0),\\ \cdot\\ \cdot\\ \cdot\\ \bar{a}_n=(1;1;1;...;1).\end{cases}$$

Якщо в n-мірному векторному просторі вибрано базис (множина n лінійно незалежних векторів $\{\bar{a}_1, \bar{a}_2, ..., \bar{a}_n\}$), то будь-який вектор $\bar{P}$ такого простору однозначно можна зобразити лінійною комбінацією векторів базису, тобто $\bar{P}=\lambda_1\bar{a}_1+\lambda_2\bar{a}_2+...+\lambda_n\bar{a}_n$. На рис. 6.2 зображено розкладання вектора $\bar{P}$ за векторами $\bar{a}_1, \bar{a}_2, \bar{a}_3$ базису в тримірному просторі.

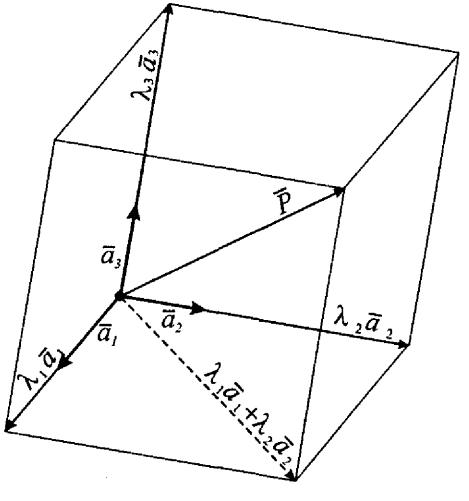


Рис. 6.2.

Питання розкладання будь-якого вектора за векторами базису та заміни базису розглянемо далі.

6.2. Матриці та визначники

Матрицею називається сукупність m×n чисел (елементів матриці), упорядкованих у m-рядків і n-стовпців. Запис $m\times n$ означає розмірність матриці. Матриці позначаються великими латинськими літерами, а їх елементи – відповідними малими літерами з двома індексами: перший індекс визначається номером рядка, а другий – номером стовпця, в яких розташований елемент, тобто якщо матрицю позначено літерою A, то її елементи позначаються як a_{ij}. Записують матриці в один із таких способів:

$$A=\begin{pmatrix}a_{11}a_{12}...a_{1n}\\ a_{21}a_{22}...a_{2n}\\ .....\\ a_{m1}a_{m2}...a_{mn}\end{pmatrix},\ \text{або}\ \begin{bmatrix}a_{11}a_{12}...a_{1n}\\ a_{21}a_{22}...a_{2n}\\ .....\\ a_{m1}a_{m2}...a_{mn}\end{bmatrix},\ \text{або}\ \begin{Vmatrix}a_{11}a_{12}...a_{1n}\\ a_{21}a_{22}...a_{2n}\\ .....\\ a_{m1}a_{m2}...a_{mn}\end{Vmatrix}$$

Використовують також скорочений запис: $[A]$, $\|A\|$, $\|a_{ij}\|$. Кількість рядків m та кількість стовпців n визначають розмірність матриці, яка позначається як $m\times n$.

Приклади доцільності використання поняття матриці.

Маємо m пунктів постачання та n пунктів споживання. Відстані між ними доцільно записати у вигляді матриці, де кожний елемент d_{ij} визначає відстань від i-го пункту постачання до j-го пункту споживання $(i=\overline{1,m};\ j=\overline{1,n})$. За умови, що відомі c_{ij} тарифи перевезення від i-го пункту постачання в j-й пункт споживання, при необхідності можна записати матрицю тарифів $\|c_{ij}\|$.

Використання термінів. Матриця називається *прямокутною*, якщо $m\neq n$. За умови $m=n$ матриця називається *квадратною*, а кількість рядків (стовпців) визначають її порядок. Матриця, що має лише один ряд, тобто один рядок або один стовпець, називається відповідно *матриця-рядок* та *матриця-стовпець*. Слушно наголосити, що вектор у багатьох випадках доцільно розглядати як матрицю одного ряду.

Елементи квадратної матриці, розташовані на місцях з однаковими номерами рядків і стовпців, називаються *діагональними*, тобто a_{ii} $(i=\overline{1,n})$, а решта елементів – позадіагональними.

Елементи a_{ii} $(i=\overline{1,n})$ утворюють головну діагональ, $a_{i,n+1-i}$ – побічну.

Матриця називається *нульовою*, якщо всі її елементи дорівнюють нулю. Нену-

льова матриця називається *діагональною,* якщо її елементи, розташовані поза головною діагоналлю, теж дорівнюють нулю.

Трикутною називається квадратна матриця, всі елементи якої вище або нижче головної діагоналі дорівнюють нулю.

Матриця, одержана з даної матриці A (квадратної або прямокутної) перестановкою розташування рядків і стовпців, називається *транспонованою* і позначається A^T.

За умови $A = A^T$ матриця A називається *симетричною.*

Розглянемо дії над матрицями. Порівнювати можна матриці лише однакової розмірності, тобто з однаковою кількістю рядків і стовпців. Дві матриці A та B вважаються рівними, якщо є рівними їх елементи, розташовані на однакових місцях, тобто за умови $a_{ij} = b_{ij}$ $(i = \overline{1,m};\ j = \overline{1,n})$.

Добутком матриці A на число λ називається матриця B тієї ж розмірності з елементами $b_{ij} = \lambda a_{ij}$ $(i = \overline{1,m};\ j = \overline{1,n})$, тобто кожний елемент вихідної матриці множиться на число λ. Названа операція має такі властивості:

1) $\lambda(\beta A) = (\lambda\beta)A$; 2) $(\lambda + \beta)A = \lambda A + \beta A$; 3) $\lambda(A + B) = \lambda A + \lambda B$.

Операція множення матриці на число може мати таке економічне тлумачення: якщо маємо матрицю тарифів перевезення вантажів із m пунктів постачання в n пунктів споживання, то зміна тарифів за умови зміни вартості грошової одиниці буде подана як результат множення вихідної матриці тарифів на відповідний коефіцієнт.

Додавання матриць. Сумою двох матриць A та B однакових розмірностей є матриця C тієї ж розмірності, елементи якої обчислюються за формулою $c_{ij} = a_{ij} + b_{ij}$ $(i = \overline{1,m};\ j = \overline{1,n})$, тобто додаються елементи матриць-доданків, розташовані на одних і тих же місцях, і результат записується в матриці суми на таке ж місце. Для введеної операції виконуються закономірності:

1) $A + B = B + A$ – комутативний закон;

2) $A + B + C = (A + B) + C = A + (B + C)$ – асоціативний закон;

3) $\lambda(A + B) = \lambda A + \lambda B$; $(\lambda + \beta)A = \lambda A + \beta A$ – дистрибутивний закон.

Операція додавання матриць може мати, наприклад, таке тлумачення. Якщо маємо систему постачання n найменувань виробів, в якій використовується m складів, то для моделювання такої системи функціонування зручно використовувати дві матриці: матрицю поточного стану системи з елементами a_{ij} – обсягами запасів на i-му складі j-го виробу та матрицю поновлення запасів виробів на складах з елементами b_{ij} – обсягами чергових поставок на i-й склад j-го виробу; сума цих матриць характеризуватиме зміну поточного стану запасів на i-му складі стосовно j-го виробу після чергової поставки.

Множення матриць. Операція множення матриць A та B визначається лише стосовно так званих *узгоджених матриць*, коли кількість стовпців першого співмножника дорівнює кількості рядків другого співмножника, тобто якщо A має розмірність $m \times n$, то B мусить мати розмірність $n \times p$.

За названих умов добутком матриць A на B називають матрицю C розмірності $m \times p$, елементи c_{ij} якої обчислюються за формулою:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \ldots + a_{in}b_{nj} \quad (i = \overline{1,m};\ j = \overline{1,n}).$$

Наведену формулу обчислення елементів матриці C можна прокоментувати так: щоб знайти елемент матриці добутку, розташований на місці перетину i-го рядка та j-го стовпця, необхідно i-й рядок матриці A помножити *скалярно* на j-й стовпець матриці B.

Одне з економічних тлумачень операції множення двох матриць може бути таким. Перший співмножник – матриця тарифів перевезення одиниці продукції з i-го пункту постачання у k-й пункт споживання $(i = \overline{1,m}; k = \overline{1,n})$. Другий співмножник – матриця обсягів j-ої продукції, яку потрібно завезти у k-й пункт споживання. За таких умов елементи матриці добутку названих матриць будуть визначати витрати на завезення в усі пункти споживання j-ої продукції в необхідних обсягах з i-го пункту постачання.

Для операції множення двох матриць *переставний закон* стосовно співмножників *не виконується*, тобто $AB \neq BA$, навіть коли матриці-співмножники квадратні.

Приклади. Перемножити A і B, якщо

$$A = \begin{bmatrix} -2 & 3 & 5 \\ 1 & 0 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 2 \\ -2 & 1 \\ 1 & 3 \end{bmatrix}.$$

$$C = AB = \begin{bmatrix} (-2)\cdot 3 + 3\cdot(-2) + 5\cdot 1 & (-2)\cdot 2 + 3\cdot 1 + 5\cdot 3 \\ 1\cdot 3 + 0\cdot(-2) + (-3)\cdot 1 & 1\cdot 2 + 0\cdot 1 + (-3)\cdot 3 \end{bmatrix} = \begin{bmatrix} -7 & 14 \\ 0 & -7 \end{bmatrix}.$$

Виконаємо множення B на A:

$$D = BA = \begin{bmatrix} 3\cdot(-2) + 2\cdot 1 & 3\cdot 3 + 2\cdot 0 & 3\cdot 5 + 2\cdot(-3) \\ (-2)\cdot(-2) + 1\cdot 1 & (-2)\cdot 3 + 1\cdot 0 & (-2)\cdot 5 + 1\cdot(-3) \\ 1\cdot(-2) + 3\cdot 1 & 1\cdot 3 + 3\cdot 0 & 1\cdot 5 + 3\cdot(-3) \end{bmatrix} = \begin{bmatrix} -4 & 9 & 9 \\ 5 & -6 & -13 \\ 1 & 3 & -4 \end{bmatrix}.$$

Порівняння результатів AB та BA свідчить про їх нетотожність.

У випадку, коли виконується умова $AB = BA$, матриці A та B називаються *переставними.*

Операція множення матриць має такі властивості:

1) $AB \neq BA$ – за винятком, коли матриці A та B переставні;

2) $A(BC) = (AB)C$ – асоціативний закон;

3) $A(B + C) = AB + AC$; $(A + B)C = AC + BC$ – дистрибутивний закон;

4) $(AB)^T = B^T \cdot A^T$ – операція транспонування добутку матриць тотожна операції добутку переставлених транспонованих співмножників.

Визначники. Кожній *квадратній числовій матриці* n-го порядку однозначно ставиться у відповідність число, яке називають *визначником* (*детермінантом*) n-го порядку і величину якого обчислюють за певними правилами через елементи матриці. Як зазначалося раніше, для позначення того, що певна таблиця чисел розглядається як матриця, стосовно таблиці використовують символи (•); [•] або ‖•‖, де "•" відповідає таблиці. Для позначення того, що стосовно таблиці чисел необхідно розглядати її визначник, використовують символ |•|, а визначник позначають так: Δ, det A, $|A|$, або $|a_{ij}|$ (A відповідна матриця). Тобто якщо маємо матрицю

$$A = \begin{bmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \dots\dots & \dots\dots \\ a_{n1} & a_{n2} \dots a_{nn} \end{bmatrix},$$

то вимога розглядати визначник матриці A буде записана в один із таких способів:

$$\det A = \Delta_n = |A| = |a_{ij}| = \begin{vmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \dots\dots & \dots\dots \\ a_{n1} & a_{n2} \dots a_{nn} \end{vmatrix}.$$

Наголосимо на двоїстості використання терміну "визначник": іноді ним називають таблицю, на підставі якої обчислюють його, а іноді відповідне число.

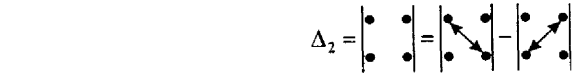
Числа a_{ij} $(i = \overline{1,n};\ j = \overline{1,n})$ називають елементами визначника. Елементи a_{ii} $(i = \overline{1,n})$ утворюють *головну* діагональ визначника, а елементи $a_{k,\,n-k+1}$ $(k = \overline{1,n})$ – *побічну* діагональ.

У курсі вищої алгебри вказано правило обчислення визначників будь-якого порядку. Наведемо це правило для визначників 2-го та 3-го порядків:

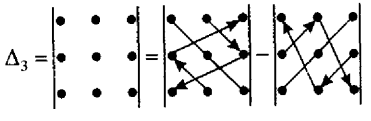
$$\Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} \cdot a_{22} - a_{12} \cdot a_{21};$$

$$\Delta_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

Правила обчислення названих визначників схематично можна зобразити в такий спосіб. Визначники 2-го порядку:



Визначники 3-го порядку:



Елементи визначників, які входять в один добуток, з'єднані відрізками. Знак "–" означає, що у відповідних добутках необхідно змінити результат їх обчислення на протилежний.

Приклад.

$$\Delta_3 = \begin{vmatrix} 3 & 7 & -2 \\ 4 & 5 & 1 \\ 0 & -3 & 2 \end{vmatrix} = 3\cdot5\cdot2+7\cdot1\cdot0+(-2)\cdot4\cdot(-3)-((-2)\cdot5\cdot0+1\cdot(-3)\cdot3+7\cdot4\cdot2) =$$

$$= 30+0+24-0+9-56=7.$$

Мінором називають визначник, одержаний з даного після вилучення у будь-якому порядку певної кількості рядків і такої ж кількості стовпців. Особливе значення мають для визначника n-го порядку мінори $(n-1)$-го порядку.

Якщо у визначнику вилучити i-й рядок та j-й стовпець, то одержимо визначник $(n-1)$-го порядку. Такий визначник називається мінором M_{ij}, що відповідає елементу a_{ij}.

Але термін „мінор" використовується і за інших умов. Якщо маємо матрицю A (таблицю чисел), то іноді доцільно розглядати визначники певного порядку, складені з елементів матриці A так, щоб зберігалося відповідне розташування взятих еле-

ментів, тобто якщо елементи були розташовані в певному ряду (рядку або стовпці) матриці A, то при складанні визначника вони мають бути в тому ж ряду. Так складені визначники називаються *мінорами матриці A*.

Алгебраїчним доповненням A_{ij}, що відповідає елементу a_{ij} визначника, називають мінор M_{ij}, помножений на $(-1)^{i+j}$:

$$A_{ij} = M_{ij}(-1)^{i+j}.$$

Основна властивість визначників. Визначник дорівнює сумі добутків елементів будь-якого ряду (рядка або стовпця) матриці на їх алгебраїчні доповнення. Цією властивістю можна користуватися для обчислення визначників будь-якого порядку, знижуючи їх порядок.

6.3. Системи лінійних алгебраїчних рівнянь

При розв'язанні прикладних задач часто виникають системи лінійних алгебраїчних рівнянь (СЛАР). Системою m лінійних алгебраїчних рівнянь з n невідомими називається система

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases}, \qquad (1)$$

де a_{ij} $(i = \overline{1,m};\ j = \overline{1,n})$ – відомі числа, які називаються коефіцієнтами системи; b_1, b_2, ..., b_m – відомі числа, які називаються вільними членами; x_1, x_2, ..., x_n – невідомі (шукані) величини.

Якщо всі b_i $(i = \overline{1,m})$ дорівнюють нулю, то система називається *однорідною*, в протилежному випадку – *неоднорідною*.

Матриця $A = \|a_{ij}\|$ $(i = \overline{1,m};\ j = \overline{1,n})$ розмірності $m \times n$ називається *матрицею коефіцієнтів СЛАР*.

Якщо до матриці A дописати справа стовпець вільних членів, то одержимо так звану *розширену матрицю коефіцієнтів* СЛАР, яку запишемо $\|A \vdots \bar{b}\|$ або $\|a_{ij} \vdots b_i\|$.

Розв'язком СЛАР (1) називається впорядкований набір чисел $(x_1, x_2, \ldots, x_n)$, який при підстановці у систему перетворює всі її рівняння на тотожності. СЛАР, яка має хоча б один розв'язок, називається *сумісною*, якщо розв'язку немає – то *несумісною*. СЛАР, яка має лише єдиний розв'язок, називається *визначеною*. При наявності

нескінченної множини розв'язків система називається *невизначеною*.

Дві системи лінійних алгебраїчних рівнянь називаються *тотожними* (*еквівалентними*), якщо множини їх розв'язків співпадають. Розв'язання системи (1) будується шляхом так званих *елементарних перетворень*, за яких вихідна та одержувана системи тотожні: 1) множення обох частин будь-якого рівняння на число, відмінне від нуля; 2) додавання до будь-якого рівняння іншого рівняння системи, помноженого на відмінне від нуля число; 3) переставляння двох рівнянь; 4) зміни послідовності нумерації невідомих.

Процедурі розв'язання СЛАР доцільно дати таке тлумачення. Множини вільних членів і шуканих невідомих, упорядковані за їх номерами, будемо розглядати як вектори $\bar{b}$ розмірності m та $\bar{x}$ розмірності n або матриці, що мають лише один ряд. За таких умов система (1) запишеться так:

$$A\bar{x} = \bar{b}. \qquad (2)$$

Якщо кожний стовпець матриці A розглядати як вектор $\bar{a}_j$ m-мірного простору, тобто $\bar{a}_j = (a_{1j}, a_{2j}, \ldots, a_{mj})$, то систему (1) можна записати так:

$$x_1\bar{a}_1 + x_2\bar{a}_2 + \ldots + x_n\bar{a}_n = \bar{b}, \text{ або } \sum_{j=1}^{n} x_j\,\bar{a}_j = \bar{b}. \qquad (3)$$

Визначальним для розв'язання СЛАР є поняття рангу матриці.

Рангом матриці називається найбільший порядок відмінних від нуля мінорів матриці. Тобто серед мінорів матриці повинен існувати хоча б один мінор порядку r, відмінний від нуля; всі мінори порядку $r+1$ та вище дорівнюють нулю або не можуть бути складені. Ранг матриці A позначається через $r\|A\|$ (іноді записують $rang\,\|A\|$).

У лінійній алгебрі доведено теорему, що *максимальна кількість лінійно незалежних стовпців матриці коефіцієнтів A СЛАР, а також максимальна кількість її лінійно незалежних рядків дорівнюють рангу матриці.*

Таким чином, якщо розширена матриця СЛАР $\|a_{ij} \vdots b_i\|$ має ранг r за умови, що $r \le m$ $(m \le n)$, то система має r лінійно незалежних рівнянь, а решта $m - r$ рівнянь є лінійно залежними. Така вихідна система m рівнянь шляхом елементарних перетворень може бути приведена до системи r лінійно незалежних рівнянь.

Стосовно розв'язання СЛАР (1) або (2) доведено теорему Кронекера-Капеллі. Для того, щоб система (1) мала розв'язок, необхідно і достатньо, щоб ранг матриці коефіцієнтів дорівнював рангу розширеної матриці:

$$r\|a_{ij}\| = r\|a_{ij} \vdots b_i\|. \qquad (4)$$

У літературі описано понад півтори тисячі методів розв'язання СЛАР. Кожний з відомих методів має свої переваги та вади в порівнянні з іншими. Розглянемо так званий метод Жордана-Гаусса, який дозволяє одночасно визначити $r\|a_{ij}\|$ та $r\|a_{ij} \vdots b_i\|$ і дати розв'язок СЛАР, якщо вона сумісна, або виявити її несумісність. Шляхом елементарних перетворень СЛАР (1) набуває вигляду тотожної системи:

$$\begin{cases} x_1 + \quad \dots \quad + a^*_{1,r+1}x_{r+1} + a^*_{1,r+2}x_{r+2} + \dots + a^*_{1,n}x_n = b^*_1 \\ \quad x_2 + \dots + a^*_{2,r+1}x_{r+1} + a^*_{2,r+2}x_{r+2} + \dots + a^*_{2,n}x_n = b^*_2 \\ \quad \ddots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ \quad x_r + a^*_{r,r+1}x_{r+1} + a^*_{r,r+2}x_{r+2} + \dots + a^*_{r,n}x_n = b^*_r, \end{cases} \tag{5}$$

коефіцієнти розширеної матриці $\|a^*_{ij} \vdots b^*_i\|$ системи (5) обчислюються через коефіцієнти розширеної матриці $\|a_{ij} \vdots b_i\|$ системи (1) в ході елементарних перетворень СЛАР. Алгоритм перетворень такий. Перший крок: переставляючи при необхідності порядок запису рівнянь в (1), маємо одержати при x_1 коефіцієнт відмінний від нуля. Помножимо перше рівняння на число, обернене до нього. Тоді в так перетвореному першому рівнянні при x_1 матимемо коефіцієнт одиницю. Так перетворене перше рівняння почергово помножимо на відповідний коефіцієнт при x_1 кожного з решти рівнянь системи (1) та віднімемо цей результат від відповідного рівняння. Внаслідок таких перетворень коефіцієнти при x_1 у всіх рівняннях, за винятком 1-го, будуть нульовими.

По завершенні першого кроку, переставляючи в одержаній СЛАР порядок запису рівнянь або змінюючи порядок нумерації невідомих, одержимо при x_2 у другому рівнянні відмінний від нуля коефіцієнт. Використовуючи цей коефіцієнт та друге рівняння, виконаємо перетворення з використанням другого рівняння за схемою, описаною при реалізації першого кроку. По завершенні другого кроку у перетворенні СЛАР всі коефіцієнти другого стовпця, за винятком другого коефіцієнта, дорівнюватимуть нулю. Виведемо формули реалізації одного кроку перетворень, тобто одержання такої системи рівнянь, у якій у певному стовпці при вибраному невідомому всі коефіцієнти, за винятком одного, дорівнюють нулю. Припустимо, маємо намір деяку невідому x_p виключити з усіх рівнянь СЛАР, за винятком q-го рівняння. Коефіцієнт a_{qp} при x_p називається *провідним*, рівняння та стовпець – *провідними*. Щоб у провідному рівнянні коефіцієнт при x_p дорівнював 1, рівняння необхідно помножити на $1/a_{qp}$, тобто нові коефіцієнти провідного рівняння обчислюються за формулою:

$$a^*_{qj} = \frac{a_{qj}}{a_{qp}} \qquad (j = \overline{1,n}). \tag{6}$$

Щоб одержати в усіх рівняннях системи, за винятком провідного, нульові коефіцієнти, необхідно провідне рівняння з коефіцієнтами (6) помножити на коефіцієнт a_{ip} при x_p поточного i-го рівняння ($i \neq q$) і результат відняти від i-го рівняння; отже, коефіцієнти поточного i-го рівняння обчислюються за формулою:

$$a^*_{ij} = a_{ij} - \frac{a_{qj}}{a_{qp}} a_{ip} \qquad (j = \overline{1,n};\ i = \overline{1,m};\ i \neq q). \tag{7}$$

Аналіз алгоритму показує, що всі перетворення СЛАР пов'язані лише з перетвореннями коефіцієнтів рядків розширеної матриці. Тому на практиці складають таблицю коефіцієнтів розширеної матриці, за стовпцем правих членів відводять місце для запису ще одного стовпця для контролю правильності виконання арифметичних дій. Назвемо таку таблицю *таблицею реалізації алгоритму*. У позиції контролюючого стовпця записують алгебраїчні суми коефіцієнтів відповідного рядка. Над елементами цього стовпця виконують дії за тими ж формулами, що й для решти елементів відповідного рядка. По завершенні одного кроку реалізації алгоритму (тобто для певного стовпця коефіцієнти при відповідному невідомому, за винятком провідного рядка, дорівнюють нулю) додаємо усі коефіцієнти кожного рядка без урахування контролюючого стовпця. Одержана сума має дорівнювати елементу контролюючого стовпця. Таблицю реалізації алгоритму записуємо за однією й тією ж схемою для кожного кроку.

Правило обчислення коефіцієнтів за формулою (7) зручно описати за "схемою прямокутника". З позиції елемента a_{ij} ($i \neq q$) таблиці реалізації "опускаємо" перпендикуляр на провідні рядок і стовпець з урахуванням позиції провідного елемента й одержуємо "прямокутник" (на рис. 6.3 зображений цятковими лініями). Щоб одержати новий елемент a^*_{ij} на позиції i, j ($i \neq q$), необхідно від елемента a_{ij} відняти добуток елементів, розташованих на бічній діагоналі, який ділиться на величину провідного елемента a_{qp}.

Якщо користуватися вже перерахованими *коефіцієнтами провідного рівняння*, то новий елемент a^*_{ij} з використанням "прямокутника" обчислюється за схемою: від елемента a_{ij} слід відняти величину добутку елементів *бічної* діагоналі. За такої схеми зменшується кількість арифметичних операцій при розв'язанні СЛАР.

Зауваження. При реалізації алгоритму Жордана-Гаусса суттєвою є побудова r-стовпців, у яких усі коефіцієнти, за винятком одного, дорівнюють нулю; розташування таких стовпців у таблиці не має значення.

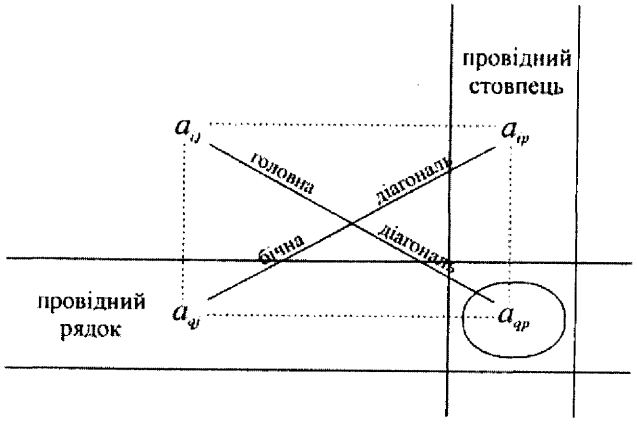


Рис. 6.3.

Приклад. Розв'язати методом Жордана-Гаусса СЛАР або довести її несумісність:

$$\begin{aligned} 3x_1 - x_2 + 4x_3 + x_4 &= 5 \\ x_1 + 2x_2 - 2x_3 - x_4 &= 23 \\ 2x_1 + x_2 + 3x_3 - 3x_4 &= 24 \\ x_1 - x_2 + 5x_3 - 2x_4 &= 1. \end{aligned} \qquad (8)$$

Запишемо вихідну таблицю реалізації алгоритму так, як наведено у верхній частині табл. 6.1; нагадаємо, що правий стовпець введено для контролю.

Таблиця 6.1.

	x_1	x_2	x_3	x_4	b	$\sum_{j=1}^{n} a_{ij}$
	3	–1	4	1	5	12
	1	2	–2	–1	23	23
	2	1	3	–3	24	27
	1	–1	5	–2	1	4
Перший крок	3	–1	4	1	5	12
	4	1	2	0	28	35
	11	–2	15	0	39	63
	7	–3	13	0	11	28

Аналізуючи вихідну таблицю, доходимо висновку, що за провідний елемент можна вибрати a_{14}, відмічений колом. Отже, невідому x_4 виключаємо з усіх рівнянь, крім першого. Для цього почергово перший рядок додаємо до другого; потім перший рядок, помножений на 3, додаємо до третього і на завершення перший рядок, помножений на 2, додаємо до четвертого згідно з (7), враховуючи, що $a_{14} = 1$. Результати названих обчислень наведені внизу табл. 6.1.

Одержавши результати першого кроку реалізації алгоритму, доходимо висновку, що за провідний елемент можна вибрати $a_{22} = 1$. Цей елемент помічений колом. На другому кроці, користуючись (7), виключимо невідому x_2 з усіх рівнянь, окрім другого. Розширена матриця результатів відповідних перетворень та контролюючий стовпець наведені у верхній частині табл. 6.2.

Таблиця 6.2.

	x_1	x_2	x_3	x_4	b	$\sum_{j=1}^{n} a_{ij}$
Другий крок	7	0	6	1	33	47
	4	1	2	0	28	35
	19	0	19	0	95	133
	19	0	19	0	95	133
Третій крок	7	0	6	1	33	47
	4	1	2	0	28	35
	19	0	19	0	95	133
	1	0	1	0	5	7
	0	0	–1	1	–2	–2
	0	1	–2	0	8	7
	0	0	0	0	0	0
	1	0	1	0	5	7

На третьому кроці за провідний елемент виберемо $a_{41} = 19$ (але можна було б вибрати $a_{31} = 19$, або $a_{33} = 19$, або $a_{43} = 19$) та помножимо четвертий рядок на 1/19 і результати запишемо четвертим рядком в середній частині табл. 6.2. Невідому x_1 виключимо з усіх рівнянь, окрім четвертого. Подальші обчислення виконаємо за правилом "чотирикутника" згідно з (7). Результати таких перетворень наведені в нижній частині табл. 6.2. Усі коефіцієнти третього рівняння після перетворень дорівнюють нулю. Такий результат означає, що вихідні рівняння системи (8) були лінійно пов'язані між собою. Вибрана послідовність елементарних перетворень обумовила залежність третього рівняння від решти рівнянь. За іншої послідовності перетворень інше рівняння одержало б нульові коефіцієнти, але за будь-якої послі-

довності перетворень одне з рівнянь системи (8) по завершенні виконання алгоритму матиме нульові коефіцієнти.

Реалізувавши алгоритм методом Жордана-Гаусса, одержимо такі результати.

1. У деяких рівняннях усі коефіцієнти зліва та справа можуть прийняти нульові значення (в таблиці реалізації алгоритму певні рядки розширеної матриці стають нульовими). Це означає, що у вихідній системі відповідні рівняння були лінійно залежними від решти рівнянь. Таким рівнянням задовольнятиме будь-який розв'язок решти рівнянь. По завершенні алгоритму за названих умов залишається система рівнянь (5), *тотожна* вихідній, і кожне рівняння (5) легко розв'язати відносно r невідомих, записавши їх значення через значення решти $(n - r)$ невідомих. Частинний випадок $r = m$.

2. Усі коефіцієнти при невідомих у деякому рівнянні прийняли нульові значення, а вільний член не одержав нульового значення. Така ситуація означає, що вихідна СЛАР несумісна, розв'язку не має.

Таким чином, метод Жордана-Гаусса або гарантує перетворення вихідної СЛАР у вид (5), що дає можливість побудувати її розв'язок, або показує її несумісність, тобто відсутність розв'язку. Подальші викладки стосуються *лише сумісних СЛАР*.

Якщо розглядати стовпці розширеної матриці СЛАР, то теорема Кронекера-Капеллі стверджує, що серед n векторів-стовпців існує r лінійно незалежних векторів, які можна обрати за базис. Таких базисів можна обрати *не більше* C_n^r, тому що деякі стовпці матриці коефіцієнтів системи можуть бути лінійно залежними. Метод Жордана-Гаусса дозволяє побудувати один з можливих базисів, а саме так званий *одиничний базис*, характерною особливістю векторів якого є те, що у кожного з них усі компоненти, окрім однієї, *дорівнюють нулю*, а компонента, яка дорівнює одиниці, у різних векторів базису розташована на різних місцях. Тобто за вектори одиничного базису обрані вектори:

$$\bar{e}_1 = \begin{bmatrix}1\\0\\\vdots\\0\\0\end{bmatrix};\quad \bar{e}_2 = \begin{bmatrix}0\\1\\\vdots\\0\\0\end{bmatrix};\quad \bar{e}_3 = \begin{bmatrix}0\\0\\\vdots\\0\\0\end{bmatrix};\quad \ldots;\ \bar{e}_{r-1} = \begin{bmatrix}0\\0\\\vdots\\1\\0\end{bmatrix};\quad \bar{e}_r = \begin{bmatrix}0\\0\\\vdots\\0\\1\end{bmatrix}.$$

За так обраного базису будь-який вектор $\bar{p} = (p_1, p_2, \ldots, p_r)^T$ відповідного простору легко розкласти по векторах одиничного базису:

$$\bar{p} = \sum_{i=1}^{r} p_i \bar{e}_i .$$

Розв'язання СЛАР (5), тотожної СЛАР (1), виконують за такою схемою. Множину невідомих $\bar{x} = (x_1, x_2, \ldots, x_n)$ розділяють на дві групи: *базисні* множники при векторах базису та *вільні*.

Перетворення СЛАР (1) на СЛАР (5) дозволяє обрати один із можливих одиничних базисів, а потім виразити кожну базисну невідому x_i $(i = \overline{1, r})$ через вільні невідомі $x_k = (x_{r+1}, x_{r+2}, \ldots, x_n)$ та вільні члени:

$$x_i = b_i^* - \sum_{k=r+1}^{n} a_{ik}^* x_k \quad (i = \overline{1, r}). \tag{9}$$

Даючи довільні значення вільним невідомим, одержуємо, користуючись (9), значення базисних невідомих. Таким чином, СЛАР (1) має численну множину розв'язків. Твердження вірне за умови істинності теореми Кронекера-Капеллі $(r\|A\| = r\|A \vdots \bar{b}\|)$ та коли $r < n$, тобто є вільні невідомі. Якщо $r = n$ (вільні невідомі відсутні), то система (1) має єдиний розв'язок. Розв'язок СЛАР (5) та їй тотожної (1) називається *базисним* за умови, що всім небазисним (вільним) невідомим надано нульові значення.

Приклади. Для спрощення запису різні невідомі будемо позначати різними літерами, не користуючись індексами. Розв'язати систему рівнянь:

$$\begin{cases} x + 2y = 5 \\ 3x - y = 1. \end{cases} \tag{10}$$

Помножимо перше рівняння на –3 та додамо його до другого, потім друге рівняння після його множення на –2 додамо до першого. Маємо послідовність елементарних перетворень:

$$\begin{cases} x + 2y = 5 \\ -7y = -14 \end{cases} \Rightarrow \begin{cases} x + 2y = 5 \\ y = 2 \end{cases} \Rightarrow \begin{cases} x = 1 \\ y = 2. \end{cases}$$

Система (10) має єдиний розв'язок: $x = 1$; $y = 2$. Розглянемо СЛАР:

$$\begin{cases} x + 2y = 6 \\ 3x + 6y = 4. \end{cases} \tag{11}$$

Помноживши перше рівняння на –3 та додавши його до другого, одержимо: $0 = -14$. Отже, СЛАР (11) не сумісна, розв'язків не має.

Змінивши вільний член у другого рівняння, розглянемо СЛАР:

$$\begin{cases} x + 2y = 6 \\ 3x + 6y = 18. \end{cases} \qquad (12)$$

Помноживши перше рівняння на –3 та додавши його до другого, одержимо: 0 = 0. Результат свідчить про те, що рівняння (12) лінійно залежні, система має нескінченну множину розв'язків, побудувати яку можна так. Не беручи до уваги друге рівняння як лінійно залежне від першого, маємо $x = 6 - 2y$. Тобто x треба взяти за базисну невідому, за вільну – y. Даючи довільні значення вільній невідомій y, одержуємо відповідні значення x. Легко переконатися, що коли x та y пов'язані залежністю $x = 6 - 2y$, то друге рівняння системи задовольняється:

$$3(6 - 2y) + 6y \equiv 18.$$

Пропонуємо, користуючись прямокутною системою координат, дати геометричне тлумачення розв'язанню СЛАР (10), (11), (12).

Як зазначалося раніше, СЛАР (1) може бути приведена до різних базисів і тому мати різні базисні розв'язки. Розглянемо, як можна замінити базис, тобто мати при інших невідомих стовпці лише з однією 1 в їх компонентах, у порівнянні з вихідним базисом.

Процедуру заміни базису побудуємо за алгоритмом так званого *однократного заміщення*, при реалізації якого певна невідома x_t вводиться в базис, а базисна невідома x_s вилучається з базису. Для реалізації такого заміщення необхідно, щоб в системі (5) коефіцієнт $a^*_{st} \neq 0$. Обчислення коефіцієнтів розширеної матриці СЛАР виконується за формулами (6) та (7) за такою ж схемою, як і при реалізації алгоритму методом Жордана-Гаусса. Реалізацію алгоритму однократного заміщення в базисі покажемо на прикладі.

$$\begin{cases} x_1 \quad + 3x_4 - 5x_5 = 4 \\ x_2 \quad + 4x_4 + 2x_5 = 6 \\ \quad x_3 - 2x_4 + 3x_5 = 3. \end{cases} \qquad (13)$$

У наведеній системі (13) необхідно невідому x_2 вивести з базису, а невідому x_5 ввести в базис. Таким же чином можна було ставити запитання стосовно інших пар змінних системи (13).

Запишемо в табл. 6. 3 результати необхідних перетворень стосовно розширеної матриці СЛАР (13) за схемою реалізації алгоритму методом Жордана-Гаусса, додавши справа ще один стовпець для запису базисного розв'язку.

За базисних змінних x_1, x_2, x_3 мали базисний розв'язок $\bar{x}$ = (4; 6; 3; 0; 0). Замінивши базисну змінну x_2 на x_5, одержали новий базисний розв'язок $\bar{x} = (19; 0; -6; 0; 3)$.

Наголосимо, що базисний розв'язок має в обох базисах п'ять компонент (за кількістю невідомих); вільні невідомі, маючи різні індекси, дорівнюють нулю.

Таблиця 6.3.

Номер ітерації	Базисні невідомі	x_1	x_2	x_3	x_4	x_5	b_i	Σ	Базисний розв'язок
Вихідна система	x_1	1	0	0	3	–5	4	3	$\bar{x}$ = (4;6;3;0;0)
	x_2	0	1	0	4	2	6	13	
	x_3	0	0	1	–2	3	3	5	
Перша ітерація (заміщення однієї змінної)	x_1	1	5/2	0	13	0	19	71/2	$\bar{x}$ = (19;0;–6;0;3)
	x_5	0	1/2	0	2	1	3	13/2	
	x_3	0	–3/2	1	–8	0	–6	–29/2	

Наведені в цій главі відомості з лінійної алгебри будуть широко використані при подальшому викладенні матеріалу.

Питання для самоконтролю

1. Що таке число?
2. За яких умов використовуються операції ≤; =; ≥ ?
3. Що таке вектор?
4. Що таке векторний простір?
5. Що таке базис векторного простору?
6. За якої умови система векторів називається лінійно незалежною?
7. Операції над векторами, правила їх виконання, властивості операцій.
8. Що таке матриця? Спеціальні види матриць.
9. Економічне тлумачення понять"вектор" і "матриця".
10. Операції стосовно матриць. Правила виконання операцій та їх властивості.
11. Економічне тлумачення операцій над матрицями.
12. Що таке визначник матриці? Які матриці мають визначники?
13. Як обчислюють визначники другого та третього порядків?
14. Що таке мінор матриці та мінор визначника?
15. Яка різниця між мінором та алгебраїчним доповненням певного елемента?
16. Мета та можливості розв'язання СЛАР методом Жордана-Гаусса.
17. Який розв'язок СЛАР називається базисним?
18. Як поділяються невідомі на базисні та вільні?

ГЛАВА 7

СИМПЛЕКСНИЙ МЕТОД РОЗВ'ЯЗАННЯ ЗАДАЧ ЛІНІЙНОГО ПРОГРАМУВАННЯ

7.1. Опорні плани задачі лінійного програмування

Як було показано у главі 3, лінійну математичну модель будь-якої практичної задачі шляхом алгебраїчних перетворень завжди можна привести до так званої канонічної форми задачі лінійного програмування (ЗЛП).

Знайти максимум цільової функції

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

за умови, що шукані величини x_j $(j = \overline{1, n})$ мають задовольняти системі лінійних алгебраїчних рівнянь:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \vdots \qquad \vdots \qquad \vdots\vdots\vdots \qquad \vdots \qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases} \tag{2}$$

та бути невід'ємними:

$$x_j \geq 0 \ (j = \overline{1, n}). \tag{3}$$

Розглядаючи упорядковані множини c_j $(j = \overline{1, n})$ та x_j $(j = \overline{1, n})$ як компоненти векторів відповідно $\bar{c}$ та $\bar{x}$, а множини коефіцієнтів розширеної матриці системи (2), розташовані в однакових стовпцях, – як компоненти векторів $\bar{a}_j$ $(j = \overline{1, n})$, задачу (1), (2) запишемо так: знайти максимум

$$f = \bar{c}\,\bar{x} \tag{1a}$$

за умови

$$\sum_{j=1}^{n} x_j\,\bar{a}_j = \bar{b} \tag{2a}$$

$$\bar{x} \geq \bar{0}. \tag{3a}$$

Формули (1а) – (3а) дозволяють записати канонічну форму ЗЛП у векторному зображенні.

Щоб був сенс говорити про пошук оптимального розв'язку задачі (1) – (3), система (2) повинна бути *сумісною* і мати не єдиний розв'язок, який би задовольняв (3). Це можливо лише за умови, що ранг r системи (2) (кількість лінійно незалежних рівнянь в системі) не більше кількості m рівнянь системи, а m не більше кількості n шуканих величин x_j $(j=\overline{1,n})$, тобто $r \le m \le n$. За умови $r = n$ СЛАР (2) має єдиний розв'язок і не має вибору. Випадок $r > m$ не має сенсу.

Таким чином, доцільно розглядати лише випадок $r \le m \le n$. За такої умови у системі векторів $\bar{a}_j$ $(j=\overline{1,n})$ завжди можна вибрати r лінійно незалежних векторів, які були названі в своїй сукупності *базисом.* Тоді множину шуканих величин x_j $(j=\overline{1,n})$ доцільно розділити на дві множини: *базисні* змінні – величини x_j, індекси яких відповідають індексам векторів, обраних за базис, і *небазисні,* або *вільні* змінні. Базисів може бути декілька, але не більше числа комбінацій з n по r (C_n^r). Таким чином, розподіл множини невідомих x_j $(j=\overline{1,n})$ на базисні та небазисні не визначається однозначно, але однозначно узгоджується з певним базисом. Використовуючи метод Жордана-Гаусса, можна шляхом елементарних перетворень привести СЛАР (2) до тотожної (еквівалентної) їй, в якій базисні змінні визначаються через небазисні, а кількість рівнянь – рангом матриці коефіцієнтів вихідної СЛАР ($m = r$). Без обмеження загальності будемо вважати, що базисні змінні відповідають першим m невідомим і визначаються через вільні змінні так:

$$x_i = b_i^* - \sum_{j=m+1}^{n} a_{ij}^* x_j \quad (i=\overline{1,m}) \tag{4}$$

(позначка * означає, що маємо коефіцієнти розширеної матриці СЛАР (2) після відповідних елементарних перетворень, тобто СЛАР записано у вигляді (4)). Розв'язок (2) у формі (4) називається *загальним рішенням,* або *загальним розв'язком* СЛАР (2). Сенс загального розв'язку такий: вільним змінним можна давати будь-які числові значення, тоді базисні змінні мають бути обчислені за (4).

Приклад. Записати загальний розв'язок СЛАР

$$\begin{cases} x_1 - 2x_2 + x_3 + x_4 = 5 \\ 2x_1 + x_2 + 3x_3 - x_4 = 7. \end{cases} \tag{5}$$

Використаємо метод Жордана-Гаусса, записавши відповідні перетворення в табл. 7.1.

Таблиця 7.1.

	x_1	x_2	x_3	x_4	ВЗ	Σ	Пояснення
Вихідна система	1	–2	1	1	5	6	
	2	1	3	–1	7	12	
Перший крок	1	–2	1	1	5	6	До другого рівняння додали перше, помножене на –2.
	0	5	1	–3	–3	0	
Другий крок	1	–7	0	4	8	6	Від першого рівняння відняли друге.
	0	5	1	–3	–3	0	

По завершенні перетворень маємо систему

$$\begin{cases} x_1 - 7x_2 \qquad + 4x_4 = 8 \\ \qquad 5x_2 + x_3 - 3x_4 = -3. \end{cases}$$

Розв'язавши її стосовно x_1 та x_3, одержимо загальний розв'язок системи (5) у формі (4):

$$\begin{cases} x_1 = 8 + 7x_2 \quad - 4x_4 \\ x_3 = -3 - 5x_2 + 3x_4. \end{cases}$$

Підставивши одержані результати в (5), переконуємося, що дійсно маємо загальний розв'язок: ліва частина рівнянь тотожна правій за будь-яких значень вільних змінних x_2 та x_4.

Вибравши за базисні змінні x_1 та x_4, загальний розв'язок (5) запишемо у вигляді:

$$\begin{cases} x_1 = 4 + \frac{1}{3}x_2 - \frac{4}{3}x_3 \\ x_4 = 1 + \frac{5}{3}x_2 + \frac{1}{3}x_3 \end{cases}$$

(x_2 та x_3 – вільні змінні).

За умови, що базисні змінні – x_2 та x_4, а вільні змінні – x_1 та x_3, загальний розв'язок буде представлено так:

$$\begin{cases} x_2 = -12 + 3x_1 + 4x_3 \\ x_4 = -19 + 5x_1 + 7x_3. \end{cases}$$

Таким чином, загальний розв'язок (5) визначається вибором базисних і вільних змінних; вільним змінним можна давати будь-які значення.

Скориставшися методом Жордана-Гаусса та побудувавши загальний розв'язок СЛАР (2) у вигляді (4), вибравши певним чином базисні та вільні змінні, переконуємося, що за умови *будь-яких значень вільних змінних величини базисних змінних визначаються однозначно.* Для розбудови алгоритмів розв'язання ЗЛП суттєве значення мають так звані *базисні розв'язки*, які одержують із загальних розв'язків за умови, що *значення вільних змінних дорівнюють нулю.* Доцільно наголосити, що при використанні методу Жордана-Гаусса маємо одиничний базис. Вибору певного базису відповідає певний однозначно визначений базисний розв'язок. А оскільки СЛАР (2) має не більше базисів, ніж C_n^r, то й для кожної системи існує не більше C_n^r базисних розв'язків. Якщо в *базисному розв'язку всі базисні змінні мають невід'ємні значення,* тобто виконується умова (3), то такий базисний розв'язок називається *опорним*, або *допустимим*. У математичному програмуванні, з огляду на економічний зміст практичних задач, такі розв'язки називаються *опорними планами.* Враховуючи, що опорні плани (розв'язки) будуються за формулою (4) при нульових значеннях вільних змінних, доходимо висновку, що *опорний план не може мати більше r додатних компонент.* Опорний план, який має рівно *r додатних компонент,* називається *невиродженим*; якщо ж хоча б одна з його компонент дорівнює нулю, то опорний план називається *виродженим.*

У теорії лінійного програмування доведено теорему, яка стверджує, що за умови сумісності СЛАР (2) *задача лінійного програмування буде мати хоча б один опорний план.* А оскільки базисних розв'язків СЛАР (2) не більше C_n^r, то й опорних планів не може бути більше C_n^r.

Основна теорема теорії лінійного програмування стверджує, що *за умови існування оптимального розв'язку ЗЛП він обов'язково співпадатиме хоча б з одним з опорних планів. Якщо оптимальний розв'язок досягається за двох або більше різних опорних планів, то він буде оптимальним за будь-якого плану, який є так званою лінійною опуклою комбінацією оптимізуючих планів.*

Приймаючи на віру перше твердження теореми (бажаючі можуть ознайомитися з доведенням у літературі*), доведемо друге твердження. Приймемо, що функція $f(\bar{x})$ досягає одного й того ж значення M за різних планів $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p$, тобто $f(\bar{x}_i) = M$ за будь-якого $i \in \{1; p\}$. За визначенням *лінійною опуклою комбінацією планів* (векторів) називається план (вектор) $\bar{y} = \sum_{i=1}^{p} \alpha_i \bar{x}_i$ за умови $\alpha_i \geq 0 \ (i = \overline{1, p})$ та $\sum_{i=1}^{p} \alpha_i = 1$.

* Вагнер Г. Основы исследования операций. – М.: Мир, 1973. – Т. 1.

$$f(\bar{y}) = \sum_{i=1}^{p} \alpha_i f(\bar{x}_i) = \sum_{i=1}^{p} \alpha_i M = M \sum_{i=1}^{p} \alpha_i = M.$$

Отже, друге твердження основної теореми вірне. Розглянемо, як можна оцінити наявний опорний план за критерієм оптимальності та поліпшити його, якщо це можливо та необхідно.

7.2. Оцінки опорних планів

Припустимо, що для канонічної форми ЗЛП, в якій необхідно знайти найбільше значення функції

$$f(\bar{x}) = \sum_{j=1}^{n} c_j x_j$$

за умови

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

$$x_j \geq 0 \quad (j = \overline{1, n}),$$

маємо певний опорний план (не обов'язково оптимальний) та відповідний m-мірний одиничний базис, тобто всі m рівнянь-обмежень канонічної форми лінійно незалежні ($r = m$). За такого припущення систему рівнянь-обмежень доцільно розв'язати відносно базисних змінних, яким даємо *умовні назви* $x_1, x_2, \ldots, x_m$, представивши їх через вільні змінні $x_{m+1}, x_{m+2}, \ldots, x_n$:

$$x_i = b_{i0} - \sum_{j=1}^{n-m} b_{ij} x_{m+j} \qquad (i = \overline{1, m}) \tag{6}$$

(літерою з відповідними індексами позначені коефіцієнти системи рівнянь-обмежень після виділення в ній одиничного базису за методом Жордана-Гаусса).

Скориставшися залежністю (6), підставимо значення x_i у цільову функцію (1) і запишемо її через вільні змінні:

$$f = b_{00} - \sum_{j=1}^{n-m} b_{0j} x_{m+j}, \tag{7}$$

(b_{00} – значення цільової функції за наявного опорного плану, бо для нього $x_{m+j} = 0$).

Залежності (6) та (7) свідчать, що величини базисних змінних і цільової функції визначаються величинами вільних змінних за певного базису. Таким чином, значення цільової функції може бути збільшене лише з вибором нового базису, тобто нового опорного плану. Розглянемо основні засади побудови алгоритму так званого симплексного методу пошуку оптимального плану, скориставшися відповідною методикою*.

7.2.1. Ознака оптимальності опорного плану

Припустимо, розглядається опорний план, тобто в рівнянні (6) $b_{i0} \geq 0$ $(i = \overline{1, m})$. Якщо $b_{i0} > 0$ для всіх $i = \overline{1, m}$, то план називається *невиродженим*. У випадку $b_{i0} = 0$ опорний план називається *виродженим*. ЗЛП також називається *виродженою*. При розв'язанні вироджених ЗЛП, як наголошувалося раніше, можливою є ситуація зациклення, але теорія дає певні рекомендації щодо уникнення зациклення, а практика розв'язання ЗЛП свідчить про малу ймовірність таких ситуацій. Тому розглянемо невироджену ЗЛП, тобто наявний опорний план є *невироджений*. Оцінимо його за критерієм оптимальності та розглянемо можливості його поліпшення, якщо він не оптимальний. Для реалізації поставленої мети зручно скористатися розташуванням наявної інформації (6) та (7) у вигляді табл. 7.2 і всі подальші результати розрахунків теж записувати в таблицях.

У табл. 7.2 записані коефіцієнти рівнянь (6) та в нижньому рядку коефіцієнти цільової функції f, представленої у формі (7). Побудовану за такою схемою симплексну таблицю назвемо T.

Таблиця 7.2.

Базисні змінні БЗ	Величини вільних членів b_{i0}	Коефіцієнти при вільних змінних						b_{i0}/b_{is} ($b_{i0} > 0$; $b_{is} > 0$)
		$-x_{m+1}$	$-x_{m+2}$	...	$-x_{m+s}$	...	$-x_n$	
$x_1 =$	b_{10}	b_{11}	b_{12}	...	b_{1s}	...	$b_{1,\,n-m}$	
$x_2 =$	b_{20}	b_{21}	b_{22}	...	b_{2s}	...	$b_{2,\,n-m}$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
$x_k =$	b_{k0}	b_{k1}	b_{k2}	...	b_{ks}	...	$b_{k,\,n-m}$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
$x_m =$	b_{m0}	b_{m1}	b_{m2}	...	b_{ms}	...	$b_{m,\,n-m}$	
$f =$	b_{00}	b_{01}	b_{02}	...	b_{0s}	...	$b_{0,\,n-m}$	

* Кузнецов А. В., Холод Н. И. Математическое программирование: Учеб. пособие. – Минск: Вышейша шк., 1984.

Розглянемо умови, за яких можна очікувати збільшення значення цільової функції (7) за рахунок зміни величин вільних змінних x_i $(i = \overline{m+1, n})$ так, щоб їх значення залишалися в області допустимих, тобто були невід'ємними. Зростання цільової функції з такою вимогою може бути лише при наявності від'ємних коефіцієнтів b_{0j}. Отже, *наявний опорний план буде оптимальним лише за умови, що всі елементи f-рядка таблиці T, за винятком b_{00}, невід'ємні.* Дійсно, якщо для певного j $(j \in \{m+1, n\})$ $b_{0j} \geq 0$, а значення x_j вибрати невід'ємними (вимога, щоб розв'язок був допустимий), то величина цільової функції не може зростати.

Тепер припустимо, що в якомусь стовпці s маємо $b_{0s} < 0$ та хоча б один коефіцієнт $b_{is} > 0$ $(i = \overline{1, m})$. За такої умови, якщо всі вільні змінні x_{m+j}, за винятком x_{m+s}, прирівняти нулю, то згідно із залежністю (7) значення цільової функції f зросте й дорівнюватиме: $b_{00} - b_{0s}\, x_{m+s}$. Таким чином, при наявності від'ємної величини в f-рядку є можливість перейти до нового базису та відповідного нового опорного плану, при якому значення цільової функції зросте (не зменшиться). Для досягнення такої мети необхідно з'ясувати:

а) яку змінну треба вилучити з попереднього базису, щоб новий базис мав у своєму складі змінну x_{m+s};

б) яке значення в новому базисі буде мати змінна x_{m+s}.

Виключивши відповідні теоретичні докази*, сформулюємо правила заміни базису та покажемо виконання відповідних обчислень для пошуку поліпшеного опорного плану. У наявний базис вводиться змінна, якій відповідає *найменше від'ємне значення* b_{0j} $(j = \overline{1; n-m})$ у f-рядку. Стовпець таблиці T, що відповідає змінній, яку залучають до базису, називається *провідним*. Надамо йому умовний індекс s. Для визначення змінної, яку необхідно вилучити з базису, оцінюють так звані симплексні відношення – відношення вільних членів СЛАР (6), розташованих у рядках *додатних елементів провідного стовпця*, до цих елементів, тобто b_{i0} / b_{is} $(b_{is} > 0)$, $i = \overline{1, m}$. Величини вільних членів записано у відповідному стовпці таблиці T. Рядок, для якого симплексне відношення досягає *найменшого значення*, визначає змінну, яку необхідно вилучити з базису. Такий рядок називається *провідним*. Надамо йому умовний індекс k. Елемент b_{ks} таблиці T, розташований на перетині k-го рядка та s-го стовпця, називається *провідним*. При наявності кількох тотожних найменших від'ємних елементів b_{0j}, або симплексних відношень, вибирають будь-який з можливих провідних елементів.

Перехід до нового базису обумовлює перерахунок коефіцієнтів таблиці T (побудову наступної таблиці T) за правилами:

1) на позиції провідного елемента записується величина, йому обернена;

2) на решту позицій номера провідного рядка записуються елементи цього рядка, розділені на величину провідного елемента;

* Вентцель Е. С. Исследование операций: задачи, принципы, методология. – М.: Наука, 1988.

3) на позиції номера провідного стовпця, за винятком позиції провідного елемента, записуються відповідні елементи провідного стовпця, розділені на провідний елемент та домножені на "–1";

4) решта коефіцієнтів (за винятком провідних рядка та стовпця) b_{ij}^* наступної таблиці T обчислюється за формулою:

$$b_{ij}^* = b_{ij} - \frac{b_{is} \cdot b_{kj}}{b_{ks}} \quad (i = \overline{0, m};\ i \neq k;\ j = \overline{0, n-m};\ j \neq s); \tag{8}$$

5) у стовпці "Базисні змінні" таблиці T найменування x_k змінюється на x_s, а у рядку "Вільні змінні" – навпаки.

Бажаючи скоротити обсяги обчислень (але поступитися зручністю в користуванні), формулу (8), за умови використання k-го рядка нової таблиці T, записують так:

$$b_{ij}^* = b_{ij} - b_{kj}^* \cdot b_{is} \tag{8a}$$

(позначка * вказує на необхідність використання відповідних елементів наступної таблиці).

На практиці, використовуючи формули (8) або (8а), зручно враховувати, що необхідні елементи таблиць розташовані у вершинах умовного прямокутника, в якому квадратом позначено розташування провідного елемента, а колом – елемента, для якого необхідно зробити розрахунки:

$$\begin{array}{ccccc} \dots & & \dots & & \dots \\ \dots & b_{ij} & \dots & b_{is} & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & b_{kj} & \dots & \boxed{b_{ks}} & \dots \\ \dots & & \dots & & \dots \end{array}$$

Окремої уваги вимагає послідовність перетворень *за наявності від'ємного елемента* у стовпці b_{i0}, тобто від'ємної компоненти опорного плану (але недопустимого). Якщо у стовпці вільних членів b_{i0} присутні від'ємні числа, то вибір провідного елемента симплексних перетворень здійснюється за такими правилами:

1) аналізують елементи одного з рядків з від'ємними b_{i0} (назвемо такий рядок t) і вибирають в ньому *від'ємні елементи* (при сумісності обмежень буде хоча б один такий елемент); стовпець з від'ємним елементом обирають за провідний (назвемо його s-стовпець);

2) обчислюють *відношення елементів стовпця вільних членів* до відповідних за рядком *елементів провідного стовпця* при тотожності їх знаків (симплексні відношення) та записують їх у крайній правий стовпець b_{i0}/b_{is};

3) розташування *найменшого із симплексних відношень визначає провідний рядок*, наприклад, k-рядок (якщо серед симплексних найменших відношень кілька тотожних, то обираємо будь-яке з таких);

4) провідний елемент розташований на перетині провідних рядка та стовпця. Після вибору провідного елемента чергова розбудова таблиці T виконується так, як описано раніше. Якщо провідний елемент виявився розташованим у t-рядку (див. п. 1), то після виконання симплексних перетворень вільний член такого рядка стане додатним. У протилежному випадку на наступному кроці знову необхідно розглядати t-рядок. Якщо задача має розв'язок, то через певну кількість симплексних перетворень опорний план стане допустимим (у стовпці вільних членів не буде від'ємних елементів).

Доцільно наголосити на спрощенні обчислень за деяких умов. Формули (8) та (8а) свідчать: якщо у провідному рядку певний елемент b_{kj} (b_{kj}^*) дорівнює нулю, то елементи стовпця j у симплексному перетворенні не змінюються. Аналогічно: якщо в провідному стовпці наявним є нульовий елемент на i-й позиції ($b_{is} = 0$), то відповідний i-й рядок також не змінюється.

Виконавши описані розрахунки та змінивши найменування змінних (із базисних вилучивши змінну k і замінивши її на s, а для вільних змінних – s замінивши на k), одержуємо симплексну таблицю, яка відповідає наступному базису й новому опорному плану, в якому цільова функція зростає (не спадає). На завершення необхідно додати, що базисні змінні в новому базисі через вільні змінні будуть обчислюватися за формулою (8) з урахуванням коефіцієнтів симплексної таблиці T.

Застосування запропонованого варіанта симплексного методу покажемо на розв'язанні задачі.

Приклад. Знайти найбільше значення функції

$$f = x_1 + x_2 \tag{9}$$

за умови:

$$\begin{cases} -x_1 + 15x_2 + x_3 = 45 & (1) \\ 7x_1 - 11x_2 + x_4 = 14 & (2) \\ -x_1 - 2x_2 + x_5 = -1 & (3) \\ x_1 \geq 0;\ x_2 \geq 0 \end{cases} \tag{10}$$

$$x_1 \text{ та } x_2 \text{ – цілі числа.} \tag{11}$$

Умову (11) буде використано далі.

Обравши x_3, x_4, x_5 за базисні змінні, вихідну симплекс-таблицю T запишемо в

табл. 7.3. Маємо в рядку x_5 від'ємний вільний член – 1, тому за алгоритмом в цьому рядку шукаємо від'ємний елемент. Із двох наявних від'ємних елементів –1 та –2 виберемо, наприклад, –1, тоді стовпець $-x_1$ буде провідним.

Таблиця 7.3.

БЗ	b_{i0}	$-x_1$	$-x_2$	b_{i0}/b_{is}
x_3	45	–1	15	–
x_4	14	7	–11	2
x_5	–1	(-1)	–2	1
f	–	–1	–1	

Для вільних членів b_{i0} та елементів провідного стовпця $-x_1$ з тотожними знаками обчислимо симплексні відношення, записані у стовпці b_{i0}/b_{is}, і визначимо рядок з найменшим відношенням x_5. Отже, провідним елементом буде –1 (позначений колом). Послідовність наступних симплексних таблиць T записана в частинах A, B, C табл. 7.4.

Таблиця 7.4.

		A		
БЗ	b_{i0}	$-x_5$	$-x_2$	b_{i0}/b_{is}
x_3	46	–1	17	–
x_4	7	(7)	–25	1
x_1	1	–1	2	–
f	1	–1	1	
		B		
БЗ	b_{i0}	$-x_4$	$-x_2$	b_{i0}/b_{is}
x_3	47	1/7	(94/7)	7/2
x_5	1	1/7	–25/7	–
x_1	2	1/7	–11/7	–
f	2	1/7	–18/7	–
		C		
БЗ	b_{i0}	$-x_4$	$-x_3$	b_{i0}/b_{is}
x_2	7/2	1/94	7/94	
x_5	27/2	17/94	25/94	
x_1	15/2	15/94	11/94	
f	11	8/47	18/94	

Аналіз завершальної таблиці T, записаної в частині C, свідчить, що в f-рядку немає від'ємних коефіцієнтів, отже, досягнутий план з компонентами $x_1 = 7\frac{1}{2}$; $x_2 = 3\frac{1}{2}$; $x_3 = x_4 = x_5 = 0$, є оптимальним; $f = 11$.

7.2.2. Ознака необмеженості цільової функції в допустимій області

Якщо у f-рядку оцінок симплекс-таблиці при певному опорному плані є хоча б один від'ємний коефіцієнт, якому відповідає *стовпець з недодатними коефіцієнтами*, то цільова функція не обмежена на множині допустимих розв'язків, тобто $f \to \infty$. Доведемо це твердження. Припустимо, що маємо певний опорний план та відповідну йому симплекс-таблицю. Приймемо, що базисні змінні мають номери від 1 до m. Після обчислення оцінок b_{0j} припустимо наявність певної від'ємної оцінки, наприклад, b_{0s}, де s має будь-який номер із номерів від $m + 1$ до n.

Прирівнявши у формулах (6) та (7) всі вільні змінні, за винятком x_{m+s}, нулю, маємо:

$$x_i = b_{i0} - b_{is}x_{m+s} \quad (i = \overline{1, m})$$
$$f = b_{00} - b_{0s}x_{m+s}. \tag{12}$$

Оскільки за припущенням, всі $b_{is} \le 0$ $(i = \overline{1, m})$, то величину змінної x_{m+s} можна необмежено збільшувати, залишаючись в області допустимих значень. За такої умови значення $f(\bar{x})$ теж буде зростати необмежено, бо $b_{0s} < 0$. Твердження доведено.

7.2.3. Ознака наявності нескінченної множини оптимальних планів

Розглядаючи графо-аналітичний спосіб розв'язання ЗЛП, ми наголошували, що за певних умов може з'явитися ситуація, коли задача має нескінченну множину оптимальних розв'язків. З'ясуємо умови можливої появи такої ситуації в загальному випадку. *Якщо в рядку оцінок симплекс-таблиці, яка відповідає оптимальному плану (тобто серед елементів f-рядка), немає від'ємних, серед оцінок вільних змінних є хоча б одна нульова, а у відповідному стовпці – хоча б один додатний коефіцієнт, то ЗЛП має нескінченну множину оптимальних планів.*

Припустимо, що маємо симплекс-таблицю, яка відповідає оптимальному пла-

ну $\bar{x}_1^0$ і в якій оцінка b_{0s} має нульове значення, а решта оцінок невід'ємні (розглядаємо оптимальний план).

При наявності додатного коефіцієнта в s-му стовпці можемо відповідну змінну вилучити з базису, а s-ту змінну ввести в базис, але оскільки оцінка b_{0s} має нульове значення, то при заміні базису й одержанні нового оптимального розв'язку $\bar{x}_2^0$ значення цільової функції не зміниться (див. (12)).

Таким чином, маємо два допустимих розв'язки ЗЛП (до того ж обидва оптимальні). За такої умови, як було доведено раніше, опукла лінійна комбінація $\lambda\bar{x}_1^0+(1-\lambda)\bar{x}_2^0$ $(0 \le \lambda \le 1)$ буде також оптимальним розв'язком. Отже, даючи величині λ будь-які значення з проміжку [0; 1], одержуємо нескінченну множину розв'язків ЗЛП. Для кожного такого розв'язку значення цільової функції буде оптимальним. Твердження доведено.

Різні опорні оптимальні плани називаються *альтернативними.* Таких планів буде стільки, скільки є нульових оцінок *при певному оптимальному плані.* Якщо оптимальний план, який відповідає s-й нульовій оцінці, позначимо $\bar{x}_s^0$, а кількість нульових оцінок t, то множину всіх можливих оптимальних планів за досліджуваних умов можна описати залежністю:

$$\bar{x}^0 = \sum_{s=1}^{t} \lambda_s \bar{x}_s^0 \quad (\lambda_s \ge 0; \quad s=\overline{1,\ t}; \quad \sum_{s=1}^{t} \lambda_s = 1).$$

Це твердження було доведено при розгляді основної теореми лінійного програмування.

7.2.4. Вироджені плани задачі лінійного програмування та проблема зациклення

Якщо при розрахунках симплексних відношень b_{i0}/b_{is} з'являється кілька найменших тотожних, то в новому опорному плані будуть нульові значення базисних змінних, тобто план буде виродженим. Припустимо, що за вибраного s-го провідного стовпця min b_{i0}/b_{is} досягається за кількох індексів, наприклад, за $i = l$ та $i = t$, тобто $b_{l0}b_{ts} = b_{t0}b_{ls}$, а провідним вибрано l-й рядок. За такої умови в новому опорному плані базисна змінна x_t згідно (7) дорівнюватиме:

$$b_{t0}^* = (b_{l0}b_{ts} - b_{t0}b_{ls})/b_{ls} = 0.$$

Таким чином, новий опорний план буде вироджений, ЗЛП теж вироджена.

Якщо при розв'язанні виродженої ЗЛП на черговому кроці симплексних перетворень з'явиться рядок, у якому вільний член дорівнює нулю, то нулю дорівнюватиме й нова базисна компонента відповідного опорного плану, але значення решти базисних компонент і цільової функції залишаться незмінними.

Припустимо, що за наявного опорного плану з нульовою компонентою на p-му місці серед базисних змінних у загальній нумерації компонент маємо деяку b_{0k} від'ємну оцінку, а у відповідному стовпці – певний додатний елемент b_{pk}, тобто є можливість побудувати інший опорний план. Якщо виберемо b_{pk} за провідний елемент, то компоненти нового опорного плану, згідно з (8), будуть визначатися так:

$$b_{p0}^* = \frac{b_{p0}}{b_{pk}} = 0; \quad b_s^* = b_s - \frac{b_p \cdot b_{sk}}{b_{pk}} \qquad (s \ne p).$$

Але, згідно з допущенням, вихідний опорний план серед базисних змінних на p-му місці мав нульову компоненту $b_{p0} = 0$, отже, величини базисних змінних нового опорного плану не змінилися стосовно тих же компонент вихідного плану. У новому опорному плані колишня вільна змінна x_k, яка дорівнювала нулю, стала базисною змінною, рівною нулю, а колишня базисна змінна x_p, яка дорівнювала нулю, стала вільною, також рівною нулю. Компоненти інших базисних змінних у новому базисі не змінили своїх значень.

За досліджуваних умов новий опорний план не змінився стосовно вихідного плану, тобто крок у розв'язанні ЗЛП зроблено, але він виявився *даремним*: розв'язок не поліпшився, але змінилися перелік базисних змінних і величини коефіцієнтів симплекс-таблиці. Для нової симплекс-таблиці зміняться оцінки вільних змінних.

За наступних кроків зміни базисів теж можлива ситуація, коли опорні плани не будуть змінюватися, тобто розв'язки не будуть поліпшуватися.

Але кількість базисів скінченна, тому після деякого числа даремних кроків можливі такі ситуації.

1. За кілька кроків знайдено оптимальний план (усі оцінки вільних змінних невід'ємні).
2. При розв'язанні з'ясувалося, що задача не має розв'язку, цільова функція за наявних обмежень на змінні необмежено зростає.
3. За кілька даремних кроків одержано новий опорний план.
4. Через кілька кроків одержано вже раніше досягнутий опорний план. Така ситуація називається *зацикленням.* У прикладній математиці розроблено заходи щодо уникнення таких ситуацій. Досвід розв'язання практичних ЗЛП свідчить, що зациклення є малоймовірною ситуацією.

7.3. Алгоритм симплексного методу розв'язання невироджених задач лінійного програмування

Узагальнення властивостей опорних планів дозволяє розбудувати один із можливих алгоритмів розв'язання ЗЛП так званого симплексного методу.

Пропонований алгоритм гарантує, що при наявності хоча б одного опорного плану за скінченне число кроків буде або одержано оптимальний план, або доведено відсутність розв'язків досліджуваної ЗЛП. На кожному кроці (ітерації) обчислюється новий опорний план, якому відповідає більше значення цільової функції, ніж вона його мала за попереднього плану. Після чергового кроку оцінюються досягнутий опорний план і доцільність виконання наступного кроку – пошуку нового опорного плану.

Розглянемо послідовність дій при реалізації алгоритму.

1. Наявну лінійну модель практичної задачі привести до канонічної форми за умови пошуку найбільшого значення цільової функції.
2. Знайти вихідний опорний план або будь-який базисний розв'язок. Перехід від довільного базисного розв'язку до опорного описано у підрозділі 7.2.1. Тому для подальшого приймемо, що маємо опорний план.
3. За наявного опорного плану представити базисні змінні та цільову функцію через вільні змінні. Записати симплекс-таблицю T (табл. 7. 2).
4. Склавши симплекс-таблицю T, дослідити наявний опорний план:

а) якщо у f-рядку немає від'ємних елементів, за винятком, можливо, вільного члена (значення цільової функції за наявного опорного плану), то план оптимальний. Якщо до того ж немає і нульових елементів, то маємо один оптимальний план. Якщо у f-рядку маємо хоча б один нульовий елемент, то оптимальних планів незліченна множина;

б) якщо у f-рядку є хоча б один від'ємний елемент, якому відповідає стовпець з недодатними елементами, то цільова функція в області допустимих значень змінних необмежена;

в) якщо у f-рядку є хоча б один від'ємний елемент, а у відповідному йому стовпці – хоча б один додатний, то можна наявний опорний план поліпшити, як це описано в 7.2.2. Одержаний на наступному кроці опорний план знову треба досліджувати на оптимальність.

Реалізація описаного алгоритму дасть змогу або знайти оптимальний план (можливо, не один), або переконатися у відсутності розв'язку досліджуваної ЗЛП. Як наголошувалося раніше, за наявності розв'язку він буде досягнутий за скінченне число кроків.

Зауваження 1. Для реалізації алгоритму симплексного методу необхідно мати вихідний опорний план, пошук якого є суттєвим етапом у розв'язанні ЗЛП і вимагає

спеціальних досліджень фахівців прикладної математики для задач з великою кількістю змінних.

Зауваження 2. Розташування базисних змінних у вихідній симплекс-таблиці може відповідати довільним їх номерам, а не обов'язково першим номерам і відповідним позиціям в основній частині таблиці.

7.4. Пошук вихідного опорного плану

Описаний алгоритм розв'язання ЗЛП у канонічній формі ефективний при наявності вихідного опорного плану. Такий план можна знайти, користуючись методом Жордана-Гаусса щодо розв'язку відповідної СЛАР, але цей підхід не завжди зручний, особливо за умови, коли задача має кілька десятків і більше невідомих. Розглянемо можливості використання симплексних перетворень для знаходження вихідного опорного плану, тобто перетворення умови ЗЛП з виду (1) – (3) у вид (6), (7).

Припустимо, що ранг матриці коефіцієнтів СЛАР канонічної форми ЗЛП дорівнює r $(r \le m)$. Запишемо систему у спеціальному вигляді:

$$\left.\begin{array}{l} 0 = a_{10} + a_{11}(-x_1) + a_{12}(-x_2) + \ldots + a_{1n}(-x_n) \\ \ldots \qquad\qquad\qquad \ldots \\ 0 = a_{m0} + a_{m1}(-x_1) + a_{m2}(-x_2) + \ldots + a_{mn}(-x_n) \end{array}\right\} \qquad (13)$$

Приймемо, що a_{i0} $(i = \overline{1,m})$ *невід'ємні,* в протилежному випадку відповідне рівняння домножимо на "–1".

Цільову функцію запишемо так:

$$f(\overline{x}) = 0 + (c_1)(-x_1) + (c_2)(-x_2) + \ldots + (-c_n)(x_n). \qquad (14)$$

Тепер запишемо для (13) та (14) симплекс-таблицю T, схема якої наведена в табл. 7.5.

Таблиця 7.5.

	Вільні члени	$-x_1$	$-x_2$	...	$-x_n$
$0 =$	a_{10}	a_{11}	a_{12}	...	a_{1n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
$0 =$	a_{m0}	a_{m1}	a_{m2}	...	a_{mn}
$f =$	0	$-c_1$	$-c_2$	...	$-c_n$

Будемо послідовно виконувати симплексні перетворення над елементами відповідних таблиць T, вибираючи провідні елементи серед додатних чисел кожної чергової таблиці T, за мінімальним симплексним відношенням. Такий вибір провідних елементів і симплексних відношень забезпечить невід'ємність вільних членів у ході перетворень. На f-рядок при виборі провідного елемента поки що не звертаємо уваги; на першому етапі перетворень (до одержання опорного плану) цей рядок дає можливість представити цільову функцію через вільні змінні, узгоджені з досягнутим планом.

На кожному кроці перетворень одна із змінних x та один із нулів лівого стовпця міняються місцями. Через r кроків ($r \le m$) симплексних перетворень одержуємо таблицю, загальний вигляд якої представлений в табл. 7.6.

Таблиця 7.6.

	Вільні члени	0	…	0	$-x_{r+1}$	…	$-x_n$
$x_1 =$	b_{10}	b_{11}	…	b_{1r}	$b_{1,r+1}$	…	b_{1n}
$x_2 =$	b_{20}	b_{21}	…	b_{2r}	$b_{2,r+1}$	…	b_{2n}
⋮	⋮	⋮	…	⋮	⋮	…	⋮
$x_r =$	b_{r0}	b_{r1}	…	b_{rr}	$b_{r,r+1}$	…	b_{rn}
$0 =$	$b_{r+1,0}$	$b_{r+1,1}$	…	$b_{r+1,r}$	0	…	0
⋮	⋮	⋮	…	⋮	⋮	…	⋮
$0 =$	b_{m0}	b_{m1}	…	b_{mr}	0	…	0
$f =$	b_{00}	b_{01}	…	b_{0r}	$b_{0,r+1}$	…	b_{0n}

Вихідна система буде сумісною, якщо в завершальній таблиці 7.6 вільні члени $b_{r+1,0}$, $b_{r+2,0}$, …, $b_{m,0}$ дорівнюватимуть нулю. Якщо хоча б один із них відмінний від нуля, система буде несумісною, тобто ЗЛП не має розв'язку. Для будь-якого рівняння при $i = \overline{1,r}$ з використанням даних табл. 7.6 маємо:

$$x_i = b_{i0} + b_{i1} \cdot 0 + b_{i2} \cdot 0 + \ldots + b_{ir} \cdot 0 + b_{i,r+1}(-x_{r+1}) + \ldots + b_{in}(-x_n).$$

Із наведеної залежності видно, що елементи стовпців, які відповідають стовпцям, поміченим "0", обчислювати не варто; такі стовпці вилучають із запису в таблиці. Вилучають і рядки з усіма нульовими елементами; наявність таких рядків свідчить про лінійну залежність відповідних рівнянь вихідної СЛАР.

Якщо в ході перетворень виникла ситуація, що для певного нульового рядка вільний член додатний, а решта елементів не більше нуля, то рівняння та вихідна система не мають додатних розв'язків, а ЗЛП – рішення.

За умови, що вихідна СЛАР (13) сумісна в області невід'ємних розв'язків, за наведених симплексних перетворень стосовно системи (13) через r ($r \le m$) кроків одержимо таблицю типу 7.6, в якій буде записано опорний план, що відповідає базису $\{x_1, x_2, \ldots, x_r\}$ (нумерація змінних базису умовна), та представлення цільової функції через відповідні вільні змінні. Наведений алгоритм дозволяє уніфікувати послідовність реалізації усіх етапів симплексного методу розв'язання ЗЛП у канонічній формі.

Приклад пошуку вихідного опорного плану за наведеним алгоритмом: знайти найбільше значення

$$f = 2x_1 - 2x_2 - x_3 + 6x_4$$

за умови

$$\left.\begin{aligned} x_1 + x_2 + x_3 + 3x_4 &= 16 \\ -x_1 + 3x_2 + x_3 - x_4 &= 8 \\ x_j \ge 0 \ \ (j = 1,\ 2,\ 3,\ 4). \end{aligned}\right\}$$

Запишемо вихідну табл. 7.7.

Таблиця 7.7.

	Вільні члени	$-x_1$	$-x_2$	$-x_3$	$-x_4$
$0 =$	16	1	1	1	3
$0 =$	8	–1	3	1	–1
$f =$	0	–2	2	1	–6

Вибравши за провідний елемент "1" в стовпці "$-x_1$", одержимо симплекс-таблицю, записану в лівій частині табл. 7.8 (оскільки "0" та "$-x_1$" міняються місцями, перший стовпець опускаємо).

Таблиця 7.8.

	Вільні члени	$-x_2$	$-x_3$	$-x_4$	b_{i0}/b_{is}		Вільні члени	$-x_2$	$-x_4$	b_{i0}/b_{is}
x_1	16	1	1	3	16	x_1	4	–1	2	2
0=	24	4	2	2	12	x_3	12	2	1	12
$f =$	32	4	3	0		$f =$	–4	–2	–3	

Вибравши за провідний елемент "2" у стовпці "$-x_3$" та опустивши цей стовпець, одержуємо вихідний опорний план та представлення функції f при відповідному базисі.

Завершимо розв'язання задачі, скориставшися симплекс-методом. У правій частині табл. 7.9 стовпці, які відповідають від'ємним елементам "–2" та "–3" f-рядка мають додатні елементи, тому досягнутий план можна поліпшити, скориставшися розташуванням найменших оцінок у f-рядку та симплексним відношенням.

Таблиця 7.9.

БЗ	Вільні члени	БЗ		b_{i0}/b_{is}	БЗ	Вільні члени	БЗ		b_{i0}/b_{is}
		$-x_2$	$-x_1$				$-x_3$	$-x_1$	
$x_4=$	2	–1/2	1/2		$x_4=$	4	1/5	2/5	
$x_3=$	10	5/2	–1/2	4	$x_2=$	4	2/5	-1/5	
$f=$	2	–7/2	3/2	–	$f=$	16	7/5	4/5	–

За нового базису $\{x_3; x_4\}$ маємо опорний план $\bar{x} = (0; 0; 10; 2)$ та відповідно значення $f = 2$, але досягнутий план не оптимальний, бо в f-рядку є від'ємний елемент "–7/2". Виконавши симплексні перетворення з провідним елементом "5/2", одержимо новий базис $\{x_2; x_4\}$ та опорний план $\bar{x} = (0; 4; 0; 4)$, за якого в f-рядку всі оцінки додатні, тобто досягнуто оптимальний план, для якого $f = 16$.

7.5. Симплексний метод з використанням фіктивного базису

У розглянутому варіанті симплексного методу розв'язання ЗЛП канонічної форми вихідний опорний план визначався шляхом пошуку одиничного базису СЛАР за методом Жордана-Гаусса або з використанням симплексних перетворень стосовно рівнянь-обмежень, представлених у спеціальному вигляді (13). На практиці широко використовується так званий метод фіктивного базису (M-метод), при якому пошук вихідного опорного плану суттєво спрощується.

Маючи канонічну форму ЗЛП (1) – (3), можна побудувати так звану *розширену* задачу. За умови, що у СЛАР (2) всі праві коефіцієнти невід'ємні (якщо це не так, то відповідне рівняння необхідно помножити на –1), в ліву частину кожного з рівнянь додамо по одній невід'ємній змінній x_{n+i} $(i = \overline{1, m})$, які назвемо *фіктивними*. Лінійну форму, яка визначає цільову функцію, зменшимо на суму фіктивних змінних, помножену на будь-яке велике додатне число M. Як наслідок таких перетворень одержуємо так звану *M-задачу*:

$$F = \sum_{j=1}^{n} c_j x_j - M \sum_{i=1}^{m} x_{n+i} \qquad (\max); \tag{15}$$

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+i} = b_i \qquad (i = \overline{1, m}) \quad ; \tag{16}$$

$$x_j \geq 0 \qquad (j = \overline{1, n+m}). \tag{17}$$

У СЛАР (16) змінні $x_{n+i} = b_i$ $(i = \overline{1, m})$ створюють базис, який називається *фіктивним*, або *штучним*. За умови $x_1 = x_2 = \ldots x_n = 0$ одержуємо вихідний опорний план $(0, 0, \ldots, 0, b_1, b_2, \ldots, b_m)$ M-задачі (15), (16), (17). Наголосимо: якщо $(x_1, x_2, \ldots, x_n, 0, \ldots, 0)$ – деякий план M-задачі, то $(x_1, x_2, \ldots, x_n)$ – план вихідної задачі і $F(x_1, x_2, \ldots, x_n, 0, \ldots, 0) = f(x_1, x_2, \ldots, x_n)$.

Розв'язок вихідної задачі (1) – (3) пов'язаний з результатами розв'язання M-задачі згідно з теоремою.

1. Якщо розширена задача має оптимальний розв'язок, у якому всі фіктивні змінні дорівнюють нулю, то, вилучивши значення фіктивних змінних, одержимо оптимальний план вихідної задачі.
2. Якщо розширена задача має *оптимальний розв'язок, в якому хоча б одна фіктивна змінна відмінна від нуля, то вихідна задача не має розв'язку*, оскільки система її обмежень не сумісна.
3. Якщо розширена задача не має розв'язку, то вихідна задача теж не має його.

Зауваження 1. Недоцільно залучати фіктивні змінні в рівняння зі змінними, яким відповідають стовпці з нульовими коефіцієнтами, за винятком одиниці для даного рівняння.

Зауваження 2. Розширена задача може бути розв'язана за будь-якої не фіксованої конкретно величини M, але для цього необхідно змінити формування симплекс-таблиці ЗЛП.

Оскільки і цільова функція, і обмеження M-задачі лінійні, то доцільно відокремити вплив дійсних і фіктивних змінних на величину цільової функції розширеної задачі. Для досягнення цієї мети необхідно врахувати, що цільова функція F розширеної задачі має дві суми стосовно вихідних та фіктивних змінних, тому в симплексній таблиці для F доцільно мати два рядки. Тоді ознаку оптимальності спочатку перевіряють по другому рядку, який відповідає сумі $M \sum x_{n+i}$, бо знак коефіцієнта при кожному вільному невідомому в цільовій функції F за фіктивного базису визначається доданком, залежним від M. Нагадуємо, що при виборі фіктивного базису F необхідно представити через вільні змінні:

$$F = \sum_{j=1}^{n} c_j x_j + M\left(\sum_{i=1}^{m} a_{i0} - \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} x_j\right) = \sum_{j=1}^{n} c_j x_j + M \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij} x_j - M \sum_{i=1}^{m} a_{i0}.$$

Після того, як внаслідок симплекс-перетворень усі елементи M - рядка стануть рівними нулю, ознаку оптимальності перевіряють за f-рядком, який відповідає $\sum_{j=1}^{n} c_j x_j$.

Увага! Коли з базису вилучають фіктивну змінну, тобто прилучають її до вільних змінних, то відповідний стовпець у симплексній таблиці вилучають.

Приклад. Використовуючи метод фіктивного базису, знайти найбільше значення

$$f = x_1 + 2x_2 + 3x_3 - x_4$$

за умови, що змінні мають задовольняти СЛАР:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 \quad &= 15 \\ 2x_1 + x_2 + 5x_3 \quad &= 20 \\ x_1 + 2x_2 + x_3 + x_4 &= 10 \\ x_j \geq 0 \quad (j = \overline{1,4}). \end{aligned}$$

Сформулюємо M-задачу. Оскільки при x_4 коефіцієнт маємо лише в третьому рівнянні, то x_4 можна вибрати за базисну змінну, тому фіктивні змінні введемо до першого та другого рівнянь. За таких умов M-задача сформулюється так: знайти найбільше значення

$$F = x_1 + 2x_2 + 3x_3 - x_4 - M(x_5 + x_6)$$

за умови:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + \quad + x_5 \quad &= 15 \\ 2x_1 + x_2 + 5x_3 + \quad + x_6 &= 20 \\ x_1 + 2x_2 + x_3 + x_4 \quad &= 10 \\ x_j \geq 0 \quad (j = \overline{1,6}). \end{aligned}$$

Одержано задачу з неповним фіктивним базисом, але це не порушує хід розв'язання.

За вільні змінні вибираємо x_1, x_2, x_3. Представимо F через вільні змінні:

$$F = x_1 + 2x_2 + 3x_3 - 10 + x_1 + 2x_2 + x_3 - M(15 - x_1 - 2x_2 - 3x_3 + 20 - 2x_1 - x_2 - 5x_3) =$$
$$= 2x_1 + 4x_2 + 4x_3 - 10 - M(35 - 3x_1 - 3x_2 - 8x_3).$$

Запишемо T-таблицю, як це показано в табл. 7.10.

Провідний стовпець вибраний за елементом "– 8" в M-рядку рядка F, провідний елемент "5" – за найменшим симплексним відношенням. Після симплексного перетворення з базису вилучається змінна x_6 та відповідний їй стовпець.

Таблиця 7.10.

БЗ		Вільні члени	Різні змінні			b_{i0}/b_{is}
			$-x_1$	$-x_2$	$-x_3$	
$x_5 =$		15	1	2	3	5
$x_6 =$		20	2	1	5	4
$x_4 =$		10	1	2	1	10
F		–10	–2	–4	–4	
	M	–35	–3	–3	–8	

Поточна T-таблиця записана в лівій частині табл. 7.11.

Таблиця 7.11.

БЗ		Вільні члени	ВЗ		b_{i0}/b_{is}	БЗ	Вільні члени	ВЗ	b_{i0}/b_{is}	БЗ	Вільні члени	ВЗ
			$-x_1$	$-x_2$				$-x_1$				$-x_4$
$x_5 =$		3	–1/5	7/5	15/7	$x_2 =$	15/7	–1/7	–	$x_2 =$	5/2	–
$x_3 =$		4	2/5	1/5	20	$x_3 =$	25/7	3/7	25/3	$x_3 =$	5/2	–
$x_4 =$		6	3/5	9/5	10/3	$x_4 =$	15/7	6/7	15/6	$x_1 =$	5/2	–
F		6	–2/5	–16/5	–	$f =$	90/7	–6/7	–	$f =$	15	1
	M	–3	1/5	–7/5	–	–	–	–	–	–	–	–

У M-рядку рядка F маємо –7/5. Знайшовши найменше симплексне відношення 15/7, вибираємо провідний елемент 7/5 та виконуємо симплексне перетворення, вилучивши з базису фіктивну змінну x_5 та відповідний стовпець. Поточна T-таблиця записана в середній частині табл. 7.11. T-таблиця відповідає вихідній ЗЛП і має опорний план цієї задачі $\bar{x} = (0; 15/7; 25/7; 15/7)$ та значення цільової функції $f = 90/7$. Але наявний план не оптимальний, бо f-рядок має від'ємний елемент –6/7, тому необхідно виконати ще одне симплексне перетворення з провідним елементом "6/7". Результати цього перетворення записані справа в табл. 7.11. Оскільки f-рядок не має від'ємних елементів, то досягнутий опорний план $\bar{x} = (5/2; 5/2; 5/2; 0)$ оптимальний, а найбільше значення цільової функції $f = 15$.

При реалізації алгоритму симплексного методу M-задачі з базисних змінних у першу чергу необхідно вилучати фіктивні змінні з найменшою від'ємною оцінкою. Фіктивний вектор, вилучений з базису, в наступних перетвореннях використовувати

недоцільно, тому стовпець симплекс-таблиці, який йому відповідає, теж вилучають із наступної таблиці.

Розрахунки з розбудовою симплекс-таблиць виконуються до тих пір, поки або всі фіктивні змінні не будуть вилучені з базису, або серед оцінок, які відповідають фіктивним змінним, не буде від'ємних. У першому випадку пошук оптимального розв'язку продовжується з використанням дійсних змінних та досягнутого опорного плану. Поява другої ситуації вимагає додаткового аналізу і в більшості випадків означає, що вихідна ЗЛП не має розв'язку, а іноді має вироджений розв'язок.

Таким чином, при використанні методу фіктивного базису для розв'язання задачі лінійного програмування необхідно:

1) вихідній задачі надати канонічної форми;
2) сформулювати M-задачу (ввести фіктивні змінні);
3) записати вихідний опорний план M-задачі;
4) використовуючи симплекс-перетворення, або вилучити фіктивні змінні з вихідного базису, або виявити відсутність розв'язку вихідної задачі;
5) використовуючи одержаний базис та досягнутий опорний план, серед дійсних змінних знайти оптимальний розв'язок.

Питання для самоконтролю

1. Як розуміти поняття базису в n-мірному векторному просторі?
2. Що таке ранг матриці та ранг системи лінійних алгебраїчних рівнянь?
3. За яких умов система лінійних алгебраїчних рівнянь буде сумісною? За яких умов СЛАР буде мати єдиний розв'язок?
4. За яких умов СЛАР можна записати як лінійну залежність між відомими векторами?
5. Що таке базисні та вільні змінні у СЛАР?
6. Як записати загальний розв'язок СЛАР за умови, коли кількість рівнянь менша кількості невідомих?
7. Який запис має канонічна форма задачі лінійного програмування?
8. Що таке опорний план ЗЛП?
9. За якими формулами виконуються перерахунки коефіцієнтів СЛАР при заміні базису?
10. За яких умов доцільно змінювати досягнутий опорний план? Критерій оптимальності опорного плану.
11. Як і з якою метою будується фіктивний базис симплексного методу розв'язання ЗЛП?

ГЛАВА 8

СПРЯЖЕНІ ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ

8.1. Поняття спряженості (двоїстості)

У теорії дослідження розв'язання практичних задач, математичні моделі яких можна привести до лінійного програмування, доведено, що кожній задачі можна поставити у відповідність деяку іншу задачу лінійного програмування, яку називають *спряженою* (двоїстою) стосовно вихідної. До того ж у спряжених задачах вимоги щодо сенсу екстремумів протилежні: якщо у вихідній (першій) задачі за практичним сенсом необхідно знайти *найбільше значення цільової функції* за умови дотримання певних обмежень, то у спряженій (другій) задачі – *найменше значення* і навпаки. Цільова функція та обмеження спряженої задачі повністю визначаються умовами вихідної задачі. Вихідна та спряжена задачі утворюють пару взаємно спряжених задач, тому будь-яку з них можна розглядати як вихідну (першу). Теорія дослідження спряженості дозволяє суттєво розширити коло задач лінійного програмування, які мають як теоретичний, так і практичний інтерес.

Розглянемо дві **економічні спряжені задачі.** Підприємство у виробництві використовує m видів ресурсів, обсяги яких обмежені величинами $b_1, b_2, ..., b_m$, та може виробляти n найменувань продукції, обсяги яких позначимо $x_1, x_2, ..., x_n$. Відомими є обсяги витрат i-го виду ресурсу на одиницю продукції j-го найменування a_{ij} $(i=\overline{1, m};\ j=\overline{1, n})$, відомий і зиск c_j від реалізації одиниці продукції j-го найменування.

Задача 1. Визначити такий план виробництва продукції $\bar{x}^0 = (x_1, x_2, ..., x_n)$, який дав би можливість одержати найбільший зиск від реалізації за умови, що затрати ресурсів не можуть перевищувати їх відомих обсягів.

Математична модель задачі формулюється так: знайти найбільше значення функції

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

за умови

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i=\overline{1, m}), \tag{2}$$

$$x_j \ge 0 \quad (j=\overline{1, n}). \tag{3}$$

Спряжена задача за тих же вихідних економічних даних може бути сформульована так.

Задача 2. При яких можливих вартостях y_i одиниць i-их ресурсів виробництво продукції буде збитковим, тобто недоцільним, за модельованих економічних умов? Якщо обсяги ресурсів необхідно придбати в кількостях $b_1, b_2, ..., b_m$ при найменших витратах, то математична модель задачі 2 формулюється так:
знайти найменше значення функції

$$F = \sum_{i=1}^{m} b_i\, y_i \tag{4}$$

за умови

$$\sum_{i=1}^{m} a_{ij} y_i \geq c_j \quad (j = \overline{1, n}), \tag{5}$$

$$y_i \geq 0 \quad (i = \overline{1, m}), \tag{6}$$

($\sum_{i=1}^{m} a_{ij} y_i$ – обсяг витрат на виробництво одиниці продукції j-го найменування; (5) – умова, що витрати на одиницю продукції не менші, ніж зиск від реалізації, тобто виробництво j-ої продукції за такої умови збиткове).

У літературі відома й інша постановка спряженої задачі до вихідної задачі 1. Наведемо її. Уявимо, що підприємство за деяких обставин змушене реалізувати свої наявні ресурси, а не продукувати товар, але прагне одержати зиск, не менший тієї суми, яку можна чекати, організувавши виробництво. Природно, що покупець ресурсів прагне придбати їх за найменшу ціну. Необхідно визначити такі ціни на ресурси, щоб задовольнити наведені суперечливі вимоги.

Якщо вартість одиниці i-го ресурсу, як і раніше, позначаємо y_i, то зиск від реалізації усіх ресурсів буде визначатися функцією (4). Щоб уникнути порівняльних збитків при реалізації ресурсів, підприємство-продавець мусить так їх продавати, щоб вартість ресурсів $\sum_{i=1}^{m} a_{ij} y_i$, використовуваних на виробництво одиниці j-ої продукції, була не менше вартості c_j збуту виробу, тобто виконувалися умови (5).

За наведеного тлумачення спряженої задачі її математична модель тотожна раніше описаній (4) – (6).

Запропоновані математичні моделі можна розширити, наприклад, залучити питання про нарахування податку, про витрати на транспортування і т. ін., але основний сенс економічних залежностей віддзеркалений у задачах 1 та 2.

Задачі (1) – (3) та (4) – (6) утворюють пару спряжених (двоїстих) задач, і кожна з них може бути розглянута як вихідна. У задачі (4) – (6) цільова функція (4) визначає затрати підприємства на закупівлю ресурсів за цінами $y_1, y_2, ..., y_m$, а нерівності (5) віддзеркалюють умову, що коли зиск від реалізації одиниці

продукції j-го найменування не перевищуватиме витрати ресурсів на її виробництво, то виробляти продукцію за таких умов недоцільно.

Співставивши математичні моделі сформульованих задач та абстрагуючись від їх конкретного змісту, доцільно наголосити на таких їх особливостях:

1) кількість невідомих в одній із задач дорівнює кількості основних обмежень іншої задачі;

2) матриці коефіцієнтів систем обмежень транспоновані одна стосовно іншої;

3) знаки операцій відношень в обмеженнях (2) та (5) мають протилежний сенс;

4) коефіцієнти цільової функції однієї із задач є вільними членами в обмеженнях-нерівностях іншої задачі;

5) у спряжених задачах досліджуються екстремальні властивості протилежного сенсу.

Перейдемо до аналізу математичних моделей спряжених задач у загальному випадку.

Симетричні задачі. До цього класу відносять задачі, в яких обмеження визначаються лише нерівностями. Змінні в спряжених задачах цього класу невід'ємні. Скориставшися матричною формою запису ЗЛП, спряжені задачі можна записати так:

Вихідна задача	**Спряжена задача**
1. Знайти *найбільше* значення $f = \sum_{j=1}^{n} c_j x_j = \bar{c}\,\bar{x}$ за умови $A\bar{x}^T \leq \bar{b}^T$, $\bar{x} \geq \bar{0}$.	1. Знайти *найменше* значення $F = \sum_{i=1}^{m} b_i y_i = \bar{b}\,\bar{y}$ за умови $A^T\bar{y}^T \geq \bar{c}^T$, $\bar{y} \geq \bar{0}$.
2. Знайти *найменше* значення $f = \sum_{j=1}^{n} c_j x_j = \bar{c}\,\bar{x}$ за умови $A\bar{x}^T \geq \bar{b}^T$, $\bar{x} \geq \bar{0}$.	2. Знайти *найбільше* значення $F = \sum_{i=1}^{m} b_i y_i = \bar{b}\,\bar{y}$ за умови $A^T\bar{y}^T \leq \bar{c}^T$, $\bar{y} \geq \bar{0}$.

Несиметричні задачі. При визначенні спряжених задач цього класу доцільно наголосити, що стосовно змінних спряженої задачі відсутня умова їх невід'ємності.

Вихідна задача	Спряжена задача
1. Знайти *найбільше* значення $f=\sum_{j=1}^{n} c_j x_j = \bar{c}\ \bar{x}$ за умови $A\bar{x}^T \le \bar{b}^T$, $\bar{x} \ge \bar{0}$.	1. Знайти *найменше* значення $F=\sum_{i=1}^{m} b_i y_i = \bar{b}\ \bar{y}$ за умови $A^T\bar{y}^T \ge \bar{c}^T$.
2. Знайти *найменше* значення $f=\sum_{j=1}^{n} c_j x_j = \bar{c}\ \bar{x}$ за умови $A\bar{x}^T \ge \bar{b}^T$, $\bar{x} \ge \bar{0}$.	2. Знайти *найбільше* значення $F=\sum_{i=1}^{m} b_i y_i = \bar{b}\ \bar{y}$ за умови $A^T\bar{y}^T \le \bar{c}^T$.

Надалі будемо розглядати лише спряжені симетричні задачі. Сформулюємо правила постановки спряженої задачі за умови, що обмеження представлені нерівностями.

1. Усі обмеження-нерівності вихідної задачі привести до нерівностей одного сенсу: за умови пошуку найбільшого значення цільової функції відношення має бути "≤", за умови найменшого значення – "≥". За таких умов усі нерівності, які не задовольняють названим, множать на –1.

2. Записати вектор невідомих спряженої задачі $\bar{y} = (y_1, y_2, ..., y_m)$ (m – кількість основних обмежень вихідної задачі).

3. Записати цільову функцію спряженої задачі. За коефіцієнти при невідомих y_i $(i=\overline{1,m})$ обираються вільні члени обмежень-нерівностей вихідної задачі після виконання п. 1.

4. Записати систему обмежень спряженої задачі. Для цього транспоновану матрицю системи обмежень вихідної задачі множать на вектор-стовпець $\bar{y}$. Названого результату можна досягти, перемноживши вектор-рядок $\bar{y}$ справа на матрицю обмежень, тому що виконується відношення $A^T\bar{y}^T = \bar{y}A$. За вільні члени в нерівностях обирають коефіцієнти при невідомих у цільовій функції вихідної задачі. Операція відношення визначається згідно з п. 1.

5. Визначити характер екстремуму цільової функції спряженої задачі як протилежний за змістом екстремуму вихідної задачі.

6. Додати умови невід'ємності невідомих спряженої задачі.

Приклад. Сформулювати спряжену задачу для задачі: знайти найбільше значення

$$f = 2x_1 + x_2$$

за умови

$$\begin{aligned} -x_1 + 3x_2 &\le 32 \\ 3x_1 + 8x_2 &\le 137 \\ 7x_1 - 5x_2 &\le 125 \\ -2x_1 + x_2 &\ge 5 \\ x_1 \ge 0;\ x_2 &\ge 0. \end{aligned}$$

Згідно з п. 1 правил формулювання спряженої задачі, слід помножити четверту нерівність на –1. У спряженій задачі мають бути чотири невідомі величини (y_1, y_2, y_3, y_4), оскільки вихідна задача має чотири обмеження-нерівності. Цільова функція спряженої задачі $F = 32y_1 + 137y_2 + 125y_3 - 5y_4$. Система обмежень-нерівностей має вигляд:

$$\bar{y}\ A = (y_1, y_2, y_3, y_4)\begin{pmatrix} -1 & 3 \\ 3 & 8 \\ 7 & -5 \\ 2 & -1 \end{pmatrix} = \begin{cases} -y_1 + 3y_2 + 7y_3 + 2y_4 \ge 2 \\ 3y_1 + 8y_2 - 5y_3 - y_4 \ge 1. \end{cases}$$

Спряжена задача формулюється так: знайти найменше значення

$$F = 32y_1 + 137y_2 + 125y_3 - 5y_4$$

за умови

$$\begin{cases} -y_1 + 3y_2 + 7y_3 + 2y_4 \ge 2 \\ 3y_1 + 8y_2 - 5y_3 - y_4 \ge 1 \end{cases}$$
$$y_1 \ge 0;\ y_2 \ge 0;\ y_3 \ge 0;\ y_4 \ge 0.$$

8.2. Відповідність між змінними спряжених задач

Для того, щоб знайти відповідність між змінними спряжених симетричних задач (1) – (3) та (4) – (6), необхідно надати їм канонічної форми. Для вихідної задачі (1) – (3) введемо додаткові змінні $x_{n+1}, x_{n+2}, ..., x_{n+m}$, для спряженої – $y_{m+1}, y_{m+2}, ..., y_{m+n}$. Вихідна задача формулюється так: знайти *найбільше* значення функції

$$f = \sum_{j=1}^{n} c_j x_j \qquad (7)$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+i} = b_i \quad (i = \overline{1, m}) \tag{8}$$

$$x_k \geq 0 \quad (k = \overline{1, n+m}). \tag{9}$$

Спряжена задача: знайти *найменше* значення функції

$$F = \sum_{i=1}^{m} b_i \, y_i \tag{10}$$

за умови

$$\sum_{i=1}^{m} a_{ij} y_i - y_{m+j} = c_j \quad (j = \overline{1, n}) \tag{11}$$

$$y_k \geq 0 \quad (k = \overline{1, n+m}). \tag{12}$$

У системах обмежень (8) та (11) кількість змінних однакова і дорівнює $n + m$. Базисними у вихідній системі (7) – (9) будуть змінні $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$, а у спряженій задачі (10) – (12) $y_{m+1}, y_{m+2}, \ldots, y_{m+n}$. Між змінними спряжених задач існує взаємо відповідність:

x_{n+1}	x_{n+2}	...	x_{n+m}	x_1	x_2	...	x_n
↑↓	↑↓		↑↓	↑↓	↑↓	.	↑↓
y_1	y_2	...	y_m	y_{m+1}	y_{m+2}	...	y_{m+n}

Взаємозв'язок між спряженими задачами (7) – (9) і (10) – (12) та їх змінними доцільно зобразити в табл. 8.1.

Таблиця 8.1.

ВЗ	БЗ	F	БЗ			
			y_{m+1}	y_{m+2}	...	y_{m+n}
		1	ВЗ			
			$-x_1$	$-x_2$	...	$-x_n$
y_1	x_{n+1}	b_1	a_{11}	a_{12}	...	a_{1n}
y_2	x_{n+2}	b_2	a_{21}	a_{22}	...	a_{2n}
⋮	⋮	⋮	⋮	⋮	...	⋮
y_m	x_{n+m}	b_m	a_{m1}	a_{m2}	...	a_{mn}
1	f		$-c_1$	$-c_2$	...	$-c_n$

Значення базисних змінних вихідної задачі визначаються з використанням

елементів стовпця F, а значення змінних спряженої задачі – рядка f. Так, для визначення базисної змінної x_{n+i} $(i = \overline{1, m})$ необхідно до величини b_i, розташованої у стовпці F, додати скалярний добуток рядка вільних змінних (ВЗ) x_j $(j = \overline{1, n})$ на i-ий рядок матриці A:

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

Такий результат узгоджується з (8). Для визначення базисної змінної y_{m+j} $(j = \overline{1, n})$ спряженої задачі необхідно до величини $-c_j$, розташованої в рядку f, додати скалярний добуток стовпця вільних змінних (ВЗ) y_i $(i = \overline{1, m})$ на j-ий стовпець матриці A:

$$y_{m+j} = -c_j + \sum_{i=1}^{m} y_i a_{ij} .$$

Цей результат узгоджується з (11).

Таким чином, з допомогою табл. 8.1 можна записати систему рівнянь-обмежень канонічних форм спряжених задач лінійного програмування.

Розв'язавши одну зі спряжених задач та зваживши на відповідність між їх змінними, одержують розв'язок іншої спряженої задачі.

8.3. Теореми спряженості

Доведемо основну нерівність теорії спряженості. Припустимо, що маємо симетричні спряжені задачі (1) – (3) та (4) – (6). Для будь-яких допустимих планів цих задач

$$\bar{x} = (x_1, x_2, \ldots, x_n) \text{ та } \bar{y} = (y_1, y_2, \ldots, y_m) \tag{13}$$

завжди виконується нерівність

$$f(\bar{x}) \leq F(\bar{y}), \text{ тобто } \sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i . \tag{14}$$

Розглянемо системи обмежень-нерівностей (2) та (5) спряжених задач. Помножимо будь-яке i-те обмеження (2) на y_i та виконаємо додавання всіх обмежень

за індексом i $(i=\overline{1,\ m})$:

$$\sum_{i=1}^{m}(a_{i1}x_1+a_{i2}x_2+\ldots+a_{in}x_n)\,y_i \le \sum_{i=1}^{m} b_i y_i\,.$$

Розкривши зліва дужки, цю нерівність можна записати так:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}y_i x_j \le \sum_{i=1}^{m} b_i y_i\,. \qquad (15)$$

Аналогічні перетворення виконаємо стосовно (5), множачи j-те обмеження на x_j та виконуючи додавання за індексом j $(j=\overline{1,\ n})$:

$$\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}y_i x_j \ge \sum_{j=1}^{n} c_j x_j\,. \qquad (16)$$

Враховуючи, що результат додавання не залежить від порядку розташування доданків, наголосимо, що ліві частини нерівностей (15) та (16) тотожні; це дає підстави встановити відношення між їх правими частинами. Отже, нерівність (14) доведена.

Доведемо важливу **теорему стосовно достатньої ознаки оптимальності**. Якщо $\bar{x}^*=(x_1^*,\ x_2^*,\ \ldots,\ x_n^*)$ та $\bar{y}^*=(y_1^*,\ y_2^*,\ \ldots,\ y_m^*)$ – допустимі плани спряжених задач (1) – (3) і (4) – (6), для яких виконується рівність

$$f(\bar{x}^*)=F(\bar{y}^*),$$

то x^* є оптимальним планом задачі (1) – (3), а y^* – задачі (4) – (6).

Доведення. Припустимо, що $\bar{x}_1$ – будь-який допустимий план задачі (1) – (3). Згідно з (14) маємо залежність $f(\bar{x}_1) \le F(\bar{y}^*)$, але, за умовою теореми, $f(\bar{x}^*)=F(\bar{y}^*)$, отже, $f(\bar{x}_1) \le f(\bar{x}^*)$. Тобто $\bar{x}^*$ є оптимальним планом задачі (1) – (3), тому що за нього цільова функція досягає не меншого значення, ніж за будь-якого іншого допустимого плану; отже, $\bar{x}^*=\bar{x}^0$.

Враховуючи, що для спряженої задачі необхідно знайти найменше значення цільової функції $F(\bar{y})$, аналогічно доводимо, що $\bar{y}^*$ є оптимальним планом задачі (4) – (6), тобто $\bar{y}^*=\bar{y}^0$.

Тепер наведемо без доведення основні теореми теорії спряженості.

Теорема існування оптимальних планів. Для того, щоб спряжені задачі мали оптимальні плани, необхідно і достатньо, щоб кожна з них мала хоча б один допустимий план.

Розглянемо **теореми обґрунтування зв'язків між планами спряжених задач**.

Теорема 1. Якщо одна зі спряжених задач має оптимальний план, то інша також має оптимальний план; до того ж за будь-яких оптимальних планів $\bar{x}^0$ і $\bar{y}^0$ виконується умова:

$$f(\bar{x}^0)=F(\bar{y}^0) \Rightarrow \sum_{j=1}^{n} c_j x_j^0 = \sum_{i=1}^{m} b_i y_i^0\,. \qquad (17)$$

З теореми 1 випливає, що коли одна зі спряжених задач не має оптимального розв'язку ($f(\bar{x}) \to +\infty$ або $F(\bar{y}) \to -\infty$), то інша теж не має розв'язку.

Для сформульованих на початку глави моделей економічних задач планування виробництва при обмеженості ресурсів та дослідження питання доцільності виробництва за умов закупівлі ресурсів ця теорема стверджує, що коли максимально можливий зиск від реалізації продукції буде дорівнювати найменшій можливій вартості використовуваних ресурсів, то зникає доцільність виробництва.

Теорема 2. Для того, щоб допустимі плани $\bar{x}^*$ та $\bar{y}^*$ вихідної (1) – (3) та спряженої (4) – (6) задач були оптимальними, необхідно і достатньо, щоб виконувалися відношення:

якщо $\sum_{j=1}^{n} a_{ij}x_j^* < b_i$ (для будь-якого i $(i=\overline{1,\ m})$), то відповідне значення $y_i^*=0$;

якщо $\sum_{i=1}^{m} a_{ij}y_i^* > c_j$ (для будь-якого j $(j=\overline{1,\ n})$), то відповідне значення $x_j^*=0$

$(\bar{x}^*=(x_1^*,\ x_2^*,\ \ldots,\ x_n^*);\ \ \bar{y}^*=(y_1^*,\ y_2^*,\ \ldots,\ y_m^*))$.

8.4. Економічний зміст оптимальних планів спряжених задач

Спряжені задачі різних лінійних математичних моделей реальних практичних задач мають різний зміст. Розглянемо з економічної точки зору більш детально зв'язок спряжених задач, сформульованих на початку цієї глави. Розв'язок вихідної задачі планування обсягів продукції за прийнятих економічних умов дозволяє одержати оптимальний план, який забезпечує максимальний зиск від реалізації продукції, а розв'язок спряженої задачі – визначити розміри граничних цін на ресурси: якщо ціни на ресурси такі, що вартість виробництва одиниці певної продукції не менше її ринкової ціни, то зникає доцільність виробництва.

Оптимальні плани названих спряжених задач пов'язані не лише умовою рівності значень відповідних цільових функцій. Між ними існують і інші зв'язки. Використовуючи теорему 2 спряженості, встановлюють зв'язок між обмеженнями

та змінними спряжених задач. Як вихідну слід розглядати задачу планування обсягів виробництва, як спряжену – оцінки вартості ресурсів.

Для дослідження взаємозв'язків між змінними ще раз стисло проаналізуємо хід пошуків оптимальних планів спряжених задач і з'ясуємо економічний зміст деяких результатів математичних перетворень. Прилучивши додаткові змінні x_{n+i} ($i=\overline{1,\ m}$), наявні обмеження-нерівності (2) вихідної задачі перетворюємо у рівності, щоб скористатися симплекс-методом для пошуків оптимального плану. Якщо маємо оптимальний план задачі (1) – (3) і його компоненти підставимо в обмеження (2), то для деякого i-го відношення (2) одержимо *рівність*, тобто свідчення того, що за оптимального плану i-й ресурс використовується повністю. Наявність *нерівності* свідчить, що відповідний ресурс за оптимального плану не використовується в заданому обсязі, маємо залишок величини x_{n+k}, де k – номер залежності в (2). Додаткові змінні y_{m+j} ($j=\overline{1,\ n}$), які використовують, щоб обмеження-нерівності (5) привести до рівностей, свідчать, на скільки собівартість виробництва одиниці j-ої продукції за заданих цін y_i ($i=\overline{1,\ m}$) на ресурси перевищує її ринкову ціну c_j. Якщо за деякого j при оптимальному розв'язку спряженої задачі маємо, що y_{m+j} більше нуля, то це свідчить, що собівартість j-ої одиниці перевищує її ринкову ціну, виробництво такої продукції (витрати ресурсів) за наявної технології з економічної точки зору недоцільне. Якщо за наявного оптимального розв'язку спряженої задачі деяке значення $y_{m+j}=0$, то це свідчить про те, що собівартість виробництва відповідної j-ої продукції дорівнює її ринковій ціні. Таким чином, якісний аналіз оптимального розв'язку спряженої задачі (4) – (6) показує, що в план виробництва доцільно включити лише "рентабельні" види продукції, але треба зважити, що критерій рентабельності надто спрощений: собівартість продукції не перевищує її ринкової вартості. За необхідності такий критерій можна "підсилити", прилучивши певні коефіцієнти.

Виконавши якісний аналіз порівняння оптимальних розв'язків спряжених задач за їх економічного тлумачення, з'ясуємо, що показують числові величини їх компонент. В околі точки оптимального плану цільова функція задачі може бути представлена залежністю (17), яка пов'язує обсяги наявних ресурсів і компоненти оптимального плану спряженої задачі.

Таким чином, маємо:

$$y_i^0=\frac{\partial f(\bar{x}^0)}{\partial b_i}\quad (i=\overline{1,\ m}). \tag{18}$$

Залежність (18) є наслідком теореми 1 та умови (14). Вона показує, що значення змінних y_i^0 в оптимальному плані спряженої задачі є оцінками впливу вільних членів рівнянь системи обмежень вихідної задачі на величину найбільшого

значення її цільової функції. Тобто приріст цільової функції $f(\bar{x})$ вихідної задачі за умови зміни i-го вільного члена b_i (обсягу i-го ресурсу) на величину Δb_i можна обчислити, скориставшись залежністю (18):

$$\Delta f(\bar{x}^*)\approx y_i^*\Delta b_i.$$

Найбільша величина оцінки відповідає найбільш "дефіцитному" ресурсу. "Недефіцитному" ресурсу, тобто такому, який за наявного оптимального плану не використовується в повному обсязі, відповідає оцінка, яка дорівнює нулю.

Компоненти оптимального розв'язку спряженої задачі називаються *оптимальними спряженими оцінками* вихідної задачі. Академік Л. В. Канторович назвав їх об'єктивно *обумовленими оцінками*. У літературі їх називають також *захованими прибутками* або *маргінальними (межовими) оцінками*[*].

Величина спряженої оцінки певного ресурсу показує, як змінилося б найбільше значення цільової функції вихідної задачі за умови приросту ресурсу на одиницю. Але необхідно наголосити, що розглядається лінійна математична модель, практична доцільність використання якої завжди обмежена певними межами зміни величин, використаних при розбудові моделі, тому спряжені оцінки дають можливість оцінити ефект впливу за порівняно невеликих приростів обсягів ресурсів. При істотно значних змінах обсягів ресурсів такі оцінки можуть не мати практичного сенсу. Спряжені оцінки визначають ефективність малих приростів обсягів ресурсів.

Приклад. Підприємство може випускати три типи продукції, використовуючи чотири види ресурсів, запаси яких складають відповідно 6; 8; 16; 18. Слід скласти план оптимального зиску від реалізації продукції за відомих витрат в грошових одиницях виміру кожного виду ресурсів на одиницю кожного типу продукції та вартості одиниці кожного типу продукції (табл. 8.2).

Таблиця 8.2.

Типи продукції	Витрати на одиницю продукції				Вартість одиниці продукції
	Види ресурсів				
1	0	1	2	1	3
2	1	1	1	2	4
3	1	0	1	1	2

Позначимо x_1, x_2, x_3 як обсяги продукції кожного типу, тобто шуканий план $\bar{x}=(x_1,\ x_2,\ x_3)$. Зиск від реалізації за певного плану випуску визначається

[*] Кузнецов А. В., Холод Н. И. Математическое программирование: Учеб. пособие. – Минск: Вышейша шк., 1984.

функцією $f = 3x_1 + 4x_2 + 2x_3$, для якої треба знайти найбільше значення.

Умови використання ресурсів визначаються системою нерівностей:

$$\begin{cases} x_2 + x_3 \le 6 \\ x_1 + x_2 \le 8 \\ 2x_1 + x_2 + x_3 \le 16 \\ x_1 + 2x_2 + x_3 \le 18 \\ x_1 \ge 0;\ x_2 \ge 0;\ x_3 \ge 0. \end{cases} \qquad (19)$$

Спряжена задача має бути сформульована так: знайти, за яких цін y_1, y_2, y_3, y_4 на ресурси виробництво продукції є недоцільним, тобто витрати в одиницях вартості на одиницю продукції будуть не меншими від зиску після її реалізації. Природно, що, закуповуючи ресурси, намагаються досягти найменших витрат, тобто необхідно знайти найменше значення функції

$$F = 6y_1 + 8y_2 + 16y_3 + 18y_4$$

за умови

$$\begin{cases} y_2 + 2y_3 + y_4 \ge 3 \\ y_1 + y_2 + y_3 + 2y_4 \ge 4 \\ y_1 + y_3 + y_4 \ge 2 \\ y_1 \ge 0;\ y_2 \ge 0;\ y_3 \ge 0;\ y_4 \ge 0. \end{cases} \qquad (20)$$

Приведемо задачу (19) до канонічної форми та знайдемо за симплексним методом її оптимальний план.

Канонічна форма задачі (19) буде записана так: знайти найбільше значення функції

$$f = 3x_1 + 4x_2 + 2x_3$$

за умови

$$\begin{cases} x_2 + x_3 + x_4 = 6 \\ x_1 + x_2 + x_5 = 8 \\ 2x_1 + x_2 + x_3 + x_6 = 16 \\ x_1 + 2x_2 + x_3 + x_7 = 18 \\ x_j \ge 0;\ (j = 1; 7). \end{cases}$$

Вилучивши проміжні розрахунки, наведемо останню симплекс-таблицю (табл. 8.3).

Таблиця 8.3.

$c_{баз}$	$x_{баз}$	0	3	4	2	0	0	0	0
		b_i^*	x_1	x_2	x_3	x_4	x_5	x_6	x_7
4	x_2	3	0	1	0	1/2	1	–1/2	0
3	x_1	5	1	0	0	–1/2	0	½	0
2	x_3	3	0	0	1	1/2	–1	½	0
0	x_7	4	0	0	0	–1	–1	0	1
f		33	0	0	0	2	2	½	0

Досягнутий план вихідної задачі в канонічній формі $\bar{x}^0 = (5; 3; 3; 0; 0; 0; 4)$ є оптимальним, оскільки відсутні від'ємні оцінки. Значення цільової функції $f(\bar{x}^0) = 33$.

Скориставшись взаємовідповідністю змінних спряжених задач (п. 8.2), маємо:

$$\begin{array}{ccccccc} x_4 & x_5 & x_6 & x_7 & x_1 & x_2 & x_3 \\ \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow \\ y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 \end{array}$$

Отже, оптимальний план спряженої задачі визначається за оцінками змінних x_4, x_5, x_6, x_7, тобто вектором $\bar{y}^0 = (3/2; 2; 1/2; 0; 0; 0; 0)$. За такого плану величина цільової функції F спряженої задачі дорівнює:

$$F(\bar{y}^0) = 6 \cdot \frac{3}{2} + 8 \cdot 2 + 16 \cdot \frac{1}{2} + 18 \cdot 0 = 33.$$

Оцінка для четвертого виду ресурсу дорівнює нулю, що свідчить про його використання не в повному обсязі. Оцінки решти ресурсів, які використовуються, додатні, що свідчить про їх використання в повному обсязі за оптимального плану виробництва.

Система обмежень вихідної задачі за оптимального плану:

$$\begin{aligned} 3 + 3 &= 6 \\ 5 + 3 &= 8 \\ 2 \cdot 5 + 3 + 3 &= 16 \\ 5 + 2 \cdot 3 + 3 &< 18. \end{aligned}$$

Аналіз обмежень свідчить, що перші три види ресурсів використовуються в повному обсязі, а четвертого виду ресурс – не повністю (після реалізації оптимального плану залишаються невикористаними чотири одиниці).

Система обмежень (20) спряженої задачі при оптимальному плані має вигляд:

$$\begin{aligned} 2+2\cdot 1/2 &= 3 \\ 3/2+2+\ 1/2 &= 4 \\ 3/2+\qquad 1/2 &= 2. \end{aligned}$$

Тобто обмеження виконуються як рівності. Проаналізуємо розв'язок сформульованої задачі дослідження доцільності організації виробництва з використанням результатів оптимального плану спряженої задачі. Збільшення ресурсу першого виду на одну одиницю дозволило б збільшити зиск після реалізації на 1,5 у. о. вартості, збільшення ресурсу другого виду на одну одиницю – на 2 у. о., а третього виду теж на одну одиницю – на 0,5 у. о. Оскільки ресурс четвертого виду не повністю використовується при оптимальному плані, то збільшення його обсягів не впливатиме на величину зиску, тому його оцінка $y_4^0 = 0$.

Оптимальний план для спряженої задачі дає можливість оцінити вплив кожного виду необхідного ресурсу на величину зиску після реалізації продукції. Найбільший вплив має той вид ресурсу, який відповідає найбільшій компоненті оптимального плану спряженої задачі. У розглянутому прикладі це є другий вид ресурсу.

Наведений приклад показує, що розв'язок спряженої задачі дає можливість ґрунтовніше висвітлити економічний сенс вихідної задачі.

Питання для самоконтролю

1. Економічний сенс задачі, спряженої до задачі планування обсягів виробництва.
2. За яких умов спряжені задачі називаються симетричними?
3. Якими правилами необхідно користуватися, формулюючи спряжену задачу симетричної форми?
4. Як пов'язані оптимальні розв'язки спряжених задач?
5. Як пов'язані змінні вихідної та спряженої задач?
6. Як побудувати оптимальний план спряженої задачі за наявності такого плану вихідної задачі?
7. Що стверджують основні теореми спряженості?

ГЛАВА 9

ОСОБЛИВІ ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ. ТРАНСПОРТНА ЗАДАЧА

9.1. Загальні зауваження

У главах 5 та 7 було описано методику постановки задачі лінійного програмування в загальній формі та основи симплексного методу – універсального методу її розв'язання. Але чимало математичних моделей важливих практичних задач, незважаючи на їх належність до класу лінійних, мають свою специфіку. Такі задачі можуть бути розв'язані загальними методами, але їх специфіка дозволяє розбудувати більш ефективні методи цілеспрямованої орієнтації.

Суттєво важливими для організації та планування господарської діяльності є так звані *задачі розподілу* – особливі задачі лінійного програмування. Такі задачі за специфікою їх математичного формулювання поділяються на два типи.

Задачі розподілу першого типу. Математична модель таких задач може бути описана так. Знайти такі *невід'ємні значення* невідомих величин x_{ij}, які мають задовольняти системі рівнянь:

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

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

забезпечуючи лінійній формі

$$f = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{2}$$

екстремальне значення (найбільше або найменше).

Коефіцієнти a_i, b_j, c_{ij} $(i=\overline{1,m};\quad j=\overline{1,n})$ – відомі величини.

До такого типу задач належать задачі планування перевезень за певних умов, задачі планування роботи кількох підприємств з метою випуску продукції певних найменувань за умов спеціалізації виробництва, задачі планування підготовки та розподілу кадрів та ін.

Задачі розподілу другого типу. Математична модель таких задач при використанні певних гіпотез може бути описана так. Знайти такі значення величин x_{ij} $(i=\overline{1,\ m};\quad j=\overline{1,\ n})$, за яких виконується система обмежень:

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

$$\sum_{i=1}^{m} b_{ij} x_{ij} \le b_j \quad (j = \overline{1, n}), \tag{3}$$

а функція

$$f = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{4}$$

досягає свого екстремального значення (найбільшого або найменшого) в залежності від практичного сенсу розв'язуваної задачі; коефіцієнти a_i, b_j, a_{ij}, b_{ij}, c_{ij} $(i = \overline{1, m};\ j = \overline{1, n})$ – відомі величини. Задачі, математичні моделі яких можна привести до вказаної, мають велике практичне значення. Назвемо деякі з таких задач: 1) оптимальне використання різних видів пального за умови різних типів агрегатів, у яких воно використовується; 2) визначення оптимальної структури використання сільськогосподарських угідь з метою одержання максимальних обсягів продукції; 3) оптимізація завантаження обладнання на виробництві з використанням різних технологій і т. ін.

Задачі розподілу, які мають математичні моделі (1) і (2), (3) і (4), можна привести до стандартних задач лінійного програмування. Введемо інше найменування невідомих. Розглянемо невідомі x_{ij} $(i = \overline{1, m};\ j = \overline{1, n})$ в такому порядку: за певного фіксованого номера індексу j, починаючи з числа 1, будемо кожного разу брати значення індексу i від 1 до m. Так упорядкованим невідомим x_{ij} дамо нові найменування з одним індексом, який має змінюватися неперервно від 1 до mn. Тобто упорядкуємо множину вихідних невідомих $\{ x_{ij} \}$ за такою схемою:

$$\begin{array}{cccccccccccc}
x_{11}, & x_{21}, & \ldots, & x_{m1}; & x_{12}, & x_{22}, & \ldots, & x_{m2}; & \ldots, & x_{1n}, & x_{2n}, & \ldots, & x_{mn} \\
z_1, & z_2, & \ldots, & z_m; & z_{m+1}, & z_{m+2}, & \ldots, & z_{m+m}; & \ldots, & z_{m(n-1)+1}, & z_{m(n-1)+2}, & \ldots, & z_{mn}.
\end{array}$$

Якщо змінити і позначення коефіцієнтів за такою ж схемою, то одержимо для названих задач розподілу задачу лінійного програмування із значною частиною нульових величин у матриці обмежень. Так, для задачі (1), (2) в кожному стовпці буде лише два відмінних від нуля елементи. Такі обставини дають можливість розбудувати для задач розподілу більш ефективні методи розв'язання, ніж симплексний метод. Розглянемо детально одну з моделей транспортної задачі та методи її розв'язання.

9.2. Транспортна задача

При плануванні значних обсягів перевезень вантажів завжди з'являється задача їх оптимальної (за певним критерієм) оцінки організації. За одних умов це означає таке планування, при якому вартість перевезення була б мінімальною, за інших – у якомога менший можливий термін, наприклад, транспортування продукції, яка може швидко псуватися, або під час збирання врожаю. Але яка б транспортна задача планування перевезення не розв'язувалася, завжди необхідно мати певні характеристики маршруту (або маршрутів, якщо їх кілька), який з'єднує відповідні пункти.

Розглянемо в загальному випадку один із можливих варіантів постановки транспортної задачі за критерієм вартості перевезення. Маємо m пунктів постачання, на кожному з яких перебуває відповідно $a_1, a_2, \ldots, a_m$ одиниць вантажу, та n пунктів призначення, на кожний з яких необхідно завезти $b_1, b_2, \ldots, b_n$ одиниць. За відомих тарифів c_{ij} (вартість перевезення одиниці вантажу) перевезення з i-го пункту постачання в j-ий пункт призначення необхідно спланувати так, щоб їх вартість була найменшою. Позначимо x_{ij} кількість вантажу, яку планується перевезти з i-го пункту постачання в j-ий пункт призначення. Умову сформульованої транспортної задачі доцільно записати так званою *розподільчою таблицею,* схему якої наведено в табл. 9.1. Коефіцієнти матриці тарифів c_{ij} позначені кружками.

Таблиця 9.1.

Пункти постачання та наявні обсяги	Пункт призначення та запити на обсяги			
	b_1	b_2	...	b_n
a_1	c_{11} x_{11}	c_{12} x_{12}	...	c_{1n} x_{1n}
a_2	c_{21} x_{21}	c_{22} x_{22}	...	c_{2n} x_{2n}
⋮	⋮	⋮	⋮⋮⋮	⋮
a_m	c_{m1} x_{m1}	c_{m2} x_{m2}	...	c_{mn} x_{mn}

Планом транспортної задачі називається матриця $\{x_{ij}\}_{m \times n}$, елементи якої визначають кількість вантажу, призначеного для перевезення з i-го пункту постачання в j-ий пункт призначення. Матриця з елементами c_{ij} $(i = \overline{1, m};\ j = \overline{1, n})$ називається *матрицею тарифів.*

Загальні затрати, обумовлені певним планом перевезення $\{x_{ij}\}$, визначаються цільовою функцією

$$f = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} . \qquad (5)$$

Змінні x_{ij} мають задовольняти умовам по обсягах наявних запасів, по запитах на обсяги вантажів та природній умові невід'ємності. Сформульовані умови аналітично слід записати так:

$$\sum_{j=1}^{n} x_{ij} \le a_i \quad (i = \overline{1, m}); \qquad (6а)$$

$$\sum_{i=1}^{m} x_{ij} = b_j \quad (j = \overline{1, n}); \qquad (6б)$$

$$x_{ij} \ge 0 \quad (i = \overline{1, m}; \quad j = \overline{1, n}). \qquad (7)$$

Задача оптимізації плану перевезення за критерієм найменшої вартості формулюється так. Знайти такий план перевезення $\{x_{ij}\}$ $(i = \overline{1, m};\ j = \overline{1, n})$, при якому цільова функція досягає найменшого значення за умови виконання обмежень (6а), (6б), (7). Задача має $m \times n$ невідомих та $m + n$ обмежень (6а), (6б).

Математична модель планування перевезення за критерієм найменшого терміну його реалізації має свої особливості. У цьому випадку необхідно мати інформацію про термін t_{ij} перевезення вантажу з i-го пункту постачання в j-ий пункт призначення. Транспортна задача оптимізації за критерієм найменшого терміну реалізації перевезення не зводиться до задачі лінійного програмування. Постановка таких задач додатково ускладнюється, якщо враховувати пропускну спроможність маршрутів транспортування, можливі обмеження по обсягах вантажів стосовно транспортних засобів і т. ін.

Далі будемо розглядати лише транспортну задачу за критерієм вартості планування перевезень.

Приклад. Припустимо, що маємо три пункти постачання з обсягами вантажів 10, 15 та 25 одиниць і чотири пункти призначення із запитами на обсяги відповідно 5, 10, 20 та 15 одиниць за відомої матриці тарифів, коефіцієнти якої наведені в табл. 9.2 й обведені кружками. Розглянемо деякі можливі плани перевезення та оцінимо вартість реалізації кожного з них.

Розглянемо план перевезень, компоненти якого записані у відповідних клітинах табл. 9.2. Згідно з таким планом, з 1-го пункту постачання вантаж в обсягах 5 і 5 одиниць необхідно відправити в 1-й та 2-й пункти призначення, з 2-го пункту постачання вантаж в обсязі 5 і 10 одиниць – в 2-й та 3-й пункти призначення, з 3-го пункту постачання – в 3-й та 4-й пункти призначення в обсягах 10 та 15 одиниць.

Таблиця 9.2.

Обсяги можливостей \ Обсяги запитів	5	10	20	15
10	5 (8)	5 (3)	(5)	(2)
15	(4)	5 (1)	10 (6)	(7)
25	(1)	(9)	10 (4)	15 (3)

Вартість перевезення при такому плані складає:

$$5 \cdot 8 + 5 \cdot 3 + 5 \cdot 1 + 10 \cdot 6 + 10 \cdot 4 + 15 \cdot 3 = 205 \text{ одиниць.}$$

Якщо розглянути план, представлений у табл. 9.3, то вартість його реалізації дорівнює:

$$10 \cdot 2 + 5 \cdot 6 + 10 \cdot 1 + 5 \cdot 3 + 15 \cdot 4 + 5 \cdot 1 = 140 \text{ одиниць.}$$

Таблиця 9.3.

Обсяги можливостей \ Обсяги запитів	5	10	20	15
10	(8)	(3)	(5)	10 (2)
15	(4)	10 (1)	5 (6)	(7)
25	5 (1)	(9)	15 (4)	5 (3)

План, представлений у табл. 9.3, має меншу вартість, ніж план табл. 9.2. Питання пошуку оптимального плану розглянемо згодом.

Тепер припустимо, що необхідно знайти такий план перевезення, при якому термін його реалізації буде найменшим. Розглядаємо цей же приклад за умови, що замість матриці тарифів маємо матрицю термінів перевезення з пунктів постачання в пункти призначення з такими ж числовими характеристиками, але в одиницях виміру часу (година, доба). За критерієм часу план перевезення, представлений у

табл. 9.3, буде реалізований за 6 умовних одиниць часу; найбільшим є термін перевезення з 2-го пункту постачання в 3-й пункт призначення.

План перевезення, представлений у табл. 9.4, буде реалізовано за 4 у. о. часу. Значення лінійної функції $\sum_{i=1}^{m}\sum_{j=1}^{n} t_{ij}x_{ij}$ при плані табл. 9.4 дорівнює 145, при плані табл. 9.3 – 140.

Таблиця 9.4.

Обсяги можливостей \ Обсяги запитів	5	10	20	15
10	(8)	(3)	(5)	(2) 10
15	(4) 5	(1) 10	(6)	(7)
25	(1)	(9)	(4) 20	(3) 5

Тобто план, кращий за критерієм вартості, може бути гіршим плану за критерієм часу.

9.3. Збалансовані та незбалансовані моделі транспортної задачі

Розрізняють два типи транспортних задач:

а) збалансовані, в яких загальний обсяг вантажу в пунктах постачання дорівнює загальному обсягу запитів пунктів призначення, тобто:

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j ; \qquad (8)$$

б) незбалансовані, в яких загальний обсяг вантажу в пунктах постачання не дорівнює загальному обсягу запитів, а може бути більшим або меншим за нього, тобто:

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j \text{ або } \sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j . \qquad (9)$$

Незбалансовану модель можна завжди привести до збалансованої. Якщо

маємо $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$, то в математичну модель за такої умови необхідно прилучити фіктивний (n+1)-й пункт призначення. До матриці умов задачі додають ще один стовпець, який відповідає фіктивному пункту призначення з обсягом запиту, який дорівнює різниці між загальним обсягом вантажу в пунктах постачання та загальним обсягом запитів:

$$b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j . \qquad (10)$$

Тарифи на перевезення до фіктивного (n+1)-го пункту призначення мають бути з нульовими значеннями. За таких умов незбалансована модель транспортної задачі перетворюється на збалансовану. Цільова функція обох задач буде одна й та ж, оскільки тарифи перевезення до фіктивного пункту призначення нульові.

За умови $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$ в математичну модель прилучають фіктивний (m+1) пункт постачання, запас вантажу для якого приймають рівним $a_{m+1} = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i$, а тарифи перевезення з фіктивного пункту мають бути нульовими. У матрицю умов задачі прилучається один рядок, а цільова функція не змінюється, оскільки фіктивні тарифи нульові, але нова задача буде збалансованою.

Для збалансованих транспортних задач умова (6а) буде представлена системою рівностей.

Ознака наявності розв'язку транспортної задачі. Теорема 1 (без доведення). Для того, щоб транспортна задача мала допустимі плани, необхідно і достатньо, щоб вона була збалансованою.

Надалі будемо розглядати лише моделі збалансованих транспортних задач за критерієм вартості перевезення. Система обмежень збалансованих задач має $m \times n$ невідомих $\{x_{ij}\}$ $(i = \overline{1, m};\ j = \overline{1, n})$ та $m + n - 1$ лінійно незалежних рівнянь, m відношень (6а), які для збалансованих задач мають бути рівняннями, та n рівнянь (6б), пов'язаних умовою (8). Оскільки збалансована транспортна задача є частковим випадком задачі лінійного програмування, то кожний її план повинен мати $m + n - 1$ базисних змінних та $mn - (m + n - 1) = (m - 1)(n - 1)$ вільних змінних.

Алгоритм розв'язання транспортної задачі (ТЗ) має три етапи:

а) побудова вихідного плану;

б) оцінка наявного плану;

в) удосконалення наявного плану, якщо він ще не оптимальний.

9.4. Побудова опорного вихідного плану

Вихідний план перевезення будемо визначати з допомогою розподільчої таблиці, побудованої за схемою, наведеною в табл. 9. 1. Якщо заплановано перевезення з пункту i до пункту j вантажу в обсязі $x_{ij} > 0$, то в клітині на перетині i-го рядка та j-го стовпця записуємо величину x_{ij}. Якщо перевезення в цьому напрямі не планується, тобто $x_{ij} = 0$, то клітину (i, j) залишаємо вільною. А оскільки система обмежень (6а), (6б) та (8) має $m + n - 1$ лінійно незалежних рівнянь, то при формуванні плану має бути заповнена $m + n - 1$ клітина.

Широко відомими є два методи побудови вихідного плану: "північно-західного кута" і "мінімального елемента" в різних варіантах. Обидва методи ґрунтуються на реалізації схеми упорядкованого почергового визначення компонент шуканого плану з урахуванням наявних обмежень по обсягах постачання та запитах у пунктах призначення. Спочатку розглянемо основи алгоритмів їх реалізації.

Припустимо, що слід визначити компоненту опорного плану, розташовану в клітині (s, p) розподільчої таблиці. Проаналізуємо всі можливі ситуації щодо можливостей s-го пункту постачання та рівня запиту в p-му пункті призначення *до формування вмісту клітини* (s, p).

1) Маємо певний запас нерозподіленого вантажу a_s^* в s-му пункті постачання та величину b_p^* невиконаного запиту в р-му пункті призначення. У клітину (s, p) записуємо найменше з a_s^* та b_p^* значення, зазначивши, що запас s-го пункту зменшився на цю величину, а обсяг вантажу в p-му пункті призначення збільшився на цю ж величину. Якщо обсяг запасів вантажу в s-му пункті постачання повністю вичерпано, то переходимо до формування клітин *наступного рядка* розподільчої таблиці, починаючи з першої. Якщо обсяг запасів s-го пункту не вичерпався, то переходимо до формування *наступної клітини* матриці в s-му рядку.

2) Маємо певний запас a_s^* нерозподіленого вантажу в s-му пункті постачання, а запит p-го пункту постачання виконано в повному обсязі. Переходимо до наступної клітини s-го рядка розподільчої матриці.

3) Запас вантажу в s-му пункті призначення вичерпано в повному обсязі, але запит p-го пункту призначення не виконано. Переходимо до розгляду p-ої клітини наступного $(s + 1)$-го рядка розподільчої матриці.

Зауваження. Якщо при реалізації алгоритму виникає ситуація, що після заповнення (s, p) клітини вичерпується запас s-го пункту постачання та повністю задовольняється запит p-го пункту призначення і це відбувається не на останньому кроці, то в кожному з таких випадків необхідно записати нуль в одну із наступних клітин по рядку або стовпцю, а саме в ту, яка має менший тариф. Якщо тарифи в обох клітинах рівні, то записати в будь-яку з них.

Ця вимога обумовлена необхідністю мати завжди опорний план ТЗ з $m + n - 1$

компонентами. Транспортна задача, яка має в опорному плані хоча б одну нульову компоненту, називається *виродженою*.

Метод "північно-західного кута". Назва методу обумовлена порядком заповнення клітин розподільчої таблиці, починаючи з лівої верхньої клітини. *З таким же успіхом можна розпочинати з правої нижньої клітини таблиці.* Реалізацію описаних раніше основних положень алгоритму розглянемо на прикладі.

Приклад пошуку вихідного плану ТЗ. Припустимо, що маємо три пункти постачання із заданими обсягами вантажу та чотири пункти призначення з відомими запитами на постачання. За відомої матриці тарифів, приведеної в розподільчій таблиці, записаній в табл. 9.5, скласти опорний план перевезення й оцінити його.

Таблиця 9.5.

a_i \ b_j	75	80	60	85	$\sum_{j=1}^{4} x_{ij}$
100	75 (6)	25 (7)	(3)	(5)	100
150	(1)	55 (2)	60 (5)	35 (6)	150
50	(3)	(10)	(20)	50 (4)	50
$\sum_{i=1}^{3} x_{ij}$	75	80	60	85	300

Формування опорного плану почнемо з клітини (1; 1). У 1-му пункті постачання маємо 100 одиниць вантажу, а запит 1-го пункту призначення дорівнює 75 одиницям, тобто цей запит можна повністю виконати. Розглядаємо клітину (1; 2). На 1-му пункті постачання залишилося 25 одиниць вантажу, заплануємо їх перевезення до 2-го пункту призначення. Оскільки запаси 1-го пункту постачання після заповнення клітини (1; 2) вичерпані повністю, розглядаємо 2-й рядок, починаючи з клітини (2; 2). У 2-му пункті постачання маємо 150 одиниць вантажу, а в 2-му пункті призначення залишився невиконаним запит в обсязі 80–25=55 одиниць. Запишемо в клітину менше з чисел 150 і 55 та перейдемо до розгляду клітини (2; 3).

У 2-му пункті постачання 150–55=95 одиниць вантажу, а обсяг запиту 3-го пункту призначення дорівнює 60. Записуємо в клітину (2; 3) найменше з чисел 95 і 60. Оскільки обсяг вантажу в 2-му пункті постачання не вичерпано, переходимо до заповнення клітини (2; 4). Тепер у 2-му пункті постачання залишилося 150–(55+60)=35 одиниць вантажу, а запит 4-го пункту призначення становить 85 одиниць. У клітину (2; 4) записуємо найменше з чисел 35 і 85 та переходимо до розгляду клітини (3; 4), оскільки запас 2-го пункту призначення

повністю вичерпано. У 3-му пункті постачання маємо 50 одиниць вантажу, а невиконаний запит 4-го пункту призначення теж дорівнює 50, тому в клітину (3; 4) записуємо 50. Наявний вантаж повністю розподілений, обсяги запитів виконані, але оптимальність такого плану не визначена. Опорний план має шість додатних компонент: $x_{11} = 75$; $x_{12} = 25$; $x_{22} = 55$; $x_{23} = 60$; $x_{24} = 35$; $x_{34} = 50$; решта компонент дорівнює нулю. Умови збалансованості виконані $\sum_{i=1}^{3} a_i = \sum_{j=1}^{4} b_j$. Вартість перевезення при побудованому плані становить:

$$f = 6 \cdot 75 + 7 \cdot 25 + 2 \cdot 55 + 5 \cdot 60 + 6 \cdot 35 + 4 \cdot 50 = 1445 \text{ у. о.}$$

Метод "мінімального елемента" ("найменшої вартості"). При реалізації цього методу перехід від однієї компоненти плану (клітини, де вона записана) до іншої виконується не лише з урахуванням наявного запасу вантажу в пункті постачання та обсягу запиту пункту призначення, але й з оцінкою тарифу. Існує кілька різновидів цього методу в залежності від того, як вибирається розташування наступної компоненти шуканого плану перевезення: клітина з найменшим тарифом по рядку, або по стовпцю, або на множині усіх незаповнених клітин розподільчої таблиці. Обсяги наявного запасу та рівень запиту враховуються за описаним раніше алгоритмом.

Обчислення компонент опорного плану (формування клітин розподільчої таблиці) починається з клітини (s, p), яка має найменший тариф; якщо таких клітин декілька, то з будь-якої з них. У клітину (s, p) з найменшим тарифом записуємо найменше з чисел a_s та b_p. У подальшому виключаємо з розгляду або стовпець, якщо запит p-го пункту призначення виконано, або рядок, якщо запаси s-го пункту постачання вичерпано. Може трапитися випадок, коли необхідно водночас вилучити відповідно і стовпець, і рядок, якщо в пункті постачання вичерпано запас і виконано запит у пункт призначення. Потім серед незаповнених клітин розподільчої таблиці знову вибирають клітину з найменшим тарифом і обчислюють відповідну компоненту плану за описаним раніше алгоритмом. Процес розподілу запасів (виконання запитів) продовжують до тих пір, поки всі запаси не будуть розподілені, а запити виконані, бо розглядаємо збалансовану ТЗ.

Приклад. Знайти опорний план методом найменшого елемента за розподільчою таблицею, представленою табл. 9. 5.

Найменший тариф відповідає клітині (2; 1), тому що $c_{21} = 1$ є найменшою величиною серед заданих тарифів. Запишемо в клітину (2; 1) число 75, тобто $x_{21} = 75$. Виключаємо з подальшого розгляду 1-й стовпець, тобто розташовані в ньому клітини, оскільки запит 1-го пункту призначення виконано в повному обсязі. Серед решти клітин найменший тариф знаходиться на позиції (2; 2), оскільки $c_{22} = 2$. У цю клітину записуємо $x_{22} = 75$. Вилучаємо 2-й рядок, оскільки запаси 2-го пункту

постачання повністю вичерпано. Тепер найменший тариф відповідає клітині (1; 3), оскільки $c_{13} = 3$. У цю клітину записуємо $x_{13} = 60$ та вилучаємо третій стовпець, оскільки запит 3-го пункту призначення виконано в повному обсязі. Наступною з найменшим тарифом буде клітина (3; 4), в яку записуємо $x_{34} = 50$, та виключаємо третій рядок із подальшого розгляду, оскільки запаси 3-го пункту постачання повністю вичерпано. Залишився неповністю розподіленим вантаж 1-го пункту постачання, залишок в якому складає 40 одиниць, а також не виконані в повному обсязі запити 2-го та 4-го пунктів призначення. У клітину (1; 4) записуємо $x_{14} = 35$, внаслідок чого 4-й стовпець виключаємо з подальшого розгляду. Решту вантажу в 5 одиниць записуємо в клітину (1; 2), тобто $x_{12} = 5$. Виконано повний розподіл наявного обсягу вантажу та запити в пунктах призначення. Пропонований план постачання записано в табл. 9.6. Вартість перевезення при даному плані дорівнює: $f = 7 \cdot 5 + 3 \cdot 60 + 5 \cdot 35 + 1 \cdot 75 + 2 \cdot 75 + 4 \cdot 50 = 815$ у. о.

Таблиця 9.6.

a_i \ b_j	75	80	60	85	$\sum_{j=1}^{4} x_{ij}$
100	(6)	5 (7)	60 (3)	35 (5)	100
150	75 (1)	75 (2)	(5)	(6)	150
50	(3)	(10)	(20)	50 (4)	50
$\sum_{i=1}^{3} x_{ij}$	75	80	60	85	300

Порівнюючи вартості планів перевезення за умов наведеної задачі, доходимо висновку, що план, одержаний методом "мінімального елемента", має істотно меншу вартість, ніж план, одержаний методом "північно-західного кута", але поки що не можемо відповісти на запитання, чи є наявний план оптимальним.

9.5. Метод потенціалів пошуку оптимального плану

Одним з ефективних методів пошуку оптимального плану транспортної задачі за критерієм вартості перевезень є метод потенціалів. Опишемо алгоритм його реалізації. *За відомого вихідного опорного плану перевезення* кожному з пунктів постачання та призначення ставиться у відповідність певна величина, яку називають потенціалом. Для зручності потенціали пунктів постачання позначимо U_i $(i = \overline{1, m})$,

пунктів призначення – V_j $(j=\overline{1,n})$. Клітини розподільчої таблиці, в яких записані величини компонент опорного плану перевезення, назвемо "завантаженими", або "базисними", навіть тоді, коли маємо вироджений опорний план, тобто план з нульовими значеннями компонент. Як наголошувалося раніше, кількість клітин опорного плану має дорівнювати $m+n-1$. Незавантажені клітини розподільчої таблиці назвемо вільними.

Величини U_i $(i=\overline{1,m})$ та V_j $(j=\overline{1,n})$ вибираємо так, щоб для кожної з базисних (завантажених) клітин виконувалась умова:

$$c_{ij}+U_i+V_j=0 \qquad (11)$$

(c_{ij} – величина тарифу відповідної базисної (завантаженої) клітини). Оскільки число потенціалів дорівнює $m+n$, а кількість завантажених компонентами опорного плану клітин становить $m+n-1$, то для обчислення величин U_i та V_j необхідно розв'язати $m+n-1$ лінійних рівнянь з $m+n$ невідомими. За такої невизначеності одній з невідомих U_i або V_j можна надати будь-яке значення, тоді решта $m+n-1$ невідомих будуть однозначно обчислені згідно з (11). Визначивши величини потенціалів, слід перевірити, чи оптимальним є наявний план перевезення.

Для кожної вільної клітини розподільчої матриці обчислюється величина так званого *псевдотарифу* c_{ij}^n за формулою:

$$c_{ij}^n=c_{ij}+U_i+V_j \qquad (12)$$

(c_{ij} – тариф перевезення одиниці вантажу з i-го пункту постачання в j-ий пункт призначення; U_i та V_j – потенціали відповідних пунктів). Якщо за певного опорного плану перевезення всі псевдотарифи є невід'ємними, то наявний план оптимальний. При наявності від'ємних псевдотарифів є можливість побудувати новий опорний план, при якому вартість перевезення буде не більшою. Для пошуку такого плану суттєвим є поняття *циклу*. Циклом у розподільчій таблиці ТЗ називається набір клітин, в яких доцільно змінювати обсяги раніше запланованих перевезень, щоб одержати план меншої (не більшої) вартості. Цикл співвідноситься з клітиною з *від'ємним псевдотарифом*. Стосовно розташування клітин такого циклу мають бути виконані певні вимоги:

а) в одному ряду (рядку або стовпці) розташовуються *дві і лише дві клітини циклу*;

б) остання (завершальна) клітина циклу знаходиться в тому ж ряду, що і перша (вихідна);

в) якщо умовно з'єднувати клітини циклу прямолінійними відрізками, то в кожній наступній клітині виконується поворот на 90°. За такого геометричного

тлумачення взаємозв'язку між клітинами циклу не важливо, через скільки завантажених або вільних клітин проходять умовні прямолінійні відрізки.

Деякі види циклів зображені на рис. 9. 1.

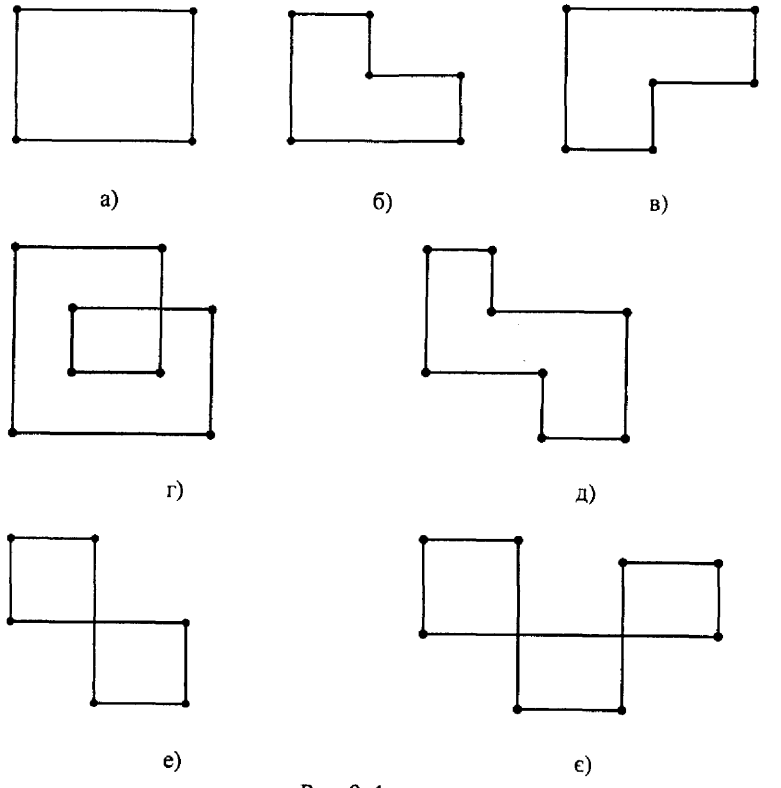


Рис. 9. 1.

З поняттям циклу пов'язані суттєві властивості опорних планів ТЗ:

1) не можна побудувати жодного циклу з вершинами лише в завантажених клітинах опорного плану;

2) у наявному опорному плані для кожної вільної клітини розподільчої таблиці можна побудувати лише один цикл, який складається з цієї клітини та завантажених клітин.

Наведемо перелік основних вимог стосовно побудови циклів.

1. У цикл можуть бути залучені *лише одна вільна* (незавантажена) клітина і певна кількість завантажених клітин.

2. Кожні дві сусідні клітини одного ряду мають об'єднуватися умовним відрізком прямої, який розташований в одному ряду.

3. Цикл як множина умовних відрізків, що з'єднують дві сусідні вершини, являє собою ламану лінію, ланцюги якої перетинаються під прямим кутом.

4. У кожному ряду (рядку або стовпці), де знаходяться клітини циклу, мають бути *лише дві вершини циклу.*

Покажемо, як, вибравши *вільну клітину з найменшим від'ємним псевдотарифом* та побудувавши цикл, однією з вершин якого є вибрана клітина, обчислити новий опорний план. Вершинам побудованого циклу, починаючи зі знаку "+", який ставиться у вільній клітині, почергово додаються знаки "+" та "–". Серед множини вантажів, розташованих у клітинах, позначених "–", знаходимо найменшу величину. Саме її додаємо до компонент опорного плану, розташованих у клітинах зі знаком "+", та віднімаємо від компонент, розташованих у клітинах зі знаком "–". Якщо, виконавши операції віднімання, маємо нульові значення більше ніж в одній клітині циклу, то необхідно залишити вільною (нічого не записувати) клітину з тарифом, найбільшим серед клітин циклу з нульовими компонентами, в решту записати нульові величини компонент опорного плану. Це так званий випадок виродженого опорного плану. За такої умови кількість клітин з компонентами опорного плану залишиться рівною $m + n - 1$, що гарантує при необхідності подальший пошук оптимального плану.

Закінчивши операцію перерахунку компонент в усіх клітинах циклу, маємо новий опорний план.

Алгоритм розв'язання транспортної задачі методом потенціалів.

1. Необхідно будь-яким методом побудувати вихідний опорний план ТЗ.
2. Заготувати розподільчу таблицю та підкреслити (або іншим чином позначити) тарифи, які відповідають розташуванню компонент наявного опорного плану.
3. Записати та розв'язати систему лінійних рівнянь (11) для обчислення величин потенціалів. Пам'ятаймо, *що одну величину з набору потенціалів можна вибрати довільно.*
4. Обчислити за формулою (12) елементи матриці псевдотарифів. *Якщо всі величини псевд тарифів є невід'ємними, то наявний план оптимальний.*
5. Якщо серед величин псевдотарифів є від'ємні, то пошуки оптимального плану продовжують. Вибирають клітину з найменшим від'ємним псевдотарифом. Якщо розташування такої клітини не визначається однозначно, то серед претендентів вибирають клітину з найменшим реальним тарифом. Вибравши клітину з від'ємним псевдотарифом, будують цикл, який їй відповідає. Використавши вершини циклу, обчислюють наступний опорний план.
6. Алгоритм продовжується, починаючи з п. 2, і завершується після обчислення оптимального опорного плану, тобто плану без від'ємних псевдотарифів.

Приклад. Знайти оптимальний план перевезення за критерієм найменшої вартості за умови, що вихідні дані представлені в табл. 9. 7.

Таблиця 9.7.

a_i \ b_j	80	120	70	30
90	5	2	3	6
60	2	3	4	5
100	6	4	2	1

Визначимо загальний обсяг наявного вантажу в пунктах постачання та обсяг запитів, щоб перевірити вимогу збалансованості транспортної задачі:

$$\sum_{i=1}^{3} a_i = 250 \text{ од.}; \quad \sum_{j=1}^{4} b_j = 300 \text{ од.}$$

Отже, обсяги вимог у пунктах призначення на 50 од. перевищують можливості пунктів постачання, тому необхідно ввести до розгляду фіктивний пункт (четвертий) постачання з обсягом запасу в 50 од. і нульовими тарифами, щоб мати збалансовану ТЗ.

Розподільчу таблицю та вихідний опорний план збалансованої ТЗ, обчислений методом мінімального елемента, представлено в табл. 9.8. Оскільки в 4-му рядку, що відповідає фіктивному пункту постачання, всі тарифи нульові, то вибрано клітину (4; 2), яка відповідає пункту запиту з найбільшим обсягом.

Згідно з побудованим планом вартість перевезення дорівнює:

$$f = 2 \cdot 70 + 3 \cdot 20 + 2 \cdot 60 + 6 \cdot 20 + 2 \cdot 50 + 1 \cdot 30 + 0 \cdot 50 = 570 \text{ у. о.}$$

Пошук оптимального плану виконуємо методом потенціалів. Для кожної завантаженої клітини згідно з (11) записуємо СЛАР для обчислення величин потенціалів:

$$2 + U_1 + V_2 = 0; \quad 3 + U_1 + V_3 = 0; \quad 2 + U_2 + V_1 = 0; \quad 6 + U_3 + V_1 = 0;$$
$$2 + U_3 + V_3 = 0; \quad 1 + U_3 + V_4 = 0; \quad 0 + U_4 + V_2 = 0.$$

Таблиця 9. 8.

a_i \ b_j	80	120	70	30	$\sum_{j=1}^{4} x_{ij}$	U_i
90	5	+ 2 70	– 3 20	6	90	$U_1 = 0$
60	2 60	3	4	5	60	$U_2 = 5$
100	– 6 20	4	2 50 +	1 30	100	$U_3 = 1$
50	+ 0	0 50 –	0	0	50	$U_4 = 2$
$\sum_{i=1}^{4} x_{ij}$	80	120	70	30	300	
V_j	$V_1 = -7$	$V_2 = -2$	$V_3 = -3$	$V_4 = -2$		

Система семи рівнянь має вісім невідомих, тому можемо значення однієї з невідомих вибрати довільно. Виберемо $U_1 = 0$, тоді решта невідомих будуть мати значення: $U_2 = 5$; $U_3 = 1$; $U_4 = 2$; $V_1 = -7$; $V_2 = -2$; $V_3 = -3$; $V_4 = -2$. Перевіримо наявний опорний план на оптимальність ($x_{12} = 70$; $x_{13} = 20$; $x_{21} = 60$; $x_{31} = 20$; $x_{33} = 50$; $x_{34} = 30$; $x_{42} = 50$), обчисливши елементи матриці псевдотарифів згідно з (12).

Для зручності обчислень випишемо матрицю тарифів і ряди значень потенціалів, підкресливши тарифи, які відповідають компонентам опорного плану (завантаженим клітинам розподільчої таблиці):

$$C = \begin{array}{c} \\ \\ \\ \\ V_j \end{array} \begin{array}{c} \begin{array}{cccc} & & & \end{array} \\ \begin{bmatrix} 5 & \underline{2} & \underline{3} & 6 \\ \underline{2} & 3 & 4 & 5 \\ \underline{6} & 4 & \underline{2} & \underline{1} \\ 0 & \underline{0} & 0 & 0 \end{bmatrix} \\ \begin{array}{cccc} -7 & -2 & -3 & -2 \end{array} \end{array} \begin{array}{c} U_i \\ 0 \\ 5 \\ 1 \\ 2 \\ \\ \end{array} \qquad C^n = \begin{bmatrix} -2 & \underline{0} & \underline{0} & 4 \\ \underline{0} & 6 & 6 & 8 \\ \underline{0} & 3 & \underline{0} & \underline{0} \\ -5 & \underline{0} & -1 & 0 \end{bmatrix}$$

Найменший елемент матриці псевдотарифів розташований на позиції (4; 1), тому для клітини (4; 1) будуємо цикл, зображений в табл. 9. 8 ламаною лінією. Починаючи з вершини у клітині (4; 1) та зі знака "+", запишемо почергово знаки "+" та "–". При цьому суттєвою є вимога неперервності перегляду вершин циклу, а не напрям, в якому починаємо їх розгляд з клітини (4; 1). Виписуємо компоненти опорного плану, розташовані у вершинах зі знаком "–" {20; 20; 50}. Найменшою

величиною серед виписаних є число 20, тому до компонент опорного плану, розташованих у клітинах зі знаком "+", додаємо величину 20, а значення компонент у клітинах зі знаком "–" зменшуємо на 20. Виконавши таку операцію, маємо одну клітину додатково (відносно наявного опорного плану) заповнену, а саме (4; 1), та дві (1; 3) і (3; 1), в яких з'являються нульові значення. Але опорний план повинен мати $m + n - 1 = 7$ компонент, тому звільняємо клітину (3; 1) як клітину з більшим, ніж у клітині (1; 3), тарифом і записуємо нуль у клітину (1; 3). Наступний опорний план, записаний в табл. 9.9, має сім компонент, але компонента $x_{13} = 0$, тобто цей план вироджений. Запишемо його компоненти: $x_{12} = 90$; $x_{13} = 0$; $x_{21} = 60$; $x_{33} = 70$; $x_{34} = 30$; $x_{41} = 20$; $x_{42} = 30$.

Таблиця 9.9.

a_i \ b_j	80	120	70	30	$\sum_{j=1}^{4} x_{ij}$	U_i
90	5	+ 2 90	– 3 0	6	90	$U_1 = 0$
60	2 60	3	4	5	60	$U_2 = 0$
100	6	4	2 70	1 30	100	$U_3 = 1$
50	0 20	0 30 –	0 +	0	50	$U_4 = 2$
$\sum_{i=1}^{4} x_{ij}$	80	120	70	30	300	
V_j	$V_1 = -2$	$V_2 = -2$	$V_3 = -3$	$V_4 = -2$		

У побудованому плані вартість перевезення складає:

$f = 2 \cdot 90 + 3 \cdot 0 + 2 \cdot 60 + 2 \cdot 70 + 1 \cdot 30 + 0 \cdot 20 + 0 \cdot 30 = 470$ у. о.,

на 100 у. о. менше, ніж у вихідному плані.

Виконаємо оцінку досягнутого опорного плану на оптимальність. Кожному ряду розподільчої табл. 9. 9 ставимо у відповідність потенціали U_i $(i = \overline{1, 4})$ та V_j $(j = \overline{1, 4})$. Скориставшись тарифами, які відповідають компонентам опорного плану, записуємо згідно з (11) СЛАР для обчислення величин потенціалів:

$2 + U_1 + V_2 = 0$; $3 + U_1 + V_3 = 0$; $2 + U_2 + V_1 = 0$; $2 + U_3 + V_3 = 0$;

$1 + U_3 + V_4 = 0$; $0 + U_4 + V_1 = 0$; $0 + U_4 + V_2 = 0$.

Вибравши $U_1 = 0$ та розв'язавши одержану СЛАР, знаходимо величини потенціалів, значення яких записані у крайньому правому та нижньому рядах

табл. 9. 9. За потенціалами рядів обчислимо елементи матриці псевдотарифів згідно з формулою (12):

$$C = \begin{array}{c} \\ \\ \\ \\ V_j \end{array} \begin{array}{c} \begin{array}{cccc} & & & \end{array} \quad U_i \\ \begin{bmatrix} 5 & \underline{2} & \underline{3} & 6 \\ \underline{2} & 3 & 4 & 5 \\ 6 & 4 & \underline{2} & \underline{1} \\ \underline{0} & \underline{0} & 0 & 0 \end{bmatrix} \begin{array}{c} 0 \\ 0 \\ 1 \\ 2 \end{array} \\ -2 \; -2 \; -3 \; -2 \end{array} \qquad C^n = \begin{bmatrix} 3 & \underline{0} & \underline{0} & 4 \\ \underline{0} & 1 & 1 & 3 \\ 5 & 3 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & -1 & 0 \end{bmatrix}$$

Матриця псевдотарифів має від'ємний елемент лише на позиції (4; 3), тому будуємо для цієї клітини цикл, зображений у табл. 9.9 ламаною лінією. На вершинах циклу, починаючи з клітини (4; 3), записуємо почергово знаки "+" та "–". Виписуємо компоненти опорного плану, розташовані у вершинах циклу, позначених "–" {30; 0}. Найменшою величиною серед виписаних є число нуль, тобто чисельно компоненти наступного опорного плану стосовно компонент плану наявного не змінюються, але змінюється розташування нульової компоненти: тепер $x_{43} = 0$, а змінна x_{13} стала вільною величиною. Опорний план має компоненти: $x_{12} = 90$; $x_{21} = 60$; $x_{33} = 70$; $x_{34} = 30$; $x_{41} = 20$; $x_{42} = 30$; $x_{43} = 0$.

Вартість перевезень не змінюється в порівнянні з її величиною в попередньому плані, тому що плани відрізняються лише компонентами з нульовими значеннями. Але величина потенціалів рядів змінюється, бо змінилося розташування нульової компоненти. Запишемо СЛАР для обчислення потенціалів рядів в одержаному плані:

$$2 + U_1 + V_2 = 0; \quad 2 + U_2 + V_1 = 0; \quad 2 + U_3 + V_3 = 0; \quad 1 + U_3 + V_4 = 0;$$
$$0 + U_4 + V_1 = 0; \quad 0 + U_4 + V_2 = 0; \quad 0 + U_4 + V_3 = 0.$$

Розв'язавши СЛАР, маємо: $U_1 = 0$; $U_2 = 0$; $U_3 = 0$; $U_4 = 2$; $V_1 = -2$; $V_2 = -2$; $V_3 = -2$; $V_4 = -1$. Обчислимо елементи матриці псевдотарифів C^n згідно з (12):

$$C = \begin{array}{c} \\ \\ \\ \\ V_j \end{array} \begin{array}{c} \begin{array}{cccc} & & & \end{array} \quad U_i \\ \begin{bmatrix} 5 & \underline{2} & 3 & 6 \\ \underline{2} & 3 & 4 & 5 \\ 6 & 4 & \underline{2} & \underline{1} \\ \underline{0} & \underline{0} & \underline{0} & 0 \end{bmatrix} \begin{array}{c} 0 \\ 0 \\ 0 \\ 2 \end{array} \\ -2 \; -2 \; -2 \; -1 \end{array} \qquad C^n = \begin{bmatrix} 3 & \underline{0} & 1 & 5 \\ \underline{0} & 1 & 2 & 4 \\ 4 & 2 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & 1 \end{bmatrix}$$

Відсутність від'ємних елементів у матриці псевдотарифів C^n свідчить, що одержаний план є оптимальним. За оптимальним планом вантаж обсягом 90 од. 1-го пункту постачання необхідно відправити в 2-й пункт призначення; вантаж обсягом 60 од. 2-го пункту постачання – в 1-й пункт призначення; вантаж обсягом 100 од. з 3-го пункту постачання в обсязі 70 од. – в 3-й пункт споживання, а решту 30 од. – в 4-й пункт, тобто оптимальний план має компоненти $x_{12} = 90$; $x_{21} = 60$; $x_{33} = 70$; $x_{34} = 30$.

Згідно з оптимальним планом, за критерієм мінімальних витрат на перевезення розв'язувана ТЗ була незбалансована, тому запити в 1-му та 2-му пунктах призначення не будуть виконані в повному обсязі. Дефіцити дорівнюють відповідно 20 од. та 30 од.

Зауваження. Оптимальний план у розв'язаній задачі визначався за критерієм найменших витрат на перевезення. За інших критеріїв оцінки плану перевезення, наприклад, виконання запитів пропорційно загальному дефіциту або з урахуванням можливої різниці міри важливості різних пунктів призначення, незбалансовану ТЗ також можна привести до збалансованої, але оптимальні плани будуть різними за різних критеріїв оптимальності та обмежень ТЗ.

Приклад для самостійного розв'язання.

Знайти оптимальний план ТЗ за критерієм найменшої вартості перевезення. Умови вказані в табл. 9.10.

Таблиця 9.10.

a_i \ b_j	60	30	40	70
80	6	5	4	2
70	2	3	5	4
50	4	1	3	6

Питання для самоконтролю

1. Сформулювати транспортну задачу за критеріями:
   а) найменшої вартості перевезення;
   б) найменшого терміну перевезення.
2. За яких умов ТЗ називається збалансованою, за яких – незбалансованою?
3. Як незбалансовану ТЗ привести до збалансованої?
4. Що таке опорний план перевезення? Методи його обчислення.
5. Скільки компонент повинно бути в опорному плані? За яких умов ТЗ називається виродженою?
6. Що таке цикл у розподільчій матриці ТЗ?
7. Метод потенціалів розв'язання транспортної задачі.
8. Матриця псевдотарифів та її використання при розв'язанні ТЗ методом потенціалів.
9. За яких умов наявний план перевезення буде оптимальним?

ГЛАВА 10

ЕЛЕМЕНТИ ТЕОРІЇ ЛІНІЙНОГО ЦІЛОЧИСЛОВОГО ТА ДИСКРЕТНОГО ПРОГРАМУВАННЯ

10.1. Особливості задачі цілочислового та дискретного програмування

У практиці розв'язання задач економіки й управління господарським процесом досить часто зустрічаються задачі, в яких частина або всі змінні мають бути лише з цілими або дискретними значеннями. Наголосимо на необхідності розрізняти два класи таких задач: а) значення змінних мають бути цілочисловими, але на множині цілих чисел змінюватися неперервно; б) значення допустимо змінювати лише дискретно на множині цілих або нецілих чисел.

Наведемо приклади таких задач. При проектуванні парку тракторів відомих типів шукані кількості мають приймати лише цілі значення, але можуть змінюватися неперервно на множині цілих чисел. Безглуздо говорити в такій задачі про $3/4$ трактора, але є сенс говорити як про кількість k, так і про $k \pm 1$. Вимоги цілочисловості до значення величин змінних будуть з'являтися завжди, коли змінні є цілими нероздільними величинами: кількість одиниць певного виду та типу техніки, обладнання, поголів'я тварин та ін.

Дещо іншою є ситуація, коли вимога цілочисловості змінних за фізичним змістом задачі має бути замінена вимогою дискретності їх можливих значень. Назвемо деякі задачі такого класу: планування перевезення вантажів в контейнерах, об'єми яких приймають дискретні значення; планування виробництва за умов, коли використовувані агрегати мають певні дискретні виробничі потужності, наприклад, потужності електромоторів; планування площ під розташування певних агрегатів з різними дискретними зонами обслуговування і т. ін.; в більш загальному випадку – планування випуску неподільних видів продукції та планування використання неподільних виробничих факторів. *Існують ґрунтовні теоретичні дослідження* стосовно дискретного програмування*.

Проблеми, обумовлені вимогами цілочисловості або дискретності стосовно множини допустимих значень змінних, з'являються не лише за умови лінійних математичних моделей, але зупинимося коротко лише на лінійних задачах.

Задача цілочислового дискретного лінійного програмування в загальному вигляді формулюється так. Знайти план $\bar{x} = (x_1, x_2, \ldots, x_n)$, за якого цільова функція

* Корбут А. А., Финкельштейн Ю. Ю. Дискретное программирование. – М.: Наука, 1969.

$$f = \sum_{j=1}^{n} c_j x_j \qquad (1)$$

досягає екстремального значення за умови виконання обмежень

$$\sum_{j=1}^{n} a_{ij} x_j \ \{\le = \ge\} \ b_i \qquad (i = \overline{1, m}), \qquad (2)$$

$$x_j \ge 0 \quad (j = \overline{1, n}); \qquad (3)$$

усі x_j або їх частина є цілими числами або можуть приймати лише дискретні значення з певної допустимої множини. (4)

Як відомо, екстремальне значення цільової функції задачі лінійного програмування досягається (за умови його наявності) у вершині опуклої множини, представленої системою обмежень (2), (3) або в точці, яка є лінійною комбінацією вершин. *Для цілочислового або дискретного лінійного програмування* точкою екстремуму може бути будь-яка точка області допустимих розв'язків (2), (3), (4). Такі обставини обумовлюють необхідність розбудови спеціальних методів розв'язання названих задач. Далі розглянемо лише задачу цілочислового лінійного програмування (ЗЦЛП).

10.2. Короткий огляд основних засад методів розв'язання задач цілочислового лінійного програмування

Порівнюючи математичну модель (1) – (4) ЗЦЛП із загальною задачею ЛП, зазначимо, що вони відрізняються умовою (4) – вимогою, щоб розв'язок був частково або повністю цілочисловим. На прикладах переконаємося, що спрощений підхід до розв'язання ЗЦЛП за схемою: розв'язати спочатку задачу ЛП, нехтуючи умовою (4) цілочисловості, а потім результати розв'язку округлити до цілих чисел, – не може гарантувати розв'язання ЗЦЛП.

Приклад. Знайти найбільше значення функції

$$f = x_1 + x_2$$

за умови

(1) $x_1 - 15x_2 \ge -45$

(2) $7x_1 - 11x_2 \le 14$

(3) $x_1 + 2x_2 \ge 1$

(4) $x_1 \ge 0;\ x_2 \ge 0$

(5) $x_1,\ x_2$ – цілі числа.

Графічний розв'язок задачі *без урахування умови цілочисловості* наведено на рис. 10. 1, де область допустимих розв'язків зображено багатокутником *ABCDE*.

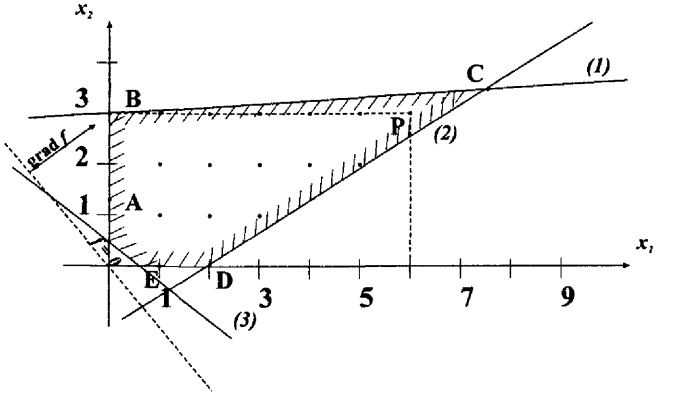


Рис. 10. 1.

Найбільше значення цільової функції досягається в точці $C(7\frac{1}{2};\ 3\frac{1}{2})$, що відповідає точці перетину ліній (1) та (2):

$$\begin{cases} x_1 - 15x_2 = -45 \\ 7x_1 - 11x_2 = 14. \end{cases}$$

За такого плану значення цільової функції $f = 11$ буде найбільшим.

При будь-якому округленні результату $\{(7; 3);\ (7; 4);\ (8; 3);\ (8; 4)\}$ не будуть виконані умови (1), (2) і (3) стосовно обмежень. За вимоги цілочисловості стосовно шуканих невідомих при наявних обмеженнях оптимальний план $x_1 = 6;\ x_2 = 3$, значення цільової функції $f = 9$. Точка P, яка відповідає оптимальному плану розв'язуваної задачі, розташована всередині допустимої області. На рис. 10. 1 зображені цілочислові "грати" області допустимих розв'язків.

Наведений приклад свідчить, що, виконавши округлення результатів розв'язку ЗЛП до цілих, не завжди можна сподіватися на розв'язок задачі цілочислового програмування. На практиці, розв'язуючи ЗЦЛП, використовують

округлення результатів розв'язку ЗЛП до цілих чисел за умови, що величини компонент оптимального плану суттєво вагомі і похибками, обумовленими округленнями, можна знехтувати. У загальному випадку для розв'язання задач дискретного програмування потрібні спеціальні методи пошуку екстремуму, які більш складні та вимагають більших обсягів розрахунків, ніж симплексний метод.

Методи цілочислового програмування можна розділити на три групи за характером їх основних положень:

а) методи відтинання, різні модифікації яких розбудовані на такій основній засаді: розв'язок ЗЛП без урахування вимог цілочисловості уточнюється за рахунок розширення кількості обмежень, обумовлених вимогами цілочисловості;

б) комбінаторні методи, розбудовані за певними правилами цілеспрямованої вибірки планів на множині скінченної кількості допустимих цілочислових розв'язків;

в) наближені методи пошуку оптимальних планів, які часто дозволяють знайти шуканий розв'язок за меншого обсягу обчислень та спрощення відповідних алгоритмів.

Ми розглядаємо цілочислове програмування як складову частину більш широкого класу задач дискретного програмування.

10.3. Метод відтинання (метод Р. Гоморі) у розв'язанні задач цілочислового програмування

Метод відтинання, який використовується для розв'язання задач як лінійного, так і опуклого цілочислового програмування, ґрунтується на засаді пошуку оптимального плану без урахування умови цілочисловості та почергового введення додаткових лінійних обмежень стосовно області допустимих розв'язків вихідної задачі. Побудовано кілька алгоритмів реалізації цього методу, основи якого були розроблені американським ученим Р. Гоморі.

Спочатку уточнимо тлумачення деяких відомих понять. Цілою частиною числа a, яку будемо позначати $[a]$, називають найбільше ціле число, яке не перевищує заданого числа a. Наприклад, якщо $a = 7.3$, то $[a] = 7$, а якщо $a = -7.3$, то $[a] = -8$, тому що *ціла частина числа має не перевищувати саме число.* Дробовою частиною числа a, яку будемо позначати $\{a\}$, називають різницею між заданим числом та його цілою частиною. Наприклад, якщо $a = -7.3$, то $\{-7.3\} = -7.3-[-7.3] = -7.3-(-8) = 0.7$. За такого визначення понять цілого числа та дробової частини *дробова частина числа не може бути від'ємною.*

Зміст різних варіантів методу відтинання Гоморі в тому, що замість розв'язання задачі (1)–(4) спочатку розв'язується задача без урахування умови (4). Якщо оптимальний план матиме необхідні цілочислові компоненти, то задача

(1)–(4) розв'язана. Інакше до обмежень розв'язуваної задачі долучають додаткове обмеження з такими властивостями:

1) має бути лінійним;

2) відтинання оптимального плану розв'язаної задачі (1) – (3) від множини допустимих планів має бути виконане так, щоб були *збережені всі допустимі плани з цілочисловими компонентами* ЗЦЛП (1) – (4).

Додаткове обмеження, яке задовольняє цим вимогам, називається *правильним відтинанням.* Долучивши до наявних обмежень (2), (3) обмеження, яке відповідає правильному відтинанню, знову розв'язують задачу, враховуючи змінену систему обмежень. Процес залучення обмежень правильного відтинання та розв'язання задачі за нових обмежень продовжують до тих пір, поки не одержать оптимальний план з необхідними цілочисловими компонентами або доведуть відсутність розв'язку.

Щодо геометричного тлумачення, залучення до наявних обмежень кожного додаткового обмеження, яке відповідає правильному відтинанню, означає, що *від області допустимих значень,* обумовленої вимогами (2) та (3), *відтинається* за допомогою гіперплощини (а у випадку двох змінних – прямої лінії) *деяка її частина,* в якій знаходиться оптимальний план *з нецілочисловими компонентами, але зберігаються всі допустимі за умов* (2), (3), (4) *плани з* цілочисловими компонентами. Різні алгоритми методу відтинання ґрунтуються *на використанні оптимальних планів відповідної ЗЛП.*

Опишемо алгоритм одного з методів розв'язання задачі цілочислового програмування. Припустимо, що для задачі ЛП (1) – (3) за використання симплексного методу одержано завершальну T-таблицю (табл. 10.1), тобто маємо оптимальний план $\overline{x^*} = (b_{10}; b_{20}; ...; b_{m0}; 0; ...; 0)$, але його компонента, наприклад, b_{k0} *не є цілим числом.* У теорії дискретного та цілочислового програмування доведено, що нерівність

$$\{b_{k,m+1}\}x_{m+1} + ... + \{b_{k,n}\}x_n \geq \{b_{k0}\}, \qquad (5)$$

сформована по k-му рядку таблиці T, має всі властивості правильного відтинання. З урахуванням такої обставини алгоритм пошуку цілочислового значення компонент оптимального плану буде таким:

1) описаним у главі 7 симплексним методом розв'язати задачу (1) – (3) тобто знехтувати вимогою (4) (вимогою цілочисловості). Якщо необхідні компоненти оптимального плану цілі числа, то знайдемо оптимальний план вихідної задачі (1) – (4). Якщо ЗЛП (1) – (3) не має розв'язку, то і вихідна задача теж не має розв'язку. Якщо серед компонент досягнутого оптимального плану є неціле числа, то необхідно виконувати п. 2;

2) серед нецілих компонент вибрати компоненту з найбільшою дробовою частиною (якщо таких компонент кілька, то одну з них) і за відповідним рядком *T*-таблиці оптимального *нецілочислового* плану записати умову відтинання (5);

3) побудовану нерівність (5) привести до рівняння шляхом введення додаткової змінної та залучити його в *T*-таблицю досягнутого нецілочислового оптимального плану;

4) продовжити розв'язання ЗЛП при додатковому обмеженні, обумовленому відтинанням.

Якщо одержаний оптимальний план буде цілочисловим, то знайдено розв'язок вихідної задачі (1) – (4), інакше необхідно повернутися до п. 2. Якщо вихідна задача (1) – (4) має розв'язок у цілих числах, то він буде одержаний за описаного алгоритму методу відтинання.

Якщо в процесі розв'язання з'явиться рядок з *нецілим* вільним членом, а решта коефіцієнтів цього рядка будуть цілими, то вихідна задача *не має розв'язку в цілих числах.* Якщо така ситуація з'явилась у практично значимій задачі, то необхідно, скориставшись оптимальним планом ЗЛП, округлити одержані результати.

Методи пошуку цілочислових розв'язків, розбудовані на засадах відтинання, дають завжди точні рішення, але не мають широкого використання через великий обсяг обчислення за наявності значної кількості змінних.

Приклад. Для закупівлі певного обладнання двох типів маємо кошти 68 тис. у. о. Для розташування обладнання можна використати площу 124 м^2. Вартість одиниці обладнання першого типу 6 тис. у. о., а площа, необхідна для розміщення та обслуговування, 10 м^2. Вартість одиниці обладнання другого типу 7 тис. у. о., а площа, необхідна для розміщення та обслуговування, 14 м^2. Продуктивність в одиницю часу обладнання першого типу 4 у. о., а другого – 5 у. о. За названих умов необхідно визначити, в яких кількостях треба закупити обладнання кожного типу, щоб мати найбільшу загальну продуктивність.

Позначимо x_1 кількість обладнання першого типу, x_2 – другого типу; загальна продуктивність f буде визначена залежністю $f = 4x_1 + 5x_2$. За названих кількостей обладнання його вартість буде рівна $6x_1 + 7x_2$, а площа, необхідна для його використання, $10x_1 + 14x_2$. Задача пошуку оптимального плану закупівлі обладнання буде сформульована так: знайти найбільше значення

$$f = 4x_1 + 5x_2$$

за умови, що величини x_1 та x_2 мають задовольняти вимогам:

$$6x_1 + 7x_2 \le 68 \qquad \text{(а)}$$
$$10x_1 + 14x_2 \le 124 \qquad \text{(б)}$$
$$x_1 \ge 0;\ x_2 \ge 0;\ x_1 \text{ та } x_2 \text{ – цілі числа.}$$

Запишемо канонічну форму ЗЛП:

$$\max\ f = 4x_1 + 5x_2$$

за умови

$$6x_1 + 7x_2 + x_3 = 68$$
$$10x_1 + 14x_2 + x_4 = 124$$
$$x_1 \ge 0;\ x_2 \ge 0;\ x_1 \text{ та } x_2 \text{ – цілі числа.}$$

Вибравши x_3, x_4 за базисні змінні, запишемо вихідну симплекс-таблицю *T* (частина *A* табл. 10.1). Вибравши найменше від'ємне число в f-рядку та оцінивши симплексні відношення b_{i0}/b_{is}, визначаємо провідний елемент 14, помічений колом.

Симплекс-таблиці *T* подальших перетворень записані в частинах *B* і *C* табл. 10.1.

Таблиця 10.1.

A				
БЗ	b_{i0}	$-x_1$	$-x_2$	b_{i0}/b_{is}
x_3	68	6	7	68/7
x_4	124	10	14	124/14
f		–4	–5	
B				
БЗ	b_{i0}	$-x_1$	$-x_4$	b_{i0}/b_{is}
x_3	6	1	–1/2	6
x_2	124/14	10/14	1/14	124/10
f	620/14	–6/14	5/14	–
C				
БЗ	b_{i0}	$-x_3$	$-x_4$	b_{i0}/b_{is}
x_1	6	1	–1/2	
x_2	64/14	–10/14	6/14	
f	656/14	6/14	2/14	–

Аналіз частини *C* табл. 10.1 свідчить, що досягнуто оптимального, *але не цілочислового* плану $x_1 = 6,\ x_2 = 4\frac{4}{7}$, за якого цільова функція $f = 46\frac{6}{7}$. Використовуючи рядок "x_2" нецілочислової компоненти плану, формуємо додаткове обмеження, яке відтинає від області допустимих розв'язків її частину *без цілочислових допустимих компонент:*

$$\frac{2}{7}x_3 + \frac{3}{7}x_4 \geq \frac{4}{7} \Rightarrow -\frac{2}{7}x_3 - \frac{3}{7}x_4 + x_5 = -\frac{4}{7}.$$

Згідно з визначенням дробова частина від'ємного числа $-\frac{5}{7}$ дорівнює $\frac{2}{7}(-\frac{5}{7} - (-1))$. Використавши додаткову змінну x_5, записуємо нову вихідну *T*-таблицю, як це показано в частині *A* табл. 10.2.

Таблиця 10.2.

A				
БЗ	b_{i0}	$-x_3$	$-x_4$	b_{i0} / b_{is}
x_1	6	1	–1/2	–
x_2	64/14	–10/14	6/14	64/6
x_5	-8/14	–4/14	–6/14	8/6
f	656/14	6/14	2/14	–
B				
БЗ	b_{i0}	$-x_3$	$-x_5$	b_{i0} / b_{is}
x_1	40/6	8/6	–7/6	–
x_2	4	–1	1	–
x_4	8/6	4/6	–14/6	–
f	140/3	1/3	1/3	–
C				
БЗ	b_{i0}	$-x_3$	$-x_5$	b_{i0} / b_{is}
x_1	40/6	8/6	–7/6	5
x_2	4	–1	1	–
x_4	8/6	4/6	–14/6	2
x_6	–4/6	–2/6	–5/6	2
f	140/3	1/3	1/3	

Рядок "x_5" частини *A* табл. 10. 2 має від'ємний вільний член –8/14, тому вибір провідних стовпця та елемента виконаємо за правилами, які відповідають наявності від'ємного вільного члена. У нашому випадку за провідний елемент виберемо –6/14, відмічений колом (можна взяти також і –4/14). Результати симплексних перетворень наведені в частині *B* табл. 10.2. Аналіз свідчить, що *T*-таблиця, записана в частині *B*, відповідає оптимальному, але нецілочисловому плану $x_1 = 6\frac{2}{3},\ x_2 = 4$, за якого $f = 280/6 = 46\frac{2}{3}$. Використовуючи рядок "$x_1$" нецілочислової компоненти досягнутого оптимального плану, формуємо додаткове обмеження (відтинання):

$$\frac{2}{6}x_3 + \frac{5}{6}x_5 \geq \frac{4}{6} \Rightarrow -\frac{2}{6}x_3 - \frac{5}{6}x_5 + x_6 = -\frac{4}{6}.$$

Арифметичні спрощення не виконано, щоб зберегти співвідношення між одиницями виміру вихідних величин.

Вихідна *T*-таблиця за умови врахування наступного додаткового відтинання наведена в частині *C* табл. 10.2. Виконавши наступне симплексне перетворення, отримаємо *T*-таблицю, записану в табл. 10.3, яка відповідає цілочисловому плану $x_1 = 4,\ x_2 = 6$, за якого значення $f = 46$, але f-рядок має від'ємний елемент. Розв'язувана задача на множині цілочислових розв'язків має два альтернативних оптимальних плани: $x_1 = 4,\ x_2 = 6$ (досягнутий) та $x_1 = 9,\ x_2 = 2$. Тому, якщо продовжити пошук, то одержимо зациклений процес, при реалізації якого відбувається перехід від плану (4; 6) до плану (9; 2), а потім навпаки.

Таблиця 10.3.

БЗ	b_{i0}	$-x_6$	$-x_5$	b_{i0} / b_{is}
x_1	4	4	–9/2	
x_2	6	–3	7/2	
x_4	0	2	–4	
x_3	2	–3	5/2	
f	46	1	–1/2	

Наявна ситуація віддзеркалена на рис. 10. 2, де зображено цілочислові "грати" області допустимих планів і компоненти альтернативних оптимальних планів; лінія рівня $f = 46$ зображена пунктиром.

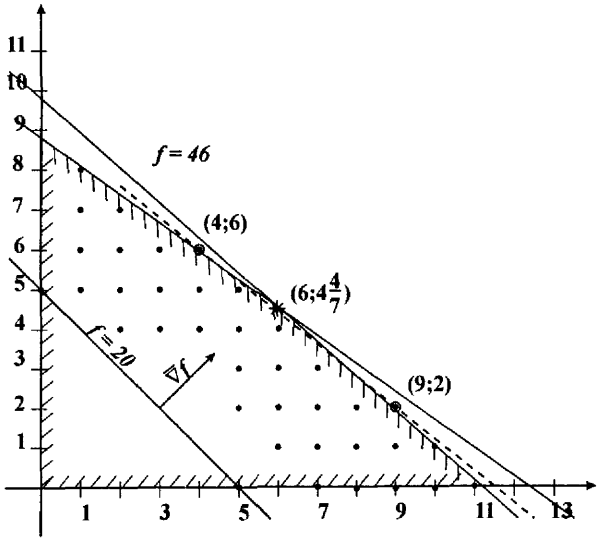


Рис. 10.2.

Практичне завершення розв'язання задачі може бути таким. Одержавши альтернативні оптимальні плани, оцінимо виконання необхідних обмежень.

При плані ($x_1 = 4$, $x_2 = 6$) маємо:

$$(1)\quad 6 \cdot 4 + 7 \cdot 6 = 66 < 68;$$
$$(2)\quad 10 \cdot 4 + 14 \cdot 6 = 124,$$

а при плані ($x_1 = 9$, $x_2 = 2$):

$$(1)\quad 6 \cdot 9 + 7 \cdot 2 = 68;$$
$$(2)\quad 10 \cdot 9 + 14 \cdot 2 = 118 < 124.$$

За обох планів загальна продуктивність обладнання становить 46. Реалізувавши перший план, одержимо економію коштів 2 тис. у. о., другий – залишок 6 м^2 від наданої площі. Маючи такі результати аналізу для завершальної вибірки плану, можна використати інші критерії: надійність обладнання, забезпеченість ремонтними деталями, витрати на експлуатацію, екологічні вимоги (шум, вібрація) і т. ін. Досягнуті цілочислові плани не можна одержати шляхом будь-якого округлення оптимального плану ($x_1 = 6$, $x_2 = 4\frac{4}{7}$) задачі без вимоги

цілочисловості. Наведений приклад задачі лише з двома змінними переконливо свідчить про складність та особливості задач цілочислового програмування.

10.4. Метод розгалужень і меж

У розв'язанні задач цілочислового та дискретного програмування широке застосування мають комбінаторні методи, розбудовані на засадах цілеспрямованого перебору множин допустимих планів. Ряд комбінаторних методів ґрунтується на методі розгалужень та меж. Розглянемо схему розбудови одного з можливих варіантів названого методу.

Необхідно знайти найбільше значення функції $f(\bar{x}) = f(x_1, x_2, ..., x_n)$ за умови, що допустимі плани $\bar{x} = (x_1, x_2, ..., x_n)$ утворюють деяку скінченну множину G, а їх компоненти мають бути повністю або частково цілочисловими або дискретними.

Пошук оптимального плану починається з розбудови певного правила обчислення величини так званої верхньої межі $\alpha(\bar{x})$ можливих значень $f(\bar{x})$ на кожній множині допустимих планів. Якщо, наприклад, розглядаємо задачу лінійного цілочислового програмування, то величина $\alpha(G)$ може бути обчислена як величина цільової функції на множині G *без урахування вимог цілочисловості.* Потім множина G певним чином поділяється на дві (або більше) підмножини G_1 та G_2, *які не перетинаються,* тобто не мають спільних планів (точок). На кожній з підмножині G_1 та G_2 є верхні межі $\alpha(G_1)$ та $\alpha(G_2)$ функції $f(\bar{x})$. Для подальших розрахунків вибирають підмножину $G_в$ з більшою верхньою межею. За умови, що обидві верхні межі тотожні, для подальших розрахунків залишають обидві підмножини G_1 та G_2. Якщо на вибраній підмножині буде знайдено такий план $\bar{x}^*$, що $f(\bar{x}^*) = \alpha G_в$, то це оптимальний план. Якщо такого плану не знайдемо, то вибрана підмножина $G_в$ знову поділяється на дві або більше підмножин, і процес пошуку оптимального плану продовжується. Отже, при кожному розподілі досягнутої підмножині розв'язання задачі оптимізації розгалужується на частини. Оскільки число допустимих планів скінченне, то оптимальний план буде знайдено за скінченної кількості розтинання вихідної множини G на частини.

Доцільно наголосити, що до методу розгалужень і меж у загальній задачі математичного програмування відносять значну кількість алгоритмів, розбудованих стосовно окремих класів задач дискретного програмування з використанням різних правил розгалуження та пошуку меж. Для задач лінійного цілочислового програмування вимоги розгалуження та пошуку верхніх меж не викликають особливих труднощів. Як наголошувалося раніше, для обчислення верхньої межі

використовують задану цільову функцію і розрахунки виконують без урахування вимоги цілочисловості.

Розглянемо реалізацію основних засад методу розгалужень і меж стосовно задачі лінійного цілочислового програмування.

Знайти найбільше значення

$$f(\bar{x}) = \sum_{j=1}^{n} c_j x_j \tag{6}$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}) \tag{7}$$

$$x_j \ge 0 \quad (j = \overline{1, n}) \tag{8}$$

$$x_j - \text{цілі числа} \quad (j = \overline{1, n}). \tag{9}$$

За критерій визначення верхньої межі планів на множині розв'язків виберемо цільову функції $f(\bar{x})$ ЗЛП (6) – (8), тобто нехтуючи умовою цілочисловості. Припустимо, що оптимальний план $\bar{x}^*$ ЗЛП (1) – (3) *знайдено*, але компонента x_p^* у цьому плані дробова, а не цілочислова, як це необхідно в задачі (6) – (9). Інтервал $[x_p^*] < x_p < [x_p^*]+1$ ($[x_p^*]$– ціла частина величини x_p^*) *не містить* у собі допустимих цілочислових компонент шуканого розв'язку задачі (6) – (9). Тому шукана величина цілочислового значення x_p має задовольняти одній із нерівностей: $x_p \le [x_p^*]$ або $x_p \ge [x_p^*]+1$ Для пошуку цілочислового значення компоненти x_p тепер необхідно розв'язувати *дві задачі лінійного нецілочислового програмування*, пов'язані з вихідною задачею (6) – (9), *з вилученою умовою цілочисловості* для компоненти x_p, але з додатковими обмеженнями:

$$\begin{array}{ll}
\max f(\bar{x}) = \sum_{j=1}^{n} c_j x_j & \max f(\bar{x}) = \sum_{j=1}^{n} c_j x_j \\
\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}) & \sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}) \\
x_j \ge 0 \quad (j = 1, n) & x_j \ge 0 \quad (j = 1, n) \\
x_p \le [x_p^*] & x_p \ge [x_p^*] + 1.
\end{array} \tag{10}$$

Остання вимога в (10) стосовно компоненти x_p обумовлює розгалуження вихідної задачі на дві інші задачі з тотожною цільовою функцією, але зміненими обмеженнями. Результати розв'язку задач (10) можуть бути такі:

а) оптимальний план однієї із задач, в порівнянні з іншою, забезпечує більше значення цільової функції; якщо такий оптимальний план має цілочислову компоненту x_p, то переходять до *пошуку цілочислового значення іншої компоненти*, інакше ту задачу з двох (10), для якої досягнуто найбільше значення цільової функції, знову розгалужують на дві задачі і т. д.;

б) якщо оптимальні плани кожної із задач дають тотожні значення цільової функції, необхідно продовжувати дослідження кожної з задач (10).

Процес розгалуження продовжують до тих пір, поки кожна задача (10) не завершиться цілочисловим розв'язком або не буде встановлено неможливість подальшого поліпшення розв'язку, досягнутого на певному кроці.

Приклад. Знайти найбільше значення функції

$$f = 5x_1 + 3x_2$$

за умови

$$\begin{aligned}
-x_1 + x_2 &\le 1 \\
x_1 - x_2 &\le 3 \\
x_1 + x_2 &\le 4 \\
x_1 \ge 0;\quad x_2 &\ge 0;\ x_1, x_2 - \text{цілі числа.}
\end{aligned}$$

Якщо вилучити умову цілочисловості, то оптимальний план $\bar{x} = (7/2;\ 1/2)$ забезпечить найбільше значення $f = 19$. Обидві компоненти дробові, будь-яку можна обрати для розгалуження.

Вибираємо x_1, тоді згідно з теорією необхідно розв'язувати дві задачі – 1. 1 та 1. 2.

Знайти найбільше значення $f = 5x_1 + 3x_2$ за умов:

задача 1. 1.	*задача 1. 2.*
$-x_1 + x_2 \le 1$	$-x_1 + x_2 \le 1$
$x_1 - x_2 \le 3$	$x_1 - x_2 \le 3$
$x_1 + x_2 \le 4$	$x_1 + x_2 \le 4$
$x_1 \le 3$	$x_1 \ge 4$
$x_1 \ge 0;\quad x_2 \ge 0$	$x_1 \ge 0;\quad x_2 \ge 0$

Розв'яжемо задачу 1.1, привівши її до канонічної форми:

$$\max \quad f = 5x_1 + 3x_2$$

$$\begin{aligned} -x_1 + x_2 + x_3 \qquad\qquad &= 1 \\ x_1 - x_2 \qquad + x_4 \qquad &= 3 \\ x_1 + x_2 \qquad\qquad + x_5 \quad &= 4 \\ x_1 \qquad\qquad\qquad + x_6 &= 3. \end{aligned}$$

Обравши x_3, x_4, x_5, x_6 за базисні змінні, записуємо їх у симплекс-таблицю T (табл. 10.4).

Таблиця 10.4.

A				
БЗ	b_{i0}	$-x_1$	$-x_2$	b_{i0}/b_{is}
x_3	1	–1	1	–
x_4	3	1	–1	3
x_5	4	1	1	4
x_6	3	1	0	3
f	0	–5	–3	
B				
БЗ	b_{i0}	$-x_6$	$-x_2$	b_{i0}/b_{is}
x_3	4	1	1	4
x_4	0	–1	–1	
x_5	1	–1	1	1
x_1	3	1	0	
f	15	5	–3	
C				
БЗ	b_{i0}	$-x_6$	$-x_5$	b_{i0}/b_{is}
x_3	3	2	–1	
x_4	1	–2	1	
x_2	1	–1	1	
x_1	3	1	0	
f	18	2	3	

Наявний план $\bar{x}$ = (0; 0; 1; 3; 4; 3) має бути поліпшений, бо f-рядок містить від'ємні величини. Оцінивши відношення b_{i0}/b_{i1}, за провідний елемент виберемо

1, що дає підстави прилучити до базису x_1 та вилучити x_6. Таблиця T, що відповідає зміненому базису, наведена в табл. 10. 4. Маємо від'ємний елемент у f-рядку та два додатних елементи в стовпці $-x_2$.

Аналіз таблиці T, наведеної в табл. 10. 4, показує, що в задачі 1.1 досягнуто оптимальний цілочисловий план $\bar{x}$ = (3; 1; 3; 1; 0; 0), за якого f =18, але *він не обов'язково буде оптимальним для вихідної задачі.* Розглянемо тепер задачу 1.2. Система її обмежень не сумісна тому, що $x_1 + x_2 \le 4$; $x_1 \ge 4$; $x_2 \ge 0 \Rightarrow x_1 = 4$; $x_2 = 0$, але за таких значень змінних x_1 та x_2 не виконується решта обмежень умови задачі 1.2.

Отже, розв'язок вихідної задачі тотожний розв'язку задачі 1.1.

Розглянемо розв'язання вихідної задачі за умови, що її розгалуження виконано з використанням компоненти $x_2^* = \frac{1}{2}$ оптимального плану без урахування умови цілочисловості. За названої умови на першому кроці розгалуження необхідно розв'язати такі дві задачі.

Знайти найбільше значення $f = 5x_1 + 3x_2$ за умов:

задача 1. 1.	*задача 1. 2.*
$-x_1 + x_2 \le 1$	$-x_1 + x_2 \le 1$
$x_1 - x_2 \le 3$	$x_1 - x_2 \le 3$
$x_1 + x_2 \le 4$	$x_1 + x_2 \le 4$
$0 \le x_2 \le 0$	$x_2 \ge 1$
$x_1 \ge 0$	$x_1 \ge 0$

За умовами задачі 1.1 визначаємо, що наявна система обмежень має єдиний розв'язок – допустимий план $x_1 = 1$; $x_2 = 0$, за якого $f = 5$.

Розв'яжемо задачу 1.2, привівши її до канонічної форми:

$$\max \quad f = 5x_1 + 3x_2$$

$$\begin{aligned} -x_1 + x_2 + x_3 \qquad\qquad &= 1 \\ x_1 - x_2 \qquad + x_4 \qquad &= 3 \\ x_1 + x_2 \qquad\qquad + x_5 \quad &= 4 \\ x_2 \qquad\qquad - x_6 &= 1 \Rightarrow -x_2 + x_6 = -1. \end{aligned}$$

Обравши x_3, x_4, x_5, x_6 за базисні змінні, вихідну симплекс-таблицю T запишемо так, як наведено в табл. 10.5, але враховуючи, що маємо від'ємний вільний член у стовпці b_{i0}, провідні стовпець і рядок обираються за іншими

правилами, ніж за умови лише додатних вільних членів (глава 7). У рядку з від'ємним вільним членом знаходимо від'ємний коефіцієнт (за умови сумісності обмежень такий коефіцієнт обов'язково буде), відповідний йому стовпець обираємо за провідний. Для елементів провідного рядка обчислюємо симплексні відношення b_{i0} / b_{is}, але використовуємо лише ті елементи, які мають однакові знаки. Провідний рядок, а потім і провідний елемент визначаємо за розташуванням найменшого симплексного відношення.

Таблиця 10.5.

A				
БЗ	b_{i0}	$-x_1$	$-x_2$	b_{i0} / b_{is}
x_3	1	–1	1	1
x_4	3	1	–1	–
x_5	4	1	1	4
x_6	-1	0	–1	1
f	0	–5	–3	
B				
БЗ	b_{i0}	$-x_1$	$-x_6$	b_{i0} / b_{is}
x_3	4	1	–1	4
x_4	3	1	1	3
x_5	1	–1	–1	
x_2	1	0	1	
f	3	–5	3	
C				
БЗ	b_{i0}	$-x_4$	$-x_6$	b_{i0} / b_{is}
x_3	1	–1	–2	
x_1	3	1	1	
x_5	4	1	1	
x_2	1	0	1	
f	18	5	8	

Для вихідної симплекс-таблиці T провідний елемент –1 розташований на перетині стовпця $-x_2$ та рядка x_6 і позначений кружком. Результати симплексних перетворень за обраного провідного елемента –1 наведені в частині B табл. 10.5. Наявність від'ємного елемента –5 в f-рядку за умови, що всі вільні члени b_{i0} невід'ємні, свідчить, що досягнутий план $\bar{x}$ = (0; 1; 4; 3; 1; 0) не оптимальний і може

бути поліпшений, якщо, скориставшися розташованим на перетині рядка x_4 та стовпця $-x_1$ провідним елементом 1, залучимо до базису x_1, вилучивши x_4. Результати відповідних обчислень наведені в табл. 10. 5. У f-рядку немає від'ємних елементів, отже, досягнутий план $\bar{x}$ = (3; 1; 1; 0; 4; 0) оптимальний і цілочисловий. Розв'язок вихідної задачі: x_1 = 3, x_2 = 1 – знайдено, і він тотожний раніше обчисленому.

Питання для самоконтролю

1. Різниця між цілочисловим та дискретним програмуванням. Навести приклади.
2. Який план називається допустимим?
3. Що таке базисний розв'язок? Що таке базисні та вільні змінні?
4. У якій формі має бути записана задача лінійного програмування, щоб скористатися симплексним методом, описаним у главі 7.
5. Схема запису вихідної таблиці симплексних перетворень.
6. Як вибирають провідні рядок, стовпець, елемент, якщо за одиничного базису всі вільні члени додатні?
7. Як вибирають провідні рядок, стовпець, елемент, якщо за одиничного базису серед вільних членів є від'ємні величини?
8. Як пересвідчитися, чи досягнутий план можна поліпшити, чи він є оптимальним?
9. Як визначаються поняття цілої та дробової частин нецілого числа? Навести приклади.
10. Що таке правильне відтинання в задачі цілочислового програмування?
11. Як записати умову правильного відтинання в задачі цілочислового програмування?
12. Як зміниться симплексна таблиця за умови правильного відтинання?
13. Основні засади методу Гоморі розв'язання задач цілочислового програмування.
14. Основні засади методу розгалужень і меж.
15. Використовуючи основну ідею методу розгалуження та меж про оцінку верхньої межі значень функції на певній множині, побудувати з допомогою терезів алгоритм пошуку фальшивої монети серед 100 монет, якщо відомо, що вона має більшу вагу.

ГЛАВА 11

НЕЛІНІЙНЕ ПРОГРАМУВАННЯ

11.1. Задача нелінійного програмування

Математичні моделі лінійного програмування ґрунтуються на гіпотезі лінійної залежності між числовими характеристиками досліджуваного об'єкта або системи та параметрами, обраними для моделювання певних властивостей системи. Лінійні моделі є ефективним засобом дослідження широкого кола задач управління та економіки за умови короткотермінових проміжків або стабільного стану системи. Але практичний досвід показує, що гіпотези про лінійну залежність можуть лише наближено характеризувати об'єкт дослідження. У багатьох випадках залежності між основними параметрами системи мають суттєво нелінійний характер. Наприклад, зиск від реалізації продукції виявляється нелінійною функцією ціни: чим нижче ціна, тим більше попит і, як наслідок, більше обсяг реалізації. Перевірений широкою практикою закон спадаючої ефективності виробництва свідчить про нелінійну залежність обсягів виробництва від собівартості продукції. Різного виду затрати праці та матеріальних ресурсів, як правило, є нелінійними функціями точності, надійності та інших характеристик якості продукції виробництва. Ретельний облік норм споживання певних ресурсів свідчить про необхідність використання моделей нелінійних обмежень як більш узгоджених з практикою.

Якщо дослідження системи або об'єкта може бути виконано з використанням n параметрів, то математична модель задачі нелінійного програмування в загальному випадку формулюється так. У n-мірному просторі змінних необхідно знайти план (вектор) $\bar{x} = (x_1, x_2, ..., x_n)$, за якого цільова функція досягає екстремуму (найбільшого або найменшого значення в залежності від практичного сенсу задачі):

$$\text{extr}\, f(\bar{x}) = f(x_1, x_2, ..., x_n) \qquad (1)$$

за умови:

$$f_i(\bar{x}) = f_i(x_1, x_2, ..., x_n)\ \{\le = \ge\} b_i \quad (i = \overline{1, m}) \qquad (2)$$

$$x_j \ge 0 \quad (j = \overline{1, n}). \qquad (3)$$

Наголошуємо, що термін "вектор" слід сприймати як упорядкований набір змінних $x_1, x_2, ..., x_n$.

У залежностях (1), (2) або цільова функція, або деякі з обмежень є

нелінійними функціями змінних. Вектор $\bar{x} = (x_1, x_2, ..., x_n)$ у n-мірному просторі називається *допустимим планом*, якщо його компоненти задовольняють умовам (2) та (3). Множина допустимих векторів $\bar{x}$ утворює *область допустимих планів D*. *Оптимальним розв'язком задачі* (1) – (3) називається план $\bar{x}$, за якого цільова функція досягає шуканого екстремуму.

До нелінійного програмування належить широке коло складних різнопланових задач управління та економіки. Розбудова такого ефективного й універсального методу, як симплексний в задачах лінійного програмування, тут поки що не мала успіху. Складність розв'язання нелінійних задач обумовлена як структурою області допустимих розв'язків, яка може мати неопуклі межі та нескінченну множину кутових точок, так і характером зміни цільової функції в області допустимих планів (розв'язків). За названих причин цільова функція може досягати екстремуму як на межі, так і в середині області допустимих розв'язків; до того ж екстремальні значення можуть бути досягнуті не в єдиній точці.

Доречно нагадати певні поняття з курсу вищої математики. Функція $f(\bar{x})$ досягає *локального максимуму (мінімуму) в точці* $\bar{x}^*$, якщо $f(\bar{x}^*) \ge f(\bar{x})$ $(f(\bar{x}^*) \le f(\bar{x}))$ *для будь-якого* $\bar{x}$ *з околу точки* $\bar{x}^*$. Якщо $f(\bar{x}^*) \ge f(\bar{x})$ $(f(\bar{x}^*) \le f(\bar{x}))$ *для будь-якого* $\bar{x}$ *з області дослідження функції*, то функція досягає в точці $\bar{x}^*$ *глобального максимуму (мінімуму)* в досліджуваній області.

Локальний мінімум може бути істотно більшим за локальний максимум. Для графіка функції рис. 11.1 локальні максимуми досягаються в точках з абсцисами, які позначені літерами A, C, E, локальні мінімуми – в точках з абсцисами B, D.

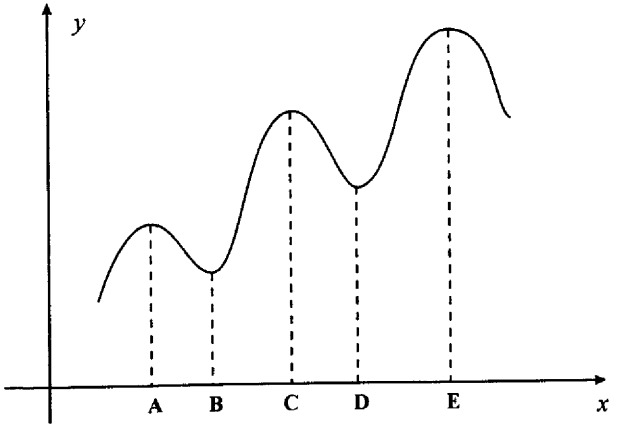


Рис. 11.1.

Локальний максимум, який відповідає точці A, менше локального мінімуму, який відповідає точці D. Глобальний максимум в області дослідження функції є одним із локальних. Твердження є вірним і стосовно локального мінімуму. Глобальні екстремуми (мінімум, максимум) можуть бути досягнуті не в єдиній точці області дослідження.

Розглянемо інші види екстремумів на прикладі рис. 11.2.

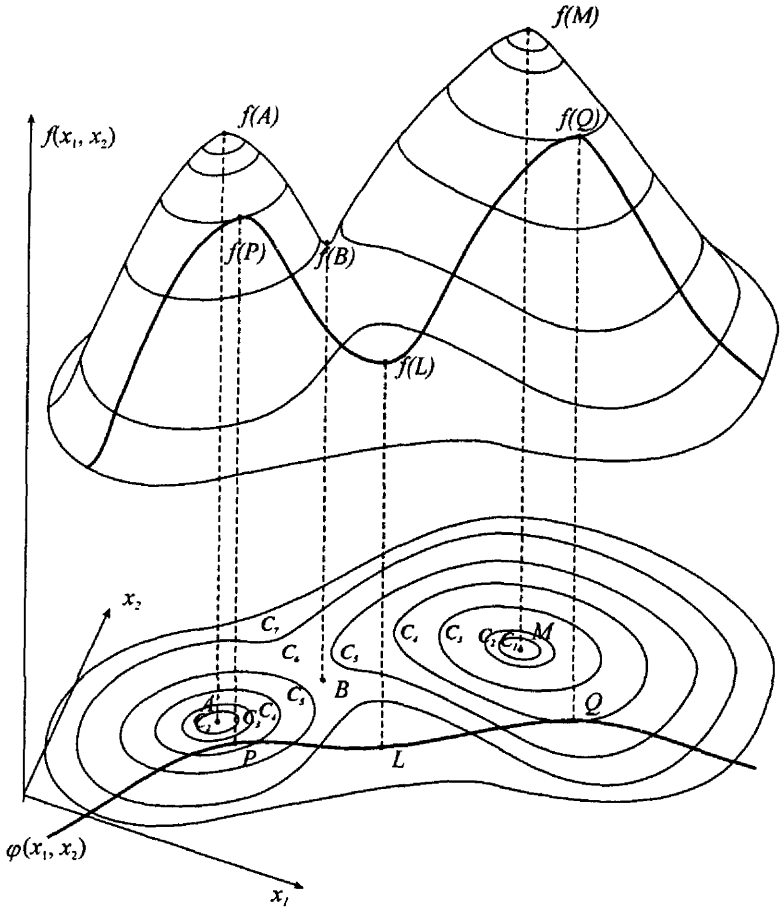


Рис. 11.2.

Якщо виконується дослідження функції кількох змінних за умови, що їх значення пов'язані між собою *певним рівнянням або системою рівнянь, то екстремальні значення* функції, досягнуті за таких обставин, називаються *умовними*, тобто екстремум $f(\bar{x})$, знайдений за умови, що $\bar{x} = (x_1, x_2, ..., x_n)$, задовольняє системі рівнянь $\varphi_i(\bar{x}) = \varphi_i(x_1, x_2, ..., x_n) = 0$ $(i = \overline{1, m})$. Локальні, глобальні та умовні екстремуми для функції двох змінних схематично зображені на рис. 11. 2. Локальні максимуми досягаються в точках A, M, умовні екстремуми – в точках P, Q (максимуми) та L (мінімум).

У задачах нелінійного програмування за наявними властивостями цільової функції та області допустимих обмежень D нами розглядаються такі класи:

- **опуклого програмування** – якщо цільова функція та обмеження опуклі;
- **квадратичного програмування** – якщо цільова функція квадратична, а обмеження представлені системою лінійних рівнянь і нерівностей;
- **багатоекстремальні задачі**, серед яких виділяють спеціальні класи, які зустрічаються при моделюванні практичних задач, наприклад, задачі пошуку екстремуму угнутих функцій на опуклих множинах.

Найбільш широке практичне використання мають математичні моделі нелінійного програмування за умови, що цільова функція опукла або угнута та множина допустимих планів опукла. Зміст названих термінів буде визначений далі.

11.2. Опуклі та угнуті функції, їх властивості

Нагадаємо, що в теорії лінійного програмування було введено поняття опуклої множини векторів (точок) у n-мірному просторі. Множина векторів (точок) у n-мірному просторі називається *опуклою*, якщо разом з будь-якими двома векторами $\bar{x}$ та $\bar{y}$ цієї множини їй належать також всі вектори $\bar{r} = \lambda\bar{x} + (1-\lambda)\bar{y}$ за умови $0 \le \lambda \le 1$.

При моделюванні широкого кола планово-виробничих та економічних задач відповідні цільові функції виявляють певні властивості, які дозволяють віднести їх до класів так званих опуклих та угнутих функцій.

Функція $f(\bar{x})$, де $\bar{x} = (x_1, x_2, ..., x_n)$, визначена на опуклій множині D, називається *опуклою*, якщо за будь-яких векторів $\bar{x}$ та $\bar{y}$ множини D і будь-якого $\lambda \in [0; 1]$ виконується нерівність

$$f(\lambda\bar{x} + (1-\lambda)\bar{y}) \le \lambda f(\bar{x}) + (1-\lambda) f(\bar{y}). \qquad (4)$$

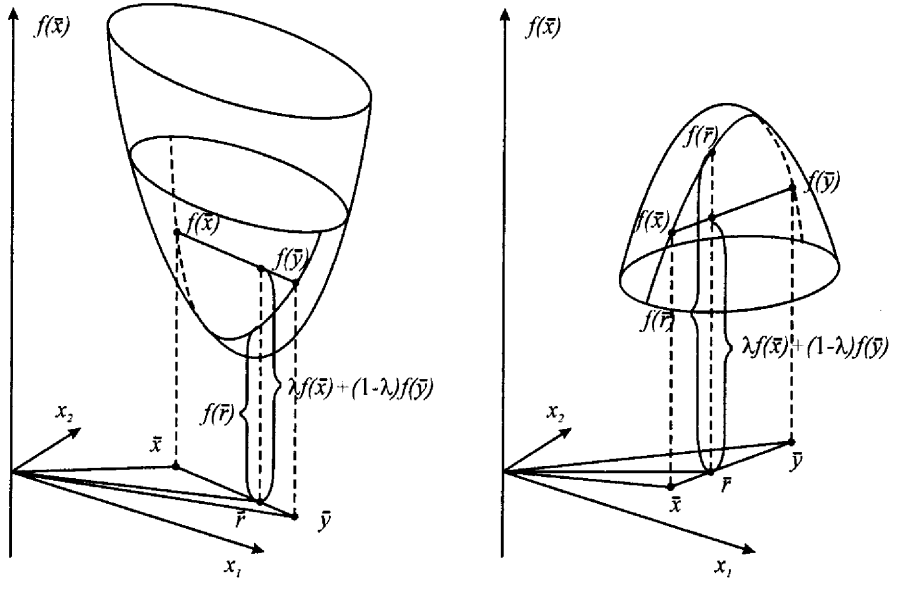
Якщо у співвідношенні (4) за умови $\lambda \in [0; 1]$ і будь-яких $\bar{x}$ та $\bar{y}$ виконується строга нерівність, то функція називається *строго опуклою*.

Функція $f(\bar{x})$, де $\bar{x} = (x_1, x_2, ..., x_n)$, визначена на опуклій множині D, називається *угнутою*, якщо за будь-яких векторів $\bar{x}$ та $\bar{y}$ множини D і будь-якого $\lambda \in [0;1]$ виконується нерівність

$$f(\lambda\bar{x}+(1-\lambda)\bar{y}) \geq \lambda f(\bar{x}) + (1-\lambda)f(\bar{y}). \qquad (5)$$

Якщо у співвідношенні (5) за умови $\lambda \in [0;1]$ і будь-яких $\bar{x}$ та $\bar{y}$ виконується строга нерівність, то функція називається *строго угнутою*.

Геометричне тлумачення визначень (4) та (5) для функцій двох змінних наведено на рис. 11.3: опукла функція (рис. 11.3 а) $f(\bar{x})$ для точок відрізку $[\bar{x}, \bar{y}]$ не може приймати більших значень, ніж лінійна функція, яка з'єднує значення $f(\bar{x})$ та $f(\bar{y})$. Аналогічно: угнута функція (рис. 11.3 б) $f(\bar{x})$ не може приймати значень менших, ніж значення лінійної функції, яка з'єднує значення $f(\bar{x})$ та $f(\bar{y})$.



опукла функція *угнута функція*

а) б)

Рис. 11.3.

Розглянемо деякі властивості опуклих та угнутих функцій, суттєві для розбудови алгоритмів пошуку екстремумів таких функцій.

Якщо функції $f_j(\bar{x})$ $(j = 1, 2, .., k)$ є опуклими на опуклій множині D, то опуклою буде і невід'ємна лінійна комбінація названих функцій, тобто функція $F(\bar{x}) = \sum_{j=1}^{k} \lambda_j f_j(\bar{x})$ за умови $\lambda_j \geq 0$.

Теорема (без доведення). Якщо $\varphi(\bar{x})$ є опуклою функцією для всіх $\bar{x} \geq \bar{0}$, то опуклою буде і множина розв'язків системи $\varphi(\bar{x}) \leq b$, $\bar{x} \geq \bar{0}$.

Аналогічна теорема справедлива і для угнутих функцій. *У загальній теорії опуклих множин доведено, що їх перетин (спільна загальна частина) теж є опуклою множиною.* Як наслідок цього твердження маємо: множина розв'язків системи нерівностей $\varphi_i(\bar{x})$ $\{\leq = \geq\} b_i$ $(i = \overline{1,m})$ за умови $\bar{x} \geq \bar{0}$ опукла, якщо в нерівностях зі знаком $\leq$ функції φ_i є опуклими, а в нерівностях зі знаком $\geq$ – угнутими.

Нагадаємо зміст поняття "градієнт функції".

Градієнтом функції $f(\bar{x})$ в точці $\bar{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ називається вектор, компоненти якого є частинними похідними функції, обчисленими в заданій точці. Градієнт функції позначається $\bar{\nabla} f$ або *grad f*. Для функції трьох змінних у прямокутній системі координат *grad f* за визначенням буде записаний так:

$$\bar{\nabla} f(\bar{x}^*) \equiv grad\, f(\bar{x}^*) = \left(\frac{\partial f(\bar{x}^*)}{\partial x_1}; \frac{\partial f(\bar{x}^*)}{\partial x_2}; \frac{\partial f(\bar{x}^*)}{\partial x_3}\right) \quad \text{або}$$

$$\bar{\nabla} f(\bar{x}^*) \equiv grad\, f(\bar{x}^*) = \frac{\partial f(\bar{x}^*)}{\partial x_1}\bar{i} + \frac{\partial f(\bar{x}^*)}{\partial x_2}\bar{j} + \frac{\partial f(\bar{x}^*)}{\partial x_3}\bar{k} \quad (\bar{i}, \bar{j}, \bar{k} \text{ – орти}).$$

Для функції в n-мірному просторі градієнтом називається вектор з компонентами, які дорівнюють значенням частинних похідних, обчисленим у точці $\bar{x}^*$:

$$\bar{\nabla} f(\bar{x}^*) \equiv grad\, f(\bar{x}^*) = \left(\frac{\partial f(\bar{x}^*)}{\partial x_1}; \frac{\partial f(\bar{x}^*)}{\partial x_2}; ...; \frac{\partial f(\bar{x}^*)}{\partial x_n}\right).$$

У подальшому будуть використані такі властивості *grad f*:

1) *градієнт визначає напрям найшвидшого зростання функції в точці*, тобто якщо в n-мірному просторі маємо точку $\bar{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ і досліджуємо величину зростання функції за умови приросту на задану величину значень $\bar{x}$ в певному напрямі, то зростання функції буде найбільшим у напрямі $grad\, f(\bar{x}^*)$;

2) *градієнт функції розташований завжди по нормалі до поверхні рівня*

$f(\bar{x}) = f(\bar{x}^*)$, для двомірного простору – по нормалі до лінії рівня (рис. 11.4). Шлях руху зміни величин x_1, x_2 від точки S до точки M – максимального значення функції, поверхня якої зображена на рис. 11.4, показано лінією SM.

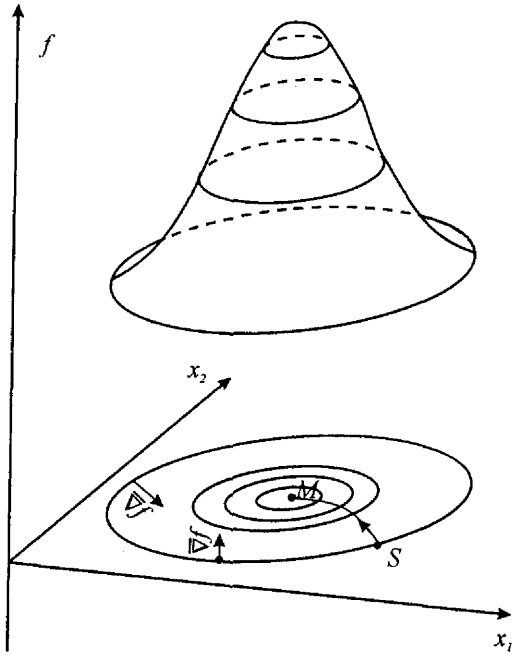


Рис. 11.4.

Напрям найшвидшого спадання функції визначається вектором, що називається *антиградієнтом*: – $grad\, f(\bar{x}^*)$.

Опукла диференційована функція $f(\bar{x})$, визначена на опуклій множині D, досягає свого *глобального мінімуму* в кожній точці $\bar{x}$, в якій градієнт функції дорівнює нулю.

Теорема (без доведення). Локальний мінімум опуклої функції $f(\bar{x})$, визначеної на опуклій множині D, співпадає з глобальним мінімумом на D.

Аналогічним є твердження стосовно локального максимуму угнутої функції, визначеної на опуклій множині.

Зауваження. У літературі іноді використовується дещо інша термінологія: функція опукла донизу та функція опукла доверху.

За допомогою наведених означень досить важко визначити опуклість або угнутість певної функції.

На практиці для двічі диференційованих функцій використовують такий спосіб. Будують матрицю H:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}. \qquad (6)$$

Двічі диференційована функція $f(\bar{x})$ є строго угнутою (опуклою доверху) в околі точки $\bar{x}^*$, якщо всі головні мінори матриці H (мінори, що відповідають головній діагоналі), обчислені в точці $\bar{x}^*$, непарного порядку від'ємні, а парного порядку додатні, тобто:

$$\frac{\partial^2 f}{\partial x_1^2} < 0; \quad \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{vmatrix} > 0; \quad \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{vmatrix} < 0; \ldots$$

Якщо всі названі мінори (парного та непарного порядку) додатні, то функція $f(\bar{x})$ строго опукла (опукла донизу) в точці $\bar{x}^*$. Якщо $f(\bar{x})$ опукла (угнута) для будь-якого значення $x \in D$, то вона називається опуклою (угнутою) в області D.

Приклад 1. Дослідити функцію $f = x_1^2 - 2x_1x_2 + 3x_2^2$ на опуклість (угнутість) в точці $\bar{x} = (2; 3)$. Знайдемо частинні похідні та побудуємо матрицю H (6) і головні мінори:

$$\frac{\partial f}{\partial x_1} = 2x_1 - 2x_2; \quad \frac{\partial^2 f}{\partial x_1^2} = 2; \quad \frac{\partial f}{\partial x_2} = -2x_1 + 6x_2; \quad \frac{\partial^2 f}{\partial x_2^2} = 6; \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -2.$$

$$H=\begin{bmatrix}2 & -2\\ -2 & 6\end{bmatrix};\quad \frac{\partial^2 f}{\partial x_1^2}=2;\quad \begin{vmatrix}2 & -2\\ -2 & 6\end{vmatrix}=12-4=8.$$

Головні мінори матриці H додатні і не залежать від змінних, тобто *досліджувана функція строго опукла в усіх областях її визначення.*

Приклад 2. Дослідити функцію $f(\bar{x})=x_1^3-2x_1x_2+x_2^3$ на опуклість (угнутість). Знайдемо частинні похідні:

$$\frac{\partial f}{\partial x_i}=3x_1^2-2x_2;\quad \frac{\partial^2 f}{\partial x_1^2}=6x_1;\quad \frac{\partial f}{\partial x_2}=-2x_1+3x_2^2;\quad \frac{\partial^2 f}{\partial x_2^2}=6x_2;\quad \frac{\partial^2 f}{\partial x_1\partial x_2}=-2;$$

Запишемо матрицю H та обчислимо головні мінори.

$$H=\begin{bmatrix}6x_1 & -2\\ -2 & 6x_2\end{bmatrix};\quad \frac{\partial^2 f}{\partial x_1^2}=6x_1;\quad \begin{vmatrix}6x_1 & -2\\ -2 & 6x_2\end{vmatrix}=36x_1x_2-4.$$

Якщо $x_1>0$ і $x_2>\dfrac{1}{9x_1}$, то функція $f(\bar{x})=x_1^3-2x_1x_2+x_2^3$ *опукла.* Якщо $x_1<0$ і $x_2>\dfrac{1}{9x_1}$, то функція строго *угнута,* бо визначник першого порядку $\left(\dfrac{\partial^2 f}{\partial x_1^2}\right)$ від'ємний, а визначник другого порядку – додатний.

Оскільки для опуклої (угнутої) функції локальний екстремум співпадає з глобальним, то в опуклій області існування функції *може бути лише один екстремум.* За таких властивостей опуклих функцій алгоритм пошуку екстремуму істотно спрощується в порівнянні із загальною задачею нелінійного програмування.

11.3. Умовний екстремум класичної теорії. Метод множників Лагранжа, їх економічний зміст

Класична постановка задачі пошуку умовного екстремуму функції кількох змінних: знайти вектор $\bar{x}=(x_1, x_2, ..., x_n)$, за якого цільова функція

$$f(\bar{x})=f(x_1, x_2, ..., x_n) \tag{7}$$

досягає екстремуму за умови, що компоненти шуканого вектора задовольняють системі рівнянь:

$$\varphi_i(\bar{x})=\varphi_i(x_1, x_2, ..., x_n)=b_i\ \ (i=\overline{1,m}). \tag{8}$$

Функції $f(\bar{x})$ та $\varphi_i(\bar{x})$ $(i=\overline{1,m})$ вважаються неперервними разом зі своїми похідними першого порядку. Особливість задачі (7), (8) в тому, що *обмеження (8) задані системою рівнянь.* Класичний підхід до розв'язання цієї задачі надає необхідні умови, представлені системою функціональних рівнянь, яким мають задовольняти компоненти шуканого оптимального вектора $\bar{x}$. Для функції двох змінних за наявності *лінійного обмеження* геометричне тлумачення умовного мінімуму наведено на рис. 11.5: поверхня опуклої функції $z=f(x_1, x_2)$ перетинається площиною $\varphi_2(x_1, x_2)=ax_1+bx_2+c=0$. За такої умови найменше значення функції $f(x_1, x_2)$ буде в точці M. При лінійному обмеженні $\varphi_1(x_1, x_2)=0$ найменше значення буде в точці N.

За методом Лагранжа *пошук умовного екстремуму функції n-змінних* заміняється *пошуком безумовного екстремуму* спеціально побудованої функції L (функції Лагранжа) $n+m$ змінних:

$$L(\bar{x},\lambda)=f(\bar{x})+\sum_{i=1}^{m}\lambda_i(b_i-\varphi_i(\bar{x})),\ \lambda=(\lambda_1, \lambda_2, ..., \lambda_m) \tag{9}$$

(λ_i $(i=\overline{1,m})$ – шукані величини, які називаються множниками Лагранжа).

За прийнятої умови диференційованості функцій $f(\bar{x})$ та $\varphi_i(\bar{x})$ $(i=\overline{1,m})$ необхідні умови існування локального екстремуму записуються системою функціональних рівнянь:

$$\begin{cases}\dfrac{\partial L}{\partial x_j}=\dfrac{\partial f(\bar{x})}{\partial x_j}-\sum\limits_{i=1}^{m}\lambda_i\dfrac{\partial\varphi_i(\bar{x})}{\partial x_j}=0;\quad (j=\overline{1,n});\\ \dfrac{\partial L}{\partial\lambda_i}=b_i-\varphi_i(\bar{x})=0\quad (i=\overline{1,m}).\end{cases} \tag{10}$$

У загальному випадку пошук розв'язку системи рівнянь (10) є дуже складним, бо рівняння нелінійні і не існує числових універсальних ефективних методів їх розв'язання. Якщо ж розв'язок (10) знайдено, то обчислені значення $\bar{x}=(x_1, x_2, ..., x_n)$ підставляють в цільову функцію $f(\bar{x})$ і аналізують результати. У загальному випадку задача (7), (8), (9) має не єдиний розв'язок. У задачах *опуклого програмування* цільова функція може мати лише один локальний екстремум, який співпадає з глобальним.

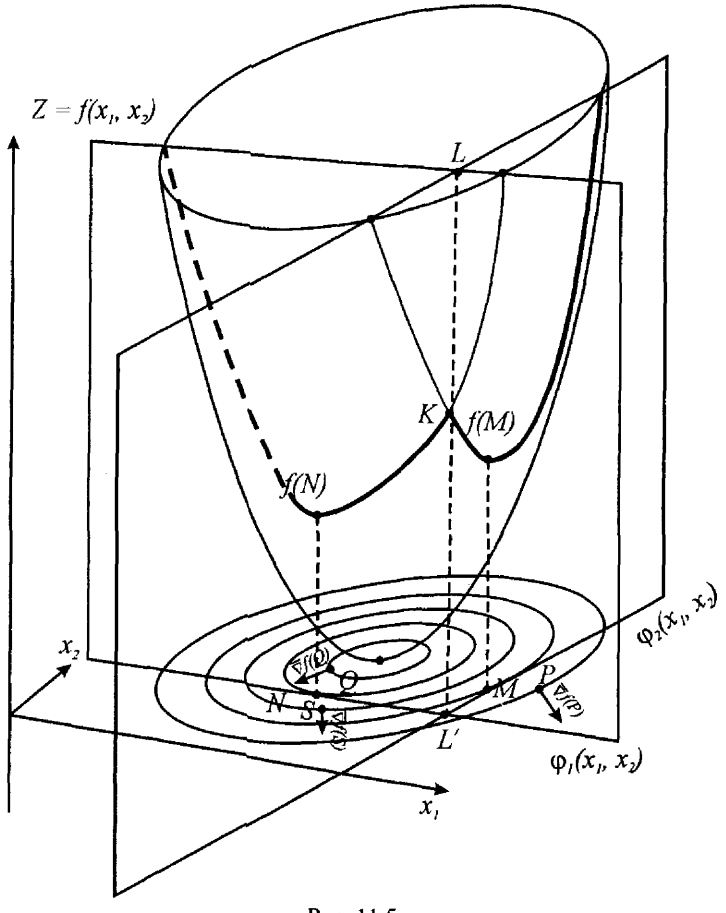


Рис. 11.5.

Для пошуку умовного екстремуму задачі (7), (8) за методом Лагранжа необхідно:

1) записати функцію Лагранжа (9);

2) записати систему рівнянь (10), дорівнюючи частинні похідні функції Лагранжа нулю;

3) знайти розв'язок системи рівнянь (10), тобто координати стаціонарних точок функції Лагранжа в об'єднаному просторі змінних $x_1, x_2, ..., x_n$ та $\lambda_1, \lambda_2, ..., \lambda_m$;

4) знайдені координати стаціонарних точок $x_1, x_2, ..., x_n$ підставити в цільову функцію та виконати аналіз результатів, щоб визначити, за яких координат і якого змісту досягаються локальні екстремуми.

Пошук умовного екстремуму методом Лагранжа має теоретично-пізнавальне значення. Основні перепони для його широкого використання такі: необхідність розв'язання системи нелінійних функціональних рівнянь (10), проблема аналізу множини локальних екстремумів.

Розглянемо **економічне тлумачення множників Лагранжа**. Скористаємося задачею пошуку умовного екстремуму у двомірному просторі: знайти

$$extr\, f(x_1, x_2) \tag{11}$$

за умови

$$\varphi(x_1, x_2) = b. \tag{12}$$

Припустимо, що шуканий умовний екстремум досягається в точці $\bar{x}^* = (x_1^*, x_2^*)$ та $\lambda = \lambda^*$, а відповідне екстремальне значення функції $f(\bar{x})$ позначимо f^*, тобто $f^* = f(x_1^*, x_2^*)$.

Приймаємо тепер, що величина b в обмеженні (12) змінна, тоді координати x_1^* і x_2^* локального умовного екстремуму та відповідне значення цільової функції (11) будуть величинами, залежними від b, тобто певними функціями від b: $x_1^* = x_1^*(b)$, $x_2^* = x_2^*(b)$, $f^* = f(x_1^*(b), x_2^*(b))$. Обчислимо похідну від цільової функції $f^*(\bar{x})$ по змінній b:

$$\frac{df^*}{db} = \frac{\partial f^*}{\partial x_1} \cdot \frac{dx_1^*}{db} + \frac{\partial f^*}{\partial x_2} \cdot \frac{dx_2^*}{db}. \tag{13}$$

Скориставшись умовою (12), визначаємо:

$$\frac{d\varphi^*}{db} = \frac{\partial \varphi^*}{\partial x_1} \cdot \frac{dx_1^*}{db} + \frac{\partial \varphi^*}{\partial x_2} \cdot \frac{dx_2^*}{db} = 1. \tag{14}$$

Але в точці умовного екстремуму мають виконуватись умови (10), тобто

$$\begin{cases} \dfrac{\partial f^*}{\partial x_1} - \lambda^* \dfrac{\partial \varphi^*}{\partial x_1} = 0 \\ \dfrac{\partial f^*}{\partial x_2} - \lambda^* \dfrac{\partial \varphi^*}{\partial x_2} = 0 \\ \varphi^* - b = 0. \end{cases} \tag{15}$$

Скориставшись, що $\frac{\partial f^*}{\partial x_1} = \lambda^* \frac{\partial \varphi^*}{\partial x_1}$, $\frac{\partial f^*}{\partial x_2} = \lambda^* \frac{\partial \varphi^*}{\partial x_2}$, та рівняннями (13) і (14), знаходимо:

$$\frac{df^*}{db} = \lambda^* \frac{\partial \varphi^*}{\partial x_1} \cdot \frac{dx_1^*}{db} + \lambda^* \frac{\partial \varphi^*}{\partial x_2} \cdot \frac{dx_2^*}{db} = \lambda^*.$$

Для загальної задачі на умовний екстремум (7), (8) маємо:

$$\frac{\partial f^*}{\partial b_i} = \lambda_i^* \quad (i = \overline{1, m}). \tag{16}$$

Якщо цільову функцію $f(x_1, x_2, ..., x_n)$ тлумачити як прибуток або вартість, а b_i – як обсяги використовуваних ресурсів (за умови (8) ресурси використовуються в повному обсязі), то множники Лагранжа λ_i^* показують, як змінюється максимальний прибуток (або мінімальна вартість) за умови, що кількість i-го ресурсу змінюється на одиницю.

11.4. Задача опуклого програмування. Графо-аналітичний розв'язок у просторі двох змінних

У теорії опуклого програмування за основну модель обрано задачу пошуку *найменшого значення опуклої функції*

$$f = f(\bar{x}) \equiv f(x_1, x_2, ..., x_n) \tag{17}$$

за умови

$$\varphi_i(\bar{x}) \equiv \varphi_i(x_1, x_2, ..., x_n) \le 0; \; x_j \ge 0; \; (i = \overline{1, m}; \; j = \overline{1, n}), \tag{18}$$

де $\varphi_i(\bar{x})$ – опуклі функції, що забезпечує опуклість множини допустимих розв'язків, визначену умовою (18).

Якщо $f(\bar{x})$ та $\varphi_i(\bar{x})$ угнуті, то розглядають задачу пошуку найбільшого значення цільової функції за умови $\varphi_i(\bar{x}) \ge 0$ і $\bar{x} \ge 0$.

У просторі двох змінних x_1 та x_2 задача (17), (18) має графо-аналітичний розв'язок. Загальна схема розв'язання така:

а) у прямокутній системі координат побудувати лінії рівня цільової функції. Доцільно побудувати і графічне зображення поверхні $z = f(x_1, x_2)$;

б) побудувати графічне зображення області допустимих розв'язків (18);

в) за графічним зображенням ліній рівня $f(\bar{x})$ та області допустимих розв'язків визначити розташування точки, в якій досягається шуканий мінімум. За необхідності виконати додаткові розрахунки.

Приклад. Знайти найменше значення функції

$$f(x_1, x_2) = \frac{1}{4}x_1^2 + \frac{1}{9}x_2^2 - 3\frac{1}{2}x_1 - 2x_2 + 21\frac{1}{4} = \frac{(x_1 - 7)^2}{4} + \frac{(x_2 - 9)^2}{9}$$

за умови

$$\begin{cases} 3x_1 + x_2 \le 18 & (1) \\ 27x_1 + 16x_2 \le 188 & (2) \\ x_1 \ge 0; \quad x_2 \ge 0. \end{cases}$$

У прямокутній системі координат $x_1 0 x_2$ (рис. 11.6) побудуємо лінії рівня у точці (7; 9), з осями, перпендикулярними осям координат, та їх довжинами $2\sqrt{c}$ і $3\sqrt{c}$ (c – значення функції на лінії рівня $\frac{(x_1 - 7)^2}{4} + \frac{(x_2 - 9)^2}{9} = c$).

Область допустимих значень обмежена прямими лініями $3x_1 + x_2 = 18$; $27x_1 + 16_2 = 188$; $x_1 = 0$; $x_2 = 0$ і зображена чотирикутником *ABDO*.

Аналіз взаємного розташування ліній рівня цільової функції та області допустимих значень свідчить, що найменше значення функції буде відповідати лінії рівня, яка дотикається межі $27x_1 + 16x_2 = 188$ області допустимих значень.

Вилучивши відповідні розрахунки, наведемо їх результати. Координати точки M: $x_1 = 4$; $x_2 = 5$. Найменше значення цільової функції за даної області допустимих розв'язків дорівнює

$$f(4; 5) = \frac{(4 - 7)^2}{4} + \frac{(5 - 9)^2}{9} = \frac{145}{36}.$$

Розв'язок перевіримо методом Лагранжа, врахувавши, що найменше значення досягається на лінії $27x_1 + 16x_2 = 188$. За такої умови необхідно розв'язати задачу: знайти найменше значення $f(x_1, x_2) = \frac{(x_1 - 7)^2}{4} + \frac{(x_2 - 9)^2}{9}$ при $27x_1 + 16x_2 = 188$. Будуємо функцію Лагранжа (9):

$$L = \frac{(x_1 - 7)^2}{4} + \frac{(x_2 - 9)^2}{9} + \lambda(188 - 27x_1 - 16x_2).$$

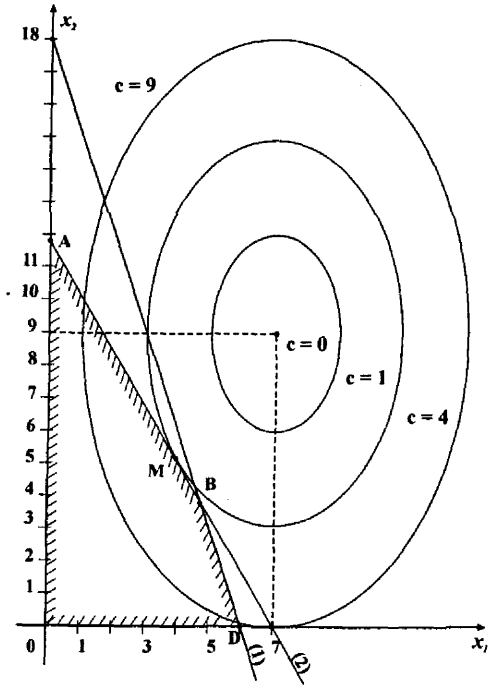


Рис. 11.6.

Система рівнянь (10), які визначають необхідні умови наявності екстремуму, має вигляд:

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = \dfrac{x_1 - 7}{2} - 27\lambda = 0 \Rightarrow x_1 - 54\lambda = 7 \\ \dfrac{\partial L}{\partial x_2} = \dfrac{2(x_2 - 9)}{9} - 16\lambda = 0 \Rightarrow 2x_2 - 144\lambda = 18 \\ \dfrac{\partial L}{\partial \lambda} = 188 - 27x_1 - 16x_2 = 0 \Rightarrow 27x_1 - 16x_2 = 188. \end{cases}$$

Одержана СЛАР має розв'язок $x_1^* = 4$; $x_2^* = 5$; $\lambda^* = -\dfrac{1}{18}$; значення x_1 та x_2 співпадають з наведеними раніше.

11.5. Необхідні та достатні умови оптимальності в задачі опуклого програмування. Теорема Куна-Таккера

Основна задача опуклого програмування визначається умовами (17) та (18). Суттєво принциповою відмінністю цієї задачі від класичної на пошук умовного екстремуму (7) та (8) є те, що в загальній задачі *обмеження можуть бути визначені нерівностями певного змісту.* У сучасній теорії пошуку екстремуму опуклих функцій на опуклих множинах узагальнено класичний метод множників Лагранжа. Результатами такого узагальнення є теорема Куна-Таккера, визначальна в теорії опуклого програмування.

Використовуючи умову (18), будують функцію Лагранжа $L(\bar{x}, \bar{l})$:

$$L(\bar{x}, \bar{l}) = f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \varphi_i(\bar{x}) \tag{19}$$

($\bar{x} = (x_1, x_2, ..., x_n)$, $\bar{l} = (\lambda_1, \lambda_2, ..., \lambda_m)$).

Точка $(\bar{x}^*, \bar{l}^*)$ у просторі векторів $(\bar{x}, \bar{l}) \equiv (x_1, x_2, ..., x_n;\ \lambda_1, \lambda_2, ..., \lambda_m)$ називається *сідловою точкою* функції Лагранжа (19), якщо n-мірна точка $\bar{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ є точкою мінімуму функції $L(\bar{x}, \bar{l}^*)$, тобто для всіх $\bar{x}$ та $\bar{l}$ з околу точки $(\bar{x}^*, \bar{l}^*)$ виконується нерівність

$$L(\bar{x}^*; \bar{l}) \le L(\bar{x}^*; \bar{l}^*) \le L(\bar{x}; \bar{l}^*). \tag{20}$$

Звертаємо увагу читача, що запис $L(\bar{x}^*, \bar{l}^*)$ визначає число, яке дорівнює значенню функції Лагранжа (19) в точці $(\bar{x}^*, \bar{l}^*)$, а записи $L(\bar{x}^*; \bar{l})$ і $L(\bar{x}; \bar{l}^*)$ визначають функції змінних $\bar{l} = (\lambda_1, \lambda_2, ..., \lambda_m)$ та $\bar{x} = (x_1, x_2, ..., x_n)$, які отримують за допомогою функції Лагранжа (19) за умови, що зафіксовані лише компоненти векторів $\bar{x}$ або $\bar{l}$.

Теорема Куна-Таккера. Якщо існує такий вектор $\bar{x} \ge \bar{0}$, що виконуються умови $\varphi_i(\bar{x}) < 0$ $(i = \overline{1,m})$, то необхідною і достатньою умовою оптимальності вектора $\bar{x}^*$, який належить області допустимих розв'язків, є наявність такого вектора $\bar{l}^* \ge \bar{0}$, що для всіх $\bar{x} \ge \bar{0}$ і $\bar{l} \ge \bar{0}$ виконується нерівність (20).

Таким чином, теорема Куна-Таккера встановлює зв'язок між оптимальним планом задачі опуклого програмування (17) і (18) та сідловою точкою функції Лагранжа (19) для цієї задачі. Тому наведену теорему називають *теоремою про сідлову точку.*

Користуючись результатами розв'язку прикладу п. 11. 4, перевіримо виконання умов теореми Куна-Таккера. Згідно з (18) та (19) запишемо функцію Лагранжа

$$L(x_1, x_2, \lambda_1, \lambda_2) = \frac{(x_1-7)^2}{4} + \frac{(x_2-9)^2}{9} + \lambda_1(3x_1 + x_2 - 18) + \lambda_2(27x_1 + 16x_2 - 188). \quad (21)$$

Наголосимо, що функція $L(x_1, x_2, \lambda_1, \lambda_2)$ відрізняється від функції Лагранжа, використаної в розв'язанні прикладу 11.4, за умовами класичного умовного екстремуму. Підставимо результати розв'язку $x_1 = 4$; $x_2 = 5$; $\lambda_2 = -\frac{1}{18}$ ($\lambda_1 = 0$, тому що відповідне обмеження не є визначальним) у функцію Лагранжа (21):

$$L(\bar{x}^*, \bar{l}^*) = \frac{145}{36};\ L(\bar{x}^*, \bar{l}) = \frac{145}{36};\ L(\bar{x}, \bar{l}^*) = \frac{(x_1-7)^2}{4} + \frac{(x_2-9)^2}{9} - \frac{1}{18}(27x_1+16x_2-188).$$

Перевіримо виконання нерівностей (20). Ліве відношення виконується як рівність. Проаналізуємо праве відношення:

$$\frac{145}{36} \le \frac{(x_1-7)^2}{4} + \frac{(x_2-9)^2}{9} - \frac{1}{18}(27x_1 + 16x_2 - 188). \quad (22)$$

Значення цільової функції $f(\bar{x}) = \frac{(x_1-7)^2}{4} + \frac{(x_2-9)^2}{9}$ в області допустимих розв'язків не менше свого мінімального значення $\frac{145}{36}$, а другий доданок, згідно з вимогою обмежень стосовно допустимих розв'язків $27x_1 + 16x_2 - 188 \le 0$ та умовою $\lambda_2 = -\frac{1}{18}$, завжди буде невід'ємним, тобто нерівність (22) виконується. Якщо функції $f(\bar{x})$ та $\varphi_i(\bar{x})$ $(i = \overline{1, m})$ диференційовані, то нерівності (20) та $\bar{x} \ge \bar{0}$, $\bar{l} \ge \bar{0}$ еквівалентні так званим "локальним умовам" теореми Куна-Таккера:

$$\begin{cases} \dfrac{\partial L(\bar{x}^*, \bar{l}^*)}{\partial x_j} \ge 0;\ x_j^* \ge 0\ (j = \overline{1, n}); \\ \sum\limits_{j=1}^{n} x_j^* \dfrac{\partial L(\bar{x}^*, \bar{l}^*)}{\partial x_j} = 0 \\ \dfrac{\partial L(\bar{x}^*, \bar{l}^*)}{\partial \lambda_i} \le 0;\ \lambda_i^* \ge 0\ (j = \overline{1, m}); \\ \sum\limits_{i=1}^{m} \lambda_i^* \dfrac{\partial L(\bar{x}^*, \bar{l}^*)}{\partial \lambda_i} = 0. \end{cases} \quad (23)$$

11.6. Задача квадратичного програмування

До квадратичного програмування належать задачі нелінійного опуклого програмування, в яких *обмеження задані лінійними залежностями, а цільова функція є сумою лінійної та квадратичної форм.* Основну задачу квадратичного програмування формулюють так: знайти найменше значення функції

$$f(\bar{x}) = f(x_1, x_2, ..., x_n) = \sum_{j=1}^{n} c_j x_j + \sum_{k=1}^{n} \sum_{j=1}^{n} d_{kj} x_k x_j \quad (24)$$

за лінійних обмежень

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}) \quad (24а)$$

$$x_j \ge 0 \quad (j = \overline{1, n}). \quad (24б)$$

Стосовно запису квадратичної форми в наведеному вигляді доречно зробити зауваження: коефіцієнти при x_j^2 позначаються d_{jj}, а при $x_k x_j$ – $2d_{kj}$ $(k \ne j)$, приймаючи, що $d_{kj} = d_{jk}$, тобто за умови наявності доданку $a_{kj}x_k x_j$ його записують як $2\left(\frac{a_{kj}}{2}\right)x_k x_j = 2d_{kj}x_k x_j$. Квадратична форма

$$Q(\bar{x}) = \sum_{k=1}^{n} \sum_{j=1}^{n} d_{kj} x_k x_j \quad (25)$$

характеризується своєю матрицею

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad (26)$$

розмірності n, побудованої з коефіцієнтів d_{kj}.

Оскільки $d_{kj} = d_{jk}$ за будовою квадратичної форми, то матриця D буде симетричною. Якщо ранг r цієї матриці (його ще називають рангом квадратичної форми) дорівнює n, тобто матриця D (26) *невироджена*, то і квадратичну форму $Q(\bar{x})$ (25) називають *невиродженою.*

Квадратична форма $Q(\bar{x})$ (25) називається *додатно визначеною,* якщо вона за будь-яких значень невідомих (x_1, x_2, ..., x_n), серед яких хоча б одне є відмінним від нуля, приймає додатні значення. Якщо серед значень $Q(\bar{x})$ немає від'ємних, але нульові її значення можуть бути не лише за нульових значень змінних, то така форма називається *невід'ємно визначеною.* Якщо за будь-яких значень змінних $Q(\bar{x})$ приймає лише від'ємні значення, то вона називається *від'ємно визначеною.*

Названі властивості квадратичної форми $Q(\bar{x})$ (25) однозначно пов'язані зі знаками головних мінорів її матриці D (26):

$$d_{11};\quad \begin{vmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{vmatrix};\quad \begin{vmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{vmatrix};\quad \ldots;\quad \begin{vmatrix} d_{11} & d_{12} & \ldots & d_{1n} \\ d_{21} & d_{22} & \ldots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \ldots & d_{nn} \end{vmatrix}.$$

Квадратична форма $Q(\bar{x})$ (25) буде додатно визначеною тоді і лише тоді, коли всі головні мінори її матриці D будуть додатними. За умови, що знаки головних мінорів чергуються, квадратична форма буде від'ємно визначеною. Якщо ранг r матриці D менше її розмірності n, тобто $r < n$, квадратична форма $Q(\bar{x})$ (25) буде невід'ємно визначеною.

Приклад. $Q(x_1x_2) = 4x_1^2 + x_2^2 - 2x_1x_2$. Матриця D має вигляд:

$$D = \begin{vmatrix} 4 & -1 \\ -1 & 1 \end{vmatrix},$$

а її головні мінори 4; $\begin{vmatrix} 4 & -1 \\ -1 & 1 \end{vmatrix} = 3$ додатні; отже, наведена квадратична форма додатно визначена, тобто для всіх ненульових x_1 та x_2 приймає лише додатні значення. У цьому легко переконатися, бо $Q(x_1x_2) = (x_1 - x_2)^2 + 3x_1^2$.

Оскільки цільова функція і обмеження в задачі квадратичного програмування (24) диференційовані, то для її розв'язання можна використати локальні умови теореми Куна-Таккера, виконання яких гарантує оптимальність розв'язку. Запишемо функцію Лагранжа для досліджуваної задачі (24):

$$L = \sum_{j=1}^{n} c_j x_j + \sum_{k=1}^{n} \sum_{j=1}^{n} d_{kj} x_k x_j + \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right). \tag{27}$$

Знайдемо частинні похідні функції L (27):

$$\frac{\partial L}{\partial x_j} = c_j + 2\sum_{k=1}^{n} d_{kj} x_k + \sum_{i=1}^{m} \lambda_i a_{ij} \quad (j = \overline{1,n}) \tag{28}$$

$$\frac{\partial L}{\partial \lambda_i} = \sum_{j=1}^{n} (a_{ij} x_j - b_i) \quad (i = \overline{1,m}). \tag{29}$$

Для скорочення записів позначимо:

$$v_j = \frac{\partial L}{\partial x_j} = c_j + 2\sum_{k=1}^{n} d_{kj} x_k + \sum_{i=1}^{m} \lambda_i a_{ij} \quad (j = \overline{1,n})$$

$$y_i = -\frac{\partial L}{\partial \lambda_i} = b_i - \sum_{j=1}^{n} a_{ij} x_j \quad (i = \overline{1,m}).$$

За таких позначень локальні умови теореми Куна-Таккера (23) у координатній формі запишуться так:

$$\begin{cases} c_j + 2\sum_{k=1}^{n} d_{kj} x_k + \sum_{i=1}^{m} \lambda_i a_{ij} = v_j \geq 0 \quad (j = \overline{1,n}) \\ x_j v_j = 0, \quad x_j \geq 0 \quad (j = \overline{1,n}) \end{cases} \tag{30}$$

$$\begin{cases} b_i - \sum_{j=1}^{n} a_{ij} x_j = y_i \quad (i = \overline{1,m}) \\ \lambda_i \; y_i = 0; \lambda \geq 0 \quad (i = \overline{1,m}). \end{cases} \tag{31}$$

Об'єднавши названі умови та змінивши порядок їх запису, маємо:

$$\sum_{j=1}^{n} a_{ij} x_j + y_i = b_i \quad (i = \overline{1,m}) \tag{32}$$

$$2\sum_{k=1}^{n} d_{kj} x_k + \sum_{i=1}^{m} a_{ij} \lambda_i - v_j = -c_j \quad (j = \overline{1,n}) \tag{33}$$

$$\begin{cases} x_j \geq 0; \quad v_j \geq 0 \quad (j = \overline{1,n}) \\ y_i \geq 0, \quad \lambda_i \geq 0 \quad (i = \overline{1,m}) \end{cases} \tag{34}$$

$$\sum_{j=1}^{n} x_j v_j + \sum_{i=1}^{m} y_i \lambda_i = 0. \qquad (35)$$

Таким чином, згідно з локальними умовами теореми Куна-Таккера план $\bar{x}^* = (x_1^*, x_2^*, ..., x_n^*)$ є оптимальним для задачі квадратичного програмування (24), (24 а, б) тоді і лише тоді, коли сумісно з розв'язком $\bar{v} = (v_1, v_2..., v_n)$ існують такі вектори $\bar{l} = (\lambda_1, \lambda_2..., \lambda_m)$ та $\bar{y} = (y_1, y_2..., y_m)$, які задовольняють системі СЛАР (32) та (33) за умови виконання рівності (35).

Використання локальних умов теореми Куна-Таккера стосовно задачі квадратичного програмування дало можливість побудувати для розв'язання нелінійної задачі систему $n + m$ лінійних алгебраїчних рівнянь з $2(n + m)$ невідомими $(x_1, x_2, ..., x_n;\ v_1, v_2..., v_n;\ \lambda_1, \lambda_2..., \lambda_m;\ y_1, y_2..., y_m)$. Але розв'язок побудованої системи має задовольняти додатковій *комбінаторній умові* (35), згідно з якою для кожної пари невід'ємних змінних x_j та v_j, y_i та λ_i *хоча б одна змінна має дорівнювати нулю.* Тобто не всі розв'язки системи (32), (33) задовольняють (35), а лише ті, в яких не менше $n + m$ компонент дорівнюють нулю. Саме такі властивості мають базисні розв'язки системи лінійних алгебраїчних рівнянь, тому пошук розв'язку, який задовольняє умові (35), необхідно виконувати лише серед них. Таким чином, для розв'язання сформульованої задачі можна використати симплексний метод, якщо побудувати належним чином цільову функцію. За таку функцію T обирають ліву частину рівняння (35):

$$T = \sum_{j=1}^{n} x_j v_j + \sum_{i=1}^{m} y_i \lambda_i,$$

яка є лінійною стосовно кожної змінної з вищеназваних сукупностей $\bar{x}^* = (x_1^*, x_2^*, ..., x_n^*)$, $\bar{v} = (v_1, v_2..., v_n)$, $\bar{l} = (\lambda_1, \lambda_2..., \lambda_m)$ та $\bar{y} = (y_1, y_2..., y_m)$, але нелінійною на об'єднанні множин змінних. Така ситуація обумовлює певні особливості в реалізації симплексного методу для задачі квадратичного програмування, які не будемо розглядати. Ознайомимося з основними поняттями більш універсальних методів.

11.7. Градієнтні методи

До градієнтних методів належать числові методи пошуку екстремумів, розбудованих з використанням властивості градієнта визначати в точці напрям найбільшого зростання відповідної функції. Такими методами можна розв'язувати будь-яку задачу нелінійного програмування, але при їх використанні визначається

лише *локальний екстремум.* Градієнтні методи доцільно використовувати для розв'язання задач опуклого програмування; вони забезпечують значну ефективність, тому що в таких задачах *локальний екстремум тотожний глобальному.* Розглянемо основні положення розбудови деяких градієнтних методів.

Необхідно знайти найбільше значення *угнутої* (опуклої догори) диференційованої функції $f(x_1, x_2, ..., x_n)$ без *будь-яких обмежень на область зміни аргументів.* Один з варіантів градієнтних методів у такому випадку будується так. Беремо будь-яку точку $\bar{x}^0 = (x_1^0, x_2^0, ..., x_n^0)$, в якій визначена функція $f(\bar{x})$. Знаходимо в цій точці $grad\ f = \bar{\nabla} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n}\right)$ і в напрямі, визначеному градієнтом, змінюємо величини аргументів:

$$\bar{x} = \bar{x}^0 + h\bar{\nabla} f(\bar{x}^0) \qquad \text{(запис у векторній формі)} \Rightarrow$$

$$\Rightarrow x_j = x_j^0 + h\frac{\partial f(\bar{x}^0)}{\partial x_j} \qquad \text{(запис у координатах)}$$

(h – числовий параметр, який визначає величину "кроку"). Вибравши h та обчисливши $\bar{\nabla} f(\bar{x}^0)$, за наведеними формулами визначаємо координати наступної точки $\bar{x}_1$. У досягнутій точці знову обчислюємо градієнт $\bar{\nabla} f$ та робимо наступний крок у його напрямі. Якщо позначити певну досягнуту точку $\bar{x}_k$, а наступну точку $\bar{x}_{k+1}$, то розрахункові формули можна записати так:

$$\bar{x}_{k+1} = \bar{x}_k + h_k \bar{\nabla} f(\bar{x}_k) \Rightarrow x_j(k+1) = x_j(k) + h_k \frac{\partial f}{\partial x_j}(\bar{x}_k). \qquad (36)$$

При реалізації будь-якого варіанта градієнтного методу суттєвими є два питання:

а) як вибирати величину кроку h_k;

б) як довго виконувати обчислення.

На практиці ці питання можуть бути розв'язані так. Вибирають певну величину h і виконують обчислення до тих пір, поки $f(\bar{x}_{k+1}) > f(\bar{x}_k)$. За протилежного відношення величину h зменшують у два рази і знову роблять крок з точки $\bar{x}_k$. Іноді величину h_k вибирають пропорційно модулю градієнта у вихідній точці $\bar{x}_k$.

У прикладній математиці розроблено чимало ефективних методів пошуку екстремуму за названих умов. Пошук екстремуму можна завершити (оптимальний розв'язок буде знайдено з деякою похибкою), скориставшись такими міркуваннями. Порівнюючи приріст функції після кожного кроку, приймемо за наближене значення шуканого екстремуму ту точку, навколо якої приріст функції $\bar{\nabla} f = f(\bar{x}_{k+1}) -$

$f(\bar{x}_k)$ за модулем менше обраної малої величини δ. Іноді за критерій завершення пошуку екстремуму використовують величину зміни аргументів. Критерій завершення пошуку екстремуму може бути визначений і в такий спосіб:

$$\sum_{j=1}^{n}\left(\frac{\partial f(\bar{x}_{k+1})}{\partial x_j}\right)^2 < \delta .$$

Такий вибір критерію обумовлений тим, що в точці екстремуму диференційованої функції частинні похідні мають дорівнювати нулю.

Опишемо ще один з можливих варіантів градієнтного методу: **пошук найшвидшого зростання (спадання)** значення угнутої (опуклої) функції $f(\bar{x})$ за умови відсутності обмежень на змінні величини.

Припустимо, що маємо певну точку $\bar{x}_k = (x_{k1}, x_{k2}, ..., x_{kn})$. Обчислимо в ній градієнт функції та розглянемо зміну функції лише в напрямі градієнта:

$$\bar{x}_{k+1} = \bar{x}_k + t\bar{\nabla} f(\bar{x}_k) \;\Rightarrow\; x_{k+1,j} = x_{k,j} + t\,\frac{\partial f(\bar{x}_k)}{\partial x_j} \quad \text{(у координатній формі)} \qquad (37)$$

(t – змінна величина).

Якщо аргументи функції, починаючи з точки $\bar{x}_k = (x_{k1}, x_{k2}, ..., x_{kn})$, змінювати за (37), то будемо мати для досліджуваної функції *залежність лише від однієї змінної t*. Знайдемо за такої умови екстремум досліджуваної функції чисельно (для функції однієї змінної наявний значний набір методів пошуку екстремумів) або аналітично, позначивши відповідну величину t_k. У точці $\bar{x}_{k+1} = \bar{x}_k + t_k\bar{\nabla} f(\bar{x}_k)$, обчисливши градієнт досліджуваної функції, визначимо новий напрям зміни аргументів і розглянемо функцію однієї змінної.

У загальному випадку побудову функції однієї змінної t по досягненні екстремуму за умови руху в напрямі градієнта від попередньої точки $\bar{x}_k$ можна записати так:

$$f(\bar{x}_{k+1}(t)) = f(\bar{x}_k + t_k\bar{\nabla} f(\bar{x}_k)) + t\bar{\nabla} f(\bar{x}_k + t_k\bar{\nabla} f(\bar{x}_k)) \qquad (38)$$

(t_k – раніше визначена величина).

Критеріями завершення обчислень можуть бути критерії, названі для попереднього варіанта градієнтного методу.

Скориставшись припущенням про диференційованість функції, знайдемо умову для обчислення величини t, за якої досягається екстремум при зміні аргументів у напрямі градієнта, обчисленого в точці $\bar{x}_k$. За такої умови функція $f(\bar{x})$

буде функцією лише однієї змінної t. Екстремум може бути досягнутий лише в стаціонарній точці, тому

$$\frac{df\left(x_{k1} + t\frac{\partial f(\bar{x}_k)}{\partial x_1}; x_{k2} + t\frac{\partial f(\bar{x}_k)}{\partial x_2}; ...; x_{kn} + t\frac{\partial f(\bar{x}_k)}{\partial x_n}\right)}{dt} = \sum_{j=1}^{n}\frac{\partial f(\bar{x}_k + t\bar{\nabla} f(\bar{x}_k))}{\partial x_j}\frac{\partial f(\bar{x}_k)}{\partial x_j} =$$

$$= \bar{\nabla} f(\bar{x}_k + t\bar{\nabla} f(\bar{x}_k)) \cdot \bar{\nabla} f(\bar{x}_k) = 0. \qquad (39)$$

Результат (39) має геометричне тлумачення: якщо з точки $\bar{x}_k = (x_{k1}, x_{k2}, ..., x_{kn})$ рухатися в напрямі градієнта $\bar{\nabla} f(\bar{x}_k)$, то екстремальне значення функції за такої умови буде в точці $(\bar{x}_k + t\bar{\nabla} f(\bar{x}_k))$, в якій градієнт перпендикулярний вихідному. Для функції двох змінних схема переміщень аргументів у площині при використанні методу найшвидшого зростання (спадання) наведена на рис. 11.7.

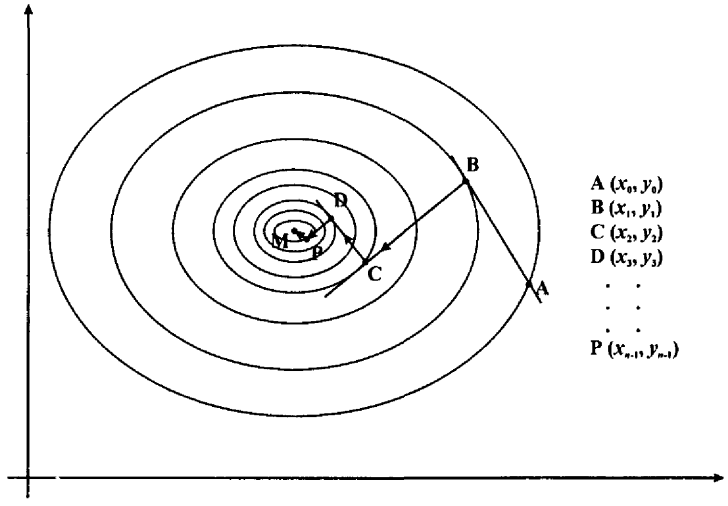


Рис. 11.7.

Приклад. Знайти найбільше значення

$$f(x, y) = 10x + 40y - x^2 - 2y^2 + 50.$$

Запишемо матрицю D квадратичної форми (26) та обчислимо її головні мінори:

$$D = \begin{vmatrix} -1 & 0 \\ 0 & -2 \end{vmatrix}; \; d_{11} = -1 < 0; \; |D| = 2 > 0.$$

Знаки головних мінорів чергуються; отже, квадратична форма від'ємно визначена, функція угнута, екстремум лише один. Запишемо:

$$\overline{\nabla} f = (10 - 2x)\bar{i} + (40 - 4y)\bar{j} = (10 - 2x; 40 - 4y).$$

За вихідну точку беремо $x = 0$; $y = 1$. Отже, $\overline{\nabla} f(0; 1) = (10; 36)$, і з вихідної точки рухаємось у напрямі $x = 0 + 10t$; $y = 1 + 36t$. За такої зміни аргументів функція від t буде визначатися залежністю:

$$f(10t; 1+36t) = 100t + 40(1 + 36t) - (10t)^2 - 2(1 + 36t)^2 + 50.$$

За алгоритмом необхідно знайти найбільше значення цієї функції. Воно буде таким у стаціонарній точці, тобто $\frac{df}{dt} = 0$. Знайшовши $\frac{df}{dt} = 1396 - 5384t$ та прирівнявши до нуля, маємо $t = \frac{1396}{5384} = 0{,}259287$. Для зменшення обсягів обчислень візьмемо $t = 0{,}26$. Обчислимо координати досягнутої точки: $x_1 = 0 + 10 \cdot 0{,}26 = 2{,}6$; $y_1 = 1 + 36 \cdot 0{,}26 = 10{,}36$. У досягнутій точці (x_1, y_1) обчислимо градієнт функції $\overline{\nabla} f(x_1, y_1) = (10 - 2 \cdot 2{,}6; 40 - 4 \cdot 10{,}36) = (4{,}8; -1{,}44)$. Від досягнутої точки (x_1, y_1) рухаємося за напрямом градієнта, тобто $x_2 = 2{,}6 + 4{,}8t$; $y_2 = 10{,}36 - 1{,}44t$. За такої зміни аргументів функція від t буде визначатися такою залежністю:

$$f(2{,}6+4{,}8t; 10{,}36-1{,}44t) = 10(2{,}6+4{,}8t) + 40(10{,}36-1{,}44t) - (2{,}6+4{,}8t)^2 - 2(10{,}36-1{,}44t)^2 + 50.$$

Знайшовши $\frac{df}{dt}$ та розв'язавши рівняння $\frac{df}{dt} = 0$, маємо $t \approx 0{,}46$. Досягнуті значення аргументів: $x_2 = 2{,}6 + 4{,}8 \cdot 0{,}46 \approx 4{,}8$; $y_2 = 10{,}36 - 1{,}44 \cdot 0{,}46 \approx 9{,}7$. Обчислимо $\overline{\nabla} f(x_2, y_2) = (10 - 2 \cdot 4{,}8; 40 - 4 \cdot 9{,}7) = (0{,}4; 1{,}2)$. Від точки $(4{,}8; 9{,}7)$ рухаємося за напрямом градієнта, тобто $x_3 = 4{,}8 + 0{,}4t$, $y_3 = 9{,}7 + 1{,}2t$. За таких значень x_3 та y_3 залежність функції від t буде визначатися так:

$$f(4{,}8 + 0{,}4t; 9{,}7 + 1{,}2t) = 10(4{,}8 + 0{,}4t) + 40(9{,}7 + 1{,}2t) - (4{,}8+0{,}4t)^2 - 2(9{,}7 + 1{,}2t)^2 + 50.$$

Вимога $\frac{df}{dt} = 0$ обумовлює значення $t = 0{,}26$. Досягнута точка $x_3 \approx 5$; $y_3 \approx 10$. Результат $x_3 = 5$; $y_3 = 10$ обумовлює нульові значення компонентів градієнта в досягнутій точці, що свідчить про одержання координат найбільшого значення функції $f(5; 10) = 10 \cdot 5 + 40 \cdot 10 - (5)^2 - 2(10)^2 + 50 = 275$, тобто оптимальний план $\bar{x} = (5; 10)$.

Розглянемо задачу пошуку екстремуму при наявності обмежень. Наголосимо, що за умови, коли цільова функція угнута (опукла), тобто вона має єдиний екстремум і достовірно відомо, що *точка екстремуму розташована в середині області, обумовленої обмеженнями,* – можна користуватися описаними градієнтними методами без будь-яких обмежень. Тому розглядаємо задачу оптимізації за відсутності відомостей про розташування точки екстремуму стосовно області допустимих значень.

Знайти найбільше значення неперервно диференційованої угнутої функції

$$f(\bar{x}) = f(x_1, x_2, ..., x_n) \tag{40}$$

за умови лінійних обмежень на допустимі значення аргументів:

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i \quad (i = \overline{1, m}) \tag{41}$$

$$x_j \ge 0 \quad (j = \overline{1, n}).$$

Пояснення пропонованого алгоритму будемо супроводжувати геометричним тлумаченням відповідних розрахунків, користуючись рис. 11.8, на якому зображено лінії рівня двомірної функції та область D допустимих значень її аргументів. Припустимо, що вихідна точка $\bar{x}_0 = (x_{10}, x_{20}, ..., x_{n0})$ є внутрішньою точкою області D (рис. 11. 8). У напряму $\overline{\nabla} f(\bar{x}_0)$ доцільно рухатися лише до тих пір, поки не буде досягнуто хоча б одного з обмежень області або найбільшого значення функції в напрямі $\overline{\nabla} f(\bar{x}_0)$. У нашому випадку треба зупинитися на межі області; якби не було такої межі, то можна було б рухатися до точки A. Досягнувши межі в точці $\bar{x}_1$, треба вибрати такий напрям руху, щоб, не виходячи з допустимої області D, забезпечити найбільш можливе зростання функції.

Напрям такого руху визначиться вектором $\bar{r}_1$, який мусить утворювати з вектором $\overline{\nabla} f(\bar{x}_1)$ (напрямом найбільшого зростання функції в точці $\bar{x}_1$) найменший гострий кут у порівнянні з будь-яким іншим вектором, що виходить із точки $\bar{x}_1$ і розташований в допустимій області. Компоненти такого вектора обчислюються за умови забезпечення найбільшої величини скалярного добутку $\overline{\nabla} f(\bar{x}_1)\bar{r}_1$. Оскільки обмеження лінійні, то напрям $\bar{r}_1$ буде визначатися напрямом досягнутої прямої лінії межі області.

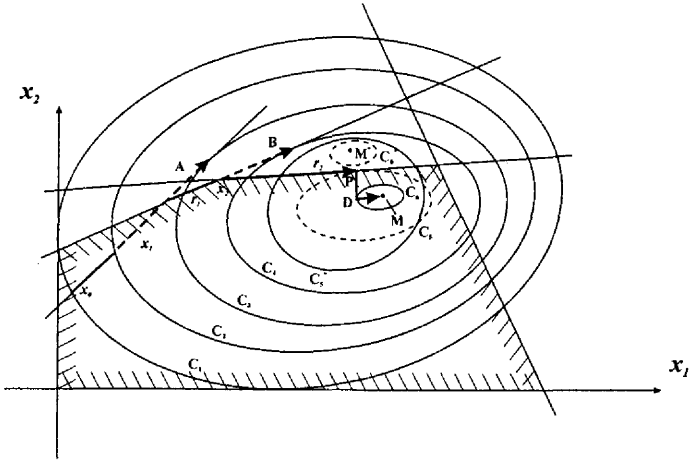


Рис. 11.8.

Таким чином, досягнувши межі, необхідно на наступному кроці рухатися по граничній прямій до тих пір, поки зростає $f(\bar{x})$ або буде досягнута інша межа, в нашому випадку – до точки $\bar{x}_2$ (рис. 11.8). Із точки $\bar{x}_2$ напрям подальшого руху визначається вектором $\bar{r}_2$ – орієнтацією наступної досягнутої частини межі. У досліджуваній ситуації в напрямі вектора $\bar{r}_2$ доцільно рухатися лише до точки P, яка розташована на прямій, що є не лише межею області D, а й дотичною до певної лінії рівня функції $f(\bar{x})$ (геометричне тлумачення умови (39)), тобто в точці P досягається максимум за умови руху із точки $\bar{x}_2$ в напрямі $\bar{r}_2$. За такої умови в подальшому може бути два випадки. Якщо найбільше значення функції $f(\bar{x})$ *без урахування допустимих обмежень на аргументи знаходиться за межами допустимої області D*, то точка P буде *точкою глобального максимуму в області допустимих розв'язків*. Лінії рівня функції такої залежності від аргументів на рис. 11.8 зображено пунктиром. Якщо найбільше значення $f(\bar{x})$ знаходиться у внутрішній точці допустимої області D (лінії рівня C_5, C_6 за такої ситуації зображені неперервною лінією), то від точки P подальший рух буде по $\overline{\nabla} f(x_{1,p};\ x_{2,p};\ \ldots;\ x_{n,p})$. Розв'язок задачі (40), (41) може бути завершений без урахування (41), якщо точка максимуму знаходиться в області D. Можливі подальші переміщення в пошуках максимуму зображені векторами $\overline{PD}$ та $\overline{DM}$.

Розглянемо алгоритм розв'язання задачі (40), (41), що ґрунтується на вищенаведених міркуваннях. Якщо пошук максимуму розпочинається від внутрішньої точки області D (41), тобто всі обмеження виконуються як строгі нерівності, то рух необхідно починати в напрямі градієнта функції згідно з (36) або (37), але вибір величини кроку h_k (відповідно t в (37)) обумовлений вимогою, щоб

наступна точка $\bar{x}_{k+1}$ не вийшла за межі області D. Аналітично ця вимога означає, що координати $\bar{x}_{k+1} = (x_{1,k+1};\ x_{2,k+1};\ \ldots;\ x_{n,k+1})$ мають задовольняти обмеженням (41):

$$\begin{cases} \sum a_{ij}\left(x_{jk} + h_k \dfrac{\partial f(\bar{x}_k)}{\partial x_j}\right) \le b_i \quad (i = \overline{1, m}) \\ x_{j,k} + h_k \dfrac{\partial f(\bar{x}_k)}{\partial x_j} \ge 0 \quad (j = \overline{1, n}). \end{cases} \tag{42}$$

Розв'язавши систему нерівностей (42), знаходимо проміжок $[h_s,\ h_f]$ допустимих значень величини кроку h_k. Величина кроку t, яка означає розташування локального максимуму за рухом у напрямі $\overline{\nabla} f(\bar{x}_k)$, визначається розв'язком рівняння (39). Якщо t є внутрішньою точкою $[h_s, h_f]$, то величина кроку h_k дорівнює t і досягнута точка $\bar{x}_{k+1} = \bar{x}_k + h_k \overline{\nabla} f(\bar{x}_k)$ буде внутрішньою точкою області D. Аналогічно виконується наступний крок. Якщо t не належить проміжку $[h_s,\ h_f]$ ($t > h_f$), то це означає, що, рухаючись з точки $\bar{x}_k$ до локального максимуму в напрямі $\overline{\nabla} f(\bar{x}_k)$, слід зупинитися на межі допустимої області D. Аналітично це означає, що необхідно взяти $h_k = h_f$, а досягнута за такої величини кроку h_k точка $\bar{x}_{k+1}$ буде такою, що в системі нерівностей (42) частина залежностей виконуватиметься як рівності. За такої умови наступний крок розпочинається з точки, розташованої на межі області D. Якщо вихідна точка наступного кроку розташована на межі області D, то необхідно розв'язати допоміжну задачу з визначення напряму руху $\bar{r}_k$.

Знайти найбільше значення величини скалярного добутку

$$w_k = \overline{\nabla} f(\bar{x}_k)\ \bar{r}_k \tag{43}$$

за умови

$$\sum a_{ij} r_{kj} \le 0 \tag{44}$$

для тих значень індексу i, для яких виконуються рівності

$$\sum_{j=1}^{n} a_{ij}\, x_j = b_i \tag{45}$$

та умови нормування вектора $\bar{r}_k$ (напрям визначається вектором одиничної довжини):

$$|\bar{r}_k| = \sqrt{\sum_{j=1}^{n} r_{kj}^2}\,. \qquad (46)$$

Розв'язок задачі (43) – (46) дає можливість визначити напрям руху, якщо вихідна точка $\bar{x}_k$ розташована на межі області D.

Виконання умови (44) свідчить, що рух від точки $\bar{x}_k$ *буде виконуватися не за межі* допустимої області (41), а по межі або в середину допустимої області (див. точку P рис. 11. 8).

Визначивши $\bar{r}_k$ (напрям руху), обчислюють величину кроку t_k в цьому напрямі. Для цього використовують вимоги необхідної умови екстремуму (39), замінивши вектор $\bar{\nabla} f(\bar{x}_k)$ на вектор $\bar{r}_k$, тобто розв'язавши рівняння

$$\bar{\nabla} f(\bar{x}_k + t\bar{r}_k) \cdot \bar{r}_k = 0. \qquad (47)$$

Маючи розв'язок (47), визначають величину кроку t_k в напрямі $\bar{r}_k$ так, як це було описано раніше. Умови завершення пошуку екстремуму були сформульовані при описанні алгоритму градієнтного методу за відсутності обмежень.

11.8. Метод штрафних функцій

Як наголошувалося раніше, для реалізації числових методів розв'язання *задач безумовної оптимізації* необхідно виконати істотно менший обсяг обчислень, ніж при розв'язанні задач нелінійного математичного програмування. Наприклад, за реалізації алгоритму пошуку *умовного екстремуму* в задачі квадратичного програмування з використанням властивостей градієнта, крім обчислень, обумовлених визначенням градієнта, необхідно розв'язувати на кожному кроці певну задачу лінійного програмування (43) – (47). З таких причин природним є прагнення розбудови таких числових методів, які б давали можливість задачу з обмеженнями привести до певної задачі пошуку безумовного екстремуму.

Одним із засобів досягнення такої мети є *метод штрафних функцій*. Під такою назвою об'єднано різні модифікації методів, які дозволяють *задачу умовної оптимізації* заданої функції $f(\bar{x})$ привести до *задачі безумовної оптимізації* деякої допоміжної функції $w(\bar{x})$, пов'язаної певним чином з вихідною:

$$w(\bar{x}) = f(\bar{x}) + s(\bar{x}) \qquad (48)$$

($s(\bar{x})$ – так звана функція штрафу, за допомогою якої прагнуть вплинути на функцію $f(\bar{x})$ за умови порушення деяких обмежень умовного екстремуму). Наявність різних

варіантів методу штрафних функцій пов'язана з особливостями методики побудови штрафної функції $s(\bar{x}) = s(x_1, x_2, ..., x_n)$.

Вибір аналітичної залежності для функції штрафу $s(\bar{x})$ узгоджують з використовуваним методом пошуку екстремуму функції $w(\bar{x})$ та вимогою найбільш простого обчислення як $w(\bar{x})$, так і її частинних похідних. Функцію штрафу $s(\bar{x})$ будують при виконанні умов:

1) $s(\bar{x}) = 0$ для всіх точок $\bar{x}$ області допустимих розв'язків D;

2) $s(\bar{x}) > 0$ для точок, що не належать області D.

Широко використовуються штрафні функції такого аналітичного вигляду:

$$s(\bar{x}) = \sum_{i=1}^{m} [\max(\varphi_i(\bar{x}); 0)]^p, \quad p = 1 \text{ або } 2,$$

$$s(\bar{x}) = \sum_{i=1}^{m} 0{,}25\ [\varphi_i(\bar{x}) + |\varphi_i(\bar{x})|]^2;$$

($\varphi_i(\bar{x})$ – функції, якими визначають межі області допустимих розв'язків: $\varphi_i(\bar{x}) = 0$).

Розглянемо один із варіантів методу штрафних функцій. Необхідно знайти план (вектор), при якому цільова функція

$$f = f(\bar{x}) \qquad (49)$$

досягає найбільшого значення за умови виконання обмежень

$$\varphi_i(x_1, x_2, ..., x_n) \geq 0; \quad (i = \overline{1, m})$$
$$x_j \geq 0 \quad (j = \overline{1, n}). \qquad (50)$$

Вважаємо, що $f(\bar{x})$ та всі $\varphi_i(\bar{x})$ є диференційованими функціями, до того ж $f(\bar{x})$ – угнута або опукла в області її визначення.

Доведено, що алгоритм пошуку найбільшого значення $f(\bar{x})$ задачі (49), (50) буде збіжним при виборі штрафної функції $s(\bar{x})$ за правилом:

$$s(\bar{x}) = \sum_{i=1}^{m} \alpha_i(\bar{x}) \cdot \varphi_i(\bar{x}), \qquad (51)$$

$\alpha_i(\bar{x})$ – функції, які приймають дискретні значення:

$$\alpha_i(\bar{x}) = \begin{cases} 0, \text{ якщо } \varphi_i(\bar{x}) \geq 0, & i = \overline{1, m} \\ \alpha, \text{ якщо } \varphi_i(\bar{x}) < 0, & i = \overline{1, m} \end{cases} \tag{52}$$

(α – деяка додатна величина, яка приймає дискретні значення і вибрана так, щоб при розташуванні точки $\bar{x} = (x_1, x_2, ..., x_n)$ за межами допустимої області (50) градієнт функції $w(\bar{x})$ (48) був направлений до межі області (50)).

Пропонований варіант методу штрафних функцій дає можливість задачу (49), (50) *пошуку умовного максимуму функції* $f(\bar{x})$ привести до задачі *пошуку безумовного максимуму функції* $w(\bar{x})$ (48) градієнтним методом.

Згідно з методом, розташування наступної точки $\bar{x}_{k+1} = (x_{k+1,1}; x_{k+1,2}; ...; x_{k+1,n})$ за відомої досягнутої точки $\bar{x}_k = (x_{k1}, x_{k2}, ..., x_{kn})$ обчислюється за формулою:

$$x_{k+1,j} = \max\left\{0; x_{k,j} + h\frac{\partial f(\bar{x}_k)}{\partial x_j} + \sum_{i=1}^{m}\left(\alpha_i(\bar{x}_k)\frac{\partial \varphi_i(\bar{x}_k)}{\partial x_j}\right)\right\} \tag{53}$$

(h визначає величину кроку градієнтного методу). Аналіз формули (53) з урахуванням умови (52) показує, що, доки точка $\bar{x}_k$ знаходиться в області допустимих значень (50), розташування наступної точки $\bar{x}_{k+1}$ визначається лише градієнтом цільової функції $f(\bar{x})$ (49), тому що другий доданок у круглих дужках дорівнює нулю. Тобто якщо обмеження не порушуються, то процес пошуку максимуму виконується так, ніби обмеження відсутні. Порушення хоча б одного, наприклад, p-го обмеження (50) на k-му кроці обумовлює появу ненульового коефіцієнта $\alpha_p(\bar{x}_k)$. За такої ситуації розташування наступної точки буде визначатися з урахуванням і другого доданку в (53). Саме цей доданок забезпечує переміщення чергової точки, координати якої обчислюються за (53), в допустиму область (50).

Розв'язання задачі (49), (50) фахівці радять починати при порівняно малих відносно одиниці значеннях α. При досягненні в першому наближенні оптимального плану треба збільшити величину α та уточнити розв'язок, починаючи з досягнутого наближеного плану. Завершення пошуку виконується за тими ж критеріями, що й для задачі безумовного екстремуму градієнтним методом.

Для реалізації алгоритму пошуку оптимального плану (розв'язку) за методом штрафних функцій в задачі нелінійного програмування необхідно:

1) вибрати певний вихідний план (вихідну точку); бажано, але не обов'язково, щоб його компоненти задовольняли обмеженням (50). Вибрати певні значення h та α. Величину h рекомендується вибрати наближено рівною одній десятій величини найбільшого з проміжків зміни аргументів задачі. Починати вибір величини α можна із значення 0,1. Вихідні значення h та α на практиці визначають за досвідом розв'язання певного класу задач;

2) за наявних компонент плану перевірити виконання обмежень та визначити аналітичний вигляд цільової функції $w(\bar{x})$ (48) з урахуванням побудови штрафної функції $s(\bar{x})$ (51);

3) обчислити градієнт функції w (48);

4) якщо всі обмеження (50) виконано, відхилення від нуля у кожному з обмежень малі в порівнянні з одиницею і довжина градієнта (корінь квадратний із суми квадратів компонент) близька до нуля, то наявний план можна вважати задовільним наближенням до оптимального і завершити розрахунки, обчисливши значення функції цілі. У протилежній ситуації виконувати п. 5. За умови, коли довжина градієнта та відхилення в (50) від нуля істотні, необхідно проаналізувати доцільність заміни величини кроку h та коефіцієнта α в (51). Визначивши h та α, перейти на п. 5;

5) обчислити компоненти наступного плану згідно з (53). Перейти на виконання п. 2.

Приклад. Знайти найбільше значення функції

$$f(x, y) = 20x - x^2 + 12y - y^2 \tag{54}$$

за умови

$$\begin{cases} \varphi_1(x, y) = 30 - 3x - 5y \geq 0 & (3x + 5y \leq 30), \\ \varphi_2(x, y) = 20 - x - 5y \geq 0 & (x + 5y \leq 20), \\ x \geq 0; y \geq 0. \end{cases} \tag{55}$$

Використовуємо найменування змінних x та y для зручності записів. Лінії рівня цільової функції (54) та область допустимих розв'язків зображені на рис. 11.9.

Допоміжна функція $w(x, y)$ (48) має вигляд:

$$w(x, y) = 20x - x^2 + 12y - y^2 + \alpha(30 - 3x - 5y) + \alpha(20 - x - 5y). \tag{56}$$

Вибираємо для $\varphi_1(x, y)$ та $\varphi_2(x, y)$ однакові α, хоч це і необов'язково. Запишемо необхідні для обчислень частинні похідні:

$$\frac{\partial f}{\partial x} = 20 - 2x; \quad \frac{\partial f}{\partial y} = 12 - 2y; \quad \frac{\partial \varphi_1}{\partial x} = -3; \quad \frac{\partial \varphi_1}{\partial y} = -5; \quad \frac{\partial \varphi_2}{\partial x} = -1; \quad \frac{\partial \varphi_2}{\partial y} = -5.$$

Виберемо вихідну точку $x_0 = 2$; $y_0 = 2$ та величини кроку $h = 0{,}1$ і $\alpha = 0{,}25$.

Підставивши координати вихідної точки (2; 2) в (55), пересвідчуємося, що умова (55) виконується. Тому координати наступної точки згідно з (53) обчислюємо за формулами:

$$x_1 = x_0 + h\frac{\partial f(x_0, y_0)}{\partial x} = \left| x_0 = 2;\ y_0 = 2;\ h = 0{,}1 \right| = 3{,}6;$$

$$y_1 = y_0 + h\frac{\partial f(x_0, y_0)}{\partial y} = \left| y_0 = 2;\ x_0 = 2;\ h = 0{,}1 \right| = 2{,}8\,.$$

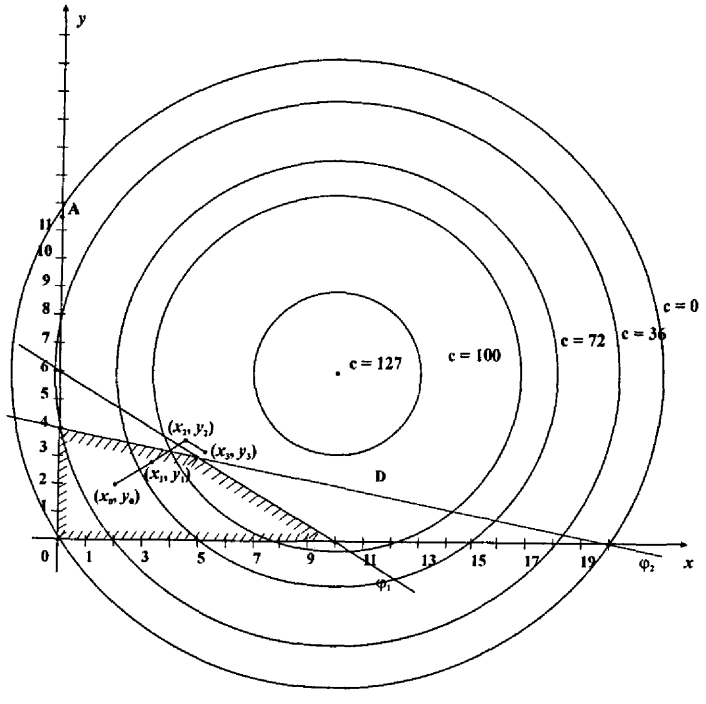


Рис. 11.9.

Координати досягнутої точки (3,6; 2,8) підставляємо в умову (55) і пересвідчуємося, що умова виконується. Отже, координати наступної точки згідно з (53) обчислюємо за формулами:

$$x_2 = x_1 + h\frac{\partial f(x_1, y_1)}{\partial x} = \left| x_1 = 3{,}6;\ y_1 = 2{,}8;\ h = 0{,}1 \right| = 4{,}88;$$

$$y_2 = y_1 + h\frac{\partial f(x_1, y_1)}{\partial y} = \left| x_1 = 3{,}6;\ y_1 = 2{,}8;\ h = 0{,}1 \right| = 3{,}44.$$

За координат досягнутої точки (4,88; 3,44) умова (55) не виконується, тому напрям руху від цієї точки буде визначатися з урахуванням функції штрафу:

$$x_3 = x_2 + h\left(\frac{\partial f(x_2, y_2)}{\partial x} + \alpha\left(\frac{\partial \varphi_1(x_2, y_2)}{\partial x} + \frac{\partial \varphi_2(x_2, y_2)}{\partial x}\right)\right) = \left| \begin{matrix} x_2 = 4{,}88;\ y_2 = 3{,}44 \\ h = 0{,}1;\ \alpha = 0{,}25 \end{matrix} \right| = 5{,}804;$$

$$y_3 = y_2 + h\left(\frac{\partial f(x_2, y_2)}{\partial y} + \alpha\left(\frac{\partial \varphi_1(x_2, y_2)}{\partial y} + \frac{\partial \varphi_2(x_2, y_2)}{\partial y}\right)\right) = \left| \begin{matrix} x_2 = 4{,}88;\ y_2 = 3{,}44 \\ h = 0{,}1;\ \alpha = 0{,}25 \end{matrix} \right| = 3{,}702.$$

Результати розв'язання прикладу наведені в табл. 11.1. Вплив функції штрафу на функцію $w(x, y)$ (56) відмічено у стовпці w.

Таблиця 11.1.

№	(x_k, y_k)	h	α	φ_1	φ_2	w	$\nabla w(x_k; y_k)$	x_{k+1}, y_{k+1}	$f(x_{k+1}, y_{k+1})$
1	(2;2)	0,1	0,25	14>0	18>0	f	$\frac{\partial W}{\partial x} = 16;\ \frac{\partial W}{\partial y} = 8$	3,6;2,8	84,8
2	3,6;2,8	0,1	0,25	5,2>0	2,4>0	f	$\frac{\partial W}{\partial x} = 12{,}8;\ \frac{\partial W}{\partial y} = 6{,}4$	4,88;3,44	103,23
3	4,88;3,44	0,1	0,25	–1,84<0	–2,08<0	$f+\alpha\varphi_1+$ $+\alpha\varphi_2$	$\frac{\partial W}{\partial x} = 9;\ \frac{\partial W}{\partial y} = 2{,}62$	5,804;3,702	–
4	5,80;3,70	0,1	0,5	–5,92<0	–4,31<0	$f+\alpha\varphi_1+$ $+\alpha\varphi_2$	$\frac{\partial W}{\partial x} = 6{,}39;\ \frac{\partial W}{\partial y} = -0{,}004$	6,443;3,702	–
5	5,44;3,70	0,1	1	–7,839<0	–4,953<0	$f+\alpha\varphi_1+$ $+\alpha\varphi_2$	$\frac{\partial W}{\partial x} = 3{,}114;\ \frac{\partial W}{\partial y} = -5{,}004$	6,754;3,162	–
6	6,75;3,16	0,2	1	–6,071<0	–2,256<0	$f+\alpha\varphi_1+$ $+\alpha\varphi_2$	$\frac{\partial W}{\partial x} = 2{,}491;\ \frac{\partial W}{\partial y} = 4{,}323$	7,004;2,729	–
7	7,00;2,72	0,2	1	–3,324<0	1,262>0	$f+\alpha\varphi_1$	$\frac{\partial W}{\partial x} = 2{,}494;\ \frac{\partial W}{\partial y} = 2{,}706$	7,752;2,778	–
8	7,75;2,77	0,2	1	–1,643<0	–7,146<0	$f+\alpha\varphi_1+$ $+\alpha\varphi_2$	$\frac{\partial W}{\partial x} = 0{,}496;\ \frac{\partial W}{\partial y} = -3{,}556$	7,851;2,067	–
9	7,85;2,06	0,2	2	–3,887<0	1,814>0	$f+\alpha\varphi_1$	$\frac{\partial W}{\partial x} = -1{,}708;\ \frac{\partial W}{\partial y} = -2{,}134$	7,570;1,640	–
10	7,57;1,64	0,2	2	–0,9106<0	4,229>0	$f+\alpha\varphi_1$	$\frac{\partial W}{\partial x} = -1{,}141;\ \frac{\partial W}{\partial y} = -1{,}28$	7,342;1,384	–
11	7,34;1,38	0,2	2	1,08>0	5,76>0	f	$\frac{\partial W}{\partial x} = 5{,}315;\ \frac{\partial W}{\partial y} = 9{,}232$	7,448;1,568	109,8
12	7,44;1,56	0,2	2	–0,188<0	4,708>0	$f+\alpha\varphi_1$	$\frac{\partial W}{\partial x} = -0{,}897;\ \frac{\partial W}{\partial y} = -1{,}138$	7,430;1,546	109,5

Заміна величин h і α в процесі розрахунку була виконана за аналізом

проміжних результатів з метою прискорення збіжності. Обчислення завершені по стабілізації першої цифри в десятковій частині шуканих величин x, y – компонент плану. Наближене значення плану, який гарантує найбільше значення цільової функції $f(x, y)$ (54) за умови (55), – 7,430; 1,546. Для досягнення (при необхідності) більш точних результатів необхідно зменшити величину h. Завершення розрахунків (див. табл. 11.1) виконано за h = 0,2. Точне значення компонент оптимального плану – 7,353; 1,588.

Питання для самоконтролю

1. За яких властивостей задачі пошуку оптимальних рішень належать до класу задач нелінійного математичного програмування?
2. Як формулюється задача квадратичного програмування?
3. Як визначаються опукла та угнута функції?
4. Яка різниця в поняттях локального, умовного та глобального екстремумів?
5. За яких умов можна користуватися класичним методом пошуку оптимальних розв'язків?
6. За яких умов використовується метод Лагранжа для пошуку оптимальних розв'язків?
7. Властивості градієнта функції та їх використання при пошуку оптимального розв'язку. Формули пошуку компонент наступного наближення до оптимального рішення.
8. Основна мета розбудови методу штрафних функцій. За якими правилами будуються штрафні функції?
9. Зміст теореми Куна-Таккера.

ГЛАВА 12

МЕТОДИ ПЛАНУВАННЯ ТА УПРАВЛІННЯ МЕРЕЖАМИ

12.1. Призначення та сфера використання

Темпи виробництва, його масштаби та спеціалізація окремих галузей, багатопрофільні зв'язки обумовлюють необхідність розробки ефективних методів планування та управління, які б давали можливість оцінити змінний стан системи та передбачити її майбутнє, щоб оптимізувати відповідний процес і керувати його перебігом. Системи об'єктів дослідження разом зі зв'язками між ними називаються *мережею*. Діапазон реального існування мереж дуже широкий: мережі електропостачання, радіо- та телекомунікацій, транспортні (залізничні, автомобільні), об'єкти господарювання як в одному господарстві, так і в їх комплексі, плани виконання робіт з реалізації певних проектів і т. ін. Але прикладами таких систем можуть бути також організація поточного виробництва, реконструкція існуючого виробництва, організація капітального будівництва, реконструкція та ремонт існуючих споруд, організація науково-дослідних робіт і т. ін., де також необхідно узгоджувати та оцінювати зв'язки між окремими елементами. Методи планування та управління мережею забезпечують:

- складання календарного плану виконання певного комплексу робіт;
- оцінку необхідних трудових, матеріальних та фінансових ресурсів, затрат часу;
- контроль комплексу робіт з прогнозуванням і запобіганням можливих зривів при виконанні робіт;
- ефективне управління при чіткому розподілі відповідальності між керівниками різних рівнів і виконавцями робіт;
- оцінку дієздатності та якості системи стосовно певних критеріїв.

Для одних систем зв'язки між об'єктами реалізовані фізично (система комунікацій між населеними пунктами), для інших мають інформаційний характер або є поєднанням як фізичної реалізації, так і інформаційної.

У цій главі розглядаються дві задачі стосовно управління мережами: планування робіт з реалізації певного проекту та оцінка пропускної спроможності даної мережі.

Математичний апарат, який використовується при дослідженні мереж, розроблений у так званій теорії графів.

12.2. Основні поняття теорії графів

Побудова математичних моделей розв'язання вказаних задач планування та управління мережами ґрунтується на дослідженні попарних (бінарних) зв'язків між об'єктами, які утворюють систему дослідження. Графічне зображення множини досліджуваних об'єктів і зв'язків між ними називається *графом*. Граф доцільно зображати у вигляді діаграми. На діаграмі об'єкти зображаються пронумерованими точками або кружками, які називаються вершинами, зв'язки між об'єктами – відрізками ліній, які з'єднують відповідні об'єкти. Якщо зв'язок між двома об'єктами *A* та *B* однобічний (від *A* до *B* є зв'язок, а зворотний зв'язок відсутній), то це зображається орієнтованим відрізком, стрілка якого відповідає напрямку зв'язку. Такий однобічний орієнтований відрізок називається *дугою*, а графічне зображення неорієнтованих попарних зв'язків між об'єктами – *ребрами* (ситуація, коли об'єкт *A* може бути пов'язаний з об'єктом *B* і навпаки). У подальшому ми не будемо в термінології відрізняти поняття графу та його діаграми. Граф, вершини якого мають лише однобічний зв'язки, називається *орієнтованим*, або *орграфом*.

Приклади графів.

1. Карта автомобільних доріг є граф, вершини якого – населені пункти, а зв'язки (ребра або дуги у випадку однобічного руху) – дороги, які з'єднують населені пункти.

2. При розв'язанні задач технічної діагностики (наприклад, при пошуку зіпсованого елемента) досліджувану систему умовно поділяють на кілька взаємопов'язаних частин, кожна з яких виконує певну функцію. Схему функціонування такої системи зручно зобразити як орієнтований граф, вершинами якого будуть виділені частини системи, а дугами – функціональні зв'язки між цими частинами.

На рис. 12.1 зображено схему використання обладнання при запуску бензинового двигуна внутрішнього згоряння (а) та її граф (б).

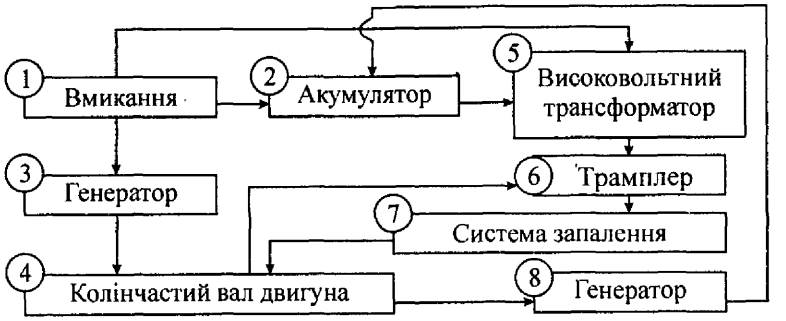
Граф вважається *завантаженим*, якщо він визначений разом з певною функцією на множині його ребер або дуг. Така функція може визначати віддаль між вершинами (карта доріг), час або вартість перевезень між населеними пунктами, пропускну спроможність лінії електропередач або каналу системи зрошення.

Маршрутом називається така послідовність ребер, коли кожна пара сусідніх ребер має одну загальну вершину.

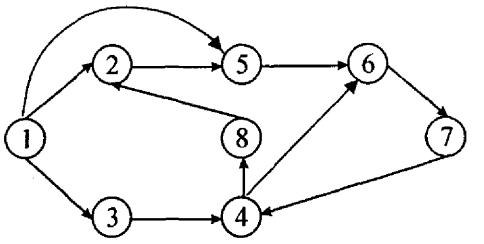
Простим ланцюгом називається маршрут, в якому вершини не повторюються. У подальшому будемо користуватися лише простими ланцюгами, опускаючи слово простий. Ланцюг визначається послідовністю вершин, через які він проходить.

Цикл – це ланцюг, початкова вершина якого співпадає з кінцевою.

Шляхом називається орієнтований ланцюг. Отже, поняття "шлях" стосується лише орієнтованих графів.



а)



б)

Рис. 12.1.

12.3. Побудова правильної нумерації вершин графу

Взагалі вершини графу можна нумерувати довільно, але для розв'язання багатьох практичних задач зручно виконати так звану правильну нумерацію вершин. За такої нумерації будь-який шлях від вершини з меншим номером до вершини з більшим номером буде проходити лише через вершини зі зростаючими номерами. Правильна нумерація вершин виконується за так званим "алгоритмом викреслення дуг".

Опишемо зміст цього алгоритму, пояснюючи його етапи стосовно завантаженого графу, зображеного на рис. 12.2, для якого у квадратах наводиться деяка довільна нумерація вершин.

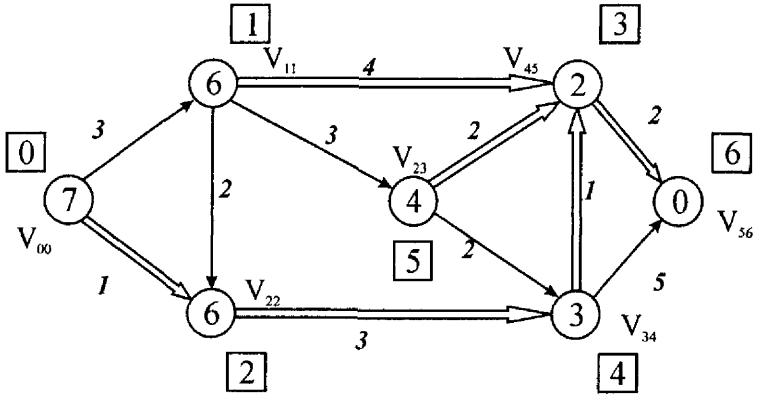


Рис. 12.2.

I. Умовно виділимо всі дуги, які виходять з початкової вершини, назвемо її вершиною нульового рангу та дамо їй номер "00" (V_{00}). Тепер розглянемо вершини, в які не заходять інші дуги, окрім викреслених. Такі вершини назвемо вершинами 1-го рангу та пронумеруємо їх у довільному порядку, дотримуючись неперервності в нумерації. За цих умов кожній вершині будемо надавати два індекси: перший – ранг вершини, другий – її порядковий номер серед множини вершин однакового рангу. На графові рис. 12.2 маємо одну вершину "11" (V_{11}) першого рангу.

II. Умовно викреслимо всі дуги, які виходять з вершин 1-го рангу. Вершини, в які не заходять інші дуги, окрім уже позначених, назвемо вершинами 2-го рангу та пронумеруємо їх у довільній послідовності, зберігаючи неперервність в нумерації стосовно раніше використаних чисел натурального ряду. Для графу рис. 12.2 це вершини "22" (V_{22}) та "23" (V_{23}).

Припустимо, пройдений n–1 етап і визначено вершини (n–1)-го рангу. Викреслимо всі дуги, які виходять з вершин (n–1)-го рангу. Розглянемо всі вершини, в яких закінчуються викреслені на цьому етапі дуги. Вершини, в які заходять лише дуги, визначені на (n–1)-му етапі, утворюють множину вершин n-го рангу. Пронумеруємо їх у довільному порядку, використовуючи неперервно числа натурального ряду, починаючи з найменшого, яке не було використане для нумерації вершин (n–1)-го рангу. Алгоритм завершується по досягненні кінцевої вершини. Для нашого прикладу (рис. 12.2) це вершина "56" (V_{56}), де індекс 5 означає ранг кінцевої вершини, а індекс 6 – її порядковий номер. Виконавши правильно нумерацію вершин графу, в подальшому індекс рангу вершини можна не вказувати.

Алгоритм правильної нумерації вершин графу може бути представлений по-іншому*.

12.4. Алгоритм пошуку найкоротшого шляху мережі (графу)

Виконавши правильну нумерацію графу мережі, можна ефективно реалізувати алгоритм пошуку найкоротшого шляху між заданими вершинами. Реалізацію алгоритму приведено до побудови послідовності шляхів з однієї дуги, двох дуг, трьох дуг і т. д., які з'єднують вершини певного рангу з кінцевою вершиною. При цьому перегляд вершин виконується в послідовності зменшення їх номерів. На кожному етапі алгоритму відбувається перехід від вершини більш високого рангу до вершини меншого рангу за умови, що із множини всіх шляхів, які починаються в одній і тій же вершині, необхідно залишити для наступного кроку реалізації алгоритму лише найкоротший шлях від цієї вершини до завершальної. За правильної нумерації вершин такий упорядкований перехід у послідовності зменшення рангів відбувається автоматично, навіть за умови, коли нумерацію рангів вершин опущено.

Розглянемо реалізацію алгоритму на прикладі графу, зображеного на рис. 12.2. Згідно з алгоритмом будемо рухатися від кінцевої вершини V_{56} до початкової V_{00}, крокуючи послідовно від вершин більш високого рангу до вершин меншого рангу. У кружках, які зображають вершини графу, будемо записувати найкоротшу віддаль від вершини (n–1)-го рангу до кінцевої вершини. Одночасно найкоротший шлях від кожної вершини до кінцевої будемо відмічати подвійною стрілкою. У кружку останньої кінцевої вершини V_{56} записуємо "0", бо звідси виконуємо відлік віддалі. Потім переходимо до вершини 4-го рангу – в прикладі це вершина V_{45}. Від цієї вершини в кінцеву маємо лише один шлях: (V_{45}, V_{56}), його довжина дорівнює двом одиницям виміру, її і запишемо у вершині V_{45}, а відповідну дугу позначимо на графові подвійною стрілкою. Переходимо до вершини 3-го рангу V_{34}. З цієї вершини в кінцеву маємо два шляхи: (V_{34}, V_{56}) та (V_{34}, V_{45}, V_{56}). Найкоротшим є останній. Його одержуємо, додаючи дугу (V_{34}, V_{45}) до вже побудованого шляху (V_{45}, V_{56}) та відмічаючи шлях (V_{34}, V_{45}) подвійною стрілкою.

Довжина шляху (V_{34}, V_{45}, V_{56}) дорівнює 1+2=3, записуємо це число у вершині V_{34}. Шлях (V_{34}, V_{56}) при наступних пошуках не використовуємо, тому що з вершини 3-го рангу вже знайдено найкоротший шлях до кінцевої вершини.

Переходимо до вершин 2-го рангу: V_{22}, V_{23}. Розглянемо шляхи, якими можна перейти від цих вершин до вершин вищого рангу. З вершини V_{22} до вершини вищого

* Кузнецов А. В., Холод Н. И. Математическое программирование: Учеб. пособие. – Минск: Вышейша шк., 1984.

рангу можна потрапити лише по дузі (V_{22}, V_{34}). Її відмічаємо подвійною лінією та записуємо в V_{22} число 3+3=6 – найкоротша віддаль від V_{22} до V_{56}, враховуючи, що найкоротша віддаль від V_{34} до V_{56} уже визначена.

З вершини V_{23} можна потрапити в вершини вищого рангу або по дузі (V_{23}, V_{34}), або по дузі (V_{23}, V_{45}). Обчислюємо шлях від V_{23} до V_{56} за умов руху із V_{23} по названим дугам. Шлях (V_{23}, V_{45}, V_{56}) має довжину 2+2=4, шлях (V_{23}, V_{34}, V_{56}) 2+3=5. Порівнявши названі величини, приходимо до висновку, що найкоротшим шляхом від V_{23} до V_{56} буде шлях (V_{23}, V_{45}, V_{56}), отже дугу (V_{23}, V_{45}) позначаємо подвійною лінією, а в вершині V_{23} ставимо "4".

Розглянувши всі вершини 2-го рангу, переходимо до вершин 1-го рангу. У наведеному прикладі це лише V_{11}, з якої виходять дуги (V_{11}, V_{22}), (V_{11}, V_{23}) та (V_{11}, V_{45}). Розглядаємо й оцінюємо шляхи, які починаються цими дугами та закінчуються в V_{56}: (V_{11}, V_{45}, V_{56}), (V_{11}, V_{23}, V_{45}, V_{56}) та (V_{11}, V_{22}, V_{34}, V_{45}, V_{56}). Решту шляхів, що починаються цими дугами й закінчуються в V_{56}, не розглядаємо, бо вони мають дуги, які не позначені подвійною стрілкою (шляхи з непозначеними подвійною стрілкою дугами не можуть бути найкоротшими до кінцевої вершини). Обчислюємо довжину кожного з трьох названих шляхів та вибираємо найкоротший: (V_{11}, V_{45}, V_{56}). Довжину цього шляху 4+2=6 записуємо в V_{11}, позначивши подвійною лінією дугу (V_{11}, V_{45}). Отже, найкоротші шляхи як від V_{11}, так і від V_{22} до V_{56} мають однакову довжину – шість одиниць.

Перейдемо до розгляду вершини V_{00}. З неї виходять дві дуги – (V_{00}, V_{11}) та (V_{00}, V_{22}), користуючись якими, можна дістатися до вершин вищого, ніж нульовий, рангу. З двох названих дуг меншу довжину має дуга (V_{00}, V_{22}). Тому найкоротший шлях від початкової вершини V_{00} до кінцевої V_{56} буде (V_{00}, V_{22}, V_{34}, V_{45}, V_{56}), його довжина – сім одиниць. Складові дуги шляху всі позначені подвійною лінією.

Якщо завантаження графу рис. 12.2 тлумачити не як відстані, а як тарифи перевезень вантажу, то ми знайшли шлях найменшої вартості перевезення з вершини V_{00} до V_{56}.

Зауваження 1. Використання описаного алгоритму дає можливість визначити найкоротший шлях та найменшу відстань від довільної вершини графу до кінцевої: найменша відстань записується в кружках, які зображають вершини, а найкоротший шлях відмічається на графові подвійною лінією. Наприклад (рис. 12. 2): найкоротший шлях з вершини V_{11} до V_{56} має довжину шість одиниць і проходить через вершини (V_{11}, V_{45}, V_{56}), з вершини V_{22} – теж шість одиниць і проходить через вершини (V_{22}, V_{34}, V_{45}, V_{56}).

Зауваження 2. Описаний алгоритм побудований на так званому "принципі оптимальності": якщо найкоротший шлях від початкової вершини до кінцевої проходить через деяку вершину V_i, то відрізок цього шляху від вершини V_i до кінцевої вершини є найкоротшим серед усіх шляхів, які з'єднують вершину V_i та кінцеву. Так, у розглянутому прикладі найкоротший шлях від V_{00} до V_{56} (V_{00}, V_{22}, V_{34}, V_{45}, V_{56}) як свої складові має найкоротші шляхи до кінцевої вершини від усіх вершин, через які він проходить.

12.5. Побудова графу планування та управління мережею (ПУМ)

Розглянемо методи планування та управління при реалізації проектів створення складних систем, якими можуть бути як виробничі, так і науково-дослідні системи. Щоб скласти план виконання робіт за такими проектами, необхідно зобразити його деякою математичною моделлю, яка називається моделлю мережі і є відображенням певних послідовностей виконання робіт і взаємозв'язків між ними з урахуванням необхідних матеріальних ресурсів. Отже, у цьому розділі під *моделлю мережі* будемо розуміти план виконання комплексу взаємопов'язаних робіт, представлений у специфічній формі графу, який називається *графіком мережі*.

Основними елементами графіка мережі є поняття події та роботи.

Термін *робота* використовується в системі планування та управління мереж (ПУМ) у широкому розумінні. По-перше, це певна реальна робота, яка потребує затрат матеріальних ресурсів та відповідного терміну виконання. Кожна така робота має бути чітко визначеною, конкретною та стосуватися відповідального виконавця, без наявності якого марно говорити про планування.

По-друге, до поняття роботи відносять чекання – процес у часі, який не потребує ніяких матеріальних затрат (наприклад, затвердіння бетону після виконання відповідних робіт, висихання фарби тощо).

По-третє, *фіктивна робота* – це природний логічний взаємозв'язок між двома або кількома роботами чи їх завершенням, який не потребує затрат праці, матеріальних ресурсів або часу. Такий взаємозв'язок показує, що можливість виконання однієї роботи безпосередньо залежить від результатів іншої. Термін виконання фіктивної роботи приймають рівним нулю.

Подія – це фіксація моменту завершення певного етапу виконання проекту. Подія може бути як результатом однієї роботи, так і підсумковим результатом декількох робіт. Подія може відбутися лише тоді, коли будуть виконані всі роботи, які передують події. Наступні роботи можуть розпочатися лише за умови, що відповідна подія відбулася. Отже, для всіх безпосередньо попередніх робіт подія фіксує момент їх закінчення, а для безпосередньо наступних – початок.

Серед подій ПУМ виділяють *вихідну та завершальні події*. Вихідна подія не має попередніх робіт та подій стосовно досліджуваного в моделі комплексу робіт. Завершальна подія не може мати наступних робіт і подій. Події, які визначають термін виконання певної роботи, назвемо *початковою та кінцевою*.

Подію на графові ПУМ зображають кружками (вершини графу), а роботи – орієнтованими дугами, які показують, які роботи необхідно виконати, щоб відбулася

певна подія, та які роботи можна виконувати, якщо подія відбувалася. Отже, будь-яка робота на графові ПУМ позначається двома подіями, між якими вона знаходиться. Подія ж може належати кільком роботам, які можна розпочинати, якщо відповідна подія відбулася. Фіктивні роботи на графові ПУМ позначаються штриховою лінією без зазначення часу. На рис. 12.3 наведено граф, який узагальнено описує комплекс будівельних робіт при спорудженні виробничого корпусу.

І на початку, і по завершенні реальної роботи повинна бути *лише одна подія*, бо роботи ідентифікуються за номерами подій, між якими виконуються, тому на графові рис.12.3 введено фіктивні роботи (3; 4) та (5; 4).

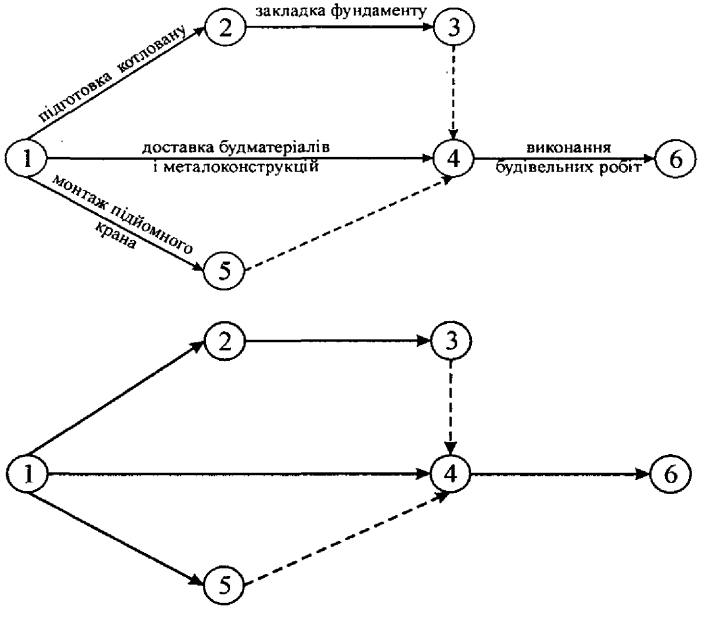


Рис. 12.3.

Використовуються два принципи побудови графів ПУМ. При реалізації одного з них роботи зображаються дугами, а вершини відповідають (співставляються) подіям, які означають завершення робіт. Це так званий граф "події – роботи". Але використовується й інший підхід до побудови графів ПУМ – без подій. За таким принципом роботи є вершинами графу, а дуги означають залежність між певними роботами в послідовності їх виконання. Побудовані за

таким принципом графи ПУМ, як правило, більш місткі та менш зручні для аналізу та реалізації в ЕОМ. З вищезазначених причин на практиці ширше використовується побудова графу ПУМ за принципом "події – роботи". Саме цей принцип будемо використовувати в подальшому.

Порядок та правила побудови графів ПУМ. План реалізації проекту виконання комплексу робіт починається з побудови відповідного графу. У планованому процесі виділяються роботи, які відповідають певним етапам, складається реєстр робіт і подій, аналізується послідовність виконання робіт і взаємозв'язки між ними, за роботами закріплюються виконавці. Про визначення термінів виконання робіт буде сказано нижче. Виходячи з розподілу процесу реалізації проекту на роботи та терміни їх виконання, будують відповідний граф ПУМ. Потім виконують аналіз побудованого графу та його оптимізацію за обраними критеріями.

При побудові графу ПУМ необхідно дотримуватися певних правил, щоб у подальшому можна було досліджувати його. Графічне зображення вимог щодо побудови графів ПУМ наведено на рис. 12.4 (позиції а–и).

1. Граф ПУМ не повинен мати "глухих кутів", тобто подій, з яких не виходить жодної роботи, окрім завершальної події (рис. 12.4 а). Поява "глухих кутів" подій свідчить про не досить ретельно виконаний аналіз робіт та їх взаємозв'язків.

2. На графові не може бути "хвостових" подій (окрім вихідної події), тобто подій, яким не передує жодна робота. На рис. 12.4 б такою "хвостовою" подією є подія 3; вона не може відбутися, отже, не можуть відбутися і наступні події.

3. Граф не може мати замкнутих контурів і петель, тобто шляхів, які з'єднують певні події з ними ж (рис. 12.4 в г). Поява замкнутих контурів вимагає перегляду складу робіт та їх взаємозв'язків, після змістовного аналізу яких завжди з'являється можливість уникнути замкнутих контурів і петель.

4. Дві довільні події повинні бути безпосередньо пов'язаними не більше ніж однією дугою – роботою. Ця вимога обумовлена тим, що роботи позначають двома індексами (i, j), які відповідають подіям "i" та "j". Якщо в дійсності треба виконати кілька робіт, які розпочинаються та завершуються одночасно, при одних і тих же подіях початкових та кінцевих, то в таких випадках необхідно ввести фіктивні події та фіктивні роботи (рис. 12. 4 д), паралельні роботи при цьому замикаються на фіктивні події (рис. 12.4 е).

5. На графові ПУМ повинна бути лише одна вихідна та лише одна завершальна подія. Якщо це об'єктивно не так (початок реалізації комплексу робіт можна розпочинати паралельно з декількох робіт - рис. 12. 4 є), то необхідно ввести фіктивні події та роботи, як це показано на рис. 12. 4 ж.

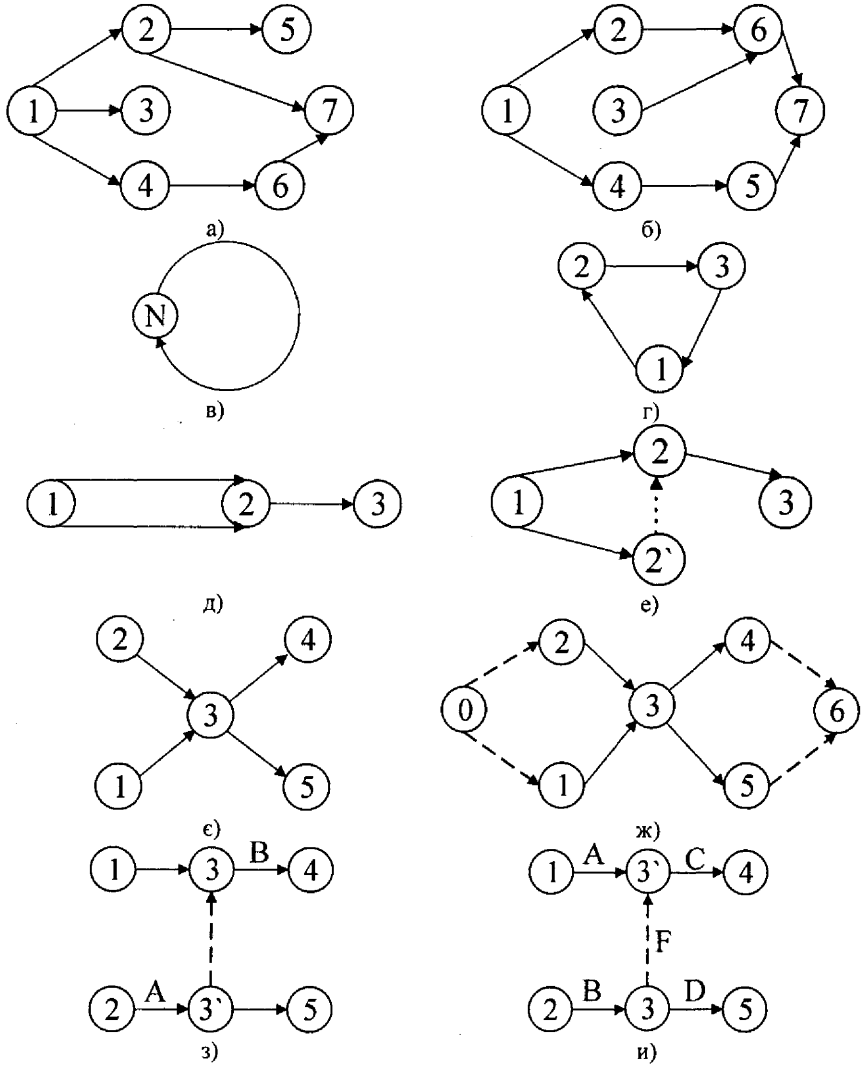


Рис. 12.4. Графічне зображення вимог до побудови графів ПУМ.

Фіктивні події та роботи можуть вводитися і через інші умови. Однією з таких умов є відображення залежності подій, не пов'язаних з реальними роботами. Наприклад, роботи *A* та *B* (рис. 12.4 з) можуть виконуватися технологічно

незалежно одна від одної, але в умовах конкретного виробництва робота *B* не може розпочатися раніше, ніж завершиться робота *A*. За таких обставин необхідно ввести фіктивну роботу *C* (рис. 12.4 и). Другий випадок – неповна залежність робіт. Наприклад, робота *C* може бути розпочата лише по завершенні робіт *A* та *B*, але робота *D* – лише по завершенні роботи *B* і не залежить від роботи *A*. Тоді необхідно ввести фіктивну роботу *F* та фіктивну подію 3′, як показано на рис. 12.4 и.

Фіктивні роботи можуть бути введені і для відображення реальних відстрочок і чекань – в цих випадках вони характеризуються відповідним часом.

12.6. Упорядкування графу ПУМ, обчислення основних параметрів

Побудувавши за всіма вимогами граф ПУМ, необхідно зробити правильну нумерацію його вершин за алгоритмом, описаним у підрозділі 12.3. При такому упорядкуванні графу всі роботи – дуги будуть орієнтовані зліва направо (якщо вихідна подія буде зліва): від подій з меншими номерами до подій з більшими номерами. Після виконання правильної нумерації індекс рангу вершини можна опустити.

Зауваження. Оскільки при виконанні правильної нумерації вершини одного рангу нумеруються в довільному порядку, то граф для однієї і тієї ж мережі може мати різну нумерацію стосовно його вершин та одних і тих же робіт.

Якщо досліджуваний комплекс робіт складний, то його граф будується частинами, які потім "зшиваються". За таких умов аналіз графу ПУМ виконується на ЕОМ.

Термін виконання робіт оцінюють, як правило, висококваліфіковані, досвідчені фахівці. У залежності від характеру та умов виробництва оцінки визначаються одним з наступних методів або їх об'єднанням:

- з урахуванням продуктивності праці за умови, що такі роботи виконувалися раніше за таких же або близьких обставин;
- за діючими нормами, з урахуванням обсягу роботи, продуктивності праці одного виконавця, кількості працюючих у зміну, кількості змін тощо;
- методом експертних оцінок, з використанням оцінок, визначених кожним експертом. Наприклад, термін виконання роботи обчислюється як середня арифметична термінів, названих кожним експертом;
- оскільки ПУМ використовується в основному при плануванні складних робіт, які, як правило, не мають аналогів у минулому, то терміни виконання робіт розглядаються як випадкові величини з деякими законами розподілу та відповідними числовими характеристиками: математичним сподіванням та дисперсією. Використовуючи досвід та статистичні дані, практично завжди в

дослідженнях ПУМ виходять з того, що розподіл термінів виконання робіт має такі властивості: а) неперервність; б) унімодальність, тобто єдиний максимум у функції щільності розподілу; в) додатну асиметрію, тобто максимум кривої розподілу, зміщений вліво відносно медіани (лінії, що ділить площу під функцією щільності розподілу на дві рівні за площею частини). Графік функції щільності розподілу, як правило, більш круто підіймається при віддаленні від мінімального значення терміну виконання певної роботи та менш круто спадає при наближенні до максимального значення терміну виконання роботи (рис. 12.5). Найпростішим розподілом з такими загальними властивостями є так званий в математичній статистиці β-розподіл.

Для визначення математичного сподівання терміну виконання певної роботи від виконавців та експертів одержують таку інформацію:

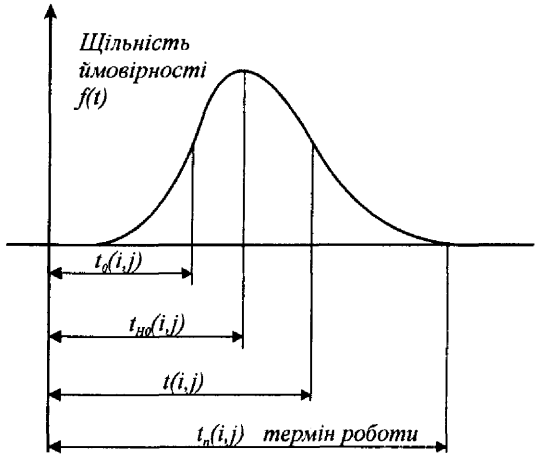


Рис. 12. 5.

оптимістичну оцінку t_0, тобто термін виконання роботи за найсприятливіших умов;

– песимістичну оцінку $t_п$, тобто термін виконання роботи за найнесприятливіших умов;

– найбільш ймовірну оцінку $t_{но}$ терміну виконання роботи.

Виходячи з гіпотези про β-розподіл величини терміну виконання певної роботи та використовуючи вказану інформацію, отриману від фахівців, математичне сподівання $\bar{t}(i, j)$ терміну виконання кожної роботи (i, j) обчислюють так:

$$\bar{t}(i, j) = \frac{t_0(i, j) + 4t_{но}(i, j) + t_п(i, j)}{6}.$$

Зважаючи на труднощі одержання оцінки $t_{но}$, на практиці використовують більш просту, хоч і менш точну, формулу:

$$\bar{t}(i, j) = \frac{2t_0(i, j) + 3t_п(i, j)}{5}.$$

Отже, для досліджень графу ПУМ необхідно визначити термін виконання кожної роботи. Завантажений граф ПУМ буде мати або детерміновані, або стохастичні величини термінів виконання робіт. Для подальшого аналізу приймемо, що завантажений граф ПУМ має детерміновані оцінки термінів виконання робіт. У літературі розроблено методи аналізу графів за допомогою стохастичних оцінок.

Параметри планування й управління мережі за критерієм часу. Одним з визначальних основних понять графу ПУМ є поняття шляху. *Шлях* – це будь-яка послідовність робіт, якщо кінцева подія кожної роботи є початковою подією наступної роботи. Таке визначення шляху в ПУМ має більш реальний сенс, ніж загальне в теорії графів (п. 12.2). Серед шляхів графу ПУМ виділяють підмножини завершених шляхів. *Завершений шлях* – це будь-який шлях, початком якого є вихідна подія, а закінченням – завершальна.

Наприклад, маємо завантажений граф ПУМ з правильно введеною нумерацією. Термін шляху дорівнює сумі всіх термінів виконання робіт, які створюють шлях. Завершений шлях з найбільшим терміном серед усіх завершених шляхів називається *критичним шляхом*. Роботи, які його створюють, називаються критичними. Граф певної ПУМ може мати не один критичний шлях.

Максимальним шляхом між двома подіями "i" та "j" називають шлях від i-ої події до j-ої, який має максимальний термін, тобто сума термінів робіт, які складають такий шлях, є не меншою ніж відповідна сума для довільного шляху від i-ої події до j-ої.

Наприклад, для графу рис. 12. 6 завершеними шляхами будуть L_1=[2; 6; 7], L_2=[4; 5; 7], L_3=[3; 6; 7] і т. д. Їх терміни $t(L_1)$=3+5+16=24; $t(L_2)$=8+7+18=33; $t(L_3)$=4+10+16=30. Критичний шлях слід визначати за алгоритмом, описаним у підрозділі 12. 2, використовуючи завантаження дуги не пропускною спроможністю, а терміном виконання роботи, що відповідає дузі. Далі буде наведено інший спосіб обчислення терміну критичного шляху. Критичний шлях для графу рис. 12.6 $L_{кр}$=[1; 2; 4; 5; 7], а його термін $t(L_{кр})$=3+9+7+18=37. На рис. 12.6 він зображений подвійною лінією. Критичними роботами будуть: (1; 2), (2; 4), (4; 5), (5; 7).

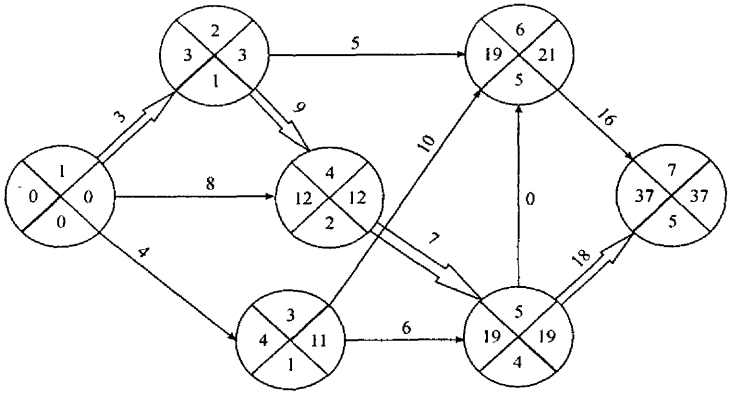


Рис. 12.6.

Критичний шлях має особливе значення для ПУМ, бо роботи, які складають цей шлях, визначають загальний термін завершення всіх робіт комплексу, які плануються в даній системі ПУМ.

Основні параметри ПУМ за критерієм часу наведені в табл. 12.1.

Таблиця 12.1.

Основні параметри планування й управління мережі

Елемент мережі, характеризований параметром	Назва параметра	Умовні позначення параметра
Подія i	Ранній термін звершення події	$t_p(i)$
	Пізній термін звершення події	$t_n(i)$
	Резерв часу події	$R_p(i)$
Робота (i, j)	Термін виконання роботи	$t\,(i, j)$
	Ранній термін початку роботи	$t_{pn}(i, j)$
	Ранній термін завершення роботи	$t_{pз}(i, j)$
	Пізній термін початку роботи	$t_{nn}(i, j)$
	Пізній термін завершення роботи	$t_{nз}(i, j)$
	Повний резерв часу роботи	$R_n(i, j)$
	Частковий резерв першого виду часу роботи	$R_1(i, j)$
	Частковий резерв другого виду часу роботи	$R_2(i, j)$
	Незалежний резерв часу роботи	$R_н(i, j)$
Шлях L	Термін шляху	$t(L)$
	Термін критичного шляху	$t_{кр}$
	Резерв часу шляху	$R_t(L)$

Розглянемо формули обчислення параметрів, указаних в табл. 12.1. Використання формул покажемо стосовно графу ПУМ, схему запису параметрів часу якого наведено на рис. 12.6.

Результати обчислення параметрів записують у таблиці на зразок табл. 12.2. А якщо граф має невелику кількість подій (до двох десятків), то результати можна записувати на графові за схемою, зображеною на рис. 12.7.

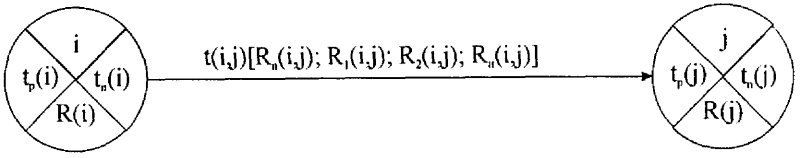


Рис. 12.7.

Розглянемо параметри подій. Певна подія j не може відбуватися раніше, ніж завершаться всі роботи, які їй передують. Отже, ранній термін $t_p(j)$ можливого звершення j-ої події визначається терміном максимального шляху:

$$t_p(j) = \max t(L_{bj}) \tag{1}$$

(L_{bj} – будь-який шлях, який передує j-й подій, тобто шлях від вихідної до j-ої події).

Якщо подія j має кілька шляхів, які їй передують, отже, кілька попередніх подій i, то ранній термін $t_p(j)$ звершення події j зручно обчислювати за формулою:

$$t_p(j) = \max\,[t_p(i) + t(i, j)]. \tag{2}$$

Формула (2) показує, що обчислення параметру t_p доцільно починати з вихідної події, для якої t_p дорівнює нулю, розглядаючи наступні події в порядку збільшення їх номерів.

Затримка зі звершенням i-ої події стосовно свого раннього терміну (строку) не буде впливати на термін звершення завершальної події (отже, і на термін виконання досліджуваного комплексу робіт), поки сума термінів (строку) звершення i-ої події та терміну максимального з усіх шляхів, які ідуть від i-ої події до завершальної, не перевищить терміну критичного шляху. Таким чином, пізній (або граничний) термін (строк) $t_n(i)$ звершення i-ої події обчислюється за формулою:

$$t_n(i) = t_{кр} - \max_{L_{iз}} t(L_{iз}) \tag{3}$$

($L_{iз}$ – будь-який шлях L від i-ої події до завершальної).

Якщо подія "i" має кілька наступних шляхів, пов'язана з кількома наступними подіями "j", то пізній строк звершення події "i" зручно обчислити за формулою:

$$t_n(i) = \min_{(i,j)}[t_n(j) - t(i, j)]. \quad (4)$$

Резерв часу $R(i)$ i-ої події обчислюється як різниця пізнього та раннього термінів звершення i-ої події:

$$R(i) = t_n(i) - t_p(i). \quad (5)$$

Резерв часу події показує, на який допустимий термін можна затримати звершення події, не гальмуючи (не збільшуючи) при цьому термін виконання всього комплексу робіт мережі.

Критичні події не мають резервів часу, бо будь-яка затримка зі звершенням подій, які розташовані на критичному шляху, викличе таку ж затримку у виконанні завершальної події.

Отже, щоб визначити термін критичного шляху, необхідно і достатньо обчислити ранній строк завершальної події, його величина і визначає цей термін.

Події з нульовим резервом часу визначають роботи, які складають критичний шлях.

Приклад. Обчислити параметри часу подій та критичний шлях графу рис. 12.6. Результати наведені в табл. 12.2.

Таблиця 12. 2.

Результати розрахунків параметрів подій графу (рис. 12.6)

Номер події	t_p – ранній термін звершення події	t_n – пізній термін звершення події	R_p – резерв часу події
1	0	0	0
2	0+3=3	$t_n = t_p$(критична подія)	0
3	0+4=4	min{19–6; 21–10}=11	11–4=7
4	max {0+8; 3+9}=12	$t_n = t_p$(критична подія)	0
5	max {4+6; 12+7}=19	$t_n = t_p$(критична подія)	0
6	max {3+5; 4+10; 19+0 }=19	37–16–21	21–19=2
7	max {19+16; 19+18}=37	37	0

Для обчислення ранніх термінів $t_p(i)$ звершення подій вершини графу розглядаємо в порядку зростання їх номерів, які записані у верхній чверті кружка вершини.

Для вихідної вершини i=1 очевидно $t_p(i)$=0. Для i=2 маємо $t_p(2)$=3, бо існує

лише одна робота, завершення якої відповідає події 2. З тієї ж причини $t_p(3)$=4. Для обчислення $t_p(4)$ скористаємося формулою (2) та заданими величинами $t(2; 4)$=9 та $t(1; 4)$=8:

$t_p(4)$=max{$t_p(1)+t(1; 4)$; $t_p(2)+t(2; 4)$}=max{0+8; 3+9}=12.

Отже, $t_p(4)$=12. Користуючись формулою (2), маємо:

$t_p(5)$=max{$t_p(3)+t(3; 5)$; $t_p(4)+t(4; 5)$}=max{4+6; 12+7}=19.

$t_p(6)$=max{$t_p(2)+t(2; 6)$; $t_p(5)+t(5; 6)$; $t_p(3)+t(3; 6)$}=max{3+5; 19+0; 4+10}=19.

$t_p(7)$=max{$t_p(5)+t(5;7)$; $t_p(6)+t(6;7)$}=max{19+18; 19+16}=37.

Отже, термін критичного шляху дорівнює 37 одиницям виміру часу.

При обчисленні пізніх термінів здійснення подій $t_n(i)$ розглядаємо вершини графу в послідовності зменшення їх нумерацій. Для завершальної події пізній термін її звершення дорівнює ранньому, тому $t_n(7)$=37. Пізні терміни інших подій обчислюємо, користуючись формулою (4).

$t_n(6)$=min{$t_n(7)-t(6;7)$}=min{37–16}=21, бо для події 6 існує лише один наступний шлях (6; 7).

$t_n(5)=t_n(7)-t(5; 7)$=37–18=19.

$t_n(4)=t_n(5)-t(4; 5)$=19–7=12, бо для події 4 існує лише один наступний шлях, який починається з дуги (4; 5).

$t_n(3)$=min{$t_n(5)-t(3; 5)$; $t_n(6)-t(3; 6)$}=min{19–6; 21–10}=11.

$t_n(2)$=min{$t_n(4)-t(2; 4)$; $t_n(6)-t(2; 6)$}=min{12–9; 21–5}=3.

$t_n(1)$=0.

За формулою (5) обчислюємо резерв часу для кожної події: $R(1)$=0–0=0; $R(2)$=3–3=0; $R(3)$=11–4=7; $R(4)$=12–12=0; $R(5)$=19–19=0; $R(6)$=21–19=2; $R(7)$=37–37=0.

Відмінний від нуля (ненульовий) резерв часу певної події означає, що термін звершення такої події може бути збільшений на величину її резерву часу без затримки терміну виконання всього комплексу робіт.

Аналізуючи результати розрахунків, знаходимо, що події, які належать критичному шляху, не мають резерву часу.

Перейдемо до розрахунків параметрів робіт.

Окрема робота може бути виконана за час відповідно до раннього, пізнього або будь-якого проміжного терміну, який відповідає завершуючій цю роботу події. Очевидно, що ранній термін $t_{pn}(i, j)$ початку роботи (i, j) співпадає з раннім терміном здійснення попередньої події i, тому:

$$t_{pn}(i, j) = t_p(i). \quad (6)$$

Таким чином, ранній термін $t_{pз}(i, j)$ завершення роботи (i, j) обчислюється так:

$$t_{pз}(i, j) = t_p(i) + t(i, j). \quad (7)$$

Для того, щоб фактичний термін виконання певної роботи не збільшив час виконання комплексу робіт, повинна виконуватися вимога: кожну роботу слід завершувати не пізніше допустимого терміну її завершальної події j. За таких умов пізній термін $t_{пз}(i, j)$ завершення роботи (i, j) визначається формулою:

$$t_{пз}(i, j) = t_п(j), \tag{8}$$

а пізній термін $t_{пп}(i, j)$ початку цієї роботи – формулою:

$$t_{пп}(i, j) = t_п(j) - t(i, j). \tag{9}$$

Таким чином, у межах певної системи ПУМ моменти початку та завершення кожної роботи тісно пов'язані з сусідніми подіями співвідношеннями (6) – (9).

Розглянемо резерви часу шляхів. Такі резерви мають усі некритичні шляхи. *Резерв часу* $R_t(L)$ шляху L визначається як різниця термінів критичного та даного шляхів:

$$R_t(L) = t_{кр} - t(L). \tag{10}$$

Величина $R_t(L)$ показує, на скільки можуть бути збільшені в сумі терміни всіх робіт, які складають шлях L. Якщо збільшити терміни виконання робіт шляху L в сумі більше ніж на величину $R_t(L)$, критичним шляхом буде L. Отже, будь-яка з робіт шляху L, якщо вона не є складовою критичного шляху, має певний резерв часу.

Розглянемо чотири різновиди резервів часу робіт.

Повний резерв $R_п(i, j)$ часу роботи (i, j) показує, на скільки можна збільшити термін виконання даної роботи за умови, що термін виконання всього комплексу робіт не зміниться. Повний резерв $R_п(i, j)$ часу роботи (i, j) обчислюється за формулою:

$$R_п(i, j) = t_п(j) - t_р(i) - t(i, j). \tag{11}$$

На рис. 12.8 показано схему розрахунків параметрів часу стосовно робіт.

Повний резерв часу роботи дорівнює резервові часу максимального зі шляхів, які містять цю роботу. Цим резервом можна скористатися при виконанні роботи за умови, що її початкова подія відбудеться найранніший термін, а її завершальна подія – в пізній термін (рис. 12.8 а).

Суттєвою властивістю повного резерву часу роботи є те, що цей резерв стосується не лише цієї роботи, а й усіх повних шляхів, які містять її. Якщо використати повний резерв часу лише для цієї однієї роботи, то резерви часу решти

робіт максимального шляху, який містить її, будуть повністю вичерпані. Резерви часу робіт, які належать іншим, але не максимальним за терміном шляхам, що містять цю роботу, відповідно зменшаться на величину використаного резерву.

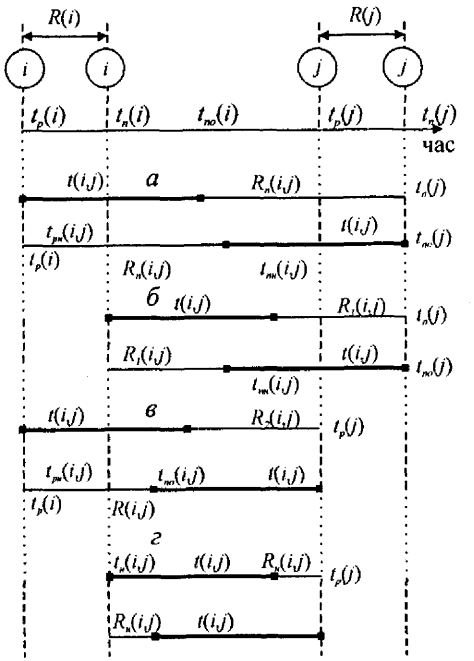


Рис. 12.8.

Решта резервів часу роботи, які розглянемо далі, є частинами його повного резерву.

Частинним резервом першого виду R_1 часу роботи (i, j) є та частина повного резерву часу цієї роботи, на яку можна збільшити термін роботи, не змінивши при цьому пізній термін її початкової події. Цим резервом можна скористатися при виконанні певної роботи за умови, що її початкова та завершальна події відбудуться в свої пізні терміни (рис. 12. 8, б). Величина $R_1(i, j)$ обчислюється за формулою:

$$R_1(i, j) = t_п(j) - t_п(i) - t(i, j) \tag{12}$$

або

$$R_1(i, j) = R_п(i, j) - R(i). \tag{13}$$

Частинним резервом другого виду R_2 часу роботи (i, j) є та частина повного резерву часу цієї роботи, на яку можна збільшити термін її виконання, не змінивши при цьому ранній термін її початкової події.

Цим резервом можна користуватися при виконанні певної роботи за умови, що початкова та завершальна події відбудуться в свої ранні терміни (рис. 12.8 в). Величина $R_2(i, j)$ обчислюється за формулою:

$$R_2(i, j) = t_p(j) - t_p(i) - t(i, j) \tag{14}$$

або

$$R_2(i, j) = R_n(i, j) - R(j). \tag{15}$$

Резервом часу R_2 можна користуватися при запобіганні випадковостям, які можуть трапитися при виконанні відповідної роботи. За умови планування виконання робіт у ранні терміни їх початку та завершення завжди маємо можливість при необхідності перейти на пізні терміни початку та завершення робіт, тобто величина $R_2(i, j)$ показує, на скільки можна перенести початок або збільшити термін виконання роботи.

Незалежним резервом часу $R_н$ роботи (i, j) є та частина повного резерву часу цієї роботи, яка визначається за умови, що попередні роботи завершуються в пізні терміни, а всі наступні розпочинаються в ранні терміни (рис. 12.8 г).

Величина $R_н$ обчислюється за формулою:

$$R_н(i, j) = t_p(j) - t_n(i) - t(i, j) \tag{16}$$

або

$$R_н(i, j) = R_2(i, j) - R(i) = R_1(i, j) - R(j). \tag{17}$$

Використання незалежного резерву часу не впливає на величини резервів часу інших робіт. Незалежні резерви часу намагаються використати тоді, коли завершення попередньої роботи відбулося в пізній термін, а наступні роботи бажано виконати в ранній термін. Таку можливість маємо, якщо величина незалежного резерву часу $R_н(i, j)$, обчислена за формулою (16) або (17), дорівнює нулю або додатна. За умови, що величина $R_н(i, j)$ від'ємна, немає можливості виконати наступні роботи в ранні терміни, бо попередня робота ще не завершена, тому від'ємні значення $R_н(i, j)$ не мають реального сенсу. Незалежний резерв часу має сенс лише для тих робіт, які не належать максимальним шляхам, що проходять через їх початкові та завершальні події.

Висновки. Частковий резерв першого виду часу роботи може бути використаний на збільшення терміну виконання даної та наступних робіт. Частковий резерв другого виду часу роботи може бути використаний на збільшення

терміну виконання даної та попередніх робіт без зміни (порушення) резерву часу наступних робіт. Незалежний резерв часу може бути використаний для збільшення терміну виконання лише даної роботи.

Роботи, які складають критичний шлях, та критичні події будь-яких резервів часу не мають. Якщо початкова подія i роботи (i, j) належить критичному шляху, то повний резерв часу

$$R_n(i, j) = R_1(i, j). \tag{18}$$

Якщо завершальна подія j роботи (i, j) належить критичному шляху, то повний резерв часу

$$R_n(i, j) = R_2(i, j). \tag{19}$$

Якщо початкова подія i та завершальна подія j належать критичному шляху, але сама робота не належить йому, то повний резерв часу визначається так:

$$R_n(i, j) = R_1(i, j) = R_2(i, j) = R_н(i, j). \tag{20}$$

Наведені співвідношення (18) – (20) доцільно використовувати для контролю вірності (правильності) розрахунків резервів термінів виконання окремих робіт.

Обчислимо часові параметри робіт графу ПУМ рис. 12.6.

Результати розрахунків представлені в табл. 12.3; для кращої орієнтації наведені проміжні викладки.

Таблиця 12.3.

Таблиця розрахунків термінів графу (рис. 12.6)

Роботи	Терміни робіт	R_n (роботи) формула (11)	R_1 (роботи) формула (12)	R_2 (роботи) формула (14)	$R_н$ (роботи) формула (16)
1; 2	3	3–0–3=0	3–0–3=0	3–0–3=0	3–0–3=0
1; 3	4	11–0–4=7	11–0–4=7	4–0–4=0	4–0–4=0
1; 4	8	12–0–8=4	12–0–8=4	12–0–8=4	12–0–8=4
2; 4	9	12–3–9=0	12–3–9=0	12–3–9=0	12–3–9=0
2; 6	5	21–3–5=13	21–3–5=13	19–3–5=11	19–3–5=11
3; 5	6	19–4–6=9	19–11–6=2	19–4–6=9	19–11–6=2
3; 6	10	21–4–10=7	21–11–10=0	19–4–10=5	19–11–10= –2
4; 5	7	19–12–7=0	19–12–7=0	19–12–7=0	19–12–7=0
5; 6	0	21–19–0=2	21–19–0=2	19–19–0=0	19–19–0=0
5; 7	18	37–19–18=0	37–19–18=0	37–19–18=0	37–19–18=0
6; 7	16	37–19–16=2	37–21–16=0	37–19–16=2	37–21–16=0

Величина $R_n(3;\ 6) = -2$ свідчить про те, що наступну роботу (6; 7) не можна виконати в ранній термін. Аналіз розрахунків показує, що виконання робіт (1; 4), (2; 6), (3; 5), (3; 6), (6; 7) та наступних за будь-якими шляхами може бути затримано на відповідну кількість часу без витрат термінів виконання попередніх робіт, бо величина R_2 для вказаних робіт більша за нуль.

12.7. Аналіз та оптимізація планування й управління мережею

Аналіз графу мережі дозволяє оцінити доцільність обраної структури графу, тобто класифікацію робіт та їх послідовність, завантаження виконавців робіт, можливості зміщень в часі виконання робіт некритичних шляхів. Якщо оцінки термінів виконання робіт мають ймовірний характер, то аналіз ПУМ дає можливість оцінити ймовірність виконання повного комплексу робіт з реалізації плану в заданий термін.

На першому етапі аналізу графу досліджується його *топологія* (взаємне узгодження елементів графу відповідно до вимог його побудови) та оцінюється доцільність вибору послідовності робіт і структури графу. На цьому етапі перевіряються доцільність виконаного рівня деталізації робіт та оцінка наявних виробничих можливостей.

На другому етапі виконується класифікація робіт і групуються роботи за величинами резервів термінів їх виконання, а також визначається ступінь напруженості своєчасного виконання робіт некритичних шляхів.

Ступінь напруженості терміну виконання роботи (i, j) прийнято характеризувати коефіцієнтом напруженості $K(i, j)$, величина якого визначається відношенням суми термінів виконання неспівпадаючих робіт максимального шляху, якому належить дана робота, та критичного шляху:

$$K(i, j) = \frac{t(L_{\max}) - t'_{кр}}{t_{кр} - t'_{кр}} \qquad (21)$$

($t(L_{\max})$ – термін виконання робіт максимального шляху, який містить роботу (i, j), $t_{кр}$ – термін виконання робіт критичного шляху;

$t'_{кр}$ – сума термінів виконання тих робіт максимального шляху, відповідного роботі (i, j), які співпадають з роботами критичного шляху).

Формулу (21) можна записати і в такому вигляді:

$$K(i, j) = 1 - \frac{R_n(i, j)}{t_{кр} - t'_{кр}}, \qquad (22)$$

($R_n(i, j)$ – повний резерв часу роботи (i, j)).

Коефіцієнт напруженості $K(i, j)$ приймає значення від 0 до 1. Чим більше його величина, тим важче своєчасно виконати дану роботу.

Третім етапом аналізу ПУМ є розрахунки запитів виробничих ресурсів та їх розподіл у часі.

На четвертому етапі визначають ймовірність своєчасного виконання планового комплексу робіт з реалізації проекту. Ці розрахунки вельми складні. Як правило, з певною мірою вірогідності вони визначають найменший та найбільший можливі терміни виконання комплексу робіт. Виділяють також роботи, які з високим рівнем надійності можуть попасти на критичний шлях. Саме цим роботам приділяється особлива увага при розподілі виробничих ресурсів і в процесі реалізації робіт.

Усі вказані етапи аналізу системи ПУМ передують її оптимізації.

Приклад. Обчислимо коефіцієнт напруженості $K(i, j)$ для деяких робіт графу рис. 12. 6. Для робіт критичного шляху (1; 2), (2; 4), (4; 5), (5; 7) величина $K(i, j)$ дорівнює 1, бо повний резерв часу $R_n(i, j)$ для названих робіт (22) дорівнює нулю (див. табл. 12. 3):

$$K(1;4) = 1 - \frac{4}{37-25} = 1 - \frac{4}{12} = \frac{2}{3};$$

$$K(1;3) = 1 - \frac{7}{37-0} = \frac{30}{37},$$

бо максимальний шлях, який містить роботу (1; 3), має роботи: (1; 3), (3; 6), (6; 7) і не має спільних робіт з критичним шляхом. Використавши відношення (21), маємо ту ж величину:

$$K(1;3) = \frac{4+10+16}{37} = \frac{30}{37}.$$

Оптимізація планування й правління мережею. Маючи граф ПУМ, виконують ретельний аналіз відображених у ньому зв'язків і послідовності виконання запланованих робіт з метою поліпшення якості необхідного комплексу робіт за умови своєчасної його реалізації та ефективного використання необхідних ресурсів. Оптимізація ПУМ виконується за рахунок перерозподілу наявних

виробничих ресурсів (виконавців робіт, матеріальних, енергетичних, фінансових ресурсів і т. д.), оцінки можливостей інтенсифікації робіт критичного шляху (залучення додаткових виконавців, фінансів, паралельне виконання робіт, якщо це доцільно, і т. д.), поліпшення топології графу ПУМ завдяки іншій організації виконання певних робіт при використанні іншого матеріально-технічного їх забезпечення.

Виконуючи оптимізацію ПУМ, необхідно враховувати вимоги, які пред'являються до системи: а) термін виконання планованого комплексу робіт; б) по можливості рівномірне в часі використання всіх видів ресурсів; в) зменшення витрат у поєднанні з ефективністю використання ресурсів.

Математичний апарат оптимізації ПУМ одночасно за кількома параметрами недостатньо розроблений, тому оптимізацію виконують почергово за кожним параметром. Спочатку виконується оптимізація за часом без урахування обмежень на ресурси та виробничі можливості. Потім оптимізують ПУМ за ресурсами і, нарешті, за витратами.

Поліпшення ПУМ за критерієм часу без урахування ресурсів. Порівняння розрахункових термінів виконання робіт з плановими, аналіз планів матеріально-технічного забезпечення, можливостей використання трудових ресурсів є основою оптимізації ПУМ комплексу робіт з реалізації проекту. Якщо директивний термін не може бути виконаний, то переглядають технологію виконання робіт, узгодженість робіт за часом, розглядають заходи щодо іншого розподілу матеріально-технічних можливостей. Якщо розроблені заходи можуть забезпечити виконання директивного терміну, що означає зменшення терміну критичного шляху, то будують новий граф ПУМ і знову виконують його розрахунки й аналіз.

Оптимізація ПУМ за ресурсами та часом. Мета виконання такої оптимізації – найліпшим чином узгодити планові терміни з виробничими можливостями та забезпечити якомога рівномірніше завантаження виконавців робіт і використання ресурсів. Задача оптимального розподілу всіх необхідних ресурсів у системі ПУМ в загальному випадку не має алгоритмів розв'язання – розроблено лише методи розв'язання оптимізації розподілу якого-небудь одного виду ресурсу.

Оптимізація за часом при обмеженнях на обсяги ресурсів зводиться до розрахунків термінів виконання робіт та обчислення необхідної кількості контрольованого ресурсу для забезпечення календарних періодів. У тих випадках, коли розрахункові обсяги ресурсів перевищують їх наявність, вживають заходи по запобіганню таких ситуацій: переглядають технології та терміни виконання робіт, доцільність зміни обсягів ресурсів, залучення додаткових фінансів і т. ін. Після такого корегування ПУМ виконують повторні розрахунки й аналізують їх результати.

Ознакою поліпшення якості ПУМ вважається більш рівномірне та неперервне використання ресурсів виробництва.

Оптимізація ПУМ за критерієм затрат і часом. На практиці поряд з оцінкою системи ПУМ за часом з необхідністю треба враховувати вартість виконання робіт. Одним із напрямків коригування й оптимізації ПУМ за критерієм вартості є використання наявних повних резервів часу виконання робіт. Такий підхід обумовлений тим, що збільшення термінів виконання робіт у певних межах зменшує вартість відповідних робіт і, таким чином, поліпшує ПУМ за рахунок допустимого збільшення терміну некритичних шляхів без зміни терміну завершення всіх робіт.

Звичайно, що оптимізація ПУМ за будь-якими критеріями не можлива без використання ЕОМ. Розроблені та використовуються різноманітні програми розв'язання задачі оптимізації ПУМ за тими або іншими критеріями чи обмеженнями.

Управління перебігом робіт (реалізації комплексу робіт) з допомогою ПУМ починається після затвердження відповідного плану та ознайомлення з ним усіх виконавців робіт.

Мета управління – контроль за термінами виконання робіт і використанням ресурсів. Забезпечення ефективності ПУМ вимагає віддзеркалення реальної ситуації та усіх змін, які відбуваються в процесі реалізації контрольованого комплексу робіт. Динамічне використання ПУМ може бути гарантоване за умов одержання об'єктивної інформації про стан і хід робіт та можливостей оперативно через виконавців робіт вносити зміни в план реалізації подальших робіт.

Вправи.

1. У табл. 12.4 наведені оцінки термінів виконання робіт з реалізації певного проекту, визначені фахівцями.

Таблиця 12.4.

Приклад оцінки термінів виконання робіт

№п/п	Робота (i, j)	Оцінки термінів виконання роботи (доба)		
		оптимістична	песимістична	Найімовірніша (найбільш вірогідна)
1	(1; 2)	5	9	6
2	(1; 3)	2	7	5
3	(1; 4)	4	10	8
4	(3; 4)	9	14	11
5	(2; 5)	7	13	10
6	(4; 5)	1	4	3

1. Необхідно:

а) побудувати граф ПУМ;

б) визначити очікуваний термін виконання робіт з реалізації проекту;

в) визначити критичний шлях та його термін;

г) обчислити параметри подій і робіт.

2. Для графу, зображеного на рис. 12.9:

а) виконати правильну нумерацію;

б) знайти повні шляхи та обчислити їх терміни;

в) знайти критичний шлях та вказати критичні роботи.

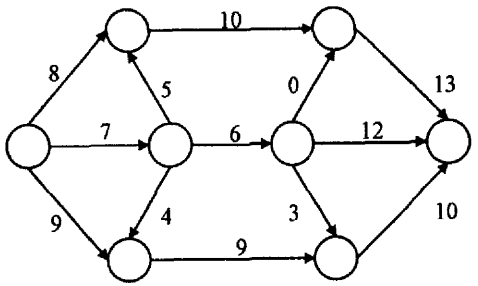


Рис. 12.9.

3. Побудувати граф ПУМ комплексу робіт по заміні колеса автомобіля двома виконавцями та назвати орієнтовні терміни виконання робіт:

1) дістати з багажника домкрат та інструменти; 2) звільнити в багажнику колесо від кріплення; 3) прилаштувати домкрат під машину; 4) підняти машину; 5) зняти гайки та колесо, яке потребує заміни; 6) дістати колесо з багажника; 7) установити на вісь запасне колесо та прилаштувати гайки; 8) опустити машину та скласти домкрат; 9) поставити домкрат у багажник; 10) закріпити колесо на вісі; 11) покласти зняте колесо та інструмент у багажник.

Питання для самоконтролю

1. Що розуміють під терміном "мережа"? Що означає абревіатура ПУМ? Приклади мереж.
2. Що таке граф та орієнтований граф? Приклади.
3. Основні поняття теорії графів: маршрут, ланцюг, цикл, шлях.
4. Що таке завантажений граф? Приклади.
5. Як виконується правильна нумерація графу?
6. Сформулювати задачу пошуку найкоротшого шляху орграфу, описати алгоритм її розв'язання.
7. Особливість побудови графу ПУМ.
8. Основні параметри графу ПУМ.
9. Що таке завершальний, максимальний та критичний шляхи ПУМ?
10. Що характеризує коефіцієнт напруженості роботи? Його обчислення.
11. Можливості практичного використання ПУМ.

ГЛАВА 13

МАТЕМАТИЧНІ МОДЕЛІ УПРАВЛІННЯ ЗАПАСАМИ

13.1. Основні поняття

Управління запасами є важливою проблемою економіки. Запаси матеріальних ресурсів необхідні в народному господарстві для забезпечення його стабільності та ритмічного функціонування усіх його галузей. Правильне визначення оптимальної стратегії управління запасами та їх нормативного рівня дає можливість звільнити значні оборотні кошти та підвищити ефективність використання ресурсів. За допомогою методів кількісного аналізу можна досягти найвигіднішого компромісу між різноплановими вимогами: з одного боку, зменшенням витрат на транспортування та зберігання, з іншого – надійним забезпеченням споживачів ресурсами.

Предметом теорії управління запасами є розробка методики організації поставок, за якої загальні витрати на доставку, зберігання та забезпечення безперебійного функціонування споживачів будуть мінімальними. Задачі різнопланові, і для їх розв'язання використовуються різні математичні моделі. Розглянемо коротко основні поняття та характеристики використовуваних моделей.

Попит на запас може бути *детермінованим* (однозначно представленим у часі та обсязі) або *випадковим*. Випадковість визначається або випадковим моментом попиту, або обсягом, або обома цими характеристиками.

Поповнення запасів може бути або з фіксованим періодом та обсягом, або в залежності від наявних обсягів, якщо їх величини досягли певного рівня.

Обсяг замовлення визначається наявністю контрольованих запасів та організацією поставок. При періодичному поповненні запасів і випадковому їх використанні обсяг замовлення може залежати від обсягу запасів на момент замовлення або від інтенсивності споживання.

Термін виконання замовлення в деяких моделях управління запасами означає, що замовлення виконується миттєво, якщо така гіпотеза має доцільність. Для інших моделей затримки поставок мають або фіксований, або випадковий термін.

Вартість поставок моделюється як величина, яка залежить від двох складових – певних разових витрат, які не залежать від обсягів поставок, і витрат, обумовлених обсягом поставок.

Збитки зберігання обумовлені тим, що зберігання запасів вимагає певних витрат на зарплату обслуговуючому персоналу, на експлуатацію складських приміщень, дотримання певних умов зберігання запасів і т. ін.

Штраф за дефіцит. Будь-яка система зберігання запасів утворюється з метою

уникнути дефіциту постачання певної продукції в системі обслуговування. Відсутність продукції в необхідний момент породжує збитки, обумовлені різними причинами: неможливістю виконання певних робіт, несвоєчасним поверненням коштів від реалізації і т. ін. Збитки такого походження називаються *штрафом за дефіцит*.

Структура складської системи являє собою співвідношення системи складів, на яких виконуються операції з формування запасів, і постачання споживачів. Складські системи поділяються на *одноланцюгові*, коли запаси зберігаються на одному складі від моменту їх завезення до поставок у систему обслуговування, і *багатоланцюгові*, коли необхідні запаси на шляху до споживача проходять систему складів, що може бути обумовлено вимогами до зберігання, невизначеністю з обсягами замовлень на запаси і т. ін. Прикладом багатоланцюгової системи є організація армійських складів і постачання. У багатоланцюговій системі, як правило, існує певна ієрархія з різними умовами функціонування на кожному щаблі.

У залежності від призначення запаси можна поділити на такі групи: поточного призначення, сезонні, довготермінові, цільового призначення.

Основними елементами системи управління запасами є: організаційна структура товаропостачання, попит, умови поновлення запасів, функція витрат, стратегія управління в залежності від мети постачання.

Моделі управління запасами визначаються попитом, організацією поставок, термінами виконання замовлень. Якщо ці характеристики можна однозначно визначити в часі, то відповідні моделі управління запасами називаються *детермінованими*. Якщо параметри моделі не змінюються в часі, вона називається *статичною*, в протилежному випадку – *динамічною*. Динамічні моделі використовуються за умови, що управління визначається з урахуванням зміни рівня запасу, моменту замовлення на поставку, терміну виконання замовлення і т. ін. Якщо витрати запасів, терміни виконання замовлень мають випадковий характер, то відповідні моделі називаються *стохастичними*.

Під стратегією управління розуміють сукупність правил, якими необхідно керуватися при виборі моменту та обсягу замовлення.

Моделі управління запасами розробляють за тих чи інших гіпотез про використання запасів, організацію їх поповнення, умови зберігання, вимоги щодо дефіциту і т. ін.

13.2. Управління однономенклатурними запасами

Статична детермінована модель без дефіциту

Розглянемо найпростішу модель управління запасами, коли зберігається

продукція одного найменування або коли можна усі витрати звести до одного показника. Приймемо гіпотезу, що обсяг поставки продукції постійний і дорівнює S, поставки виконуються миттєво з періодом T, витрати на зберігання продукції безперервні з постійною інтенсивністю. За прийнятих гіпотез однозначності обсягів S і періоду T поставок природно поставити вимогу про відсутність дефіциту (запобігання зайвих витрат).

Якщо відлік часу вести з моменту одержання першої поставки, то рівень запасу дорівнюватиме обсягу S поставки. За умови рівномірного використання запасів на проміжку T зміни рівнів запасу зображені на рис. 13.1. На проміжку $[0; T]$ рівень запасу змінюється по прямій $s(t) = S - \frac{S}{T} \cdot t$ від S до нуля. У момент T рівень запасу миттєво відновлюється до величини S. Такий характер зміни обсягів запасу повторюється на кожному з проміжків величини T (рис. 13.1).

Якщо відомі загальні потреби Q на певному проміжку θ (квартал, рік), то задача управління запасами зводиться до визначення такого обсягу S кожної поставки, за якої сумарні витрати на поставки та зберігання запасу на проміжку θ були б мінімальними.

Позначимо *сумарні* витрати C, витрати за проміжок θ на поставку – C_n, витрати на зберігання за одиницю часу – $C_з$ та обчислимо ці величини за проміжок T. Припустимо, що витрати на поставку, які не залежать від її обсягу, дорівнюють c_1, а витрати на зберігання одиниці продукції за одиницю часу – c_2. Оскільки за контрольований проміжок θ необхідно мати обсяг запасу Q, який рівномірно витрачається в часі та забезпечується поставками в обсязі S, то кількість поставок k дорівнює:

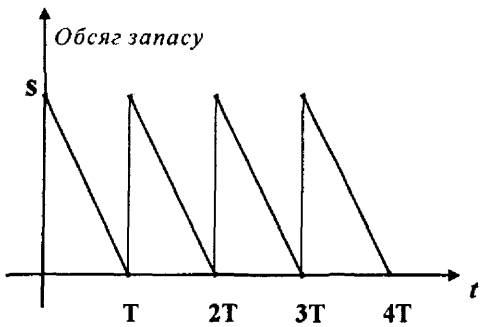


Рис. 13.1.

$$k = \frac{Q}{S} = \frac{S\theta}{ST} = \frac{\theta}{T}. \qquad (1)$$

Звідси маємо:

$$C_n = c_1 k = c_1 \cdot \frac{Q}{S}. \qquad (2)$$

Поточний обсяг запасу $s(t) = S - \frac{S}{T}t$, на його зберігання маємо витрати $c_2 s(t)$, а на проміжку T маємо:

$$\int_0^T c_2 (S - \frac{S}{T}t)dt = c_2 (St - \frac{S}{T}\frac{t^2}{2}) \Big|_0^T = c_2 \frac{ST}{2}. \qquad (3)$$

Враховуючи, що величина обсягу запасу є періодичною функцією з періодом T (рис. 13.1) і проміжок θ має k періодів, то витрати на зберігання $C_з$ дорівнюють:

$$C_з = c_2 \frac{ST}{2} k = c_2 \frac{S\theta}{2}. \qquad (4)$$

Порівнявши (2) та (4), доцільно наголосити, що витрати на поставку C_n обернено пропорційні, а витрати на зберігання $C_з$ прямо пропорційні обсягу однієї поставки S. Графіки функцій $C_n(S)$ та $C_з(S)$, а також функції сумарних витрат

$$C(S) = C_n(S) + C_з(S) = c_1 \frac{Q}{S} + \frac{1}{2} c_2 S\theta \qquad (5)$$

наведені на рис. 13.2.

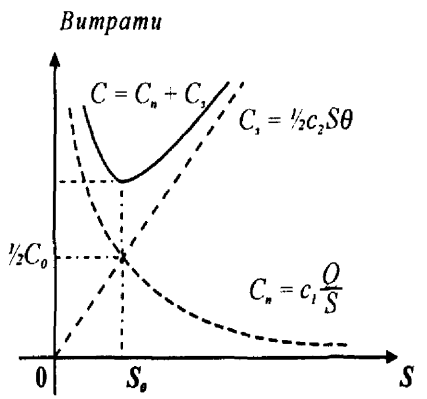


Рис. 13.2.

Мета управління запасами – вибрати такий обсяг поставки S_0, при якому функція сумарних витрат (5) досягає найменшого значення на досліджуваному проміжку θ.

Для обчислення величини мінімуму знайдемо $C'(S)$ та прирівняємо її до нуля:

$$C'(S) = -c_1 \frac{Q}{S^2} + \frac{1}{2} c_2 \theta = 0.$$

Таким чином, оптимальний обсяг поставок S_0 дорівнює:

$$S_0 = \sqrt{\frac{2c_1 Q}{c_2 \theta}}. \qquad (6)$$

Підносимо обидві частини (6) до квадрата та використовуємо залежності (2) і (4):

$$c_1 \frac{Q}{S_0} = \frac{1}{2} c_2 S_0 \theta. \qquad (7)$$

Залежність (7) показує, що мінімум витрат досягається за умови, коли витрати на формування запасу дорівнюють витратам на його зберігання. За цих умов мінімальні сумарні витрати складають:

$$C_0 = \frac{2c_1 Q}{S}. \qquad (8)$$

Кількість оптимальних поставок k_0 за проміжок θ дорівнює:

$$k_0 = \frac{Q}{S_0} = \sqrt{\frac{c_2 Q \theta}{2c_1}}.$$

Оптимальний період поставок T_0 (термін використання поставки оптимального обсягу) становить:

$$T_0 = \frac{\theta}{k_0} = \frac{\theta S_0}{Q} = \sqrt{\frac{2c_1 \theta}{c_2 Q}}. \qquad (9)$$

Обчислимо оптимальні витрати на функціонування системи постачання за проміжок θ. Скориставшися залежностями (5), (2), (4) та (6), маємо:

$$C_0(S) = c_1 \frac{Q}{S_0} + c_2 \frac{S_0 \theta}{2} = \left| \; S_0 = \sqrt{\frac{2c_1 Q}{c_2 \theta}} \; \right| = \sqrt{2c_1 c_2 Q \theta}. \qquad (10)$$

Приклад. Потреба в деяких деталях складає на рік 480000 шт. У процесі виробництва вони використовуються безперервно та рівномірно. Деталі замовляються один раз на рік і постачаються в обсязі, вказаному в замовленні. Зберігання однієї деталі коштує 0.35 у. о. за добу, а постачання однієї партії – 10000 у. о. Визначити оптимальний обсяг однієї поставки та проміжок між поставками, які необхідно вказати в замовленні, за умови, що затримки в поставках недопустимі.

За формулою (6) маємо:

$$S_0 = \sqrt{\frac{2c_1 Q}{c_2 \theta}} = \left| \begin{array}{l} c_1 = 10000; Q = 480000 \\ c_2 = 0.35; \theta = 365 \;\; діб \end{array} \right| = \sqrt{\frac{2 \times 10000 \times 480000}{0.35 \times 365}} = 8670 \; дет.$$

Оптимальний період поставки T_0 визначаємо за формулою (9):

$$T_0 = \frac{\theta S_0}{Q} = \left| \begin{array}{l} \theta = 365; S_0 = 8670 \\ Q = 480000 \end{array} \right| = \frac{365 \times 8670}{480000} = 6.6.$$

Таким чином, найбільш економний обсяг поставки - 8670 деталей, а проміжок між поставками – 7 діб.

Обчислимо витрати на функціонування системи постачання за наведених у прикладі умов. Скориставшись формулами (5), (2) і (4), маємо:

$$C_0 = c_1 \frac{Q}{S} + c_2 \frac{S\theta}{2} = \left| \begin{array}{l} c_1 = 10000 \;\; у.о.; \;\; Q = 480000 \;\; дет. \\ S = 8670 \;\; дет; \;\; c_2 = 0.35 \;\; у.о.; \;\; \theta = 365 \;\; діб \end{array} \right| =$$

$$\frac{10000 \times 480000}{8670} + 0.35 \frac{8670 \times 365}{2} = 559332 \;\; у.о.$$

Статична детермінована модель з дефіцитом

Розглянемо одну з можливих моделей управління однономенклатурними запасами, допускаючи можливість дефіциту. Це означає, що при відсутності запасу попит зберігається, і дефіцит накопичується з деякою швидкістю b, тобто з тією ж інтенсивністю. Графік зміни запасу в часі для такого випадку представлений

на рис. 13.3. Розташування графіка під віссю абсцис в області від'ємних значень величини запасу визначає, на відміну від графіка рис. 13.2, накопичення дефіциту. Якщо період між поставками сталий і дорівнює $T = \frac{S}{b}$, то його необхідно розділити на два проміжки: T_1 – термін, коли є можливість скористатися наявним запасом; T_2 – термін, коли запас відсутній і накопичується дефіцит, який потім буде ліквідований у момент чергової поставки.

Необхідність компенсації дефіциту обумовлюється тим, що максимальний рівень запасу S^* в момент виконання поставки не дорівнює її обсягу S, а менший на величину дефіциту $S - S^*$, який виник за час T_2 (рис. 13.3).

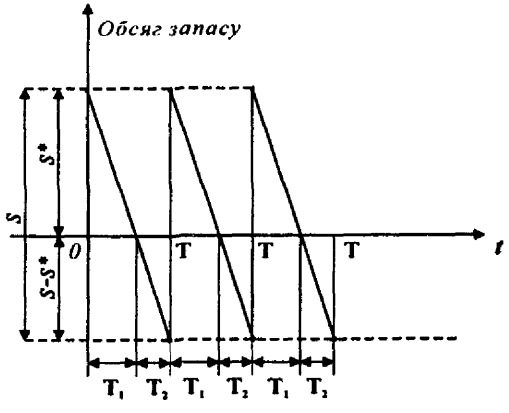


Рис. 13.3.

Оскільки користуємося гіпотезою про лінійну залежність обсягу споживання запасу від часу, то

$$T_1 = \frac{S^*}{S}T; \qquad T_2 = \frac{S - S^*}{S}T. \tag{11}$$

Функція сумарних витрат C буде визначатися трьома доданками: витратами $C_н$ і $C_з$, обумовленими необхідністю поставки та зберігання запасу, та обсягами C_c штрафів через дефіцит:

$$C = C_н + C_з + C_c. \tag{12}$$

Витрати $C_н$ обчислюємо за формулою (2), витрати $C_з$ – за формулою (4),

враховуючи, що зберігається запас в обсязі S^* терміном T_1:

$$C_з = \frac{c_2 S^* T_1}{2} k = \left| \; k = \frac{\theta}{T}; \quad T_1 = \frac{S^* T}{S} \right| = c_2 \frac{(S^*)^2 \theta}{2S}. \tag{13}$$

Обчислення стягнення за дефіцит C_c виконується за умови, що за кожну одиницю продукції в дефіциті сплачується штраф c_3 за кожну одиницю часу. Оскільки середній рівень дефіциту за термін T_2 дорівнює $\frac{1}{2}(S - S^*)T_2$, то штраф за вказаний період становить $\frac{1}{2}c_3(S - S^*)T_2$, а за весь період θ, враховуючи (11), маємо:

$$C_c = \frac{1}{2}c_3(S - S^*)T_2 k = \left| \; T_2 = \frac{S - S^*}{S}T; \quad k = \frac{\theta}{T} \right| = \frac{c_3 \theta (S - S^*)^2}{2S}. \tag{14}$$

Таким чином, сумарні витрати C з урахуванням формул (2), (13) і (14) обчислюються так:

$$C = c_1 \frac{Q}{S} + c_2 \frac{\theta (S^*)^2}{2S} + c_3 \frac{\theta (S - S^*)^2}{2S}. \tag{15}$$

Якщо S дорівнює S^*, то формули (15) і (5) тотожні, коли не враховувати дефіцит.

Досліджувана задача управління запасами за умови дефіциту приведена до визначення таких значень S та S^*, за яких функція C (15) досягає найменшого значення.

Значення функції C (15) може бути екстремальним лише в стаціонарній точці, тобто якщо частинні похідні $\frac{\partial C}{\partial S}$ та $\frac{\partial C}{\partial S^*}$ дорівнюють нулю.

$$\begin{cases} \dfrac{\partial C}{\partial S} = -c_1 \dfrac{Q}{S^2} - c_2 \dfrac{\theta (S^*)^2}{2S^2} - c_3 \dfrac{\theta (S - S^*)^2}{2S^2} + c_3 \dfrac{\theta (S - S^*)}{S} = 0 \\ \dfrac{\partial C}{\partial S^*} = c_2 \dfrac{\theta S^*}{S} - c_3 \dfrac{\theta (S - S^*)}{S} = 0. \end{cases}$$

Виконавши алгебраїчні перетворення, систему рівнянь перепишемо так:

$$\begin{cases} S^2 c_3 - (c_2 + c_3)(S^*)^2 = 2c_1 Q / \theta \\ S^* = S \dfrac{c_3}{c_2 + c_3}. \end{cases} \quad (16)$$

Розв'язавши систему (16) та переконавшися за допомогою необхідної умови, що стаціонарна точка функції C (15) буде точкою мінімуму, одержуємо величини параметрів оптимального управління для пропонованої моделі. Обсяг оптимальної поставки S_0 дорівнює:

$$S_0 = \sqrt{\frac{2c_1 Q}{c_2 \theta}} \sqrt{\frac{c_2 + c_3}{c_3}}, \quad (17)$$

а обсяг оптимального запасу S_0^* –

$$S_0^* = \sqrt{\frac{2c_1 Q}{c_2 \theta}} \sqrt{\frac{c_3}{c_2 + c_3}}. \quad (18)$$

Величина $\rho = \dfrac{c_3}{c_2 + c_3}$ називається *коефіцієнтом збитків через дефіцит* і має суттєве значення в управлінні запасами. Наголосимо, що за умови, коли c_3 – мала величина відносно c_2, величина ρ близька до нуля; якщо ж c_3 суттєво перевищує значення c_2, то ρ має значення близьке до одиниці. У загальному випадку виконується умова: $0 \le \rho \le 1$. Недопустимість дефіциту рівносильна вимозі, що $c_3 = \infty$ або $\rho = 1$.

Скориставшися формулою (11), маємо $\dfrac{T_1}{T} = \dfrac{S_0^*}{S_0} = \rho$ і $\dfrac{T_2}{T} = \dfrac{(S_0 - S_0^*)}{S_0} = 1 - \rho$.

Порівнюючи формули (17) та (6), знаходимо, що оптимальні обсяги поставок для задач з дефіцитом $\tilde{S}_0$ і без дефіциту S_0 за однакових параметрів пов'язані залежністю:

$$\tilde{S}_0 = S_0 \frac{1}{\sqrt{\rho}}. \quad (19)$$

Таким чином, оптимальний обсяг поставки за умов дефіциту завжди більший в $1/\sqrt{\rho}$ разів, ніж у задачі без дефіциту.

13.3. Статична модель управління багатономенклатурними запасами

Господарська діяльність у багатьох випадках визначається потоком обсягів різноманітної продукції, тому важливе значення мають задачі управління багатономенклатурними запасами.

На думку фахівців, оптимізацію управління запасами доцільно проводити лише на ті найменування, які складають за вартістю не менше 60% загального потоку запасів. Наголосимо, що в деяких випадках багатономенклатурну задачу можна привести до однієї або кількох задач однономенклатурних за таких умов:

– поставки різної продукції виконуються незалежно;

– поставки виконуються упорядкованими комплектами, завдяки чому комплекти можна вважати за як одну номенклатуру;

– попит на певний набір продукції суттєво взаємопов'язаний, і таку множину можна вважати за одну номенклатуру.

Розглянемо задачі управління багатономенклатурними запасами за умов:

– періоди поставок тотожні;

– наявними є певні обмеження.

Припустимо, що досліджуємо задачу на проміжку θ і всі m-номенклатур мають один і той же період поставки T. За таких умов сумарні витрати D_1 на зберігання запасу дорівнюватимуть:

$$D_1 = \frac{1}{2} \sum_{i=1}^{m} c_{2i} S_i \theta \quad (20)$$

(S_i – обсяг i-ої продукції в поставці; c_{2i} – витрати на зберігання одиниці i-ої продукції протягом одиниці часу).

Урахувавши, що

$$S_i = \frac{Q_i}{k} = \frac{Q_i}{\theta} T, \quad (21)$$

обчислимо

$$D_1 = \frac{1}{2} \sum_{i=1}^{m} c_{2i} Q_i T \quad (22)$$

(Q_i – обсяг поставок i-ої продукції за контрольований термін θ).

Припустимо, що оцінку витрат d на одну поставку можна виконати за такою залежністю:

$$d = d_1 + d_2 m \tag{23}$$

(d_1 – величина витрат на поставку, яка не залежить від обсягів поставки; d_2 – величина, яка дозволяє врахувати вплив кількості номенклатур, що входять у поставку, на її вартість).

За таких умов вартість всіх поставок D_2 на проміжку θ дорівнюватиме:

$$D_2 = (d_1 + d_2 m)k = (d_1 + d_2 m)\frac{\theta}{T}. \tag{24}$$

Сумарні витрати D на управління запасами на проміжку θ дорівнюють:

$$D = D_1 + D_2 = \frac{1}{2}\sum_{i=1}^{m} c_{2i} Q_i T + (d_1 + d_2 m)\frac{\theta}{T}. \tag{25}$$

Обчислимо оптимальні параметри управління запасами. Знайдемо $D_T^{/}$:

$$D_T^{/} = \frac{1}{2}\sum_{i=1}^{m} c_{2i} Q_i - \frac{(d_1 + d_2 m)\theta}{T^2}.$$

Оптимальна величина періоду поставки T_0 дорівнює:

$$T_0 = \sqrt{\frac{2(d_1 + d_2 m)\theta}{\sum_{i=1}^{m} c_{2i} Q_i}}. \tag{26}$$

Оптимальні кількість k_0 та обсяги S_{i0} поставок будуть відповідно дорівнювати:

$$k_0 = \frac{\theta}{T_0} = \sqrt{\frac{\sum_{i=1}^{m} c_{2i} Q_i \theta}{2(d_1 + d_2 m)}}, \tag{27}$$

$$S_{i0} = \frac{Q_i}{k_0} = Q_i \sqrt{\frac{2(d_1 + d_2 m)}{\sum_{i=1}^{m} c_{2i} Q_i \theta}}. \tag{28}$$

Підставивши у функцію витрат D (25) оптимальне значення періоду поставки T_0 (26), одержимо величину оптимальних витрат на функціонування системи постачання:

$$D_{\min} = \sqrt{2(d_1 + d_2 m)\sum_{i=1}^{m} c_{2i} Q_i \theta}. \tag{29}$$

Виконані дослідження в межах побудованих математичних моделей дали можливість визначити оптимальні параметри функціонування системи постачання за критерієм мінімальної вартості витрат. За умови управління однономенклатурними запасами функція витрат $C(S)$ у бездефіцитній моделі визначається залежністю (5):

$$C(S) = c_1 \frac{Q}{S} + \frac{1}{2} c_2 S\theta.$$

Тепер з'ясуємо питання функціонування системи постачання за умови її фінансування в сумі D:

$$D = c_1 \frac{Q}{S} + \frac{1}{2} c_2 S\theta. \tag{30}$$

Обчислимо обсяг поставок S, який задовольнятиме умові (30), тобто ця величина має бути коренем відповідного квадратного рівняння:

$$S_{1,2} = \frac{D \pm \sqrt{D^2 - 2c_1 c_2 Q\theta}}{c_2 \theta}.$$

Дійсні корені (30) (реальні обсяги поставок) можуть існувати за умови $D \geq \sqrt{2c_1 c_2 Q\theta}$. Порівнявши (10) й одержану умову, доходимо висновку, що система постачання існуватиме, якщо її фінансування на проміжку θ буде не в меншому обсязі, ніж оптимальні витрати на неї (10).

Для багатономенклатурної бездефіцитної моделі, функція витрат в якій визначається формулою (25), маємо аналогічний результат: система постачання може функціонувати лише за умови, якщо обсяг її фінансування буде не менше величини мінімальних витрат (29) на проміжку θ.

При розбудові будь-якого типу моделі управління запасами досить часто виникає необхідність урахування певних обмежень на параметри. Такі обмеження можуть бути на періоди поставок, обсяги поставок, проміжки від замовлення до його реалізації і т. ін. Розглянемо ситуацію, коли необхідно враховувати обмеження за об'ємом складських приміщень. Це може стримувати накопичення запасів або підвищувати вартість їх зберігання. Припустимо, що необхідно обчислити обсяги $S_1, S_2, \ldots, S_m$ одночасних поставок m-номенклатур за умови, що відомі витрати c_{2i}

на зберігання одиниці кожної номенклатури та c_{1i} на постачання кожної партії, обсяги поставок на проміжку θ та обмеження за об'ємом V складських приміщень. Скориставшися формулами (2) і (20), визначаємо, що загальні витрати D обчислюються за формулою:

$$D = \frac{1}{2}\sum_{i=1}^{m} c_{2i} S_i \theta + \sum_{i=1}^{m} c_{1i} \frac{Q_i}{S_i}.$$

При цьому враховуємо обмеження $V = \sum_{i=1}^{m} V_i S_i \le V^*$ (V_i – об'єм складського приміщення для зберігання одиниці i-ої номенклатури; V^* – об'єм складу для зберігання обсягу всієї поставки). Оптимальне управління запасами приведено до задачі пошуку умовного екстремуму, сформульованої так: обчислити величини обсягів поставок S_i так, щоб функція загальних витрат D досягала найменшого значення за умови, що будуть виконані обмеження на величину складського приміщення.

13.4. Стохастичні моделі управління запасами

Розглянуті раніше моделі управління запасами були розбудовані відповідно до гіпотез, що рівень споживання запасів можна описати лінійною функцією в часі і що терміни та обсяги поставок можуть бути детермінованими. Але на практиці і попит, і термін поставки, як правило, є випадковими величинами. Урахування цих обставин суттєво впливає на принципи побудови відповідних моделей та обумовлює необхідність введення нових понять.

Страховий запас – обсяг запасу, який необхідно мати, щоб гарантувати безперебійне постачання (при нерівномірностях попиту та різних випадкових проміжках) між замовленням на поставку і його виконанням.

Момент замовлення визначається часом, коли рівень запасу досягає такого значення $S_{кр}$, за якого необхідно робити замовлення на чергову поставку, щоб запобігти дефіциту на проміжку між черговим замовленням і виконанням поставки.

Рівень обслуговування $U_{об}$ – ймовірність P того, що планового запасу $S_{кр}$ вистачить для виконання запиту в обсязі q на випадковому проміжку τ від замовлення поставки до його виконання, тобто $U_{об} = P(q < S_{кр})$. Інакше $U_{об}$ – це ймовірність відсутності дефіциту. Ймовірність дефіциту $P_{деф}$ дорівнюватиме:

$$P_{деф} = 1 - U_{об} = P(q > S_{кр}). \qquad (31)$$

На практиці обсяги запасів науково не обґрунтовані, що призводить до невиправдано значної витрати обігових коштів та "заморожування" матеріальних цінностей. Така ситуація досить негативно впливала на соціалістичну економіку, а в умовах ринкової економіки може бути згубною. Обсяги страхового запасу $S_{кр}$ необхідно економічно обґрунтувати, гарантуючи певний рівень обслуговування. Розглянемо деякі моделі дослідження вказаної ситуації. Для спрощення вважатимемо, що маємо однономенклатурну задачу. Розбудовуючи модель, будемо виходити з того, що знайдено орієнтовно оптимальні детерміновані параметри управління системою, тобто задача розв'язана за умови вибору деяких середніх значень величин необхідних параметрів (йдеться про уточнення моделі).

Обсяг поставки S маємо згідно з формулою (6). Момент замовлення визначатиметься рівнем запасу $S_{кр}$:

$$S_{кр} = \bar{q} + R \qquad (32)$$

($\bar{q} = M(q)$ – середній попит (математичне сподівання) за проміжок τ від моменту замовлення поставки до його виконання;
R – страховий (резервний) запас, необхідний для виконання замовлення у випадку, якщо фактичний попит q за термін реалізації перевищуватиме очікуваний $\bar{q}$, тобто $q > \bar{q}$).

Сумарні витрати C на управління при врахуванні страхового запасу R обчислимо, скориставшися формулою (10):

$$C = C_0(S) + c_2 R. \qquad (33)$$

Як відомо, за міру відхилення можливих значень випадкової величини від її математичного сподівання (середнього значення) прийнято середнє квадратичне відхилення σ, тому доцільно прийняти, що резервний запас пропорційний $\sigma(q)$:

$$R = \lambda\sigma(q). \qquad (34)$$

Тоді момент замовлення визначається за умови:

$$S_{кр} = \bar{q} + \lambda\sigma(q). \qquad (35)$$

Величина λ має бути вибрана так, щоб забезпечити певний рівень обслуговування:

$$U_{об} = P(q < S_{кр}) = P(q < \bar{q} + \lambda\sigma(q)) \quad . \qquad (36)$$

Середній попит $\overline{q}$ на проміжку τ і його середнє квадратичне відхилення $\sigma(q)$ за умови випадкових величин q і τ знайдемо в такий спосіб.

Позначимо: $\overline{\tau} = M(\tau)$ – математичне сподівання (середнє значення) і $\sigma(\tau)$ – середнє квадратичне відхилення випадкової величини τ-проміжку між замовленням та виконанням поставки; q_τ – інтенсивність попиту (обсяг попиту в одиницю часу) на проміжку τ. Приймаючи природне допущення, що інтенсивність попиту q_τ та проміжок τ незалежні випадкові величини, маємо:

$$q = q_\tau \tau; \qquad \overline{q} = M(q_\tau \tau) = \overline{q}_\tau \overline{\tau} \tag{37}$$

($\overline{q}_\tau$– математичне сподівання інтенсивності попиту на проміжку τ).

Дисперсія $D(q)$ попиту на проміжку τ дорівнює:

$$D(q) = \sigma^2(q) = M((q_\tau \cdot \tau)^2) - (\overline{q}_\tau \cdot \overline{\tau})^2 = M(q_\tau^2 \cdot \tau^2) - \overline{q}_\tau^2 \cdot \overline{\tau}^2 = M(q_\tau^2) \cdot M(\tau^2) - \overline{q}_\tau^2 \cdot \overline{\tau}^2.$$

Урахувавши, що $D(q_\tau) = \sigma^2(q_\tau) = M(q_\tau^2) - \overline{q}_\tau^2$; $D(\tau) = \sigma^2(\tau) = M(\tau^2) - \overline{\tau}^2$, обчислимо

$$\sigma^2(q) = \left[\sigma^2(q_\tau) + \overline{q}_\tau^2\right] \cdot \left[\sigma^2(\tau) + \overline{\tau}^2\right] - \overline{q}_\tau^2 \overline{\tau}^2 = \sigma^2(q_\tau)\sigma^2(\tau) + \overline{\tau}^2 \sigma^2(q_\tau) + \overline{q}_\tau^2 \sigma^2(\tau).$$

Отже,

$$\sigma(q) = \sqrt{\sigma^2(q_\tau)\sigma^2(\tau) + \overline{\tau}^2 \sigma^2(q_\tau) + \overline{q}_\tau^2 \sigma^2(\tau)}; \tag{38}$$

$$R = \lambda \sqrt{\sigma^2(q_\tau)\sigma^2(\tau) + \overline{\tau}^2 \sigma^2(q_\tau) + \overline{q}_\tau^2 \sigma^2(\tau)} \quad . \tag{39}$$

Практично формулою (39) можна користуватися, якщо відомі закони розподілу випадкових величин q_τ – інтенсивності попиту і τ– проміжку між черговими поставками. Розглянемо деякі окремі випадки.

1. Інтенсивність попиту q_τ є детермінованою величиною, проміжок τ – випадковою. За таких умов $q_\tau = q^*$ стала, а $\sigma(q_\tau) = 0$. Попит q та його середнє квадратичне відхилення визначаються залежностями:

$$q = q^* \tau; \quad \overline{q} = q^* \overline{\tau}; \tag{40}$$

$$\sigma(q) = q^* \sigma(\tau). \tag{41}$$

Резервний запас (34) визначається так:

$$R = \lambda q^* \sigma(\tau). \tag{42}$$

Момент замовлення (35) визначається умовою:

$$S_{кр} = q^* \left[\overline{\tau} + \lambda \sigma(\tau)\right]. \tag{43}$$

2. Інтенсивність попиту q_τ – випадкова величина, а проміжок τ – детермінована:

$$\tau = \tau^* \text{ стала, } \sigma(\tau) = 0;$$

$$\overline{q} = \overline{q}_\tau \cdot \tau^*; \quad \sigma(q) = \sigma(q_\tau)\tau^*; \tag{44}$$

$$S_{кр} = \tau^* (\overline{q}_\tau + \lambda \sigma(q_\tau)). \tag{45}$$

3. Інтенсивність попиту q_τ і проміжок τ – детерміновані величини, тобто $\tau = \tau^*$ стала, $q_\tau = q^*$ стала. За таких умов $\sigma(q) = 0$, $R = 0$.

Момент замовлення визначається обсягом запасу

$$S_{кр} = q^* \cdot \tau^*, \tag{46}$$

який достатній для забезпечення попиту до одержання наступної поставки, тобто резервний запас непотрібний.

Тепер розглянемо управління запасами за умов дії деяких законів розподілу випадкових величин інтенсивності попиту та проміжку виконання замовлення.

Управління запасами за умови, що попит характеризується нормальним законом розподілу. Приймаємо гіпотезу про нормальний закон розподілу попиту з параметрами q та σ. Функція щільності має вигляд:

$$f(q) = \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(q-\overline{q})^2}{2\sigma^2}}.$$

Як відомо, при дії нормального закону розподілу ймовірність події, що випадкова величина (попит) відхилиться від свого математичного сподівання на величину $\varepsilon > 0$, обчислюється так:

$$p(|q-\bar{q}|<\varepsilon)=\Phi(\frac{\varepsilon}{\sigma})$$

($\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ - функція Лапласа).

Якщо $\varepsilon=\lambda\sigma$, то $p(|q-\bar{q}|<\lambda\sigma)=\Phi(\lambda)$. Таким чином, ймовірність події, що величина попиту q відхилиться від свого середнього значення (математичного сподівання) за термін виконання замовлення менше ніж на $\lambda\sigma$, дорівнює $\Phi(\lambda)$. Отже, якщо взяти страховий запас у розмірі $R=\lambda\sigma$, а момент замовлення визначити за величиною запасу $S_{кр}=\bar{q}+\lambda\sigma$, то при використанні формули (31) матимемо:

$$P_{деф}=P(q>\bar{q}+\lambda\sigma)=\frac{1}{2}(1-\Phi(\lambda)). \qquad (47)$$

Рівень обслуговування $U_{об}$ можна обчислити за формулою:

$$U_{об}=1-P_{деф}=\frac{1}{2}(1+\Phi(\lambda)). \qquad (48)$$

Формула (48) дозволяє при вибраному рівні обслуговування $U_{об}$ обчислити коефіцієнт λ, рівень страхового запасу R та величину запасу на момент замовлення поставки. Якщо $\lambda=2$, то $R=2\sigma$, $S_{кр}=\bar{q}+2\sigma$, $P_{деф}=\frac{1}{2}(1-0{,}9545)$, $U_{об}=\frac{1}{2}(1+0{,}9545)=0{,}9772$. Якщо $\lambda=3$, то $R=3\sigma$, $S_{кр}=\bar{q}+3\sigma$, $P_{деф}=\frac{1}{2}(1-0{,}9973)=0{,}00135$, $U_{об}=\frac{1}{2}(1+0{,}9973)=0{,}9986$.

Приклад. Склад поповнюється деталями певного типу зі сталим проміжком від моменту замовлення до поставки. Попит на такі деталі на проміжку одержання поставок характеризується нормальним законом розподілу з параметрами $\bar{q}=10000$ од. і $\sigma=2000$ од. Треба визначити момент замовлення та величину страхового запасу так, щоб рівень обслуговування був не менше 0,98.

Скориставшися формулою (48), вибираємо значення λ так, щоб досягти бажаного рівня обслуговування. Якщо $\lambda=1$, то $U_{об}=\frac{1}{2}(1+\varphi(1))=0{,}8413$, що менше бажаної величини. Якщо $\lambda=2$, то $U_{об}=\frac{1}{2}(1+\varphi(2))=0{,}9773$. Збільшимо λ до 2,1. Тоді $U_{об}=\frac{1}{2}(1+\varphi(2{,}1))=0{,}9821$, що задовольняє бажаному рівневі і буде

досягнуто, якщо страховий запас $R=2{,}1\cdot 2000=4200$ од. Замовлення на чергову поставку треба подавати за умови, що запас досяг рівня $S_{кр}=10000+4200=14200$ од.

Управління запасами за умови, що попит характеризується розподілом за законом Пуассона. Приймемо, що проміжок від замовлення до поставки є сталою величиною, але обсяг попиту q описується розподілом за законом Пуассона з математичним сподіванням $\bar{q}$. За таких умов ймовірність події, що обсяг попиту q прийме значення m, обчислюється за формулою:

$$P(q=m)=\frac{\bar{q}^m}{m!}e^{-\bar{q}}. \qquad (49)$$

Коли випадкова величина розподілена за законом Пуассона, то квадрат її середнього квадратичного відхилення дорівнює її математичному сподіванню, тобто $\sigma^2=\bar{q}$. Якщо прийняти $R=\lambda\sqrt{\bar{q}}$, то рівень запасу $S_{кр}$ на момент замовлення визначиться так: $S_{кр}=\bar{q}+\lambda\sqrt{\bar{q}}$, а рівень обслуговування

$$U_{об}=P(q<S_{кр})=P(q<\bar{q}+\lambda\sqrt{\bar{q}}). \qquad (50)$$

Використавши таблиці розподілу за законом Пуассона та формулу (50), можна підібрати величину λ, при якій досягається певний рівень обслуговування, а потім обчислити відповідний страховий запас. Наголосимо, що при $\lambda=2$ $U_{об}\approx 0{,}98$, а це означає, що рівень страхового запасу в обсязі $R=2\sqrt{\bar{q}}$ гарантує досить високий рівень обслуговування.

Управління запасами за умови штрафу за дефіцит. Розглянемо стохастичну модель управління запасами за умови випадкового попиту на фіксованому проміжку T та наявності штрафу за дефіцит. Припустимо, що обсяг попиту r за проміжок T є випадковим, і маємо його закон (ряд) розподілу $p(r)$ або функцію щільності $f(r)$. Закон розподілу або функцію щільності будують за статистичними даними. Якщо обсяг попиту r менше рівня запасу S, то закупівля та зберігання надлишку вимагають певних додаткових витрат c_2 на одиницю продукції на проміжку T. Якщо попит буде більший, ніж обсяг запасу, то це обумовить штраф за дефіцит у розмірі c_3 за кожну одиницю обсягу. *Задача управління запасами полягає в тому, щоб вибрати такий обсяг запасу S_0, за якого математичне сподівання сумарних витрат на закупівлю, зберігання та штраф за можливий дефіцит матиме найменше значення.*

За умови, що обсяг попиту r є дискретною величиною з відомим законом розподілу $p(r)$, математичне сподівання загальних витрат $C(S)$ обчислюється за формулою:

$$C(S) = c_2 \sum_{r=0}^{S} (S-r)p(r) + c_3 \sum_{r=S+1}^{\infty} (r-S)p(r). \qquad (51)$$

Перший доданок дозволяє оцінити витрати на закупівлю та зберігання надлишків $S-r \quad (r \le S)$.

Якщо попит розглядати як неперервну випадкову величину з відомою функцією щільності $f(r)$, то математичне сподівання загальних витрат $C(S)$ можна обчислити за формулою:

$$C(S) = c_2 \int_0^S (S-r)f(r)dr + c_3 \int_S^{\infty} (r-S)f(r)dr. \qquad (52)$$

Доведено, що за умови дискретного випадкового попиту r функція $C(S)$ (51) досягає свого найменшого значення при S_0, яка задовольняє нерівностям:

$$F(S_0) < \rho < F(S_0+1), \qquad (53)$$

а за умови неперервного випадкового попиту r функція $C(S)$ (52) досягає свого мінімального значення при S_0, яка задовольняє рівнянню:

$$F(S_0) = \rho \qquad (54)$$

($F(S) = p(r < S)$ – *функція розподілу попиту,* ρ – коефіцієнт збитків через дефіцит, який визначається за формулою (19)).

Приклад. Підприємство бажає придбати агрегат і запасні блоки до нього. Вартість одного блоку 6 у. о. Якщо агрегат вийде з ладу через поломку блоку, то простій і термінове замовлення нового блоку будуть коштувати 120 у. о. Необхідно визначити кількість блоків, які треба замовити разом з агрегатом, якщо відомий емпіричний закон розподілу кількості агрегатів, для яких за термін планової експлуатації необхідно було замінити певну кількість блоків (табл. 13.1).

Таблиця 13.1.

Кількість замінених блоків r	0	1	2	3	4	5	6
Частка агрегатів, яким потрібна була заміна r блоків	0,90	0,05	0,02	0,01	0,01	0,01	0,00

За умовою $c_2 = 6$, $c_3 = 120$. Обчислимо коефіцієнт збитків через дефіцит (19): $\rho = 120/(6+120) \approx 0{,}952$.

Скориставшися даними табл. 13.1, побудуємо емпіричну функцію розподілу $F(S) = p \ (r < S)$. Відповідні числові дані наведені у табл. 13.2 та представлені на рис. 13.4.

Таблиця 13.2.

S	0	1	2	3	4	5	6	>6
r	0	1	2	3	4	5	6	>6
$F(S)$	0,00	0,90	0,95	0,97	0,98	0,99	1	1

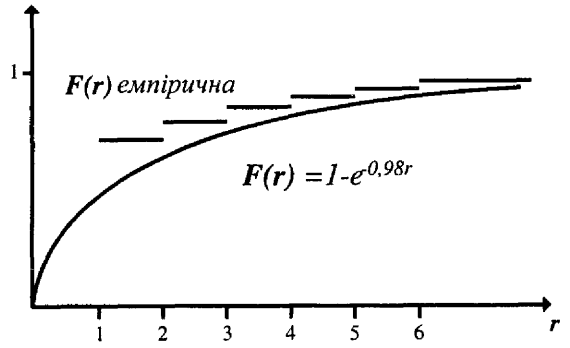


Рис. 13.4.

Скориставшись умовою обчислення оптимальної кількості запасу (53) та даними табл. 13.2, знаходимо, що оптимальний запас $S_0 = 2$, тому що виконується умова $F(2) < 0{,}952 < F(3)$. Таким чином, оптимальна кількість запасних блоків дорівнює 2. Природно чекати, що на практиці буде замовлено три блоки.

Тепер розглянемо сформульовану задачу за умови, що попит на запасні блоки характеризується функцією розподілу $F(r) = 1 - e^{-\lambda r}$ за $\lambda = 0{,}98$. За такої умови розподілу кількості блоків, які необхідно замінити, оптимальну кількість запасних блоків S_0 обчислюємо, скориставшися формулою (54):

$$F(S_0) = S \Rightarrow 1 - e^{-\lambda S_0} = \rho.$$

$$e^{-\lambda S_0} = 1 - \rho,$$

$$S_0 = -\frac{1}{\lambda}\ln(1-\rho) = \left|\lambda = 0{,}98; \rho = 0{,}952\right| = S_0 = -\frac{1}{0{,}98}\ln\left|1 - 0{,}952\right| \approx 3.$$

Тобто кількість запасних блоків за прийнятого закону розподілу $F(r) = 1 - e^{-0{,}98r}$ дорівнює 3.

Різниця в результатах розрахунків обумовлена тим, що були прийняті різні закони розподілу попиту на запасні блоки. Відповідні функції розподілів наведені на рис. 13.4.

На завершення доцільно наголосити, що знайти аналітично оптимальне управління (обсягами необхідного запасу, обсягами та термінами поставок) можна лише в простих випадках. Якщо система зберігання запасів складна (багато номенклатур, ієрархічна структура), використовувані стохастичні моделі складні, їх параметри змінні в часі, то єдиним ефективним засобом аналізу та управління такими системами є *імітаційне моделювання*, яке дозволяє з допомогою комп'ютерів досліджувати та керувати системами запасів. Засади розбудови імітаційних моделей описані у гл. 17.

Питання для самоконтролю

1. Як треба розуміти проблему управління запасами?
2. Які основні поняття використовуються в теорії управління запасами?
3. Основні відмінності між детермінованими та стохастичними моделями управління запасами.
4. Які відомості про систему постачання запасами потрібні для побудови певної моделі управління?
5. Які гіпотези використані в розглянутій детермінованій моделі управління однономенклатурними запасами?
6. Які гіпотези використані в побудові статичної детермінованої моделі управління запасами з урахуванням дефіциту?
7. Основні поняття, використані при побудові пропонованої стохастичної моделі управління запасами.
8. Відмінності між функціями розподілу та щільністю розподілу випадкової величини.
9. З яких міркувань визначаються момент замовлення та обсяг страхового запасу?
10. Зміст поняття "рівень обслуговування".

ГЛАВА 14

ЕЛЕМЕНТИ ТЕОРІЇ ГРИ В РОЗВ'ЯЗАННІ ГОСПОДАРСЬКИХ ЗАДАЧ

14.1. Основні поняття теорії гри

Перші дослідження деяких питань теорії гри були виконані французьким математиком Е. Борелем у 20-ті роки 20-го століття. Монографію Дж. Неймана та О. Моргенштерна "Теорія гри й економічна поведінка", яка вийшла з друку в 1944 р., приймають за початок становлення теорії гри як самостійної науки. У часи другої світової війни та по її завершенні теорією гри дуже зацікавилися військові спеціалісти, які побачили в ній ефективний апарат дослідження стратегічних питань. Згодом у працях учених знову почали розглядатися питання економіки та планування. Кількість досліджень з різнопланового використання теорії гри весь час розширюється.

Різноманітні методи теорії оптимізації тих чи інших процесів використовуються тоді, коли перебіг може бути керований з одного певного центру і ефективність цього процесу оцінюється лише з урахуванням інтересів такого центру. У даній главі розглядається розв'язання задач прийняття рішень при наявності кількох зацікавлених сторін, коли інтереси цих сторін не співпадають, а часто й протилежні, тобто за умови так званих конфліктних ситуацій.

Предметом теорії ігор є дослідження конфліктних ситуацій методом математики. Однією з характерних і суттєвих рис громадського, соціально-економічного процесу є розмаїття та різноплановість інтересів і наявність сторін, які є носіями таких інтересів. Класичними прикладами є ринкові відносини: продавець-покупець; кілька виробників товару, які можуть об'єднуватися та встановлювати ціну товару; кілька конкуруючих між собою виробників. Більш складні конфліктні ситуації обумовлені наявністю об'єднань та груп, інтереси яких не співпадають: визначення рівня заробітної плати об'єднанням профспілок і підприємців та ін.

Конфлікт може виникнути також через розбіжності цілей, які віддзеркалюють не лише несполучні інтереси різних осіб або сторін, а також різноманітні інтереси однієї і тієї ж особи. Наприклад, планування економічної політики на певний період вимагає узгодження протилежних і несполучних вимог: зростання обсягів виробництва, збільшення прибутків, покращання екології і т. ін. Протидія зацікавленій стороні може бути не лише наслідком усвідомлених дій, а й результатом об'єктивно існуючих, але непередбачених в повному обсязі умов, наприклад, цілеспрямованих дій по одержанню високого врожаю, природних умов, які не завжди цьому сприяють. Багато таких ситуацій можна зустріти в різних сферах діяльності: економіці, господарюванні, біології, соціології, військовій справі і т. ін. Моделювання таких

ситуацій прийнято називати "гра з природою". Характерною особливістю відповідних моделей є розробка математичного апарата з прийняття рішень в умовах так званої природної невизначеності.

Прикладами гри є також різноманітні карточні та спортивні ігри, доміно тощо.

Будь-яка математична модель перебігу економічного, соціального та інших процесів має адекватно віддзеркалювати притаманні їм особливості конфлікту та методи його вирішення:

а) перелік зацікавлених сторін, які будемо називати гравцями; в літературі з теорії гри користуються й іншими назвами: сторони, учасники і т. ін.;

б) перелік можливих дій усіх учасників гри в залежності від ситуації; кожна можлива дія називається ходом або стратегією;

в) інтереси сторін, представлені певною мірою; залежності таких інтересів від ситуації описуються так званими функціями виграшу.

Віддзеркалення змісту конфлікту в аналітичних залежностях створює математичну модель, яку називають грою.

14.2. Класифікація моделей ігор

Велике розмаїття реальних конфліктних ситуацій обумовлює побудову різних моделей ігор. Моделі ігор класифікуються за кількістю гравців, визначеністю стратегій, умовами взаємовідносин між гравцями, характером моделювання наслідків гри, способом визначення функцій виграшу і т. ін. Розглянемо деякі класи ігор.

У залежності від кількості гравців маємо ігри з двома, трьома і більше гравцями. Різні теорії оптимізації можна тлумачити як варіанти гри одного гравця за умови відсутності протидії.

За кількістю можливих стратегій ігри поділяються на скінчéнні та нескінчéнні. Гра називається скінченною, якщо вона має скінченне число кроків (ходів) за умови, що на кожному з них виконується вибір із скінченного числа можливостей зробити певний хід (прийняти певне рішення). У протилежному випадку гра називається нескінченною за умови, що хоч один із гравців має нескінченну кількість вибору можливих стратегій. Наприклад, модель гри "покупець-продавець": кожний з учасників може назвати і будь-яку ціну з певного інтервалу, і кількість товару.

У залежності від кількості дій, які необхідно виконати для досягнення певного результату, ігри поділяються на одноходові (однокрокові) та багатоходові. Прикладом гри першого виду є гра в "орлянку" (після кожного кидання монети можна оцінювати результат), другого виду – гра в шахи або шашки.

За критерієм взаємовідносин між гравцями ігри поділяються на безкоаліційні, кооперативні та коаліційні. У *безкоаліційній* грі учасники не мають або можливостей, або права організовувати коаліції. Якщо гравці можуть організовувати коаліції, домовлятися про певні спільні дії, то за таких умов гра називається *коаліційною*. Гра називається *кооперативною*, якщо до її початку гравці створюють коаліції та домовляються про спільні дії. Усі антагоністичні ігри є прикладом некооперативних.

Важливою характеристикою класифікації гри є властивості функції виграшу (функції платежів). У теорії гри досить ґрунтовно досліджені ситуації, коли виграш одного гравця чисельно дорівнює програшу другого учасника, тобто наявний безпосередній конфлікт між гравцями. Такі ігри утворюють клас ігор *з нульовою сумою*. Прикладом є гра в "орлянку". До названого класу належать моделі дослідження багатьох військових проблем. Протилежністю ігор цього класу є ігри зі *сталою різницею*, в яких гравці як виграють, так і програють одночасно, а тому їм доцільно та корисно об'єднуватися для спільних дій. Між цими крайніми класами знаходяться ігри з *ненульовою сумою*, коли маємо моделювати як конфлікти, так і можливі узгодження дій гравців. Прикладом таких ситуацій є моделювання економічних стосунків між державами, якщо відкидаються політичні амбіції.

14.3. Математичне моделювання конфліктних ситуацій

Множину гравців позначимо I; елементи цієї множини будемо визначати їх переліком, наприклад, $I = \{1; 2\}$ за гру в "орлянку" або в шахи. Множину допустимих стратегій (ходів) i-го гравця позначимо X_i. За гру в "орлянку" множина $X_i = \{$ціна, герб$\}$, бо кожний гравець може скористатися лише цими стратегіями. При моделюванні взаємовідносин "продавець-покупець" кожний з учасників може вибрати будь-яку довільну ціну товару з деякого інтервалу, тобто множину X_i можна характеризувати як $p_i \in (\alpha_i, \beta_i)$, де p_i – ціна товару, α_i, β_i – межі можливих значень ціни.

Надалі розглядаємо моделювання гри двох сторін, як умовно називаємо двох гравців. Позначаємо множину стратегій першого гравця X, другого – Y. Будь-яка пара x_i, y_j, $x_i \in X$, $y_j \in Y$ називається *ситуацією*. Міра зацікавленості гравців у ситуаціях характеризується певними числами. Число, пов'язане з певною ситуацією, називається *виграшем гравця*. Закон відповідності між набором можливих ситуацій певної гри та виграшем конкретного гравця називається функцією виграшу або *функцією платежів*. За умови, що множини X та Y скінченні, функцію виграшу кожного з гравців зручно записати як матрицю виграшів H, в котрій рядки відповідають стратегіям того з гравців, для якого будують *матрицю виграшів*, а стовпці – стратегіям його напарника. Елементи h_{ij} так побудованої матриці визначають виграші гравця за умови певної ситуації, обумовленої i-ою стратегією першого гравця та j-ою стратегією його напарника. Така форма запису функції виграшу скінченних ігор двох гравців обумовлює назву таких ігор – *матричні*.

Приклад. Гра в "орлянку" побудована за такою схемою: кожний із гравців має по дві стратегії – назвати "герб", або "ціна". Якщо обидва гравці обирають то-

тожні стратегії, тобто або "герб" або "ціна", то перший гравець має виграш обсягом h, а його напарник такого ж обсягу програш. У ситуації, коли гравці вибирають різні стратегії, то перший має програш того ж обсягу h, а другий – виграш h. Матриця $H(1)$ виграшів першого гравця буде записана за такою схемою:

$$\begin{array}{cc} & \textit{Стратегії другого гравця} \\ & \begin{array}{cc} \textit{герб} & \textit{ціна} \end{array} \\ \begin{array}{r} \textit{Стратегії} \\ \textit{першого гравця} \end{array} & \begin{array}{l} \left\{\begin{array}{l} \textit{герб} \\ \textit{ціна} \end{array}\right. \end{array} \begin{bmatrix} h & -h \\ -h & h \end{bmatrix} \end{array}$$

Матриця $H(2)$ виграшів другого гравця відповідно буде записана так:

$$\begin{array}{cc} & \textit{Стратегії першого гравця} \\ & \begin{array}{cc} \textit{герб} & \textit{ціна} \end{array} \\ \begin{array}{r} \textit{Стратегії} \\ \textit{другого гравця} \end{array} & \begin{array}{l} \left\{\begin{array}{l} \textit{герб} \\ \textit{ціна} \end{array}\right. \end{array} \begin{bmatrix} -h & h \\ h & -h \end{bmatrix} \end{array}$$

У наведеному прикладі сума виграшів за будь-якої ситуації дорівнює нулю. У загальному випадку ігри з нульовою сумою моделюють гранично антагоністичні зіткнення інтересів в економіці, у військовій справі і т. ін. Матриці виграшів двох гравців у таких іграх пов'язані залежністю: $H(1) + H(2) = 0$.

Іноді має сенс сумістити матриці виграшів обох гравців, тоді в клітині відповідної ситуації зліва записують виграші першого гравця, справа – другого гравця:

$$\begin{array}{cc} & \textit{Стратегії другого гравця} \\ & \begin{array}{cc} \textit{герб} & \textit{ціна} \end{array} \\ \begin{array}{r} \textit{Стратегії} \\ \textit{першого гравця} \end{array} & \begin{array}{l} \left\{\begin{array}{l} \textit{герб} \\ \textit{ціна} \end{array}\right. \end{array} \begin{bmatrix} h;-h & -h;\ h \\ -h;\ h & h;-h \end{bmatrix} \end{array}$$

У більш загальному випадку при скінченній множині стратегій у кожного з гравців функція виграшів для першого гравця буде представлена прямокутною матрицею $m \times n$ (m – кількість стратегій першого гравця, n – кількість стратегій другого гравця), функція виграшів другого гравця – матрицею $n \times m$.

Наведемо приклад з нескінченною множиною стратегій у грі. Маємо дві фірми, які продають однаковий товар. Кожна з них може призначити ціну одиниці товару відповідно P_1 або P_2. Природно, що за умови однакової якості товарів споживачі купують товар з меншою ціною або купують порівну товар кожної фірми, якщо ціни рівні. Приймемо, що функція обсягу попиту $d(P)$ залежить від ціни на товар, тоді функція виграшу H_1 для першої фірми буде записана так:

$$H_1(P_1, P_2) = \begin{cases} P_1 d(P_1), & \textit{якщо}\ \ P_1 < P_2; \\ P_1 d(P_1)/2, & \textit{якщо}\ \ P_1 = P_2; \\ 0, & \textit{якщо}\ \ P_1 > P_2. \end{cases}$$

Аналогічно буде записана функція виграшу для другої фірми.

14.4. Розв'язання матричних ігор у чистих стратегіях

Розглянемо розв'язання задачі теорії гри за умови, що маємо гру двох сторін і можливі стратегії кожної сторони представлені скінченними множинами. Як було зазначено раніше, функцію виграшів можна представити прямокутною матрицею. Приймемо умову, що виграш першого гравця дорівнює програшу другого гравця, тобто маємо *гру з нульовою сумою.*

У теорії гри використовують основне природне припущення: *кожний гравець намагається досягти найбільшого виграшу* за будь-яких дій партнера, котрий необов'язково може бути суперником. Метою теоретичних побудов теорії гри є пошук оптимальних стратегій для кожного з гравців. *Оптимальною стратегією* гравця називають таку стратегію, яка гарантує йому найбільш можливий виграш за будь-яких дій його партнера, тобто такий виграш, величину якого в грі ніяк не можна зменшити. Поняття оптимальної стратегії визначається за такими критеріями.

1. Кожний гравець має партнера не менш досвідченого, ніж він сам. Партнер має протилежну мету, тобто не враховуються можливі помилки або прорахунки кожного з гравців та елементи ризику.

2. Кожний гравець робить виважені дії, намагаючись досягти гарантованого виграшу за будь-яких можливих ситуацій (протидій партнера).

Тепер розглянемо зміст поняття гарантованого рівня виграшу. Оскільки розглядаємо матричну гру з нульовою сумою, то перший гравець намагається (зацікавлений) вибрати таку стратегію (що відповідає вибору певного рядка матриці платежів), за якої величина функції платежів h_{ij} досягає найбільш можливого значення. Партнер діє так, щоб ця величина досягла якомога меншого значення (вибір відповідного стовпця матриці платежів).

Але значення h_{ij} залежить від вибору номера як рядка, так і стовпця. Цей вибір виконується різними учасниками гри з протилежними цілями. Перший гравець, обравши i-ту стратегію, може досягти лише того, що його партнер дозволить йому мати найменший виграш серед усіх можливих за обраної i-ої стратегії. Нижнім гарантованим рівнем виграшу α_i першого гравця буде величина

$$\alpha_i = \min_j h_{ij} \quad (j=1, 2, \ldots, n).$$

Природно, що перший гравець буде вибирати таку i-ту стратегію, за якої він *напевне* одержить якийсь максимально можливий виграш, тобто вибір стратегії першого гравця визначається умовою :

$$\max_i \alpha_i = \max_i \min_j h_{ij} \quad (i = 1, 2, \ldots, m).$$

Найбільший гарантований виграш $h(1)$ першого гравця визначається так:

$$h(1) = \alpha = \max_i \alpha_i = \max_i \min_j h_{ij} \quad (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n). \qquad (1)$$

Оптимальною стратегією першого гравця буде та, для якої виконується умова (1), тобто стратегія, за якої нижній гарантований рівень досягає найбільшого значення за наявної матриці виграшів.

Протилежна сторона моделі конфліктної ситуації – другий гравець, враховуючи, що гра є з нульовою сумою, прагне, щоб його програш був якнайменший, отже, його оптимальною стратегією буде вибір такого стовпця матриці платежів, для якого величина платежу h_2 (величина його програшу) буде найменшою серед усіх найбільших платежів, яких прагне досягти перший гравець.

$$h(2) = \beta = \min_j \beta_j = \min_j \max_i h_{ij} \quad (i = 1, 2, \ldots, m; j = 1, 2, \ldots, n). \qquad (2)$$

Оптимальна стратегія першого гравця називається *максимінною*. За такої стратегії його виграш при будь-якій протидії партнера буде не менше величини $\alpha = \max_i \min_j h_{ij}$, яку називають *нижньою ціною гри*. Оптимальна стратегія другого гравця називається *мінімаксною*, а величина його програшу за такої стратегії визначається величиною $\beta = \min_j \max_i h_{ij}$ – *верхньою ціною гри*. Доцільно наголосити, що величини $h(1)$, α, $h(2)$ та β мають різний зміст і, як буде показано далі, не завжди співпадають чисельно.

У випадку, коли величини нижньої та верхньої ціни гри співпадають, таку величину називають *чистою ціною гри*, або *ціною гри;* за таких умов величина виграшу першого гравця дорівнює величині програшу другого гравця. *Чистою стратегією гравця* називають вибір мінімаксної стратегії в матриці платежів. Отже, коли нижня та верхня границі гри співпадають, то задача має розв'язок у чистих стратегіях, гра називається *грою з сідловою точкою*. Сідловою точкою називають пару чистих стратегій, за яких виграш дорівнює ціні гри. Матриця виграшів і сідлова точка схематично представлені на рис. 14.1.

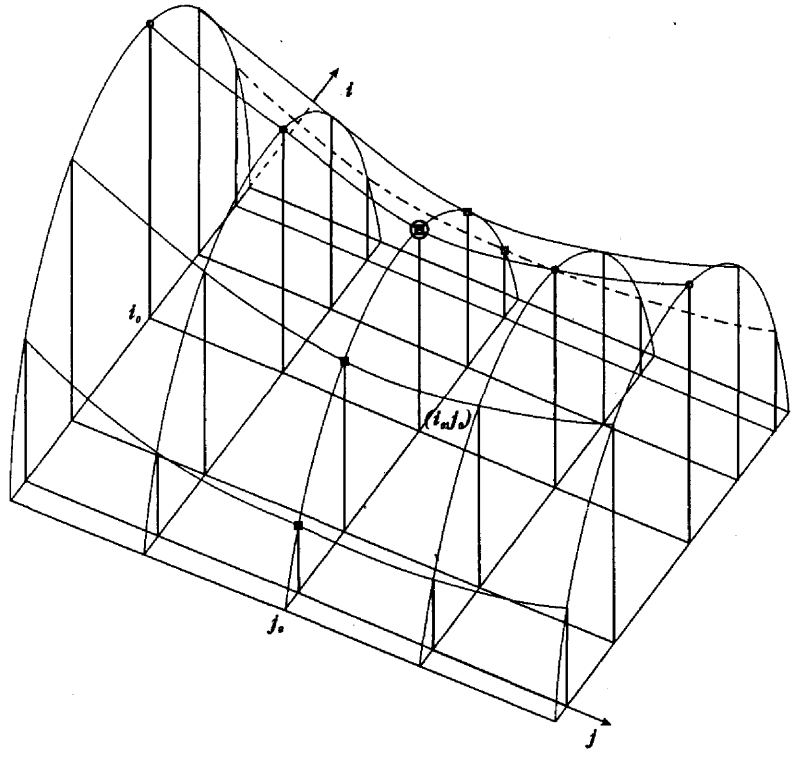


Рис. 14.1.

Сідлова точка в розв'язанні задачі має таке значення: якщо один із гравців обрав стратегію, яка відповідає сідловій точці, то у партнера немає кращого вибору, ніж вибір тієї своєї стратегії, яка відповідає сідловій точці. Для стратегій $i_0 j_0$, які породжують сідлову точку, справедливе відношення:

$$h_{ij_0} \le h_{i_0 j_0} \le h_{i_0 j} \qquad (3)$$

(i, j – будь-які чисті стратегії гравців).

Згідно з (3) елемент $h_{i_0 j_0}$ матриці платежів є найменшим в i_0-му рядку та найбільшим в j_0-му стовпці (не більше будь-якого елемента i_0-го рядка та не менше будь-якого елемента j_0-го стовпця).

Пошук сідлової точки виконується так. У матриці платежів по черзі (послідовно) в кожному рядку знаходять мінімальний елемент. Якщо такий елемент до того ж буде максимальним для свого стовпця, то він називається *сідловим елементом*, а стратегії, що його породжують, визначають сідлову точку. Чисті стратегії, які визначають сідлову точку, називаються *оптимальними чистими стратегіями*. Сенс *гри з нульовою сумою* – у визначенні оптимальних чистих стратегій та відповідного елемента $h_{i_0 j_0}$ матриці платежів, якщо це можна виконати за відомої матриці платежів.

Зауваження. Гра (матриця платежів) може мати кілька сідлових точок.

Приклад. Знайти розв'язок гри, визначеної такою матрицею платежів (табл. 14.1).

Таблиця 14.1.

A_i \ B_j	1	2	3	4	$\alpha_i(min)$
1	1	6	2	–4	–4
2	2	3	4	2	2
3	–1	7	-5	1	–5
$\beta_j(max)$	2	7	4	2	

Розглядаємо по черзі рядки матриці та знаходимо в кожному рядку мінімальний елемент α_i, записуємо його величини в крайній правий стовпець табл. 14. 1. Для першого рядка найменший елемент $\alpha_1 = h_{1,4} = -4$ не може бути сідловим, бо він не є найбільшим у четвертому стовпці.

Для другого рядка мінімальне значення виграшу $\alpha_2 = 2$ може бути досягнуто за умови, що партнер вибере або першу, або четверту стратегію, але величина 2 є найбільшою для першого та четвертого стовпців, а це означає, що елементи $h_{2,1}$ та $h_{2,4}$ – сідлові і створюють розв'язок:

– оптимальна стратегія першого гравця – друга; партнер має за цієї умови дві оптимальні стратегії - першу та четверту;

– ціна модельованої гри дорівнює 2.

Проаналізувавши третій рядок матриці виграшів табл. 14.1, необхідно переконатися, що цей рядок не може мати сідлових елементів. Отже, гра з представленою в табл. 14.1 матрицею виграшів має дві сідлові точки в чистих стратегіях $h_{2,1}$ та $h_{2,4}$; оптимальні стратегії названі вище.

Досвід моделювання скінченних антагоністичних ігор показує, що розв'язок у чистих стратегіях – це виняток, а не закономірність. Розглянемо моделі ігор при відсутності сідлових точок.

14.5. Розв'язання матричних ігор у змішаних стратегіях

Проаналізуємо можливий перебіг гри з матрицею платежів

$$H = \begin{bmatrix} 8 & 1 & 5 \\ 3 & 9 & 7 \end{bmatrix}, \qquad (4)$$

враховуючи той факт, що за відомої матриці платежів невідомо, які стратегії оберуть гравці. Виконаємо оцінку можливого виграшу (програшу) гравців. Якщо гравець A вибере першу стратегію, сподіваючись, що найгірше буде, якщо гравець B вибере свою другу стратегію, за якої його програш дорівнюватиме 1, то відповідно виграш гравця A теж дорівнюватиме 1. Але гравець B, аналізуючи матрицю H, очікує від гравця A вибору другої стратегії, бо за такого вибору гравець A матиме щонайменше виграш 3. Якщо гравець B вибере першу стратегію, за інших стратегій гравця B виграш A буде лише збільшений. Для гравця B вибір першої стратегії ризикований, бо якщо гравець A уже вибрав свою першу стратегію, то гравець B програє 8.

Питання вибору гравцями своїх найкращих стратегій за алгоритмом, описаним в 14. 3, не може бути вирішене однозначно. Іншими словами, модель гри за наведеної матриці H не має розв'язків у чистих стратегіях. У такій ситуації стає важливим, щоб супротивник не вгадав, яка стратегія буде використана.

У загальному випадку в іграх, які моделюють конфліктні ситуації, нижня та верхня ціни гри, як правило, не співпадають; розв'язок таких задач не може бути знайдений у чистих стратегіях. Виникає питання: чи існує розв'язок за таких умов? як його знайти, якщо він є?

Якщо гравець A, виконавши аналіз матриці платежів, скористається своєю максимінною стратегією, то він матиме виграш не менше α (нижньої ціни гри) (1). Гравець B за своєї мінімаксної стратегії буде мати програш не більший β (верхньої ціни гри) (2). Проміжок $[\alpha; \beta]$ – це область невизначеності, в межах якої гравець A має певну ймовірність збільшити свій виграш, а гравець B – зменшити програш. Тобто за наведеної матриці виграшів (4) немає можливості однозначно визначити кожному гравцеві свої стратегії.

Якщо досліджувана гра одноразова, то теоретично гравець A не має можливості збільшити виграш, якщо не ризикувати. Зовсім іншою буде ситуація, коли гра повторюється багаторазово; за цих умов можна збільшити середній гарантований виграш гравця A за рахунок певної випадкової комбінації чистих стратегій.

Вектор, компоненти якого показують ймовірності доцільного використання гравцем своїх чистих стратегій у багаторазовій грі, називається змішаною стратегією гравця.

Позначимо в загальному випадку змішані стратегії першого гравця

$\overline{p} = (p_1, p_2, \ldots, p_m)$, другого гравця $\overline{q} = (q_1, q_2, \ldots, q_n)$, де p_i $(i = 1, 2, \ldots, m)$, q_j $(j = 1, 2, \ldots, n)$ – ймовірності використання відповідними гравцями своїх чистих стратегій у багаторазовій грі. Оскільки p_i та q_j, за означенням, є ймовірностями, то вони задовольняють умовам:

$$p_i \geq 0; \quad \sum_{i=1}^{m} p_i = 1;$$

$$q_j \geq 0; \quad \sum_{j=1}^{n} q_j = 1. \tag{5}$$

За такого означення змішаних стратегій чисті стратегії i та j гравців можна тлумачити як змішані стратегії, в яких усі компоненти векторів $\overline{p}$ і $\overline{q}$ дорівнюють нулю, за винятком компонент на i-му та j-му місцях, які дорівнюють 1. Тобто якщо використовується i-та чиста стратегія, то $\overline{p} = (0, 0, \ldots, p_i = 1; \ldots, 0)$, відповідно $\overline{q} = (0, 0, \ldots, q_j = 1; \ldots, 0)$.

Оскільки гра антагоністична, то природно допустити, що використання гравцями своїх стратегій незалежне, тому ймовірність події, що перший гравець вибере i-ту стратегію, а другий – j-ту, дорівнює $p_i q_j$.

Функція змішаних стратегій

$$F(\overline{p}, \overline{q}) = \sum_{i=1}^{m} \sum_{j=1}^{n} p_i h_{ij} q_j \tag{6}$$

називається *функцією платежів гри з матрицею* $[h_{ij}]_{m \times n}$. Отже, $F(\overline{p}, \overline{q})$ є функцією змінних p_i та q_j.

Функція платежів визначає величину математичного сподівання виграшу $H(1)$ гравця A (його "середнього виграшу" за відомих змішаних стратегій обох гравців).

Залежність (6) у матричній формі можна записати так:

$$F(\overline{p}, \overline{q}) = \overline{p} H \overline{q}. \tag{6a}$$

На множині своїх змішаних стратегій перший гравець, прагнучи досягти найбільшого значення величини середнього виграшу, мусить так вибрати вектор $\overline{p}$ (ймовірності використання певних чистих стратегій), щоб максимум мінімальних значень середнього виграшу $H(1)$ був найбільшим за будь-якої протидії суперника, тобто:

$$H(1) = \max_{\overline{p}} \min_{\overline{q}} \overline{p} H \overline{q}.$$

Стратегія $\overline{p}^{\,0}$, за якої величина $H(1)$ досягає найбільшого значення, називається максимінною, або оптимальною для гравця A.

Аналогічно метою другого гравця є досягнення ситуації, коли максимальний програш $H(2)$ досягає мінімуму, тобто:

$$H(2) = \min_{\overline{q}} \max_{\overline{p}} \overline{p} H \overline{q}.$$

Стратегія $\overline{q}^{\,0}$, за якої величина $H(2)$ досягає свого мінімального значення, називається мінімаксною, або оптимальною для гравця B.

Для оптимальних стратегій $\overline{p}^{\,0}$ і $\overline{q}^{\,0}$ виконується умова:

$$H(2)(\overline{p}, \overline{q}^{\,0}) \leq F(\overline{p}^{\,0}, \overline{q}^{\,0}) \leq H(1)(\overline{p}^{\,0}, \overline{q}). \tag{7}$$

Залежність (7) показує, що використання у грі оптимальних змішаних стратегій гарантує першому гравцеві виграш, не менший величини $F(\overline{p}^{\,0}, \overline{q}^{\,0})$, а другому – програш, не більший величини $F(\overline{p}^{\,0}, \overline{q}^{\,0})$. Величина $V = F(\overline{p}^{\,0}, \overline{q}^{\,0})$ називається ціною гри.

Залежність (7) можна тлумачити і так:

а) жодна стратегія першого гравця не забезпечить йому виграшу більшого, ніж ціна гри, якщо другий гравець використає свою мінімаксну стратегію; середній очікуваний виграш першого гравця за названих умов обчислюється як $H(1) = \overline{p} H \overline{q}^{\,0}$;

б) за будь-якої стратегії другого гравця його програш не може бути меншим, ніж ціна гри, якщо перший гравець використає свою максимінну стратегію; середній очікуваний програш другого гравця за названих умов обчислюється як $H(2) = \overline{p}^{\,0} H \overline{q}$.

Сформулюємо основну теорему теорії гри.

Теорема Неймана (без доведення). *У змішаних стратегіях будь-яка скінчéнна матрична гра має хоча б одну сідлову точку.*

Теорема Неймана гарантує, що за будь-якої скінченної матричної гри існує хоча б одна пара оптимальних (максимінної та мінімаксної) змішаних стратегій, до того ж за таких стратегій виконується умова:

$$H(1) = V = H(2).$$

Якщо $V = F(\overline{p}^{\,0}, \overline{q}^{\,0}) = 0$, тобто ціна гри дорівнює нулю, то гра називається *справедливою*.

Розглянемо деякі особливості розв'язання матричних ігор. Наголосимо, що в оптимальній змішаній стратегії деякі можливі чисті стратегії гравців можуть бути і не використаними за умови, що ймовірності таких чистих стратегій дорівнюють нулю.

Активними називаються такі чисті стратегії гравця, які входять в оптимальну стратегію з ймовірностями, відмінними від нуля.

Пошук оптимальних стратегій суттєво ускладнюється, якщо розміри матриці платежів зростають. Розглянемо умови, які дозволяють зменшити розмірність матриці платежів, не порушуючи розв'язку задачі.

Ряди (рядки або стовпці) матриці платежів називаються *дублюючими*, якщо їх елементи тотожні. Для розрахунків залишають лише один з дубльованих рядів.

У теорії матричних ігор доведено, що коли елементи одного рядка не менше відповідних елементів другого рядка, то в оптимальну змішану стратегію можуть бути введені лише стратегії домінуючих рядків, тобто рядки, над якими є домінуючі, можна вилучити.

Для гравця, який контролює величину свого можливого програшу, напевно, недоцільними є ті стратегії, величини програшів за яких більші, ніж за його інших стратегій і за будь-яких стратегій гравця *A*. Тому, якщо в матриці платежів є стовпці, елементи яких не менші елементів інших стовпців, такі стовпці доцільно опустити (гравцеві *B*, напевно, недоцільно користуватися відповідними стратегіями).

Ймовірностями чистих стратегій, номери яких відповідають номерам опущених рядів матриці платежів, необхідно дати нульові значення.

Приклад. Маємо матрицю платежів

$$H = \begin{bmatrix} 4 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 6 & 4 \end{bmatrix},$$

у якій третій рядок домінує над другим. Вилучимо другий рядок:

$$\begin{bmatrix} 4 & 3 & 2 \\ 3 & 6 & 4 \end{bmatrix}.$$

В одержаній матриці другий стовпець домінує над третім. Вилучивши другий стовпець, одержуємо матрицю платежів

$$\begin{bmatrix} 4 & 2 \\ 3 & 4 \end{bmatrix}.$$

Якщо знайти розв'язок гри для одержаної матриці, то його легко використати для побудови розв'язку гри з вихідною матрицею *H*, присвоївши номерам стратегій, які відповідають номерам вилучених рядів, нульові значення.

Другий спосіб спрощення матриці виграшів ґрунтується на доведеній у теорії гри властивості, за якою *афінне перетворення* матриці платежів (тобто перетворення (перерахунок) усіх елементів матриці *H* за правилом $h^H_{ij} = ah_{ij} + b$, де $a \neq 0$) не змінює розв'язків гри, до того ж ціна V^H гри з перетвореною матрицею дорівнює $aV + b$. Наведена властивість означає, що для побудови матриці платежів не має значення, в яких одиницях вимірюються виграші (програші), а додавання (віднімання) певної величини *b* на таку ж величину змінить виграш (програш), не змінивши розв'язок гри.

Нагадаємо, що під розв'язком гри розуміють упорядковану сукупність ймовірностей використання певних чистих стратегій. Величина виграшу (програшу) буде залежати від одиниці виміру елементів матриці платежів.

Афінне перетворення матриці платежів доцільно використати для її спрощення: якщо в клітинах матриці є дроби зі спільним знаменником, то всю матрицю можна помножити на деяку сталу величину, щоб одержати цілі числа; якщо багато елементів матриці є тотожними, то їх величину можна відняти від елементів усієї матриці, щоб мати нулі на відповідних позиціях. До того ж будь-яку антагоністичну гру можна перевести у клас справедливих – для цього необхідно від кожного елемента вихідної матриці платежів відняти ціну гри. Але слід пам'ятати, що, знайшовши розв'язок, треба повернутися до змісту вихідної задачі.

Приклад. Маємо закласти сад, користуючись двома видами фруктових дерев, за умови, що щільність дерев на одиниці площі однакова, а чистий зиск через певний час від кожного дерева на одиницю витрат щорічно визначатиметься в залежності від умов експлуатації (табл. 14.2). Під несприятливими умовами слід розуміти не лише природні, а й рівень догляду в залежності від матеріально-технічних можливостей. Несприятливі умови будемо розглядати як протидіючого гравця, наголосивши, що несприятливі умови для одного виду дерев можуть бути сприятливими для іншого.

Таблиця 14.2.

Вид дерева \ Умови	Сприятливі	Несприятливі
1	0,7	0,4
2	0,3	0,6

Випишемо матрицю виграшів:

$$H = \begin{bmatrix} 0{,}7 & 0{,}4 \\ 0{,}3 & 0{,}6 \end{bmatrix}.$$

Скориставшись можливостями афінних перетворень, помножимо всі елементи матриці на 10 і запишемо у вигляді:

$$H = \begin{bmatrix} 7 & 4 \\ 3 & 6 \end{bmatrix}.$$

Дослідимо задачу на пошук розв'язку в чистих стратегіях, тобто доцільності скористатися лише одним видом дерев. Величина $H(1) = \alpha = \max\limits_i \min\limits_j h_{ij} = 4$, величина $H(2) = \beta = \min\limits_j \max\limits_i h_{ij} = 6$. Нижня та верхня ціни гри не співпадають, розв'язку в чистих стратегіях немає. Отже, розв'язок необхідно шукати у змішаних стратегіях, тобто шукати оптимальне співвідношення між кількостями дерев першого та другого видів, розглядаючи таке співвідношення як випадкову величину. Садіння дерев першого виду будемо вважати за першу стратегію, дерев другого виду – за другу стратегію. Позначимо p випадкову частину дерев першого виду, тоді частка дерев другого виду становитиме $q = 1 - p$, а середній очікуваний зиск H_{1C} (математичне сподівання величини зиску) за сприятливих умов дорівнюватиме:

$$H_{1C} = 7p + 3(1 - p) = 4p + 3; \qquad (8)$$

середній очікуваний зиск за несприятливих умов H_{1H}:

$$H_{1H} = 4p + 6(1 - p) = -2p + 6. \qquad (9)$$

Виберемо прямокутну систему координат (p, H_1) та побудуємо в ній графіки (8), (9), обравши певний масштаб.

Кожний із графіків буде зображений відрізками прямої, які перетинаються в точці T. Позначимо відрізки $A_1A'_1$ та $A_2A'_2$. Якщо перший гравець (садівник) вибере певну величину p (частку дерев першого виду), то за сприятливих умов (перша стратегія "партнера") його очікуваний середній виграш (зиск) буде зображений точкою H_{1C} перетину прямої p з відрізком $A_1A'_1$, за несприятливих умов – точкою перетину H_{1H} цієї ж прямої з відрізком $A_2A'_2$. Проміжні значення величини середнього очікуваного зиску розташовані між точками H_{1C} та H_{1H}. На ці значення можна сподіватися, якщо "партнер" буде користуватися змішаними стратегіями (певне співвідношення сприятливих і несприятливих умов). Але перший гравець (садівник) прагне мати найбільший, але гарантований зиск, величина якого на рис 14.2 досягається в точці T за умови, що $p = 0{,}5$. За інших величин p найбільший гарантований зиск графічно буде зображатися точками відрізку $A_1TA'_2$. Практичний висновок: за даних умов задачі доцільно висадити порівну дерева першого та другого видів.

Найбільший гарантований зиск за цих умов буде при $p = 0{,}5$. Пропонуємо довести, що поставлена задача дає розв'язок $p = 0{,}75$ за умови, що матриця H має вигляд:

$$H = \begin{bmatrix} 5 & 4 \\ 3 & 6 \end{bmatrix}.$$

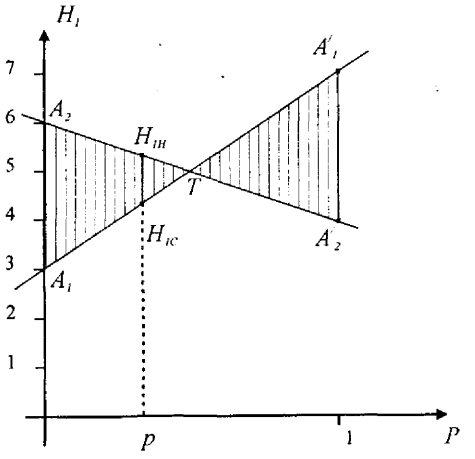


Рис. 14.2.

Слід дати тлумачення розв'язку.

14.6. Приведення матричної гри до задачі лінійного програмування

Для побудови алгоритмів розв'язання задач матричних ігор використовується властивість оптимальних змішаних стратегій: *оптимальна змішана стратегія першого гравця гарантує йому виграш не менший ціни гри за будь-яких стратегій другого гравця і рівний ціні гри за оптимальної стратегії другого гравця*. Використання названої властивості матричних ігор дає можливість привести задачу пошуку оптимальних змішаних стратегій до задачі лінійного програмування. Розглянемо одну з методик побудови відповідного алгоритму.

Маємо матричну гру за відомої матриці платежів $H = \{h_{ij}\}$ ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$). Перший гравець має чисті стратегії $A_1, A_2, \ldots, A_m$, другий – $B_1, B_2, \ldots, B_n$. Необхідно знайти оптимальні змішані стратегії $S_A^0 = (p_1^0, p_2^0, \ldots, p_m^0)$ та $S_B^0 = (q_1^0, q_2^0, \ldots, q_m^0)$, де p_i^0 та q_j^0 – ймовірності використання відповідних чистих стратегій гравцями. Використовуючи можливість афінних перетворень матриці платежів, без застережень можна вважати, що $h_{ij} \geq 0$ і $V > 0$.

Якщо перший гравець використовує свою оптимальну змішану стратегію $S_A^0 = (p_1^0, p_2^0, \ldots, p_m^0)$ проти будь-якої чистої стратегії B_j другого гравця, то його очікуваний гарантований виграш, тобто математичне сподівання виграшу,

$H_j = \sum_{i=1}^{m} p_i^0 h_{ij}$ $(j = 1, 2, \ldots, n)$. Наголосимо, що значення $\{p_i^0\}$ поки що невідомі. За оптимальної стратегії $S_A^0 = (p_1^0, p_2^0, \ldots, p_m^0)$ гравця A його середні виграші за будь-якої чистої стратегії супротивника будуть не менше ціни гри V, тому можемо записати і гарантувати виконання системи нерівностей:

$$\begin{cases} h_{11}p_1^0 + h_{21}p_2^0 + \ldots + h_{m1}p_m^0 \geq V, \\ h_{12}p_1^0 + h_{22}p_2^0 + \ldots + h_{m2}p_m^0 \geq V, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ h_{1n}p_1^0 + h_{2n}p_2^0 + \ldots + h_{mn}p_m^0 \geq V. \end{cases} \tag{10}$$

Кожну з наведених нерівностей помножимо на $1/V$ і введемо нові змінні:

$$x_1 = p_1^0/V;\ x_2 = p_2^0/V; \ldots\ldots;\ x_m = p_m^0/V. \tag{11}$$

За нових змінних система (10) запишеться так:

$$\begin{cases} h_{11}x_1 + h_{21}x_2 + \ldots + h_{m1}x_m \geq 1, \\ h_{12}x_1 + h_{22}x_2 + \ldots + h_{m2}x_m \geq 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ h_{1n}x_1 + h_{2n}x_2 + \ldots + h_{mn}x_m \geq 1. \end{cases} \tag{12}$$

Скориставшись умовою $p_1^0 + p_2^0 + \ldots + p_m^0 = 1$ (сума всіх ймовірностей дорівнює ймовірності достовірної події) та (11), маємо: $x_1 + x_2 + \ldots + x_m = 1/V$. Мета першого гравця – максимізувати свій гарантований виграш, тобто ціну гри V. Але максимізація ціни гри V обумовлює мінімізацію величини $1/V$, тому розв'язання задачі пошуку величин $x_1, x_2, \ldots, x_m$ можна сформулювати так: *знайти величини змінних* $x_i \geq 0$ $(i=1, 2, \ldots, m)$ *так, щоб вони задовольняли системі лінійних обмежень* (12) *за умови досягнення найменшого значення лінійної функції:*

$$Z = x_1 + x_2 + \ldots + x_m. \tag{13}$$

Розв'язавши сформульовану задачу лінійного програмування (12), (13) та скориставшись (11), одержуємо оптимальну змішану стратегію для першого гравця $S_A^0 = (p_1^0, p_2^0, \ldots, p_m^0)$ та максимальний гарантований виграш V.

Для розв'язання задачі пошуку оптимальних змішаних стратегій $S_B^0 = (q_1^0, q_2^0, \ldots, q_n^0)$ другого гравця використовуються умови:

а) середній очікуваний програш другого гравця за умови використання ним оптимальної стратегії не перевищить (не буде більшим) ціни гри, які б стратегії не використовував перший гравець;

б) гравець B прагне мінімізувати свій гарантований програш, тобто досягти максимуму величини $1/V$.

За названих умов змінні $q_1^0, q_2^0, \ldots, q_n^0$ задовольняють нерівностям:

$$\begin{cases} h_{11}q_1^0 + h_{21}q_2^0 + \ldots + h_{m1}q_m^0 \geq V, \\ h_{12}q_1^0 + h_{22}q_2^0 + \ldots + h_{m2}q_m^0 \geq V, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ h_{1n}q_1^0 + h_{2n}q_2^0 + \ldots + h_{mn}q_m^0 \geq V. \end{cases} \tag{14}$$

Якщо ввести нові змінні

$$y_j = q_{ij}^0 / V \quad (j = 1, 2, \ldots, n), \tag{15}$$

то система (14) матиме вигляд:

$$\begin{cases} h_{11}y_1 + h_{21}y_2 + \ldots + h_{m1}y_m \geq 1, \\ h_{12}y_1 + h_{22}y_2 + \ldots + h_{m2}y_m \geq 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ h_{1n}y_1 + h_{2n}y_2 + \ldots + h_{mn}y_m \geq 1. \end{cases} \tag{16}$$

До того ж змінні y_j $(j = 1, 2, \ldots, n)$ мають задовольнити умові:

$$y_1 + y_2 + \ldots + y_n = 1/V. \tag{17}$$

Таким чином, розв'язання задачі пошуку оптимальних стратегій для другого гравця приведено до задачі лінійного програмування: *знайти величини змінних* $y_j \geq 0$ $(j = 1, 2, \ldots, n)$ *так, щоб вони задовольняли системі лінійних обмежень* (16) *за умови досягнення найбільшого значення лінійної функції:*

$$Z^* = y_1 + y_2 + \ldots + y_n. \tag{18}$$

Порівнявши задачі лінійного програмування (12), (13) та (16), (18), до яких зведені задачі пошуку оптимальних стратегій задачі матричної гри, робимо висновок, що названі задачі лінійного програмування спряжені. Доцільно наголосити, що

для пошуку оптимальних змішаних стратегій у конкретній моделі задачі гри треба вибирати ту із взаємноспряжених задач, розв'язок якої легше знайти, а розв'язок другої (іншої) знаходити, використовуючи теореми спряженості.

Приклад. Підприємство може випускати чотири види продукції A_1, A_2, A_3, A_4, одержуючи прибуток у залежності від попиту, який умовно може бути визначений трьома різними станами B_1, B_2, B_3. Побудуємо матрицю виграшів H, елементи якої h_{ij} визначають прибуток підприємства за умови випуску i-ої продукції при j-му попиті на неї.

$$H = \begin{array}{c} \\ A_1 \\ A_2 \\ A_3 \\ A_4 \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & B_2 & B_3 \end{array} \\ \begin{bmatrix} 3 & 6 & 8 \\ 9 & 4 & 2 \\ 7 & 5 & 4 \\ 4 & 4 & 3 \end{bmatrix} \end{array} \qquad (19)$$

Визначити оптимальні пропорції у виробництві продукції, які б гарантували деяку середню величину прибутку за будь-якого попиту, вважаючи його невизначеним.

Проаналізувавши матриці, доходимо висновку, що третій рядок домінує над четвертим, тому для подальшого четвертий рядок вилучимо. Економічно це означає недоцільність виробництва продукції четвертого виду за даних умов. Розв'язання задачі побудуємо як дослідження моделі гри виробника з невідомими умовами попиту.

Визначимо верхню та нижню ціни гри. Розрахунки наведені в табл. 14.3.

Таблиця 14.3.

Попит / Вид продукції	B_1	B_3	B_4	$\alpha_i = \min\limits_j h_{ij}$
A_1	3	6	8	3
A_2	9	4	2	2
A_3	7	5	4	4
$\beta_j = \max\limits_i h_{ij}$	9	6	8	4 / 6

Оскільки $\max\limits_i \min\limits_j h_{ij} = 4 \neq \min\limits_j \max\limits_i h_{ij} = 6$, то задача не має розв'язку в чистих стратегіях і оптимальний розв'язок будемо шукати у змішаних стратегіях:

$$S_A{}^0 = (p_1{}^0, p_2{}^0, p_3{}^0), \quad S_B{}^0 = (q_1{}^0, q_2{}^0, q_3{}^0).$$

Введемо нові змінні $x_i = p_i{}^0/V$ $(i = 1, 2, 3)$ та $y_j = q_j{}^0/V$ $(j = 1, 2, 3)$. Використовуючи залежності (12), (13) і (16), (18), запишемо дві спряжені задачі лінійного програмування:

$$\begin{array}{ll} \text{задача 1} & \text{задача 2} \\ 3x_1 + 9x_2 + 7x_3 \geq 1 & 3y_1 + 6y_2 + 8y_3 \leq 1 \\ 6x_1 + 4x_2 + 5x_3 \geq 1 & 9y_1 + 4y_2 + 2y_3 \leq 1 \\ 8x_1 + 2x_2 + 4x_3 \geq 1 & 7y_1 + 5y_2 + 4y_3 \leq 1 \\ x_i \geq 0; \quad i = 1, 2, 3 & y_j \geq 0; \quad j = 1, 2, 3 \\ Z = x_1 + x_2 + x_3 \to \min; & Z^* = y_1 + y_2 + y_3 \to \max. \end{array} \qquad (20)$$

Розв'яжемо симплекс-методом задачу 2, бо для неї легше знайти допустимий базисний розв'язок. Приведемо задачу 2 до стандартного вигляду, скориставшись додатковими змінними:

$$\begin{array}{l} 3y_1 + 6y_2 + 8y_3 + y_4 = 1 \\ 9y_1 + 4y_2 + 2y_3 + y_5 = 1 \\ 7y_1 + 5y_2 + 4y_3 + y_6 = 1 \\ y_j \geq 0; \quad j = 1, 2, 3, 4, 5, 6 \\ Z^* = y_1 + y_2 + y_3 \to \max. \end{array}$$

У табл. 14.4 записані вихідна та симплекс-таблиці задачі 2. Її базисний розв'язок $\overline{Y} = (0; 0; 0; 1; 1; 1)$. Введемо в базис y_2 та вилучимо з базису y_4. Виконавши аналіз симплекс-таблиці першого кроку симплексних перетворень, доходимо висновку, що з базису доцільно вилучити y_6 та ввести y_1.

Оскільки елементи останнього рядка симплекс-таблиці другого кроку перетворень додатні, то оптимальним розв'язком задачі 2 (20) є вектор з такими компонентами $\overline{Y}^0 = (1/27; 4/27; 0; 0; 2/27; 0)$.

Використовуючи теореми про властивості розв'язків спряжених задач, визначимо оптимальний розв'язок задачі 1 (20):

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow & \uparrow\downarrow \\ y_4 & y_5 & y_6 & y_1 & y_2 & y_3 \\ 2/27 & 0 & 1/9 & 0 & 0 & 1/27 \end{array}$$

Отже, оптимальним розв'язком задачі 1 є $\overline{x}^0 = (2/27; 0; 1/9; 0; 0; 1/27)$, а $\min Z = \max Z^* = 5/27$. Скориставшись (17), обчислюємо ціну гри $V = 1/\max Z^* = 1/\min Z = 27/5 = 5{,}4$.

Оптимальну стратегію $S_A^0 = (p_1^0, p_2^0, p_3^0)$ (план випуску продукції) знаходимо, використовуючи (11): $p_i^0 = x_i^0 V$ ($i = 1, 2, 3$). Отже, $p_1^0 = 5{,}4 \times 2/27 = 0{,}4$; $p_2^0 = 5{,}4 \times 0 = 0$; $p_3^0 = 5{,}4 \times 1/9 = 0{,}6$. Таким чином, підприємству доцільно випускати 40% продукції A_1, 60% продукції A_3 і не випускати продукцію A_2.

Знайшовши за аналогією (15) оптимальну стратегію попиту S_B^0, маємо: $q_1^0 = 0{,}2$; $q_2^0 = 0{,}8$; $q_3^0 = 0$. Таким чином, оптимальний попит знаходиться на 20% у стані B_1 та на 80% у стані B_2. Оптимальним результатам можна дати різне тлумачення, виходячи з конкретної ситуації, наприклад, сезонний попит, попит територіальний і т. ін. Треба врахувати й очікуваний попит, якщо необхідно зберігати продукцію.

Таблиця 14.4.

Вихідна симплекс-таблиця					
БЗ		ВЗ			b_{i0}/b_{is}
		$-y_1$	$-y_2$	$-y_3$	
y_4	1	3	6	8	1/6
y_5	1	9	4	2	1/4
y_6	1	7	5	4	1/5
f	0	–1	–1	–1	
Перший крок симплексних перетворень					
БЗ		ВЗ			b_{i0}/b_{is}
		$-y_1$	$-y_4$	$-y_3$	
y_2	1/6	3/6	1/6	8/6	1/3
y_5	1/3	7	–4/6	–10/3	1/27
y_6	1/6	9/2	–5/6	–8/3	1/27
f	1/6	–1/2	1/6	1/3	
Другий крок симплексних перетворень					
БЗ		ВЗ			b_{i0}/b_{is}
		$-y_6$	$-y_4$	$-y_3$	
y_2	4/27	–1/9	7/27	44/27	
y_5	2/27	–14/9	17/27	22/27	
y_1	1/27	2/9	–5/27	–16/27	
f	5/27	1/9	2/27	1/27	

При розв'язанні задач скінченних ігор розмірності $m \times n$ доцільно дотримуватися такої схеми:

1) виключити з вихідної платіжної матриці явно невигідні стратегії;

2) знайти верхню та нижню ціни гри та перевірити, чи має гра сідлову точку. Якщо така точка є, то відповідні їй стратегії будуть оптимальними, гра має розв'язок у чистих стратегіях, а ціна гри дорівнює верхній (нижній) ціні;

3) якщо сідлова точка відсутня, розв'язання гри необхідно шукати у змішаних стратегіях, приводячи, наприклад, до задачі лінійного програмування.

На практиці розв'язок у змішаних стратегіях може бути реалізований порізному. Наприклад, чисті стратегії використовуються в послідовності, заданій відповідними ймовірностями за умови, що гра може повторюватися багато разів. За інших умов виконується певний фізичний розподіл у пропорціях, визначених ймовірностями, як це було показано на прикладах планування саду та виробництва.

14.7. Ігри з ненульовою сумою та кооперативні. Моделювання проблем мікроекономіки з використанням математичного апарата теорії гри

Моделі гри з ненульовою сумою використовуються для дослідження ситуації, коли учасники процесу не лише мають антагоністичні цілі, а, навпаки, можуть і вигравати, і програвати водночас. Якщо в моделі гри з *нульовою сумою* кожному учаснику невигідно було інформувати партнера про свою стратегію, бо це могло призвести до зменшення суми виграшу, то за гри з *ненульовою сумою* часто доцільно партнерам координувати свої дії з метою одержання найбільшого виграшу.

Ігри з ненульовою сумою поділяються на *кооперативні* та *некооперативні*. У некооперативних іграх гравці приймають рішення незалежно один від одного або тому, що рішення не можна узгодити або узгодження заборонено правилами гри. Опишемо коротко зміст основних положень розв'язання задачі некооперативної гри. Один із методів розв'язання таких задач ґрунтується на визначенні *точок рівноваги гри*. Пара стратегій $\bar{p}^0$ та $\bar{q}^0$ першого та другого гравців називається *точкою рівноваги за Нешем*, якщо жодному з учасників гри невигідно самостійно відхилятися від тієї стратегії, яка визначає точку рівноваги. Ця умова записується так: $H_1(\bar{p}, \bar{q}^0) \le H_1(\bar{p}^0, \bar{q}^0)$ для будь-яких p і $H_2(\bar{p}^0, \bar{q}) \le H_1(\bar{p}^0, \bar{q}^0)$ для будь-яких q (H_1 та H_2 – виграші першого та другого гравців).

Розглянемо приклад, наголосивши, що маємо гру з ненульовою сумою, тому виграші першого гравця не дорівнюють програшам другого гравця і в матриці виграшів у кожній клітині має бути відповідно дві величини:

$$H = \begin{bmatrix} (5;1) & (0;0) \\ (0;0) & (1;5) \end{bmatrix}.$$

За наведеної матриці гри H пари стратегій $\bar{p}^0 = (1;0)$, $\bar{q}^0 = (1;0)$ та $\bar{p}^0 = (0;1)$, $\bar{q}^0 = (0;1)$ визначають дві точки рівноваги. Дійсно, якщо перший гравець вибрав першу стратегію, то другому гравцеві доцільно вибрати теж першу стратегію. За таких умов виграш першого гравця буде 5, а другого – 1. Якщо ж другий гравець, знаючи, що перший вибрав першу стратегію, вибере другу стратегію, то виграш у обох гравців дорівнюватиме нулю. Аналогічний аналіз підтверджує, що пара $\bar{p}^0 = (0;1)$ і $\bar{q}^0 = (0;1)$ теж визначає точку рівноваги та недоцільність відхилятися від названих стратегій будь-кому із гравців. За наведеної матриці платежів маємо дві точки рівноваги з різними величинами виграшів кожного гравця (5 і 1) та (1 і 5).

У загальній теорії дослідження моделей гри доведено: *для будь-якої скінченної*

некооперативної гри з ненульовою сумою (біматричні ігри) завжди існує хоча б одна точка рівноваги серед змішаних стратегій. У загальному випадку таких точок може бути декілька, і відповідні величини виграшів різні.

Кооперативною грою називається гра з ненульовою *сумою*, в якій гравцям дозволяється узгоджувати перед грою свої дії (стратегії), тобто створювати коаліції. Основна задача кооперативної гри зводиться до розподілу виграшу між членами коаліції. Деякі суттєві поняття теорії кооперативних ігор розглянемо на прикладі моделей економічних відносин двох продавців, які реалізують два види товарів.

Позначимо через X і Y загальну кількість товарів першого та другого видів. Кожний із продавців має власну функцію зиску $h_1(x, y)$ і $h_2(x, y)$, задану на наборі названих товарів. Приймемо, що ці функції неперервні, монотонні по кожній із змінних та угнуті. Припустимо, що вихідна кількість товару кожного виду розподілена між продавцями: перший має x_1, y_1, а другий – x_2, y_2; $x_1 + x_2 = X$, $y_1 + y_2 = Y$. За відомих функцій зиску $h_1(x, y)$ та $h_2(x, y)$ необхідно дослідити, чи можуть продавці шляхом обміну кількостями товару збільшити величини свого зиску в порівнянні з вихідним розподілом кількості товару між ними.

Для наочного зображення моделі економічних стосунків за умови сформульованої задачі використовують так званий *ящик Еджворта*. Вибирають дві системи прямокутних координат, розташованих так, як зображено на рис. 14.3.

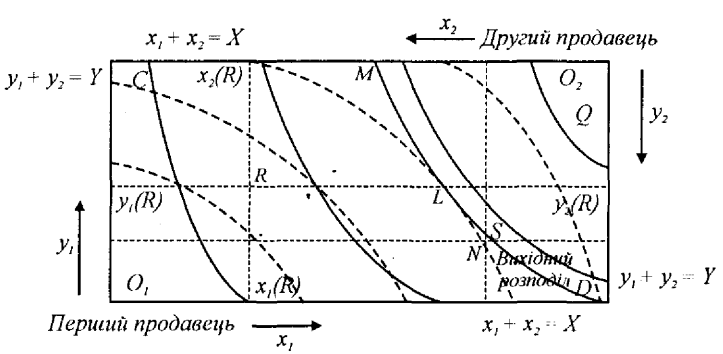


Рис. 14.3.

Кожна система координат відповідає певному продавцеві. Перетин взаємоперпендикулярних осей названих координат за обраного масштабу (мірила) зображає кількість першого товару (горизонтальна вісь) та другого товару (вертикальна вісь) в обох системах координат. Побудований прямокутник O_1CO_2D є геометричним тлумаченням всіх можливих розподілів товарів між продавцями. Наприклад, точка R у системі координат CO_1D має координати, які відповідають кількостям першого та

другого товару в першого продавця. Координати цієї ж точки R у системі координат DO_2C відповідають кількостям названих товарів у другого продавця. Уявимо, що у вибраних системах координат побудовано сімейства ліній рівня функцій зисків першого та другого продавців: $h_1(x, y) = \text{const}$ і $h_2(x, y) = \text{const}$. Деякі лінії названих сімейств зображені на рис. 14.3 неперервною та пунктирною лініями. Стосовно ліній сімейств необхідно наголосити, що через кожну точку проходить одна і лише одна лінія певного сімейства, тобто лінії одного сімейства не перетинаються. Лінії різних сімейств розташовані відносно одна одної довільно. Наголосимо, що коли лінії різних сімейств дотикаються одна одної, наприклад, у точці L, то це не означає, що відповідні функції зисків у цій точці мають тотожні значення.

Аналіз пошуку оптимізації значень функцій зисків $h_1(x, y)$ та $h_2(x, y)$ розпочнемо з визначення *Парето-оптимального розподілу*. Розподіл називається Парето-оптимальним, якщо зиск (виграш) одного з учасників не можна покращити, не зменшуючи величину зиску іншого партнера. Множину точок Парето-оптимального розподілу зручно зобразити, використовуючи ящик Еджворта, якщо маємо лише двох учасників. Розглянемо певну лінію рівня зиску одного з партнерів, наприклад, першого (відрізок лінії MLN на рис. 14.4).

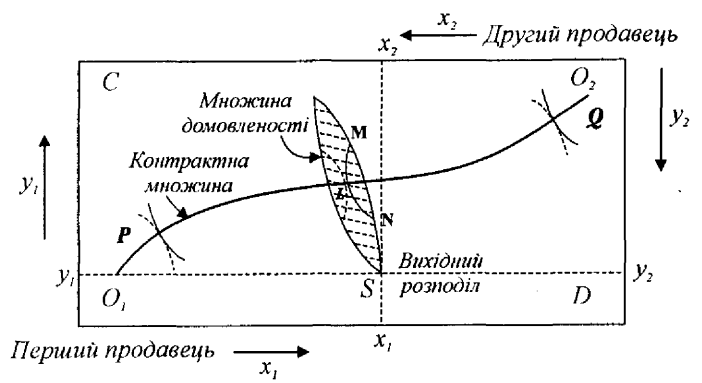


Рис. 14.4.

Уявимо, що розподіл товарів виконується так, що величини x_1 та y_1, узгоджені з фіктивною лінією рівня MLN для першого продавця, такі, що їх координати зворотно задовольняють рівнянню певної лінії рівня. За такої умови величина зиску першого продавця не буде змінюватися. Тому лінію рівня зиску називають ще лінією байдужості. Якщо величини x_1 та y_1 змінюються, то другий партнер теж матиме змінну величину зиску. Точкою Парето-оптимального розподілу буде точка L, в якій лінії рівня зисків обох учасників дотикаються одна одної. Розподіл товарів саме в

точці дотику досягає таких значень, що його зміни в напрямі спільної дотичної не будуть впливати на зміни зисків обох партнерів. Наголосимо, що досліджуємо не величини зисків, а їх можливу зміну. У точці Парето-оптимального розподілу маємо рівні диференціали функцій зисків обох партнерів за різних у загальному випадку значень величин зиску.

Таким чином, якщо існують Парето-оптимальні точки, то їх множина співпадає з множиною точок дотику ліній рівня функцій зисків обох партнерів. Множина Парето-оптимальних розподілів у просторі товарів називається *контрактною множиною*, бо партнерам доцільно домовлятися між собою за таких величин розподілу товарів; за інших розподілів у одного з партнерів буде втрата величини зиску. Множина Парето-оптимальних розподілів на рис. 14.4 зображена лінією *PLQ*.

Розглянемо розв'язок задачі про пошук оптимальних зисків для обох партнерів за умови, що кожний з них мав вихідну (початкову) кількість товарів кожного виду – x_1, y_1 та x_2, y_2 (рис. 14.4). Для розв'язання задачі скористаємося її тлумаченням з використанням ящика Еджворта. Пам'ятаючи, що суми $x_1+x_2=X$ і $y_1+y_2=Y$ є сталі величини, через точку S з координатами (x_1, y_1) в системі координат першого партнера проведемо лінії рівня $h_1(x, y)$ та $h_2(x, y)$, які обмежують заштриховану частину точок площини за умови, що початковий розподіл не є Парето-оптимальним. Якщо товари перерозподіляти так, щоб відповідна зображаюча точка переміщалася в середину заштрихованої області, величини зисків кожного з учасників будуть змінюватися, частина контрактної множини буде розташована в середині заштрихованої області. Ця частина контрактної множини породжує переговорну множину домовленостей, обумовлену вихідним розподілом S. Для того, щоб із вихідного положення S перейти в множину домовленостей, перший продавець повинен передати другому продавцеві частину товару першого виду в обмін на деяку кількість товару другого виду, яку йому необхідно одержати від другого продавця. Обсяги обмінів товарами визначаються відносним розташуванням точки S вихідного розподілу та точки, в якій дотикаються між собою лінії зисків обох партнерів в Парето-оптимальній множині, тобто точки L (рис. 14.4).

Висновок. Аналіз показує, що партнери можуть шляхом домовленості про перерозподіл вихідних кількостей наявних товарів підвищити величини своїх зисків. Один із можливих способів перерозподілу – обмін певними кількостями товарів.

14.8. Позиційні ігри як моделювання проблеми вступу до ринку

Розглянемо одну з можливих моделей розв'язання проблеми доцільності вступу до ринку. Проведемо дослідження такої ситуації. Ринком деякого товару повністю володіє монополіст-виробник – фірма 1, яка за монопольного стану має прибуток у 14 умовних одиниць (у. о.). Високий рівень прибутку підбурює фірму 2, яка вирішує для себе дилему: будувати завод з виробництва такого ж товару чи ні? Фірмі 2 відомо, що фірма 1 може здійснити певні дії на противагу діям фірми 2:

1) знизити удвічі рівень власного виробництва, поступившись частиною ринку фірмі 2; у такому випадку обидві фірми матимуть прибуток по 7 у. о.;

2) зберегти рівень виробництва; насичення ринку товарами обох виробників знизить ціну товару, і як наслідок фірма 1 одержує лише 6 у. о. прибутку, але фірма 2, витративши кошти на організацію виробництва нового товару, матиме не прибуток, а збитки в сумі 2 у. о.

Якщо ж фірма 2 утримається від спроби вступу до ринку, то вона не матиме збитку, але її прибуток дорівнюватиме нулю, фірма ж 1 одержить свій попередній прибуток у 14 у. о. Якщо ж фірма 1 під тиском можливої загрози від фірми 2 знизить рівень виробництва, то її прибуток становитиме 9 у. о. Ненасиченість ринку спровокує підвищення цін на монопольний товар.

Модель можливих стосунків між фірмами 1 та 2 може бути описана матрицею виграшів, поданою в табл. 14.5 (першими записані виграші фірми 1).

Таблиця 14.5.

Стратегії фірми 1		Стратегії фірми 2
Зберегти обсяги виробництва	Зменшити обсяги виробництва	
(6; –2)	(7; 7)	Налаштувати нове виробництво
(14; 0)	(9; 0)	Утриматися

Слушно наголосити, що запропонована модель гри за своїми умовами відрізняється від уже розглянутих. Якщо раніше вважалося, що учасники приймають свої рішення одночасно, не знаючи про рішення свого партнера (це було вельми суттєво), то за умов пропонованої гри фірма 1 може приймати рішення, уже маючи інформацію про рішення фірми 2; саме ця обставина суттєво впливає на розв'язання задачі.

Ігри, які можна реалізувати як певну *послідовність дій* гравців, називаються позиційними; кількість гравців і кроків реалізації може бути довільною. До *позиційних багатокрокових ігор* двох гравців, де кожний партнер приймає рішення, маючи певну інформацію про попередні дії партнера, належать ігри в шахи та шашки.

Оскільки позиційні ігри реалізуються послідовністю певних дій, то їх доцільно представляти не матрицею виграшів, а графом розв'язання, який відображає послідовність дій гравців від вихідної позиції до завершальної. Так, досліджувану гру *вступу до ринку* можна описати графом (рис. 14.5), ребра якого віддзеркалюють рішення партнерів, а біля кожної завершальної позиції записані відповідні виграші гравців (на першому місці записані виграші фірми 1).

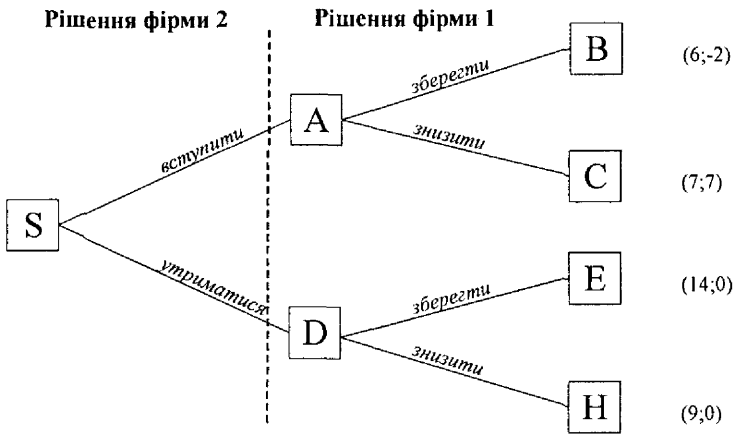


Рис. 14.5.

Вершини графу гри називаються *позиціями*. Позиції, які безпосередньо можуть відбутися після певної позиції, називаються *альтернативами*. Позиції, які не мають альтернатив, називаються *завершальними*. Шлях від вихідної позиції до завершальної називається *партією*. Досліджувана гра вступу до ринку має чотири партії (рис. 14.5): *SAB*; *SAC*; *SDE*; *SDH*.

Партії (стратегії) *SAC* та *SDE* урівноважені: жодному з партнерів невигідно відхилятися самостійно від обраних стратегій. Партія *SDE* означає, що фірма 2 утримується від вступу до ринку, в той же час фірма 1 зберігає рівень свого виробництва і має прибуток 14 у. о., а фірма 2 – 0 у. о. Партія *SAC* показує, що, коли фірма 2 вступає до ринку, а фірма 1 знижує обсяги виробництва, прибутки обох фірм за цих умов однакові – 7 у. о. Виникає питання: яка з двох партій найбільш ймовірна? У *непозиційній грі*, в якій партнери приймають рішення одночасно та незалежно один від другого, реалізація обох партій була б рівноймовірною; немає ніяких підстав віддати перевагу одній з партій (набору стратегій).

Але в позиційній грі мусимо врахувати, що фірма 1 приймає рішення, уже знаючи про рішення фірми 2. Менеджер фірми 2, яка робить перший вибір, має дилему:

а) якщо до ринку не вступати, то нічого не втрачаємо (але й нічого не одержуємо);

б) якщо вступити до ринку, то не виключено, що фірма 1 збереже обсяги виробництва, а ми матимемо збитки в 2 у. о.

Таким чином, якщо користуватися принципом максимізації свого мінімального виграшу, то фірма 2 має обрати стратегію "утриматися від вступу до ринку".

Такі міркування не враховують дуже важливу обставину – умову раціональної поведінки партнерів в їх прагненні одержати максимально можливий прибуток. За такої обставини менеджер фірми 2 мусить вирішити питання "А яка ймовірність того, що фірма 1 збереже обсяги свого виробництва, якщо ми вступимо до ринку? За цих умов фірма 1 одержить менший прибуток (6 у. о.), ніж тоді, коли вона зменшить обсяги виробництва та поділиться частиною ринку з нами (7 у. о.)". Отже, якщо фірма 1 керуватиметься принципом раціональності, то за умови вступу фірми 2 до ринку фірма 1 мусить зменшити обсяги свого виробництва, а не втратити 7 – 6 = 1 у. о. прибутку, аби подавити фірму 2. (Такі міркування не враховують наступні часи та економічний стан фірми 1 після можливого вступу фірми 2 до ринку).

У теорії гри розглянута ситуація класифікується як правдоподібна загроза. У запропонованій моделі гри за обраних величин виграшів загроза фірми 1 зберегти рівень свого виробництва у відповідь на вступ фірми 2 до ринку є неправдоподібною, бо реалізація такої загрози зменшує прибуток фірми 1 у порівнянні з іншими стратегіями. Є підстави стверджувати, що найбільш ймовірною партією буде *SAC*: фірма 2 вступає до ринку, а фірма 1 зменшує обсяги свого виробництва. Такий розв'язок гри дає матриця виграшів, представлена в табл. 14. 5.

Тепер розглянемо приклад моделі гри *вступу до ринку* з дещо іншими оцінками стратегій. Граф розв'язання цієї гри представлено на рис. 14. 6. Аналіз гри за наведеної величини виграшів у кожній партії показує, що загроза фірми 1 зберегти рівень виробництва підтверджується економічно: прибуток фірми 1 – 5 у. о. буде більший, ніж за умови зниження рівня (4 у. о.) та вступу фірми 2 до ринку. Стратегія

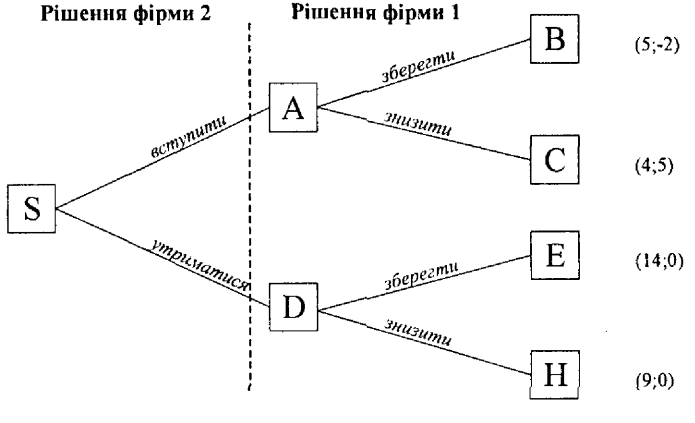


Рис. 14. 6.

фірми 1 "зберігати рівень виробництва" є домінуючою. Фірма 1 за такої стратегії одержить більший прибуток, ніж за стратегії "знизити обсяги виробництва", незалежно від стратегії фірми 2. Враховуючи таку ситуацію, фірма 2 змушена буде утриматися від вступу до ринку. Фірма 1 зберігає обсяги виробництва та залишається монополістом на ринку.

Приклад стратегічної гри за умов, представлених на рис. 14.6, моделює так звану ситуацію *стабільної* монополії, за якої фірма-монополіст може ефективно реалізувати загрозу знищення своїх потенційних конкурентів. Цьому сприяють чимало факторів: апробована технологія, якість виробництва, реклама, організація збуту, вміння працювати на перспективу і т. ін. За таких обставин в умовах соціально орієнтованої економіки природним є бажання держави контролювати в певних межах діяльність фірм-монополістів.

Математичний апарат теорії ігор дозволяє ефективно моделювати та досліджувати чимало економічних проблем господарської діяльності.

14.9. Теорія гри та прийняття управлінських рішень

У моделях теорії гри прийняття управлінських рішень пов'язане з розв'язанням задач, які умовно можна розділити на класи:

1) прийняття рішень за умов однозначності вихідних даних і перебігу керованого процесу в майбутньому; це так звані *детерміновані задачі*;

2) прийняття рішень за умов неоднозначності як вихідних даних, так і перебігу процесу в майбутньому, коли умови можна оцінити лише з певною мірою ймовірності; це *стохастичні задачі* управління;

3) прийняття певних рішень за умов невизначеності.

Математичні моделі обґрунтування рішень у задачах першого класу досить широко та змістовно розроблені з використанням певних методів оптимізації: лінійного та нелінійного математичного програмування, теорії управління процесами, перебіг яких моделюється диференціальними та інтегральними рівняннями, динамічного програмування.

Стохастичні моделі управління використовуються за умов, коли відомі можливі стратегії досягнення мети та ймовірні наслідки використання певної стратегії, але модельовані процеси такі, що не можна гарантувати однозначності їх перебігу. Для пошуку стратегій, оптимальних за обраним критерієм якості, використовуються методи, які в цілому можна характеризувати як "оптимізацію в середньому": оцінюється математичне сподівання показника ефективності кожної з можливих стратегій. Такий підхід дозволяє стохастичну задачу обґрунтування управлінського рішення звести до певної детермінованої задачі. Приклади таких задач було розглянуто в теорії управління запасами.

У моделях третього класу досліджується розв'язання задач, для яких характерно те, що про умови перебігу досліджуваного процесу принципово не можна мати певної інформації ні детермінованої, ні стохастичної. Умови невизначеності поділяються на два класи:

а) перебіг процесу визначається до певної міри наявністю факторів, обумовлених цілеспрямованою дією свідомо протидіючих учасників, які мають певну, але не відому конкретно кожному мету; наприклад, умови конкуренції на вільному ринку, військове або політичне протистояння;

б) процес, стосовно якого необхідно прийняти управлінські рішення, має відбуватися за умов певної невизначеності, але без активної цілеспрямованої протидії та певної байдужості як до умов, так і до наслідків процесу; це так звана природна невизначеність.

Для розбудови моделей ігор з природою необхідно врахувати, що "партнер" не має на меті активну протидію та байдужий до наслідків керованого процесу. Ці обставини обумовлюють певну специфіку побудови моделей гри. Рівень знань законів природи часто недостатній для побудови навіть статистичних моделей щодо багатьох керованих людиною процесів, а іноді бажане сприймається за дійсне.

Розглянемо деякі особливості побудови моделей гри "людина – природа". Приймаючи рішення, людина може скористатися кількома стратегіями – $A_1, A_2, \ldots, A_m$. Вплив природи на досліджуваний процес можна також змоделювати як використання певної множини стратегій $\Pi_1, \Pi_2, \ldots, \Pi_n$. Якщо з попереднього досвіду відомі ймовірності можливих станів природи, то такі ймовірності називаються *апріорними* (доекспериментальними). Якщо людина шляхом експерименту може вдосконалити свої знання про відповідні стани природи стосовно керованого процесу та їх ймовірності, то такі ймовірності називаються *апостеріорними* (післяекспериментальними). Але необхідно пам'ятати, що експерименти вимагають і коштів, і часу (особливо в економіці) саме тоді, коли рішення необхідно приймати терміново. Тому розглянемо *моделі гри з природою без експериментів*.

Плануючи свої дії, людина може користуватися як деякими чистими стратегіями A_i ($i = 1, 2, \ldots, m$), так і змішаними:

$$\overline{S} = (p_1, p_2, \ldots, p_m), \text{ де } p_i \geq 0;\ i = 1, 2, \ldots, m;\ \sum_{i=1}^{m} p_i = 1,$$

за умови, що є можливість оцінити наслідки використання будь-якої чистої стратегії в залежності від будь-якого довільного стану природи Π_j, тобто для кожної допустимої сполуки (A_i, Π_j) відомий чисельний результат a_{ij} і можна задати матрицю виграшів A досліджуваної гри (табл. 14.6). Елементи матриці A позначимо a_{ij}, наголошуючи на умовності поняття "виграшів".

Таблиця 14 6.

A_i \ Π_j	Π_1	Π_2	……	Π_n	p_i
A_1	a_{11}	a_{12}	……	a_{1n}	p_1
.	.	.	……	.	.
.	.	.	……	.	.
A_m	a_{m1}	a_{m2}	……	a_{mn}	p_m
β_j	β_1	β_2	……	β_n	
q_i	q_1	q_2	……	q_n	

Природно, що перш ніж обрати оптимальну стратегію необхідно проаналізувати матрицю A та спростити її, враховуючи можливі домінуючі стратегії людини. Відхиляти стратегії (стани) природи недопустимо, бо природа може перебувати в своєму довільному стані незалежно від того, чи корисно це людині. У задачах досліджуваного типу за критерій ефективності стратегій A_i приймають математичне сподівання a_i виграшу за умови використання i-ої стратегії, тобто:

$$a_i = a_{i1}q_1 + a_{i2}q_2 + \ldots + a_{in}q_n = \sum_{j=1}^{n} a_{ij}q_j \qquad (21)$$

($\{q_j\}, j = 1, 2, \ldots, n$ – апріорні ймовірності можливих станів природи Π_j). За оптимальну обирається стратегія, за якої величина a_i (21) досягає найбільшого значення.

Оскільки оптимальну стратегію задач гри з природою доцільно шукати серед чистих стратегій (природа не може комбінувати свої стани), то користуються іншим методом пошуку, ніж з використанням математичного сподівання величини виграшу (21). Маючи матрицю виграшів (табл. 14.6), обчислюють так звану *матрицю ризиків*, яка дозволяє більш чітко виявити переваги певної стратегії за даного можливого стану природи.

Ризиком r_{ij}, якщо користуватися чистою стратегією A_i за Π_j стану природи, називають різницю між максимальним виграшем $\max\limits_i a_{ij}$, який можна було б одержати, якби природа достовірно була в стані Π_j, та виграшем a_{ij}, який можна одержати, використовуючи стратегію A_i та приймаючи Π_j за можливий стан природи. Таким чином, елементи r_{ij} матриці R ризиків обчислюються за формулою:

$$r_{ij} = \beta_j - a_{ij} \quad (i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n) \qquad (22)$$

(β_j – максимально можливий виграш за Π_j стану природи (тобто максимальний елемент j-го стовпця матриці платежів $\beta_j = \max\limits_i a_{ij}$)).

Приклад. На опалювальний сезон потрібно від 10 до 12 т вугілля в залежності від його якості. Якщо запасу вугілля буде недостатньо, то його можна докупити за

умови додаткових витрат у 5 умовних одиниць на 1 т. Якщо ж запас вугілля перевищить необхідні витрати, то затрати на зберігання залишків складуть 2 у.о. на 1 т. Побудувати модель гри та скласти матрицю платежів і ризику.

Можна замовити 10, 11 або 12 т вугілля – це будуть відповідно стратегії A_1, A_2, A_3. Вплив природи на перебіг процесу опалення будемо моделювати як витрати 10, 11 або 12 т у залежності як від якості вугілля, так і від теплового режиму атмосфери, тому вважатимемо, що стани природи Π_1, Π_2, Π_3 – це витрати вугілля відповідно в обсягах 10, 11, 12 т.

Елементи матриці платежів (табл. 14.7) характеризують додаткові затрати, обумовлені або поставкою вугілля у випадку його нестачі, або зберіганням, якщо з'явився його надлишок.

Таблиця 14.7.

A_i \ Π_j	$\Pi_1(10)$	$\Pi_2(11)$	$\Pi_3(12)$	$\alpha_i = \min\limits_j a_{ij}$
$A_1(10)$	0	–5	–10	–10
$A_2(11)$	–2	0	–5	–5
$A_3(12)$	–4	–2	0	–4
$\beta_j = \max\limits_i a_{ij}$	0	0	0	

Так, якщо замовляти 11 т, а буде витрачено в опалювальний сезон менше 10 т (ситуація A_2, Π_1), то доведеться зберігати 11 – 10 = 1 т, що обумовить додаткові витрати 2 одиниці, а в ситуації (A_1, Π_3), коли було замовлено 10 т, а можливі витрати – 12 т, то на закупівлю додаткових 12 – 10 = 2 т необхідно витратити 10 одиниць.

Використовуючи дані табл. 14.7 та формулу (22), побудуємо матрицю ризиків, записану в табл. 14.8.

Таблиця 14. 8.

A_i \ Π_j	$\Pi_1(10)$	$\Pi_2(11)$	$\Pi_3(12)$	$\max\limits_j r_{ij}$
$A_1(10)$	0	5	10	10
$A_2(11)$	2	0	5	5
$A_3(12)$	4	2	0	4

Розглянемо розв'язок задачі пошуку оптимальних стратегій за критерієм вибору найбільшого математичного сподівання (21) виграшу та найменшого середнього очікуваного ризику за умови, що ймовірності q_1, q_2, q_3 використання обсягів вугілля 10, 11, 12 т дорівнюють відповідно 0,3; 0,1; 0,6. Маємо:

$$a_1 = 0\times0{,}3 + (-5)\times0{,}1 + (-10)\times0{,}6 = -6{,}5;$$
$$a_2 = (-2)\times0{,}3 + 0\times0{,}1 + (-5)\times0{,}6 = -3{,}6;$$

$$a_3 = (-4)\times 0{,}3 + (-2)\times 0{,}1 + 0\times 0{,}6 = -1{,}4.$$

Обчислення показують, що очікувані збитки будуть найменшими – 1,4 одиниць, якщо замовити 12 т.

Обчислимо математичні сподівання величини ризиків r_i, користуючись табл. 14. 8:

$$r_1 = 0\times 0{,}3 + 5\times 0{,}1 + 10\times 0{,}6 = 6{,}5;$$
$$r_2 = 2\times 0{,}3 + 0\times 0{,}1 + 5\times 0{,}6 = 3{,}6;$$
$$r_3 = 4\times 0{,}3 + 2\times 0{,}1 + 0\times 0{,}6 = 1{,}4.$$

Величина очікуваного ризику буде найменшою за вибору стратегії A_3 (замовлення 12 т). У наведеному прикладі стратегія, за якої максимізувався середній виграш, співпала зі стратегією, за якої мінімізувався середній очікуваний ризик.

У теорії гри "людина – природа" цей результат доведено в загальному випадку: *стратегії оптимізації максимального середнього виграшу та мінімізації середнього ризику співпадають.*

Якщо немає об'єктивної інформації про апріорні ймовірності q_j стану природи Π_j ($j = 1, 2, ..., n$) і всі вони уявляються однаково правдоподібними, то їх ймовірності приймаються рівними між собою, тобто $q_j = 1/n$. Такий підхід називають *"принципом недостатнього обґрунтування Лапласа"*. Його використовують у пошуку оптимальної стратегії за критерієм Лапласа:

$$\max_i a_i = \max_i \left(\frac{1}{n}\right)\sum_{j=1}^{n} a_{ij}\,. \qquad (23)$$

Досі ми розглядали методи пошуку оптимальних стратегій за умови певної інформації про ймовірності станів природи, коли, згідно з теорією, оптимальну стратегію можна знайти в множині чистих стратегій.

Розглянемо критерії пошуку оптимальних стратегій за умови відсутності будь-яких відомостей про ймовірності станів природи. Методи, які використовуються при дослідженні таких ситуацій, дозволяють визначити деяку гарантовану величину виграшу на підставі аналізу матриць виграшу або ризику.

Критерій Вальда (крайнього песимізму) – це максимінний критерій, який використовується як для чистих, так і для змішаних стратегій. За критерієм Вальда, вибір оптимальної стратегії виконується з припущенням, що природа діє супроти найгіршим чином, тобто за будь-якої стратегії A_i в природі очікується такий стан Π_j, за якого виграш приймає найменше значення a_{ij}. За таких умов необхідно обирати ту чисту стратегію A_i, за якої найменший виграш $\min_j a_{ij}$ буде максимальним, тобто оптимальною стратегією буде максимінна чиста стратегія. Для наведеного прикладу $\max_i \min_j a_{ij} = \max_i (-10;\ -5;\ -4) = -4$. Отже, оптимальною чистою стратегією, за Вальдом, буде A_3, тобто запас вугілля має бути 12 т, а додаткові витрати дорівнюють 4 одиницям.

Для змішаних стратегій критерій Вальда формулюється так: за оптимальну змішану стратегію приймають ту, за якої мінімальний середній виграш $\min_j \sum_{i=1}^{m} a_{ij} p_i$ буде максимальним, тобто змішана оптимальна стратегія $\overline{p}^{\,0} = (p^0_1, p^0_2, \ldots, p^0_m)$ визначається за умови:

$$\max_{p_i^0} \min_j \sum_{i=1}^{m} a_{ij} p_i\,.$$

У наведеному прикладі для визначення оптимальної стратегії за Вальдом необхідно розв'язати відповідну задачу лінійного програмування (12), (13). Результат її розв'язку такий:

$$\overline{p}^{\,0} = (2/7;\ 0;\ 5/7;) \quad V = -20/7.$$

З практичної точки зору результат необхідно тлумачити так: очікувані додаткові витрати будуть найменшими та дорівнюватимуть 2,86 одиницям (20/7), а запас вугілля повинен складати 11,43 т ($10\times 2/7 + 11\times 0 + 12\times 5/7$). Розв'язок у змішаних стратегіях більш доцільний, ніж у чистих: рівень запасу та додаткові витрати менші.

Критерій Севіджа теж будується на допущенні, що від природи очікуються найнесприятливіші стани. За цим критерієм оптимальною стратегією обирають ту з чистих стратегій A_i ($i = 1, 2, \ldots, m$), за якої мінімізується величина максимального ризику, тобто на оптимальній стратегії досягається $\min_i \max_j r_{ij}$.

Необхідно наголосити ще раз, що для розв'язання задачі у змішаних стратегіях необхідно мати компоненти вектора $\overline{p}^{\,0}$, які можна обчислити, розв'язавши відповідну задачу лінійного програмування (12), (13).

Для наведеного прикладу оптимальною, за Севіджем, чистою стратегією A_i буде та, для якої виконується умова:

$$\min_i \max_j r_{ij} = \min_i (10;\ 5;\ 4) = 4,$$

тобто стратегія A_3.

Якщо критерій Севіджа використовується на множині змішаних стратегій, то

оцінюється середній ризик $\sum_{i=1}^{m} r_{ij} p_i$ $(j = 1, 2, \ldots, n)$. Найнесприятливішим приймається такий стан природи $П_j$, за якого величина середнього ризику досягає найбільшого значення $\max_j \sum_{i=1}^{m} r_{ij} p_i$. За критерієм Севіджа, оптимальною змішаною стратегією буде та, за якої максимальне значення ризику мінімізується, тобто має бути виконана умова:

$$\min_{\bar{p}} \max_j \sum_{i=1}^{m} r_{ij} p_i .$$

Критерій Гурвіца, який називають критерієм песимізму-оптимізму, радить вибирати дещо середнє між результатами найбільш і найменш можливих виграшів. На множині чистих стратегій, за критерієм Гурвіца, за оптимальну приймають ту стратегію, для якої виконується умова:

$$\max_i \{\lambda \min_j a_{ij} + (1 - \lambda) \max_j a_{ij}\}, \qquad (24)$$

де $\lambda \in [0; 1]$ і вибирається за *суб'єктивними міркуваннями*. За умови $\lambda = 1$ критерій Гурвіца тотожний критерію Вальда (крайнього песимізму): $\max_i \min_j a_{ij}$; за умови $\lambda = 0$ маємо критерій крайнього оптимізму: $\max_i \max_j a_{ij}$. Якщо ж виберемо за λ певне проміжне значення від 0 до 1, то одержимо дещо середнє між стратегіями крайнього песимізму та крайнього оптимізму. Коефіцієнт λ є ніби мірою песимізму того, хто приймає рішення про вибір стратегій: чим більш небезпечна ситуація, тим доцільніше обрана λ ближче до 1.

Для наведеного прикладу виберемо $\lambda = 0{,}6$ (це наше бажання), тоді з (24) маємо: $\max_i \{0{,}6 \min_j a_{ij} + 0{,}4 \max_j a_{ij}\} = \max_i h_j$.

Розрахунки наведені в табл. 14.9.

Таблиця 14.9.

A_i \ $П_j$	$П_1$	$П_2$	$П_3$	$\min_j a_{ij}$	$0{,}6 \min_j a_{ij}$	$\max_j a_{ij}$	$0{,}4 \max_j a_{ij}$	h_j
A_1	0	–5	–10	–10	–6,0	0	0	–6,0
A_2	–2	0	–5	–5	–3,0	0	0	–3,0
A_3	–4	–2	0	–4	–2,4	0	0	–2,4

Використовуючи результати, записані в останньому стовпці, маємо

$\max_i h_j = -2{,}4$, що відповідає чистій стратегії A_3, яка і буде оптимальною, за критерієм Гурвіца, при $\lambda = 0{,}6$.

Зауваження. У моделях гри з природою за умови невизначеності її можливих станів доцільно обирати оптимальну стратегію, використовуючи усі три вищеназвані критерії. Якщо при використанні усіх критеріїв оптимальні стратегії тотожні, то це переконлива аргументація для практичного використання результату. Якщо розв'язки не співпадають, необхідно ретельно проаналізувати результати та використати один із них.

Питання для самоконтролю

1. Які питання досліджуються в теорії гри?
2. Яка гра називається матричною?
3. Навести приклад моделювання конфліктної ситуації.
4. Що розуміють під термінами:
   а) чиста стратегія;
   б) змішана стратегія?
5. Що таке матриця виграшів? Як знайти верхню та нижню ціни гри?
6. Зміст основної теореми теорії гри з нульовою сумою.
7. Приведення задачі пошуку оптимальної стратегії до задачі лінійного програмування.
8. Різниця між некооперативними, кооперативними та позиційними іграми.
9. Що таке ящик Еджворта?
10. Що таке Парето-оптимальний розподіл?
11. Що називається в теорії гри ризиком?
12. Зміст критеріїв Вальда, Гурвіца, Севіджа.
13. Практичне використання теорії гри.
14. Принципова різниця між задачами математичного програмування та теорії гри.

ГЛАВА 15

МОДЕЛІ ДИНАМІЧНОГО ПРОГРАМУВАННЯ

15.1. Історична довідка

Ефективне планування й управління в АПК не можливе без використання сучасних систем управління, математичного моделювання та обчислювальної техніки.

У моделях математичного програмування не враховується час. Це так звані одноетапні моделі, які допомагають аналізувати статичні, не залежні від часу умови й ефективно використовуються, коли змінами досліджуваного процесу в часі можна знехтувати без спотворення його моделі. Оптимальний розв'язок за такого моделювання має сенс або за умов стабільності господарського процесу, або на короткий проміжок у майбутньому. Але практика господарювання досить часто вимагає розв'язання таких задач, в яких необхідно враховувати можливі зміни економічних ситуацій у часі за умов обов'язкового контролювання певних проміжків (рік, квартал, місяць тощо). Такі багатоетапні процеси можуть бути як детермінованими, так і стохастичними, як неперервними, так і дискретними.

Динамічне програмування (динамічне планування) є особливим математичним методом розв'язання задач певної структури – задач пошуку оптимального розв'язку за умов, що досліджуваний процес істотно залежить від часу і його перебіг може бути поетапно контрольований. Тобто динамічне програмування (ДП) є методом дослідження операцій, з допомогою якого моделюються процеси, змінювані в часі, якщо їх перебігом можна керувати на певних проміжках. ДП доцільно розглядати як певну єдину теорію, об'єднану на засадах тотожних ідей і засобів, які ефективно використовуються при математичному моделюванні задач, різних за походженням і призначенням:

- 1) розподіл капіталовкладень між можливими напрямами їх використання;
- 2) управління запасами;
- 3) календарне планування виробництва при коливанні попиту на продукцію;
- 4) закупівля запасних частин для забезпечення неперервної експлуатації обладнання;
- 5) планування профілактичного ремонту обладнання економічно значимих підприємств (ТЕЦ, АЕС та ін.);
- 6) вибір методів проведення реклами;
- 7) планування вилучення основних фондів з експлуатації.

Динамічне програмування з його специфічним математичним апаратом сформувалося у 50-ті роки XX століття на ґрунті вимог практики планування й управління. Основи методу ДП були розроблені американським ученим Р. Белманом та його

співробітниками, але перші роботи з використанням основних ідей ДП з'явилися раніше: Массе (Франція, 1946), Вальд (США, 1947), Ероу і Блекуел (США, 1949).

Моделі математичного програмування використовуються при дослідженні широкомасштабних проблем макроекономіки. У мікроекономіці вони ефективні, якщо розглядаються короткотермінові процеси, які можна вважати стабільними на досліджуваних проміжках. Моделі динамічного програмування використовуються переважно в дослідженні задач мікроекономіки і, на відміну від моделей математичного програмування, дають можливість враховувати перебіг процесу в часі, оцінювати його стан і моделювати поетапне управління процесом.

Деякі досліджувані в ДП операції розпадаються на етапи природним шляхом, наприклад, процес планування господарської діяльності групи підприємств (природними етапами є місяці, квартали, роки). В інших ситуаціях поділ на етапи може мати умовний характер. Але особливістю усіх задач, які розв'язуються методом ДП, є те, що *на кожному етапі можна керувати перебігом досліджуваного процесу та оцінювати якість такого управління.* Перш ніж розглядати математичний апарат ДП, опишемо розв'язання двох конкретних задач для з'ясування специфіки методу ДП.

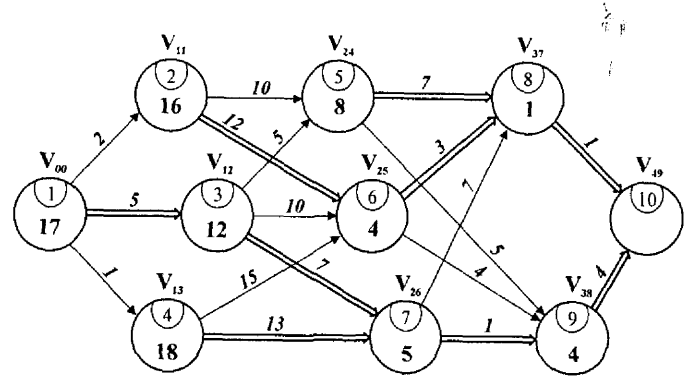
15.2. Приклади розв'язання задач методом динамічного програмування

15.2.1. Задача пошуку найекономнішого маршруту доставки вантажу

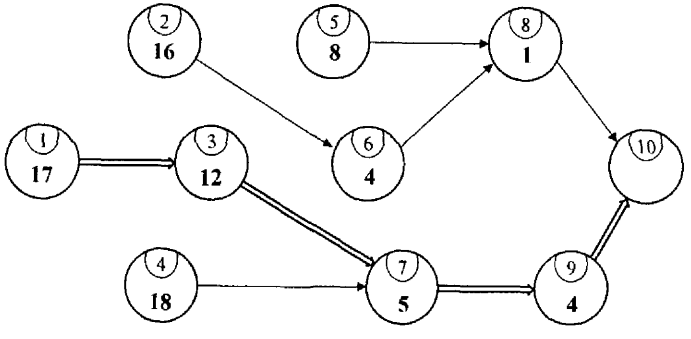
Маємо мережу доріг (схематично представлені на рис. 15.1 а) кількох маршрутів, якими можна перевезти вантаж з пункту 1 у пункт 10. Вартість перевезення вантажу між окремими проміжними пунктами мережі відома (записана на відповідних ребрах графу мережі). Необхідно так спланувати перевезення вантажу з пункту 1 у *пункт 10, щоб вартість була мінімальною.*

Виконаємо так звану "правильну" нумерацію вершин графу, використовуючи неперервність у нумерації та надавши номер 1 вихідній вершині, бо в цю вершину не заходить жодна дуга. Правильна нумерація вершин виконується за такою схемою. Умовно викреслюємо всі дуги, які виходять з вершин, що вже одержали номери. Наступну нумерацію виконуємо для вершин, в які заходять *лише викреслені дуги.* Потім викреслюємо дуги, які виходять з уже пронумерованих вершин, і продовжуємо нумерацію вершин, в які заходять лише викреслені дуги, і т. д. Для наведеного графу крайня ліва вершина має лише дуги, які виходять з неї, тому їй було надано номер 1. Умовно викресливши дуги, які виходять з першої вершини, одержуємо три вершини, в які заходять лише викреслені дуги. Цим вершинам даємо номери 2, 3, 4 у

будь-якому порядку, наприклад, так, як зображено на рис. 15.1 а). Викреслюємо умовно дуги, які виходять з цих вершин, і знову одержуємо три вершини, в які заходять лише викреслені дуги. Розподіляємо між ними три чергових номери 5, 6, 7, наприклад, так, як наведено на рис. 15.1 а). Процес нумерації завершуємо тоді, коли досягнемо вершини, з якої не виходить жодна дуга; в наведеному прикладі це вершина з номером 10. Мета виконання правильної нумерації – *уникнути ситуації, за якої вершина, віддалена від вихідної меншою кількістю прогонів, має більший номер, ніж вершина, віддалена більшою кількістю прогонів.* У теорії графів доведено, що коли граф не має замкнених ланцюгів, то можна завжди виконати правильну нумерацію його вершин.



а)



б)

Рис. 15.1.

Правильно виконана нумерація вершин показана на рис. 15.1 а) і виділена меншими кружками в середині більших, якими зображено вершини. За умови правильної нумерації переміщення транспорту з вантажем з пункту 1 у пункт 10 розглядаємо як поетапну операцію: на кожному етапі транспорт переміщується від пункту з меншим номером до будь-якого пункту з більшим номером. Разом з цим пошуки найекономнішого маршруту з пункту 1 у пункт 10 доцільно розбити на кроки. На кожному кроці треба так обирати маршрут транспорту з досягнутого пункту до наступного, щоб по всьому шляху від пункту 1 до пункту 10 витрати були мінімальними. "Правильно" виконана нумерація враховує особливості мережі, граф якої наведений на рис. 15.1 а): *недоцільність руху від вершини з більшим номером до вершини з меншим номером.*

Принцип оптимальності ДП, який буде в загальному вигляді розглянуто далі, стосовно розв'язуваної задачі можна сформулювати так: *оптимальний маршрут (шлях) перевезення вантажу з пункту* 1 *у пункт* 10 *має ту властивість, що незалежно від того, яким був шлях досягнення з пункту* 1 *певного проміжного пункту мережі, подальший оптимальний шлях від досягнутого пункту до кінцевого пункту* 10 *має співпадати з оптимальним шляхом на всій мережі.* Оптимальний шлях від будь-якого пункту до завершального називається *локально оптимальним,* або *умовно оптимальним.*

Отже, пошук оптимального шляху в розв'язуваній задачі зведемо до чотирьох кроків за кількістю повторень операцій умовного викреслення дуг і нумерації вершин, досягнутих лише викресленими дугами. Пошук оптимального шляху на кожному кроці виконуємо, починаючи з останнього – четвертого. Доцільно наголосити, що за виконання правильної нумерації при розгляді вершин у зворотному напрямі від завершальної (10) до вихідної (1) неможливою стає ситуація, коли вершина з меншим номером передує вершині з більшим номером.

Якщо на передостанньому кроці вантаж потрапив у пункти 8 або 9, то в будь-якому випадку подальший рух у пункт 10 вибирається однозначно, проблеми вибору немає. Локально оптимальні шляхи (8; 10) та (9; 10) позначені на графі подвійною стрілкою, а відповідні витрати на транспортування записані числами поруч зі стрілками та в нижній половині кружків вершин 8 та 9 (рис. 15.1 а)).

Виконаємо умовну оптимізацію стосовно передостаннього – третього кроку. Умовна оптимізація означає, що з кожної вершини 5, 6, 7 треба знайти локально оптимальний шлях до кінцевої вершини 10. Шлях з пункту 5 до цієї вершини може проходити через вершини (8; 10) або (9; 10). Вартість перевезення на відповідних шляхах визначається сумами 7 + 1 = 8 (5; 8; 10); 5 + 4 = 9 (5; 9; 10). Вибираємо умовно оптимальний шлях для п'ятої вершини (5; 8; 10), вартість перевезення цим шляхом вказуємо внизу кружка п'ятої вершини та відзначаємо шлях подвійною стрілкою на графі (рис. 15.1 а)). Ребро (5; 9) не буде мати відмітки подвійною стрілкою як таке, що не належить локально оптимальному шляху.

З вершини 6 маємо два шляхи до вершини 10 – (6; 8; 10) та (6; 9; 10). Оцінивши вартість перевезення для кожного з цих шляхів: 3 + 1 = 4 та 4 + 4 = 8, обираємо локально оптимальний шлях (6; 8; 10). Вартість перевезення з пункту 6 у пункт 10 оптимальним шляхом записуємо внизу кружка вершини 6, а ребро (6; 8) позначаємо подвійною стрілкою. Ребро (6; 9) не буде мати подвійної стрілки. Аналогічно для пункту 7 знаходимо локально оптимальний шлях (7; 9; 10) з вартістю перевезень у п'ять одиниць, тому число 5 записуємо внизу кружка вершини 7 і позначаємо подвійною стрілкою ребро (7; 9). Таким чином, для кожної з вершин 5; 6; 7 визначено локально оптимальні шляхи від названих вершин до кінцевої (їх позначено подвійними стрілками на рис. 15.1 а)). Вартості перевезень по кожному з цих шляхів записані відповідними числами внизу кружків, які зображають названі вершини.

Оптимізуємо вартість перевезень з кожної вершини 2, 3, 4, дуги з яких заходять до вже розглянутих вершин 5, 6, 7. При пошуку умовного оптимуму для кожної з вершин 2, 3, 4 слід проаналізувати усі можливі шляхи, які починаються в цих вершинах і завершуються в раніше розглянутих. Для пункту 2 умовно оптимальним буде шлях (2; 6; 8; 10) з витратами 12 + 4 = 16 одиниць, для пункту 3 – (3; 7; 9; 10) з витратами 7 + 5 = 12 одиниць, для пункту 4 – (4; 7; 9; 10) з витратами 13 + 5 = 18 одиниць. Названі умовно оптимальні шляхи на рис. 15.1 а) позначені подвійною лінією.

Оптимізуємо останній крок – перехід від пункту 1 до вершин 2, 3, 4, які можна досягти з цього пункту за один прогін. Для вибору умовного оптимального шляху необхідно розглянути три шляхи, які з пункту 1 проходять через пункти 2, 3, 4. Знаходимо, що найкращий (оптимальний) шлях від пункту 1 до завершального пункту проходить через пункт 3 з витратами в 17 одиниць, позначений на рис. 15. 1, а) подвійною лінією.

Таким чином, етап умовної оптимізації завершений, його результати позначені подвійними рисками та цифрами у кожній із вершин графу (рис. 15.1 а)). Тепер необхідно розглянути пункти у напрямі від пункту 1 до кінцевого 10 та назвати оптимальний шлях на всій мережі: в 1; 3; 7; 9;10 (рис. 15.1 б)).

Наголосимо, що від пункту 1 до одного з наступних (2, 3, 4) для оптимізації задачі в цілому необхідно обрати шлях (1; 3), вартість перевезень на якому в 2,5 раза перевищує витрати на шляху (1; 2) та у п'ять разів – на шляху (1; 4). Розв'язок задачі в цілому (дослідження чотирьох кроків розв'язання), а не лише на першому кроці, показує, що оптимізація вартості доставки вантажу в кінцевий, 10-ий пункт вимагає певної "жертви" (перевитрат) на першому етапі, щоб забезпечити мінімум витрат на всьому чотирикроковому шляху. Це яскрава та переконлива ілюстрація однієї з головних особливостей методу динамічного програмування: *вибирати розв'язок на кожному кроці, керуючись не зиском на цьому кроці, а загальним зиском, який можна одержати, лише дослідивши задачу в цілому.*

Наведений спосіб міркування дозволив визначити не лише оптимальний шлях

перевезення вантажу з початкового пункту 1 у кінцевий пункт 10, але і всю множину локально оптимальних шляхів від кожного пункту мережі до кінцевого пункту. Елементи цієї множини умовно оптимальних шляхів позначені на рис. 15.1 б) одинарними стрілками. Наприклад, оптимальний шлях від пункту 4 є (4; 7; 9; 10), а вартість перевезень дорівнює 18 одиницям. Цей результат іноді більш важливий, ніж знайдений один оптимальний шлях на всій мережі.

15.2.2. Задача розширення виробничих потужностей за умови відрахувань від прибутку з метою одержання максимального зиску

Розглянемо задачу оптимізації розширення виробництва за таких умов. На початку п'ятирічного періоду роботи підприємство одержало кошти в сумі C для закупівлі нового обладнання. Один комплект обладнання коштує 10 умовних одиниць. Закуплене обладнання зразу ж використовується у виробництві. Використання одного комплекту обладнання на кінець року дає чотири умовні одиниці прибутку. По закінченні кожного року підприємство може використати деяку частку λ прибутку на розширення виробничих потужностей, тобто на закупівлю додатково деякої кількості комплектів обладнання, яке буде використовуватися в наступні роки та приносити прибуток. Частка λ відрахувань може змінюватися кожного року та задовольняє природній умові: $0 \le \lambda \le 1$. Необхідно так спланувати розширення виробництва, щоб за п'ять років мати максимальний зиск.

Досліджувана задача має чітко виражений п'ятикроковий процес, станом якого можна керувати, витрачаючи на розширення виробництва певну частку річного прибутку . Розв'язком на кожному кроці (по завершенні кожного року) є оптимальне значення величини λ – частки прибутку, якою визначають відрахування для закупівлі комплектів додаткового обладнання. Припустимо, λ_i $(i = \overline{1,5})$ – частка відрахувань по закінченні i-го року, m_i – кількість комплектів обладнання, які використовувались у виробництві в i-му році, φ_i – фактичний прибуток підприємства в i-му році після відрахувань на розширення виробництва. За умовою задачі маємо: $\varphi_i = 4m_i - 4\lambda_i m_i = 4m_i(1 - \lambda_i)$.

Стосовно питання пошуку оптимального прибутку виробничий процес буде відомий, якщо на початку кожного року буде задано кількість комплектів використовуваного обладнання та *сумарний прибуток*, накопичений на цей момент.

У перший рік у виробничому процесі буде використано:

$$m_1 = C/10 \qquad (1)$$

комплектів обладнання, що забезпечить по завершенні року прибуток $4m_1$ одиниць.

Якщо частку λ_1 цього прибутку, тобто $4m_1\lambda_1$, планується використати на розширення виробництва, то по завершенні першого року *фактичний прибуток* φ_1 дорівнюватиме:

$$\varphi_1 = (1 - \lambda_1)\, 4m_1. \tag{2}$$

На відраховану для розширення виробництва суму $4m_1\lambda_1$ може бути закуплено $4m_1\lambda_1/10$ комплектів обладнання. Отже, протягом другого року буде використано:

$$m_2 = m_1 + 0{,}4m_1\lambda_1 \tag{3}$$

комплектів обладнання, що забезпечить по завершенні другого року прибуток

$$\varphi_2 = 4m_2 - 4\lambda_2 m_2 = (1 - \lambda_2)4m_2 \tag{4}$$

($4\lambda_2 m_2$ – величина прибутку, відрахована на розширення виробництва, що дасть можливість протягом третього року використати додатково $4\lambda_2 m_2/10$ комплектів обладнання).

Всього буде використано протягом третього року

$$m_3 = m_2 + 0{,}4\lambda_2 m_2 = m_2(1 + 0{,}4\lambda_2) \tag{5}$$

комплектів обладнання, що забезпечить по завершенні третього року фактичний прибуток

$$\varphi_3 = 4(1 - \lambda_3)\, m_3 \tag{6}$$

(враховано кошти $4\lambda_3 m_3$, які будуть використані для розширення виробництва).

Аналогічно одержуємо для четвертого року:

$$m_4 = (1 + 0{,}4\lambda_3)\, m_3 \tag{7}$$

$$\varphi_4 = 4(1 - \lambda_4)\, m_4, \tag{8}$$

для п'ятого року:

$$m_5 = (1 + 0{,}4\lambda_4)\, m_4, \tag{9}$$

$$\varphi_5 = 4(1 - \lambda_5)\, m_5. \tag{10}$$

Підсумковий прибуток за п'ять років будемо розглядати як деяку функцію $Z(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$. Наша мета – знайти такі величини $\lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, \lambda_5^*$, за яких величина Z досягає найбільшого значення. Використовуючи принцип оптимальності

методу динамічного програмування та розглядаючи процес у зворотному напрямі – від останнього до попереднього кроку і т. д., побудуємо *стратегію оптимальних відрахувань*. Позначимо $Z(1)$, $Z(1–2)$, $Z(1–3)$, $Z(1–4)$ сумарний прибуток відповідно за перший рік; за перший і другий роки; за перший, другий і третій роки; за перший, другий, третій і четвертий роки.

Якщо на початок п'ятого року стан виробництва характеризувався кількістю комплектів обладнання m_5 та обсягом прибутку $Z(1–4)$, то прибуток за п'ять років можна обчислити так: $Z = Z(1 - 4) + \varphi_5$, а за умови (10) –

$$Z = Z(1–4) + 4(1 - \lambda_5)m_5, \text{ або } Z = Z(1–4) + 4m_5 - 4\lambda_5 m_5. \tag{11}$$

Отже, величина Z досягне найбільшого значення за умови $\lambda_5 = 0$, тобто недоцільно робити відрахування на розширення виробництва. Цей беззаперечний висновок можна одержати і за інших міркувань: прибуток буде максимальним без будь-яких відрахувань по завершенні контрольованого проміжку в п'ять років. Отже, маємо:

$$Z = Z(1–4) + 4m_5. \tag{12}$$

Оптимізуємо прибуток за попередній, четвертий рік. Стан виробництва за рік визначається величинами m_4 та $Z(1–3)$. Прибуток за перші чотири роки дорівнює:

$$Z(1–4) = Z(1–3) + \varphi_4. \tag{13}$$

Враховуючи, що $\lambda_5^* = 0$ та залежність (12), маємо $Z = Z(1–3) + \varphi_4 + 4m_5$. Використавши (8) та (9), одержимо залежність:

$$Z = Z(1–3) + 4(1 - \lambda_4)m_4 + 4(1 + 0{,}4\lambda_4)m_4 = Z(1–3) + 8m_4 - 2{,}4\lambda_4 m_4. \tag{14}$$

Для визначення оптимального розв'язку для цього кроку скористаємося принципом оптимальності: величину λ_4 знаходимо, використовуючи залежність (14), в якій враховано вибір оптимального значення $\lambda_5^* = 0$ для останнього кроку. З (14) маємо, що значення Z буде найбільшим за умови $\lambda_4^* = 0$ по завершенні четвертого кроку. Отже, цільова функція Z, яка визначає сумарний прибуток, дорівнює:

$$Z = Z(1–3) + 8m_4. \tag{15}$$

Тепер прийдемо до початку третього року, в якому стан виробництва визначається параметрами m_3 та $Z(1–2)$. Враховуючи, що $Z(1–3) = Z(1–2) + \varphi_3$, і скориставшись (15), а потім (6) та (7), за умови $\lambda_4^* = 0$ і $\lambda_5^* = 0$ маємо:

$$Z = Z(1–2) + 12m_3 – 0,8\lambda_3 m_3. \quad (16)$$

Згідно з принципом оптимальності, необхідно знайти оптимальний розв'язок для третього року. Аналізуючи (16), маємо $\lambda_3^* = 0$. Отже, цільова функція Z, яка визначає максимальний прибуток за п'ять років, визначається залежністю:

$$Z = Z(1–2) + 12m_3. \quad (17)$$

На завершення другого року виробничий процес характеризується величинами m_2 та $Z(1)$. Скориставшися, що $Z(1–2) = Z(1) + \varphi_2$, та залежністю (17), а потім (5) і (4), маємо:

$$Z = Z(1) + 16m_2 + 0,8\lambda_2 m_2. \quad (18)$$

З (18) видно, що умовно оптимальний розв'язок для другого року буде досягнуто за умови $\lambda_2^* = 1$, а цільова функція Z представлена залежністю

$$Z = Z(1) + 16,8m_2. \quad (19)$$

Тепер проаналізуємо п'ятий від кінця крок зворотного ходу, тобто перший рік. З (2) маємо:

$$Z(1) = \varphi_1 = 4(1 - \lambda_1)m_1. \quad (20)$$

Щоб знайти оптимальний розв'язок, необхідно, згідно з принципом оптимальності, використати умовно оптимальні розв'язки попередньої оптимізації, представлені залежністю (19). Враховуючи (20) та (3), маємо:

$$Z = 20,8m_1 + 2,72\lambda_1 m_1. \quad (21)$$

Цільова функція Z, яка характеризує прибуток за п'ять років (21), досягне найбільшого значення за умови $\lambda_1^* = 1$:

$$Z^* = 23,52m_1. \quad (22)$$

Таким чином, знайдено умовно оптимальні розв'язки на всіх п'яти кроках (роках), наведено формули обчислення відповідних їм умовно максимальних прибутків (12), (15), (17), (22). Залишилося проаналізувати процес розширення виробництва за умови одержання максимального зиску в прямому напрямі та вказати оптимальну стратегію: $\lambda_1^* = \lambda_2^* = 1$; $\lambda_3^* = \lambda_4^* = \lambda_5^* = 0$. За такої стратегії використання прибутку для розширення виробництва максимальний прибуток буде $Z^* = 2,352C$, де C визначає величину початкової інвестиції.

На рис. 15.2 вказано закономірності зміни кількості одиниць обладнання та обсягів прибутку по роках.

Z					5-й рік $m_5 = m_4(1+0,4\lambda_4)$	$\varphi_5=4m_5(1-\lambda_5)=4m_1(1+0,4\lambda_1)(1+0,4\lambda_2)(1+0,4\lambda_3)(1+0,4\lambda_4)(1-\lambda_5)$
	Z(1-4)				4-й рік $m_4 = m_3(1+0,4\lambda_3)$	$\varphi_4=4m_4(1-\lambda_4)=4m_1(1+0,4\lambda_1)(1+0,4\lambda_2)(1+0,4\lambda_3)(1-\lambda_4)$
		Z(1-3)			3-й рік $m_3 = m_2(1+0,4\lambda_2)$	$\varphi_3=4m_3(1-\lambda_3)=4m_1(1+0,4\lambda_1)(1+0,4\lambda_2)(1-\lambda_3)$
			Z(1-2)		2-й рік $m_2 = m_1(1+0,4\lambda_1)$	$\varphi_2=4m_2(1-\lambda_2)=4m_1(1+0,4\lambda_1)(1-\lambda_2)$
				Z(1)	1-й рік $m_1 = C/10$	$\varphi_1=4m_1(1-\lambda_1)$
					Стан виробництва кожного року	Оцінка стану по завершенні кожного року

Рис. 15.2.

Тепер розв'яжемо сформульовану задачу з використанням класичних методів математичного аналізу. Спочатку будуємо функцію сумарного прибутку Z за п'ять років як функцію, яка залежить від часток відрахувань на розширення виробництва по завершенні кожного року.

Формули (2), (4), (6), (8) і (10) визначають залежність величини прибутку від часток відрахувань λ_1, λ_2, λ_3, λ_4, λ_5 та обсягів обладнання, яке було використане протягом відповідного року. Почергове використання формул (3), (5), (7) і (9) дає можливість визначити обсяги обладнання поточного року m_i $(i = \overline{2, 5})$ через обсяг обладнання m_1 вихідного року. Запишемо функцію $Z(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ сумарного прибутку за п'ять років:

$$Z = 4m_1(1 - \lambda_1) + 4m_1(1 + 0,4\lambda_1)(1 - \lambda_2) + 4m_1(1 + 0,4\lambda_1)(1 + 0,4\lambda_2)(1 - \lambda_3) + \\ + 4m_1(1 + 0,4\lambda_1)(1 + 0,4\lambda_2)(1 + 0,4\lambda_3)(1 - \lambda_4) + \\ + 4m_1(1 + 0,4\lambda_1)(1 + 0,4\lambda_2)(1 + 0,4\lambda_3)(1 + 0,4\lambda_4)(1 - \lambda_5).$$

Частки відрахувань λ_1, λ_2, λ_3, λ_4, λ_5 необхідно вибрати так, щоб функція Z досягала *найбільшого* значення за умови, що кожна з величин λ_i $(i = \overline{1, 5})$ може бути в межах [0; 1]. Знайдемо частинні похідні $Z(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ та виконаємо їх дослідження.

$$\frac{\partial Z}{\partial \lambda_1} = 4m_1(-1+0{,}4(1-\lambda_2)+0{,}4(1+0{,}4\lambda_2)(1-\lambda_3)+0{,}4(1+0{,}4\lambda_2)(1+0{,}4\lambda_3)(1-\lambda_4)+$$

$$+0{,}4(1+0{,}4\lambda_2)(1+0{,}4\lambda_3)(1+0{,}4\lambda_4)(1-\lambda_5));$$

$$\frac{\partial Z}{\partial \lambda_2} = 4m_1(-(1+0{,}4\lambda_1)+0{,}4(1+0{,}4\lambda_1)(1-\lambda_3)+0{,}4(1+0{,}4\lambda_1)(1+0{,}4\lambda_3)(1-\lambda_4)+$$

$$+0{,}4(1+0{,}4\lambda_1)(1+0{,}4\lambda_3)(1+0{,}4\lambda_4)(1-\lambda_5));$$

$$\frac{\partial Z}{\partial \lambda_3} = 4m_1(-(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)+0{,}4(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)(1-\lambda_4)+$$

$$+0{,}4(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)(1+0{,}4\lambda_4)(1-\lambda_5));$$

$$\frac{\partial Z}{\partial \lambda_4} = 4m_1(-(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)(1+0{,}4\lambda_3)+$$

$$+0{,}4(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)(1+0{,}4\lambda_3)(1-\lambda_5));$$

$$\frac{\partial Z}{\partial \lambda_5} = 4m_1(-(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)(1+0{,}4\lambda_3)(1+0{,}4\lambda_4)).$$

За умови $\lambda_i \in [0;1]$ $(i=\overline{1;5})$ $\frac{\partial Z}{\partial \lambda_5}$ буде мати від'ємні значення, тобто якщо розглядати зміну $Z(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)$ лише від аргументу λ_5, а решту аргументів зафіксувати, то при збільшенні величини λ_5 значення функції Z будуть спадати. Тому її значення буде екстремальним при $\lambda_5 = 0$. Врахувавши значення $\lambda_5 = 0$, для $\frac{\partial Z}{\partial \lambda_4}$ одержимо залежність

$$\frac{\partial Z}{\partial \lambda_4} = 4m_1(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)(1+0{,}4\lambda_3)(-1+0{,}4).$$

За наведених міркувань (таких, як і для $\frac{\partial Z}{\partial \lambda_5}$) одержимо $\lambda_4 = 0$. Використавши $\lambda_5 = 0$, $\lambda_4 = 0$, для $\frac{\partial Z}{\partial \lambda_3}$ одержимо залежність $\frac{\partial Z}{\partial \lambda_3} = 4m_1(1+0{,}4\lambda_1)(1+0{,}4\lambda_2)(-1+0{,}8)$. За міркувань, наведених раніше стосовно $\frac{\partial Z}{\partial \lambda_5}$ та $\frac{\partial Z}{\partial \lambda_4}$, одержимо $\lambda_3 = 0$. Підставивши обчислені значення $\lambda_3 = \lambda_4 = \lambda_5 = 0$ у формули, які визначають $\frac{\partial Z}{\partial \lambda_1}$ та $\frac{\partial Z}{\partial \lambda_2}$, маємо:

$$\frac{\partial Z}{\partial \lambda_1} = 4m_1(-1+0{,}4(1-\lambda_2+1+0{,}4\lambda_2+1+0{,}4\lambda_2+1+0{,}4\lambda_2)) = 4m_1(0{,}6+0{,}08\lambda_2);$$

$$\frac{\partial Z}{\partial \lambda_2} = 4m_1(-1+0{,}4(-\lambda_1+1+0{,}4\lambda_1+1+0{,}4\lambda_1+1+0{,}4\lambda_1)) = 4m_1(0{,}2+0{,}08\lambda_1).$$

При невід'ємних значеннях λ_1 та λ_2 похідні $\frac{\partial Z}{\partial \lambda_1}$ та $\frac{\partial Z}{\partial \lambda_2}$ будуть додатними, отже, функція прибутку Z зростатиме при збільшенні λ_1 та λ_2. Таким чином, необхідно взяти найбільш можливі значення $\lambda_1 = 1$; $\lambda_2 = 1$. Одержаний результат – величини $\lambda_1 = \lambda_2 = 1$, $\lambda_3 = \lambda_4 = \lambda_5 = 0$ відрахувань від щорічних обсягів прибутку тотожні значенням, досягнутим при використанні методу динамічного програмування, але обсяг роботи при використанні методу пошуку оптимального розв'язку засобами класичного аналізу істотно більший.

Наведена задача наочно показує, як, використовуючи метод динамічного програмування, замінити розв'язання вихідної складної задачі на пошук максимуму функції п'яти змінних розв'язанням кількох простих задач з однією змінною.

При аналізі розв'язку розглянутої задачі наголошувалося, що при використанні методу ДП побічні результати часто мають не менш важливе значення, ніж основний результат. *Основний результат:* прибуток за п'ять років буде максимальним при даній ефективності виробництва, якщо в перші два роки весь прибуток використовувати для розширення виробництва, а в наступні три роки не розширювати його. *Читачеві пропонується дати мотивоване пояснення доцільності такої інвестиційної програми.* Розглянемо тепер *додаткові результати розв'язання задачі.* Якщо в (18) вибрати $\lambda_2 = 0$, а в (20) $\lambda_1 = 0$ (це означає, що в перші два роки виробництво не розширюється через відрахування прибутку), то одержимо $Z^0 = 20m_1 = 2C$. Порівнюючи цю величину з максимальною $Z^* = 2{,}352C$, доходимо висновку: якщо не розширювати виробництво, то прибуток підприємства буде на 17,6% менше Z^*, який воно одержить за умов оптимальної стратегії. Прийнявши $\lambda_1 = \lambda_2 = 1$ і $\lambda_3 = \lambda_4 = \lambda_5 = 0$ у формулах (3), (5), (7) і (9), одержуємо $m_2 = 1{,}4m_1$; $m_3 = m_4 = m_5 = 1{,}96m_1$. Таким чином, використання оптимальної стратегії забезпечує розширення виробничих потужностей з даного виду обладнання на 96%.

Необхідно наголосити, що розв'язання оптимізаційної задачі будувалось як послідовність розв'язання взаємопов'язаних оптимізаційних задач однієї змінної – частки відрахування прибутку на розширення виробництва. Результат для попереднього року використовувався для розв'язання рівняння для наступного року і т. д. Така методика не випадкова. Це є математичне втілення принципу оптимальності і становить основу визначального методу розв'язання задач ДП – методу рекурентних функціональних залежностей.

15.3. Загальна постановка задачі динамічного програмування

Розглянемо умови доцільності користування методом динамічного програмування та відповідні основні поняття. Досліджується перебіг деякого *керованого процесу*, тобто такого, на стан і розвиток якого можна впливати через певні проміжки. В економічних процесах управління означає перерозподіл коштів, заміну обладнання, визначення обсягів поставок сировини на певний період і т. ін. Приймається, що процес управління можна реалізувати дискретно за N кроків (етапів). Віділення етапів управління може бути як природним (рік, квартал, місяць тощо), так і навмисним. Будь-яку багатоетапну задачу можна розв'язувати по-різному: або відразу шукати всі елементи розв'язку для всіх N етапів, або знаходити оптимальне управління поетапно, на кожному етапі визначаючи розв'язок стосовно лише цього етапу. Як правило, поетапну оптимізацію легше реалізувати, ніж глобальну, особливо при значній кількості етапів і змінних параметрів.

Сама *ідея* поступової, *поетапної* (покрокової) *оптимізації* визначає метод динамічного програмування. Параметри математичних моделей, які використовуються при дослідженні таких керованих процесів, доцільно розбити на дві множини: параметри стану та параметри управління. За допомогою *параметрів стану* визначають ті властивості процесу, для дослідження яких була розбудована відповідна математична модель. За допомогою *параметрів управління* визначають ті фактори, які можуть впливати на стан процесу. Наприклад, у розглянутій задачі про одержання максимального прибутку завдяки розширенню виробництва кількість комплектів виробничого обладнання була параметром стану, а величина частки прибутку, яку використовували для збільшення обсягів обладнання та прибутку, – параметром управління. У загальному випадку розглядаємо такі процеси, математичні моделі перебігу яких можна представити як послідовність деяких дискретних етапів (кроків). У реально діючих великих економічних системах такі процеси обумовлені необхідністю планування в часі більш-менш незалежних практичних проблем.

Для зручності викладу введемо певні позначення. Кількість етапів (кроків) позначимо символом N, використовуючи індекс i для виділення поточного етапу. На кожному i-му етапі процес може бути в різних станах $\{\bar{S}_j\}_i$, кожний з яких характеризується скінченною множиною параметрів. Множину параметрів, використовуваних для характеристики певного стану $\bar{S}_j$, доцільно розглядати як компоненти деякого вектора $\bar{S} = (S_1, S_2, \ldots, S_l)$, де l – кількість параметрів, обраних для характеристики стану. На кожному з N досліджуваних етапів система може бути в кількох станах.

Перебіг процесу визначається певною послідовністю переходів з одного стану в інший. Якщо процес на i-му етапі перебував у деякому стані $\bar{S}_{j,i}$, то наступний стан $\bar{S}_{k,i+1}$ на (i+1)-му кроці визначається не лише попереднім станом, а й вибором

певного управління при досягненні $\bar{S}_{j,i}$ ($j=\overline{1,m_i}$; $i=\overline{1,N}$). У загальному випадку будь-яке управління на будь-якому кроці (етапі) доцільно розглядати як p-мірний вектор $\bar{u} = (u_1, u_2, \ldots, u_p)$. Числові значення компонент вектора управління будуть залежати як від вихідного стану $\bar{S}_{j,i}$ на i-му кроці, так і від наступного стану на (i+1)-му кроці $\bar{S}_{k,i+1}$ ($k=\overline{1,m_{i+1}}$), тобто вектор $\bar{u}$ визначається чотирма індексами j, i, k, i+1 і має бути вибраний з певної множини допустимих управлінь.

У подальшому для спрощення записів вектори можливих поточного стану та управління будемо позначати лише одним індексом, співставляючи їх певному кроку (етапу), тобто щодо стану $\bar{S}_i$, маємо на увазі один із можливих станів множини $\{\bar{S}_j\}_i$, а щодо вектора $\bar{u}_i$ - один із можливих векторів множини $\{\bar{u}_{j,i}\}_{k,i+1}$, ($i=\overline{1,N}$, $j=\overline{1,m_i}$, $k=\overline{1,m_{i+1}}$). За уважного читання таке спрощення в позначеннях не спотворює змісту викладу.

У розглянутій задачі пошуку оптимальних стратегій інвестування вектор $\bar{S}_i$ мав лише одну компоненту – кількість комплектів використовуваного обладнання, але допустимі значення цієї величини були 1, 2, 3, …

У задачі пошуку найкоротшого шляху вектор $\bar{S}_i$ характеризував кількість пунктів, які можна досягти з певного пункту за один перегін на i-му етапі розв'язання задачі, тобто кількість шляхів, які починаються в досягнутому пункті.

На рис. 15.3 схематично кругами зображені можливі стани на кожному етапі,

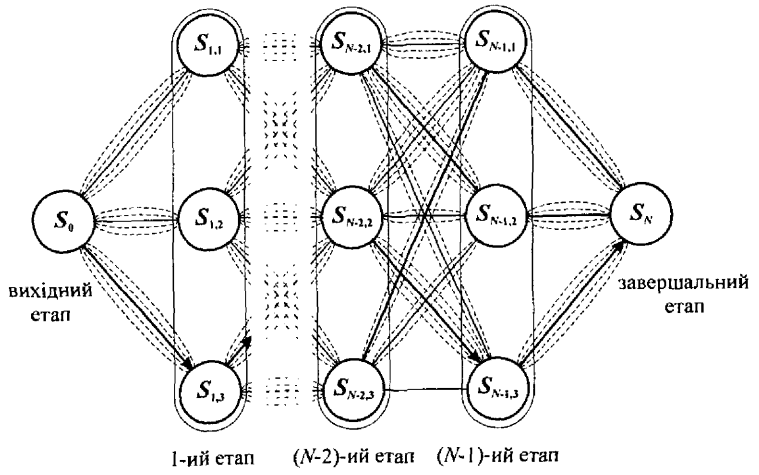


Рис. 15.3.

пунктирними або суцільними лініями – можливі переходи від одного стану до іншого за вибору певного управління, тобто лінія відповідає певному управлінню, за якого система переходить від стану до стану.

У задачі розподілу коштів був лише один параметр управління: λ – частка відрахувань на розширення виробництва, можливі значення якої належали проміжку [0; 1]. У задачі пошуку найкоротшого шляху параметри управління ототожнювалися з вибором певного шляху з множини шляхів, які починалися в досягнутому пункті.

Таким чином, стан процесу на i-му кроці (етапі) визначається певною функціональною залежністю від стану на попередньому кроці та значеннями параметрів управління на початку чергового кроку, що в загальному випадку можна записати так: $\overline{S}_i = \varphi\,(\overline{S}_{i-1}, \overline{u}_i)$. Процес управління моделюється як вибір за кожного можливого j-го стану на i-му етапі певного k мірного вектора $\overline{u}_{j,i}$ з деяких допустимих множин векторів $\{\overline{u}\}_{j,i}$. Для подальшого спрощення записів цей вибраний вектор слід позначити $\overline{u}_i$. Позначимо $\overline{U} = (\overline{u}_1, \overline{u}_2, ..., \overline{u}_N)$ множину послідовності управлінь, які переводять систему зі стану $\overline{S}_0$ у стан $\overline{S}_N$, що схематично представлено на рис. 15.4.

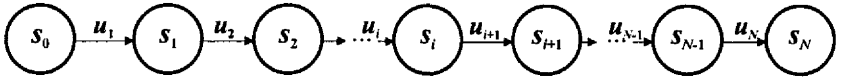


Рис. 15.4.

Будь-яку послідовність $\overline{U} = (\overline{u}_1, \overline{u}_2, ..., \overline{u}_N)$, яка переводить систему зі стану $\overline{S}_0$ у стан $\overline{S}_N$, називають *стратегією*, а вектори $\overline{u}_i$ – її складовими. Схематично взаємозв'язок між послідовністю станів і управлінь представлений на рис. 15.3.

Ефективність вибору послідовності управлінь $\overline{U}$ (стратегії) оцінюється за вибраним критерієм певною цільовою функцією Z:

$$Z = F\,(\overline{S}_0, \overline{U}). \tag{23}$$

Модель динамічного програмування можна використовувати в тих випадках, коли є підстави прийняти такі допущення стосовно досліджуваної системи.

1. Стан $\overline{S}_i$ системи в кінці i-го кроку визначається лише попереднім станом $\overline{S}_{i-1}$ та управлінням $\overline{u}_i$ на i-му кроці і не залежить від попередніх станів та управлінь. Ця властивість називається *"відсутність післядії"*. Сформульоване припущення записується як залежність

$$\overline{S}_i = \varphi_i(\overline{S}_{i-1}, \overline{u}_i) \;\; i = 1, 2, \ldots, N, \tag{24}$$

яку називають *рівнянням стану*.

2. *Цільова функція* (23) *є адитивною стосовно кожного кроку* (етапу) і залежить від того, яким був стан системи $\overline{S}_{i-1}$ на початку етапу та яке було обране управління. Якщо позначити показник ефективності i-го кроку через

$$Z_i = f_i\,(\overline{S}_{i-1}, \overline{u}_i),\; i = 1, 2, \ldots, N, \tag{25}$$

то цільова функція (23) буде представлена так:

$$Z = \sum_{i=1}^{N} f_i\,(\overline{S}_{i-1}, \overline{u}_i). \tag{26}$$

Метод ДП можна використовувати також при розв'язанні задач з так званою *"мультиплікативною" цільовою функцією*, представленою як добуток показників ефективності на кожному кроці управління, тобто

$$Z = \prod_{i=1}^{N} f_i\,(\overline{S}_{i-1}, \overline{u}_i).$$

Ми будемо розглядати лише адитивні цільові функції.

За названих умов *задача поетапної оптимізації*, тобто задача ДП, формулюється так: *визначити таку допустиму стратегію управління* $U^* = (\overline{u}_1, \overline{u}_2, ..., \overline{u}_N)$, *яка переводить систему S зі стану* $\overline{S}_0$ *у стан* $\overline{S}_N$ *і за якої цільова функція* (26) *досягає екстремального значення* (або найменшого, або найбільшого, як того вимагає зміст задачі).

Доцільно ще раз наголосити на взаємозв'язок використовуваних понять стану та управління станом. За відомого вихідного стану $\overline{S}_0$ системи *вибір будь-якої допустимої послідовності управлінь* при переході від стану до стану *породжує відповідну послідовність станів системи.* Кожний наступний стан за відомого попереднього стану та обраного управління визначається залежністю (24), функціональне представлення якої залежить від природи досліджуваної системи. У загальному вигляді мовою аналітичних співвідношень задачу ДП можна сформулювати так: знайти стратегію $U^* = (\overline{u}_1, \overline{u}_2, ..., \overline{u}_N\}$, при якій функція

$$Z = \sum_{i=1}^{N} f_i\,(\overline{S}_{i-1}, \overline{S}_i)$$

досягає екстремуму за таких умов: $\overline{S}_0$ – відомий вектор вихідного стану системи; $\overline{S}_i = \varphi_i(\overline{S}_{i-1}, \overline{u}_i)$ – відомі залежності визначення стану системи на будь-якому кроці за будь-якого допустимого управління $\overline{u}_i$. Множини можливих станів і управлінь на кожному етапі $\overline{S}_i$ та $\overline{u}_i$ можуть моделюватися як множини чисел або n-мірних векторів (упорядкованих наборів чисел).

Наголосимо на **особливостях моделі ДП**:

1) задача оптимізації може бути розв'язана за N керованих кроків;
2) цільова функція адитивна на кожному кроці можливого управління (дорівнює сумі значень цільової функції на кожному кроці);
3) вибір управління на кожному i-му кроці залежить лише від досягнутого на попередньому етапі стану системи і не впливає на її стан за попередні кроки (відсутній зворотний зв'язок);
4) стан $\overline{S}_i$ після i-го кроку управління залежить тільки від попереднього стану $\overline{S}_{i-1}$ та управління $\overline{u}_i$ (відсутність післядії);
5) на кожному кроці як управління $\overline{u}_i$, так і стан $\overline{S}_i$ залежать лише від скінченної множини параметрів.

15.4. Принцип оптимальності та структура рівняння Белмана

Принцип оптимальності для задач ДП уперше був сформульований Р. Белманом у 1953 р. *За будь-якого стану системи, досягнутого за певну кількість кроків управління, на наступному кроці треба вибирати управління так, щоб воно разом з оптимальним управлінням для всіх майбутніх кроків, необхідних для завершення процесу , починаючи з даного кроку, забезпечувало оптимальний результат до завершення всього процесу.* Белман обґрунтував і умови, за яких буде виконуватися цей принцип. Основна вимога полягає в тому, щоб процес *управління був без зворотного зв'язку, тобто управління на даному кроці не може впливати на стан системи на попередніх кроках.* Процес переходу від будь-якого довільного проміжного стану в інший проміжний стан називається *частинним процесом. Принцип оптимальності* стверджує, що для будь-якого процесу без зворотного зв'язку оптимальне управління системою в цілому містить у собі оптимальне управління для будь-якого частинного процесу стосовно його вихідного стану. Тобто якщо відомо, що певний стан, досягнутий на будь-якому кроці, є станом, який відповідає оптимальному управлінню процесом у цілому, то вимога оптимізації процесу від досягнутого стану до завершального робить на відповідних кроках подальше управління тотожним оптимальному. Множина оптимальних управлінь на етапах від можливих станів системи до завершального завжди містить у собі ті частинні послідовності оптимального управління системою в цілому, які відповідають станам, досягнутим за оптимального управління. Саме з цієї причини при розв'язанні *задачі пошуку найкоротшого шляху з кожного пункту мережі доріг* обирався оптимальний шлях досягнення кінцевого пункту (рис. 15.1 б)). Такі шляхи були названі умовно оптимальними. *Умовно оптимальні шляхи* для всіх пунктів, *через які проходить оптимальний шлях від початкового пункту до кінцевого, співпадали з відповідними частинами оптимального шляху.*

Рівняння Белмана. Задачу ДП із заданими або вибраними кількістю кроків і вихідними та завершальними умовами розщеплюють на послідовність простих задач. Розв'язання кожної такої спрощеної задачі називається кроком або етапом розв'язання досліджуваної задачі ДП. Домовимося відлік кроків завжди виконувати *від вихідного стану до завершального*, якщо не вказано протилежне.

На кожному кроці при досягненні будь-якого допустимого стану $\overline{S}_{i-1}$ системи наступне управління $\overline{u}_i$ необхідно вибирати з урахуванням усіх його наслідків для подальшого перебігу процесу, тому що цей вибір впливає на наступний стан $\overline{S}_i$ і на подальше управління $\overline{u}_{i+1}$. Ця вимога обумовлена принципом оптимальності і стосується кожного кроку, за винятком останнього. На останньому, N-му кроці за будь-якого стану $\overline{S}_{N-1}$ оптимальне управління з множини допустимих управлінь вибирають таким, щоб досягти стану $\overline{S}_N$, враховуючи ситуацію лише на цьому останньому кроці, якщо стан $\overline{S}_N$ заданий.

Розглянемо саме останній, N-ий крок: $\overline{S}_{N-1}$ – один із можливих станів системи на початку цього кроку; $\overline{S}_N$ – заданий завершальний стан; $\{\overline{u}_N\}$ – множина управлінь на цьому кроці; $f_N(\overline{S}_{N-1}, \overline{u}_N)$ – відповідний доданок цільової функції N-го кроку. Відповідно до принципу оптимальності $\overline{u}_N$ необхідно вибирати так, щоб за будь-якого стану $\overline{S}_{N-1}$ одержати екстремум для відповідного доданку цільової функції на цьому кроці (покажчика якості управління).

Позначимо через $Z^*_{j,N}(\overline{S}_{N-1})$ екстремум цільової функції – покажчик ефективності N-го кроку за обраного критерію оцінки якості управління системою (25). Приймаємо, що перед початком останнього (завершального) кроку система може бути в будь-якому стані $\overline{S}_{N-1}$ з множини допустимих станів $\overline{S}_{j,N-1}$ $(j = \overline{1, m_{N-1}})$, а управління на останньому кроці було оптимальним.

$Z^*_{j,N}$ називають *умовним екстремумом цільової функції на N-му кроці*, тобто

$$Z^*_{j,N}(S_{N-1}) = \underset{\{\overline{u}_j\}_N}{extr}\, f_N(\overline{S}_{N-1}, \overline{u}_{j,N}). \qquad (27)$$

Екстремум визначається на множині всіх допустимих управлінь і станів на останньому кроці. Управління $\overline{u}_N$, за якого досягається екстремум (27), залежить від стану $\overline{S}_{j,N}$ $(j = \overline{1, m_{N-1}})$ і називається *умовним оптимальним управлінням на N-му*

кроці (див. рис. 15. 4, на якому умовні оптимальні управління позначені суцільною лінією). Позначимо його як $\bar{u}^*_{j,N}(\bar{S}_{N-1})$. Який вигляд матиме $u^*_{j,N}(\bar{S}_{N-1})$, залежить від того, як у конкретній задачі ДП моделюється стан системи. Умовно оптимальних управлінь і умовно оптимальних екстремумів $Z^*_{j,N}$ буде стільки, скільки можливих станів $\bar{S}_{j,N-1}$ на N-му кроці. Наприклад, для задачі 15.2.1 умовне оптимальне управління на останньому кроці (досягнення кінцевого пункту з пунктів 8 та 9) буде представлено вибіркою серед двох елементів – шляхів (8; 10) та (9; 10). Для пунктів 5, 6 та 7 умовне оптимальне управління на передостанньому кроці – це вибірка відповідних найкращих шляхів (5; 8; 10), (6; 8; 10) та (7; 9; 10) з множини шляхів {(5; 8; 10); (5; 9; 10); (6; 8; 10); (6; 9; 10); (7; 8; 10); (7; 9; 10)}.

У загальному випадку можливі стани системи на кожному кроці та умовні оптимальні управління не можуть бути представлені певними аналітичними залежностями, тому що математичний апарат, використаний при постановці відповідних задач ДП, різний. Але для зручності будемо користуватися термінологією, що результатом реалізації пошуку умовного оптимального розв'язку будуть множини значення функції $Z^*_{j,N}(\bar{S}_{N-1})$ та векторів $\{\bar{u}^*_j\}_N = (u^*_{1,j}, u^*_{2,j}, ..., u^*_{k,j})_N$ $(j = \overline{1, m_{N-1}})$. У цьому випадку термін "функція" означає певний алгоритм, який дозволяє за певних вхідних даних одержати однозначний набір вихідних. Маємо вхідні дані $\bar{S}_{j,N-1}$ (множину можливих станів системи). По виконанні умов (27), використовуючи множину допустимих управлінь $\{\bar{u}_j\}_N$ $(j = \overline{1, m_{N-1}})$, необхідно вибрати управління $\bar{u}^*_{j,N}$, яке буде оптимізувати $Z_N(\bar{S}_{j,N-1})$ за кожного допустимого стану.

Визначивши множину умовних оптимальних управлінь $\bar{U}^*_N = \{\bar{u}_j\}_N$ $(j = \overline{1, m_{N-1}})$ на N-му кроці, розглянемо розв'язання задачі пошуку умовного оптимального управління з урахуванням передостаннього кроку N-2. За будь-яких станів $\bar{S}_{N-2}$ системи та управлінь $\bar{u}_{N-1}$, з урахуванням оптимального управління на N-му кроці, значення цільової функції визначається залежністю:

$$Z = f_{N-1}(\bar{S}_{N-2}, \bar{u}_{N-1}) + Z^*_N(\bar{S}_{N-1}). \qquad (28)$$

Згідно з принципом оптимальності, для будь-якого стану $\bar{S}_{N-2}$ розв'язок на (N–1)-му кроці треба знайти так, щоб він разом з розв'язком на останньому, N-му кроці гарантував екстремум цільової функції для двох останніх кроків. Тобто знайти екстремум (28) на множині всіх допустимих управлінь $\bar{u}_{N-1}$ з урахуванням усіх можливих станів $(\bar{S}_j)_{N-2}$. Екстремум цієї суми залежить від кожного можливого стану $\bar{S}_{N-2}$ і називається умовним екстремумом цільової функції за умови оптимального управління на двох останніх кроках. Позначимо названий екстремум

як $Z^*_{N-1}(\bar{S}_{N-2})$, а відповідне управління як $\bar{u}^*_{N-1}(\bar{S}_{N-2})$ і назвемо його умовним оптимальним управлінням на (N–1)-му кроці (див. рис. 15. 3), де множини можливих управлінь від певного стану до наступного умовно позначені штриховою лінією, умовно оптимальні управління – суцільною лінією, а оптимальне – стрілками суцільних ліній. Отже:

$$Z^*_{N-1}(\bar{S}_{N-2}) = \underset{\{\bar{u}_{N-1}\}}{extr}\{f_{N-1}(\bar{S}_{N-2}, \bar{u}_{N-1}) + Z^*_N(\bar{S}_{N-1})\}. \qquad (29)$$

Доцільно наголосити, що вираз у фігурних дужках (29) залежить від $\bar{S}_{N-2}$ та $\bar{u}_{N-1}$, бо $\bar{S}_{N-1}$ можна визначити, скориставшися рівнянням стану (24) за умови $i = N$–1: $\bar{S}_{N-1} = \varphi_{N-1}(\bar{S}_{N-2}, \bar{u}_{N-1})$, і підставити у $Z^*_N(\bar{S}_{N-1})$.

По завершенні пошуків умовно оптимальних розв'язків (29) для кожного можливого стану $(\bar{S}_j)_{N-2}$ $(j = \overline{1, m_{N-2}})$ одержуємо:

а) двоелементну послідовність умовно оптимальних управлінь $\bar{u}^*_{N-1}(\bar{S}_{N-2})$, $\bar{u}^*_N(\bar{S}_{N-1})$, за допомогою яких виконується перехід від кожного зі станів $(\bar{S}_j)_{N-2}$ $(j = \overline{1, m_{N-2}})$ до завершального стану $\bar{S}_N$ так, щоб функція $Z_{N-1}(\bar{S}_{N-2})$, яка являє собою суму значень показника ефективності (25) на названих переходах, досягла екстремального значення;

б) множину відповідних екстремальних значень $Z^*_{N-1}(\bar{S}_{N-2})$.

На рис. 15.1, а такі умовно оптимальні управління, як результат вибору переходів від пунктів 5, 6, 7, позначені стрілками з подвійною лінією, а оптимальні значення функції Z^*_{N-1} наведені жирним шрифтом внизу кругів, які зображають пункти мережі.

По визначенню умовно оптимальних управлінь для двох останніх кроків розглядаємо використання принципу оптимізації для трьох останніх кроків і т. д.

Перейдемо до побудови розв'язувальних рівнянь *у загальному випадку*.

Позначимо $Z^*_{N-i+1}(\bar{S}_{N-i})$ умовний екстремум цільової функції (26), отриманий за умовно оптимального управління на останніх кроках за умови, що на початок i-го кроку система перебувала в одному з допустимих станів $(\bar{S}_j)_{N-i}$ $(j = \overline{1, m_{N-i}})$. Пам'ятаючи про цільову функцію, маємо

$$Z^*_{N-i+1}(\bar{S}_{N-i}) = extr \sum_{k=N-i+1}^{N} f_k(\bar{S}_{k-1}, \bar{u}_k).$$

Умовно оптимальне значення цільової функції $Z^*_{N-i+1}(\bar{S}_{N-i})$ (26) як функції на останніх i+1 кроках (відлік кроків виконуємо у зворотному напрямку, але хід процесу – у природному) доцільно представити так:

$$Z^*_{N-i+1}(\bar{S}_{N-i+1}) = \underset{\{\bar{u}_{N-i+1}\}}{extr}\left\{f_{N-i+1}(\bar{S}_{N-i-1};\bar{u}_{N-i+1}) + Z^*_{N-i+1}(\bar{S}_{N-i})\right\} \ (i=\overline{1;N-1}). \quad (30)$$

За принципом оптимальності для кожного можливого стану $\bar{S}_{j,N-i-1}$ ($j=\overline{1;m_{N-i-1}}$) на (i+1)-му кроці від завершального стану $\bar{S}_N$ умовно оптимальне управління u^*_{N-i} необхідно вибирати так, щоб, починаючи з (i+1)-го кроку до завершення процесу, цільова функція $Z_{N-i}(\bar{S}_{N-(i+1)})$ (26) у представленні (30) досягла екстремуму на множині з допустимих управлінь, обумовлених множиною можливих станів.

Рівняння, побудовані за залежністю (30), називаються **рівняннями Белмана**. Вони являють собою математичний запис принципу оптимальності. Рівняння, побудовані за залежністю (30) з урахуванням залежності (24), дозволяють процедуру розв'язання задачі ДП привести до послідовності пошуків умовних оптимальних розв'язків на кожному кроці, виконуючи відлік від завершення процесу, і на завершення знайти оптимальну стратегію управління процесом у цілому. За залежності (30) при визначенні "наступного" значення функції Z_{N-i} як аргумент використовується "попереднє" значення Z_{N-i+1}. Нагадуємо, що відлік кроків "i" виконується у зворотному напрямку. Співвідношення, пов'язані таким чином, називаються рекурентними. Вся послідовність обчислень, яка завершується визначенням $Z_1(\bar{S}_0)$, може бути реалізована, якщо є відомими значення функції Z_{N-i+1} для $i = 1, \ldots, N-1$.

Рівняння (30) мають символічний характер, дають лише загальну принципову схему обчислень при використанні методу динамічного програмування. Для кожної конкретної задачі рівняння (30) має свою специфіку, але в ньому обов'язково мусить зберігатися рекурентність як віддзеркалення основної ідеї розв'язання задач ДП. Для кожної конкретної задачі запис (30) вимагає майстерності та певних навичок, що обумовлює труднощі у використанні методу ДП.

Розв'язок (30) для кожної конкретної задачі ДП, названий *умовно оптимальним*, показує, що слід робити далі в залежності від стану системи, досягнутого на попередньому кроці. Припустимо, що на кожному кроці для кожного можливого стану знайдено умовно оптимальне управління, в тому числі для 1-го кроку. Враховуючи, що вихідний стан $\bar{S}_0$ системи відомий, *умовно оптимальне управління на 1-му кроці*, яке відповідає $\bar{S}_0$, буде *оптимальним управлінням на 1-му кроці*. Якщо використати його на 1-му кроці, то система перейде в певний (конкретний) стан $\bar{S}_1$, але для нього вже знайдено умовно оптимальне управління: воно буде оптимальним для 2-го кроку за умови вихідного стану $\bar{S}_1$. Використавши оптимальне управління на 2-му кроці, переведемо систему в стан $\bar{S}_2$, який буде вихідним для 3-

го кроку, а умовно оптимальне управління, яке відповідало цьому стану, буде оптимальним і т. д. Для задачі 15.2.1 вибір оптимальних управлінь і відповідних оптимальних шляхів серед множин умовно оптимальних показаний на рис. 15.1 б). Загальна схема ситуацій вибору оптимального розв'язку серед множин умовно оптимальних представлена на рис. 15.4.

Таким чином, процедура реалізації методу ДП починається від моменту завершення досліджуваного процесу (операції) до його початку. Це перший хід розв'язання задачі. За цього ходу для кожного кроку знаходять умовні оптимальні управління, які залежать від можливого стану досліджуваного процесу на початок кожного кроку. На першому ході визнається також умовно оптимальний екстремум на послідовності кроків, починаючи з даного і до завершального.

На другому ході (зворотний хід) процедури розв'язання задачі ДП для кожного кроку визначається безумовне оптимальне управління, бо вже однозначно визначено перед кожним кроком такий стан системи, який відповідає оптимальному управлінню від цього стану до завершального. Досліджуваний процес розглядається в послідовності від першого кроку до завершального. Схематично реалізацію етапів і наслідків кожного етапу показано на рис. 15. 4.

15.5. Задача оптимального розподілу коштів на реконструкцію та модернізацію між кількома об'єктами

Реконструкція та модернізація виробництва є важливим економічним завданням. Критерієм його оптимальності служить або збільшення обсягу зиску (задача на пошук максимуму – задача максимізації), або зменшення експлуатаційних витрат на певний плановий період (задача на пошук мінімуму – задача мінімізації). Рішення про реконструкцію та модернізацію приймається на початку певного проміжку експлуатації об'єктів виробництва, і оцінка доцільності заходу стосується саме цього проміжку. З цього виходимо при побудові відповідної моделі задачі.

Припустимо, що для реконструкції та модернізації N об'єктів маємо C коштів. Для кожного об'єкта відома можлива величина збільшення зиску $f_i(x)$ (i = 1, 2, …, N) після виконання запланованого заходу в залежності від витрачених для цього коштів x. Необхідно так розподілити кошти C між об'єктами, щоб величина сумарного зиску від усіх об'єктів після виконання запланованого заходу була максимальною.

З'ясуємо основні поняття ДП стосовно конкретної задачі. Система – це N об'єктів. *Стан системи* – заданий розподіл коштів між об'єктами, оцінка якості – можливий сумарний зиск за відомого розподілу коштів, управління – перерозподіл коштів між об'єктами.

Стосовно кроків розв'язання задачі пропонується така схема: 1-й крок – спрямування певної суми коштів на один об'єкт, 2-й крок – спрямування коштів на два

об'єкти, 3-й крок – на три об'єкти, N-й крок – на N об'єктів. За такої схеми визначення кроків розв'язання задачі його зворотний процес доцільно тлумачити так: на N–1 кроці вже виконано певний розподіл між N–1 об'єктами і необхідно спрямувати кошти на N-й об'єкт, тобто на один об'єкт, який зручно назвати першим, а відповідну суму коштів позначити як y. На N–2 кроці уже виконано певний розподіл коштів між N–2 об'єктами, і необхідно спрямувати кошти на (N–1)-й та N-й об'єкти, які назвемо відповідно другим і першим. Аналогічно тлумачимо N–3 крок і т. д. Зрозуміло, що нумерація об'єктів є умовною. Проаналізуємо функцію:

$$Z_N^*(\bar{S}_{N-1}) = \underset{\{\bar{u}_N\}}{extr}\{f_N(\bar{S}_{N-1}, \bar{u}_N)\}. \qquad (31)$$

Якщо процес завершується на N-му кроці, то $Z_{N+1}^*(\bar{S}_N) = 0$ (немає наступного кроку і немає зиску). Величина $\bar{S}_{N-1}$ у співвідношенні (31) визначає ту величину коштів, яка забезпечує максимум зиску, якщо її спрямовують на останній, N-ий об'єкт.

Зважаючи на те, що кошти – величина скалярна, то будемо записувати їх символами без риски, а оскільки управління є дією з перерозподілу коштів, то його символ опускаємо. Припустимо, що на два останніх об'єкти направлено кошти в сумі y_{N-2}, а на об'єкт N–1 – кошти x, що може забезпечити зиск $f_{N-1}(x)$. Загальний зиск від розподілу коштів між двома останніми об'єктами дорівнюватиме $f_{N-1}(x) + Z_N^*(y_{N-2} - x)$. Екстремальне значення величини названого зиску при розподілі суми y_{N-2} між N та N–1 об'єктами буде досягнуто за такої величини x, при якій

$$Z_{N-1}^*(y_{N-2}) = \underset{0 \le x \le y_{N-2}}{extr}\{f_{N-1}(x) + Z_N^*(y_{N-2} - x)\}. \qquad (32)$$

Знаючи Z_{N-1}^*, можна визначити умови та знайти відповідне значення Z_{N-2}^* при розподілі коштів y між трьома останніми об'єктами і т. д.

Загальне рівняння розв'язання сформульованої задачі можна записати так:

$$Z_{N-i}^*(y_{N-i-1}) = \underset{0 \le x \le y_{N-i-1}}{extr}\{f_{N-i}(x) + Z_{N-i+1}^*(y_{N-i-1} - x)\}, \qquad (33)$$

(i – кількість "останніх" підприємств, між якими виконується розподіл коштів).

Використовуючи залежності (32) та (33), можна послідовно визначити Z_N^*, Z_{N-1}^*, Z_{N-2}^*, Z_2^* та Z_1^* за будь-яких розподілів коштів. Для пошуку оптимального розподілу коштів між об'єктами визначимо величину коштів x_1^*, виділених для першого об'єкта, потім, знаючи суму $C - x_1^*$, величину x_2^* для другого об'єкта і т. д.

Розглянемо конкретний числовий приклад. Оскільки нумерація об'єктів є умовною, а для розв'язання задачі методом ДП за тільки-но розглянутою схемою необхідно починати розподіл коштів для одного об'єкта, потім для двох і т. д., за умови, що для решти об'єктів кошти уже розподілені, - почнемо процес з умовно обраного першого об'єкта, потім розглянемо два об'єкти і т. д.

Припустимо, що між чотирма об'єктами необхідно розподілити 100 у. о. коштів. Обсяги зисків $f(x)$ від величини асигнувань x кожного об'єкта наведені в табл. 15.1 (обсяги асигнувань кратні числу 20).

Таблиця 15.1.

Обсяги асигнувань	Приріст прибутку на підприємствах (f)			
	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_4(x)$
20	10	12	11	16
40	31	26	36	37
60	42	36	45	46
80	62	54	60	63
100	76	78	77	80

За схемою розв'зання задач ДП умовна оптимізація розпочинається з пошуку оптимізації процесу на останньому кроці. Для нашої задачі це означає асигнування одного об'єкта (умовно названого як перший). Для одного об'єкта пошук умовних оптимальних розв'язків виконується згідно з (31). За цієї умови, використовуючи дані табл. 15.1, обчислюємо значення $Z_1^*(x)$, наведені в табл. 15.2 (для першого об'єкту $Z_1^*(x) = f_1(C)$) тобто визначаємо приріст зиску, якщо кошти в певному обсязі будуть направлені лише першому підприємству.

Таблиця 15.2.

Обсяги асигнувань	$f_1(C)$
20	10
40	31
60	42
80	62
100	76

Побудуємо тепер умовно оптимальний пошук для двох об'єктів, враховуючи оптимізацію на першому кроці (для першого об'єкта). За умовою задачі це означає, що кошти направлені на два об'єкти. Відповідне рівняння одержуємо з (32) за умови, що y_{N-2} визначає обсяг коштів, асигнованих для двох об'єктів:

$$Z_2^*(y_2) = \underset{0 \le x \le y_2}{extr}\{f_2(x) + Z_1^*(y_2 - x)\}. \qquad (34)$$

На цьому кроці необхідно знайти $Z_2^*(y_2)$ для всіх можливих комбінацій x та $y_2 \le C$. Розрахунки даного кроку наведені в табл. 15.3.

Таблиця 15 3.

y_2 \ x	0	20	40	60	80	100	$Z_2^*(y_2)$	x_2^*
20	0+10	12+0					12	20
40	0+31	12+10	26+0				31	0
60	0+42	12+31	26+10	36+0			43	20
80	0+62	12+42	26+31	36+10	54+0		62	0
100	0+76	12+62	26+42	36+31	54+10	78+0	78	100

Для кожного можливого обсягу асигнування y_2 двох об'єктів виділено рядок, а для кожного можливого обсягу асигнування другого об'єкта – стовпчик, тобто якщо маємо рядок $y_2 = 40$, то це означає, що на два об'єкти виділено 40 у. о.; якщо маємо стовпчик $x = 20$, то це означає, що на другий об'єкт виділено 20 у. о., а решту 20 на перший об'єкт; за такого розподілу коштів очікуваний зиск буде 22 (12 за рахунок другого об'єкта та 10 за рахунок першого об'єкта). Деякі клітини табл. 15. 3 залишаються пустими, тому, що загальна сума коштів не може бути більшою за виділений обсяг асигнувань. У кожну клітину табл. 15. 3 записуємо суму $f_2(x)$ та $Z_1^*(y_2 - x)$; перший доданок береться з табл. 15. 1, другий – з табл. 15. 2. Наприклад, при розподілі коштів $y_2 = 80$ один із можливих варіантів такий: для другого об'єкта виділяють 60 у. о., тобто $x = 60$, а для першого – решту коштів, тобто 80 – 60 = 20 у.о. За такого розподілу на другому об'єкті очікуваний зиск буде 36 у. о. (табл. 15. 1), а на першому 10 у. о. (табл. 15. 2); загальний зиск з цих двох об'єктів визначається як (36 + 10) у. о., що і записано у відповідній клітині табл. 15. 3. У двох правих стовпчиках табл. 15. 3 записані: величина найбільшого зиску $Z_2^*(y_2)$ при можливому значенні y_2 та відповідна величина коштів, виділених для другого об'єкта x_2^*. Якщо сума коштів, виділених для двох об'єктів, дорівнює 60 у. о., то за такої умови найбільша величина зиску складе 43 у. о. (12 + 31), і цю величину буде досягнуто, якщо для другого об'єкта буде виділено 20 у. о., а для першого 60 – 20 = 40 у. о.

Розрахунок значень $Z_3^*(y_3)$ для третього кроку пошуку умовних оптимальних розв'язків наведений у табл. 15.4.

Розрахунки виконано за формулою:

$$Z_3^*(y_3) = \underset{0 \le x \le y_3}{extr} \left\{ f_3(x) + Z_2^*(y_3 - x) \right\}.$$

Таблиця 15 4.

y_3 \ x	0	20	40	60	80	100	$Z_3^*(y_3)$	x_3^*
20	0+12	11+0					12	0
40	0+31	11+12	36+0				36	40
60	0+43	11+31	36+12	45+0			48	40
80	0+62	11+43	36+31	45+12	60+0		67	40
100	0+76	11+62	36+43	45+31	60+12	77+0	79	40

При побудові табл. 15.4 перші доданки вибрані згідно з табл. 15.1, другі – з табл. 15.3. Аналогічно виконуються розрахунки пошуку $Z_4^*(y_4)$, які не наводимо, але радимо читачеві побудувати таблицю та порівняти результати з наведеними у двох останніх стовпчиках табл. 15.5, яку записано з використанням даних табл. 15.2 – 15.4.

Таблиця 15.5.

C	$x_1^*(C)$	$Z_1^*(C)$	$x_2^*(C)$	$Z_2^*(C)$	$x_3^*(C)$	$Z_3^*(C)$	$x_4^*(C)$	$Z_4^*(C)$
20	20	10	20	12	0	12	20	16
40	40	31	0	31	40	36	40	37
60	60	42	20	43	40	48	20	52
80	80	62	0	62	40	67	40	73
100	100	76	100	78	40	79	40	85

Проаналізуємо одержані результати. Нижній рядок двох останіх стовпчиків табл. 15.5 свідчить, що найбільший зиск – 85 у. о. від чотирьох об'єктів одержуємо за умови, що заплановану суму коштів 100 у. о. розподілено так: для четвертого об'єкта виділено 40 у. о. ($x_4^*(100) = 40$), а для перших трьох 100 – 40 = 60 у. о. Оптимальний розподіл цих коштів ($y_3 = 60$) між трьома об'єктами гарантує зиск в обсязі 48 у. о. ($Z_3^*(60) = 48$) за умови, що для третього об'єкта буде виділено 40 у. о. ($x_3^*(60) = 40$), а для перших двох 60 – 40 = 20 у. о. Ці кошти при оптимальному розподілі й дають зиск в обсязі 12 у. о. ($Z_2^*(20) = 12$) за умови, що для другого об'єкта буде виділено 20 у. о. ($x_2^*(20) = 20$).

Схема розподілу коштів у сумі 60 у. о. зображена на рис. 15. 5.

Розподіл коштів, наданих підприємствам:

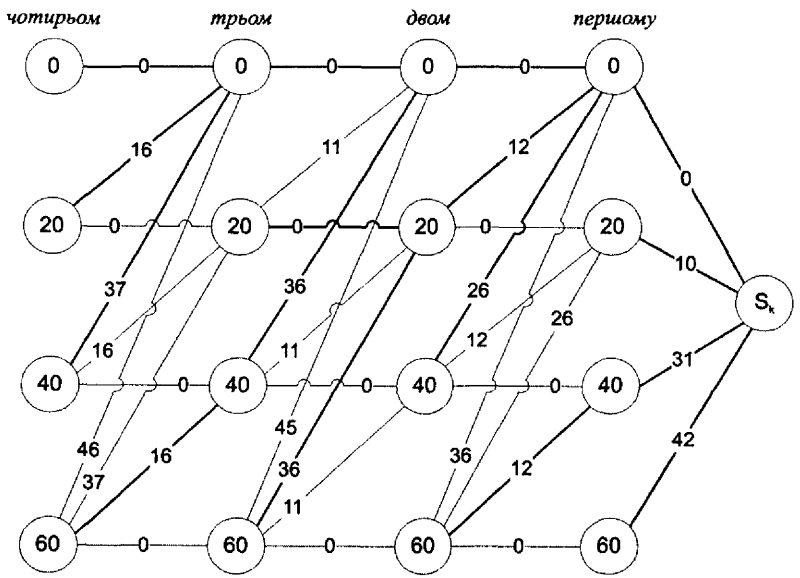


Рис. 15.5 .

Висновок. За економічними даними ефективності реконструкції та модернізації чотирьох об'єктів, представлених у табл. 15.1, максимальний очікуваний зиск буде 85 у. о. за умови розподілу між об'єктами коштів в обсязі 100 у. о. за такою схемою: для першого об'єкта кошти не виділяти; для другого об'єкта виділити 20 у. о.; для третього та четвертого – по 40 у. о. Це основний результат розв'язку сформульованої задачі. У табл. 15.5 наведені також розв'язки поставленої задачі з оптимальним розподілом коштів для різних обсягів від 20 до 100 у. о. між двома, трьома та чотирма об'єктами за умови ефективності асигнування відповідно до даних табл. 15. 1.

Наприклад, кошти 100 у. о. необхідно оптимально розподілити між трьома об'єктами. Скориставшися даними табл. 15. 5, знаходимо, що максимальний зиск $Z_3^*(100)$ становить 79 у. о. Такий зиск можна одержати, якщо виділити для третього об'єкта 40 у. о., бо $x_3^*(100) = 40$, для останніх двох залишається 100 - 40 = 60 у. о. Такі кошти за умови оптимального розподілу між двома об'єктами забезпечать зиск в обсязі $Z_2^*(60) = 43$ у. о., але цього можна досягти, якщо $x_2^*(60) = 20$ у. о., тобто виділити для другого об'єкта 20 у. о., а решту коштів 60 – 20 = 40 у. о. виділити для першого об'єкта.

Читачеві пропонується переконатися, що за умови оптимального розподілу коштів в обсязі 80 у. о. між трьома об'єктами відповідні асигнування складуть $x_1^* = 40$; $x_2^* = 0$; $x_3^* = 40$; $Z_3^*(80) = 67$ у. о.

Питання для самоконтролю

1. Навести приклади однокрокової та багатокрокових економічних операцій.
2. Навести загальну постановку задачі ДП.
3. Зміст принципу оптимальності Р. Белмана.
4. Сформулювати задачу пошуку оптимального за певним критерієм шляху між пунктами мережі.
5. Указати основні етапи розв'язання задачі ДП.
6. Сформулювати задачу оптимального розподілу інвестицій між підприємствами за критерієм максимального зиску.

ГЛАВА 16

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСІВ МАСОВОГО ОБСЛУГОВУВАННЯ

Не можна уявити життя сучасної людини без її контактів з різними системами обслуговування: бібліотеками, транспортними засобами, медициною, телефонним зв'язком, ринками, банками тощо. При моделюванні таких контактів необхідно враховувати наявність певних запитів і вимог щодо обслуговування та засобів, здатних у певній мірі виконати ці запити. Методи математичного моделювання роботи систем масового обслуговування (СМО) дають змістовні й ефективні рекомендації щодо проектування, реалізації та керування роботою цих систем у напрямі їх цільового призначення. Характерною особливістю СМО є *наявність двох взаємопов'язаних сторін: сторони, яка потребує обслуговування, та сторони призначеної виконати необхідне обслуговування.* Уникнути випадковостей у цих відносинах неможливо. Їх поява обумовлена природою взаємодії вказаних сторін: запити на обслуговування як у часі, так і в певному діапазоні вимог є випадковими; термін обслуговування теж має випадковий характер (наприклад, поява вимоги на певний вид технічного обслуговування або ремонту та термін реалізації такого запиту). Будемо дотримуватися такого користування термінами: запит на обслуговування, прийнятий у СМО (в чергу або обслуговування), назвемо вимогою.

Існує велике розмаїття математичних моделей СМО. Тому в цій главі маємо намір висвітлити основні засади найбільш типових і суттєвих елементів аналізу СМО та їх моделювання, а також розглянути кілька конкретних моделей, які є одними з основних. Розглянемо моделі СМО, в яких використано методи кількісної оцінки якості їх роботи. СМО, представлені такими моделями, є спрощенням реально існуючих систем, але вони дозволяють з певною мірою відповідності або адекватності охарактеризувати їх якість і наближені кількісні характеристики. У багатьох ММ математичний апарат процесів масового обслуговування є досить складним. Виходячи з практичної значимості розуміння основних засад теорії СМО, будемо користуватися результатами теоретичних обгрунтувань з відповідними поясненнями. Основну увагу буде приділено так званим *операційним характеристикам* (середня "довжина" черги, середній термін очікування обслуговування, ймовірність того, що всі компоненти системи в певний момент задіяні), а також іншим показникам кількісних оцінок ефективності функціонування.

Поки що не існує єдиного підходу до розв'язання задач оптимізації у СМО. У більшості випадків для розв'язання певної задачі використовується метод вузького цільового призначення.

16.1. Основні поняття та класифікація моделей СМО

Ознайомимося з основними поняттями та термінологією, які використовуються в теорії СМО. Будь-яка СМО, що являє собою спосіб реалізації взаємодії двох випадкових у часі процесів – появи запитів на обслуговування та реалізації цих запитів, характеризується: 1) вхідним потоком, тобто потоком надходження запитів на обслуговування; 2) дисципліною черги, тобто правилами узгодженості виконання запитів у часі ; 3) механізмом обслуговування (засобами реалізації запитів).

Під терміном **запит** розуміємо будь-яку вимогу на виконання певної роботи: ремонт апаратури, обслуговування покупців у роздрібній торгівлі, виконання певних робіт, обслуговування клієнтів у сфері послуг і т. ін. **Черга** – сукупність запитів, які необхідно задовольнити.

Кожна СМО має певну кількість засобів виконання запитів (технічні засоби; люди, які надають послуги, використовуючи відповідні технічні засоби), які називаються **каналами обслуговування.** За кількістю каналів обслуговування СМО поділяються на *одноканальні* та *багатоканальні.*

Вхідний потік вимог визначається послідовністю запитів у часі й характеризується певним законом розподілу ймовірної кількості запитів на одиницю часу. Запити на обслуговування можуть бути як одиничними, так і груповими. Якщо запити надходять через рівні проміжки часу, то такий потік називається *регулярним.* СМО з регулярними потоками зустрічаються досить рідко й не викликають особливих труднощів у їх дослідженні. Будемо розглядати СМО з *нерегулярними вхідними потоками* за умови, що термін виконання (обслуговування) однієї вимоги є випадковою величиною. За таких ситуацій можуть з'являтися черги на виконання запитів.

Дисципліна черги характеризує порядок виконання запитів, тобто визначається правилами формування черги. Досить часто дисципліна черги реалізується за схемою: *першим надійшов – першим обслуговується.* Така схема є досить простою з точки зору математичного моделювання, але стосується лише таких систем, в яких запити на обслуговування можуть бути упорядковані в певну послідовність у часі.

За деяких умов порядок обслуговування є випадковим, наприклад, опитування студентів на заняттях. Досить часто дисципліна обслуговування будується за певною системою пріоритетів або за умови обмеженого терміну чекання: запит вилучається з черги, якщо він не виконується за певний проміжок.

Механізм обслуговування, як і вхідний потік, характеризується терміном обслуговування та кількістю виконаних вимог в одиницю часу. Термін обслуговування залежить як від якості запитів, так і від організації процедури обслуговування. Як правило, він характеризується використанням незалежних випадкових величин з певними законами їх розподілу, з урахуванням ймовірності виходу каналів обслуговування з ладу.

Предметом теорії систем масового обслуговування є розбудова їх математичних моделей і дослідження (при використанні цих ММ) взаємозв'язків між кількістю каналів обслуговування, термінами обслуговування, режимами роботи каналів, характеристиками вхідних потоків запитів і критеріями оцінок ефективності функціонування системи. Доцільно наголосити, що критеріїв оцінок якості СМО завжди декілька, тому задачі оптимізації таких систем є багатокритеріальними, а їх розв'язання обумовлює системний підхід.

Як критерії оцінки якості СМО використовуються: середня кількість запитів, яку може виконати система за одиницю часу; середнє число запитів у черзі; середній термін перебування вимоги в черзі; середнє число зайнятих каналів обслуговування й інші оцінки.

Якщо менеджер не знає теорію СМО, може виникнути ілюзія ідеальної роботи керованої ним системи за умови узгодженості вхідного потоку з максимальною спроможністю обслуговування, якщо вони співпадають, та оцінки такої системи як оптимально ефективної. Економічно ж доцільно мати систему з *переважаючою спроможністю обслуговування.*

Математичні моделі СМО використовуються для розв'язання різнопланових задач цільового призначення систем. Наприклад, необхідно, дослідивши характеристики вхідного потоку та процесу обслуговування, визначити при якій кількості каналів обслуговування сумарні затрати, обумовлені перебуванням запитів у черзі та простоєм каналів обслуговування, будуть щонайменшими. За оцінку ефективності СМО іноді приймають мінімум втрат часу при перебуванні в черзі. За такої умови режим функціонування СМО визначається так, щоб кількість запитів або термін їх перебування в черзі не перевищували деякої заданої величини.

Оскільки потоки запитів у СМО, терміни їх перебування в черзі та обслуговування можна характеризувати лише законом розподілу випадкових величин, то для математичного моделювання процесів функціонування СМО використовується апарат теорії ймовірностей та математичної статистики. Рівень складності задач дослідження й оптимізації СМО залежить від структурних властивостей системи та від діапазону альтернатив, які необхідно проаналізувати.

16.2. Коротка історична довідка

Поява теорії СМО на початку XX ст. була обумовлена вимогами практики. Засновником теорії є датський математик А. К. Ерланг (1878-1929 рр.), який успішно використав методи теорії ймовірностей у розв'язанні задач проектування й обґрунтування експлуатаційних можливостей автоматичних телефонних станцій (АТС). Термін "теорія масового обслуговування" був запропонований і розвинутий радянським математиком О. Я. Хінчиним (1894-1959 рр.).

Ідеї та методи теорії СМО широко розбудовувалися в роботах А. Н. Колмогорова, Б. В. Гнеденка, М. П. Бусленка, Б. А. Севастьянова, І. М. Коваленка та інших учених. Особливо вагомий внесок у розвиток теорії СМО був зроблений дослідженнями А. А. Маркова. Широко відомими є також роботи зарубіжних учених К. Пальма, Т. Сааті, Л. Такача, Д. Кендалла, Ф. Поллачека.

16.3. Класифікація систем масового обслуговування

Класифікацію СМО розглянемо за критерієм становлення системи до одержаних нею запитань, при умови відповідності вільних каналів обслуговування. За названою ознакою у СМО виділяють *системи з відмовою (втратою)* та *системи з очікуванням (чергою).* У системі з відмовою запит вилучають із СМО, якщо його було одержано в момент, коли всі канали обслуговування зайняті. Прикладом такої системи є АТС (сучасні, а не майбутні), в яких каналами обслуговування є канали зв'язку. Якщо в момент запиту на підключення певного абонента відповідний номер зайнятий, запит залишає систему. У нових АТС запит можна буде ставити в чергу на обслуговування. Прикладом системи з відмовою може бути також система торгівлі за умови відсутності в ній товару, який запитує покупець.

Системи з очікуванням (чергою) читачеві добре знайомі: каси будь-якого цільового призначення, бензоколонки і т. ін.

За характером організації потоку запитів СМО поділяються на *замкнуті* та *незамкнуті.* У незамкнутих системах характеристики вхідного потоку запитів не залежать від кількості зайнятих каналів обслуговування, а в замкнутих СМО навпаки. Прикладами замкнутих СМО є системи організації ремонтного обслуговування певного обладнання, в яких інтенсивність вхідного потоку безпосередньо залежить від числа працюючих одиниць обладнання, але це число у свою чергу залежить від кількості обладнання, яке потребує ремонту і знаходиться у СМО. Для незамкнутих СМО характерною особливістю є те, що інтенсивність вхідного потоку не залежить від кількості запитів, які перебувають у системі й очікують обслуговування. Прикладами таких систем є бензоколонки, каси пасажирських перевезень.

За критерієм організації черги на обслуговування запитів у СМО з очікуванням виділяють два типи: *чисті* СМО та *змішані* СМО (системи з обмеженнями). Прикладами чистих СМО з очікуванням є магазини системи торгівлі. Системи переробки сільськогосподарської продукції належать до змішаних СМО з характерними для них обмеженнями на терміни обслуговування (переробки продукції).

Іноді необхідно враховувати, що в деяких СМО процес обслуговування є багатофазним, наприклад, у системах постачання обладнання та деталей немасового користування.

Однією з ознак класифікації СМО є *дисципліна обслуговування* – правила від-

бору запиту при наявності черги та незайнятого каналу обслуговування. За способом відбору запитів для обслуговування виділяють такі види дисципліни черги:

а) перший у черзі обслуговується першим (таке правило позначається англійською абревіатурою FIFO – First Input First Output); це часто використовується в пакетах прикладних програм ПЕОМ;

б) останній у черзі обслуговується першим (правило LIFO – Last Input First Output);

в) обмежений термін перебування запиту в черзі;

г) обслуговування за пріоритетом.

СМО типу LIFO часто є економічно більш зручними та вигідними. Прикладами таких систем можуть бути навантажувально-розвантажувальні операції при контейнерному транспортуванні дискретних одиниць вантажу.

Якщо у СМО надходять запити різної категорії значимості, то до їх обслуговування має бути різний підхід стосовно послідовності виконання запитів: більш значимі запити доцільно обслуговувати раніше менш значимих. СМО з такою дисципліною черги називаються *системами з пріоритетом.* Правила визначення пріоритетів можуть бути різними. Пріоритет називається *абсолютним*, якщо після надходження відповідного запиту СМО розпочинає його виконання, навіть призупинивши виконання запитів менш значимих, які перебували в каналі обслуговування. Пріоритет називається *відносним*, якщо в каналі обслуговування виконується запит меншого пріоритету й відповідна операція завершується, а запит більш високого пріоритету має право лише на переваги в черзі запитів на обслуговування. Прикладами СМО з відносним пріоритетом є: а) поліклініка, де хворі з гострим зубним болем обслуговуються поза чергою; б) каси продажу квитків, в яких поза чергою обслуговуються пільговики.

Наведена класифікація СМО має умовний характер. Існуючі системи не є однорідними і поєднують у собі різні властивості.

16.4. Основні засади математичного моделювання роботи систем масового обслуговування

Для оцінки дієспроможності СМО використовують такі характеристики, як кількість запитів в одиницю часу, пропускна спроможність системи, термін обслуговування запитів, термін очікування обслуговування і т. ін. Ці характеристики СМО є результатом прояву певних випадкових процесів: розподілу кількості запитів у часі, випадковості термінів обслуговування, наявності в черзі вимог, які очікують обслуговування і т. ін. Тому робота СМО може бути досліджувана лише як випадковий стохастичний процес, перебіг якого визначається перебігом інших процесів: потоку вхідних запитів, організацією черги, термінами обслуговування вимог.

Під *випадковим (стохастичним) процесом* розуміють процес перебігу в часі стану певної системи у відповідності до ймовірних закономірностей, тобто можливі стани такого процесу є випадковими, але не хаотичними, а певним чином унормованими.

Процес називається *процесом дискретних станів,* якщо його можливі стани $S_1, S_2, S_3, ..., S_n$ можна перелічити наперед, а перехід від одного стану до іншого відбувається за такі малі проміжки часу в порівнянні з терміном перебування в певному стані, що ними можна знехтувати, тому при моделюванні таких процесів переходи до різних станів вважаються миттєвими. Процес називається *процесом з неперервним часом*, якщо моменти можливих переходів від стану до стану не фіксовані, а випадкові. Доцільно наголосити, що вказані визначення характеризують різні властивості процесів, а саме можливість чітко визначати ймовірні стани та не передбачувані наперед моменти переходів від стану до стану. Процес роботи СМО являє собою стохастичний процес з дискретними ймовірними станами та неперервним часом. Це означає, що стан системи змінюється стрибком (миттєво) у випадкові моменти часу появи певних подій, наприклад, нового запиту або завершення операції обслуговування.

Досвід математичного моделювання СМО свідчить, що їх роботу можна ефективно описувати, використовуючи апарат теорії марківських процесів. Це поняття базується на уявленні про змінні в часі стохастичні стани системи, для яких характерною особливістю є відсутність післядії (амнезія). Перехід системи з одного ймовірного стану в інший можливий у будь-який момент часу t, а *ймовірність* $p_{ik}(t, \tau)$ переходу системи зі стану S_i, в якому вона знаходилася в момент t, у стан S_k в момент τ $(\tau \geq t)$ *не залежить від часу та способу досягнення системою стану* S_i, тобто оцінка ймовірності майбутнього стану системи визначається лише наявним станом, а не тим перебігом процесу, при якому система досягла цього стану.

Стан будь-якої СМО визначається переважно трьома випадковими потоками: вхідним потоком запитів, потоком їх обслуговування та потоком запитів, які залишають систему без обслуговування. У теорії СМО широко використовується поняття *елементарного потоку*. Однією з визначальних характеристик потоку випадкових подій є його *інтенсивність,* тобто середнє число подій, які можуть відбутися в одиницю часу. Для вхідного потоку запитів інтенсивність у подальшому будемо позначати літерою λ, а інтенсивність обслуговування запитів одним каналом при неперервній його роботі – μ. Потік випадкових подій доцільно класифікувати як елементарний, якщо він має такі властивості:

1) *стаціонарність,* якщо ймовірні характеристики потоку, зокрема його інтенсивність, не залежать від часу, тобто ймовірні характеристики такого потоку не змінюються в часі;

2) *ординарність*, тобто ймовірність появи більше однієї події за нескінченно малий проміжок часу dt є нескінченно малою величиною більш високого порядку, ніж dt, що означає неможливість одночасної появи двох і більше подій;

3) *відсутність післядії*, тобто ймовірність появи наступної події не залежить від кількості та часу появи попередніх подій.

У теорії ймовірностей доведено: для елементарного потоку подій за відомої інтенсивності λ ймовірність того, що на проміжку t досліджувана подія з'явиться k раз, підлягає розподілу за законом Пуассона (французький механік, фізик і математик 1781-1840):

$$P_k(t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \qquad (1)$$

Формула (1) характеризує закон розподілу ймовірностей дискретної випадкової величини: кількості k появ події на проміжку t. Використовуючи названий закон (1), можна повністю описати потік вимог (запитів або їх обслуговування), а саме визначити ймовірність будь-якого стану потоку: по-перше, що за проміжок t надійде k вимог (k=0; 1; 2; ...); по-друге, що за проміжок t в систему надійде не більше та не менше певних кількостей вимог. Названі ймовірності дають можливість оцінити ефективність функціонування СМО та визначити при необхідності шляхи її поліпшення.

На практиці часто замість закону розподілу числа вимог k на проміжку t (1) доцільно користуватися законами розподілу випадкової величини терміну τ між двома послідовними вимогами. Функція розподілу $F(t)$ випадкової величини терміну τ між двома послідовними вимогами за елементарного потоку визначається так:

$$F(t) = P(\tau < t) = 1 - e^{-\lambda t}. \qquad (2)$$

Функція щільності $f(t)$ розподілу названої випадкової величини визначається залежністю:

$$f(t) = \lambda e^{-\lambda t} \qquad (3)$$

(λ – величина інтенсивності потоку (середня кількість вимог за одиницю часу).

Згідно з визначенням функції щільності розподілу випадкової величини ймовірність події, що величина проміжку τ між двома послідовними вимогами прийме значення з інтервалу $(t; t+\Delta t)$, обчислюється за формулою:

$$P(\tau \in (t; t+\Delta)) = f(t)\Delta t. \qquad (4)$$

Математичне сподівання $M(\tau)$ випадкової величини τ за функції щільності розподілу (3), тобто середній термін, який характеризує розподіл у часі появи двох послідовних вимог, обчислюється за формулою:

$$M(\tau) = \int_0^\infty t f(t) dt = \int_0^\infty t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}. \qquad (5)$$

Закон розподілу випадкової величини, представлений функцією щільності розподілу (3) (рис. 16.1) або функцією розподілу (2), називається *показниковим*, або *експоненціальним*.

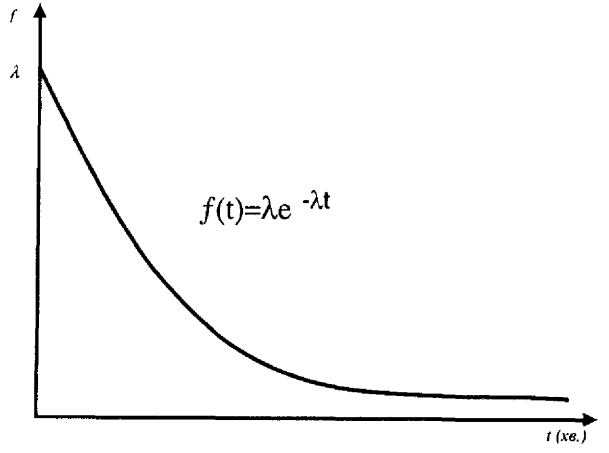


Рис. 16.1.

За показниковим законом математичне сподівання (5) та середнє квадратичне відхилення σ чисельно рівні між собою і дорівнюють величині $1/\lambda$, тобто обернені до величини інтенсивності потоку.

Показниковий закон (3) розподілу випадкової величини τ означає, що ймовірність надходження вимоги при будь-якому достатньо малому проміжку $(T; T+h)$ залежить лише від його величини h і не залежить ні від відліку часу T, ні від перебігу процесу надходження вимог у моменти, які передують T. Тобто показниковий розподіл моделює випадковий процес "позбавлений пам'яті" (відсутність післядії). Ймовірність відсутності вимоги на проміжку (0; T) рівновелика ймовірності події, що вимога надійде після моменту T:

$$P(\tau \geq T) = \int_T^{\infty} \lambda e^{-\lambda t} dt = e^{-\lambda T} . \qquad (6)$$

Умовна ймовірність відсутності вимоги на проміжку $(0; T+h)$ за умови, що не було жодного надходження вимог на проміжку $(0; T)$ при визначенні поняття умовної ймовірності, обчислюється так:

$$\frac{\int_{T+h}^{\infty} \lambda e^{-\lambda t} dt}{\int_T^{\infty} \lambda e^{-\lambda t} dt} = \frac{e^{-\lambda (T+h)}}{e^{-\lambda T}} \int_{T+h}^{\infty} = e^{-\lambda h} = P\ (\tau \geq h). \qquad (7)$$

Таким чином, дійсною є ймовірність події, що проміжок τ між двома послідовними вимогами, більший за деяку величину h, залежить від її значення і не залежить від її розташування у відліку часу.

Аналізу роботи будь-якої СМО має передувати дослідження вхідного потоку запитів.

Як з'ясувати, чи можна досліджуваний потік моделювати як елементарний, і як розбудувати розподіл за законом Пуассона за даними спостережень? Наведемо приклад розв'язання сформульованої задачі. У табл. 16.1 наведені результати спостережень за кількістю запитів на телефонні переговори на АТС протягом 500 хв. Заміри виконувалися на *проміжку кожної наступної хвилини*, а потім були упорядковані за наведеною в табл. 16.1 схемою. У першому стовпці вказано кількість k запитів, які надходили протягом однієї хвилини спостереження, у другому стовпці - кількість тих хвилинних проміжків, на яких надійшло відповідне число запитів. За 500 хв. спостереження надійшло 1000 запитів (сума добутків відповідних значень першого та другого стовпців).

Таким чином, середня кількість запитів протягом однієї хвилини дорівнює $\lambda = 1000/500 = 2$ зап./хв. Обчислимо незміщену оцінку дисперсії $D(\lambda(t))$ випадкової величини $\lambda(t)$ кількості запитів, які надійшли у СМО за одну хвилину:

$$D(\lambda(t)) = \frac{1}{n-1} \sum_{k=0}^{8} m_k (\lambda_k(t) - 2)^2 . \qquad (8)$$

Величина $D(\lambda(t))$ для наведеного прикладу дорівнює 1,91.

Таблиця 16.1.

Число запитів, які надійшли протягом однієї хвилини за довільного її розташування на проміжку спостережень, k	Кількість хвилинних проміжків, на яких було k запитів, m_k	Величини ймовірностей, обчислених за законом Пуассона (1), P_k	Величини теоретично розрахованих значень частот, $m_{kp} = nP_k$	$\frac{(m_k - nP_k)^2}{n \cdot P_k}$
0	70	0,135335	68	0,059
1	130	0,270671	135,3	0,185
2	140	0,270671	135,3	0,185
3	90	0,180447	90	0
4	40	0,090224	45	0,556
5	22	0,036089	18	0,222
6	6	0,012030	6	0
7	2	0,003437	2	0
8	0	0,000859	0,4	0,4
	500		**500**	$\chi^2_P = \sum_{k=0}^{\infty} \frac{(m_k - n \cdot P_k)^2}{n \cdot P_k} =$ **= 1,607**

З курсу математичної статистики відомо: якщо випадкова величина має цілі додатні значення та близькі за величиною математичне сподівання й дисперсію, то доцільно скористатися гіпотезою, що закон розподілу такої випадкової величини відповідає закону Пуассона (1). За даними прикладу маємо проміжок t, рівний одній хвилині:

$$P_k = \frac{2^k e^{-2}}{k!} . \qquad (9)$$

Величини P_k, обчислені за формулою (9), наведені у третьому стовпці табл. 16.1. Перевірку узгодженості прийнятої гіпотези про розподіл за законом Пуассона та експериментальних даних виконаємо за критерієм Пірсона, тобто обчислимо величину:

$$\chi^2_p = \sum_{k=0}^{8} \frac{(m_k - nP_k)^2}{nP_k}$$

(n – загальна кількість проміжків спостереження (500); nP_k – теоретичні значення частот, обчислені при використанні (1)).

Дані, використані при виконанні розрахунків, наведені в четвертому та п'ятому стовпцях табл. 16.1. По виконанні розрахунків маємо χ^2_p=1,607. Скористаємося таблицями розподілу χ^2 з числом степенів волі $\nu = l - r$ (l – кількість груп експериментального розподілу; r – кількість параметрів гіпотетичного теоретичного закону (для дослідження прикладу маємо один параметр λ), додана до числа додаткових залежностей, якими пов'язані експериментальні частоти (у прикладі лише одна залежність $\sum_{k=0}^{8} m_k = n = 500$)).

Для досліджуваного прикладу маємо: $l = 9$; $r = 1 + 1 = 2$; $\nu = 9 - 2 = 7$. За $\nu = 7$ найближче табличне значення χ^2 розподілу дорівнює 1,690, а йому відповідає величина ймовірності 0,975. Отже, з ймовірністю 0,975 можна вважати, що відхилення між експериментальними даними та вибраним теоретичним розподілом (9) є випадковими і ними можна знехтувати. Тобто наявний потік запитів є елементарним інтенсивності $\lambda = 2$.

16.5. Терміни виконання вимог (обслуговування)

Термін виконання вимоги визначається величиною часу, витраченого каналом обслуговування на її виконання, і, як правило, є випадковою величиною, що обумовлено багатьма чинниками. Наприклад: запити на постачання комплектуючих для ремонту мають різний перелік найменувань деталей; термін обслуговування касиром одного покупця в магазині визначається кількістю покупок.

За названих причин у теорії СМО досліджують закони розподілу термінів обслуговування. Досвід показує, що в багатьох випадках доцільно користуватися показниковим законом розподілу з функцією щільності

$$f(t) = \mu \cdot e^{-\mu t} \quad (t > 0) \tag{10}$$

($\mu = 1/\bar{t}_{обс}$; $\bar{t}_{обс}$ – середній термін обслуговування однієї вимоги; μ – середня кількість вимог для обслуговування одним каналом за одиницю часу).

Питання моделювання роботи каналу обслуговування при розв'язанні практичних задач аналізу СМО вирішується або шляхом обробки відповідних статистичних даних, або з використанням певних теоретичних міркувань. Як було наголошено раніше, однією з ознак доцільності використання показникового закону розподілу є наближена рівність значень величин математичного сподівання та середнього квадратичного відхилення. Вибір певного закону розподілу має бути перевірений з використанням відомих статистичних критеріїв узгодження.

Результати спостереження термінів обслуговування запитів на базі постачання ремонтних комплектуючих наведені в табл. 16.2 (відлік часу у хвилинах).

Таблиця 16.2.

Номер проміжку об'єднання даних, i	Проміжки об'єднання даних, хв	Кількість вимог на проміжку, m_i	Середина проміжку, хв., $\bar{t}_i$	$m_i\,\bar{t}_i$	$\mu\bar{t}_i$	$e^{-\mu\bar{t}_i}$	$P_i(t_i<\tau<t_{i+1})$	$m_{ip} = n \cdot P_i$	$\frac{(m_i - m_{ip})}{m_{ip}}$
1	0–10	102	5	510	0	1,0	0,4885	98,19	0,1478
2	10–20	45	15	675	–0,666	0,5115	0,2390	48,04	0,1923
3	20–30	26	25	650	–1,339	0,2725	0,1372	27,58	0,0905
4	30–40	15	35	525	–2,0	0,1353	0,0681	13,69	0,1253
5	40–50	7	45	315	–2,666	0,0672	0,0303	6,09	0,1360
6	50–60	3	55	165	–3,333	0,0369	0,0278	5,56	0,3729
7	60–70	4	65	260	–4,0	0,0183	–	–	–

n=202; $\sum_{i=1}^{7} m_i \bar{t}_i = 3100$; $\chi^2_p = \sum_{i=1}^{7} \frac{(m_i - m_{ip})^2}{m_{ip}} = 1{,}0648$.

Значення частот кількостей виконаних вимог за контрольовані проміжки часу співставлено їх серединам та зображено на рис. 16.2.

Порівнявши зміни наведених експериментальних даних і показникового закону розподілу (рис. 16.1), доходимо висновку про доцільність використання показникового закону розподілу для моделювання зміни експериментальних даних. Обчислимо параметр μ (середню кількість вимог, виконаних за одну хвилину), скориставшися тим, що його величина чисельно дорівнює величині, оберненій середньому терміну обслуговування вимог:

$$\bar{t}_{обс} = \sum_{i=1}^{l} m_i \frac{t_i + t_{i+1}}{2} \Big/ \sum_{i=1}^{l} m_i = \frac{3100}{202} \approx 15{,}35 \tag{11}$$

(m_i – число вимог на i-му проміжку об'єднання даних спостереження, такі проміжки засвідчують, що термін виконання вимоги був саме в цих межах; t_i та t_{i+1} – відпові-

дні межі i-го проміжку; l – число проміжків об'єднання даних). Для подальших розрахунків приймемо $\bar{t}_{обс} = 15$.

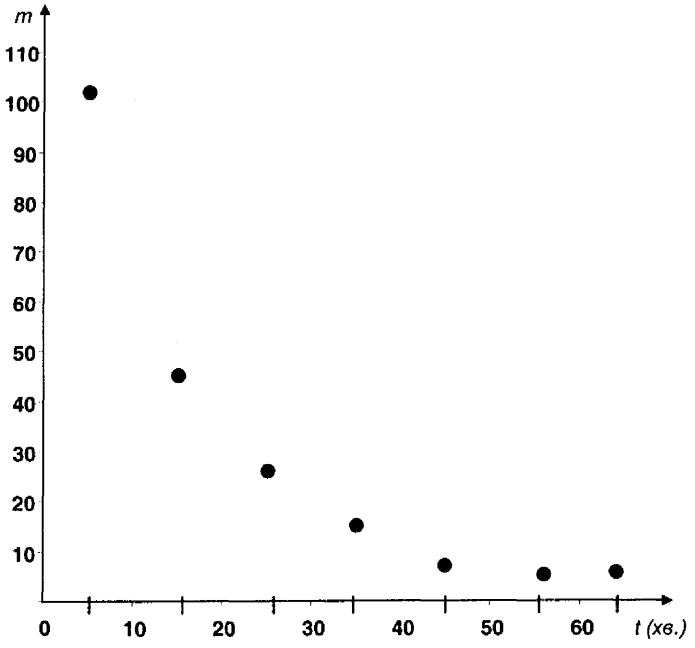


Рис. 16.2.

Таким чином, середня кількість вимог, виконаних за одну хвилину, буде дорівнювати:

$$\mu = \frac{1}{\bar{t}_{обс}} = \frac{1}{15} \text{ вимог / хв.}$$

Маємо підстави прийняти гіпотезу про дію показникового закону розподілу термінів виконання вимог (обслуговування):

$$f(t) = \mu e^{-\mu t} = \left| \mu = \frac{1}{15} \right| = \frac{e^{-t/15}}{15}. \qquad (12)$$

Оцінку доцільності використання вибраного закону розподілу виконаємо за критерієм узгодження Пірсона (χ^2 – розподіл).

Ймовірність події, що випадкова величина – термін виконання вимоги τ прийме значення з певного інтервалу $(a; b)$, обчислюємо, користуючись загальною формулою $P(a \le \tau \le b) = \int_a^b f(t)dt$. За показникового закону розподілу $P(a \le \tau \le b) = \int_a^b \mu e^{-\mu t} dt = e^{-\mu a} - e^{-\mu b}$. Необхідні результати числових розрахунків наведено в табл. 16. 2. Кількість груп поєднання даних спостереження рівна $l = 7$. Маємо лише один параметр μ теоретичного розподілу та одну залежність, що пов'язує дані спостереження: їх сума дорівнює 202. Отже, кількість степенів волі $\nu = l - r = 7 - (1+1) = 5$, величина $\chi^2 = 1{,}456$. При розрахунку χ^2, згідно з вимогою правил обробки статистичних даних (на проміжку мусить бути не менше п'яти спостережень), об'єднано два останніх проміжки. Скориставшися таблицями розподілу за Пірсоном, маємо, що з ймовірністю 0,9 (90%) можна прийняти показниковий закон розподілу $f(t) = e^{-t/15}/15$.

16.6. Графічне моделювання систем масового обслуговування

Розглянемо СМО, стан якої не впливає на вхідний потік запитів. Мета розбудови пропонованої моделі – дослідження змін у часі (динаміки) СМО. Тому мусимо показати:

1) множину можливих станів системи;
2) можливі зміни станів;
3) закономірності, які обумовлюють зміни станів.

При наявності у СМО певної кількості n каналів обслуговування стан системи будемо виділяти за кількістю вимог, які обслуговуються каналами, та числом m $(0 \le m \le \infty)$ запитів на обслуговування. За такої умови у СМО в будь-який момент може перебувати від 0 до $R=n+m$ вимог. На рис. 16.3 стан СМО зображено різними геометричними фігурами, щоб наочно виділити ситуацію, коли число запитів не перевищує числа каналів обслуговування (прямокутники) та коли в системі накопичується черга запитів (овали). Стан системи при відсутності запитів позначено колом, у середині якого записано нуль. Стан найбільшого завантаження СМО помічено прямокутником із записом "$n+m$".

Можливі напрямки зміни стану зображено стрілками, що поєднують фігури зображення станів, з показом напрямку зміни. Будь-яка зміна певного стану обумовлена зміною числа вимог у СМО на одиницю.

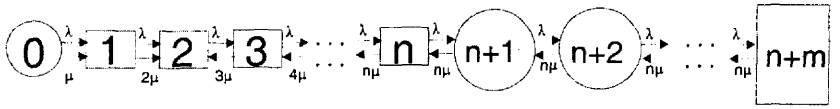


Рис. 16.3.

Кожна зміна числа вимог обумовлена однією з двох можливих подій: надходженням у систему ще одного запиту, за рахунок чого кількість вимог збільшиться у СМО на одиницю, або завершенням обслуговування вимоги і відповідним зменшенням їх кількості в системі на одиницю. Так моделюється ситуація, коли ймовірністю одномоментної появи двох і більше з названих подій можна нехтувати. За такої умови, якщо СМО знаходилася в стані за номером i (сума числа запитів і числа вимог, які обслуговуються, дорівнює i), вона може перейти лише в один із двох можливих станів: $(i-1)$ або $(i+1)$. Для вихідного та завершального станів можливий лише один із названих переходів.

Описана система моделі роботи СМО є орієнтованим графом, якщо за його вершини прийняти стани (події в теорії графів), а за дуги – переходи від стану до стану. Граф схеми, представленої на рис. 16.3, іноді називають ланцюгом "*розмноження – вимирання*". Назва обумовлена його використанням у біології для досліджень динаміки чисельності популяції, що буде розглянуто згодом.

Множина стрілок графу 16.3 відповідає множині можливих змін станів у модельованій СМО, тобто множині подій, які можуть у ній відбутися.

Для завершення розбудови моделі СМО необхідно додати числові характеристики можливих станів. Для цього необхідно врахувати, що ймовірність подій, які відповідають верхнім стрілкам, обумовлена впливом на систему вхідного потоку запитів і пов'язана з його інтенсивністю λ. Перехід системи в менш завантажений стан визначається інтенсивністю обслуговування (μ – середня кількість вимог, які можуть бути виконані в системі за одиницю часу одним каналом обслуговування) за умови використання i каналів. Якщо номер стану не менше числа каналів обслуговування, то інтенсивність обслуговування стабілізується на рівні $n\mu$. Таким чином, орієнтований граф, зображений на рис. 16.3, являє собою графічну модель роботи СМО. Конкретне зображення графічної моделі залежить як від числа каналів обслуговування, так і від домовленості стосовно довжини черги m. За названими двома ознаками виділяються деякі типи СМО, вказані в табл. 16.3.

Для розглянутих у табл. 16.3 СМО характерним є те, що інтенсивність вхідного потоку запитів як величина стабільна не залежить від стану самої системи. Системи такого типу вище було названо *незамкнутими*.

Побудову графу *замкнутої* СМО розглянемо на схемі аналізу роботи конкретної системи. Припустимо, що моделюємо організацію роботи ремонтної служби з n

каналами обслуговування на підприємстві, де використовується одне типове обладнання, яке в повному обсязі перебуває в експлуатації і резерв якого не передбачений. Таке припущення спрощує розбудову відповідної моделі.

Таблиця 16.3.

Порядковий номер	Параметри СМО		Класифікація СМО (назва)
	n	m	
1	1	0	Одноканальна без черги
2	$n>1$	0	Багатоканальна без черги
3	1	$1<m<\infty$	Одноканальна з обмеженою чергою
4	$n>1$	$1<m<\infty$	Багатоканальна з обмеженою чергою
5	1	$m=\infty$	Одноканальна з необмеженою чергою
6	$n>1$	$m=\infty$	Багатоканальна з необмеженою чергою

Позначимо через U загальне число наявного обладнання, частина якого, в середньому в кількості m за вибрану одиницю часу, вимагає негайного ремонту. Надходження кількості запитів на ремонт залежатиме від числа одиниць обладнання, яке вже перебуває в ремонті (позначимо його літерою i).

Інтенсивність λ_i потоку запитів на ремонт може бути обчислена за формулою:

$$\lambda_i = \frac{m}{U}(U-i) \quad (i=\overline{0,U-1})$$

($U-i$ – кількість обладнання в експлуатації; m/U – частка обладнання, яке вимагає ремонту за одиницю часу).

Граф моделі досліджуваної СМО представлений на рис. 16.4.

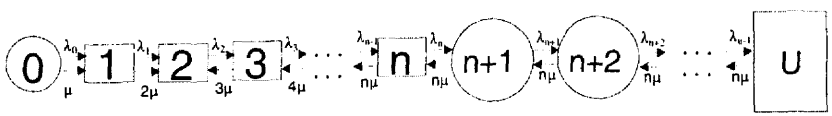


Рис. 16.4.

Якщо кількість обладнання, яке вимагає ремонту, не перевищує кількості каналів обслуговування (майстрів або їх бригад), то черга не може виникнути, і виконання запиту, якщо він з'являється, реалізується без затримки. Специфічною властивістю досліджуваної системи є те, що чим більше в ній вимог, тим менше інтенсивність вхідного потоку. Якщо у СМО буде U вимог (кількість обладнання, яке вимагає ремонту, додана до кількості обладнання, яке перебуває в ремонті), інтенсивність вхідного потоку дорівнюватиме нулю (на рис. 16.4 ця інтенсивність позначена λ_{U-1}).

За наявності резервного обладнання необхідно внести відповідні корективи в розбудову графу. Дотепер вважалося, що канали обслуговування ідентичні відносно їх спроможності в обслуговуванні. За умови різної інтенсивності обслуговування на кожному з наявних каналів необхідно враховувати можливі різні правила їх роботи.

Загальною властивість замкнутих СМО є скінченне число їх можливих станів.

Перш ніж описувати загальні принципи розбудови математичних моделей СМО, розглянемо моделювання найпростіших із них.

16.7. Чисті системи масового обслуговування за необмеженої черги

Досліджуємо СМО з n однотиповими каналами обслуговування за умови елементарного вхідного потоку інтенсивності λ запитів в одиницю часу. Термін обслуговування вимог *одним каналом* характеризується показниковим законом розподілу з інтенсивністю μ. Як наголошувалося раніше, випадковий стан СМО визначається кількістю k вимог у ній, тобто сумою числа запитів і кількістю вимог на обслуговування. Проаналізуємо можливі стани досліджуваної СМО:

1) у системі відсутні запити, тобто $k = 0$; усі її канали вільні;

2) кількість вимог k менше кількості каналів обслуговування n ($k < n$); в системі $n - k$ каналів є вільними;

3) кількість вимог k більша або дорівнює кількості числу каналів обслуговування n, тобто $k \geq n$, отже, всі канали обслуговування зайняті.

Робота *чистої* СМО з очікуванням моделюється так: якщо в момент надходження запиту хоч один канал обслуговування не зайнятий, то система миттєво починає обслуговувати вимогу; якщо запит надійшов у систему в момент, коли усі канали обслуговування зайняті, то запит розташовується в чергу, чекаючи обслуговування. Приймаємо, що СМО не втрачає запитів ні за яких умов, тобто термін очікування обслуговування не обмежений.

Прикладами таких систем, у яких доцільно використовувати охарактеризовану модель, можуть бути майстерні по ремонту сезонного сільськогосподарського знаряддя, магазини самообслуговування за умови обліку лише тих відвідувачів, які визначилися з покупками і чекають касових розрахунків.

16.7.1. Одноканальні СМО з очікуванням

Для детального ознайомлення з особливостями моделювання чистих СМО розглянемо одноканальну систему з необмеженою чергою. Позначимо $P_k(t)$ та

$P_k(t+\Delta t)$ – ймовірності подій, що в моменти часу t і $(t+\Delta t)$ (Δt – малий проміжок у порівнянні з одиницею часу) в системі знаходиться k вимог ($k = 0; 1; 2; ...$). Обчислимо величину $P_k(t)$ в залежності від величини інтенсивності λ вхідного потоку та середньої пропускної спроможності μ каналу обслуговування (інтенсивність обслуговування). Обчислимо ймовірність події, що за проміжок Δt у СМО з'явиться одна вимога. Оскільки за допущенням вхідний потік елементарний, то ймовірністю появи на проміжку Δt більше однієї вимоги можна знехтувати, тобто вважаємо, що маємо можливість появи *лише двох подій*: відсутності запиту або появи одного запиту. За такої умови запишемо рівняння

$$P_0(\Delta t) + P_1(\Delta t) = 1;$$

$$P_1(\Delta t) = 1 - P_0(\Delta t).$$

За допущенням вхідний потік елементарний, тому $P_0(\Delta t)$ можна обчислити, скориставшися формулою Пуассона (1):

$$P_0(\Delta t) = \frac{(\lambda \Delta t)^0}{0!} e^{-\lambda \Delta t} = e^{-\lambda \Delta t}. \tag{13}$$

Нагадаємо, що при використанні операції обчислення факторіалу маємо 0!=1. Якщо функцію e^x розгорнути в ряд Маклорена: $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + ... + \frac{x^n}{n!} + ...$, одержуємо:

$$P_0(\Delta t) = e^{-\lambda \Delta t} \approx 1 - \lambda \Delta t, \tag{14}$$

а ймовірність

$$P_1(\Delta t) \approx \lambda \Delta t. \tag{15}$$

Оскільки досліджуємо одноканальну СМО за показникового закону розподілу (10) термінів обслуговування, то ймовірність події, що за проміжок Δt буде виконано вимогу, яка знаходиться в каналі обслуговування, дорівнює $1 - e^{-\mu \Delta t} \approx \mu \Delta t$. А ймовірність події, що вимога не буде виконана, дорівнює $1 - \mu \Delta t$.

Тепер можемо виконати дослідження станів СМО, якщо на момент t у ній було k вимог; позначимо цей стан S_k, а його ймовірність – $P_k(t)$.

Розглянемо чотири можливості збереження стану S_k на момент $t + \Delta t$.

Позначимо A_1 подію, що при наявності в системі k вимог за проміжок Δt не

надійшло жодного запиту (ймовірність відсутності запиту за проміжок Δt дорівнює $1-\lambda\Delta t$), а також не було завершено обслуговування жодної вимоги (ймовірність події незавершення обслуговування вимоги дорівнює $1-\mu\Delta t$). Оскільки названі події незалежні між собою, то за теоремою обчислення ймовірності суміщення незалежних подій одержуємо:

$$P(A_1) = P_k(t)(1-\lambda\Delta t)(1-\mu\Delta t). \qquad (16)$$

Розглянемо подію A_2, сенс якої визначимо так: у момент t в системі маємо $(k-1)$ вимогу (ймовірність цього стану $P_{k-1}(t)$), за проміжок Δt надійшов один запит (імовірність події надходження одного запиту за проміжок Δt дорівнює $\lambda\Delta t$), але незавершено виконання жодної вимоги (ймовірність такої події $1-\mu\Delta t$)). За названих раніше умов імовірність події A_2 дорівнюватиме:

$$P(A_2) = P_{k-1}(t)\,\lambda\Delta t\,(1-\mu\Delta t). \qquad (17)$$

Сенс події A_3 визначаємо так: у момент t в системі маємо k вимог (імовірність цього стану $P_k(t)$), за проміжок Δt з'явився один запит (імовірність такої події $\lambda\Delta t$), було завершено обслуговування однієї вимоги (ймовірність такої події $\mu\Delta t$). Імовірність події A_3 як події суміщення трьох названих подій дорівнює:

$$P(A_3) = P_k(t)\,\lambda\Delta t\,\mu\Delta t. \qquad (18)$$

Розглянемо ще одну подію A_4. У момент t у СМО була $(k+1)$ вимога (ймовірність такого стану $P_{k+1}(t)$), за проміжок Δt додаткових запитів не надходило (ймовірність події – $(1-\lambda\Delta t)$), але в каналі обслуговування було завершено виконання однієї вимоги (ймовірність такої події $\mu\Delta t$). Імовірність події A_4 дорівнює:

$$P(A_4) = P_{k+1}(t)(1-\lambda\Delta t)\mu\Delta t. \qquad (19)$$

А оскільки події A_1, A_2, A_3, A_4 не сумісні та утворюють повну групу подій (для досліджуваної СМО вони і лише вони на момент $t+\Delta t$ забезпечують стан S_k), то можемо записати:

$$P_k(t+\Delta t) = P(\sum_{i=1}^{4} A_i) = \sum_{i=1}^{4} P(A_i). \qquad (20)$$

Скориставшись імовірністю подій A_i (16, 17, 18, 19), маємо:

$$P_k(t+\Delta t) = P_k(t)(1-\lambda\Delta t)(1-\mu\Delta t) + P_{k-1}(t)\lambda\Delta t(1-\mu\Delta t) + \\ + P_k(t)\lambda\Delta t\,\mu\Delta t + P_{k+1}(t)(1-\lambda\Delta t)\mu\Delta t.$$

Нехтуючи величинами малості другого порядку стосовно Δt, запишемо:

$$P_k(t+\Delta t) = P_k(t)\cdot(1-\lambda\Delta t-\mu\Delta t) + P_{k-1}(t)\cdot\lambda\Delta t + P_{k+1}(t)\cdot\mu\Delta t.$$

Перенісши в ліву частину рівності величину $P_k(t)$ та поділивши обидві її частини на Δt, одержуємо:

$$\frac{P_k(t+\Delta t)-P_k(t)}{\Delta t} = -P_k(t)\cdot(\lambda+\mu) + P_{k-1}(t)\cdot\lambda + P_{k+1}(t)\cdot\mu.$$

Виконавши перехід до межі за умови $\Delta t \to 0$, одержуємо диференціальне рівняння, яке пов'язує похідну $P_k'(t)$ імовірності наявного стану S_k системи з імовірностями усіх можливих станів: збереження в системі контрольованої кількості вимог, збільшення або зменшення кількості вимог на одиницю:

$$P_k'(t) = -P_k(t)(\lambda+\mu) + P_{k-1}(t)\cdot\lambda + P_{k+1}(t)\cdot\mu. \qquad (21)$$

У теорії СМО найбільше уваги приділяють дослідженню ситуацій стабільного стану їх функціонування, коли ймовірні характеристики процесу не змінюються в часі. Отже, розглянемо стан роботи, за якого $P_k'(t)=0$, тобто

$$P_{k+1}(t) = -\frac{\lambda}{\mu}P_{k-1}(t) + \left(\frac{\lambda+\mu}{\mu}\right)\cdot P_k(t). \qquad (22)$$

Залежність (22) дає можливість моделювати роботу СМО за різних можливих стабільних станів, тобто за будь-якого додатного значення k. Якщо $k=0$, маємо:

$$P_0(t+\Delta t) = P_0(t)\cdot(1-\lambda\Delta t) + P_1(t)\cdot(1-\lambda\Delta t)\cdot\mu\Delta t. \qquad (23)$$

Залежність (23) характеризує подію відсутності вимог у системі на момент $t+\Delta t$, яка може бути наслідком об'єднання лише таких двох подій: у системі не було жодної вимоги на момент t і жодного запиту за проміжок Δt не надійшло (ймовірність такої події $P_0(t)(1-\lambda\Delta t)$); в системі була одна вимога на момент t, за

проміжок Δt було завершено її виконання, але нового запиту не надійшло (ймовірність такої події $P_1(t)(1-\lambda\Delta t)\mu\Delta t$).

Нехтуючи величинами більш високого порядку малості, ніж Δt, маємо:

$$P_0(t+\Delta t) = P_0(t)(1-\lambda\Delta t) + P_1(t)\mu\Delta t.$$

Перенісши $P_0(t)$ вліво та виконавши граничний перехід після того, як обидві частини залежності поділено на Δt, одержали диференціальне рівняння.

При стабільному режимі роботи $P_0'(t) = 0$, тому

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t), \tag{24}$$

$$P_1(t) = \rho P_0(t) \tag{25}$$

($\rho = \lambda/\mu$ позначено *інтенсивність навантаження* системи). Скориставшися залежністю (22), одержуємо ймовірність будь-якого стану S_k досліджуваної системи. Оскільки розглядаємо стабільний процес, то запис аргументу t вилучаємо. Названі ймовірності стабільних станів СМО визначаються за формулою:

$$P_k(t) = \rho^k P_0. \tag{26}$$

Обчислимо величину P_0. Оскільки система обов'язково буде в одному зі станів S_k, а події, які відповідають знаходженню системи у певному стані, несумісні, то сума таких подій буде достовірною подією:

$$\sum_{k=0}^{\infty} P_k = 1.$$

Скориставшися формулою (26), одержуємо:

$$P_0\left(1+\rho+\rho^2+\rho^3+\ldots+\rho^k+\ldots\right) = 1.$$

Вираз у дужках є геометричною прогресією зі знаменником ρ, сума якої дорівнює $1/(1-\rho)$. Величина P_0 дорівнюватиме:

$$P_0 = 1-\rho. \tag{27}$$

Імовірності стабільних станів досліджуваної СМО за наявності в них k вимог ($k > 0$) обчислюються за формулою:

$$P_k = \rho^k(1-\rho) = \left(\frac{\lambda}{\mu}\right)^k\left(1-\frac{\lambda}{\mu}\right). \quad (k>0). \tag{28}$$

Формули (27) та (28) обчислення ймовірностей можливих станів одноканальних СМО за необмеженої черги дозволяють обчислювати основні характеристики експлуатаційної спроможності таких систем:

1) середню кількість вимог у такій системі в будь-який момент;
2) середнє число запитів у черзі ($\bar{V}$);
3) імовірність утворення черги (P_u);
4) імовірність очікування обслуговування запиту не менше t_0 одиниць часу;
5) середній термін очікування обслуговування в черзі ($\bar{t}_{оч}$);
6) середній термін $\bar{t}_{смо}$ перебування вимоги в СМО;
7) коефіцієнт простою каналу обслуговування ($k_{пр}$);
8) середні втрати часу на очікування обслуговування усіх запитів за певний проміжок ($\bar{t}_{вт}$).

Наголосимо, що усі названі величини є середніми характеристиками випадкових величин.

1. Середнє число вимог у СМО. Маючи ймовірності знаходження системи в певних станах і враховуючи, що стан системи визначається кількістю вимог, одержуємо:

$$\begin{aligned}\bar{k} &= \sum_{k=0}^{\infty} kP_k = \sum_{k=0}^{\infty} k\rho^k(1-\rho) = \rho(1-\rho)\left(1+2\rho+3\rho^2+\ldots\right) = \\ &= \rho(1-\rho)\left(\rho+\rho^2+\rho^3+\ldots\right)'_\rho = \rho(1-\rho)\left(\frac{\rho}{1-\rho}\right)'_\rho = \frac{\rho}{1-\rho}.\end{aligned} \tag{29}$$

За алгебраїчних перетворень у формулі (29) було використано припущення, що $\rho = \frac{\lambda}{\mu} < 1$. Лише за такої умови ряд $\sum_{n=1}^{\infty}\rho^n$ буде збіжний і його можна диференціювати. Таким чином, середня кількість вимог дорівнює:

$$\bar{k} = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}. \tag{30}$$

2. Середнє число запитів у черзі ($\bar{V}$) обчислимо аналогічно. Зауважимо, що за $k = 0$ число вимог у черзі дорівнює нулю. Якщо $k > 1$, то запитів у черзі буде $k-1$, бо один запит обслуговується. Отже,

$$\overline{V} = \sum_{k=1}^{\infty}(k-1)P_k = \sum_{k=1}^{\infty}kP_k - \sum_{k=1}^{\infty}P_k = \overline{k} - (1-P_0) = \overline{k} - \rho .$$

Скориставшися формулою (30), маємо:

$$\overline{V} = \frac{\lambda^2}{\mu(\mu-\lambda)}. \qquad (31)$$

3. Імовірність утворення черги ($P_ч$) – це ймовірність того, що $k > 1$, тобто $P_ч = P_{k>1} = 1 - P_0 - P_1$. Скориставшися формулами (25) та (27), одержуємо:

$$P_ч = 1-(1-\rho)-\rho(1-\rho) = \rho^2 = \left(\frac{\lambda}{\mu}\right)^2. \qquad (32)$$

4. Імовірність очікування обслуговування не менше t_0 одиниць часу $P(t_{оч} > t_o)$ обчислюється за формулою, яку наводимо без доведення:

$$P(t_{оч} > t_o) = \rho\, e^{-t_0(\mu-\lambda)}. \qquad (33)$$

5. Середній термін очікування в черзі ($\bar{t}_{оч}$) обчислюється за таких міркувань. Якщо маємо ймовірність події, що термін очікування в черзі перевищить задану величину t (33), то ймовірність протилежної події, що термін очікування в черзі буде менше t, обчислюється за формулою:

$$P(t_{оч} < t) = 1 - P(t_{оч} > t) = 1 - \rho^{-t(\mu-\lambda)} = F_{оч}(t)$$

($F_{оч}(t)$– функція розподілу випадкової величини – терміну очікування обслуговування).

Функція щільності розподілу $f_{оч}(t)$ названої випадкової величини має вигляд:

$$f_{оч}(t) = F'_{оч}(t) = \rho(\mu-\lambda)e^{-t(\mu-\lambda)}.$$

Математичне сподівання (середній термін очікування в черзі) $\bar{t}_{оч}$ дорівнює:

$$\bar{t}_{оч} = \int_0^{\infty} t f_{оч}(t)dt = \rho(\mu-\lambda)\int_0^{\infty} t e^{-t(\mu-\lambda)}dt = \frac{\rho}{\mu-\lambda}. \qquad (34)$$

Це співвідношення можна записати й іншим чином, якщо скористатися формулою (31):

$$\bar{t}_{оч} = \frac{\overline{V}}{\lambda}. \qquad (35)$$

6. Середній термін перебування вимоги в СМО ($\bar{t}_{смо}$) дорівнює сумі величин середнього терміну очікування в черзі $\bar{t}_{оч}$ та середнього терміну обслуговування $\bar{t}_{обс} = \frac{1}{\mu}$:

$$\bar{t}_{смо} = \bar{t}_{оч} + \frac{1}{\mu} = \frac{1}{\mu-\lambda}. \qquad (36)$$

7. Коефіцієнт простою каналу обслуговування ($k_{пр}$) визначається як імовірність події, що канал обслуговування не зайнятий, тобто в системі немає вимоги. За означенням:

$$k_{пр} = P_0 = 1 - \rho. \qquad (37)$$

8. Середні втрати часу очікування обслуговування ($\bar{t}_{вт}$) усіх запитів за проміжок T роботи СМО обчислюються як добуток середнього числа запитів, які надходять у систему за одиницю часу (λ), на середній термін перебування вимог у черзі $(\bar{t}_{оч})$ та на величину контрольного часу T роботи системи:

$$\bar{t}_{вт} = \lambda\bar{t}_{оч}T = \overline{V}T. \qquad (38)$$

Приклад. Досліджується робота бази постачання отрутохімікатів. Аналіз роботи показав, що вхідний потік запитів доцільно моделювати як елементарний з інтенсивністю $\lambda = 18$ за шестигодинний робочий день, обслуговування можна характеризувати показниковим законом розподілу з інтенсивністю п'яти вимог за годину. Визначити основні характеристики СМО.

За одиницю часу обираємо годину, тоді інтенсивність вхідного потоку дорівнюватиме $\lambda = 3$ зап./год. Інтенсивність навантаження $\rho = \lambda/\mu = 3/5 = 0{,}6 < 1$. Величина $\rho < 1$ свідчить, що під час роботи будуть проміжки відсутності запитів, а черга не буде накопичуватися.

Середнє число вимог у системі протягом години $\bar{k} = \frac{\rho}{1-\rho} = \frac{0,6}{1-0,6} = 1,5$.

Середнє число запитів у черзі $\bar{V} = \bar{k} - \rho = 1,5 - 0,6 = 0,9$.

Імовірність утворення черги $P_ч = \rho^2 = 0,36$.

Середній термін очікування обслуговування в черзі для всіх запитів (без урахування терміну обслуговування) $\bar{t}_{оч} = \frac{\bar{V}}{\lambda} = \frac{0,9}{3} = 0,3$ години.

Оцінюючи цей результат, треба враховувати, що ймовірність утворення черги порівняно невелика $(P_ч = 0,36)$.

Середній термін перебування вимоги в СМО (з урахуванням термінів очікування й обслуговування) $\bar{t}_{СМО} = \bar{t}_{оч} + \frac{1}{\mu} = 0,5$ години.

Коефіцієнт простою СМО $k_{пр} = 1 - \rho = 1 - 0,6 = 0,4$.

Одержаний результат свідчить, що в середньому 40 % робочого часу система буде простоювати.

Середні витрати часу на очікування обслуговування усіх запитів за шестигодинний робочий день $\bar{t}_{вт} = T \cdot \bar{V} = 6 \cdot 0,9 = 5,4$ год.

Для того, щоб краще зрозуміти вплив параметрів вхідного потоку λ та інтенсивності обслуговування μ на досліджувані характеристики СМО, радимо виконати відповідні розрахунки за $\lambda = 27$ та $\lambda = 36$ запитів за робочий день, залишивши незмінною середню інтенсивність обслуговування $\mu = 5$ вимог/годину.

16.7.2. Багатоканальні СМО з очікуванням

Вилучаючи обґрунтування наведених результатів аналізу СМО з n однотипними каналами, нагадаємо, що середня величина терміну обслуговування в кожному з каналів дорівнює $1/\mu$ (μ – середнє число вимог, які може обслуговувати один канал за одиницю часу). Приймаємо, що вхідний потік є елементарним інтенсивності λ. Наведемо результати для стабілізованого процесу.

1. Імовірність події, що в системі знаходиться k вимог. Імовірність відсутності в системі вимог обчислюється за формулою:

$$P_0 = \left[\frac{\rho^n}{(n-1)!(n-\rho)} + \sum_{m=0}^{n-1} \frac{\rho^m}{m!}\right]^{-1}, \quad \rho = \lambda/\mu. \tag{39}$$

Імовірність наявності в системі k вимог за умови, що $1 \le k \le n$, обчислюється так:

$$P_k = \frac{\rho^k}{k!} P_0$$

або так:

$$P_k = \frac{\rho}{k} P_{k-1}.$$

Імовірність наявності в системі числа вимог k, більшого за кількість n каналів обслуговування $(k > n)$, визначається так:

$$P_k = \left(\frac{\rho}{n}\right)^{k-n} P_n. \tag{41}$$

2. Середнє число запитів у черзі обчислюється за формулою:

$$\bar{V} = \sum_{k=n}^{\infty} (k-n) P_k = \frac{\rho \cdot n \cdot P_n}{(n-\rho)^2}. \tag{42}$$

(40)

3. Середнє число зайнятих каналів обслуговування:

$$\overline{n_з} = \rho. \tag{43}$$

4. Середнє число вимог у системі:

$$\bar{k} = \bar{V} + \overline{n_з} = \bar{V} + \rho. \tag{44}$$

5. Середнє число незайнятих каналів:

$$\overline{n_в} = n - \rho. \quad (45)$$

6. Коефіцієнт простою каналів:

$$k_{пр} = \frac{\overline{n_в}}{n} = 1 - \frac{\rho}{n}. \quad (46)$$

7. Імовірність наявності черги:

$$P_ч = \sum_{k=n+1}^{\infty} P_k = \frac{(n-\rho)}{n}\overline{V} = \frac{\rho}{n \cdot \rho} P_n. \quad (47)$$

8. Імовірність очікування обслуговування:

$$P_{оч} = P_ч + P_n. \quad (48)$$

9. Імовірність очікування не менше t_0 одиниць часу:

$$P(t_{оч} > t_0) = P_{оч} \cdot e^{-\mu t_0 (n-\rho)}. \quad (49)$$

10. Середній час надходження запитів:

$$\overline{t_\lambda} = \frac{1}{\lambda}. \quad (50)$$

11. Середній термін обслуговування одного запиту:

$$\overline{t_\mu} = \frac{1}{\mu}. \quad (51)$$

12. Середній термін чекання в черзі:

$$\bar{t}_{оч} = \frac{\overline{V}}{\lambda}. \quad (52)$$

13. Середній термін перебування вимоги в СМО:

$$\bar{t}_{смо} = \bar{t}_{оч} + \bar{t}_\mu. \quad (53)$$

14. Середні загальні втрати часу:

$$\bar{t}_{вт} = \lambda T \bar{t}_{оч} = \overline{V} T. \quad (54)$$

Так можна обчислювати основні характеристики роботи СМО за наявності n каналів обслуговування та умови необмеженої черги*.

16.8. Рівняння Колмогорова

Ознайомивши читача з математичним моделюванням роботи деяких СМО, розглянемо розбудову універсальної математичної моделі дослідження СМО як марківського процесу за *дискретних станів та неперервного перебігу часу.* Будемо вважати, що всі можливі зміни станів системи відбуваються лише під впливом елементарних вхідних потоків з відомими їх інтенсивностями та роботи каналів обслуговування. При малому проміжку Δt *імовірність переходу* системи з деякого i-го стану до іншого j-го стану *пропорційна, у першому наближенні*, величині проміжку з коефіцієнтом λ_{ij}, який характеризує щільність імовірності зміни станів у залежності від вхідних потоків відомих інтенсивностей:

$$P_{ij}(\Delta t) = \lambda_{ij} \Delta t. \quad (55)$$

Вважаємо, що в довільний момент t система перебуває *в будь-якому i-му* стані з можливих дискретних станів скінченної їх множини в кількості N. Імовірність події, що в момент t СМО перебуватиме в i-му стані, позначимо $P_i(t)$. За досить малий проміжок Δt система під дією зовнішнього впливу, обумовленого вхідними потоками, та зміни внутрішнього стану (перебігу процесу обслуговування вимог) перейде до певного стану за номером s. Подія перебування системи в цьому стані може відбутися за будь-якого вихідного стану СМО. Події перебування СМО в певному стані є несумісними. Події, які обумовлюють зміни станів, також несумісні. Скориставшися теоремою обчислення повної ймовірності появи події, яка може відбутися як наслідок однієї несумісної події, маємо:

$$\begin{aligned} P_s(t+\Delta t) = \sum_{i=1}^{N} P_i(t) P_{is}(\Delta t) = P_1(t) P_{1s}(\Delta t) + P_2(t) P_{2s}(\Delta t) + \ldots + \\ + P_s(t) P_{ss}(\Delta t) + \ldots + P_N(t) P_{Ns}(\Delta t). \end{aligned} \quad (56)$$

* Крутовой Ж. А. Экономико-математические методы в торговле и общественном питании / ХИОП. - Ч. 1. – Харьков, 1989.

Враховуючи формулу (55), цю залежність запишемо так:

$$P_s(t+\Delta t)=P_1(t)\lambda_{1s}\Delta t+P_2(t)\lambda_{2s}\Delta t+\ldots+P_s(t)\lambda_{ss}\Delta t+\ldots+P_N(t)\lambda_{Ns}\Delta t.$$

Скористаємось умовою, що за будь-якого вихідного стану системи в момент t вона завжди в момент $t+\Delta t$ перебуватиме в одному з можливих для неї станів, тобто сума можливих станів є достовірною подією, отже:

$$\sum_{i=1}^{N}P_{si}(\Delta t)=1\Rightarrow \Delta t\sum_{i=1}^{N}\lambda_{si}=1.$$

Остання залежність дозволяє обчислити ймовірність події, що стан системи за проміжок Δt не зміниться:

$$P_{ss}(\Delta t)=1-\sum_{i=1(i\neq s)}^{N}P_{si}(\Delta t)=1-\Delta t\sum_{i=1(i\neq s)}^{N}\lambda_{si}.$$

Підставивши вираз для $P_{ss}(\Delta t)$ у формулу (56) та перенісши вліво $P_s(t)$, обчислюємо приріст імовірності збереженого стану s за проміжок Δt:

$$\Delta P_s=P_s(t+\Delta t)-P_s(t)=\left[\sum_{i=1(i\neq s)}^{N}P_i(t)\lambda_{is}-\sum_{i=1(i\neq s)}^{N}P_s(t)\lambda_{si}\right]\Delta t.$$

Розділивши в цій залежності обидві частини на Δt та виконавши перехід до межі за умови $\Delta t\to 0$, одержуємо диференціальне рівняння, в якому пов'язано похідну від часу функції ймовірності s-го стану з функціями ймовірностей $P_i(t)$ інших станів СМО та функціями щільностей імовірностей переходів від i-го довільного стану до s-го (λ_{is}) та від s-го стану до довільного i-го (λ_{si}). За умови зміни s від 1 до N (множина дискретних можливих станів СМО) одержуємо систему диференційних рівнянь, названу іменем радянського математика А. Н. Колмогорова:

$$\frac{dP_s(t)}{dt}=\sum_{i=1(i\neq s)}^{N}P_i(t)\lambda_{is}-\sum_{i=1(i\neq s)}^{N}P_s(t)\lambda_{si}\quad (s=\overline{1,N}). \tag{57}$$

Розв'язок задачі Коші для системи (57), тобто за умови, що при $t=0$ ймовірності усіх можливих станів відомі, дозволяє визначити функції ймовірностей станів СМО за будь-якого моменту t. За умови, що на СМО діють лише елементарні потоки, множина можливих змін станів зменшується і система рівнянь Колмогорова (57) спрощується:

$$\frac{dP_s(t)}{dt}=P_{s-1}(t)\lambda_{s-1,s}+P_{s+1}(t)\lambda_{s+1,s}-P_s(t)(\lambda_{s,s-1}+\lambda_{s,s+1})\quad (s=\overline{1;N}). \tag{58}$$

Математична модель СМО у вигляді (58) системи диференціальних рівнянь відповідає системі, граф якої зображено на рис. 16. 5, і буде використана при дослідженні умов стабільної роботи таких СМО.

Приклад. Маємо агрегат S із двох різних вузлів, кожний з яких у будь-який момент може вийти з ладу, після чого *миттєво* розпочнеться його ремонт з невідомим терміном завершення. Побудувати граф станів агрегату S і скласти систему відповідних диференціальних рівнянь.

Можливі стани агрегату: S_1 – обидва вузли справні; S_2 – перший вузол ремонтується, а другий вузол справний; S_3 – перший вузол справний, а другий ремонтується; S_4 – обидва вузли ремонтуються. Граф станів агрегату наведено на рис. 16.5.

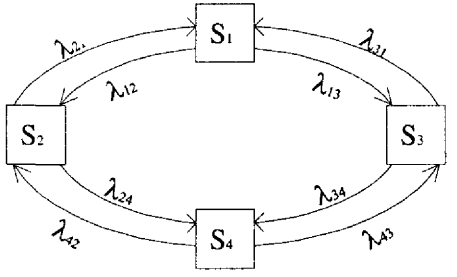


Рис. 16.5.

На графові відсутні переходи із S_1 у S_4 та із S_2 у S_3 і навпаки. Це результат природного припущення, що виходи кожного вузла з ладу або завершення їх ремонту – події незалежні, тому *ймовірностями їх одночасності* можна знехтувати. Перехід від стану до стану відбувається під впливом елементарних потоків подій: виходу вузлів з ладу або завершення їх ремонту. Біля стрілок записані інтенсивності відповідних переходів зі щільністю їх імовірностей.

Імовірністю і-го стану назвемо ймовірність $P_i(t)$ події, що в момент t агрегат (система) перебуває в стані S_i (i = 1; 2; 3; 4). Сума названих подій є достовірною подією, а події несумісні, тому для будь-якого моменту t дорівнює:

$$\sum_{i=1}^{4}P_i(t)=1.$$

Припустимо, що в момент t система була в стані S_1. Вибравши малий проміжок Δt, обчислимо ймовірність $P_1(t+\Delta t)$ події, що в момент $t+\Delta t$ система залишається в тому ж стані S_1. Названа подія може буди наслідком таких подій.

1. Система в момент *t була в стані* S_1, імовірність якого $P_1(t)$ і за проміжок Δt *не вийшла* з цього стану. Система може вийти з цього стану (див. граф на рис. 16. 5), якщо відбудеться одна з подій: зіпсується або перший вузол (імовірність такої події $\lambda_{12}\Delta t$), або другий вузол (імовірність цієї події $\lambda_{13}\Delta t$). Оскільки названі події несумісні, то ймовірність події, що відбудеться хоча б одна з них, дорівнює сумі ймовірностей названих подій ($\lambda_{12}\Delta t+\lambda_{13}\Delta t$). Тоді ймовірність події, що система не вийде зі стану S_1, дорівнює $1-(\lambda_{12}+\lambda_{13})\Delta t$. Імовірність події, що система, перебуваючи в момент t у стані S_1, залишається в цьому ж стані, дорівнюватиме добутку $P_1(t)(1-(\lambda_{12}+\lambda_{13})\Delta t)$.

2. Система може бути в момент $t+\Delta t$ у стані S_1, якщо вона в момент t була в стані або S_2, або S_3. Імовірності названих станів у момент t дорівнюють відповідно $P_2(t)$ та $P_3(t)$. За відомих значень функцій щільностей їх імовірних переходів відповідно λ_{21} та λ_{31}. Імовірність події, що система в момент $t+\Delta t$ буде в стані S_1, якщо в момент t вона перебувала в стані або S_2, або S_3, дорівнює $P_2(t)\lambda_{21}\Delta t+P_3(t)\lambda_{31}\Delta t$. Оскільки всі названі можливі переходи є несумісними подіями ($S_1\to S_1$; $S_2\to S_1$; $S_3\to S_1$), то ймовірність події, що система, перебуваючи в момент t у стані S_1, залишиться в цьому ж стані в момент $t+\Delta t$, дорівнюватиме:

$$P_1(t+\Delta t)=P_2(t)\lambda_{21}\Delta t+P_3(t)\lambda_{31}\Delta t+P_1(t)(1-(\lambda_{12}+\lambda_{13}))\Delta t).$$

Після раніше названих відповідних аналітичних перетворень одержуємо:

$$\frac{dP_1}{dt}=P_2(t)\lambda_{21}+P_3(t)\lambda_{31}-P_1(\lambda_{12}+\lambda_{13}).$$

За таких же міркувань стосовно визначення ймовірностей інших можливих станів системи одержуємо систему лінійних диференціальних рівнянь Колмогорова, які описують зміни ймовірності можливих станів системи в часі:

$$\begin{aligned}\frac{dP_1}{dt}&=\lambda_{21}P_2(t)+\lambda_{31}P_3(t)-(\lambda_{12}+\lambda_{13})P_1,\\ \frac{dP_2}{dt}&=\lambda_{12}P_1(t)+\lambda_{42}P_4(t)-(\lambda_{21}+\lambda_{24})P_2,\\ \frac{dP_3}{dt}&=\lambda_{13}P_1(t)+\lambda_{43}P_4(t)-(\lambda_{31}+\lambda_{34})P_3,\end{aligned}\qquad(59)$$

$$\frac{dP_4}{dt}=\lambda_{24}P_2(t)+\lambda_{34}P_3(t)-(\lambda_{42}+\lambda_{43})P_4.$$

За вихідні умови (початкові умови задачі Коші) для системи (59) треба вибрати будь-який початковий стан системи. Природно розпочати розв'язання (59) за умови, що в системі (агрегаті) за $t=0$ працювали обидва вузли, тобто $P_1(0)=1$; $P_2(0)=P_3(0)=P_4(0)=0$.

Розв'язок системи рівнянь Колмогорова дає можливість обчислити можливі ймовірності всіх станів як *функції часу*. Але щоб скористатися ними, необхідно мати величини λ_{ij} (в загальному випадку функції часу), які характеризують функції щільності ймовірностей зміни станів. Для кожної конкретної задачі λ_{ij} визначаються або шляхом теоретичних міркувань, або в результаті аналізу відповідних статистичних даних, або за результатами експертних оцінок.

16.9. Граничні ймовірності стабільної роботи СМО

Розв'язок задачі Коші (задачі за відомих вихідних умов) для системи диференціальних рівнянь (57) породжує систему функцій від часу $P_i(t)$ $\left(i=\overline{1,N}\right)$. Це дозволяє одержати дискретний розподіл імовірності перебування СМО в будь-якому стані з множин допустимих її станів за будь-якого значення $t>0$. Тобто за будь-якого допустимого вихідного стану (розподілу ймовірностей за $t=0$) розв'язок системи (57) дозволяє обчислити ймовірності відповідних станів у будь-який момент часу. Аналіз розв'язків таких систем диференціальних рівнянь свідчить, що для досить великих значень t незалежно від вихідного стану (розподілу ймовірностей) розподіли $P_i(t)$ стабілізуються і практично стають не залежними від часу. Названа асимптотична властивість розв'язку є віддзеркаленням факту, який спостерігається на практиці, що з плином часу функціонування СМО переходить у певний режим роботи, тобто розподіл імовірностей перебування СМО в кожному з допустимих станів не залежить від часу.

Наголосимо на розумінні термінів "*стаціонарний*" і "*стабільний*". У техніці стаціонарним називають процес, досліджувані характеристики якого не змінюються в часі, наприклад, кількість обертів парової турбіни ТЕЦ або температура кипіння води за атмосферного тиску. Для стабільних процесів *характерною є їх нечутливість до певних змін величин тих факторів*, які обумовлюють перебіг процесу. Для СМО дискретними можливими станами за умови *значного часу їх експлуатації* при врахуванні тієї обставини, що вхідний потік є випадковим процесом, можна оцінити лише ймовірності перебування системи в кожному з допустимих станів. Тому, на нашу думку, краще користуватися терміном "стабільний".

У теорії випадкових процесів доведено, що коли число можливих станів системи скінченне і з кожного стану можна за скінченну множину кроків перейти в будь-який інший стан, то існують так звані *граничні ймовірності* стабільної роботи СМО. Отже, маємо:

$$\lim_{t\to\infty} P_i(t) = P_i = const \quad \left(i = \overline{1, N}\right).$$

Гранична ймовірність P_i стану S_i має чіткий зміст: вона показує середній відносний термін перебування СМО в i-му стані за тотожних умов вхідного потоку. Наприклад, якщо гранична ймовірність стану S_2, тобто P_2, дорівнює 0,75, то це означає, що коли система буде працювати досить довго, то орієнтовно біля 75% часу вона перебуватиме у стані S_2, якщо вхідний потік буде мати незмінні характеристики.

Оскільки величини ймовірностей не залежать від часу, то відповідні похідні дорівнюють нулю. Прирівнявши ліві частини системи рівнянь (57) до нуля, для обчислення граничних імовірностей одержимо систему *лінійних однорідних алгебраїчних рівнянь:*

$$P_{s-1}\lambda_{s-1,s} + P_{s+1}\lambda_{s+1,s} - P_s(\lambda_{s,s-1} + \lambda_{s,s+1}) = 0. \tag{60}$$

Для одержання однозначного розв'язку системи (60) необхідно врахувати сенс величин P_i (імовірностей станів СМО):

$$\sum_{i=0}^{N} P_i = 1. \tag{61}$$

Для кращого розуміння сенсу граничних імовірностей запишемо кілька рівнянь системи (60) за різних значень s:

$$\begin{aligned} &s = 0 \Rightarrow P_1\lambda_{10} - P_0\lambda_{01} = 0 \text{ (решта величин дорівнює нулю)},\\ &s = 1 \Rightarrow P_0\lambda_{01} + P_2\lambda_{21} - P_1(\lambda_{10} + \lambda_{12}) = 0,\\ &s = 2 \Rightarrow P_1\lambda_{12} + P_3\lambda_{32} - P_2(\lambda_{21} + \lambda_{23}) = 0. \end{aligned} \tag{62}$$

Скориставшися першим рівнянням, запишемо залежність P_1 від P_0:

$$P_1 = P_0\frac{\lambda_{01}}{\lambda_{10}}. \tag{63}$$

Скориставшися другим рівнянням, маємо:

$$P_2\lambda_{21} = P_0\frac{\lambda_{01}}{\lambda_{10}}(\lambda_{10} + \lambda_{12}) - P_0\lambda_{01} \Rightarrow P_2 = P_0\frac{\lambda_{01}}{\lambda_{10}}\frac{\lambda_{12}}{\lambda_{21}}.$$

Враховуючи особливості системи рівнянь (57), для будь-якого стану СМО за номером s маємо залежність P_s від P_0:

$$P_s = P_0\frac{\lambda_{01}}{\lambda_{10}}\frac{\lambda_{12}}{\lambda_{21}}\cdots\frac{\lambda_{s-1,s}}{\lambda_{s,s-1}} \quad \left(s = \overline{0, N}\right). \tag{64}$$

Доцільно ввести нові позначення:

$$d_i = \frac{\lambda_{i-1,i}}{\lambda_{i,i-1}} \quad \left(i = \overline{1, N}\right) \tag{65}$$

стосовно відношення щільностей можливих переходів *від попереднього* (i–1)*-го стану до i-го* та *від i-го стану до попереднього* i–1. Аналізуючи невеликі за кількістю можливі стани СМО, величини d_i доцільно при розбудові її графу записувати поруч зі стрілками орієнтації зміни станів.

При використанні введеного позначення залежність (64) буде записана так:

$$P_s = P_0 d_1 d_2 \ldots d_s \Rightarrow P_s = P_0\prod_{i=1}^{s} d_i \tag{66}$$

(П – операція обчислення добутку названих співмножників).

Для обчислення величини P_0 використаємо залежність (66), оскільки сума ймовірностей усіх станів СМО має дорівнювати одиниці. Маємо рівняння:

$$P_0 + P_0\sum_{s=1}^{N}\prod_{i=1}^{s} d_i = 1, \tag{67}$$

з якого одержуємо формулу обчислення величини P_0:

$$P_0 = \left[1 + d_1 + d_1 d_2 + d_1 d_2 d_3 + \ldots + d_1 d_2 \cdots d_n\right]^{-1}. \tag{68}$$

За відомого значення P_0 обчислюємо ймовірності інших станів стабільної роботи СМО, скориставшися залежністю (64).

Аналіз одержаних залежностей показує, що доцільно користуватися рекурентною формулою:

$$P_i = P_{i-1} d_i \qquad \left(i = \overline{1, N}\right). \tag{69}$$

За відомих граничних імовірностей стабільної роботи СМО можна обчислити її різні характеристики як відповідні математичні сподівання (середні значення).

Розрахунки за формулою (66) зручно виконувати, якщо кількість запитів у черзі не перевищує 10. Тому має сенс одержати формули, дещо краще пристосовані для практичних розрахунків. Наведемо їх для незамкнутих систем, для замкнутих систем з ними можна ознайомитися в спеціальній літературі*.

Розглянемо n-канальну СМО незамкнутого типу з допустимою чергою в m запитів, граф якої наведено на рис. 16.3. Скориставшись формулою (65) та позначенням $\rho = \lambda/\mu$, при врахуванні сенсу величини d_i можемо записати:

$$d_i = \lambda/(i\mu) = \rho/i \ \left(i = \overline{1, n}\right); d_i = \rho/n \ \left(i = \overline{n+1, N}\right).$$

За наведених позначень формула має вигляд:

$$P_0 = 2\left[\sum_{i=0}^{n} \frac{\rho^i}{i!} + \frac{\rho^n}{n!} \sum_{i=1}^{m} \left(\frac{\rho}{n}\right)^i\right]^{-1}. \tag{70}$$

За умови $\rho = \lambda/\mu < 1$ обчислимо суму геометричної прогресії $\sum_{i=1}^{m} (\rho/n)^i$:

$$\sum_{i=1}^{m} (\rho/n)^i = \frac{\rho}{n} \cdot \frac{1-(\rho/n)^m}{1-\rho/n}.$$

Таким чином, формулу (70) можна записати у вигляді:

$$P_0 = 2\left[\sum_{i=0}^{n} \frac{\rho^i}{i!} + \frac{\rho^{n+1}}{n!n} \frac{1-(\rho/n)^m}{1-\rho/n}\right]^{-1}. \tag{71}$$

* Бешелев С. Д., Гурвич Ф. Г. Математико-статистические методы экспертных оценок. – М.: Статистика, 1974.

За умови необмеженої черги ($m \to \infty$) імовірність незавантаженого (вільного) стану СМО обчислюється за більш простою формулою:

$$P_0 = 2\left[\sum_{i=0}^{n} \frac{\rho^i}{i!} + \frac{\rho^{n+1}}{n!(n-\rho)}\right]^{-1}. \tag{72}$$

Названа умова означає, що сумарна інтенсивність обслуговування $n\mu$ на всіх каналах обслуговування водночас мусить бути більше за інтенсивність λ вхідного потоку запитів. Невиконання такої умови не лише унеможливлює доцільність наведених математичних перетворень, а й ефективне функціонування СМО: ніякого ефективного та стабільного процесу в такій системі бути не може.

Наголошуємо, що використання формули (71) можливе лише за умови, коли СМО може досягти стабільного режиму:

$$\frac{\lambda}{n\mu} = \frac{\rho}{n} < 1. \tag{73}$$

Названа умова означає, що сумарна інтенсивність обслуговування $n\mu$ на всіх каналах водночас має бути більше за інтенсивність λ вхідного потоку запитів. Невиконання такої умови робить не тільки недоцільними наведені математичні перетворення, а й неможливим ефективне функціонування СМО: ніякого ефективного та стабільного процесу в такій системі бути не може.

Використовуючи формулу (71), запишемо формули розрахунку ймовірностей інших станів СМО відповідно до наявності кількості запитів у ній:

$$P_i = P_0 \cdot \rho^i / i!, \ \left(i = \overline{1, n}\right); \quad P_n = P_0 \cdot \rho^n / n!; \tag{74}$$

$$P_{n+k} = P_n \cdot \left(\rho/n\right)^k, \quad \left(k = \overline{1, m}\right). \tag{75}$$

Залежність (75) дозволяє записати рекурентну формулу:

$$P_{n+k} = P_{n+k-1} \, \rho/n \qquad (k = \overline{1, m}).$$

Використання формул (72) - (75) дозволяє визначити характеристики економічної ефективності СМО.

Приклад. Знайти граничні ймовірності для системи, граф можливих станів якої наведено на рис. 16. 3, за умови:

$$\lambda_{12} = 1; \lambda_{13} = 2; \lambda_{21} = 2; \lambda_{24} = 2; \lambda_{31} = 3; \lambda_{42} = 3; \lambda_{34} = 1; \lambda_{43} = 2.$$

Скориставшися групою формул (59) за умови $\frac{dp_i}{dt} = 0 \quad (i = 1, 2, 3, 4)$, маємо:

$$\begin{cases} 3p_1 = 2p_2 + 3p_3 \\ 4p_2 = p_1 + 3p_4 \\ 4p_3 = 2p_1 + 2p_4 \\ 5p_4 = 2p_2 + p_3 \\ p_1 + p_2 + p_3 + p_4 = 1. \end{cases}$$

Ураховуючи вимоги до точності розрахунків (імовірності), маємо: $p_1 = 0{,}4$; $p_2 = 0{,}2$; $p_3 = 0{,}27$; $p_4 = 0{,}13$. Одержаний результат свідчить, що при досить тривалому терміні експлуатації системи, приблизно 40% часу вона перебуватиме в стані S_1 (обидва вузли справні); 20% – у стані S_2 (перший вузол ремонтується, а другий працездатний); 27% – у стані S_3 (другий вузол ремонтується, а перший працездатний); 13% часу – в стані S_4 (обидва вузли ремонтуються).

16.10. Процес розмноження та вимирання

У теорії масового обслуговування широко використовується математична модель випадкових процесів, які називаються процесами *розмноження та вимирання.* Назва обумовлена використанням моделі при дослідженні чисельності біологічних популяцій. Граф можливих станів процесу розмноження та вимирання зображено на рис. 16. 6.

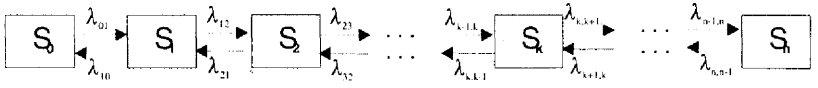


Рис. 16. 6.

Упорядженість множини станів виконаємо за показником чисельності популяції k, а перехід від стану S_k до сусідніх станів S_{k+1} та S_{k-1} відбувається за умови відповідного народження або загибелі одного члена популяції. Отже, особливістю

досліджуваного процесу є те, що від кожного можливого стану перехід може бути лише до одного з *двох сусідніх станів.* За умови, що потоки подій, які переводять систему зі стану в стан, елементарні з відомими інтенсивностями $\lambda_{k,k+1}$ та $\lambda_{k+1,k}$, складемо та розв'яжемо систему алгебраїчних рівнянь обчислення граничних імовірностей стабільного процесу. Як наголошувалося раніше, такий стабільний процес обов'язково досягається, що обумовлено можливостями переходів від стану до стану та скінченною множиною можливих станів. За характером розбудови математичної моделі враховується зміна чисельності популяції, що віддзеркалюється врахуванням станів та інтенсивності їх зміни. Названі зміни характеризуються коефіцієнтами інтенсивності $\lambda_{k,l}$. Скориставшися формулою (60) за умови зміни індексу s від 0 до N та можливістю використання рекурентних відношень, одержуємо систему:

$$\begin{cases} \lambda_{01}\, p_0 = \lambda_{10}\, p_1 \\ \lambda_{12}\, p_1 = \lambda_{21}\, p_2 \\ \dots\dots\dots\dots\dots\dots \\ \lambda_{k-1,k}\, p_{k-1} = \lambda_{k,k-1}\, p_k \\ \dots\dots\dots\dots\dots\dots \\ \lambda_{N-1,N}\, p_{N-1} = \lambda_{N,N-1}\, p_N. \end{cases} \tag{76}$$

До цієї системи необхідно долучити рівняння:

$$p_0 + p_1 + p_2 + \ldots + p_N = 1. \tag{77}$$

З використанням раніше введених позначень $d_i = \frac{\lambda_{i-1,i}}{\lambda_{i,i-1}}$ (65) отримаємо:

$$p_0 = \left[1 + \sum_{k=1}^{N} \prod_{i=1}^{k} d_i\right]^{-1}; \quad p_i = p_{i-1} d_i \tag{78}$$

(П – добуток записаних справа від літери співмножників). Моделюючи реальні біологічні процеси розмноження та вимирання популяції, необхідно мати відомості про вихідний стан S_0 чисельності популяції та коефіцієнти інтенсивності народження і загибелі членів популяції в одиницю часу, що може бути обумовлено багатьма причинами, починаючи з кліматичних та екологічних.

Приклад. Процес розмноження та вимирання популяції моделюється графом на рис. 16.6. Можна прийняти, що за відомого вихідного стану чисельності S_0 про-

цес зміни визначається теплою та холодною частинами року. Знайти граничні ймовірності можливих чисельностей за відомих $\lambda_{01} = 1, \lambda_{10} = 4, \lambda_{12} = 2, \lambda_{21} = 3$.

За формулами (78) маємо:

$$p_0 = \left(1 + \frac{1}{4} + \frac{2}{3} \cdot \frac{1}{4}\right)^{-1} = 0{,}706;\ p_1 = \frac{1}{4} 0{,}706 = 0{,}176;\ p_2 = \frac{2}{3} \cdot \frac{1}{4} \cdot 0{,}706 = 0{,}118.$$

При відомій чисельності в моменти виконання спостережень S_0, S_1, S_2 та інтенсивностях її зміни кількість S_0 членів популяції спостерігатиметься в середньому в 70,6% часу, S_1 – в 17,6%, S_2 – в 11,8%, якщо спостереження виконуватимуться тривалий час.

Питання для самоконтролю

1. Основна мета розбудови систем масового обслуговування (СМО).
2. Основні характеристики формалізації роботи СМО.
3. Поняття вхідного потоку запитів, його основна характеристика.
4. Поняття каналу обслуговування, його основна характеристика.
5. Які особливості організації СМО використовуються при їх класифікації?
6. Як визначається пропускна спроможність СМО? Що є умовою стабільного функціонування СМО?
7. Роботу яких СМО можна описувати за допомогою рівнянь Колмогорова?
8. За яких умов СМО може досягнути стабільного режиму роботи?
9. Моделювання процесу розмноження та вимирання, його графічне зображення та умови зміни можливих станів.

ГЛАВА 17

ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ ОПЕРАЦІЙНИХ СИСТЕМ

17.1. Імітаційний підхід

Моделі оптимізації, розбудовані з використанням математичного апарату лінійного, нелінійного та динамічного програмування, теорії гри, теорії масового обслуговування й управління запасами, є ефективним засобом дослідження операцій за умови існування чітко визначеної кінцевої мети та можливостей сформулювати критерії, які дозволяють порівняти між собою й оцінити різні варіанти досягнення мети досліджуваної операції. Говорячи про існування як мети, так і оцінки (за певними критеріями) шляхів її досягнення, маємо на увазі їх представлення з використанням певного математичного апарату, тобто засобами і можливостями математики. Такий підхід до дослідження операцій дозволяє цілеспрямовано й ефективно промоделювати та дослідити певні аспекти економічних і соціально-економічних процесів.

Ці процеси надто складні за своєю суттю: їх перебіг відбувається за нестабільних, важко передбачуваних умов, коли змінюються в часі як обставини, так і цілі, тому управління ними вимагає врахування значних обсягів різноманітної інформації, змінної в часі. У таких випадках для одержання більш-менш достовірної інформації при прийнятті управлінських рішень використовуються певні засоби моделювання ситуацій з урахуванням змінних у часі чинників.

Наявний математичний апарат моделювання в теорії операцій забезпечує розбудову моделей різного рівня – від частинних до загальних. Вони являють собою множину математичних залежностей, з використанням яких можна обчислити або якісно оцінити результати моделювання різних варіантів можливого перебігу досліджуваного процесу. Моделі такого змісту називаються *імітаційними* (лат. imitatio – підробка, наслідування). З їх допомогою імітують, "програють" можливі варіації перебігу процесу й одержують інформацію для оцінки таких варіацій.

Метою розбудови та виконання розрахунків з використанням імітаційної моделі є накопичення інформації, аналіз якої в поєднанні з можливими впливами на характер перебігу процесу дозволив би спрямувати процес у бажане русло або ж зробити неможливими негативні наслідки. Тобто імітаційна модель є засобом (іноді єдино можливим) аналізу "конфліктної ситуації". Результати такого аналізу, як свідчить досвід, значно надійніші за результати відповідних експертиз, оскільки при використанні значної кількості необхідних математичних моделей у комп'ютері є реальна можливість швидко оцінити різноманітні допустимі варіанти перебігу процесу.

Імітаційна модель будь-якого призначення може бути ефективною при використанні усіх її можливостей лише за умови оперативного спілкування: людина – комп'ютер – людина. Використання імітаційних моделей дозволяє наочно оцінити роль чинника часу. Проектування та запровадження технологій імітації набуває все більшого значення в управлінні різноплановими процесами – як соціально-економічними, так і науковими, інженерно-технічними.

При використанні імітаційних моделей не розв'язуються складні екстремальні задачі – користувач мусить діяти раціонально в межах свого уявлення про можливі ситуації в перебігу імітованого процесу. За такої умови розбудова імітаційної моделі можлива лише при значному сервісному програмному забезпеченні. Зміст і методика розбудови мають бути чітко й однозначно окреслені, щоб користувач мав змогу працювати в діалоговому режимі.

17.2. Задачі економіки, при розв'язанні яких доцільно використовувати імітаційне моделювання

Моделі задач економіки й управління, які використовують методи математичного програмування та стохастичного моделювання, є ефективними засобами дослідження операцій за певної стабільності у виконанні тих умов, при наявності яких формуються такі моделі. Можливості будь-якого моделювання обмежуються віддзеркаленням реально існуючих явищ та процесів. Імітаційне моделювання дозволяє об'єктивніше та глибше дослідити певний процес. Наведемо перелік деяких значимих задач управління й економіки, які можуть бути ефективно розв'язані з використанням імітаційного моделювання.

1. *Визначення інвестиційної політики при перспективному плануванні.* Інвестиційна політика значних промислово-економічних об'єднань (фірм) з необхідністю має враховувати: фінансове забезпечення науково-дослідних і конструкторських робіт при створенні нових видів продукції; можливості ринку збуту та його розширення; оцінку доцільності й ефективності втілення проектів (давно розроблено й апробовано двигун внутрішнього згоряння з кераміки, але найближчим часом масове його виробництво не планується); оцінку рівня ризику при плануванні комплексу робіт; джерела й обсяги фінансування (кредит, залучення капіталу за допомогою продажу акцій і т. ін.); зміни фонду заробітної плати; фінансово-економічні стосунки з іншими об'єктами і т. ін. При розбудові ефективної імітаційної операційної моделі, за допомогою якої можна було б аналізувати й оцінювати різні можливі варіанти інвестиційної політики, необхідно враховувати стохастичну природу та динамічний характер інвестування, забезпечити оцінку ефективності кожного з можливих альтернативних проектів.

2. *Вибір і комплектація засобів обслуговування при поточному плануванні.*

Деякі задачі такого класу було розглянуто в теорії масового обслуговування. Моделі цієї теорії ґрунтуються на певних гіпотезах, представлених аналітичними залежностями стосовно вхідного потоку запитів на обслуговування та і його реалізації. Такі гіпотези дають можливість розбудови відповідних моделей у завершеному математичному представленні, але вельми умовно дозволяють віддзеркалити реальності процесу обслуговування. Використання імітаційного моделювання в розв'язанні таких задач дає можливість більш об'єктивно і оперативно досліджувати як проектування систем обслуговування, так і поточне управління.

3. *Планування з урахуванням зворотних інформаційних зв'язків і вимог поточного управління.* До задач названого класу належать задачі поточного планування на підприємствах малооб'ємного різнопланового серійного виробництва, ремонту техніки і т. ін. При плануванні організації роботи таких підприємств необхідно враховувати змінювані терміни виконання замовлень, запити на замовлення, наявні ресурси, виробничі потужності, рівень кваліфікації виконавців, можливості постачання необхідної сировини і т. ін. У залежності від поточної інформації при виконанні та запитів на нові замовлення необхідно оперативно коригувати існуючі графіки роботи або складати нові. Ще одним прикладом є планування роботи багатогалузевого приміського сільськогосподарського об'єкта, поточна ефективна робота якого тісно пов'язана зі щоденною реалізацією продукції на ринку або ж її переробкою.

Використання імітаційного моделювання при розв'язанні задач поточного управління у вищенаведених випадках обумовлене необхідністю одночасного врахування чинників невизначеності в поєднанні динамічних взаємообумовлених поточних рішень та їх майбутніх наслідків. До якої б послідовності відомих математичних моделей не було приведено розв'язання цих задач, необхідність оперативного управління в залежності від досягнутого наявного стану робить неможливим єдиний для всього часу розв'язок. При моделюванні функціонування таких систем слід виконувати багаторазово повторювані розв'язки при змінних у часі параметрах для дискретної та чітко визначеної послідовності проміжків часу.

Розв'язання задач організаційного управління шляхом імітації реальних умов і відповідних дій на зразок військових навчань, уявляється занадто незручним і затратним, а іноді й принципово неможливим. З багатьох точок зору доцільно й ефективно імітувати функціонування складної системи за допомогою множини певних математичних моделей, записаних у комп'ютері. Імітаційний експеримент за допомогою математичної імітаційної моделі вимагає відповідних математичних досліджень та значного обсягу обчислень, але це економніше, надійніше, ефективніше, ніж використання інших методів дослідження.

17.3. Кроки практичної реалізації імітаційного моделювання

Наведемо в загальних рисах характеристики кожного з основних етапів робіт, які необхідно виконати для практичного втілення методу імітаційного моделювання.

1. *Розбудова моделі.* Призначення та зміст цього етапу нічим іншим, за винятком більш широкого та глибокого аналізу досліджуваного процесу, не відрізняються від відповідного етапу при будь-якому математичному моделюванні.

Структуру імітаційної моделі в більшості випадків доцільно розбудувати, визначивши зміст і призначення наявних у досліджуваній системі динамічних процесів. Стисло розглянемо основні вимоги до розбудови імітаційної моделі функціонування виробничої системи:

1) планування поточного виробництва продукції, ефективного за цінами (що неможливо без високоефективних технологій);

2) урахування термінів виробництва, обсягів збуту, прибутку;

3) визначення затрат на транспортування та можливості зберігання як ресурсів виробництва, так і продукції;

4) визначення доцільності інвестування в подальше виробництво цієї ж продукції за умови його економічної ефективності або ж використання засобів виробництва й одержаного прибутку для інших цілей.

Розв'язання кожної з указаних задач виконується з використанням відповідної математичної моделі за умови вибору певних значень необхідних вхідних параметрів. Результати функціонування реальної системи завжди мають певне фізичне втілення. При використанні певної імітаційної моделі вибір значень вхідних параметрів теж визначає стан імітаційної моделі. Перебіг реального процесу для вказаних задач обумовить *певні значення* параметрів, хоч і вимірюваних з деякою точністю; досягнутий стан реальної системи більш чітко визначить подальший перебіг, але при цьому необхідно враховувати екзогенні (зовнішні) чинники.

При дослідженні відповідних реальних задач із використанням імітаційних моделей необхідно вибирати певні значення параметрів, дійсні величини яких можна оцінити лише з певною достовірністю. Тому, розбудовуючи імітаційну модель, доцільно мати відомості (або приймати припущення) про межі змін використовуваних параметрів.

2. *Розбудова проекту експерименту.* Значення та якість імітаційної моделі мають бути перевіреними експериментально до її практичного використання. Для виконання експерименту необхідно чітко визначити, які функціональні характеристики імітованої системи та як будуть досліджуватися, як будуть оцінюватися можливі відхилення значень використовуваних у моделі параметрів, рівень відповідності результатів розрахунків дійсному стану системи. Тому по розбудові імітаційної моделі постає важлива проблема розробки схеми перевірки пропонованої моделі й умов співставлення реальних даних, які визначають процес, з

результатами розрахунків. Для цього необхідно визначити вихідні умови функціонування модельованої системи, параметри характеристики різних аспектів системи, вплив часу та методику його врахування, змінні, значення яких необхідно співставляти з експериментальними даними, та способи одержання достовірних даних про стан реальної системи.

Труднощі в аналізі результатів імітаційного моделювання пов'язані з необхідністю врахування впливу випадкових величин значущих чинників і можливих випадкових процесів.

3. *Розбудова програмного забезпечення.* У загальних рисах його методику можна описати так. Оскільки в імітаційних моделях обов'язково необхідно враховувати плин часу, то слід мати "хребет" програми, який називають **основною програмою.** У такій програмі передбачається перехід до використання допоміжних програм або процедур у залежності від результатів, одержаних при використанні попередніх процедур (програм), за допомогою кожної з яких моделюють аспекти роботи досліджуваної реальної системи. Іншими словами: в основній програмі мають аналізуватися результати розрахунків за певною методикою, реєструватися можливі виклики необхідних процедур і формуватися вхідні параметри для них. В основній програмі слід обов'язково передбачити діалоговий режим її експлуатації. Перебіг імітаційного процесу забезпечується переходами від однієї процедури до іншої в залежності від аналізу результатів попередньої процедури (або процедур). Особлива увага має бути приділена проміжним розрахункам при реалізації діалогового режиму та завершальних результатів, які використовуються при прийнятті відповідних управлінських рішень.

17.4. Приклад блок-схеми імітаційної моделі

Представлення будь-якої значимої імітаційної моделі на рівні програми вимагає значного обсягу інформації як програмної, так її супроводжуючої, тому обмежимося розглядом лише блок-схеми. На рис. 17.1 стисло зображено блок-схему імітаційної моделі балансового середньотермінового макроекономічного прогнозування, яка використовується в Республіці Беларусь*. Наголосимо на деяких її особливостях, які віддзеркалюють суть імітаційного моделювання.

Модель має шість блоків різного цільового призначення, але використовуваних послідовно: блок вартості виробництва та витрат, блок прибутків і витрат держави, блок прибутків і витрат населення, блок розрахунків національного прибутку та продукту, блок платіжного балансу, блок кредитно-грошового сектора.

* Экономико-математические методы и модели: Учеб. пособие / Н. И. Холод, А. В. Кузнецов, Я. Н. Жихар и др.; Под общ. ред. А. В. Кузнецова. – 2-е изд. - Минск: БГЭУ, 2000.

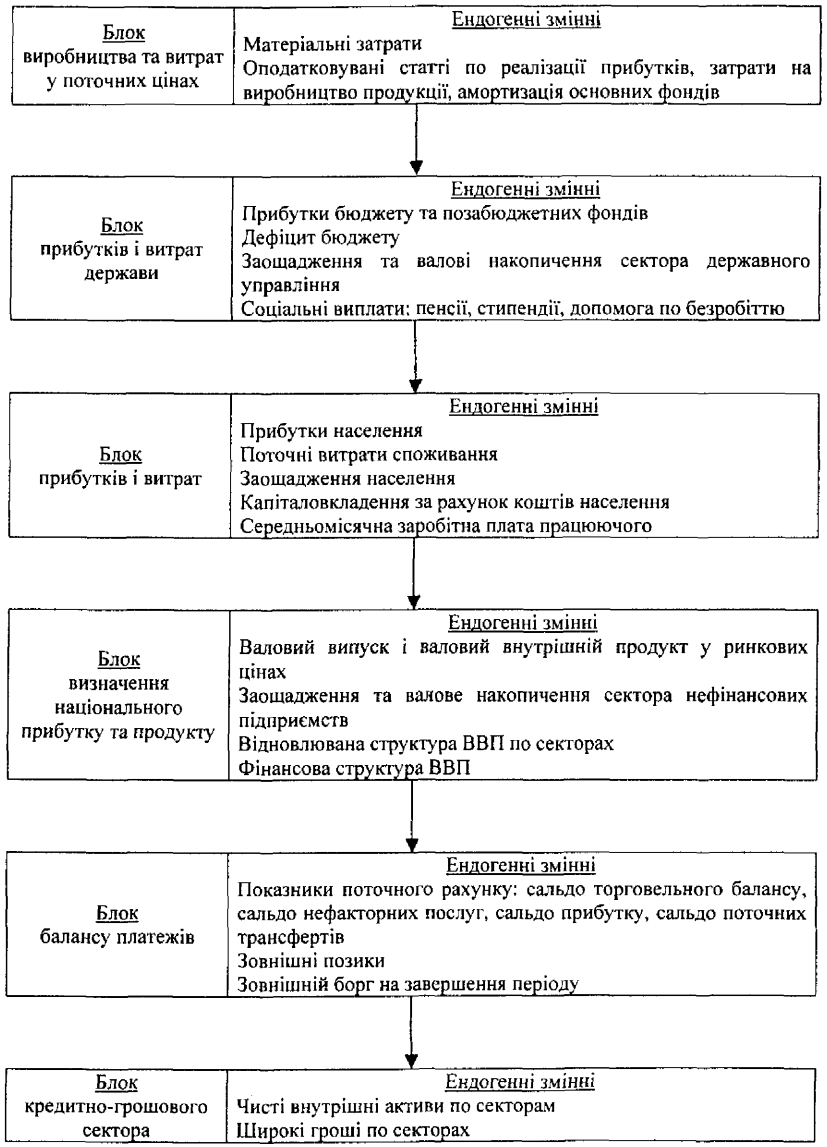


Рис. 17.1. Блок-схема макромоделі середньотермінового прогнозування

Модель представлена 300 аналітичними залежностями в поточних цінах, 50 – в постійних і програмно реалізована з використанням засобів EXCEL.

Множину *екзогенних* параметрів моделі створюють три групи показників: показники перебігу економічного процесу (індикатори зовнішньої та внутрішньої кон'юнктури), цільові параметри економіки (що планується досягти), параметри економічної політики (ставки податків, витрати бюджету та позабюджетних фондів, ставки за депозитом). Прямі зв'язки між показниками моделюються за допомогою побудованих аналітичних залежностей. Зворотні залежності досліджуються в такий спосіб. Наприклад, якщо за вибраних величин вхідних параметрів по виконанні розрахунків одержано, що у фінансово-кредитному покритті запитів приватного сектора визначальна роль відводиться зовнішнім позичкам, доступ до яких у цього сектора вельми складний, то це означає, що розрахунки кінцевого запиту, з точки зору його кредитної забезпеченості, є нереальними. Така ситуація викличе спад на товари та послуги з боку приватного сектора, і це обумовлює необхідність коригування відповідних вхідних параметрів.

Модель дозволяє також досліджувати в режимі імітації вплив грошової маси на перебіг інфляції. Індикатором адекватності такої залежності реальному стану є швидкість обігу грошей, динаміка якої тісно пов'язана з інфляцією: при підвищенні інтенсивності інфляційних процесів швидкість обігу грошей збільшується. Якщо результати розрахунків динаміки швидкості обігу грошей та темпу інфляції співпадають якісно, то є підстави вважати, що прогнозований попит на грошові активи є реальним. За протилежних результатів обсяги запиту на грошову масу коригуються вибором належних величин вхідних параметрів. Розбудована відповідно наведеній блок-схемі імітаційна модель з успіхом використовується для прогнозування перебігу економічного процесу в Республіці Бєларусь і засвідчила свою практичну доцільність*.

У літературі є відомості про широке використання імітаційного моделювання при аналізі планових військових операцій, науково-проектних робіт і т. ін. У теорії дослідження операцій імітаційне моделювання вважається однією з перспективних розробок.

* Экономико-математические методы и модели: Учеб. пособие / Н. И. Холод, А. В. Кузнецов, Я. Н. Жихар и др.; Под общ. ред. А. В. Кузнецова. – 2-е изд. - Минск: БГЭУ, 2000.

Питання для самоконтролю

1. Що розуміють під терміном "імітаційна модель"?
2. Принципові відмінності між формалізованими математичними моделями та імітаційним моделюванням.
3. Мета розбудови імітаційної моделі.
4. Які чинники обумовлюють доцільність та ефективність використання імітаційного моделювання?
5. Основні задачі управління й економіки, при розв'язанні яких доцільно користуватися імітаційними моделями.
6. Основні етапи розбудови імітаційної моделі.
7. Імітаційне моделювання та прийняття управлінських рішень.

ГЛАВА 18

ДОСЛІДЖЕННЯ ІНВЕСТИЦІЙНИХ ПРОЕКТІВ

18.1. Задача вибору інвестиційного проекту

Під терміном *інвестиційний проект* будемо розуміти план будь-якого використання грошей з метою одержання зиску. Приклади інвестиційних проектів розташовані в надзвичайно широкому діапазоні: комерційна діяльність, модернізація виробництва, створення нового виробництва за принципово нових технологій, виконання певних проектно-конструкторських робіт з наступною їх реалізацією, планування науково-дослідницьких робіт і т. ін. Задача вибору інвестиційного проекту за певним критерієм або критеріями може мати різний зміст: вибір кращого серед проектів досягнення однієї й тієї ж мети різними шляхами, наприклад, налагодження виробництва товарів певного найменування; вибір серед планів організації виробництва різних номенклатур товарів; раціональний розподіл коштів між різними роботами, які виконуються в ході реалізації якогось проекту; визначення доцільних галузей господарювання за певних природно-технічних і фінансових умов і т. ін.

У загальному випадку кожне діюче підприємство для збільшення обсягів прибутку, підвищення рентабельності своєї роботи має постійно розробляти проекти, які передбачають використання коштів для розробки нових технологій, організації нового виробництва, реконструкції діючого з метою підвищення конкурентоспроможності продукції на внутрішньому та зовнішньому ринках, і чимало інших проектів. Кожний з них за умов ринкової економіки має бути глибоко обґрунтованим стосовно обсягу прибутку, який очікується після втілення проекту.

Розглянемо деякі загальні аспекти оцінки ефективності інвестиційного проекту. Проблеми вибору проекту та визначення пріоритетів при необхідності виконання проектів у певній послідовності завжди доводиться розв'язувати при обмеженні коштів (яких завжди і всюди не вистачає) і строків, тому в реальних ситуаціях це виключно складне завдання. Необхідно також враховувати, що відповідальність за реалізацію проекту та за його фінансування стосуються, як правило, різних організаційних структур.

Стисло розглянемо моделі та методи, використовувані для *економічної оцінки* певного проекту за умови як стабільності грошової одиниці, так і наявності інфляційних процесів. Розв'язання задачі оцінки проектів будемо досліджувати за двох різних умов: а) ефективність розміщення наявних коштів; б) доцільність використання банківського кредиту для інвестування.

18.2. Дисконтування грошових потоків

Грошовим потоком називається *розподіл у часі* коштів, які необхідно інвестувати в процесі реалізації проекту, та прибутків, очікуваних від втілення проекту на певному проміжку часу й обумовлених проектом. Для того, щоб оцінити ефективність грошових потоків різних інвестиційних проектів у поточний момент, слід якось порівняти потоки, враховуючи різну зміну в часі обсягів інвестування і реінвестування та прибутків, а також певний рівень ризику в досягненні як обсягів, так і строків надходження планових грошових потоків у майбутньому.

Для розв'язання проблеми порівняння грошових потоків, обумовлених досліджуваним проектом, у сучасному фінансовому аналізі використовується **метод дисконтування.** Терміном "дисконтування" визначається приведення вартостей, очікуваних у майбутньому, до сьогочасного моменту. За ринкової економіки інвестування будь-якого проекту має сенс лише за умови, що очікуваний прибуток буде не менше середньоринкової ставки доходності або банківського депозиту. Цим і обумовлена необхідність дисконтування.

Приймемо, що обсяги вартостей будемо оцінювати в грошових одиницях, а одним з альтернативних досліджуваних проектів є банківський депозит за відомого коефіцієнта β зиску від банківського депозиту. Банківський відсоток дорівнює 100β. Доцільно наголосити, що в сучасній економічній літературі термінологічно все ширше використовуються не відсотки, а відповідні їм коефіцієнти, іноді навіть не відрізняючись у тексті. Перехід до відповідних коефіцієнтів спрощує аналітичні викладення. Надзвичайно суттєві питання врахування інфляції та ризику поки що вилучаємо.

Припустимо, що в банку мусимо розмістити початкові кошти в обсязі P з відомим коефіцієнтом β банківського відсотка. Обсяг коштів C, які одержимо за n періодів капіталізації, обчислюємо за формулою:

$$C = P(1+\beta)^n. \qquad (1)$$

У наведеній формулі коефіцієнт $(1+\beta)^n$ визначає обсяг коштів, які будуть одержані за n періодів капіталізації при вихідному капіталі обсягом в одну грошову одиницю. Цей коефіцієнт називається *коефіцієнтом акумуляції* (накопичення), позначимо його літерою α.

Використання формули (1) дозволяє обчислити обсяг коштів, який необхідно розмістити в банк, щоб через n періодів капіталізації одержати кошти в обсязі C. Такий обсяг вихідного капіталу називають *сьогочасною (або приведеною) вартістю суми C* і позначають символом PV (від англ. present value – сьогочасна величина):

$$PV = C/(1+\beta)^n. \qquad (2)$$

Процес визначення сьогочасної вартості називається дисконтуванням. При використанні банківського депозиту процес дисконтування реалізується згідно з формулою (2). Величина $1/(1+\beta)^n$ називається *коефіцієнтом дисконтування*; вона показує, яку суму коштів необхідно покласти на банківський депозит, з відомим коефіцієнтом банківського відсотка, щоб за n періодів капіталізації одержати обсяг коштів в одну грошову одиницю. Позначимо цей коефіцієнт δ:

$$\delta = 1/(1+\beta)^n. \qquad (3)$$

Приклад. Які кошти необхідно покласти на банківський депозит за відомого коефіцієнта банківського відсотка 0,12, щоб через п'ять років одержати 1000 у. о.? Скориставшися формулою (3), маємо $\delta = 1/(1+0{,}12)^5 = 0{,}56743$; отже, обсяг депозиту (сьогочасна вартість очікуваної через п'ять років суми коштів в обсязі 1000 у. о.) $PV = 1000 \cdot 0{,}56743 = 567{,}43$ *у. о.*

Приймемо, що *рівні ризиків одержати кошти обсягом C через n періодів капіталізації за рахунок реалізації проекту та банківського депозиту є тотожними.* За такої умови для порівняння ефективності інвестування досліджуваного проекту можна скористатись альтернативним банківським депозитом. При відомому коефіцієнті банківського відсотка β замість того, щоб інвестувати проект, після реалізації якого через n періодів очікують прибуток обсягом C, можна покласти на банківський депозит кошти обсягом $C/(1+\beta)^n$.

Обсяги коштів при відомому банківському відсотку, необхідних для банківського депозиту з метою одержання *очікуваного прибутку від проекту,* називаються *сьогочасною (приведеною) вартістю проекту.* Коефіцієнт відсоткової ставки, використовуваний при дисконтуванні грошових потоків, називається коефіцієнтом *норми дисконтування.* Таким чином, за умови тотожності ризиків оцінка сьогочасної вартості проекту виконується при порівнянні обсягів інвестиції та банківського депозиту, необхідного для одержання такого ж прибутку, як і очікуваний по реалізації проекту.

Приклад. Для реалізації проекту необхідні 700 у. о. інвестицій. Очікуваний прибуток через два роки дорівнює 1000 у. о., а банківський щорічний відсоток - 12%. Визначити сьогочасну вартість проекту.

Скориставшися формулою (2), маємо:

$$PV = 1000/(1+0{,}12)^2 = 797{,}19 \ \textit{у.о.}$$

Результат розрахунку означає: для того, щоб одержати через два роки кошти в

обсязі, очікуваному від реалізації проекту (1000 у. о.), для банківського депозиту потрібні кошти в обсязі 797,19 у. о. Тобто досліджуваний проект вимагає початкової інвестиції на 97,19 у. о. менше, ніж банківський депозит (797,19-700=97,19).

У загальному випадку, якщо коефіцієнт норми дисконтування вибрано з урахуванням величини коефіцієнта банківського відсотка, то вводиться поняття так званої *чистої сьогочасної (або чистої приведеної) вартості* проекту, яку позначають *NPV* (початкові літери net present value) і обчислюють за формулою:

$$NPV = PV - I_0 \qquad (4)$$

(PV – сьогочасна вартість проекту, обчислена за формулою (2); I_0 – обсяг початкової інвестиції в проект).

Чиста сьогочасна вартість *NPV* проекту показує як відрізняються обсяги початкових інвестицій в проект і банківського депозиту, необхідного для одержання такого ж прибутку, як і очікуваний по реалізації проекту. Якщо величина *NPV* додатна, то це означає, що гроші краще інвестувати в проект, ніж покласти на банківський депозит. Якщо величина *NPV* від'ємна, то доцільно скористатися банківським депозитом.

Тепер розглянемо характеристику динаміки грошових потоків стосовно досліджуваного проекту, яка називається *власною нормою прибутковості проекту – IRR* (від internal rate of return). Величина *IRR* визначає *коефіцієнт норми дисконтування r* грошових потоків, при якому обсяг початкового інвестування I_0 дорівнюватиме сьогочасній вартості *PV* очікуваного прибутку *C* по реалізації проекту, тобто $C = I_0(1+r)^n$. Таким чином, величина *IRR* є розв'язком рівняння:

$$I_0 = \frac{C}{(1+r)^n} \qquad (5)$$

(n – кількість проміжків часу від початкового інвестування до одержання очікуваної величини прибутку *C* по втіленні проекту). Величина проміжку часу, використовувана в формулі (5), дорівнює проміжку капіталізації за банківського депозиту.

Приклад. Обчислимо власну норму прибутковості *IRR* проекту за даними, наведеними в попередньому прикладі. Нагадаємо, що $I_0 = 700\ у.о.$, $C = 1000\ у.о.$, $n = 2$. Для обчислення величини *IRR* скористаємося рівнянням (5):

$$700 = 1000/(1+r)^2.$$

Розв'язок цього рівняння $r = 0{,}1952$, тобто $IRR = 0{,}1952 = 19{,}52\%$. Одержаний

результат свідчить, що досліджуваний проект має власну норму прибутковості *IRR*, тотожну банківському депозиту з коефіцієнтом банківського щорічного відсотка 0,1952 (19,52%).

Як наголошувалося раніше, коефіцієнт банківського відсотка доцільно брати за коефіцієнт норми дисконтування грошових потоків проекту, лише за умови тотожності ризиків, пов'язаних з очікуваним обсягом прибутку по втіленні проекту та банківського депозиту. У загальному випадку *за коефіцієнт норми дисконтування грошових потоків проекту необхідно брати власну норму прибутковості альтернативних проектів з такими ж фінансовими ризиками, як і досліджуваний.* Доцільно наголосити, що проекти з більшим фінансовим ризиком на фінансовому ринку для забезпечення конкурентноздатності мають бути з більшою власною нормою прибутковості.

Тепер розглянемо на конкретному прикладі, як оцінити два альтернативних проекти за тотожними фінансовими ризиками, *вилучаючи банківський депозит.* Припустимо, що проекту, фінансові характеристики якого були наведені раніше (обсяг початкових інвестицій $I_0 = 700$ у. о., очікуваний від реалізації проекту прибуток через два роки 1000 у. о.), має альтернативний проект з такими ж ризиком та очікуваним прибутком з коефіцієнтом власної норми прибутковості *IRR* 0,18 (18%) щорічно. У досліджуваній ситуації для порівняння альтернативних проектів треба брати коефіцієнт норми дисконтування 0,18. Знайдемо сьогочасну *PV* та чисту сьогочасну *NPV* вартість досліджуваного проекту за названої умови:

$$PV = 1000/(1+0{,}18)^2 = 718{,}18\ у.о.; \quad NPV = 718{,}19 - 700 = 18{,}18\ у.о.$$

Одержаний результат *NPV* =18,18 у. о. свідчить, що альтернативний проект порівняно до досліджуваному вимагає більше початкових інвестицій на 18,18 у. о.

Для більш чіткого з'ясування змісту параметрів порівняння проектів ми розглянули проекти з одноразовими обсягами інвестування та очікуваного прибутку. У загальному випадку проект може мати кілька різночасових обсягів інвестування (реінвестування) та кілька обсягів прибутків, очікуваних у різні терміни. Грошові потоки таких проектів зручно зображати у вигляді діаграми, побудованої за такою схемою: на горизонтальній осі у певному масштабі помічають контрольовані проміжки часу, під нею проти відповідних точок часу вказують зі знаком "–" обсяги поточного інвестування (реінвестування), над віссю – обсяги очікуваних прибутків для відповідних моментів.

Приклад. Досліджуваний проект має такі грошові потоки: вихідний обсяг інвестування – 2000 у. о., додатковий обсяг інвестування (реінвестування) через рік 400 у. о., на проміжку в чотири роки через кожний рік очікувані обсяги прибутку в такій послідовності: 600; 1200; 1400; 1000.

Діаграма грошових потоків за названих умов має вигляд (рис. 18.1):

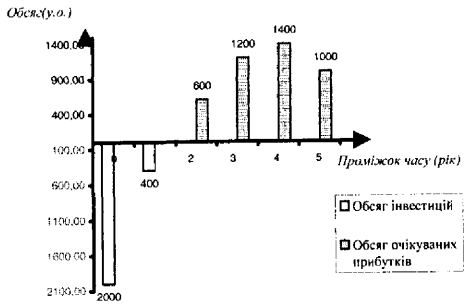


Рис. 18.1.

Якщо проект генерує кілька грошових потоків, то сьогочасна вартість його PV визначається за формулою, аналогічною (2):

$$PV = \sum_{k=1}^{n} C_k /(1+r)^k \qquad (6)$$

(C_k – обсяг очікуваного прибутку по завершенні k-го проміжку; r –коефіцієнт норми дисконтування). Економічний зміст PV такий же, як і для випадку, коли проект генерує лише один потік. Нагадаємо, що коефіцієнт норми дисконтування r визначається як величина коефіцієнта власної норми прибутковості альтернативного проекту за такого ж ризику, як і в досліджуваного.

Для проектів, генеруючих кілька грошових потоків, коефіцієнт власної норми прибутковості IRR дорівнює такому *коефіцієнту дисконтування* r, за якого приведені (сьогочасні) обсяги інвестування дорівнюють сьогочасній вартості проекту (6). Отже, IRR є коренем рівняння:

$$\sum_{l=0}^{p} I_l /(1+r)^l = \sum_{k=p+1}^{n} C_k /(1+r)^k . \qquad (7)$$

(I_l – обсяги вихідного потоку та реінвестування в l-ті проміжку часу; C_k – обсяги очікуваних прибутків по завершенні k-их проміжків).

При дослідженні ефективності інвестування необхідно враховувати не лише його природну певного рівня невизначеність у майбутньому, а й інфляційний процес, у різній мірі характерний для всіх країн. Оскільки розглядаємо економічний аспект інвестування, основною метою якого є одержання прибутку, то принагідно стисло розглянути деякі аспекти впливу інфляції на прибуток.

18.3. Прибуток та інфляція

Під терміном "*інфляція*" (від лат. inflation – розбухання) будемо розуміти будь-яке *підвищення* цін на товари та послуги, зумовлене зміною загальної економічної рівноваги. Зниження цін називається *дефляцією* (від лат. deflare – здувати). Одним із вимірів інфляції стосовно будь-якої національної валюти є пряме зіставлення цін певного кошику товарів через певний період часу.

За методикою, використаною у США, інфляція визначається за *середньою ринковою вартістю* товарів 3000 найменувань. Середню вартість товарів кошику, використовуваних при дослідженні інфляції, позначимо c і назвемо *вартістю одиниці умовного товару*. При такому тлумаченні величина c характеризує перебіг інфляції.

Зміна купівельної спроможності національної грошової одиниці певної країни може вимірюватися також зіставленням цін у даній країні з цінами в країні з більш стабільною національною валютою.

Для дослідження інфляції будують динамічний ряд, рівні якого відповідають перебігу вартості одиниці умовного товару. *За числову характеристику* ***інтенсивності інфляції*** *вибирають темп зміни вартості одиниці умовного товару.* Якщо позначимо рівні відповідного динамічного ряду c_i, то інтенсивність інфляції v_i в момент часу t_i буде визначатися відношенням $c_i(t_i)/c_{i-1}(t_{i-1})$, тобто відношенням *поточної* величини вартості одиниці умовного товару та її величини на початку кожного проміжку розбудови динамічного ряду. У залежності від темпу знецінення купівельної спроможності грошової одиниці виділяють кілька різних за економічною значимістю видів інфляції. Якщо темпи знецінення грошей не перевищують зростання цін в межах 10-15% на рік, то таку інфляцію називають "*повзучою*". Така інфляція, на думку О. М. Мельника, не тільки не несе в собі суттєвих негативних наслідків, а є стимулом економічного розвитку та пожвавлення ділової активності*. Повзучу інфляцію навмисно стимулює держава, прагнучи посилити розподіл реальних доходів на свою користь. Інфляція такого рівня досить легко контролюється.

При "*галопуючій*" інфляції спостерігається більш значне зростання темпів знецінення грошей. Темп знецінення може досягти 100% на рік, що різко негативно впливає на всі сфери економічного та соціального життя країни.

Якщо темпи зростання цін перевищують 100% на рік, то починається "*гіперінфляція*" (зверхінфляція). За такого процесу гроші починають втрачати здатність виконувати свої функції, що порушує неперервність процесу відтворення

* Мельник О. М. Інфляція: теорія і практика регулювання. – К., 1999.

та баланс інтересів його суб'єктів, породжує надмірну інтенсифікацію перерозподільчих процесів, посилює галузеві диспропорції, стимулює стихійні та некеровані процеси в економіці та соціальному житті.

Для спрощення розрахунків користуються також **коефіцієнтом темпу інфляції**, який визначають як величину темпу інфляції, зменшену на одиницю.

Для теоретичних досліджень та прогнозування доцільно користуватися функцією інфляції $\gamma(t)$, яка визначається як зміна в часі відносної вартості одиниці умовного товару й обчислюється як добуток послідовних величин темпу інфляції за умови, що вихідне значення (базове) функції дорівнює одиниці. Офіційно коефіцієнт темпу інфляції визначають у відсотках. При відомому коефіцієнті темпу інфляції ν_i, який враховують на завершальному моменті поточного проміжку обліку, функцію інфляції $\gamma(t_k)$ буде представлено так:

$$\gamma(t_k) = \prod_{i=1}^{k}(1+\nu_i) \tag{8}$$

(символом $\prod_{i=1}^{k}$ означає, що необхідно обчислити добуток усіх співмножників від $i=1$ до $i=k$).

Оцінюючи величину очікуваного прибутку, необхідно враховувати не тільки природну певного рівня невизначеність майбутнього, а й інфляційний процес, у різній мірі характерний для всіх країн. Розбудову функції $\gamma(t)$ прогнозованого рівня інфляції на проміжок t виконуємо за даними відповідного динамічного ряду, вилучаючи дослідження соціально-економічних причин*, але не забуваючи про їх визначальний вплив, що до певної міри враховується в експертних оцінках. Вартість суми капіталу визначаємо його купівельною спроможністю. Якщо на даний час вартість одиниці умовного товару була c, то на суму S капіталу можна придбати S/c одиниць умовного товару. За умови, що через проміжок t функція інфляції досягне величини $\gamma(t)$, обсяг товару, який можна придбати на суму S, дорівнюватиме $S/(c\gamma(t))$. За рахунок інфляції купівельна спроможність суми S зменшилася на величину

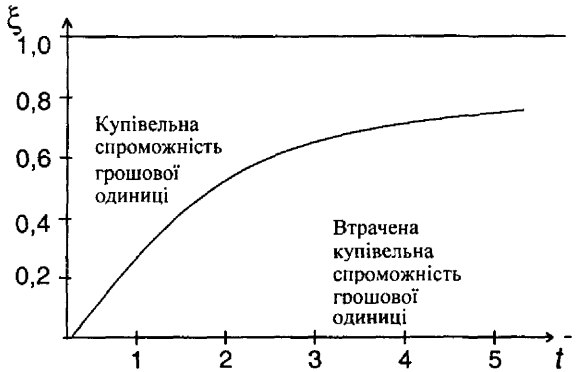
$$\frac{S}{c} - \frac{S}{c\gamma(t)} = \frac{S}{c}\left(\frac{\gamma(t)-1}{\gamma(t)}\right). \tag{9}$$

Коефіцієнт $\xi(t) = \dfrac{\gamma(t)-1}{\gamma(t)}$ визначає втрату купівельної спроможності грошової одиниці в залежності від перебігу інфляції і дорівнює частині обсягу умовного

* Степанюк В. В., Скрипка А. Г. Исследование операций: Учеб. пособие. – К., 1972.

товару, "вимитого" інфляцією. Зміни в часі коефіцієнта $\xi(t)$ втрати купівельної спроможності грошової одиниці за умови лінійного закону перебігу інфляції $\gamma(t) = 1 + \nu t$ наведено на рис. 18.2 а), за умови галопуючої інфляції $\gamma(t) = (1+\nu)^t$ (один із можливих її варіантів) - на рис. 18.2 б) (ν – стала величина).

а)



б)

Рис. 18. 2.

Оцінка втрати купівельної спроможності грошової одиниці за умов гіперінфляції втрачає сенс.

Обчислимо номінальну ставку банківської позики з умовою прагнення банку мати зиск при інфляції. Припустимо, що при відсутності інфляції банк має одержати від позики в сумі S зиск kS, тобто сума, яку треба повернути за термін нарахування

відсотка, дорівнює $S + kS = S(1+k)$. На такі кошти за відсутності інфляції можна придбати $S(1+k)/c$ одиниць умовного товару. За наявності інфляції з коефіцієнтом її темпу v за проміжок нарахування банківського відсотка ціна одиниці умовного товару збільшиться в $(1+v)$ разів за умови тотожності проміжків нарахування банківської ставки та визначення темпу інфляції. Якщо проміжок нарахування банківської ставки кратний проміжку обліку інфляції з коефіцієнтом k, то ціна збільшиться в $(1+\mu\tau v)$ разів при лінійній інфляції або в $(1+v)^{\mu\tau}$ разів при галопуючій інфляції, де τ – поточний номер проміжку обліку інфляції.

Для залучення депозитів банк мусить забезпечити певний прибуток на депозит за термін капіталізації. Позначимо літерою λ коефіцієнт прибутку на одиницю депозиту за відсутності інфляції. За такої умови депозит у сумі S забезпечить приріст капіталу λS, а по завершенні періоду капіталізації загальна сума коштів дорівнюватиме $S(1+\lambda)$. За таку суму при відсутності інфляції можна придбати $S(1+\lambda)/c$ одиниць умовного товару (c – вартість його одиниці). Якщо за період капіталізації коефіцієнт темпу інфляції досяг величини v, то вартість одиниці умовного товару дорівнюватиме $c(1+v)$. Ураховуючи інфляцію, коефіцієнт d банківського депозиту визначимо за умови виконання вимоги, що не зміниться обсяг умовного товару, який можна було придбати по завершенні періоду капіталізації як за відсутності інфляції, так і при її наявності:

$$S(1+\lambda)/c = S(1+d)/c(1+v).$$

Коефіцієнт банківського депозиту d за умови збереження купівельної спроможності депозиту по завершенні періоду капіталізації при наявності інфляції дорівнює:

$$d=\lambda+v+\lambda v \qquad (10)$$

(λ – коефіцієнт банківського депозиту за відсутності інфляції; v – коефіцієнт темпу інфляції за один період капіталізації; величина $(v+\lambda v)$ обумовлює протиінфляційну надбавку).

Проаналізуємо вплив інфляції на розрахунки раніше визначених характеристик інвестиційних проектів. Такий аналіз виконаємо за умови збереження купівельної спроможності коштів грошових потоків при наявності інфляції або при безінфляційній ситуації. За такої умови обсяги інвестицій необхідно збільшувати в $\gamma(t_j)$ разів ($\gamma(t_j)$ – очікувана величина функції інфляції на момент реінвестування), а обсяги очікуваних доходів зменшувати відповідно в $\gamma(t_k)$ разів (t_k – поточні моменти надходження очікуваних прибутків). Якщо, як і раніше, за альтернативний проект вибирати банківський депозит, ураховуючи

протиінфляційний коефіцієнт d (10), то чиста сьогочасна вартість проекту NPV за відомої функції $\gamma(t)$ перебігу інфляції обчислюється за формулою:

$$NPV = -\sum_{j=0}^{n} I_j\,\gamma(t_j)/(1+d)^{t_j} + \sum_{k=1}^{n} C_k\,/\left(\gamma(t_k)(1+d)^k\right) \qquad (11)$$

(t з відповідним індексом – фіксовані номери проміжків термінів реінвестування та надходження очікуваних прибутків по втіленні проекту;
I_j – обсяги інвестування та реінвестування, віднесені до базового моменту;
C_k – обсяги очікуваних прибутків від втілення проекту по завершенні k-го проміжку в цінах базового моменту;
$\gamma(t)$ – прогнозовані значення функції інфляції у наведені моменти часу).

При обчисленні першого доданку у формулі (11) необхідно врахувати обсяги всіх реінвестицій, починаючи з вихідної, співставлених відповідним моментам часу, при обчисленні другого доданку – обсяги очікуваних прибутків.

Доцільно наголосити, що величина d, яку необхідно обчислити згідно з формулою (10), теж залежить від інфляції.

Як і раніше, внутрішня норма прибутковості IRR проекту – норма r дисконтування його грошових потоків, за якої чиста сьогочасна вартість прибутку NPV дорівнює нулю, визначається величиною кореня рівняння:

$$-\sum_{j=0}^{m} I_j\gamma(t_j)/(1+r)^{t_j} + \sum_{k=1}^{n} C_k\,/\left(\gamma(t_k)(1+r)^k\right) = 0. \qquad (12)$$

Зміст використаних у рівнянні (12) символів указаний у запису формули (11).

Приклад. Послідовність грошових потоків за відсутності інфляції наведено в діаграмі (рис. 18.3).

Згідно з рівнянням (7) IRR обчислюється як корінь рівняння:

$$-1000 - \frac{200}{1+r} + \frac{300}{(1+r)^2} + \frac{600}{(1+r)^3} + \frac{700}{(1+r)^4} + \frac{500}{(1+r)^5} = 0. \qquad (13)$$

Обчисливши перший корінь рівняння (13), маємо, що величина IRR дорівнює 0,1772 (17,72%).

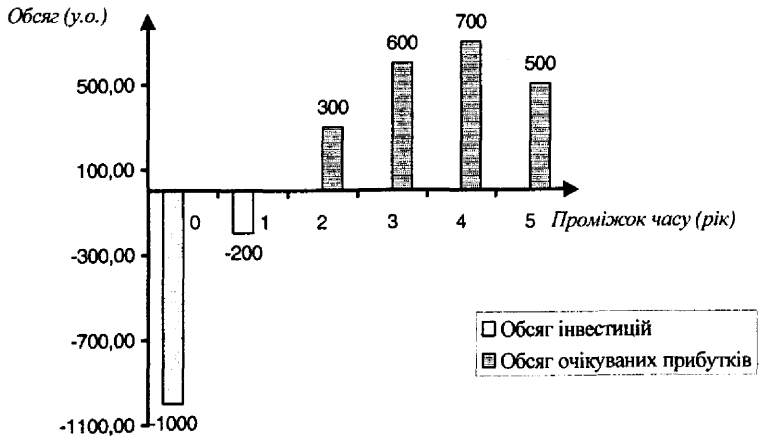


Рис. 18.3.

Тепер обчислимо величину *IRR*, використавши наведений за відсутності інфляції (в базових цінах вихідного інвестування) грошовий потік з припущенням, що в перші три роки прогнозований коефіцієнт щорічного темпу інфляції дорівнює 0,1, а в завершальні два роки – 0,2. Запишемо рівняння (12) за прийнятих припущень:

$$-1000-\frac{200}{1+r}1{,}1+\frac{300}{(1+r)^2}(1{,}1)^2+\frac{600}{(1+r)^3}(1{,}1)^3+ \\ +\frac{700}{(1+r)^4}((1{,}1)^3 1{,}2)+\frac{500}{(1+r)^5}((1{,}1)^3(1{,}2)^2)=0. \qquad (14)$$

Розв'язавши наведене рівняння, маємо *IRR* рівну 0,19.

Чисельно значення *IRR* збільшилося, але одержана величина містить у собі наближену протиінфляційну складову, більшу за 0, 1. При порівнянні альтернативних проектів така протиінфляційна складова рівнозначно впливає на величину *IRR* кожного проекту.

Питання прогнозування перебігу інфляції надзвичайно складне. Труднощі визначення рівня інфляції $\gamma(t)$ обумовлені в основному не розбудовою тренда для відповідного часового ряду, а оцінкою достовірності екстраполяції тих соціально-економічних умов на майбутнє, які породили інфляцію в минулому. Тому поряд з відомими методами статистичних оцінок математичного сподівання відповідної величини

$\gamma(t)$ та інтервалу довіри в залежності від часу при використанні тренда необхідно певним чином враховувати експертні оцінки та використовувати імітаційне моделювання.

Наведені міркування являють собою лише узагальнені вимоги щодо оцінки інвестиційних проектів, і їх необхідно більш детально розробляти при дослідженні більш конкретних проектів.

Варто додати, що на практиці, як правило, грошові потоки можна оцінити з певним рівнем надійності їх величин лише на обмежену кількість років наперед. Такий термін називається *горизонтом оцінки грошових потоків проекту.*

Як приклад можливого уточнення розглянутої моделі оцінки чистої сьогочасної вартості *NPV* проекту розглянемо інвестування поточного виробництва за умови, що номінальна виручка від втілення проекту та обсяги номінальних витрат (сировина, оплата праці) "інфлюють" у нетотожних співвідношеннях; до того ж необхідно враховувати ставку оподаткування. За альтернативний проект прийнято можливий банківський депозит з урахуванням протиінфляційної надбавки. За вищеназваних умов величину *NPV* доцільно обчислити так:

$$NPV-I_0=\frac{(C_1\gamma_1(t_1)-I_1\gamma_2(t_1)-A_1)(1-m)+A_1}{1+d_1}+ \\ +\frac{(C_2\gamma_1(t_2)-I_2\gamma_2(t_2)-A_1)(1-m)+A_2}{(1+d_1)(1+d_2)}+\ldots+ \\ +\frac{(C_n\gamma_1(t_n)-I_n\gamma_2(t_n)-A_n)(1-m)+A_n}{(1+d_1)(1+d_2)\ldots(1+d_n)}. \qquad (15)$$

(C_j $(j=\overline{1,n})$ – очікуваний валовий дохід (виручка) після реалізації продукції за відповідні роки у цінах базового року;

I_j $(j=\overline{1,n})$ – планові валові експлуатаційні витрати за відповідні роки у цінах базового року;

A_j $(j=\overline{1,n})$ – обсяги амортизаційних відрахувань за відповідні роки;

$\gamma_1(t)$ – функція прогнозованого перебігу інфляції на вартість продукції, яка формує доходи від втілення проекту;

$\gamma_2(t)$ – функція прогнозованого перебігу вартості на ресурси, обумовленої поточною експлуатацією;

m – ставка оподаткування прибутку;

d_j $(j=\overline{1,n})$ – коефіцієнти банківського депозиту з урахуванням очікуваного перебігу інфляції).

У наведеному прикладі використання більш детальної інформації стосовно інвестиційного проекту дало можливість розбудувати більш змістовну математичну модель оцінки чистої сьогочасної вартості проекту.

18.4. Оцінка інвестиційного ризику

На значення очікуваних у майбутньому грошових потоків C_k впливають багато різнопланових чинників, тому природно кожну з величин C тлумачити як випадкову за того чи іншого закону розподілу, а плановане значення – як її математичне сподівання (індекси вилучені для спрощення запису). *Розуміючи під ризиком можливість появи несприятливої події в майбутньому*, а саме недосягнення (з певними відхиленнями) очікуваної величини грошового потоку, *ризик R доцільно характеризувати ймовірністю події*, що очікувана величина грошового потоку не буде в певному околі її математичного сподівання:

$$R = 1 - P(C \in |\mu - \delta|) \qquad (16)$$

(δ - величина, вибрана для виміру відхилення можливого значення C грошового потоку від його математичного сподівання μ).

Залежність (16) свідчить, що за рівності очікуваних величин грошових потоків (рівності їх математичних сподівань) ризик вищий у того, для кого величина R більша, тобто меншою є ймовірність події, що фактична величина грошового потоку буде в заданому околі математичного сподівання. Основні труднощі при визначенні ризику обумовлені розбудовою математичних моделей прогнозу перебігу досліджуваного економічного процесу в майбутньому, оскільки він у значній мірі обумовлений соціально-політичними обставинами.

Щоб визначити ймовірність ризику, пов'язану з грошовим потоком, необхідно мати закон його розподілу як певної випадкової величини. Оскільки грошовий потік проекту залежить від багатьох чинників, наприклад, від узгодженості обсягів виробництва та збуту в майбутньому, собівартості та ринкової ціни продукції, податкової політики тощо, то закон його розподілу визначається законами розподілу відповідних чинників та їх взаємозв'язками, що враховується прийняттям певних гіпотез у поєднанні з експертними оцінками. Оскільки при наявності багатьох випадкових чинників розбудова теоретичних розподілів грошових потоків для проектів практичної значимості є надто складним завданням, то доцільно скористатися відповідним емпіричним розподілом при використанні імітаційного моделювання з урахуванням експертних оцінок.

При визначенні доцільності впровадження ризикового проекту коефіцієнт дисконтування має бути більшим, ніж у випадку гарантованих (безризикових) проектів розміщення капіталу, тобто сьогочасна вартість проекту менша. У практиці США "премія" за ризик визначається за експертними оцінками при умовному врахуванні трьох рівнів ризику: мало-, середньо- та високоризикових.

Відсоток "премії" змінюється відповідно від 3 до 9. Названі величини враховуються шляхом збільшення на їх значення коефіцієнта дисконтування.

Таким чином, оцінюючи визначення оптимального інвестування необхідно враховувати динаміку зміни вартості коштів і природну невизначеність величин грошового потоку. При одній з можливих моделей врахування ризику змінюється коефіцієнт дисконтування. Очікувані величини грошових потоків і параметри, використовувані при оцінці інвестиційного проекту, доцільно сприймати як математичні сподівання випадкових величин, з'ясування законів розподілу яких вимагає додаткових досліджень. На практиці при оцінці рівня ризику інвестиційних проектів, особливо при складанні бізнес-планів, широко використовуються експертні оцінки, а для економічно значимих – імітаційне моделювання.

Основні вимоги алгоритму визначення сьогочасної вартості проекту стисло можна сформулювати так:

1) спрогнозувати перебіг інфляції;
2) спрогнозувати обсяги грошових потоків;
3) спрогнозувати рівень ризику впровадження інвестиційного проекту;
4) визначити поетапно величини коефіцієнтів дисконтування на кожний термін горизонту оцінки грошових потоків;
5) обчислити очікувану сьогочасну чисту вартість проекту;
6) визначити норми власної прибутковості проекту;
7) розбудувати закони розподілу параметрів, які визначають сьогочасну чисту вартість проекту, та обчислити її ймовірний інтервал відхилення від математичного сподівання; при розв'язанні цих питань доцільно використовувати експертні оцінки та імітаційне моделювання.

18.5. Одна з інвестиційних моделей при використанні банківського кредиту

Умови господарювання в перехідний період породжують необхідність використання суб'єктами підприємницької діяльності запозичених фінансових ресурсів поряд із власними коштами. За значимістю перше місце серед запозичених ресурсів у більшості випадків посідає *банківський кредит*. Необхідність кредиту як джерела фінансових ресурсів підприємства обумовлена особливостями кругообігу основних та оборотних коштів. Об'єктивні та суб'єктивні чинники ринкової економіки породжують порушення ритму зміни функціональних форм кругообігу коштів, вкладених у засоби виробництва та засоби обігу. Наприклад, підприємство виробило певну продукцію, тобто частина його виробничих засобів перейшла у товарну форму, але ще до реалізації цієї продукції у підприємства з'являється необхідність у закупівлі сировини матеріалів, що вимагає авансування коштів у

новий кругообіг. До цього необхідно додати потреби у додаткових коштах для розширення виробництва або подолання тимчасових труднощів у реалізації продукції і т. ін. Вимоги ефективного господарювання обумовлюють необхідність створення на підприємствах за рахунок власних оборотних коштів мінімальні запаси товарно-матеріальних цінностей. Але потреби в коштах для створення таких запасів коливаються протягом року. Особливо це стосується підприємств з сезонним характером виробництва. Розглянемо розбудову однієї з математичних моделей оцінки ефективності інвестування коштів у виробництво з використанням банківського кредиту.

За відомого прогнозованого перебігу інфляції номінальний коефіцієнт банківського кредиту β буде обчислюватися за такою ж формулою, як (10), але перший її додаток, який відповідає коефіцієнту реального банківського кредиту, завжди буде більший за реальний коефіцієнт банківського депозиту. Різниця між величинами названих коефіцієнтів породжує одне з основних джерел прибутку банку. Таким чином, коефіцієнт банківського кредиту β в умовах інфляції за один термін кредитування (як правило, один рік) обчислюється так:

$$\beta = \rho + \nu + \rho\nu \tag{17}$$

(ρ – коефіцієнт реального банківського кредиту (без наявності інфляції);
ν – коефіцієнт темпу інфляції за один термін кредитування).

Розглянемо одну з можливих моделей організації виробництва з використанням банківського кредиту за відомих прогнозованих функцій інфляції $\gamma(t)$ та коефіцієнту банківської позики. Припустимо, що планується виробництво продукції n найменування в обсязі x_j по кожному з t контрольованих проміжків часу. Для виробництва використовуються m видів ресурсів за відомих в матеріальному вимірі витрат обсягом a_{ij} на одиницю продукції кожного найменування ($j = \overline{1,n}$; $i = \overline{1,m}$) за відомих вартостей υ_i одиниці i-го ресурсу в базових цінах (на час дослідження питання інвестування). При моделюванні режиму оплати ресурсів доцільно розмежовувати вимоги необхідності попередньої оплати та по завершенні певного проміжку обліку часу. *За умови рівномірних витрат ресурсів у часі,* якщо виробничий процес триває t проміжків обліку часу та відомою є функція інфляції $\gamma(t)$, – за виробничий цикл вартість i-го ресурсу при необхідності передоплати обчислюється так:

$$\sum_{\tau=0}^{t-1}\sum_{j=1}^{n}\frac{1}{t}a_{ij}x_j\upsilon_i\gamma(\tau).$$

За умови оплати i-го ресурсу по завершенні одиниці обліку часу його вартість за виробничий цикл буде обчислюватися так:

$$\sum_{\tau=1}^{t}\sum_{j=1}^{n}\frac{1}{t}a_{ij}x_j\upsilon_i\gamma(\tau).$$

Різниця в обчисленні названих сум *обумовлена терміном урахування функції інфляції.* Надалі будемо користуватися другою формою розрахунків вартостей ресурсів. За термін циклу виробництва в t одиниць обліку часу вартість ресурсів U визначається величиною:

$$U = \frac{1}{t}\sum_{i=1}^{m}\sum_{\tau=1}^{t}\sum_{j=1}^{n}a_{ij}x_j\upsilon_i\gamma(\tau). \tag{18}$$

За перебігу індексу i треба враховувати ситуації різновидності оплат за ресурси. Моделі розрахунків за ресурси при відомих схемах їх використання в часі можуть бути різними, основні труднощі обумовлені точністю передбачення перебігу інфляції.

Для розбудови подальшого дослідження умов інвестування необхідно домовитися: стосовно інвестування нового чи поточного виробництва; стосовно співвідношення проміжків часу щодо циклів виробництва та реалізації продукції; стосовно системи оподаткування.

Наведемо використовувані далі позначення та домовимося стосовно термінів:

t – проміжок часу від початку становлення циклу виробництва до одержання продукції;

T – проміжок часу від початку циклу організації виробництва до одержання зиску від реалізації продукції;

a_{ij} – обсяг i-ої складової витрат виробництва на одиницю j-ої продукції;

u_i – вартість одиниці i-ої складової витрат на час розгляду питання інвестування;

$c_j(t)$ – ринкова прогнозована вартість одиниці продукції на момент завершення циклу виробництва;

x_j – обсяг виробництва j-ої продукції;

$c_j^0(t)$ – прогнозовані обмеження ринкової вартості одиниці j-ої продукції;

b_j – виробничі обмеження обсягів j-ої продукції;

d_i – обмеження обсягів i-ої складової витрат (i-го ресурсу);

r_j – обмеження на обсяги виробництва j-ої продукції, обумовлені кон'юнктурою ринку.

Якщо припустити, що в часі реалізація обсягів продукції кожного найменування є рівномірною, то зиск від реалізації при врахуванні перебігу інфляції обчислюється за формулою:

$$\sum_{j=1}^{n}\sum_{\tau=t+1}^{T}\frac{1}{(T-t)}x_j c_j(\tau).$$

Приймемо, що податок нараховується за схемою "по відвантаженню", а його обсяг пропорційний різниці між зиском від реалізації та затратами на виробництво, його ставка вимірюється коефіцієнтом p. За названих припущень, якщо кредит повертати рівними частинами, починаючи з моменту $t+1$, чистий прибуток Z буде обчислюватися згідно з формулою:

$$Z=\left(\frac{1}{T-t}\sum_{j=1}^{n}\sum_{\tau=t+1}^{T}x_j c_j(\tau)-\frac{1}{t}\sum_{i=1}^{m}\sum_{\tau=1}^{t}\sum_{j=1}^{n}a_{ij}x_j u_i\gamma(\tau)\right)(1-p)-\frac{1}{T-t}\sum_{\tau=t+1}^{T}S(1+\beta)^{\tau}. \quad (19)$$

Задачу оптимального використання банківського кредиту для інвестування одного циклу виробництва з прийняттям умов стосовно використання ресурсів, реалізації продукції та системи оподаткування можна сформулювати так: знайти такий обсяг кредиту S та план виробництва $\{x_j\}$ $(j=\overline{1,n})$ і так визначити ціни за певних обмежень на їх величини, щоб функція Z чистого прибутку досягала найбільшого значення при відомому прогнозованому перебігу інфляції $\gamma(t)$ та обмеженнях:

1) $\frac{1}{t}\sum_{i=1}^{m}\sum_{\tau=1}^{t}\sum_{j=1}^{n}a_{ij}x_j u_i\gamma(\tau)\le\frac{1}{t}\sum_{\tau=0}^{t-1}S\gamma(\tau)$ – обмеження суми інвестиції, якщо кредит брати на кожний період закупівлі ресурсів. Якщо кредит одноразовий, то в правій частині необхідно записати величину кредиту S;
2) $\sum_{j=1}^{n}a_{ij}x_j\le d_i$ $(i=\overline{1,m})$ – обмеження обсягів витрат ресурсів;
3) $0\le x_j\le b_j$ $(j=\overline{1,n})$ – обмеження, обумовлені виробничими можливостями;
4) $0\le x_j\le R_j$ $(j=\overline{1,n})$ – обмеження, обумовлені кон'юнктурою ринку за обсягом можливих реалізацій продукції;
5) $0\le c_j(\tau)\le c_j^*(\tau)$ $(j=\overline{1,n})$ – обмеження на ціни продукції, обумовлені кон'юнктурою ринку або умовами оподаткування.

Легко довести, що при нарахуванні податку пропорційно обсягу прибутку ціни доцільно вибирати максимально допустимими. Ситуація зміниться, якщо ставка податку залежить від обсягу прибутку: питання вибору цін за обмежень значень їх величин вимагає додаткових досліджень.

За вибраних припущень стосовно циклу відтворень, умов оподаткування та обліку наявних можливостей виробництва запропоновану математичну модель інвестування приведено до задачі лінійного програмування при відомих цінах на продукцію та до задачі квадратичного програмування, якщо ціни на продукцію підлягають визначенню з урахуванням певних обмежень.

Питання для самоконтролю

1. Що таке інвестиційний проект?
2. З якою метою виконується дисконтування коштів?
3. Зміст понять "сьогочасна вартість проекту", "чиста сьогочасна вартість проекту", "власна прибутковість проекту".
4. Чим відрізняються поняття "інфляція" та "дефляція"?
5. Як визначається на практиці інфляція?
6. Основні види класифікації темпів інфляції та їх характерні відмінності.
7. Як впливає темп інфляції на банківські депозит та кредит?
8. Як впливає темп інфляції на визначення чистої сьогочасної вартості проекту та власної прибутковості?
9. Чим обумовлена наявність ризику при інвестуванні? Як оцінюється рівень ризику? Як він впливає на сьогочасну вартість проекту?

ГЛАВА 19

ЕКСПЕРТНІ ОЦІНКИ В МЕНЕДЖМЕНТІ

19.1. Задача прийняття рішень

Розглядаючи задачі дослідження операцій, ми завжди мали справу з інформацією в кількісному вимірі. Певна математична модель розбудовується за певних гіпотез, використання яких дозволяє узгодити в термінах аналітичних залежностей взаємозв'язки між чинниками, визначальними для досліджуваного явища або процесу. Але дійсність надто складна й часто складно прогнозована, тому необхідно враховувати принципову неможливість формалізувати (розбудувати математичні моделі) все розмаїття практичної діяльності людини, зокрема: вибір професії, вибір супутника життя, прийняття керівних рішень щодо перспективного розвитку певної виробничої структури і т. ін. У цих ситуаціях необхідно враховувати багато якісних чинників при певній невизначеності вихідної інформації, взаємодію чинників, можливі наслідки рішень та багатокритеріальність при оцінці наявних альтернатив. З цих причин вимоги обґрунтування та прийняття рішень не можуть бути задоволені в межах можливостей математичних моделей, розбудовуваних у теорії дослідження операцій, тому використовуються методи та можливості так званого неформального аналізу з урахуванням досвіду, інтуїції та певних суб'єктивних оцінок.

Необхідність прийняття рішень, тобто вибору однієї з можливих альтернатив, є постійним атрибутом життя людини. Тому неформальний аналіз завжди існував і буде існувати, які б досягнення не були в математиці, логіці та обчислювальній техніці. Для кращого розуміння важливості проблем неформального аналізу доцільно наголосити, що будь-які математичні моделі – це наслідок неформального аналізу, виконуваного з метою пізнання та використання явищ і процесів навколишнього світу в корисних цілях. Розбудова відповідних математичних теорій – прагнення знайти засоби підвищення ефективності та достовірності неформального аналізу, але, прагнучи логічної завершеності, ми неминуче звужуємо діапазони використання.

Важливим засобом об'єднання формального та неформального методів аналізу є імітаційне моделювання. Але існують і набагато простіші методи поєднання, зокрема метод експертних оцінок. Термін *експерт* походить від латинського expertus – досвідчений, той, хто має знання та досвід. Під *експертизою* розуміють оцінку групою експертів деяких властивостей та особливостей стану певної системи або процесу. Специфіка такої процедури, як експертиза, в тому, що вимірювальними засобами виступають люди. Причини залучення експертів до оцінок об'єктів або процесів обумовлені багатьма чинниками: відсутністю

відповідних вимірювальних приладів, складністю досліджуваних явищ, відсутністю необхідних обсягів достовірної інформації, або суб'єктивністю досліджуваних характеристик. Класичним прикладом можливих оцінок лише за допомогою експертизи є дегустація вин, парфумів, суддівство спортивних змагань.

Назвемо деякі причини, які обумовлюють всезростаючий інтерес до експертиз при дослідженні соціально-економічних процесів. Для нашого часу характерною є постійно зростаюча складність управління такими процесами, що пояснюється як ускладненням об'єктів управління, так і усвідомленням необхідності враховувати не лише об'єктивні тенденції перебігу процесів, а й наслідки прийняття рішень. До цього варто додати вимоги обов'язковості прийняття рішень при відсутності необхідної інформації про наявний стан і про можливі тенденції розвитку. Експертиза є одним із засобів зменшення ризику прийняти помилкове рішення.

Великі можливості сучасних комп'ютерів в обробці інформації дали якісно нове життя методу експертиз, відомому людині з давніх часів. Розроблено та ефективно використовується чимало раціональних процедур оцінки перебігу складних соціально-економічних процесів з використанням результатів експертиз.

До експертних методів відносять комплекс логічних і математичних процедур, з реалізацією яких прагнуть досягти таких цілей: 1) отримати від висококваліфікованих фахівців (експертів) інформацію про об'єкт дослідження в найбільш можливому обсязі; 2) проаналізувати інформацію, використовуючи певні математичні методи її обробки з метою розбудови рекомендацій для прийняття керівних рішень по управлінню об'єктом дослідження.

Експертні методи використовуються при неможливості або економічній недоцільності використання формалізованих математичних методів. Цариною експертних методів є прогнозування та планування довготермінових соціально-економічних процесів, визначення можливих шляхів розвитку науково-технічного прогресу, оцінки якості товарів та послуг, оцінки якості підготовки фахівців і т. ін. Експертні методи використовуються також в поєднанні з іншими, більш точними та формалізованими методами.

Стисло розглянемо основні засади розбудови експертних методів.

19.2. Експертні оцінки

Розглянемо приклад використання експертних оцінок при обґрунтуванні певного рішення.

Приклад. Маємо чотири проекти можливого розвитку підприємства на п'ять років. Кожний проект за попередніми розрахунками забезпечує певний грошовий прибуток на проміжку в п'ять років. Створено групу з трьох експертів для оцінки достовірності практичного втілення гіпотез, використовуваних у розрахунках

відповідних грошових потоків. Навіть якщо прийняти припущення, що гіпотези стосовно розрахунків грошових потоків безсумнівні, то питання вибору "найкращого" проекту залишається невизначеним, тому що практично неможливою є така ситуація, що якийсь із проектів гарантує щорічний найбільший прибуток на одиницю інвестиційних коштів.

Ураховуючи ситуацію, що проекти стосуються майбутнього, перебіг якого не може бути визначений однозначно, при оцінці проектів необхідно розв'язати такі основні задачі: 1) рівень вірогідності гіпотез, використаних у розрахунках кожного проекту; 2) як порівняти між собою проекти? за якими критеріями необхідно виконувати порівняння, якщо проекти дають різні грошові потоки? 3) як вибрати "найкращий" проект з урахуванням невизначеності майбутнього?

Стисло розглянемо коротко види експертних оцінок.

Кожний зустрічався з кількісними оцінками, які були представлені числами за певної одиниці виміру. Про кількісні переваги оцінок говоримо в тому випадку, коли кількісні оцінки характеризують тотожну якість і доцільно порівнювати за принципом: *на скільки* або *у скільки* разів одна оцінка різниться від іншої. Кількісні оцінки, як правило, відповідають об'єктивним замірам об'єктивних показників.

Оцінки експертів суб'єктивні, вони здійснюються за певною *бальною шкалою*. Всім відомий приклад бальних оцінок – шкільні оцінки. Оцінки за балами поділяються на два види. До першого виду належать бальні оцінки при наявності об'єктивних, більш-менш точно сформульованих критеріїв і певних еталонів визначення величини бала оцінки. Чим точніше визначені межі відхилення від еталону, тим більше довіри до таких оцінок. Наприклад, оцінка диктанту визначається за кількістю морфологічних і синтаксичних помилок. Другим прикладом бальних оцінок першого виду є визначення розряду робітника.

Бальна оцінка другого виду ставиться при відсутності не лише загальноприйнятих еталонів, але й єдиного об'єктивного критерію, суб'єктивним віддзеркаленням якого є оцінка. Порівняння об'єктів виконується за принципом: "більш-менш" або "краще-гірше". У таких випадках розбудовують так звану *порядкову (або рангову) шкалу*. Наприклад, порівняння смаку різних страв, фасонів одягу одного призначення.

Третій вид оцінки – ранжування: упорядкування множини об'єктів відповідно до зменшення їх значимості (допускається можливість рівнозначимості). Прикладом ранжування є визначення призерів будь-якого конкурсу. Ранжування можна тлумачити як оцінювання за ранговою шкалою: рангом об'єкта, тобто оцінкою, можна вважати номер місця, яке займає цей об'єкт в нумерації.

У процедурі пошуку об'єкта, найкращого за певним критерієм, користуються таким видом оцінки, як *попарне порівняння:* визначають перевагу одного об'єкта над іншим у кожній парі (визначають також ситуацію їх рівноцінності або неможливість порівняння). Метод попарного порівняння використовується тому, що

якісне порівняння об'єктів, як правило, легше виконати, ніж оцінити відношення в бальній або ранговій шкалах.

Для аналізу кількісними методами якісної експертної інформації використовують спеціальні так звані **вербально-числові шкали.** Поширеною є вербально-числова шкала Харінгтона (від лат. verbalis – усний, словесний), тобто шкала відповідностей словесних оцінок числовим (табл.19.1).

Таблиця 19.1.

Градація інтенсивності якості	Числові інтервали
Надвисока	1,0–0,8
Висока	0,8–0,63
Середня	0,63–0,37
Низька	0,37–0,2
Надзвичайно низька	0,2–0

Вище вказано лише основні типи експертних оцінок, що свідчить про розмаїття та складнощі оцінювання при виконанні експертиз.

19.3. Оцінювання об'єктів при проведенні експертиз

Розглянемо можливості бальних оцінок, які використовують в залежності від ситуації як кількісні, так і якісні показники при відповідних методах обробки інформації. Якщо наявні бальні оцінки проаналізовано як кількісно, так і якісно, то за умови тотожності висновків з великою мірою надійності можна стверджувати, що результати є об'єктивними.

При кількісному підході бальні оцінки x_{ij}, які виставляє i-му об'єкту j-ий експерт можуть бути проаналізовані по-різному. Якщо припустити, що експерти мають однакову компетентність, користуються *найпростішою* груповою оцінкою x_i:

$$x_i = \frac{1}{n}\sum_j x_{ij} \qquad (j = \overline{1, n}) \tag{1}$$

(n – кількість експертів).

Якщо бажано або необхідно врахувати компетентність, об'єктивність, інформованість експертів, то вводяться так звані вагові коефіцієнти компетентності q_j для кожного j-го експерта. Тоді групова оцінка x_i i-го об'єкта обчислюється за формулою:

$$x_i = \sum_j q_j x_{ij} \qquad (j = \overline{1, n}). \tag{2}$$

Часто досліджуваний показник являється надто складним для безпосередньої оцінки. Наприклад, показник якості виробу. За такої умови в показнику виділяються більш прості показники (як складові досліджуваного), кожний з яких стосується певної характеристики досліджуваного показника. Наприклад, споживацька якість товару є об'єднуючою для таких показників, як естетичні, соціальні (потреба у товарі), функціональні та ергономічні (зручність в експлуатації). Іншим прикладом може бути порівняння сортів сільськогосподарської культури одразу за декількома ознаками: величина врожаю, термін визрівання, вміст білка, кількість вітамінів, умови зберігання і т. ін.

У такому випадку оцінка визначається показником з трьох індексів: x_{ijk} - оцінка i-го об'єкта j-им експертом за k-ою ознакою (показником). Використовується ще одна величина: Z_{kj} – оцінка значимості k-ої ознаки j-им експертом. Тоді приймають, що оцінка значимості k-ої ознаки Z_k обчислюється так:

$$Z_k = \sum_j q_j Z_{kj} \qquad (j = \overline{1, n}), \tag{3}$$

оцінка x_i i-го об'єкта:

$$x_i = \sum_j \sum_k Z_k q_j x_{ijk} \qquad ((j = \overline{1, n});\ k = (\overline{1, m})) \tag{4}$$

(m – число досліджуваних ознак).

Залежність (4) можна записати в іншому вигляді:

$$x_i = \sum_j q_j \sum_k Z_k x_{ijk} = \sum_j q_j x_{ij}. \tag{5}$$

Практика виконання експертиз свідчить, що середні оцінки, обчислювані за наведеними формулами, є досить стабільними величинами при зміні складу експертних груп у випадках, якщо існує досить чітко окреслена й добре апробована система оцінок.

При формуванні й аналізі експертної інформації широко використовується такий спосіб оцінки, як **ранжування.** Під ранжуванням розуміють таку послідовність упорядкування оцінок об'єктів, за якої міра досліджуваної якості змінюється монотонно з урахуванням рівноцінності деяких об'єктів. Мірою якості

може бути бальна оцінка, яка відбиває суб'єктивну оцінку експерта. Ранжування доцільно тлумачити як оцінювання за ранговою шкалою (від франц. rang – ряд). Експерт надає кожному досліджуваному об'єкту номер натурального ряду 1; 2; ..., n в залежності від результатів співставлення досліджуваної ознаки в різних об'єктів; міра ознаки визначається експертом суб'єктивно з урахуванням досвіду, знань і припущень. Якщо кільком об'єктам надано тотожні номери (тобто такі об'єкти мають рівну міру якості за оцінкою експерта), після ранжування необхідно виконати **стандартизацію рангів.** Якщо s об'єктів за ранговою шкалою одержали номери $n_1, n_2, \ldots, n_s$, то таким об'єктам надається стандартизований ранг $\overline{n}_s$, рівний середньому арифметичному номерів, які вони одержали:

$$\overline{n}_s = \frac{n_1 + n_2 + \ldots n_s}{s}. \tag{6}$$

Ця формула має сенс за умови, що кількість рангів дорівнює кількості об'єктів.

Приклад. Маємо десять об'єктів (a, b, c, d, e, f, g, h, i, j). Для зручності надамо їм послідовно номери від 1 до 10. Експерт у послідовності *зменшення* досліджуваної якості надав об'єктам ранги, наведені у третьому стовпці табл. 19.2. Об'єктам за номерами 4, 9, 10 тотожний ранг 6, а об'єктам 7 і 8 - 7, тому для названих об'єктів необхідно виконати стандартизацію рангів.

Таблиця 19.2.

Об'єкти	Номери об'єктів	Ранги об'єктів	Розташування об'єктів згідно з рангом	Номери об'єктів згідно з рангом	Стандартизовані ранги об'єктів
a	1	4	b	1	1
b	2	1	e	2	2
c	3	5	f	3	3
d	4	6	a	4	4
e	5	2	c	5	5
f	6	3	d	6	7
g	7	7	i	7	7
h	8	7	j	8	7
i	9	6	g	9	9,5
j	10	6	h	10	9,5

Для об'єктів d, i, j стандартизований ранг дорівнює

$$\bar{n}_s = \frac{6+7+8}{3} = 7.$$

Для об'єктів *g, h* при врахуванні місць їх розташування за ранговою шкалою стандартизований ранг дорівнює

$$\bar{n}_s = \frac{9+10}{2} = 9{,}5.$$

Послідовність розташування об'єктів за ранговою шкалою наведено в останньому стовпці табл. 19. 2.

Якщо ранжування виконується кількома експертами, то процедурою спочатку передбачається розрахунок суми стандартизованих рангів, пропонованих групою експертів для кожного з досліджуваних об'єктів, за формулою:

$$R_i = \sum_j r_{ij} \qquad (j = \overline{1, n}) \tag{7}$$

(n – кількість експертів; r_{ij} – стандартизований ранг i-го об'єкта, визначений j-им експертом).

По завершенні розрахунків виконується ранжування об'єктів.

До задач ранжування об'єктів належать всі задачі, пов'язані з відносною оцінкою будь-яких об'єктів або явищ (таких, наприклад, як оцінка популярності митця чи громадсько-політичного діяча і т. ін.). Задачі ранжування дуже важливі в конкретних соціальних дослідженнях, і їх вірний розв'язок може мати важливі практичні соціально-економічні наслідки. Наприклад, розв'язується проблема підвищення ефективності роботи певного виробничого колективу. По завершенні відповідних досліджень визначено певні чинники, які дадуть можливість досягти бажаного результату. Запроваджувати відповідні заходи необхідно в послідовності їх значимості. Тобто перш ніж почати їх реалізацію, необхідно їх ранжувати за критерієм ефективності.

Метод безпосередньої оцінки. Приймається, що проміжок можливих змін досліджуваного чинника доцільно розділити на кілька несполучних частин, кожна з яких співставляється з певним балом із наперед вибраного діапазону, наприклад, від 0 до 10 або від 0 до 20 і т. ін. Знання навчального матеріалу в школі оцінюється за 12-бальною шкалою.

Експерт мусить кожному досліджуваному параметру дати певний бал в межах вибраної шкали та у відповідності до його думки стосовно значимості параметра. Якщо p_{ij} є оцінкою i-го параметра j-им експертом, то середня оцінка P_i такого параметра з урахуванням відповідних оцінок усіх експертів обчислюється так:

$$P_i = \frac{1}{m}\sum_j p_{ij} \qquad (j = \overline{1, n}). \tag{8}$$

Коефіцієнт значимості (ваги) K_i i-го параметра обчислюють за формулою:

$$K_i = \frac{P_i}{\sum_i P_i} = \frac{\sum_j p_{ij}}{\sum_i \sum_j p_{ij}} \qquad (i = \overline{1, m};\ j = \overline{1, n}). \tag{9}$$

Легко переконатися, що сума коефіцієнтів значимості усіх параметрів дослідження дорівнює 1.

Метод попарних порівнянь (переваг). Стисло розглянемо два варіанти цього методу.

а) *Метод частинних попарних порівнянь.* Ранжування або оцінювання з використанням балів, особливо при значній кількості показників (чинників), можна істотно полегшити, якщо виконувати попарне порівняння значимості показників з метою визначення показника більш важливого в кожній парі .

Для реалізації такої процедури будується матриця (таблиця) методу попарних порівнянь, в якій порівнювані параметри записуються у верхньому рядку та у крайньому лівому стовпці (табл. 19.3)

Таблиця 19.3.

Показники		...	v_i
⋮			
μ_k			

Заповнюються лише клітинки, що розташовані *справа* від діагоналі, тобто ($i; k$) за умови $i < k$. У таку клітинку експерт записує або i, або k – номер того показника, якому він надає перевагу при обов'язковому виконанні умови однозначного визначення переваги. При формуванні матриці кожним j-им експертом у крайній правий стовпець записують частоту переваг $v_i(j)$ i-го показника в i-му рядку, а в нижній рядок таблиці частоту $\mu_k(j)$ переваг k-го показника в k-ому стовпці над рештою показників. Потім для кожного i-го показника обчислюється величина

$$m_i(j) = \mu_i(j) + v_i(j), \tag{10}$$

тобто частота переваг i-го показника, визначена j-им експертом. Середня для усіх експертів частота переваг i-го показника обчислюється за формулою:

$$m_i = \frac{1}{m}\sum_j m_i(j) \quad (j = \overline{1, n}) \tag{11}$$

(n – кількість експертів).

Загальна кількість співставлень N, виконаних кожним експертом при одноразовому частинному попарному порівнянні, дорівнює:

$$N = \frac{1}{2}n(n-1).$$

Коефіцієнт значимості K_i показника (параметра) обчислюється за формулою:

$$K_i = \frac{m_i}{N} = \frac{2m_i}{n(n-1)}. \tag{12}$$

Приклад. Результати частинного попарного порівняння п'яти параметрів чотирма експертами наведено відповідно в табл. 19.4 - 19.7.

Таблиця 19.4.

	1	2	3	4	5	v_i
1	–	1	1	4	1	3
2	–	–	3	4	2	1
3	–	–	–	3	5	1
4	–	–	–	–	5	0
5	–	–	–	–	–	0
μ_k	0	0	1	2	2	×

Таблиця 19.5.

	1	2	3	4	5	v_i
1	–	1	3	1	5	2
2	–	–	2	4	2	2
3	–	–	–	3	5	1
4	–	–	–	–	4	1
5	–	–	–	–	–	0
μ_k	0	0	1	1	2	×

Таблиця 19.6.

	1	2	3	4	5	v_i
1	–	2	1	4	1	2
2	–	–	3	4	2	1
3	–	–	–	3	3	2
4	–	–	–	–	5	0
5	–	–	–	–	–	0
μ_k	0	1	1	2	1	×

Таблиця 19 7.

	1	2	3	4	5	v_i
1	–	1	1	1	5	3
2	–	–	3	4	2	1
3	–	–	–	4	3	1
4	–	–	–	–	5	0
5	–	–	–	–	–	0
μ_k	0	0	1	2	2	×

Величини $m_i(j)$, обчислені згідно з (10), наведено в табл. 19.8.

Таблиця 19.8.

Показники	$m_i(j)$				m_i
	$j = 1$	$j = 2$	$j = 3$	$j = 4$	
1	3	2	2	3	10/5
2	1	2	2	1	6/5
3	2	2	3	2	9/5
4	2	2	2	2	8/5
5	2	2	1	2	7/5

Коефіцієнти значимості K_i, згідно з формулою (12) мають такі величини:

$$K_1 = \frac{2\cdot 10}{5\cdot 5\cdot 4} = \frac{1}{5}; \quad K_2 = \frac{2\cdot 6}{5\cdot 5\cdot 4} = \frac{3}{25}; \quad K_3 = \frac{2\cdot 9}{5\cdot 5\cdot 4} = \frac{9}{50};$$

$$K_4 = \frac{2\cdot 8}{5\cdot 5\cdot 4} = \frac{4}{25}; \quad K_5 = \frac{2\cdot 7}{5\cdot 5\cdot 4} = \frac{7}{50}.$$

б) *Метод повного попарного порівняння* відрізняється від методу частинних парних порівнянь лише тим, що кожна пара показників порівнюється не один раз, а двічі, тобто експерти мусять заповнювати всі клітинки матриці (таблиці). Ця вимога обумовлена вимогою по можливості нейтралізувати можливу помилку експертів, яка з'являється тому, що можна без достатнього обґрунтування віддати перевагу показнику, розташованому на першому місці. Чим менша об'єктивно існуюча різниця між двома порівнюваними показниками, тим більша ймовірність названої похибки.

Коефіцієнт вагомості i-го показника обчислюється за формулою

$$K_i = \frac{m_i}{n(n-1)} \tag{13}$$

(m_i – середня для всіх експертів частота переваги i-го показника, обчислювана за формулою (11); n – кількість експертів).

19.4. Аналіз узгодженості експертних оцінок

Якщо міркування експертів з одного й того ж питання різні, то чи варто брати їх до уваги та користуватися в подальшому? Ця проблема має дві складові: як розуміти узгодженість оцінок та оцінити таку узгодженість? як організувати

експертизу, щоб забезпечити об'єктивність і найкращу узгодженість оцінок?

Перше питання стосується математичних методів обробки інформації. Необхідність його ефективного розв'язання сприяла розбудові одного з методів обробки інформації в багатовимірному просторі – так званого **кластерного аналізу** (від англ. classter – гроно, щільна сукупність). Цей метод досить складний, може бути практично реалізований лише з використанням комп'ютерів, тому не будемо його розглядати*. Наголосимо лише, що для розбудови процедури узгодженості треба вибрати оцінку міри узгодженості. Особливості втілення процедури узгодженості розглянемо на умовних прикладах. Припустимо, що деякий об'єкт за m різними показниками (ознаками) оцінюється (n) експертами, кожний з яких указав у характеристиці об'єкта міру значимості досліджуваних показників . Вихідний матеріал доцільно представити матрицею A:

$$A = \begin{pmatrix} \alpha_{11} \dots \alpha_{1n} \\ \alpha_{21} \dots \alpha_{2n} \\ \dots\dots\dots\dots \\ \alpha_{m1} \dots \alpha_{mn} \end{pmatrix} \qquad (14)$$

(α_{ij} $(i = \overline{1,m};\ j = \overline{1,n})$ – міра i-го показника, вказана j-им експертом).

Розглядаючи елементи кожного стовпця матриці A як компоненти вектора $\overline{\alpha}_j$ $(j = \overline{1;n})$, за міру узгодженості оцінок, наданих двома експертами, умовно позначеними k та l, можна взяти також величину $d_1(\overline{\alpha}_k;\ \overline{\alpha}_l)$, обчислювану за формулою:

$$d_1(\overline{\alpha}_k;\ \overline{\alpha}_l) = \sqrt{\sum_i (\alpha_{ik} - \alpha_{il})^2}, \quad (i = \overline{1,m}) \qquad (15)$$

і величину d_2, обчислювану за формулою:

$$d_2(\overline{\alpha}_k, \overline{\alpha}_l) = \sum_i |\alpha_{ik} - \alpha_{il}| \quad (i = \overline{1,m}). \qquad (16)$$

У кластерному аналізі використовуються й інші оцінки міри узгодженості. Використовуючи і (15), і (16), одержимо квадратну матрицю попарної узгодженості оцінок експертів. Потім необхідно вибрати оцінку якості експертизи в цілому. У літературі описано декілька різних методів оцінки узгодженості результатів різними експертами.

Вибравши методику оцінки, необхідно виконати ряд організаційних заходів з

* Царенко О. М., Злобін Ю. А., Скляр В. Г., Панченко С. М. Комп'ютерні методи в сільському господарстві та біології. – Суми, 2000.

метою забезпечення достатнього рівня надійності результатів експертизи. Такі заходи обумовлені спеціальними вимогами до експертів і процедури експертиз. Звичайно, що відповідні вимоги та їх втілення залежать від змісту та призначення конкретних експертиз, але можна дати перелік досить загальних міркувань, які мають бути реалізовані.

19.5 Підготовка та проведення експертиз

Доцільність використання експертних оцінок та їх значення для прийняття керівних рішень у великій мірі обумовлені не лише метою експертизи, а й програмою її реалізації. Для процедури втілення експертизи обов'язковими є такі етапи: 1) визначення мети експертизи; 2) розробка анкет опитування експертів; 3) формування групи експертів; 4) формування групи фахівців-аналітиків з питань цілей експертизи й аналізу відповідної інформації; 5) опитування експертів; 6) обробка й аналіз наслідків опитування; 7) розробка рекомендацій щодо прийняття керівних рішень.

Група фахівців-аналітиків визначає найбільш доцільний, виходячи з мети експертизи, спосіб опитування, формує склад груп експертів, організує їх опитування, аналізує результати опитування та готує на їх підставі рекомендації щодо прийняття керівних рішень. Збирання інформації від експертів виконується за допомогою анкет із переліком питань, обумовлених метою експертизи, та зразком відповіді на них.

Очевидно, що якість оцінок, які дають експерти, визначається фахом і кількістю експертів. Тому дуже важливо якнайкраще сформувати групи експертів. Доцільно, щоб у таку групу входили фахівці зі споріднених питань, бажано працюючі в різних організаціях.

Визначення компетентності експерта. Безсумнівно, компетентність є дуже важливою рисою експерта, що враховується у відповідних формулах за допомогою так званого вагового коефіцієнта. Розроблено спеціальні методи визначення таких коефіцієнтів, але відповідний математичний апарат досить складний і специфічний, тому обмежимося стислим викладом питання.

Для визначення рівня компетентності окремого експерта використовуються різні методики оцінки: евристичні (оцінки, висловлювані фахівцями), статистичні (опитування експертів за спеціальними тестами) та ін.

До евристичних оцінок належать оцінки одержувані шляхом як *самооцінки*, так і *взаємооцінки*. У першому випадку кожний експерт заповнює спеціальну анкету самооцінки; у другому – кожний експерт за спеціальною анкетою оцінює своїх колег. Оцінки визначаються в бальній або вербально-числовій шкалах. Після аналізу таких анкет одержують усереднену (групову оцінку) кожного експерта. Однією із

суттєвих проблем при такій процедурі оцінювання є тотожність розуміння експертами градації шкал.

Різновидом самооцінки є так званий диференційний метод, при якому, як правило, оцінка надається за двома групами критеріїв: за критеріями, які віддзеркалюють рівень обізнаності експерта з основними джерелами інформації стосовно змісту та мети експертизи, та за критеріями, які характеризують обізнаність експерта з об'єктами експертизи. Спеціальні дослідження показали, що самооцінки як засіб визначення компетентності експертів досить об'єктивні в багатьох важливих випадках.

При статистичному методі оцінки фаху експерта приймають середні групові оцінки. Найпростіше втілення цього методу передбачає, що кожний експерт із певної групи експертів дає перелік фахівців, яких він вважає компетентними в досліджуваному питанні. Коефіцієнт компетентності експерта визначається як відношення кількості реєстрів, в яких названо даного експерта, до загальної їх кількості. Розбудовано й більш складні методи оцінки фаху експертів.

Ефективним способом оцінки фаху експерта є тестовий метод, який широко використовується при можливості неупереджено оцінити результати тестування. Тестові методи орієнтовані на визначення рівня фаху експерта, навичок і досвіду, необхідних для цільової експертизи. При тестових випробуваннях необхідно враховувати такі аспекти: 1) має бути розбудований з урахуванням конкретного змісту та мети експертизи; 2) має існувати шкала, яка б забезпечувала точність оцінок фаху експерта; 3) мають бути розбудовані так, щоб зробити максимально неможливими випадкові вгадування при відповідях на запитання.

Природно, що визначення рівня фаху експерта істотно спрощується, якщо він брав участь у подібних експертизах.

Стосовно числа експертів будь-які загальновизнані рекомендації відсутні, є кілька точок зору. Доцільно виказати загальні зауваження. Чим менша кількість експертів, тим нижчі точність оцінок та ймовірність якісної експертизи. Але при значній кількості експертів зменшується вплив високих оцінок на групову оцінку, якщо вони надані незначним числом експертів. Визначення оптимальної кількості експертів є складним завданням.

Вибравши певну методику оцінки об'єкта експертизи, необхідні певні організаційні заходи, щодо забезпечення максимальної достовірності результатів експертизи. Відповідні заходи обумовлені як змістом і метою експертизи, так і лімітом часу та можливостями виконання відповідного аналізу результатів експертизи.

У сучасній літературі широко пропагується так званий *метод Дельфи* як один із найбільш відомих методів формування неупереджених групових оцінок експертів при проведенні експертиз. Метод названо на честь міста Дельфи, де, за легендою, у Стародавній Греції мешкав відомий оракул.

Метод Дельфи

Для методу характерні три особливості: анонімність, регульований зворотний зв'язок, групові оцінки. Анонімність досягається завдяки використанню спеціально розбудованої системи питань або інших способів особистого опитування. Колективні обговорення скасовані. Регульований зворотний зв'язок досягається виконанням кількох турів експертизи. Результати кожного туру аналізуються за певною методикою та повідомляються особисто кожному експерту. Результатом аналізу особистих оцінок є групові оцінки.

Метод Дельфи має реєстр рекомендацій і вимог щодо проведення експертиз. Деякі з таких вимог очевидні: 1) експертами можуть бути лише крупні фахівці у відповідній галузі; 2) оскільки кожний експерт може припускатися помилок, то лише міркування досить значного числа експертів можуть визначити задовільно досліджуване питання; 3) питання до експертів мають бути відносно простими та чітко сформульованими, щоб зробити неможливою неоднозначність їх тлумачення; 4) для проведення експертиз необхідно мати у складі відповідних керівних органів стабільні групи організаторів експертиз, бо лише добре організована і якісно проведена експертиза може гарантувати важливі рекомендації для прийняття керівних рішень.

Але вимоги методу Дельфи суттєво значимі та до певної міри незвичайні, наприклад, заборона загальних дискусій експертів. Автори методу вважають, що таким чином вони практично вилучають вплив психологічних та емоційних чинників, які завжди виникають під час загальних дискусій.

Організатори експертизи завчасно складають реєстр запитань особисто для кожного експерта. Експерт мусить висловити свою точку зору на кожне запитання, користуючись пропонованою *числовою шкалою.*

У рекомендаціях до організації експертизи за методом Дельфи велике значення мають процедури одержання й аналізу експертами додаткової інформації, в тому числі обґрунтованої інформації щодо думок інших експертів (але така інформація анонімна). Автори методу особливо наголошують на значимості оцінок узгодженості та пропонують певну процедуру визначення міри узгодженості.

Згідно з процедурою організації методу опитування експертів виконується в чотири тури. Після першого туру запитань відповіді експертів аналізуються, і кожному з них повідомляють величину середньої групової оцінки (нагадуємо, що відповіді на запитання необхідно давати за певною числовою шкалою) по кожному запитанню та максимальні відхилення від неї.

На другому турі експертам, оцінки яких значно відхилилися від середньої, пропонують обґрунтувати свою точку зору. Потім з аргументами (без імені автора) знайомлять усіх експертів. Ознайомившись з результатами першого туру й

одержавши додаткову інформацію стосовно оцінок по кожному запитанню, експерти знову дають відповіді на запитання. Одержані відповіді знову аналізуються та порівнюються за відповідною шкалою і т. д.

За групову оцінку кожного запитання приймають усереднену оцінку відповідей по завершенні четвертого туру запитань-відповідей. Якщо середня групова оцінка певного запитання досягається раніше, то відповідне запитання в наступному турі вилучається.

Доцільно наголосити, що наявність зворотного зв'язку, анонімність, аргументація значних відхилень при їх появі в середніх групових оцінках, ітерацій сприяють тому, що надійність та ефективність оцінок за методом Дельфи значно підвищуються. Метод досить широко використовується при дослідженнях, які передують прийняттю керівних рішень у багатьох галузях соціально-економічної, політичної, військової діяльності.

У літературі описано також інші методи організації та проведення експертиз, наприклад, метод тестової діагностики Ю.І.Журавльова, метод представлення складних проблем, які досліджуються при експертизі, сукупністю більш простих, запропонований Г.С.Поспєловим*.

Питання для самоконтролю

1. Чому при прийнятті керівних рішень необхідно за певних умов звертатися до експертиз?
2. Основна мета проведення експертиз.
3. З якою метою розбудована вербально-числова шкала?
4. Розбудова числових оцінок при проведенні експертиз.
5. Як виконується процедура ранжування експертних оцінок?
6. Основні засади розбудови аналізу узгодженості експертних оцінок.
7. Як визначається компетентність експерта?
8. Призначення та основні засади розбудови методу Дельфи.

* Литвок Б. Г. Экспертные оценки и принятие решений. – М., 1996.

ГЛАВА 20

ФІНАНСОВИЙ СТАН ПІДПРИЄМСТВА; ДІАГНОСТИКА БАНКРУТСТВА

20.1. Необхідність і значимість оцінки фінансового стану підприємства

Всі підприємства сфери матеріального виробництва за умов переходу до ринкової економіки мають працювати на засадах комерційних розрахунків, що обумовлює істотне розширення їх прав у сфері фінансово-економічної діяльності. Комерційний розрахунок вимагає чіткої організації фінансової діяльності підприємства, оскільки суб'єкти господарювання, маючи високій рівень фінансової незалежності, несуть реальну економічну відповідальність за ефективність свого функціонування, своєчасність виконання зобов'язань перед іншими діловими партнерами та державою. Про здатність підприємства успішно діяти в умовах комерційного розрахунку свідчить його фінансовий стан.

Фінансовий стан – це відображення результату взаємодії всіх елементів системи фінансових відносин підприємства, які визначаються сукупністю виробничо-господарських чинників і системою показників, обумовлених наявністю, розміщенням і використанням фінансових ресурсів.

Фінансова діяльність підприємства має бути спрямована на забезпечення систематичного надходження та ефективного використання фінансових ресурсів, дотримання розрахункової та кредитної дисципліни, досягнення оптимального співвідношення власних і залучених коштів з метою ефективного господарювання. Саме цим обумовлюється необхідність і практична значимість оцінки фінансового стану підприємства, що відіграє істотну роль у забезпеченні його стабільності. Отже, фінансовий стан – це одна з найважливіших характеристик діяльності будь-якого підприємства. Нормативні вимоги до характеристик фінансового стану підприємств України наведено у відповідних державних документах.

Фінансовий стан підприємства оцінюється з метою пошуку резервів підвищення рентабельності виробництва та зміцнення комерційного розрахунку як основи стабільної, ефективної роботи та виконання зобов'язань перед бюджетом, банками й іншими суб'єктами фінансово-господарської діяльності.

20.2. Фінансовий аналіз

Фінансовий аналіз (від гр. розклад, розчленування) – це метод оцінювання та прогнозування фінансового стану підприємства. У якісному фінансовому аналізі

підприємства зацікавлене широке коло суб'єктів ринку:

– керівники підприємства, які мусять знати неупереджену думку про свою діяльність і розробляти заходи щодо її поліпшення;

– інвестори, яким потрібна інформація про ризиковість або ж ефективність інвестицій;

– кредитори та постачальники, які бажають впевнитися у платоспроможності підприємства;

– партнери по бізнесу, які прагнуть встановити з підприємством стабільні ділові відносини (постачальники, споживачі, страхові компанії, транспортувальники та ін.);

– сторонні щодо підприємства структури (державна податкова адміністрація, профспілки, благодійні організації).

Мета фінансового аналізу – інформаційне забезпечення прийняття керівних рішень, на які істотно впливають фактичні та прогнозовані дані про фінансовий стан підприємства. Наслідки фінансового аналізу являють собою незначну кількість ключових, найінформативніших показників, які об'єктивно, всебічно характеризують фінансовий стан підприємства: платоспроможність, фінансову стійкість і незалежність, рентабельність діяльності, ймовірність банкрутства тощо.

Інформаційною базою для оцінювання фінансового стану підприємства є такі дані: баланс (форма №1), звіт про фінансові результати (форма №2), звіт про рух грошових коштів (форма №3), звіт про власний капітал (форма №4), дані статистичної й оперативної звітності.

Інформація, яка використовується при аналізі фінансового стану підприємства, за доступністю поділяється на відкриту та закриту (таємну). Інформація, яка міститься в бухгалтерській та статистичній звітностях, виходить за межі підприємства, отже, є відкритою. Загальні вимоги до фінансової звітності викладено в положенні (стандарті) бухгалтерського обліку 1 (П(С)БО–1), затвердженому наказом Міністерства фінансів України від 31.03.99 р. за № 87. Основою П(С)БО–1 є міжнародний стандарт бухгалтерського обліку комітету з міжнародних стандартів бухгалтерського обліку.

Кожне підприємство розробляє свої планові та прогнозовані показники, норми, нормативи, тарифи та ліміти, систему їх оцінки та регулювання фінансової діяльності. Така інформація є комерційною таємницею, а іноді й "ноу-хау". Відповідно до чинного законодавства України підприємству надано право тримати таку інформацію в таємниці. Перелік її визначає керівник підприємства.

Оскільки для зовнішнього фінансового аналізу (П(С)БО–1:4) використовуються агреговані показники відповідної діяльності підприємства, то це обмежує можливості виявити усі її особливості та причини успіхів або невдач. Тому проводиться внутрішній фінансовий аналіз, призначений для керівників підприємства, з метою виявити чинники, які впливають на рівень прибутковості,

рентабельності, собівартості виробництва за видами продукції та витрат, визначити критичні обсяги виробництва, зробити фінансовий аналіз ефективності інвестицій тощо.

20.3. Основні показники фінансового стану підприємства

Стабільний фінансовий стан підприємства обумовлюється всією його виробничо-господарською діяльністю. Тому не можливо об'єктивно та достовірно оцінити його, використовуючи тільки якийсь один показник. До того ж показники оцінки мають бути такими, щоб усі, хто пов'язаний з підприємством економічними стосунками, могли одержати відповідь на питання, наскільки надійне підприємство як партнер у фінансовому відношенні, отже, прийняти рішення про доцільність продовження або встановлення фінансово-економічних відносин з підприємством. У кожного партнера (акціонерів, банків, податкової адміністрації і т. ін.) свої критерії зацікавленості у співпраці. Отже, і показники оцінки фінансового стану мають бути такими, щоб кожний партнер міг зробити вибір за власними інтересами.

Зрозуміло, що в якісному фінансовому стані має бути зацікавленим перш за все саме підприємство. Але відповідний стан формується в процесі взаємовідносин підприємства з постачальниками, покупцями, акціонерами, банками й іншими юридичними та фізичними особами. З іншого боку, від якості роботи самого підприємства залежить міра його економічної привабливості для всіх суб'єктів, які завжди мають можливість вибору між багатьма іншими підприємствами, здатними задовольнити економічний інтерес. Тому необхідно систематично, ґрунтовно та в динаміці аналізувати фінансовий стан підприємства, оскільки від його якості залежить економічна перспектива. У залежності від особливостей фінансово-господарської діяльності підприємства та мети аналізу користуються кількома десятками показників. Спеціалісти пропонують 65 показників, з яких вибирається необхідний реєстр. Коротко охарактеризуємо основні показники, використовувані при оцінці фінансово-господарської діяльності підприємств.

20.3.1. Показники оцінки майнового стану

1. *Сума господарських коштів, які є у розпорядженні підприємства.* Цей показник дає загальну оцінку вартості активів, які перебувають на балансі підприємства. Зростання цього показника свідчить про збільшення майнового потенціалу підприємства.

2. *Питома вага активної частини основних засобів.* Згідно з нормативними документами, активну частину основних засобів складають машини,

обладнання, транспорт. Зростання цього показника в динаміці є позитивною тенденцією.

3. *Коефіцієнт зносу основних засобів* характеризує частку зносу основних засобів у загальній їх вартості. Використовується в аналізі для характеристики стану основних засобів. Доповненням до цього показника є так званий *коефіцієнт придатності.*

4. *Коефіцієнт оновлення основних засобів* показує, яку частину основних засобів, наявних на кінець звітного періоду, складають нові основні засоби.

5. *Коефіцієнт вилучення основних засобів* показує, яка частина основних засобів, з якими підприємство почало діяльність у звітному періоді, вилучена через знос та з інших причин.

20.3.2. Показники оцінки фінансової стійкості

Однією з найважливіших вимог до фінансового стану підприємства є стабільність його діяльності в майбутньому. Вона пов'язана із загальною фінансовою структурою підприємства, його залежністю від кредиторів та інвесторів.

1. *Коефіцієнт концентрації власного капіталу* характеризує частку власності самого підприємства в загальній сумі коштів, інвестованих у його діяльність. Чим вищий цей коефіцієнт, тим більш фінансово стійким і незалежним від кредиторів є підприємство. Доповненням до цього показника є *коефіцієнт концентрації залученого (позикового) капіталу.* Сума обох коефіцієнтів дорівнює 1 (100%).

2. *Коефіцієнт фінансової залежності* є оберненим до попереднього показника. Коли його значення наближається до 1 (100%), це означає, що власники повністю фінансують своє підприємство.

3. *Коефіцієнт маневреності власного капіталу* показує, яка частина власного капіталу використовується для фінансування поточної діяльності, тобто яку частину вкладено в оборотні кошти, а яку капіталізовано.

4. *Коефіцієнт довгострокових вкладень* показує, яку частину основних коштів та інших позаоборотних активів профінансовано зовнішніми інвесторами, тобто яка частина належить їм, а не власникам підприємства.

5. *Коефіцієнт довгострокового залучення позикових коштів* характеризує структуру капіталу. Зростання цього показника є негативною тенденцією, яка означає, що підприємство починає все сильніше залежати від зовнішніх інвесторів.

6. *Коефіцієнт співвідношення позикових і власних коштів.* Зростання цього показника в динаміці також свідчить про посилення залежності підприємства від кредиторів, тобто про зниження його фінансової стійкості.

Інші не менш важливі показники оцінки фінансово-господарської діяльності підприємства буде розглянуто далі.

20.4. Аналіз фінансової стійкості підприємства

20.4.1 Поняття фінансової стійкості

Фінансова стійкість підприємства визначається рівнем його фінансової незалежності у володінні власним майном і його використанні. Показники фінансової стійкості визначаються нормативними вимогами до характеристики фінансового стану підприємств в Україні. Сукупність абсолютних і відносних показників фінансової стійкості доцільно об'єднати за призначенням у такі групи:

– можливості покриття обігових коштів (запасів) надійними джерелами фінансування;

– платоспроможність підприємства (потенційна спроможність покрити термінові зобов'язання мобільними активами);

– частка власних або стабільних джерел у сукупних джерелах фінансування.

Стисло розглянемо основні засади розрахунків за вказаними критеріями.

20.4.2. Фінансова стійкість стосовно джерел покриття обігових коштів

За названим критерієм фінансова стійкість визначається в такий спосіб. Вартість запасів (*З*) послідовно порівнюється з обсягами таких джерел фінансування: власні обігові кошти (*ВОК*); власні обігові кошти в поєднанні з довготерміновими кредитами та позиками ($ВОК + К_Д$); власні обігові кошти, довго- та короткотермінові кредити разом з позиками ($ВОК + К_Д + К_К$). Обсяги власних обігових коштів визначаються різницею між власними капіталами (*ВК*) та вартістю необоротних активів (*НА*), на покриття якої першочергово використовується власний капітал:

$$ВОК = ВК - НА.$$

Величина *ВОК* характеризує фінансову стійкість підприємства. Умова її наявності – додатне значення цієї величини.

Розрізняють чотири рівні фінансової стійкості*.

* Шеремет А. Д., Сайфулин Р. С. Финансы предприятий. – М, 1998.

1. *Абсолютна стійкість.* Власних обігових коштів вистачає для забезпечення запасів (*З*):

$З < ВОК$, платоспроможність гарантовано.

2. *Достатня стійкість.* Для забезпечення запасів необхідно до власних обігових коштів залучити довготермінові кредити та позики:

$З < ВОК + К_Д$, платоспроможність гарантовано.

3. *Нестійкий фінансовий стан.* Для забезпечення запасів необхідно до власних обігових коштів залучити довготермінові кредити та позики разом з короткотерміновими кредитами та позиками:

$З < ВОК + К_Д + К_К$, платоспроможність на межі критичного стану, але ще є наявні можливості поновити її.

4. *Кризовий фінансовий стан.* Вичерпано всі джерела фінансування, але їх обсягів недостатньо для забезпечення запасів:

$З > ВОК + К_Д + К_К$, підприємству загрожує банкрутство.

При визначенні рівня фінансової стійкості підприємства складають, використовуючи дані відповідного поточного балансу, агрегований баланс, позиції активу і пасиву якого за ступенем агрегації відповідають меті аналізу. Таким чином, рівень забезпеченості запасів (матеріальних обігових коштів) є сутністю фінансової стійкості, її реальним проявом, що обумовлює платоспроможність суб'єкта фінансово-господарської діяльності.

20.4.3. Показники платоспроможності (ліквідності)

Оцінивши рівень фінансової стійкості підприємства, доцільно мати сукупність показників, за допомогою яких можна визначити його кількісну міру. Основними з таких показників є коефіцієнти поточної, швидкої та абсолютної ліквідності. У наведених далі формулах знаменник один і той же – обсяги поточних пасивів, тобто обсяги термінових зобов'язань.

1. Коефіцієнти $К_{л.п.}$ поточної ліквідності:

$$К_{л.п.} = \frac{Поточні\ активи}{Поточні\ пасиви} = \frac{ПА}{ПП}.$$

Величина $К_{л.п.}$ показує, скільки одиниць обігових коштів припадає на кожну одиницю короткотермінових, тобто невідкладних, зобов'язань. Критичне значення $К_{л.п.}$ дорівнює 1. Якщо $1 < К_{л.п.} < 1{,}5$, то це свідчить про те, що підприємство вчасно ліквідує свої борги.

2. Коефіцієнти $К_{л.ш.}$ швидкої ліквідності:

$$К_{л.ш.} = \frac{Поточні\ активи - Запаси}{Поточні\ пасиви} = \frac{ПА - З}{ПП}.$$

3. Коефіцієнти $К_{л.а.}$ абсолютної ліквідності:

$$К_{л.а.} = \frac{Грошові\ кошти}{Поточні\ пасиви} = \frac{К}{ПП}.$$

За вимогами методики*, значення коефіцієнта абсолютної ліквідності має бути в межах [0,2; 0,35]. Цей коефіцієнт показує здатність підприємства негайно ліквідувати короткотермінову заборгованість.

20.5. Банкрутство: фінансові та правові питання

20.5.1. Поняття банкрутства. Фінансовий стан підприємства та банкрутство

Фінансово-економічний стан суб'єкта економічної діяльності, який робить неможливим продовження його прибуткової фінансової діяльності, визначається як банкрутство. Банкрут (юридична або фізична особа) у своїй господарській діяльності пов'язаний з іншими її суб'єктами фінансово, економічно, соціально. Такий системний характер банкрутства обумовлює необхідність розглядати його з різних точок зору, що породжує безліч визначень цього поняття та різноплановість його наслідків. Розглянемо фінансово-економічний аспект банкрутства та деякі його юридичні особливості.

На сучасному етапі розвитку суспільства за умов розбудови економіки на

* Методика проведення поглибленого фінансово-господарського стану неплатоспроможних підприємств та організацій: Затвердж. наказом Агентства з питань запобігання банкрутству від 27 червня 1997р. // Держ. інформ. бюл. про приватизацію. – 1997. - № 12.

ринкових засадах з різних причин виникають ситуації, коли деякі суб'єкти господарювання не можуть продовжувати беззбиткову фінансово-економічну діяльність, не здатні виконати взяті на себе зобов'язання, як правило, грошового характеру. Як свідчить багатовікова світова практика, банкрутство – одна з невід'ємних ознак ринкових відносин, обумовлена самою суттю підприємництва. Використання фінансово-економічних та юридичних заходів проти не здатних беззбитково господарювати суб'єктів з метою їх вилучення з економічної діяльності є інструментом перерозподілу капіталу та засобів виробництва, відбиває об'єктивні процеси структурної перебудови економіки в напрямі її ефективного функціонування.

Процедури, обумовлені банкрутством, як перерозподіл капіталу на користь більш ефективно господарюючих власників на макро- і мікроекономічному рівнях впроваджуються в Україні спільно з новими економічними відносинами. За часів СРСР в моделі так званої соціалістичної економіки (насправді командно-адміністративної) не передбачалася необхідність такого заходу. Внаслідок цього законодавча база з питань банкрутства, не кажучи вже про практичний досвід, в нашій країні тільки розбудовується. Законодавчі та підзаконні акти, які стосуються банкрутства, постійно удосконалюються. Одночасно уточнюється та доповнюється саме поняття стану банкрутства.

Згідно із Законом України "Про відновлення платоспроможності боржника або визнання його банкрутом" від 14 травня 1992 р., банкрутство розглядається як визнана арбітражним судом нездатність боржника поновити свою платіжну спроможність та задовольнити визнані судом вимоги кредиторів *не інакше як через застосування ліквідаційної процедури.*

В умовах командно-адміністративної системи управління економікою банкрутство не визнавалось як особливий стан господарської діяльності. Вважалося, що явище характерне лише для країн з приватним капіталом. Стверджувалося, що суспільна форма власності робить неможливою не лише будь-яку форму експлуатації, а й банкрутство юридичної або фізичної особи. Але рівень ефективності господарювання за часів СРСР та руйнація суспільно-політичної системи без будь-яких зовнішніх розворушень засвідчує, що дійсність була зовсім іншою. Значна частина підприємств навіть за системи радянського ціноутворення, коли ціни визначалися в основному директивно, були збитковими, вимагали додаткових коштів для продовження виробничого процесу. Банкрутували не лише сільськогосподарські, а й промислові підприємства, передусім вугільної промисловості, яка щорічно дотувалася на сотні мільйонів карбованців. Але в офіційних джерелах такі підприємства характеризувалися не як банкрути, а як "малорентабельні", "збиткові", "нерентабельні". При відсутності відповідних законів справи про банкрутство не порушувалися, бо за умов радянської моделі господарювання пов'язані з появою банкрутства заходи щодо більш ефективного

перерозподілу капіталу між суб'єктами господарської діяльності не могли мати попиту взагалі.

У Росії до сумнозвісних подій жовтня 1917 р. більше століття існувало та широко використовувалося законодавство про неплатоспроможність. У більшості пострадянських країн однією з характерних рис економічної ментальності суспільства є майже всеохоплююче негативне сприйняття процедур банкрутства суб'єктів підприємницької діяльності. Психологічно процедура банкрутства підприємства, а тим більше фізичної особи, сприймається як трагедія, остаточне припинення господарських справ, загибель всіх надій. Таке сприйняття банкрутства обумовлене нерозумінням того, що не завжди воно супроводжується повною ліквідацією бізнесу, а, навпаки, частіше означає перерозподіл економічних ресурсів від недієздатних до більш ефективно працюючих власників. *Тому процедуру заходів, викликаних банкрутством, доцільно розглядати як одну із форм структурування та реорганізації бізнесу.*

В умовах ринкової економіки банкрутство підприємств є природним явищем. З кожних 100 підприємств на ринку залишається в середньому біля 20. У США, наприклад, у 1997 р. загальна кількість підприємств, оголошених банкрутами, становила близько 30000. У ФРН, починаючи з 1990 р., постійно збільшувалася кількість підприємств-банкрутів:

Показники	1990 р.	1991 р.	1992 р.	1993 р.	1994 р.	1995 р.	1996 р.
Кількість банкрутів, од.	8730	8837	10920	15148	18837	22344	25530
Питома вага у загальній кількості, %	–	0,34	0,41	0,58	0,68	0,8	0,92

Протягом кількох сторіч інститут банкрутства компаній з подальшою ліквідацією або реструктуризацією їх активів є невід'ємним та ефективним засобом ринкового регулювання у переважній більшості країн з ринковою економікою. Цей інститут водночас виконує кілька суттєво важливих для суспільства функцій. Першою та найголовнішою його функцією є правовий захист майнових прав суб'єктів ринку. Наявність ефективного дієздатного механізму повернення (при необхідності) боргів значно зменшує ризик господарської діяльності, підвищує відповідальність та стимулює додаткову активність суб'єктів ринку і за своєю суттю є однією з головних ознак приватної власності.

У макроекономічному значенні інститут банкрутства та ліквідації забезпечує ефективність функціонування ринкового середовища, відтинаючи недієздатних суб'єктів. Кожне підприємство, яке опинилось у складній фінансовій ситуації і не

має змоги вчасно розрахуватися за своїми зобов'язаннями, зменшує ефективність економічної діяльності своїх ділових партнерів.

В особливих умовах ведення господарської діяльності в пострадянських країнах цей негативний вплив фінансово не спроможних підприємств на більш ефективно діючі викликав і поширив так звану кризу неплатежів, коли внаслідок суцільної дії такої кризи механізм банкрутства втратив свій оздоровлюючий економічний зміст, оскільки занадто багато суб'єктів опинилися в стані банкрутства через тісні фінансово-економічні взаємозв'язки і практично неможливо відрізнити потенційно економічно спроможних від неспроможних.

У стабільних соціально-економічних умовах та конкретній ситуації на мікрорівні заходи, обумовлені наявністю банкрутства, доцільно розуміти та використовувати*:

- як механізм призупинення діяльності підприємств, економічно-фінансовий стан яких об'єктивно не дозволяє їм продовжувати функціонування як господарських суб'єктів, що не мають потенціалу на майбутнє;
- як пошук шляхів розв'язання фінансових проблем управління, можливих альтернатив розвитку при інших власниках;
- як юридичний механізм (можливо, що для підвищення ефективності діяльності банкрутуючого підприємства слід позбутися тягаря попередніх боргів і нездібних керівників);
- як форму приватизації державного майна, якщо фінансові проблеми виникли у державного підприємства, планованого для передачі у приватну власність.

20.5.2. Діагностика банкрутства

Математичні моделі дослідження зв'язків між показниками фінансового стану з урахуванням їх змін у часі дали змогу одержати рівняння регресії для прогнозування ймовірності банкрутства. У країнах з розвинутою ринковою економікою широко використовується методика діагностики банкрутства, запропонована у 1968 р. відомим західним економістом Е. Альтманом. За двофакторною моделлю ймовірність банкрутства (індекс Z) визначається коефіцієнтами покриття K_n та автономії (фінансової незалежності) $K_{АВТ}$ за формулою:

$$Z = -0{,}3877 - 1{,}0736K_n + 0{,}0579\,K_{АВТ}.$$

* Ізмайлова К. В. Фінансовий аналіз: Навч. посібник / МАУП. - К., 2000.

Якщо Z=0, то ймовірність банкрутства становить 50%. Від'ємні значення Z свідчать про зменшення ймовірності банкрутства, а додатні – про те, що ймовірність перевищує 50% і зростає зі збільшенням Z. Модель приваблива своєю простотою, але має невелику точність прогнозу.

Для більш точних розрахунків імовірності банкрутства в розвинутих країнах використовується п'ятифакторна модель обчислення індексу Z:

$$Z = 3{,}3K_1 + 1{,}0K_2 + 0{,}6K_3 + 1{,}4K_4 + 1{,}2K_5$$

$$\left(K_1 = \frac{\textit{Прибуток до сплати процентів і податків}}{\textit{Активи}};\right.$$

$$K_2 = \frac{\textit{Виручка від реалізації}}{\textit{Активи}};$$

$$K_3 = \frac{\textit{Ринкова вартість власного капіталу}}{\textit{Залучений капітал}};$$

$$K_4 = \frac{\textit{Реінвестований прибуток}}{\textit{Активи}};$$

$$\left.K_5 = \frac{\textit{Власні обігові кошти}}{\textit{Активи}}\right).$$

Критичне значення $Z = 2{,}675$. Якщо для конкретного підприємства $Z < 2{,}675$, то підприємству в найближчі два–три роки загрожує банкрутство. Умова $Z > 2{,}675$ свідчить про сталий фінансовий стан підприємства. Описану методику доцільно використовувати при можливості оцінити ринкову вартість власного капіталу, що можна зробити за розвинутих ринкових відносин, та й то лише для значних корпорацій, акції яких мають обіг на фондових біржах.

При відсутності розвинутого фондового ринку в Україні оцінку фінансового стану підприємства можна виконати, застосовуючи критерій, сформульований Е. Альтманом у 1983 р. Величина Z^* обчислюється за формулою:

$$Z^* = 3{,}107\,K_1 + 0{,}995\,K_2 + 0{,}42\,K_3^* + 0{,}847\,K_4 + 0{,}717\,K_5$$

$$\left(K_3^* = \frac{\textit{Балансова вартість власного капіталу}}{\textit{Залучений капітал}}\right).$$

Критичне значення $Z^* = 1{,}23$.

У літературі описано також інші методики тестування на ймовірність банкрутства.

Питання діагностики банкрутства та підтримки платоспроможності дуже важливе в Україні, про що свідчить тривала криза неплатежів та її негативні наслідки. Відповідно до Закону України "Про відновлення платоспроможності боржника або визнання його банкрутом" від 14.05.1995 р. № 2343-XII розроблена і діє методика проведення поглибленого аналізу фінансово-господарського стану неплатоспроможних підприємств та організацій, затверджена наказом Агентства з питань запобігання банкрутства підприємств та організацій 27 червня 1997 р.

Джерелами інформації при здійсненні аналізу є форми фінансової звітності, розрахунки нормативу власних обігових коштів, розшифровка дебіторської та кредиторської заборгованостей, бізнес-плани, матеріали маркетингових досліджень та інша інформація.

Результати аналізу заносять у зведену форму "Основні фінансово-економічні показники та коефіцієнти". Після цього спеціальна комісія робить висновки про доцільність віднесення підприємства до Реєстру неплатоспроможних підприємств та організацій. Щоб уникнути вказаного Реєстру, доцільно заздалегідь здійснювати самоаналіз фінансово-господарської діяльності "власного" підприємства.

Наведемо основні параметри оцінки фінансового стану за вказаною методикою. В умовах ринкової економіки найбільше значення мають показники величини співвідношень частини активів, які можуть бути реалізовані в певний строк, і суми зобов'язань, які за цей час необхідно сплатити. Найважливішими є такі показники.

Коефіцієнт покриття поточних пасивів $K_п$ визначає співвідношення між загальною сумою ліквідних активів і сумою поточних зобов'язань. Цей коефіцієнт обчислюється за формулою:

$$K_п = \frac{S - S_{нл}}{П_п}$$

(S – загальна сума коштів, вкладених в обігові активи; $S_{нл}$ – сума обігових коштів, вкладених в активи із сумнівною ліквідністю або в неліквідні активи; $П_п$ – обсяги поточних зобов'язань (пасиви)).

Коефіцієнт показує, чи спроможне підприємство за рахунок наявних коштів (залучення боргів на свою користь (депозитів), реалізації запасів та інших активів) виконати вимоги кредиторів (постачальників, бюджету, банків тощо), якщо строки повернення кредитів настають у близькому майбутньому, розрахуватися з робітниками та службовцями за нарахованою заробітною платою, сплатити за всіма іншими фінансовими короткотерміновими зобов'язаннями. Вважається, що

поточний фінансовий стан можна визнати задовільним, якщо значення коефіцієнта $K_п$ близьке до 2, тобто якщо сума ліквідних обігових активів приблизно у два рази перевищує обсяги поточних зобов'язань.

Другий важливий коефіцієнт, який використовується для визначення фінансового стану підприємства, – співвідношення між так званими "чистими" ліквідними активами, тобто такими, які можна негайно перетворити у гроші (це товари, поточна заборгованість покупців, цінні папери, залишки на розрахунковому та інших банківських рахунках, касова готівка), та поточними фінансовими зобов'язаннями. Цей **коефіцієнт негайного покриття поточних зобов'язань** $K_{нп}$ обчислюється за формулою:

$$K_{нп} = \frac{S_л}{П_п}$$

($S_л$ – обсяг обігових коштів, вкладених у "чисті" ліквідні активи; $П_п$ – обсяги поточних зобов'язань (пасиви)).

Нормальною з точки зору ліквідності та фінансового стану підприємства вважається ситуація, якщо величина коефіцієнта $K_{нп}$ близька до одиниці.

Статистика свідчить, що підприємства у своїй фінансово-господарській діяльності в той чи інший спосіб використовують позичені ресурси. Але залучення позикових коштів має певні межі в залежності від обсягу власних фінансових ресурсів. Питома вага позичених фінансових ресурсів підприємства в загальній сумі його фінансових ресурсів визначається **коефіцієнтом фінансової напруги** $K_{фн}$, який обчислюється за формулою:

$$K_{фн} = \frac{ПЗ}{ПЗ + ВР}$$

($ПЗ$ – обсяги позичених фінансових ресурсів (пасиви); $ВР$ – власні фінансові ресурси підприємства).

Якщо цей коефіцієнт становить 0,5 і більше, то вважається, що підприємство знаходиться у складному фінансовому стані, а кредитори мають обґрунтовані підстави вважати інвестування такого підприємства ризикованим.

Низька ліквідність підприємств спричиняється збитковою діяльністю або низьким рівнем рентабельності, нераціональним розміщенням фінансових ресурсів – усім, що може призвести до *банкрутства*.

Рішення про банкрутство підприємств приймає, згідно з чинним законодавством України, арбітражний суд. Заяву до арбітражного суду про

банкрутство боржника подають кредитори, але і сам боржник має право на таку заяву.

Після порушення справи про банкрутство треті юридичні особи (державні та недержавні) мають право подати свої пропозиції щодо задоволення вимог кредиторів до боржника в певні строки та на певних умовах.

Визнання підприємства банкрутом тягне за собою *негайне* припинення його функціонування як юридичної особи, а всі права щодо розпоряджання майном банкрута, його майнові права та обов'язки переходять до ліквідаційної комісії, яка розпоряджається майном банкрута та виконує інші заходи щодо задоволення вимог кредиторів.

Протягом 1996 р. в арбітражних судах України було розглянуто 6552 справи про банкрутство, з них 1691 справа завершилася прийняттям постанови про визнання боржника банкрутом. У 1997 р. ці показники становили відповідно 9645 та 4107; у 1998 р. – 12281 та 4525, у 1999 р. – 12618 та 6244.

Банкрутство підприємства та його ліквідація означають не лише збитки для акціонерів, кредиторів, виробничих партнерів, споживачів продукції, а й зменшення податкових надходжень у бюджет і підвищення безробіття, що теж є одним із чинників впливу на стан макроекономічної стабільності. Суттєвим є те, що серед підприємств, справи яких розглядаються судами, значний відсоток таких, які тимчасово потрапили в скрутне фінансове становище. Вартість активів таких підприємств буває набагато вищою за їх кредиторську заборгованість. За умови якісно проведеної санації (від лат. sanare – оздоровлення, видужання) або реструктуризації такі підприємства зможуть розрахуватися з боргами та більш ефективно продовжити діяльність.

Санація: соціально-економічні аспекти. Провідні зарубіжні економісти вважають, що санація – це система фінансово-економічних, організаційно-правових, виробничо-технічних і соціальних заходів, спрямованих на досягнення або поновлення платоспроможності, ліквідності, прибутковості та конкурентоспроможності підприємства-боржника на довготерміновий період. Тобто санація – це сукупність усіх різнопланових можливих заходів для фінансового оздоровлення підприємства.

Природно, що визначальне місце в санації належить заходам фінансово-економічного спрямування, пов'язаним з фінансовими відносинами, які виникають у процесі мобілізації та використання як внутрішніх, так і зовнішніх фінансових джерел. Метою фінансової санації є покриття поточних збитків, поновлення ліквідності та платоспроможності, поліпшення структури обігового капіталу та формування фондів фінансових ресурсів, необхідних для втілення санаційних заходів виробничо-технічного характеру.

При проведенні фінансової санації підприємств враховуються три основних функціональних блоки: 1) розробка плану санації; 2) санаційний аудит; 3) менеджмент санації.

Досить часто підприємство доходить до банкрутства через невміле або навмисно неправильне управління. Тому важливе значення для успішного фінансового оздоровлення має організація ефективного менеджменту санації. У Великобританії, наприклад, щорічно з 60 тис. порушених справ про банкрутство близько 90% завершуються заміною керівництва. Керуючих санацією, консультантів, аудиторів та інших осіб для проведення санації призначають відповідно до рішення арбітражного суду.

Важливим елементом санаційних заходів є аналіз виробничо-господарської діяльності підприємства щодо узгодженості фактичного стану з даними, відображеними в санаційній концепції:

- загальна виробнича структура підприємства;
- оцінка рівня існуючої технології виробництва;
- аналіз виробничих затрат;
- аналіз руху основних фондів та амортизаційних відрахувань;
- різнопланові оцінки ефективності організації праці (динаміка чисельності працівників і рівня заробітної плати, плинність кадрів, втрати робочого часу, маркетингові дослідження тощо).

Участь кредиторів у санації боржника може здійснюватися в таких напрямках:

- продовження та реструктуризація наявної заборгованості;
- повна або часткова відмова від своїх вимог;
- надання додаткових кредитних ресурсів;
- надання кредитного забезпечення (поручительство, гарантії);
- трансформація боргу у власність.

Безперечно, активної та конструктивної участі кредиторів у санації підприємства доцільно очікувати лише за умови, що внаслідок збереження підприємства - боржника більш повно будуть задоволені їх вимоги, ніж при його ліквідації.

Якщо фінансових ресурсів, мобілізованих із децентралізованих джерел, не вистачає для успішного втілення заходів, то за певних умов може бути прийнято рішення про надання державної фінансової підтримки, якщо держава визнає продукцію відповідних підприємств суспільно необхідною. *Існують критерії відбору підприємств для надання цільової комплексної державної підтримки.* Підтримку зорієнтовано передусім на підприємства, здатні використати її з максимальною віддачею та забезпечити збільшення виробництва продукції, що може позитивно вплинути на прибуткову частину бюджету. Серед підприємств - потенційних претендентів на державну підтримку доцільно назвати такі, які займаються:

- розробкою та використанням нових, ефективних ресурсозберігаючих та екологічно безпечних технологій;

- експортом конкурентоспроможної продукції;
- проблемами заміни імпортної продукції, сировини та матеріалів вітчизняними;
- проблемами енергозабезпечення та енергозбереження;
- проблемами збереження науково-технічного потенціалу (дослідження та розробки, які мають пріоритетне значення для країни).

Як санатор держава керується передусім народногосподарською доцільністю: протидія спаду виробництва та зростанню безробіття.

Одним із головних чинників обмеженого доступу фінансово не спроможних підприємств до кредитних ресурсів є цілковитий брак або недостатній обсяг кредитного забезпечення. Вихід на фінансовий ринок у такому разі можливий за одержання санаційної підтримки у вигляді державних гарантій або зобов'язань здійснити (при необхідності) сплату боргів підприємства. Така форма фінансової підтримки підприємств в Україні дуже широко використовувалася протягом 1993–1999 рр. Але за браком ефективних механізмів відбору кандидатів для надання гарантій, через невикористання інституту санаційного аудиту у переважної більшості підприємств-боржників не виявилося коштів для погашення кредитів, гарантованих урядом, внаслідок чого кредити доводиться погашати за рахунок коштів державного бюджету.

Використання певного методу державної підтримки санації та реструктуризації обумовлено народногосподарською та регіональною значимістю конкретних підприємств. На завершення доцільно додати, що навмисне доведення до банкрутства, якщо це завдало істотної шкоди державним або громадським інтересам, правам та інтересам кредиторів, карається штрафом з позбавленням права обіймати певні посади або здійснювати певну діяльність на строк до п'яти років. Названі дії, якщо вони завдали великої матеріальної шкоди, караються позбавленням волі на строк до п'яти років з конфіскацією майна.

В умовах формування нових соціально-економічних відносин в Україні питання діагностики банкрутства та платоспроможності є надзвичайно актуальним, про що свідчить тривала криза неплатежів та її негативні наслідки.

Питання для самоконтролю

1. Зміст поняття "фінансовий стан підприємства".
2. Інформаційна база оцінки фінансового стану підприємства.
3. Основні показники фінансового стану підприємства.
4. Чинники, які впливають на фінансову стійкість підприємства.
5. Показники платоспроможності підприємства.
6. Що таке банкрутство підприємства?
7. При використанні яких показників і в яких їх межах досліджується фінансовий стан підприємства з визначенням наявності банкрутства?
8. Що таке санація?
9. Роль прогнозування фінансового стану в управлінні підприємством.
10. Які існують директивні документи стосовно контролю за фінансовим станом підприємства?
11. Юридична відповідальність у сфері фінансової діяльності.

ГЛАВА 21

ВИКОРИСТАННЯ ПАКЕТІВ ПРИКЛАДНИХ ПРОГРАМ (ППП) У ДОСЛІДЖЕННІ ОПЕРАЦІЙ

Інтенсивний розвиток обчислювальної техніки обумовив її проникнення в усі сфери суспільного життя. Сучасні ЕОМ будь-якого класу мають різнобічне і якісне математичне забезпечення. Якщо спеціалісту певної галузі необхідно розв'язати або регулярно розв'язувати задачі будь-якого типу, то доцільно користуватися наявними, уже розробленими програмами реалізації певних алгоритмів. Сукупність програм розв'язання задач певного класу називається пакетом програм.

У повсякденному житті, бізнесі, наукових дослідженнях ми часто стикаємося з подіями та явищами, наслідки яких неможливо визначити заздалегідь. Наприклад, бізнесмен не може знати сьогодні, який буде завтра чи через місяць курс долара, банкір – чи повернуть йому позику. У побуті людина, спираючись на здоровий глузд, досвід, інтуїцію, часто приймає рішення в ситуаціях, на завершення яких можуть вплинути випадкові фактори. У серйозному бізнесі рішення приймаються на основі ретельного аналізу наявної інформації, оскільки вони повинні бути обґрунтованими та доказовими.

Для розв'язання задач, пов'язаних з аналізом даних, математики, біологи, інженери, економісти й інші фахівці розробили могутній та гнучкий арсенал методів, які так і називаються – аналіз даних. Ці методи дозволяють виявити закономірності на фоні випадковостей, зробити обґрунтовані висновки та прогнози, оцінити ймовірності їх здійснення. Поєднання математичних методів і змістовного економічного аналізу відкриває нові можливості для економічної науки.

Пакети прикладних програм зробили методи аналізу даних більш доступними та наочними: всю чорнову роботу з виконання складних розрахунків, побудови таблиць і графіків взяв на себе комп'ютер, а досліднику залишилася в основному творча праця: постановка задачі, вибір методу її розв'язання, інтерпретація результатів. Зараз спеціалісту не обов'язково володіти професійними навичками програміста для розв'язання задач певного типу, знати декілька мов програмування. Цілком досить підібрати вже готову програму, яка реалізує певний алгоритм. Для тих, хто не вміє програмувати, створені ППП.

21.1. Загальна характеристика ППП

ППП – це сукупність взаємопов'язаних прикладних програм для виконання типових розрахунків у досить формалізованих галузях знань. Хороший пакет дозволяє зробити всі необхідні дії (або принаймні їх значну частину) з дослідження

проблеми, аналізу даних, моделювання, документування й оформлення результатів, не виходячи з пакета. Це дає можливість не занурюватися у складання математичних моделей, у вибір методів статистичного аналізу, у труднощі програмування, а зосередитись на самій проблемі. Але це не означає, що користувач не повинен знати основних понять математики, статистики, володіти англійською мовою (бо не всі пакети русифіковані) та основами комп'ютерної грамотності.

ППП повинні задовольняти загальним вимогам, якими є:

– простота освоєння та використання;

– відповідність високим вимогам до введення, перетворення та зберігання даних;

– широкий набір засобів графічного подання даних і результатів їх обробки;

– можливість включення таблиць і графіків у звіт;

– широкий набір методів аналізу даних;

– докладність документації (з урахуванням інтересів як початківця-користувача, так і спеціаліста).

ППП, використовувані для аналізу даних, можна умовно поділити на математичні (Maple, Mathcad та ін.) та статистичні (StatGraphics, Stadia, Statistica та ін). Внаслідок великої популярності та необхідності використання методів аналізу даних відповідні засоби почали включатися в табличні процесори (Quattro Pro, Lotus, Excel та ін.), а також у деякі бази даних. Найчастіше у таких пакетах зустрічаються засоби описової статистики, методи регресійного аналізу, засоби аналізу часових рядів, згладжування та прогнозування.

21.2. Засоби пакетів статистичних програм для обробки та аналізу даних

Статистична обробка даних є основою створення математичної моделі розв'язуваної задачі.

Статистичні пакети можна поділити на спеціалізовані пакети та пакети загального призначення.

Спеціалізовані пакети містять методи, використовувані в конкретній предметній галузі. Це досить повний набір традиційних методів, який інколи доповнюється оригінальними методами й алгоритмами, створеними розробниками пакета. Як правило, такі пакети розраховані на спеціалістів, добре знайомих з цими методами, і застосовуються у випадках, коли необхідно систематично розв'язувати задачі з цієї галузі, а можливостей пакета загального призначення невистачає.

У пакетах загального призначення відсутня пряма орієнтація на специфічну предметну галузь, але вони мають широкий діапазон методів аналізу даних. До цієї групи пакетів належать системи Statistica, StatGraphics, Stadia, SPSS.

Перші версії статистичної діалогової системи **Stadia** з'явилися на ринку програмної продукції в 1989 р. Пакет орієнтований на масового користувача з невеликим досвідом як у статистичному аналізі, так і в спілкуванні з комп'ютером. Пакет поєднує діалоговий характер інтерфейсу з ієрархічною системою меню та контекстово-орієнтованою допомогою.

У пакеті представлені не тільки прості процедури статистичної обробки інформації, але й складні процедури аналізу даних:

– параметричні тести (кореляція, тести Стьюдента та Фішера, гістограма, нормальність);

– непараметричні тести;

– аналіз часових рядів (кореляційний, спектральний, згладжування, фільтрація);

– дисперсійний аналіз (одно-двофакторний параметричний і непараметричний);

– регресійний аналіз (порівняння двох регресій, проста регресія (тренд), множинна лінійна регресія, покрокова регресія, загальна регресія);

– багатовимірні методи.

Пакет Stadia має також досить повний набір двомірних і тримірних діаграм розсіяння, функціональних графіків однієї або декількох змінних, стовпців і кругових діаграм тощо.

У цій системі відсутній текстовий редактор для підготовки звітів, але передбачене все для того, щоб одержані результати могли бути оброблені за допомогою зовнішнього редактора.

Пакет має внутрішню експертну систему з вибору та використання методів аналізу, для чого використовується спеціальна класифікація різноманітних можливих типів статистичних даних і для кожного з них вказані можливі методи обробки, призначення та порядок роботи з ними.

Пакет **StatGraphics** (Statistical Graphics System) розроблений корпорацією STSC (USA). Безперечними достоїнствами пакета є повнота представлених у пакеті методів аналізу даних, добра двовимірна та тривимірна графіка, широкі можливості оперування даними. Пакет розрахований на спеціалістів, добре знайомих з концепціями використовуваних процедур, а це в певній мірі знижує можливості застосування пакета широким колом користувачів.

Крім процедур аналізу даних, представлених у пакеті й аналогічних процедурам пакета Stadia, до StatGraphics включений великий запас загальноматематичних алгоритмів (розв'язання систем лінійних рівнянь, чисельне диференціювання та інтегрування тощо). Графічний редактор пакета дозволяє настроювати всі можливі елементи графіка від області визначення, розмірів, масштабу по кожній з вісей, рамки заголовка до типів і кольорів ліній, точок, графіка та детального оформлення вісей координат. У режимі інтерактивного

редагування графіка пакет StatGraphics дозволяє розміщувати мітки точок, змінюючи при цьому їх вигляд і розташування, виводити значення координат точок тощо. Пакет дозволяє здійснити повну підготовку підсумкового звіту. Документація StatGraphics досить широка та змістовна.

21.3. Математичні ППП

Типовими можливостями математичних пакетів є розв'язання таких проблем:

– тотожні перетворення алгебраїчних, тригонометричних, логарифмічних, експоненційних, дробово-раціональних виразів;

– обчислення точні та наближені (розв'язування рівнянь і систем рівнянь із заданою точністю);

– символьне та чисельне диференціювання та інтегрування;

– операції над векторами та матрицями;

– обчислення імовірнісних і статистичних функцій;

– побудова та перетворення графіків функцій (двомірні та тримірні графіки, графіки функцій, задані в різних системах координат).

Ядро будь-якого математичного пакета складається з великої кількості функцій. До нього часто додають змінний набір пакетів підпрограм, які розширюють спроможність пакету та дозволяють розв'язувати типові задачі інших предметних галузей.

Усі математичні пакети мають загальний набір:

– файлових операцій (Load, Save, Save As,… та ін);

– редагування;

– вставляння та вилучення об'єктів і фрагментів;

– форматування тексту та графіків;

– настроювання системи;

– управління вікнами;

– забезпечення довідковою інформацією (систематичний та алфавітний каталоги).

Спілкування з ЕОМ у середовищі математичних пакетів програм відбувається на досить природній мові. Для запису команд та окремих виразів у математичних пакетах використовується вхідна мова, яка нагадує Бейсік із додаванням Фортрану та Паскалю. Вона застосовується або безпосередньо (Matlab, Derive), або в комбінації з "кнопково-шаблонним" інтерфейсом (Maple, Mathematica, Mathcad).

Як уже вказувалося вище, у склад математичних пакетів входять засоби побудови графіків, що дозволяє представити розрахунки в максимально наочній формі. Графіки можуть бути побудовані у двомірному та тримірному просторі (у декартовій, циліндричній та сферичній системах координат). Користувач може

задавати частоту сітки, оцифровку вісей, різні надписи та ін. Для тримірних графіків є можливість побудови "дротяних" поверхонь, заховання невидимих ліній, переміщення "точки зору", освітлення та забарвлення окремих частин графіку і т.ін.

Вихідним документом математичного пакета є робочий аркуш (**Worksheet**), який являє собою уривок майбутнього документа з мінімумом тексту. Звичайно робочий аркуш створюється у специфічному форматі, що обмежує можливість перенесення інформації. Однак, якщо навіть вставляння робочого аркуша з виконаним розрахунком у документ MS Word пройшло успішно, залишається проблема редагування вставки. Власні ж засоби MS Word для набору математичних виразів дуже обмежені. Цю проблему можна розв'язати, використовуючи видавницьку систему LaTex, створену для набору та редагування технічних текстів.

Застосування математичних методів для дослідження цілеспрямованої людської діяльності називається дослідженням операцій. Сучасні методи дослідження операцій є важливим інструментом діяльності економіста. Це такі методи:

– лінійне програмування; використовується для розв'язання задач з лінійною цільовою функцією при лінійних обмеженнях на змінні та вимозі їх невід'ємності (задачі на складання оптимальних сумішей, транспортні проблеми і т. ін.);

– нелінійне програмування; застосовується для розв'язання задач з нелінійною цільовою функцією та/або обмеженнях (зиск від реалізації продукції, різного виду затрати праці та матеріальних ресурсів і т. ін.);

– динамічне програмування; використовується для пошуку оптимальних багатоетапних розв'язків (планування господарської діяльності групи підприємств, вибір найекономнішого маршруту доставки вантажу та ін.);

– задачі на графах; застосовується для розв'язання мережевих задач (складання календарного плану виконання певного комплексу робіт, оцінка пропускної спроможності даної мережі);

– теорія гри;

– теорія запасів;

– теорія масового обслуговування та ін.

Всі ці методи повністю або частково можна реалізувати за допомогою математичних пакетів.

21.3.1. Пакет Maple

Maple V6.01 – середовище для виконання математичних розрахунків на комп'ютері. Maple дозволяє аналізувати та моделювати найрізноманітніші процеси, розв'язувати рівняння, складати звіти.

21.3.1.1. Загальна характеристика

Програма розроблена дослідницькою групою (The Symbolic Computation Group) відділу обчислювальної техніки університету Waterloo (Канада).

Необхідні вимоги до комп'ютера:

– Intel 386, 486, Pentium або повністю сумісний процесор;

– від 18 до 42 МБ вільного дискового простору;

– мінімум 8 МБ оперативної пам'яті;

– операційна система Microsoft Windows 9x, Windows NT.

Maple може виконувати складні алгебраїчні перетворення та спрощення над полем комплексних чисел, знаходити скінченні та нескінченні суми, добутки, границі та інтеграли, розв'язувати в символьному вигляді та чисельно алгебраїчні системи рівнянь і нерівностей, знаходити всі корені многочленів, розв'язувати аналітично та чисельно системи звичайних диференційних рівнянь та деякі класи рівнянь у частинних похідних. До Maple включені пакети підпрограм для розв'язання задач лінійної та тензорної алгебри, Евклідової та аналітичної геометрії, теорії ймовірностей і математичної статистики, чисельної апроксимації та лінійної оптимізації (симплекс-метод), задачі фінансової математики та ін.

21.3.1.2. Робоча панель Maple

Виклик пакета здійснюється через **Меню Пуск – Программы – Maple – Maple 6**. Після цього відкривається робочий аркуш Maple із звичайним для Windows-додатків набором пунктів меню та кнопкових функцій (рис. 21.1).

Програма дозволяє водночас працювати з декількома робочими аркушами та встановлювати між ними динамічні зв'язки (тобто переводити обчислення з одного аркуша на інший).

У залежності від того, на якому об'єкті встановлений курсор, змінюється вигляд рядка меню та кнопок панелі інструментів у відповідності до операцій, які можна виконати над об'єктом. Для вивчення призначення кнопок панелі інструментів досить активізувати опцію **Balone help** у розділі **Help** рядка меню.

Відзначимо лише призначення основних кнопок:

[Σ] – текстуальне вставлення команди;

[T] – вставлення та форматування текстового коментаря;

[>] – вставлення нової зони набору після курсору;

[Stop] – припинення теперішнього обчислення;

– збільшення екранного зображення на 100, 150 та 200% від оригіналу.

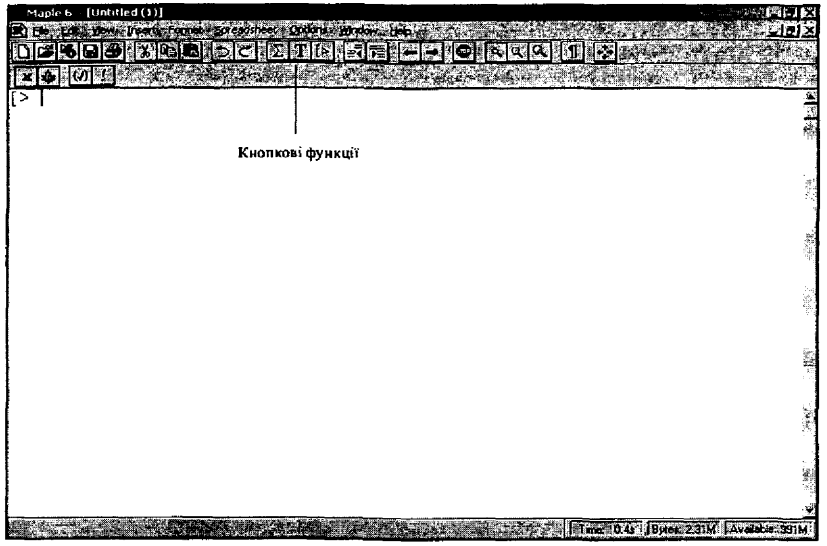


Рис. 21.1. Робочий аркуш Maple.

Система має добрий текстовий редактор, який дозволяє використовувати всі інстальовані на комп'ютері шрифти, змінювати їх тип (напівжирний, курсив і т. ін.), форматувати розділи тексту.

На екрані дисплея в середовищі Maple рядки введення мають червоний колір, відповідь системи – блакитний, коментарі користувача – чорний.

Команди Maple закінчуються символами "**:**" або "**;**" (перший варіант заглушає виведення команди на екран). Частина рядка, взята у зворотні лапки, сприймається як коментар (рис. 21. 2).

21.3.1.3. Довідкова підсистема Maple

Одержати інформацію про конкретну команду або функцію Maple можна, якщо ввести її назву в командний рядок після знака питання. Іншим способом

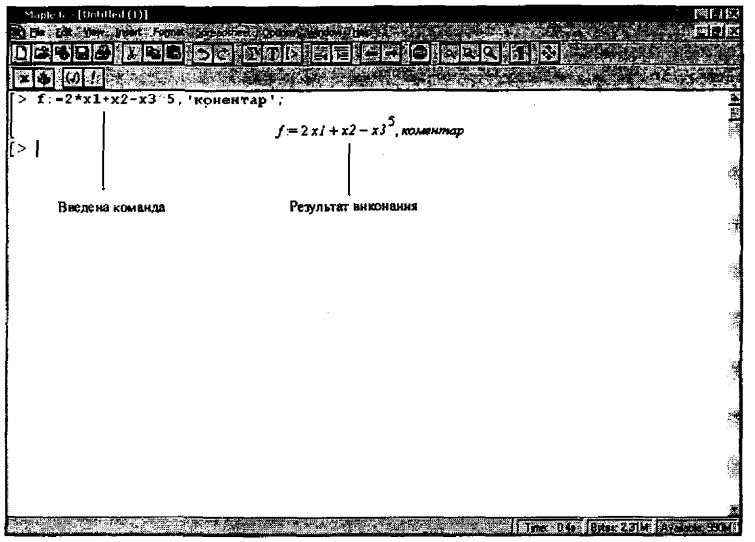


Рис. 21.2. Введення коментаря.

пошуку інформації є виклик пункту **Contents** (зміст) з розділу **Help** рядка меню. При цьому можна переглянути назву всіх розділів довідки. Можна також здійснити тематичний пошук,використовуючи пункт **Topic search** (тематичний пошук) і вводячи в діалоговому вікні назву теми пошуку. Для звертання до контéкстно залежного **Help** необхідно виділити в робочому аркуші ключове слово та натиснути клавішу **[F1]**. У відповідь з'явиться потрібна сторінка **Help**.

21.3.1.4. Типи даних

Maple не передбачає обов'язкового оголошення типів змінних. До вбудованих типів даних належать раціональні, дійсні, логічні та символьні числа. Раціональні числа в загальному випадку мають вигляд дробів із цілими компонентами. Дійсні числа записуються у формі <мантиса> *10^ (<порядок>). Символьні змінні містяться у зворотних лапках. Як об'єкти цього вигляду найчастіше використовуються імена файлів для операцій введення-виведення.

Імена (ідентифікатори) записуються так само, як і в усіх алгоритмічних мовах. В іменах маленькі та великі літери відрізняються.

З елементарних об'єктів можуть бути сформовані складені – множини, списки, масиви, матриці, вектори. Елементи множин записуються у фігурних дужках,

порядок переліку не грає ролі. Списки містяться у квадратних дужках; тут можливі повторення, а позиція елементів суттєва.

Для складання виразів використовуються знаки операцій +, -, *, /, ^, !. Засобом запам'ятовування результатів обчислення є оператор присвоєння ": =". Набрана команда, якщо вона закінчується крапкою з комою, виконується після натискання клавіші [Enter], результат її виконання виводиться на екран. Знак ":" у кінці команди заглушає виведення результату на дисплей.

21.3.1.5. Підпрограма *Networks*

До складу Maple входять 32 додаткових пакети підпрограм. Для виклику процедур підпрограми слід відкрити до неї доступ командою

>with (<ім'я пакета підпрограм>):

(при закінченні команди символом ";" виводиться перелік процедур пакета підпрограм).

Розглянемо підпрограму ***Networks***.

Процедури цієї підпрограми служать для розв'язання задач теорії графів. Виклик підпрограми *Networks* здійснюється за допомогою команди

>with (networks):

або

>with (networks, <назва процедури>):

для виклику конкретної процедури підпрограми. *Networks* містить 75 процедур роботи з графами.

Граф складається з вершин і ребер (дуг), може бути спрямованим, припускає кратні дуги та цикли. У підпакеті *Networks* граф створюється за допомогою процедури *new* або процедурами, які створюють графи окремих виглядів. Утворений граф можна перетворювати за допомогою різних процедур, які буде розглянуто нижче.

За допомогою процедур *addvertex* та *addedge* можна додавати відповідно вершини та ребра. Усі вершини і ребра графу нумеруються. Ребра задаються множинами з'єднаних вершин, дуги (спрямовані ребра) – аналітичними списками. По умовчанню ребра мають одиничну вагу. Для визначення властивостей графу *G* використовуються процедури *edges* (*G*) або *vertices* (*G*).

Крім вказаних вище, пакет містить ще такі процедури:

– знаходження найкоротших шляхів між усіма парами вершин – *allpairs* (*G*);
– находження з'єднаних компонентів графу – *components* (*G*);
– з'єднання двох наборів вершин графу – *connect* (*S*1, *S*2, *G*);
– стягування дуг графу – *contract* (<набір дуг>, *G*);
– знаходження найменшого розрізу – *mincut* (*G*, *S*, *T*);
– знаходження дерев найменшого вигляду – *counttrees* (*G*);
– обчислення найменшої відстані між двома вершинами – *diameter* (*G*);
– відображення графу – *draw* (*G*);
– знаходження ваги ребер – *eweight* (*G*);
– знаходження найкоротшого циклу неспрямованого графу – *girth* (*G*);
– створення графу із заданим набором вершин і ребер – *graph* (<набір вершин>,<набір з'єднаних пар вершин>);
– вилучення простого графу зі складеного – *gsimp* (*G*);
– перевірка площинності графу – *isplanar* (*G*);
– демонстрація відомої інформації про граф – *show* (*G*);
– знаходження зв'язності графу – *connectivity* (*G*);
– знаходження дерев найменшої ваги – *spantree* (*G*)

та інші.

Розглянемо приклад побудови графу та визначення його властивостей у підпрограмі *Networks* (рис. 21.3):

– підпрограму викликаємо командою

>with (*networks*):

– для створення графу використовуємо процедуру *new*(*G*), яку завершуємо двокрапкою;

– задаємо вершини графу процедурою *addvertex*. У дужках записуємо список вершин і через кому, назву графу *G*. Завершуємо процедуру крапкою з комою, на екрані висвітлюються назви вершин графу;

– задаємо ребра графу процедурою *addedge*. У дужках записуємо список пар з'єднаних вершин графу, після коми – процедуру *weights*, яка задає вагу ребер. Вагу ребер графу записуємо як список чисел, у тому ж порядку, що й самі ребра. Через кому пишемо назву графу *G*. Процедуру завершуємо символом ";", після чого на екрані висвітлюються назви ребер графу.

– відобразимо граф за допомогою процедури *draw* (*G*) – рис. 21.4.

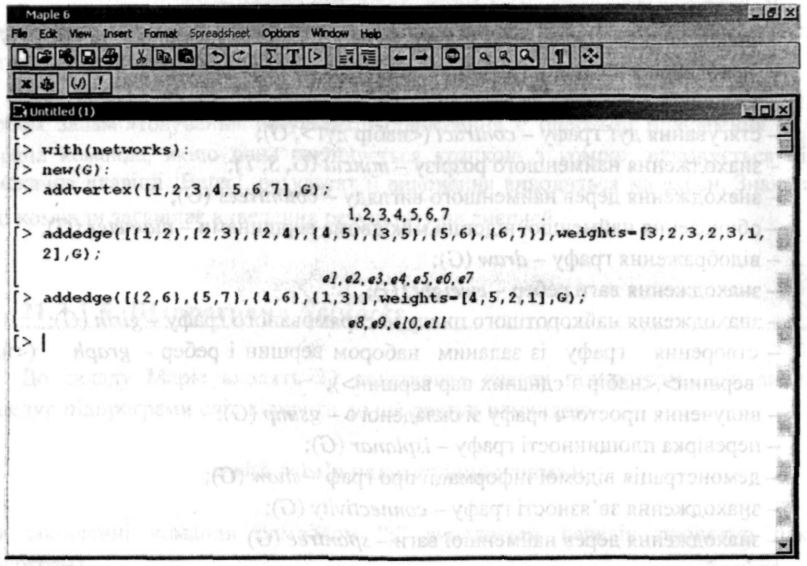


Рис. 21.3. Задання графу.

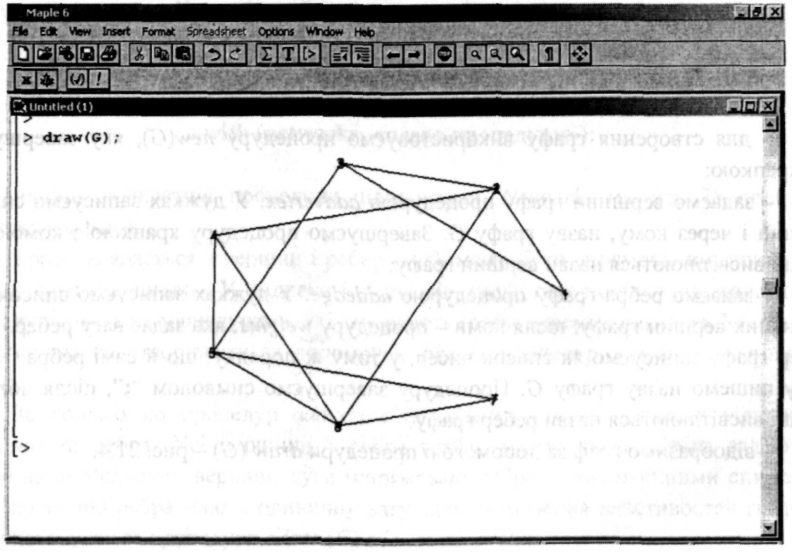


Рис. 21.4. Відображення графу.

Властивості створеного графу можна досліджувати за допомогою різних процедур підпрограми. Продемонструємо деякі з них.

Знайдемо найкоротші шляхи між кожною парою вершин графу за допомогою процедури *allpairs* (*G*, *V*), де *G* – назва графу, *V* – умовна назва вершини (викликає список усіх вершин графу) – рис. 21.5.

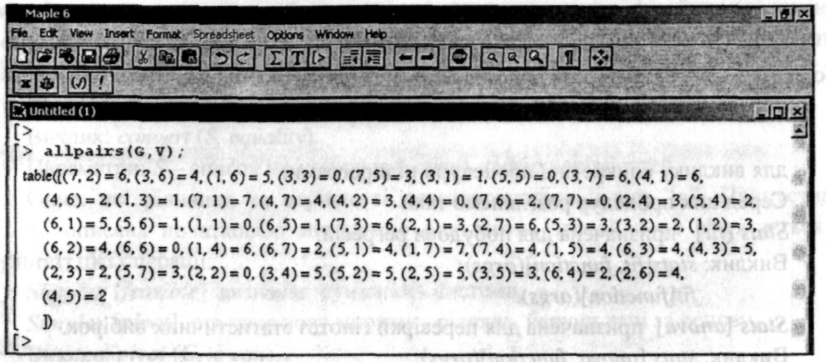


Рис. 21.5. Знаходження найкоротших шляхів між вершинами графу.

За допомогою процедури *arrivals*(*G*) можна знайти списки усіх вершин графу, які з'єднуються ребрами з кожною з вершин.

І нарешті, знайдемо найкоротший шлях графу з першої до останньої вершини. Для цього використовуємо процедуру *diameter* (*G*) – рис. 21.6.

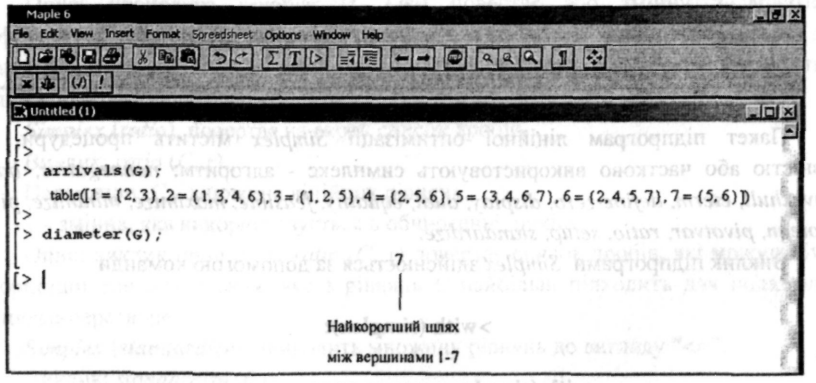


Рис. 21.6. Знаходження з'єднаних вершин графу та найкоротшого шляху між вершинами 1 – 7.

21.3.1.6. Підпрограма *Stats*

Підпрограма *Stats* містить процедури для обробки й аналізу даних. На відміну від інших спеціалізованих пакетів, цей пакет підпрограм складений, тобто він містить процедури, які у свою чергу містять команди.

Виклик підпрограми *Stats* здійснюється за допомогою команди

>with (stats):

або

>with (stats, <назва процедури>):

для виклику конкретної процедури підпрограми.

Серед цих процедур розглянемо такі:

***Stats* [*fit*]** призначена для побудови регресій.

Виклик: *stats*[*fit, function*](*args*);

fit[*function*](*args*).

***Stats* [*anova*]** призначена для перевірки гіпотез статистичних вибірок.

Виклик: *stats* [*anova, function*](*args*);

anova [*function*](*args*).

***Stats* [*statsplots*]** використовується для побудови графіків та гістограм.

***Stats*[*describe*]** призначена для розрахунку чисельних характеристик статистичного ряду; дозволяє обчислювати моду та розмах розподілу, середні значення (звичайне, геометричне, гармонічне), перерізи статистичного розподілу (медіани, квартилі та децилі, а також квантилі для довільного аргументу) та ін.

21.3.1.7. Підпрограма *Simplex*

Пакет підпрограм лінійної оптимізації *Simplex* містить процедури, які повністю або частково використовують симплекс - алгоритм: *nonnegative, basis, convexhull, cterm, define-zero, display, dual, equality, feasible, maximize, minimize, pivot, pivoteqn, pivotvar, ratio, setup, standardize.*

Виклик підпрограми *Simplex* здійснюється за допомогою команди

>with (simplex):

або

>with (simplex, <назва процедури>):

для виклику конкретної процедури підпрограми.

Опишемо деякі з цих процедур.

***Simplex* [*setup*]** створює множину рівнянь з ізольованими змінними.

Виклик: *setup* (*C*).

Параметри: *C* – множина лінійних рівнянь.

Опис: процедура *setup* (*C*) відтворює множину рівнянь, змінні в яких розташовані в лівій частині. Ці змінні складають базис відповідної лінійної системи і не зустрічаються у правій частині кожного рівняння. Необмежені змінні перетворюються на різницю двох змінних. Одержана система еквівалентна початковій системі *C*, її розв'язок може бути перетворений у розв'язок *C*.

***Simplex* [*equality*]** перетворює нерівності в рівняння.

Виклик: *convert* (*S, equality*).

Параметри: *S* – список чи множина рівнянь та/або нерівностей.

Опис: всі рівняння та нерівності перетворюються у тип "=". Припустима область, описана початковою множиною відношень, звичайно не зберігається в результаті цієї операції.

***Simplex* [*feasible*]** визначає сумісність системи.

***Simplex* [*pivot*]** створює нову множину рівнянь, беручи дані за основу.

Виклик: *pivot* (*C*, *x*, *eqn*).

Параметри: *C* – множина або список рівнянь, *x* – ім'я змінної, *eqn* – рівняння або список рівнянь.

***Simplex* [*pivotvar*]** повертає на екран змінну з додатним коефіцієнтом.

Виклик: *pivotvar* (*f*, *List*)

pivotvar (*f*)

Параметри: *f* – цільова функція, виражена через небазисні змінні; *List* – список змінних.

Опис: процедура pivotvar (*f*, *List*) повертає або змінну з додатним коефіцієнтом, або FAIL (відсутність такої). Опція *List* використовується для визначення порядку, в якому змінні повинні бути розглянуті. На екран виводиться перший позитивний результат.

***Simplex* [*ratio*]** повертає на екран список дробів.

Виклик: *ratio* (*C*, *x*)

Параметри: *C* – множина лінійних рівнянь;

x – змінна, яка використовується в обчисленні дробів.

Опис: виклик процедури *ratio* (*C*, *x*) повертає список дробів, які можуть бути використані для визначення, яке з рівнянь *C* найбільш підходить для подальших симплекс-перетворень.

***Simplex* [*standardize*]** приводить множину рівнянь до вигляду "<=".

Виклик: *standardize* (*C*)

Параметри: *C* – множина або список лінійних обмежень.

Опис: процедура *standardize(C)* повертає множину або список нерівностей вигляду "<=", еквівалентних початковій множині.

Convert/std приводить дані до стандартної форми для симплекс-перетворень.

Simplex [maximize] – програма лінійної оптимізації.

Виклик: *maximize (f, C)*

maximize (f, C, vartype)

maximize (f, C, vartype, 'NewC', 'transform')

Параметри: *f* – цільова функція (лінійна); *C* – множина або список лінійних обмежень;

vartype – NONNEGATIVE чи UNRESTRICTED; *NewC* – ім'я; *transform* – ім'я.

Опис: вираз *f* – лінійна цільова функція, для якої треба знайти найбільше значення за умови лінійних обмежень *C*. Процедура *maximize* повертає або множину рівностей, які описують оптимальний розв'язок, або порожню множину у випадку, якщо розв'язок системи *C* не існує, або NULL, якщо розв'язок необмежений. Рівності можуть бути підставлені в цільову функцію *f* для обчислення найбільшого значення функції.

Третій параметр (*vartype*) може бути використаний для уточнення того, що всі змінні повинні бути невід'ємними (NONNEGATIVE). Параметр UNRESTRICTED означає, що змінна повинна бути одержана без знаку.

Четвертий і п'ятий параметри можуть використовуватися при уточненні імен для одержання оптимального опису.

Наведемо приклад використання підпрограми *Simplex* для розв'язання задачі лінійної оптимізації.

Приклад. Знайти найбільше значення функції $f = 4x_1 + 9x_2 + x_3 + x_4$ за умови:

$$-2x_1 + x_2 + x_3 = 4,$$
$$2x_1 + 7x_2 + x_4 = 92,$$
$$2x_1 - 3x_2 \le 12,$$
$$x_1 \ge 0, x_2 \ge 0, x_3 \ge 0, x_4 \ge 0.$$

Вводимо умову задачі (рис. 21.7). Для введення тексту використовуємо текстуальний режим (натискаємо кнопку [T] панелі інструментів), а для введення формул – текстуальну вставку формули (натискаємо кнопку [Σ] панелі інструментів).

Викликаємо підпрограму *Simplex* (рис. 21.8).

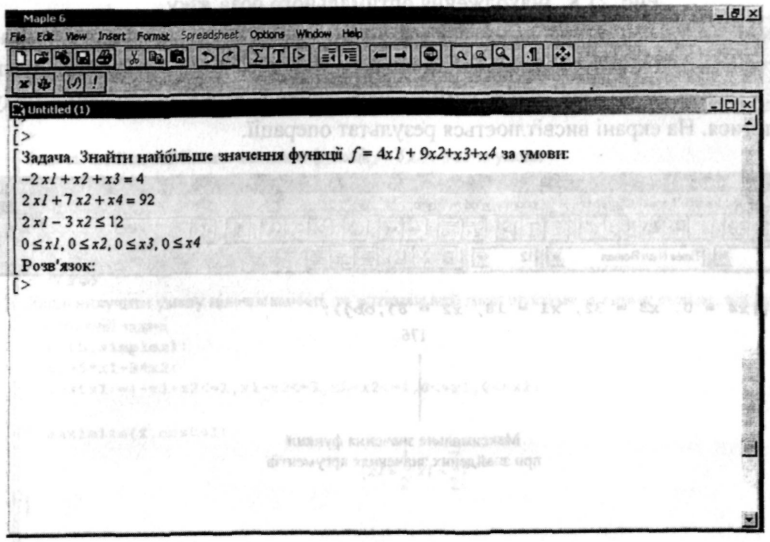


Рис. 21.7. Введення умови задачі.

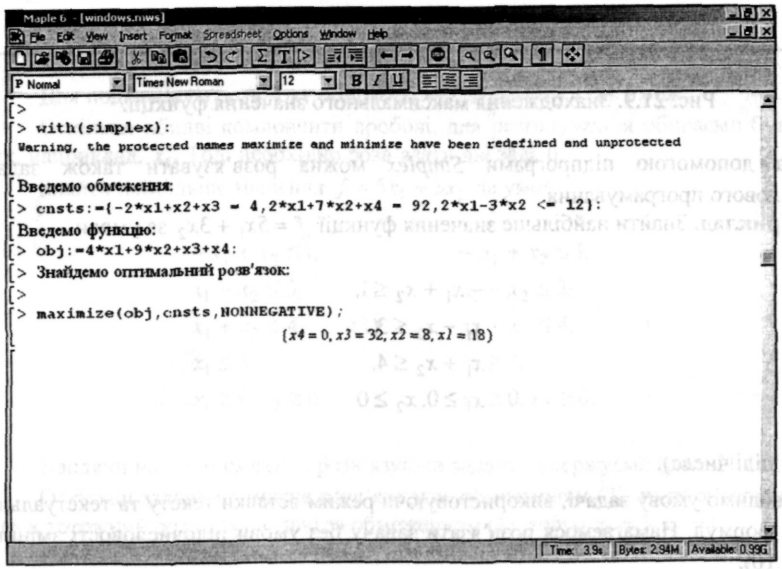


Рис. 21.8. Знаходження оптимального розв'язку.

Для знаходження значення функції при одержаних значеннях змінних використовуємо процедуру *subs* (підстановка) – рис. 21.9. У фігурних дужках вказуємо значення змінних, після коми – ім'я функції, в яку ці значення будуть підставлятися. На екрані висвітлюється результат операції.

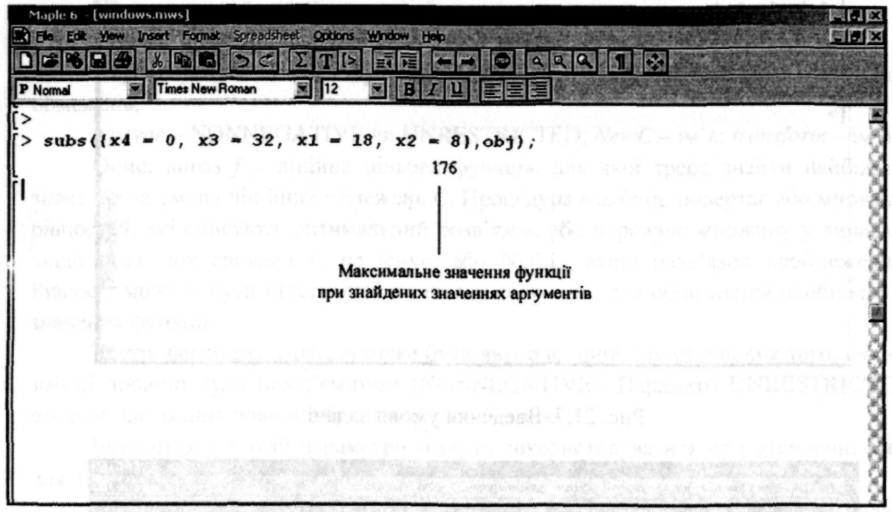


Рис. 21.9. Знаходження максимального значення функції.

За допомогою підпрограми *Simplex* можна розв'язувати також задачі цілочислового програмування.

Приклад. Знайти найбільше значення функції $f = 5x_1 + 3x_2$ за умови

$$\begin{aligned} &-x_1 + x_2 \le 1, \\ &x_1 - x_2 \le 3, \\ &x_1 + x_2 \le 4, \\ &x_1 \ge 0, x_2 \ge 0 \end{aligned}$$

(x_1, x_2 - цілі числа).

Вводимо умову задачі, використовуючи режим вставки тексту та текстуальної вставки формул. Намагаємося розв'язати задачу без умови цілочисловості змінних (рис. 21.10).

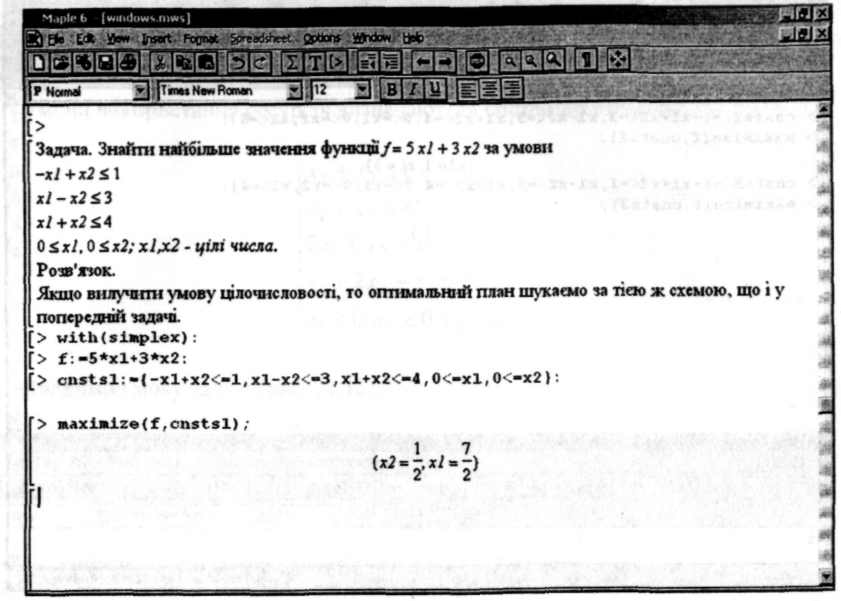


Рис. 21.10. Перший етап розв'язку задачі.

Для подальшого розв'язку обираємо метод розгалужень і меж.

Оскільки обидві компоненти дробові, для розгалуження обираємо будь-яку з них, наприклад, x_1. Тоді необхідно розв'язати дві задачі.

Знайти найбільше значення $f = 5x_1 + 3x_2$ за умов

$$\begin{aligned} &-x_1 + x_2 \le 1, \\ &x_1 - x_2 \le 3, \\ &x_1 + x_2 \le 4, \\ &x_1 \le 3, \\ &x_1 \ge 0, x_2 \ge 0, \end{aligned} \quad (1) \qquad \begin{aligned} &-x_1 + x_2 \le 1, \\ &x_1 - x_2 \le 3, \\ &x_1 + x_2 \le 4, \\ &x_1 \ge 4, \\ &x_1 \ge 0, x_2 \ge 0. \end{aligned} \quad (2)$$

Вводячи нові обмеження і розв'язуючи задачу , одержуємо (рис. 21.11).

Оскільки задачу не можна розв'язати за обмеженням (2), то розв'язок вихідної задачі тотожний розв'язку задачі за обмеженням (1), тобто $x_1 = 3, x_2 = 1$.

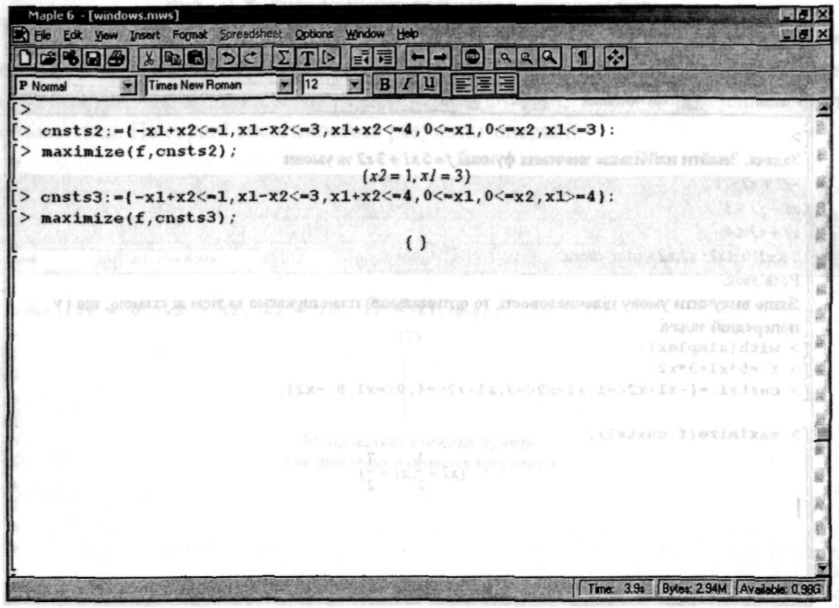


Рис. 21.11. Другий етап розв'язку задачі.

У дослідженні операцій симплекс-метод використовується при вирішенні оптимізаційних задач, тому розглянемо вирішення однієї з них з використанням цього методу.

Приклад. Підприємство може випускати три типи продукції, використовуючи чотири види ресурсів, запаси яких складають відповідно 6;8;16;18. Слід скласти план оптимального зиску від реалізації продукції за відомих витрат у грошових одиницях виміру кожного виду ресурсів на одиницю кожного типу продукції та вартості одиниці кожного типу продукції (табл. 21.1).

Таблиця 21.1.

Тип продукції	Витрати на одиницю продукції				Вартість одиниці продукції
	Види ресурсів				
	1	2	3	4	
1	0	1	2	1	3
2	1	1	1	2	4
3	1	0	1	1	2

Позначимо x_1, x_2, x_3 як обсяги продукції кожного типу, тобто шуканий план $\bar{x} = (x_1, x_2, x_3)$. Зиск від реалізації за певного плану випуску визначається функцією $f = 3x_1 + 4x_2 + 2x_3$, для якої треба знайти найбільше значення.

Умови використання ресурсів визначаються системою нерівностей:

$$\begin{cases} x_2 + x_3 \le 6, \\ x_1 + x_2 \le 8, \\ 2x_1 + x_2 + x_3 \le 16, \\ x_1 + 2x_2 + x_3 \le 18, \\ x_1 \ge 0, x_2 \ge 0, x_3 \ge 0. \end{cases}$$

Вводимо умову задачі (рис. 21.12).

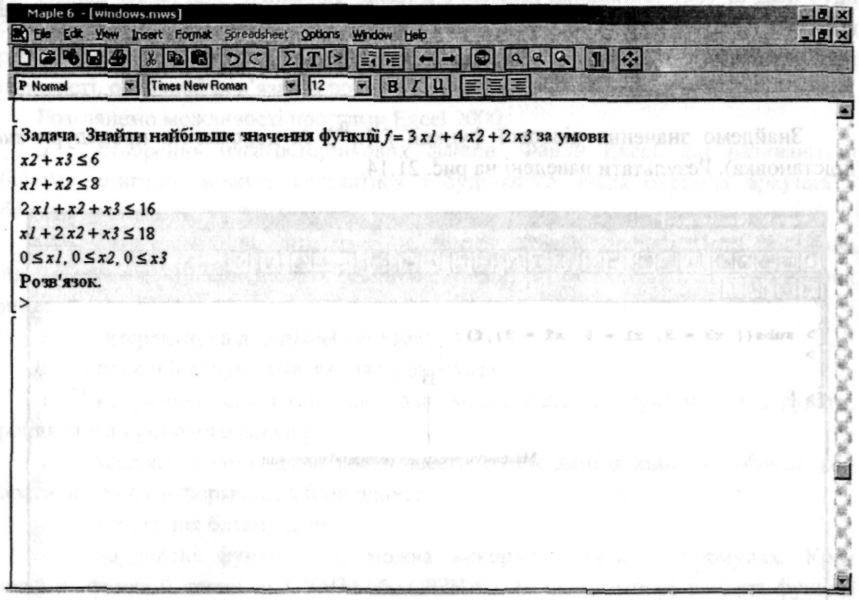


Рис. 21.12. Введення умови задачі.

Викликаємо підпрограму *Simplex* і вводимо обмеження як список нерівностей, який має ім'я *cnsts*, а функції присвоюємо ім'я f (рис. 21.13).

Отже, оптимальним планом задачі є $\bar{x}^0 = (5; 3; 3)$.

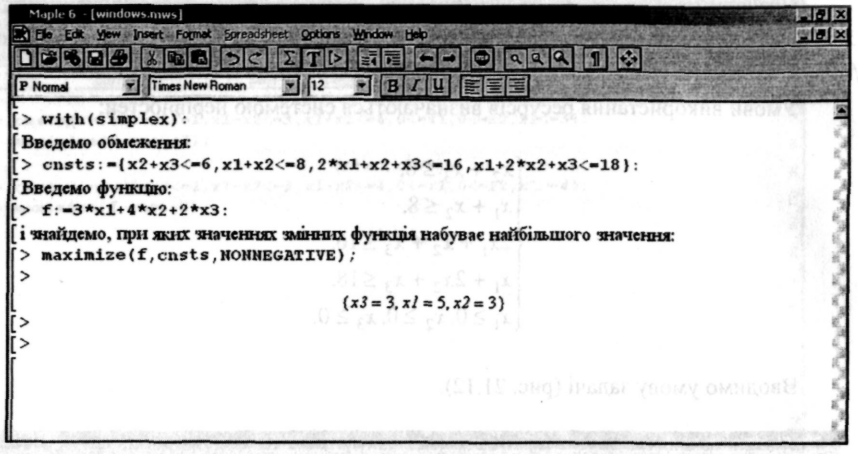


Рис. 21.13. Знаходження оптимального плану.

Знайдемо значення цільової функції $f(\overline{x^0})$ за допомогою процедури *subs* (підстановка). Результати наведені на рис. 21.14.

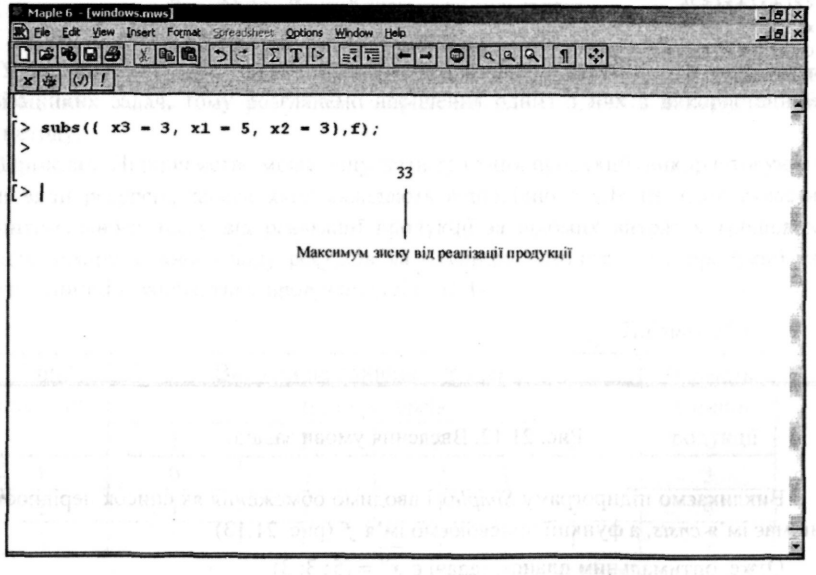


Рис. 21.14. Знаходження максимального зиску.

21.4. Microsoft Excel 2000 як інструмент для вирішення задач математичного програмування

Розглянемо цей розділ з урахуванням того, що читач володіє основними навичками роботи в середовищі Windows, а також працював з якою-небудь версією програми Microsoft Excel*. Але все ж таки нагадаємо про можливості цієї програми та її основні поняття.

Excel – це програмний продукт, який належить до категорії *електронних таблиць*. Ця програма дає необмежені можливості в обчисленні даних у вигляді таблиць, у виконанні бухгалтерських розрахунків, аналізі даних, статистичних обчисленнях, вирішенні наукових задач, пов'язаних з виконанням багатьох математичних операцій. Усі ці розрахунки можна представити також у вигляді різноманітних графіків, діаграм.

Перша версія для Windows – Excel 2 з'явилася наприкінці 1987 р. Далі були версії Excel 3 (4, 5, 7, 8) і остання версія Excel 9 – більше відома як Excel 2000. Ця версія зберегла все найкраще від попередніх версій, але й придбала цілковито нові можливості, особливо пов'язані з роботою в Internet.

Розглянемо можливості програми Excel 2000:

- створення багатосторінкових файлів. Файли Excel, які називаються робочими книгами, можуть складатися з будь-якого числа окремих аркушів – робочих або аркушів діаграм;
- багатовіконний інтерфейс. В Excel можна працювати з кількома файлами одночасно і здійснювати обмін інформацією між даними в різних робочих книгах;
- інтерактивна довідкова система;
- панелі інструментів, які настроюються;
- керування списками, що дає можливість сортувати, фільтрувати, обробляти дані робочого аркуша;
- зведені таблиці дозволяють звести разом дані з кількох таблиць або вибрати потрібну інформацію з бази даних;
- керування базами даних;
- вбудовані функції, які можна використовувати у формулах. Крім звичайних функцій, таких, як СУММ або СРЗНАЧ, до цього набору входять функції більш складних операцій, які важко або неможливо виконати іншим способом. Наприклад, функція КОРРЕЛ обчислює коефіцієнт кореляції між двома наборами даних;
- розширені засоби аналізу, такі, як алгоритм підбору параметрів,

* Щербаков П. А., Ульянченко О. В., Мартьянова Н. М., Бутенко Т. А. Інформатика та комп'ютерна техніка: програмне забезпечення ЕОМ: Навч. посібник / Харк. держ. аграр. ун-т ім. В.В. Докучаєва. – Харків, 2001.

процедура пошуку оптимальних рішень, а також надбудова **Пакет анализа**, яка включає в себе безліч статистичних, фінансових, інженерних і наукових функцій та процедур;

- різноманітність діаграм – можливість не тільки створювати діаграми та графіки різних типів, але й модифікувати існуючі та поповнювати їх асортимент. Розташовувати діаграму можна в будь-якому місці робочого аркуша або на спеціальному аркуші діаграми;
- використання можливостей картографії. Excel дозволяє розмістити числові дані, прив'язавши їх до географічної карти. Наприклад, дані про обсяги продажу зерна в різних країнах, зображені поруч з географічною картою, справлять велике враження;
- *інструменти малювання* дозволяють створювати діаграми та малюнки безпосередньо на робочому аркуші або діаграмі. Наприклад, щоб полегшити виконання аналізу даних, можна включити блок-схему в таблицю з даними;
- Visual Basic for Applications (VBA). Мова програмування, вбудована в Excel, розкриває нові можливості;
- підтримка Internet. Excel 2000 має широкий набір нових засобів, які дозволяють мати доступ до інформації в Internet, зберігати документи у форматі HTML і створювати гіперзв'язки безпосередньо у своїй робочій книзі.

Входячи до пакета Microsoft Office 2000, програма обробки таблиць Excel 2000 ставить такі ж високі вимоги до комп'ютера, як і інші програми цього пакета: об'єм оперативної пам'яті не менше 32 Мб, наявність процесора Pentium з робочою частотою від 100 МГц, розмір вільного місця на жорсткому диску для установки програми більше 25 Мб.

Після інсталяції програми завантажити її можна, викликавши **Меню Пуск – Программы – Office – Microsoft Excel.**

Після завантаження Excel на екрані з'являється вікно інтерфейса цієї програми зі стандартними елементами вікон Windows (рис. 21.15).

Нагадаємо, що основне робоче поле вікна програми Excel складається з набору рядків і стовпчиків. Це вікно можна прокручувати вверх і вниз, праворуч і ліворуч, використовуючи смуги прокручування. Область, що знаходиться на перехресті рядку та стовпця, називається *чарункою*. Якщо її позначити табличним курсором, вона стає *активною чарункою*. У чарунку можна ввести число, текст або формулу. За допомогою формул можна виконати обчислення над значеннями, які зберігаються в інших чарунках. При копіюванні або переміщенні чарунок змінюються і формули, які в них зберігаються, відповідно до нових адрес.

Найбільш важливою перевагою електронної таблиці є те, що при зміні вмісту будь-якої чарунки, значення якої використовується у формулі, результат буде автоматично перераховуватися. Це означає: якщо створено електронну таблицю з відповідними формулами, то можна використовувати її завжди, коли потрібно

виконати обчислення відповідно до цього алгоритма, але з іншими значеннями. Для цього потрібно лише внести нові значення у відповідні чарунки.

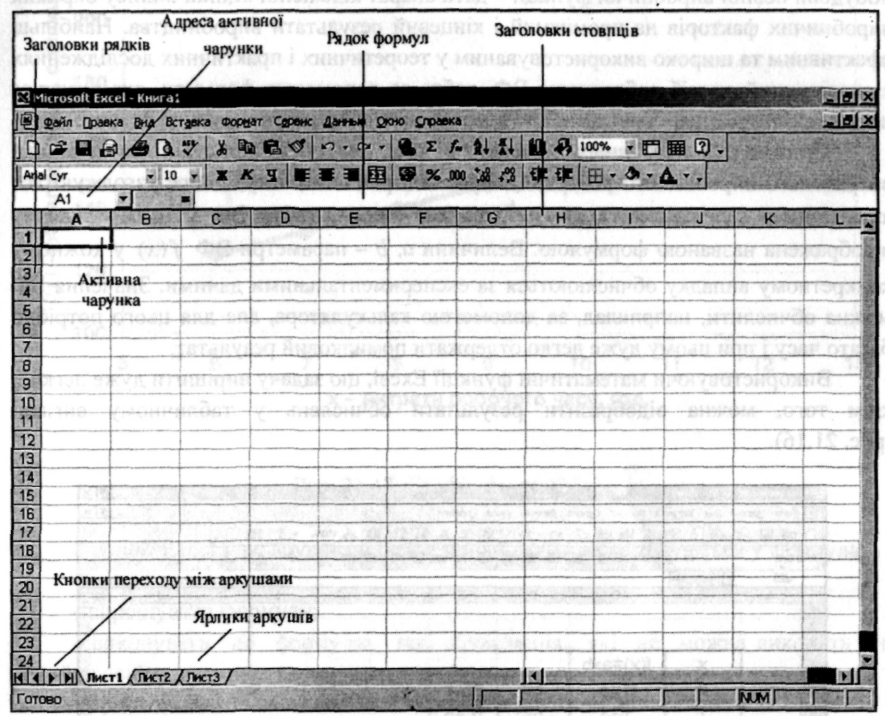


Рис. 21.15. Вікно інтерфейса Microsoft Excel.

Чарунки електронної таблиці можна форматувати самими різними засобами та друкувати для ведення документації. Крім того, окремі групи чарунок з даними можна використовувати для створення діаграм або карт.

Електронну таблицю можна зберігати у файлі для подальшого використання, як і будь-який інший документ.

21.4.1. Функції Excel

У дослідженні операцій широко використовуються математичні моделі, які грунтуються на формально-логічному апараті математики. Пошуки методів

ефективного використання математичного апарату в економічних дослідженнях обумовили необхідність побудови так званих *виробничих функцій* (ВФ). Мета побудови певної виробничої функції – дати апарат кількісної оцінки впливу окремих виробничих факторів на проміжний і кінцевий результати виробництва. Найбільш ефективним та широко використовуваним у теоретичних і практичних дослідженнях є аналітичний спосіб зображення ВФ, тобто за допомогою формули, яка визначає послідовність дій з аргументами для одержання відповідного значення функції.

Приклад. З практичного досвіду відомо, що обсяг продукції та величина витраченого виробничого ресурсу (наприклад, робочого часу) добре узгоджуються аналітичною залежністю $f(x) = ax^b$ $(a > 0;\ 0 < b \le 1)$, тобто ВФ у цьому випадку відображена названою формулою. Величини *a, b* – параметри ВФ $f(x)$ у кожному конкретному випадку обчислюються за експериментальними даними. Значення ВФ можна обчислити, наприклад, за допомогою калькулятора, але для цього потрібно багато часу і при цьому дуже легко отдержати помилковий результат.

Використовуючи математичні функції Excel, цю задачу вирішити дуже легко і, крім того, можна відобразити результати обчислень у табличному вигляді (рис. 21.16).

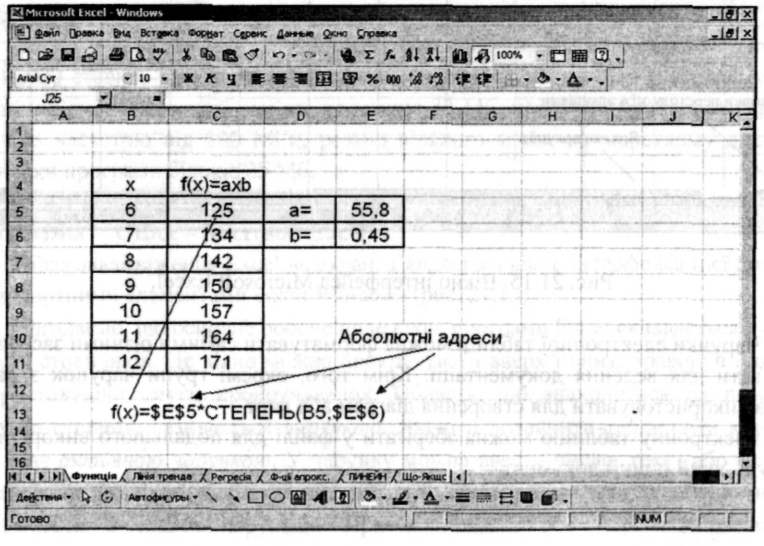


Рис. 21.16. Приклад використання функції СТЕПЕНЬ.

А при необхідності можна простежити на діаграмі залежність обсягу продукції від витраченого на її виробництво робочого часу (рис. 21. 17).

Рис. 21.17. Графік функції *f*(*x*).

Функції Excel – це вбудовані інструменти, які використовуються у формулах. Функції дозволяють:

– спрощувати формули;

– виконувати по формулах такі обчислення, які не можна виконати в інший спосіб;

– прискорювати виконання деяких задач редагування;

– виконувати "умовне" обчислення по формулах, реалізуючи прості алгоритми прийняття рішень.

У поставку Excel входять понад 300 функцій, які згруповані по категоріях – усього десять категорій:

– математичні / тригонометричні;

– інженерні;

– логічні;

– статистичні;

– функції роботи з базами даних;

– фінансові;

– інформаційні;

– текстові;

– функції роботи з датою і часом;

– функції посилань та автопідстановки.

Для розв'язання задач математичного моделювання у пригоді можуть стати такі математичні функції, як СУММ, СУМЕСЛИ, ПРОИЗВЕД, СУММПРОИЗВ, СТЕПЕНЬ (рис. 21. 16), КОРЕНЬ, ОКРУГЛ та інші, а також тригонометричні функції (SIN, COS і т. ін.), які входять до цієї категорії.

При будь-яких економічних дослідженнях поєднуються теоретичні розробки, представлені математичними моделями, та практичний досвід, віддзеркалений у статистичних даних. Щоб одержати такі дані, треба зібрати й упорядкувати необхідну інформацію, виконавши спеціальні експерименти й обробивши відповідні статистичні матеріали. Для цього в Excel є широкий набір статистичних функцій, наприклад, МАКС, МИН, СРЗНАЧ, СРОТКЛ, КОРРЕЛ, ЛИНЕЙН та ін.

Формули, які містять у собі функції, задають дії з величинами (аргументами) у вказаному порядку (називається синтаксисом функції).

Аргументами функції можуть бути:

- числові значення;
- текст;
- логічні величини;
- масиви;
- значення помилок;
- посилання, імена чарунок або діапазону чарунок;
- інша функція.

Написання функції починається з імені функції, а потім у круглих дужках вказуються аргументи (від одного до кількох):

ІМ'Я ФУНКЦІЇ (АРГУМЕНТИ).

Формула разом з функцією може мати різні посилання та будь-які оператори. Отже, можна створювати навіть дуже складну формулу.

У процесі роботи з Excel немає необхідності вивчати опис окремої функції. Програма має так званий **Мастер функций** (рис. 21.18). Запустити його можна одним із таких способів:

- вибрати команду **Вставка – Функция**;
- натиснути кнопку **Вставка функции** f_x, яка знаходиться на панелі інструментів *Стандартная*;
- натиснути комбінацію клавіш [Shift+F3].

У розділі *Категория* вікна *Мастер функций* подається список категорій функцій. При виборі однієї з них у списку *Функция* з'являється перелік функцій, які входять у цю категорію.

Слід звернути увагу, що при виборі функції у списку *Функция* унизу діалогового вікна з'являється список її аргументів разом із коротким описом.

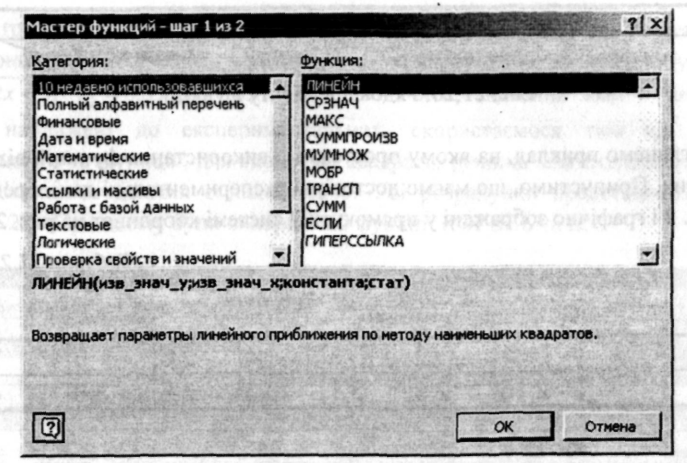


Рис. 21.18. Діалогове вікно *Мастер функций.*

Коли знайдено потрібну функцію, треба натиснути кнопку *ОК*. З'явиться вікно введення аргументів функції (рис. 21.19).

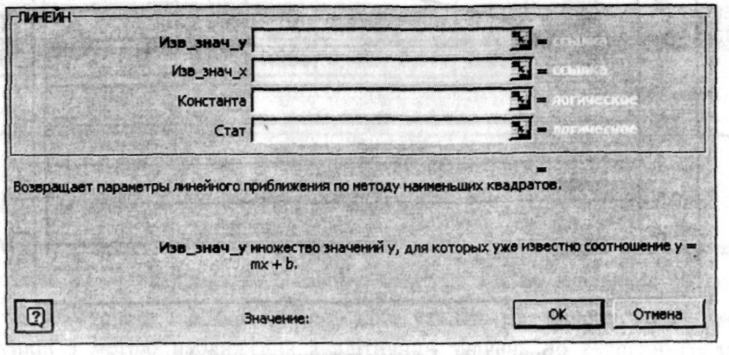


Рис. 21.19. Вікно введення аргументів функції.

Якщо це вікно закриває дані на аркуші, треба клацнути по кнопці згортання діалового вікна у правій частині поля введення аргумента. Вікно згорнеться в один рядок (рис. 21.20), після цього можна вибрати в таблиці аргумент.

Повторне натискання кнопки згортання повертає вікно введення аргументів функції. Для редагування функції необхідно зробити чарунку з функцією активною і в рядку формул натиснути кнопку .

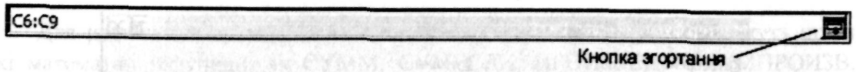


Рис. 21.20. Рядок аргументу функції.

Розглянемо приклад, на якому простежимо використання функцій від простих до складних. Припустимо, що маємо достовірні експериментальні дані, представлені в табл. 21. 2 і графічно зображені у прямокутній системі координат на рис. 21.21.

Таблиця 21.2.

Експериментальні дані

X	1	2	3	4	5	6	7
Y	2,15	2,8	3,1	3,3	3,5	3,8	4

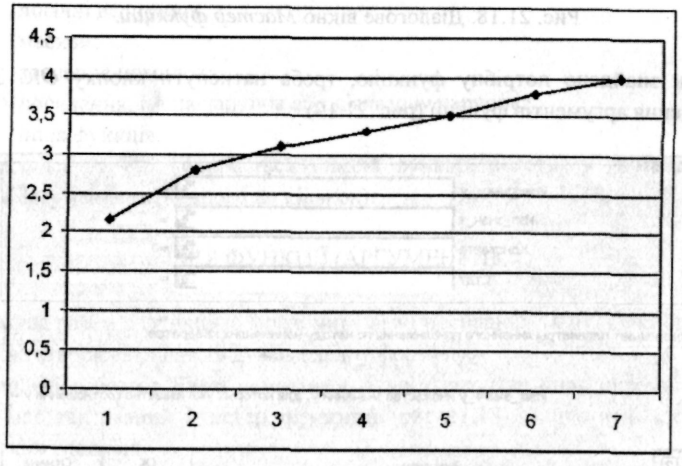
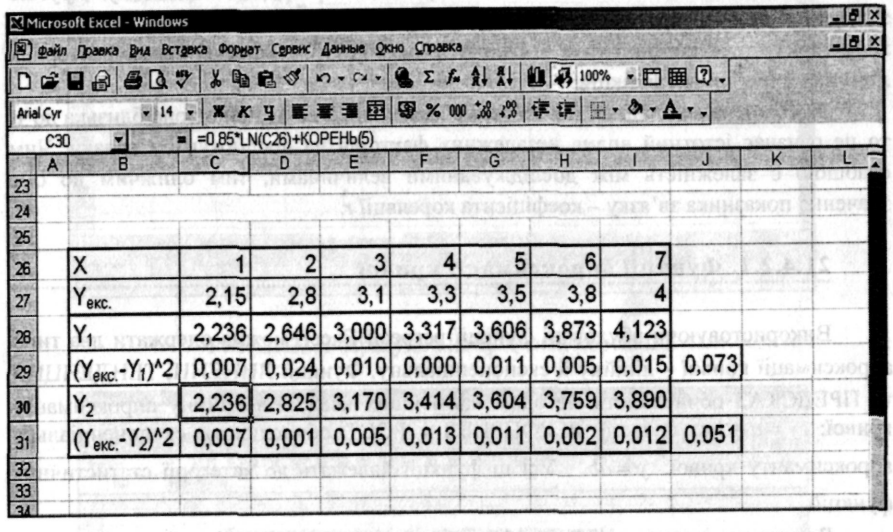


Рис. 21.21. Графік залежності $y=f(x)$.

Необхідно побудувати аналітичне зображення виробничої функції, тобто виконати задачу апроксимації експериментальних даних аналітичним зображенням. Аналізуючи графік, доходимо висновку, що аналітичне зображення відповідної функції доцільно шукати у вигляді $y_1 = \sqrt{ax+b}$ або $y_2 = a\ln x + b$. Ці види функцій за відповідного вибору параметрів a та b мають такий же вигляд залежності від свого аргумента, як і наявні експериментальні дані. Припустимо, що для обчислення параметрів вибраного виду функцій скористалися методом найменших квадратів.

Параметри a і b для функції кожного виду матимуть свої значення, а для запропонованого прикладу зображення конкуруючих функцій буде таким: $y_1 = \sqrt{2x+3}$, $y_2 = 0{,}85\ \ln x + \sqrt{5}$. Щоб визначити, значення якої з цих функцій більш наближені до експериментальних, скористаємося тим же критерієм: найкращою є та функція, для якої сума квадратів різниць між експериментальними та аналітичними даними є найменшою. Усі розрахунки представлені в таблиці (рис. 21.22). При цьому були використані функції КОРЕНЬ та LN.



X	1	2	3	4	5	6	7	
$Y_{екс.}$	2,15	2,8	3,1	3,3	3,5	3,8	4	
Y_1	2,236	2,646	3,000	3,317	3,606	3,873	4,123	
$(Y_{екс.}-Y_1)^2$	0,007	0,024	0,010	0,000	0,011	0,005	0,015	0,073
Y_2	2,236	2,825	3,170	3,414	3,604	3,759	3,890	
$(Y_{екс.}-Y_2)^2$	0,007	0,001	0,005	0,013	0,011	0,002	0,012	0,051

Рис. 21.22. Розрахунки по методу найменших квадратів.

Аналіз розрахунків показує, що серед двох конкуруючих функцій за кращу треба взяти $y_2 = 0{,}85\ln x + \sqrt{5}$. У цьому прикладі ми не вдавалися в подробиці вибору параметрів a і b, але можемо собі уявити, що використання численних розрахунків з метою наближення аналітичних значень до експериментальних – процес дуже довгий і складний. Але в Excel є спеціальні функції та надбудови, які значно спрощують задачу апроксимації даних.

21.4.2. Апроксимація даних

Існує декілька способів використання Excel для апроксимації:

- вбудовані функції регресії;
- інструменти пакета аналізу;

- лінії тренда;
- надбудова **Поиск решения** для більш точного підбору параметрів нелінійного рівняння.

Засоби апроксимації даних Excel базуються на поняттях кореляційно-регресійного аналізу. Нагадаємо, що рівняння, одержані для характеристики залежності впливаючих факторів на досліджувану ознаку, в статистиці називаються кореляційними, або рівняннями регресії. Вибір виду рівняння та визначення його параметрів є однією з основних задач кореляційно-регресійного аналізу. Другою задачею є оцінка міри впливу незалежних факторів на досліджувану ознаку, яка характеризується коефіцієнтом кореляції *r* для однофакторних і багатофакторних лінійних виробничих функцій.

Абсолютна величина *r* змінюється в межах від 0 до 1. Якщо вона близька до 1, то це означає істотний вплив незалежних факторів на досліджувану ознаку. Чим слабшою є залежність між досліджуваними величинами, тим ближчим до 0 є значення показника зв'язку – коефіцієнта кореляції *r*.

21.4.2.1. Функції апроксимації кривої

Використовуючи вбудовані функції регресії Excel, можна одержати два типи апроксимації кривої – лінійну й експоненціальну. Функції ЛИНЕЙН, ТЕНДЕНЦИЯ та ПРЕДСКАЗ обчислюють методом найменших квадратів лінійну апроксиманту кривої: $y = a + bx$, функції ЛГРФПРИБЛ і РОСТ обчислюють експоненціальну апроксиманту кривої: $y = ab^x$. Усі ці функції належать до категорії статистичних функцій.

Розглянемо використання цих функцій на нашому прикладі (табл. 21.2).

Спочатку обчислимо значення коефіцієнта кореляції *r*, щоб виявити тісноту зв'язку між *x* і *y*. Для цього розташуємо наші дані на робочому аркуші так, як показано на рис. 21.23, і скористаємося статистичною функцією КОРРЕЛ.

Коефіцієнт кореляції *r=0,977* свідчить про дуже тісний зв'язок між двома наборами даних.

Тепер для одержання лінійного рівняння знайдемо його параметри *a* і *b*, використовуючи функцію ЛИНЕЙН.

Для цього необхідно:

– визначити на робочому аркуші місце для таблиці коефіцієнтів регресії та статистики (рис. 21.24);

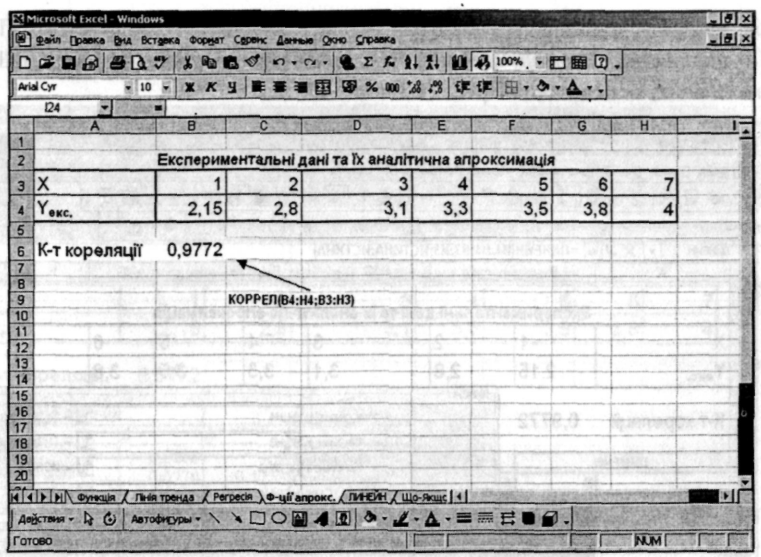


Рис. 21.23. Розрахунок коефіцієнта кореляції.

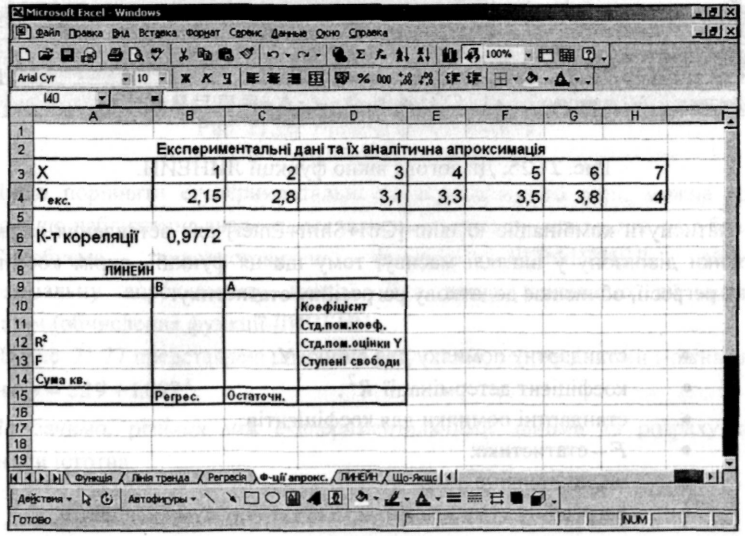


Рис. 21.24. Підготовка таблиці для розрахунку функції ЛИНЕЙН.

– виділити діапазон чарунок $B10:C14$;

– викликати *Мастер функций* і в діалоговому вікні функції ЛИНЕЙН заповнити поля, як показано на рис. 21.25;

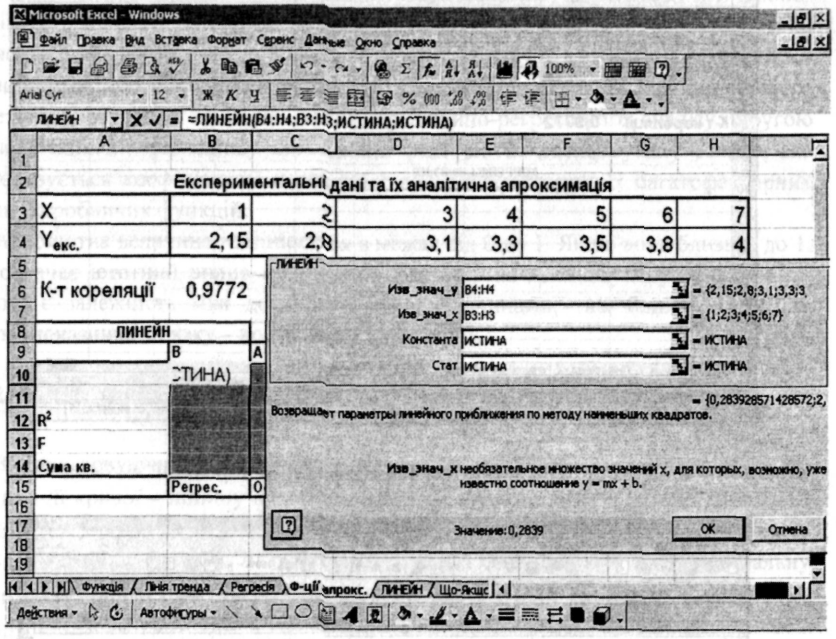


Рис. 21.25. Діалогове вікно функції ЛИНЕЙН.

– натиснути комбінацію клавіш [Ctrl+Shift+Enter] для вставлення формули у всі чарунки діапазону у вигляді масиву, тому що ця функції, окрім коефіцієнтів рівняння регресії, обчислює додаткову регресійну статистику:

- стандартну помилку для оцінки Y;
- коефіцієнт детермінації R^2;
- стандартні помилки для коефіцієнтів;
- F – статистика;
- число степенів волі;
- регресійну суму квадратів та остаточну суму квадратів.

По одержаних результатах виконання функції ЛИНЕЙН можна скласти рівняння $y = 2{,}1 + 0{,}283x$ та обчислити аналітичне значення функції (рис. 21.26).

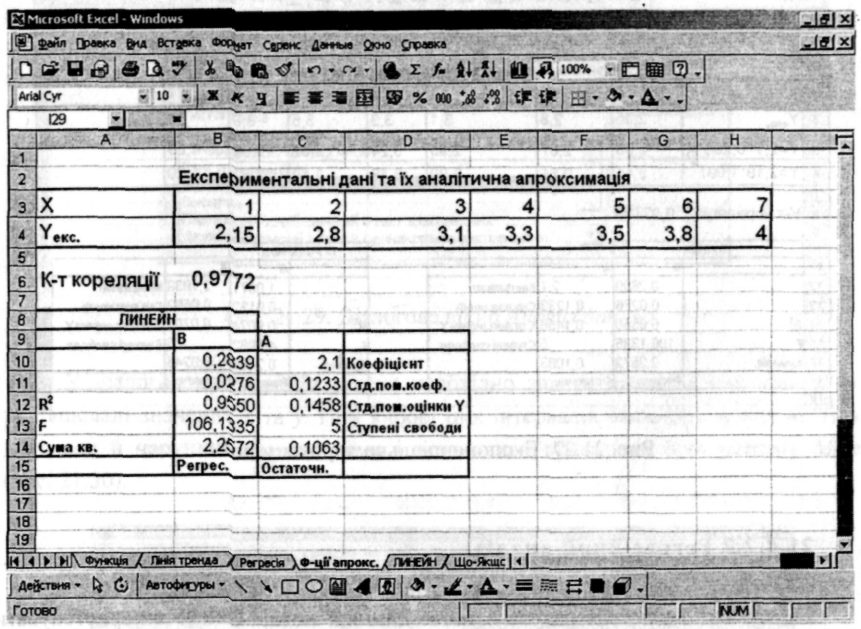


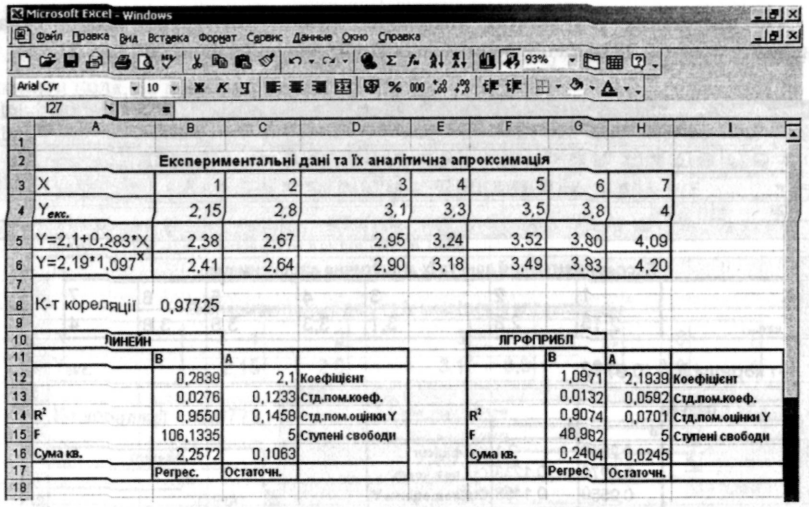
Рис. 21.26. Лінійна апроксимація.

Якщо порівняти експериментальні та апроксимовані дані, можна зробити висновок, що наближення аналітичних значень до експериментальних не є бажаним.

Спробуємо, використовуючи функцію ЛГРФПРИБЛ, обчислити експоненціальну апроксиманту кривої. Для цього виконаємо дії аналогічні попереднім (обчислення функції ЛИНЕЙН).

На рис. 21.27 представлені одержані результати, згідно з якими можна скласти рівняння $y = 2{,}19 + 1{,}097^x$.

Як бачимо, різниця між експериментальними даними та розрахунковими значеннями істотна.



	A	B	C	D	E	F	G	H
2	Експериментальні дані та їх аналітична апроксимація							
3	X	1	2	3	4	5	6	7
4	$Y_{екс.}$	2,15	2,8	3,1	3,3	3,5	3,8	4
5	Y=2,1+0,283*X	2,38	2,67	2,95	3,24	3,52	3,80	4,09
6	$Y=2,19*1,097^X$	2,41	2,64	2,90	3,18	3,49	3,83	4,20
8	К-т кореляції	0,97725						

10	ЛИНЕЙН			
11		B	A	
12		0,2839	2,1	Коефіцієнт
13		0,0276	0,1233	Стд.пом.коеф.
14	R^2	0,9550	0,1458	Стд.пом.оцінки Y
15	F	106,1335	5	Ступені свободи
16	Сума кв.	2,2572	0,1063	
17		Регрес.	Остаточн.	

	ЛГРФПРИБЛ		
	B	A	
	1,0971	2,1939	Коефіцієнт
	0,0132	0,0592	Стд.пом.коеф.
R^2	0,9074	0,0701	Стд.пом.оцінки Y
F	48,982	5	Ступені свободи
Сума кв.	0,2404	0,0245	
	Регрес.	Остаточн.	

Рис. 21.27. Експоненціальна апроксимація.

21.4.2.2 Регресійний аналіз

Інструмент *Регрессия* пакета **Анализ данных** можна також використовувати для одержання лінійної апроксимації аналогічно функції ЛИНЕЙН. Будуть одержані ті ж результати, але цей інструмент обчислює набагато більше статистичних параметрів. Скористатись інструментом *Регрессия* можна, якщо встановити пакет аналізу командою **Сервис – Надстройки** (рис. 21.28).

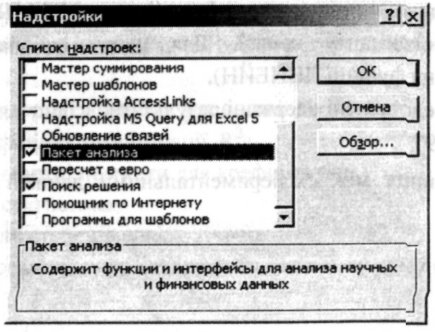


Рис. 21.28. Діалогове вікно *Надстройки*.

Після встановлення пакета аналізу треба вибрати команду **Сервис – Анализ данных** і необхідний інструмент *Регрессия* (рис. 21.29).

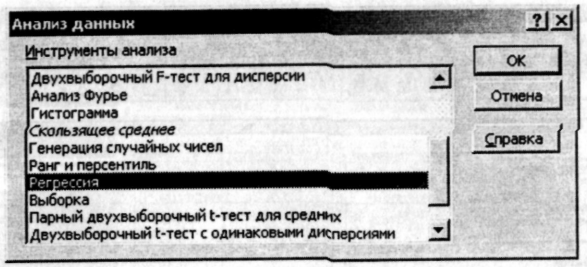


Рис. 21.29. Діалогове вікно *Анализ данных*.

У діалоговому вікні *Регрессия* необхідно задати діапазон чарунок, в яких розташовані значення *x* та *y*. При виборі цих інтервалів включити в них не тільки дані, але й чарунки з заголовками і встановити прапорець біля режиму *Метки* (рис. 21.30).

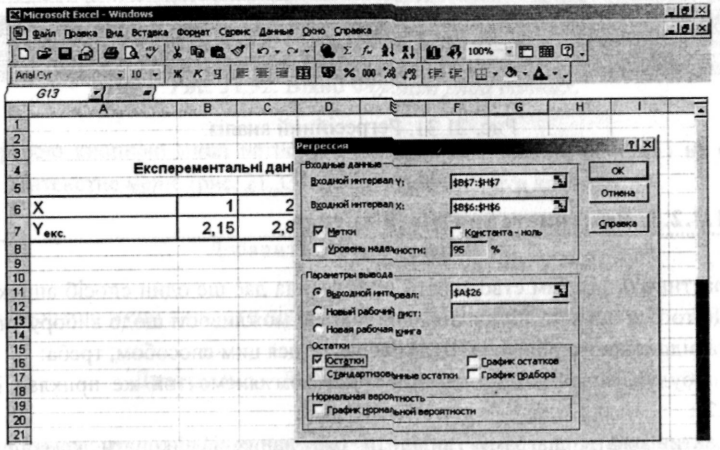


Рис. 21.30. Діалогове вікно *Регрессия*.

Результати регресійного аналізу показані на рис. 21.31. Одержані коефіцієнти *a* і *b* мають такі ж значення, як і при використанні функції ЛИНЕЙН. Але між двома цими способами є різниця. При використанні функції ЛИНЕЙН її значення

автоматично перераховуються при кожній зміні даних, в той час як інструмент аналізу *Регрессия* необхідно кожного разу запускати вручну.

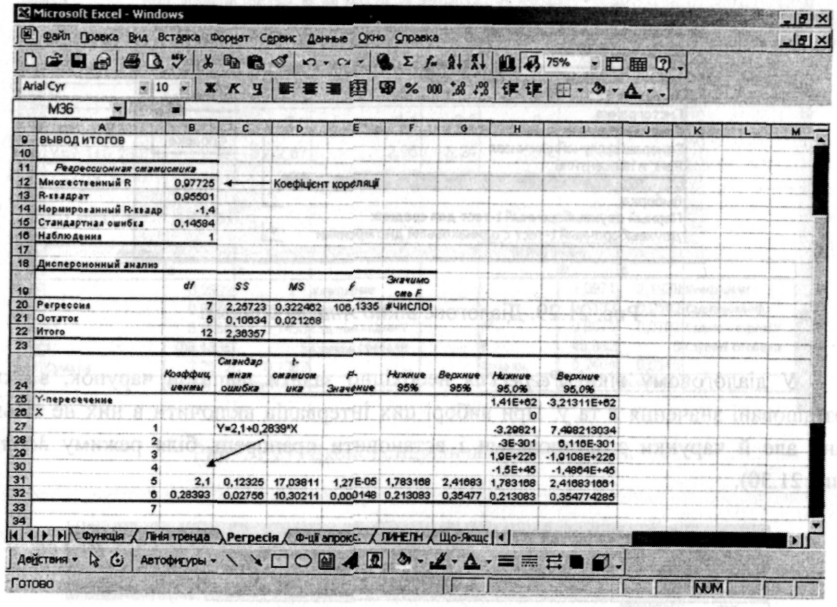


Рис. 21.31. Регресійний аналіз.

21.4.2.3. Використання ліній тренда

Властивість діаграм створювати лінії тренда дає ще один спосіб апроксимації даних. До того ж цей спосіб має більш поширені можливості щодо вибору типу лінії тренда та відповідного рівняння. Щоб скористатися цим способом, треба:

– побудувати діаграму типа *Точечная* (розглянемо той же приклад і дані з табл. 21.2);

– активізувати діаграму, виділити ряд даних і виконати команду меню **Формат – Выделенный ряд**;

– у вікні *Формат ряда данных* (рис. 21.32) на укладинці *Вид* у групі *Линия* поставити прапорець біля режиму

✓ *отсутствует*, тобто відключити відображення лінії, залишивши тільки маркери;

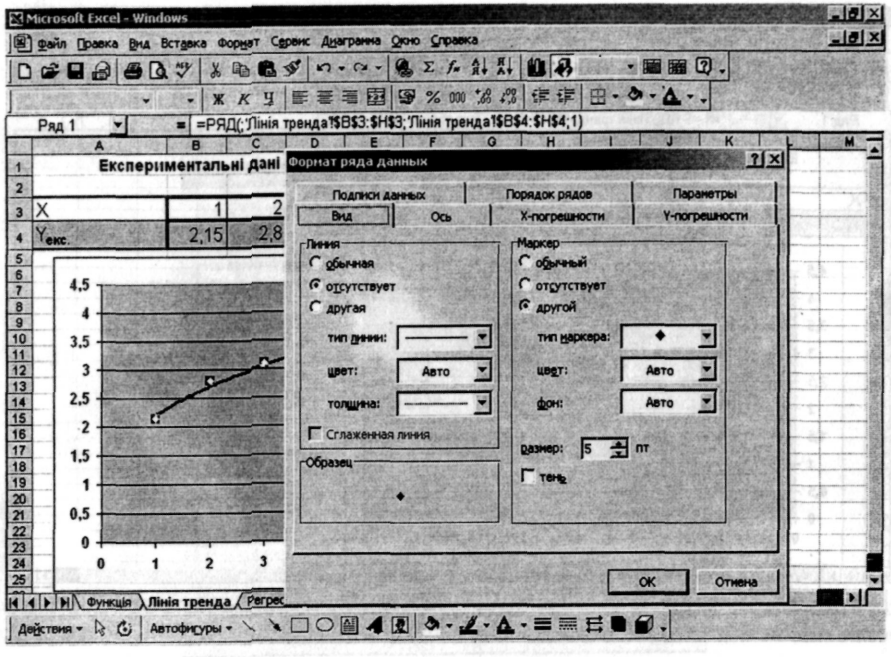


Рис. 21.32. Вікно *Формат ряда данных*.

– правою кнопкою миші натиснути на будь-який маркер даних, і на екрані з'явиться контекстне меню (рис. 21.33), де потрібно вибрати команду

Добавить линию тренда,

або лівою кнопкою миші виділити діаграму і вибрати команду меню

Диаграмма – Добавить линию тренда;

– у діалоговому вікні *Линия тренда* (рис. 21.34) на укладинці *Тип* вибрати тип лінії.

Є можливість створювати лінії типу *Линейная, Степенная, Логарифмическая, Экспоненциальная, Скользящее среднее* та будувати відповідне рівняння. Для нашого прикладу виберемо тип лінії *Степенная*;

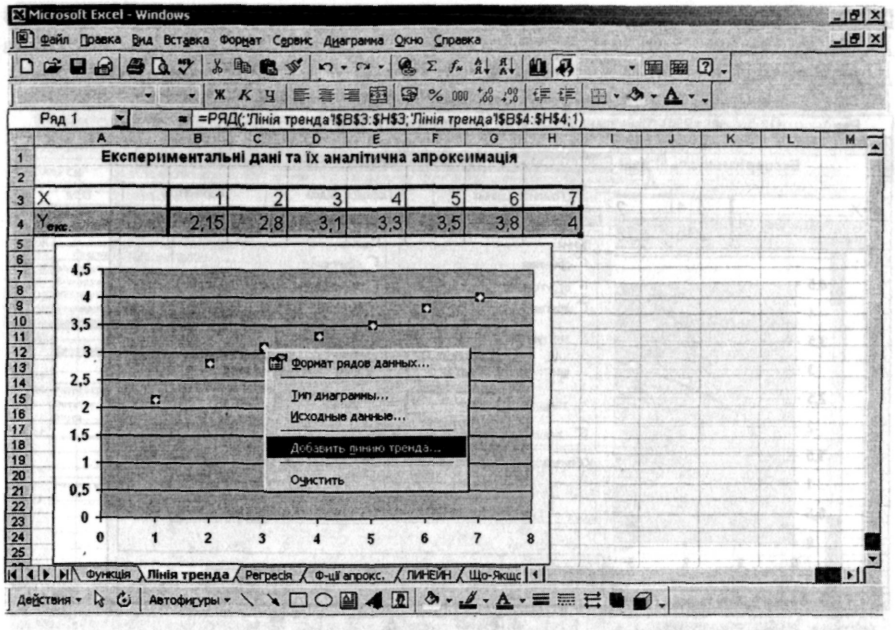


Рис. 21.33. Вікно для побудови лінії тренда.

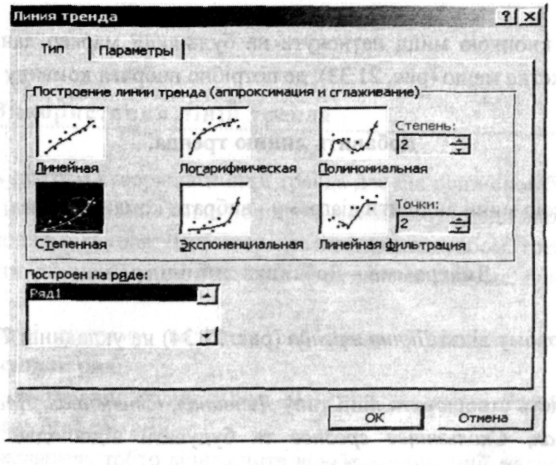


Рис. 21.34. Діалогове вікно *Линия тренда.*

– на укладинці *Параметры* (рис. 21.35) встановити прапорці біля режимів:

✓ показывать уравнение на диаграмме;

✓ поместить на диаграмму величину достоверности аппроксимации (R^2).

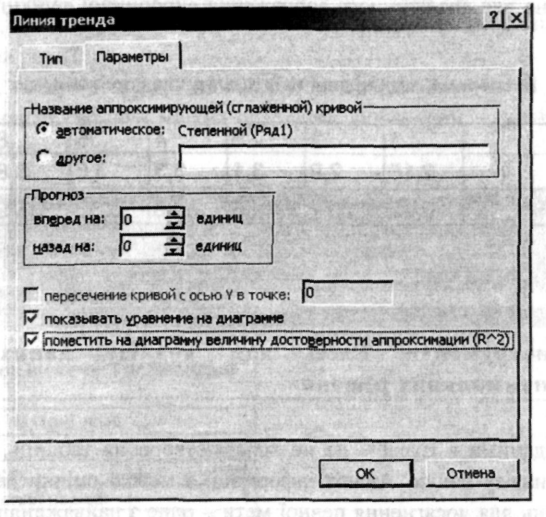


Рис. 21.35. Укладинка *Параметры.*

Графік з побудованою лінією тренда та відповідним рівнянням показано на рис. 21.36.

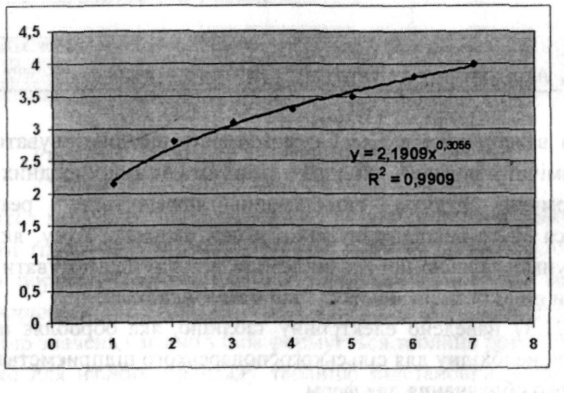


Рис. 21.36. Графік з лінією тренда.

Обчисливши значення функції згідно з одержаним рівнянням регресії $f(x) = 2{,}19x^{0{,}305}$ і порівнявши аналітичні значення функції з експериментальними, доходимо висновку, що розбіг між цими даними не є істотним і це рівняння може бути використане для аналітичного зображення виробничої функції досліджуваних величин:

Експериментальні дані та їх аналітична апроксимація

X	1	2	3	4	5	6	7
$Y_{екс.}$	2,15	2,8	3,1	3,3	3,5	3,8	4
$Y=2{,}1909*x^{0{,}3055}$	2,19	2,71	3,06	3,35	3,58	3,79	3,97

21.4.3. Інструменти Excel для аналізу даних і пошуку оптимальних рішень

Робота з даними в Excel – це не тільки створення таблиць, а й можливість повноцінного аналізу даних. Аналіз інформації з метою оцінки певних ситуацій і прийняття рішень для досягнення певної мети – одне з найважливіших завдань для сучасного керівника.

Аналіз даних в Excel – це потужний інструмент, який засновано на використанні функцій більш складних, ніж ті, які було розглянуто раніше. Excel дає можливість проведення аналізу даних від простого, наприклад, за сценарієм "що буде – якщо…", до складного – пошуку оптимальних рішень.

21.4.3.1. Аналіз за сценарієм "що буде – якщо…"

Однією з важливіших в Excel є можливість використовувати формули для створення динамічних моделей, які одразу реагують на зміну вхідних даних. Іншими *словами, формули* будуть автоматично перелічувати результат, якщо змінюватиметься зміст використовуваних у них чарунок. Тому, якщо змінювати значення в чарунках за якою-небудь системою, а потім відстежувати результати, то можна виконати один із видів аналізу: ***"що буде – якщо…"***.

На рис. 21.37 наведено електронну таблицю, яка обробляє інформацію про іпотечну позику, необхідну для сільськогосподарського підприємства, наприклад, на придбання нового обладнання для ферм.

Аркуш розділено на дві секції, в яких розташовані чарунки з вхідними даними і чарунки результатів. За допомогою цієї електронної таблиці легко відповісти на питання "що буде – якщо…", подібні таким:

– що буде, якщо обладнання буде придбано за нижчою вартістю?

– що буде, якщо кредитор запросить перший внесок у розмірі 15%?

– що буде, якщо керівник підприємства зможе домовитися взяти позику не на один рік, а на два роки?

– що буде, якщо відсоток річних знизиться до 7,5?

Відповіді на ці питання можна одержати, якщо просто підставляти потрібні значення для вхідних даних.

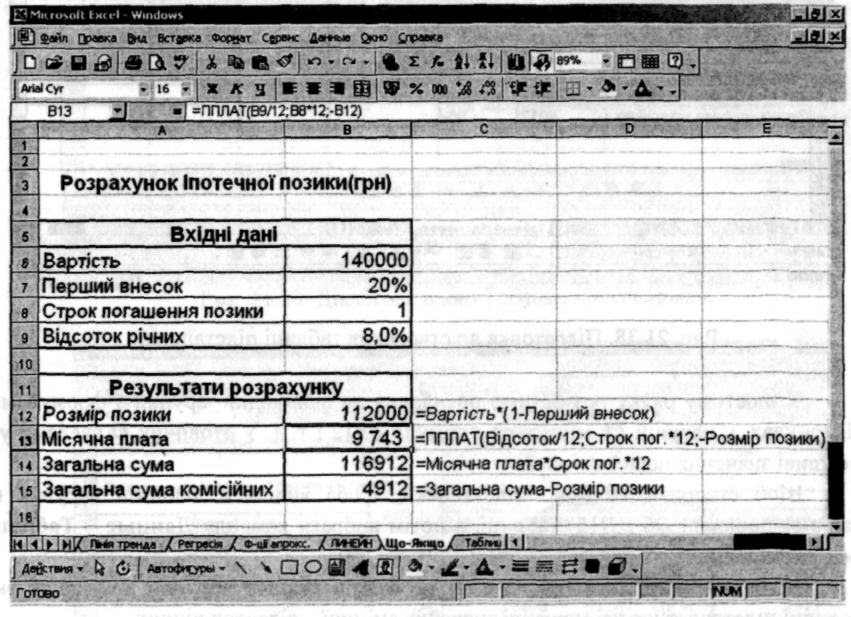


Рис. 21.37. Приклад аналізу "*що буде – якщо…*".

При роботі з моделлю "що буде – якщо…" у відповідний момент часу можна використовувати тільки один сценарій (тільки один набір даних). Але що робити, якщо потрібно порівняти результати декількох сценаріїв? Для цього в Excel є можливість створювати так звані ***Таблицы подстановки***, коли для однієї змінної створюється набір значень і згідно з цим формується таблиця результатів.

Побудуємо для нашого прикладу таблицю підстановки, якщо змінною буде відсоток річних – від 7% до 9% з кроком 0,25% (рис. 21.38).

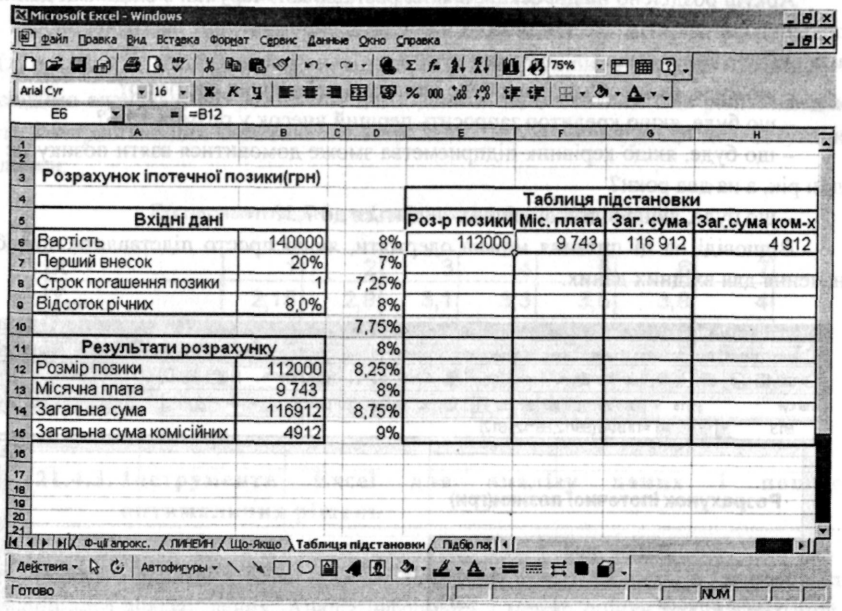


Рис. 21.38. Підготовка до створення таблиці підстановки.

У шостому рядку розмістимо посилання на відповідні чарунки з формулами. Наприклад, у чарунці *E*6 міститься формула =*B*12 і т. д. У стовпчик *D* введемо усі можливі значення нашої змінної.

Щоб створити таблицю підстановки, треба виділити діапазон чарунок (у нашому прикладі *D*6 : *H*15) таблиці, а потім вибрати команду **Данные – Таблица подстановки…**

У діалоговому вікні (рис. 21.39) необхідно вказати чарунку аркуша, в яку повинні підставлятися усі можливі значення змінної – відсоток річних.

Оскільки всі вхідні дані нашої змінної знаходяться у стовпчику, ім'я чарунки *B*9 в діалоговому вікні таблиці підстановки потрібно ввести у поле *Подставлять значения по строкам в:* та натиснути кнопку *OK*.

Якщо розглянути чарунки таблиці, в яких знаходяться результати після виконання команди **Данные – Таблица подстановки…**, можна відзначити, що програма Excel розмістила тут не прості формули, а формули масивів, в яких використовується функція ТАБЛИЦЯ (рис. 21.40).

Цікаво, що набір значень для змінної можна ввести інший, одержавши таблицю підстановки з перерахованими результатами.

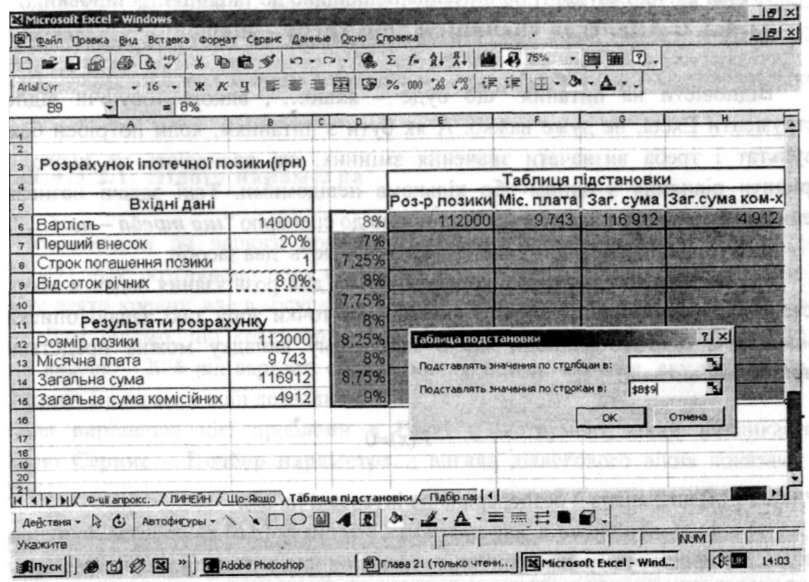


Рис. 21.39. Діалогове вікно *Таблиця подстановки.*

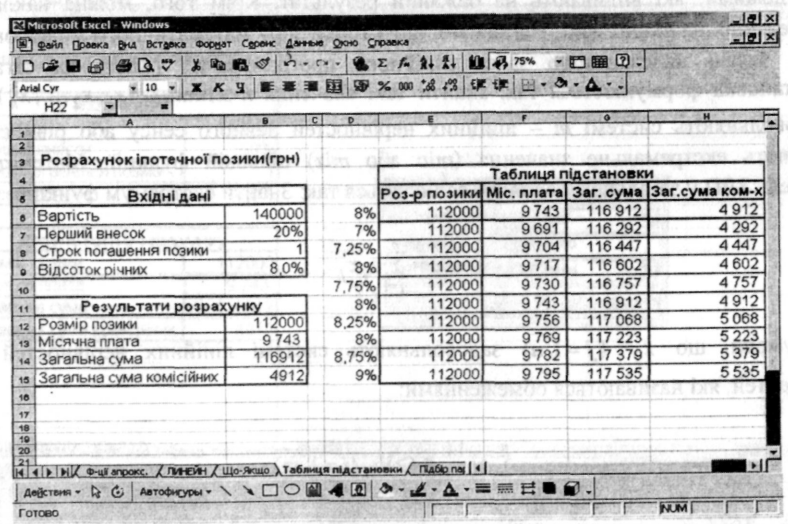


Рис. 21.40. Результат обчислення відповідно до таблиці підстановки.

21.4.3.2. Аналіз за сценарієм "що треба – щоб..."

Відповісти на питання "що буде – якщо...", використовуючи відповідні інструменти Excel, не дуже важко. А як бути з питанням, коли потрібен бажаний результат і треба визначати значення змінних, які впливають на нього, тобто вирішити рівняння з одним або кількома невідомими. Такі задачі називаються оберненими, а аналіз виконується відповідно до сценарію "***що треба – щоб...***".

В Excel для проведення такого аналізу існують два засоби:

– **Подбор параметра** використовується для розв'язання задач в оберненій постановці з одним невідомим. З формальної точки зору такі задачі описуються рівнянням з однією змінною, яке в загальному випадку можна представити в канонічному вигляді:

$$F(x)=0$$

(функція $F(x)$ визначена й безперервна на інтервалі $[a, b]$).

Будь-яке число x, яке належить інтервалу $[a,b]$ та обертає функцію $F(x)$ на нуль, називається коренем рівняння. Таким чином, можна сказати, що інструмент *Подбор параметра* використовується для визначення кореня рівняння x;

– **Поиск решения** використовується для розв'язання задач з кількома невідомими, які впливають на бажаний результат. Крім того, можна накладати обмеження на вхідні дані. Нагадаємо, що з точки зору математичного моделювання такі задачі належать до задач лінійного програмування (ЗЛП), їх загальна постановка формулюється так: знайти такі значення n-змінних $(x_1, x_2, ..., x_n)$, які задовольняють системі m – лінійних нерівностей певного сенсу або рівностей і надають екстремальне значення (*min* або *mix*) цільовій функції. Математична модель ЗЛП у загальному випадку записується так: Знайти екстремум функції

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

за умови, що x_j $(j=\overline{1,n})$ задовольняють системі лінійних нерівностей або рівностей, які називаються обмеженнями:

$$\sum_{j=1}^{n} a_{ij} x_j \le (\ge) b_i \quad (i=\overline{1,m}).$$

Приймається, що c_j, b_i, a_{ij} – відомі сталі, а в кожному з обмежень наявний

лише один знак $\le$, $=$ або $\ge$. Величина n і m кількісно між собою не пов'язані. У багатьох задачах змінні x_j мають задовольняти вимозі невід'ємності, тобто $x_j \ge 0$ $(j=\overline{1,n})$. Набір числових значень величин $(x_1, x_2, ..., x_n)$ називається планом задачі.

21.4.3.2.1. Підбір параметра

Повернемося до нашого прикладу з розрахунком іпотечної позики (див. рис. 21. 37). Але розглянемо іншу ситуацію. Вам необхідно нове обладнання, і Ви плануєте взяти кредит, але в місяць Ви в змозі погашати не більше 5000 грн. взятої позики. Вам відомо, що кредитор визначив фіксований відсоток річних (8%), а перший внесок – 20% від вартості обладнання. Виникає питання, яку максимальну суму позики Ви можете собі дозволити.

Для вирішення цієї проблеми в Excel є інструмент, який викликається командою **Сервис – Подбор параметра...**; вигляд діалогового вікна показано на рис. 21.41.

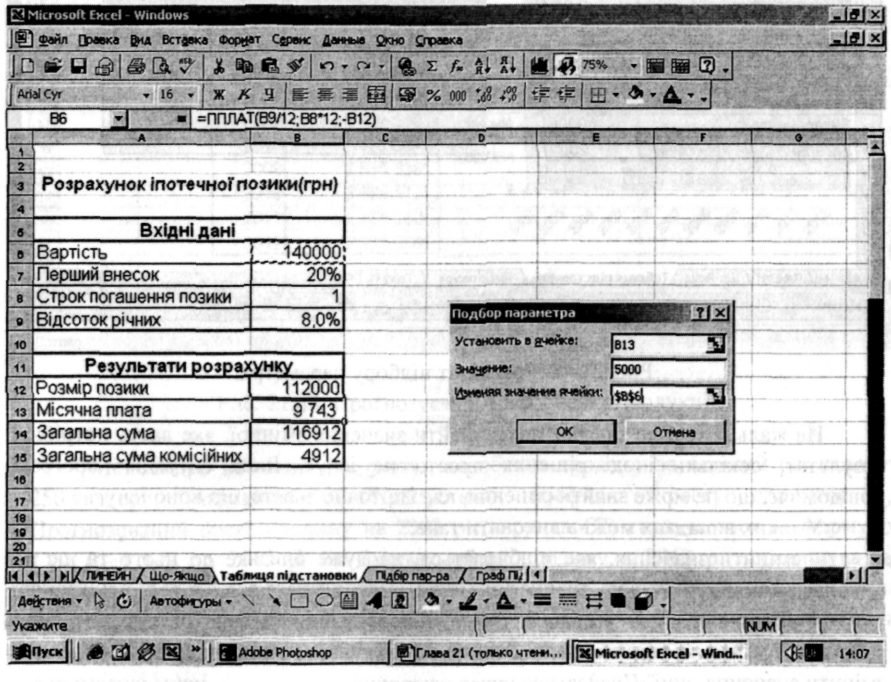


Рис. 21.41. Діалогове вікно *Подбор параметра.*

У відповіді поля цього вікна треба ввести необхідні значення для одержання бажаного результату (рис. 21.42).

Як бачимо, можна дозволити собі придбати обладнання на суму не більше 71849 грн., при цьому 20% від вартості сплатити одразу, а позику у розмірі 57479 грн. погасити протягом року, сплачуючи по 5000 грн. щомісячно.

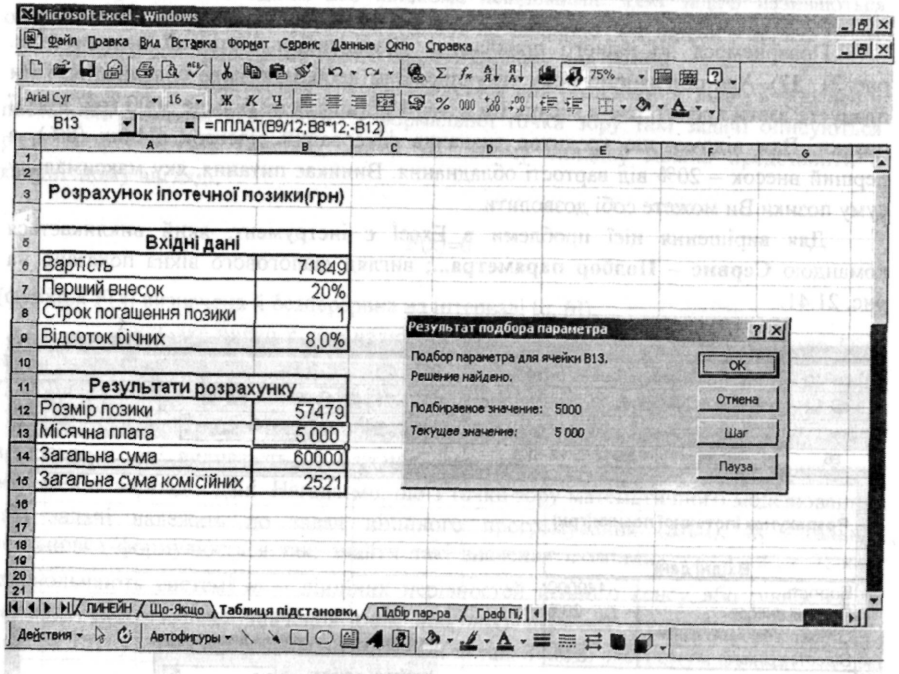


Рис. 21.42. Результат підбору параметра.

На жаль, Excel не завжди може знайти значення змінної, яке дасть потрібний результат, оскільки іноді рішення просто не існує. Інша ситуація: програма повідомляє, що не може знайти рішення, але Ви точно знаєте, що воно існує.

У таких випадках можна виконати таке:

– змінити значення, яке підбирається, на дуже близьке до нього та ще раз виконати команду;

– збільшити число ітерацій, тим самим збільшивши ймовірність потрібного рішення. Треба виконати команду **Сервис – Параметры** і на укладинці *Вычисления* змінити значення опції *Предельное число итераций.*

21.4.3.2.2. Графічний підбір параметра

Excel має й інший спосіб проведення підбору параметра – за допомогою маніпулювання діаграмами. На рис. 21.43 показано робочий аркуш, на якому відображено прогнозований обсяг продажу компаній, працюючих у цій галузі, який може збільшуватися відповідно до показового закону $y \cdot b^x$ (y – обсяги продажу за перший рік; b – коефіцієнт росту; x – змінна часу).

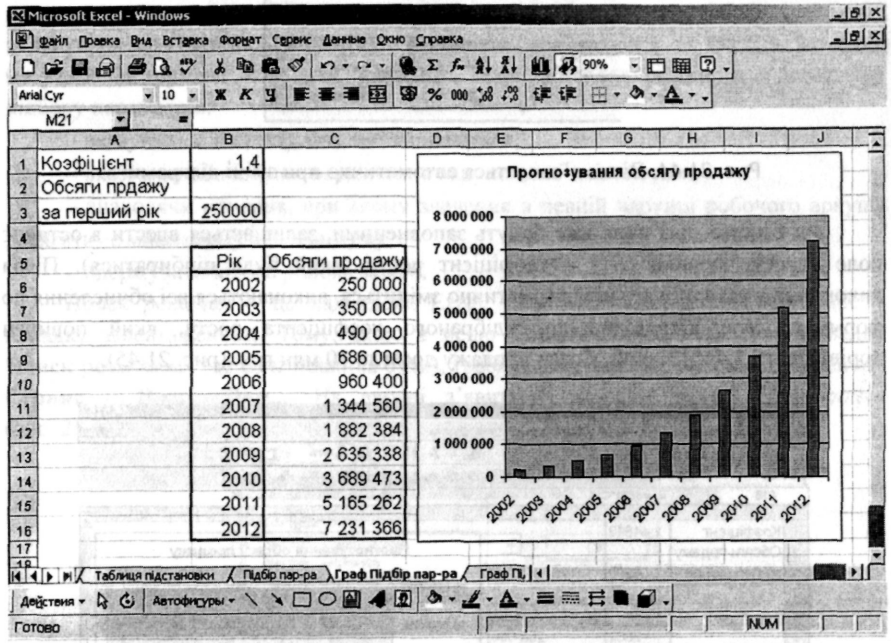


Рис. 21.43. Прогнозуємий зріст обсягу продукції.

Менеджери компанії планують, що обсяги продажу за перший рік складуть 250 тис. грн. (чарунка *C*6). Припустимо, коефіцієнт росту дорівнює 1,4. Прогнозований обсяг продажу на 2003 р. (чарунка *C*7) визначається за формулою $C6 \times B1$ і т.д. до 2012 р. Як бачимо, обсяг продажу в 2012 р. перевищуватиме 7 млн грн. На скільки треба збільшити коефіцієнт росту, щоб довести обсяги продажу до 10 млн грн.? Для розв'язання цієї задачі побудуємо гістограму по змінній "обсяг продажу" (рис. 21.43). Виділимо цю діаграму для редагування за допомогою миші.

Виберемо для редагування стовпчики діаграми і за допомогою миші виділимо останній стовпчик. Помістимо курсор миші на вершину стовпчика, при цьому він змінить форму, і потягнемо край стовпчика вгору до потрібного значення (10млн.) у полі адреси чарунки. Відпустимо кнопку миші, і на екрані з'явиться вже відоме діалогове вікно підбору параметра (рис. 21.44).

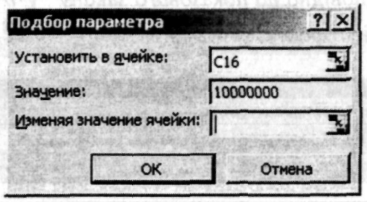


Рис. 21.44. Вікно з'являється автоматично при зміні діаграми.

Як бачимо, два поля вже будуть заповненими, залишається ввести в останнє поле адресу чарунки (*B*1) – коефіцієнт росту (який буде підбиратися). Після виконання команди діаграма автоматично зміниться, виконаються всі обчислення по формулах, але відповідно до підібраного коефіцієнта росту, який повинен дорівнювати 1,44613, щоб обсяги продажу досягли 10 млн грн. (рис. 21.45).

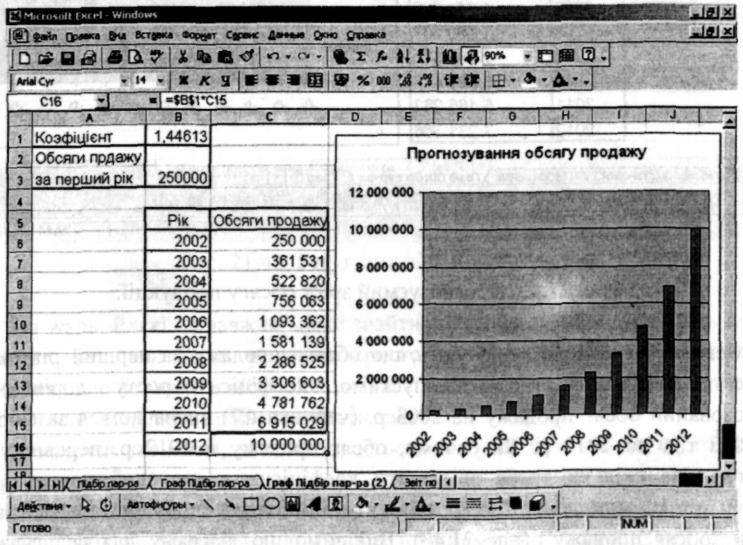


Рис. 21.45. Графічний підбір параметра.

Наведений приклад зовсім не означає, що треба завжди користуватися графічним методом підбору параметра. Але якщо його порівняти із звичайним використанням команди *Подбор параметра*, він є більш наочним.

21.4.3.2.3. Пошук рішення

Підбір параметра, безперечно, є дуже корисним інструментом Excel, але він має деякі обмеження. Наприклад, можна знаходити значення тільки для однієї чарунки, яка змінюється, і одержувати тільки одне рішення. В Excel є ще один могутній інструмент – **Поиск решения**, який дозволяє розширити процедуру підбору параметрів:

– вказуючи кілька чарунок, які змінюються;

– вказуючи обмеження на значення чарунок, які змінюються;

– знаходячи рішення, при якому значення в певній чарунці робочого аркуша досягатиме максимуму або мінімуму;

– одержувати кілька рішень задачі.

Поиск решения – це надбудова Excel, її можна використовувати лише у випадку завантаженості. Тому, якщо в меню **Сервис** Ви не знаходити команду **Поиск решения...** (де вона повинна бути), треба її завантажити, виконавши команду **Сервис – Надстройки...**. На екрані з'явиться діалогове вікно *Надстройки* (рис. 21.46).

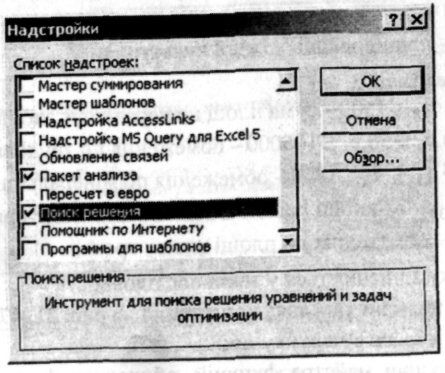


Рис. 21.46. Діалогове вікно *Надстройки*.

Встановити прапорець біля надбудови *Поиск решения*. Після натискання на кнопку *OK* її буде завантажено.

Розглянемо декілька задач.

Розподіл ресурсів. Цю задачу можна віднести до категорії ідеальних задач для розв'язання за допомогою методу пошуку рішень. Потрібно визначити структуру посівів озимої пшениці, проса та гречки, щоб прибуток від реалізації продукції був максимальним за таких умов:

– загальна площа посівів не перевищує 1500 га;

– запаси мінеральних добрив 1800 у д.р.;

– трудові ресурси 105000 люд.-год.;

– площа гречки не більше 300 га.

Нормативи затрат праці, добрив і розмір прибутку в розрахунку на 1 га посівів наведені в табл. 21.3.

Таблиця 21 3.

Показники	Одиниця виміру	Припадає на 1 га посівів		
		оз. пшениці	проса	гречки
Витрати праці	люд.-год.	75	60	56
Витрати добрив	ц	1,6	1,5	1,0
Прибуток	ум. од.	165	72	280

Для розв'язання цієї задачі складемо її математичну модель.

Знайти MAX цільової функції:

$$C = 165x_1 + 72x_2 + 280x_3$$

(де змінні x_1, x_2, x_3 - площі посівів кожної культури).

При таких *обмеженнях*:

1) $x_1 + x_2 + x_3 \le 1500$ – сума площ не перевищує загальну площу;

2) $75x_1 + 60x_2 + 56x_3 \le 105000$ – обмеження по трудових ресурсах;

3) $1{,}6x_1 + 1{,}5x_2 + x_3 \le 1800$ – обмеження по мінеральних добривах;

4) $x_1,\ x_2,\ x_3 \ge 0$ – *площі посівів є величинами позитивними;*

5) $x_3 \le 300$ – обмеження по площі висіву гречки.

Пошук розв'язку здійснюється у такій послідовності:

– розташувати вхідні дані так, як показано на рис. 21. 47;

– встановити курсор у чарунку *E*6;

– використовуючи майстра функцій, обчислити функцію СУММПРОИЗВ (рис. 21.48), де у поле Массив 1 ввести діапазон чарунок *B*5:*D*5 (розшукувані значення змінних) і натиснути клавішу [F4], щоб зробити цю адресу абсолютною, у поле Массив 2 – відповідно *B*6:*D*6 (згідно з обмеженням першим) і натиснути кнопку *OK*;

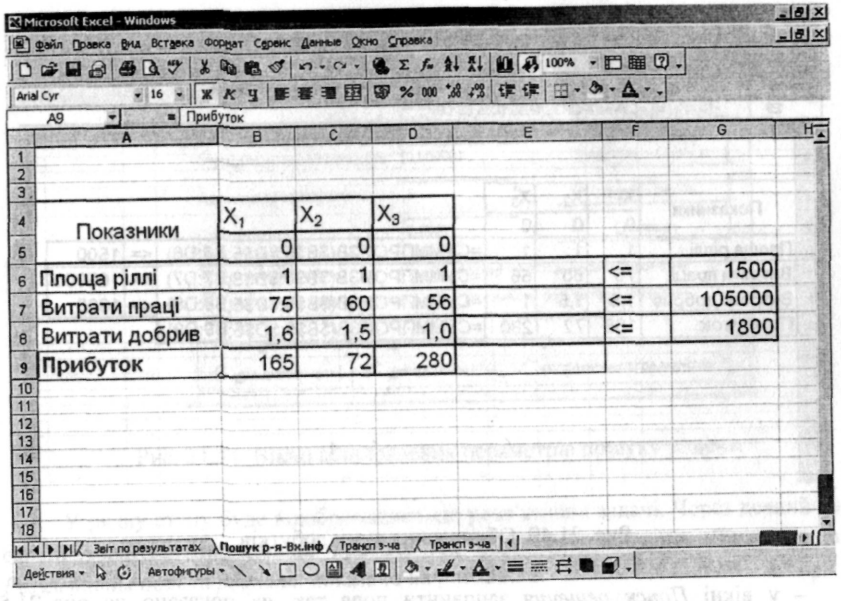


Рис. 21.47. Вхідна інформація.

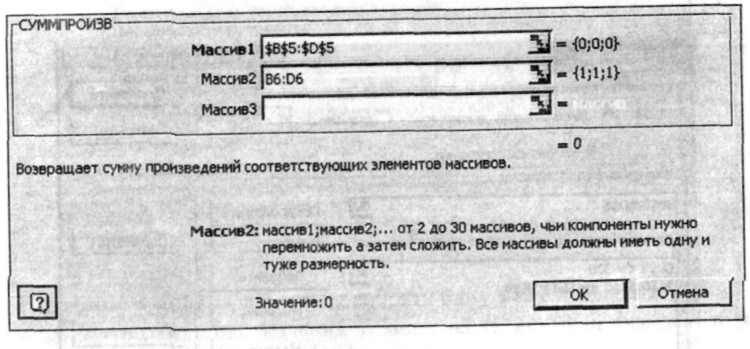
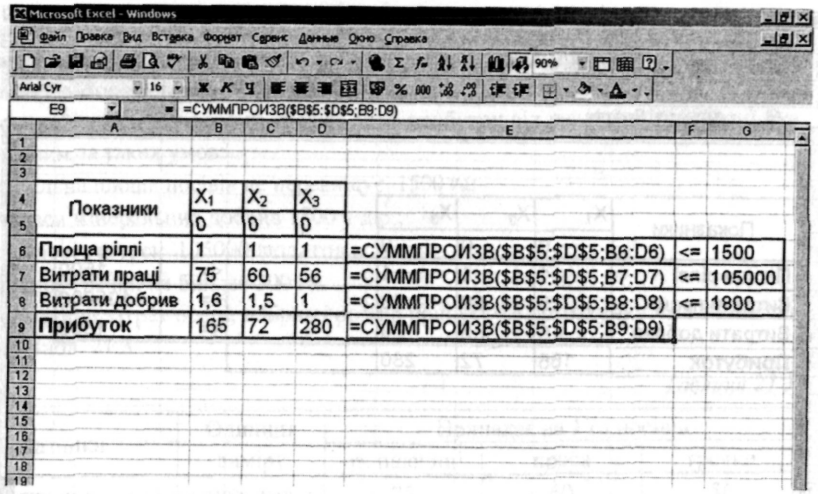


Рис. 21.48. Майстер функції **СУММПРОИЗВ.**

– продублювати цю формулу у чарунки *E*7:*E*9 (згідно з обмеженнями 2,3 і цільовою функцією) (рис. 21.49);

– установити курсор у чарунку *E*9 – цільова функція;

– вибрати команду **Сервис – Поиск решения...**;



Показники	X_1	X_2	X_3			
	0	0	0			
Площа ріллі	1	1	1	=СУММПРОИЗВ($B$5:$D$5;B6:D6)	<=	1500
Витрати праці	75	60	56	=СУММПРОИЗВ($B$5:$D$5;B7:D7)	<=	105000
Витрати добрив	1,6	1,5	1	=СУММПРОИЗВ($B$5:$D$5;B8:D8)	<=	1800
Прибуток	165	72	280	=СУММПРОИЗВ($B$5:$D$5;B9:D9)		

Рис. 21.49. Обчислення суми добутків.

– у вікні *Поиск решения* заповнити поля так, як показано на рис. 21.50 (обмеження додавати за допомогою кнопки *Добавить*) і натиснути кнопку *Параметры;*

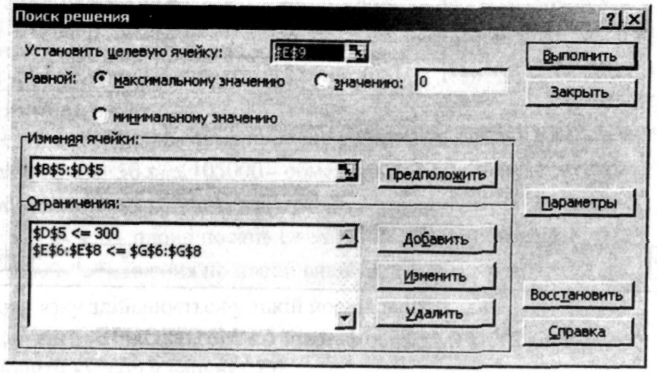


Рис. 21.50. Вікно *Поиск решения.*

– установити параметри, як вказано на рис. 21.51 і натиснути кнопку *ОК*;
– у вікні *Поиск решения* натиснути кнопку *Выполнить.*

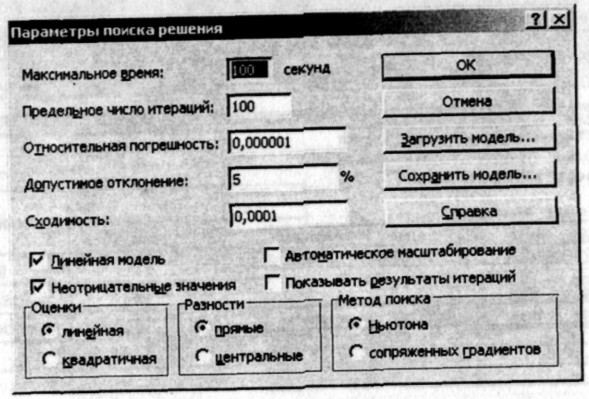


Рис. 21.51. Вікно встановлення параметрів пошуку рішень.

У рядку стану буде відображатися хід розв'язання задачі. Через деякий час на екрані з'явиться інформація про те, що розв'язок знайдено. Діалогове вікно *Результаты поиска решения* показано на рис. 21.52.

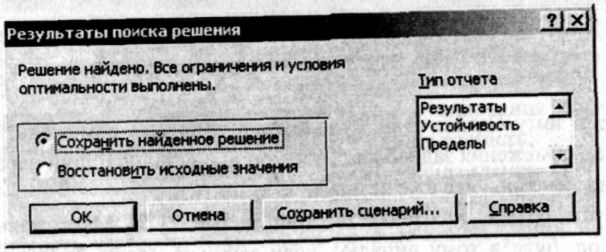


Рис. 21.52. Вікно *Результаты поиска решения.*

Після розв'язання задачі можна вибрати одну з таких можливостей:

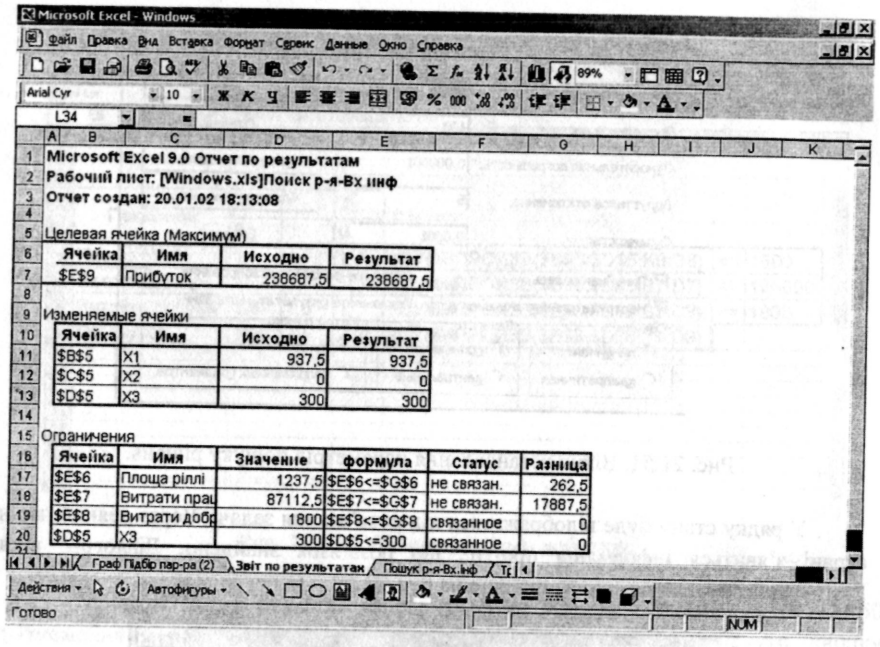
– замінити початкові значення змінних на ті, які були знайдені в результаті рішення задачі;

– поновити початкові значення змінних;

– створити кілька звітів (виділити потрібні звіти можна за допомогою клавіші [Shift]);

– натиснути кнопку *Сохранить сценарий* для збереження рішення у вигляді сценарію, який можна використовувати для розв'язання подібних задач.

Якщо треба створити звіт, він буде розташований на новому робочому аркуші, який одержить відповідне ім'я (рис. 21.53).



Microsoft Excel 9.0 Отчет по результатам
Рабочий лист: [Windows.xls]Поиск р-я-Вх.инф
Отчет создан: 20.01.02 18:13:08

Целевая ячейка (Максимум)

Ячейка	Имя	Исходно	Результат
$E$9	Прибуток	238687,5	238687,5

Изменяемые ячейки

Ячейка	Имя	Исходно	Результат
$B$5	X1	937,5	937,5
$C$5	X2	0	0
$D$5	X3	300	300

Ограничения

Ячейка	Имя	Значение	формула	Статус	Разница
$E$6	Площа ріллі	1237,5	$E$6<=$G$6	не связан.	262,5
$E$7	Витрати прац	87112,5	$E$7<=$G$7	не связан.	17887,5
$E$8	Витрати добр	1800	$E$8<=$G$8	связанное	0
$D$5	X3	300	$D$5<=300	связанное	0

Рис. 21.53. Звіт по результатах.

У розділі звіту *Ограничения* буде вказано стан усіх обмежень. Тип *связанное* означає, що це обмеження задоволено, але при цьому відповідний параметр прийняв своє граничне значення, яке вже неможливо змінити.

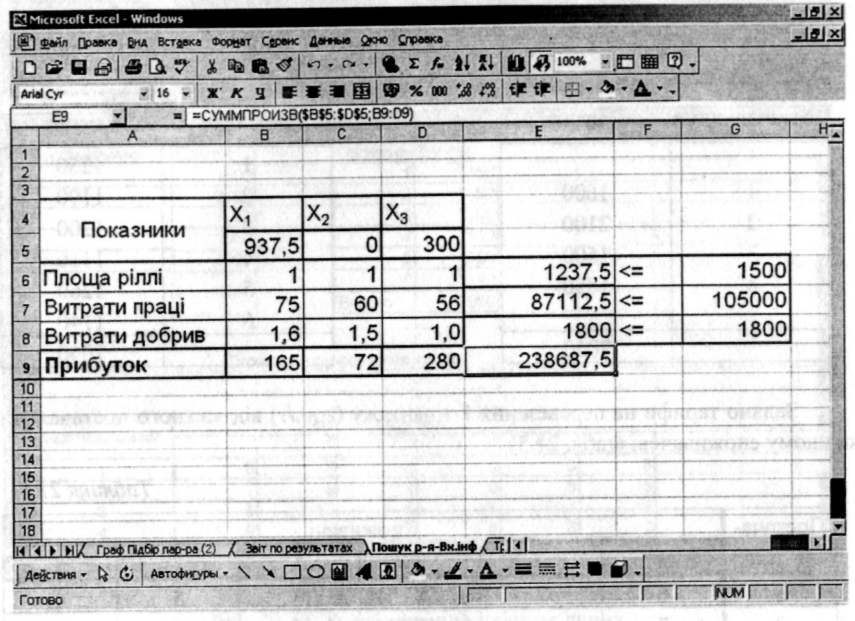
Результати розв'язання задачі показані на рис. 21.54. Максимальний прибуток 238687,5 ум.од. буде в тому випадку, коли загальна площа буде розподілена між двома культурами: озима пшениця – 937,5 га, гречка – 300 га. При цьому ресурси по добривах будуть використані всі, але залишилися недовикористаними площа та трудові ресурси.

Якщо в результаті виконання процедури пошуку рішення саме рішення не знайдено, але відомо, що воно є, треба змінити деякі опції в діалоговому вікні *Параметры поиска решения* (див. рис. 21.51) і ще раз виконати процедуру пошуку рішення.

До таких параметрів належать:

– Максимальное время дає можливість обмежити максимальний час розв'язання задачі;

– Предельное число итераций призначене для введення максимального числа проміжних рішень;



Показники	X_1	X_2	X_3			
	937,5	0	300			
Площа ріллі	1	1	1	1237,5	<=	1500
Витрати праці	75	60	56	87112,5	<=	105000
Витрати добрив	1,6	1,5	1,0	1800	<=	1800
Прибуток	165	72	280	238687,5		

Рис. 21.54. Результат пошуку рішення.

– Относительная погрешность задає точність виконання обмежень; задачу буде вирішено більш швидко, якщо встановити меншу точність;

– Допустимое отклонение – максимальне відхилення для цілочислових рішень;

– Сходимость застосовується тільки до розв'язання нелінійних задач; коли відносне змінювання значення у цільовій чарунці за останні п'ять ітерацій стає менше числа, яке вказано в полі *Сходимость*, пошук припиняється;

– Линейная модель використовується для прискорення пошуку рішення, застосовується до задач оптимізації; нелінійні моделі припускають використання нелінійних функцій, що сповільнює обчислення.

Мінімізація витрат на перевезення. При плануванні значних обсягів перевезень вантажів завжди з'являється задача оптимальної (за певним критерієм) оцінки їх організації. Розглянемо транспортну задачу за критерієм вартості перевезення.

Відомі наявність вантажу у постачальників і потреби споживачів (табл. 21.4).

Таблиця 21.4.

а)

Постачальники	Вантаж
1	1600
2	2100
3	1500
4	1850
	7050

б)

Споживачі	Вантаж
1	1250
2	1190
3	1200
4	1110
5	1260
6	1140
	7050

Задано тарифи на перевезення 1 т вантажу (грн./т) від кожного постачальника кожному споживачеві (табл. 21.5).

Таблиця 21.5.

Постачальники	Споживачі					
	1	2	3	4	5	6
1	6	7	8	9	5	6
2	7	8	9	6	8	5
3	8	10	5	7	7	8
4	9	6	7	8	8	7

Складемо математичну модель цієї задачі: необхідно визначити напрями та обсяги вантажоперевезень (x_{ij}) з мінімальними загальними витратами на перевезення

$$f(x) = \sum_{i=1}^{4}\sum_{j=1}^{6} c_{ij}x_{ij} \to \min$$

(c_{ij} – тарифи на перевезення).

При цьому всі вантажі повинні бути вивезені ($\sum_{j=1}^{6} x_{ij} = a_i$; a_i – наявність вантажу у постачальників, $i = \overline{1,4}$), а потреби задоволені ($\sum_{i=1}^{4} x_{ij} = b_j$; b_j – потреби споживачів, $j = \overline{1,6}$).

Розташуємо вхідні дані в електронних таблицях, як показано на рис. 21.55.

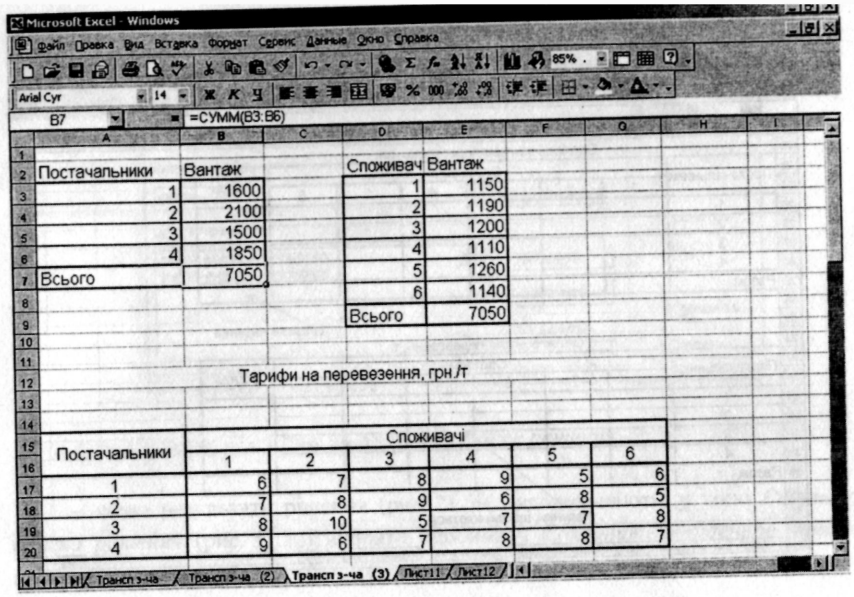


Рис. 21.55. Розташування вхідних даних.

Для розв'язання задачі необхідно:

– побудувати таблиці для розрахунків, як показано на рис. 21.56;

– у чарунку *B*25 ввести формулу =*B*17**B*35 (вартість перевезення вантажу від першого постачальника до першого споживача) і продублювати цю формулу, як показано на рис. 21.57;

– загальні витрати, тобто вартість перевезення всього вантажу, обчислити у чарунці *H*29 як загальну суму витрат усіх постачальників (цільова чарунка повинна мати мінімальне значення);

– вибрати команду **Сервис – Поиск решения...**;

– у вікні *Поиск решения* заповнити поля, як показано на рис. 21.58, встановши обмеження – обсяг вантажу, перевезеного споживачеві, повинен дорівнювати його потребам; обсяг вантажу, перевезеного постачальником, повинен дорівнювати його можливостям;

– натиснути кнопку *Параметры* та встановити їх значення, як показано на рис. 21.51 і натиснути клавішу *OK*;

– у вікні *Поиск решения* натиснути кнопку *Выполнить;*

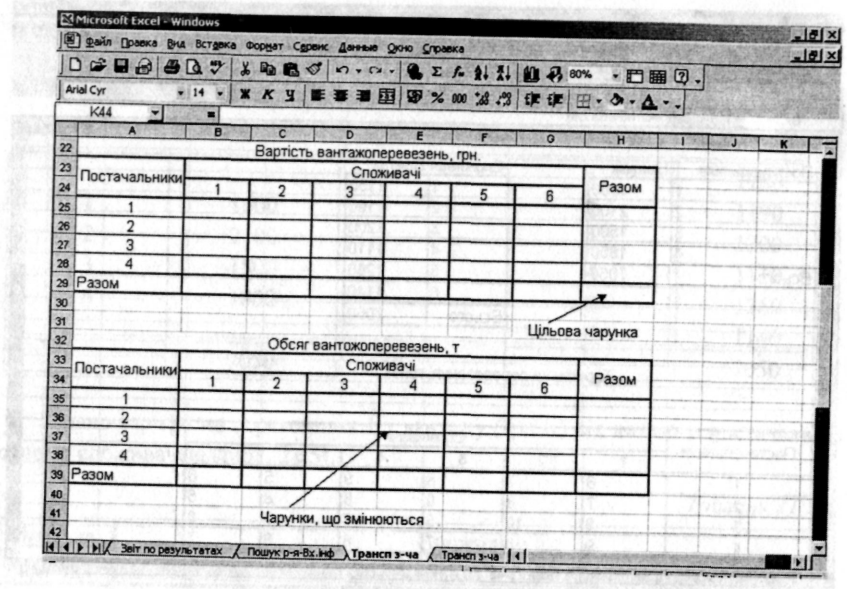


Рис. 21.56. Таблиці для розрахунків.

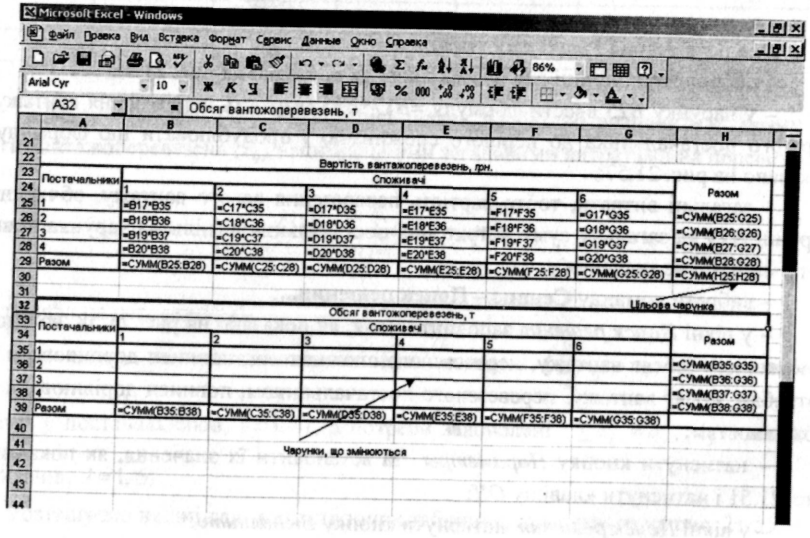


Рис. 21.57. Обчислення в таблицях.

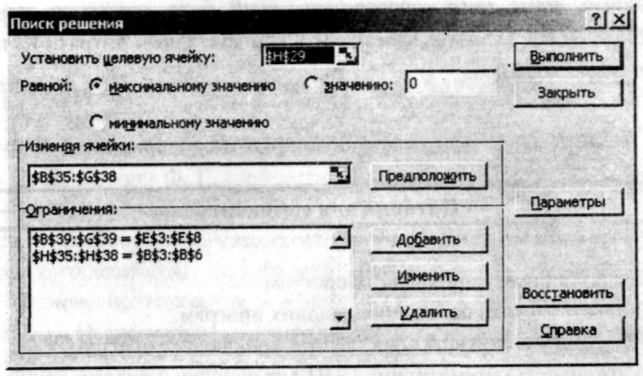


Рис. 21.58. Вікно *Поиск решения.*

– якщо результати рішення (рис. 21.59) задовольняють, у вікні *Результаты поиска решения* (рис. 51.52) вибрати перемикач *Сохранить найденное решение* і натиснути кнопку *ОК.*

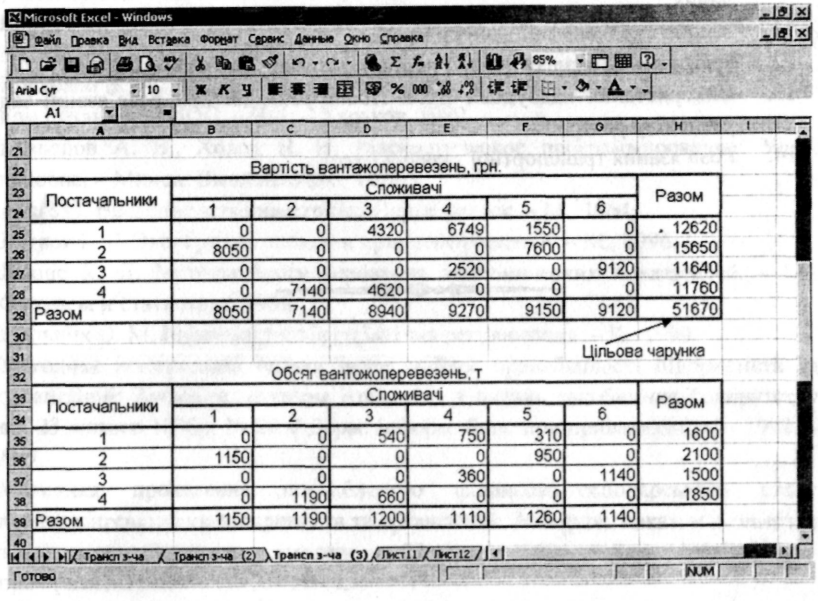


Рис. 21.59. Результат розв'язання транспортної задачі.

Як бачимо, план вантажоперевезень, який було одержано за допомогою надбудови Excel *Поиск решения*, забезпечить нам мінімальні витрати на перевезення вантажу в розмірі 51670 грн.

Питання для самоконтролю

1. Поняття пакету прикладних програм.
2. Характеристика пакетів прикладних програм.
3. Особливості статистичних пакетів прикладних програм.
4. Можливості математичних ППП для обробки й аналізу даних.
5. Загальна характеристика ППП Maple.
6. Характеристика підпрограм пакету Maple.
7. Розв'язання задач теорії графів у пакеті Maple.
8. Можливості підпрограми Simplex для розв'язання задач оптимізації.
9. Загальна характеристика Microsoft Excel 2000.
10. Використання функцій Excel для математичної та статистичної обробки даних.
11. Засоби апроксимації даних в Excel.
12. Інструменти аналізу даних і пошуку оптимальних рішень в Excel.
13. Використання надбудови Excel *Поиск решения* при розв'язанні задач розподілу ресурсів.
14. Розв'язання транспортної задачі в Excel.

БІБЛІОГРАФІЯ

1. Антикризисное управление: от банкротства – к финансовому оздоровлению / Под ред. Г. П. Иванова. – М., 1995.
2. Артеменко В. Г., Беллендир М. В. Финансовый анализ. – М., 1998.
3. Бешелев С. Д., Гурвич Ф. Г. Математико-статистические методы экспертных оценок. – М.: Статистика, 1974.
4. Браславец М. Е. Экономико-математические методы в организации планирования сельского хозяйства. – К.: Урожай, 1968.
5. Вагнер Г. Основы исследования операций. – Тт. 1 – 3.– М.: Мир, 1973.
6. Вентцель Е. С. Исследование операций: задачи, принципы, методология. – М.: Наука, 1988.
7. Гранберг А. Г. Моделирование социалистической экономики. – М.: Экономика, 1988.
8. Ізмайлова К. В. Фінансовий аналіз: Навч. посібник / МАУП. – К., 2000.
9. Інструкція про застосування Плану рахунків бухгалтерського обліку активів, капіталу, зобов'язань і господарських операцій: Затвердж. наказом Міністерства фінансів України від 30 листопада 1999 р. № 291 // Бухгалтер. учет и аудит. – 2000. –№ 1.
10. Корбут А. А., Финкельштейн Ю. Ю. Дискретное программирование. – М.: Наука, 1969.
11. Крутовой Ж. А. Экономико-математические методы в торговле и общественном питании / ХИОП. – Ч. 1. – Харьков, 1989.
12. Кузнецов А. В., Холод Н. И. Математическое программирование: Учеб. пособие. – Минск: Вышейша шк., 1984.
13. Ланге О. Введение в эконометрию / Пер. с польск. – М., 1964.
14. Литвок Б. Г. Экспертные оценки и принятие решений. – М., 1996.
15. Льюис К. Д. Методы прогнозирования экономических показателей. – М.: Финансы и статистика, 1986.
16. Мельник О. М. Інфляція: теорія і практика регулювання. – К., 1999.
17. Методика інтегральної оцінки інвестиційної привабливості підприємств та організацій: Затвердж. наказом Агентства з питань запобігання банкрутству від 23 лютого 1998р. № 22 // Держ. інформ. бюл. про приватизацію. – 1998. – № 7.
18. Методика проведення поглибленого фінансово-господарського стану неплатоспроможних підприємств та організацій: Затвердж. наказом Агентства з питань запобігання банкрутству від 27 червня 1997р. // Держ. інформ. бюл. про приватизацію. – 1997. – № 12.
19. Царенко О. М., Злобін Ю. А., Скляр В. Г., Панченко С. М. Комп'ютерні методи в сільському господарстві та біології. – Суми, 2000.

20. Степанюк В. В., Скрипка А. Г. Исследование операций: Учеб. пособие. – К., 1972.
21. Тейл Г. Экономические прогнозы и принятие решений. – М.: Статистика, 1971.
22. Ульянченко О. В. Методи оптимізацій в економіці: Навч. посібник / Харк. держ. аграр. ун-т ім. В. В. Докучаєва. – Харків, 2001.
23. Ульянченко О. В. Сучасні моделі дослідження операцій в економіці: Навч. посібник / Харк. держ. аграр. ун-т ім. В. В. Докучаєва. – Харків, 2000.
24. Финансы / Под ред. А. М. Ковалевой . – М.: Финансы и статистика, 1996.
25. Четыркин Е. М. Статистические методы прогнозирования. – М.: Статистика, 1975.
26. Шеремет А. Д., Сайфулин Р. С. Финансы предприятий. – М, 1998.
27. Щербаков П. А., Ульянченко О. В., Мартьянова Н. М., Бутенко Т. А. Інформатика та комп'ютерна техніка: програмне забезпечення ЕОМ: Навч. посібник / Харк. держ. аграр. ун-т ім. В.В. Докучаєва. – Харків, 2001.
28. Экономико-математические методы и модели: Учеб. пособие / Н. И. Холод, А. В. Кузнецов, Я. Н. Жихар и др.; Под общ. ред. А. В. Кузнецова. – 2-е изд. – Минск: БГЭУ, 2000.

ЗМІСТ

Від автора

ВСТУП 5

Глава 1

ОСНОВНІ ПОНЯТТЯ ТА ПРИНЦИПИ ДОСЛІДЖЕННЯ ОПЕРАЦІЙ 16

1.1. Об'єкт дослідження 16
1.2. Мета та цілі дослідження 18
1.3. Моделювання в теорії дослідження операцій 19
1.4. Математичне моделювання 21
1.5. Операційні системи, призначені для людей 25
Питання для самоконтролю 26

Глава 2

МЕТОДИКА ПРОВЕДЕННЯ ДОСЛІДЖЕННЯ ОПЕРАЦІЙ 28

2.1. Визначення мети та значимості цілей 28
2.2. Дослідження стратегій 31
2.3. Планування етапів розбудови проекту 32
2.4. Визначення проблеми дослідження 34
2.5. Розбудова математичної моделі 38
2.6. Інформаційне забезпечення та вибір числових методів 43
2.7. Розробка технічного завдання, програмування та відладка 45
2.8. Накопичення даних 44
2.9. Перевірка дієздатності моделі 45
Питання для самоконтролю 47

Глава 3

ВИРОБНИЧІ ФУНКЦІЇ 49

3.1. Поняття виробничої функції 49
3.2. Виробничі функції однієї змінної 52
3.3. Багатофакторні виробничі функції 61
3.4. Загальні властивості виробничих функцій 64
3.5. Визначення параметрів виробничих функцій 66
3.5.1. Елементи теорії обробки даних спостереження 69
3.5.2. Лінійна парна регресія 72
3.5.3. Оцінки параметрів лінійної парної регресії 74

3.5.4. Нелінійна парна регресія ... 77
3.5.5. Перевірка значимості рівняння регресії ... 80
3.5.6. Поняття про багатовимірні кореляцію та регресійний аналіз ... 84
3.6. Граничні та середні значення виробничих функцій ... 89
3.7. Врахування часу при розбудові виробничих функцій ... 93
Питання для самоконтролю ... 94

Глава 4

ПРОГНОЗУВАННЯ ЕКОНОМІЧНИХ ПОКАЗНИКІВ З ВИКОРИСТАННЯМ ЧАСОВИХ РЯДІВ ... 95

4.1. Загальні поняття ... 95
4.2. Часові ряди ... 97
4.2.1. Різновиди часових рядів ... 97
4.2.2. Тренд ... 97
4.2.3. Проміжки переоцінки та уточнення прогнозу ... 99
4.2.4. Вихідні дані ... 100
4.3. Аналіз тенденцій зміни рівнів часового ряду ... 101
4.3.1. Дослідження тенденції зміни середнього рівня ... 101
4.3.2. Елементарні засоби статистичного аналізу динаміки часових рядів ... 105
4.3.3. Плинні середні ... 106
4.3.4. Зважені плинні середні ... 108
4.3.5. Середні прирости ... 109
4.3.6. Темпи зростання та середні темпи приросту ... 111
4.4. Короткотермінове прогнозування ... 112
4.4.1. База та горизонт прогнозу ... 114
4.4.2. Прогнозування стаціонарних показників ... 115
4.4.2.1. Експоненціальна зважена середня ... 116
4.4.3. Прогнозування нестаціонарних показників ... 118
4.4.3.1. Характери трендів ... 118
4.4.3.2. Типи трендів ... 119
4.4.4. Лінійно-адитивна модель прогнозування ... 119
4.4.5. Визначення точності короткотермінового прогнозу ... 121
4.4.6. Контроль при короткотерміновому прогнозуванні ... 122
4.4.7. Поняття про адаптивне прогнозування ... 124
4.5. Методи середньо- та довготермінового прогнозування ... 126
4.5.1. Вихідні положення ... 126
4.5.2. Автокореляція ... 129
4.5.3. Екстраполяція тренда для розрахунків прогнозу ... 134
4.5.4. Оцінки точності та надійності прогнозу ... 135
4.5.4.1. Оцінка дисперсії прогнозу ... 137

4.5.4.2. Інтервал довіри прогнозу для лінійного тренда ... 141
4.5.4.3. Інтервали довіри для трендів, аналітичні форми яких можна привести до лінійних ... 145
4.6. Використання індикаторів та індексів ... 146
Питання для самоконтролю ... 152

Глава 5

ЗАГАЛЬНА ПОСТАНОВКА ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ (ЗЛП) ... 153

5.1. Приклади задач лінійного програмування ... 153
5.2. Різні форми запису задачі лінійного програмування ... 156
5.3. Геометричне тлумачення задачі лінійного програмування ... 160
Питання для самоконтролю ... 169

Глава 6

НЕОБХІДНІ ВІДОМОСТІ З ЛІНІЙНОЇ АЛГЕБРИ ... 170

6.1. Числа, вектори, *n*-мірний простір Евкліда ... 170
6.2. Матриці та визначники ... 175
6.3. Системи лінійних алгебраїчних рівнянь ... 180
Питання для самоконтролю ... 190

Глава 7

СИМПЛЕКСНИЙ МЕТОД РОЗВ'ЯЗАННЯ ЗАДАЧ ЛІНІЙНОГО ПРОГРАМУВАННЯ ... 191

7.1. Опорні плани задачі лінійного програмування ... 191
7.2. Оцінки опорних планів ... 195
7.2.1. Ознака оптимальності опорного плану ... 196
7.2.2. Ознака необмеженості цільової функції в допустимій області ... 201
7.2.3. Ознака наявності нескінченної множини оптимальних планів ... 201
7.2.4. Вироджені плани задачі лінійного програмування та проблема зациклення ... 202
7.3. Алгоритм симплексного методу розв'язання невироджених задач лінійного програмування ... 204
7.4. Пошук вихідного опорного плану ... 205
7.5. Симплексний метод з використанням фіктивного базису ... 208
Питання для самоконтролю ... 212

Глава 8

СПРЯЖЕНІ ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ 213

8.1. Поняття спряженості (двоїстості) 213
8.2. Відповідність між змінними спряжених задач 217
8.3. Теореми спряженості 219
8.4. Економічний зміст оптимальних планів спряжених задач 221
Питання для самоконтролю 226

Глава 9

ОСОБЛИВІ ЗАДАЧІ ЛІНІЙНОГО ПРОГРАМУВАННЯ. ТРАНСПОРТНА ЗАДАЧА 227

9.1. Загальні зауваження 227
9.2. Транспортна задача 229
9.3. Збалансовані та незбалансовані моделі транспортної задачі 232
9.4. Побудова опорного вихідного плану 234
9.5. Метод потенціалів пошуку оптимального плану 237
Питання для самоконтролю 246

Глава 10

ЕЛЕМЕНТИ ТЕОРІЇ ЛІНІЙНОГО ЦІЛОЧИСЛОВОГО ТА ДИСКРЕТНОГО ПРОГРАМУВАННЯ 247

10.1. Особливості задачі цілочислового та дискретного програмування 247
10.2. Короткий огляд основних засад методів розв'язання задач цілочислового лінійного програмування 248
10.3. Метод відтинання (метод Р. Гоморі) у розв'язанні задач цілочислового програмування 250
10.4. Метод розгалужень і меж 257
Питання для самоконтролю 263

Глава 11

НЕЛІНІЙНЕ ПРОГРАМУВАННЯ 264

11.1. Задача нелінійного програмування 264
11.2. Опуклі та угнуті функції, їх властивості 267
11.3. Умовний екстремум класичної теорії. Метод множників Лагранжа, їх економічний зміст 272
11.4. Задача опуклого програмування. Графо-аналітичний розв'язок у просторі двох змінних 276

11.5. Необхідні та достатні умови оптимальності в задачі опуклого програмування. Теорема Куна-Таккера 279
11.6. Задача квадратичного програмування 281
11.7. Градієнтні методи 284
11.8. Метод штрафних функцій 292
Питання для самоконтролю 298

Глава 12

МЕТОДИ ПЛАНУВАННЯ ТА УПРАВЛІННЯ МЕРЕЖАМИ 299

12.1. Призначення та сфера використання 299
12.2. Основні поняття теорії графів 300
12.3. Побудова правильної нумерації вершин графу 301
12.4. Алгоритм пошуку найкоротшого шляху мережі (графу) 303
12.5. Побудова графу планування та управління мережею (ПУМ) 305
12.6. Упорядкування графу ПУМ, обчислення основних параметрів 309
12.7. Аналіз та оптимізація планування й управління мережею 320
Питання для самоконтролю 325

Глава 13

МАТЕМАТИЧНІ МОДЕЛІ УПРАВЛІННЯ ЗАПАСАМИ 326

13.1. Основні поняття 326
13.2. Управління однономенклатурними запасами 327
13.3. Статична модель управління багатономенклатурними запасами 335
13.4. Стохастичні моделі управління запасами 338
Питання для самоконтролю 346

Глава 14

ЕЛЕМЕНТИ ТЕОРІЇ ГРИ В РОЗВ'ЯЗАННІ ГОСПОДАРСЬКИХ ЗАДАЧ 347

14.1. Основні поняття теорії гри 347
14.2. Класифікація моделей ігор 348
14.3. Математичне моделювання конфліктних ситуацій 349
14.4. Розв'язання матричних ігор у чистих стратегіях 351
14.5. Розв'язання матричних ігор у змішаних стратегіях 355
14.6. Приведення матричної гри до задачі лінійного програмування 361
14.7. Ігри з ненульовою сумою та кооперативні. Моделювання проблем мікроекономіки з використанням математичного апарата теорії гри 367
14.8. Позиційні ігри як моделювання проблеми вступу до ринку 370
14.9. Теорія гри та прийняття управлінських рішень 374
Питання для самоконтролю 381

Глава 15

МОДЕЛІ ДИНАМІЧНОГО ПРОГРАМУВАННЯ 382

15.1. Історична довідка 382
15.2. Приклади розв'язання задач методом динамічного програмування 383
15.2.1. Задача пошуку найекономнішого маршруту доставки вантажу 383
15.2.2. Задача розширення виробничих потужностей за умови відрахувань від прибутку з метою одержання максимального зиску 387
15.3. Загальна постановка задачі динамічного програмування 394
15.4. Принцип оптимальності та структура рівняння Белмана 398
15.5. Задача оптимального розподілу коштів на реконструкцію та модернізацію між кількома об'єктами 403
Питання для самоконтролю 409

Глава 16

МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ПРОЦЕСІВ МАСОВОГО ОБСЛУГОВУВАННЯ 410

16.1. Основні поняття та класифікація моделей СМО 411
16.2. Коротка історична довідка 412
16.3. Класифікація систем масового обслуговування 413
16.4. Основні засади математичного моделювання роботи систем масового обслуговування 414
16.5. Терміни виконання вимог (обслуговування) 420
16.6. Графічне моделювання систем масового обслуговування 423
16.7. Чисті системи масового обслуговування за необмеженої черги 426
16.7.1. Одноканальні СМО з очікуванням 426
16.7.2. Багатоканальні СМО з очікуванням 434
16.8. Рівняння Колмогорова 437
16.9. Граничні ймовірності стабільної роботи СМО 441
16.10. Процес розмноження та вимирання 446
Питання для самоконтролю 448

Глава 17

ІМІТАЦІЙНЕ МОДЕЛЮВАННЯ ОПЕРАЦІЙНИХ СИСТЕМ 449

17.1. Імітаційний підхід 449
17.2. Задачі економіки, при розв'язанні яких доцільно використовувати імітаційне моделювання 450
17.3. Кроки практичної реалізації імітаційного моделювання 452
17.4. Приклад блок-схеми імітаційної моделі 453
Питання для самоконтролю 456

Глава 18

ДОСЛІДЖЕННЯ ІНВЕСТИЦІЙНИХ ПРОЕКТІВ 457

18.1. Задача вибору інвестиційного проекту 457
18.2. Дисконтування грошових потоків 458
18.3. Прибуток та інфляція 463
18.4. Оцінка інвестиційного ризику 470
18.5. Одна з інвестиційних моделей при використанні банківського кредиту 471
Питання для самоконтролю 475

Глава 19

ЕКСПЕРТНІ ОЦІНКИ В МЕНЕДЖМЕНТІ 476

19.1. Задача прийняття рішень 476
19.2. Експертні оцінки 477
19.3. Оцінювання об'єктів при проведенні експертиз 479
19.4. Аналіз узгодженості експертних оцінок 485
19.5. Підготовка та проведення експертиз 487
Питання для самоконтролю 490

Глава 20

ФІНАНСОВИЙ СТАН ПІДПРИЄМСТВА, ДІАГНОСТИКА БАНКРУТСТВА 491

20.1. Необхідність і значимість оцінки фінансового стану підприємства 491
20.2. Фінансовий аналіз 491
20.3. Основні показники фінансового стану підприємства 493
20.3.1. Показники оцінки майнового стану 493
20.3.2. Показники оцінки фінансової стійкості 494
20.4. Аналіз фінансової стійкості підприємства 495
20.4.1. Поняття фінансової стійкості 495
20.4.2. Фінансова стійкість стосовно джерел покриття обігових коштів 495
20.4.3. Показники платоспроможності (ліквідності) 496
20.5. Банкрутство: фінансові та правові питання 497
20.5.1. Поняття банкрутства. Фінансовий стан підприємства та банкрутство 497
20.5.2. Діагностика банкрутства 500
Питання для самоконтролю 507

Глава 21

ВИКОРИСТАННЯ ПАКЕТІВ ПРИКЛАДНИХ ПРОГРАМ (ППП) У ДОСЛІДЖЕННІ ОПЕРАЦІЙ 508
21.1. Загальна характеристика ППП 508
21.2. Засоби пакетів статистичних програм для обробки та аналізу даних 509
21.3. Математичні ППП 511
21.3.1. Пакет Maple 512
21.3.1.1. Загальна характеристика 513
21.3.1.2. Робоча панель Maple 513
21.3.1.3. Довідкова підсистема Maple 514
21.3.1.4. Типи даних 515
21.3.1.5. Підпрограма Networks 516
21.3.1.6. Підпрограма Stats 520
21.3.1.7. Підпрограма Simplex 520
21.4. Microsoft Excel 2000 як інструмент для вирішення задач математичного програмування 529
21.4.1. Функції Excel 531
21.4.2. Апроксимація даних 537
21.4.2.1. Функції апроксимації кривої 538
21.4.2.2. Регресійний аналіз 542
21.4.2.3. Використання ліній тренда 544
21.4.3. Інструменти Excel для аналізу даних і пошуку оптимальних рішень 548
21.4.3.1. Аналіз за сценарієм «що буде - якщо...» 548
21.4.3.2. Аналіз за сценарієм «що треба - щоб...» 552
21.4.3.2.1. Підбір параметра 553
21.4.3.2.2. Графічний підбір параметра 555
21.4.3.2.3. Пошук рішення 557
Питання для самоконтролю 568

БІБЛІОГРАФІЯ 569

Навчальне видання

Олександр Вікторович Ульянченко

ДОСЛІДЖЕННЯ ОПЕРАЦІЙ В ЕКОНОМІЦІ

Підручник для студентів вузів

Комп'ютерний набір і верстка В.О.Бабенко, Г.Е.Греченко, О.І.Листопад, Ю.А.Сотніков, А.І.Листопад.
Редактор С.О.Рамакаєва.
Коректори Я.І.Лагоза, О.Ю.Потапова.

Підписано до друку 14.06.2002. Формат 60×90/16. Папір офсет.
Гарнітура Таймс. Ум.-друк. арк. 34,5. Облік.-вид. арк. 40,0.
Наклад 5000. Зам.

61022, Харків, вул. А.Барбюса, 3а
ВКФ "Гриф"