МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ УКРАЇНИ
ХАРКІВСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ
МІСЬКОГО ГОСПОДАРСТВА

А. Є. АЧКАСОВ, О. О. ВОРОНКОВ

КОНСПЕКТ ЛЕКЦІЙ
З КУРСУ

ДОСЛІДЖЕННЯ ОПЕРАЦІЙ

(для студентів галузі знань 0306 – «Менеджмент і адміністрування» напряму 6.030601 – «Менеджмент» заочної форми навчання)

**ХАРКІВ
ХНАМГ
2012**

Ачкасов, А. Є. Конспект лекцій з курсу «Дослідження операцій» (для студентів галузі знань 0306 – «Менеджмент і адміністрування» напряму 6.030601 – «Менеджмент» заочної форми навчання) / А. Є. Ачкасов, О. О. Воронков; Харк. нац. акад. міськ. госп-ва. – Х.: ХНАМГ, 2012. – 132 с.

Автор: А. Є. Ачкасов
О. О. Воронков

Рецензент: доцент кафедри ЕПМГ, к.е.н. А. І. Базецька

Рекомендовано кафедрою „Економіки підприємств міського господарства", протокол № 12 від 07.06.12 р.

ЗМІСТ

ЗМ 1. МОДЕЛІ І МЕТОДИ СИСТЕМНОГО АНАЛІЗУ, СПОСОБИ ДОСЛІДЖЕННЯ І ОПТИМІЗАЦІЇ ОПЕРАЦІЙ

ТЕМА 1
ДОСЛІДЖЕННЯ ОПЕРАЦІЙ – НАУКА ПРО ОБҐРУНТУВАННЯ ТА ПРИЙНЯТТЯ РІШЕНЬ

Дослідження операцій є наукою, що займається розробкою і практичним застосуванням методів найбільш оптимального управління організаційними системами. В даний час методи дослідження операцій знаходять широке застосування у розв’язанні самих різних практичних задач, починаючи від перспективного планування наукових розробок та закінчуючи прогнозуванням розвитку сфери обслуговування. Це пов'язано з тим, що будь-яка операція є сукупністю цілеспрямованих дій, а дослідження операцій - пошук шляхів досягнення однієї або кількох цілей.

Витоки дослідження операцій йдуть у далеку історію. Різке збільшення розмірів виробництва, розподіл праці зумовив поступову диференціацію управлінської праці. З'явилася необхідність в плануванні матеріальних, трудових та грошових ресурсів, в обліку та аналізі праці і виробленні прогнозу на майбутнє. В управлінському апараті почали виділятися підрозділи: відділ фінансів, збуту, бухгалтерії, планово-економічний відділ та інші, що прийняли на себе окремі управлінські функції.

До цього періоду належать перші роботи з досліджень у області організації праці та управління. Як самостійний науковий напрям дослідження операцій оформилося на початку 40-х років. Перші публікації з дослідження операцій належать 1939-1940 рокам, в яких ці методи застосовані для вирішення військових завдань, зокрема для аналізу й дослідження військових операцій. Звідси і пішла назва дисципліни. Пізніше принципи і методи дослідження операцій стали застосовуватися у сфері промислово-фінансового управління. Із збільшенням масштабів виробництва розширювалися масштаби операційних досліджень, коло вирішуваних задач, удосконалювалися методи нової науки. Виникла необхідність в підготовці фахівців з дослідження операцій - операціоністів. У провідних університетах США та Англії вперше було почато систематичне викладання курсу дослідження операцій. Виникла необхідність в координації роботи операціоністів, у регулярному обміні теоретичними дослідженнями та прикладними розробками. З цією метою в 1957 р. було створено Міжнародну федерацію дослідження операцій IFORS, до складу якої входили національні товариства і комітети з дослідження операцій багатьох країн.

Нерідко дослідження операцій розуміють як застосування математичних кількісних методів для обґрунтування рішень в усіх областях людської діяльності. При цьому випускають з уваги, що дослідження операцій як наукова дисципліна знаходиться на стику наук, що оперують не тільки кількісними, але й якісними факторами. Саме зневагою до якісних факторів пояснюється основна кількість негативних результатів застосування методів дослідження операцій до розв’язання практичних задач, оскільки змістовні і методологічні аспекти в будь-якому операційному дослідженні грають анітрохи не меншу роль, ніж формальні.

Предметом дослідження операцій є системи організаційного управління або організації, які складаються з великої кількості взаємодіючих між собою підрозділів тих, інтереси яких не завжди узгоджуються між собою і можуть бути протилежні.

Мета дослідження операцій - кількісне обґрунтування ухвалюваних рішень з управління організаціями. Рішення, яке виявляється найбільш вигідним для всієї організації називають оптимальним, а рішення найбільш вигідне одному або кільком підрозділам називають субоптимальним.

Відзначимо основні особливості дослідження операцій. По-перше - це системний підхід до аналізу поставленої проблеми. Системний аналіз є основним методологічним принципом дослідження операцій, який полягає в тому, що будь-яке завдання, яким би частковим воно не здавалося, розглядається з погляду його впливу на критерій ефективності функціонування всієї системи. Другою характерною особливістю дослідження операцій є те, що під час розв'язання кожної проблеми виникають все нові і нові завдання. Якщо спочатку покладають вузькі цілі, застосування операційних методів є неефективним. Найбільшого ефекту можна досягнути тільки при безперервному дослідженні, що забезпечує спадкоємність в переході від одного завдання до іншого.

Однією з істотних особливостей дослідження операцій є прагнення знайти оптимальне рішення поставленої задачі. Проте, часто таке рішення опиняється недосяжним із-за обмежень, що накладаються наявними ресурсами або рівнем сучасної науки. Наприклад, для комбінаторних завдань, зокрема завдань з календарного планування при кількості верстатів понад 4, оптимальне рішення при сучасному рівні розвитку математики можливо знайти лише простим перебором варіантів. Проте навіть при невеликих n кількість можливих варіантів настільки велика, що перебір всіх варіантів при існуючих обмеженнях на швидкодію ЕОМ та припустимий машинний час практично немислимі. Тоді доводиться обмежуватися пошуком досить хорошого або субоптимального рішення.

Особливість операційних досліджень полягає ще в тому, що їх проводять комплексно, за багатьма напрямами. Для проведення такого дослідження створюють операційну групу, у склад якої входять фахівці різних областей: інженери, математики, економісти, соціологи, психологи.

1.1. Основні поняття. Моделі дослідження операцій. Етапи дослідження операцій.

Нагадаємо, що будь-яка математична модель є описом (часто наближеним) будь-якого класу явищ реального світу, вираженим за допомогою математичних символів.

Дослідження операцій - комплексна математична дисципліна, що займається побудовою, аналізом та застосуванням математичних моделей ухвалення оптимальних (якнайкращих в якомусь сенсі) рішень. Основним завданням дослідження операцій є завдання вибору в заданій множині елементу, що задовольняє тим або іншим критеріям (від грецького kriterion - засіб для порівняння). При цьому будь-який елемент множини називають допустимим рішенням, а вибраний елемент - оптимальним рішенням.

Побудова математичної моделі будь-якої задачі дослідження операцій завжди починається з опису множини D припустимих розв'язків та критеріїв оптимальності. При цьому під критерієм оптимальності розуміють ознаку, на підставі якої проводиться порівняльна оцінка припустимих розв'язків і вибір оптимального розв'язку.

На наступному етапі досліджень з використанням побудованої моделі треба знайти оптимальний розв'язок рішення з множини D припустимих. розв'язків. Але, як правило, ухвалює рішення не дослідник операції, який лише готує інформацію для ухвалення рішення а «особа, що приймає рішення» (ОПР). Саме ОПР формулює вимоги і воно ж ухвалює остаточне рішення.

Під ОПР розуміють групу людей, яка виробляє узгоджені вимоги до припустимих рішень та критеріїв оптимальності, ухвалює рішення. При цьому потрібно пам'ятати про те, що будь-яке рішення завжди ухвалюється відповідно до інформаційного стану ОПР, тобто відповідно до його змістовних уявлень про можливі і доцільні дії в певних умовах. У математичній моделі це відображається на множині D припустимих рішень і критеріях оптимальності.

Треба зазначити, що побудова будь-якої математичної моделі задачі дослідження операцій майже завжди пов'язана з необхідністю задоволення двох, по суті, суперечливих вимог: а) якомога точніше відобразити реальні явища, що вивчаються; б) побудувати достатньо просту математичну модель. що дозволяє вирішити початкове завдання і одержати осяжні результати. Саме тому на етапі побудови математичної моделі необхідна тісна співпраця дослідника операції (математика) і ОПР (фахівця).

Охарактеризуємо основні етапи операційного дослідження:

1. Постановка завдання. Спочатку завдання формулюють з погляду замовника. Під час аналізу системи завдання поступово уточнюється.

2. Формалізація завдання. Одержавши доволі строгу і логічно несуперечливу змістовну постановку завдання, потрібно побудувати її математичну модель.

3. Знаходження методу рішення. Для знаходження оптимального рішення залежно від структури завдання застосовують ті або інші методи теорії оптимальних рішень, звані також методами математичного програмування.

4. Перевірка і коректування моделі. У складних системах, до яких належать системи організаційного типа, модель лише частково відображає реальний процес. Тому необхідна перевірка ступеня відповідності або адекватності моделі і реального процесу. Перевірку проводять порівнянням передбаченої поведінки з фактичною при зміні значень зовнішніх некерованих дій. Коректування може зажадати додаткових досліджень об'єкту, уточнення структури математичної моделі, численних змін змінних. Отже, 4 етапи багато разів повторюються, поки не буде досягнуто задовільної відповідності між виходами об'єкту і моделі.

5. Реалізація знайденого рішення на практиці. Впровадження можна розглядати як самостійне завдання, застосувавши до нього системний підхід і аналіз.

1.2. Класифікація задач та методів дослідження операцій

Зупинимося на класифікації можливих постановок завдань дослідження операцій, яка може проводитися за різними ознаками.

Класифікація завдань дослідження операцій за видом інформаційного стану ОПР. Якщо прийняття рішення відбувається в наперед відомому інформаційному стані ОПР, що не змінюється в часі, то завдання дослідження операцій називають статичним, а всю процедуру прийняття рішення можна реалізувати за один етап (крок). Класу статичних завдань належать завдання з кількома різними, але не змінними у часі інформаційними станами, це характерно для групи ОПР.

Якщо у процесі прийняття рішення інформаційний стан ОПР змінюється в часі, то завдання дослідження операцій називають динамічним. У цьому випадку найбільш доцільною є поетапна (багатокрокова) процедура прийняття рішення.

Класифікація завдань дослідження операцій за структурою інформаційного стану ОПР. Інформаційний стан ОПР може відповідати або єдиному фізичному стану об'єкту досліджень, або множині фізичних станів. У першому випадку завдання дослідження операцій називають детермінованим, в другому - стохастичним (або завданням ухвалення рішень в умовах ризику), якщо відомі апріорні імовірності перебування об'єкту досліджень в кожному із станів, і невизначеним (або завданням ухвалення рішень в умовах невизначеності), якщо відсутня інформація про апріорну імовірність. Завдання ухвалення рішень в умовах невизначеності є предметом дослідження теорії ігор.

Отже, детермінований рівень - найпростіший рівень інформації про ситуацію, коли умови, в яких ухвалюються рішення, відомі повністю. Стохастичний рівень - рівень, при якому відома множина можливих варіантів умов та їх імовірнісний розподіл. Невизначений рівень - рівень, коли відома множина можливих варіантів, але без будь-якої інформації про їх імовірність.

Класифікація завдань дослідження операцій за видом критерію оптимальності. Класифікація завдань дослідження операцій може бути пов'язаною з видом використовуваного критерію оптимальності, який може мати будь-який вигляд, у тому числі і вигляд, що не формалізується.

Критерієм оптимальності може бути вимога про максимізацію або мінімізацію певної скалярної функції f, визначеної на множині припустимих рішень, її називають цільовою функцією. В цьому випадку завдання дослідження операцій називають завданням математичного програмування. Якщо ж критерієм оптимальності є вимога про максимізацію або мінімізацію кількох скалярних функцій, то говорять про завдання багатокритеріальної (векторної) оптимізації.

У математичному програмуванні частіше за інші розглядають завдання, в яких множина D припустимих рішень є підмножиною R^n, що задовольняє системі лінійних нерівностей

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

Таку множину, якщо вона непорожня і обмежена, називають опуклим багатогранником, оскільки цей клас опуклих множин відповідає геометричному уявленню про багатогранники в просторі.

Якщо множина припустимих рішень $D \in R^n$ є опуклим багатогранником, а цільова функція f лінійна, то початкове завдання називають завданням лінійного програмування. Якщо множина $D \in R^n$ є опуклим багатогранником, а цільова функція f є квадратичною, то початкове завдання називають завданням квадратичного програмування. Якщо $D \in R^n$ - опукла множина, а f - опукла функція, то початкове завдання називають завданням опуклого програмування. Теорія рішення стохастичних задач лінійного програмування є предметом досліджень стохастичного програмування.

У інших завданнях математичного програмування множина D припустимих рішень може бути кінцевою множиною. Такі завдання належать до дискретного програмування. У них припустимі рішення можуть бути точками цілочисельних решіток Z^n (цілочисельне програмування) або векторами, кожна координата яких може приймати лише два значення (булеве програмування). У окремих завданнях елементи множини D припустимих рішень можуть бути перестановками кінцевої кількості символів та ін.

Множина D припустимих рішень може бути підмножиною певного функціонального простору. В цьому випадку маємо завдання варіаційного числення або завдання оптимального управління.

Особливим випадком завдань математичного програмування є задачі про знаходження максиміну:

$$\max_{Y} \min_{Z} f(Y, Z), \quad \begin{pmatrix} Y \\ Z \end{pmatrix} \in D$$

і мінімаксу:

$$\min_{Y} \max_{Z} f(Y, Z), \quad \begin{pmatrix} Y \\ Z \end{pmatrix} \in D .$$

Зробимо зауваження, щодо математичних методів дослідження операцій. Як математичні методи дослідження операцій зазвичай розуміють математичний апарат (спеціально розроблений або адаптований), що призначений для вирішення завдань дослідження операцій. Розробленість математичних методів для різних класів завдань дослідження операцій є далеко не однаковою. В даний час найбільш розроблена теорія лінійного та опуклого програмування. Вкажемо основні напрями, за якими велися і ведуться математичні дослідження.

Розвиток симплекс-методу. В даний час існує багато модифікацій цього методу, що дозволяють істотно скоротити час розрахунку, зробити алгоритм нечутливим до виродженості опорних планів, підвисити розмірність вирішуваних задач, вирішувати так звані блокові задачі та ін. Не дивлячись на велику кількість цих модифікацій, продовжують з'являтися все нові й нові його варіанти.

Цілочисельне лінійне програмування. У великій кількості вирішуваних задач лінійного програмування на змінні накладається додаткова умова їх цілочисельності. Коли накладено додаткову умову цілочисельності змінних, відповідне завдання носить назву завдання цілочисельного лінійного програмування. Просте округлення до цілих чисел тут не допомагає, план може вийти не оптимальним. Тому доводиться розробляти спеціальні алгоритми рішення таких задач, до найвідоміших з яких належать алгоритми Гоморі, засновані на так званій ідеї відсікання.

Булеве програмування. До окремого випадку завдання цілочисельного лінійного програмування належать завдання, де змінні можуть приймати лише два значення 0 та 1. Відповідні завдання часто називають завданнями булевого програмування. Найвідоміші з цих завдань - це завдання про призначення (якого працівника на яку роботу поставити), завдання вибору маршруту (завдання комівояжера, завдання листоноші), завдання про максимальне паросполучення та ін. Для вирішення завдань цього типа розробляються дуже специфічні алгоритми, засновані на комбінаториці, графах та ін.

Стохастичне лінійне програмування. Буває багато практичних ситуацій, коли коефіцієнти цільової функції, коефіцієнти в матриці коефіцієнтів, коефіцієнти обмежень є випадковими величинами. В цьому випадку сама цільова функція стає випадковою величиною, і обмеження типу нерівностей можуть виконуватися лише з певною імовірністю. Доводиться змінювати постановку самих завдань з урахуванням цих ефектів і розробляти абсолютно нові методи їх рішення.

Квадратичне програмування. Завдання лінійного програмування є окремим випадком цих завдань, вони виходять при A=0. Способи рішення цих задач багато в чому визначаються видом матриці A: якщо A - позитивно визначена матриця, то цільова функція буде опуклою і будь-який її локальний мінімум буде глобальним. Якщо A - негативно визначена матриця, то може бути кілька локальних мінімумів, але глобальний мінімум, якщо він існує, досягається обов'язково на вершині припустимої області. У загальному випадку, коли власні числа матриці A мають різні знаки, завдання дуже сильно ускладнюється, оскільки глобальний мінімум може досягатися де завгодно - і усередині області і на її межі.

Опукле програмування. Для цих завдань характерним є те, що будь-який локальний мінімум виявляється глобальним, і все збігається до знаходження цього єдиного мінімуму. Для вирішення завдань цього типу розроблені численні чисельні методи, пристосовані для вирішення на ЕОМ, в основному пов'язані з поняттям градієнта цільової функції і основною ідеєю про те, що функція найшвидше убуває, якщо рухатися в напрямі, протилежному градієнту. До них належать метод градієнтного спуску, метод пов'язаних градієнтів та ін. Але є і методи, засновані на інших ідеях - метод штрафних функцій, численні варіанти методу випадкового пошуку та ін.

Геометричне програмування. Під завданнями геометричного програмування розуміють завдання найбільш щільного розташування деяких об'єктів в заданій двовимірній або тривимірній області. Такі завдання зустрічаються в завданнях розкрою матеріалу для виробництва якихось виробів та ін. Це ще недостатньо розроблена область математичного програмування і алгоритми, що є тут, в основному орієнтовані на скорочення перебору варіантів з пошуком локальних мінімумів.

Дискретне програмування. Багато завдань дослідження операцій такі як розподіл ресурсів, мережеве планування, календарне планування описуються математичними моделями дискретного програмування. У завданнях дискретного програмування область допустимих рішень є неопуклою і незв'язною, тому відшукання рішення в таких завданнях пов'язане із значними труднощами. Зокрема неможливе застосування стандартних прийомів, використовуваних при заміні дискретного завдання його безперервним аналогом, що полягають в подальшому округленні знайденого рішення до найближчого цілочисельного.

Динамічне програмування - це обчислювальний метод для вирішення завдань певної структури. Виникло і сформувалося в 1950-1953 рр. завдяки роботам Р. Беллмана над динамічними завданнями управління запасами. У спрощеному формулюванні динамічне програмування є спрямований послідовний перебір варіантів, який обов'язково призводить до глобального максимуму. Основні необхідні властивості завдань, до яких можливо застосувати цей принцип, наступні:

1. Завдання повинне допускати інтерпретацію як n-кроковий процес ухвалення рішень;
2. Завдання має бути визначеним для будь-якої кількості кроків і мати структуру, не залежну від їх кількості;
3. Під час розгляду k-крокового завдання має бути заданою певна множина параметрів, що описують стан системи, від яких залежать оптимальні значення змінних. Причому ця множина не повинна змінюватися при збільшенні кількості кроків;
4. Вибір рішення (управління) на k-му кроці не повинен робити впливу на попередні рішення, окрім необхідного перерахунку змінних.

Завдання про вибір траєкторії, завдання послідовного ухвалення рішення, завдання про використання робочої сили, завдання управління запасами є класичними завданнями динамічного програмування.

Контрольні запитання

1. Що розуміють під дослідженням операцій?
2. Сформулюйте основне завдання дослідження операцій.
3. Чи існує зв'язок між критерієм оптимальності та принципом оптимальності? Відповідь аргументуйте.
4. У чому полягає принципова відмінність статичних і динамічних завдань дослідження операцій?
5. Приведіть класифікацію завдань дослідження операцій за структурою інформаційного стану ОПР.
6. Яке завдання дослідження операцій називають параметричним, в чому його принципова відмінність від завдання ухвалення рішень в умовах невизначеності?
7. Приведіть класифікацію завдань дослідження операцій за видом критерію оптимальності.
8. Який зв'язок існує між завданнями математичного програмування і векторної оптимізації?

ТЕМА 2
ЗАГАЛЬНА ЗАДАЧА ЛІНІЙНОГО ПРОГРАМУВАННЯ ТА ТЕОРІЯ ДВОЇСТОСТІ

2.1. Лінійні економіко-математичні моделі. Постановка задачі лінійного програмування (ЛП)

Серед задач математичного програмування найпростішими й найкраще розробленими є задачі лінійного програмування. Характерним для них є те, що цільова функція F(x) лінійно залежить від елементів розв'язку $x_1,x_2,\dots,x_n$. Її називають ***лінійною формою*** й позначають L; обмеження, що накладають на елементи розв'язку, мають вигляд лінійних рівностей і нерівностей щодо $x_1, x_2,\dots,\dots,x_j,\dots,x_n$.

Такі задачі часто зустрічаються на практиці. До задач лінійного програмування належить, зокрема, найпростіша задача виробничого планування (задача про оптимальне використання ресурсів). Також до задач лінійного програмування належать задачі про використання інвестицій, про мінімізацію витрат, транспортна задача, задача про складання раціону та ін.

Математичну модель задачі лінійного програмування завжди записують у двох формах – у ***загальній формі*** (ЗЗЛП) і ***канонічній формі*** (КЗЛП).

У загальному вигляді задачу лінійного програмування (ЗЛП) формулюють в такий спосіб:

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \to \text{extr}. \qquad (2.1)$$

на деякій множині D, де $x \in D$ задовольняє системі обмежень

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \le b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \le b_2 \\ \dots\dots\dots\dots\dots\dots \\ a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kn}x_n \le b_k \\ a_{(k+1)1}x_1 + a_{(k+1)2}x_2 + \dots + a_{(k+1)n}x_n = b_{(k+1)} \\ \dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \\ x_j \ge 0, \quad j = \overline{1,n} \end{array} \qquad (2.2)$$

Відзначимо, що в системі перші k обмежень є нерівностями, а наступні m-k - рівняннями. Цього можна домогтися простим переупорядкуванням виразів.

Щодо напрямку знака нерівності вважатимемо, що ліва частина менша або дорівнює правій частині. Домогтися цього можна, помноживши на (-1) обидві частини нерівностей із протилежним знаком.

Вибір типу шуканого екстремуму цільової функції також не принциповий, оскільки задача пошуку максимуму функції $L = \Sigma c_j x_j$ є еквівалентною задачі пошуку мінімуму функції $-L = \Sigma(-c_j x_j)$.

Задачу лінійного програмування, записану в такій формі, називають загальною задачею лінійного програмування (ЗЗЛП).

2.2. Структура множини припустимих розвязків задач ЛП. Геометрична інтерпретація задачі ЛП

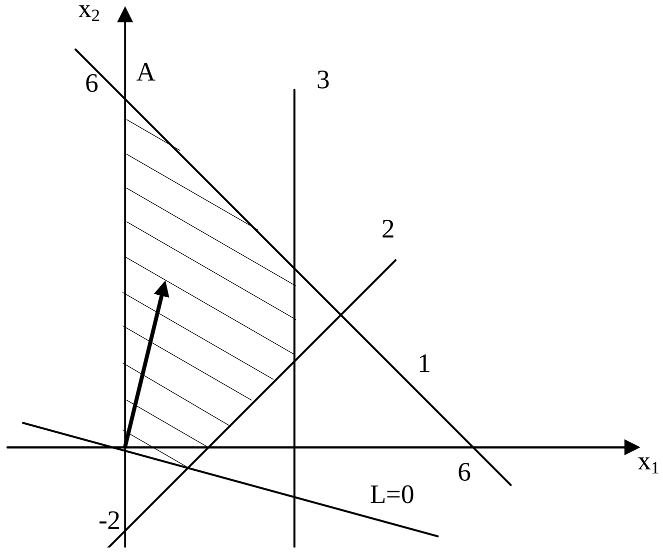

Рис. 2.1 – Геометрична інтерпретація ЗЗЛП

У тому випадку, коли ЗЗЛП містить дві змінні x_1 і x_2, її можна зобразити на координатній площині й одержати розв'язок графічним методом. Графічне розв'язання ЗЗЛП носить ілюстративний характер, але основний зміст і термінологія розповсюджуються на задачі великої розмірності.

Розглянемо приклад. Нехай цільова функція представлена виразом $L = x_1 + 3x_2 \to \max$, а обмеження задані системою нерівностей:

$$x_1 + x_2 \le 6$$
$$x_1 - x_2 \le 2$$
$$x_1 \le 3$$
$$x_1 \ge 0, x_2 \ge 0.$$

Будемо зображувати пару значень x_1 і x_2 точкою на координатній площині $x_1 0 x_2$ з координатами (x_1, x_2), що показано на рис. 2.1.

Кожна нерівність визначає деяку напівплощину. Перетинання трьох напівплощин є множиною припустимих планів D, тому що кожна точка його множини належить одночасно кожній із трьох напівплощин, а отже задовольняє обмеженням ЗЗЛП. Помітимо, що припустимих розв'язків - нескінченна множина.

Для визначення оптимального плану задачі, тобто такого розв'язку (x_1, x_2), що обертає цільову функцію на максимум, скористаємося визначеннями:

- градієнтом функції f(x) називають вектор

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \ldots, \frac{\partial f(x)}{\partial x_n} \right),$$

який вказує напрямок найскорішого зростання функції f(x);

- лінією рівня функції f(x) називають множину точок з області її визначення, у яких функція приймає те саме фіксоване значення.

У нашому прикладі $\nabla L = (1, 3)$. Лінії рівня L перпендикулярні напрямку градієнта. Побудуємо опорну пряму L=0, що проходить через початок координат, і будемо переміщувати її в напрямку ∇L. Очевидно, що переміщати лінію рівня в напрямку зростання цільової функції має сенс тільки в межах області припустимих розв'язків. Точкою, у якій цільова функція дістане максимальне значення, у нашому прикладі є точка A з координатами (0, 6). Отже, отриманий оптимальний план задачі

$$x^* = (0, 6),$$

при якому цільова функція приймає максимальне значення

$$L = 1*0 + 3*6 = 18.$$

Теоретично можливі також такі окремі випадки розв'язку ЗЗЛП:

- цільова функція L не обмежена зверху, тобто не має максимуму (рис. 2.2);
- коли лінія рівня збігається із гранню області припустимих розв'язків (рис. 2.3). У цьому випадку всі точки, що лежать на грані множини D, є оптимальними планами й кажуть, що має місце ***альтернативний оптимум***.

У розглянутих ілюстраціях припустимі плани ЗЗЛП мають вигляд опуклої багатогранної множини. Таке подання множини припустимих планів називають першою геометричною інтерпретацією ЗЗЛП.

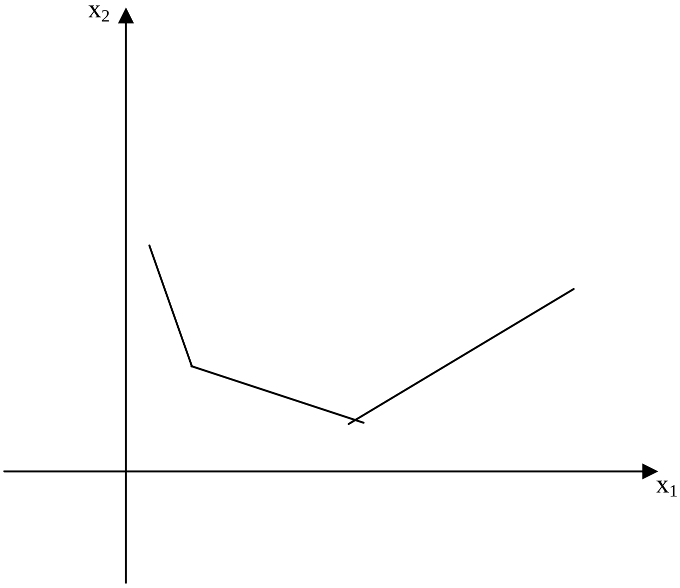

Рис. 2.2 - цільова функція L не обмежена зверху

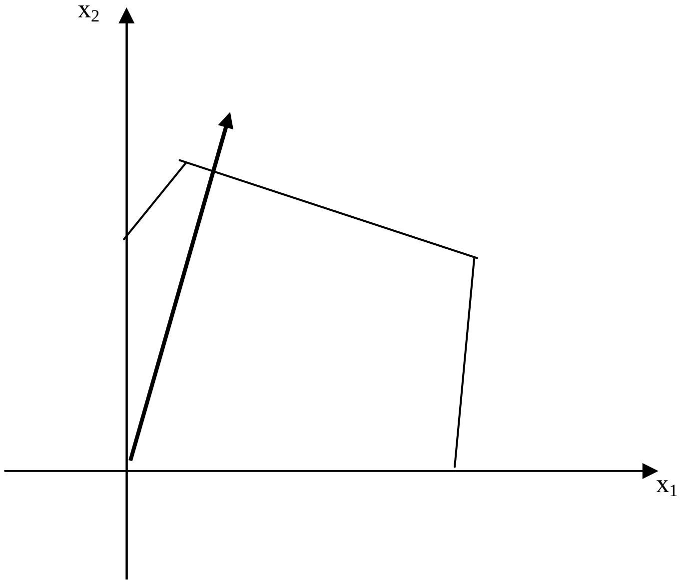

Рис. 2.3 – Альтернативний оптимум

Розглянемо деякі теореми, що відбивають фундаментальні властивості задач лінійного програмування і полягають в основі методів їх розв'язання. Вони узагальнюють на випадок задач із довільною кількістю змінних ті властивості, які ми спостерігали у двовимірному випадку.

Теорема 2.1. Якщо цільова функція L приймає максимальне значення в деякій точці множини припустимих планів D, то вона приймає це значення й у деякій кутовій точці даної множини.

Доказ.

Щоб не ускладнювати виклад, обмежимося тим випадком, коли множина D обмежена, і, отже, є опуклим багатогранником:

Для доказу скористаємося наступною відомою властивістю обмежених опуклих множин:

Якщо D - замкнена обмежена опукла множина, що має кінцеве число кутових точок, то будь-яка точка $x \in D$ може бути поданою у вигляді опуклої комбінації кутових точок D.

Нехай $x_1, x_2, \ldots, x_m$ - кутові точки множини D, а x^* - точка, у якій цільова функція L досягає максимуму. Точку x^* можна уявити у вигляді опуклої комбінації кутових точок $x_1, x_2, \ldots, x_m$

$$x^* = \sum_{i=1}^{m} \lambda_i x_i \,,$$

де $\sum \lambda_i = 1, \quad \lambda_i \ge 0, \quad i = \overline{1, m}$.

Оскільки x^* - точка максимуму, то для будь-якого x $cx^* \ge cx$. У тому числі й для cx_r (x_r – кутова точка).

Функція L(x) - лінійна, тому

$$L\left(\sum_{i=1}^{m} \lambda_i x_i\right) = \sum_{i=1}^{m} \lambda_i L(x_i)\,,$$

а тоді

$$cx^* = c\sum_{i=1}^{m} \lambda_i x_i = \sum \lambda_i (cx_i) \le \sum \lambda_i (cx_r) = cx_r \sum \lambda_i = cx_r \,,$$

де x_r – кутова точка множини D, що задовольняє умові

$$cx_r = \max_{1<i<m} \{cx_i\}.$$

Отже, $cx^* \le cx_r$. У той самий час $cx^* \ge cx_r$, звідки випливає $cx^* = cx_r$.

Тобто існує принаймні одна кутова точка x_r, у якій цільова функція приймає максимальне значення.

Теорема 2.2. Якщо цільова функція L приймає максимальне значення в кількох точках множини D, то вона приймає це саме значення в будь-якій точці, що є їх опуклою комбінацією.

Доказ.

Нехай максимальне значення цільової функції L(x) досягається в точках $x_1, x_2, \ldots, x_s$, тобто $L^* = cx_i, \quad i = \overline{1, s}$. Розглянемо довільну опуклу комбінацію цих точок

$$x^* = \sum_{i=1}^{s} \lambda_i x_i \,,$$

де $\sum \lambda_i = 1, \quad \lambda_i \ge 0, \quad i = \overline{1, s}$.

Знайдемо значення цільової функції в точці x^*

$$L(x^*) = cx^* = c\sum_{i=1}^{s} \lambda_i x_i = \sum \lambda_i (cx_i) = \sum \lambda_i L^* = L^* \sum \lambda_i = L^*.$$

Теорема 2.1 є фундаментальною, тому що вона вказує принциповий шлях розв'язання ЗЛП. Замість дослідження нескінченної множини припустимих розв'язків для знаходження серед них оптимального, необхідно досліджувати лише кінцеве число кутових точок багатогранника розв'язків.

2.3. Симплекс-метод розв'язування ЗЛП

Переважна більшість методів розв'язання задач лінійного програмування призначена для канонічних задач. Тому початковий етап розв'язання будь-якої загальної ЗЛП завжди пов'язаний із приведенням її до еквівалентної КЗЛП.

Канонічною називають задачу лінійного програмування, якщо всі її обмеження є рівняннями. Загальна ідея переходу від ЗЗЛП до КЗЛП досить проста. Обмеження у вигляді нерівностей перетворюють на рівняння за рахунок додавання фіктивних невід'ємних змінних, які одночасно входять до цільової функції з коефіцієнтом 0, тобто не надають впливу на її значення.

$$x'_{n+i}, \quad i = \overline{1,k} .$$

Змінні, на які не накладено умову невід'ємності, подають у вигляді різниці двох нових невід'ємних змінних.

Додатково треба помітити, що вибір типу шуканого екстремуму (максимуму або мінімуму) носить відносний характер. Так, задача пошуку максимуму функції

$$L(x) = \sum_{j=1}^{n} c_j x_j$$

еквівалентна задачі пошуку мінімуму функції

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

У результаті таких перетворень одержують канонічну задачу:

Знайти найбільше або найменше значення функції

$$L = з_1x_1 + з_2x_2 + \ldots + c_nx_n + 0 \cdot x'_{n+1} + \ldots + 0 \cdot x'_{n+k} \tag{2.3}$$

на деякій множині D, де $x \in D$ задовольняє системі обмежень

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x'_{n+1} &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x'_{n+2} &= b_2 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + x'_{n+k} &= b_k \\
a_{k+1,1}x_1 + a_{k+1,2}x_2 + \ldots + a_{k+1,n}x_n &= b_{k+1} \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \\
x_j \geq 0, \quad j = \overline{1, n+k} .
\end{aligned} \tag{2.4}$$

Можна показати, що отримана задача еквівалентна вихідній задачі.

Отже, «платою» за перехід від загальної форми задачі лінійного програмування до канонічної є зростання її розмірності, що, за інших рівних умов, ускладнює процес розв'язання.

Запишемо КЗЛП у матричній формі:

знайти максимум

$$\overline{C}^T \overline{X} \tag{2.5}$$

за умови

$$\overline{A}^T \overline{X} = \overline{B}, \qquad \overline{X} \geq 0 , \tag{2.6}$$

де $\overline{\overline{A}}$, $\overline{X}$ і $\overline{B}$ – вектори.

$$\overline{A_1}=\begin{pmatrix}a_{11}\\a_{21}\\\vdots\\a_{m1}\end{pmatrix};\qquad \overline{A_2}=\begin{pmatrix}a_{12}\\a_{22}\\\vdots\\a_{m2}\end{pmatrix}\qquad \overline{A_n}=\begin{pmatrix}a_{1n}\\a_{2n}\\\vdots\\a_{mn}\end{pmatrix};\qquad \overline{B}=\begin{pmatrix}b_1\\b_2\\\vdots\\b_m\end{pmatrix}. \tag{2.7}$$

Розглядатимемо стовпці матриці $\overline{A}_j$ як вектори простору R^m. Тоді кожному припустимому плану КЗЛП (n-мірному вектору $\overline{X}$) відповідає лінійна комбінація стовпців $\overline{A}_j$, рівна стовпцю $\overline{B}\in R^m$.

Вектори $\overline{A}_j$, $j=\overline{1,n}$ називають векторами вимог задачі. Вектор $\overline{B}$ називають вектором обмежень. При цьому компоненти певного припустимого плану $x \in D$ - x_j – є коефіцієнтами розкладання вектора обмежень $\overline{B}$ задачі за векторами вимог $\overline{A}_j$.

Очевидно, що кількість рівнянь, які задають множину D, менша або дорівнює кількості змінних задачі ($m \le n$). Якщо це не так, то або система рівнянь

$$\overline{A}\ \overline{X} = \overline{B}$$

є несумісною, або вона містить надлишкові (лінійно залежні) рівняння.

Якщо деякі m стовпців $A_1, A_2, \ldots, A_m$ матриці $\overline{A}$ є лінійно незалежними, то вони утворюють ***базис*** у просторі R^m і їх буде досить для уявлення вектора $\overline{B}$ у вигляді лінійної комбінації зазначених стовпців. Це означає, що інші стовпці ввійдуть у дане розкладання з нульовими коефіцієнтами.

Якщо коефіцієнти лінійної комбінації опиняться невід'ємними, то ми отримаємо ***базисний припустимий план*** *x*, в якого не більш за m елементів відмінні від нуля. Або іншими словами. Система обмежень КЗЛП являє собою систему m рівнянь з n змінними, причому $m \le n$.

У такій системі m змінних називаються ***базисними***, а інші (n- m) змінних – ***вільними***.

Базисним розв'язком системи m лінійних рівнянь з n змінними називають розв'язок, у якому всі n-m вільних змінних дорівнюють нулю.

Базисними розв'язками можуть бути різні групи з n змінних. У загальному випадку кількість таких груп не перевершує C_n^m.

Отже, система з m лінійних рівнянь з n змінними ($m<n$) є ***невизначеною***, тому що кожному довільному набору вільних змінних відповідає один базисний розв'язок системи.

Опорним планом КЗЛП називають припустимий базисний розв'язок, компоненти якого більші за нуль.

Базисний розв'язок, у якому хоча б одна з базисних змінних дорівнює нулю, називають ***виродженим***.

Теорема 2.3. Кожний припустимий базисний план є кутовою точкою множини припустимих планів D.

Доказ:

Вважатимемо, що базисними є перші m стовпців матриці $\overline{A}$

$A_1, A_2, \ldots, A_m$...

Тоді формулювання теореми можна записати у вигляді:

Якщо існує такий n-мірний вектор

$$x = \left(x_1, x_2, ..., x_k, \underbrace{0,...,0}_{n-k} \right)$$

$$x_j > 0; \quad j = \overline{1,k}; \quad k \le m,$$

що $x_1 A_1 + x_2 A_2 + ... + x_k A_k = B_0$, то x є кутовою точкою множини D.

Припустимо, що базисний план x *не є* кутовою точкою множини D. У цьому випадку його можна уявити у вигляді опуклої комбінації двох різних припустимих планів x^1 і x^2

$$x = \lambda x^1 + (1-\lambda)x^2, \quad 0 < \lambda < 1.$$

Оскільки останні (n-k) компонент вектора x дорівнюють нулю, а λ і (1-λ) більші за нуль, то ці самі (n-k) компонент у векторах x^1 і x^2 так само дорівнюють нулю.

Оскільки плани x^1 і x^2 – припустимі плани,

$$x_1^1 A_1 + x_2^1 A_2 + ... + x_k^1 A_k = B$$

$$x_1^2 A_1 + x_2^2 A_2 + ... + x_k^2 A_k = B \cdot$$

Віднімемо з першого рівняння друге

$$(x_1^1 - x_1^2)A_1 + (x_2^1 - x_2^2)A_2 + ... + (x_k^1 - x_k^2)A_k = 0 .$$

Отримали нульовий вектор. Оскільки A_j лінійно незалежні, нулю дорівнюють коефіцієнти

$$\begin{aligned} &(x_1^1 - x_1^2) = 0 \\ &(x_2^1 - x_2^2) = 0 \\ &\vdots \\ &(x_k^1 - x_k^2) = 0 \end{aligned} .$$

Звідки випливає, що $x^1 = x^2$. Це суперечить припущенню, що x^1 і x^2 є різними кутовими точками множини D.

Отже, план x не може бути поданий у вигляді опуклої комбінації двох інших точок множини, і є кутовою точкою даної множини.

Справедливо й зворотне ствердження: Якщо x - кутова точка множини D, то вона є припустимим базисним планом задачі лінійного програмування.

На підставі розглянутих теорем про властивості лінійних екстремальних задач можна побачити, що пошук їх розв'язків збігається до послідовного перебору кутових точок множини припустимих планів.

Проте такий перебір для реальних багатомірних завдань на практиці нездійсненний або вкрай неефективний. Наприклад, кількість базисних планів у задачі з 10 змінними й 30 обмеженнями становить

$$C_{30}^{10} = \frac{30!}{10!*20!} = 1489411 .$$

Класичним методом розв'язання задач лінійного програмування є симплекс-метод, що ще називають методом послідовного поліпшення плану. Він розроблений у 1947 році американським математиком Джорджем Данцигом.

Симплекс-метод є послідовним перебором кутових точок області припустимих розв'язків, при якому значення цільової функції зростає від ітерації до ітерації (від однієї кутової точки до іншої). Критерій оптимальності в симплекс-методі реалізують шляхом визначення спеціальних оцінок Δ_j для небазисних векторів-стовпців матриці $\overline{A}$, щодо поточного базису (симплекс-різниць). Симплекс-різниці обчислюють за формулою

$$\Delta_j = L_j - c_j, \tag{2.8}$$

де L_j – індекси векторів, що відповідають поточному базисові

$$\mathrm{L_j = \sum_{i=1}^{m} c_j a_{ij}}\,. \tag{2.9}$$

Сформулюємо критерій оптимальності припустимого базисного плану:

план є оптимальним, якщо для всіх $\mathrm{j = \overline{1, n}}$ $\Delta_j \geq 0$, і неоптимальним у протилежному випадку, тобто якщо існує таке $\mathrm{j = \overline{1, n}}$, що $\Delta_j < 0$.

Якщо симплекс-різниці показують неоптимальність плану, здійснюють перехід до наступного базису. При цьому один стовпець виводять з базису, а інший вводять. Для забезпечення покращення значення цільової функції в базис має бути введеним вектор-стовпець з від'ємною оцінкою. Якщо таких стовпців кілька, то для введення рекомендують обирати стовпець, що має максимальний за модулем добуток оцінки Δ_j на відношення $\Theta_r = \min\limits_i \left\{ \dfrac{b_i}{a_{ij}} \right\}$. Одночасно на цьому кроці потрібно ухвалити рішення щодо того, який стовпець треба вивести з базису. Зробити це потрібно так, щоб знову сформований базис опинився припустимим.

Можна довести, що припустимість базису, до якого здійснюють перехід, забезпечується наступним правилом виводу стовпця з поточного базису:

для стовпця, що претендує на введення до базису, і вектора обмежень $\overline{\mathrm{B}}$ розглядають відношення

$$\Theta_i = \frac{b_i}{a_{ij}}\,. \tag{2.10}$$

і визначають такий рядок r, що

$$\Theta_r = \min_i \{\Theta_i\}. \tag{2.11}$$

Отриманий індекс r визначає номер рядка, який відповідає вектору, виведеному з базису.

Отже, якщо базис на q-ї ітерації містив стовпці з номерами

$$\{j_1, j_2,.., j_{r-1}, j_r, j_{r+1},.., j_m\},$$

то базис на ітерації $q + 1$ складатиметься зі стовпців з номерами:

$$\{j_1, j_2,.., j_{r-1}, j_{r+1},.., j_m\}.$$

Окремо треба домовитися про випадок, коли стовпець, що претендує на введення до базису, не містить додатних компонентів ($a_{ij}<0$). Це означає, що цільова функція в задачі не обмежена на множині припустимих значень, тобто може досягати як завгодно великого значення, отже оптимальний план відсутній.

Після переходу до нового базису можна заново сформувати матрицю $\overline{A}$ й дослідити новий план на оптимальність. З погляду техніки обчислень раціонально безпосередньо переходити від матриці q-ї ітерації до матриці (q+1)-ї ітерації. Справа в тому, що в цих матриць стовпці, які відповідають базисним векторам, складаються з нулів, за винятком одного елемента, що дорівнює одиниці. Позиція цього ненульового (одиничного) елемента визначається порядковим номером базисного стовпця. Тому для одержання матриці (q+1)-ї ітерації досить за допомогою лінійних операцій над рядками матриці q-ї (попередньої) ітерації привести її стовпець, що відповідає вектору, який вводиться до базису, до «базисного» вигляду.

Для цього застосовують перетворення Жордана-Гауса (так званий метод повного виключення). У цьому випадку воно полягає в тому, що ми повинні дістати одиницю на місці елемента a_{rj} (його зазвичай називають ведучим) і нулі на місці інших елементів стовпця a_{ij}. Перше досягається за допомогою поділення r-го рядка на ведучий елемент, друге - шляхом додавання знову отриманого r-го рядка, помноженого на відповідний коефіцієнт, до інших рядків матриці q-ї (попередньої) ітерації.

Формально результат виконання даного перетворення над елементами матриці можна виразити у наступному виді:

$$a_{rj}^{q+1} = \frac{a_{rj}^{q}}{a_{rk}^{q}}, \qquad b_{r}^{q+1} = \frac{b_{r}^{q}}{a_{rk}^{q}}, \tag{2.12}$$

де $j = \overline{1, n}$; k – номер стовпця, що вводять до базису;

$$a_{ij}^{q+1} = a_{ij}^{q} - a_{ik}^{q}\frac{a_{rj}^{q}}{a_{rk}^{q}}, \tag{2.13}$$

де $i = \overline{1, m}$, $i \neq r$;

$$b_{i}^{q+1} = b_{i}^{q} - a_{ik}^{q}\frac{b_{r}^{q}}{a_{rk}^{q}}, \tag{2.14}$$

де $i = \overline{1, m}$, $i \neq r$.

Треба особливо зазначити зміст елементів вектора $\overline{B}$. Його нульовий компонент b_0 містить значення цільової функції, що досягається нею на поточному плані, а інші елементи - ненульові компоненти цього плану.

Приведемо схему алгоритму симплекс-методу для розв'язання задачі максимізації.

1. Знаходять припустимий базисний план.

2. Перевіряють оптимальність поточного базисного плану: здійснюють перегляд рядка оцінок Δ_j . Можливі два варіанти:

- $\Delta_j \geq 0$ — план, що відповідає поточному базису задачі, оптимальний. Обчислювальний процес закінчений. Виписують оптимальний план задачі x^* і значення цільової функції $L(x^*)$.

- у рядку оцінок Δ_j існує щонайменше один елемент Δ_j<0, тобто оцінка якого є від'ємною. Отже, план неоптимальний. Вибирають стовпець із номером k, для якого добуток $\Delta_j\ \Theta$ є найбільшим за абсолютною величиною. Він називається ведучим і повинен бути введений до чергового базису

3. Визначають стовпець, що треба вивести з базису. Досліджують ведучий стовпець, можливі два варіанти:

- для всіх $i=\overline{1,m}$ a_{ik}^{q} < 0. Роблять висновок про необмеженість цільової функції й завершують обчислювальний процес.

- існує принаймні один рядок з номером $i=\overline{1,m}$, для якого a_{ik}^{q} >0. Відповідно до правила (2.11) визначають номер r стовпця, виведеного з базису.

4. Перераховують елементи матриці $\overline{A}$ й стовпця $\overline{B}$ щодо нового базису відповідно до формул (2.12)-(2.14). Переходять до пункту 2 алгоритму.

З погляду забезпечення раціональності й наочності обчислювального процесу виконання алгоритму симплекс-методу зручно оформляти у вигляді послідовності таблиць. У різних джерелах наводять різні модифікації симплекс-таблиць, що відрізняються одна від одної розташуванням окремих елементів. Однак всі вони базуються на тих самих принципах. Зупинимося на наступній структурі таблиці:

Базис	$\boldsymbol{C_{баз}}$	$\boldsymbol{C_j}$					
		B	$\boldsymbol{A_1}$	$\boldsymbol{A_2}$	$\boldsymbol{A_j}$	**…**	$\boldsymbol{A_n}$
$\boldsymbol{A_j}$							
	$\boldsymbol{L_j}$ $\boldsymbol{\Delta_j}$						

Стовпець «Базис» містить найменування базисних векторів (у тій послідовності, у якій вони входять до базису), стовпець $C_{баз}$ – містить коефіцієнти цільової функції при базисних змінних, стовпець В — компоненти вектора обмежень щодо поточного базису, A_1-A_n — компоненти матриці задачі щодо поточного базису. У рядку $\boldsymbol{L_j}$ записують індекси, визначені за формулою (2.9).

У рядку Δ_j містяться поточні оцінки стовпців. Рядок C_j містить коефіцієнти при компонентах поточного плану в цільовій функції.

Слід зазначити, що таблична модифікація симплекс-методу має важливе практичне значення не стільки як зручна форма організації ручного рахунку, скільки як основа для реалізації цього алгоритму в рамках програмного забезпечення ЕОМ.

Розглянемо приклад розв'язання ЗЛП за симплекс-методом. Нехай дана канонічна задача ЛП:

$$L(x)= 50x_1 - 10x_2 + 6x_3 + 40x_4 - 30x_5 \to \max$$
$$x_1 + x_4 + 3x_5 = 12,$$
$$2x_1 + x_2 + 3x_4 = 14,$$
$$-2x_1 + 3x_3 - 4x_4 = 17,$$
$$x_1>0,\ x_2>0,\ x_3>0,\ x_4>0,\ x_5>0.$$

Як видно, стовпці матриці з номерами {5, 2, 3} є лінійно незалежними. І можна одержати розкладання за цими стовпцями вектору обмежень із додатними коефіцієнтами. Останнє означає, що стовпці {5, 2, 3} утворюють припустимий базис, з якого можна почати розв'язання задачі. Початковий опорний план має вигляд x^1 = {0, 14, 17/3, 0, 4}. Заповнюємо симплекс-таблицю, що відповідає першої ітерації (q=1):

Базис	$C_{баз}$	C_j	*50*	*-10*	*6*	*40*	*-30*
		B	A_1	A_2	A_3	A_4	A_5
A_2	*-10*	*14*	*2*	*1*	*0*	*3*	*0*
A_3	*6*	*17/3*	*-2/3*	*0*	*1*	*-4/3*	*0*
A_5	*-30*	*4*	*1/3*	*0*	*0*	*1/3*	*1*
	L_j	*-226*	*-34*	*-10*	*6*	*-48*	*-30*
	Δ_j		*-84*	*0*	*0*	*-88*	*0*

Оскільки рядок оцінок у першому й четвертому стовпцях містить від'ємні елементи Δ_1 = -84, Δ_4= -88, план x^1 = {0, 14, 17/3, 0, 4} не є оптимальним, і значення цільової функції L(x^1) = -226 можна покращити. Перейдемо до нового опорного плану.

Визначимо вектор, що будемо вводити до базису (A_1 або A_4).

Відповідно до алгоритму симплекс-методу визначимо відношення Θ. Для вектора A_1: $\Theta_r = \min\limits_i \left\{ \frac{14}{2}; \frac{4}{1/3} \right\} = \min\{7;12\} = 7$, r=2; для вектора A_4: $\Theta_r = \min\limits_i \left\{ \frac{14}{3}; \frac{4}{1/3} \right\} = \min\{4{,}66;12\} = 4{,}66$, r=2. Добутки для вектора A_1: -34*7= -238; для вектора A_4: -48*4,66= -223,7. Вважаємо номер стовпця, що вводиться в черговий базис, k = 1 (тому що |-238|>|-223,7|). З базису повинен бути виведеним стовпець із номером 2. Отримуємо черговий припустимий базис задачі {1, 3, 5}. Елемент, що перебуває на перетинанні виділених стовпця й рядка таблиці є ведучим, він дорівнює 2. Застосувавши формули (2.12)—(2.14), переходимо до симплекс-таблиці, що відповідає другої ітерації, і вважаємо індекс поточної ітерації q = 2.

Базис	$C_{баз}$	C_j	*50*	*-10*	*6*	*40*	*-30*
		B	A_1	A_2	A_3	A_4	A_5
A_1	*50*	*7*	*1*	*1/2*	*0*	*3/2*	*0*
A_3	*6*	*10,3*	*0*	*0,2*	*1*	*-0,3*	*0*
A_5	*-30*	*1,7*	*0*	*-0,2*	*0*	*-0,2*	*1*
	L_j	*360,8*	*50*	*32,2*	*6*	*79,2*	*-30*
	Δ_j		0	*42,2*	*0*	*39,2*	*0*

Одержуємо новий план x^2 ={7; 0; 10,3; 0; 1,7}. Як видно з таблиці, рядок оцінок містить тільки невід'ємні значення, тому досягнутий базис {1, 3, 5} є оптимальним. Отже, вектор x^*={7; 0; 10,3; 0; 1,7} є оптимальним планом (точкою максимуму) задачі, максимальне значення цільової функції дорівнює $L^* = L(x^*)$ = 360,8.

Найважливішою властивістю будь-якого обчислювального алгоритму є збіжність, тобто можливість одержання в ході його застосування шуканих результатів (із заданою точністю) за кінцеве число кроків (ітерацій).

Легко помітити, що проблеми зі збіжністю симплекс-методу потенційно можуть виникнути на етапі вибору значення r у випадку, коли однакові мінімальні значення відношення $\Theta_r = \min_i \left\{ \frac{b_i}{a_{ij}} \right\}$ будуть досягнуті для кількох рядків таблиці одночасно. Тоді на наступній ітерації стовпець b^{q+1}-й міститиме нульові елементи. Нагадаємо, що припустимий базисний план канонічної задачі ЛП, що відповідає поточному базису, називають виродженим, якщо деякі його базисні компоненти дорівнюють нулю, тобто вектор b^q містить нульові елементи.

Задачу ЛП, що має вироджені плани, називають виродженою. При виході на вироджений план ми фактично одержуємо розкладання вектора $\overline{B}$ на меншу чим *m*, кількість векторів $\overline{A}_j$ і, отже, втрачаємо можливості коректно визначити, введення якого стовпця до базису призведе до зростання значення цільової функції. Подібні ситуації, в остаточному підсумку, можуть призвести до зациклення обчислювального процесу, тобто до нескінченного перебору тих самих базисів.

З погляду першої геометричної інтерпретації ЗЛП ситуація виродженості означає, що через певну кутову точку багатогранної множини припустимих планів задачі, що відповідає поточному базисному плану, проходить більш чим m гіперплощин обмежень задачі. Іншими словами, одне або кілька обмежень у цій точці є надлишковими. Останнє дає повід для міркувань про економічну сторону проблеми виродженості як проблеми наявності надлишкових обмежень і в деяких випадках визначає шляхи її розв'язання.

Ідея методу розв'язання вироджених задач ЛП, що дістала назви ***методу збурювань***, полягає в тому, що при виході на вироджений план здійснюють незначний зсув вектора $\overline{B}$, і виродженість усувається.

Необхідно сказати, що розглянута проблема зациклення для більшості практично значущих задач є досить рідкою й малоймовірною. Більше того, вона дуже часто вирішується за рахунок помилок округлень при виконанні розрахунків на ЕОМ.

Знаходження припустимого базисного плану. У розглянутому вище прикладі вихідний базисний план, необхідний для початку обчислень за симплекс-методом, був підібраний за рахунок особливостей матриці умов. Дійсно, ця матриця вже містила необхідну кількість «майже базисних» стовпців. Очевидно, що для переважної більшості задач ЛП неможливо подібним чином відразу й у явному вигляді вказати вихідний припустимий базисний план. Існують різні прийоми розв'язання цієї задачі. Ми зупинимося на одному з них, що отримав назву ***методу мінімізації нев'язань***. Його сильною стороною, безумовно, є універсальність. Хоча, у деяких окремих випадках, він може опинитися занадто громіздким.

Ідея методу мінімізації нев'язань полягає в побудові відповідної допоміжної задачі, для якої можна в явному вигляді вказати вихідний базисний план, і розв'язанні її за допомогою процедури симплекс-методу.

Можна вважати, що вектор обмежень у вихідній задачі невід’ємний, тобто $b_i > 0, \quad i=\overline{1,m}$. Тоді для КЗЛП максимізації функції

$$L = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max \tag{2.15}$$

на множині, зумовленій обмеженнями

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \\ x_j \ge 0, \quad j=\overline{1,n} \end{gathered} \tag{2.16}$$

будують допоміжну задачу

$$L = 0x_1 + 0x_2 + \ldots + 0x_n - 1x_{n+1} - \ldots - 0 \cdot x_{n+m} \tag{2.17}$$

на новій множині D'={x ∈ R^{n+m}}, зумовленій системою обмежень

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + 1 * x_{n+1} + \ldots + 0 * x_{n+m} = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + 0 * x_{n+1} + 1 * x_{n+2} + \ldots + 0 * x_{n+m} = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + 0 * x_{n+1} + 0 * x_{n+2} + \ldots + 1 * x_{n+m} = b_m \\ x_j \ge 0, \quad j=\overline{1,n+m}. \end{gathered} \tag{2.18}$$

Як видно, допоміжна задача (2.17) - (2.18) відрізняється від основної задачі (2.15) - (2.16) матрицею, що отримана в результаті додавання до вихідної матриці $\overline{A}$ одиничної матриці. Їй відповідають додаткові (фіктивні) змінні $x_{n+1},\ldots,x_{n+m}$. Ці змінні (їх називають нев'язаннями) у цільовій функції мають коефіцієнти (-1), а змінні $x_1,\ldots,x_n$, які є спільними для обох задач, входять до цільової функції з нульовими коефіцієнтами. Істотною особливістю допоміжної задачі є наявність в неї очевидного вихідного базису, утвореного доданими стовпцями, і відповідного плану з базисними компонентами $x_{n+i}=b_i \ge 0$, $i=\overline{1,m}$. В силу структури цільової функції в процесі пошуку її максимуму за процедурою симплекс-методу фіктивні змінні (нев'язання) x_{n+i} повинні мінімізуватися, бажано до нуля, чого можна досягнути за рахунок послідовного переведення їх у небазисні компоненти плану. Якщо в оптимальному плані x^*, отриманому в результаті розв'язання допоміжної задачі, останні m компонент (тобто нев'язання) дорівнюють нулю, то його початкові n компонент задовольняють системі обмежень, які визначають область D, і x^* простим відкиданням нев'язань перетворюють у припустимий план основної задачі. При цьому всі рядки симплекс-таблиці, отриманої на останній ітерації допоміжної задачі, переносять до першої таблиці основної задачі.

У випадку, коли оптимальний план допоміжної задачі все-таки містить не рівні нулю фіктивні компоненти, припустимі плани у вихідної задачі відсутні, тобто D = ∅.

2.4. Поняття двоїстості. Правило побудови двоїстих задач

Більшість задач лінійного програмування можна розглядати як економічні задачі про розподіл ресурсів $b_1, b_2, ..., b_m$ між різними споживачами, зокрема, між технологічними процесами $\overline{A}_1, \overline{A}_2, ..., \overline{A}_n$.

Будь-який припустимий розв'язок задачі лінійного програмування $x_1, x_2, ..., x_n$ дає конкретний розподіл ресурсів, що вказує частку кожного з них, яка буде витрачена під час реалізації певного технологічного процесу.

Загальну задачу лінійного програмування формулюють в такий спосіб:

знайти

$$L = c_1x_1 + c_2x_2 + ... + c_nx_n \to \max, \tag{2.19}$$

де *x* задовольняє обмеженням

$$\begin{array}{l} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \le b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \le b_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \le b_m \\ \qquad x_j \ge 0 \end{array} \tag{2.20}$$

Тут a_{ij} – витрати i-го ресурсу на j-й технологічний процес. Зміст обмежень – сумарні витрати i-го ресурсу повинні не перевищувати кількість цього ресурсу, наявного в розпорядженні b_i.

Для будь-якого технологічного процесу актуальним є запитання про те, щоб витрати ресурсів були мінімальними. Позначимо u_i ціну одиниці i-го ресурсу. Тоді можна записати вимогу, щоб вартість ресурсів була мінімальною

$$L' = b_1u_1 + b_2u_2 + ... + b_mu_m \to \min,$$

а величина витрат на кожний вид технологічного процесу не повинна бути меншою за його вартість

$$\begin{array}{l} a_{11}u_1 + a_{12}u_2 + ... + a_{1m}u_m \ge c_1 \\ a_{12}u_1 + a_{22}u_2 + ... + a_{m2}u_m \ge c_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}u_1 + a_{2n}u_2 + ... + a_{mn}u_m \ge c_n \\ \qquad u_i \ge 0 \end{array}.$$

Отже, одержали задачу:

знайти такий розв'язок u*, що перетворює на мінімум цільову функцію

$$L' = \sum_{i=1}^{m} b_iu_i \to \min \tag{2.21}$$

і задовольняє обмеженням

$$\begin{array}{l} a_{11}u_1 + a_{12}u_2 + ... + a_{1m}u_m \ge c_1 \\ a_{12}u_1 + a_{22}u_2 + ... + a_{m2}u_m \ge c_2 \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{1n}u_1 + a_{2n}u_2 + ... + a_{mn}u_m \ge c_n \\ \qquad u_i \ge 0 \end{array}. \tag{2.22}$$

Цю задачу називають *двоїстою* щодо вихідної задачі, яку називають *прямою* задачею.

Порівняємо умови прямої та двоїстої задач у матричній формі:

Пряма	Двоїста
Знайти таке $\overline{X}^*$, що	Знайти таке $\overline{U}^*$, що
$\overline{C}^T\overline{X}^* \to \max$	$\overline{B}^T\overline{U}^* \to \min$
при обмеженнях	при обмеженнях
$\overline{A}\overline{X} \le \overline{B}, \quad \overline{X} \ge 0$	$\overline{A}^T\overline{U} \ge \overline{C}, \quad \overline{U} \ge 0$

Якщо задано загальну задачу ЛП, де множина припустимих планів D визначається системою рівнянь і нерівностей (2.20), то двоїстою щодо неї називають загальну задачу ЛП, де D' визначається системою рівнянь і нерівностей (2.22).

При переході від прямої задачі ЛП до двоїстої:

- тип оптимуму змінюється на протилежний, тобто якщо пряма задача є задачею максимізації, то двоїста буде задачею мінімізації й навпаки;
- вектор коефіцієнтів цільової функції $\overline{C}$ й стовпець обмежень $\overline{B}$ міняються місцями. Тобто коефіцієнти цільової функції прямої задачі c_1, c_2, ..., c_n стають вільними членами обмежень двоїстої задачі, а вільні члени обмежень прямої задачі b_1, b_2, ..., b_m стають коефіцієнтами цільової функції двоїстої задачі;
- матрицю обмежень двоїстої задачі одержують транспонуванням матриці обмежень прямої задачі $\overline{A}$;
- множину індексів змінних, на які накладено умову невід'ємності в прямій задачі (наприклад, $x_j \ge 0$ або $u_i \ge 0$), визначають номера обмежень, що мають форму нерівностей у двоїстій задачі ($a_j u \ge c_j$ або $a_i x \le b_i$);
- множину номерів обмежень, що мають форму нерівностей у прямій задачі (наприклад, $a_i x \le b_i$ або $a_j u \ge c_j$), визначає множина індексів змінних, на які накладено умову невід'ємності, у двоїстій задачі ($u_i \ge 0$ або $x_j \ge 0$).

З наведених властивостей пари задач випливає важлива властивість — ***симетричність відносини подвійності***, тобто задача, двоїста щодо двоїстої, збігається з прямою (вихідною) задачею.

Тим самим можна говорити про пару взаємно двоїстих задач.

Поняття про двоїсті оцінки. Теореми двоїстості та їх економічний зміст

Фундаментальні властивості двоїстих задач лінійного програмування формулюють у вигляді тверджень, що наведені нижче. Їх зазвичай називають ***теоремами подвійності***.

Теорема 2.3. (***перша теорема подвійності***). Якщо $\overline{X}_0$ та $\overline{U}_0$ - припустимі плани для пари двоїстих задач, тобто якщо

$$\overline{A}^T X_0 \le \overline{B} \qquad і \qquad \overline{A}^T \overline{U}_0 \ge \overline{C},$$

то

$$\overline{C}^T \overline{X_0} \le \overline{B}^T \overline{U_0},$$

тобто значення цільової функції прямої задачі ніколи не перевищують значень цільової функції двоїстої задачі.

Доказ:

Оскільки $\overline{U}_0$ - припустимий план, то

$$\overline{A}^T \overline{U}_0 \geq \overline{C}; \qquad (2.23)$$

оскільки $\overline{X}_0$ - припустимий план, то

$$\overline{A}\ \overline{X}_0 \leq \overline{B}. \qquad (2.24)$$

Помножимо (2.23) на $\overline{X}_0^T$

$$\overline{X}_0^T \overline{A}^T \overline{U}_0 \geq \overline{X}_0^T \overline{C}; \qquad (2.25)$$

помножимо (2.24) на $\overline{U}_0^T$

$$\overline{X}_0 \overline{A}\overline{U}_0^T \leq \overline{U}_0^T \overline{B} \qquad (2.26)$$

і порівняємо (2.25) і (2.26). Оскільки

$$\overline{X}_0 \overline{A}\overline{U}_0^T = \left(\overline{X}_0^T \overline{A}^T \overline{U}_0\right)^T, \qquad \text{то } \overline{U}_0^T \overline{B} \geq \overline{X}_0^T \overline{C},$$

або

$$\overline{U}_0 \overline{B}^T \geq \overline{X}_0 \overline{C}^T.$$

Зауваження. Теорема 2.3, зрозуміло, вірна й для оптимальних планів взаємно двоїстих задач: $\overline{U}*\overline{B}^T \geq \overline{X}*\overline{C}^T$, де $\overline{X}*$ і $\overline{U}*$ — будь-які оптимальні плани задач. Насправді, як буде видно з подальшого, справедлива рівність $\overline{U}*\overline{B}^T = \overline{X}*\overline{C}^T$.

Теорема 2.4. (***друга теорема подвійності***). Якщо для деяких припустимих планів $\overline{X}_0$ і $\overline{U}_0$ взаємно двоїстих задач виконується рівність

$$\overline{C}^T \overline{X}_0 = \overline{B}^T \overline{U}_0,$$

то $\overline{X}_0$ і $\overline{U}_0$ є оптимальними планами цих задач.

Доказ:

Відповідно до теореми 2.3 для всіх припустимих розв'язків $\overline{X}$ і $\overline{U}$ справедлива нерівність

$$\overline{C}^T \overline{X} \leq \overline{B}^T \overline{U},$$

але оскільки з умови теореми

$$\overline{C}^T \overline{X}_0 = \overline{B}^T \overline{U}_0,$$

то $\overline{C}^T \overline{X}_0$ – найбільше з можливих значень цільової функції, тобто

$$\overline{C}^T \overline{X} \leq \overline{C}^T \overline{X}_0$$

цільова функція від $\overline{X}_0$ – найбільша, таким чином, $\overline{X}_0$ є оптимальним значенням.

Аналогічно для цільової функції двоїстої задачі

$$\overline{B}^T \overline{U}_0 \leq \overline{B}^T \overline{U},$$

тобто $\overline{B}^T \overline{U}_0$ – найменше з можливих значень цільової функції двоїстої задачі, отже $\overline{U}_0$ – оптимальне значення.

Отже, якщо для деяких припустимих планів $\overline{X}_0$ і $\overline{U}_0$ прямої й двоїстої задач їх цільові функції рівні, то $\overline{X}_0$ й $\overline{U}_0$ – оптимальні плани пари сполучених задач.

Теорема 2.5. Якщо цільова функція в прямій задачі не обмежена зверху, то двоїста до неї задача не має припустимих планів.

Доказ.

Якщо припустити, що у двоїстій задачі існує хоча б один припустимий план $\overline{U}_0$, то відповідно до теореми 2.4, для будь-якого припустимого плану $\overline{X}_0$ прямої задачі справедлива нерівність $\overline{C}^T\overline{X} \le \overline{B}^T\overline{U} < +\infty$. Останнє означає, що цільова функція прямої задачі обмежена зверху. Оскільки це суперечить умові теореми, припущення про існування припустимих планів двоїстої задачі є невірним.

Наступне ствердження, відоме як ***теорема рівноваги***, використовують при перевірці оптимальності планів ЗЛП.

Теорема 2.6. Нехай $\overline{X}*$ і $\overline{U}*$ — оптимальні плани канонічної й двоїстої щодо неї задач. Якщо j-та компонента плану $\overline{X}*$ строго додатна ($x^*_j>0$), то відповідне j-е обмеження двоїстої задачі виконується як рівність: $a_{1j}u_1+\ldots+a_{mj}u_m=c_j$; якщо j-та компонента плану $\overline{X}*$ має нульове значення ($x^*_j = 0$), то j-е обмеження двоїстої задачі виконується як нерівність: $a_{1j}u_1+\ldots+a_{mj}u_m>c_j$.

Доказ.

Вектори $\overline{X}*$ і $\overline{U}*$ як припустимі плани, задовольняють обмеженням

$$\overline{A}\overline{X}* = \overline{B}; \qquad \overline{X}* > 0 \text{ - прямої задачі}$$

$$\overline{A}^T\overline{U}*-\overline{C}^T \ge 0; \qquad \overline{U}* > 0 \text{ - двоїстої задачі.}$$

Запишемо скалярний добуток

$$\left(\overline{A}^T\overline{U}*-\overline{C}^T\right)\overline{X}* = \overline{A}^T\overline{X}*\overline{U}*-\overline{C}^T\overline{X}* = \overline{B}^T\overline{U}*-\overline{C}^T\overline{X}*.$$

Одержали різницю цільових функцій прямої і двоїстої задач. На підставі другої теореми подвійності оптимальні значення цільових функцій взаємно двоїстих задач збігаються. Отже скалярний добуток

$$\left(\overline{A}^T\overline{U}*-\overline{C}^T\right)\overline{X}* = 0.$$

Але скалярний добуток двох невід'ємних векторів може дорівнювати нулю тільки в тому випадку, якщо всі попарні добутки їх відповідних координат дорівнюють нулю. Тоді, якщо $x_j > 0$, то $\sum_i a_{ij}u_i - c_j = 0$ або $\sum_i a_{ij}u_i = c_j$. А якщо $x_j = 0$, то можливо, що $\sum_i a_{ij}u_i - c_j > 0$ або $\sum_i a_{ij}u_i > c_j$.

Практичне значення теорем подвійності полягає в тому, що вони дозволяють замінити процес розв'язання основної задачі на розв'язання двоїстої, який в певних випадках може виявитися простішим. Наприклад, задачу, область припустимих значень якої описується двома рівняннями, що зв'язують шість змінних (*m*

= 2, n = 6), не можна вирішити графічним методом. Однак цей метод можна застосовувати для розв'язання двоїстої до неї задачі, що має тільки дві змінні.

Оптимальний розв'язок двоїстої задачі можна одержати з таблиці, отриманої на фінальній ітерації процедури симплекс-методу. Елементи індексного рядка цієї таблиці L_j обчислюють відповідно до виразу (2.9) $L_j = \sum_{i=1}^{m} c_j a_{ij}$, де c_j – елементи вектора-рядка, що містить коефіцієнти цільової функції прямої задачі при змінних, які є базисними в оптимальному плані; a_{ij} - елементи матриці $\overline{D}^{-1}$, що зворотна до матриці $\overline{D}$. Матриця $\overline{D}$ складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану задачі. Оптимальний план двоїстої задачі визначається співвідношенням

$$\overline{U}^* = \bar{c}_{j\text{баз}} \overline{D}^{-1}. \qquad (2.27)$$

Зворотна матриця $\overline{D}^{-1}$ завжди міститься в останній симплекс-таблиці в тих стовпчиках, де в першій таблиці задачі перебувала одинична матриця. (Нагадаємо, що добуток матриці на її зворотну дає одиничну матрицю, в якої діагональні елементи дорівнюють 1, а всі інші дорівнюють 0).

Отже, зв'язок між оптимальними розв'язками прямої й двоїстої задач і елементами індексних рядків L_j симплекс-таблиць, що відповідають цим розв'язкам, виражається наступними співвідношеннями:

$$a_{0,n+i}^{\text{ПР}} = u_i^*, \quad i = \overline{1,m},$$

$$-a_{0,m+j}^{\text{ДВ}} = x_j^*, \quad j = \overline{1,n},$$

де n – кількість змінних прямої задачі; m – кількість обмежень прямої задачі; $a_{0,n+i}^{\text{ПР}}$ - $(n+i)$-й елемент індексного рядка симплекс-таблиці прямої задачі, що містить оптимальний план; $a_{0,m+j}^{\text{ДВ}}$ - $(m+j)$-й елемент індексного рядка симплекс-таблиці двоїстої задачі, що містить оптимальний план.

2.5. Інтерпретація двоїстих оцінок в задачах техніко-економічного планування

У різних джерелах компоненти оптимального плану двоїстої задачі називають ***двоїстими оцінками*** або ***тіньовими цінами***. На підставі теорем подвійності для пари задач ЛП у загальній формі формулюють певні важливі з погляду економічної інтерпретації слідства. Зокрема, з теореми 2.6 випливає, що якщо під час реалізації оптимального плану прямої задачі i-е обмеження виконується як строга нерівність, то оптимальне значення відповідної двоїстої змінної дорівнює нулю:

$$a_{i1}x_1^* + \cdots + a_{in}x_n^* < b_i, \text{ то } u_i^* = 0.$$

Це означає, що якщо деякий ресурс b_i є в надлишковій кількості (не витрачається повністю під час реалізації оптимального плану), то i-е обмеження стає неістотним, і тіньова оцінка такого ресурсу дорівнює нулю. Отже, тіньові оцінки характеризують ***дефіцитність ресурсів***.

Якщо під час реалізації оптимального плану двоїстої задачі j-е обмеження виконується як строга нерівність, то оптимальне значення відповідної змінної в оптимальному плані прямої задачі має дорівнювати нулю

$$a_{1j}u_1^* + \cdots + a_{mj}u_m^* > c_j, \text{ то } x_j^* = 0.$$

Цей факт виражає ***принцип рентабельності виробництва***. З огляду на економічний зміст двоїстих оцінок $u_1^*, \cdots, u_m^*$, вираз $a_{1j}u_1^* + \cdots + a_{mj}u_m^*$ можна інтерпретувати як питомі витрати на j-й технологічний процес. Отже, якщо ці витрати перевищують прибуток від реалізації одиниці j-го продукту, виробництво j-го продукту є нерентабельним і не повинне бути присутнім в оптимальному виробничому плані ($x_j^* = 0$).

Незважаючи на можливі аналогії, які можуть виникнути у зв'язку з такими фундаментальними поняттями економічної теорії, як граничні витрати й граничний дохід, двоїсті оцінки не можна однозначно ототожнювати із цінами (хоча такі спроби іноді вживали на початковій стадії становлення дослідження операцій як науки).

2.6. Аналіз двоїстих оцінок. Параметричні зміни вектору обмежень

Традиційна економічна інтерпретація двоїстої задачі ЛП базується на моделі найпростішої задачі виробничого планування. В неї кожний j-й елемент вектора $\overline{X}$ розглядається як план випуску продукції певного виду в натуральних одиницях, c_j - ціна одиниці продукції j-го виду, $\overline{A}_j$ - вектор, що визначає технологію витрати наявних m ресурсів на виробництво одиниці продукції j-го виду, $\overline{B}$ - вектор обмежень на обсяги цих ресурсів.

Припустимо, що для певних значень $\overline{A}_j$, $\overline{B}$ і c_j знайдений оптимальний план x^* прямої задачі, що максимізує сумарний дохід $\max_{x \in D}\{cx\} = cx^*$, і визначено оптимальні оцінки сировини, тобто оптимальний план двоїстої задачі u*. З виразу цільової функції двоїстої задачі $L'^* = b_1u_1^* + b_2u_2^* + \ldots + b_mu_m^*$ видно, що величина двоїстої оцінки u_i^* показує, наскільки зросте максимальне значення цільової функції прямої задачі при збільшенні кількості сировини відповідного виду на одну одиницю. Отже, змінні двоїстої задачі $u_1^*, \cdots, u_m^*$ за своїм змістом є оцінками потенційної можливості одержання додаткового прибутку за рахунок збільшення обсягу відповідного ресурсу в умовах оптимального функціонування керованого економічного об'єкту.

Виникає питання про те, як буде змінюватися оптимальний план x^* під час зміни компонентів вектора обмежень $\overline{B}$ і, зокрема, за яких варіацій $\overline{B}$ оптимальний план x^* залишиться оптимальним. Ця задача одержала назву проблеми ***стійкості оптимального плану***. Очевидно, що дослідження стійкості x^* має й практичне значення, тому що в реальному виробництві обсяги доступних ресурсів b_i можуть істотно коливатися після ухвалення планового розв'язку x^*.

Коли вектор обмежень $\overline{\mathrm{B}}$ отримує приріст Δb, виникають відповідні варіації для оптимального плану прямої задачі x*(b+Δb) і значення цільової функції L[x*(b+Δb)]. Припустимо, приріст Δb такий, що він не призводить до зміни оптимального базису задачі, тобто *x**(b+Δb)≥0. Введемо функцію F(b), що повертає оптимальне значення цільової функції задачі для різних значень вектору обмежень $\overline{B}$

$$F(b) = \max_{x \in D(b)} \{cx\}. \tag{2.28}$$

Розглянемо відношення її приросту F(b+Δb)-F(b) до приросту аргументу Δb. Якщо для деякого і спрямувати Δb→0, одержимо

$$\lim_{\Delta b_i \to 0} \frac{F(b+\Delta b) - F(b)}{\Delta b_i} = \frac{\partial F(b)}{\partial b_i}. \tag{2.29}$$

З огляду на те, що відповідно до теореми 2.4 цільові функції пари сполучених задач при оптимальних планах дорівнюють одна одной, запишемо

$$F(b) = \sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i u_i^*. \tag{2.30}$$

Підставимо (2.30) до (2.29) і одержимо вираз

$$\frac{\partial F(b)}{\partial b_i} = \frac{\partial \left(\sum_{i=1}^{m} b_i u_i^* \right)}{\partial b_i} = u_i^* \tag{2.31}$$

Теорема 2.7. В оптимальному плані двоїстої задачі значення змінної u_i* чисельно дорівнює частинній похідній цільової функції L'* за відповідним аргументом b_i.

Звідси випливає економічна інтерпретація оптимальних змінних двоїстої задачі:

Кожний елемент $u_1^*, \cdots, u_m^*$ може розглядатися як гранична (миттєва) оцінка внеску і-го ресурсу в сумарний дохід L* при оптимальному розв'язку $x_1^*, \cdots, x_n^*$.

Інакше кажучи, u_i^* дорівнює приросту доходу, що виникає при збільшенні ресурсу і на одиницю за умови оптимального використання ресурсів.

Отже, якщо знайдено оптимальний план прямої задачі, можна провести аналіз стійкості двоїстих оцінок щодо змін компонентів вектору $\overline{B}$. Це дозволяє оцінити стійкість оптимального плану двоїстої задачі щодо зміни обмежень прямої задачі й ступінь впливу зміни $\overline{B}$ на максимальне значення цільової функції, а так само визначити найдоцільніший варіант можливих змін $\overline{B}$.

План двоїстої задачі не змінюється для всіх значень $b_i + \Delta b_i$, за яких стовпець вектора $\overline{B}*$ останньої симплекс-таблиці не містить від'ємних чисел, тобто коли серед компонентів вектору

$$\overline{B}^* = \begin{vmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \dots\dots\dots \\ b_m + \Delta b_m \end{vmatrix} \qquad (2.32)$$

немає від'ємних. $\overline{B}^*$ – матриця, складена з компонентів векторів базису, що визначає оптимальний план задачі, оскільки базисні компоненти оптимального плану перебувають у стовпці вектора $\overline{B}$ останньої симплекс-таблиці.

Елементи (n+i)-го стовпця a_{ij} останньої симплекс-таблиці, що містить оптимальну оцінку i-го ресурсу u_i*, показують, на скільки одиниць зміняться компоненти оптимального плану x* під час збільшення обсягу цього ресурсу на одиницю, тобто $x_j^*(b_i+\Delta b_i)=x_j^*(b_i)+a_{i,n+i}\Delta b_i$. Припустимі інтервали зміни для i-го ресурсу можна визначити з умови:

$$\overline{B}^* = \begin{vmatrix} x_1 + a_{1,n+i}\Delta b_i \geq 0 \\ x_2 + a_{2,n+i}\Delta b_i \geq 0 \\ \dots\dots\dots\dots\dots\dots \\ x_n + a_{n,n+i}\Delta b_i \geq 0 \end{vmatrix}. \qquad (2.33)$$

Розглянемо приклад. Нехай остання симплекс-таблиця, що містить оптимальний план, має вигляд

Базис	***Cj**баз*	***Cj***	***6***	***5***	***0***	***0***	***0***
		B	***A_1***	***A_2***	***A_3***	***A_4***	***A_5***
A_1	6	83	1	0	4/7	-3/7	0
A_2	5	36	0	1	-3/7	4/7	0
A_5	0	21	0	0	11/28	-11/7	1
індексний рядок	***Lj***	*678*	*6*	*5*	*9/7*	*2/7*	*0*
	Δj		0	*0*	*9/7*	*2/7*	*0*

Звідки оптимальний план прямої задачі x*=(83; 36; 0; 0; 21), оптимальний план двоїстої задачі u*=($\frac{9}{7}$; $\frac{2}{7}$; 0). Дефіцитнішою є сировина S_1, тому що її тіньова оцінка вища й відповідно сильніше впливає на величину цільової функції.

Очевидно, що збільшення доходу можна отримати тільки шляхом зміни оптимального плану прямої задачі. З таблиці видно, що при збільшенні на 1 одиницю кількості сировини S_1, дохід збільшиться на $\frac{9}{7}$ грн. Це відбудеться якщо виробництво виробів А збільшити на $\frac{4}{7}$ одиниці, а виробництво виробів В знизити на $\frac{3}{7}$ одиниці, при цьому витрата сировини S_3 зросте на $\frac{11}{28}$ одиниці. Новий оптимальний план прямої задачі матиме вигляд x*'=($83\frac{4}{7}$; $35\frac{4}{7}$; 0; 0; $20\frac{17}{28}$), а прибуток становитиме $L^*=6*83\frac{4}{7} +5*35\frac{4}{7} + 0*0+ +0*20\frac{17}{28}$ = 679,286 грн.

Визначимо інтервали стійкості двоїстих оцінок. Для ресурсу 1 відповідно до елементів стовпчика A_3 маємо

$$x^* = (83 + 0{,}57\Delta b_1; 36 - 0{,}43\Delta b_1; 0; 0; 21 + 0{,}39\Delta b_1).$$

Запишемо вектор $\overline{B}^*$ з умовами його невід'ємності й визначимо межі припустимих значень Δb_1

$$\overline{B}^* = \left| \begin{array}{c} 83+0{,}57\Delta b_1 \geq 0 \\ 36-0{,}43\Delta b_1 \geq 0 \\ 21+0{,}39\Delta b_1 \geq 0 \end{array} \right| = \left| \begin{array}{c} \Delta b_1 \geq -66{,}7 \\ \Delta b_1 \leq 83{,}7 \\ \Delta b_1 \geq -53{,}8 \end{array} \right|$$

Оптимальний план двоїстої задачі залишиться незмінним, якщо Δb_1 належить інтервалу $-53{,}8 \leq \Delta b_1 \leq 83{,}7$, а перший ресурс $440 - 53{,}7 \leq b_1 \leq 440 + 83{,}7$ або $386{,}3 \leq b_1 \leq 523{,}7$.

Аналогічно для ресурсу 2 відповідно до елементів стовпчика A_4 запишемо

$$x^* = (83 - 0{,}43\Delta b_2; 36 + 0{,}57\Delta b_2; 0; 0; 21 - 1{,}57\Delta b_2).$$

$$\overline{B}^* = \left| \begin{array}{c} 83-0{,}43\Delta b_2 \geq 0 \\ 36+0{,}57\Delta b_2 \geq 0 \\ 21-1{,}57\Delta b_2 \geq 0 \end{array} \right| = \left| \begin{array}{c} \Delta b_2 \leq 193 \\ \Delta b_2 \geq -63{,}2 \\ \Delta b_2 \leq 13{,}4 \end{array} \right|$$

Оптимальний план двоїстої задачі залишиться незмінним, якщо Δb_2 належить інтервалу $-63{,}2 \leq \Delta b_2 \leq 13{,}4$, а другий ресурс $393 - 63{,}2 \leq b_2 \leq 393 + 13{,}4$ або $329{,}8 \leq b_2 \leq 406{,}4$.

Отже, якщо збільшення кількості ресурсів S_1 належить проміжку $-53{,}8 < \Delta b_1 < 83{,}7$, а кількість інших ресурсів незмінна, або збільшення кількості ресурсів S_2 належить проміжку $-63{,}2 < \Delta b_2 < 83{,}7$, а кількість інших ресурсів незмінна, то двоїста задача має той самий оптимальний план

$$u^* = (1{,}286; 0{,}286; 0).$$

Стосовно прямої задачі, можна показати, що під час зміни кількості першого ресурсу S_1 у межах $386{,}3 \leq b_1 \leq 523{,}7$ можливий дохід підприємства лежить у межах $609{,}7 \leq L^* \leq 784{,}3$ а оптимальний план прямої задачі

$$(52{,}3; 59{,}1; 0; 0; 0{,}02) \leq x^* \leq (131; 0; 0; 0; 53{,}6).$$

Піл час зміни кількості другого ресурсу S_2 у межах $329{,}8 \leq b_2 \leq 406{,}4$ можливий дохід підприємства лежить у межах $661 \leq L^* \leq 681{,}6$ а оптимальний план прямої задачі

$$(110; 0; 0; 0; 53{,}6) \leq x^* \leq (77; 43{,}6; 0; 0; 0).$$

Розраховані інтервали належать випадкам, коли змінюється тільки один ресурс. У випадку одночасної зміни обсягу всіх або кількох ресурсів для визначення нового оптимального плану використовують одне з основних співвідношень обчислювальної процедури симплекс-методу:

$$x^* = \overline{D}^{-1} * \overline{B}, \tag{2.34}$$

де $\overline{D}$ - матриця, що складається з базисних векторів оптимального плану, компоненти яких узяті з початкового опорного плану; $\overline{B}$ - вектор обмежень.

У розглянутому числовому прикладі матриці $\overline{D}$ й $\overline{D}^{-1}$ відповідно мають вигляд

$$\overline{D} = \begin{vmatrix} 4 & 3 & 0 \\ 3 & 4 & 0 \\ 3 & 5 & 1 \end{vmatrix}; \qquad \overline{D}^{-1} = \begin{vmatrix} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{vmatrix}$$

Нехай новий вектор обмежень

$$\overline{B} = \begin{vmatrix} 440+84 \\ 393+13,4 \\ 450+0 \end{vmatrix} = \begin{vmatrix} 524 \\ 406,4 \\ 450 \end{vmatrix};$$

тоді новий оптимальний план визначиться в такий спосіб

$$x^* = \begin{vmatrix} 4/7 & -3/7 & 0 \\ -3/7 & 4/7 & 0 \\ 11/28 & -11/7 & 1 \end{vmatrix} * \begin{vmatrix} 524 \\ 406,4 \\ 450 \end{vmatrix} = \begin{vmatrix} 125,26 \\ 7,66 \\ 17,2 \end{vmatrix}$$

Тобто x^*=(125,26; 7,66; 0; 0; 17,2), при якому прибуток дорівнюватиме 790 грош. од.

2.7. Параметричні зміни вектора цільової функції

З погляду економічної інтерпретації задачу дослідження параметричної стійкості можна розглядати як вивчення тих меж коливання цін на продукцію керованого підприємства (фірми), за яких прийнятий план випуску продукції залишається оптимальним. Отже, питання стійкості оптимального плану ЗЛП можна поставити для випадку варіації коефіцієнтів цільової функції

$$c_j, \quad j = \overline{1, n}.$$

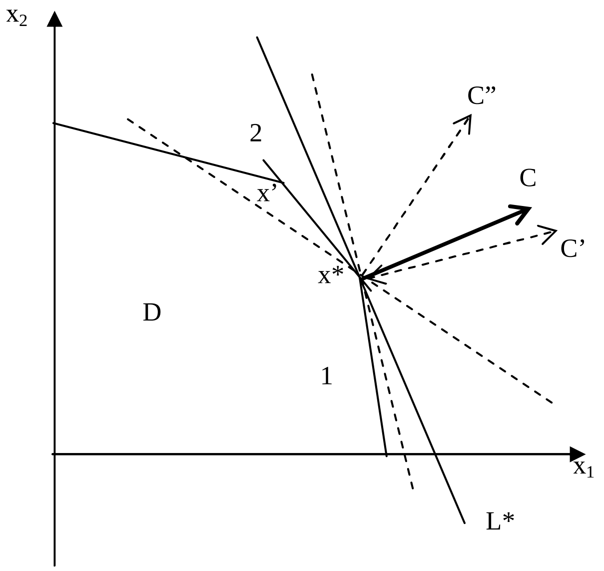

Рис. 2.4 - Графічна інтерпретація параметричної стійкості оптимального плану

Зміст проблеми стійкості оптимального плану ЗЛП стосовно варіацій цільової функції можна проілюструвати за допомогою першої геометричної інтерпретації. На рис. 5.1 зображено множину припустимих планів D деякої задачі ЛП. Як видно з рисунка, цільова функція L досягає екстремального значення в точці x^*, а зміні її коефіцієнтів від c до c' або c'' на рисунку відповідає поворот лінії рівня відносно x^*. Активним обмеженням (тобто таким, що звертаються на рівність) у точці x^* відповідають лінії 1 та 2. Доти, поки під час повороту, викликаному зміною вектора c, лінія рівня цільової функції не виходить за межі утворені лініями обмежень множини, x^* залишається оптимальним планом. Як показано на рис. 5.1, цей план не змінюється під час переходу від c до c', і, навпаки, під час переходу від c до c'' лінія рівня цільової функції L(x)=c''x перетинає лінію 2, що викличе зміну оптимального базисного плану, яким тепер стане точка x'. Використовуючи умови оптимальності плану ЗЛП

$$\Delta_j = L_j - c_j \geq 0, \tag{2.35}$$

можна одержати кількісні оцінки для меж коливань коефіцієнтів цільової функції, за яких не відбувається зміна оптимального плану. Припустимо, що варіації піддався певний елемент $c_r' = c_r + \Delta c_r$. Можливі два випадки:

1. Стовпець *r* не входить до оптимального базису. Тоді для незмінності оптимального плану необхідно й достатньо виконання умови

$$\Delta_r' = L_r - c_r' \geq 0.$$

Звідси можна одержати значення для припустимої варіації

$$\Delta c_r \leq L_r - c_r. \tag{2.36}$$

2. Стовпець r входить до оптимального базису. У цьому випадку для збереження оптимальності поточного плану потрібно виконання для всіх небазисних стовпців умов (2.35) або

$$\Delta_j = \sum_{i=1}^{m} c_{j_{баз}} a_{ij} - c_j \geq 0, \tag{2.37}$$

оскільки у цьому випадку зміни відбуваються так само у стовпчику $C_{баз}$ симплекс-таблиці, а це, у свою чергу, стосується всіх ненульових оцінок Δ_j.

Отже, у цьому випадку припустима варіація має задовольняти умовам

$$\Delta c_r \leq \sum_{i=1}^{m} c_{j_{баз}} a_{ij} - c_j. \tag{2.38}$$

Повернемося до числового прикладу й визначимо межі зміни параметрів цільової функції, за яких знайдений план $x^*=(83;\ 36;\ 0;\ 0;\ 21)$ залишається оптимальним. У цій задачі інтерес представляють варіації коефіцієнтів c_1 та c_2, які стоять при базисних змінних в оптимальному плані.

Запишемо умови (2.38) для коефіцієнта c_1

$$\begin{cases} \Delta_3 = L_3 - c_3 = (6+\Delta c_1)*4/7 + 5*(-3/7) + 0*11/28 - 0 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \\ \Delta_4 = L_4 - c_4 = (6+\Delta c_1)*(-3/7) + 5*4/7 + 0*11/7 - 0 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \dfrac{24}{7} + \dfrac{4}{7}\Delta c_1 - \dfrac{15}{7} \geq 0 \\ \Delta_4 = -\dfrac{18}{7} - \dfrac{3}{7}\Delta c_1 + \dfrac{20}{7} \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_1 \geq -\dfrac{9}{4} \\ \Delta c_1 \leq \dfrac{2}{3} \end{cases}.$$

$$-\frac{9}{4} \leq \Delta c_1 \leq \frac{2}{3} \qquad 3{,}75 \leq c_1 \leq 6{,}67.$$

Аналогічно визначимо варіацію коефіцієнта c_2.

$$\begin{cases} \Delta_3 = L_3 - c_3 = 6*4/7 + (5+\Delta c_2)*(-3/7) + 0*11/28 - 0 = \dfrac{9}{7} - \dfrac{3}{7}\Delta c_2 \\ \Delta_4 = L_4 - c_4 = 6*(-3/7) + (5+\Delta c_2)*4/7 + 0*(-11/7) - 0 = \dfrac{2}{7} + \dfrac{4}{7}\Delta c_2 \end{cases} \Rightarrow$$

$$\begin{cases} \Delta_3 = \dfrac{9}{7} - \dfrac{3}{7}\Delta c_2 \geq 0 \\ \Delta_4 = -\dfrac{2}{7} + \dfrac{4}{7}\Delta c_2 \geq 0 \end{cases} \Rightarrow \begin{cases} \Delta c_2 \leq 3 \\ \Delta c_2 \geq \dfrac{1}{2} \end{cases}.$$

$$\frac{1}{2} \leq \Delta c_2 \leq 3 \qquad 5{,}5 \leq c_2 \leq 8.$$

Наведений приклад дослідження чутливості оптимального плану щодо зміни параметрів задачі є простим. Існують і складніші задачі, в яких, наприклад, досліджуються спільні варіації параметрів різних типів. Вони складають предмет спеціального розділу дослідження операцій, що одержав назву параметричного програмування.

Контрольні запитання

1. Сформулюйте задачу лінійного програмування.
2. Дайте визначення для наступних понять: план, припустимий план, оптимальний план, розв'язок задачі.
3. Чим відрізняється загальна задача лінійного програмування від канонічної?
4. Чи завжди загальну задачу лінійного програмування можна привести до канонічного виду?
5. Яку точку опуклої множини називають кутовою?
6. В чому полягає перша геометрична інтерпретація задачі лінійного програмування?
7. Який план ЗЛП називають базисним?
8. Як пов’язані базисні плани й кутові точки області визначення задачі лінійного програмування?
9. Який план задачі лінійного програмування називають виродженим?
10. Сформулюйте критерій оптимальності припустимого базисного плану, застосовуваний у симплекс-методі.
11. Сформулюйте основні етапи стандартної ітерації симплекс-методу.
12. Для чого застосовують перетворення Жордана-Гаусса?
13. Який елемент симплекс-таблиці називають ведучим?
14. За яких умов роблять висновок про необмеженість цільової функції в розв'язуваній задачі?
15. Чи можна заздалегідь точно визначити кількість ітерацій, що потрібна для розв'язання задачі симплекс-методом? Чи можна знайти верхню границю для даної величини?
16. Яку задачу називають виродженою? За якими ознаками можна впізнати, що поточний план є виродженим?
17. Поясніть, в чому полягає основна ідея методу збурювань?
18. Для чого призначений метод мінімізації нев'язань?
19. Поясніть суть подвійності в лінійному програмуванні.
20. Складіть просту економіко-математичну модель та запишіть до неї двоїсту. Дайте економічну інтерпретацію двоїстих оцінок.
21. Скільки змінних і обмежень має двоїста задача щодо прямої задачі?
22. Поясніть економічний зміст першої теореми подвійності.
23. Поясніть економічний зміст другої теореми подвійності.
24. У чому полягає економічний зміст третьої теореми подвійності?
25. Сформулюйте правила побудови двоїстих задач.
26. Як на підставі оптимального розв’язку прямої задачі одержати оптимальний розв’язок двоїстої задачі?
27. У чому полягає економічна інтерпретація прямої та двоїстої задач лінійного програмування?
28. Як визначити, чи є ресурс дефіцитним?
29. Як визначити, що продукція є рентабельною або нерентабельною?
30. У чому полягає економічний зміст змінних двоїстої задачі?
31. Який зміст вкладають у поняття «параметрична стійкість»?
32. Сформулюйте умови для припустимих змін цільової функції задачі, за яких її оптимальний план залишається незмінним.
33. Як визначити статус ресурсів прямої задачі?
34. Як визначити інтервали стійкості двоїстих оцінок щодо зміни запасів дефіцитних ресурсів?
35. Як визначити оптимальний план виробництва продукції й зміну доходу підприємства під час збільшення або зменшення обсягу ресурсів?
36. Як розрахувати інтервали можливої зміни ціни одиниці кожного виду продукції?

ТЕМА 3
ТРАНСПОРТНА ЗАДАЧА ТА ЗАДАЧА ПРО ПРИЗНАЧЕННЯ
3.1. Економічна постановка транспортної задачі. Математична модель

Нехай є m пунктів відправлення (постачальників) $A_1, A_2,... A_m$, у яких знаходиться однорідна продукція в кількостях $a_1, a_2,...,a_m$ відповідно. Нехай є n пунктів призначення (споживачів) $B_1, B_2,..., B_n$, що подали заявки відповідно на $b_1, b_2,...,b_n$ одиниць вантажу. Сума всіх заявок дорівнює сумі всіх запасів.

Відомі вартості c_{ij} перевезення одиниці продукції з i-го пункту відправлення в j-й пункт призначення. Вважається, що вартість перевезення кількох одиниць вантажу пропорційна їх кількості.

Потрібно скласти такий план перевезень (звідки, куди й скільки одиниць везти), щоб всі заявки були виконані, а загальна вартість всіх перевезень була мінімальною.

Поставимо цю задачу як задачу лінійного програмування. Позначимо x_{ij} — кількість одиниць вантажу, що відправляється з i-го ПВ A до j-го ПП B.

Сукупність чисел x_{ij} називатимемо «планом перевезень», а саме величини x_{ij} — «перевезеннями». Ці невід'ємні змінні повинні задовольняти наступним умовам:

1. Сумарна кількість вантажу, що направляється з кожного ПВ до усіх ПП, повинна дорівнювати запасу вантажу в даному пункті.
2. Сумарна кількість вантажу, що доставляється до кожного ПП із усіх ПВ, повинна дорівнювати заявці, поданої даним пунктом.
3. Сумарна вартість всіх перевезень, тобто сума величин x_{ij}, помножених на відповідні вартості c_{ij}, має бути мінімальною.

Отже, задача збігається до визначення такого плану перевезень певного продукту з пунктів його виробництва до пунктів споживання $x= (x_{11}, \ldots, x_{1n}, x_{21},\ldots,x_{2n}, \ldots, x_{i1}, \ldots, x_{in}, \ldots, x_{m1}, \ldots, x_{mn})$, який мінімізує цільову функцію

$$f(x)=\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}\,. \tag{3.1}$$

на множині припустимих планів

$$\sum_{j=1}^{n} x_{ij}=a_i\,;\ \sum_{i=1}^{m} x_{ij}=b_j\,;\ x_{ij}\ge 0\,;\ i=\overline{1,m};\quad j=\overline{1,n} \tag{3.2}$$

при дотриманні умови балансу

$$\sum_{i=1}^{m} a_i=\sum_{j=1}^{n} b_j\,. \tag{3.3}$$

Якщо умова (3.3) виконується, то задачу називають *збалансованою* або *закритою*, а інакше задачу називають *незбалансованою* або *відкритою*.

Транспортна задача є представником класу задач лінійного програмування й тому має всі якості лінійних оптимізаційних задач, але одночасно вона має й ряд додаткових корисних властивостей, які дозволили розробити спеціальні методи для її розв'язання.

Кількість лінійно незалежних серед рівнянь в умовах-обмеженнях транспортної задачі (ТЗ) дорівнює

$$m + n - 1,$$

отже кількість базисних змінних так само дорівнює

$$m + n - 1.$$

Загальна кількість змінних x_{ij} у транспортній задачі дорівнює $m * n$, а кількість вільних змінних

$$k = mn - (m + n - 1) = (m - 1)(n - 1).$$

Відомо, що в задачі лінійного програмування оптимальний розв'язок досягається в одній з вершин ОДР, в опорній точці, де принаймні k змінних дорівнюють нулю. Виходить, у нашому випадку для оптимального плану принаймні $(m-1)(n-1)$ перевезень матимуть дорівнювати нулю (з відповідних ПВ у відповідні ПП нічого не перевозиться).

Називатимемо будь-який план перевезень *припустимим*, якщо він задовольняє умовам-обмеженням (всі заявки задоволені, всі запаси вичерпані).

Припустимий план будемо називати *опорним*, якщо в ньому відмінні від нуля не більш за $m + n-1$ базисних перевезень, а інші $(m-1)(n-1)$ перевезень дорівнюють нулю.

План називатимемо *оптимальним*, якщо він, серед всіх припустимих планів, призводить до мінімальної сумарної вартості перевезень (L=min).

В силу особливої структури ТЗ під час її розв'язання не приходиться довго розв'язувати систему рівнянь. Всі операції із знаходження оптимального плану збігаються до маніпуляцій безпосередньо з таблицею, де в певному порядку записані умови транспортної задачі: перелік ПВ й ПП, заявки й запаси, а також вартості перевезень c_{ij}. У міру заповнення цієї таблиці в її клітках проставляють самі перевезення x_{ij}. Транспортна таблиця складається з m рядків і n стовпців. Рядки транспортної таблиці відповідають пунктам виробництва (в останній клітці кожного рядка зазначений обсяг запасу продукту a_i), а стовпці — пунктам споживання (остання клітка кожного стовпця містить значення заявки b_j). У правому верхньому куті кожної клітки ставлять вартість c_{ij} перевезення одиниці продукту з A_i до B_j, а центр клітки залишають вільним, щоб поміщати до нього саме перевезення x_{ij}. Клітки, які містять нульові перевезення (x_{ij}=0), називають вільними, а ненульові - зайнятими (x_{ij}>0).

3.2. Методи знаходження опорного плану транспортної задачі

За аналогією з іншими задачами лінійного програмування розв'язання транспортної задачі починається з побудови припустимого базисного плану. Найпростіший спосіб його знаходження ґрунтується на так званому ***методі північно-західного кута***. Суть методу полягає в послідовному розподілі всіх запасів, наявних у першому, другому і т.д. пунктах виробництва, до першого, другого і т.д. пунктів споживання. Кожний крок розподілу зводиться до спроби повного вичерпання запасів у черговому пункті виробництва або до спроби повного задоволення потреб у черговому пункті споживання. На кожному кроці q величини поточних нерозподілених запасів позначають a_i^q, а поточних незадоволених потреб — b_j^q. Побудову припустимого початкового плану, відповідно до методу північно-західного кута, починають з лівого верхнього кута транспортної таблиці. Для чергової клітки, розташованої в рядку і і стовпці j, розглядають значення нерозподіленого запасу в і-ому пункті виробництва й незадоволеної заявки в j-ому пункті споживання, з них вибирають мінімальне й призначають як обсяг перевезення між даними пунктами: $x_{ij} = \min\{a_i^q, b_j^q\}$. В результаті цього значення нерозподіленого запасу й незадоволеної потреби у відповідних пунктах зменшуються:

$$a_i^{(q+1)} = a_i^q - x_{ij}; \qquad b_j^{(q+1)} = b_j^q - x_{ij}.$$

Очевидно, що на кожному кроці виконується хоча б одна з рівностей: $a_i^{(q+1)} = 0$ або $b_j^{(q+1)} = 0$. Якщо справедливо $a_i^{(q+1)} = 0$, це означає, що весь запас i-го пункту виробництва вичерпаний і необхідно перейти до розподілу запасу в пункті виробництва i + 1, тобто переміститися до наступної клітки униз за стовпцем. Якщо $b_j^{(q+1)} = 0$, то повністю задоволена заявка для j-го пункту, після чого виконують перехід на клітку, розташовану праворуч за рядком. Знову обрана клітка стає поточною, і для неї повторюють всі перелічені операції.

Ґрунтуючись на умові балансу запасів і заявок (3.3), неважко довести, що за кінцеве число кроків буде отриманий припустимий план. У силу тієї самої умови число кроків алгоритму не може бути більшим за $m+n$-1, тому завжди залишаться вільними (нульовими) mn-($m+n$-1) кліток. Отже, отриманий план є базисним. Не виключено, що на певному проміжному кроці поточний нерозподілений запас опиниться рівним поточній незадоволеній заявці ($a_i^q = b_j^q$). У цьому випадку перехід до наступної клітки відбувається в діагональному напрямку (одночасно змінюються поточні пункти виробництва й призначення), а це означає «втрату» одного ненульового компонента в плані або, виродженість побудованого плану.

Розглянемо приклад. З 3-х пунктів виробництва необхідно вивезти однорідний продукт в 5 пунктів споживання. Транспортні витрати, обсяг виробництва й споживання наведені в таблиці 3.1.

Таблиця 3.1 - Вихідні дані

	B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
A_1	7	5	2	8	7	125
A_2	8	9	4	6	9	60
A_3	5	1	9	2	3	115
Заявки	30	50	100	40	80	300

Зауважимо, що запаси дорівнюють заявкам. Отже, задача є збалансованою.

Визначимо кількість базисних змінних

$$m + n - 1 = 3 + 5 - 1 = 7.$$

Кількість вільних змінних

$$(n-1)*(m-1) = (5-1)*(3-1) = 8.$$

Іншими методами визначення початкового опорного плану є метод мінімальної вартості, метод подвійної переваги або метод апроксимації Фогеля.

У таблиці 3.2 показаний процес пошуку припустимого плану за ***методом північно-західного кута***, включаючи послідовну зміну обсягу нерозподілених запасів і незадоволених потреб. Стрілки відображають траєкторію переходу по клітках транспортної таблиці, а цифри, що знаходяться за її межами, - поточні нерозподілені залишки після призначення обсягу для чергової клітки.

Таблиця 3.2 - Визначення опорного плану методом північно-західного кута

	B_1	B_2	B_3	B_4	B_5	Запаси				
A_1	30 $\Rightarrow$	50 $\Rightarrow$	45 $\Downarrow$			125	95	45	0	
A_2			55 $\Rightarrow$	5 $\Downarrow$		60			5	0
A_3				35 $\Rightarrow$	80	115			80	0
Заявки	30	50	100	40	80	300				
	0	0	55							
			0	35						
				0	0					

Знайдений опорний план x = (30, 50, 45, 0, 0, ...). Значення цільової функції при цьому плані перевезень

$$L = 30*7+50*5+45*2+55*4+5*6+35*2+80*3 = 1110.$$

Особливістю припустимого плану, побудованого за методом північно-західного кута, є те, що цільова функція на ньому приймає значення, як правило, далеке від оптимального. Це відбувається тому, що при його побудові ніяк не враховуються значення c_{ij}. У зв'язку із цим на практиці для одержання вихідного плану використовують інший спосіб - метод мінімального елемента, у якому при розподілі обсягів перевезень у першу чергу займають клітки з найменшими цінами.

Метод мінімальної вартості полягає в тому, що в таблиці вартостей обирають найменшу, і в цій клітці записують найменше з чисел (a_i, b_j), таблиця 3.3.

Таблиця 3.3 - Визначення опорного плану за методом мінімальної вартості

	B_1	B_2	B_3	B_4	B_5	Запаси
A_1	7 **25**	5	2 **100**	8	7	125
A_2	8 **5**	9	4	6	9 **55**	60
A_3	5	1 **50**	9	2 **40**	3 **25**	115
Потреби	30	50	100	40	80	300

Значення цільової функції при цьому плані перевезень

$$L = 25*7+100*2+5*8+55*9+50*1+40*2+25*3 = 1115.$$

Метод подвійної переваги показаний у таблиці 3.4. Він полягає в тому, що в кожному стовпці позначають знаком V клітку з найменшою вартістю, потім те саме роблять у кожному рядку. У клітки з W заносять найбільші обсяги перевезень. Потім розподіляють перевезення по клітках, позначеним V. У частині таблиці, що залишилася, перевезення розподіляють за методом найменшої вартості.

Таблиця 3.4 - Визначення опорного плану за методом подвійної переваги

	B_1	B_2	B_3	B_4	B_5	Запаси
A_1	7 **25**	5	2 **100** W	8	7	125
A_2	8 **5**	9	4 V	6	9 **55**	60
A_3	5 V	1 **50** W	9	2 **40** V	3 **25** V	115
Заявки	30	50	100	40	80	300

Отриманий опорний план збігається із планом за методом мінімальної вартості, L = 1115.

Приймати як опорний треба той план, для якого транспортні витрати опинилися найменшими. Отже, за опорний треба прийняти план, отриманий за методом північно-західного кута.

Цей план є ***припустимим***, тому що суми за рядками дорівнюють запасам, а суми за стовпцями дорівнюють заявкам.

Отриманий план є ***опорним,*** тому що число ненульових перевезень дорівнює $m + n - 1 = 3 + 5 - 1 = 7$, а число нульових перевезень дорівнює $(n\text{-}1)*(m\text{-}1) = (5\text{-}1)*(3\text{-}1) = 8$.

План можна покращити (табл. 3.2), якщо зменшити перевезення в дорогій клітці, наприклад (1.1), і збільшити в дешевій (3.1).

Щоб при цьому план залишався опорним, необхідно одну з вільних кліток зробити базисною, а одну з базисних - вільною.

3.3. Метод потенціалів

Одним з методів розв'язання транспортної задачі є метод, що одержав назву ***методу потенціалів***. Вперше він був запропонований в 1949 р. Л. В. Канторовичем і М. К. Гавуріним. Пізніше на базі загальних ідей лінійного програмування аналогічний метод був запропонований Дж. Данцигом.

Так само як транспортна задача є окремим випадком задачі ЛП, так і метод потенціалів, загалом кажучи, може трактуватися як різновид симплексних процедур. Він являє собою ітеративний процес, на кожному кроці якого розглядають деякий поточний базисний план, перевіряють його оптимальність, і при необхідності здійснюють, перехід до «кращого» базисного плану.

Алгоритм методу потенціалів починається з вибору деякого припустимого базисного плану. Якщо початковий опорний план має ***m+n-1*** додатних перевезень, то його називають ***невиродженим***. Якщо опорний план має менше за ***m+n-1*** додатних перевезень, то його називають ***виродженим***.

У цьому початковому опорному плані кожному пункту ставлять у відповідність деяке число, називане його ***попереднім потенціалом***. Якщо даний план ***невироджений*** (число ненульових базисних кліток дорівнює m+n-1) то за ним можна так визначити потенціали u_i і v_j, щоб для кожної базисної клітки (тобто для тієї, в якій $x_{ij}>0$) виконувалася умова

$$v_j - u_i = c_{ij}, \quad \text{якщо } x_{ij} > 0. \tag{3.4}$$

Теорема 3.1. Якщо план ТЗ є оптимальним, то йому відповідає система з $m+n$ чисел, що задовольняють умовам:

$$u_i + v_j = c_{ij} \quad \text{для} \quad x_{ij} > 0,$$

$$u_i + v_j \le c_{ij} \quad \text{для} \quad x_{ij} = 0, \qquad i = \overline{1,m}; \quad j = \overline{1,n}.$$

де u_i і v_j - потенціали постачальників і споживачів відповідно.

Отже, якщо хоча б для однієї вільної клітки

$$\Delta_{ij} = u_i + v_j - c_{ij} > 0, \tag{3.5}$$

то план не оптимальний і вимагає поліпшення.

Оскільки система (3.4) містить $m+n$-1 рівнянь й $m+n$ невідомих, то один з потенціалів можна задати довільно (наприклад, дорівняти v_j або u_i, до нуля). Після цього інші невідомі u_i і v_j- визначаються однозначно.

Розглянемо процес визначення потенціалів поточного плану транспортної задачі на прикладі. У таблиці 3.3 переписані умови задачі з таблиці 3.1 і її припустимий базисний план, побудований за методом північно-західного кута з таблиці 3.2.

Потенціал першого пункту виробництва приймаємо рівним нулю (u_1=0). Тепер, знаючи його, можна визначити потенціали для всіх пунктів споживання, пов'язаних з першим пунктом виробництва ненульовими перевезеннями. У цьому випадку їх три (це перший, другий і третій пункти), отримаємо:

$$v_1 = c_{11}-u_1= 7 - 0 = 7;\ v_2 = c_{12}-u_1=5-0 = 5;\ v_3 = c_{13}-u_1=2-0 = 2.$$

Маючи v_3 і з огляду на те, що в другому рядку таблиці існують ненульові компоненти x_{23} і x_{24}, можна визначити $u_2 = c_{23}- v_3$ =4 - 2 =2, $v_4= c_{24}- u_2$ =6 - 2 = 4, після чого з'являється можливість розрахувати $u_3=c_{34}-v_4$ =2-4 = -2 і, нарешті, $v_5=$ $=c_{35}-u_3$ =3 - (-2) = 5. В результаті отримали повну систему потенціалів, показану в таблиці 3.5.

Таблиця 3.5 - Визначення потенціалів для початкового опорного плану

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=5	
		B$_1$	**B$_2$**	**B$_3$**	**B$_4$**	**B$_5$**	**Запаси**
u_1=0	**A$_1$**	7 30	5 50	2 45	8	7	125
u_2=2	**A$_2$**	8	9	4 55	6 5	9	60
u_3=-2	**A$_3$**	5	1	9	2 35	3 80	115
	Заявки	30	50	100	40	80	300

Для вільних кліток транспортної таблиці обчислюють величини $\Delta_{ij}=v_j+u_i- -c_{ij}$. У таблиці 3.6 вони виписані для всіх небазисних кліток під цінами.

Таблиця 3.6 - Перевірка оптимальності поточного плану (обчислення оцінок Δ_{ij})

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=5	
		B$_1$	**B$_2$**	**B$_3$**	**B$_4$**	**B$_5$**	**Запаси**
u_1=0	**A$_1$**	7 30	5 50	2 45	8 0	7 -2	125
u_2=2	**A$_2$**	8 1	9 -2	4 55	6 5	9 -2	60
u_3=-2	**A$_3$**	5 0	1 2	9 -9	2 35	3 80	115
	Заявки	30	50	100	40	80	300

Відповідно до теореми 3.1, якщо всі $\Delta_{ij}\leq 0$, план є оптимальним, у протилежному випадку, якщо існує хоча б одна клітка, для якої $\Delta_{ij} > 0$, його можна поліпшити. Процес «поліпшення» плану полягає у визначенні клітки, що вводиться, й клітки, що виводиться. У цьому простежується змістовна аналогія методу з відповідними пунктами симплекс-процедур.

Кандидатом на введення може бути будь-яка клітка, у якій $\Delta_{ij} > 0$, оскільки після введення її до базису буде забезпечена рівність $v_j + u_i = c_{ij}$. Для визначеності рекомендується брати ту клітку, у якій оцінка Δ_{ij} максимальна. У розглянутому нами прикладі це буде клітка (3, 2).

Виведена клітка визначається за допомогою так званого ланцюжка перетворення плану, що описує характер перерозподілу вантажних потоків. У відповідності із властивостями транспортної задачі для невиродженого базисного плану в поточній таблиці можна утворити замкнутий ланцюжок, що складається з вертикальних і горизонтальних ланок, однією з вершин якого є обрана вільна клітка, а інші - зайняті клітки. У таблиці 3.7 показаний ланцюжок перетворення поточного плану щодо клітки, яка вводиться до нього, (3, 2).

Таблиця 3.7 - Перетворення поточного плану

		v_1=7	v_2=5	v_3=2	v_4=4	v_5=5	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
u_1=0	**A_1**	7 30	5 50 -Θ	2 45 +Θ	8 4	7 5	125
u_2=2	**A_2**	8 5	9 3	4 55 -Θ	6 5 +Θ	9 3	60
u_3=-2	**A_3**	5 9	1 +Θ 7	9 4	2 35 -Θ	3 80	115
	Заявки	30	50	100	40	80	300

Логіка алгоритму побудови ланцюжка досить проста: «вийшовши» із клітки (3, 2) у горизонтальному напрямку, ми повинні «зупинитися» у тій зайнятій клітці плану, з якої зможемо рухатися далі за вертикаллю. У даному прикладі цій вимозі задовольняють як клітка (3, 4), так і клітка (3, 5). Однак ланцюжок від (3, 5) не може бути продовжений далі, у той час як рухаючись від (3, 4) за вертикаллю до (2, 4) і далі до (2.3), ми вертаємося через клітки (1, 3) і (1, 2) до вихідної клітки (3, 2) і утворюємо замкнений цикл.

У побудованому ланцюжку, починаючи із клітки, що вводиться (яка вважається першою), позначаються вершини: непарні — «+Θ», а парні «-Θ». Знаком «+» позначають ті клітки, у яких обсяги перевезень повинні збільшитися (такою, зокрема, є клітка, що вводиться до плану, оскільки вона повинна стати базисною). Знаком «-» — ті клітки, у яких перевезення зменшуються з метою збереження балансу. Серед множини кліток, позначених знаком «-», обирають клітку з найменшим значенням x_{ij}. Вона й стає кандидатом на вивід, тому що зменшення обсягу перевезень на більшу величину може призвести до від'ємних значень x_{ij} в інших «мінусових» клітках. Потім провадиться перерахування плану за ланцюжком: до обсягів перевезень у клітках, позначених знаком «+», додається обсяг Θ, а з обсягів кліток, позначених знаком «-», він віднімається. У результаті вводу однієї клітки й виводу іншої утворюється новий базисний план, для якого на наступній ітерації описані вище дії повторюються.

У нашому прикладі знаком «-» позначені клітки (3, 4), (2, 3) і (1, 2), причому x_{34}=35, x_{23} =55, x_{12} =50. Обчисливши значення Θ= min{x_{34}, x_{23}, x_{12}} = 35, здійснюємо перетворення й переходимо до наступного базисного плану, показаному в таблиці 3.8.

Таблиця 3.8 - Оцінка оптимальності наступного базисного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=7$	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
$u_1=0$	**A_1**	7 30	5 15	2 80	8 -4	7 0	125
$u_2=2$	**A_2**	8 1	9 -2	4 20	6 40	9 0	60
$u_3=-4$	**A_3**	5 -2	1 35	9 -11	2 -2	3 80	115
	Заявки	30	50	100	40	80	300

Для знов отриманого плану повторюють дії стандартної ітерації: розраховують потенціали й оцінки для небазисних кліток транспортної таблиці. Як можна бачити, план у таблиці 3.6 також не є оптимальним (у клітці (2, 1) Δ_{21} =1>0), тому знову будуємо ланцюжок перетворення плану й переходимо до наступного базисного плану за ланцюжком в таблиці 3.9.

Таблиця 3.9 - Перетворення поточного базисного плану

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=4$	$v_5=7$	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
$u_1=0$	**A_1**	7 30 -Θ	5 15	2 80 +Θ	8 -4	7 0	125
$u_2=2$	**A_2**	8 +Θ 1	9 -2	4 20 -Θ	6 40	9 0	60
$u_3=-4$	**A_3**	5 -2	1 35	9 -11	2 -2	3 80	115
	Заявки	30	50	100	40	80	300

Визначивши Θ=20, одержимо

Таблиця 3.10 - Оптимальний базисний план

		$v_1=7$	$v_2=5$	$v_3=2$	$v_4=5$	$v_5=7$	
		B_1	**B_2**	**B_3**	**B_4**	**B_5**	**Запаси**
$u_1=0$	**A_1**	7 10	5 15	2 100	8 -3	7 0	125
$u_2=1$	**A_2**	8 20	9 -3	4 -1	6 40	9 -1	60
$u_3=-4$	**A_3**	5 -2	1 35	9 -11	2 -1	3 80	115
	Заявки	30	50	100	40	80	300

З транспортної таблиці 3.10 видно, що отриманий план оптимальний, тому що всі оцінки для небазисних кліток $\Delta_{ij} \le 0$, тобто $u_i + v_j$ не перевищують відповідних цін c_{ij}. За цим планом обчислюють оптимальне (найменше) значення сумарних витрат на перевезення

$$L^{*}=10x_7+15x_5+100x_2+20x_8+40x_6+35x_1+80x_3=1020.$$

Зупинимося на ситуації виникнення виродженого плану. Якщо задача вироджена, то на якомусь етапі розв'язання може виявитися, що таблиця містить менше за $m+n$-1 заповнених кліток. Це, зокрема, може відбутися, якщо однакове мінімальне значення буде досягнуто відразу на кількох клітках, позначених знаком «-». Для подолання виродженості в транспортній задачі поточний план доповнюють необхідною кількістю нульових кліток (фіктивними перевезеннями) так, щоб вони дозволяли розрахувати повну систему потенціалів, і далі діяти відповідно до правил описаного вище алгоритму. Тобто, якщо не вистачає k заповнених кліток, то їх вважають фіктивно заповненими. Фактично такий прийом є аналогом методу збурювань для транспортної задачі як окремого випадку ЗЛП. До такого висновку легко прийти, якщо покласти, що фіктивні клітки, що додаються, містять певний малий обсяг ε.

Розглянемо приклад. З трьох кар'єрів до чотирьох склозаводів возять глину. Вартості перевезень, потужності кар'єрів і потреби заводів наведені в таблиці 3.11.

Таблиця 3.11 - Вихідні дані

	B_1	**B_2**	**B_3**	**B_4**	**Запаси**
A_1	3	9	7	4	50
A_2	6	8	10	6	65
A_3	5	4	7	6	50
Потреби	50	20	65	30	165

Опорний план визначимо за методом найменшої вартості.

Таблиця 3.12 - Початковий опорний план

		$v_1=3$	$v_2=5$	$v_3=8$	$v_4=4$	
		B_1	**B_2**	**B_3**	**B_4**	**Запаси**
$u_1=0$	**A_1**	3 50	9 -4	$+\Theta$ 7 1	$-\Theta$ 4 [0]	50
$u_2=2$	**A_2**	6 -1	8 -1	$-\Theta$ 10 35	$+\Theta$ 6 30	65
$u_3=-1$	**A_3**	5 -3	4 20	7 30	6 -3	50
	Потреби	50	20	65	30	165

У таблиці 3.12 заповнених кліток 5, а необхідно $m+n-1 = 3+4-1 = 6$. Потрібна одна фіктивно заповнена клітка. Вважатимемо клітку (1, 4) заповненою. Перевіримо план на оптимальність.

План не оптимальний, тому що $\Delta_{13} = 1$. $\Theta = \min(35,0) = 0$. План залишається таким самим, але фіктивно заповненою буде клітка (1, 3). Перевіримо його на оптимальність.

Таблиця 3.13 - Перевірка базисного плану на оптимальність

		$v_1=3$	$v_2=4$	$v_3=7$	$v_4=3$	
		B_1	**B_2**	**B_3**	**B_4**	**Запаси**
$u_1=0$	**A_1**	3 50	9 -5	7 0	4 -1	50
$u_2=3$	**A_2**	6 0	8 -1	10 35	6 30	65
$u_3=0$	**A_3**	5 -2	4 20	7 30	6 -3	50
	Потреби	50	20	65	30	165

Всі $\Delta_{ij} \leq 0$, отже план є оптимальним. Вартість перевезень при такому плані становить $L = 50*3+20*4+0*7+35*10+30*7+30*6 = 970$.

3.4. Змістовна постановка задачі про призначення. Математична модель задачі вибору

Нехай є n різних робіт, кожну з яких може виконати будь-який з n виконавців. Вартість виконання i-ї роботи j-м виконавцем відома і дорівнює c_{ij} (у умовних грошових одиницях). Необхідно розподілити виконавців за роботами (призначити одного виконавця на кожну роботу) так, щоб мінімізувати сумарні витрати, пов'язані з виконанням всього комплексу робіт.

У дослідженні операцій завдання, сформульоване вище, відоме як завдання про призначення. Введемо змінні x_{ij}, де x_{ij} приймає значення 1 у разі, коли i-у роботу виконує j-й виконавець, і значення 0 у всій решті випадків, $i, j = \overline{1, n}$. Тоді обмеження

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = \overline{1, n}$$

гарантує виконання кожної роботи лише одним виконавцем, обмеження

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = \overline{1, n}$$

гарантує, що кожен виконавець виконуватиме лише одну роботу. Вартість виконання всього комплексу робіт дорівнює

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \quad .$$

Отже, завдання про призначення можна записати таким чином

$$\begin{cases} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \to \min \\ \sum_{j=1}^{n} x_{ij} = 1, \quad i = \overline{1, n}; \quad \sum_{i=1}^{n} x_{ij} = 1, \quad j = \overline{1, n};. \\ x_{ij} \in \{0,1\}, \quad i = \overline{1, n}, \quad j = \overline{1, n} \end{cases} \tag{3.6}$$

Задача про призначення (3.6) є окремим випадком класичної транспортної задачі), в якій треба покласти n = m, $\sum_{j=1}^{n} x_{ij} = 1$, $i = \overline{1, n}$, $\sum_{i=1}^{n} x_{ij} = 1$, $j = \overline{1, n}$. При цьому умова $x_{ij} \in \{0,1\}$ означає виконання вимоги цілочисельності змінних x_{ij}. Це пов'язано з тим, що потужності всіх джерел і стоків дорівнюють одиниці, звідки випливає, що в припустимому цілочисельному рішенні значеннями змінних можуть бути тільки 0 та 1.

Як окремий випадок класичної транспортної задачі, задачу про призначення можна розглядати як задачу лінійного програмування. Тому в даному випадку використовують термінологію і теоретичні результати лінійного програмування.

У задачі про призначення змінні x_{ij}, можуть приймати значення 0 або 1. При цьому, відповідно до (3.6), в будь-якому припустимому рішенні лише n змінних можуть приймати значення 1. Отже, будь-яке припустиме базисне рішення задачі про призначення буде виродженим.

На практиці зустрічаються задачі про призначення в постановках яких параметр c_{ij} розуміють як ефектність виконання i-ї роботи j-м виконавцем. У цих задачах треба так розподілити роботи між виконавцями, щоб сумарна ефектність їх виконання була максимальною, тобто

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} \to \max, \qquad (3.7)$$

де максимум шукають при обмеженнях, вказаних у (3.6).

Параметри c_{ij} задачі про призначення (3.6) зручно подавати матрицею $C = (c_{ij})$, яку називають матрицею вартості. Припустимо, що $C^* = (c^*ij)$ і $C = (c_{ij})$ - дві матриці вартості, елементи яких пов'язані таким чином:

$$c^*_{ij} = c_{ij}+d_i+l_j,$$

де d_i та l_j - певні постійні. Отже, для отримання матриці C^* треба до елементів кожного i-го рядка матриці C додати число d_i, а до елементів її кожного j-го стовпця - число l_j. В цьому випадку, якщо X - припустиме рішення, що задовольняє обмеженням з (3.6), і

$$f(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}, \qquad f^*(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} c^*_{ij} x_{ij},$$

то з урахуванням обмежень з (3.6) типу рівностей маємо

$$f^*(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij} + d_i + l_j) x_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij} + \\ + \sum_{i=1}^{n} d_i \left(\sum_{j=1}^{n} x_{ij} \right) + \sum_{j=1}^{n} l_j \left(\sum_{i=1}^{n} x_{ij} \right) = f(X) + \gamma,$$

де $\gamma = \sum_{i=1}^{n} d_i + \sum_{j=1}^{n} l_j$.

Отже, для будь-якого припустимого рішення X відповідні йому значення функцій f та f* відрізнятимуться на постійну γ,. яка не залежить від X. Тому, якщо є дві задачі про призначення з тією самою множиною D припустимих рішень і цільовими функціями f та f* відповідно, то їх оптимальні рішення співпадають. Неважко переконатися у наявності аналогічної властивості і в класичної транспортної задачі.

Якщо задача про призначення є задачею максимізації, тобто шукається максимум цільової функції на множині D припустимих рішень, яка задається системою обмежень з (3.6), то еквівалентну їй задачу мінімізації

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (-c_{ij}) x_{ij} \to \min \qquad (3.8)$$

формально не можна віднести до задач про призначення, оскільки коефіцієнти її цільової функції не є позитивними. Цю невідповідність можна подолати, замінивши (3.8) еквівалентною задачею

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (-c^*_{ij}) x_{ij} \to \min,$$

у якій

$$c_{ij}^* = c_{ij} - \max_{i=\overline{1,n}} c_{ij}, \quad j = \overline{1,n},$$

оскільки в цьому випадку для всіх i та j має місце нерівність $-c_{ij}^* \geq 0$.

Приклад 3.1. Припустимо, що невелика компанія одержує замовлення на виконання чотирьох видів робіт, для кожної з яких відома продуктивність чотирьох її штатних робітників. Продуктивність i-го робітника під час виконання їм j-ї роботи визначає відповідний елемент наступної матриці C денних доходів компанії (матриці вартості):

$$C = \begin{pmatrix} 1 & 8 & 4 & 1 \\ 5 & 7 & 6 & 5 \\ 3 & 5 & 4 & 2 \\ 3 & 1 & 6 & 3 \end{pmatrix}.$$

Сформулюємо задачу про розподіл штатних робітників фірми за видами робіт, що дає максимальний сумарний денний дохід, як задачу про призначення.

В цьому випадку

$$l_1 = \max_i c_{i1} = 5, \qquad l_3 = \max_i c_{i3} = 6,$$

$$l_2 = \max_i c_{i2} = 8, \qquad l_4 = \max_i c_{i4} = 5,$$

Таким чином, маємо n = 4, а коефіцієнти при змінних моделі в цільовій функції представлені матрицею

$$-C^* = \left(-c_{ij}^*\right) = -\left(c_{ij} - l_j\right) = \begin{pmatrix} 4 & 0 & 2 & 4 \\ 0 & 1 & 0 & 0 \\ 2 & 3 & 2 & 3 \\ 2 & 7 & 0 & 2 \end{pmatrix}.$$

3.5. Угорський метод розв'язання задачі про призначення

Нагадаємо, що ця задача є окремим випадкам класичної транспортної задачі і, як наслідок, є задачею транспортного типу. Стосовно задачі про призначення симплексний метод не ефективний, оскільки будь-яке її припустиме базисне рішення є виродженим. Специфічні особливості задачі про призначення дозволили розробити ефективний метод її рішення, відомий як «угорський метод».

Основна ідея угорського методу полягає в переході від початкової матриці вартості C до еквівалентної їй матриці вартості C* з невід'ємними елементами і системою n незалежних нулів, тобто з сукупністю нульових елементів матриці, з яких ніякі два не належать тому самому рядку або тому самому стовпцю. Зрозуміло, що квадратна матриця порядку n не може мати систему більш ніж з n незалежних нулів. Згідно (3.6), можна стверджувати, що для будь-якого припустимого рішення задачі про призначення матриця змінних X = (x_{ij}) містить n(n - 1) нулів та n одиниць, з яких ніякі дві не належать тому самому рядку або тому самому стовпцю. Отже, якщо ці n одиниць в матриці X розставити відповідно до розташування елементів системи n незалежних нулів еквівалентної матриці вартості C*, то одержимо припустиме рішення даної задачі про призначення. Більш того, знайдене припустиме рішення є оптимальним рішенням, оскільки йому відповідає нульове значення цільової функції, визначуваною матрицею C*, яке не може бути зменшене через позитивність елементів еквівалентної матриці вартості C*.

Алгоритм угорського методу розв’язання задачі про призначення складається з підготовчого етапу та не більше за n-2 ітерацій, що послідовно повторюються. На підготовчому етапі одержують матрицю вартості C_0, еквівалентну матриці вартості цієї задачі про призначення і таку, що містить первинну систему незалежних нулів. На кожній ітерації кількість незалежних нулів в перетвореній еквівалентній матриці вартості збільшується не менше ніж на одиницю. Через кінцеву кількість ітерацій система незалежних нулів в перетвореній еквівалентній матриці вартості C* складатиметься з n елементів, що означає завершення процесу рішення даної задачі.

Підготовчий етап полягає в послідовному виконанні наступних трьох кроків.

Крок 1. Для кожного стовпця матриці вартості C задачі про призначення знаходять мінімальний елемент

$$l_j = \min_{i=1,n} c_{ij}, \quad j = \overline{1,n}$$

Звернемося до прикладу. Проілюструємо підготовчий етап алгоритму угорського методу для наступної матриці вартості:

$$C = \begin{pmatrix} 5 & 6 & 7 & 1 \\ 10 & 4 & 6 & 7 \\ 8 & 5 & 3 & 5 \\ 12 & 5 & 9 & 8 \end{pmatrix}.$$

Крок 1. У першому стовпці мінімальним є елемент l_1=c_{11} = 5, в другому l_2=c_{22} = 4, в третьому l_3=c_{33} = 3, і четвертому l_4=c_{14} = 1. Віднімаємо ці значення з відповідних стовпців і приходимо до матриці

$$C' = \begin{pmatrix} 0 & 2 & 4 & 0 \\ 5 & 0 & 3 & 6 \\ 3 & 1 & 0 & 4 \\ 7 & 1 & 6 & 7 \end{pmatrix}.$$

Крок 2. Знаходимо мінімальні елементи в рядках матриці C’. У першому рядку d_1 = c'_{11}=0, у другому d_2 = c'_{22}=0, 0, у третьому d_3 = c'_{33}=0, у четвертому d_4 = c'_{42}=0. Віднімаючи ці значення з відповідних рядків, одержуємо матрицю

$$C_0 = \begin{pmatrix} 0 & 2 & 4 & 0 \\ 5 & 0 & 3 & 6 \\ 3 & 1 & 0 & 4 \\ 6 & 0 & 5 & 6 \end{pmatrix}.$$

Крок 3. Відзначаємо систему незалежних нулів:

$$C_0 = \begin{pmatrix} 0* & 2 & 4 & 0 \\ 5 & 0* & 3 & 6 \\ 3 & 1 & 0* & 4 \\ 6 & 0 & 5 & 6 \end{pmatrix}, \text{ або } C_0 = \begin{pmatrix} 0* & 2 & 4 & 0 \\ 5 & 0 & 3 & 6 \\ 3 & 1 & 0* & 4 \\ 6 & 0* & 5 & 6 \end{pmatrix}.$$

Тепер перейдемо до опису ітерації алгоритму угорського методу рішення задачі про призначення. Нехай проведено m ітерацій (m≥0) в результаті яких одержана еквівалентна матриця вартості C_m.

Крок 1. У матриці C_m підраховуємо кількість елементів в системі незалежних нулів, яку позначимо k. Якщо k = n, то переходимо до кроку 2. Якщо k < n, то переходимо до кроку 3.

Крок 2. Відповідно до системи n незалежних нулів еквівалентної матриці вартості $C^*=C_m$ у матриці змінних моделі X розставляємо n одиниць, а решту її елементів замінюємо нулями. Рішення завершене.

Крок 3 . Стовпці матриці C_m, що містять 0*, виділяємо знаком «+», їх елементи називають виділеними (решта елементів матриці називається невиділеними). Переходимо до кроку 4.

Крок 4. Якщо серед невиділених елементів матриці C_m є хоч би один нуль, то переходимо до кроку 5. Інакше переходимо до кроку 9.

Крок 5. Якщо рядок, що містить невиділений нуль, містить так само і 0*, то переходимо до кроку 6. Інакше переходимо до кроку 7.

Крок 6. Знайдений невиділений нуль позначаємо через 0', рядок, що містить його, відзначаємо знаком «+» і всі її елементи називаємо виділеними. Знімаємо знак «+» із стовпця, в якому розташований 0* з виділеного рядка, і переходимо до кроку 4.

Крок 7. Знайдений невиділений нуль позначаємо 0' і, починаючи з нього, будуємо так званий L-ланцюжок за наступним правилом: початковий 0', далі 0*, розташований з ним у одному стовпці (якщо такий є); потім 0', розташований в одному рядку з передуючим 0*; далі 0*, розташований в одному стовпці з передуючим 0' (якщо такий є) і т.д. Переходимо до кроку 8.

Побудова L-ланцюжка здійснюється однозначно. Дійсно, в кожному стовпці не може бути більш одного 0*, а в кожному рядку не може бути більш одного 0' (після того, як в рядку один нуль виділений штрихом, цей рядок виділяють і ніякі інші його нулі не можуть бути виділені). L-ланцюжок завжди починається з 0' і закінчується 0'. Принципову схему побудови L-ланцюжка подано на рис. 3.1.

Дійсно, нехай $C_m = (c_{ij}^m)$ та існує L-ланцюжок вигляду

$$L = (c_{i1j1}^m, c_{i2j1}^m, c_{i2j2}^m, ..., c_{ikjk}^m, c_{ik+1jk}^m),$$

$$0' \xrightarrow{\text{За стовпцем}} 0' \xrightarrow{\text{За рядком}} 0'$$

Рис. 3.1 - Схема побудови L-ланцюжка

який не можна продовжити, причому c_{ikjk}^m відзначений як 0', c_{ik+1jk}^m відзначений як 0*. В цьому випадку в матриці C_m стовпець з номером j_k не виділений знаком «+», оскільки в цьому стовпці є елемент c_{ikjk}^m типу 0'. В той же час цей стовпець містить 0* (так відзначений елемент c_{ik+1jk}^m). Отже, знак виділення стовпця з номером j_k був знятий і рядок з номером j_{k+1} містить 0', оскільки на кроці 6 його було виділено. Таким чином, L-ланцюжок можна продовжити, що суперечить зробленому припущенню. Зауважимо, що кількість елементів в L-ланцюжку є непарною. При цьому L-ланцюжок може складатися з одного елементу, якщо в одному стовпці з тим, що розглядається 0' немає 0*.

Крок 8. У L-ланцюжку усі 0* замінюємо нулями, а усі 0' - символами 0*, внаслідок чого в еквівалентній матриці вартості C_m одержуємо нову систему незалежних нулів, кількість елементів якої на одиницю більша за кількість елементів в попередній системі незалежних нулів (в L-ланцюжку кількість елементів 0' завжди на одиницю перевищує кількість елементів 0*). Поза L-ланцюжком усі 0' замінюємо нулями та знімаємо всі виділення рядків і стовпців матриці C_m. Переходимо до кроку 1.

Крок 9. Серед невиділених елементів матриці C_m знаходимо мінімальний елемент h (h>0 через позитивність елементів еквівалентної матриці вартості C_m та відсутність невиділених нулів). Значення h віднімаємо з елементів невиділених рядків та додаємо до елементів виділених стовпців. Знов одержану еквівалентну матрицю вартості з невід'ємними елементами, в якій щонайменше один з невиділених елементів є нулем, позначаємо через C_m та переходимо до кроку 5.

Приклад 3.2. Повернемося до задачі про призначення, для якої в прикладі 3.1 виконаний підготовчий етап угорського методу та одержана еквівалентна матриця вартості C_0. Розглянемо ітераційний процес угорського методу рішення цієї задачі.

Перша ітерація.

Крок 1. Кількість k елементів в системі незалежних нулів матриці C_0 дорівнює трьом. А оскільки n =4 та k < n, переходимо до кроку 3.

Крок 3. Виділяємо стовпці матриці C_0, що містять 0*:

$$\begin{array}{cc} & \begin{array}{cccc} + & + & + & \end{array} \\ C_0 = & \begin{pmatrix} 0^* & 2 & 4 & 0 \\ 5 & 0^* & 3 & 6 \\ 3 & 1 & 0^* & 4 \\ 6 & 0 & 5 & 6 \end{pmatrix}. \end{array}$$

Переходимо до кроку 4.

Крок 4. Один з елементів невиділеного четвертого стовпця матриці C_0 є нульовим (у першому рядку). Переходимо до кроку 5.

Крок 5. У першому рядку матриці C_0 разом з невиділеним нулем ($c_{14}^0 = 0$) є елемент c_{11}^0, виділений як 0*. Переходимо до кроку 6.

Крок 6. Знайдений невиділений нуль позначаємо через 0' і виділяємо перший рядок, що містить його. Знімаємо знак виділення з першого стовпця, в якому розташований 0* з першого рядка:

$$\begin{array}{ccc} & & \begin{array}{cccc} & + & + & \end{array} \\ & + & \\ C_0 = & & \begin{pmatrix} 0^* & 2 & 4 & 0 \\ 5 & 0^* & 3 & 6 \\ 3 & 1 & 0^* & 4 \\ 6 & 0 & 5 & 6 \end{pmatrix}. \end{array}$$

Переходимо до кроку 4.

Крок 4. Серед невиділених елементів матриці C_0 немає нулів. Переходимо до кроку 9.

Крок 9. Серед невиділених елементів матриці C_0 знаходимо мінімальний елемент h = min (5, 3, 6, 6, 4, 6) = 3. Значення h = 3 віднімаємо з елементів невиділених рядків з номерами 2, 3, 4 та додаємо до елементів виділених стовпців з номерами 2, 3. Знов одержану еквівалентну матрицю вартості. позначаємо C_0:

$$C_0=\begin{array}{cc} & \begin{array}{cccc} & + & + & \end{array} \\ \begin{array}{c} + \\ \\ \\ \\ \end{array} & \begin{pmatrix} 0^* & 5 & 7 & 0' \\ 2 & 0^* & 3 & 3 \\ 0 & 1 & 0^* & 1 \\ 3 & 0 & 5 & 3 \end{pmatrix} \end{array}.$$

Переходимо до кроку 5.

Крок 5. У третьому рядку, що містить невиділений нуль ($c'_{31} = 0$), є й виділений нуль ($c'_{33} = 0*$). Переходимо до кроку 6.

Крок 6. Знайдений невиділений нуль позначаємо 0', третій рядок, що містить його, виділяємо. Знімаємо знак виділення з третього стовпця, в якому розташований 0* з третього рядка:

$$C_0=\begin{array}{cc} & \begin{array}{cccc} & + & + & \end{array} \\ \begin{array}{c} + \\ \\ + \\ \\ \end{array} & \begin{pmatrix} 0^* & 5 & 7 & 0' \\ 2 & 0^* & 3 & 3 \\ 0 & 1 & 0^* & 1 \\ 3 & 0 & 5 & 3 \end{pmatrix} \end{array}.$$

Переходимо до кроку 4.

Крок 4. Серед невиділених елементів матриці C_0 немає нулів. Переходимо до кроку 9.

Крок 9. Серед невиділених елементів матриці C_0 знаходимо мінімальний елемент h = min{2, 3, 3, 3, 5, 3}= 2. Значення h = 2 віднімаємо з елементів невиділених рядків з номерами 2 і 4 та додаємо до елементів виділеного другого стовпця. Знов одержану еквівалентну матрицю вартості позначаємо C_0.

$$C_0=\begin{array}{cc} & \begin{array}{cccc} & + & + & \end{array} \\ \begin{array}{c} + \\ \\ + \\ \\ \end{array} & \begin{pmatrix} 0^* & 7 & 7 & 0' \\ 0 & 0^* & 1 & 1 \\ 0' & 3 & 0^* & 1 \\ 1 & 0 & 3 & 1 \end{pmatrix} \end{array}.$$

Переходимо до кроку 5.

Крок 5. У другому рядку, що містить невиділений нуль ($c^0_{21} = 0$), є виділений нуль ($c^0_{22} = 0*$). Переходимо до кроку 6.

Крок 6. Знайдений невиділений нуль позначаємо 0’, другий рядок, що містить його, виділяємо. Знімаємо знак виділення з другого стовпця, в якому розташований 0* з другого рядка:

$$C_0 = \begin{array}{ccccc} + & 0^* & 7 & 7 & 0' \\ + & 0' & 0^* & 1 & 1 \\ + & 0' & 3 & 0^* & 1 \\ & 1 & 0 & 3 & 1 \end{array}$$

Переходимо до кроку 4.

Крок 4. Серед невиділених елементів матриці C_0 є нуль ($c_{42}^0 = 0$). Переходимо до кроку 5.

Крок 5. Оскільки четвертий рядок, що містить невиділений нуль, не містить 0*, то переходимо до кроку 7.

Крок 7. Знайдений невиділений нуль позначаємо символом 0’ і будуємо L-ланцюжок ($c_{42}^0, c_{22}^0, c_{21}^0, c_{11}^0, c_{14}^0$):

$$C_0 = \begin{array}{ccccc} + & 0^* & 7 & 7 & 0' \\ + & 0' & 0^* & 1 & 1 \\ + & 0' & 3 & 0^* & 1 \\ & 1 & 0 & 3 & 1 \end{array}$$

Переходимо до кроку 8.

Крок 8. У L-ланцюжку всі символи 0* замінюємо нулями, а символи 0' замінюємо на 0*. Поза L-ланцюжком всі 0' замінюємо нулями та знімаємо всі виділення рядків і стовпців. Одержану матрицю вартості позначаємо C_1:

$$C_1 = \begin{pmatrix} 0 & 7 & 7 & 0^* \\ 0^* & 0 & 1 & 1 \\ 0 & 3 & 0^* & 1 \\ 1 & 0^* & 3 & 1 \end{pmatrix}.$$

Переходимо до кроку 1 наступної ітерації (m = 2).

Друга ітерація.

Крок 1. Кількість елементів k в системі незалежних нулів матриці C_1 дорівнює чотирьом. А оскільки k =4= n, то переходимо до кроку 2.

Крок 2. У відповідності з еквівалентною матрицею вартості $C^*=C_1$ виписуємо оптимальне рішення даної задачі, представлене у вигляді матриці її змінних

$$X = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

На практиці зустрічаються задачі, для яких параметр c_{ij} має смисл ефективності виконання i-ї роботи j-м виконавцем. У таких завданнях сумарна ефективність виконання всіх робіт має бути максимальною, що призводить до задачі (3.7). Для перетворення задачі (3.7) на стандартну задачу (3.6) можна скористатися формулою (3.8), або змінити перший крок підготовчого етапу угорського методу, який в цьому випадку приймає наступний вигляд.

Крок 1. Для кожного стовпця матриці C задачі про призначення (3.7) знаходимо максимальний елемент

$$l_j = \max_{i=1,n} c_{ij}, \quad j = \overline{1,n}$$

і формуємо еквівалентну матрицю вартості C’, в якій

$$c_{ij}' = l_j - c_{ij}, \quad i, j = \overline{1,n}.$$

Приклад 3.3. Розглянемо задачу про призначення вигляду (3.7) з матрицею

$$C=\begin{pmatrix}10 & 7 & 6 & 9\\ 5 & 9 & 7 & 3\\ 7 & 8 & 10 & 5\\ 3 & 8 & 4 & 2\end{pmatrix}.$$

У цьому випадку

$$l_1=\max\{10,5,7,3\}=10,\quad l_{12}=\max\{6,7,10,4\}=10,$$

$$l_3=\max\{7,9,8,8\}=9,\quad l_{14}=\max\{9,3,5,2\}=9.$$

і еквівалентна матриця вартості має вигляд

$$C'=\begin{pmatrix}0 & 2 & 4 & 0\\ 5 & 0 & 3 & 6\\ 3 & 1 & 0 & 4\\ 7 & 1 & 6 & 7\end{pmatrix}.$$

Ця матриця співпадає з матрицею, одержаною на першому кроці підготовчого етапу угорського методу в прикладі 3.1 (завдання мінімізації витрат). А оскільки подальші кроки підготовчого етапу не зміняться, то зрозуміло, що рішення X, одержане в прикладі 3.2 буде оптимальним рішенням і для даної задачі про призначення (задачі максимізації сумарної ефективності).

Контрольні запитання

1. Які специфічні властивості дозволяють виділити транспортні задачі в окремий клас з множини задач лінійного програмування?
2. Опишіть методи побудови припустимого плану транспортної задачі.
3. Скільки ненульових елементів має містити невироджений базисний план транспортної задачі?
4. Сформулюйте критерій оптимальності для припустимого плану транспортної задачі.
5. Поясніть, на чому заснований метод потенціалів?
6. З чого випливає критерій оптимальності припустимого плану транспортної задачі?
7. Перелічіть основні етапи методу потенціалів.
8. Які умови повинні бути дотримані під час побудови ланцюжка перетворення плану в методі потенціалів?
9. Що треба робити під час виникнення ситуації виродженості поточного плану в транспортній задачі?
10. Як формулюють задачу про призначення?
11. Які значення мають приймати змінні задачі про призначення?
12. У чому полягає зміст угорського методу розв’язання задачі про призначення?
13. Охарактеризуйте особливості алгоритму угорського методу розв’язання задачі про призначення.

ТЕМА 4

МЕРЕЖЕВІ МОДЕЛІ

4.1. Основні поняття теорії графів та мереж

Багато економічних завдань, такі як перевезення вантажів, перекачування нафти й газу трубопроводами, керування запасами та ін., зручно моделювати й вирішувати в термінах сіток і потоків. Основою подібних моделей служать орієнтовані або неорієнтовані графи. Уведемо визначення.

Орієнтованим графом називають трійку (I,D,G), у якій I — непуста множина вершин; D - множина дуг; G - відображення, що кожній дузі $d \in D$ ставить у відповідність упорядковану пару вершин (i, j), де $i, j \in I$.

Неорієнтованим графом називають трійку (I,D,G), у якій I - непуста множина вершин; D - множина ребер; G - відображення, що кожному ребру $d \in D$ ставить у відповідність неупорядковану пару вершин $[i, j]$, де $i, j \in I$.

Граф (I, D, G) називають ***кінцевим***, якщо множини I і D кінцеві.

Геометрично граф можна представити у вигляді множини точок, що зображують вершини, і з'єднуючих їх ліній, що відповідають ***дугам*** (рис. 4.1). Очевидно, що з кожним орієнтованим графом можна однозначно пов’язати неорієнтований, замінивши дуги на ***ребра***. Якщо будь-які дві вершини графа з'єднуються не більш ніж однією дугою (ребром), то граф називають ***простим*** і його можна задати за допомогою пари (I, D). У цьому випадку кожна дуга (ребро) d повністю визначається парою вершин(i, j), що з'єднуються, і це умовно записують у вигляді: $d=(i, j)$. Упорядкована пара вершин (i, j), що ставиться у відповідність певній дузі d, задає її орієнтацію: і називають ***початком дуги***, а j - її ***кінцем***. Саму дугу вважають ***інцидентною*** цим вершинам.

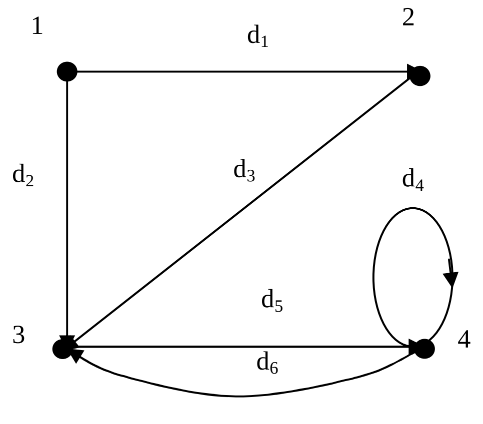

Рис. 4.1 – Орієнтований граф

Шляхом довжини n в орієнтованому графі (I, D) називають впорядковану послідовність різних дуг $(d_1, d_2,..., d_n)$, для яких початок кожної наступної збігається з кінцем попередньої. Кінцевий шлях, в якого початкова вершина збігається з кінцевою, називають ***контуром***.

Для неорієнтованого графа аналогом поняття шляху є ***ланцюг***, а контуру - ***цикл***.

Якщо дві будь-які вершини неорієнтованого графа можна з'єднати ланцюгом, то його називають ***зв'язним***. Орієнтований граф називають зв'язним, якщо йому відповідає зв'язний неорієнтований граф.

Зв'язний неорієнтований граф, що не містить циклів, називають ***деревом***.

Розглянемо задачу. Є кінцевий граф (I, D, G), кожній вершині і якого зіставлене певне число b_i, називане інтенсивністю вершини. Граф (I, D, G), вершинам якого зіставлені значення інтенсивностей b_i, називатимемо ***сіткою***. Якщо $b_i>0$, то вершину і називають ***джерелом***, якщо $b_i<0$, то - ***стоком***, а якщо

b_i=0, то - ***нейтральною вершиною***. Множину джерел, стоків та нейтральних вершин позначимо відповідно I^+, I^-, I^0.

Для визначеної вище сітки ***потоком*** називають таку сукупність величин, заданих на множині дуг, $X = \{x_d\}_{d\in D}$, що

$$\sum_{d\in D_i^+} x_d - \sum_{d\in D_i^-} x_d = b_i, \quad i\in I \tag{4.1}$$

$$x_d \geq 0, \quad d\in D, \tag{4.2}$$

де D_i^+ - множина дуг, що виходять з вершини i, а D_i^-. - множина дуг, що входять до неї. Величину x_d називають ***значенням потоку*** за дугою d і змістовно інтерпретують як кількість продукту, що пропускають за даною дугою.

Співвідношення (4.1) означає, що для будь-якої вершини сітки різниця вихідного й вхідного потоків дорівнює її інтенсивності.

На базі введеної термінології можна сформулювати багато різних задач. Розглянемо найвідоміші з них. Для кожної дуги $d\in D$ визначимо значення $c_d\geq 0$, називані ***вартістю переміщення одиниці продукту*** за дугою, тоді сумарна вартість потоку X прийме вигляд

$$f(X) = \sum_{d\in D} c_d x_d\ . \tag{4.3}$$

Задачу мінімізації функції (4.3) за обмежень (4.1)-(4.2) зазвичай називають ***лінійною сітковою задачею***. Очевидно, що вона є задачею лінійного програмування. Якщо додатково для кожної дуги сітки $d\in D$ визначити величини $r_d\geq 0$, називані ***пропускними здатностями***, то, додавши обмеження

$$0\leq x_d \leq r_d, \quad d\in D, \tag{4.4}$$

одержимо задачу про потік у сітці з обмеженими пропускними здатностями.

Наведені формулювання задач спеціально надані в абстрактному вигляді, що дозволяє підкреслити їх універсальність. До очевидної сфери їх використання належить організація вантажоперевезень у транспортній сітці. У таких моделях вершини трактують як пункти, по'єднані сіткою доріг, і характеризуються потребами в певному продукті (b_i <0) або його запасів (b_i >0). Задачі з визначення плану, що мінімізує витрати на перевезення, які з математичної точки зору повністю ідентичні (4.1)-(4.4), ще називають ***транспортними задачами в сітковій постановці***.

4.2. Транспортна задача у мережевій формі. Дослідження мережевої моделі як задачі лінійного програмування

Розглянемо задачу з визначення оптимального потоку X у певній сітці (I, D, G), для якого

$$f(X) = \sum_{d\in D} c_d x_d \to \min \tag{4.5}$$

за обмежень

$$\sum_{d\in D_i^+} x_d - \sum_{d\in D_i^-} x_d = b_i, \quad i\in I \tag{4.6}$$

$$0\leq x_d \leq r_d, \quad d\in D, \tag{4.7}$$

де $r_d \geq 0$. Передбачається так само, що сітка є збалансованою, тобто

$$\sum_{i \in I} b_i = 0 . \qquad (4.8)$$

Для задачі (4.5)-(4.8) справедливим є критерій оптимальності:

Щоб припустимий потік $X = \{x_d\}_{d \in D}$, (тобто такий, що задовольняє умовам (4.6)-(4.7)) був оптимальним, необхідно й достатньо існування для кожної вершини $i \in I$ такого числа v, називаного потенціалом, що для всіх дуг $d = (i, j)$

$$v_j - v_i \leq c_d, \quad \text{якщо } x_d = 0, \qquad (4.9)$$

$$v_j - v_i = c_d, \quad \text{якщо } 0 < x_d < r_d, \qquad (4.10)$$

$$v_j - v_i \geq c_d, \quad \text{якщо } x_d = r_d. \qquad (4.11)$$

Для розв'язання транспортної задачі в сітковій постановці (4.5)-(4.7) можна застосовувати метод потенціалів, який є узагальненням розглянутого у попередній темі методу потенціалів для транспортної задачі в матричній постановці.

Оскільки задача (4.5)-(4.7) є окремим випадком задачі лінійного програмування, її можна привести до канонічної форми. При цьому можна встановити, що ранг матриці задачі дорівнює m-1, де m - кількість вершин у сітці. Введемо додатково ще деякі поняття, використовувані під час опису властивостей сіткових задач.

Остовом сітки (I, D, G) називають будь-яке її часткове дерево (частковий граф, що є деревом). Справедливе ствердження:

Довільному остову сітки (I, D, G) відповідає базис задачі (4.5)-(4.7) і навпаки.

Нехай є певний потік $X = \{x_d\}_{d \in D}$. Розглянемо множину дуг $D(X) = =\{d \in D | 0 < x_d < r_d\}$. ***Опорою*** потоку X називають частковий граф (I, $D(X)$, G). Говорять, що потік X ***невироджений***, якщо його опора (I, $D(X)$, G) є остовом сітки (I, D, G). Іншими словами, використовуючи термінологію транспортної задачі, у невиродженому потоці, якому відповідає припустимий базисний план задачі, дороги, якими здійснюється перевезення вантажу, що не досягають за обсягом обмеження на пропускну здатність, утворюють остов (пов'язану підсітку без циклів) розглянутої транспортної сітки.

Тепер дамо короткий опис схеми методу потенціалів для транспортної задачі в сітковій постановці.

1. Передбачають, що на початку чергової ітерації q є певний ***припустимий невироджений потік*** $X^{(q)} = \{x_d^{(q)}\}_{d \in D}$.

За наявним потоком $X^{(q)}$ будують систему потенціалів пунктів сітки. Для цього обирають довільний пункт i_0, потенціал якого покладається $v_{i_0} = 0$. Множину вершин, суміжних з i_0, позначимо через $I(i_0)$. Тоді для будь-якої вершини $j \in I(i_0)$ потенціали розраховують за правилом

$$v_j = v_{i_0} + c_{i_0 j} , \qquad (4.12)$$

якщо $(i_0,j) \in G(D(X^{(q)}))$ (дуга спрямована від i_0), і

$$v_j = v_{i_0} - c_{ji_0}\,, \tag{4.13}$$

якщо (j, i_0) ∈ G(D($X^{(q)}$)) (j, i_0) ∈ D($X^{(q)}$) (дуга спрямована до i_0).

Діставши чергову групу вершин з відомими потенціалами, ми маємо можливість на підставі (4.12) та (4.13) обчислити потенціали для наступної групи суміжних вершин і т.д., поки не будуть визначені всі потенціали. Можливість зробити це єдиним чином випливає з відсутності циклів в остова сітки.

Маючи повну систему потенціалів, для всіх дуг треба перевірити умови критерію оптимальності (4.9)-(4.11). Якщо вони виконуються, то поточний потік $X^{(q)}$ - оптимальний і, отже, алгоритм завершений; у протилежному випадку, переходимо до побудови наступного «поліпшеного» потоку.

2. За аналогією з іншими методами послідовного поліпшення плану черговий потік виходить за рахунок введення до нього однієї дуги й виведення іншої. Якщо умови критерію оптимальності порушуються відразу для кількох дуг, то для введення має сенс вибрати ту, на якій досягається максимальне відхилення ціни від різниці потенціалів вершин, що з'єднуються. Нехай для введення обрана певна дуга $d_l = (s, t)$, спрямована з вершини s у вершину t. З правил побудови потенціалів випливає, що в остові існують два ланцюги, один з яких з'єднує базову вершину i_0, потенціал якої був прийнятий рівним нулю, з s, а інший — i_0 з t. Якщо доповнити остов дугою d_l, утвориться єдиний цикл. Побудований цикл є аналогом ланцюжка перетворення плану в методі потенціалів для транспортної задачі в матричній постановці. Позначимо через D_{st}^{+} множину дуг даного циклу, орієнтація яких збігається з орієнтацією дуги $d_l = (s, t)$, а через D_{st}^{-} — множину дуг, що мають протилежну орієнтацію. Визначимо величину можливого коректування обсягів вантажоперевезень, переміщуваних за циклом

$$\Theta = \min\left\{ \min_{D_{st}^{-}} x_d^{(q)}, \quad \min_{D_{st}^{+}} (r_d - x_d^{(q)}) \right\} \tag{4.14}$$

Формула (4.14) показує, що при циклічному перетворенні поточного потоку збільшуються обсяги вантажоперевезень на тих дугах, які однаково спрямовані з дугою, що уводиться, і зменшуються на дугах, що мають зворотну орієнтацію. Під час додавання треба стежити за тим, щоб не перевищити обмеження на пропускні здатності ($\Theta \le r_d - x_d^{(q)}$), а при відніманні - за невід'ємністю $x_d^{(q)}$. Після визначення Θ здійснюють перерахунок компонентів поточного потоку за формулою

$$x_d^{(q+1)} = \begin{cases} x_d^{(q)} + \Theta, & \text{если} \quad d \in D_{st}^{+} \\ x_d^{(q)} - \Theta, & \text{если} \quad d \in D_{st}^{-} \\ x_d^{(q)}, & \text{если} \quad d \notin D_{st}^{+} \cup D_{st}^{-} \end{cases}. \tag{4.15}$$

В результаті ми дістанемо новий припустимий потік $x_d^{(q+1)}$, покладаємо номер поточної ітерації q+l та переходимо до п. 1.

В описаному алгоритмі, як і у випадку з матричною транспортною задачею, ми не гарантовані від виникнення виродженого потоку, такому потоку відповідатиме незв'язна опора. Для подолання виродженості рекомендується включити до поточного плану фіктивні компоненти з нульовими обсягами так, щоб відповідні їм дуги доповнювали опору до остова сітки. Побудований за таким способом план дозволяє виконати всі дії, що входять у стандартну ітерацію методу потенціалів.

Окремо треба зупинитися на методах генерації вихідного припустимого потоку. Найпростіший з них (хоча, можливо, і найменш раціональний) заснований на ідеях, подібних ідеям методу мінімізації нев'язань, використовуваного для побудови припустимого базисного плану ЗЛП. Цей метод припускає розв'язання відповідної допоміжної задачі, що виходить з основної в результаті наступних перетворень:

1. До множини вершин сітки додають фіктивну нульову вершину з нульовою інтенсивністю ($b_0 = 0$).

2. Всі вершини, що мають від'ємну інтенсивність (попит) $b_i < 0$, з'єднують з доданою вершиною 0 вхідними дугами (0, i); а вершини, що мають додатну інтенсивність (запасом) $b_i > 0$, - вихідними дугами (i, 0). Обмеження на пропускні здатності для дуг, що додаються, відсутні.

3. Вартості переміщення одиниці продукту для знову доданих дуг покладають рівними 1, а для дуг, що відповідають транспортній сітці основної задачі, - 0.

Побудована допоміжна задача має очевидний припустимий невироджений потік, одержуваний призначенням обсягів, що дорівнюють інтенсивностям вершин, за всіма доданими дугами. Вирішивши допоміжну задачу, ми або одержимо припустимий потік для основної задачі, або прийдемо до висновку про відсутність в неї припустимих планів.

4.3. Задача про максимальний потік. Визначення критичного шляху

Класичним прикладом сіткових задач є визначення найкоротшого шляху між вершинами сітки. Нехай заданий граф (I, D, G), кожній дузі якого поставлено у відповідність число c_d, називане ***довжиною***. Нехай виділені дві вершини графа s та t, і потрібно знайти шлях найменшої довжини, що веде з вершини s у вершину t.

Якщо в графі є «кратні» дуги, що з'єднують однакові початок та кінець, то досить залишити одну - з найменшою довжиною, а інші відкинути. Отже, досить розглядати задачу про найкоротший шлях для простого графа (I, D), у якому дуги визначаються впорядкованими парами вершин d = (i, j). Тоді природно шлях L, що йде з вершини s у вершину t, задавати у вигляді впорядкованого набору вершин, через які проходить даний шлях:

$$L=(s=i_0, i_1, ..., i_{p-1}, i_p=t),$$

а довжини дуг позначати як $c_d = c_{ij}$.

Довжину описаного вище довільного шляху L визначають за формулою

$$f(L) = \sum_{k=1}^{p} c_{i_{k-1} i_k} \ . \tag{4.16}$$

Задача про найкоротший шлях є окремим випадком транспортної задачі в сітковій постановці (або, що те саме, задачі про оптимальний потік). Для цього досить привласнити вершині s одиничний запас, вершині t одиничну потребу, всі інші вершини покласти нейтральними, а дугам привласнити необмежені пропускні здатності. Проте, як правило, раціональнішим опиняється використання конкретних властивостей даної задачі й розв'язання її спеціальними методами. До їх числа належить, наприклад, метод Мінті.

Метод Мінті з розв'язання задачі про найкоротший шлях у сітці є ітеративним процесом, у ході якого будують шлях $L = (s = i_0, i_1, ..., i_{p-1}, i_p=t)$.

На попередньому (нульовому) етапі алгоритму формують масив значень так званих ***модифікованих довжин*** $\tilde{c}_{ij}$, які на початку першої ітерації покладають рівними $c_{ij}\geq 0$. Потім здійснюють оцінку вершини $i_0=s$ числом $m_{i_0} = 0$.

Стандартна ітерація включає наступні етапи:

1. Оцінка вершин сітки. Позначимо множину вершин сітки, позначених на попередніх ітераціях як $\tilde{I}$ (на першій ітерації $\tilde{I} = \{i_0\}$). Для кожної вершини $i \in \tilde{I}$ шукають дуги, що з'єднують її з ще не позначеними вершинами-нащадками j, модифікована довжина яких $\tilde{c}_{ij} = 0$. Знайдені за таким способом вершини j позначають числом $m_j = i$, що вказує на «батька». У тому випадку, коли відразу кілька дуг, що мають $\tilde{c}_{ij} = 0$, закінчуються в тій самій вершині j, значення і для її позначки обирають довільно.

Якщо серед знову позначених вершин виявиться вершина t, то знайдений шуканий шлях $(i_0, i_1, ..., i_{p-1}, i_p)$, де

$$i_p = t, \quad i_{p-1} = m_{i_p}, \quad i_{p-2} = m_{i_{p-1}}, \ldots, i_1 = m_{i_2}, \quad i_0 = m_{i_1} = s .$$

На цьому алгоритм завершується.

У випадку якщо вершини t немає серед позначених, і одночасно не можна позначити ні однієї нової вершини, то переходимо до етапу 2.

2. Перетворення значень модифікованих довжин дуг. Для кожної вершини $i \in \tilde{I}$ шукають дуги, що з'єднують її з ще не позначеними вершинами j

$$\tilde{\Delta}_i = \min_{j \notin \tilde{I}} \tilde{c}_{ij} .$$

Далі модифіковані довжини всіх дуг, які з'єднують позначені вершини з непозначеними ($i \in \tilde{I}, \quad j \notin \tilde{I}$), зменшують на величину

$$\tilde{\Delta} = \min_{j \in \tilde{I}} \tilde{\Delta}_i ,$$

в результаті чого найкоротші невикористані дуги одержують нульову модифіковану довжину.

Потім відбувається перехід до наступної ітерації.

Шлях, побудований за методом Мінті, буде найкоротшим. Це можна довести за допомогою індукції за номером ітерації, на якій була позначена вершина t, або, що те саме, за кількістю дуг, що складають найкоротший шлях. Якщо це відбулося на першому кроці (що можливо тільки у випадку, якщо початкова й кінцева вершини з'єднані дугою нульової довжини), то доказуване твердження очевидне. Припустимо, що воно вірно для всіх пунктів, позначених за перші r ітерацій, тобто тих, які досягаються переходом за r дугами. Тоді, якщо кінцева вершина t позначена на (r+1)-ої ітерації, отриманий шлях так само буде найкоротшим, тому що дана вершина позначається в результаті мінімально можливого продовження одного з шляхів, отриманого за попередні r ітерацій і такого, що за припущенням є найкоротшим.

Відзначимо, що описаний алгоритм придатний для побудови найкоротших шляхів на неорієнтованих графах.

Розглянемо приклад. Визначимо найкоротший шлях з вершини 1 у вершину 6 для неорієнтованої сітки, що показана на рис. 4.2.

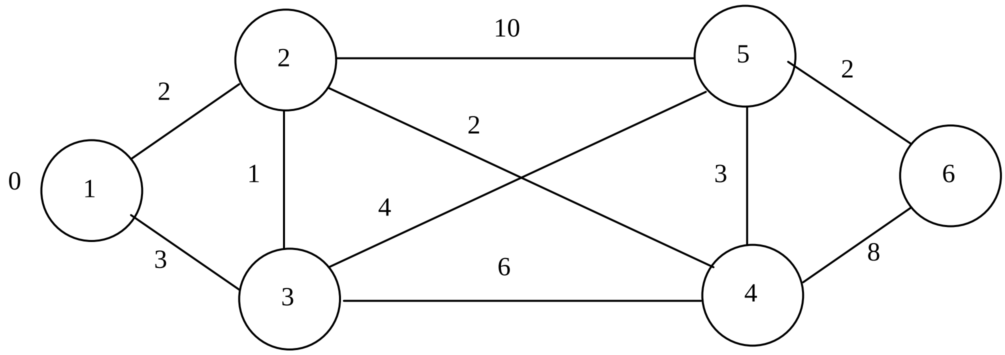

Рис. 4.2 - Вихідна схема неорієнтованої сітки

На попередньому етапі вершину 1 позначають числом $m_1 = 0$, а модифіковані довжини збігаються із заданими довжинами дуг.

Ітерація 1. Оскільки з вершини 1 не виходять дуги нульової довжини, подальша оцінка вершин неможлива. Переходимо до етапу 2. Суміжними з вершиною 1 є вершини 2 та 3. Для них визначаємо $\Delta = \min\{\tilde{c}_{12}, \quad \tilde{c}_{13}\} = 2$ й віднімемо її з $\tilde{c}_{12}, \quad \tilde{c}_{13}$, у результаті чого одержимо $\tilde{c}_{12} = 0, \quad \tilde{c}_{13} = 1$.

Ітерація 2 (рис. 4.3). Позначаємо вершину 2 m_2=1. Подальша позначка неможлива, тому переходимо до етапу 2. Суміжними з позначеними вершинами 1 та 2 є вершини 3, 4, 5. З чого визначаємо $\Delta = \min\{\tilde{c}_{13}, \quad \tilde{c}_{23}, \quad \tilde{c}_{24}, \quad \tilde{c}_{25}\} = 1$ й після відповідного перетворення маємо $\tilde{c}_{13} = 0, \quad \tilde{c}_{23} = 0, \quad \tilde{c}_{24} = 1, \quad \tilde{c}_{25} = 9$.

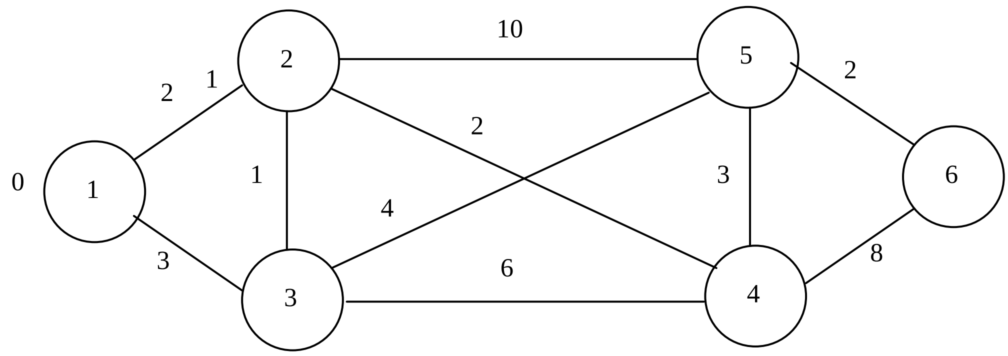

Рис. 4.3 - Схема неорієнтованої сітки на 2-ї ітерації

Ітерація 3. У вершину 3 ведуть дуги нульової довжини як з вершини 1, так і з вершини 2. Оскільки вибір тут може бути довільним, позначимо вершину 3 числом m_3 = 1 (рис. 4.4). Подальша позначка неможлива, тому переходимо до етапу 2. Суміжними з раніше позначеними вершинами є вершини 4, 5. Звідки визначаємо $\Delta = \min\{\tilde{c}_{24}, \quad \tilde{c}_{25}, \quad \tilde{c}_{34}, \quad \tilde{c}_{35}\} = 1$ й після перетворення маємо $\tilde{c}_{24} = 8, \quad \tilde{c}_{25} = 0, \quad \tilde{c}_{34} = 3, \quad \tilde{c}_{35} = 5$.

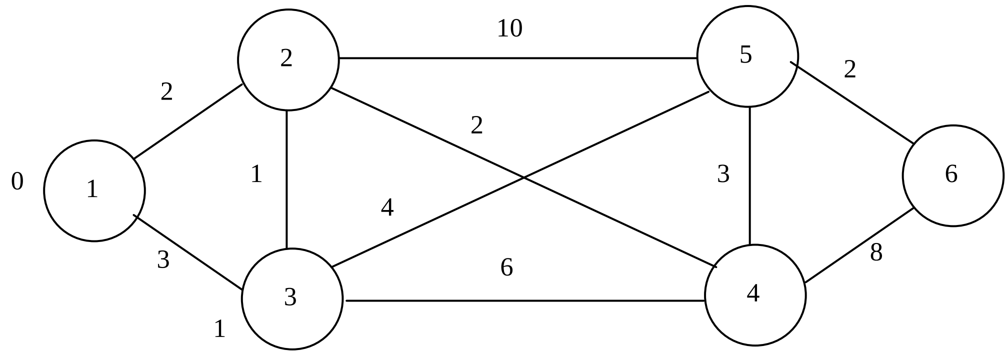

Рис. 4.4 - Схема неорієнтованої сітки на 3-й ітерації

Ітерація 4. Позначаємо вершину 4 $m_4 = 2$ (див. рис. 4.5). Подальша позначка неможлива, тому переходимо до етапу 2. Суміжними з раніше позначеними вершинами є вершини 5, 6. Звідки визначаємо $\Delta = \min\{\tilde{c}_{25}, \quad \tilde{c}_{35}, \quad \tilde{c}_{45}, \quad \tilde{c}_{46}\} = 3$ та після перетворення маємо $\tilde{c}_{25} = 5, \quad \tilde{c}_{35} = 0, \quad \tilde{c}_{45} = 0, \quad \tilde{c}_{46} = 5$.

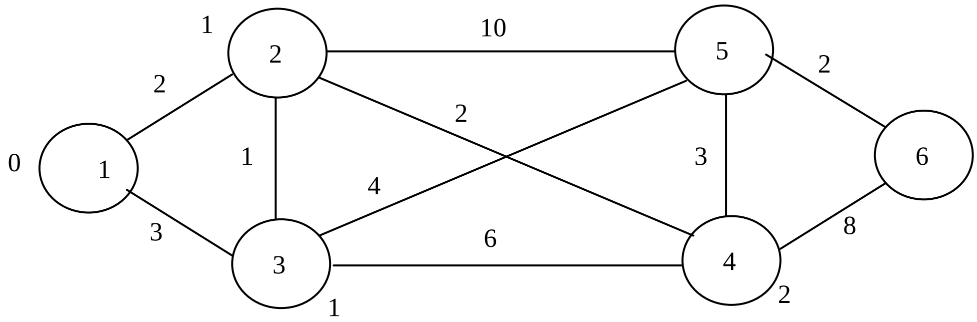

Рис. 4.5 - Схема неорієнтованої сітки на 4-й ітерації

Ітерація 5. У вершину 5 ведуть дуги нульової довжини як з вершини 3, так і з вершини 4. Керуючись тими самими міркуваннями, що й на ітерації 3, позначимо вершину 5 числом $m_5 = 3$ (рис. 4.6). Подальша позначка неможлива,

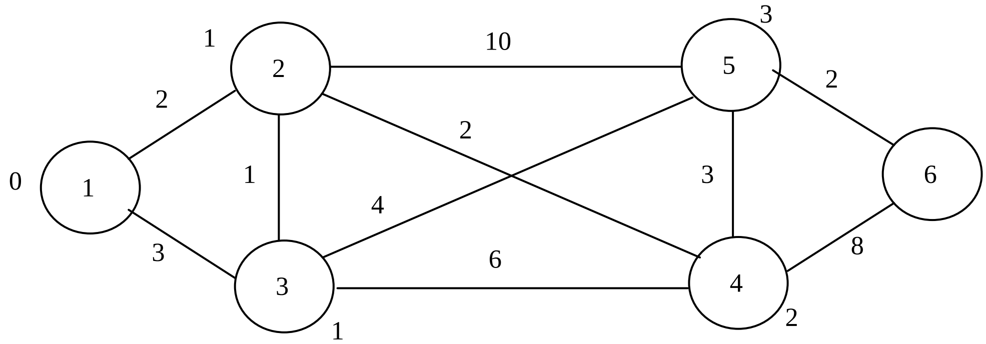

Рис. 4.6 - Схема неорієнтованої сітки на 5-ї ітерації

тому переходимо до етапу 2. Суміжною з раніше позначеними вершинами є вершина 6. Звідки визначаємо $\Delta = \min\{\tilde{c}_{46}, \quad \tilde{c}_{56}\} = 2$ й після перетворення маємо $\tilde{c}_{46} = 3, \quad \tilde{c}_{56} = 0$.

Ітерація 6. У вершину 6 веде дуга нульової довжини з вершини 5, тому позначаємо її числом m_6 =5 (рис. 4.7). Оскільки ми позначили кінцеву вершину маршруту, то алгоритм завершений, можна, використовуючи значення оцінок для «батьків», виписати шуканий найкоротший шлях (1, 3, 5, 6).

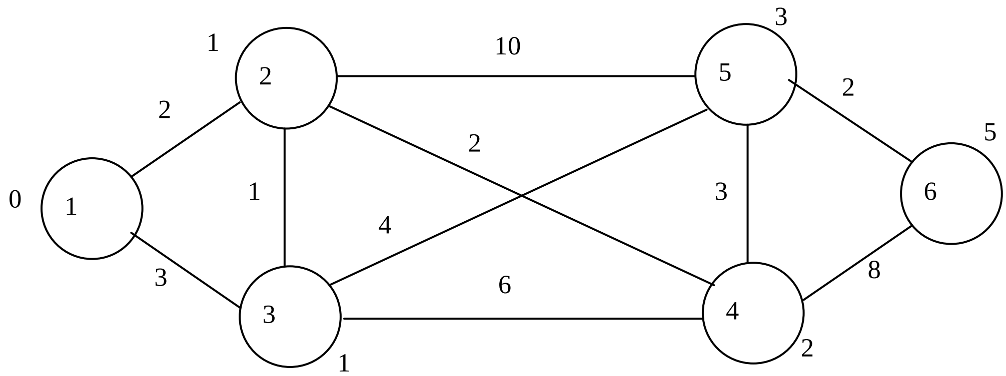

Рис. 4.7 - Схема неорієнтованої сітки на 6-й ітерації

Треба додати, що якби наш вибір на ітераціях 3 та 5 був іншим, то ми одержали би альтернативний шлях тієї самої довжини (1, 2, 4, 5, 6), тобто розглянута задача має кілька розв'язань.

Контрольні запитання

1. Які специфічні властивості дозволяють виділити транспортні задачі в окремий клас з множини задач лінійного програмування?
2. Опишіть методи побудови припустимого плану транспортної задачі.
3. Скільки ненульових елементів має містити невироджений базисний план транспортної задачі?
4. Сформулюйте критерій оптимальності для припустимого плану транспортної задачі.
5. Поясніть, на чому заснований метод потенціалів?
6. З чого випливає критерій оптимальності припустимого плану транспортної задачі?
7. Перелічіть основні етапи методу потенціалів.
8. Які умови мають бути дотриманими під час побудови ланцюжка перетворення плану в методі потенціалів?
9. Що треба робити під час виникнення ситуації виродженості поточного плану в транспортній задачі?
10. Приведіть формулювання лінійної сіткової задачі.
11. Покажіть, що транспортна задача в матричній постановці є окремим випадком транспортної задачі в сітковій постановці.
12. Дайте визначення поняття «остов сітки». Який зв'язок існує між остовом сітки й базисом транспортної задачі в сітковій постановці?
13. Перелічіть основні етапи методу потенціалів для транспортної задачі в сітковій постановці.
14. За яким способом можна одержати припустимий потік у транспортній сітці?
15. У чому полягає задача про найкоротший шлях?
16. Перелічіть основні етапи методу Мінті.

ТЕМА 5
ДИСКРЕТНЕ ПРОГРАМУВАННЯ

5.1. Класичні задачі цілочислової оптимізації, класифікація та характеристика методів

Цілочислове лінійне програмування орієнтоване на розв'язання задач лінійного програмування, в яких всі або деякі змінні повинні приймати цілочислові (або дискретні) значення. Багато економічних задач характеризуються тим, що обсяги керованих ресурсів можуть приймати тільки цілі значення. До цілочислового програмування належать так само задачі, у яких змінні можуть приймати тільки два значення – 0 або 1 (булеві або бінарні змінні). До задач цілочислового програмування належать задачі про призначення, про найкоротший шлях та інші. У загальному вигляді задачу цілочислового програмування можна сформулювати як задачу знаходження максимуму (або мінімуму) цільової функції $L(x_1, x_2,..., x_n)$ на множині D, зумовленій системою обмежень. Загальну задачу цілочислового програмування формулюють в такий спосіб

Знайти найбільше або найменше значення лінійної функції

$$L = \sum_{j=1}^{n} c_j x_j \to \text{extr} \tag{5.1}$$

за умови

$$\sum_{j=1}^{n} a_{ij} x_j = b_i; \quad i = \overline{1,m} \tag{5.2}$$

$$x_j \geq 0, \quad x_j - цілі, \quad j = \overline{1,n}.$$

Умову $x_j - цілі$ називають умовою дискретності. Особливе місце серед дискретних задач займає цілочислова задача лінійного програмування в канонічній формі (ЦКЗЛП). У певних ситуаціях вимога «цілочислове» може бути накладена лише на певні змінні x_j, що кардинально не змінює характеру задачі.

Принципова складність, викликувана наявністю умов цілочисловості в системі обмежень оптимізаційної задачі, полягає в тому, що в значній кількості випадків неможливо замінити дискретну задачу її безперервним аналогом і, знайшовши відповідний розв'язок, округлити його компоненти до найближчих цілих значень. На рис. 5.1 показано, що при округленні оптимального плану x^* звичайної задачі ЛП до цілих значень виходить точка B, що не належить області припустимих планів. Якщо навіть оптимальний план безперервної задачі, округлений до цілих значень компонент, виявиться припустимим, то цільова функція може поводитися так, що її значення буде на ньому істотно «гірше», ніж на оптимальному плані цілочислової задачі.

Перелічені проблеми визначили необхідність розробки спеціальних методів розв'язання дискретних і цілочислових задач.

У літературі, як правило, виділяють наступні класи дискретних оптимізаційних задач:

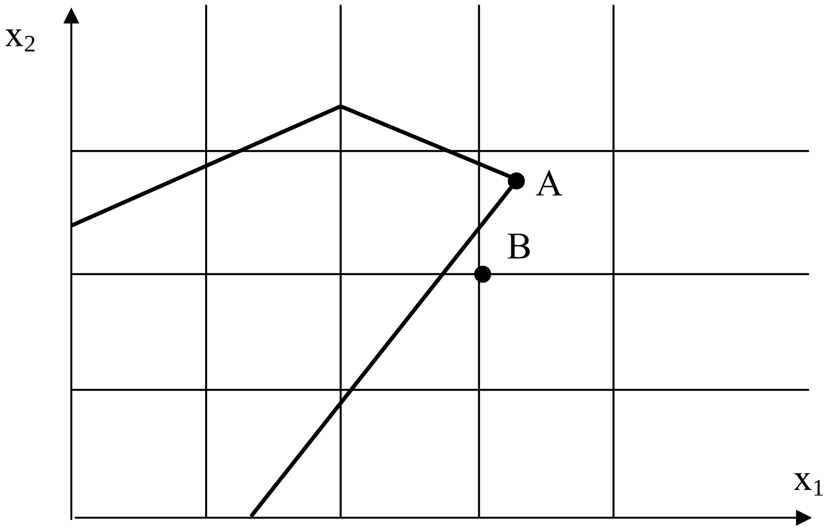

Рис. 5.1 – Розв'язок цілочислової задачі

- задача з неділимостями;
- екстремальні комбінаторні задачі;
- задачі з розривними цільовими функціями;
- задачі на незв'язних і неопуклих областях та ін.

Задача з неділимостями. У переважній більшості випадків наявність умов неділимості визначається фізичними властивостями об'єктів, наприклад, вони можуть з'явитися як додаткові обмеження в задачі виробничого планування, якщо в ній здійснюється керування випуском великої штучної продукції. Класичним представником задач даного класу є так звана задача про ранець. Вона полягає в тому, що солдат (або турист), що збирається в похід, може нести вантаж вагою не більш за W кг. Цей вантаж може складатися з набору предметів n типів, кожний предмет j-го типу важить w кг і характеризується певною «корисністю». Треба визначити, скільки предметів кожного виду потрібно покласти в ранець, щоб його сумарна корисність була максимальною. До такого формулювання можуть бути зведені багато економічних задач. У літературі ця задача також відома як задача про завантаження судна.

Комбінаторні задачі. До цього класу належать задачі з оптимізації функції, заданої на кінцевій множині, елементами якої служать вибірки з n об'єктів. Класичним представником такої задачі є задача про комівояжера. Вона полягає в складанні маршруту відвідування торговим агентом, що перебуває в деякому початковому пункті, n інших міст за умови, що задано матрицю вартостей переїздів з одного міста до іншого. Причому припустимим є маршрут, що передбачає однократне відвідування всіх міст і повернення у вихідний пункт. Найкращий маршрут повинен мінімізувати сумарну вартість переїздів. Кожний припустимий маршрут можна ототожнити з перестановкою n чисел. Задача комівояжера має велику кількість змістовних аналогів. Зокрема, задача розробки графіка переналагодження устаткування, що може випускати різні типи виробів, але вимагає певних витрат (часових або матеріальних) при переході з одного технологічного режиму на інший.

Задача з розривними цільовими функціями. Багато економічних систем характеризуються наявністю так званих постійних витрат, які мають місце незалежно від обсягу виробництва. Врахування у моделях цих і подібних факторів призводить до появи в них цільових функцій, що не є безперервними. Як приклад можна навести транспортну задачу з фіксованими доплатами. У цьому випадку цільова функція сумарних витрат на перевезення містить «стрибкоподібні» розриви, що істотно утрудняє її мінімізацію.

Методи розв'язання задач цілочислового лінійного програмування засновані на використанні обчислювальних можливостей методів лінійного програмування. Зазвичай алгоритми цілочислового програмування включають три кроки:

1. «Ослаблення» простору припустимих розв'язків задачі шляхом відкидання вимоги цілочисловості. У результаті виходить звичайна задача лінійного програмування.

2. Визначення оптимального розв'язку задачі лінійного програмування.

3. Маючи отриманий оптимальний розв'язок, додають спеціальні обмеження, які ітераційним шляхом змінюють простір припустимих розв'язків задачі лінійного програмування так, щоб в остаточному підсумку вийшов оптимальний розв'язок, що задовольняє вимогам цілочисловості.

Розроблено два методи побудови спеціальних обмежень:

- метод Гоморі (або метод площин, що відсікають, він був уперше запропонованим Р.Гоморі у 1957-1958 р.);

- метод віток і границь.

Помітимо також, що досить ефективний і широко застосовуваний підхід до розв'язання цілочислових задач заснований на зведенні їх до задач транспортного типу. Це пояснюється тим, що якщо в умовах транспортної задачі значення запасів і потреб є цілочисловими, то цілочисловим буде й оптимальний план.

5.2. Метод відсікань Гоморі

Сутність методу Гоморі полягає в тому, що спочатку задачу вирішують без урахування умови цілочисловості. Якщо отриманий у результаті оптимальний план x^* містить тільки цілі компоненти, задачу вирішено. У протилежному випадку до системи обмежень задачі додають нове обмеження, що має наступні властивості:

- воно має бути лінійним;

- має відсікати знайдений нецілочисловий оптимальний план x^*;

- не повинне відсікати жодного цілочислового плану.

Додаткове обмеження, що має зазначені властивості, називають ***правильним відсіканням***. Далі задачу вирішують з урахуванням нового обмеження. Після цього, якщо буде потреба, додається ще одне обмеження й т.д.

Геометрично додавання кожного лінійного обмеження відповідає проведенню прямої (гіперплощини), що відсікає від багатокутника (багатогранника) розв'язків певну його частину разом з оптимальною точкою з нецілими координатами, але не зачіпає ні однієї із цілих точок цього багатогранника. В результаті новий багатогранник розв'язків містить всі цілі точки, утримувані в первісному багатограннику розв'язків, і відповідно отриманий при цьому багатограннику оптимальний розв'язок буде цілочисловим.

Один з алгоритмів розв’язання задачі лінійного цілочислового програмування, запропонований Гоморі, заснований на симплексному методі й використовує досить простий спосіб побудови правильного відсікання.

Нехай задача лінійного програмування має кінцевий оптимум і на останньому кроці її розв’язання за симплексним методом отримані наступні рівняння, у яких базисні змінні x_1, x_2, …, x_m виражені через вільні змінні x_{m+1}, x_{m+2}, …, x_n оптимального розв’язку

$$
\begin{aligned}
x_1 &= b_1 - a_{1\,m+1}x_{m+1} - \ldots - a_{1n}x_n, \\
x_2 &= b_2 - a_{2\,m+1}x_{m+1} - \ldots - a_{2n}x_n, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
x_m &= b_m - a_{m\,m+1}x_{m+1} - \ldots - a_{mn}x_n,
\end{aligned}
\tag{5.3}
$$

де оптимальним розв’язком задачі є $x^* = (b_1, b_2, \ldots, b_m, 0, 0, \ldots, 0)$, в якому, наприклад, b_i – нецілий компонент.

Нагадаємо, що цілою частиною числа b називають найбільше ціле число $[b]$, що не перевищує b, а дробовою частиною числа b - число $\{b\}$, що дорівнює різниці між цим числом і його цілою частиною, тобто $\{b\}= b-[b]$.

Сформувавши за і-м рівнянням системи (5.3) додаткове обмеження

$$\{b_i\} - \{a_{i\,m+1}\}x_{m+1} - \ldots - \{a_{in}\}x_n \le 0, \tag{5.4}$$

можна довести, що воно має всі властивості правильного відсікання.

Для розв’язання задачі цілочислового програмування методом Гоморі використовують наступний алгоритм:

1. Вирішити задачу за симплексним методом без урахування умови цілочисловості. Якщо всі компоненти оптимального плану цілі, то він є оптимальним і для задачі цілочислового програмування. Якщо перша задача є нерозв'язною (тобто не має кінцевого оптимуму або умови її суперечливі), то й друга задача також нерозв'язна.

2. Якщо серед компонентів оптимального розв’язку є нецілі, то вибрати компоненту з найбільшою цілою частиною і за відповідним їй рівнянням системи (5.3) сформувати правильне відсікання (5.4).

3. Отриману нерівність перетворюють на рівносильне рівняння шляхом введення додаткової невід’ємної цілочислової змінної

$$\{b_i\} - \{a_{i\,m+1}\}x_{m+1} - \ldots - \{a_{in}\}x_n + x_{n+1} \tag{5.5}$$

і включають його до системи обмежень. Необхідно пам'ятати, що після включення до системи обмежень додаткового рівняння, яке відповідає правильному відсіканню, завжди буде виходити неприпустимий базисний розв’язок. Для одержання припустимого базисного розв’язку потрібно перевести в базисні змінні одну з вільних змінних.

4. Отриману розширену задачу вирішити симплексним методом. Якщо знайдений оптимальний план буде цілочисловим, то задачу цілочислового програмування вирішено. У протилежному випадку треба повернутися до п.2 алгоритму.

Якщо задачу розв'язано в цілих числах, то після кінцевого числа кроків оптимальний цілочисловий план буде знайденим.

Якщо в процесі розв'язання з'явиться рівняння (що виражає базисну змінну через вільні) з нецілим вільним членом і цілими іншими коефіцієнтами, то відповідне рівняння не має розв'язання в цілих числах. У цьому випадку й задача не має цілочислового оптимального розв'язку.

Розглянемо особливості застосування методу Гоморі на прикладі. Нехай дано задачу з наступними умовами:

$$L = 3x_1 + x_2 + 2x_3 \to \max$$
$$x_1 + 2x_2 + 4x_3 = 9$$
$$3x_1 + x_2 + 5x_3 = 10 \qquad .$$
$$x_j \geq 0; \quad x_j - цілі; \quad j = \overline{1{,}3}$$

Використовуючи звичайний симплекс-алгоритм, вирішуємо безперервний аналог вихідної задачі, в якої ігноруються умови цілочисловості. Як вихідний базис можна взяти перший і другий стовпці. На його основі заповнюємо симплекс-таблицю:

Базис	$C_{баз}$	C_j	3	1	2
		B	A_1	A_2	A_3
A_1	3	11/5	1	0	6/5
A_2	1	17/5	0	1	7/5
	L_j	10	3	1	5
	Δ_j		0	0	3

Як видно з рядка оцінок, цей базис є оптимальним, проте відповідний йому план $x^* = \left(\frac{11}{5}; \frac{17}{5}; 0\right)$ не є цілочисловим, тому обираємо рядок, що містить перший нецілий елемент, і відповідно до формули (5.4) будуємо відсікаюче обмеження:

$$\left\{\frac{11}{5}\right\} - \left\{\frac{6}{5}\right\} x_3 \leq 0 \, .$$

Увівши додаткову цілочислову змінну $x_4 \geq 0$, одержимо рівносильне нерівності рівняння

$$-\frac{1}{5} x_3 + x_4 = -\frac{1}{5} \, ,$$

яке необхідно включити до системи обмежень. Введемо його до симплекс-таблиці

Базис	$C_{баз}$	C_j	3	1	2	0
		B	A_1	A_2	A_3	A_4
A_1	3	11/5	1	0	6/5	0
A_2	1	17/5	0	1	7/5	0
A_4	0	-1/5	0	0	-1/5	1
	L_j	10	3	1	5	0
	Δ_j		0	0	3	0

Врахуємо, що отриманий базисний розв'язок є неприпустимим (компонента b_3 від'ємна). Для одержання припустимого базисного розв'язку переведемо в базисні змінні вільну змінну x_3.

Базис	$C_{баз}$	C_j	3	1	2	0
		B	A_1	A_2	A_3	A_4
A_1	3	1	1	0	0	6
A_2	1	2	0	1	0	7
A_3	2	1	0	0	1	-5
	L_j	7	3	1	2	15
	Δ_j		0	0	0	15

Отриманий план x^*=(1, 2, 1, 0) є не тільки оптимальним (всі b_j>0), але й повністю складається з цілочислових компонент, тобто розв'язок задачі знайдено, max L = 7.

5.3. Метод гілок та границь

Метод віток і границь - один з комбінаторних методів. Його суть полягає у впорядкованому переборі варіантів і розгляді тих з них, які за певними ознаками є перспективними. Вперше цей метод запропонували у 1960 р. Ленг та Дойг, а його «друге народження» відбулося у 1963 р. у зв'язку з виходом роботи Літтла, Мурті, Суїні й Керел, присвяченої розв'язанню задачі про комівояжера.

Метод віток і границь полягає в тому, що множину припустимих розв'язків (планів) за певним способом розбивають на підмножини, кожну з яких за тим самим способом знову розбивають на підмножини. Процес триває доти, поки не буде отриманий оптимальний цілочисловий розв'язок вихідної задачі.

Якщо задача максимізації лінійної функції L вирішена за симплексним методом без урахування цілочисловості змінних, то стають відомими нижня й верхня границя для кожної цілочислової змінної x_j: $[x_j]<x_j<[x_j]+1$ і нижня границя лінійної функції L_0, тобто при будь-якому плані x $L(x)\geq L_0$. Тоді з області припустимих значень, наприклад, змінної x_r виключають область $[x_r]<x_r<[x_r]+1$, де $[x_r]$ – ціла частина числа x_r. В результаті це дозволяє сформулювати дві задачі, що відрізняються тим, що в одну з них додано обмеження $x_r \leq [x_r]$, а в іншу – обмеження $x_r \geq [x_r]+1$.

Сформульовані задачі вирішують у будь-якому порядку. Залежно від отриманого розв'язку список таких задач може розширюватися або зменшуватися. Якщо в ході розв'язання будь-якої із задач отриманий нецілочисловий оптимальний план, для якого $L(x) \leq L_0$, то цю задачу виключають зі списку. Якщо $L(x) \geq L_0$, то з цієї задачі формулюють дві нові задачі.

Розглянемо приклад:

$$L = 5x_1 + 4x_2 \to \max$$
$$x_1 + x_2 \leq 5$$
$$10x_1 + 6x_2 \leq 45$$
$$x_j \geq 0; \quad x_j - цілые; \quad j = \overline{1,2}$$

Вирішимо задачу шляхом відкидання умов цілочисловості. Використаємо функцію Excel «Пошук рішення» і одержимо

x_1=3,75; x_2=1,25; L=23,75.

Оберемо одну з цілочислових змінних, значення якої в оптимальному розв'язку не є цілочисловим, наприклад, x_1. Очевидно, що область $3<x_1<4$ простору припустимих розв'язків не містить цілочислових значень змінної x_1, і отже її можна виключити з розгляду. Це еквівалентно заміні вихідної задачі двома новими задачами. В одну з них додамо обмеження $x_1 \le 3$, одержимо

$$\begin{aligned} &L = 5x_1 + 4x_2 \to \max \\ &x_1 + x_2 \le 5 \\ &10x_1 + 6x_2 \le 45 \\ &x_1 \le 3 \\ &x_j \ge 0; \quad x_j - цілі; \quad j = \overline{1,2} \end{aligned}.$$

Використовуючи функцію Excel «Пошук рішення», дістанемо оптимальний цілочисловий план

$$x_1=3;\ x_2=2;\ L=23.$$

Якість отриманого цілочислового розв'язку оцінити неможливо, скільки розв'язок другої задачі може призвести до більшого значення цільової функції. Це треба перевірити. Сформулюємо другу задачу, додавши до неї обмеження $x_1 \ge 4$:

$$\begin{aligned} &L = 5x_1 + 4x_2 \to \max \\ &x_1 + x_2 \le 5 \\ &10x_1 + 6x_2 \le 45 \\ &x_1 \ge 4 \\ &x_j \ge 0; \quad x_j - цілі; \quad j = \overline{1,2} \end{aligned}.$$

В оптимальному розв'язку другої задачі змінна x_2 не є цілим числом:

$$x_1=4;\ x_2=0{,}83;\ L=23{,}33.$$

Оскільки значення змінної $x_2=0{,}83$ не є цілим числом, другу задачу необхідно досліджувати далі, розділивши її у свою чергу ще на дві задачі і т.д. Проілюструємо хід розв'язання схемою, наведеною на рисунку 5.2.

У ході розв'язання можна було спочатку як змінну розгалуження прийняти x_2. При цьому наступні обчислення можуть значно відрізнятися.

Очевидним недоліком алгоритму методу віток і границь під час розв'язання задач великої розмірності є необхідність перебрати занадто велику кількість варіантів, перш ніж буде знайдений оптимальний план.

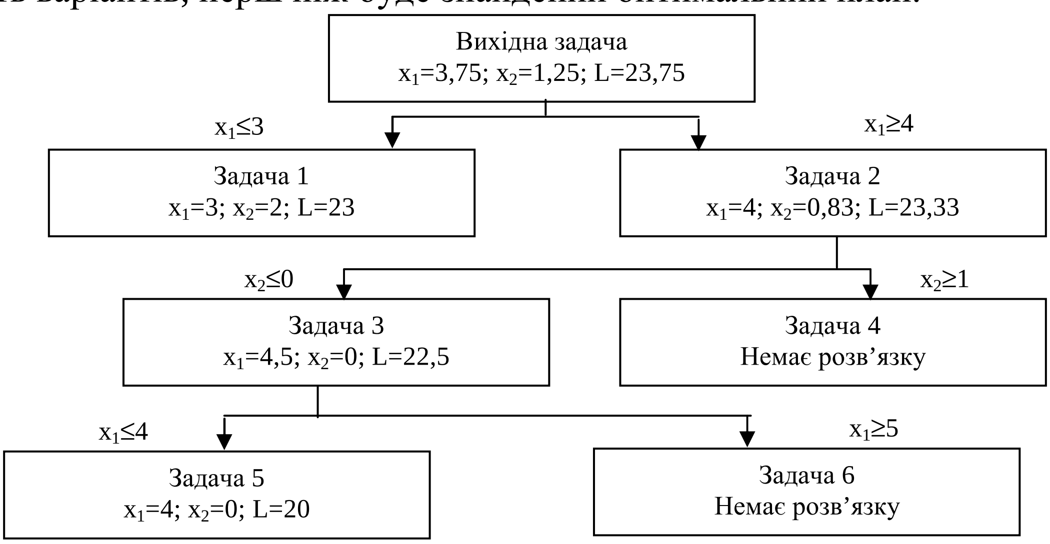

Рис. 5.2 – Схема методу віток і границь

Контрольні запитання

1. Які основні проблеми виникають при розв'язанні дискретних задач?
2. Сформулюйте задачу про ранець.
3. Які економіко-математичні моделі можна звести до задачі про комівояжера?
4. Наведіть приклади моделей з розривними цільовими функціями.
5. Який принцип використовують для побудови правильного відсікання в методі Гоморі?
6. Яку роль відіграє алгоритм двоїстого симплекс-методу при розв'язанні цілочислової лінійної задачі за методом Гоморі?
7. Перелічіть принципові ідеї, що лежать в основі методів віток і границь.
8. Як провадиться побудова відсікання при розв'язанні цілочислової лінійної задачі за методом віток і границь?
9. Опишіть схему розв'язання цілочислової задачі лінійного програмування за методом віток і границь.
10. За рахунок яких перетворень вдається побудувати сполучений базис під час додавання відсікаючого обмеження?

ТЕМА 6
НЕЛІНІЙНЕ ПРОГРАМУВАННЯ

6.1. Загальна постанова задачі нелінійного програмування

Залежності між керованими змінними далеко не завжди можна описати за допомогою адекватної лінійної моделі. Наприклад, у лінійних моделях ціну товару вважають незалежною від кількості реалізованого продукту, у той час як вона може залежати від обсягу партії товару. З приводу технологічних обмежень можна помітити, що витрата певних видів сировини й ресурсів відбувається не лінійно, а стрибкоподібно (залежно від обсягу виробництва). Спроби врахувати ці фактори призводять до формулювання загальніших та складніших оптимізаційних задач. Вивчення методів їх розв'язання складає предмет наукової області, що одержала назву *нелінійного програмування*.

Для задач нелінійного програмування не існує універсального методу розв'язання, тому щораз необхідно доводити існування розв'язку задачі, а так само його одиничність. Відомі точні методи розв'язання нелінійних задач, але алгоритми їх розв'язання є трудомісткими навіть для сучасного програмного забезпечення ЕОМ. На практиці частіше користуються наближеними методами, проблема яких пов'язана з пошуком локальних і глобальних оптимумів. Більшість наближених методів дозволяють визначити локальний оптимум. Визначивши всі локальні оптимуми й порівнявши їх, можна знайти глобальний оптимум. Але такий підхід не є ефективним для практичних розрахунків. Треба зазначити, що якщо в задачах лінійного програмування оптимальний розв'язок завжди перебував на границі області обмежень, то у задачі нелінійного програмування він може перебувати й усередині цієї області.

У класичній теорії оптимізації для пошуку точок максимуму й мінімуму (екстремальних точок) функцій, як за відсутності, так і за наявності обмежень на змінні використовують апарат диференціального обчислення. Екстремальна точка функції $f(x)$ визначає або її максимальне, або мінімальне значення. З математичної точки зору точка $x_0=(x_1, x_2, \ldots, x_n)$ є точкою максимуму, якщо значення функції f в оточенні точки x_0 не перевищують $f(x_0)$. На рис. 6.1 показані точки максимуму й мінімуму функції однієї змінної $f(x)$. Точки x_1, x_2, x_3, x_4 і x_6 складають множину екстремальних точок функції $f(x)$. Причому точка x_6 є точкою ***глобального (абсолютного)*** максимуму, тому що $f(x) = max\{f(x_1), f(x_3), f(x_6)\}$, а точки $f(x_1)$ та $f(x_3)$ – точками ***локального (відносного)*** максимуму. Необхідною умовою існування екстремуму є рівність нулю похідних від $f(x)$. Але похідні дорівнюють нулю і в точках перегину функції $f(x)$ у випадку однієї змінної (точка x_5), і в сідлових точках у випадку функції двох змінних. Тому рівність нулю похідних від $f(x)$ є необхідною, але недостатньою умовою існування екстремуму, а точки, у яких виконується ця умова, називають ***стаціонарними***.

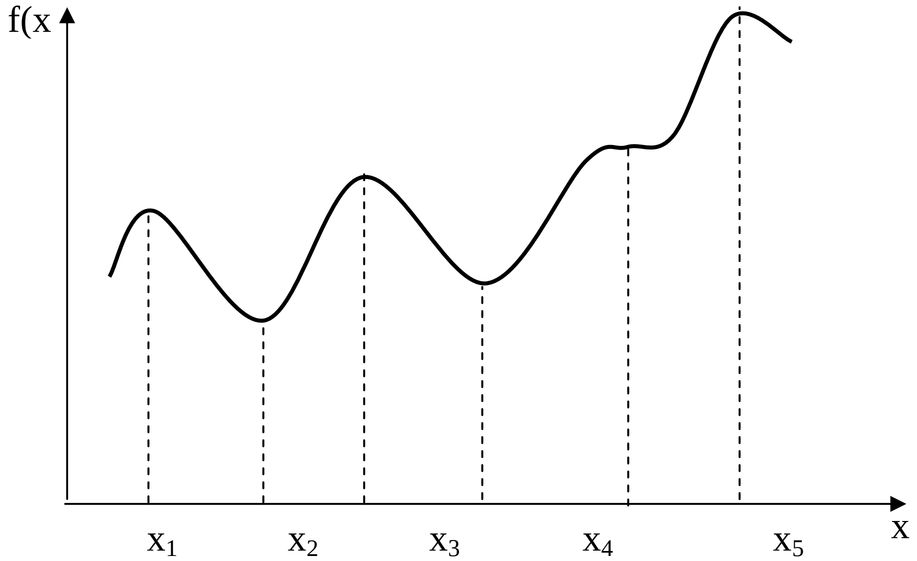

Рис. 6.1 – Точки максимуму та мінімуму функції

Більшість методів, використовуваних для розв'язання задач нелінійного програмування, базуються на теорії диференціального обчислення. Серед них розрізняють ***прямі*** й ***непрямі*** методи. За прямими методами пошук оптимуму ведеться в напрямі найскорішого зростання або убування цільової функції. До таких методів належать градієнтні методи, зокрема, метод найшвидшого спуску й метод припустимих напрямків. Непрямі методи припускають перетворення вихідної задачі до вигляду, що дозволяє спростити пошук екстремуму цільової функції. До них належить квадратичне програмування й сепарабельне програмування, а також геометричне й стохастичне програмування.

Загальну задачу нелінійного програмування (ЗЗНП) визначають як задачу із знаходження максимуму (або мінімуму) цільової функції $f(x_1, x_2,..., x_n)$ на множині D, зумовленій системою обмежень

$$\begin{gathered} g_1(x_1, x_2,..., x_n) \le 0 \\ g_2(x_1, x_2,..., x_n) \le 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ g_i(x_1, x_2,......., x_n) \le 0 \\ g_{i+1}(x_1, x_2,..., x_n) \le 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ g_m(x_1, x_2,......., x_n) = 0 \\ x_j \ge 0, \quad j = \overline{1,n} \end{gathered} \tag{6.1}$$

де *хоча б одна* з функцій f або g_i є нелінійною.

Очевидно, що питання про тип оптимізації не є принциповим. Тому, для визначеності, надалі будемо розглядати задачі максимізації.

Як і в ЗЛП, вектор $x^*=(x_1, x_2,..., x_n)\in D$ називають *припустимим планом,* а якщо для будь-якого $x\in D$ виконується нерівність $f(x^*)>f(x)$, то x^* називають *оптимальним планом.* У цьому випадку x^* є точкою глобального максимуму.

З погляду економічної інтерпретації $f(x)$ може розглядатися як дохід, що одержує підприємство при плані випуску x, а $g_i(x)<0$ як технологічні обмеження на можливості випуску продукції.

Набір обмежень, що визначають множину D, при необхідності завжди можна звести або до системи, що складається з одних нерівностей, або, додавши фіктивні змінні, до системи рівнянь. Перелічимо властивості ЗНП, які істотно ускладнюють процес їх розв'язання порівняно із задачами лінійного програмування.

1. Множина припустимих планів D може бути дуже складною за структурою (наприклад, бути неопуклою або незв'язною).

2. Глобальний максимум (мінімум) може досягатися як усередині множини D, так і на її границях (де він, загалом кажучи, буде не збігатися з жодним з локальних екстремумів).

3. Цільова функція f може бути недиференційованою, що утруднює застосування класичних методів математичного аналізу.

У чинність названих факторів задачі нелінійного програмування настільки різноманітні, що для них не існує загального методу розв'язання.

6.2. Класичний метод оптимізації з використанням множників Лагранжа

Одним з найзагальніших підходів до розв'язання задач з пошуку екстремуму функції за наявності сполучних обмежень на її змінні (*задача умовної оптимізації*) є ***метод Лагранжа***. Багатьом студентам він має бути відомим з курсу диференціального обчислення. Ідея цього методу полягає у зведенні задачі пошуку умовного екстремуму цільової функції

$$f(x_1, x_2,..., x_n) \tag{6.2}$$

на множині припустимих значень D, описуваній системою рівнянь

$$\begin{aligned} &g_1(x_1, x_2,..., x_n)= 0\\ &g_2(x_1, x_2,..., x_n)= 0\\ &\dots\dots\dots\dots\dots\dots\\ &g_m(x_1, x_2,......., x_n)= 0 \end{aligned} \tag{6.3}$$

до задачі безумовної оптимізації функції

$$\Phi(x, \lambda) = f(x_1, x_2,..., x_n)+ \lambda_1 g_1(x_1, x_2,..., x_n) + ...+ \lambda_m g_m(x_1, x_2,..., x_n), \tag{6.4}$$

де λ_i - вектор додаткових змінних, називаних *невизначеними множниками Лагранжа*. Рівняння (6.3) називають *рівняннями зв'язку*, а функціюя $\Phi(x, \lambda)$ - *функцією Лагранжа*.

Для функції $\Phi(x, \lambda)$ вирішують систему рівнянь

$$\begin{cases} \dfrac{\partial \Phi(x,\lambda)}{\partial x_j} = 0, \ (j = \overline{1,n}) \\ \dfrac{\partial \Phi(x,\lambda)}{\partial \lambda_i} = g_i(x) = 0, \ (i = \overline{1,m}) \end{cases} \tag{6.5}$$

щодо змінних x та λ.

Метод Лагранжа складається з наступних етапів.

1. Складання функції Лагранжа $\Phi(x,\lambda)$.

2. Знаходження частинних похідних

$$\frac{\partial \Phi(x,\lambda)}{\partial x_j}, \ j = \overline{1,n} \qquad \text{і} \qquad \frac{\partial \Phi(x,\lambda)}{\partial \lambda_i}, \ i = \overline{1,m}.$$

3. Розв’язання системи рівнянь (6.5) щодо змінних x та λ.

4. Дослідження точок, що задовольняють системі (6.5), на максимум (мінімум) за допомогою достатньої ознаки екстремуму.

Наявність останнього (четвертого) етапу пояснюється тим, що розглянутий алгоритм виконує необхідну, але не достатню умову екстремуму. Для визначення достатніх ознак умовного екстремуму та його типу існує спеціальний алгоритм, як правило, важко застосовний на практиці.

Основне практичне значення методу Лагранжа полягає в тому, що він *дозволяє перейти від умовної оптимізації до безумовної*. Проте задача розв’язання системи рівнянь (6.5), до якої збігається цей метод, у загальному випадку не простіше вихідної проблеми пошуку екстремуму (6.2)-(6.3).. Методи, що припускають таке розв’язання, називають *непрямими*. Їх можна застосовувати для досить вузького класу задач, для яких вдається одержати лінійну або таку, що збігається до лінійної систему рівнянь (6.5). Їх застосування пояснюється необхідністю одержати розв’язок екстремальної задачі в аналітичній формі. Під час розв’язання конкретних практичних задач зазвичай використовують *прямі* методи, засновані на ітеративних процесах обчислення й порівнянні значень функцій, що оптимізують.

6.3. Градієнтні методи розв’язання задач безумовної оптимізації

Провідне місце серед прямих методів розв’язання екстремальних задач займає *градієнтний метод* (точніше, сімейство градієнтних методів) пошуку стаціонарних точок диференційованої функції. Нагадаємо, що *стаціонарною* називають точку, у якій $\nabla f(x)=0$ і яка відповідно до необхідної умови оптимальності є «підозрілою» на наявність локального екстремуму. Отже, застосовуючи градієнтний метод, знаходять множину точок локальних максимумів (або мінімумів), серед яких визначають максимум (або мінімум) глобальний.

Ідея цього методу базується на тому, що градієнт функції вказує напрям її *найскорішого зростання* в оточенні тієї точки, в якій він обчислений. Тому, якщо з певної поточної точки $x^{(1)}$ переміщатися в напрямі вектору $\nabla f(x^{(1)})$, функція f зростатиме, принаймні, у певному оточенні $x^{(1)}$. Отже, для точки $x^{(2)} = x^{(1)} + k\nabla f(x^{(1)})$, $(k>0)$, що лежить у такому оточенні, справедлива нерівність $f(x^{(1)}) \le f(x^{(2)})$. Продовжуючи цей процес, ми поступово наближатимемося до точки певного локального максимуму.

Проте як тільки визначається напрям руху, відразу встає питання, як далеко треба рухатися в цьому напрямі, тобто виникає проблема вибору кроку *r* у рекурентній формулі

$$x^{(k+1)}=x^{(k)}+r\,\nabla f(x^{(k)}), \tag{6.6}$$

що задає послідовність точок, які прагнуть до точки максимуму.

Залежно від способу її розв’язання розрізняють різні варіанти градієнтного методу. Зупинимося на найвідоміших з них.

Метод найскорішого спуску. Назву методу можна було б розуміти буквально, якби мова йшла про мінімізацію цільової функції. Проте, за традицією таку назву використовують і під час розв’язання задачі на максимум.

Нехай $f(x) = f(x_l ,x_2,..,.x_n)$ - диференційована функція, а $x^{(k)}=(x_1^{(k)},x_2^{(k)},...,x_n^{(k)})$ - певна поточна точка. Будь-яких загальних рекомендацій щодо вибору вихідної точки (початкового наближення) $x^{(0)}$ не існує, але за можливістю вона повинна перебувати близько від шуканого оптимального плану x^*. Якщо $x^{(k)}$ - нестаціонарна точка, то під час руху в напрямі $\nabla f(x^{(k)})$ функція $f(x)$ на певному проміжку обов'язково зростатиме. Звідси виникає необхідність такого вибору кроку, щоб рух у зазначеному напрямі тривав доти, поки зростання не припиниться. Виразимо залежність значення $f(x)$ від крокового множника $r > 0$, покладаючи $x = x^{(k)} + r\,\nabla f(x^{(k)})$

$$f(x) = f(x^{(k)} + r\,\nabla f(x^{(k)})) = \varphi(r), \tag{6.7}$$

або в координатній формі,

$$\varphi(r) = f\left(x_1^{(k)} + r\frac{\partial f(x^{(k)})}{\partial x_1},\ldots,x_n^{(k)} + r\frac{\partial f(x^{(k)})}{\partial x_n}\right). \tag{6.8}$$

Щоб домогтися найбільшого з можливих значень *f* під час руху у напрямі $\nabla f(x^{(k)})$, потрібно вибрати таке значення *r*, що максимізує функцію $\varphi(r)$. Для обчислення *r* використовують необхідну умову екстремуму $\frac{d\varphi(r)}{dr}=0$. Помітимо, що якщо для будь-якого $r>0$ $\frac{d\varphi(r)}{dr}>0$, то функція $f(x)$ не обмежена зверху (тобто не має максимуму). У протилежному випадку, на підставі (6.8) одержуємо

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1}\times\frac{dx_1}{dr}+\cdots+\frac{\partial f(x)}{\partial x_n}\times\frac{dx_n}{dr}, \tag{6.9}$$

що, у свою чергу, дає

$$\frac{d\varphi(r)}{dr} = \frac{\partial f(x)}{\partial x_1}\times\frac{\partial f(x^{(k)})}{\partial x_1}+\cdots+\frac{\partial f(x)}{\partial x_n}\times\frac{\partial f(x^{(k)})}{\partial x_n} = \nabla f(x)\nabla f(x^{(k)}). \tag{6.10}$$

Якщо вважати, що наступна точка $x^{(k+1)}$ відповідає оптимальному значенню, то в ній має виконуватися умова $\frac{d\varphi(r)}{dr}=0$, і *r* треба знаходити з умови $\nabla f(x^{(k+1)})\nabla f(x^{(k)})=0$ або

$$\nabla f(x^{(k)} + r\nabla f(x^{(k)}))\nabla f(x^{(k)}) = 0. \tag{6.11}$$

Умова (7.11) означає рівність нулю скалярного добутку градієнтів функції

f в точках $x^{(k+1)}$ та $x^{(k)}$. Геометрично її можна інтерпретувати як перпендикулярність векторів градієнтів функції f у зазначених точках. Продовжуючи геометричну інтерпретацію методу найскорішого спуску, відзначимо, що в точці $x^{(k+1)}$ вектор $\nabla f(x^{(k+1)})$, будучи градієнтом, перпендикулярний до лінії рівня, що проходить через цю точку. Відповідно вектор $\nabla f(x^{(k)})$ є дотичним до цієї лінії. Отже, рух у напрямі градієнту $\nabla f(x^{(k)})$ треба продовжувати доти, поки він перетинає лінії рівня функції, що оптимізують.

Після того як точку $x^{(k+1)}$ знайдено, вона стає поточною для чергової ітерації. На практиці ознакою досягнення стаціонарної точки служить досить мала зміна координат точок, розглянутих на послідовних ітераціях. Одночасно з цим координати вектору $\nabla f(x^{(k)})$ мають наближатись до нуля.

6.4. Метод дроблення кроку

Для знаходження кроку r у методі найскорішого спуску потрібно вирішити рівняння (7.11), що може опинитися досить складним. Тому часто обмежуються «підбором» такого значення r, що $\varphi(r) > \varphi(0)$. Для цього задаються певним початковим значенням r_1 (наприклад, $r_1=1$) і перевіряють умову $\varphi(r_1) > \varphi(0)$. Якщо вона не виконується, то покладають

$$r_2 = \frac{r_1}{2}$$

і т.д. доти, поки не вдається знайти підходящий крок, з яким переходять до наступної точки $x^{(k+1)}$. Критерій завершення алгоритму буде таким самим, як і в методі найскорішого спуску.

6.5. Опукле програмування

Основний недолік методів нелінійного програмування полягає в тому, що за їх допомогою не вдається знайти глобальний екстремум за наявності кількох локальних екстремумів. Теоретично нелінійне програмування розроблене тільки для одного окремого випадку опуклих функцій, і відповідно цей розділ названий ***опуклим програмуванням***.

*Функцію $f(x_1, x_2,..., x_n)$ називають **опуклою** в області D, якщо для будь-яких двох точок* $x^{(1)}, x^{(2)} \in D$ *і будь-якого* $\lambda = \overline{1{,}0}$ *виконується нерівність*

$$f((1-\lambda)x^{(1)} + \lambda x^{(2)}) \le (1-\lambda) f(x^{(1)}) + \lambda f(x^{(2)})\,, \tag{6.12}$$

якщо

$$f((1-\lambda)x^{(1)} + \lambda x^{(2)}) \ge (1-\lambda) f(x^{(1)}) + \lambda f(x^{(2)})\,, \tag{6.13}$$

*то функція називається **увігнутою**.*

Геометричний зміст понять опуклості й увігнутості для функції однієї змінної поданий на рис. 6.2. З нього видно, що графік опуклої функції лежить нижче відрізка, що з'єднує точки $(x^{(1)}, f(x^{(1)}))$ і $(x^{(2)}, f(x^{(2)}))$, а графік увігнутої - вище.

Можна довести, що *достатньою умовою опуклості* функції $f(x_1, x_2,..., x_n)$ є *позитивна визначеність* матриці

$$H(x)=\left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{n\times n}, \tag{6.14}$$

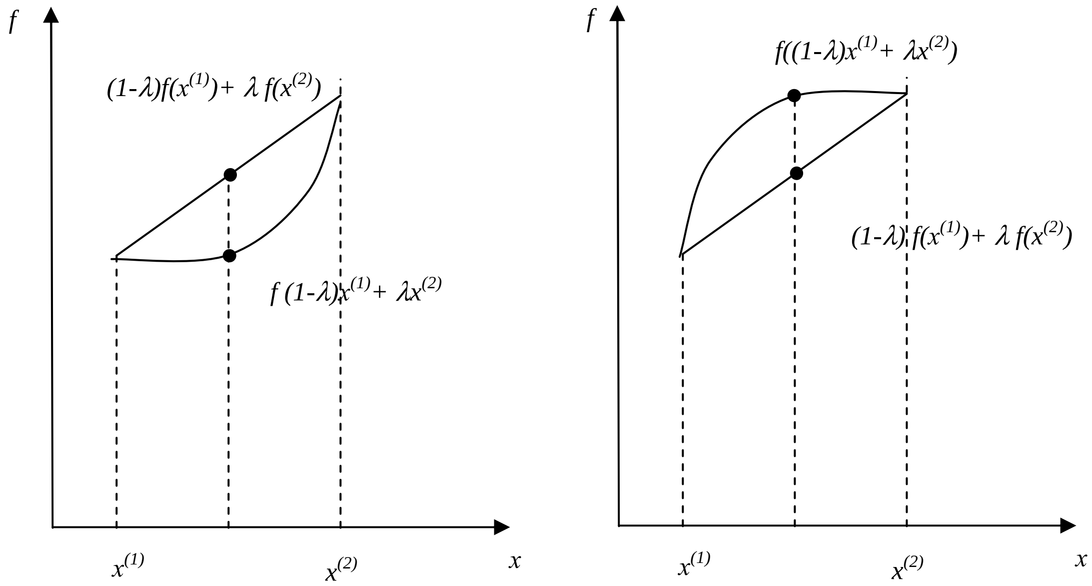

Рис. 6.2 - Графіки опуклої й увігнутої функцій

називаної *матрицею Гессе,* у всіх точках $x\in D$. Відповідно, *достатньою умовою ввігнутості* є *негативна визначеність* матриці Гессе. Зокрема, для функцій однієї змінної достатньою умовою опуклості (увігнутості) є виконання нерівності $f''(x)\geq 0$ ($f''(x)\leq 0$).

Як видно з геометричної інтерпретації, для опуклої функції локальний екстремум, якщо він існує, збігається із глобальним. Для функції багатьох змінних точка x_0 являє собою вектор $f'(x)$ − вектор перших похідних (градієнт) функції $f(x)$, а матриця Гессе (гессіан) $f''(x)$ − симетричну матрицю других частинних похідних функції $f(x)$. Характер стаціонарної точки x_0 пов'язаний зі знаковизначеністю матриці Гессе $f''(x^*)$.

Знаковизначеність матриці А залежить від знаків квадратичної форми

$$Q(x)=X^T AX=\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j\ . \tag{6.15}$$

Квадратична форма $Q(x)$ є позитивно визначеною (напіввизначеною), якщо значення всіх кутових мінорів визначника $|A|$ додатні (невід'ємні). У цьому випадку матрицю А називають позитивно визначеною (напіввизначеною). k-м кутовим мінором визначника матриці $A_{n\times n}$ називають визначник виду

$$\begin{vmatrix} a_{11}\ a_{12}\ldots a_{1k} \\ a_{21}\ a_{22}\ldots a_{2k} \\ \vdots\quad \vdots\quad \ldots \vdots \\ a_{k1}\ a_{k2}\ldots a_{kk} \end{vmatrix},\quad k=\overline{1,n}. \tag{6.16}$$

Квадратична форма $Q(x)$ є негативно визначеною, якщо значення k-х кутових мінорів визначника $|A|$ відмінні від нуля й мають знак $(-1)^k$. У цьому випадку матрицюА називають негативно визначеною.

Квадратична форма $Q(x)$ є негативно напіввизначеною, якщо значення k-х кутових мінорів визначника $|A|$ дорівнюють нулю або мають знак $(-1)^k$.

Розглянемо приклад. Нехай є функція

$$f(x_1, x_2, x_3) = x_1 + 2x_3 + x_2 x_3 - x_1^2 - x_2^2 - x_3^2,$$

необхідно визначити її екстремум. Визначимо градієнт функції $f(x)$ (необхідна умова екстремуму)

$$\frac{\partial f}{\partial x_1} = 1 - 2x_1 = 0$$
$$\frac{\partial f}{\partial x_2} = x_1 - 2x_2 = 0 \quad ,$$
$$\frac{\partial f}{\partial x_3} = 2 + x_2 - 2x_3 = 0$$

звідки одержимо стаціонарну точку $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$. Визначимо характер стаціонарної точки. Для цього перевіримо достатність умови, склавши матрицю Гессе

$$H(x_0) = \begin{pmatrix} \frac{\partial^2 f}{\partial^2 x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial^2 x_2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial^2 x_3} \end{pmatrix}_{x_0} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{pmatrix}$$

і кутові мінори матриці $H(x_0)$

$$M_1 = -2;\ M_2 = (-2)*(-2)-0*0 = 4;\ M_3 = -2*[(-2)*(-2)-1*1] = -6.$$

Отже, матриця $H(x_0)$ є негативно визначеною, звідки випливає, що функція $f(x)$ – увігнута, а стаціонарна точка $x_0 = \left(\frac{1}{2}, \frac{2}{3}, \frac{4}{3}\right)$ є точкою максимуму.

6.6. Теорема Куна-Таккера

Зупинимося на певних фундаментальних моментах теорії нелінійного програмування. Відправною точкою для них є поширення методу Лагранжа на розв'язання ЗНП із обмеженнями у формі нерівностей:

$$f(x_1, x_2, \ldots, x_n) \rightarrow max$$
$$g_1(x_1, x_2, \ldots, x_n) \le 0$$
$$g_2(x_1, x_2, \ldots, x_n) \le 0$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$g_m(x_1, x_2, \ldots\ldots, x_n) \le 0 \tag{6.17}$$

Визначимо функцію Лагранжа:

$$\Phi(x, \lambda) = f(x_1, x_2, \ldots, x_n) - \sum_{i=1}^{m} \lambda_i g_i(x_1, x_2, \ldots, x_n). \tag{6.18}$$

Пару векторів (x, λ) називають сідловою точкою функції Φ(x, λ) у певній області, якщо для будь-яких x і λ, що належать цієї області,

$$\Phi(x, \lambda^*) \le \Phi(x^*, \lambda^*) \le \Phi(x^*, \lambda)\,. \tag{6.19}$$

Нерівності (6.19) називають *нерівностями сідлової точки.*

Як приклад розглянемо функцію $\Phi(x,\lambda) = -x^2 + \lambda^2$. Сідловою точкою для цієї функції є точка (0, 0). Справді, Φ(0,0) = 0, $\Phi(x,0) = -x^2$, $\Phi(0, \lambda)=\lambda^2$, а для будь-яких x і λ виконуються нерівності $-x^2 \le 0$ і $0 \le \lambda^2$.

На рис. 6.3 зображений графік функції *Φ(x,λ)* (гіперболічний параболоїд), і, як видно, в оточенні точки (0,0) він дійсно за формою нагадує сідло, чим і пояснюється походження відповідного терміна.

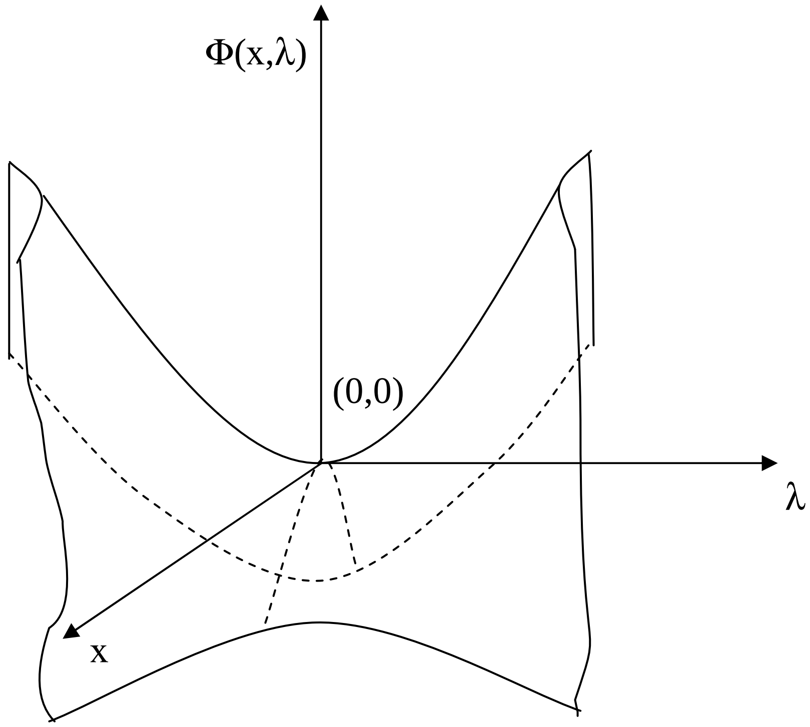

Рис. 6.3 - Графік функції Φ(x,λ)

Центральне місце в теорії нелінійного програмування займає ***теорема Куна-Таккера***, що пов'язує розв'язання ЗНП із наявністю сідлової точки у відповідної функції Лагранжа.

Теорема 6.1. (Достатня умова екстремуму).

Якщо (x^*, λ^*) *— сідлова точка функції Лагранжа в області* $x \in X \supseteq D$, *λ≥0, то* x^* *є оптимальним планом задачі* (6.17), причому справедливе так зване правило додаткової нетвердості (умова Слейтера):

$$\sum_{i=1}^{m} \lambda_i^* g_i(x^*) = 0\,. \tag{6.20}$$

Доказ.

Скористаємося визначенням сідлової точки й запишемо

$$f(x) - \sum_{i=1}^{m} \lambda_i^* g_i(x) \le f(x^*) - \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \le f(x^*) - \sum_{i=1}^{m} \lambda_i g_i(x^*) \tag{6.21}$$

при усіх $x \in X$, *λ≥0.* Із другої нерівності в (6.21) випливає, що

$$\sum_{i=1}^{m}\lambda_i g_i(x^*) \le \sum_{i=1}^{m}\lambda_i^* g_i(x^*) \text{ для } \lambda \ge 0. \qquad (6.22)$$

Проте (6.22) може мати місце тільки тоді, коли $g(x^*) \le 0$ при всіх $i=\overline{1,m}$. Дійсно, якщо існує таке k, що $g_k(x^*) > 0$, то поклавши $\lambda_i = 0$ для всіх $i \ne k$ і вибравши досить велике $\lambda_k > 0$, можна домогтися того, що значення $\sum_{i=1}^{m}\lambda_i g_i(x^*) = \lambda_k g_k(x^*)$ опиниться більшим за постійний вираз $\sum_{i=1}^{m}\lambda_i^* g_i(x^*)$.

З того, що для всіх $i=\overline{1,m}$ виконуються нерівності $g_i(x^*) \le 0$, випливає, що x^* є припустимим планом задачі (6.17).

Якщо до лівої частини нерівності (6.22) підставити значення $\lambda_i = 0$ для всіх $i=\overline{1,m}$, одержимо

$$\sum_{i=1}^{m}\lambda_i^* g_i(x^*) \ge 0.$$

Разом з тим з того що, $g_i(x^*) \le 0$ і $\lambda_i^* \ge 0$ випливає оцінка

$$\sum_{i=1}^{m}\lambda_i^* g_i(x^*) \le 0.$$

Спільний розгляд останніх двох нерівностей призводить до правила додаткової нетвердості у *точці* x^*:

$$\sum_{i=1}^{m}\lambda_i^* g_i(x^*) = 0.$$

Тоді на підставі лівої частини нерівності сідлової точки (6.21) маємо, що для всіх $x \in X$ (у тому числі й для $x \in D$)

$$f(x^*) \ge f(x) - \sum_{i=1}^{m}\lambda_i^* g_i(x).$$

За умовою ЗНП для будь-яких $x \in D$ вірні нерівності $g_i(x^*) \le 0$, що в сполученні з умовою $\lambda_i^* = 0$ дозволяє записати

$$\sum_{i=1}^{m}\lambda_i^* g_i(x) \le 0.$$

Виходить,

$$f(x) - \sum_{i=1}^{m}\lambda_i^* g_i(x) \ge f(x).$$

Остаточно одержуємо, що для будь-яких $x \in D$ справедливим є співвідношення $f(x^*) \ge f(x)$, тобто x^* - оптимальний план задачі (6.17).

Значення теореми Куна-Таккера полягає в тому, що вона дозволяє зв'язати процес розв'язання оптимізаційної задачі з пошуком сідлових точок функції Лагранжа, тобто з максимізацією цієї функції за x і мінімізацією за λ..

Нехай $F(x)$ є функцією, що ставить у відповідність кожному значенню x

мінімальне значення функції $\Phi(x,\lambda)$ за λ:

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda)$$

і за аналогією

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda).$$

Розглянемо задачу відшукання максимуму функції $F(x)$

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda) \to \max, \quad x \in X \tag{6.23}$$

і задачу мінімізації $G(\lambda)$

$$G(\lambda) = \max_{x \in X} \Phi(x, \lambda) \to \min, \quad \lambda \geq 0. \tag{6.24}$$

Очевидно, що

$$F(x) = \min_{\lambda \geq 0} \Phi(x, \lambda) = \begin{cases} f(x), \; x \in D, \\ \infty, \quad x \notin D \end{cases}.$$

Звідси випливає, що максимум $F(x)$ перебуває в припустимій області D і збігається з максимумом цільової функції $f(x)$ задачі (7.11):

$$\max_{x \in X} F(x) = \max_{x \in X} \min_{\lambda \geq 0} \Phi(x, \lambda) = \max_{x \in D} f(x).$$

Отже, задача (6.23), у певному змісті, рівносильна (6.17). Аналогічні висновки можна отримати і для (6.24). Задачі (6.23) та (6.24) утворюють двоїсту пару. Це відношення є узагальненням відношення подвійності для задач лінійного програмування. Відповідно, за певних умов пара двоїстих задач нелінійного програмування має властивості, аналогічні властивостям двоїстих лінійних задач. Зокрема, за будь-яких $x \in X$, $\lambda \geq 0$

$$F(x) \leq G(\lambda). \tag{6.25}$$

Умова (6.25) знаходить широке застосування під час побудови оцінок в ітеративних методах розв’язання оптимізаційних задач. Наприклад, якщо є можливість приблизно вирішити пряму й двоїсту задачі та одержати послідовності наближень, то за допомогою нерівностей виду

$$f(x^{(k)}) \leq f(x^*) \leq G(\lambda^{(k)})$$

можна визначити момент зупинки обчислювальної процедури.

Можливим є варіант виведення виразів для цільових функцій та обмежень пари двоїстих задач лінійного програмування із загального визначення відносини подвійності для нелінійних задач. Відзначимо, що в процесі формування нелінійних двоїстих задач існує велика неоднозначність: їх вид можна варіювати, включаючи до множини X частину обмежень $g_i(x) \leq 0$.

6.7. Квадратичне програмування

До задач квадратичного програмування належить спеціальний клас задач нелінійного програмування, для яких цільова функція $f(x)$ є квадратичною, а всі обмеження - лінійні. Застосувавши до цієї задачі теорему Куна-Таккера, одержують умову для оптимального розв’язку у вигляді системи лінійних рівнянь, вирішити які можна за симплекс-методом. У матричному вигляді задачу квадратичного програмування формулюють в такий спосіб:

Максимізувати (мінімізувати) цільову функцію $f = \overline{C}\,\overline{X} + \overline{X}^T\,\overline{D}\,\overline{X}$ за обмежень

$$\overline{A}\,\overline{X} \le b, \quad \overline{X} \ge 0, \tag{6.26}$$

де

$$\overline{X} = (x_1, x_2, ..., x_n)^T,$$

$$\overline{C} = (c_1, c_2, ..., c_n),$$

$$b=(b_1, b_2, \ldots, b_m)^T,$$

$$\overline{A} = \begin{pmatrix} a_{11}\, a_{12} \ldots a_{1n} \\ a_{21}\, a_{22} \ldots a_{2n} \\ \vdots \quad \vdots \quad \ldots \vdots \\ a_{m1}\, a_{m2} \ldots a_{mn} \end{pmatrix},$$

$$\overline{D} = \begin{pmatrix} d_{11}\, d_{12} \ldots d_{1n} \\ d_{21}\, d_{22} \ldots d_{2n} \\ \vdots \quad \vdots \quad \ldots \vdots \\ d_{n1}\, d_{n2} \ldots d_{nn} \end{pmatrix}.$$

Функція $\overline{X}^T\,\overline{D}\,\overline{X}$, де D – симетрична матриця, що є квадратичною формою (6.15). Матриця D є негативно визначеною в задачі максимізації й позитивно визначеною – у задачі мінімізації. Це означає, що функція f є строго опуклою за змінними $\overline{X}$ у задачі мінімізації й строго ввігнутою – у задачі максимізації. Обмеження задачі є лінійними, що гарантує опуклість області припустимих розв’язків.

Складемо функцію Лагранжа для задачі (6.26)

$$\Phi(\overline{X}, \overline{\lambda}) = \overline{C}\,\overline{X} + \overline{X}^T\,\overline{D}\,\overline{X} + \overline{\lambda}(b - \overline{A}\,\overline{X}) \tag{6.27}$$

Відповідно до теореми Куна-Таккера для існування сідлової точки необхідно й достатньо, щоб похідні функції Лагранжа за x_j були менші за нуль, а за λ_i – більші за нуль, маємо:

$$\frac{\partial\Phi(\overline{X}, \overline{\lambda})}{\partial x_j} = \overline{C} + 2\overline{D}\,\overline{X} - \overline{\lambda}\,\overline{A} \le 0, \tag{6.28}$$

причому, якщо $\frac{\partial\Phi}{\partial x_j} < 0$, відповідні x_j дорівнюють нулю;

$$\frac{\partial\Phi(\overline{X}, \overline{\lambda})}{\partial \lambda_i} = b - \overline{A}\,\overline{X} \ge 0, \tag{6.29}$$

якщо $\frac{\partial\Phi}{\partial \lambda_i} > 0$, відповідні λ_i дорівнюють нулю.

Щоб нерівності (6.28) та (6.29) призвести до рівностей, введемо два допоміжних вектори: $\overline{V} = (v_1, v_2, \ldots, v_n) \ge 0$ та $\overline{S} = (s_1, s_2, \ldots, s_m) \ge 0$, причому оберемо $v_j>0$, якщо $\frac{\partial\Phi}{\partial x_j} < 0$ й $v_j=0$, якщо $\frac{\partial\Phi}{\partial x_j} = 0$, а також $\lambda_i>0$, якщо

$\frac{\partial \Phi}{\partial \lambda_i} > 0$ й λ_i=0, якщо $\frac{\partial \Phi}{\partial \lambda_i} = 0$, одержимо

$$\overline{C} + 2\overline{D}\overline{X} - \overline{\lambda}\overline{A} + \overline{V} = 0 \qquad (6.30)$$

$$b - \overline{A}\overline{X} - \overline{S} = 0 \qquad (6.31)$$

Порівнявши компоненти векторів $\overline{X}$ та $\overline{V}$, а також $\overline{\lambda}$ та $\overline{S}$, одержимо дві умови додаткової нетвердості

$$\overline{X}^T \overline{V} = 0 \text{ і } \overline{\lambda}^T \overline{S} = 0 \qquad (6.32)$$

З умов (6.32) випливає, що хоча б n змінних з $\overline{X}$ та $\overline{V}$, а також m змінних з $\overline{\lambda}$ та $\overline{S}$ обертаються на нуль. Якщо існує оптимальний розв’язок задачі (6.26), то він є одним з базисних розв’язків системи (6.30), (6.31). Для знаходження припустимого базисного розв’язку використовується симплекс-метод.

Розглянемо приклад. Знайти

$$\max f(x_1,x_2)=\max(10x_1+20x_2+x_1x_2-2x_1^2-2x_2^2)$$

за обмежень:

$$8-x_2 \geq 0;$$
$$10-x_1-x_2 \geq 0;$$
$$x_1 \geq 0;\ x_2 \geq 0.$$

Можна показати, що функція $f(x_1,x_2)$ є увігнутою функцією. Складемо функцію Лагранжа

$$\Phi(x_1,x_2,\lambda_1,\lambda_2) = 10x_1+20x_2+x_1x_2-2x_1^2-2x_2^2+\lambda_1(8-x_2)+\lambda_2(10-x_1-x_2).$$

Застосувавши теорему Куна-Таккера, одержимо умови існування сідлової точки:

$$\frac{\partial \Phi}{\partial x_1} = 10 + x_2 - 4x_1 + \lambda_2 \leq 0$$
$$\frac{\partial \Phi}{\partial x_2} = 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \leq 0 \, .$$
$$\frac{\partial \Phi}{\partial \lambda_1} = 8 - x_2 \geq 0$$
$$\frac{\partial \Phi}{\partial \lambda_2} = 10 - x_1 - x_2 \geq 0$$

Одержали систему лінійних нерівностей:

$$\begin{cases} 10 + x_2 - 4x_1 + \lambda_2 \leq 0 \\ 20 + x_1 - 4x_2 - \lambda_1 - \lambda_2 \leq 0 \\ 8 - x_2 \geq 0 \\ 10 - x_1 - x_2 \geq 0 \end{cases} .$$

Введемо вільні змінні $\overline{V}$ й $\overline{S}$, що обертають систему нерівностей на систему лінійних рівнянь, вирішивши яку за симплекс-методом, знайдемо оптимум

$$\begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 = 10 \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 = 20 \\ x_2 + s_1 = 8 \\ x_1 + x_2 + s_2 = 10 \end{cases} .$$

Для знаходження припустимого базисного плану отриманої задачі скористаємося методом мінімізації нев'язань. Для цього побудуємо допоміжну задачу, увівши до обмежень 1 і 2 фіктивні змінні x_3 та x_4, які входять до цільової функції допоміжної задачі з коефіцієнтами M (а інші змінні – з нульовими коефіцієнтами). Отже, допоміжна задача має вигляд

$$f'(x)=Mx_3+Mx_4,$$

$$\begin{cases} 4x_1 - x_2 - \lambda_2 - v_1 + x_3 = 10 \\ -x_1 + 4x_2 + \lambda_1 + \lambda_2 - v_2 + x_4 = 20 \\ x_2 + s_1 = 8 \\ x_1 + x_2 + s_2 = 10 \end{cases}.$$

Складемо симплекс-таблицю

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_7	M	10	4	-1	0	-1	-1	0	1	0	0	0
A_8	M	20	-1	4	1	1	0	-1	0	1	0	0
A_9	0	8	0	1	0	0	0	0	0	0	1	0
A_{10}	0	10	1	1	0	0	0	0	0	0	0	1
	L_j	30M	3M	3M	M	0	-M	-M	M	M	0	0
	Δ_j		3M	3M	M	0	-M	-M	0	0	0	0

Оскільки в першу чергу необхідно мінімізувати нев'язання, виведемо з базису вектор $\overline{A_7}$ і введемо вектор $\overline{A_1}$ (тим самим змінна x_3 дорівнюватиме нулю, а змінна x_1 увійде до базису). Складемо нову симплекс-таблицю.

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0	10/4	1	-1/4	0	-1/4	-1/4	0	1/4	0	0	0
A_8	M	90/4	0	15/4	1	3/4	-1/4	-1	1/4	1	0	0
A_9	0	8	0	1	0	0	0	0	0	0	1	0
A_{10}	0	30/4	0	5/4	0	¼	¼	0	-1/4	0	0	1
	L_j	90M/4	0	15M/4	M	3M/4	-M/4	-M	M/4	M	0	0
	Δ_j		0	15M/4	M	3M/4	-M/4	-M	3M/4	0	0	0

Тепер виведемо з базису вектор $\overline{A}_8$ і введемо $\overline{A}_2$

Базис	$C_{баз}$	C_j	0	0	0	0	0	0	M	M	0	0
		B	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0	4	1	0	1/15	-1/5	-4/15	-1/15	4/15	1/15	0	0
A_2	0	6	0	1	4/15	3/15	-1/15	-4/15	1/15	4/15	0	0
A_9	0	2	0	0	-4/15	-1/5	1/15	4/15	-1/15	-4/15	1	0
A_{10}	0	0	0	0	-1/3	0	1/3	1/3	-1/3	-1/3	0	1
	L_j	0	0	0	0	0	0	0	0	0	0	0
	Δ_j	0	0	0	0	0	0	0	-M	-M	0	0

Дістали припустимий базисний план вихідної задачі:

$$x_1=4;\ x_2=6;\ \lambda_1=0;\ \lambda_2=0;\ v_1=0;\ v_2=0;\ s_1=2;\ s_2=0.$$

Перевіримо виконання умов правила додаткової нетвердості (6.32)

$$\overline{X}^T\overline{V}=0 \quad \text{і} \quad \overline{\lambda}^T\overline{S}=0:$$

$$x_1*v_1=0;\ x_2*v_2=0;\ \lambda_1*s_1=0;\ \lambda_2*s_2=0.$$

Оскільки умови додаткової нетвердості виконуються, отриманий план є оптимальним. Оптимальне значення цільової функції

$$f(x_1,x_2)=10*4+20*6+4*6-2*4^2-2*6^2=80.$$

Контрольні запитання

1. За яких умов оптимізаційну задачу можна віднести до класу нелінійних?
2. Наведіть приклад економічної моделі, що зводиться до задачі нелінійного програмування.
3. Перелічіть основні труднощі, що виникають у процесі розв’язання задачі нелінійного програмування.
4. Який зміст вкладають в поняття «умовна оптимізація»?
5. Для чого призначений метод множників Лагранжа й у чому він полягає?
6. Яку точку множини розв’язків називають стаціонарною?
7. Які принципові етапи входять у градієнтні методи?
8. Для розв’язання яких задач призначений метод найскорішого спуску й метод дроблення кроку?
9. Дайте визначення опуклої (увігнутої) функції.
10. Сформулюйте достатню умову опуклості (увігнутості) функції.
11. У чому полягає специфіка задач опуклого програмування?
12. Дайте визначення сідлової точки. Наведіть приклад функції, що має сідлову точку.
13. Сформулюйте необхідну й достатню умови теореми Куна-Таккера. Яке значення вони мають для розв’язання задач нелінійного програмування?
14. У чому полягає умова регулярності Слейтера? Поясніть її зміст.
15. Наведіть приклад пари двоїстих задач нелінійного програмування.
16. Які властивості пари нелінійних двоїстих задач можна застосувати для їх розв’язання?

ТЕМА 7
ДИНАМІЧНЕ ПРОГРАМУВАННЯ

7.1. Загальна схема методів динамічного програмування

Деякі задачі математичного програмування мають специфічні особливості, які дозволяють звести їх розв’язання до розгляду низки простіших підзадач. В результаті глобальна оптимізація певної функції збігається до поетапної оптимізації проміжних цільових функцій. У динамічному програмуванні розглядають методи, що дозволяють шляхом поетапної (багатокрокової) оптимізації дістати загальний оптимум.

За методами динамічного програмування оптимізують роботу керованих систем, ефективність яких оцінюється адитивною, або мультиплікативною, цільовою функцією. Адитивною називають таку функцію кількох змінних $f(x_1, x_2, \ldots, x_n)$, значення якої обчислюється як сума функцій f_j, що залежать тільки від однієї змінної x_j:

$$f(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} f_j(x_j) \tag{7.1}$$

Доданки адитивної цільової функції відповідають ефективності розв’язків, прийнятих на окремих етапах керованого процесу. За аналогією, мультиплікативна функція розпадається на добуток позитивних функцій різних змінних:

$$\mathrm{f}(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n) = \prod_{j=1}^{n} \mathrm{f}_j(\mathrm{x}_j). \tag{7.2}$$

Оскільки логарифм функції (7.2) є адитивною функцією, обмежимося розглядом функцій виду (7.1).

Обчислення в динамічному програмуванні виконують рекурентно (рекурентний означає поворотний), так що оптимальне розв’язання однієї підзадачі використовують як вихідні дані для наступної. Вирішивши останню підзадачу, одержують оптимальний розв’язок вихідної задачі - глобальний оптимум.

Розглянемо приклад. Нехай необхідно обрати найкоротший шлях між двома містами. Сітка доріг (рис. 7.1) утворює можливі маршрути між вихідним містом, що перебуває у вузлі 1, та кінцевим пунктом, що перебуває у вузлі 7. Маршрути проходять через проміжні міста, позначені вузлами з номерами 2-6.

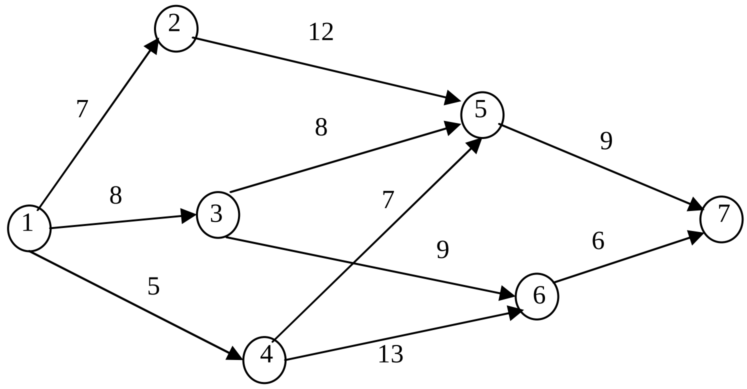

Рис. 7.1 - Сітка доріг, що утворює можливі маршрути

Цю задачу можна вирішити шляхом перебору всіх маршрутів між вузлами 1 та 7. Проте у великій сітці такий підхід є неефективним. Для розв’язання задачі за методом динамічного програмування розділимо її на три етапи (рис. 7.2) та виконаємо обчислення для кожного етапу окремо.

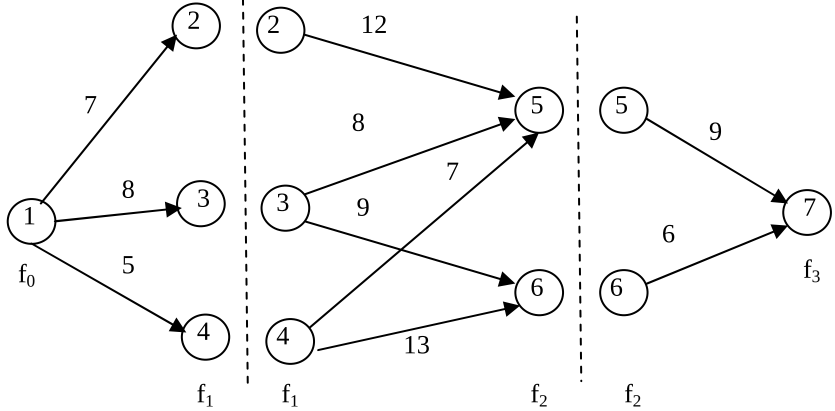

Рис. 7.2 - Декомпозиція задачі

Обчислимо найкоротші відстані до всіх вузлів першого етапу з наступним використанням цих відстаней як вихідних даних для другого етапу. На першому етапі кожний з вузлів 2, 3 та 4 пов'язаний з початковим вузлом 1 єдиною дугою, маємо

найкоротший шлях (1-2) дорівнює f_1=7 км;

найкоротший шлях (1-3) дорівнює f_1=8 км;

найкоротший шлях (1-4) дорівнює f_1=5 км.

На другому етапі обчислимо найкоротші відстані до вузлів 5 і 6. Розглянемо вузол 5. До нього ведуть три маршрути (2, 5), (3, 5) і (4, 5). Найкоротша відстань до вузла 5 з урахуванням відстаней до вузлів 2, 3 і 4 визначиться в такий спосіб

$$f_2=\min_{i=2,3,4}\left\{\begin{matrix}f_1+\text{відстань від}\\ \text{вузла i до вузла 5}\end{matrix}\right\}=\min\begin{pmatrix}7+12=19\\ 8+8=16\\ 5+7=12\end{pmatrix}=12\ (\text{з вузла 4}).$$

Аналогічно для вузла 6 одержимо

$$f_2=\min_{i=2,3,4}\left\{\begin{matrix}f_1+\text{відстань від}\\ \text{вузла i до вузла 6}\end{matrix}\right\}=\min\begin{pmatrix}8+9=17\\ 5+13=18\end{pmatrix}=17\ (\text{з вузла 3}).$$

Звернемося до третього етапу. Кінцевий вузол 7 можна досягти як з вузла 5 так і з вузла 6. Використовуючи результати другого етапу, одержимо

$$f_3=\min_{i=5,6}\left\{\begin{matrix}f_2+\text{відстань від}\\ \text{вузла i до вузла 7}\end{matrix}\right\}=\min\begin{pmatrix}12+9=21\\ 17+6=23\end{pmatrix}=21\ (\text{з вузла 5}).$$

Остаточно, найкоротший шлях до вузла 7 становить 21 км, оптимальний маршрут - послідовність вузлів 1, 4, 5, 7.

Рекурентні обчислення динамічного програмування виражають математично в такий спосіб. Нехай $f_i(x_i)$ – найкоротша відстань до вузла x_i на етапі i, $d(x_{i-1}, x_i)$ – відстань від вузла x_{i-1} до вузла x_i, тоді f_i обчислюється на підставі рекурентного виразу

$$f_i(x_i)=\min\{d(x_{i-1},x_i)+f_{i-1}(x_{i-1})\},\quad i=\overline{1,n}. \tag{7.1}$$

Рівняння (7.1) показує, що функція цілі $f_i(x_i)$ на i-му етапі має бути вираженою як функція попереднього етапу. У термінології динамічного програмування x_i називають ***станом системи*** на етапі i. Стан системи на етапі i – це інформація, що пов'язує етапи між собою, при цьому оптимальні розв'язки для етапів, що залишилися, можуть прийматися без повторної перевірки того, як були отримані розв'язки на попередніх етапах. Таке визначення стану системи дозволяє розглядати кожний етап окремо й гарантує, що розв'язок є припустимим на кожному етапі. Отже, можна сформулювати ***принцип оптимальності***:

> *На кожному етапі оптимальна стратегія визначається незалежно від стратегій, використованих на попередніх етапах. (Оптимальна стратегія керування не залежить від передісторії системи, а залежить від стану системи в поточний момент часу й цілі керування).*

Послідовність обчислень, використовану в прикладі, називають ***алгоритмом прямого прогону***. Цю саму задачу можна вирішити за допомогою ***алгоритму зворотного прогону***, за яким обчислення проводять від третього етапу до першого. Алгоритми прямого й зворотного прогону призведуть до того самого результату, але на практиці частіше використовують алгоритм зворотного прогону, оскільки він ефективніше з обчислювальної точки зору.

Для алгоритму зворотного прогону рекурентне рівняння (7.1) має вигляд

$$f_i(x_i)=\min\{d(x_i,x_{i+1})+f_{i+1}(x_{i+1})\},\quad i=\overline{1,n}, \tag{7.2}$$

Вирішимо задачу за алгоритмом зворотного прогону.

Етап 3. Вузол 7 пов'язаний з вузлами 5 і 6 тільки одним маршрутом. Результати третього етапу занесемо в таблицю

	$d(x_3, x_4)$	Оптимальний розв'язок	
x_3	$x_4=7$	$f_3(x_3)$	x^*_4
5	9	9	7
6	6	6	7

Етап 2. Використовуючи значення $f_3(x_3)$, отримані на третьому етапі, порівняємо припустимі альтернативні розв'язки в наступній таблиці

	$d(x_2, x_3)+ f_3(x_3)$		Оптимальний розв'язок	
x_2	$x_3=5$	$x_3=6$	$f_2(x_2)$	x^*_3
2	12+9=21	-	21	5
3	8+9=17	9+6=15	15	6
4	7+9=16	13+6=19	16	5

Оптимальний розв'язок другого етапу означає, що з вузла 2 або 4 найкоротший шлях до вузла 7 проходить через вузол 5, а з вузла 3 - через вузол 6.

Етап 1. Використовуючи значення $f_2(x_2)$, отримані на другому етапі, порівняємо припустимі альтернативні розв'язки в наступній таблиці

	$d(x_2, x_3)+ f_3(x_3)$			Оптимальний розв'язок	
x_1	$x_2=2$	$x_2=3$	$x_2=4$	$f_1(x_1)$	x^*_2
1	7+21=28	8+15=23	5+16=21	21	4

7.2. Основні типи задач і моделі динамічного програмування

Задача про завантаження. Це задача про раціональне завантаження судна, що має обмеження за обсягом або вантажопідйомністю. Кожний розміщений на судні вантаж приносить певний прибуток. Задача полягає у визначенні завантаження судна такими вантажами, які принесуть найбільший сумарний прибуток. Ця задача відома також як *задача про ранець*, в якої треба визначити найцінніші предмети, що підлягають завантаженню в ранець.

Нехай вантажопідйомність судна W предметів n найменувань, m_i – кількість предметів i-го найменування, що підлягають завантаженню, r_i – прибуток, що приносить один завантажений предмет i-го найменування, w_i – вага одного предмета i-го найменування. Формулювання задачі має вигляд:

Максимізувати $f=r_1m_1+r_2m_2+\ldots+r_nm_n$

за умови

$$w_1m_1+w_2m_2+\ldots+w_nm_n\leq W;$$
$$m_1, m_2,\ldots,m_n\geq 0 \quad \text{і цілі.}$$

Елементи моделі динамічного програмування визначають в такий спосіб. Етап i ставиться у відповідність предмету i-го найменування; варіанти розв'язку на етапі i описують кількістю m_i предметів i-го найменування, що підлягають завантаженню. Відповідний прибуток дорівнює r_im_i. Значення m_i лежить у межах від 0 до $\frac{W}{w_i}$. Стан x_i на етапі i виражає сумарну вагу предметів, рішення про завантаження яких прийняті на етапах i.

Рекурентне рівняння має вигляд

$$f_i(x_i) = \max\{r_i m_i + f_{i+1}(x_i - w_i m_i)\}, \quad i = \overline{1,n} \tag{7.3}$$

Задача про завантаження є типовим представником задачі розподілу ресурсів, у якій обмежений ресурс розподіляють між кінцевим числом видів економічної діяльності. При цьому ціллю є максимізація прибутку. В таких моделях станом на етапі i є сумарна кількість ресурсу, що розподіляється на відповідному етапі.

Розглянемо приклад. У 4-тонне судно завантажують предмети трьох найменувань. Дані про вагу одного предмета й прибутку, одержуваного від одного завантаженого предмета, наведені в таблиці

Предмет i	$\mathbf{w_i}$	$\mathbf{r_i}$
1	2	31
2	3	47
3	1	14

Етап 3. Вага, що може бути завантаженою на етапі 3 (предмет 3) може приймати одне із значень 0, 1, 2, 3, 4 (оскільки вантажопідйомність судна 4 тонни). Основою для порівняння варіантів на даному етапі є співвідношення

$$f_3(x_3) = \max\{14m_3\}, \quad \max\{m_3\} = \frac{4}{1} = 4.$$

Припустимі розв'язки для кожного значення x_3 порівняємо в таблиці

	14 m_3					**Оптимальний розв'язок**	
x_3	m_3=0	m_3=1	m_3=2	m_3=3	m_3=4	$f_3(x_3)$	m^*_3
0	0	-	-	-	-	0	0
1	0	14	-	-	-	14	1
2	0	14	28	-	-	28	2
3	0	14	28	42	-	42	3
4	0	14	28	42	56	56	4

Етап 2. $f_2(x_2) = \max\{47m_2 + f_3(x_2 - 3m_2)\}, \quad \max\{m_2\} = \frac{4}{3} = 1$

	47 m_2+f_3(x_2-3m_2)		**Оптимальний розв'язок**	
x_2	m_2=0	m_2=1	$f_2(x_2)$	m^*_2
0	0+0=0	-	0	0
1	0+14=14	-	14	0
2	0+28=28	-	28	0
3	0+42=42	47+0=47	47	1
4	0+56=56	47+14=61	61	1

Етап 1. $f_1(x_1) = \max\{31m_1 + f_2(x_1 - 2m_1)\}, \quad \max\{m_1\} = \frac{4}{2} = 2$.

	31 m_3 + $f_2(x_1$-2$m_1)$			**Оптимальний розв'язок**	
x_1	m_1=0	m_1=1	m_1=2	$f_1(x_1)$	m^*_1
0	0+0=0	-	-	0	0
1	0+14=14	-	-	14	0
2	0+28=28	31+0=31	-	31	1
3	0+47=47	31+14=45	-	47	0
4	0+61=61	31+28=59	62+0=62	62	2

З умови W=4 випливає, що перший етап при x_1=4 дає оптимальний розв'язок, що означає, що будуть завантажені два предмети першого найменування. Це завантаження залишає для предмета другого найменування $x_2= x_1\text{-}2m^*_1$=4-2*2=0. Розв'язання на другому етапі при x_2=0 приводить до оптимального розв'язку m^*_2=0, що дає $x_3= x_2\text{-}3m^*_2$=0+3*0=0. Далі етап 3 при x_3=0 приводить до m^*_3=0. Отже, оптимальним розв'язком задачі є m^*_1=2, m^*_2=0 і m^*_3=0. Відповідний прибуток становитиме 62 грошові одиниці.

Задача планування робочої сили. При виконанні певних проектів число робітників, необхідне для реалізації будь-якого проекту, регулюють шляхом їх наймання й звільнення. Наймання й звільнення робітників пов'язане з додатковими витратами. Треба визначити, як треба регулювати чисельність робітників.

Припустимо, що проект буде виконуватися протягом n тижнів і мінімальна потреба в робочій силі протягом i-го тижня складе b_i робітників. Проте залежно від вартісних показників може бути вигіднішим відхилення чисельності робітників як в одну так і в іншу сторону від мінімальної потреби. Якщо x_i - кількість робітників протягом i-го тижня, то можливі витрати двох видів: $C_1(x_i\text{-}b_i)$ - витрати пов'язані з необхідністю утримувати надлишок $x_i\text{-}b_i$ робочої сили й $C_2(x_i\text{-}x_{i-1})$ - витрати, пов'язані з необхідністю додаткового наймання $x_i\text{-}x_{i-1}$ робітників.

Елементи моделі динамічного програмування в цій задачі визначають в такий спосіб. Етап i представляють порядковим номером тижня i, $i = \overline{1,n}$. Варіантами розв'язку на i-му етапі є значення x_i - кількість робітників протягом і-го тижня. Станом на і-му етапі є x_{i-1} - кількість робітників протягом (і-1)-го тижня (етапу).

Рекурентне рівняння динамічного програмування має вигляд

$$f_i(x_{i-1}) = \min\{C_1(x_i - b_i) + C_2(x_i - x_{i-1}) + f_{i+1}(x_i)\}, \quad i = \overline{1,n}, \qquad (7.4)$$

Обчислення починають з етапу n при x_n=b_n і закінчують на етапі 1.

Задача заміни устаткування. Чим довше механізм експлуатують, тим вище витрати на його обслуговування й нижче його продуктивність. Коли термін експлуатації механізму досягає певного рівня, може виявитися вигіднішою його заміна. Задача заміни устаткування зводиться до визначення оптимального строку експлуатації механізму.

Нехай на початку кожного року приймають рішення або про експлуатацію механізму ще один рік, або про заміну його новим. Позначимо $r(t)$ і $c(t)$ прибуток від експлуатації t-літнього механізму й витрати на його обслуговування протягом року. Нехай *s(t)* - вартість продажу механізму, що експлуатувався *t* років. Вартість придбання нового механізму залишається незмінною протягом всіх *n* років і дорівнює *I*.

Елементи моделі динамічного програмування визначають в такий спосіб. Етап *i* представляють порядковим номером року *i*, $i=\overline{1,n}$. Варіантами розв'язку на і-му етапі є альтернативи: продовжити експлуатацію або замінити механізм на початку і-го року. Станом на і-му етапі є тривалість експлуатації *t* механізму до початку і-го року.

Нехай $f_i(t)$ – максимальний прибуток, одержуваний за роки від *i* до *n* за умови, що на початку і-го року є механізм t-літнього віку.

Рекурентне рівняння має такий вигляд

$$f_i(t)=\max\left\{\begin{array}{ll} r(t)-c(t)+f_{i+1}(t+1), & \text{якщо експлуатувати механізм,} \\ r(0)+s(t)-I-c(0)+f_{i+1}(1), & \text{якщо замінити механізм.} \end{array}\right\} \qquad (7.5)$$

Задача інвестування. Припустимо, що на початку кожного з наступних *n* років необхідно зробити інвестиції $P_1, P_2, \ldots, P_n$. Є можливість вкласти капітал у два банки, перший з яких виплачує річний складний відсоток r_1, а другий – r_2. Для заохочення депозитів обидва банки виплачують новим інвесторам премії у вигляді відсотка від вкладеної суми. Преміальні змінюються від року до року й для і-го року дорівнюють q_{i1} і q_{i2} у першому й другому банках відповідно. Їх виплачують наприкінці року, і вони можуть бути інвестовані в один із двох банків на наступний рік. Розміщений у банку внесок повинен перебувати там до кінця розглянутого періоду. Необхідно розробити стратегію інвестицій на наступні *n* років.

Елементи моделі динамічного програмування визначають в такий спосіб. Етап *i* представляють порядковим номером року і, $i=\overline{1,n}$. Варіантами розв'язку на і-му етапі є суми I_i^1 і I_i^2 інвестицій у перший і другий банк відповідно. Станом x_i на і-му етапі є сума грошей на початок і-го року, які можуть бути інвестовані.

Оскільки $I_i^2=x_i=I_i^1$, то

$$x_i=P_i+q_{i-1,1}I_{i-1}^1+q_{i-1,2}(x_{i-1}-I_{i-1}^1)=P_i+(q_{i-1,1}+q_{i-1,2})I_{i-1}^1+\quad+q_{i-1,2}x_{i-1}.$$

Нехай $f_i(x_i)$ – оптимальна сума інвестицій для інтервалу від і-го до *n*-го року за умови, що на початку і-го року є грошова сума x_i. Далі позначимо через s_i суму, накопичену до кінця *n*-го року за умови, що I_i^1 і $(x_i - I_i^1)$ – обсяги інвестицій протягом і-го року в перший і другий банк відповідно. Позначивши $\alpha_i=(I^1+r_i)$, i=1,2, сформулюємо задачу:

Максимізувати z=s_1,+s_2+…+s_n, де

$$s_i=I_i^1\alpha_1^{n+1-i}+(x_i-I_i^1)\alpha_2^{n+1-i}=\left(\alpha_1^{n+1-i}-\alpha_2^{n+1-i}\right)I_i^1+\alpha_2^{n+1-i}x_i,\quad i=\overline{1,n-1},$$

$$s_n=(\alpha_1+q_{n1}-\alpha_2-q_{n2})I_n^1+(\alpha_2+q_{n2})x_n.$$

Оскільки преміальні за *n*-й рік є частиною накопиченої грошової суми, у вираз для s_n додані q_{n1} і q_{n2}. В цьому випадку рекурентне рівняння для зворотного прогону має вигляд

$$f_i(x_i) = \max\{s_i + f_{i+1}(x_{i+1})\}, \quad i = \overline{1, n-1}. \tag{7.6}$$

Задача керування запасами. Однією з найвідоміших сфер використання методів динамічного програмування є така область як теорія керування запасами. Її предметом є розробка й дослідження математичних моделей систем, що займають проміжне положення між джерелами (виробниками) тих або інших ресурсів та їх споживачами. Під час математичноїй формалізації процесів керування запасами дуже часто доводиться використовувати стрибкоподібні, недиференційовані та кусочно-безперервні функції. Як правило, це зумовлено необхідністю обліку ефектів концентрації, фіксованих витрат і плати за замовлення. У зв'язку з цим одержувані задачі трудно піддаються аналітичному розв’язанню класичними методами, однак можуть бути успішно вирішені за допомогою апарата динамічного програмування. Найпростіші задачі характеризуються постійним у часі попитом, миттєвим поповненням запасу та відсутністю дефіциту. Розглянемо модель з витратами на оформлення замовлення, у якій передбачається, що дефіцит не допускається. Введемо позначення: z_i - кількість замовленої продукції, D_i - потреба в продукції (попит), x_i - обсяг запасу на початок етапу і, K_i - витрати на оформлення замовлення, h_i - витрати на зберігання одиниці продукції, що переходить з етапу i в етап $i+1$.

Відповідну функцію витрат для етапу і задають формулою

$$C_i(z_i) = \begin{cases} 0, & z_i = 0 \\ K_i + c_i(z_i), & z_i > 0 \end{cases},$$

де $c_i(z_i)$ - функція граничних витрат при заданому значенні z_i.

Оскільки дефіцит не допускається, задача керування запасами зводиться до обчислення значень z_i, що мінімізують сумарні витрати, пов'язані з розміщенням замовлень, закупівлею та зберіганням продукції протягом n етапів. Витрати на зберігання на і-му етапі передбачаються пропорційними величині

$$x_{i+1}=x_i+z_i-D_i,$$

яка є обсягом запасу, що переходить з етапу i в етап $i+1$.

Для рекурентного рівняння процедури прямого прогону стан на етапі і визначають як обсяг запасу x_{i+1} на кінець етапу

$$0 \le x_{i+1} \le D_{i+1}+\ldots+D_n.$$

Ця нерівність означає, що в граничному випадку запас x_{i+1} може задовольнити попит на всіх наступних етапах.

Нехай $f_i(x_{i+1})$ - мінімальні загальні витрати на етапах 1, 2, … і при заданій величині запасу x_{i+1} на кінець етапу і. Тоді рекурентне рівняння алгоритму прямого прогону можна записати у вигляді

$$f_1(x_2) = \min\{C_1(z_1) + h_1 x_2\},$$

$$f_i(x_{i+1}) = \min\{C_i(z_i) + h_i x_{i+1} + f_{i-1}(x_{i+1} + D_i - z_i)\}, \quad i = \overline{2, n}. \tag{7.7}$$

Контрольні запитання

1. Для розв'язання яких задач призначений метод динамічного програмування?
2. Поясніть, що являють собою адитивна й мультиплікативна функції.
3. У чому полягає сутність методу динамічного програмування?
4. Яким умовам має задовольняти задача, щоб для її розв'язання міг бути застосований алгоритм динамічного програмування?
5. Сформулюйте принцип оптимальності Беллмана, поясніть відсутність післядії.
6. Як визначають напрям розв'язання задачі в алгоритмах динамічного програмування?
7. Сформулюйте математичну модель для задачі про планування робочої сили.
8. Випишіть основне рекурентне співвідношення, використовуване під час розв'язання задачі про планування робочої сили.
9. Сформулюйте математичну модель для задачі про заміну устаткування.
10. Випишіть основне рекурентне співвідношення, використовуване під час розв'язання задачі про заміну устаткування.
11. З якими особливостями задач керування запасами пов'язане застосування під час їх розв'язання апарату динамічного програмування?
12. Який вигляд має цільова функція в динамічній задачі керування запасами?
13. Випишіть основне рекурентне співвідношення, використовуване під час розв'язання динамічної задачі керування запасами.

ТЕМА 8
ЕЛЕМЕНТИ ТЕОРІЇ ІГОР

8.1. Основні поняття теорії ігор

Під час вивчення процесів прийняття рішень кількома суб'єктами, інтереси яких можуть не збігатися, виникають задачі з багатьма цільовими функціями (критеріями). Область математики, що вивчає ці проблеми, одержала назву ***теорії ігор***. Задачі теорії ігор належать до області прийняття рішень в умовах невизначеності, а їх специфіка полягає в тому, що, як правило, мають на увазі невизначеність, яка виникає в результаті дій двох або більше «розумних» супротивників, здатних оптимізувати свою поведінку за рахунок інших. Такі ситуації є типовими в практичній діяльності менеджерів, маркетологів та інших фахівців, що приймають рішення в умовах гострої конкуренції, неповноти інформації та ін.

Одним з основних питань у задачах з колективним вибором рішень є питання про визначення оптимальності, тобто питання, які рішення можна визнавати найкращими в ситуації оптимізації за кількома критеріями, що відбиває різні інтереси. Багато методів вирішення проблем теорії ігор ґрунтуються на зведенні їх до задач математичного програмування. Теорія ігор бере початок від робіт Е. Бореля (1921 р.), а принциповим етапом у її становленні як самостійного наукового напряму стала монографія Дж. Неймана, що вийшла у 1944 р.

Термін ***гра*** застосовують для позначення сукупності правил і угод, якими керуються суб'єкти, поведінку яких ми вивчаємо. Кожний такий суб'єкт k, де $k=\overline{1,K}$, або ***гравець***, характеризується наявністю індивідуальної системи цільових настанов і ***стратегій*** $s_1^k, s_2^k, \ldots, s_{m_k}^k$, тобто можливих варіантів дій у грі.

Розповсюджений спосіб математичного опису гри ґрунтується на завданні функцій $f_k(s_{i1}^1, s_{i2}^2, \ldots, s_{ik}^k, \ldots, s_{iK}^K)$, кожна з яких визначає результат (платіж, виграш), одержуваний k-м гравцем залежно від набору стратегій $S = (s_{i1}^{1}, s_{i2}^{2}, ..., s_{ik}^{k}, ..., s_{iK}^{K})$, застосованого всіма учасниками гри. Функції f_k також називають ***функціями виграшу***, або ***платіжними функціями***. У тому випадку, якщо для будь-яких S

$$\sum_{k=1}^{K} f_k(s) = 0,$$

гру називають ***грою з нульовою сумою***. Гру із двома учасниками та нульовою сумою називають ***антагоністичною***. Антагоністичні ігри, тобто ігри, в яких виграш одного учасника дорівнює програшу іншого, у зв'язку з відносно простою постановкою задачі є найбільш вивченим розділом теорії ігор. Проте зміст теорії ігор, безумовно, не вичерпується ними. У класифікації ігрових моделей виділяють ігри з кінцевими та нескінченними наборами стратегій у гравців, виділяють ігри за можливими кількостями ходів учасників. Так само ігри поділяють на некооперативні та кооперативні, тобто ті, в яких функції виграшу учасників залежать від утворених ними коаліцій. Окрім цього ігри можна розрізняти за обсягом інформації, наявної в гравців щодо минулих ходів. В цьому зв'язку їх поділяють на ігри з повною та неповною інформацією.

8.2. Матричні ігри двох осіб з нульовою сумою

Розглянемо докладніше антагоністичні ігри та їх основні властивості. Зручним способом завдання гри двох учасників з нульовою сумою є ***платіжна матриця***. Звідси, до речі, випливає ще одна їх назва — ***матричні ігри***. Кожний елемент платіжної матриці a_{ij} містить числове значення виграшу гравця *I* (програшу гравця *II*), якщо перший застосовує стратегію *i*, а другий - стратегію *j*. Терміни виграш та програш треба розуміти в широкому змісті, оскільки вони можуть приймати від'ємні значення й означати протилежне. Нетривіальність задачі, насамперед, полягає в тому, що кожний з гравців робить свій вибір, не знаючи про вибір іншого, що істотно ускладнює процес оптимізації обираної стратегії.

Класичним прикладом антагоністичної гри є гра з двома учасниками, що загадують незалежно один від одного числа. Передбачають, що якщо їх сума опиниться парною, то виграш, рівний 1, дістається першому гравцеві, а якщо непарною, то другому. Поклавши, що для обох гравців загадування непарного числа є першою стратегією, а парного - другою, можемо записати платіжну матрицю цієї гри:

	н/п	**п**
н/п	1	-1
п	-1	1

(8.1)

Рядки матриці (8.1) відповідають стратегіям гравця I, стовпці - стратегіям гравця II, а її елементи - результатам першого гравця. Також з визначення гри випливає, що елементи цієї матриці, узяті із зворотним знаком, відповідають виграшам другого гравця.

Складнішу і змістовнішу платіжну матрицю можна отримати, якщо трохи модифікувати запропоновану гру. Припустимо, що обоє учасники мають право загадувати числа від 1 до 4, що становить їх відповідні стратегії. У випадку якщо результат додавання задуманих чисел буде парним, то другий гравець виплачує першому суму, що вийшла, а якщо непарним, то перший - другому. Запишемо платіжну матрицю для такої гри:

$$A=\begin{array}{|cccc|c} 2 & -3 & 4 & -5 & 1 \\ -3 & 4 & -5 & 6 & 2 \\ 4 & -5 & 6 & -7 & 3 \\ -5 & 6 & -7 & 8 & 4 \\ \multicolumn{1}{c}{1} & 2 & 3 & \multicolumn{1}{c}{4} & \end{array} \qquad (8.2)$$

Як відзначено вище, найважливішим у теорії ігор є питання про оптимальність розв'язку (вибору стратегії) для кожного з гравців. Проаналізуємо з цього погляду певну матричну гру, для якої задано платіжну матрицю $A=\left\|a_{ij}\right\|_{m\times n}$. Під час вибору гравцем *I* стратегії *i* його гарантований дохід незалежно від дій гравця *II* складе $\min\limits_{j} a_{ij}$. Оскільки він може обирати *i* самостійно, доцільно цей вибір зробити таким, щоб він при будь-якій стратегії супротивника максимізував величину гарантованого доходу, тобто забезпечував одержання $\max\limits_{i}\left(\min\limits_{j} a_{ij}\right)$. Такий принцип вибору стратегії одержав назву ***принцип максиміна***, а розмір гарантованого виграшу – ***нижньої ціни гри***. З іншого боку, аналогічні міркування можна провести з приводу дій другого гравця. Його найбільший програш під час вибору стратегії *j* складе $\max\limits_{i} a_{ij}$, і, отже, йому треба обирати стратегію так, щоб мінімізувати величину програшу при будь-яких діях суперника, тобто забезпечити $\min\limits_{j}\left(\max\limits_{i} a_{ij}\right)$. у цьому полягає суть ***принципу мінімакса***, розмір програшу називають ***верхньою ціною гри***.

Можна довести справедливість наступного співвідношення:

$$\max_{i}\min_{j} a_{ij} \le \min_{j}\max_{i} a_{ij}. \qquad (8.3)$$

Проте очевидний інтерес представляє ситуація, при якій значення виграшу (платежу), одержуваного гравцем *I* під час вибору ним максимінної стратегії, дорівнює платежу (програшу) *П*-го гравця під час мінімаксній стратегії

$$\max_{i}\min_{j} a_{ij} = \min_{j}\max_{i} a_{ij}. \qquad (8.4)$$

В цьому випадку говорять, що гра має ***сідлову точку***. Збіг значень гарантованих виграшів гравців при максимінній та мінімаксній стратегії означає можливість досягнення в грі певного оптимального (стабільного, рівноважного) стану, від якого невигідно відхилятися жодному з учасників. Поняття «оптимальність» тут означає, що жоден розумний (обережний) гравець не прагне змінити свою стратегію, оскільки його супротивник, зможе обрати таку стратегію, що дасть гірший для першого результат. Стратегії i^* та j^*, що утворюють сідлову точку, називають ***оптимальними***, а значення $v = a_{i^* j^*}$ називають ***ціною гри***. Трійку (i*,j*,v) вважають ***розв'язком матричної гри із сідловою точкою***.

Неважко помітити, що не усяка гра має сідлову точку. Зокрема, як гра (8.1), так і гра (8.2) сідлової точки не мають. Прикладом гри, що має сідлову точку, є гра з платіжною матрицею (8.5).

$$A= \begin{vmatrix} 7 & 10 & 4 & 1 \\ 6 & 8 & 5^* & 17 \\ 8 & -3 & 2 & 11 \end{vmatrix} \qquad (8.5)$$

У цій матриці мінімальні (гарантовані) виграші першого гравця за рядками дорівнюють 1, 5 і (-3). Отже, його максимінному вибору відповідатиме стратегія 2, що гарантує виграш 5. Для другого гравця максимальні програші за стовпцями матриці складуть 8, 10, 5, 17, тому має сенс зупинитися на стратегії 3, при якої він програє тільки 5. Отже, друга стратегія першого гравця й третя стратегія другого утворюють сідлову точку із значенням 5, тобто гра з матрицею (10.5) має розв'язання (2; 3; 5).

8.3. Чисті та змішані стратегії

Подальший розвиток теорії матричних ігор ґрунтується на дослідженні гри як певного повторюваного процесу. Дійсно, навряд чи можна дати змістовні рекомендації з такого питання, як треба поводитися учасникам однократно проведеної гри, що не має сідлової точки. У випадку ж її багаторазових повторів природною уявляється ідея *рандомізації* вибору стратегій гравцями, тобто внесення до процесу вибору елемента випадковості. Дійсно, систематичне відхилення, наприклад, гравця *I* від максимінної стратегії з метою збільшення виграшу може бути зафіксованим другим гравцем і покарано. В той самий час абсолютно хаотичний вибір стратегій не принесе в середньому найкращого результату.

Змішаною стратегією гравця *I* у грі з матрицею $A = \left\| a_{ij} \right\|_{m \times n}$ називають впорядкований набір дійсних чисел x_i, $i = \overline{1,m}$, що задовольняють умовам

$$x_i \ge 0, \quad i = \overline{1,m}; \quad \sum_{i=1}^{m} x_i = 1. \qquad (8.6)$$

Числа x_i інтерпретують як імовірності застосування гравцем *I* стратегій 1,2,…,*m*, які, на відміну від змішаних, називають ***чистими стратегіями***.

Аналогічно вводять поняття змішаних стратегій гравця *II*, які визначають як набір чисел y_j, $j = \overline{1,n}$, що задовольняють умовам

$$y_j \ge 0, \quad j = \overline{1,n}; \quad \sum_{j=1}^{n} y_j = 1. \qquad (8.7)$$

Тоді, якщо гравець *I* застосовує змішану стратегію $x=(x_1, x_2,..., x_m)$, а гравець *II* змішану стратегію $y = (y_1, y_2, \dots, y_n)$, математичне сподівання виграшу гравця I (програшу гравця II) визначають співвідношенням

$$F(x,y)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\ . \qquad (8.8)$$

Надалі через X позначатимо множину припустимих змішаних стратегій гравця I, зумовлену умовою (8.6), а через Y - зумовлену умовою (8.7) множину припустимих змішаних стратегій гравця II.

До пошуку розв'язку гри в змішаних стратегіях можна застосовувати критерії максиміна-мінімакса. Відповідно до них гравець I обиратиме свою змішану стратегію $x=(x_1, x_2,..., x_m)$ так, щоб максимізувати найменший середній виграш:

$$\max_{x\in X}\left[\min_{y\in Y}\ \sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right], \qquad (8.9)$$

який дорівнює

$$\max_{x\in X}\left[\min\left\{\sum_{i=1}^{m}a_{i1}x_i,\sum_{i=1}^{m}a_{i2}x_i,...,\sum_{i=1}^{m}a_{in}x_i\right\}\right], \qquad (8.10)$$

а гравець II - свою змішану стратегію так, щоб мінімізувати найбільший середній програш:

$$\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right], \qquad (8.11)$$

також рівний

$$\min_{y\in Y}\left[\max\left\{\sum_{j=1}^{n}a_{1j}y_j,\sum_{j=1}^{n}a_{2j}y_j,...,\sum_{j=1}^{n}a_{mj}y_j\right\}\right]. \qquad (8.12)$$

За аналогією з (10.3) для будь-яких $x\in X$ та $y\in Y$ справедлива нерівність

$$\max_{x\in X}\left[\min_{y\in Y}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right]\le\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right] \qquad (8.13)$$

Стратегії $x^*\in X$ та $y^*\in Y$ називають ***оптимальними змішаними стратегіями***, якщо для будь-яких $x\in X$ та $y\in Y$ справедлива рівність

$$F(x^*,y^*)=\max_{x\in X}\left[\min_{y\in Y}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right]=\min_{y\in Y}\left[\max_{x\in X}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}x_i y_j\right]. \qquad (8.14)$$

$v = F(x^*,y^*)$ називають ціною гри, і якщо x^* та y^* існують, говорять, що ***гра має розв'язання у змішаних стратегіях (x*, в*, v)***.

Справедлива фундаментальна теорема Дж. Неймана, (приведемо без доказу).

Теорема 8.1 (основна теорема матричних ігор). Будь-яка матрична гра має розв'язання в змішаних стратегіях.

Значення й нетривіальність теореми (8.1) зумовлені насамперед тим, що у загальному випадку матричні ігри в чистих стратегіях розв'язання не мають.

8.4. Зведення гри двох гравців до задачі лінійного програмування

Задачу, розв'язувану першим гравцем, (8.10) було сформульовано як максимізацію найменшої з сум

$$\sum_{i=1}^{m} a_{ij} x_i \ ,$$

але якщо визначити деяке x_{m+1}, для якого виконується

$$x_{m+1} \le \sum_{i=1}^{m} a_{ij} x_i \ , \tag{8.15}$$

її можна звести до задачі лінійного програмування:

знайти $$F(x) > max \tag{8.16}$$

при обмеженнях

$$\sum_{i=1}^{m} a_{ij} x_i \ge x_{m+1}, \quad j = \overline{1,n}; \quad \sum_{i=1}^{m} x_i = 1; \quad x_i \ge 0; \quad i = \overline{1,m} \, . \tag{8.17}$$

Провівши аналогічні міркування, приходимо до того, що задача мінімізації найбільшого сподіваного програшу, розв'язувана гравцем *II* (8.12), збігається до задачі лінійного програмування

$$F(x) > min, \tag{8.18}$$

$$\sum_{j=1}^{n} a_{ij} y_j \le y_{n+1}, \quad i = \overline{1,m}; \quad \sum_{j=1}^{n} y_j = 1; \quad y_j \ge 0; \quad j = \overline{1,n} \tag{8.19}$$

Отже, ми одержуємо можливість застосовувати всі можливості апарата лінійного програмування для пошуку оптимальних стратегій обох гравців.

Досить легко перевірити, що задачі (8.16)-(8.17) та (8.18)-(8.19) утворюють двоїсту пару. Тут у певному змісті ми повернулися до взаємозв'язку між наявністю розв’язку в деякої оптимізаційної задачі та існуванням сідлової точки у відповідної функції Лагранжа. В цьому випадку аналогічний зв'язок простежується між сідловою точкою гри й розв’язком пари задач оптимізації.

Графічний метод вирішення гри. Треба зазначити, що застосування для розв’язання задач (8.16)-(8.17), (8.18)-(8.19) стандартних алгоритмів лінійного програмування далеко не завжди є раціональним. Окрім цього існують інші методи, які ґрунтуються на використанні специфіки цих задач. Зупинимося на дуже простому класичному способі пошуку оптимальних змішаних стратегій у матричних іграх, де один з учасників має тільки дві стратегії (це так звані $2{\times}n$ та $m{\times}2$ ігри).

Для визначеності покладемо, що гравець *I* має можливість вибирати між двома стратегіями з імовірностями x_1 та x_2=l-x_1, тоді його очікувані виграші, що відповідають чистим стратегіям гравця *II*, приймуть вигляд

$$a_{11}x_1 + a_{21}(1-x_1),\ a_{12}x_1 + a_{23}(1-x_1), ..., a_{1n}x_1 + a_{2n}(1-x_1)$$

або

$$(a_{11} - a_{21})x_1 + a_{21},\ (a_{12} - a_{22})x_1 + a_{22}, ..., (a_{1n} - a_{2n})x_1 + a_{2n}\,,$$

тобто очікувані виграші можна представити у вигляді графіків лінійних функцій, що залежать від змінної x_1 (рис. 8.1, де передбачається, що гравець *II* має три стратегії).

Лінії, зображені на рис. 8.1, задають залежності середнього виграшу гравця *I* від значення імовірності x_1, з якою він обирає свою першу стратегію, для випадків, коли його супротивник обирає першу, другу або третю чисту стратегію. Тоді значенням мінімального гарантованого доходу першого гравця відповідає нижня огинаюча всіх трьох прямих. Відповідно до принципу максиміна, оптимальному вибору гравця *I* відповідатиме найвища точка, що лежить на даній огинаючій, позначена на рисунку як (x_1^*, f^*). Знаючи її, можна визначити оптимальну змішану стратегію першого гравця $x^*=(x_1^*, 1-x_2^*)$ та ціну гри, що дорівнює f^*.

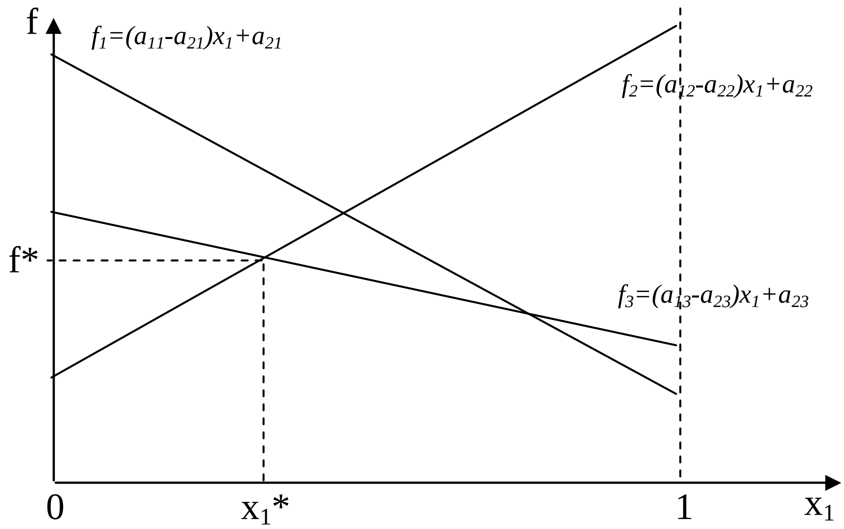

Рис. 8.1 - Графічне вирішення гри

Виходячи з відношення подвійності, яким пов'язані задачі обох гравців, за оптимальною стратегією першого учасника x^* однозначно визначається оптимальна стратегія його супротивника y^*. Оскільки y є результатом розв'язання задачі лінійного програмування, він має всі властивості припустимого базисного плану, тобто у випадку $2\times n$ гри має не більш за два ненульових компоненти й не менш за (n-2) нульових. Номера ненульових елементів y^* визначають номерами ліній, перетинання яких визначило оптимальну стратегію першого гравця. Дійсно, гравець *II* знає оптимальну стратегію суперника, і застосування ним стратегій, що відповідають прямим, які проходять вище точки (x_1^*, f^*), тільки збільшить його програш. У розглянутому прикладі це лінії f_2 та f_3, і, отже, у своїй оптимальній стратегії другий гравець повинен з ненульовими імовірностями застосовувати другу і третю чисті стратегії ($y_2 >0$, $y_3 >0$). На основі цього, а також з огляду на умову нормування

$$\sum_{j=1}^{n} y_j = 1,$$

можна виразити: $y_3=1-y_2$, тоді оптимальне значення y^*_2 можна знайти з умови

$$a_{11}\times 0 + a_{21}y_2^* + a_{31}(1-y_2^*) = a_{12}\times 0 + a_{22}y_2^* + a_{32}(1-y_2^*)$$

або

$$(a_{21}-a_{31})y_2^* + a_{31} = (a_{22}-a_{32})y_{21}^* + a_{32}.$$

В результаті дістанемо оптимальну стратегію гравця *II* $y^*=(0,y^*_2,y^*_3)$.

Очевидно, що пошук розв'язання в грі $m\times 2$ здійснюють аналогічно: будують графіки очікуваного програшу гравця *II*, знаходять їх верхню огинаючу й так далі.

Безумовно, графічний спосіб в силу обмеженості кола задач, до яких його можна застосовувати, має скоріше теоретичне, чим практичне значення. Проте він добре ілюструє змістовну сторону процесу пошуку розв'язку в грі.

8.5. Статистичні ігри (ігри з природою)

Апарат теорії ігор розрізняє стратегічні й ***статистичні ігри***. В основі стратегічних ігор лежить припущення, що кожен з гравців діє активно, та їх інтереси є протилежними. У теорії статистичних ігор одним з гравців є природа, тобто сукупність зовнішніх обставин, в умовах яких доводиться приймати рішення. Неминучою платою за спробу одержати рішення в умовах неповної інформації про закони природи є прийняття помилкових рішень. Теорія статистичних ігор дозволяє виробити таку стратегію щодо прийняття рішень, що хоча й не виключає можливості прийняття невірних рішень, але зводить до мінімуму пов'язані з цим небажані наслідки.

Розглянемо приклад. Нехай необхідно зробити запас n товарів у кількості $X=(x_1, x_2, \ldots, x_n)$, на які є випадковий попит $\omega=(\omega_1, \omega_2, \ldots, \omega_n)$... Нестачу одиниці j-*го* товару обкладають штрафом c_j, тобто $C=(c_1, c_2, \ldots, c_n)$, а витрати на зберігання одиниці відповідного товару, що не вдалося збути, задані вектором $d=(d_1, d_2, \ldots, d_n)$. Функція збитку, що відповідає розв'язку x, має вигляд

$$f(x,\omega)=\sum_{j=1}^{n}\left\{Cj\max\left(0,\omega_j-x_j\right)+d_j\max\left(0,x_j-\omega_j\right)\right\},$$

де $Cj\max\left(0,\omega_j-x_j\right)$ - штраф за незадоволення попиту на j-*й* вид товару; $d_j\max\left(0,x_j-\omega_j\right)$ - витрати на зберігання j-*го* виду товару. Для знаходження оптимального розв'язку цієї задачі необхідно знати функцію розподілу випадкової величини ω. Оскільки така функція невідома, вважають, що ω розподілена рівномірно. При цьому необхідно враховувати, що це припущення може призвести до ухвалення неправильного рішення.

Відмінність між прийняттям рішень в умовах ризику й невизначеності полягає в тому, що в умовах невизначеності імовірнісний розподіл, що відповідає станам ω_j невідомий. Ця обставина зумовила розвиток спеціальних критеріїв для аналізу ситуації, пов'язаної з прийняттям рішень: критерію Лапласа, мінімаксного критерію, критерію Севіджа та критерію Гурвіца. Ці критерії відрізняються за ступенем консерватизму особи, що приймає рішення.

Критерій Лапласа спирається на принцип недостатньої підстави, який говорить, що оскільки розподіл імовірностей станів $P\{\omega_j\}$ невідомий, немає підстав вважати їх різними. Отже, роблять припущення, що імовірності всіх станів природи дорівнюють одна одній, тобто $P\{\omega_1\}= P\{\omega_2\}=\ldots=P\{\omega_n\}=1/n$. Якщо при цьому $v(a_i, \omega_j)$ є прибутком, то найкращим рішенням є те, що забезпечує

$$\max_{ai}\left\{\frac{1}{n}\sum_{j=1}^{n} v(a_i, \omega_j)\right\}.$$

Якщо величина $v(a_i, \omega_j)$ є видатками, то максимум замінюють на мінімум.

Максимінний (мінімаксний) критерій заснований на консервативній обережній поведінці особи, що приймає рішення, і збігається до вибору найкращої альтернативи з найгірших. Якщо величина $v(a_i, \omega_j)$ є прибутком, то відповідно до максимінного критерію за оптимальне обирають рішення, що забезпечує

$$\max_{ai}\left\{\min_{\omega j} v(a_i, \omega_j)\right\}.$$

Якщо величина $v(a_i, \omega_j)$ є втратами, використовують мінімаксний критерій

$$\min_{ai}\left\{\max_{\omega j} v(a_i, \omega_j)\right\}.$$

Критерій Севіджа знижує консерватизм мінімаксного (максимінного) критерію шляхом заміни матриці платежів $v(a_i, \omega_j)$ матрицею втрат $r(a_i, \omega_j)$, що визначають в такий спосіб

$$r(a_i, \omega_j) = \begin{cases} \max\limits_{ak}\{v(a_k, \omega_j)\} - v(a_i, \omega_j), & \text{якщо } v - \text{доход}, \\ v(a_ш, \omega_j) - \min\limits_{ak}\{v(a_k, \omega_j)\}, & \text{якщо } v - \text{втрати}. \end{cases}$$

Розглянемо приклад. Є матриця платежів

	ω_1	ω_2
a_1	11000	90
a_2	10000	10000

Максимум рядка 1 становить 11000, а рядка 2 – 10000 (мінімакс). Застосування мінімаксного критерію призводить до того, що рішення a_2 з фіксованими втратами 10000 є кращим. Проте можна обрати й a_1, оскільки в цьому випадку існує можливість втратити лише 90, якщо реалізується стан ω_2 при потенційному виграші 11000.

Визначимо, який результат вийде, якщо в мінімаксному критерії замість матриці платежів $v(a_i, \omega_j)$ використати матрицю втрат $r(a_i, \omega_j)$.

	ω_1	ω_2
a_1	1000	0
a_2	0	9910

Максимум рядка 1 становить 1000 (мінімакс), а рядка 2 – 9910. Як бачимо, мінімаксний критерій, застосований до матриці втрат, призводить до вибору рішення a_1.

Критерій Гурвіца охоплює низку різних підходів до прийняття рішень – від найоптимістичнішого до найпесимістичнішого. Нехай $0 \leq \alpha \leq 1$ і величини $v(a_i, \omega_j)$ є доходами. Тоді рішенню, обраному за критерієм Гурвіца, відповідає

$$\max_{a_i}\left\{\alpha \max_{\omega_j} v(a_i,\omega_j) + (1-\alpha)\min_{\omega_j} v(a_i,\omega_j)\right\}.$$

Параметр α - показник оптимізму. Якщо α=0, критерій Гурвіца стає еквівалентним застосуванню мінімаксного критерію. Якщо α=1, критерій Гурвіца стає занадто оптимістичним, тому що розраховує на найкраще з найкращих рішень. Можна конкретизувати ступінь оптимізму вибором величини α з інтервалу [0, 1]. Найрозумнішим уявляється вибір α=0,5.

Якщо величини $v(a_i, \omega_j)$ є втратами, то критерій приймає наступний вигляд

$$\min_{a_i}\left\{\alpha \min_{\omega_j} v(a_i,\omega_j) + (1-\alpha)\max_{\omega_j} v(a_i,\omega_j)\right\}.$$

Контрольні запитання

1. Коротко сформулюйте предмет теорії ігор як наукової дисципліни.
2. Який зміст вкладають в поняття «гра»?
3. Для опису яких економічних ситуацій можна застосовувати апарат теорії ігор?
4. Яку гру називають антагоністичною?
5. Чим однозначно визначаються матричні ігри?
6. В чому полягають принципи максиміна та мінімакса?
7. За яких умов можна говорити про те, що гра має сідлову точку?
8. Наведіть приклади ігор, які мають сідлову точку й у яких вона відсутня.
9. Які підходи існують до визначення оптимальних стратегій?
10. Що називають «ціною гри»?
11. Дайте визначення поняттю «змішана стратегія».
12. Сформулюйте основну теорему матричних ігор.
13. Поясніть розходження між стратегічною й статистичною іграми.

ТЕМА 9
БАГАТОКРИТЕРІАЛЬНА ОПТИМІЗАЦІЯ

9.1. Постановка проблеми багатокритеріальної оптимізації

Під дослідженням операцій розуміють застосування математичних, кількісних методів для обґрунтування рішень у всіх галузях цілеспрямованої людської діяльності. Основними етапами рішення будь-якої задачі в дослідженні операцій є:

1) побудова моделі;
2) вибір критерію оптимальності;
3) знаходження оптимального рішення.

Для підходу дослідження операцій характерні наступні риси.

1. Використовувані моделі носять об'єктивний характер. Побудову моделей розглядають в рамках дослідження операцій як засіб відбиття об'єктивно існуючої реальності. Коли модель, що правильно відбиває дійсність, знайдено, критерій оптимальності встановлений, оптимальне рішення може бути отриманим ***єдино можливим*** чином. Інакше кажучи, спираючись на ті самі дані, різні фахівці-аналітики мають одержувати однакові результати. Ця вимога означає, що діяльність людей, описувана моделлю, підлегла вимогам ***доцільності***.

2. Керівник одержує науково обґрунтоване рішення. За замовленням keрі-

вника аналітик досліджує організацію, зовнішнє середовище й намагається побудувати адекватну модель. У цій роботі сама ОПР найчастіше не потрібна. Можна сказати, що керівник дає замовлення й одержує готове рішення. Все інше роблять аналітики-фахівці з дослідження операцій.

3. Існує об'єктивний критерій успіхів у застосуванні методів дослідження операцій. Якщо проблема, що вимагає рішення, ясна і критерій визначений, то аналітичний метод відразу показує, наскільки нове рішення краще старого. Оптимальне рішення проблеми безглуздо заперечувати.

Аналіз багатьох реальних практичних проблем, з якими зіштовхувалися фахівці з дослідження операцій, природно призвів до появи класу ***багатокритеріальних завдань***. У загальному випадку постановки завдань багатокритерійної оптимізації є коректнішими за постановки завдань математичного програмування. Це пов'язано з тим, що будь-яка операція є сукупністю цілеспрямованих дій і проведення практично будь-якої операції, як правило, припускає досягнення не однієї, а кількох цілей.

Для багатокритеріальних задач характерним є те, що модель, яка описує множину припустимих рішень, є об'єктивною, але якість рішення оцінюється за багатьма критеріями. Для вибору найкращого варіанту рішення необхідний компроміс між оцінками за різними критеріями. В умовах задачі відсутня інформація, що дозволяє знайти такий компроміс. Отже, його не можна визначити на основі об'єктивних розрахунків.

Під час появи багатьох критеріїв задачі вибору найкращого рішення мають наступні особливості. Задача має унікальний характер - немає статистичних даних, що дозволяють обґрунтувати співвідношення між різними критеріями. На момент ухвалення рішення принципово відсутня інформація, що дозволяє об'єктивно оцінити можливі наслідки вибору того або іншого варіанту рішення. Але оскільки рішення так чи інакше треба прийняти, то нестачу інформації необхідно заповнити. Це може бути зроблено лише людьми на основі їх досвіду й інтуїції.

У загальному випадку математичне формулювання задачі багатокритеріальної оптимізації з множиною припустимих розв'язків D і векторною цільовою функцією

$$f(X)=(f_1(X),\ldots,f_m(X))^T$$

можна записати так:

$$f_k(X) \to \underset{X \in D}{extr}, \quad k=\overline{1,m}.$$

Будь-який скалярний критерій вигляду

$$f_j(X) \to \max_{X \in D}$$

можна замінити еквівалентним скалярним критерієм

$$-f_j(X) \to \min_{X \in D}.$$

Отже задачу багатокритеріальної оптимізації можна сформулювати таким чином:

$$f_k(X) \to \min_{X \in D}, \quad k=\overline{1,m}, \tag{9.1}$$

або у векторній формі:

$$f(X) \to \min_{X \in D}. \qquad (9.2)$$

Задачу дослідження операцій називають некоректною, якщо вона не має розв'язку. На початковому етапі розвитку дослідження операцій некоректними вважали задачі багатокритеріальної оптимізації, а як обґрунтування приводилися наступні міркування.

Нехай для кожного $k = \overline{1, m}$ елемент X_k множини D є розв'язком задачі математичного програмування

$$f(X) \to \min_{X \in D}.$$

В цьому випадку, згідно (1.1), якщо

$$X_k \equiv X_0, \quad k = \overline{1, m},$$

то X_0 - розв'язок задачі багатокритеріальної оптимізації (9.2). Але, як правило $X_k \neq X_j, \quad k \neq j$. Тому в загальному випадку треба очікувати, що задача багатокритеріальної оптимізації (9.2) не має розв'язку.

Приклад 9.1. Розглянемо просту задачу багатокритеріальної оптимізації з множиною припустимих розв'язків

$$D = \left\{ (x_1 x_2)^T : (x_1 - 2)^2 + (x_2 - 2)^2 \leq 1 \right\}$$

та критеріями оптимальності

$$f_k(X) = x_k \to \min_{X \in D}, \quad k = \overline{1,2}.$$

З геометричних міркувань видно (рис. 9.1), що

$$X_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

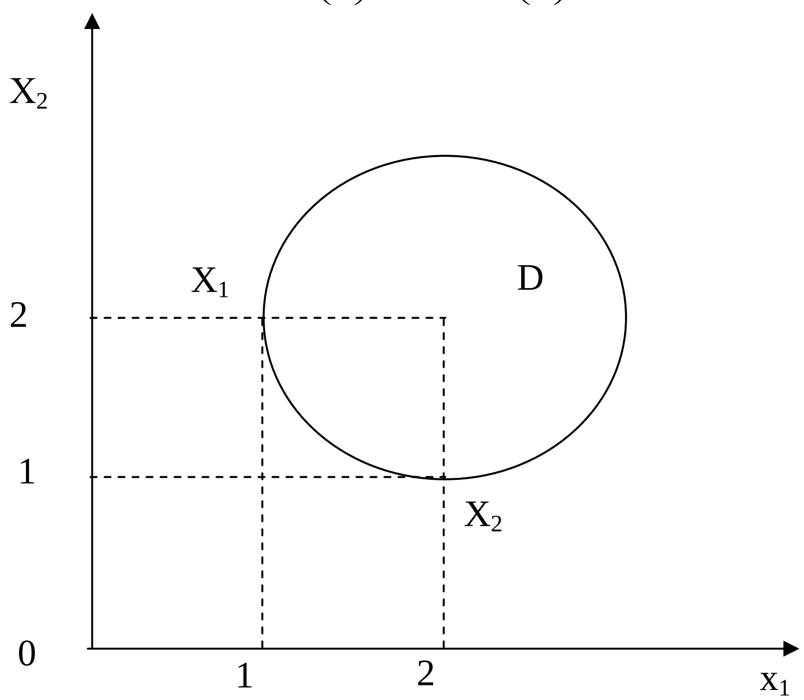

Рис. 9.1 – Ілюстрація прикладу 9.1

Оскільки $X_1 \neq X_2$, початкова задача не має розв’язання.

Отже, одночасне досягнення мінімуму за всіма скалярними критеріями $f_k(X)$ на одному розв’язку X в загальному випадку неможливе. Вихід полягає в пошуку певного компромісу в досягненні локальної мети. ОПР має сформулювати певний принцип компромісу і дотримуватися його під час вибору оптимального розв’язку. Принцип компромісу має визначити властивості оптимального розв’язку і дати відповідь на головне запитання: у якому смислі оптимальний розв’язок кращий за всі інші розв’язки? Окрім того, цей оптимальний розв’язок має належати множині припустимих розв’язків задачі. Кількість можливих принципів компромісу дуже велика. Тому під час розв’язання багатокритеріальних задач виникає низка проблем, що мають не обчислювальний, а концептуальний характер.

На перший погляд простим виходом з ситуації, що склалася, є зведення некоректної задачі багатокритеріальної оптимізації (9.2) до відповідної задачі математичного програмування шляхом виділення з множини скалярних цільових функцій однієї основної і використання інших для формування додаткових обмежень, що накладають на множину D припустимих розв’язків (тобто шляхом виділення з множини скалярних критеріїв одного основного критерію оптимальності і переведення решти критеріїв в розряд обмежень).

Трудомісткість та проблематичність коректної реалізації розглянутого підходу в загальному випадку пов'язана як з труднощами вибору однієї основної цільової функції $f_i(X)$ з множини скалярних цільових функцій, так і з обґрунтованим призначенням верхніх меж для скалярних критеріїв, що переводяться в обмеження.

9.2. Основні множини ефективних рішень (альтернатив): повна множина альтернатив, множина Парето

Розглянемо один з підходів до розв’язання задач багатокритеріальної оптимізації. Нехай припустимий розв’язок X_1 задачі багатокритеріальної оптимізації (9.2) є переважнішим за припустимий розв’язок X_2, тобто $X_1 \succ X_2$, якщо $f(X_1) < f(X_2)$. Остання нерівність означає, що $f_k(X_1) \leq f_k(X_2)$, $k = \overline{1, m}$, причому серед цих m нерівностей є хоч би одна строга. Припустимі розв’язки X_1, X_2 задачі багатокритеріальної оптимізації (9.2) називатимемо еквівалентними розв’язками $X_1 \sim X_2$, якщо $f(X_1) = f(X_2)$. Відзначимо, що:

а) у загальному випадку з $X_1 \sim X_2$, не випливає $X_1 = X_2$;

б) якщо припустимий розв’язок X_1 є строго переважнішим за припустимий розв’язок X_2, а припустимий розв’язок X_2 є строго переважнішим за припустимий розв’язок X_3, то припустимий розв’язок X_1 є строго переважнішим за, припустимий розв’язок X_3, тобто

$$\left((X_1 \succ X_2) \wedge (X_2 \succ X_3)\right) \Rightarrow (X_1 \succ X_3);$$

в) якщо X_1 та X_2 - еквівалентні розв'язки, X_2 та X_3 – так само еквівалентні розв'язки, то X_1 та X_3 будуть еквівалентними розв'язками, тобто

$$\left((X_1 \sim X_2) \wedge (X_2 \sim X_3)\right) \Rightarrow (X_1 \sim X_3) .$$

Повна множина альтернатив Ω = f(D) - множина можливих значень векторної цільової функції f у задачі багатокритеріальної оптимізації (9.2), породжувана множиною D припустимих розв'язків. Нехай m=2, тобто цільова функція $f(X)=(f_1(X),\ f_2(X))$, довільно обираємо припустимий розв'язок X_0, якому відповідає значення цільової функції $f(X_0)=(f_1(X_0),\ f_2(X_0))^T$ та підмножина $\Omega^1 \in \Omega$, породжувана підмножиною D^1 припустимих розв'язків. Отже, будь-який припустимий розв'язок з множини D^1 є строго переважнішим за припустимий розв'язок $X_0 \in D$, тобто

$$X_1 \succ X_0, \quad X_1 \in D_1 \subset D .$$

Продовжуючи ці міркування приходимо до висновку, що з повної множини альтернатив Ω можна виділити підмножину Ω*, яке має дуже цінні властивості:

1) для припустимих розв'язків з множини Ω* у множині D припустимих розв'язків задачі багатокритеріальної оптимізації (9.2) вже немає строго переважніших припустимих рішень;

2) припустимі розв'язки з множини D* або еквівалентні, або незіставні в смислі строгої переваги.

Множину Ω* називають множиною Парето або множиною компромісу. На множині Парето зменшення значень одного з скалярних критеріїв можна досягнути лише ціною збільшення значень іншого. Це спостерігається і в загальному випадку, оскільки будь-які припустимі розв'язки X_1, X_2, або еквівалентні, і для них $f(X_1)= f(X_2)$, або незіставні, тобто для пар координат $f_k(X_1)$, $f_k(X_2)$ векторів $f(X_1)$, $f(X_2)$ мають місце хоч би два з наступних трьох співвідношень: $f_i(X_1)>f_i(X_2)$, $f_j(X_1)=f_j(X_2)$, $f_n(X_1)<f_n(X_2)$.

Підмножина D* множини D припустимих розв'язків є узагальненим розв'язком задачі багатокритеріальної оптимізації (9.2). Цей узагальнений розв'язок в загальному випадку може складатися більш ніж з одного елементу, і тоді ОПР стикається з проблемою вибору одного припустимого розв'язку з множини еквівалентних і незіставних рішень.

9.3. Підходи до розв'язування задачі багатокритеріальної оптимізації

Розв'язки, що утворюють множину Парето, характерні тим, що для жодного з них не існує домінуючого розв'язку. Множину таких точок називають множиною точок оптимальних за Парето або називають областю компромісів (переговорною множиною). Множину векторних оцінок, що відповідають множині неефективних точок (домінуємих розв'язків), називають областю згоди $\Omega_з$. У області $\Omega_з$ немає суперечності між частковими критеріями оптимальності, оскільки кожну точку $X \in D$ можна змінити таким чином, що одночасно будуть поліпшені всі часткові критерії. Якщо область критеріїв Ω складається тільки з області згоди $\Omega_з$, то існує єдина точка $X_{opt} \in D$, в якій всі часткові критерії узгоджені один з одним в тому смислі, що під час руху до точки X_{opt} усі $f_i(X)$ одночасно поліпшуються. Усі часткові критерії досягають мінімуму в точці X_{opt}.

Проте така ситуація зустрічається украй рідко. Найбільш типовим є випадок, коли часткові критерії є суперечливими і мінімум за кожним з них досягається в різних точках. В цьому випадку зменшення одного часткового критерію призводить до збільшення інших часткових критеріїв.

Виділення множини Парето багатокритеріальної задачі оптимізації часто не є задовільним рішенням. Це пов'язано з тим, що при достатньо великій початковій множині варіантів множина Парето виявляється неприпустимо великою для того, щоб ОПР була б в змозі здійснити вибір самостійно. Отже, виділення множини Парето можна розглядати лише як попередній етап оптимізації, і виникає проблема подальшого скорочення цієї множини.

Для вибору однієї оптимальної стратегії з множини ефективних рішень в кожній конкретній багатокритеріальній задачі необхідно використовувати додаткову інформацію про мету операції, тобто ту інформацію, яка під час завдання векторного критерію залишилася неформалізованою і тому невикористаною.

Розглянемо деякі найпростіші способи скорочення Парето-оптимальної множини.

Лексикографічна оптимізація (ранжирування критеріїв). Додаткова інформація, що допомагає у виборі розв'язку, може полягати в тому, що скалярні цільові функції $f_k(X)$ у задачі векторної оптимізації (1.1) впорядковані відповідно до їх значущості. В цьому випадку номер цільової функції відображає ранг (пріоритет) відповідного скалярного критерію.

Нехай Ω^* - множина Парето для задачі векторної оптимізації (1.1) і номер скалярної цільової функції $f_k(X)$ відповідає її рангу. Процедуру вибору єдиного рішення з підмножини D^* почнемо з використання критерію першого рангу. Вважаємо

$$q_1 = \min_{X \in D^*} f_1(X),$$

тобто D_1^* містить всі припустимі розв'язки з D^*, які мінімізують в D^* цільову функцію першого рангу. Далі переходимо до цільової функції другого рангу і вважаємо

$$q_2 = \min_{X \in D^*_1} f_2(X),$$

тобто множина D_2^* містить всі припустимі розв'язки з D_1^*, які мінімізують цільову функцію другого рангу і т.д. Переходимо до цільової функції (m-1)-го рангу і вважаємо

$$q_{m-1} = \min_{X \in D^*_{m-2}} f_{m-1}(X),$$

тобто множина D_{m-1}* містить всі припустимі розв'язки з D_{m-2}*, які мінімізують критерій (m-1)-го рангу. Оскільки $D_{m-1}^* \subset D_{m-2}^* \subset \ldots \subset D_1^* \subset D^*$, то для завершення процедури розв'язання задачі (1.1) в умовах ранжирування критеріїв залишилося розв'язати задачу математичного програмування

$$f_m(X) \to \min_{X \in D^*_{m-1}} .$$

Для реалізації запропонованої процедури розв'язання задачі багатокритеріальної оптимізації підмножині D_{m-1}* множини D має відповідати підмножина множини Парето, що складається більш ніж з одного елементу. Оскільки в загальному випадку ця умова може не виконуватися, то на практиці для розвязання задачі багатокритеріальної оптимізації частіше використовують метод, відомий як метод компромісів.

Припустимо, що для скалярної цільової функції $f_k(X)$ призначено припустиму поступку $\delta_k>0$, яка визначає величину припустимого відхилення значення критерію k-го рангу від його мінімального значення

$$\rho_k = \min_{X \in D^*} f_k(X)$$

на множині D*. Очевидно, що для кожного $k=\overline{1,m-1}$ поступка δ_k визначає певну підмножину в множині D*. Якщо m>2 ця підмножина може виявитися порожньою. В цьому випадку поступки обрані невдало і необхідна їх корекція.

Відзначимо, що практична реалізація розглянутого методу пов'язана, принаймні з двома принциповими труднощами:

1) необхідністю ранжирування скалярних критеріїв;

2) призначенням поступок та їх корекцією.

Обидві операції формально не визначаються і виконуються експертним шляхом. Звернення до експертів неминуче під час розв'язання задач багатокритеріальної оптимізації, оскільки необхідна додаткова інформація, що дозволяє ввести відношення переваги на підмножині D*, яка відповідає множині Парето $\Omega^* \subset \Omega$. Проте, процедури ранжирування скалярних критеріїв та визначення поступок за ними не завжди прості для експертів. Тому застосування розглянутого підходу, як правило, обмежують лише тими ситуаціями, в яких експерти можуть кваліфіковано подолати обидві відзначені труднощі.

Принцип справедливої абсолютної поступки. Згідно цьому принципу, справедливим є такий компроміс, при якому сумарний абсолютний рівень підвищення одного або кількох скалярних критеріїв не перевершує сумарного абсолютного рівня зниження інших критеріїв.

Розглянемо дві точки A і B множини Парето. Через принцип справедливої абсолютної поступки під час переходу від A до B зміна вектору f(X) характеризується величиною

$$\Delta_{абс} = \sum_{k=1}^{m} \Delta_k = \sum_{k=1}^{m} (f_k^B - f_k^A) = \sum_{k=1}^{m} f_k^B - \sum_{k=1}^{m} f_k^A ,$$

де f_k^A, f_k^B - значення скалярних критеріїв в точках A і B. Якщо $\Delta_{абс}<0$, то розв’язок, що відповідає точці B, вважають кращим порівняно з розв’язком, що відповідає точці A. Тому якнайкращим буде такий розв’язок, для якого $\Delta_{абс}\geq 0$ під час переходу в будь-яку іншу точку.

Отже, принцип справедливої абсолютної поступки збігається до мінімізації суми скалярних критеріїв на множині D:

$$\sum_{k=1}^{m} f_k(X) \to \min_{X\in D^*}.$$

Недолік цього принципу в тому, що він припускає диференціацію за окремими критеріями: низьке значення суми $f_1 + f_2 + . + f_m$ може досягатися, коли одні критерії мають порівняно низький рівень, тоді як інші - порівняно високий рівень. Потенційна можливість такої диференціації є характерною для задач, в яких критерії виражені в різних одиницях вимірювання. У таких задачах критерії необхідно нормалізувати, тобто призвести до єдиного, бажано безрозмірному, масштабу вимірювання.

Синтез глобального критерію. Ідея цього підходу дуже проста: для задачі багатокритеріальної оптимізації (1.1) будують глобальний скалярний критерій з цільовою функцією

$$F(X) = \Phi[f_1(X),\ldots, f_m(X)] = \Phi[f(X)], \tag{9.3}$$

що залежить від початкових скалярних цільових функцій так, щоб розв’язок задачі математичного програмування

$$F(X) \to \min_{X\in D}$$

був розв’язком початкової задачі (9.2) у смислі цього принципу компромісу.

Обмежимося лише коротким обговоренням синтезу глобального скалярного критерію (греч. synthesis - з'єднання, поєднання, складання) для задач багатокритеріальної оптимізації. Оскільки система нерівностей

$$f_k(X_0) \leq f_k(X), \quad k=\overline{1,m}, \quad X\in D$$

еквівалентною системі нерівностей

$$l_k f_k(X_0)+c_k \leq l_k f_k(X)+c_k, \quad k=\overline{1,m}, \quad X\in D,$$

де $l_k > 0, \quad k=\overline{1,m}$ і c_k можна вибрати довільно, то припустимий розв’язок X_0 є розв’язком задачі багатокритеріальної оптимізації (9.1) тоді і тільки тоді, коли X_0 є розв’язком задачі векторної оптимізації.

$$l_k f_k(X)+c_k \to \min_{X\in D}, \quad k=\overline{1,m}.$$

Виходячи з цього, розглянемо вимоги, яким має задовольняти функція $\Phi[f(X)]$ (9.3).

По-перше, функція $\Phi[f_1(X),\ldots, f_m(X)]$ має бути інваріантною щодо перетворення зрушення, тобто для будь-якого набору дійсних чисел і для будь-якого припустимого розв'язку X має виконуватися рівність

$$\Phi[f_1(X),\ldots,f_m(X)]=\Phi[f_1(X)+c_1,\ldots,f_m(X)+c_m]. \qquad (9.4)$$

По-друге, функція $\Phi[f_1(X),\ldots,f_m(X)]$ має бути інваріантною щодо до зміни масштабу будь-якого скалярного критерію f_k, тобто для будь-якого набору дійсних позитивних чисел і для будь-якого припустимого розв'язку X має виконуватися рівність

$$\Phi[f_1(X),\ldots,f_m(X)]=\Phi[l_1 f_1(X),\ldots,l_m f_m(X)]. \qquad (9.5)$$

Сформульовані вимоги означають, що який би не був набір дійсних чисел c_k, набір дійсних позитивних чисел l_k і припустимий розв'язок X, вірною є рівність

$$\Phi[f_1(X),\ldots,f_m(X)]=\Phi[l_1 f_1(X)+c_1,\ldots,l_m f_m(X)+c_m]. \qquad (9.6)$$

Виконання вимог (9.4), (9.5) або еквівалентної їм вимоги (9.6), що накладається на функцію $\Phi[f(X)]$ під час синтезу глобального скалярного критерію, визначуваного згідно (9.3), може забезпечуватися за різними способами. Зокрема, функція

$$F(X)=\sum_{k=1}^{m} l_k \overset{0}{f}_k(X), \qquad (9.7)$$

де

$$\overset{0}{f}_k(X)=\frac{f_k(X)-\min\limits_{X\in D} f_k(X)}{\max\limits_{X\in D^*} f_k(X)-\min\limits_{X\in D^*} f_k(X)}$$

відповідає вимогам (9.4), (9.5). Окрім того, ця функція враховує і вимогу нормалізації критеріїв, оскільки замість абсолютних значень скалярних критеріїв розглядаються безрозмірні величини їх відносних відхилень від мінімальних значень. Критерій (9.7) називають нормованим скалярним критерієм.

У спеціальній літературі розглядають інший вид нормованого скалярного критерію, в якому врахована важливість окремих складових:

$$F(X)=\sum_{k=1}^{m} \lambda_k \overset{0}{f}_k(X), \qquad (9.8)$$

де λ_k - певні параметри, для яких

$$\sum_{k=1}^{m} \lambda_k = 1, \quad 0\le \lambda_k \le 1, \quad k=\overline{1,m}. \qquad (9.9)$$

Ці параметри називають ваговими коефіцієнтами. Вагові коефіцієнти можна визначати різними способами, але будь-який підхід збігається до використання експертних оцінок. Відзначимо, що для експерта задача визначення вагових коефіцієнтів не простіше за задачу ранжирування скалярних критеріїв, оскільки в цьому випадку пріоритет кожного скалярного критерію він має виразити кількісно.

Метод «ідеальної» точки. Розглядається m-мірний простір (де - m кількість локальних критеріїв), в якому апріорі обирається вектор, що відображає «ідеальний» розв'язок (або, що теж саме, «ідеальна» точка, координатами якої є «ідеальні» значення, наприклад, мінімальні або максимальні значення локальних критеріїв). У цьому просторі вводиться певна метрика, з метою обчислення відстані між вектором розвязку та "ідеальним". За якнайкраще вибирають такий розв’язок, векторна оцінка якого найбільш близька до "ідеальної" точки. Недоліками методу є свавілля під час вибору ідеальної точки і введення метрики.

Визначимо узагальнений критерій таким чином. Покладемо $a_i = max f_i(X)$, $i = \overline{1, m}$, тобто a_i є максимально (мінімально) можливим значенням за і-м критерієм. Покладемо $a=(a_1, a_2, \ldots, a_m)$. Точку a називають *ідеальною*. Смисл назви пов'язаний з тим, що такі точки оптимальні відразу за всіма критеріями, отримати більшого (меншого) значення за жодним критерієм неможливо. Як правило, точка $a \in \Omega$. Задамо для всіх точок функцію, що є евклідовою відстанню між точками $f_i(X)$ і a:

$$\rho(f_i, a) = \left[\sum_{i=1}^{m}(a_i - f_i)^2\right]^{\frac{1}{2}}.$$

За цільову функцію (узагальнений критерій) беруть вираз

$$f(X) = \sum_{i=1}^{m}\lambda_i\left[a_i - f_i(X)\right]^2,$$

де λ_i - вагові коефіцієнти.

Отже, задача оптимізації формулюється наступним чином

$$\min\sum_{i=1}^{m}\lambda_i\left[a_i - f_i(X)\right]^2, \quad X \in D.$$

Тут принцип оптимальності виражається функцією вибору визначуваною близькістю до ідеальної точки. Як ідеальну точку беруть директивні значення параметрів, задані ОПР.

Контрольні запитання

1. Що називають множиною Парето і чому його ще називають множиною компромісу?
2. У чому полягає основна ідея підходу до розв’язання задач багатокритеріальної оптимізації, пов'язаного з ранжируванням скалярних критеріїв? Які достатні умови реалізації цього підходу Ви знаєте?
3. У яких ситуаціях для розв’язання задач багатокритеріальної оптимізації використовують метод компромісів?
4. Сформулюйте вимоги, яким має задовольняти глобальний скалярний критерій задачі векторної оптимізації, і приведіть їх обґрунтування.
5. Доведіть еквівалентність вимог (1.6), (1.7) і (1.8), що пред'являються до глобального скалярного критерію задачі векторної оптимізації.
6. Викладіть загальну ідею синтезу глобального скалярного критерію для задачі векторної оптимізації.
7. Які можна запропонувати принципи оптимальності для задач багатокритеріальної оптимізації?
8. Які завдання оптимального проектування приводять до використання методу ідеальної точки?

ЗМ 2. МЕТОДИ МОДЕЛЮВАННЯ БІЗНЕС-ПРОЦЕСІВ

ТЕМА 10
МАТЕМАТИЧНІ МОДЕЛІ МАКРОЕКОНОМІКИ

10.1. Напрями економіко-математичного моделювання процесів бізнесу

Економіко-математичне моделювання процесів бізнесу потрібне, перш за все, для однозначного формулювання проблем. Воно є інструментом, що дозволяє приймати адекватні рішення. Основними напрямами моделювання є:

- оцінка стану економічного об'єкту;
- прогнозування стану економічного об'єкту та зовнішнього середовища, в якому він знаходиться;
- планування стану економічного об'єкту.

Знаючи майбутній стан зовнішнього середовища та оцінюючи, наскільки цей стан сприятливий для економічного об'єкту, менеджер може проектувати свої дії з метою зменшення несприятливих наслідків такого стану та його покращення.

Моделювання є одним з найпоширеніших способів вивчення процесів та явищ. За властивостями моделі можна судити про властивості об'єкту, що вивчається. Але не про всі властивості, а лише про ті, які наявні у моделі та в об'єкті і при цьому важливі для дослідження. Такі властивості називають ***істотними***. Отже, економіко-математична модель є основним засобом експериментального дослідження економіки. Вона має наступні властивості:

- імітує реальний економічний процес (або поведінку об'єкту);
- має відносно низьку вартість;
- може багато разів використовуватися;
- враховує різні умови функціонування об'єкту.

Економіко-математична модель - це математичне відображення досліджуваного економічного об'єкту (процесу), за допомогою якого вивчають його функціонування та оцінюють зміну його ефективності за можливих змін вхідних характеристик.

Моделі можна класифікувати за низкою ознак. Розрізняють моделі ***статичні***, в яких описується моментний стан економіки, та ***динамічні***, які показують розвиток об'єкту моделювання.

Змінними моделі є величини, що включені у модель та приймають різні значення в процесі рішення економіко-математичної задачі. Незалежні змінні моделі приймають значення координат модельованої системи. Вони можуть бути керованими або супутніми.

Керовані змінні - це змінні моделі, значення яких піддаються зміні в процесі пошуку рішення. У будь-якій моделі завжди, окрім керованих змінних, присутні фактори, серед яких необхідно виділити ***керовані фактори*** і ***параметри, що управляють.*** Керований фактор - фактор, рівні якого цілеспрямовано обираються менеджером. Значення керованого фактору завжди фіксується під час вирішення задачі, а значення змінної моделі обчислюється. Параметри, що управляють, - змінні величини (зазвичай функції часу), що визначають напрям та швидкість зміни керованої системи. Параметри, що управляють, характеризують рішення, які треба здійснювати в кожен момент часу, виходячи з інтервалу між початковим та кінцевим станом системи. Наприклад, підприємству доцільно показувати наявність прибутку тільки в певні періоди часу, коли має відбутися виплата дивідендів. У решту періодів, виходячи з основних положень фіскальної політики, величина прибутку має бути мінімальною, щоб мінімізувати податкові платежі.

Прийнято розрізняти ***екзогенні*** або вхідні (що розраховуються поза моделлю) змінні та ***ендогенні*** або вихідні (невідомі, визначувані в процесі рішення задачі) змінні, траєкторія зміни яких визначається в результаті реалізації моделей.

Сутність використання економіко-математичних моделей в практичних дослідженнях в основному полягає в прогнозуванні поведінки ендогенних змінних за певних припущень щодо поведінки екзогенних змінних.

Макроекономічною є модель, що відображає функціонування економіки країни або регіону як єдиного цілого. Макромоделі оперують, як правило, крупно агрегованими показниками: валовий національний продукт, інвестиції. Макромоделі використовують для теоретичного аналізу найбільш загальних закономірностей функціонування та розвитку економіки країни або регіону. У практичній діяльності їх застосовують для прогнозування економічних процесів.

Мікроекономічна модель відображає функціонування та структуру окремого елементу економічної системи, його взаємодію з іншими елементами системи в процесі функціонування. Чітке розмежування між макромоделями та мікромоделями відсутнє. Але до перших, як правило, належать найбільш узагальнені глобальні моделі. Для мікромоделей характерна велика залежність від зовнішнього середовища, дезагрегація показників.

Отже, економіко-математичні моделі є результатом формалізації реально існуючих систем, проте, основною метою більшості описаних в літературі моделей є не вивчення поведінки реальних економічних об'єктів, а пошук ефективних методів рішення конкретних ситуацій в процесі функціонування цих об'єктів.

Більшість економіко-математичних моделей характеризуються статичним підходом до вивчення економіки, коли її стан вивчається на заданий момент часу. Під статичною економічною системою розуміють таку систему, координати якої на відрізку часу, що вивчається, можна розглядати як постійні. Відповідно, при формулюванні статичної економіко-математичної моделі передбачають, що всі залежності належать одному моменту часу, а модельована система незмінна в часі. Оскільки статичні моделі у формалізованому вигляді не містять фактора часу, вони завжди простіші, ніж динамічні моделі тих самих економічних систем.

У статичних моделях виділяють групу макроекономічних моделей, до яких належать моделі народно-господарського рівня, призначені для опису великих секторів економіки або економіки країни в цілому. Метою макроекономічного моделювання є вивчення економічних законів, що пов'язують найважливіші показники. В цілому, розроблені до теперішнього часу математичні моделі народного господарства можна умовно розбити на дві великі групи. До першої належать ***моделі економічного зростання*** (часто це динамічні моделі). Ці моделі оперують крупноагрегованими показниками (валовий суспільний продукт, національний дохід, об'єм основних фондів, фонд накопичення, фонд споживання). Вони призначені для вивчення основних тенденцій розвитку економіки протягом тривалих періодів часу (близько кількох десятиріч) та часто подаються виробничими функціями. До другої групи належать ***міжгалузеві балансові моделі***, що є матричними моделями, які відображають співвідношення між витратами на виробництво та його результатами. Матричні моделі застосовують в міжгалузевому балансі, під час розв'язання галузевих задач оптимального планування розвитку та розміщення виробництва, в еколого-економічному моделюванні та ін.

У разі використання виробничих функцій економіку розглядають як «чорний ящик», структура якого невідома. Звідси випливає, що в цій моделі економіка виступає як цілісна неструктурована одиниця, на вході якої ресурси, а на виході, як результат функціонування - валовий випуск або валовий внутрішній продукт. Ресурси розглядають як аргументи, а валовий випуск або валовий внутрішній продукт - як функцію.

У моделі міжгалузевого балансу економіка структурована і складається з кінцевої кількості чистих галузей, кожна з яких виробляє тільки один продукт. А для виробництва одиниці кожного продукту в галузі потрібно витратити певні обсяги інших продуктів, включаючи даний. Модель міжгалузевого балансу часто називають ***моделлю Леонтьева***, основу якої складає матриця коефіцієнтів прямих витрат. Ця модель дозволяє за заданим кінцевим продуктом в галузевому розрізі визначити валові випуски галузей.

Як динамічну систему розуміють систему, що змінюється в часі. Математично це прийнято виражати через змінні, які називають ***координатами***. Процес зміни змінних характеризується траєкторією:

$$Q(t)=[q_1(t), q_2(t),..., q_n(t)].$$

де координати $q_1(t),...,q_n(t)$ є функціями часу t.

Серед таких систем найпростішими є лінійні динамічні системи, в яких зв'язки між вхідними величинами, параметрами стану та вихідними величинами носять характер лінійних залежностей.

У економіко-математичних моделях динамічні системи можуть відображатися за допомогою опису стану системи в певні моменти часу. Виходять як би моментальні знімки, звані статичними моделями. А також за допомогою динамічних моделей економіки, що описують сам процес розвитку системи.

Один з принципових підходів до побудови таких моделей - оптимізаційний. Він полягає у виборі такої траєкторії економічного розвитку з числа можливих, при якій забезпечується максимальне зростання одного або кількох показників. Другий підхід полягає в дослідженні рівноваги в економічній системі. В цьому випадку, переходячи до економічної динаміки, використовують поняття «рівноважна траєкторія», тобто врівноважене збалансоване зростання. Такий результат виходить за рахунок взаємодії множини факторів та об'єктів економічної системи.

Як екзогенні фактори можуть виступати виявлені за допомогою статистики макроекономічні залежності, дані прогнозу про демографічні процеси. Як ендогенні фактори - темпи зростання, показники економічної ефективності.

Математичний опис динамічних моделей проводиться, як правило:

- системами диференціальних рівнянь (де час є безперервною змінною);

- різницевими рівняннями (де час - дискретна величина);
- системами звичайних алгебраїчних рівнянь.

За допомогою динамічних моделей, зокрема, розв'язують завдання з планування та прогнозування економічних процесів: визначення траєкторії розвитку економічної системи та її станів в задані моменти часу; аналіз економічної системи на стійкість; аналіз структурних зрушень.

З погляду теоретичного аналізу велике значення придбали ***динамічна модель фон Неймана*** та теореми про магістралі. Окрім того, в практичній діяльності використовують багатогалузеві (багатосекторні) динамічні моделі розвитку економіки (динамічні моделі міжгалузевого балансу), виробничі функції, теорію економічного зростання.

10.2. Поняття та призначення виробничих функцій

Виробничою функцією (функцією виробництва) є рівняння, що пов'язує змінні величини витрат (ресурсів, факторів виробництва) з величиною випуску продукції. Випуск може вимірюватися в натуральних або вартісних показниках, в реальних або потенційних величинах. А ресурси можуть розглядатися або фактично витрачені, або наявні на початок періоду виробництва. Кількість факторів у виробничій функції не обов'язково обмежується наперед, проте потрібна їх зіставність за характером впливу на випуск та рівнем агрегації.

У економічному моделюванні найширше представлені ***макроекономічні виробничі функції***. Ці функції характеризують залежність показника сукупного суспільного продукту або іншого узагальнюючого показника від основних факторів виробництва. Як основні фактори виробництва зазвичай розглядають обсяг капіталу, робочої сили, а також землі. У ряді макроекономічних виробничих функцій як окремий фактор враховується ще вплив науково-технічного прогресу. Макроекономічні виробничі функції досліджуються самостійно або включаються у складні економетричні моделі. Вони, як правило, містять 2-4 фактори виробництва, наприклад, живу працю, основні засоби, науково-технічний прогрес, показник природних ресурсів та ін.

Виробничу функцію, що встановлює залежність обсягу виробництва від наявності або споживання ресурсів, називають функцією випуску. Окремими випадками виробничої функції є:

- функція витрат, що описує зв'язок між обсягом випуску та витратами виробництва;
- інвестиційна функція, що описує залежність необхідних інвестицій від виробничої потужності майбутнього підприємства.

Формально виробничу функцію можна записати таким чином:

$$y = f(x_1, x_2, x_n),$$

де y - обсяг випуску; x_j - обсяг ресурсу j.

Параметри функції оцінюють за методами кореляційно-регресійного аналізу. Одержані таким чином виробничі функції є статистичними залежностями між ресурсами та випуском. Причому, часто оцінка погрішності є такою, що користуватися одержаними залежностями на практиці не можливо, особливо у разі множинної регресії. Тому одержані залежності відображають тільки передбачувані тенденції розвитку та мають низьку достовірність. У роботах західних економістів неокласичного напряму значення параметрів виробничої функції часто визначають виходячи з наступних гіпотез:

- про рівність відношення граничних продуктивностей ресурсів та відношення цін на них. Наприклад, як «ціну праці» розглядають середню ставку заробітної платні, а як «ціну капіталу» - норму відсотку;
- про рівність еластичностей випуску за ресурсами та часткою їх власників у доході.

Виробнича функція дозволяє розрахувати ряд важливих характеристик, що описують різні сторони досліджуваної виробничої одиниці. Найчастіше розраховують наступні характеристики:

- граничну продуктивність (граничний продукт) фактору j $\frac{\partial f(x)}{\partial x_j}$, $j=\overline{1,n}$.

Цей фактор показує, наскільки збільшується випуск при збільшенні витрат фактору j на одну одиницю, при незмінній кількості решти факторів;

- часткову еластичність випуску за фактором j (часткова еластичність фактору) $\frac{\partial f(x)}{\partial x_j} \cdot \frac{x_j}{f(x)}$, $j=\overline{1,n}$. Показує, на скільки відсотків збільшиться випуск при збільшенні витрат фактору j на 1% та при незмінній кількості решти факторів. Часткова еластичність є відношенням граничної продуктивності до середньої;
- еластичність виробництва

$$\pi(x_1, x_2, \ldots, x_n) = \lim_{\lambda \to 1} \frac{\lambda}{f(\lambda x_1, \ldots, \lambda x_n)} \cdot \frac{\partial f(\lambda x_1, \ldots, \lambda x_n)}{\partial \lambda}.$$

Еластичність виробництва показує, на скільки відсотків збільшиться випуск при збільшенні на 1% витрат кожного фактора. Цей показник є локальною характеристикою ефекту масштабу виробництва. Очевидно, що

$$\pi(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{n} \frac{\partial Y}{\partial x_j} \cdot \frac{x_j}{Y}.$$

- гранична норма заміщення фактору j фактором i. Цей показник визначає кількість фактору j, яка потрібна для заміни однієї одиниці фактора j при збереженні на незмінному рівні обсягу випуску та кількості решти факторів. Зазвичай позначається R_{ij} і, за визначенням, дорівнює:

$$R_{ij} = -\frac{\partial x_i}{\partial x_j}, \text{ при } Y=\text{const}, x_k=\text{const}, k\neq i,j.$$

Очевидно, що $R_{ij} = -\frac{\partial Y}{\partial x_j} : \frac{\partial Y}{\partial x_i}$.

- еластичність заміщення фактору j фактором i. Разом з граничною нормою заміщення цей показник характеризує можливості заміни одного фактору іншим. У простому випадку визначається як

$$\sigma_{ij} = \left(\frac{\partial R_{ij}}{\partial\left(\frac{x_i}{x_j}\right)} \cdot \frac{x_i / x_j}{R_{ij}} \right)^{-1}$$, при Y=const, x_k=const, k≠i,j.

Існує низка інших визначень еластичності заміщення для багатофакторних виробничих функцій. Всі існуючі визначення є еквівалентними тільки для двофакторних лінійно однорідних виробничих функцій. В цьому випадку всі вони призводять до формули:

$$\sigma_{12} = \sigma_{21} = \sigma = \frac{\frac{\partial Y}{\partial x_1} \cdot \frac{\partial Y}{\partial x_2}}{Y \frac{\partial^2 Y}{\partial x_1 \partial x_2}}.$$

Часто конкретний вид виробничої функції виводять, виходячи з гіпотез про значення та характер зміни будь-яких з вказаних п'яти характеристик.

Отже, за допомогою виробничих функцій вивчають взаємозамінюваність факторів виробництва, яка може бути незмінною або змінною (тобто залежною від обсягів ресурсів). Відповідно, функції поділяють на два види:

- з постійною еластичністю заміни (CES - Constant Elasticity of Substitution);
- із змінною еластичністю заміни (VES - Variable Elasticity of Substitution).

10.3. Основні форми подання виробничих функцій

В цей час математиками-аналітиками запропоновано багато конкретних виробничих функцій. Найчастіше використовують наступні:

- лінійну $Y = a_1x_1 + ... + a_nx_n$;
- леонтьєвську $Y = \min\left(\frac{x_1}{a_1}, \ldots, \frac{x_n}{a_n}\right)$;
- Кобба-Дугласа $Y = Ax_1^{\alpha 1} \cdot x_2^{\alpha 2} \cdot \ldots \cdot x_n^{\alpha n 1}$;
- з постійною еластичністю заміщення. У найпростішому варіанті ця функція має вигляд:

$$Y = A\left[a_1 x_1^{-\rho} + \ldots + a_n x_n^{-\rho}\right]^{-\frac{\lambda}{\rho}}.$$

Найпоширеною в теоретичних та прикладних дослідженнях є функція Кобба-Дугласа. Вона поєднує простоту математичного запису, очевидну економічну інтерпретацію та відносну легкість визначення чисельних значень її параметрів. Особливість цієї мультиплікативно-степеневої форми виробничої функції полягає в тому, що якщо один співмножник дорівнює нулю, то результат перетворюється також на нуль. Ця властивість відповідає тому факту, що в більшості випадків для виробництва необхідні всі фактори, і за відсутності одного з них випуск продукції неможливий. Наприклад, навіть в самому автоматизованому виробництві не можна обійтися без відповідного персоналу. У найзагальнішій формі (форма називається канонічною) мультиплікативно-степеневу функцію записують в наступному вигляді:

$$Y = Ax_1^{\alpha 1} \cdot x_2^{\alpha 2} \cdot \ldots \cdot x_n^{\alpha n} \quad \text{або} \quad Y = A\prod_j x_j^{\alpha j}.$$

Коефіцієнт А враховує розмірність, що залежить, у свою чергу, від обраної одиниці вимірювань витрат та випуску. Співмножники від першого до n-го можуть мати різний зміст залежно від того, які фактори роблять вплив на загальний результат (тобто випуск продукції). Наприклад, у виробничій функції, яка застосовується для вивчення економіки в цілому, як результативний показник можна прийняти обсяг кінцевого продукту, а як співмножники - основні фактори виробництва:

- чисельність зайнятого населення x_1;
- величину основного та оборотного капіталу x_2;
- площу використовуваної землі x_3.

За допомогою функції Кобба-Дугласа було зроблено спробу оцінити зв'язок таких факторів, як праця і капітал, із зростанням національного доходу США в 20-30 роки XX сторіччя:

$$N = A \cdot L^{\alpha} \cdot K^{\beta},$$

де N - національний дохід; A - коефіцієнт розмірності; L і K - відповідно обсяги прикладеної праці і капіталу; α, β- коефіцієнти еластичності виробництва за працею L та капіталом K.

Функцію Кобба-Дугласа використовують для опису обсягу виробництва залежно від кількості зайнятих (поряд з капіталом):

$$Y = cK^{a} \cdot L^{b},$$

де Y - обсяг виробництва; K - величина капіталу; L - чисельність зайнятих; c - постійний параметр продуктивності; a - коефіцієнт еластичності виробництва за величиною капіталу; b - коефіцієнт еластичності виробництва за чисельністю зайнятих. Сума коефіцієнтів еластичності a + b характеризує ефект масштабу виробництва:

- зростаючий, якщо $a + b > 1$;
- постійний, якщо $a + b = 1$;
- що убуває, якщо $a + b < 1$.

Хоча сума a + b може приймати будь-які значення, найчастіше передбачається незмінний масштаб виробництва. У зв'язку з цим припущенням, один параметр визначається через інший: $b = 1 - a$. У «класичній» виробничій функції Кобба-Дугласа $\alpha = 0,33$, $\beta = 0,67$.

Серед моделей, що характеризують вплив демографічного фактора на економічний розвиток, виділяють динамічні моделі, засновані на припущенні, що технологічні зміни впливають на обсяг виробництва безпосередньо (модель Р. Солоу):

$$Y_t = c \cdot e^{r-t} \cdot (K_t)^{a} \cdot (L_t)^{b},$$

де t - календарний рік; r - постійний темп технічного розвитку.

Другий метод обліку технічного розвитку припускає зміну впливу окремих факторів виробництва, яку моделюють за допомогою динамічної зміни коефіцієнтів еластичності (модель М. Брауна):

$$Y_t = c \cdot (K_t)^{a(t)} \cdot (L_t)^{b(t)}.$$

Третій метод заснований на тому, що технічний розвиток призводить до якісної зміни усередині факторів виробництва (модель Р. Солоу):

$$Y_t = c \cdot (K^*{}_t)^a \cdot (L^*{}_t)^b,$$

де індекс «*» відбиває якісні зміни у фізичному або людському капіталі.

Степеневі коефіцієнти (параметри) мультиплікативно-степеневої виробничої функції показують ту частку у відсотковому прирості кінцевого продукту, яку вносить кожний із співмножників (або на скільки відсотків зросте продукт, якщо витрати відповідного ресурсу збільшити на один відсоток). Ці параметри є коефіцієнтами еластичності виробництва щодо витрат відповідного ресурсу. Якщо сума коефіцієнтів складає 1, то це означає однорідність функції: вона зростає пропорційно зростанню кількості ресурсів. Але можливі і такі випадки, коли сума параметрів більше (або менше) одиниці. Це показує, що збільшення витрат призводить до непропорційно більшого (або непропорційно меншого) зростання випуску (ефект від масштабу виробництва).

У динамічному варіанті застосовуються різні форми виробничої функції. Наприклад, у двофакторному випадку:

$$Y(T) = A(t) \cdot L^{\alpha}(t) \cdot K^{\beta}(t),$$

де множник A(t) зазвичай зростає в часі, відображаючи загальне зростання ефективності факторів виробництва в динаміці.

Логарифмуючи, а потім диференціюючи за t вказану функцію, можна одержати співвідношення між темпами приросту кінцевого продукту (або, наприклад, національного доходу) та приросту факторів виробництва (темпи приросту змінних прийнято описувати у відносних величинах, у відсотках). Подальша адаптація виробничої функції може полягати у використанні змінних коефіцієнтів еластичності.

Найбільш гнучкою та змістовною вважається CES-функція, окремим випадком якої є функції Кобба-Дугласа, проте в загальному випадку оцінка її параметрів утруднена.

Приклади інших виробничих функцій наводяться для випадку двох факторів Y = f(K, L), де K - капітал, а L - обсяги прикладеної праці (витрати живої праці). Значну кількість виробничих функцій одержано в результаті комбінації різних варіантів наведених вище чотирьох функцій. Серед них:

- функция з лінійною еластичністю заміщення

$$Y_t = AK^{\alpha} \cdot (\beta K + L^{1-\alpha}).$$

Ця функція виводиться з припущення, що еластичність заміщення лінійно залежить від фондоозброєності. Для цієї виробничої функції еластичність заміни фактору K фактором L дорівнює $\sigma = 1 + \frac{\beta}{\alpha} \cdot \frac{K}{L}$;

- багаторежимна функція

$$Y = A\prod_j \left[\alpha_j K^{-\rho} + (1-\alpha_j)L^{-\rho}\right]^{-\frac{\gamma_j}{\rho}}.$$

Ця функція виводиться з припущення, що еластичності випуску за ресурсами є n-рівневими ступінчастими функціями фондоозброєності (для еластичності за капіталом - убуваюча, для еластичності за працею - зростаюча).

Серед неоднорідних виробничих функцій найчастіше використовують квадратичну функцію

$$Y = aK+bL+cKL-dK^2-eL2,$$

та функцію

$$Y = A\left[\alpha K^{-\beta 1} + (1-\alpha)L^{-\beta 2}\right]^{-\frac{1}{\beta 3}},$$

називану функцією Солоу, або функцією Хілхорста. Переваги цієї виробничої функції полягають в тому, що її верифікація дозволяє перевірити гіпотезу про однорідність. Якщо β_1 та β_2 опиняються близькими, ця гіпотеза приймається, інакше - відкидається.

Описувані виробничою функцією співвідношення носять статистичний характер, тобто виявляються тільки в середньому, у великій масі спостережень, оскільки реально на результат виробництва впливають не тільки аналізовані фактори, але й множина неврахованих в цьому виді моделі. Окрім того, вживані показники як витрат, так і результатів неминуче є продуктами складної агрегації. Наприклад, узагальнений показник витрат праці в макроекономічній функції вбирає витрати праці різної продуктивності, інтенсивності, кваліфікації та ін.

На практиці застосовують три основні методи визначення параметрів макроекономічних виробничих функцій:

- на основі обробки рядів динаміки (часових рядів);
- на основі даних про структурні елементи агрегатів;
- на основі даних про розподіл національного доходу (розподільний метод).

Під час побудови виробничих функцій необхідно позбавлятися від явищ мультиколлінеарності параметрів та автокореляції, інакше неминучі грубі помилки.

10.4. Моделі споживання

Моделі споживання описують взаємозв'язки між споживанням та факторами, що його визначають. Їх використовують для аналізу динаміки споживання за минулий період та побудови прогнозів на перспективу.

Найбільшого поширення в практиці аналітичних та прогнозних розрахунків набули моделі споживання, побудовані на основі регресійного аналізу. У рівнянні регресії як функція виступає обсяг споживання, а як незалежні змінні – фактори, що його визначають.

Для випадку лінійної залежності між споживанням та визначальними факторами модель подають в наступному вигляді:

$$y = a_0 + \sum_{j=1}^{m} a_j \cdot x_j ,$$

де y - споживання конкретного товару або агрегату товарів; x_j ($j \in 1:m$) - враховані в моделі фактори; $a_0,..., a_j$ - параметри моделі.

Якщо аналізу підлягає загальний обсяг споживання населенням товарів та послуг, то як визначальні фактори в модель можна включити:

- загальний обсяг валового внутрішнього продукту;
- величину грошових доходів населення;
- індекс роздрібних цін на споживчі товари;
- індекс тарифів на платні послуги.

Якщо моделюють споживання окремих товарів, то в модель досить включати фактори, що характеризують доходи населення, ціни відповідних товарів і, можливо, ще деякі фактори, що відбирають відповідно до особливостей даного товару.

Залежно від джерела інформації розрізняють моделі, засновані на даних суцільного статистичного обліку, та на даних вибіркового обстеження бюджетів сімей. Перевагою перших є те, що вони можуть врахувати динаміку споживання та факторів, що його визначають. Перевага других - явний облік розподілу населення за рівнем середньодушового доходу, який робить сильний вплив на обсяги і структуру споживання як в цілому, так і за окремими товарами. Проте за допомогою певних математико-статистичних процедур існує можливість додавати моделям споживання, заснованим на даних бюджетної статистики, динамічний характер, а в моделях, що будують на даних часових рядів споживання та визначають низку факторів, враховувати розподіл населення за доходом. Це істотно розширює можливості цих двох типів моделей.

Для аналізу і прогнозування споживання населення застосовують і складніші економіко-математичні моделі. Проте, як показує досвід, ускладнення моделей споживання не завжди призводить до підвищення точності одержаних за їх допомогою результатів, а часто точність прогнозу навіть знижується. До того ж для ускладнених моделей не завжди вдається знайти достатню та достовірну первинну інформацію, що призводить до необхідності використовувати оцінкові дані.

Контрольні запитання

1. Охарактеризуйте основні напрями економіко-математичного моделювання бізнес-процесів.
2. Які властивості об'єкту, що вивчається, називають істотними? Як їх визначають?
3. Перелічіть переваги економіко-математичної моделі та назвіть, за якими ознаками їх розрізняють.
4. Поясніть визначення керованих факторів, параметрів, що управляють екзогенних та ендогенних змінних.
5. З якою метою використовують макроекономічні моделі? Які процеси вони моделюють?
6. Наведіть визначення виробничої функції. Поясніть, що являє собою макроекономічна виробнича функція? Наведіть приклади.
7. Які характеристики виробничої одиниці найчастіше розраховують на підставі виробничих функцій?
8. Проаналізуйте, як відрізняються еластичність виробництва та еластичність заміщення факторів.
9. Які форми виробничих функцій використовують найчастіше?
10. Наведіть вираз мультиплікативно-степеневої форми виробничої функці Кобба-Дугласа.
11. Поясніть, як у функції Кобба-Дугласа, що описує обсяг виробництва залежно від чисельності зайнятих, коефіцієнти еластичності характеризують ефект масштабу виробництва.
12. За якими методами визначають параметри виробничих функцій?
13. Наведіть лінійну модель споживання та охарактеризуйте її складові.

ТЕМА 11

МАТЕМАТИЧНІ МОДЕЛІ МІКРОЕКОНОМІКИ

11.1. Особливості моделювання мікроекономічних процесів

Чіткої відмінності між макромоделями та мікромоделями немає. Як правило, термін «мікроекономічна модель» відносять до вивчення діяльності таких провідних ланок економіки як домашнє господарство (споживач) та фірма (виробник). Домашнє господарство прагне до максимізації корисності, фірма - до максимізації прибутку. Відповідно, до мікроекономічних моделей належать, наприклад, моделі попиту та споживання, поведінки фірми, ціноутворення, ринку товарів, ринку капіталів та інших приватних товарів.

Мікроекономічна модель описує поведінку конкретних економічних об'єктів (аж до окремої особи - споживача або виробника), що ухвалюють рішення (здійснюють вибір можливих альтернатив) в умовах функціонування соціально-економічної системи. Кожен об'єкт одержує, або купує, або здобуває якимсь іншим шляхом потрібну йому інформацію, розподіляє наявні ресурси, розробляє правила вибору альтернатив та стратегію подальших дій. Виходячи з цього, можна виділити три істотні області застосування мікроекономічних моделей:

- ціноутворення;
- ухвалення рішень про обсяг виробництва та продажів;
- розподіл доходів.

Характеризуючи відмінності мікроекономічних моделей від макромоделей, треба зазначити велику залежність перших від зовнішнього середовища і дезагрегацию показників.

Так само як і макроекономічні моделі, мікроекономічні моделі можуть бути статичними і динамічними, детермінованими та імовірнісними, дискретними та безперервними.

11.2. Основні принципи та етапи моделювання попиту і споживання

Розглядатимемо попит як представлену на ринку потребу в товарах і послугах, що дорівнює величині грошових коштів, яка є у населення. Обсяг попиту формується:

- з поточних доходів, що включають заробітну платню, пенсії, допомогу, стипендії, підприємницький дохід;
- грошових заощаджень, що утворилися в минулі періоди.

Суму грошей, що фактично витрачені на покупку товарів і послуг, прийнято називати ***реалізованим попитом***. Проте останній не завжди відображає дійсні купівельні наміри. Він може включати прихований незадоволений попит, тобто випадки, коли покупець замість відсутнього продажу товару вимушений придбати інший, що в меншій мірі задовольняє його потреби, або «відкласти» гроші, але не з метою накопичення, а саме через відсутність прийнятних способів їх витратити. Прихований незадоволений попит свідчить про порушення необхідної відповідності між структурою попиту і товарної пропозиції.

Для дослідження динаміки обсягу і структури грошових доходів населення на придбання товарів і послуг використовують різні способи моделювання попиту. Вони базуються на вивченні кількісної залежності попиту від факторів, що його зумовлюють. Багатофакторний аналіз попиту здійснюється із застосуванням методів математичної статистики. Зазвичай модель попиту подається регресійною функцією $y = f(x_1, ..., x_n)$, де y - залежна змінна - величина попиту на той або інший товар, а $x_1, x_2,..., x_n$ - фактори (аргументи), що формують попит.

Моделювання попиту включає чотири етапи.

1-й етап - відбір факторів, що вводяться в модель. При відборі факторів враховують стандартні вимоги: фактори, що включаються в модель, мають відображати істотні особливості явища, що вивчається:

- характеристики факторів повинні бути кількісно сумірними;
- фактори повинні бути незалежними один від одного, тобто між ними не має бути функціонального зв'язку.

Найбільш істотним фактором, що визначає структуру попиту, є розмір сукупного доходу. При однаковому рівні доходу структура попиту розрізняється в окремих груп населення залежно від їх соціального та професійного складу. Тому закономірності попиту необхідно вивчати диференційовано, в розрізі відповідних соціальних груп.

Великий вплив на попит роблять:

- склад доходу (співвідношення грошової та натуральної його частин);
- рівень і співвідношення роздрібних цін;
- ступінь насичення потреб в різних товарах;
- взаємозалежність та взаємозамінюваність в споживанні окремих благ.

На структуру попиту в тій чи іншій мірі впливають й інші фактори, але не всі вони можуть бути враховані в моделі. Пояснюється це тим, що не кожен фактор піддається точному вимірюванню, а про деякі з них немає досить повної та достовірної інформації. Врахувати вплив таких факторів в часі можна за допомогою введення в модель спеціального фактору - тренду.

Разом з факторами, загальними для всіх або більшості товарів (дохід, ціна та ін.), попит на кожен товар залежить ще від специфічних, властивих тільки йому факторів. Тому багатофакторні моделі попиту будуються диференційовано, тобто для різних товарів у них включають різні аргументи.

2-й етап - встановлення математичної форми зв'язку між величиною попиту та факторами, що його зумовлюють. Рівняння множинної регресії можуть бути лінійні, нелінійні та комбіновані. На практиці найчастіше зустрічаються лінійні (типу: $y = a_0+a_1x_1+a_2x_2 + ... + a_nx_n$) та приведені до лінійних форм зв'язки, зокрема степеневі залежності (наприклад $\log y=a_0+a_1\log x_1+a_2\log x_2+...+a_n\log x_n$).

3-й етап - рішення сформульованої задачі, яке збігається до визначення значень параметрів регресійних рівнянь. Способи рішення можуть бути різні, частіше за інші застосовують метод найменших квадратів.

Для перевірки ступеня відповідності розрахованих в результаті розв’язання задачі теоретичних значень попиту емпіричним даним можна використовувати наступний прийом. За одержаною моделлю проводять екстраполяцію на період, за яким є звітні дані. Якщо розраховані результати достатньо близькі до емпіричних значень залежної змінної, модель вважають придатною для перспективних розрахунків.

4-й етап - на завершальному етапі здійснюють прогноз попиту шляхом підстановки прогнозних значень факторів, врахованих в моделі, у рівняння регресії. Точність виявлених тенденцій попиту та його прогнозу великою мірою залежить від інформаційної бази розрахунків, а також інтуїції та досвіду дослідника. Як інформаційна база можуть служити дані статистики товарообігу та інші джерела загальних народно-господарських статистичних показників, матеріали вибіркових обстежень сімейних бюджетів.

Економіко-математичні побудови, що описують взаємозв'язки та залежності між споживанням та факторами, що його визначають, належать до моделей споживання. Моделі споживання використовують для аналізу динаміки споживання за минулий період, побудови його прогнозів на перспективу різної тривалості. Найбільшого поширення в практиці аналітичних та прогнозних розрахунків набули моделі споживання, що побудовані на основі регресійного аналізу: у рівнянні регресії як функція виступає обсяг споживання, а як незалежні змінні – фактори, що його визначають.

Моделі споживання для випадку лінійного зв'язку між споживанням та факторами, що його визначають, мають вигляд

$$y = a_0 + \sum_{j=1}^{m} a_j x_j ,$$

де y - споживання конкретного виду товару або агрегату товарів; x_j, $j \in 1:m$ - враховані в моделі фактори; a_j - параметри моделі.

Якщо аналізу підлягає загальний обсяг споживання населенням товарів і послуг, то як визначальні фактори в модель можуть бути включені:

- загальний обсяг валового внутрішнього продукту;
- величина грошових доходів населення;
- індекс роздрібних цін на споживчі товари та індекс тарифів на платні послуги.

Якщо ж моделюється споживання окремих товарів, то в модель досить включати фактори, що характеризують доходи населення, ціни відповідних товарів, і, можливо, ще деякі фактори, що відбираються відповідно до особливостей даного товару.

Залежно від джерела інформації розрізняються моделі, засновані:

- на даних суцільного статистичного обліку;
- на даних вибіркового обстеження бюджетів сімей.

Перевагою перших моделей є те, що вони можуть врахувати динаміку споживання та факторів, що його визначають. Перевагою других моделей є явний облік розподілу населення за рівнем середньодушового доходу, який робить сильний вплив на обсяги та структуру споживання, як в цілому, так і за окремими товарами. Проте за допомогою певних математичних процедур існує можливість додавати моделям споживання, заснованим на даних бюджетної статистики, динамічний характер, а в моделях, що будуються на даних динамічних рядів споживання і факторів, що його визначають, враховувати розподіл населення за доходом. Це значно розширює можливості цих двох типів моделей. Спеціальну групу представляють моделі споживання, засновані на просторовій інформації.

11.3. Функції корисності та споживання

Для моделювання поведінки споживача доцільно використовувати цільову функцію споживання. Такого роду цільова функція є окремим випадком функції корисності, яка, у свою чергу, представляє математичну модель споживчих переваг.

Народно-господарська функція корисності може бути представлена у вигляді суми індивідуальних функцій корисності за кожним продуктом, вироблюваним в народному господарстві

$$U = u_1(x_1)+... + u_n(x_n),$$

де U - загальна сума господарської користі, а $u_i(x_i)$ - функції корисності за кожним продуктом.

Гранична корисність буде першою похідною функції корисності, вона завжди буде додатною: $u'_i(x_i)$. Загальною обмежуючою умовою є сукупні трудові ресурси суспільства

$$N = T_1x_1+... + T_nx_n,$$

де N - ресурси праці, приурочені до певного моменту часу, а T_1, ..., T_n - кількість праці, що витрачається на виробництво кожного виду продукції.

Під час аналізу функції корисності навіть в такій простій формі можна одержати цікаві та змістовні висновки. Нехай потрібно визначити, яке співвідношення має бути між x_1, ..., x_n, за яких би то не було часткових їх значеннях, якщо виробництво ведеться доцільно. Справа збігається до відшукання (за правилами відносних максимумів) функції Лагранжа

$$W = U+\lambda(T_1x_1+... + T_nx_n-N).$$

Якщо узяти перший та другий диференціали функції W, то одержимо

$$dW = [u'_1(x_1) + \lambda T_1]\, dx_1+...+[u'_n(x_n) + \lambda T_n]\, dx_n;$$

$$d^2W = u''_1(x_1)dx_1^2+... + u''_n(x_n)dx_n^2.$$

Оскільки d^2W істотно від'ємна через властивості функцій $u_i(x_i)$ - ці функції мають від'ємні другі похідні, то, очевидно, досягається максимум, якщо задовольняється умова вигляду

$$u'_1(x_1) + \lambda T_1 =0;...;u'_n(x_n)+ \lambda T_n =0,$$

яку інакше можна переписати так:

$$-\lambda = \frac{u'_1}{T_1} = \ldots = \frac{u'_n(x_n)}{T_n}.$$

Під час розгляду останньої рівності можна виявити, що чисельники дробів в цій формулі є не що інше, як граничні корисності економічних благ, а знаменники - трудові витрати на їх виробництво. З цієї формули виходить, що граничні корисності вільно відтворених економічних благ пропорційні їх трудовим вартостям.

Якнайповніший доказ умов досягнення максимуму функції корисності був даний Е. Е. Слуцким. Він показав, що корисність будь-якого поєднання благ має тим більшу величину, чим в більшій мірі це поєднання опиняється переважним для певного індивіда. Функція корисності тим самим предстає у вигляді функції переваги. За такій постановці питання аналіз функції корисності збігається не до визначення її абсолютного рівня, а до аналізу змін поверхонь її рівня, оскільки тільки рух рівнів в тому або іншому напрямі дозволяє врахувати реальний зміст самої функції. Відповідно до цього функція корисності одержує вигляд:

$$U=f(x_1,.,x_n),$$

де x_i - кількість різних благ, споживаних суб'єктом за даний інтервал часу; U - корисність, що одержується суб'єктом за допомогою даного поєднання благ.

Гранична корисність будь-якого блага визначається за формулою першої часткової похідної $\frac{\partial U}{\partial x_i}$, яка завжди буде додатною.

Під час аналізу бюджету споживача Слуцкий виходив з рівняння $S=p_1x_1+...+p_nx_n$, де S - дохід індивіда, p_i - ціна на благо і x_1, ..., x_n - куплена індивідом кількість благ.

Завдання збігається до того, щоб знайти похідні обсягів споживаних благ за цінами і доходом, тобто визначити через коефіцієнти еластичності попиту перші та другі похідні функції корисності при вказаному вище обмеженні.

Основний висновок Слуцкого в аналізі попиту можна представити в наступній формі. Якщо через n позначити кількість продуктів, x_i - суму попиту, p_i - ціни ($i \in 1{:}n$), M - величину доходу, то, при даному напрямі зміни цін (dp_i) для підтримки рівноваги компенсуюча зміна доходу (dM) складе:

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

У такому разі зміни попиту приймуть вигляд:

$$dx_i = \sum_{j=1}^{n} \frac{\partial x_i}{\partial p_j} dp_j + \frac{\partial x_i}{\partial M} dM \ ,$$

або

$$dx_i = \sum_{j=1}^{n} \left(\frac{\partial x_i}{\partial p_j} + x_r \frac{\partial x_i}{\partial p_j} \right) dp_j \ .$$

Тут ясно виражені залежності коливань попиту для компенсації змін ціни (вираз в дужках - ефект заміщення).

Пізніше було підтверджено основне правило стійкості бюджету: споживач одержить максимум задоволення, коли співвідношення цін стане рівним співвідношенню граничних корисностей кожних двох товарів, що входять до бюджету споживача.

Згідно цієї моделі, вся множина споживчих наборів (векторів можливого споживання x) впорядкована перевагою, що є у споживача. У даному контексті x -= (x_1, x_2,..., x_n), де n - кількість споживчих благ. Передбачається, що якщо набір x переважніший, чим набір y, то значення функції корисності U для вектора x має бути більшим, ніж для вектора y: U(x)> U(y). Відповідно до кейнсианськой моделі поведінки споживача на ринку споживчих благ зумовлено прагненням одержати максимум корисності. При заданих цінах p = (p_1, p_2,..., p_n) і бюджеті (наміченій витраті грошей) M функції x_i(p, M) попиту на товар і повинні бути рішеннями оптимізаційної задачі:

$$U(x) \rightarrow \max, \ x \geq 0, \ (p, x) \leq M.$$

Криві байдужості (або поверхні, гіперповерхні), що задаються співвідношенням U(x)= const не залежать від монотонного перетворення F(U) функції корисності: V(x)= F(U(x)) F'>0. Функція корисності, визначена тільки з точністю до своїх поверхонь байдужості, називається функцією порядкової (ординальной) корисності. Такі функції корисності можуть використовуватися при описі та прогнозуванні споживчої поведінки у ряді простих випадків, але не годяться для багатьох завдань розподілу благ між соціальними групами або завдань динамічної оптимізації, що викликало з боку ряду економістів критику всієї теорії корисності.

Якщо перевагами споживача впорядковані не тільки пари наборів x, y, але і зміни наборів x→x', y→y', то функція корисності визначається з точністю до позитивного лінійного перетворення F: V(x)= F(U(x)= AU(x)+B, A>0 (за певних умов щодо цих двох упорядкованостей). При цьому ця функція корисності є індикатором як наборів x,x, так і переходів x→x', y→y'. Таку функцію корисності природно називати інтервальною, відповідно до термінології математичної теорії вимірювань, оскільки відносини різниць шкальних значень корисності не змінюються при припустимому перетворенні. Для споживача зв'язок між корисністю благ U(x) та корисністю грошей v(s) може даватися формулою: v(s)= U(x(s)), де x(s) - функція попиту від доходу. В цьому випадку гранична корисність доходу (множник Лагранжа на бюджетне обмеження в даному завданні співпадає з похідною функції v(s) в точці s.

У економічному аналізі як функції корисності часто використовують лінійні, квадратичні, логарифмічні:

$$U(x) = \sum_i a_i \ln(x_i + b_i)$$

функції CES - з постійною еластичністю заміщення

$$U(x) = \left(\sum_i a_i x_i^{-b} \right)^{\frac{m}{b}} \text{ та ін.}$$

У прикладних дослідженнях попиту та споживання набула поширення лінійна модель Стоуна-Джірі, цільова функція якої задається у формі

$$u(x) = \sum_i h_i \ln(x_i + c_i) \,,$$

де h_i - коефіцієнт, що задає пріоритет споживання блага i, c_i - мінімальний обсяг його споживання.

11.4. Моделювання виробничих можливостей

Під виробництвом в найзагальнішому випадку розуміють з'єднання факторів виробництва для здійснення випуску товарів і послуг відповідно до наявної технології. Будь-яке конкретне співвідношення між витратами та випуском, яке технологічно можливо на даному підприємстві, називають виробничим процесом. Множину виробничих процесів називають множиною виробничих можливостей, або технологічною множиною підприємства. Виробничі можливості формально відображають не досяжність необхідних для виробництва ресурсів, а стан технологічних знань про можливості перетворення одних продуктів на інші.

Для формалізації цих двох понять обирається простір продуктів - звичайний евклідовий простір R^n, де n - число найменувань продуктів. Кожен виробничий процес можна уявити парою невід'ємних векторів (x,x'), де x - вектор витрат x' - вектор випуску. Кожна компонента вектору x(x') є кількістю відповідного продукту, що витрачається (що випускається) в даному виробничому процесі (x,x'). Множина виробничих можливостей підприємства повністю описується переліком всіх таких пар векторів. Це уявлення зручно, коли аналіз ведеться в термінах величин, під якими розуміють запаси певних видів капіталу, змінних протягом виробничого циклу. У такому разі говорять, що технологічна множина задана в термінах запасів. У цих термінах описують моделі економічної динаміки з дискретним часом. Проте, якщо припустити, що продукти одночасно вводяться та виводяться з виробництва, то зручніше розглядати виробничий процес як вектор чистого випуску $x'-x\in R^n$. Він має характер потоку, тобто кількостей продуктів, що вводяться (якщо від'ємні) або виводяться (якщо додатні), в одиницю часу. Відповідно і технологічна множина - сукупність векторів чистого випуску - є підмножиною R^n. В цьому випадку говорять, що воно задано в термінах потоків. В термінах потоків описують моделі економічної рівноваги та моделі динаміки з безперервним часом.

Вивчення технології з економічної точки зору збігається до виявлення і визначення структурних характеристик технологічної множини. Розглянемо найбільш поширені стандартні характеристики. При цьому неважливо, в яких термінах (запасів або потоків) задано технологічну множину - і в тому, і в іншому випадку - це підмножина евклідова простору. Для визначеності можна вважати, що її задано в термінах потоків. Позначимо її через Y. Однією з самих елементарних та безумовних властивостей технології є відсутність «рогу достатку». Це означає, що неможливий виробничий процес, в якому в необмеженій кількості виробляються блага «з нічого», тобто без витрат. Вказану властивість можна подати у вигляді формули: $Y\cap R^n_+$ - обмежено.

Часто під час моделювання, коли є значний надлишок продукту і неможливо точно збалансувати пропозицію виробника та попит споживача, допускається вільне витрачання. Це означає, що припустимими є процеси, що характеризуються вилученням з системи будь-якої кількості продуктів. Формально це можна записати у вигляді: $y\in Y,\ x\le y\rightarrow x\in Y$.

Подільність виробництва припускає, що разом з кожним виробничим процесом $y\in Y$ можливий і виробничий процес $\lambda y\in Y$, $0\le\lambda\le 1$. Іншими словами, в будь-якому виробничому процесі можна пропорційно зменшити сукупні витрати та випуск. Це неможливо, якщо у виробничому процесі беруть участь неподільні продукти. Зокрема, подільність виробництва припускає бездіяльність (y=0), яка фізично можлива за будь-якої технології, але, разом з тим, припускає суверенітет виробника.

Аддитивність виробництва означає, що будь-які два виробничі процеси $y_1+y_2\in Y$ можуть функціонувати одночасно та незалежно (тобто не впливаючи один на одного), утворюючи новий виробничий процес $y_1+y_2\in Y$. Часто замість аддитивності вимагають опуклості технології. Це та сама подільність виробництва, але що має місце між двома будь-якими виробничими процесами, тобто $y_1+y_2\in Y\rightarrow\alpha y_1+(l-\alpha)y_2\in Y$ для будь-якого $\alpha\in[0,1]$.

Технологічна множина окремих підприємств Y_i в сумі складає сукупну технологічну множину економіки ΣY_i. Основне завдання економіки у сфері виробництва - це розподіл його за окремими підприємствами, називаний розподілом виробництва. Він здійснюється вибором ефективних процесів на підприємствах y_i, які в сумі дають сукупний виробничий процес (ΣY_i). Під ефективним виробничим процесом розуміють такий вектор $y \in Y$, що не існує іншого вектору $x \in Y$ $x>y$. Іншими словами, виробничий процес ефективний, якщо неможливо виробляти більше продуктів з меншими витратами. Іноді ефективний виробничий процес розуміють як оптимальний за Парето.

Для існування ефективних векторів потрібна замкнутість технологічної множини. Якщо певний вектор витрат та випуску можна за будь-яким ступенем точності апроксимувати технологічно можливим виробничим процесом, то і сам вектор є технологічно можливим.

Коли ціни p відіграють визначальну роль в економічній ситуації, то будь-який виробничий процес y одержує свою оцінку - прибуток py, що дорівнює перевищенню вартості випуску над вартістю витрат у виробничому процесі y. Він може бути від'ємним для деяких y. Тоді вибір ефективного виробничого процесу в кожному Y здійснюється, виходячи з принципу максимізації прибутку, тобто з рішення задачі:

$$\max_{y \in Y} py,$$

де $p \in R^n$ - вектор цін.

До цього спонукають виробника суб'єкти управління, що розпоряджаються прибутком підприємства. Відповідність Y_i, яке значенням цін p ставить у відповідність множину найбільш прибуткових векторів $Y_i(p)$ в Y_i, називається пропозицією виробника i; $\Sigma Y_i(p)$ - сукупна пропозиція в економіці. Для встановлення економічної рівноваги сукупна пропозиція має бути в певному розумінні безперервною в своїй області визначення. Це забезпечується вимогою опуклості, замкнутості Y_i та обмеженості $Y_i \cap R^n_+$.

Якщо виробництво ділимо, аддитивно та замкнено, то технологічна множина є опуклим, замкнутим конусом. Ця властивість технології робить її виключно зручною для математичного аналізу виробництва. Для прикладу розглянемо опис технології в трактуванні В. Леонтьева. Мабуть, цей найпростіший опис технології. Вивчаючи структуру національної економіки (США), Леонтьев подав її у вигляді галузей, кожна з яких виробляє єдиний продукт, витрачаючи інші в певних фіксованих співвідношеннях: a_{ij} $j \in 1{:}n$ - кількість продукту i, що витрачається для виробництва одиниці продукту галузі j. Для виробництва x_j одиниць продукту j потрібно $a_{ij}x_j$ одиниць продукту i. Отже, технологічну множину галузі j можна подати у вигляді променю:

$$\begin{pmatrix} -a_{1j} \\ \vdots \\ 1-a_{jj} \\ \vdots \\ -a_{nj} \end{pmatrix} x_j, \quad x_j \ge 0 .$$

Керівнику такої галузі залишається лише обрати інтенсивність процесу x_j. Припущення про постійність технологічного коефіцієнта a_{ij}, тобто про єдиність або незамінність даного виробничого процесу в галузі, має під собою певну підставу. Припустимо, що в галузях є альтернативні виробничі процеси, що виробляють той самий продукт і що витрачають певний фіксований ресурс, наприклад, працю. Позначимо замкнутий технологічний конус галузі j через Y_j. Тоді так звана теорема про незамінність стверджує: у кожній з множин Y_j можна виділити такий виробничий процес y^*_j, що сумісне функціонування цих виділених процесів, тобто $\Sigma y_j{}^*x_j$, $x_j \geq 0$ дозволяє одержати всі ефективні вектори в опуклому конусі сукупної технологічної множини ΣY_i. Отже, використання тільки цих виділених процесів y^*_j у кожній галузі j не призводить до втрати жодного ефективного виробничого процесу в національній економіці.

Якщо можливо сумісне виробництво, тобто випуск кількох продуктів в одному виробничому процесі, то теорема про незамінність не має місця. У загальному випадку конічні технології представляють особливий інтерес з погляду аналізу економічного розвитку. З'явилися цілі напрями досліджень, пов'язані з конічними технологіями: існування променів максимально збалансованого зростання, магістральні властивості таких променів, ефективні траєкторії, характеристики ефективних траєкторій, асимптотичні властивості.

Контрольні запитання

1. Охарактеризуйте області застосування мікроекономічних моделей та їх відмінність від макромоделей.
2. Охарактеризуйте зміст чотирьох етапів моделювання попиту.
3. Яким вимогам мають задовольняти фактори, що вводяться в модель? Який з цих факторів є найбільш істотним?
4. Перелічіть фактори, що роблять найбільший вплив на структуру попиту.
5. Наведіть лінійну модель споживання та охарактеризуйте її складові.
6. Поясніть, що таке функція корисності.
7. Які змістовні висновки можна одержати з аналізу функції корисності?
8. Поясніть, що таке пропозиція виробника та сукупна пропозиція. Яку властивість повинна мати сукупна пропозиція для встановлення економічної рівноваги?

СПИСОК ДЖЕРЕЛ

1. Ачкасов А. Є. Конспект лекцій з курсу «Економіко-математичне моделювання» (для студентів 3 курсу заочної форми навчання бакалаврів за галуззю знань 0305 «Економіка і підприємництво», напрями підготовки 6.030504 «Економіка підприємства», 6.030509 «Облік і аудит») / А. Є. Ачкасов, О. О. Воронков; Харк. нац. акад. міськ. госп-ва. – Х.: ХНАМГ, 2011.– 204 с.

2. Вітлінський В. В., Наконечний С. І., Терещенко Т. О. Математичне програмування. - К.: КНЕУ, 2001.

3. Кузнецов Ю. Н., Кузубов В. А., Волощенко А. В. Математическое программирование. - М.:Высш.школа,1980. - 240с.

4. Таха Х. А. Введение в исследование операций. - М.: Изд.дом «Вильямс», 2005.

5. Исследование операций в экономике: Уч. пособие для вузов / Н.Ш. Кремер, Б.А. Путко, И.М. Тришин, М.Н. Фридман./ Под ред. проф. Н.Ш. Кремера. - М.: ЮНИТИ, 2003. - 407 с.

6. Трояновский В. М. Математическое моделирование в менеджменте: Учебное пособие. – М.: ЮНИТИ, 2000.

7. Математические модели трансформационной экономики: Учебное пособие / Клебанова Т. С., Раевнева Е. В., Стрижиченко К. А. и др.. – Х.: ИД «ИНЖЭК», 2006. – 280 с.

8. Крушевский Ф. В. Справочник по экономико-математическим методам и моделям. – М.: Экономика, 1982. – 196 с.

9. Власов М. П., Моделирование экономических процессов / М. П. Власов, П. Д. Шимко. – Ростов н/Д: Феникс, 2005. – 409.

10. Математические модели трансормационной экономики: Учебное пособие / Клебанова Т. С., Раевнева Е. В. и др. – Х.: ИДЖ ИНЖЭК, 2006. – 280 с.

11. Цифровий репозиторій ХНАМГ: http://eprints.ksame.kharkov.ua

12. Національна парламентська бібліотека України: http://ukrlibrary.org

НАВЧАЛЬНЕ ВИДАННЯ

АЧКАСОВ Анатолій Єгорович
ВОРОНКОВ Олексій Олександрович

Конспект лекцій
з курсу
«ДОСЛІДЖЕННЯ ОПЕРАЦІЙ»
(для студентів галузі знань 0306 – «Менеджмент і адміністрування» напряму 6.030601 – «Менеджмент» заочної форми навчання)

Відповідальний за випуск *А. Є. Ачкасов*

За авторською редакцією

Комп'ютерне верстання *К. А. Алексанян*

План 2012, поз. 185Л

Підп. до друку 02.10.2012 Формат 60х84/16
Друк на ризографі. Ум. друк. арк. 7,7
Зам. № Тираж 50 пр.

Видавець і виготовлювач:
Харківська національна академія міського господарства,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@ksame.kharkov.ua
Свідоцтво суб’єкта видавничої справи:
ДК № 4064 від 12.05.2011р.