

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ И МЕТОДЫ ОПТИМИЗАЦИИ В ЭКОНОМИКЕ

Часть 1. Лекционный курс

Учебное пособие

ТОМСК 2019

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Томский государственный университет систем управления и
радиоэлектроники

Факультет систем управления
Кафедра автоматизированных систем управления

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ И МЕТОДЫ ОПТИМИЗАЦИИ

Часть 1. Лекционный курс

Учебное пособие

УДК 330.4(076.5)

**Исследование операций и методы оптимизации. Часть 1.
Лекционный курс**

Составитель А.А. Мицель.

ТОМСК: Изд. ТУСУР, 2019. – 167 с.

В пособии описаны различные модели математического программирования – линейного, нелинейного, квадратичного, целочисленного и динамического программирования, задачи многокритериальной оптимизации, а также методы безусловной оптимизации функций многих переменных.

Пособие подготовлено для студентов, обучающихся по специальности 09.03.03 – прикладная информатика (бакалавриат)

Содержание

| | |
|--|----|
| Введение | 7 |
| Тема 1. Линейное программирование | 12 |
| 1.1. Постановка задачи линейного программирования | 12 |
| 1.2 Примеры задач линейного программирования | 13 |
| Вопросы для самопроверки | 19 |
| Тема 2. Решение задач линейного программирования | 20 |
| 2.1 Графический метод решения задач линейного программирования | 20 |
| 2.2 Формы записи задач (ЗЛП) | 21 |
| 2.3 Основы симплекс-метода | 24 |
| 2.4. Алгоритм симплекс метода | 26 |
| 2.5 Поиск начального базиса | 31 |
| 2.5.1. Метод симплексного преобразования | 32 |
| 2.5.2. Метод искусственного базиса | 33 |
| Вопросы для самопроверки | 37 |
| Тема 3. Двойственная задача линейного программирования | 38 |
| 3.1 Постановка двойственной задачи | 38 |
| 3.2 Свойства взаимно двойственных задач | 39 |
| 3.3 Теоремы двойственности | 40 |
| Вопросы для самопроверки | 45 |
| Тема 4. Транспортная задача | 46 |
| 4.1 Экономико – математическая модель транспортной задачи | 46 |
| 4.2 Решение транспортной задачи симплексным методом | 51 |
| 4.3 Первоначальное закрепление потребителей за поставщиками | 52 |
| 4.4 Метод потенциалов | 54 |
| 4.5 Улучшение оптимального плана перевозок (циклы перераспределения) | 56 |
| 4.6 Открытая модель транспортной задачи | 58 |
| Вопросы для самопроверки | 60 |
| Тема 5. Целочисленное программирование | 61 |
| 5.1 Графический метод решения ЗЦП | 61 |
| 5.2 Метод Гомори (МГ) | 62 |
| 5.2.1 Алгоритм МГ с использованием СМ | 62 |
| 5.2.2. Решение частично целочисленных задач методом Гомори | 65 |
| 5.3. Метод ветвей и границ | 67 |
| 5.3.1 Алгоритм МВГ | 71 |
| 5.4 Задача о назначениях | 73 |
| 5.5 Задача о коммивояжере | 74 |
| 5.6. Венгерский метод | 75 |
| Вопросы для самопроверки | 78 |

| | |
|--|-----|
| Тема 6. Задачи многокритериальной оптимизации | 79 |
| 6.1. Метод уступок | 80 |
| 6.2. Метод справедливого компромисса | 83 |
| Вопросы для самопроверки | 84 |
| Тема 7. Методы оптимизации функций | 85 |
| 7.1. Основные понятия и определения | 85 |
| 7.2. Классификация задач оптимизации | 88 |
| 7.3. Необходимые и достаточные условия существования экстремума | 89 |
| 7.3.1 Скалярный случай $x \in R^1$ | 89 |
| 7.3.2 Векторный случай | 89 |
| 7.3.3 Минимизация при ограничениях | 90 |
| 7.4. Критерии останова | 90 |
| 7.5. Характеристики алгоритмов оптимизации | 91 |
| Вопросы для самопроверки | 92 |
| Тема 8. Методы поиска экстремумов функции одной переменной | 93 |
| 8.1 Прямые методы оптимизации | 93 |
| 8.1.1 Метод равномерного поиска (МРП) | 93 |
| 8.1.2 Метод деления отрезка пополам (МДОП) | 93 |
| 8.1.3 Метод Фибоначчи (МФ) | 94 |
| 8.1.4 Метод золотого сечения (МЗС) | 95 |
| 8.2 Сравнение прямых методов оптимизации | 96 |
| 8.3 Полиномиальная аппроксимация и методы точечного оценивания | 97 |
| 8.3.1 Квадратичная аппроксимация | 97 |
| 8.3.1 Метод Пауэлла | 99 |
| 8.4 Методы с использованием производных | 101 |
| 8.4.1 Метод Ньютона-Рафсона | 101 |
| 8.4.2 Метод средней точки (поиск Больцано) | 102 |
| 8.4.3 Другие методы поиска экстремума функций | 103 |
| 8.4.4 Метод оптимизации с использованием кубичной аппроксимации | 104 |
| 8.5 Сравнение методов одномерной оптимизации | 106 |
| Вопросы для самопроверки | 106 |
| Тема 9. Поиск экстремумов функции нескольких переменных (безусловная оптимизация) | 108 |
| 9.1 Классификация методов безусловной оптимизации | 108 |
| 9.2. Методы прямого поиска | 108 |
| 9.2.1 Симплексный метод | 109 |
| 9.2.2 Метод Хука-Дживса | 111 |
| 9.3 Градиентные методы | 113 |
| 9.3.1 Метод сопряженных направлений | 114 |
| 9.3.2 Метод наискорейшего спуска (метод Коши) | 118 |
| 9.3.3 Метод Ньютона (МН) | 120 |

| | |
|---|-----|
| 9.3.4 Модифицированный метод Ньютона | 120 |
| 9.3.5 Метод Флетчера–Ривза (МФР) | 121 |
| 9.3.6 Вариант Полака-Рибьера (МПР) | 122 |
| 9.4 Квазиньютоновские методы (КМ) (методы с переменной метрикой) | 123 |
| 9.4.1 Метод Дэвидона–Флетчера–Пауэлла (ДФП) | 124 |
| Вопросы для самопроверки | 125 |
| Тема 10. Нелинейное программирование | 126 |
| 10.1. Задачи с ограничениями в виде равенств | 126 |
| 10.1.1 Метод замены переменных (МЗП) | 126 |
| 10.1.2 Метод множителей Лагранжа (ММЛ) | 127 |
| 10.2. Необходимые и достаточные условия оптимальности задач с ограничениями общего вида | 129 |
| Вопросы для самопроверки | 131 |
| Тема 11. Методы штрафов | 132 |
| 11.1. Общая схема метода штрафов | 132 |
| 11.2. Основные типы штрафов | 132 |
| 11.2.1 Квадратичный штраф | 133 |
| 11.2.1 Бесконечный барьер | 133 |
| 11.2.3 Логарифмический штраф | 134 |
| 11.2.4 Штраф типа обратной функции | 134 |
| 11.2.5 Штраф типа квадрата срезки | 135 |
| 11.3. Примеры использования штрафов | 137 |
| Вопросы для самопроверки | 140 |
| Тема 12. Квадратичное программирование | 141 |
| 12.1. Задача квадратичного программирования | 141 |
| 12.2 Задача выбора портфеля ценных бумаг | 141 |
| 12.3. Условие Куна-Таккера для ЗКП | 144 |
| 12.4 Решение ЗКП методом симплексного преобразования коэффициентов уравнений | 145 |
| 12.5 Метод решения ЗКП с помощью искусственного базиса | 146 |
| Вопросы для самопроверки | 147 |
| Тема 13. Модели динамического программирования | 148 |
| 13.1. Общая постановка задачи динамического программирования | 148 |
| 13.2. Принцип оптимальности и уравнения Беллмана | 150 |
| 13.3. Задача о распределении средств между предприятиями | 154 |
| 13.4. Задача об оптимальном распределении ресурсов между отраслями на N лет | 158 |
| 13.5. Задача о замене оборудования | 161 |
| Вопросы для самопроверки | 166 |
| Литература | 167 |

Введение

Исследование операций – научная дисциплина прикладного направления кибернетики, занимающаяся разработкой и практическим применением методов наиболее эффективного управления различными организационными (в том числе экономическими) системами. Предмет исследования операций очень широк. Теоретический аспект исследования операций состоит в построении и исследовании математических моделей принятия оптимальных решений.

Приведем некоторые определения.

Операция – любое управляемое мероприятие, направленное на достижение цели. Результат операции зависит от способа ее проведения, т.е. от выбора некоторых параметров.

Всякий определенный выбор параметров называется *решением*. *Оптимальными* считают те решения, которые по тем или иным соображениям предпочтительней других.

Модель операции. Для применения количественных методов исследования операций необходимо построить *математическую модель операции*. При построении модели операция, как правило, упрощается, схематизируется и схема операции описывается с помощью того или иного математического аппарата. Модель *операции* — это достаточно точное описание операции с помощью математического аппарата (различного рода функций, уравнений, систем уравнений и неравенств и т.п.). Составление модели операции требует понимания сущности описываемого явления и знания математического аппарата.

Эффективность операции — степень ее приспособленности к выполнению задачи — количественно выражается в виде критерия эффективности — целевой функции. Например, в задаче об использовании ресурсов критерий эффективности — прибыль от реализации произведенной продукции, которую нужно максимизировать, в транспортной задаче — суммарные затраты на перевозку грузов от поставщиков к потребителям, которые нужно минимизировать. Выбор критерия эффективности определяет практическую ценность исследования.

Общая постановка задачи исследования операции. В дальнейшем важно усвоить методологию построения моделей задач исследования операции. Все факторы, входящие в описание операции, можно разделить на две группы:

- *постоянные факторы* (условия проведения операции), на которые мы влиять не можем. Обозначим их через $\alpha_1, \alpha_2, \dots$;
- *зависимые факторы* (элементы решения) x_1, x_2, \dots , которые в известных пределах мы можем выбрать по своему усмотрению.

Например, в задаче об использовании ресурсов к постоянным факторам следует отнести запасы ресурсов каждого вида, производственную матрицу, элементы которой определяют расход сырья каждого вида на единицу выпускаемой продукции каждого вида. Элементы решения — план выпуска продукции каждого вида.

Критерий эффективности, выражаемый некоторой функцией, называется *целевой*, зависит от факторов обеих групп, поэтому целевую функцию Z можно записать в виде

$$Z = f(x_1, x_2, \dots, \alpha_1, \alpha_2, \dots).$$

Выделяют следующие классы моделей исследования операций.

- 1) *класс оптимизационных моделей.* Задачи этого класса возникают при попытке оптимизировать планирование и управление сложными экономическими

системами. Оптимизационную задачу можно сформулировать в общем виде: *найти переменные* x_1, x_2, \dots, x_n , *удовлетворяющие системе неравенств*

$$\varphi_i(x_1, x_2, \dots, x_n) \leq b_i, \quad i = 1, 2, \dots, m \quad (1)$$

и обращающие в максимум (или минимум) целевую функцию

$$Z = f(x_1, x_2, \dots, x_n) \rightarrow \max(\min). \quad (2)$$

Как известно, упорядоченная совокупность значений n переменных x_1, x_2, \dots, x_n представляется точкой n -мерного пространства. В дальнейшем эту точку будем обозначать $X = (x_1, x_2, \dots, x_n)$, а само оптимальное решение $X^* = (x_1^*, x_2^*, \dots, x_n^*)$.

Рассмотрим еще одну, характерную для исследования операций задачу — *классическую задачу потребления*, имеющую важное значение в экономическом анализе.

Пусть имеется n видов товаров и услуг, количества которых (в натуральных единицах) x_1, x_2, \dots, x_n по ценам соответственно p_1, p_2, \dots, p_n за единицу. Суммарная

стоимость этих товаров и услуг составляет $\sum_{i=1}^n p_i x_i$.

Уровень потребления Z может быть выражен некоторой функцией $Z = f(x_1, x_2, \dots, x_n)$, которую называют функцией полезности. Необходимо найти такой набор товаров и услуг x_1, x_2, \dots, x_n при данной величине доходов I , чтобы обеспечить максимальный уровень потребления, т.е.

$$Z = f(x_1, x_2, \dots, x_n) \rightarrow \max \quad (3)$$

при условии

$$\sum_{i=1}^n p_i x_i \leq I, \quad (4)$$

$$x_i \geq 0 \quad (i = 1, 2, \dots, n). \quad (5)$$

Решения этой задачи, зависящие от цен p_1, p_2, \dots, p_n и величины дохода I , называются *функциями спроса*.

Очевидно, что рассмотренная задача потребления (3) – (5) является частным случаем сформулированной выше задачи (1) – (2) на определение экстремума функции n переменных при некоторых ограничениях, т.е. задачей на условный экстремум.

Если критерий эффективности $Z = f(x_1, x_2, \dots, x_n)$ (2) представляет линейную функцию, а функции $\varphi_i(x_1, x_2, \dots, x_n) \leq b_i$ в системе ограничений (1) также линейны, то такая задача является задачей *линейного программирования*. Если, исходя из содержательного смысла, ее решения должны быть целыми числами, то эта задача *целочисленного линейного программирования*. Если критерий эффективности и (или) система ограничений задаются нелинейными функциями, то имеем, задачу *нелинейного программирования*. В частности, если указанные функции обладают свойствами выпуклости, то полученная задача является задачей *выпуклого программирования*.

Если в задаче математического программирования имеется переменная времени и критерий эффективности (2) выражается не в явном виде как функция

переменных, а косвенно — через уравнения, описывающие протекание операций во времени, то такая задача является задачей *динамического программирования*.

Если критерий эффективности (2) и система ограничений (1) задаются функциями вида $cx_1^{\alpha_1} x_2^{\alpha_2} \dots x_n^{\alpha_n}$, то имеем задачу *геометрического программирования*. Если функции f и (или) Φ_i в выражениях (2) и (1) зависят от параметров, то получаем задачу *параметрического программирования*, если эти функции носят случайный характер — задачу *стохастического программирования*. Если точный оптимум найти алгоритмическим путем невозможно из-за чрезмерно большого числа вариантов решения, то прибегают к методам *эвристического программирования*, позволяющим существенно сократить просматриваемое число вариантов и найти если не оптимальное, то достаточно хорошее, удовлетворительное с точки зрения практики, решение.

Из перечисленных методов математического программирования наиболее распространенным и разработанным является линейное программирование. В его рамки укладывается широкий круг задач исследования операций.

Пример В.1. Рассмотрим пример из области оптимального проектирования. Пусть коробка изготавливается из прямоугольного листа материала размером $a \times b$, $a < b$. Для этого из четырех углов прямоугольника вырезаются квадраты со стороной x и материал сгибается вдоль линий, отмеченных на рис. В.1 штриховыми линиями.

В результате получается коробка с основанием в виде прямоугольника размером $(a - 2x)(b - 2x)$ и высотой x . Здесь стратегия $x \in X = (0, a/2)$ может быть оценена целевой функцией $V(x) = x(a - 2x)(b - 2x)$ - объемом коробки. Максимум функции $V(x)$ на множестве X достигается в точке

$$x^* = \frac{a + b - \sqrt{a^2 - ab + b^2}}{6}.$$

2) *Задачи сетевого планирования и управления* рассматривают соотношения между сроками окончания крупного комплекса операций (работ) и моментами начала всех операций комплекса. Эти задачи состоят в нахождении минимальных продолжительностей комплекса операций, оптимального соотношения величин стоимости и сроков их выполнения.

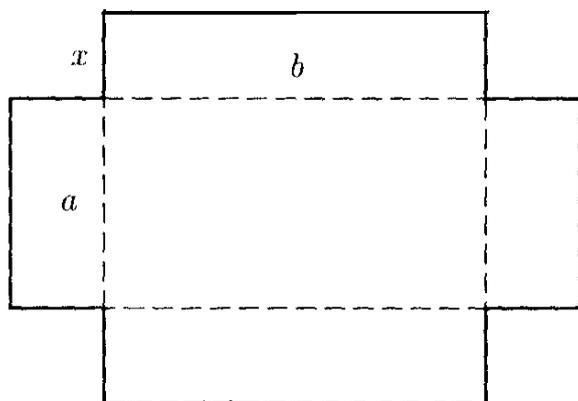


Рис. В1

3) *Задачи массового обслуживания* посвящены изучению и анализу систем обслуживания с очередями заявок или требований и состоят в определении

показателей эффективности работы систем, их оптимальных характеристик, например, в определении числа каналов обслуживания, времени обслуживания и т.п.

4) *Задачи управления запасами* состоят в отыскании оптимальных значений уровня запасов (точки заказа) и размера заказа. Особенность таких задач заключается в том, что с увеличением уровня, с одной стороны, увеличиваются затраты на их хранение, с другой стороны, уменьшаются потери вследствие возможного дефицита запасаемого продукта.

5) *Задачи распределения ресурсов* возникают при определенном наборе операций (работ), которые необходимо выполнять при ограниченных наличных ресурсах, и требуется найти оптимальные распределения ресурсов между операциями или состав операций.

6) *Задачи ремонта и замены оборудования* актуальны в связи с износом и старением оборудования и необходимостью его замены с течением времени. Задачи сводятся к определению оптимальных сроков, числа профилактических ремонтов и проверок, а также моментов замены оборудования модернизированным.

7) *Задачи составления расписания (календарного планирования)* состоят в определении оптимальной очередности выполнения операций (например, обработки деталей) на различных видах оборудования.

8) *Задачи планировки и размещения* состоят в определении оптимального числа и места размещения новых объектов с учетом их взаимодействия с существующими объектами и между собой.

9) *Задачи выбора маршрута, или сетевые задачи*, чаще всего встречаются при исследовании разнообразных задач на транспорте и в системе связи и состоят в определении наиболее экономных маршрутов.

10) Среди моделей исследования операций особо выделяются модели принятия оптимальных решений в конфликтных ситуациях, изучаемые *теорией игр*. К конфликтным ситуациям, в которых сталкиваются интересы двух (или более) сторон, преследующих разные цели, можно отнести ряд ситуаций в области экономики, права, военного дела и т. п. В задачах теории игр необходимо выработать рекомендации по разумному поведению участников конфликта, определить их оптимальные стратегии.

На практике в большинстве случаев успех операции оценивается не по одному, а сразу по нескольким критериям, один которых следует максимизировать, другие — минимизировать. Математический аппарат может принести пользу и в случаях *многокритериальных задач исследования операции*, по крайней мере, помочь отбросить заведомо неудачные варианты решений.

Для того чтобы из множества критериев, в том числе и противоречащих друг другу (например, прибыль и расход), выбрать целевую функцию, необходимо установить *приоритет* критериев. Обозначим $f_1(x), f_2(x), \dots, f_n(x)$ (здесь x — условный аргумент). Пусть они расположены в порядке убывания приоритетов. В зависимости от определенных условий возможны в основном два варианта:

- в качестве целевой функции выбирается критерий $f_1(x)$, обладающий наиболее высоким приоритетом;
- рассматривается комбинация

$$f(x) = \omega_1 f_1(x) + \omega_2 f_2(x) + \dots + \omega_n f_n(x), \quad (6)$$

где $\omega_1, \omega_2, \dots, \omega_n$ — некоторые коэффициенты (веса).

Величина $f(x)$, учитывающая в определенной степени все критерии, выбирается в качестве целевой функции.

В условиях определенности ω_i — числа, $f_i(x)$ — функции. В условиях неопределенности $f_i(x)$ могут оказаться случайны и вместо $f(x)$ в качестве целевой функции следует рассматривать математическое ожидание суммы (6).

Попытка сведения многокритериальной задачи к задаче с одним критерием эффективности (целевой функцией) в большинстве случаев не дает удовлетворительных результатов. Другой путь состоит в отбрасывании ("выбраковке") из множества допустимых решений заведомо неудачных решений, уступающих по другим *всем критериям*. В результате такой процедуры остаются так называемые *эффективные* (или "*паретовские*") решения, множество которых обычно существенно меньше исходного. А окончательный выбор "компромиссного" решения (не оптимального по всем критериям, которого, как правило, не существует, а *приемлемого* по этим критериям) остается за человеком — лицом, принимающим решение.

В создание современного математического аппарата и развитие многих направлений исследования операций большой вклад внесли российские ученые Л.В. Канторович, Н.П. Бусленко, Е.С. Вентцель, Н.Н. Воробьев, Н.Н. Моисеев, Д.Б. Юдин и многие другие. Особо следует отметить роль академика Л.В. Канторовича, который в 1939 г., занявшись планированием работы агрегатов фанерной фабрики, решил несколько задач: о наилучшей загрузке оборудования, о раскрое материалов с наименьшими потерями, о распределении грузов по нескольким видам транспорта и др. Л.В. Канторович сформулировал новый класс условно-экстремальных задач и предложил универсальный метод их решения, положив начало новому направлению прикладной математики — линейному программированию.

Значительный вклад в формирование и развитие исследования операций внесли зарубежные ученые Р. Акоф, Р. Беллман, Г. Данциг, Г. Кун, Дж. Нейман, Т. Саати, Р. Черчмен, А. Кофман и др.

Вопросы для самопроверки

1. Что такое операция?
2. Что такое эффективность операции?
3. Что понимают под критерием эффективности операции?
4. Перечислите классы моделей исследования операций

Тема 1. Линейное программирование

1.1. Постановка задачи линейного программирования

Значительная часть задач принятия решения — это задачи распределения ресурсов между объектами.

Пусть имеется m видов ресурсов. Наличие каждого i -го вида ресурса составляет b_i ($i = 1, \dots, m$) в соответствующих единицах измерения. Эти ресурсы предназначены для производства n видов продукции. Для выпуска единицы j -го вида продукции необходимо a_{ij} единиц i -го вида ресурса. Требуется определить, какого вида и сколько продукции следует произвести, чтобы такой выпуск был наилучшим для принятого критерия оптимальности.

Обозначим через x_j количество выпускаемой j -го вида продукции. Тогда для i -го вида ресурса можно записать

$$\sum_{j=1}^n a_{ij} x_j \leq b_i, \quad i = 1, \dots, m \quad (1.1)$$

где левая часть неравенства выражает потребность в ресурсе i -го вида, правая — располагаемое количество этого ресурса.

Если номенклатуру продукции ограничить предельными значениями объемов производства и продаж, то запишутся следующие граничные условия

$$N_j^L \leq x_j \leq N_j^U, \quad j = 1, \dots, n \quad (1.2)$$

где N_j^L, N_j^U — соответственно минимально- и максимально-допустимые объемы производства и продаж продукции j -го вида.

В зависимость (1) можно ввести дополнительные переменные. Тогда

$$\sum_{j=1}^n a_{ij} x_j + y_i = b_i, \quad i = 1, \dots, m \quad (1.3)$$

В этой системе уравнений дополнительные переменные означают разность между располагаемым ресурсом и потребным и, следовательно, равны недоиспользуемым величинам ресурсов, то есть это **резервы** каждого вида ресурсов.

В реальных задачах суммарное количество основных x_j ($j = 1, \dots, n$) и дополнительных y_i ($i = 1, \dots, m$) переменных всегда больше, чем число зависимостей m , поэтому система (1) имеет бесчисленное множество решений. Из этого бесчисленного множества следует выбрать одно — оптимальное, соответствующее критерию — цели решения задачи.

Цель задачи распределения ресурсов устанавливается какой-либо одной из двух взаимоисключающих постановок:

- 1) при заданных ресурсах максимизировать получаемый результат;
- 2) при заданном результате минимизировать потребные ресурсы.

Первая постановка аналитически запишется:

$$f_1(x) = \sum_{j=1}^n c_j x_j \rightarrow \max_x \quad (1.4)$$

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

$$N_j^L \leq x_j \leq N_j^U, \quad j = 1, \dots, n \quad (1.6)$$

где c_j — величина, показывающая, какой вклад в результат дает единица продукции j -го вида.

Решение задачи состоит в нахождении значений x_j , обеспечивающих при заданных ресурсах получение максимального результата.

Вторая постановка задачи будет иметь вид

$$f_2(y) = \sum_{i=1}^m y_i \rightarrow \max_y \quad (1.7)$$

$$\sum_{j=1}^n c_j x_j \geq C \quad (1.8)$$

$$\sum_{j=1}^n a_{ij} x_j + y_i = b_i, \quad i = 1, \dots, m \quad (1.9)$$

$$N_j^L \leq x_j \leq N_j^U, \quad j = 1, \dots, n \quad (1.10)$$

$$y_i \geq 0, \quad i = 1, \dots, m \quad (1.11)$$

где C — минимально допустимое значение потребного результата. Максимизация переменных y_i означает максимизацию резервов ресурсов, а значит минимизация потребных ресурсов.

Первую и вторую задачи, в которые переменные x_j и y_i входят в первой степени, то есть в виде линейных зависимостей, называют **задачами линейного программирования**.

Каждая задача линейного программирования содержит **целевую функцию** (1.4) или (1.7), **ограничения** (1.5), (6) или (1.8)-(1.11).

Для решения задач линейного программирования используют графический и аналитический методы.

1.2 Примеры задач линейного программирования

1. Задача об использовании ресурсов (задача планирования производства).

Для изготовления двух видов продукции P1 и P2 используют четыре вида ресурсов S1, S2, S3, S4. Запасы ресурсов, число единиц ресурсов, затрачиваемых на изготовление единицы продукции, приведены в табл. 1 (цифры условные).

Таблица 1.1

| Вид ресурса | Запас ресурса | Число единиц ресурсов, затрачиваемых на изготовление единицы продукции | |
|-------------|---------------|--|----|
| | | P1 | P2 |
| | | | |

| | | | |
|-------------------------|----|--------|--------|
| S1. | 18 | 1 | 3 |
| S2 | 16 | 2 | 1 |
| S3 | 5 | — | 1 |
| S4 | 21 | 3 | — |
| Стоимость ед. продукции | | 2 д.е. | 3 д.е. |

Необходимо составить такой план производства продукции, при котором прибыль от ее реализации будет максимальной.

Решение. Составим экономико-математическую модель задачи.

Обозначим x_1, x_2 — число единиц продукции соответственно P1 и P2, запланированных к производству. Для их изготовления (см. табл. 2.1) потребуется ($x_1 + 3x_2$) единиц ресурса S1, ($2x_1 + x_2$) единиц ресурса S2, (x_2) единиц ресурса S3 и ($3x_2$) единиц ресурса S4. Так как потребление ресурсов S1, S2, S3 и S4 не должно превышать их запасов, соответственно 18, 16, 5 и 21 единицы, то связь между потреблением ресурсов и их запасами выразится системой неравенств:

$$\begin{cases} x_1 + 3x_2 \leq 18, \\ 2x_1 + x_2 \leq 16, \\ x_2 \leq 5, \\ 3x_1 \leq 21 \end{cases} \quad (1.12)$$

По смыслу задачи переменные

$$x_1 \geq 0, \quad x_2 \geq 0 \quad (1.13)$$

Суммарная прибыль f составит $2x_1$ руб. от реализации продукции P1 и $3x_2$ руб. — от реализации продукции P2, т.е.

$$f = 2x_1 + 3x_2. \quad (1.14)$$

Итак, экономико-математическая модель задачи: найти такой план выпуска продукции $X = (x_1, x_2)$, удовлетворяющий системе (1.11) и условию (1.12), при котором функция (1.13) принимает **максимальное** значение.

Задачу легко обобщить на случай выпуска n видов продукции с использованием m видов ресурсов.

Обозначим x_j ($j=1, \dots, n$) — число единиц продукции Pj, запланированной к производству; b_i ($i=1, \dots, m$) — запас ресурсов Si, a_{ij} — число единиц ресурса Si затрачиваемого на изготовление единицы продукции Pj (числа a_{ij} часто называют технологическими коэффициентами); c_j — прибыль от реализации единицы продукции Pj.

Тогда экономико-математическая модель задачи об использовании ресурсов в общей постановке примет вид: найти такой план $X = (x_1, x_2, \dots, x_n)$ выпуска продукции, удовлетворяющий системе

Пример 1.1

Для изготовления брусьев длиной 1,2 м, 3 м и 5 м в соотношении 2:1:3 на распил поступают 195 бревен длиной 6 м. Определить план распила, обеспечивающий максимальное число комплектов. Составить экономико-математическую модель задачи.

Решение. Прежде всего определим всевозможные способы распила бревен, указав соответствующее число получаемых при этом брусьев (табл. 1.3).

Таблица 1.3.

| Способ распила j | Число получаемых брусьев длиной, м | | |
|-----------------------|------------------------------------|-----|-----|
| | 1,2 | 3,0 | 5,0 |
| 1 | 5 | — | — |
| 2 | 2 | 1 | — |
| 3 | — | 2 | — |
| 4 | — | — | 1 |

Обозначим: x_j — число бревен, распиленных j -ым способом ($j=1,2,3,4$); x — число комплектов брусьев.

Учитывая, что все бревна должны быть распилены, а число брусьев каждого размера должно удовлетворять условию комплектности, экономико-математическая модель задачи примет вид:

$$f = x \rightarrow \max$$

при ограничениях:

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 195, \\ 5x_1 + 2x_2 = 2x, \\ x_2 + 2x_3 = x, \\ x_4 = 3x, \\ x_j \geq 0 \quad (j=1,2,3,4) \end{cases}$$

5. Задача технического контроля

В отделе технического контроля (ОТК) некоторой фирмы работают контролеры 1-го и 2-го разрядов. Норма выработки ОТК за 8-ми часовой рабочий день составляет не менее 1800 изделий. Контролер 1-го разряда проверяет 25 изделий в час, причем не ошибается в 98 % случаев. Контролер 2-го разряда проверяет 15 изделий в час; его точность – 95 %.

Зарплата контролера К1 – 4 доллара в час, контроля К2 – 3 доллара в час. При каждой ошибке контролера фирма несет убыток в размере 2 долларов. Фирма может использовать не более 8 К1 и не более 10 К2. Определить оптимальный состав ОТК, при котором общие затраты на контроль будут минимальны.

Модель ОТК

- 1) пусть x_1, x_2 обозначают количество К1 и К2 соответственно
- 2) число контролеров каждого разряда ограничено, т.е.
 $x_1 \leq 8$ (разряд 1)
 $x_2 \leq 10$ (разряд 2)

Ежедневно необходимо проверять не менее 1800 изделий. Поэтому должно выполняться неравенство

$$8 \cdot 25 \cdot x_1 + 8 \cdot 15 \cdot x_2 = 200x_1 + 120x_2 \geq 1800$$

или $5x_1 + 3x_2 \geq 45$

3) При построении ЦФ следует иметь в виду, что расходы фирмы, связанные с контролем, включают две составляющие

- зарплату контролеров
- убытки, вызванные ошибками контролеров

Расходы на одного контролера К1 составляют

$$4_{\text{долл}} + 2 \cdot 25 \cdot 0.02 = 5_{\text{долл/час}}$$

Расходы на одного контролера К2 составляют

$$3_{\text{долл}} + 2 \cdot 15 \cdot 0.05 = 4.5_{\text{долл/час}}$$

Следовательно, ЦФ, выражающая ежедневные расходы на контроль, имеет вид

$$f(x_1, x_2) = 8(5x_1 + 4.5x_2) = 40x_1 + 36x_2$$

Экономико-математическая модель отдела технического контроля примет вид:

$$\min f(x) = 40x_1 + 36x_2$$

при ограничениях: $5x_1 + 3x_2 \geq 45$

$$0 \leq x_1 \leq 8; \quad 0 \leq x_2 \leq 10$$

Вопросы для самопроверки

1. Сформулируйте общую постановку задачи линейного программирования
2. Сформулируйте задачу планирования производства
3. Сформулируйте задачу составления рациона
4. Сформулируйте задачу о загрузке оборудования
5. Сформулируйте задачу о раскрое материалов
6. Сформулируйте задачу технического контроля

Тема 2. Решение задач линейного программирования

2.1 Графический метод решения задач линейного программирования

Графический метод применяют, когда число переменных не превосходит 2. На практике задачи такой размерности встречаются редко, однако графический метод хорошо иллюстрирует основные понятия, используемые при решении ЗЛП большой размерности. Графический метод рассмотрим на примере задачи 5 (задача технического контроля)

Пример 2.1

$$\begin{aligned} \min f &= 40x_1 + 36x_2 \\ 5x_1 + 3x_2 &\geq 45, \\ x_1 &\leq 8, \quad x_2 \leq 10, \\ x_1 &\geq 0, \quad x_2 \geq 0. \end{aligned}$$

В качестве первого шага решения следует определить допустимую область или область допустимых решений (ОДР).

Для изображения ОДР следует начертить графики всех ограничений. Все допустимые решения лежат в первом квадранте, т.к. $x_1, x_2 \geq 0$ (см. рис. 2.1)

ОДР является треугольник ABC (в общем случае многогранник M при $m > 2$), который содержит бесконечное число допустимых точек. Нужно найти точку, в которой $f(x) = \min$.

Если зафиксировать значение ЦФ $f = 40x_1 + 36x_2$, то соответствующие ему точки будут лежать на некоторой прямой. При изменении величины f прямая подвергается параллельному переносу. Рассмотрим прямые, соответствующие различным значениям f , имеющие с ОДР хотя бы одну общую точку. Положим $f_0 = 680$. Тогда прямая ЦФ пройдет через точку $B(8,10)$. При приближении прямой ЦФ к началу координат значение f уменьшается. Минимальное значение ЦФ достигается в узловой точке $A(8, 5/3)$.

Следовательно, $x^* = (8, 5/3)$; $f^* = 380$ - оптимальное решение задачи. Таким образом

$x^* = (8, 5/3)$ - оптимальный план.



Рис. 2.1

Итак, для оптимального режима работы ОТК необходимо использовать 8 контролеров 1-го разряда и 1,6 контролера 2-го разряда. Дробное значение $x_2 = 1.6$ соответствует ситуации, когда один из двух К2 используется в течении неполного рабочего дня. При недопустимости неполной загрузки контролеров дробные значения обычно округляются, получая приближенное оптимальное целочисленное решение $x_1 = 8, x_2 = 2$.

Итак, мы рассмотрели линейные уравнения и неравенства с двумя переменными. Если перейти к линейным зависимостям с тремя переменными, то тогда они будут описывать *плоскость* в трехмерном пространстве; линейное неравенство характеризует *полупространство*, а система линейных неравенств — *многогранник* как область допустимых решений в трехмерном пространстве.

С увеличением числа переменных выше **трех**, геометрическая интерпретация невозможна, так как система неравенств определяет область допустимых решений в k -мерном пространстве.

При этом мерность пространства определяют так: если ограничения заданы неравенствами, то $k = n$, где n — число переменных; если ограничения в виде уравнений, то $k = n - m$, где m — число уравнений.

Из рассмотренного примера решения задачи ОТК мы видели, что оптимальное решение задачи соответствует одной из вершин многогранника.

- Отсюда можно сделать исключительно важный вывод: оптимальное решение — координаты вершины области допустимых решений.

Из этого вывода следует метод **решения** задач линейного программирования, который заключается в следующем:

1. Найти вершины области допустимых решений как точки пересечения ограничений.
2. Определить последовательно значения целевой функции в вершинах.
3. Вершина, в которой целевая функция приобретает оптимальное (\max или \min) значение, является оптимальной вершиной.
4. Координаты оптимальной вершины есть оптимальные значения искомых переменных.

2.2 Формы записи задач (ЗЛП)

В самом общем виде ЗЛП записывается в следующем виде

Найти \min (\max)

$$f(x) = \sum_{j=1}^n c_j x_j \quad (2.1)$$

при ограничениях

$$\sum_{j=1}^n a_{ij} x_j \begin{cases} \leq \\ = \\ \geq \end{cases} b_i, \quad i = \overline{1, m} \quad (2.2)$$

$$x_j \geq 0, \quad j = \overline{1, n} \quad (2.3)$$

ЗЛП (2.1) – (2.3) можно записать, например таким образом

Здесь n - число переменных; m - число ограничений.

Запишем ЗЛП в матричном виде

$$\begin{aligned} \min f(x) &= \min c^t x, \\ Ax &\begin{cases} \leq \\ = \\ \geq \end{cases} b, \\ x &\geq 0, \end{aligned}$$

где $A = \{a_{ij}\}$ - действительная матрица $m \cdot n$ ограничений; $c = (c_1, c_2, \dots, c_n)^t$ - вектор строка; $b = (b_1, \dots, b_m)$ - вектор ограничений ($b_j \geq 0, j = 1, \dots, m$);

$$f(x) = c^t x = \sum_{j=1}^n c_j x_j - \text{целевая функция.}$$

Неотрицательность элементов вектора b легко получить, если это необходимо, умножив соответствующие неравенства в (2.2) на (-1).

Стандартная форма записи ЗЛП (каноническая форма)

ЗЛП имеет стандартную форму, если все ее ограничения имеют форму равенства (кроме ограничений неотрицательности $x_j \geq 0$)

$$\begin{aligned} f(x) &= c^t x \rightarrow \min_x \\ Ax &= b \\ x &\geq 0 \end{aligned} \tag{2.4}$$

Замечание. Если ищется максимум ЦФ $f(x)$, то заменой ЦФ на $-f(x)$ сводят исходную задачу к минимизации функции $-f(x)$, то есть $\min(-f(x)) = \max f(x)$.

Основные методы решения ЗЛП ориентированы на использование стандартной (канонической) формы записи.

Любую ЗЛП всегда можно представить в стандартной форме, введя дополнительные переменные.

Пример 2.2

$$\begin{aligned} \min (5x_1 - 3x_2) \\ x_1 - x_2 &\geq 2 \\ 2x_1 + 3x_2 &\leq 4 \\ -x_1 + 6x_2 &= 10 \\ x_1 \geq 0; \quad x_2 &\geq 0 \end{aligned} \tag{2.5}$$

Введя дополнительные переменные $x_3 \geq 0, x_4 \geq 0$ в 1-ом и 2-ом ограничениях, получим задачу (2.5) в эквивалентной форме

$$\begin{cases} \min (5x_1 - 3x_2 + 0x_3 + 0x_4) \\ x_1 - x_2 - x_3 = 2 \\ 2x_1 + 3x_2 + x_4 = 4 \\ -x_1 + 6x_2 = 10 \\ x_1 \geq 0; \quad x_2 \geq 0; \quad x_3 \geq 0; \quad x_4 \geq 0 \end{cases} \tag{2.6}$$

(2.6) – стандартная форма записи задачи (2.5).

Следует подчеркнуть, что дополнительные переменные столь же необходимы, как и исходные переменные задачи.

В некоторых случаях возникает необходимость рассмотрения переменных x_j , на которые не наложены ограничения, т. е. x_j может принимать и положительные, и отрицательные значения. Поскольку переменные задачи линейного программирования в стандартной форме предполагаются неотрицательными, неограниченные переменные заменяют выражением:

$$x_j = u_j - v_j, \quad \text{где } u_j \geq 0, v_j \geq 0.$$

Значения x_j могут быть > 0 или < 0 в зависимости от соотношения величин u_j и v_j .

Пример 2.3. Рассмотрим задачу линейного программирования в нестандартной форме или в общем виде:

$$f(x) = x_1 - 2x_2 + 3x_3 \rightarrow \max;$$

ограничения:

$$x_1 + x_2 + x_3 \leq 7, \quad (2.7)$$

$$x_1 - x_2 + x_3 \geq 2, \quad (2.8)$$

$$3x_1 - x_2 - 2x_3 = -5, \quad (2.9)$$

$$x_1 \geq 0, x_2 \geq 0, x_3 (\geq / \leq) 0,$$

где x_3 – переменная, неограниченная по знаку.

Преобразуем эту задачу к стандартной форме (СФ):

1) заменим x_3 на $x_4 - x_5$, где $x_4, x_5 \geq 0$;

2) умножим обе части уравнения (2.9) на (-1);

3) введем дополнительные переменные x_6, x_7 в ограничения (2.7) и (2.8) соответственно;

4) припишем нулевой коэффициент переменным x_6, x_7 , а целевую функцию умножим на (-1).

Таким образом, рассматриваемая задача сводится к следующей задаче линейного программирования в СФ:

$$-f(x) = -x_1 + 2x_2 - 3x_4 + 3x_5 - 0 \cdot x_6 - 0 \cdot x_7 \rightarrow \min;$$

при ограничениях:

$$x_1 + x_2 + x_4 - x_5 + x_6 = 7,$$

$$x_1 - x_2 + x_4 - x_5 - x_7 = 2,$$

$$-3x_1 + x_2 + 2x_4 - 2x_5 = 5;$$

$$x_j \geq 0; j = 1, \dots, 7.$$

Ранее мы отмечали, что если b — вектор ресурсов, то в (2.6) дополнительная переменная x_4 означает резерв ресурса $b_2 = 4$. В первом ограничении введена переменная x_3 со знаком (-). Это значит, что переменная x_3 представляет собой количество дополнительного ресурса, необходимого для выполнения плана.

Сформулируем основные определения ЛП.

Определение 1. Вектор x , удовлетворяющий всем ограничениям ЗЛП, называется допустимым решением (планом).

Определение 2. Множество всех допустимых решений образует допустимую область S : $S = \{x | Ax \leq b, x \geq 0\}$. Если $S = \emptyset$ - пустое множество, то ЗЛП называется противоречивой.

Определение 3. План x^* , для которого $f(x)$ достигает минимума, называется оптимальным решением $f(x^*) \leq f(x)$.

Неединственность оптимального решения Существует более одного допустимого решения со значениями ЦФ равными оптимальному f^*

Неограниченный оптимум $\min (\max) f \rightarrow -\infty (+\infty)$

Рассмотрим несколько теорем, позволяющих провести анализ решения ЗЛП

Теорема 1. Множество всех решений ЗЛП выпукло.

Теорема 2. Линейная форма $f(x) = c^t x$ ЗЛП достигает своего минимума (максимума) в крайней точке выпуклой области решений S (в вершине многогранника). Если $f(x)$ достигает $\min (\max)$ более, чем в одной крайней точке, то она достигает того же значения в любой точке, являющейся выпуклой комбинацией этих точек.

2.3 Основы симплекс-метода

Рассмотрим общую ЗЛП с m ограничениями и n переменными, записанную в стандартной (канонической) форме

$$\begin{aligned} f(x) = c^t x &= \sum_{j=1}^n c_j x_j \rightarrow \min \\ a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ \dots \quad \dots & \quad \dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \\ x_j \geq 0; \quad j &= \overline{1, n} \end{aligned} \tag{2.10}$$

Как правило, число уравнений задачи меньше числа переменных (т.е. $m < n$), поэтому множество ее допустимых решений равно ∞ . Задача состоит в том, чтобы найти наилучшее решение в смысле принятого критерия (минимума целевой функции).

Из теоремы 2 следует, что оптимальное решение представляет собой одну из вершин многогранника допустимой области. Другими словами, оптимальное решение это одно из базисных решений.

Получение одного из базисных решений основано на известном классическом методе решения систем линейных уравнений – методе Гаусса – Жордана. Основная идея этого метода состоит в сведении системы m уравнений с n неизвестными к каноническому или ступенчатому виду при помощи элементарных операций над строками

- 1) умножение любого уравнения системы на положительное или отрицательное число
- 2) прибавление к любому уравнению другого уравнения системы, умноженного на положительное или отрицательное число

При использовании первых m переменных такая система имеет следующий вид

$$\begin{aligned}
 x_1 + a_{1,m+1}^0 x_{m+1} + \dots + a_{1,s}^0 x_s + \dots + a_{1,n}^0 x_n &= b_1^0 \\
 x_k + a_{k,m+1}^0 x_{m+1} + \dots + a_{k,s}^0 x_s + \dots + a_{k,n}^0 x_n &= b_k^0 \\
 x_m + a_{m,m+1}^0 x_{m+1} + \dots + a_{m,s}^0 x_s + \dots + a_{m,n}^0 x_n &= b_m^0
 \end{aligned} \tag{2.11}$$

Подставляя эти переменные в целевую функцию, получим

$$f(x) = p_0^0 + \sum_{j=m+1}^n p_j^0 x_j \tag{2.12}$$

$$\text{где } p_0^0 = \sum_{i=1}^m c_i b_i^0; \quad p_j^0 = -\sum_{i=1}^m c_i a_{ij}^0 + c_j, \quad j = m+1, \dots, n \tag{2.13}$$

Систему (2.11) можно переписать следующим образом

$$x_i = b_i^0 - \sum_{j=m+1}^n a_{ij}^0 x_j, \quad i = \overline{1, m} \tag{2.14}$$

Переменные x_1, \dots, x_m входящие с единичными коэффициентами в одно уравнение системы (2.11) (или (2.14)) и с нулевыми в остальные, называются базисными или зависимыми. В канонической системе (2.11), (2.14) каждому уравнению соответствует ровно одна базисная переменная.

Остальные $(n - m)$ переменные (x_{m+1}, \dots, x_n) называются небазисными или независимыми переменными.

При записи системы в каноническом виде все ее решения можно получить, присваивая независимым переменным произвольные значения и решая затем получающуюся систему относительно зависимых переменных.

Определение 4 Базисным решением системы (2.11) называется решение, полученное при нулевых значениях небазисных переменных.

Например, в системе (2.11) одно из базисных решений задается как

$$\left. \begin{aligned}
 x_1 = b_1^0, \quad x_2 = b_2^0, \dots, \quad x_m = b_m^0 \\
 x_{m+1} = 0; \dots; \quad x_n = 0
 \end{aligned} \right\} \tag{2.15}$$

Определение 5 Базисное решение называется допустимым базисным решением, если значения входящих в него базисных переменных $x_j \geq 0, \quad j = \overline{1, m}$, что эквивалентно

условию $b_j^0 \geq 0, \quad j = \overline{1, m}$.

Т.к. различные базисные решения системы (2.11) соответствуют различным вариантам выбора (m) из общего числа (n) переменных x_j , то число допустимых базисных решений (угловых точек) не превышает

$$c_n^m = \frac{n!}{m!(n-m)!}$$

Поэтому ЗЛП можно решать посредством перебора конечного числа угловых точек допустимого множества S , сравнивая значения ЦФ в этих точках. Это наихудший вариант решения ЗЛП, т.к. требует огромного объема вычислений.

Пример 2.4. Если $n = 50, \quad m = 25$ (задача небольшой размерности), то количество переборочных составит 1.26×10^{14} (кол-во вариантов). Далее если ЭВМ будет

выполнять 10^5 переборов вариантов в 1 сек., то ей потребуется 40 лет для поиска решения.

Обычно $n = 1500 \div 2000$; $m = 1000 \div 1500$

Идея СМ состоит в направленном переборе угловых точек допустимого множества S с последовательным уменьшением ЦФ $f(x)$.

СМ разработал Дж. Данциг (американский ученый) в 1947 г. Этот метод называют так же методом *последовательного улучшения решения (плана)*.

Гарантии результативности СМ обеспечиваются следующей теоремой.

Теорема 3 (о конечности алгоритма симплекс метода). Если существует оптимальное решение ЗЛП, то существует и базисное оптимальное решение. Это решение может быть получено через конечное число шагов симплекс-методом, причем, начать можно с любого исходного базиса.

2.4. Алгоритм симплекс метода

Ручные вычисления по симплекс-методу удобно оформлять в виде так называемых симплекс-таблиц. Для удобства составления симплекс-таблицы (СТ) будем считать, что ЦФ и система ограничений переписаны в форме (2.11), (2.12)

$$f(x) = p_0^0 + \sum_{j=m+1}^n p_j^0 x_j$$

$$\text{где } p_0^0 = \sum_{i=1}^m c_i b_i^0; \quad p_j^0 = -\sum_{i=1}^m c_i a_{ij}^0 + c_j, \quad j = m+1, \dots, n$$

$$x_i + \sum_{j=m+1}^n a_{ij}^0 x_j = b_i^0, \quad i = \overline{1, m}$$

По исходным данным заполняется исходная СТ. В первый столбец выписывают базисные переменные, в последний столбец свободные члены. В первую строку выписывают небазисные переменные, в последнюю строку – коэффициенты при переменных в целевой функции. В последней клетке последней строки заносим p_0^0 со знаком “минус”.

В остальные клетки заносятся коэффициенты ограничений

Таблица 2.1

| ↓ | | | | | | |
|----------------------------|--------------|-----|------------|-----|------------|--------------------|
| Базисные переменны е | x_{m+1} | ... | x_k | ... | x_n | Свободные члены |
| x_1 | a_{1m+1}^0 | | a_{1k}^0 | | a_{1n}^0 | b_1^0 |
| ... | | | | | | ... |
| → x_j | a_{jm+1}^0 | | a_{jk}^0 | | a_{jn}^0 | b_j^0 |

| | | | | | | |
|---------------------|--------------|-----|------------|-----|------------|----------|
| ... | | | | | | ... |
| x_m | a_{mm+1}^0 | | a_{mk}^0 | | a_{mn}^0 | b_m^0 |
| Целевая функция f | p_{m+1}^0 | ... | p_k^0 | ... | p_n^0 | $-p_0^0$ |

Сформулируем сейчас алгоритм СМ применительно к данным, внесенным в табл.

1.

Шаг 1. Выяснить, имеются ли в последней строке таблицы отрицательные числа (p_0^0 не принимается во внимание). Если все числа неотрицательны, то процесс закончен; базисное решение $(b_1^0, \dots, b_m^0, 0, \dots, 0)$ является оптимальным; $f^* = p_0^0$.

Если в последней строке имеются отрицательные числа $p_{m+1}^0, p_{m+2}^0, \dots, p_n^0$, перейти к Шаг 2.

Шаг 2.

Просмотреть столбец, соответствующий наименьшему отрицательному числу p_k^0 , $m+1 \leq k \leq n$ и выяснить, имеются ли в нем положительные числа a_{ik} , $i = 1, m$. Если ни в одном из таких столбцов нет положительных чисел, то оптимального решения не существует.

Если найден столбец, содержащий хотя бы один положительный элемент (если таких столбцов несколько взять любой из них), отметить его вертикальной стрелкой (см. табл. 2.1) и перейти к Шаг 3.

Шаг 3. Разделить свободные члены на соответствующие положительные числа из выделенного столбца и выбрать наименьшее частное. Отметить строку, соответствующую наименьшему частному горизонтальной стрелкой. Выделить разрешающий элемент a_{jk}^0 , стоящий на пересечении отмеченных строки и столбца. Перейти к Шаг 4.

Шаг 4.

1. Поменять местами переменные x_k и x_j , остальные переменные оставить на прежних местах.
2. На место опорного элемента поставить число $\frac{1}{a_{jk}^0}$.
3. На остальных местах разрешающей (опорной) строки записать соответствующие элементы исходной таблицы, делённые на опорный элемент.
4. На свободные места разрешающего столбца поставить со знаком минус соответствующие элементы исходной таблицы, делённые на опорный элемент.

Шаг 5

Оставшиеся свободные места в новой СТ заполнить построчно следующим образом: из строки элементов исходной таблицы вычесть произведение ее элемента из разрешающего столбца на уже заполненную разрешающую строку новой таблицы.

Например, для строки i – ой базисной переменной имеем:

$$(a_{i,m+1}^1, \dots, \neg, \dots, a_{i,n}^1, b_i^1) = (a_{i,m+1}^0, \dots, \neg, \dots, a_{i,n}^0, b_i^0) - a_{i,k}^0 (a_{k,m+1}^1, \dots, \neg, \dots, a_{k,n}^1, b_k^1)$$

Знак \rightarrow стоит на месте элемента разрешающего столбца, заполненного согласно определению операции 4.

На этом заполнение новой таблицы заканчивается и происходит переход к Шаг 1.

Замечание. Если целевую функцию необходимо максимизировать, то предварительно нужно умножить ее на (-1).

Пример 2.5. Решить ЗЛП симплекс-методом

$$x_1 - x_4 + x_5 = 2$$

$$x_2 + 2x_4 + 3x_5 = 7$$

$$x_3 + x_4 - 2x_5 = 1$$

$$f = 3 + x_4 - 2x_5 \rightarrow \min$$

Из системы ограничений видно, что переменные x_1, x_2, x_3 являются базисными. Эти переменные используем в качестве начального опорного плана (начального базиса). Поэтому 1, 2 и второй столбцы исключаем из таблицы. Результаты расчета представлены в табл.2.2

Таблица 2.2

| Базисные переменные | x_4 | x_5 ↓ | Свободные члены |
|---------------------|------------|------------|-----------------|
| $\rightarrow x_1$ | -1 | 1 | 2 |
| x_2 | 2 | 3 | 7 |
| x_3 | 1 | -2 | 1 |
| f | 1 | -2 | -3 |
| | x_4 ↓ | x_1 | |
| x_5 | -1 | 1 | 2 |
| $\rightarrow x_2$ | 5 | -3 | 1 |
| x_3 | -1 | 2 | 5 |
| f | -1 | 2 | 1 |
| | x_2 | x_1 | |
| x_5 | 0.2 | 0.4 | 2,2 |
| x_4 | 0.2 | -0.6 | 0.2 |
| x_3 | 0.2 | 1.4 | 5,2 |
| f | 0.2 | 1.40 | 1,2 |

Ответ: $x^* = (0, 0, 5.2, 0.2, 2.2)$; $f^* = -1.2$.

Пример 2.6. Решить ЗЛП симплекс методом

$$f(x) = 2x_1 + 3x_2 \rightarrow \max$$

при ограничениях

$$\left. \begin{array}{l} x_1 + 3x_2 \leq 18 \\ 2x_1 + x_2 \leq 16 \\ x_2 \leq 5 \\ 3x_1 \leq 21 \end{array} \right\}$$

Решение. Приведем задачу к стандартному виду с помощью дополнительных переменных.

$$\left. \begin{array}{l} x_1 + 3x_2 + x_3 = 18 \\ 2x_1 + x_2 + x_4 = 16 \\ x_2 + x_5 = 5 \\ 3x_1 + x_6 = 21 \end{array} \right\}$$

Дальше решаем симплексным методом.

Таблица 2.3

| Базисные переменные | x_1 | x_2 ↓ | Свободные члены |
|---------------------|----------|------------|-----------------|
| x_3 | 1 | 3 | 18 |
| x_4 | 2 | 1 | 16 |
| → x_5 | 0 | 1 | 5 |
| x_6 | 3 | 0 | 21 |
| f | -2 | -3 | 0 |
| x_1 ↓ | | x_5 | |
| x_3 | 1 | -3 | 3 |
| x_4 | 2 | -1 | 11 |
| x_2 | 0 | 1 | 5 |
| x_6 | 3 | 0 | 21 |
| f | -2 | 3 | 15 |
| x_3 ↓ | | x_5 | |
| x_1 | 1 | -3 | 3 |
| → x_4 | -2 | 5 | 5 |
| x_2 | 0 | 1 | 5 |
| x_6 | -3 | 9 | 12 |
| f | 2 | -3 | 21 |
| x_3 ↓ | | x_4 | |
| x_1 | -1/5 | 3/5 | 6 |
| x_5 | -2/5 | 1/5 | 1 |

| | | | |
|-------|-------|--------|----|
| x_2 | $2/5$ | $-1/5$ | 4 |
| x_6 | $3/5$ | $-9/5$ | 3 |
| f | $4/5$ | $3/5$ | 24 |

Ответ: $x^* = (6, 4, 0, 0, 1, 3)$; $f^* = 24$.

Таким образом, мы можем записать

$$\begin{cases} x_1 = 6 + \frac{1}{5}x_3 - \frac{3}{5}x_4, \\ x_2 = 4 - \frac{2}{5}x_3 + \frac{1}{5}x_4, \\ x_5 = 3 + \frac{2}{5}x_3 - \frac{1}{5}x_4, \\ x_6 = 3 - \frac{3}{5}x_3 + \frac{9}{5}x_4. \end{cases}$$

$$f = 24 - \frac{4}{5}x_3 - \frac{3}{5}x_4.$$

Коэффициенты целевой функции, полученные в последней строке, мы взяли со знаком «минус», так как мы умножали целевую функцию на (-1) . Оптимальное значение целевой функции берем со знаком «плюс» (потому что в исходной задаче мы ищем максимум f).

Пример 2.7. Решить ЗЛП симплекс методом (эта задача является двойственной к задаче 2.6)

$$Z = 18y_1 + 16y_2 + 5y_3 + 21y_4 \rightarrow \min$$

при ограничениях

$$\begin{cases} y_1 + 2y_2 + 3y_4 \geq 2, \\ 3y_1 + y_2 + y_3 \geq 3, \\ y_i \geq 0, \quad i = 1, 2, 3, 4. \end{cases}$$

Решение. Вводим дополнительные переменные y_5, y_6 со знаком «минус». Получим стандартную ЗЛП

$$\begin{cases} y_1 + 2y_2 + 3y_4 - y_5 = 2, \\ 3y_1 + y_2 + y_3 - y_6 = 3, \\ y_i \geq 0, \quad i = 1, 2, 3, 4, 5, 6. \end{cases}$$

В качестве базисных переменных возьмем y_3, y_4 . Для этого первое уравнение разделим на 3. Выразим базисные переменные через свободные

$$\begin{aligned} y_3 &= 3 - 3y_1 - y_2 + y_6, \\ y_4 &= \frac{2}{3} - \frac{1}{3}y_1 - \frac{2}{3}y_2 + \frac{1}{3}y_5 \end{aligned}$$

и подставим в целевую функцию, получим

$$Z = 29 - 4y_1 - 3y_2 + 7y_5 + 5y_6$$

Дальше решаем симплексным методом.

Таблица 2.4

| Базисные переменные | y_1 ↓ | y_2 | y_5 | y_6 | Свободные члены |
|---------------------|------------|------------|-------|-------|-----------------|
| $\rightarrow y_3$ | 3 | 1 | 0 | -1 | 3 |
| y_4 | 1/3 | 2/3 | -1/3 | 0 | 2/3 |
| Z | -4 | -3 | 7 | 5 | -29 |
| | y_3 | y_2 | y_5 | y_6 | |
| y_1 | 1/3 | 1/3 | 0 | -1/3 | 1 |
| y_4 | -1/9 | 5/9 | -1/3 | 1/9 | 1/3 |
| Z | 4/3 | -5/3 | 7 | 11/3 | -25 |
| | y_3 | y_4 | y_5 | y_6 | |
| y_1 | 2/5 | -3/5 | 1/5 | -2/5 | 4/5 |
| y_2 | -1/5 | 9/5 | -3/5 | 1/5 | 3/5 |
| Z | 1 | 3 | 6 | 4 | -24 |

Ответ: $y^* = (4/5, 3/5, 0, 0, 0, 0)$; $Z^* = 24$.

Таким образом, мы можем записать

$$\begin{cases} y_1 = \frac{4}{5} - \frac{2}{5}y_3 + \frac{3}{5}y_4 - \frac{1}{5}y_5 + \frac{2}{5}y_6, \\ y_2 = \frac{3}{5} + \frac{1}{5}y_3 - \frac{9}{5}y_4 + \frac{3}{5}y_5 - \frac{1}{5}y_6 \end{cases}$$

$$Z = 24 + y_3 + 3y_4 + 6y_5 + 4y_6$$

2.5 Поиск начального базиса

Для решения задачи ЛП симплексным методом, необходимо получить начальный опорный план (начальный базис).

Рассмотрим способ получения начального базиса (начальной угловой точки многогранника допустимой области). Если в исходной задаче ЛП ограничения заданы в виде неравенств, например,

$$\sum_{j=1}^n a_{ij}x_j \leq b_i, \quad i = \overline{1, m} \quad (2.16)$$

то введение дополнительных переменных $y_1 = x_{n+1}, y_2 = x_{n+2}, \dots, y_m = x_{n+m}$ приведет (1) к виду

$$y_i = b_i - \sum_{j=1}^n a_{ij} x_j, \quad i = \overline{1, m}, \quad (2.17)$$

Вектор $X = (y_1, y_2, \dots, y_m, 0, 0, \dots, 0)$ будет являться начальным опорным планом (начальным базисным решением).

2.5.1. Метод симплексного преобразования

Если ограничения исходной задачи ЛП заданы в виде равенств (то есть имеем сразу стандартную форму ЗЛП), то для получения начального базиса можно воспользоваться методом симплексного преобразования, который является одной из модификаций метода Гаусса – Жордана.

Запишем систему $\sum_{j=1}^n a_{ij} x_j = b_i, \quad i = \overline{1, m}$ в виде таблицы.

Таблица 2.5

| x_1 | x_2 | ... | x_n | Свободные члены |
|----------|----------|-----|----------|-----------------|
| a_{11} | a_{12} | ... | a_{1n} | b_1 |
| a_{21} | a_{22} | ... | a_{2n} | b_2 |
| ... | ... | ... | ... | \vdots |
| a_{m1} | a_{m2} | ... | a_{mn} | b_m |

При этом предполагается, что все $b_i \geq 0, \quad i = \overline{1, m}$. Алгоритм симплексного преобразования основан на идее исключения переменных методом, во многом схожим с методом преобразования симплекс-таблиц.

Шаг 1

Среди столбцов из коэффициентов при неизвестных выбирается столбец, в котором имеется хотя бы один положительный элемент. Если в таком столбце несколько положительных элементов, то из них выбирается тот, который отвечает наименьшему частному при делении соответствующих свободных членов на положительные элементы выбранного столбца. Выделенный таким образом элемент называется разрешающим.

Шаг 2.

Элементы разрешающей строки делятся на разрешающий элемент и переписываются в новую таблицу.

Шаг 3.

Каждая новая строка новой таблицы образуется следующим образом: из строки элементов исходной таблицы вычитается разрешающая строка, полученная в Шаг 2, которая предварительно умножается на соответствующий элемент разрешающего столбца. При этом в клетках выделенного (разрешающего) столбца появятся нули. На этом заполнение новой таблицы заканчивается и происходит переход к Шаг 1.

Процесс продолжается до тех пор, пока не будет получено неотрицательное базисное решение.

Пример 2.8 Найти неотрицательное базисное решение системы

$$x_1 \quad -5x_5 + 2x_6 = -3$$

$$x_2 \quad -2x_5 - x_6 = 5$$

$$x_3 \quad -3x_5 + x_6 = -7$$

$$x_4 + x_5 - 3x_6 = -4$$

Воспользуемся методом симплексного преобразования. Перед составлением начальной таблицы с целью получения неотрицательных свободных членов первое, третье и четвертое уравнения умножим на (-1) . Дальнейшие преобразования по нахождению начального базиса приведены в табл. 2.6.

Таблица 2.6.

| Этап преобразований | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 | Свободные члены |
|---------------------|------------|-------|-------|-------|----------|------------|-----------------|
| Этап 1 | -1 | 0 | 0 | 0 | <u>5</u> | -2 | 3 |
| | 0 | 1 | 0 | 0 | -2 | -1 | 5 |
| | 0 | 0 | -1 | 0 | 3 | -1 | 7 |
| | 0 | 0 | 0 | -1 | -1 | 3 | 4 |
| Этап 2 | -1/5 | 0 | 0 | 0 | 1 | -2/5 | 3/5 |
| | -2/5 | 1 | 0 | 0 | 0 | -9/5 | 31/5 |
| | <u>3/5</u> | 0 | -1 | 0 | 0 | 1/5 | 26/5 |
| | -1/5 | 0 | 0 | -1 | 0 | 13/5 | 23/5 |
| Этап 3 | 0 | 0 | -1/3 | 0 | 1 | -1/3 | 7/3 |
| | 0 | 1 | -2/3 | 0 | 0 | -5/3 | 29/3 |
| | 1 | 0 | -5/3 | 0 | 0 | 1/3 | 26/3 |
| | 0 | 0 | -1/3 | -1 | 0 | <u>8/3</u> | 19/3 |
| Этап 4 | 0 | 0 | -3/8 | -1/8 | 1 | 0 | 25/8 |
| | 0 | 1 | -7/8 | -5/8 | 0 | 0 | 109/8 |
| | 1 | 0 | -13/8 | 1/8 | 0 | 0 | 63/8 |
| | 0 | 0 | -1/8 | -3/8 | 0 | 1 | 19/8 |

Итак, получили базисное решение $(63/8, 109/8, 0, 0, 25/8, 19/8)$, что при необходимости позволяет составить первую симплекс таблицу с начальным базисом

$$B = (x_1, x_2, x_5, x_6)$$

В канонической системе (2.17) каждому уравнению соответствует ровно одна базисная переменная.

2.5.2. Метод искусственного базиса

Другим методом поиска начального базиса является метод искусственного базиса.

В систему ограничений вводятся искусственные переменные $x_{n+i} \geq 0$ ($i = 1, \dots, m$) и вводится искусственная целевая функция $z(x) = \sum_{i=1}^m x_{n+i}$. Таким образом, приходим к следующей задаче линейного программирования

| | | | | | | | |
|----------|-------|-------|-------|-------|-------|----------|--------|
| x_7 | -1 | 0 | 0 | 0 | 5 | -2 | 3 |
| x_8 | 0 | 1 | 0 | 0 | -2 | -1 | 5 |
| x_9 | 0 | 0 | -1 | 0 | 3 | -1 | 7 |
| x_{10} | 0 | 0 | 0 | -1 | -1 | 3 | 4 |
| z | 1 | -1 | 1 | 1 | -5 | 1 | -19 |
| | x_1 | x_8 | x_3 | x_4 | x_5 | x_6 | |
| x_7 | -1 | 0 | 0 | 0 | 5 | -2 | 3 |
| x_2 | 0 | 1 | 0 | 0 | -2 | -1 | 5 |
| x_9 | 0 | 0 | -1 | 0 | 3 | -1 | 7 |
| x_{10} | 0 | 0 | 0 | -1 | -1 | 3 | 4 |
| z | 1 | 1 | 1 | 1 | -7 | 0 | -14 |
| | x_1 | x_8 | x_3 | x_4 | x_7 | x_6 | |
| x_5 | -1/5 | 0 | 0 | 0 | 1/5 | -2/5 | 3/5 |
| x_2 | -2/5 | 1 | 0 | 0 | 2/5 | -9/5 | 31/5 |
| x_9 | 3/5 | 0 | -1 | 0 | -3/5 | 1/5 | 26/5 |
| x_{10} | -1/5 | 0 | 0 | -1 | 1/5 | 13/5 | 23/5 |
| z | -2/5 | 1 | 1 | 1 | 7/5 | -14/5 | -49/5 |
| | x_1 | x_8 | x_3 | x_4 | x_7 | x_{10} | |
| x_5 | -3/13 | 0 | 0 | -2/13 | 3/13 | 2/13 | 17/13 |
| x_2 | -7/13 | 1 | 0 | -9/13 | 7/13 | 9/13 | 122/13 |
| x_9 | 8/13 | 0 | -1 | 1/13 | -8/13 | -1/13 | 63/13 |
| x_6 | -1/13 | 0 | 0 | -5/13 | 1/13 | 5/13 | 23/13 |
| z | -8/13 | 1 | 1 | -1/13 | 21/13 | 14/13 | -63/13 |
| | x_9 | x_8 | x_3 | x_4 | x_7 | x_{10} | |
| x_5 | 3/8 | 0 | -3/8 | -1/8 | 0 | 1/8 | 25/8 |
| x_2 | 7/8 | 1 | -7/8 | -5/8 | 0 | 5/8 | 109/8 |
| x_1 | 13/8 | 0 | -13/8 | 1/8 | -1 | -1/8 | 63/8 |
| x_6 | 1/8 | | -1/8 | -3/8 | 0 | 3/8 | 19/8 |
| z | 1 | 1 | 0 | 0 | 1 | 1 | 0 |

Итак, получили базисное решение $(63/8, 109/8, 0, 0, 25/8, 19/8)$. Это базисное решение совпало с решением, полученным симплексным преобразованием таблицы ограничений. Значение искусственной целевой функции равно нулю. В результате получим следующую систему уравнений-ограничений

$$x_1 - \frac{13}{8}x_3 + \frac{1}{8}x_4 = \frac{63}{8}$$

$$x_2 - \frac{7}{8}x_3 - \frac{5}{8}x_4 = \frac{109}{8}$$

$$x_5 - \frac{3}{8}x_3 - \frac{1}{8}x_4 = \frac{25}{8}$$

$$x_6 - \frac{1}{8}x_3 - \frac{3}{8}x_4 = \frac{19}{8}$$

Рассмотрим случай, когда часть ограничений задана в виде неравенств, а часть ограничений – в виде равенств

$$\begin{aligned} \sum_{j=1}^n a_{ij}x_j &\leq b_i, \quad i = 1, \dots, k, \\ \sum_{j=1}^n a_{ij}x_j &= b_i, \quad i = k + 1, \dots, m, \\ x_j &\geq 0, \quad j = 1, \dots, n \end{aligned} \quad (2.21)$$

Дополним неравенства дополнительными переменными x_{n+i} , $i = 1, \dots, k$, а ограничения равенства дополним искусственными переменными x_{n+k+i} , $i = 1, \dots, m - k$. В результате мы получим систему ограничений в стандартной форме

$$\begin{aligned} \sum_{j=1}^n a_{ij}x_j + x_{n+i} &= b_i, \quad i = 1, \dots, k, \\ \sum_{j=1}^n a_{ij}x_j + x_{n+k+i} &= b_i, \quad i = k + 1, \dots, m, \\ x_j &\geq 0, \quad j = 1, \dots, n + m \end{aligned} \quad (2.22)$$

Затем рассматривается вспомогательная задача

$$\begin{aligned} z(x) &= \sum_{i=k+1}^m x_{n+i} \rightarrow \min \\ \sum_{j=1}^n a_{ij}x_j + x_{n+i} &= b_i, \quad i = 1, \dots, k, \\ \sum_{j=1}^n a_{ij}x_j + x_{n+k+i} &= b_i, \quad i = k + 1, \dots, m, \\ x_j &\geq 0, \quad j = 1, \dots, n + m \end{aligned} \quad (2.23)$$

В качестве начального базиса используются переменные x_{n+i} , $i = 1, \dots, m$. В процессе решения искусственные переменные выводятся из базиса. При этом дополнительные переменные x_{n+i} , $i = 1, \dots, k$ должны остаться в базисе. На последнем шаге симплекс-метода значение искусственной целевой функции будет равно нулю.

Вопросы для самопроверки

1. Как построить на графике область допустимых решений?
2. В чем состоит суть графического метода решения задач линейного программирования
3. Где расположено оптимальное решение на допустимой области?
4. Приведите стандартную форму записи задач линейного программирования
5. Как привести ЗЛП к стандартной форме?
6. Приведите основные определения и теоремы линейного программирования
7. Приведите алгоритм симплекс метода
8. Сформулируйте алгоритм поиска начального базиса в задаче линейного программирования
9. В чем состоит суть метода симплексного преобразования таблицы ограничений?
10. В чем состоит суть метода искусственного базиса?
11. Каким образом найти начальное приближение, когда часть ограничений задана в виде равенств, а часть – в виде неравенств?

| | |
|---|--|
| $x = (x_1, x_2, \dots, x_n)$, при котором прибыль (выручка) от реализации продукции будет максимальной при условии, что потребление ресурсов по каждому виду продукции не превзойдет имеющихся запасов | $y = (y_1, y_2, \dots, y_m)$, при котором общие затраты на ресурсы будут минимальными при условии, что затраты на ресурсы при производстве каждого вида продукции будут не менее прибыли от реализации этой продукции |
|---|--|

Цены ресурсов y_1, y_2, \dots, y_m в экономической литературе получили различные названия: **учетные, неявные, теневые**. Смысл этих названий состоит в том, что это условные, "ненастоящие" цены. В отличие от "внешних" цен c_1, c_2, \dots, c_n на продукцию, известных, как правило, до начала производства, цены ресурсов y_1, y_2, \dots, y_m являются внутренними, ибо они задаются не извне, а определяются непосредственно в результате решения задачи, поэтому их чаще называют оценками ресурсов.

3.2 Свойства взаимно двойственных задач

Рассмотрим формально две задачи I и II линейного программирования, представленные в табл. 3.1, абстрагируясь от содержательной интерпретации параметров, входящих в их экономико-математические модели. Обе задачи обладают следующими свойствами:

1. В одной задаче ищут максимум целевой функции, в другой — минимум.
2. Коэффициенты при переменных в целевой функции одной задачи являются свободными членами системы ограничений в другой.
3. В задаче максимизации все неравенства ограничений вида " \leq ", а в задаче минимизации — все неравенства вида " \geq ".
4. Матрицы коэффициентов при переменных в системах ограничений обеих задач являются транспонированными друг к другу:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}$$

5. Число неравенств в системе ограничений одной задачи совпадает с числом переменных в другой задаче.
 6. Условия неотрицательности переменных имеются в обеих задачах.
- Две задачи I и II линейного программирования, обладающие указанными свойствами, называются симметричными взаимно двойственными задачами. В дальнейшем для простоты будем называть их просто двойственными задачами.

Примечание. Если в прямой задаче ограничения заданы в форме равенств, то в двойственной задаче ограничение (3.6) отсутствует, т.е. переменные y_i могут быть и положительными и отрицательными.

Пример 3.1. Составить задачу, двойственную следующей задаче:

$$f = -x_1 + 2x_2 \rightarrow \max$$

при ограничениях:

$$\begin{cases} 2x_1 - x_2 \geq 1, \\ -x_1 + 4x_2 \leq 24, \\ x_1 - x_2 \leq 3, \\ x_1 + x_2 \geq 5, \\ x_1 \geq 0, x_2 \geq 0. \end{cases}$$

Решение. 1. Так как исходная задача на максимизацию, то приведем все неравенства системы ограничений к виду " \leq ", для чего обе части первого и четвертого неравенства умножим на -1. Получим

$$\begin{cases} -2x_1 + x_2 \leq -1, \\ -x_1 + 4x_2 \leq 24, \\ x_1 - x_2 \leq 3, \\ -x_1 - x_2 \leq -5, \end{cases}$$

2. Сформулируем двойственную задачу:

$$Z = -y_1 + 24y_2 + 3y_3 - 5y_4 \rightarrow \max$$

при ограничениях

$$\begin{cases} -2y_1 - y_2 + y_3 - y_4 \geq -1, \\ y_1 + 4y_2 - y_3 - y_4 \geq 2, \\ y_1 \geq 0, y_2 \geq 0, y_3 \geq 0, y_4 \geq 0. \end{cases}$$

3.3 Теоремы двойственности

Связь между оптимальными решениями двойственных задач устанавливается с помощью теорем двойственности. Вначале рассмотрим вспомогательное утверждение.

Основное неравенство теории двойственности. Пусть имеется пара двойственных задач I и II. Покажем, что для любых допустимых решений $x = (x_1, \dots, x_n)$ и $y = (y_1, \dots, y_m)$ исходной и двойственной задач справедливо неравенство

$$f(x) \leq Z(y) \text{ или } \sum_{j=1}^n c_j x_j \leq \sum_{i=1}^m b_i y_i \quad (3.7)$$

Доказательство. Умножив неравенства системы ограничений (3.2) исходной задачи

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

соответственно на переменные y_1, \dots, y_m и сложив правые и

левые части полученных неравенств, имеем

$$\sum_{i=1}^m y_i \sum_{j=1}^n a_{ij} x_j \leq \sum_{i=1}^m b_i y_i \quad (3.8)$$

Аналогично преобразуем систему ограничений (3.5) двойственной задачи $\sum_{i=1}^m a_{ij} y_i \geq c_j$

($j = 1, \dots, n$) путем умножения обеих частей ее неравенства на переменные x_1, \dots, x_n и последующего их сложения

$$\sum_{j=1}^n c_j x_j \leq \sum_{j=1}^n x_j \sum_{i=1}^m a_{ij} y_i \quad (3.9)$$

Из (3.8) и (3.9) следует неравенство (3.7).

Существует ряд важных соотношений между решениями взаимодвойственных задач, которые следуют из следующей теоремы.

Первая (основная) теорема двойственности. Для взаимодвойственных ЗЛП имеет место один из взаимоисключающих случаев:

1. В прямой и двойственных задачах имеются оптимальные решения, при этом значения целевых функций на оптимальных решениях совпадают

$$f(x^*) = Z(y^*) \quad (3.10)$$

2. В прямой задаче допустимое множество не пусто, а целевая функция не ограничена сверху. При этом у двойственной задачи будет пустое допустимое множество
3. В двойственной задаче допустимое множество не пусто, а целевая функция не ограничена снизу. При этом у прямой задачи будет пустое допустимое множество
4. Обе из рассматриваемых задач имеют пустые допустимые множества.

Покажем (3.10). Пусть x^1 - любое допустимое решение исходной задачи I. Тогда на основании неравенства (3.7) получим $f(x^1) \leq Z(y^*)$. Однако x^1 - произвольное решение прямой задачи, отсюда в силу равенства (1) следует, что $f(x^1) \leq f(x^*)$, т.е. x^* - оптимальное решение задачи I. Аналогично доказывается, что решение оптимально для задачи II.

Утверждение второй части легко доказывается методом от противного. Предположим, что в исходной задаче линейная функция не ограничена, т.е. $f_{\max} = \infty$, а условия двойственной задачи не являются противоречивыми, т.е. существует хотя бы одно допустимое решение y^1 . Тогда в силу основного неравенства теории двойственности (7) $f(x) \leq Z(y^1)$, что противоречит условию неограниченности $f(x)$. Следовательно, при $f_{\max} = \infty$ в исходной задаче допустимых решений в двойственной задаче быть не может. •

Рассмотрим задачу об использовании ресурсов (см. тема 1).

| Прямая задача | Двойственная задача |
|--|---|
| $f(x) = 2x_1 + 3x_2 \rightarrow \max$ <p>при ограничениях</p> $\left. \begin{aligned} x_1 + 3x_2 &\leq 18 \\ 2x_1 + x_2 &\leq 16 \\ x_2 &\leq 5 \\ 3x_1 &\leq 21 \end{aligned} \right\}$ <p>Оптимальное решение прямой задачи $x^* = (6, 4, 0, 0, 1, 3); f^* = 24$</p> | $Z = 18y_1 + 16y_2 + 5y_3 + 21y_4 \rightarrow \min$ <p>при ограничениях</p> $\left\{ \begin{aligned} y_1 + 2y_2 + 3y_4 &\geq 2, \\ 3y_1 + y_2 + y_3 &\geq 3, \\ y_i &\geq 0, i = 1, 2, 3, 4. \end{aligned} \right.$ <p>Оптимальное решение двойственной задачи $y^* = (0.8, 0.6, 0, 0, 0, 0); Z^* = 24$</p> |

Как видим, значения целевых функций обеих задач совпадают.

Исходя из анализа оптимальных двойственных оценок, можно сделать следующие выводы.

Ресурсы первого и второго видов используются полностью (оптимальные оценки Y_1^* , Y_2^* отличны от нуля). Ресурс третьего вида недоиспользуется на 1 ед., ресурс четвертого вида недоиспользуется на 3 ед.

Экономический смысл первой (основной) теоремы двойственности. План производства $X^* = (X_1^*, \dots, X_n^*)$ и набор цен (оценок) ресурсов $Y^* = (Y_1^*, \dots, Y_m^*)$ оказываются оптимальными тогда и только тогда, когда прибыль (выручка) от продукции, найденная при "внешних" (известных заранее) ценах C_1, C_2, \dots, C_n равна затратам на ресурсы по "внутренним" (определяемым только из решения задачи) ценам Y_1, Y_2, \dots, Y_m . Для всех же других планов X и Y обеих задач в соответствии с основным неравенством (3.7) теории двойственности прибыль (выручка) от продукции всегда меньше (или равна) затрат на ресурсы.

Так, в рассмотренной задаче о ресурсах оптимумы прибыли от продукции f_{\max} и затрат на ресурсы Z_{\min} равны 24 д.е., для всех остальных планов $f(x) \leq 24$, $Z(y) \geq 24$.

Экономический смысл первой теоремы двойственности можно интерпретировать и так: предприятию безразлично, производить ли продукцию по оптимальному плану $X^* = (X_1^*, \dots, X_n^*)$ и получить максимальную прибыль (выручку) f_{\max} либо продавать ресурсы по оптимальным ценам $Y^* = (Y_1^*, \dots, Y_m^*)$ и возместить от продажи равные ей минимальные затраты на ресурсы Z_{\min} .

Вторая теорема двойственности. Компоненты оптимального решения двойственной задачи равны значениям частных производных целевой функции $f_{\max}(b_1, b_2, \dots, b_m)$ по соответствующим аргументам, т.е.

$$\frac{\partial f_{\max}}{\partial b_i} = y_i^*, \quad i = 1, \dots, m \quad (3.11)$$

Доказательство следует из первой теоремы двойственности. Действительно, для оптимального решения имеем равенство $f_{\max} = f(X^*) = Z(Y^*) = \sum_{i=1}^m b_i y_i^*$. Отсюда следует (3.11).

Из соотношения (3.11) следует, что двойственные оценки ресурсов показывают, на сколько денежных единиц изменится максимальная прибыль (выручка) от реализации продукции при изменении запаса соответствующего ресурса на одну единицу.

- *Двойственные оценки определяют дефицитность используемых ресурсов. Величина двойственной оценки показывает, на сколько возрастает максимальное значение целевой функции прямой задачи при увеличении количества соответствующего ресурса на единицу (при условии, что оптимальный план двойственной задачи не изменится)*

Так, увеличение количества ресурса первого вида на 1 ед. приведет к тому, что появится возможность найти новый оптимальный план производства изделий, при котором общий доход возрастает на 0.8 д.е. и станет равным $24 + 0.8 = 24.8$ д.е. Анализ полученных оптимальных значений новой прямой задачи показывает, что это увеличение общего дохода достигается за счет увеличения производства изделий P2 на 0.4 ед, и сокращения выпуска изделий P1 на 0,2 ед.

Точно также увеличение на 1 ед. количества ресурсов второго вида позволит перейти к новому оптимальному плану производства, при котором доход возрастает на 0.6 д.е. и составит $24+0.6 = 24.6$ д.е., что достигается за счет увеличения выпуска изделий P1 на 0,2 ед. и уменьшения выпуска P2 на 0,2 ед.. При изменении запасов ресурсов S3, S4 максимальная прибыль не изменится.

Третья теорема двойственности (теорема о дополняющей нежесткости)

Тесная связь между двумя взаимно двойственными задачами проявляется не только в равенстве оптимальных значений их линейных функций, о чем утверждалось в первой (основной) теореме двойственности.

Пусть даны две взаимно двойственные задачи I — (3.1)—(3.3) и задачи II — (3.4)—(3.6). Если каждую из этих задач решать симплексным методом, то необходимо привести их к каноническому виду, для чего в систему ограничений (3.2) задачи I (в краткой записи

$\sum_{j=1}^n a_{ij}x_j \leq b_i, i = 1, \dots, m$) следует ввести m неотрицательных переменных

$x_{n+1}, x_{n+2}, \dots, x_{n+i}, \dots, x_{n+m}$, а в систему ограничений (3.5) задачи II ($\sum_{i=1}^m a_{ij}y_i \geq c_j,$

$j = 1, \dots, n$) — n неотрицательных переменных $y_{m+1}, y_{m+2}, \dots, y_{m+j}, \dots, y_{m+n}$, где $i(j)$ — номер неравенства, в которое введена дополнительная переменная $x_{n+i} \geq 0, y_{m+j} \geq 0$.

Системы ограничений каждой из взаимно двойственных задач примут вид:

$$\sum_{j=1}^n a_{ij}x_j + x_{n+i} = b_i, i = 1, \dots, m \quad (3.12)$$

$$\sum_{i=1}^m a_{ij}y_i - y_{m+j} = c_j, j = 1, \dots, n \quad (3.13)$$

Установим соответствие между первоначальными переменными одной из двойственных задач и дополнительными переменными другой задачи (табл. 3.2).

Таблица 3.2. Связь переменных исходной и двойственной задачи

| Переменные исходной задачи I | | | | | | | | | | | |
|-----------------------------------|----------------|-----|----------------|-----|----------------|----------------|----------------|-----|----------------|-----|----------------|
| Первоначальные | | | | | Дополнительные | | | | | | |
| x_1 | x_2 | ... | x_j | ... | x_n | x_{n+1} | x_{n+2} | ... | x_{n+i} | ... | x_{n+m} |
| \updownarrow | \updownarrow | | \updownarrow | | \updownarrow | \updownarrow | \updownarrow | | \updownarrow | | \updownarrow |
| y_{m+1} | y_{m+2} | ... | y_{m+j} | ... | y_{m+n} | y_1 | y_2 | ... | y_i | ... | y_m |
| Дополнительные | | | | | Первоначальные | | | | | | |
| Переменные двойственной задачи II | | | | | | | | | | | |

Теорема 3. Положительным (ненулевым) компонентам оптимального решения одной из взаимно двойственных задач соответствуют нулевые компоненты оптимального решения другой задачи, т.е.

для любых $i = 1, \dots, m$ и $j = 1, \dots, n$ выполняются следующие соотношения:

если $x_j^* > 0$, то $y_{m+j}^* = 0$; если $x_{n+i}^* > 0$ то $y_i^* = 0$, и аналогично

если $y_i^* > 0$, то $x_{n+i}^* = 0$; если $y_{m+j}^* > 0$ то $x_j^* = 0$, и аналогично

Эти условия можно выразить в краткой форме: для того чтобы решения прямой и двойственной задачи были оптимальными, необходимо и достаточно, чтобы выполнялись соотношения

$$x_{n+i}^* y_i^* = 0, \quad i = 1, \dots, m \quad (3.16)$$

$$x_j^* y_{m+j}^* = 0, \quad j = 1, \dots, n \quad (3.17)$$

Из доказанной теоремы следует важный вывод о том, что введенное ранее соответствие (3.14), (3.15) между переменными взаимно двойственных задач при достижении оптимума (т.е. на последнем шаге решения каждой задачи симплексным методом) представляет соответствие между основными (не равными нулю) переменными одной из двойственных задач и неосновными (равными нулю) переменными другой задачи, когда они образуют допустимые базисные решения.

Пример. Рассмотрим опять задачу о распределении ресурсов.

Решение. На основании выражений (3.14) и (3.15) установим следующее соответствие между переменными:

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_5 & y_6 & y_1 & y_2 & y_3 & y_4 \end{array}$$

В теме 2 обе задачи решены симплексным методом. На последнем шаге решения каждой задачи получено:

| В исходной задаче I | В двойственной задаче II |
|---|---|
| $f = 24 - \frac{4}{5}x_3 - \frac{3}{5}x_4 \quad (18)$ | $Z = 24 + y_3 + 3y_4 + 6y_5 + 4y_6 \quad (19)$ |
| $f(x^*) = f_{\max} = 24$ при оптимальном базисном решении | $Z(y^*) = Z_{\min} = 24$ при оптимальном базисном решении |
| $x^* = (6, 4, 0, 0, 1, 3).$ | $y^* = (4/5, 3/5, 0, 0, 0, 0).$ |

Четвертая теорема двойственности

Компоненты оптимального решения двойственной задачи ($y_1^* = \frac{4}{5}$, $y_2^* = \frac{3}{5}$, $y_3^* = 0$,

$y_4^* = 0$, $y_5^* = 0$, $y_6^* = 0$) равны (по абсолютной величине) коэффициентам при соответствующих переменных целевой функции прямой задачи (записанной для

оптимального плана (3.18), $f = 24 - \frac{4}{5}x_3 - \frac{3}{5}x_4 - 0 \cdot x_5 - 0 \cdot x_6 - 0 \cdot x_1 - 0 \cdot x_2$), а

компоненты оптимального решения прямой задачи

($x_1^* = 6$, $x_2^* = 4$, $x_3^* = 0$, $x_4^* = 0$, $x_5^* = 1$, $x_6^* = 3$) равны коэффициентам при соответствующих переменных целевой функции двойственной задачи (записанной также для оптимального плана (3.19), $Z = 24 + 6y_5 + 4y_6 + 0 \cdot y_1 + 0 \cdot y_2 + 1 \cdot y_3 + 3y_4$).

Замечание. Если в одной из взаимно двойственных задач нарушается единственность оптимального решения, то оптимальное решение двойственной задачи вырожденное. Это связано с тем, что при нарушении единственности оптимального решения исходной задачи в выражении линейной функции ее оптимального решения через неосновные переменные отсутствует хотя бы одна из основных переменных.

Вопросы для самопроверки

1. Сформулируйте двойственную задачу ЛП
2. Свойства взаимно двойственных задач
3. Основное неравенство теории двойственности
4. Первая теорема двойственности и ее экономический смысл
5. Вторая теорема двойственности
6. Третья теорема двойственности
7. Четвертая теорема двойственности

Тема 4. Транспортная задача

Важным частным случаем задачи линейного программирования является так называемая транспортная задача, которую можно сформулировать следующим образом.

4.1 Экономико – математическая модель транспортной задачи

В m пунктах отправления A_1, A_2, \dots, A_m , которые в дальнейшем будем называть поставщиками, сосредоточено определенное количество единиц некоторого однородного продукта, которое обозначим a_i ($i=1, \dots, m$). Данный продукт потребляется в n пунктах B_1, B_2, \dots, B_n , которые будем называть потребителями; объем потребления обозначим b_j ($j=1, \dots, n$). Известны расходы на перевозку единицы продукта из пункта A_i в пункт B_j , которые равны c_{ij} и приведены в матрице транспортных расходов $C = (c_{ij})$.

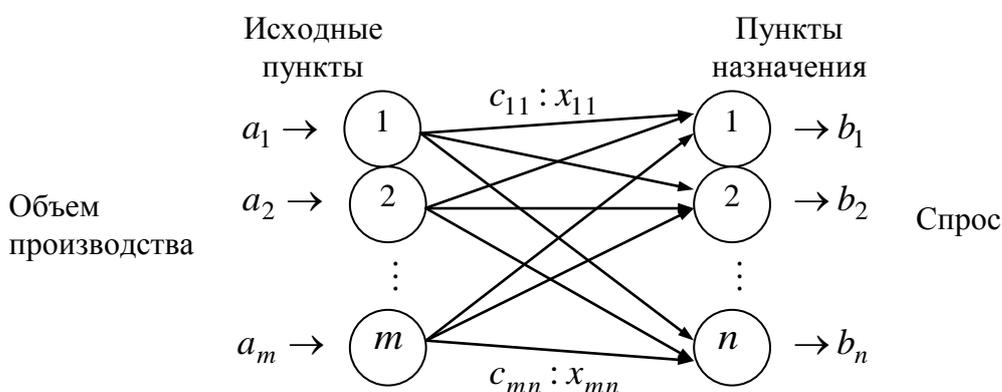


Рис. 4.1. Геометрическая интерпретация транспортной задачи

Требуется составить такой план прикрепления потребителей к поставщикам, т.е. план перевозок, при котором весь продукт вывозится из пунктов A_i в пункты B_j в соответствии с потребностью и общая величина транспортных издержек будет минимальной.

Обозначим количество продукта, перевозимого из пункта A_i в пункт B_j , через x_{ij} . Совокупность всех переменных x_{ij} для краткости обозначим x , тогда целевая функция задачи будет иметь вид

$$f(x) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \min \quad (4.1)$$

а ограничения выглядят следующим образом:

$$\sum_{j=1}^n x_{ij} \leq a_i; \quad i = 1, \dots, m, \quad (4.2)$$

$$\sum_{i=1}^m x_{ij} \geq b_j; \quad j = 1, \dots, n \quad (4.3)$$

$$x_{ij} > 0. \quad (4.4)$$

| | | | | | |
|-----|------|----------|-----|----------|-----|
| A | 1000 | | 80 | | 215 |
| | | x_{11} | | x_{12} | |
| B | 1500 | | 100 | | 108 |
| | | x_{21} | | x_{22} | |
| C | 1200 | | 102 | | 68 |
| | | x_{31} | | x_{32} | |

Пример 4.2. (Несбалансированная модель)

Изменим условия в примере 4.1. Предположим, что завод B производит не 1500, а 1300 автомобилей. Это приведет к дисбалансу, поскольку суммарный объем производства (3500) не равен суммарному спросу (3700). Другими словами, дисбаланс означает, что спрос в центрах распределения (в п. назначения) полностью удовлетворить не удастся. В этом случае необходимо изменить транспортную модель таким образом, чтобы недостаток автомобилей ($3700-3500=200$) оптимально распределился между D, E .

Введем фиктивный исходный пункт (фиктивный завод) с производительностью 200 автомобилей. Стоимость перевозок с фиктивного завода в п. D, E естественно положить равной нулю (т.к. никакие перевозки не осуществляются). Таким образом, сбалансированная модель имеет вид

Таблица 4.3

| | | D | E |
|--------|------|------|------|
| | | 2300 | 1400 |
| A | 1000 | 80 | 215 |
| B | 1300 | 100 | 108 |
| C | 1200 | 102 | 68 |
| Φ | 200 | 0 | 0 |

Здесь Φ - фиктивный завод

Если объем производства превышает спрос, можно ввести дополнительные фиктивные пункты назначения. Пусть, например, в п. D спрос упал с 2300 до 1900 автомобилей. В таблице представлена измененная модель

Таблица 4.4

| | | D | E | Φ |
|-----|------|------|------|--------|
| | | 1900 | 1400 | 400 |
| A | 1000 | 80 | 215 | 0 |
| B | 1500 | 100 | 108 | 0 |
| C | 1200 | 102 | 68 | 0 |

Здесь Φ - фиктивный пункт назначения

Автомобили, поступающие с некоторого завода в фиктивный пункт назначения, представляют собой избыток производства на этом заводе. Соответствующая стоимость перевозки равна нулю. Однако, можно назначить штраф за хранение автомобилей на складе завода, тогда стоимость перевозки одного автомобиля (перепроизведенного) будет равна штрафу за его хранение. Аналогично в модели с недопроизведенным количеством автомобилей. Каждую недопоставленную единицу продукции в п. назначения можно обложить штрафом. Тогда транспортные расходы на единицу недопроизведенной продукции равны штрафу за недополученную продукцию.

Пример 4.3 (Многопродуктовая транспортная модель)

Пусть автомобильная компания (заводы A, B, C) производят автомобили четырех различных марок $M1, M2, M3, M4$, причем завод A выпускает модели

Таблица 4.5

| | Всего | $M1$ | $M2$ | $M3$ | $M4$ |
|------------|-------|------|------|------|------|
| Завод | | | | | |
| A | 1000 | | | 700 | 300 |
| B | 1500 | 500 | 600 | | 400 |
| C | 1200 | 800 | 400 | | |
| Покупатель | | | | | |
| D | 2300 | 700 | 500 | 500 | 600 |
| E | 1400 | 600 | 500 | 200 | 100 |

$M3, M4$; завод B - $M1, M2, M4$; завод C - $M1, M2$. В таблице 5 приведены объемы выпуска и спроса автомобилей всех марок.

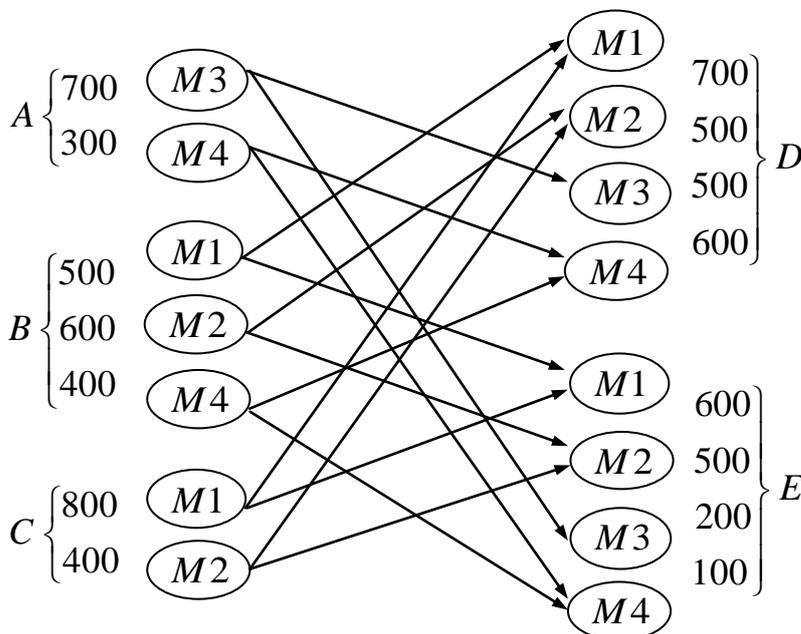


Рис. 4.2 Геометрическая интерпретация многопродуктовой транспортной задачи

Предположим, что стоимость перевозок автомобилей любой марки одинакова (стоимость перевозки на одну милю).

Видоизменим транспортную модель. Для этого каждый завод-изготовитель разобьем на несколько заводов (несколько пунктов), и каждого покупателя разобьем также на несколько пунктов

В результате получаем семь исходных пунктов и восемь пунктов назначения.

В таблице 4.6 приведена полная транспортная таблица. Заметим, что некоторые маршруты недопустимы (на рис отсутствуют стрелки, а в табл. заштрихованы клетки), они соответствуют очень высокой стоимости перевозки. Запрещенные маршруты означают, что автомобили различных марок нельзя заменять друг другом. Например, нельзя осуществлять перевозки из п. $M1$ в п. $M2$ и т.д.

Если внимательно изучить таблицу, то можно заметить, что на самом деле задачу необязательно описывать одной моделью.

Таблица 4.6

| | | | | | | | |
|------|------|------|------|------|------|------|------|
| D | | | | E | | | |
| $M1$ | $M2$ | $M3$ | $M4$ | $M1$ | $M2$ | $M3$ | $M4$ |

| | | | | | | | | | | | | | |
|---|---|----|------|--------|-----|-----|------|-----|-----|-----|-----|-----|---|
| | | | 700 | 500 | 500 | 600 | 600 | 500 | 200 | 100 | | | |
| | | | 2300 | | | | 1400 | | | | | | |
| A | { | M3 | 700 | } 1000 | | M | M | 80 | M | M | M | 215 | M |
| | | M4 | 300 | | | M | M | 80 | M | M | M | 215 | |
| B | { | M1 | 500 | } 1500 | | 100 | M | M | M | 108 | M | M | M |
| | | M2 | 600 | | | M | 100 | M | M | 108 | M | M | |
| | | M4 | 400 | | | M | M | 100 | M | M | 108 | M | |
| C | { | M1 | 800 | } 1200 | | 102 | M | M | M | 68 | M | M | M |
| | | M2 | 400 | | | M | 102 | M | M | 68 | M | M | |

В силу независимости поставок можно было представить задачу по каждой марке автомобилей в виде отдельной таблицы перевозок, но только существенно меньшего размера. В результате получим следующие таблицы

Таблица 4.7

| | | | | | | | |
|---|-----|-----------|----------|---|-----|-----------|----------|
| | | <u>M3</u> | | | | <u>M4</u> | |
| | | <u>D</u> | <u>E</u> | | | <u>D</u> | <u>E</u> |
| | | 500 | 200 | | | 600 | 100 |
| A | 700 | 80 | 215 | A | 300 | 80 | 215 |
| | | | | B | 400 | 100 | 108 |
| | | <u>M1</u> | | | | <u>M2</u> | |
| | | <u>D</u> | <u>E</u> | | | <u>D</u> | <u>E</u> |
| | | 700 | 600 | | | 500 | 500 |
| B | 500 | 100 | 108 | B | 600 | 100 | 108 |
| C | 800 | 102 | 68 | C | 400 | 102 | 68 |

Такое разбиение на задачи меньшей размерности оказалось возможным из-за независимости различных марок автомобилей (т.е. невозможностью замены одной марки другой). Если бы можно было заменять одну марку другой, то такое разбиение не получилось бы.

Пример 4.4 (Модель производства с запасами)

Некоторая фирма собирается выпускать изделия определенного вида в течение четырех месяцев. Величина спроса в течение этих четырех месяцев составляет 100, 200, 180 и 300 изделий соответственно. В каждый месяц спрос можно удовлетворить за счет

- 1) избытка произведенных в прошлом месяце изделий, сохраняющихся для реализации в будущем
- 2) производства изделий в течение месяца
- 3) избытка производства изделий в более поздние месяцы в счет невыполненных заказов

Затраты на одно изделие в каждый месяц составляют 4 доллара. Изделие, произведенное для более поздней реализации, влечет за собой дополнительные издержки за хранение 0,5 дол/мес. С другой стороны, каждое изделие, выпускаемое в счет невыполненных заказов, облагается штрафом 2 дол/мес.

Объем производства изделий меняется от месяца к месяцу. В рассматриваемые четыре месяца предполагается выпуск 50, 180, 280 и 270 изделий соответственно.

Требуется составить план, имеющий минимальную стоимость производства и хранения изделий.

Задачу можно сформулировать как транспортную. Эквивалентность между элементами производственной и транспортной систем устанавливается следующим образом (см. табл. 4.8)

Таблица 4.8

| Транспортная система | Производственная система |
|----------------------------------|---|
| 1. Исходный пункт i | 1. Период производства i |
| 2. Пункт назначения j | 2. Период потребления j |
| 3. Предложение в п. i | 3. объем пр-ва за период i |
| 4. Спрос в п. j | 4. Реализация за период j |
| 5. Стоимость перевозки i в j | 5. Стоимость пр-ва и хранения за период от i до j |

Стоимость «перевозки» изделия из периода i в период j выражается как

$$c_{ij} = \begin{cases} \text{стоимость пр - ва в } i \text{ период, } i = j \\ \text{стоимость пр - ва в } i \text{ период} + \text{стоимость хранения} \\ \quad \text{(задержки) от } i \text{ до } j, i < j \\ \text{стоимость пр - ва в } i \text{ период} + \text{штраф за} \\ \quad \text{нарушение срока, } i > j \end{cases}$$

Из определения c_{ij} следует, что затраты в период i при реализации в тот же период ($i = j$) оцениваются только стоимостью производства. Если в период i производится продукция, которая будет потребляться позже ($i < j$), то имеют место дополнительные затраты связанные с хранением. Аналогично производство в i период в счет невыполнения заказов ($i > j$) влечет за собой дополнительные расходы в виде штрафа.

В таблице 4.9 приведена структура модели

| Объем пр-ва | Спрос | Период | | | | |
|-------------|--------|--------|----|-----|-----|-----|
| | | 1 | 2 | 3 | 4 | |
| 50 | Период | 1 | 4 | 4,5 | 5 | 5,5 |
| 180 | | 2 | 6 | 4 | 4,5 | 5 |
| 280 | | 3 | 8 | 6 | 4 | 4,5 |
| 270 | | 4 | 10 | 8 | 6 | 4 |

4.2 Решение транспортной задачи симплексным методом

Закрытая транспортная задача может быть решена симплексным методом (методом последовательного улучшения плана). Для этого необходимо переобозначить переменные следующим образом: $x_1 = x_{11}, x_2 = x_{12}, \dots, x_n = x_{1,n}; x_{n+1} = x_{21},$

$$x_{n+2} = x_{22}, \dots, x_{2n} = x_{2,n}; \dots$$

$$x_{(m-1)n+1} = x_{m1}, x_{(m-1)n+2} = x_{m2}, \dots, x_{mn} = x_{m,n}$$

Точно также переобозначаем коэффициенты c_{ij} : $c_1 = c_{11}, c_2 = c_{12}, \dots, c_n = c_{1,n}$; $c_{n+1} = c_{21}, c_{n+2} = c_{22}, \dots, c_{2n} = c_{2,n}; \dots, c_{(m-1)n+1} = c_{m1}, c_{(m-1)n+2} = c_{m2}, \dots, c_{mn} = c_{m,n}$. Задача (1) с ограничениями (5), (6) примет следующий вид:

$$f(X) = \sum_{j=1}^{mn} c_j x_j \rightarrow \min \quad (4.9)$$

$$\sum_{i=1}^m x_{(i-1)n+j} = b_j, \quad j=1, \dots, n \quad (4.10)$$

$$\sum_{j=1}^n x_{(i-1)n+j} = a_i, \quad i=1, \dots, m, \quad (4.11)$$

$$x_k \geq 0, \quad k=1, \dots, mn$$

Таким образом, пришли к стандартной форме ЗЛП.

Для получения начального базисного решения можно использовать либо метод искусственного базиса, либо метод симплексного преобразования таблицы ограничений. Однако для транспортной задачи разработаны свои методы поиска начального базиса и методы последовательного улучшения решения. Это связано с следующими особенностями транспортной задачи.

- Система ограничений закрытой транспортной задачи представляет собой систему уравнений (т.е. транспортная задача задана в канонической форме) (см. (4.6), (4.7));
- Коэффициенты при переменных системы ограничений равны единице;
- Каждая переменная входит в систему ограничений два раза: один раз — в систему (4.6) и один раз — в систему (4.7).

4.3 Первоначальное закрепление потребителей за поставщиками

Первым этапом решения закрытой транспортной задачи является составление начального распределения (начального плана перевозок или начального базиса). При этом следует отметить, что благодаря условию (5) ранг системы линейных уравнений (6), (7) равен $m+n-1$; таким образом из общего числа $m \cdot n$ неизвестных базисных неизвестных будет $m+n-1$. Вследствие этого при любом допустимом базисном распределении в матрице перевозок (таблице поставок), представленной в общем виде в табл. 4.10, будет занято ровно $m+n-1$ клеток, которые будем называть *базисными* в отличие от остальных *свободных* клеток; занятые клетки будем отмечать диагональной чертой.

Таблица 4.10

| Мощности поставщиков | Мощности потребителей | | | |
|----------------------|-----------------------|----------|-----|----------|
| | b_1 | b_2 | ... | b_n |
| a_1 | c_{11} | c_{12} | ... | c_{1n} |
| | x_{11} | x_{12} | | x_{1n} |

| | | | | |
|-------|----------|----------|-----|----------|
| a_2 | c_{21} | c_{22} | ... | c_{2n} |
| | x_{21} | x_{22} | | x_{2n} |
| ... | ... | ... | ... | ... |
| a_m | c_{m1} | c_{m2} | ... | c_{mn} |
| | x_{m1} | x_{m2} | | x_{mn} |

Рассмотрим два метода получения начального распределения (начального опорного плана): метод северо-западного угла и метод наименьших стоимостей. При каждом из этих методов при заполнении некоторой клетки, кроме последней, вычеркивается или только строка матрицы перевозок, или только столбец; лишь при заполнении последней клетки вычеркиваются и строка, и столбец. Такой подход будет гарантировать, что базисных клеток будет ровно $m+n-1$. Если при заполнении некоторой (не последней) клетки одновременно удовлетворяются мощности и поставщика, и потребителя, то вычеркивается, например, только строка, а в соответствующем столбце заполняется незанятая клетка так называемой «нулевой поставкой», после чего вычеркивается и столбец. Для идентификации клетки обычно в скобках указываются номера ее строки и столбца. В методе северо-западного угла всегда в первую очередь заполняется клетка (из числа невычеркнутых), стоящая в верхнем левом (северо-западном) углу матрицы перевозок.

Пример составления начального распределения методом северо-западного угла показан в табл. 10: заполняется клетка (1;1) и вычеркивается первый столбец, заполняется клетка (1;2) и вычеркивается первая строка; заполняется клетка (2;2) и вычеркивается второй столбец; заполняется клетка (2;3) и вычеркивается вторая строка; заполняется клетка (3;3) и вычеркивается третий столбец; наконец, заполняется клетка (3;4) и вычеркиваются последние строка и столбец. Число занятых клеток равно $m+n-1=3+4-1=6$. Суммарные затраты на реализацию данного плана перевозок составят

$$f(X) = 4 \cdot 30 + 5 \cdot 30 + 3 \cdot 70 + 6 \cdot 30 + 7 \cdot 10 + 4 \cdot 110 = 1170.$$

Таблица 4.11

| Мощности поставщиков | Мощности потребителей | | | |
|----------------------|-----------------------|-----|----|-----|
| | 30 | 100 | 40 | 110 |
| 60 | 4 | 5 | 2 | 3 |
| | 30 | 30 | | |
| 100 | 1 | 3 | 6 | 2 |
| | | 70 | 30 | |
| 120 | 6 | 2 | 7 | 4 |
| | | | 10 | 110 |

Недостатком данного метода является то, что он не учитывает значения элементов c_{ij} матрицы транспортных расходов, в результате чего полученное этим методом

начальное распределение (начальный опорный план перевозок) может быть достаточно далеко от оптимального.

В методе наименьших стоимостей заполнение клеток матрицы перевозок проводится с учетом значений величин c_{ij} . Отмечают клетки с наименьшими стоимостями перевозок сначала по каждой строке, а затем по каждому столбцу. Клетки, имеющие две отметки, заполняют в первую очередь, затем заполняют клетки с одной отметкой, а данные о нераспределенном грузе записывают в неотмеченные клетки с наименьшими стоимостями. При этом из двух клеток с одинаковой стоимостью перевозок предпочтение отдается клетке, через которую осуществляется больший объем перевозок. Вычеркивание строк и столбцов при заполнении клеток проводится по описанным выше правилам. Пример начального распределения методом наименьших стоимостей для тех же исходных данных, что и ранее, представлен в табл. 4.12.

Таблица 4.12

| Мощности поставщиков | Мощности потребителей | | | |
|----------------------|-----------------------|-----|----|-----|
| | 30 | 100 | 40 | 110 |
| 60 | 4 | 5 | 2 | 3 |
| | | | 40 | 20 |
| 100 | 1 | 3 | 6 | 2 |
| | 30 | | | 70 |
| 120 | 6 | 2 | 7 | 4 |
| | | 100 | | 20 |

Порядок заполнения клеток: (2;1), (3;2), (1;3), (2;4), (1;4), (3;4). Суммарные затраты на перевозки, представленные в табл. 3, составляют

$$f(X) = 1 \cdot 30 + 2 \cdot 100 + 2 \cdot 40 + 2 \cdot 70 + 3 \cdot 20 + 4 \cdot 20 = 590.$$

Следовательно, данный план перевозок значительно ближе к оптимальному, чем план, составленный по методу северо-западного угла.

4.4 Метод потенциалов

Вторым этапом служат построение системы потенциалов.

Введем специальные показатели u_i для каждой строки, которые можно интерпретировать как цену продукта в пункте поставщика, а каждому столбцу j (j -му потребителю) устанавливается потенциал v_j , который можно принять условно за цену продукта в пункте потребителя. В простейшем случае цена продукта в пункте потребителя равна его цене в пункте поставщика плюс транспортные расходы на его доставку, т.е.

$$v_j = u_i + c_{ij} \quad (4.12)$$

Совокупность уравнений вида (4.12), составленных для всех заполненных клеток (всех базисных неизвестных), образует систему $m+n-1$ линейных уравнений с $m+n$ неизвестными u_i и v_j . Эта система всегда совместна, причем

значение одного из неизвестных можно задавать произвольно (например, $u_1 = 0$), тогда значения остальных неизвестных находятся из системы однозначно.

Рассмотрим процесс нахождения потенциалов для базисного начального распределения по методу северо-западного угла, представленного в табл. 4.11. Задав $u_1 = 0$ и используя формулу (4.12) для заполненных клеток (1;1) и (1;2), находим $v_1 = 4$ и $v_2 = 5$. Зная v_2 по заполненной клетке (2;2) находим $u_2 = 2$, а зная u_2 , по заполненной клетке (2;3) находим $v_3 = 8$. Зная v_3 по заполненной клетке (3;3) находим $u_3 = 1$, а затем по заполненной клетке (3;4) находим $v_4 = 5$. Результаты представлены в табл. 4.13, где потенциалы поставщиков приведены в последнем столбце, а потенциалы потребителей — в последней строке.

Таблица 4.13.

| Мощности поставщиков | Мощности потребителей | | | | |
|----------------------|-----------------------|-------------------------|-------------------------|----------|-------|
| | 30 | 100 | 40 | 110 3 | u_i |
| 60 | 4 30 | 5 – 30 | 2 + | 3 | 0 |
| 100 | 1 | 3 + 70 | 6 – 30 | 2 | 2 |
| 120 | 6 | 2 | 7 10 | 4 110 | 1 |
| v_j | 4 | 5 | 8 | 5 | |

Смысл выделенных клеток в табл. 4.13, и знаков при его вершинах пояснен далее при описании этапа улучшения плана метода потенциалов.

Аналогичные результаты для начального распределения по методу наименьших стоимостей, приведенного в табл. 4.12, представлены в табл. 4.14.

Таблица 4.14.

| Мощности поставщиков | Мощности потребителей | | | | |
|----------------------|-----------------------|----------|---------|----------|-------|
| | 30 | 100 | 40 | 110 3 | u_i |
| 60 | 4 | 5 | 2 40 | 3 20 | 0 |
| 100 | 1 30 | 3 | 6 | 2 70 | 1 |
| 120 | 6 | 2 100 | 7 | 4 20 | —1 |
| v_j | 2 | 1 | 2 | 3 | |

Чтобы оценить оптимальность распределения, для всех клеток (i; j) матрицы перевозок определяются их *оценки*, которые обозначим через d_{ij} , по формуле:

$$d_{ij} = (u_i + c_{ij}) - v_j. \quad (4.13)$$

Используя ранее принятую интерпретацию, выражение $(u_i + c_{ij})$ можно трактовать как сумму цены продукта у поставщика и стоимости перевозки; эта сумма путем вычитания сравнивается с ценой продукта у соответствующего потребителя v_j . Очевидно, оценки заполненных клеток равны нулю (цена потребителя покрывает цену поставщика и стоимость перевозок). Таким образом, об оптимальности распределения можно судить по величинам оценок свободных клеток. Если оценка некоторой свободной клетки отрицательна, это можно интерпретировать так: цена, предлагаемая соответствующим потребителем, больше суммы цены поставщика и стоимости перевозки, т.е. если бы эта клетка была занята, то можно было бы получить дополнительный экономический эффект. Следовательно, *условием оптимальности распределения служит условие неотрицательности оценок свободных клеток матрицы перевозок.*

Оценки клеток по формуле (4.13) удобно представить в виде *матрицы оценок*. Для ранее рассматриваемого распределения, полученного методом северо-западного угла (см. табл. 1), матрица оценок клеток имеет вид

$$d = \begin{pmatrix} 0 & 0 & -6 & -2 \\ -1 & 0 & 0 & -1 \\ 3 & -2 & 0 & 0 \end{pmatrix}. \quad (4.14)$$

Наличие большого числа отрицательных оценок свободных клеток свидетельствует о том, что данный план перевозок далек от оптимального (напомним, что суммарные затраты на перевозку по этому плану равны 1170).

Для распределения, полученного методом наименьших стоимостей (табл. 4.14), матрица оценок клеток имеет вид:

$$d = \begin{pmatrix} 2 & 4 & 0 & 0 \\ 0 & 3 & 5 & 0 \\ 3 & 0 & 4 & 0 \end{pmatrix}$$

Так как все оценки неотрицательны, то не имеется возможности улучшить данный план перевозок, т.е. он оптимален (суммарные затраты на перевозку по этому плану равны 590). Кроме того, следует отметить, что в данном случае оценки всех свободных клеток строго больше нуля, т.е. любой другой план, предусматривающий занятие хотя бы одной из этих клеток, будет менее оптимален. Это говорит о том, что найденный оптимальный план является *единственным*. Наличие нулевых оценок свободных клеток в оптимальном плане перевозок, наоборот, свидетельствует о *неединственности* оптимального плана.

4.5 Улучшение оптимального плана перевозок (циклы перераспределения)

Чтобы улучшить неоптимальный план перевозок, выбирается клетка матрицы перевозок с отрицательной оценкой; если таких клеток несколько, то обычно (но необязательно) выбирается клетка с наибольшей по абсолютной величине отрицательной оценкой. Например, для распределения, представленного в табл. 4.13, такой клеткой может служить клетка (1,3) (см. матрицу оценок (4.14)).

Для выбранной клетки строится замкнутая линия (*контур*, в табл. 4.13 эти клетки «затенены»), начальная вершина которой лежит в выбранной клетке, а все

остальные вершины находятся в занятых клетках; при этом направления отдельных отрезков контура могут быть только горизонтальными и вертикальными. Вершиной контура, кроме первой, является занятая клетка, где отрезки контура образуют один прямой угол (нельзя рассматривать как вершины клетки, где горизонтальные и вертикальные отрезки контура пересекаются). Очевидно, число отрезков контура, как и его вершин, будет четным. В вершинах контура расставляются поочередно знаки «+» и «—», начиная со знака «+» в выбранной свободной клетке. Пример простого контура показан пунктиром в табл. 4.13, хотя вид контура может быть самым разнообразным.

Величина перераспределяемой поставки определяется как наименьшая из величин поставок в вершинах контура со знаком «—», и на эту величину увеличиваются поставки в вершинах со знаком «+» и уменьшаются поставки в вершинах со знаком «-». Это правило гарантирует, что в вершинах контура не появятся отрицательных поставок, начальная выбранная клетка окажется занятой, в то время как одна из занятых клеток при этом обязательно освободится. Если величина перераспределяемой поставки равна поставкам не в одной, а в нескольких вершинах контура со знаком «-» (это как раз имеет место в контуре перераспределения в табл. 4.13), то освобождается только одна клетка, обычно с наибольшей стоимостью перевозки, а все другие такие клетки остаются занятыми с нулевой поставкой.

Результат указанных операций для представленного в табл. 4.13 распределения поставок показан в табл. 4.15, Суммарные затраты на перевозки по этому плану составляют

$$f(X) = 4 \cdot 30 + 5 \cdot 0 + 2 \cdot 30 + 3 \cdot 100 + 7 \cdot 10 + 4 \cdot 110 = 990,$$

что значительно меньше предыдущей суммы затрат 1170, хотя план перевозок в табл. 4.15 еще не является оптимальным. Об этом свидетельствует наличие отрицательных значений в матрице оценок клеток этого плана (соответствующие потенциалы u_i и v_j) найдены способом, изложенным при описании этапа 2):

$$d = \begin{pmatrix} 0 & 0 & 0 & 4 \\ -1 & 0 & 6 & 5 \\ -3 & -8 & 0 & 0 \end{pmatrix}.$$

Таблица 4.15.

| Мощности поставщиков | Мощности потребителей | | | | u_i |
|----------------------|-----------------------|-----|----|-----|-------|
| | 30 | 100 | 40 | 110 | |
| 60 | 4 | 5 | 2 | 3 | 0 |
| | 30 | 0 | 30 | | |
| 100 | 1 | 3 | 6 | 2 | 2 |
| | | 100 | | | |
| 120 | 6 | 2 | 7 | 4 | —5 |
| | | | 10 | 110 | |
| v_j | 4 | 5 | 2 | -1 | |

Транспортные задачи, в базисном плане перевозок которых имеют место занятые клетки с нулевой поставкой (или в первоначальном распределении, или в процессе итераций), называются *вырожденными*; пример такой задачи представлен в табл. 3. В случае вырожденной транспортной задачи существует опасность *зацикливания*, т.е. бесконечного повторения итераций (бесконечного перебора одних и тех же базисных комбинаций занятых клеток). Как правило, в практических задачах транспортного типа зацикливание не встречается; тем не менее следует знать, что существуют специальные правила, позволяющие выйти из цикла, если зацикливание все же произойдет. При отсутствии вырождения метод потенциалов конечен и приводит к оптимальному плану перевозок за конечное число шагов.

4.6 Открытая модель транспортной задачи

Если суммарная мощность поставщиков не равна суммарной мощности потребителей, (т. е. если нарушается условие $\sum_{i=1}^m a_i = \sum_{j=1}^n b_j$), то имеем открытую модель транспортной задачи. Открытая транспортная задача решается сведением ее к закрытой транспортной задаче.

Пример 4.5. Найти оптимальное распределение поставок для транспортной задачи, представленной в таблице 4.16.

Таблица 4.16.

| Мощности поставщиков | Мощности потребителей | | | |
|----------------------|-----------------------|----|----|----|
| | 45 | 35 | 55 | 65 |
| 40 | 4 | 1 | 2 | 5 |
| 60 | 3 | 2 | 3 | 7 |
| 90 | 4 | 4 | 5 | 2 |

Решение. В данном случае суммарный спрос потребителей больше, чем суммарная мощность поставщиков ($45+35+55+65 = 200 > 40+60+90 = 190$). Введем «фиктивного поставщика» и в таблицу поставок добавим дополнительную строку (табл. 4.17) так, чтобы задача стала закрытой. Для этого мощность фиктивного поставщика следует принять равной $10 = 200 - 190$. Коэффициенты затрат этой добавленной строки определяются издержками ввиду недогрузки мощностей потребителей. Если информация об этих издержках отсутствует, то их принимают равными одному и тому же числу (например, нулю, как в табл. 4.17). Конкретное значение этого числа не влияет на оптимальное распределение поставок.

Таблица 4.17.

| Мощности поставщиков | Мощности потребителей | | | |
|----------------------|-----------------------|----|----|----|
| | 45 | 35 | 55 | 65 |
| 40 | 4 | 1 | 2 | 5 |
| 60 | 3 | 2 | 3 | 7 |

| | | | | |
|----|---|---|---|---|
| 90 | 4 | 4 | 5 | 2 |
| 10 | 0 | 0 | 0 | 0 |

Решение

Первоначальное распределение поставок для сформулированной закрытой транспортной задачи найдем, например, по методу наименьших затрат. Для удобства укажем последовательность заполнения таблицы поставок: $x_{44} = 10$, $x_{12} = 35$, $x_{34} = 55$, $x_{13} = 5$, $x_{23} = 50$, $x_{21} = 10$, $x_{31} = 35$. В результате приходим к следующему базисному распределению поставок (табл. 4.18)

Таблица 4.18

| Мощности поставщиков | Мощности потребителей | | | | u_i |
|----------------------|-----------------------|----|-----|-----|-------|
| | 45 | 35 | 55 | 65 | |
| 40 | 4 | 1 | 2 | 5 | 0 |
| | | 35 | 5 | | |
| 60 | 3 + | 2 | 3 - | 7 | -1 |
| | 10 | | 50 | | |
| 90 | 4 - | 4 | 5 | 2 + | -2 |
| | 35 | | | 55 | |
| 10 | 0 | 0 | 0 + | 0 - | 0 |
| | | | | 10 | |
| v_j | 2 | 1 | 2 | 0 | |

Рассчитаем матрицу оценок по формуле $d_{ij} = (u_i + c_{ij}) - v_j$

$$d = \begin{pmatrix} 2 & 0 & 0 & 5 \\ 0 & 0 & 0 & 6 \\ 0 & 1 & 1 & 0 \\ -2 & -1 & -2 & 0 \end{pmatrix}.$$

Так как есть отрицательные оценки свободных клеток, то полученный план перевозок неоптимальный. Строим цикл перераспределения для клетки (4; 3). В результате получим следующий цикл (выделенные клетки в таблице 6). Далее производим перераспределение в соответствии с описанной ранее схемой. Результат представлен в таблице 4.18.

Таблица 4.18

| Мощности поставщиков | Мощности потребителей | | | | u_i |
|----------------------|-----------------------|----|----|----|-------|
| | 45 | 35 | 55 | 65 | |
| | | | | | |

| | | | | | |
|-------|----|----|----|----|----|
| 40 | 4 | 1 | 2 | 5 | 0 |
| | | 35 | 5 | | |
| 60 | 3 | 2 | 3 | 7 | -1 |
| | 20 | | 40 | | |
| 90 | 4 | 4 | 5 | 2 | -2 |
| | 25 | | | 65 | |
| 10 | 0 | 0 | 0 | 0 | 2 |
| | | | 10 | | |
| v_j | 2 | 1 | 2 | 0 | |

Матрица оценок для этой таблицы

$$d = \begin{pmatrix} 2 & 0 & 0 & 5 \\ 0 & 0 & 0 & 6 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}$$

не содержит отрицательных элементов свободных клеток, следовательно полученный план перевозок оптимальный. Суммарная стоимость перевозок составит

$$f(X) = 1 \cdot 35 + 2 \cdot 5 + 3 \cdot 20 + 3 \cdot 40 + 4 \cdot 25 + 2 \cdot 65 + 0 \cdot 10 = 455$$

В случае, когда суммарная мощность поставщиков больше суммарной мощности потребителей, в рассмотрение вводится «фиктивный потребитель», а к таблице поставок присоединяется дополнительный столбец. Коэффициенты затрат этого добавленного столбца соответствуют затратам на хранение не отправленного груза (поставки последнего столбца – не отправленный груз для каждого из поставщиков). Если информация об этих затратах отсутствует, то их принимают равными одному и тому же числу (например, нулю).

Вопросы для самопроверки

1. Сформулируйте экономико – математическую модель транспортной задачи
2. Приведите открытую модель транспортной задачи
3. Опишите многопродуктовую модель транспортной задачи
4. Сформулируйте модель производства с запасами
5. Укажите эквивалентность элементов производственной и транспортной системы
6. Как решить транспортную задачу симплексным методом?
7. В чем состоит суть метода северо-западного угла поиска начального базиса при решении транспортной задачи?
8. В чем состоит суть метода наименьших стоимостей поиска начального базиса при решении транспортной задачи?
9. Опишите суть метода потенциалов
10. Как улучшить оптимальный план перевозок в транспортной задаче (циклы перераспределения)?
11. Как решить открытую транспортную задачу?

Тема 5. Целочисленное программирование

Целочисленным (иногда его называют также *дискретным*) программированием называется раздел математического программирования, изучающий экстремальные задачи, в которых на искомые переменные накладывается условие целочисленности, а область допустимых решений конечна. Изучение этого раздела вызывается тем, что огромное количество экономических задач носит дискретный, чаще всего целочисленный характер, что связано, как правило, с физической неделимостью многих элементов расчета: например, нельзя построить два с половиной завода, купить полтора автомобиля и т.д. В ряде случаев такие задачи решаются обычными методами, например, симплексным методом, с последующим округлением до целых чисел. Однако такой, подход оправдан, когда отдельная единица составляет очень малую часть всего объема (например, товарных запасов); в противном случае он может внести значительные искажения в действительно оптимальное решение.

5.1 Графический метод решения ЗЦП

Приводимый ниже пример ясно показывает недостаточность таких методов и позволяет лучше понять неуловимую трудность ЗЦП.

Пример 5.1.

$$\begin{aligned} f(x) &= x_1 - 20x_2 \rightarrow \min \\ -x_1 + 10x_2 &\leq 40 \\ 4x_1 + 2x_2 &\leq 29 \\ x_j &\geq 0; \quad x_j \in \mathbf{Z} \quad j=1,2 \end{aligned}$$

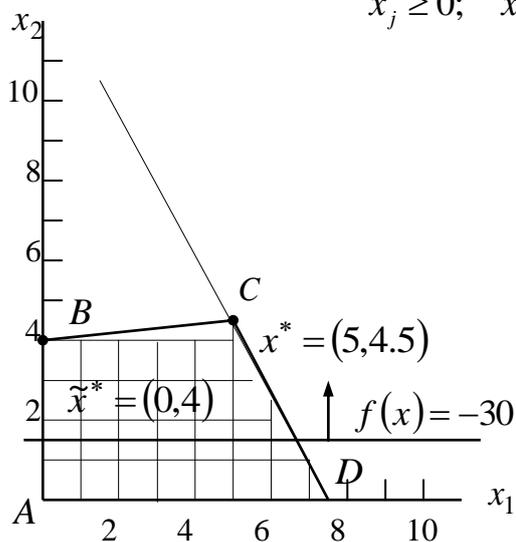


Рис. 5.1

На плоскости (x_1, x_2) (см. рис. 5.1) допустимое множество S — многоугольник $ABCD$ и отметим точки множества S с целочисленными координатами.

Здесь множество \tilde{S} отмечено отдельными точками внутри многогранника $ABCD$.

Перемещая линию уровня ЦФ $f(x)$ в направлении антиградиента $\nabla f = (-1, 20)$ (убывания f), находим крайнее положение этой линии, в котором она имеет непустое пересечение с множеством \tilde{S} . В этом положении линия уровня проходит через точку

$$B(0,4), \text{ поэтому решение задачи имеет вид } \tilde{x}^* = (0,4); \quad \tilde{f}^* = \min_{x \in \tilde{S}} f(x) = -80.$$

Из рисунка видно, что в случае непрерывной переменной оптимум есть точка $C(5,4.5)$, т.е.

$$x^* = (5,4.5); \quad f^* = -85$$

Простой метод округления привел бы к решению $x_1 = 5; x_2 = 5$, которое не удовлетворяет ограничениям. отсюда следует, что точка минимума ЦФ на допустимом множестве \tilde{S} целочисленной задачи не обязательно является ближайшей к решению x^* обычной задачи ЛП.

Замечание. Если округлить в сторону уменьшения, т.е. взять $\tilde{x}(5,4), \tilde{f} = -75$, то эта точка будет удовлетворять ограничениям задачи.

Для решения целочисленных задач разработаны специальные методы, которые можно разделить на две группы: методы отсечения (отсекающих плоскостей) и комбинаторные методы.

Метод отсекающих плоскостей состоит в построении дополнительных ограничений и применении модифицированного симплексного метода (метод Гомори). Представление о *комбинаторных методах* дает широко используемый на практике *метод ветвей и границ*.

По методу Гомори первый этап решения целочисленных задач не отличается от обычного расчета по симплексному алгоритму. Если среди значений переменных в оптимальном плане есть дробные, то составляется дополнительное ограничение, отсекающее дробную часть решения, но оставляющее в силе все прочие условия, которым должен удовлетворять оптимальный план. Это дополнительное ограничение присоединяется к исходным ограничениям задачи, и вновь применяется процедура симплексного метода. Алгоритм Гомори позволяет прийти к оптимальному целочисленному решению за конечное число шагов.

5.2 Метод Гомори (МГ)

МГ используется для решения задач ЦП с произвольным числом переменных.

Суть МГ: последовательное отсечение от допустимого множества S нецелочисленной задачи частей, не содержащих точек с целыми координатами. Эти отсечения производятся включением в задачу дополнительных ограничений на переменные x_j .

Задача ЦП имеет вид

$$f(x) \rightarrow c^t x \rightarrow \min \quad (5.1)$$

$$Ax = b \quad (5.2)$$

$$x \geq 0; \quad x \in \mathbf{Z} \text{ - множество целых чисел} \quad (5.3)$$

5.2.1 Алгоритм МГ с использованием СМ

Шаг 1 С помощью СМ, находится решение x^* ЗЛП без учета требования целочисленности (5.3). Если для x^* условие (5.3) выполняется, то задача решена. В противном случае среди чисел \bar{b}_i последнего столбца СТ, определяющей решение x^* , есть такие, что $\{\bar{b}_i\} > 0$.

Шаг 2 Среди нецелых элементов \bar{b}_i выбирается произвольный элемент \bar{b}_r (например, с максимальной дробной частью $\{\bar{b}_r\}$). По r -ой строке СТ составляется дополнительное ограничение вида

$$- \sum_{j=m+1}^n \{\bar{a}_{rj}\} x_j \leq -\{\bar{b}_r\}$$

(здесь для определенности полагаем, что свободные переменные x_j имеют номера $m+1, m+2, \dots, n$).

С помощью вспомогательной переменной $x_{n+1} \geq 0$ это ограничение представляется в виде равенства

$$x_{n+1} - \sum_{j=m+1}^n \{a_{rj}\} x_j = -\{\bar{b}_r\}$$

и вводится в СТ дополнительной строкой.

$$x_{n+1} \mid \alpha_{n+1,m+1}, \dots, \alpha_{n+1,n} \mid \beta_{n+1} \quad (5.4)$$

где $\alpha_{n+1,j} = -\{\bar{a}_{r,j}\}, \quad j = m+1, \dots, n$

$$\beta_{n+1} = -\{\bar{b}_r\}$$

Так как $\beta_{n+1} = -\{\bar{b}_r\} < 0$, то после дополнения строкой (5.4) СТ перестает соответствовать допустимому базисному решению ЗЛП, которую она описывает.

Шаг 3 Для перехода к допустимому базисному решению производятся следующие операции:

а) строка с отрицательным свободным членом β_{n+1} считается опорной

б) если все коэффициенты $\alpha_{n+1,j} > 0, j = m+1, \dots, n$, то задача не имеет решения, в противном случае номер ℓ разрешающего столбца находится из условия

$$\frac{\beta_{n+1}}{|\alpha_{n+1,\ell}|} = \min_{j: \alpha_{n+1,j} < 0} \frac{\beta_{n+1}}{|\alpha_{n+1,j}|} \quad (5.5)$$

в) совершается преобразование СТ с опорным элементом $\alpha_{n+1,\ell}$. Если в новой СТ по-прежнему есть хотя бы один отрицательный свободный член, то описанная процедура повторяется, начиная с операции (а), необходимое число раз.

Если все элементы \bar{b}_i новой СТ ≥ 0 , то допустимое базисное решение найдено. Отметим, что выбор опорного элемента $\alpha_{n+1,\ell}$ гарантирует неотрицательность коэффициентов \bar{b}_i новой СТ. Поэтому найденное допустимое решение является и оптимальным.

Шаг 4 Если найденное на Шаг 3 решение ЗЛП удовлетворяет условию целочисленности, то – «останов», если нет \rightarrow Шаг 2.

Описанный алгоритм позволяет найти решение полностью целочисленной ЗЛП или установить отсутствие решений за конечное число итераций.

Пример 5.2. Пусть для приобретения оборудования, размещаемого на производственной площади 38 м^2 , фирма выделяет 20 млн. руб. Имеются единицы оборудования двух типов: типа А стоимостью 5 млн. руб., требующее производственную площадь 8 м^2 и имеющее производительность 7 тыс. единиц продукции за смену, и типа Б — стоимостью 2 млн. руб., занимающее площадь 4 м^2 и дающее за смену 3 тыс. единиц продукции. Требуется рассчитать оптимальный вариант приобретения оборудования, обеспечивающий максимум производительности участка.

Сформулируем экономико-математическую модель задачи. Пусть x_1, x_2 — количество приобретаемых машин типа А и типа Б соответственно. Тогда целевая функция задачи будет иметь вид:

$$f(X) = 7x_1 + 3x_2 \rightarrow \max$$

при ограничениях:

$$5x_1 + 2x_2 \leq 20,$$

$$8x_1 + 4x_2 \leq 38,$$

$$x_{1,2} \geq 0; x_{1,2} \in \mathbf{Z}.$$

Сформулирована задача линейного целочисленного программирования .

Введем дополнительные переменные x_3, x_4 , с помощью которых исходные неравенства преобразуются в равенства:

$$5x_1 + 2x_2 + x_3 = 20,$$

$$8x_1 + 4x_2 + x_4 = 38,$$

$$x_1, x_2, x_3, x_4 \geq 0; x_i \in \mathbf{Z}.$$

из которых следует, что переменные x_3, x_4 могут принимать только неотрицательные целочисленные значения.

Далее решаем задачу симплексным методом (без учета целочисленности):

| СТ-1 | | | | СТ-3 | | | |
|-------|-------|-------|-------|-------|-------|-------|-------|
| | x_1 | x_2 | b_i | | x_3 | x_4 | b_i |
| x_3 | 5 | 2 | 20 | x_1 | 1 | -1/2 | 1 |
| x_4 | 8 | 4 | 38 | x_2 | -2 | 5/4 | 15/2 |
| f | -7 | -3 | 0 | f | 1 | 1/4 | 59/2 |

| СТ-2 | | | |
|-------|-------|-------|-------|
| | x_3 | x_2 | b_i |
| x_1 | 1/5 | 2/5 | 4 |
| x_4 | -8/5 | 4/5 | 6 |
| f | 7/5 | -1/5 | 28 |

Из таблицы СТ-3 видно, что в оптимальном плане $x_1 = 1; x_2 = 7,5$ и максимум целевой функции равен $f(X) = 7 \cdot 1 + 3 \cdot 7,5 = 29,5$. Полученное решение не удовлетворяет условию целочисленности, поэтому дополняем последнюю СТ строкой (4)

| СТ-4 | | | |
|-------|-------|-------|-------|
| | x_3 | x_4 | b_i |
| x_1 | 1 | -1/2 | 1 |
| x_2 | -2 | 5/4 | 15/2 |
| x_5 | 0 | -1/4 | -1/2 |
| f | 1 | 1/4 | 59/2 |

СТ - 5

| | x_3 | x_5 | b_i |
|-------|-------|-------|-------|
| x_1 | 1 | 2 | 2 |
| x_2 | -2 | 5 | 5 |
| x_4 | 0 | 4 | 2 |
| f | 1 | 1 | 29 |

Повторив процесс решения симплексным методом для данной расширенной системы ограничений, получим новый оптимальный план, в котором переменные, входящие в базис, принимают целые значения: $x_1 = 2$; $x_2 = 5$; $x_4 = 2$. Таким образом, приобретение двух машин типа А и пяти машин типа Б обеспечивает максимум производительности участка, равный 29 тыс. единиц продукции в смену. Заметим, что если бы в качестве плана был выбран вариант, получаемый в результате округления первоначального решения задачи симплексным методом ($x_1 = 1$; $x_2 = 7^*$), то суммарная производительность была бы равна лишь 28 тыс. единиц продукции.

5.2.2. Решение частично целочисленных задач методом Гомори

Если требованию целочисленности подчинены не все переменные ЗЛП, то такая задача называется частично целочисленной.

Для решения частично целочисленных задач также используется метод Гомори, но его алгоритм в этом случае отличается видом коэффициентов $\alpha_{n+1,j}$ в дополнительной строке

$$x_{n+1} \mid \alpha_{n+1,m+1}, \dots, \alpha_{n+1,n} \mid \beta_{n+1}.$$

Если переменная x_i подчинена требованию целочисленности, то имеем

$$a) \alpha_{n+1,j} = \begin{cases} -\bar{a}_{rj}, & \text{если } \bar{a}_{rj} \geq 0 \\ \frac{\{\bar{b}_r\}}{1 - \{\bar{b}_r\}} \cdot \bar{a}_{rj}, & \text{если } \bar{a}_{rj} < 0 \end{cases} \quad (5.6)$$

Если же переменная x_i не подчинена требованию целочисленности, тогда

$$b) \alpha_{n+1,j} = \begin{cases} \{\bar{a}_{rj}\}, & \text{если } \{\bar{a}_{rj}\} \leq \{\bar{b}_r\} \\ \frac{\{\bar{b}_r\}}{1 - \{\bar{b}_r\}} \cdot (\{\bar{a}_{rj}\} - 1) & \text{если } \{\bar{a}_{rj}\} > \{\bar{b}_r\} \end{cases} \quad (5.7)$$

Вычисления заканчиваются, когда целыми являются не обязательно все коэффициенты \bar{b}_i , а только те, которым соответствуют переменные x_i подчиненные требованию целочисленности.

Пример 5.3. Решить частично целочисленную задачу линейного программирования.

$$f(x) = x_1 - 10x_2 \rightarrow \min$$

$$3x_1 + x_2 \leq 12$$

$$-8x_1 + 3x_2 \leq 24$$

$$x_j \geq 0; \quad x_2 \in \mathbf{Z}$$

Запишем задачу в каноническом виде:

$$f(x) = x_1 - 10x_2 \rightarrow \min$$

$$3x_1 + x_2 + x_3 = 12$$

$$-8x_1 + 3x_2 + x_4 = 24$$

$$x_2 \in \mathbf{Z}$$

| | x_1 | x_2 | |
|-------|-------|-------|----|
| x_3 | 3 | 1 | 12 |
| x_4 | -8 | 3 | 24 |
| | 1 | -10 | 0 |

| | x_1 | x_4 | |
|-------|-------|-------|----|
| x_3 | 17/3 | -1/3 | 4 |
| x_2 | -8/3 | 1/3 | 8 |
| | -77/3 | 10/3 | 80 |

| | x_3 | x_4 | |
|-------|-------|-------|---------|
| x_1 | 3/17 | -1/17 | 12/17 |
| x_2 | 8/17 | 3/17 | 168/17 |
| | 77/17 | 31/17 | 1668/17 |

Найденное решение $x^* = (12/17, 168/17)$ по второй переменной не удовлетворяет условию целочисленности.

По второй строке составим дополнительное ограничение (см. (5.6)).

| | x_3 | x_4 | |
|-------|-------|-------|---------|
| x_1 | 3/17 | -1/17 | 12/17 |
| x_2 | 8/17 | 3/17 | 168/17 |
| x_5 | -8/17 | -3/17 | -15/17 |
| | 77/17 | 31/17 | 1668/17 |

| | x_3 | x_5 | |
|-------|-------|-------|------|
| x_1 | 1 | -1/3 | 1 |
| x_2 | 0 | 1 | 9 |
| x_4 | 8/3 | -17/3 | 15/3 |
| | -1/3 | 31/3 | 89 |

| | x_1 | x_5 | |
|-------|-------|-------|-------|
| x_3 | 1 | -1/3 | 1 |
| x_2 | 0 | 1 | 9 |
| x_4 | -8/3 | -43/9 | 7/3 |
| | 1/3 | 92/9 | 268/3 |

Таким образом, получим: $\tilde{x}^* = (0, 9)$; $f^* = -268/3$.

5.3. Метод ветвей и границ

Метод ветвей и границ (МВГ) широко используется на практике для решения как полностью целочисленных задач, так и смешанных задач ЦЛП. Он применяется в большинстве коммерческих программ решения задач ЦП. По существу МВГ представляет собой эффективную процедуру перебора всех целочисленных допустимых решений.

Как известно, в методе Гомори решение основано на отсечении нецелочисленных допустимых решений. В МВГ используется округление нецелочисленного оптимального решения ЗЛП.

Пример 5.4. Пусть оптимальное решение двумерной ЗЛП есть $x^* = (3,5; 4,4)$. В качестве кандидатов на роль приближенного целочисленного оптимального решения необходимо рассматривать решения $(3;4)$, $(4;4)$, $(4;5)$, $(3;5)$, полученные в результате округления. Истинное оптимальное целочисленное решение может не совпадать ни с одним из 4-х, поскольку целое значение x_1 в оптимальном решении может быть >4 или <3 . Таким образом, для получения истинного оптимального целочисленного решения приходится рассматривать все возможные значения x_1 большие и меньшие 3,5. Другими словами, оптимальное целочисленное значение x_1 должно удовлетворять неравенству $x_1 \leq 3$ либо неравенству $x_1 \geq 4$. Аналогично по переменной x_2 – либо $x_2 \leq 4$, либо $x_2 \geq 5$. При наличии в задаче ЦЛП большого количества переменных важно иметь процедуру, позволяющую систематически перебирать все возможные целочисленные решения, получаемые при округлении оптимального решения ЗЛП. МВГ по существу представляет собой такую процедуру эффективного перебора целочисленных решений, получаемых при округлении оптимального решения.

Для иллюстрации основных принципов МВГ рассмотрим следующую задачу ЦП

Пример 5.5.

$$F = 3x_1 + 2x_2 \rightarrow \max$$

$$x_1 \leq 2; x_2 \leq 2;$$

$$x_1 + x_2 \leq 3,5 \quad (\text{ЛП-1})$$

$$x_1, x_2 \geq 0$$

$$x_1, x_2 \in \mathbf{Z}$$

Начальный шаг решения этой задачи состоит в нахождении решения ЗЛП без учета целочисленности x_1 и x_2 . На рис. 5.2 представлено графическое решение ЗЛП. Оптимальное решение задачи ЛП-1 имеет вид $x^* = (2; 1,5)$; $F^* = 9$.

Так как $x_2 = 1,5$, то найденное решение не может быть оптимальным решением исходной задачи ЦЛП. Но найденное значение $F^* = 9$ представляет собой верхнюю границу истинного оптимального решения, поскольку при выполнении требования целочисленности x_2 значение F может лишь уменьшиться.

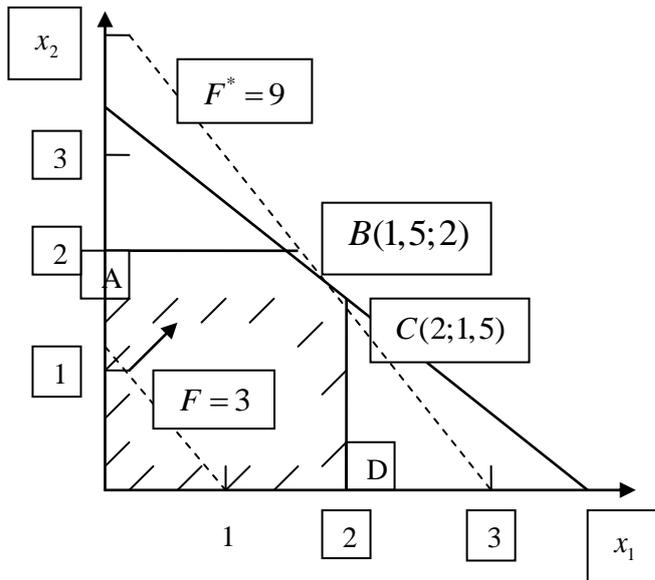


Рис. 5.2. Решение задачи ЛП-1

Следующий шаг МВГ состоит в просмотре целочисленных значений x_2 больших или меньших 1,5. Это делается путем добавления к задаче ЛП-1 нового ограничения: либо $x_2 \leq 1$, либо $x_2 \geq 2$. Таким образом, из задачи ЛП-1 получаются 2 задачи следующего вида

ЛП-2

$$F = 3x_1 + 2x_2 \rightarrow \max$$

$$x_1 \leq 2; x_2 \leq 2;$$

$$x_2 \leq 1 \quad \text{и} \quad x_1 + x_2 \leq 3,5$$

$$x_1 + x_2 \leq 3,5$$

$$x_1, x_2 \geq 0$$

$$x_1, x_2 \in \mathbf{Z}$$

ЛП-3

$$F = 3x_1 + 2x_2 \rightarrow \max$$

$$x_1 \leq 2; x_2 \leq 2;$$

$$x_2 \geq 2 \quad \text{и} \quad x_1 + x_2 \leq 3,5$$

$$x_1 + x_2 \leq 3,5$$

$$x_1, x_2 \geq 0$$

$$x_1, x_2 \in \mathbf{Z}$$

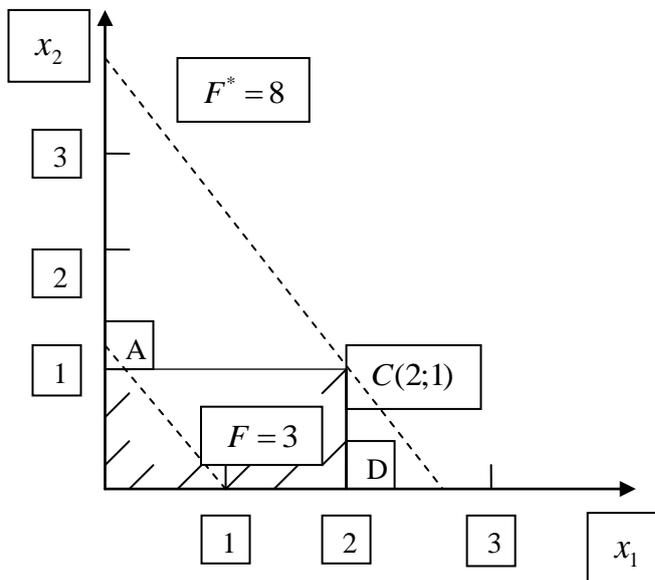


Рис. 5.3. Решение задачи ЛП-2

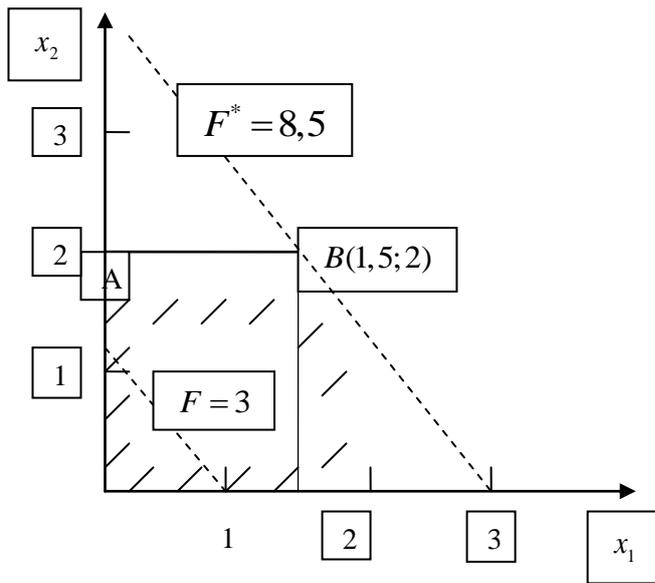


Рис. 5.4. Решение задачи ЛП-3

На рис.5.3 и рис.5.4 изображены допустимые области задач ЛП-2 и ЛП-3 соответственно ($S_{\bar{E}\bar{I}-3} \equiv \partial \bar{\Delta} \hat{c} \hat{t} \hat{e} AB$). $S_{\bar{E}\bar{I}-2}$ и $S_{\bar{E}\bar{I}-3}$ обладают следующими свойствами

1. Оптимальное решение задачи ЛП-1 ($x_1 = 2; x_2 = 1,5$) недопустимо для обеих задач ЛП-2 и ЛП-3. Таким образом, это решение не повторяется.
2. Любое целочисленное (допустимое) решение исходной задачи допустимо и для ЛП-2 и ЛП-3. Таким образом, при введении этих задач не происходит потери допустимых целочисленных решений исходной задачи.

Оптимальное решение задачи ЛП-2 (рис. 5.3) – это точка $C(2;1)$, т.е. $x^* = (2;1)$, $F^* = 8$. Таким образом, получено допустимое (целочисленное) решение исходной задачи ЦП. Даже если ЛП-2 имеет другие целочисленные решения, значение ЦФ в них не может быть больше 8. Т.е. $F^* = 8$ – нижняя граница максимального значения F для исходной задачи ЦЛП. Другими словами, оптимальное значение F исходной задачи не может быть < 8 . А так как ранее получена лишь верхняя граница, равная 9, то нельзя утверждать, что решение ЛП-2 оптимально для исходной задачи. Следовательно, необходимо также рассмотреть задачу ЛП-3.

Оптимальное решение задачи ЛП-3 (рис. 5.4) $x_1 = 1,5; x_2 = 2; F^* = 8,5$.

Для исходной задачи это недопустимо, так как x_1 принимает дробное значение. Оптимальное решение ЛП-3 $F^* = 8,5 > 8$ (нижней границы). Поэтому необходимо проверить существование в допустимой области $S_{\bar{E}\bar{I}-3}$ целочисленного решения, дающего значение ЦФ $F \geq 8$. Для этого рассмотри задачи ЛП-4 и ЛП-5, получающиеся при добавлении к ЛП-3 ограничений $x_1 \leq 1$ и $x_1 \geq 2$ соответственно. $S_{\bar{E}\bar{I}-4}$ состоит из отрезка DE (рис. 5), а задача ЛП-5 не имеет допустимых решений, т.е. $S_{\bar{E}\bar{I}-5} = \emptyset$.

Оптимальное решение ЛП-4 $x^* = (1;2); F^* = 7$.

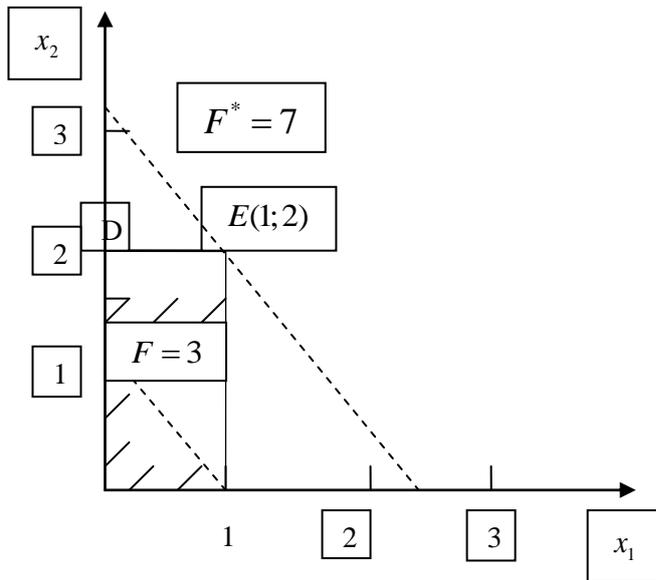


Рис. 5.5. Решение задачи ЛП-4

Следовательно, для любого целочисленного решения в $S_{\text{цел}}^{\text{LП-4}}$ значение ЦФ ≤ 7 . Таким образом, точка $x^*(2;1)$ задачи ЛП-2 представляет собой оптимальное целочисленное решение исходной задачи; оптимальное значение ЦФ в этой точке равно 8.

Удобно представить последовательность задач ЛП, возникающих при использовании процедуры МВГ в виде дерева (рис. 5.6)

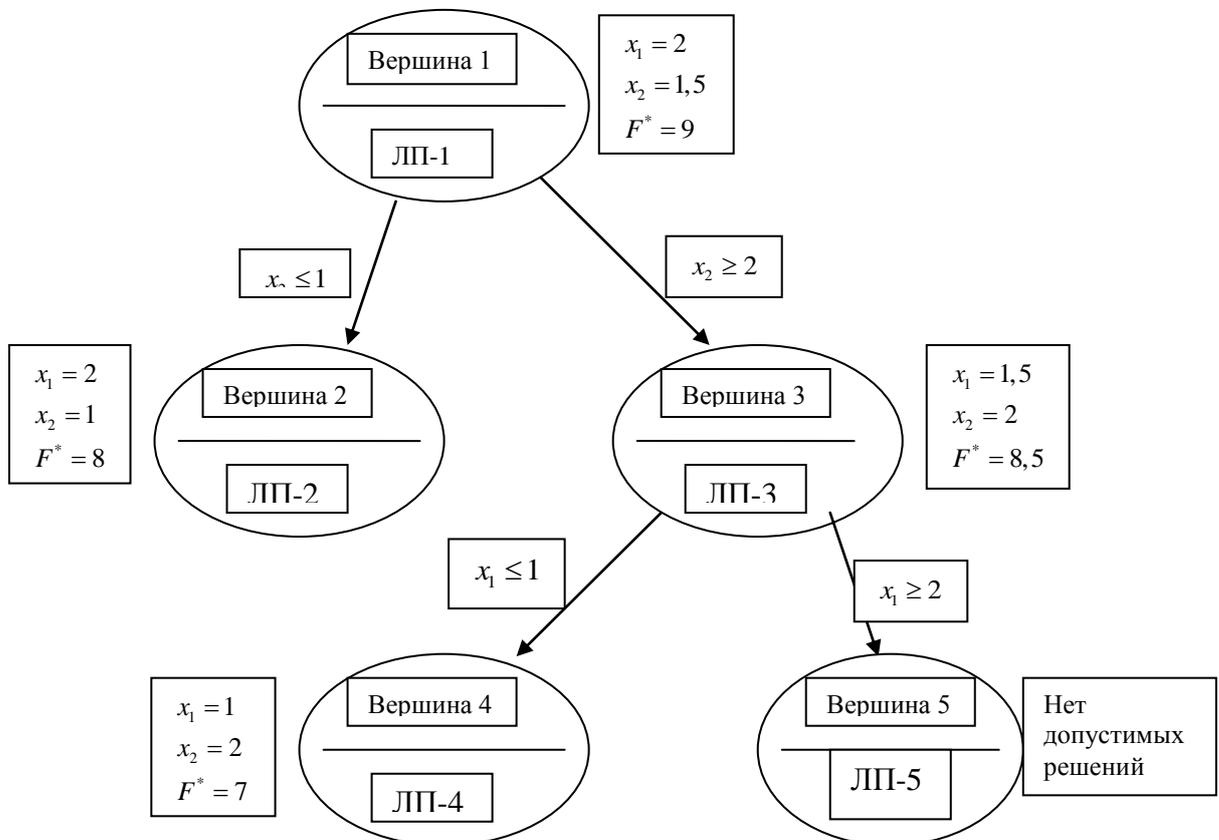


Рис. 5.6. Дерево последовательности ЗЛП, возникающих при использовании МВГ

Дерево состоит из множества вершин и соединяющих их дуг или ветвей. Вершина 1 соответствует ЛП-1 без учета требований целочисленности. Ветвление в вершине 1 определяемое целочисленной переменной x_2 с помощью ограничения $x_2 \leq 1$ приводит к вершине 2 (ЛП-2). Так как ЛП-2 имеет оптимальное целочисленное решение, то нет необходимости производить ветвление в вершине 2. Такая вершина называется *прозондированной*.

Ветвление в вершине 1. по ограничению $x_2 \geq 2$ дает ЛП-3. Так как оптимальное решение ЛП-3 дробное, происходит дальнейшее ветвление в вершине 3 по переменной x_1 . Это приводит к появлению вершин 4 и 5. Эти вершины прозондированы, поскольку ЛП-4 обладает оптимальным целочисленным решением, а ЛП-5 не имеет допустимых решений.

Наилучшее решение в прозондированных вершинах и есть оптимальное решение исходной задачи.

5.3.1 Алгоритм МВГ

Рассмотрим частично целочисленную задачу вида

$$F = c^T x \rightarrow \max$$

$$Ax = b, \quad x \geq 0,$$

$$x_j \in \mathbf{Z} (j \in I)$$

где I – множество индексов целочисленных переменных.

Шаг 1. На первом шаге решается задача ЛП-1, где все ее переменные рассматриваются как непрерывные. Пусть в оптимальном решении F_1 ЛП-1 некоторые целочисленные переменные принимают дробные значения, тогда оптимальное решение исходной задачи не совпадает с F_1 . В этом случае F_1 представляет собой верхнюю границу оптимального значения F^* исходной задачи ЛП-1.

Шаг 2. Производится ветвление по одной из целочисленных переменных, имеющей дробное значение в оптимальном решении ЛП-1. Для определения переменной, по которой производится ветвление, разработан ряд правил. Приведем некоторые из них.

- 1) Выбор целочисленной переменной, значение которой в оптимальном решении ЛП-1 имеет наибольшее дробное значение.
- 2) Приписывание целочисленным переменным приоритетов и ветвление по переменной с наибольшим приоритетом, например:
 - А) данная переменная представляет собой важное решение, принимаемое в рамках рассматриваемой модели;
 - Б) ее коэффициент стоимости или прибыли в ЦФ существенно превосходит остальные;
 - С) значение данной переменной играет ключевую роль для модели с точки зрения разработчиков и пользователей.
- 3) Произвольные правила выбора. Например, можно выбирать переменную с наименьшим номером.

Пусть ветвление происходит по целочисленной переменной x_j , дробное значение которой в оптимальном решении ЛП-1 равно β_j . Далее рассматриваются две новые задачи ЛП-2 и ЛП-3, получаемые путем введения ограничений $x_j \leq \lfloor \beta_j \rfloor$ и $x_j \geq \lceil \beta_j \rceil$, соответственно, где $\lfloor \beta_j \rfloor$ – наименьшее целое $\leq [x_j]$; здесь $[x_j]$ – целое значение переменной x_j ; $\lceil \beta_j \rceil$ – наибольшее целое $> [x_j]$.

Условия ЛП-2 и ЛП-3 можно записать следующим образом (см. рис. 7)

$$\begin{aligned} & \text{ЛП-2} \\ & F = c^T x \rightarrow \max \end{aligned}$$

$$Ax = b, \quad x_j \leq \underline{\beta}_j$$

$$x \geq 0$$

ЛП-3

$$F = c^T x \rightarrow \max$$

$$Ax = b, \quad x_j \geq \overline{\beta}_j$$

$$x \geq 0$$

Допустим, что оптимальные решения задач ЛП-2 и ЛП-3 также содержат дробные значения целочисленных переменных и поэтому не являются допустимыми для исходной задачи.

Шаг 3. Выбрать задачу ЛП-2 или ЛП-3 и произвести ветвление в соответствующей вершине, вводя новое ограничение. Выбор вершины (задачи ЛП) осуществляется с помощью специальных правил

- 1) Использование оптимального значения ЦФ. Для дальнейшего ветвления следует выбирать вершину, соответствующую наибольшему оптимальному значению ЦФ задачи ЛП.
- 2) Правило «последним пришел – первым обслужил». Для дальнейшего ветвления выбирается задача ЛП, решавшаяся последней.

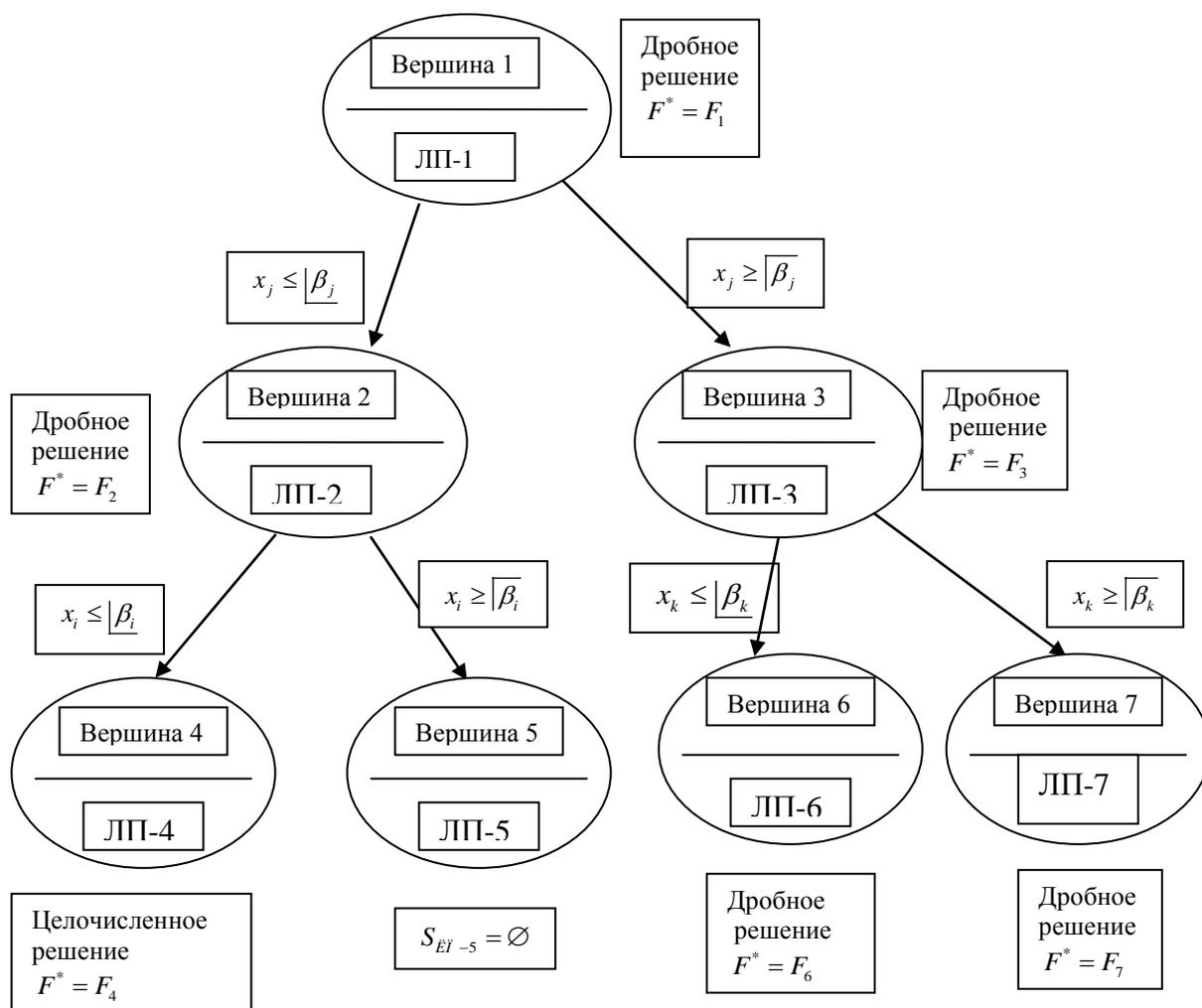


Рис. 5.7. Алгоритм МВГ

После выбора вершины для дальнейшего ветвления выбирается целочисленная переменная, которая имеет в оптимальном решении соответствующей ЗЛП дробное значение, и производится ветвление по этой переменной. Процесс ветвления и решения задач ЛП продолжается до получения целочисленного оптимального решения одной из задач ЛП. Значение F в полученной точке представляет собой *истинную границу* оптимального значения ЦФ исходной задачи ЦЛП. На этом этапе отбрасываются все вершины ЗЛП, для которых оптимальное значение F не превосходит полученной нижней границы. Про такие вершины говорят, что они являются *прозондированными*, поскольку в соответствующих им допустимых областях нет целочисленных решений, лучших, чем уже полученные.

В качестве иллюстрации рассмотрим следующее дерево.

Целочисленное оптимальное решение ЛП-4 дает нижнюю границу F_4 ЦФ исходной задачи (ЛП-1). Другими словами, оптимальное решение исходной задачи (ЛП-1) не может давать значение F меньше, чем F_4 . Дальнейшее ветвление в вершине В4 излишнее, так как для любой из последующих задач оптимальное значение не может быть $> F_4$. Таким образом, вершина В4 является *прозондированной*.

Вершина В5 также прозондирована, поскольку $S_{\bar{e}\bar{i}-5} = \emptyset$. Следовательно, в дальнейшем ветвление можно производить только в В6 и В7.

Предположим, что $F_6 < F_4$, а $F_7 > F_4$. Это значит, что В6 также прозондирована (неявным образом). Но так как $F_7 > F_4$, в $S_{\bar{e}\bar{i}-7}$ может найтись целочисленное решение со значением F , большим F_4 . Таким образом, для дальнейшего ветвления необходимо выбрать В7.

Вершина (ЗЛП) является прозондированной (явным или неявным образом) в том случае, если она удовлетворяет хотя бы одному из условий:

- 1) Оптимальное решение, соответствующее данной вершине, целочисленно. В этом случае полученное решение допустимо для исходной задачи ЦЛП.
- 2) ЗЛП, соответствующая рассматриваемой вершине, не имеет допустимых решений.
- 3) Оптимальное значение F соответствующей ЗЛП не превосходит нижней границы.

При использовании МВГ выбор вершин для дальнейшего ветвления происходит до тех пор, пока остается хотя бы одна непрозондированная вершина. Прозондированная вершина с наилучшим значением F дает оптимальное решение исходной задачи ЦЛП.

Рассмотрим далее ряд специальных оптимизационных задач, сводящихся к задачам линейного целочисленного программирования. Одной из таких задач является *задача о назначениях*, с помощью которой можно получить ответ на вопросы типа: как распределить рабочих по станкам, чтобы общая выработка была наибольшей или затраты на заработную плату наименьшими; как наилучшим образом распределить экипажи самолетов; как назначить людей на различные должности (отсюда и название задачи) и т.д.

5.4 Задача о назначениях

Математически такие задачи относятся к тому же типу распределительных задач, что и рассмотренная в лекции 6 транспортная задача, с той особенностью, что в них объемы наличных и требующихся ресурсов для выполнения каждой работы равны единице ($a_i = b_j = 1$), а все переменные x_{ij} либо равны единице, если i -й работник

назначен на j -ю работу, либо равны нулю в других случаях. Исходные данные задачи о назначениях группируются в таблице, которая называется *матрицей оценок*, а результаты — в *матрице назначений*.

Задача о назначениях в общем виде может быть сформулирована следующим образом. Имеется n работников, которые могут выполнять n работ, причем использование i -го работника на j -й работе, например, приносит доход c_{ij} . Требуется поручить каждому из работников выполнение одной вполне определенной работы, чтобы максимизировать в данном случае суммарный доход.

Введем переменные:

$$x_{ij} = \begin{cases} 1, & \text{если } i\text{-й работник выполняет } j\text{-ю работу} \\ 0 & \text{в противном случае} \end{cases}$$

Задача состоит в том, чтобы найти распределение $X = (x_{ij})$ работников по работам (т. е. найти матрицу назначений), которое максимизирует целевую функцию

$$f(X) = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} \rightarrow \max \quad (5.6)$$

при ограничениях

$$\sum_{j=1}^n x_{ij} = 1, \quad i = 1, \dots, n; \quad (5.7)$$

$$\sum_{i=1}^n x_{ij} = 1, \quad j = 1, \dots, n. \quad (5.8)$$

причем x_{ij} равно либо 0, либо 1 (так называемые *булевы переменные*) для всех $i, j = 1, \dots, n$.

Ограничения (5.7) отражают условие того, что за каждым работником может быть закреплена только одна работа, а ограничения (5.8) означают, что для выполнения каждой работы может быть выделен только один работник.

Если в задаче о назначениях элементы матрицы оценок представляют собой, например, время выполнения каждым работником любой из работ, то целевая функция этой задачи будет минимизироваться.

5.5 Задача о коммивояжере

Другой задачей подобного рода является *задача о коммивояжере*, которая может быть сформулирована следующим образом. Имеется n городов, пронумерованных числами от 1 до n . Коммивояжер, выезжая из города 1, должен побывать в каждом городе ровно один раз и вернуться в исходный пункт. Пусть известны расстояния c_{ij} между городами ($i, j = 1, \dots, n; i \neq j$). Требуется найти самый короткий маршрут.

Составим экономико – математическую модель. Введем переменные:

$$x_{ij} = \begin{cases} 1, & \text{если в маршрут входит переезд из города } i \text{ в город } j \\ 0 & \text{в противном случае } (i, j = 1, \dots, n; i \neq j) \end{cases}$$

Требование однократного въезда и выезда в города запишется в виде следующих ограничений:

$$\sum_{i=1}^n x_{ij} = 1, \quad j = 1, \dots, n, \quad (5.9)$$

$$\sum_{j=1}^n x_{ij} = 1, \quad i = 1, \dots, n \quad (5.10)$$

Однако ограничения (5.9) полностью не определяют допустимые маршруты, так как не исключают возможности разрыва пути, т. е. появления нескольких не связанных между собой подмаршрутов для части городов. Поэтому следует ввести дополнительно n переменных $u_i \geq 0; u_i \in \mathbf{Z} (i = 1, \dots, n)$, принимающих только целые неотрицательные значения, и записать для них специальные ограничения:

$$u_i - u_j + n \cdot x_{ij} \leq n - 1; \quad i, j = 2, \dots, n; \quad i \neq j \quad (5.11)$$

Общее число таких ограничений равно $(n - 1) \cdot (n - 2)$ и они, не исключая допустимый маршрут, исключают возможность существования подмаршрутов.

Таким образом, задача о коммивояжере состоит в минимизации целевой функции:

$$f(X, U) = \sum \sum c_{ij} x_{ij} \rightarrow \min$$

при условиях (5.9), (5.10), (5.11), где переменные x_{ij}, u_i принимают только неотрицательные целые значения.

К задачам целочисленного программирования приводят также многие оптимальные задачи *теории расписаний*, в которой рассматриваются методы оптимизации оперативно-календарного планирования. В качестве примера таких задач можно привести задачу определения оптимальной очередности обработки изделий на различных станках или других рабочих местах, задачу составления программы «диспетчер» для управления работой ЭВМ в мультипрограммном режиме и др.

5.6. Венгерский метод

Рассмотрим метод решения задачи о назначениях. Основная идея этого метода: оптимальность решения задачи не нарушается при уменьшении (увеличении) элементов строки (столбца) на одну и ту же величину $d_i; (d_j)$. Решение считают оптимальным, если все измененные искусственно затраты $c'_{ij} \geq 0 (i, j = 1, \dots, n)$ и можно отыскать такой набор x_{ij} , при котором достигается экстремум целевой функции (6).

Пример 5.6.

Пусть для монтажа четырех объектов ($n = 4$) требуется четыре крана ($n = 4$). Известно время монтажа каждым i -м краном каждого j -го объекта (табл. 5.1).

Необходимо так распределить краны по объектам, чтобы суммарное время монтажа всех объектов было минимально.

Решение

Соответственно исходным данным задача формализуется:

$$f = 3x_{11} + 7x_{12} + 5x_{13} + 8x_{14} + 2x_{21} + 4x_{22} + 4x_{23} + 5x_{24} + \\ + 4x_{31} + 7x_{32} + 2x_{33} + 8x_{34} + 9x_{41} + 7x_{42} + 3x_{43} + 8x_{44} \rightarrow \min$$

при ограничениях

$$x_{11} + x_{12} + x_{13} + x_{14} = 1,$$

.....

$$x_{41} + x_{42} + x_{43} + x_{44} = 1,$$

$$x_{11} + x_{21} + x_{31} + x_{41} = 1,$$

.....

$$x_{14} + x_{24} + x_{34} + x_{44} = 1$$

Таблица 5.1

| Код крана (i) | Затраты времени на монтаж по объектам (c _{ij}) | | | | a _i | d _i |
|----------------|---|---|---|---|----------------|----------------|
| | 1 | 2 | 3 | 4 | | |
| 1 | 3 | 7 | 5 | 8 | 1 | 3 |
| 2 | 2 | 4 | 4 | 5 | 1 | 2 |
| 3 | 4 | 7 | 2 | 8 | 1 | 2 |
| 4 | 9 | 7 | 3 | 8 | 1 | 3 |
| b _j | 1 | 1 | 1 | 1 | | |

Алгоритм метода включает следующие основные этапы (шаги).

Шаг 1. Получение нулей в каждой строке.

1.1. Находят наименьший элемент d_i в каждой строке (табл. 1), который вычитают из всех ее элементов и получают новую матрицу (табл. 5.2).

Таблица 5.2

| Код крана (i) | Затраты времени на монтаж по объектам (c _{ij}) | | | | a _i |
|----------------|---|---|---|---|----------------|
| | 1 | 2 | 3 | 4 | |
| 1 | 0 | 4 | 2 | 5 | 1 |
| 2 | 0 | 2 | 2 | 3 | 1 |
| 3 | 2 | 5 | 0 | 6 | 1 |
| 4 | 6 | 4 | 0 | 5 | 1 |
| b _j | 1 | 1 | 1 | 1 | |
| d _i | 0 | 2 | 0 | 3 | |

1.2. Аналогично в каждом столбце определяют его минимальный элемент d_j , который вычитают из всех его элементов с получением следующей матрицы (табл. 5.3)

Таблица 5.3

| Код крана (i) | Затраты времени на монтаж по объектам (c _{ij}) | | | | a _i |
|---------------|---|---|----|---|----------------|
| | 1 | 2 | 3* | 4 | |

| | | | | | |
|-------|----|----|----|---|---|
| 1 | 0* | 2 | 2 | 2 | 1 |
| 2 | 0 | 0* | 2 | 0 | 1 |
| 3* | 2 | 3 | 0* | 3 | 1 |
| 4* | 6 | 2 | 0 | 2 | 1 |
| b_j | 1 | 1 | 1 | 1 | |
| d_i | 0 | 2 | 0 | 3 | |

Шаг 2. Поиск оптимального решения.

2.1. Рассматривается одна из строк табл. 3, имеющая меньшее число нулей (строка 1); отмечается звездочкой (*) один из нулей этой строки ($c_{11} = 0$) и зачеркиваются все остальные нули этой строки и того столбца, в котором находится этот нуль (c_{21}).

2.2. Аналогичные операции выполняют последовательно для всех строк.

2.3. Если назначения, которые получены при всех нулях, отмеченных звездочками, являются полными, то есть число нулей, отмеченных звездочками, равно n , то решение является оптимальным. В противном случае переходят к шагу 3.

Шаг 3. Поиск минимального набора строк и столбцов, содержащих нули.

3.1. Отмечают звездочкой:

3.1.1) все строки, в которых нет ни одного отмеченного звездочкой нуля (строка 4, табл. 3);

3.1.2) все столбцы, содержащие перечеркнутый нуль хотя бы в одной из отмеченных звездочкой строк (столбец 3, табл. 3);

3.1.3) все строки, содержащие отмеченные звездочкой нули хотя бы в одном из отмеченных звездочкой столбцов (строка 3, табл. 3);

3.2. Шаги 3.1.2) и 3.1.3) повторяют поочередно до тех пор, пока есть что отмечать.

3.3. После этого зачеркивают каждую непомяченную строку и каждый помеченный столбец (строки 1, 2 и столбец 3, табл. 3) с целью провести минимальное число горизонтальных и вертикальных прямых, пересекающих по крайней мере один раз все нули (табл. 3)

Шаг 4. Перестановка некоторых нулей.

4.1. Определяют наименьшее число из тех клеток, через которые не проведены прямые (не зачеркнуты), то есть число 2 в табл. 3.

Это число вычитают из каждого числа невычеркнутых столбцов и прибавляют к каждому числу вычеркнутых строк в вычеркнутых столбцах с получением табл. 5.4.

Если эти операции не приводят к оптимальному решению, то цикл повторяется, начиная с шага 2 до получения оптимума.

В данном случае число нулей, отмеченных точкой, оказалось равным 4, значит назначение является полным, а решение оптимальным, то есть

Таблица 5.4

| Код крана (i) | Затраты времени на монтаж по объектам (c_{ij}) | | | | a_i |
|---------------|---|----|----|----|-------|
| | 1 | 2 | 3* | 4 | |
| 1 | 0* | 2 | 4 | 2 | 1 |
| 2 | 0 | 0* | 4 | 0 | 1 |
| 3* | 0 | 1 | 0* | 1 | 1 |
| 4* | 4 | 0 | 0 | 0* | 1 |

| | | | | | |
|-------|---|---|---|---|--|
| b_j | 1 | 1 | 1 | 1 | |
| d_i | 0 | 2 | 0 | 3 | |

$$x_{11}^* = x_{22}^* = x_{33}^* = x_{44}^* = 1,$$

$$f_{\min} = c_{11} + c_{22} + c_{33} + c_{44} = 3 + 4 + 2 + 8 = 17.$$

Вопросы для самопроверки

1. Сформулируйте задачу целочисленного линейного программирования
2. В чем суть графического метода решения задачи ЦЛП?
3. В чем суть метода Гомори решения задачи ЦЛП?
4. Решение частично-целочисленных задач.
5. В чем суть метода ветвей и границ решения задачи ЦЛП?
6. Рассмотреть пример. Решение задачи ЛП-1.
7. Решение задачи ЛП-2 и ЛП-3.
8. Решение задачи ЛП-4 и ЛП-5.
9. Сформулировать алгоритм метода ветвей и границ.
10. Сформулируйте задачу о назначениях
11. Сформулируйте задачу о коммивояжере
12. Раскройте суть венгерского метода решения задачи о назначениях

Тема 6. Задачи многокритериальной оптимизации

В практической деятельности часто встречаются задачи, заключающиеся в поиске лучшего (оптимального) решения при наличии различных несводимых друг к другу критериев оптимальности. Например, принятие решения о строительстве дороги в объезд города должно учитывать такие факторы, как выигрыш города в целом по соображениям экологии, проигрыш отдельных предприятий и фирм, например, из-за уменьшения проезжающих через город потенциальных покупателей и многие другие. Если такого рода задачи решаются методами математического программирования, то говорят о *задачах многокритериальной оптимизации*. Эти задачи могут носить как линейный, так и нелинейный характер. Далее будем рассматривать только линейные задачи.

Задачи многокритериальной оптимизации возникают в тех случаях, когда имеется несколько целей, которые не могут быть отражены одним критерием (например, стоимость и надежность). Требуется найти допустимое решение, которое минимизирует или максимизирует все такие критерии. Если в подобного рода задачах речь идет не о разнородных критериях некоторой системы, а о сопоставлении однородных критериев разных ее подсистем (например, отрасли, группы населения и т.п.), то эти *задачи называются задачами векторной оптимизации*.

Обозначим i -й частный критерий через $Z_i(X)$, где X — допустимое решение, а область допустимых решений — через S . Задачу многокритериальной оптимизации можно сформулировать следующим образом:

$$Z(X) = \{Z_1(X), Z_2(X), \dots, Z_m(X)\} \rightarrow \max, \quad (6.1)$$

$$X \in S$$

Некоторые частные критерии могут противоречить друг другу, другие действуют в одном направлении, третьи — индифферентны, безразличны друг к другу. Поэтому процесс решения многокритериальных задач неизбежно связан с экспертными оценками как самих критериев, так и взаимоотношений между ними. Известен ряд методов решения задач многокритериальной оптимизации:

- оптимизация одного признанного наиболее важным критерия, остальные критерии при этом играют роль дополнительных ограничений;
- упорядочение заданного множества критериев и последовательная оптимизация по каждому из них (этот подход рассмотрен ниже на примере *метода* последовательных уступок);
- сведение многих критериев к одному введением экспертных весовых коэффициентов для каждого из критериев таким образом, что более важный критерий получает более высокий вес;
- метод справедливого компромисса, который допускает одинаковую важность всех частных критериев и не требует их нормализации и упорядоченности по степени важности.

Для решения задач многокритериальной оптимизации используют критерий оптимальности Парето, суть которого состоит в улучшении одних показателей при условии, чтобы другие не ухудшались.

Определение. Вектор $X^* \in S$ называется эффективным (оптимальным по Парето) решением задачи (9.1), если для любого вектора $X \in S$ выполняется соотношение

$$Z_i(X) \leq Z_i(X^*), \quad i = 1, \dots, m \quad (6.2)$$

причем хотя бы для одного значения i имеет место строгое неравенство.

Множество допустимых решений, для которых невозможно одновременно улучшить все частные показатели эффективности (т.е. улучшить хотя бы один из них, не ухудшая остальных), принято называть **областью Парето**, или областью компромиссов, а принадлежащие ей решения — эффективными, или оптимальными по Парето.

6.1. Метод уступок

Рассмотрим один из методов решения многокритериальных задач – метод последовательных уступок.

Метод последовательных уступок решения задач многокритериальной оптимизации применяется в случае, когда частные критерии могут быть упорядочены в порядке убывания их важности. Предположим, что все частные критерии максимизируются и пронумерованы в порядке убывания их

важности. Находим максимальное значение Z_1^* первого по важности критерия в области допустимых решений путем решения однокритериальной задачи

$$\begin{aligned} Z_1(X) &\rightarrow \max, \\ X &\in S. \end{aligned}$$

Затем, исходя из практических соображений и принятой точности, назначается величина допустимого отклонения $\Delta_1 > 0$ (экономически оправданной уступки) критерия Z_1 , и далее находится максимальное значение второго критерия Z_2^* при условии, что значение первого критерия не должно отклоняться от своего максимального значения более, чем на величину допустимой уступки, т. е. решается задача:

$$\begin{aligned} Z_2(X) &\rightarrow \max, \\ Z_1(X) &\geq Z_1^* - \Delta_1, \\ X &\in S. \end{aligned}$$

Снова назначается величина уступки $\Delta_2 > 0$ по второму критерию, которая вместе с первой уступкой используется для нахождения условного максимума третьего частного критерия:

$$\begin{aligned} Z_3(X) &\rightarrow \max, \\ Z_1(X) &\geq Z_1^* - \Delta_1, \\ Z_2(X) &\geq Z_2^* - \Delta_2, \\ X &\in S. \end{aligned}$$

Аналогичные процедуры повторяются до тех пор, пока не будет выявлено максимальное значение последнего по важности критерия Z_m при условии, что значение каждого из первых $m-1$ частных критериев отличается от соответствующего условного максимума не более чем на величину допустимой уступки по данному критерию. Полученное на последнем этапе решение считается оптимальным.

Пример 6.1. Решение задачи многокритериальной оптимизации методом последовательных уступок.

Пусть задача трехкритериальной оптимизации имеет вид:

$$Z_1 = -x_1 + 2x_2 \rightarrow \max \quad (6.3)$$

$$Z_2 = 2x_1 + x_2 \rightarrow \max \quad (6.4)$$

$$Z_3 = x_3 - 3x_2 \rightarrow \max \quad (6.5)$$

$$x_1 + x_2 \leq 6$$

$$1 \leq x_1 \leq 3 \quad (6.6)$$

$$1 \leq x_2 \leq 4$$

Заметим, что так как коэффициенты при одних и тех же переменных в данных частных критериях имеют разные знаки, то в заданной области допустимых решений невозможно одновременно улучшить все частные критерии, т. е. в рассматриваемом случае область компромиссов (область Парето) совпадает с областью допустимых решений (6.6).

Для определенности будем считать, что допустимые уступки по первым двум критериям заданы: $\Delta_1 = 3$; $\Delta_2 = 5/3$.

Максимизируем функцию Z_1 в области допустимых решений, т. е. решаем однокритериальную задачу (6.3), (6.6). Это несложно сделать рассмотренным ранее графическим методом решения задач линейного программирования (см. рис 6.1).

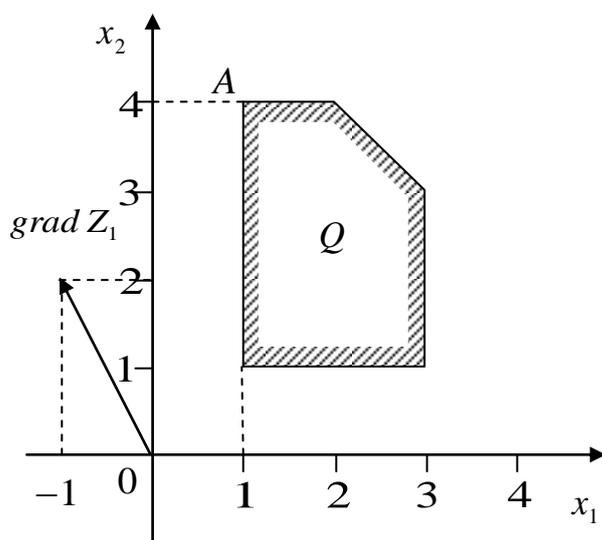


Рис. 6.1

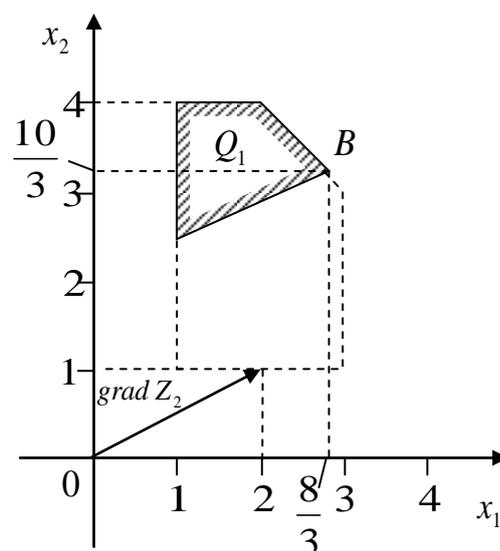


Рис. 6.2

Максимум функции Z_1 при условиях (9.6) достигается в точке A области S с координатами (1;4), так что в данном случае

$$x_1^* = 1; x_2^* = 4; Z_1^* = Z_1(A) = 7.$$

Переходим к максимизации функции Z_2 при условиях (6.6) и дополнительном ограничении, позволяющем учесть, что по критерию Z_1 нельзя уступить более чем на Δ_1 . Так как в нашем примере $Z_1^* - \Delta_1 = 4$, то дополнительное ограничение будет иметь вид:

$$-x_1 + 2x_2 \geq 4 \quad (6.7)$$

Задачу (6.4), (6.6), (6.7) также решаем графически (рис. 6.2).

Получаем, что максимум функции Z_2 при условиях (6.6), (6.7) достигается в точке В части S_1 области S , так что

$$x_1^{**} = 8/3; x_2^{**} = 10/3; Z_2^* = Z_2(B) = 26/3.$$

Теперь уступаем по критерию Z_2 на величину уступки $\Delta_2 = 5/3$ ($Z_2^* - \Delta_2 = 26/3 - 5/3 = 7$) и получаем второе дополнительное ограничение:

$$2x_1 + x_2 \geq 7 \quad (6.8)$$

Максимизируем функцию Z_3 при условиях (6.6), (6.7) и (6.8). Решение этой задачи графическим методом представлено на рис. 6.3.

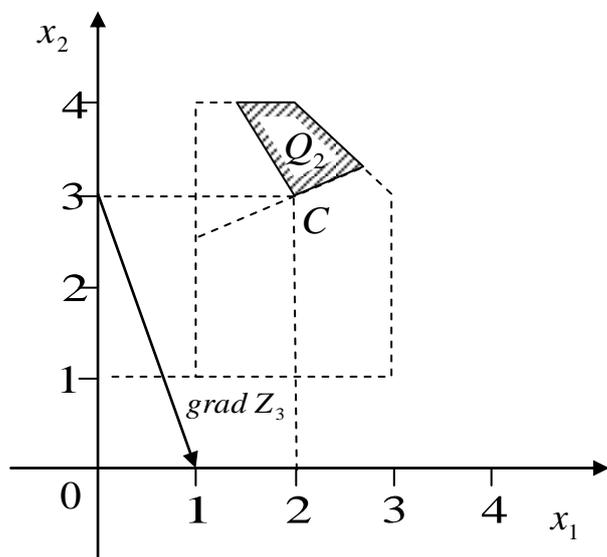


Рис. 6.3

Таким образом, получаем оптимальное решение рассматриваемой трехкритериальной задачи (точка С на рис. 6.3):

$$x_1^{***} = 2; x_2^{***} = 3$$

Соответствующие значения частных критериев при этом составляют:

$$Z_1 = 4; Z_2 = 7; Z_3 = -7.$$

6.2. Метод справедливого компромисса

Справедливым компромиссом будем называть такой компромисс, при котором относительный уровень уменьшения величины одного или нескольких частных критериев не превосходит относительного уровня увеличения величины остальных частных критериев (меньше или равен) [1].

Положим, что ставится задача минимизации каждого из частных критериев оптимальности $\phi_1(X), \phi_2(X), \dots, \phi_s(X)$ в одной и той же области допустимых значений $D_X \in R^n$. Здесь $\phi_i(X) = -Z_i(X)$, $i = 1, \dots, m$.

Задачу многокритериальной оптимизации будем записывать в виде

$$\min_{X \in D_X} \Phi(X) = \Phi(X^*), \quad (6.9)$$

где D_X — множество допустимых значений вектора варьируемых параметров X .

Для формализации понятия справедливого компромисса нам понадобится ввести отношение превосходства \blacktriangleright на множестве Парето (напомним, что множество Парето можно определить как множество, в котором значение любого из частных критериев оптимальности можно улучшить только за счет ухудшения других частных критериев — любое из решений, принадлежащее множеству Парето, не может быть улучшено одновременно по всем частным критериям) [1].

Пусть во множестве Парето задачи (6.9) даны две точки $X^1 \in D_X^*$, $X^2 \in \overline{D_X^*}$ и значения всех частных критериев оптимальности в них $\phi_k(X^1)$, $\phi_k(X^2)$, $k \in \overline{1; m}$. Введем меру относительного изменения (снижения — знак "минус" или повышения — знак "плюс") качества решения по каждому из этих критериев [1]

$$Z_k(X^1, X^2) = \frac{\Delta Z_k(X^1, X^2)}{\left| \max_{X \in \{X^1, X^2\}} Z_k(X) \right|}, \quad k = \overline{1, m}, \quad (6.10)$$

где $\Delta \phi_k(X^1, X^2) = \phi_k(X^1) - \overline{\phi_k(X^2)}$ — абсолютные изменения значений частных критериев оптимальности $Z_k(X)$, $k = \overline{1, m}$ при переходе от решения X^1 к решению X^2 . Вычислим максимальное снижение качества решения при переходе от решения X^1 к решению X^2

$$Z_{\min}(X^1, X^2) = \min_{k=1, m} Z_k(X^1, X^2). \quad (6.11)$$

Аналогично вычислим максимальное повышение качества решения при переходе решения X^1 к решению X^2

$$\tilde{\phi}_{\max}(X^1, X^2) = \max_{k=1, m} \phi_k(X^1, X^2). \quad (6.12)$$

Будем говорить, что решение X^2 превосходит решение X^1 , и писать $X^2 \blacktriangleright X^1$, если

$$\left| \tilde{\phi}_{\max}(X^1, X^2) \right| > \left| \tilde{\phi}_{\min}(X^1, X^2) \right|. \quad (6.13)$$

С другой стороны, будем говорить, что решение X^1 превосходит решение X^2 , если

$$\left| \tilde{\phi}_{\max}(X^1, X^2) \right| \leq \left| \tilde{\phi}_{\min}(X^1, X^2) \right|. \quad (6.14)$$

Выбор решений $X^r \in D_X^*$ будем производить с помощью полного перебора узлов сетки, покрывающей эффективное множество задачи (6.9). Поскольку метод справедливого компромисса использует относительные изменения частных критериев оптимальности, этот метод инвариантен к масштабу измерения частных критериев, т.е. не требуется их нормализация [1].

Вопросы для самопроверки

1. Что такое многокритериальная оптимизация
2. Какое решение называют оптимальным по Парето?
3. Какие методы используют для решения многокритериальных задач?
4. В чем состоит суть метода последовательных уступок?
5. В чем состоит суть метода справедливого компромисса?

ТЕМА 7. МЕТОДЫ ОПТИМИЗАЦИИ ФУНКЦИЙ

7.1. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Математическое программирование в самом общем виде можно определить как задачу оптимизации с ограничениями в пространстве R^n :

$$\begin{aligned} \min f(x), \\ g_k(x) \geq 0, \quad k = 1, \dots, K \\ l_j(x) = 0, \quad j = 1, \dots, J \\ x \in D \subset R^n \end{aligned} \quad (7.1)$$

Вектор $x \in D$ имеет компоненты x_1, x_2, \dots, x_n , которые являются неизвестными задачи (7.1).

Функции $f(x)$ называется **целевой функцией** (ЦФ) (функцией качества, критерием оптимальности), а множество условий $g_k(x), l_j(x), x \in D$ – **ограничениями задачи**.

В качестве D часто выступает выпуклое множество.

Выпуклым множеством $D \subset R^n$ называют такое множество, которое вместе с любыми двумя (соседними) точками x^1 и x^2 и произвольного числа $\mu \in [0, 1]$ содержит и точку $x \in D$

$$x = \mu \cdot x^1 + (1 - \mu) \cdot x^2 \text{ - отрезок при } \mu \in [0, 1],$$

т.е. выпуклое множество должно содержать отрезок, соединяющий любые две точки такого множества.

Примеры: куб, квадрат, круг, шар и др.

Решением задачи (1.1) называют любой вектор x , удовлетворяющий ограничениям.

Локальным экстремумом ЦФ $f(x)$ называют вектор \bar{x} , если имеет место неравенство

$$f(\bar{x}) \leq f(x), \quad \{x : \|x - \bar{x}\| < \varepsilon, \varepsilon > 0\}$$

Оптимальным решением или **глобальным экстремумом** задачи (7.1) называют вектор x^* , минимизирующий значение $f(x)$ на множестве всех решений: $f(x^*) \leq f(x)$ для всех $x \in D$.

Задача максимизации функции сводится к задаче поиска минимума функции $F(x) = -f(x)$.

Замечание: всякая точка глобального минимума ЦФ $f(x)$ является одновременно и точкой локального минимума этой функции. Обратное утверждение, вообще говоря, не верно.

Точность. Характеристикой точности полученного решения может служить величина абсолютного отклонения значения минимизируемой функции, достигнутого в точке $x^n \in D$, от точного значения ее минимума на множестве D :

$$\Delta(x) = \left| f(x^n) - f(x^*) \right|, \quad x^n, x^* \in D. \quad (7.2)$$

Ясно, что чем меньше неотрицательная величина δ , тем точнее полученное решение. Недостатком использования абсолютной погрешности является то обстоятельство, что она меняется при умножении ЦФ на положительную константу α : $f(x) \rightarrow \alpha \cdot f(x)$.

Целесообразнее использовать следующую оценку точности:

$$\left| \frac{f(x^n) - f(x^*)}{f(x^*)} \right| = \delta, \quad (7.3)$$

где $f(x^*)$ – либо точное значение минимума ЦФ, либо полученное "точным" алгоритмом. Алгоритм называется " ε -приближенным", если выполняется неравенство $c \leq \varepsilon$.

Унимодальные функции. Многие методы оптимизации применимы только тогда, когда целевая функция $f(x)$ является унимодальной, т. е. любой локальный минимум ЦФ одновременно является и глобальным.

Множество функций, унимодальных на отрезке $[a, b]$, обозначим $Q[a, b]$.

Определение. Функция $f(x)$ называется унимодальной на отрезке $[a, b]$, если она непрерывна на $[a, b]$ и существуют числа x^1 и x^2 , удовлетворяющие условию $a \leq x^1 \leq x^2 \leq b$, такие, что:

- 1) на отрезке $[a, x^1]$ функция $f(x)$ монотонно убывает;
- 2) отрезке $[x^2, b]$ функция $f(x)$ монотонно возрастает;
- 3) при $x \in [x^1, x^2]$ имеем $f(x^*) = f^* = \min_{[a, b]} f(x)$.

Отметим, что возможно вырождение в точку одного или двух из отрезков $[a, x^1]$, $[x^1, x^2]$ и $[x^2, b]$. Некоторые варианты расположения и вырождения в точку отрезков монотонности и постоянства унимодальных функций представлены на нижеприведенных рисунках.

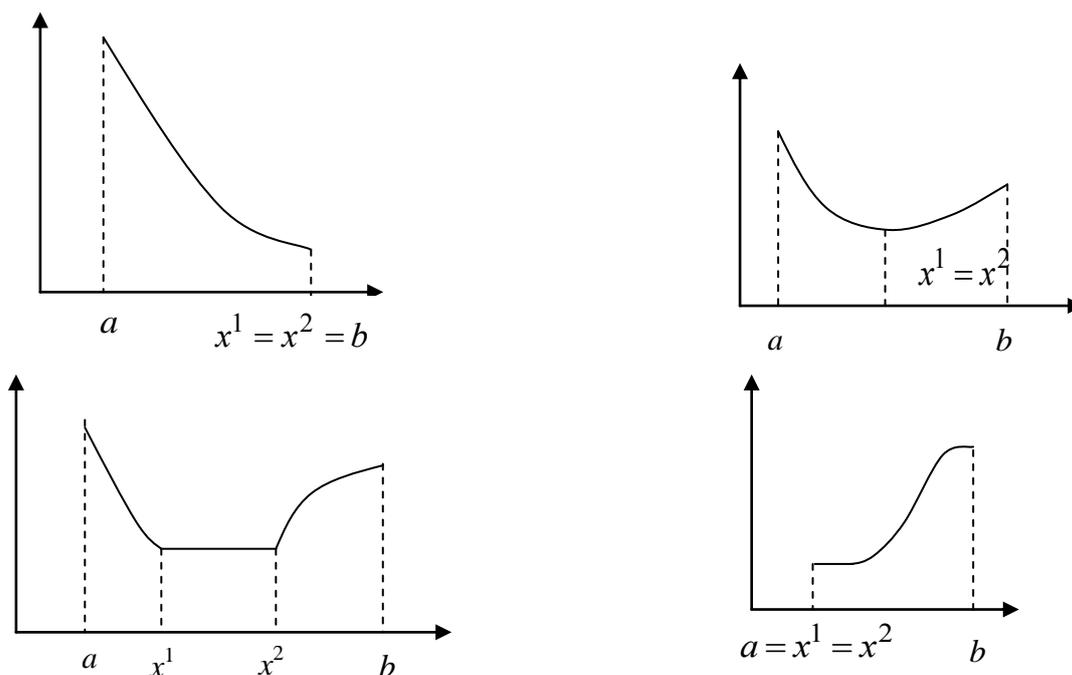


Рис. 7.1. Примеры унимодальных функций

Очевидно, что в некоторых вырожденных случаях ЦФ $f(x)$ не является дифференцируемой на отрезке $[a, b]$.

Для проверки унимодальности функции $f(x)$ на практике обычно используют следующие критерии:

– если функция $f(x)$ дифференцируема на отрезке $[a, b]$ и производная $f'(x)$ не убывает на этом отрезке, то $f(x) \in Q[a, b]$;

– если функция $f(x)$ дважды дифференцируема на отрезке $[a, b]$ и вторая производная $f''(x) \geq 0$ при $x \in [a, b]$, то $f(x) \in Q[a, b]$.

Если эти критерии не выполняются, то функция $f(x)$ является мультимодальной или многоэкстремальной.

Выпуклые функции. Функция $f(x)$ выпукла ("вниз") на $[a, b]$, если справедливо неравенство Йенссена:

$$f[\mu x^1 + (1 - \mu)x^2] \leq \mu f(x^1) + (1 - \mu)f(x^2),$$

для произвольных $x^1, x^2 \in [a, b]$ и $\mu \in [0, 1]$.

Геометрически это означает, что все точки функции $f(x)$ для всех $x \in [a, b]$ лежат ниже прямой $\mu f(x^1) + (1 - \mu)f(x^2)$.

На практике обычно используют следующие критерии:

одномерный случай:

1) дважды дифференцируемая функция $f(x)$ выпукла на $[a, b]$, если вторая производная $f''(x) \geq 0$ при всех $x \in [a, b]$;

многомерный случай (критерий Сильвестра):

2) если функция $f(x)$ дважды дифференцируема на выпуклом множестве $D \subset R^n$ и все угловые миноры матрицы Гессе – матрицы вторых производных $H_f(x) = (\partial^2 f / \partial x_i \partial x_j)$ – положительны при $x \in D$, то функция $f(x)$ выпукла на множестве D .

Квадратичные функции. Во многих задачах оптимизации рассматриваются функции вида:

$$f(x) = \sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j + \sum_{j=1}^n b_j x_j + c$$

или в матричном виде:

$$f(x) = \frac{1}{2} x^T A x + b^T x + c, \quad (7.4)$$

где x и b есть векторы – столбцы размерности n ;

A – симметричная матрица ($n \times n$);

c – константа.

Градиент и матрица Гессе функции (4) равны соответственно

$$\text{grad } f(x) = Ax + b, \quad H_f(x) = A = (a_{ij}).$$

Таким образом, для того чтобы функция (1.4) была выпуклой в R^n , достаточно, чтобы матрица A была положительно определена.

Критерии определенности матрицы (теорема Сильвестра).

Положительная определенность:

– все диагональные элементы матрицы должны быть положительны;

- все ведущие главные определители должны быть положительны.

Положительная полуопределенность:

- все диагональные элементы неотрицательны;
- все главные определители неотрицательны.

Главный определитель – это определитель главного минора.

Для положительной определенности квадратичной формы – матрицы A – необходимо и достаточно, чтобы все угловые миноры матрицы A были положительны, т. е.

$$\Delta_1 = a_{11} > 0, \quad \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \dots, \Delta_n = |A| > 0.$$

Стационарной точкой функции $f(x)$, где $x \in R^n$, называется такая точка x^* , в которой градиент функции обращается в нуль:

$$\text{grad } f(x)|_{x=x^*} = 0.$$

Если стационарная точка не соответствует локальному экстремуму, то она является **седловой точкой**.

7.2. Классификация задач оптимизации

Значительный интерес представляет деление задач оптимизации по виду целевой функции и ограничений, поскольку специфические особенности тех или иных задач играют важную роль при разработке методов их решений.

Задача безусловной оптимизации с одной переменной.

$$\min f(x), \quad x \in R^1. \quad (7.5)$$

Задача безусловной оптимизации функции многих переменных.

$$\min f(x), \quad x \in R^n. \quad (7.6)$$

Задача оптимизации с ограничениями (условная оптимизация).

$$\begin{aligned} \min f(x), \\ g_k(x) \geq 0, \quad k = 1, \dots, K \\ l_j(x) = 0, \quad j = 1, \dots, J \\ x_i^L < x_i < x_i^M, \quad i = \overline{1, n}; \end{aligned} \quad (7.7)$$

т.е. компоненты вектора x ограничены сверху и снизу.

Задача условной оптимизации с линейными ограничениями. Функции-ограничения $g_k(x)$, $l_j(x)$ в задаче (7.7) являются линейными, а целевая функция может быть либо линейной, либо нелинейной.

Задача линейного программирования (ЛП).

$$f(x) = \sum_{j=1}^n c_j x_j \rightarrow \min, \quad (7.8)$$

$$\sum_{j=1}^n d_{ij} x_j \leq z_i, \quad i = \overline{1, m}; \quad x_j \geq 0,$$

т.е. и целевая функция, и ограничения являются линейными функциями переменных x_j .

Задача целочисленного программирования. В задаче ЛП (7.8) компоненты вектора x принимают только целые значения.

Задача нелинейного программирования с линейными ограничениями. Целевая функция $f(x)$ – нелинейная, а все ограничения – линейные функции.

Оптимизационные задачи такого рода можно классифицировать на основе структурных особенностей нелинейных целевых функций.

Задача квадратичного программирования. Найти минимум $f(x)$, где целевая функция $f(x)$ является квадратичной.

$$f(x) = \frac{1}{2} x^T A x + b^T x + c, \quad (7.9)$$

$$\sum_{j=1}^n d_{ij} x_j \leq z_i, \quad i = \overline{1, m}; \quad x_j \leq 0, \quad j = \overline{1, n}.$$

Задача дробно-линейного программирования. Целевая функция есть отношение линейных функций:

$$f(x) = \frac{\sum_{j=1}^n c_j x_j + c_0}{\sum_{j=1}^n d_j x_j + d_0} \rightarrow \min. \quad (7.10)$$

7.3. Необходимые и достаточные условия существования экстремума

7.3.1 СКАЛЯРНЫЙ СЛУЧАЙ $x \in \mathbf{R}^1$

Если в точке x^* первые $(n-1)$ производные функции $f(x)$ обращаются в нуль, а производная порядка n отлична от нуля, т.е.

$$f^{(n)} = d^{(n)} f(x) / dx^n \Big|_{x=x^*} \neq 0,$$

то:

- 1) если n – нечетное, то x^* – точка перегиба;
- 2) если n – четное, то x^* – локальный экстремум; причем, если
 - $f^{(n)}(x) > 0$, то x^* – точка локального минимума;
 - $f^{(n)}(x) < 0$, то x^* – точка локального максимума.

Примечание. При поиске глобального минимума функции $f(x)$, заданной на отрезке $[a, b]$, необходимо найти все точки локальных минимумов, вычислить в них значения функции, а также вычислить значения функции на границах отрезка $f(a)$ и $f(b)$ и выбрать среди этих значений наименьшее.

7.3.2 ВЕКТОРНЫЙ СЛУЧАЙ

Функция $f(x)$ имеет в точке x^* локальный минимум, если выполняется равенство $grad f(x^*) = 0$, т.е. x^* – стационарная точка, и матрица Гессе

$H_f(x^*)$ положительно определена.

Для выпуклых функций найденный локальный минимум будет являться одновременно и глобальным.

7.3.3 МИНИМИЗАЦИЯ ПРИ ОГРАНИЧЕНИЯХ

Рассматривается задача:

$$\begin{aligned} f(x) &\rightarrow \min, \\ g_k(x) &\leq 0, \quad k = \overline{1, m}, \\ x &\in D \equiv R^n. \end{aligned} \quad (7.11)$$

Обозначим $L(x, \lambda)$ функцию Лагранжа

$$L(x, \lambda) = f(x) + \sum_{k=1}^m \lambda_k g_k(x), \quad (7.12)$$

где λ_k – неопределенные множители Лагранжа.

Здесь все функции f , g_k , L предполагаются непрерывно дифференцируемыми.

Точка x^* будет *глобальным экстремумом* задачи (7.12), если в точке x^* выполняются условия Куна-Таккера: существует такое $\lambda^* \geq 0$, что

$$\begin{aligned} \text{grad}L(x, \lambda^*) &= 0; \\ \lambda_k^* g_k(x) &= 0; \quad k = \overline{1, m}. \end{aligned}$$

Пара векторов (x^*, λ^*) образует седловую точку функции Лагранжа, если при всех $x \in D$ и $\lambda \geq 0$ выполняется неравенство:

$$L(x^*, \lambda) \leq L(x^*, \lambda^*) \leq L(x, \lambda^*).$$

Таким образом, точка x^* является точкой глобального минимума задачи (7.12), если пара векторов (x^*, λ^*) является седловой точкой функции $L(x, \lambda)$.

7.4. Критерии останова

Приведем несколько наиболее распространенных критериев останова оптимизационных методов (ε – заданная точность):

$$1) |\Delta x| = |x^{k+1} - x^k| \leq \varepsilon; \quad 2) |f(x^{k+1}) - f(x^k)| \leq \varepsilon; \quad 3) \|\Delta x\| \leq \varepsilon;$$

$$4) \max_{1 \leq i \leq n} \left| \frac{\partial f}{\partial x_i} \right| \leq \varepsilon; \quad 5) \|\text{grad}f(x)\|^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \leq \varepsilon;$$

$$6) \|\text{grad}f(x)\| = \left[\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \right]^{1/2} \leq \varepsilon; \quad 7) \|x^k - x^{k/2}\| \leq \varepsilon,$$

где k – четное число (данный критерий используется в методах штрафных и барьерных функций).

При выполнении неравенств полагают: $x^* \approx x^k$ и $f^* \approx f(x^k)$.

7.5. Характеристики алгоритмов оптимизации

Прежде чем приступить к решению конкретной задачи оптимизации (ЗО), следует выяснить вопрос, какому из методов оптимизации отдать предпочтение в данном случае. Для этого *необходимо уметь сравнивать методы*, например, *на определенных классах задач*.

Очень важна роль *априорных характеристик* методов: трудоемкость вычислений, скорость сходимости, устойчивость метода к ошибкам в вычислениях, чувствительность метода к значениям параметров алгоритма и ряд других.

Сходимость алгоритма является наиболее важным вопросом оптимизации. Большинство методов решения ЗО имеют итерационную природу: $\{x_k\}$, т. е. исходя из некоторой начальной точки x_0 , они порождают последовательность $\{x_0, x_1, \dots, x_k, \dots\}$ на основе алгоритма A , сходящуюся к точке экстремума x^* .

Глобальная сходимость. Алгоритм A обладает свойством глобальной сходимости, если для любой начальной точки x_0 , последовательность $\{x_k\}$, определяемая выражением $x_{k+1} = A(x_k)$, сходится к точке, удовлетворяющей необходимым условиям оптимальности. Данное свойство отражает надежность работы алгоритма.

Асимптотическая сходимость и скорость сходимости. С практической точки зрения эффективность алгоритма зависит от числа итераций, необходимых для получения приближенного решения x^* с заданной точностью ε .

Для получения критерия с некоторым абсолютным значением необходимо прибегнуть к другому типу анализа, взяв за объект исследования *асимптотическую сходимость* – поведение последовательности точек $\{x_k\}$ в окрестности предельной точки x^* . Это значит, что каждому алгоритму приписывается *индекс эффективности – скорость сходимости*.

Предположим, что последовательность $\{x_k\}$ сходится к точке x^* . Обозначим $\|x\|$ евклидову норму в пространстве R^n .

Линейная сходимость. Выполняется неравенство:

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \alpha < 1, \quad (7.13)$$

где α - коэффициент сходимости.

Суперлинейная сходимость:

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \rightarrow 0. \quad (7.14)$$

Сходимость порядка p . Если существует такое число $p > 1$ (2,3...), что

$$\limsup_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^p} < c < \infty, \quad (7.15)$$

где c – константа, значит, имеем сходимость порядка p ;

$p = 2$ – квадратичная сходимость;

$p = 3$ – кубическая сходимость;

$p = n$ – сходимость порядка n .

Исследование скорости сходимости алгоритма позволяет оценить его эффективность и осуществить его сравнение с другими алгоритмами.

Замечание. Иногда к выражению скорости сходимости последовательности $\{x_k\}$ приходят и при исследовании способа сходимости последовательности $f(x_k)$ к $f(x^*)$.

Для оценки эффективности выбранных методов можно рекомендовать три характеристики: **время**, затраченное на получение решения; **точность** решения; **чувствительность** к изменению параметра сходимости. Для одномерных методов оптимизации первые две характеристики можно исследовать на типовых тестовых функциях, например,

$$f(x) = \sin^k(x), \quad 0 \leq x \leq 2\pi$$

путем варьирования значений показателя степени k на множестве нечетных чисел от 1 до 79. Отметим, что для всех k $\min f(x)$ достигается в точке $x^* = 3\pi/2 = 4,71239$, при этом $f(x^*) = -1,0$. Однако с увеличением k степень гладкости функции, которая обладает узкими впадинами в окрестности x^* , уменьшается. Проверку методов на чувствительность можно осуществить путем варьирования значений параметра сходимости k и параметров алгоритма.

Точность решения можно измерить как относительную (в процентах) ошибку оценивания координаты истинного минимума.

Вопросы для самопроверки

1. Запишите задачу оптимизации (ЗО) общего вида, укажите целевую функцию (ЦФ); ограничения.
2. Что понимают под оптимальным решением ЗО? Как определить точность.
3. Локальный и глобальный экстремум функции.
4. Классификация ЗО по виду ЦФ и ограничений.
5. Унимодальные функции (УФ). Критерии для проверки унимодальности.
6. Выпуклые множества. Критерий проверки выпуклости множества.
7. Выпуклые функции. Критерии проверки выпуклости функции.
8. Квадратичные функции (КФ). Критерии определенности КФ (теорема Сильвестра). Градиент и матрица Гессе КФ.
9. Необходимые и достаточные условия (Н и ДУ) существования экстремума - скалярный случай. Что такое "точка перегиба" и как ее идентифицировать?
10. Необходимые и достаточные условия существования экстремума многомерной функции.
11. Необходимые и достаточные условия существования экстремума функции при ограничениях.
12. Критерии завершения итерационных процедур поиска экстремума функций.
13. Понятие сходимости алгоритма.
14. Характеристики оценки эффективности методом поиска минимума функций.

Тема 8. Методы поиска экстремумов функции одной переменной

Задача оптимизации функций одной переменной относится к наиболее простому классу оптимизационных задач. Тем не менее, анализ задач такого типа занимает центральное место в оптимизационных исследованиях. Такие задачи обычно решаются в инженерной практике. Вместе с тем, одномерные методы оптимизации часто используются для анализа подзадач, которые возникают при решении многомерных задач оптимизации. Существует большое число методов решения задач с одной переменной.

8.1 Прямые методы оптимизации

Большую группу приближенных методов минимизации составляют прямые методы, основанные на вычислении только значений минимизируемой функции в некоторых точках и не использующие значений ее производной.

К данному классу относятся такие методы, как метод равномерного поиска, метод дихотомии, метод золотого сечения, метод Фибоначчи (псевдоним Леонарда Пизанского, первым изучившего эту последовательность в 1202 году).

8.1.1 Метод равномерного поиска (МРП)

Пусть $f(x) \in Q[a, b]$ и требуется найти какую-либо из точек минимума x^* на $[a, b]$ с абсолютной погрешностью $\varepsilon > 0$. Разобьем $[a, b]$ на n равных частей точками

$$x_i = a + i \cdot h, \quad i = 0, 1, \dots, n; \quad h = \frac{b-a}{n}, \quad \text{где } n = \frac{b-a}{\varepsilon}. \quad \text{Найдем точку } x_m, \text{ для которой}$$

$$f(x_m) = \min_{0 \leq i \leq n} f(x_i) \quad (8.1)$$

и полагаем $x^* \approx x_m, f^* \approx f(x_m)$.

8.1.2 Метод деления отрезка пополам (МДОП)

Пусть задана $f(x) \in Q[a, b]$. Надо найти x^* и f^* . Пусть $\varepsilon > 0$ – требуемая точность определения точки x^* . Выбрав $\delta \in (0, \varepsilon)$ (здесь δ – параметр метода) построим последовательность $\{a_n\}, \{b_n\}, \{x_1^{(n)}\}, \{x_2^{(n)}\}, n = 1, \dots$, используя рекуррентные формулы:

$$\begin{cases} a_0 = a; b_0 = b; \\ x_1^{(n-1)} = (a_{n-1} + b_{n-1} - 2\delta)/2; \\ x_2^{(n-1)} = (a_{n-1} + b_{n-1} + 2\delta)/2; \\ a_n = a_{n-1}, b_n = x_2^{(n-1)} \quad \text{если } f(x_1^{(n-1)}) \leq f(x_2^{(n-1)}); \\ a_n = x_1^{(n-1)}, b_n = b_{n-1} \quad \text{если } f(x_1^{(n-1)}) > f(x_2^{(n-1)}). \end{cases} \quad (8.2)$$

За точку x^* принимаем $x^* = (a_n + b_n)/2$, погрешность равна $\varepsilon_n = \frac{(b_n - a_n)}{2}$.

Условие завершения процесса $\varepsilon_n \leq \varepsilon$. Обычно берут $\delta = \varepsilon/2$.

8.1.3 Метод Фибоначчи (МФ)

Этот метод основан на использовании чисел Фибоначчи F_n , задаваемых рекуррентной формулой $F_n = F_{n-1} + F_{n-2}$ ($n \geq 2$), $F_0 = 1$, $F_1 = 1$.

Шаг 1.

Задаем точность ε .

По таблице проверяем условие

$$\Delta_n = \frac{\Delta_0}{F_n} \leq \varepsilon; \quad \Delta_0 = (b - a) \quad (8.3)$$

и определяем число n . В таблице 8.1 приведены числа Фибоначчи F_n .

Таблица 8.1

| | | | | | | | | | | |
|-------|---|---|---|---|---|----|----|----|----|----|
| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| F_n | 1 | 2 | 3 | 5 | 8 | 13 | 21 | 34 | 55 | 89 |

| | | | | | | | | | | |
|-------|-----|-----|-----|-----|-----|------|------|------|------|-------|
| n | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| F_n | 144 | 233 | 377 | 610 | 987 | 1597 | 2584 | 4181 | 6765 | 10946 |

Шаг 2.

Определяем точку d_0 из условия $\frac{\Delta_0}{\Delta_1} = \frac{F_n}{F_{n-1}}$, $n = 4, 5, 6, \dots$

$$d_0 = a_0 + \Delta_1. \quad (8.4)$$

Шаг 3.

Определяем точку c_0 так, чтобы она была симметрична d_0 относительно середины интервала Δ_0 , т.е. $d_0 - a_0 = b_0 - c_0$, откуда $c_0 = b_0 + a_0 - d_0$.

Шаг 4.

Вычисляем значение функции в точках (a_0, b_0, c_0, d_0) . Сравниваем значения функции, возможны ситуации:

$$\begin{aligned} 1a) \quad & f(a_0) \geq f(c_0) \leq f(d_0) \leq f(b_0), \\ 1b) \quad & f(a_0) \leq f(c_0) \leq f(d_0) \leq f(b_0); \\ 2a) \quad & f(a_0) \geq f(c_0) \geq f(d_0) \leq f(b_0), \\ 2b) \quad & f(a_0) \geq f(c_0) \geq f(d_0) \geq f(b_0) \end{aligned} \quad (8.5)$$

Шаг 5.

Определяем новый (приведенный) интервал. :

Ситуация 1) $a_1 = a_0$; $b_1 = d_0$; $d_1 = c_0$; $x_1 = d_1$.

Точка c_1 находится из условия $b_1 - c_1 = d_1 - a_1$, откуда

$$c_1 = b_1 + a_1 - d_1. \quad (8.6)$$

Ситуация 2) $a_1 = c_0$; $b_1 = b_0$; $c_1 = d_0$; $x_1 = c_1$

Точка d_1 находится из условия $b_1 - c_1 = d_1 - a_1$, откуда

$$d_1 = b_1 + a_1 - c_1 \quad (8.7)$$

Итерируем до тех пор, пока не достигнем значения $i = n$. Полагаем $x^* \approx x_n$

Скорость сходимости линейная с коэффициентом

$$\alpha = \frac{F_n}{F_{n-1}} \rightarrow \frac{1}{\gamma} = 0,618$$

8.1.4 Метод золотого сечения (МЗС)

Идея метода золотого сечения такая же, как и идея метода Фибоначчи. Отличие состоит в том, что здесь не надо заранее задавать n , чтобы производить поиск точек d_0 и c_0 .

Длины последовательных интервалов в МЗС берутся в фиксированном отношении:

$$\frac{\Delta_0}{\Delta_1} = \frac{\Delta_1}{\Delta_2} = \dots = \frac{\Delta_{n-1}}{\Delta_n} = \gamma, \quad \Delta_0 = b - a, \quad (8.8)$$

где $\gamma = \frac{\sqrt{5} + 1}{2} \approx 1,618$ – золотое сечение. Заметим, что в методе Фибоначчи

$$\frac{\Delta_{n-1}}{\Delta_n} = \frac{F_{n-1}}{F_n}; \quad \lim_{n \rightarrow \infty} \frac{F_{n-1}}{F_n} = \gamma.$$

Шаг 1.

Определяем c_0 и d_0 из условий: $\frac{\Delta_0}{\Delta_1} = \gamma$ и $\frac{\Delta_1}{\Delta_2} = \gamma$

$$c_0 = a_0 + \Delta_2, \quad d_0 = a_0 + \Delta_1. \quad (8.9)$$

Шаг 2.

Вычисляем значение функции в точках (a_0, b_0, c_0, d_0) . Сравниваем значения функции, возможны ситуации:

- 1) $f(a_0) \geq f(c_0) \leq f(d_0) \leq f(b_0)$,
- 2) $f(a_0) \geq f(c_0) \geq f(d_0) \leq f(b_0)$.

Шаг 3. Определяем новый (приведенный) интервал:

Ситуация 1) $a_1 = a_0$; $b_1 = d_0$; $d_1 = c_0$; $x_1 = d_1$.

Точка c_1 находится из условия $\frac{\Delta_2}{\Delta_3} = \gamma$, откуда

$$c_1 = a_0 + \Delta_3. \quad (8.10)$$

Ситуация 2) $a_1 = c_0$; $b_1 = b_0$; $c_1 = d_0$; $x_1 = c_1$

Точка d_1 находится из соотношения

$$d_1 = b_0 - \Delta_3 \quad (8.11)$$

Шаг 4.

Продолжаем процесс, пока не достигнем заданной точности

$$\Delta_n = \Delta_0 \left(\frac{1}{\gamma} \right)^n \leq \varepsilon. \quad (8.12)$$

На самом деле, погрешность будет равна $\frac{1}{2} \Delta_n = \frac{1}{2} \cdot \Delta_0 \left(\frac{1}{\gamma} \right)^n \leq \varepsilon$

8.2 Сравнение прямых методов оптимизации

Проведем сравнение относительных эффективностей прямых методов поиска экстремума функций одной переменной.

Обозначим длину исходного интервала неопределенности через $\Delta_0 = b - a$; Δ_n – длина полученного интервала в результате n -вычислений функции; $FR(n) = \frac{\Delta_n}{\Delta_0}$ – относительное уменьшение первоначального интервала; $FR(n)$ – показатель эффективности.

В табл. 8.2 приведены значения показателей эффективности методов прямого поиска.

Таблица 8.2 Сравнение прямых методов оптимизации функции одной переменной

| Метод прямого поиска | Показатель эффективности $FR(n)$ | Гарантированная оценка погрешности |
|--|---|--|
| Метод равномерного поиска | $\frac{2}{n+1}$ | $\frac{\Delta_0}{n+1}$ |
| Метод деления отрезка пополам (n – четное, величиной δ пренебрегаем) | $\frac{1}{2^{n/2}}$ | $\frac{1}{2} \cdot \frac{\Delta_0}{2^{n/2}}$ |
| Метод Фибоначчи | $\frac{2}{F_{n+1}}$ | $\frac{\Delta_0}{F_{n+1}}$ |
| Метод золотого сечения | $\left(\frac{1}{\gamma} \right)^{n-1}$ | $\frac{\Delta_0}{2} \cdot \left(\frac{1}{\gamma} \right)^{n-1}$ |

Как видно из таблицы, метод равномерного поиска является самым трудоемким, метод Фибоначчи и метод золотого сечения по эффективности практически совпадают.

С другой стороны, можно также сравнивать количество вычислений значений функций, требуемых для достижения заданной точности ε . Если величина интервала $\Delta_0 = (b - a)$ задана, то значение n вычисляется по следующим формулам:

$$\text{для МРП: } n = \frac{\Delta_0}{\varepsilon} - 1.$$

$$\text{для метода деления отрезка пополам: } n = \frac{2 \ln(\Delta_0 / 2\varepsilon)}{\ln 2};$$

$$\text{для МФ: } n = F_{n+1}^{-1}(\Delta_0 / \varepsilon) - 1;$$

$$\text{для МЗС: } n = 1 + \frac{\ln(\Delta_0 / 2\varepsilon)}{\ln \gamma};$$

В табл. 8.3 приведено количество вычислений функции f для достижения заданной точности ε .

Таблица 8.3 . Количество вычислений функции

| Метод прямого поиска | Число вычислений n при заданном ε , равном | | | | |
|--|--|-----------|-----------|-----------|-----------|
| | 10^{-1} | 10^{-2} | 10^{-3} | 10^{-4} | 10^{-5} |
| Метод равномерного поиска | 9 | 99 | 999 | 9999 | 99999 |
| Метод деления отрезка пополам (n – четное, величиной δ пренебрегаем) | 6 | 12 | 18 | 26 | 32 |
| Метод Фибоначчи | 5 | 10 | 15 | 19 | 24 |
| Метод золотого сечения | 5 | 10 | 15 | 20 | 24 |

Видно, что метод деления отрезка пополам и два последних метода (МФ и МЗС) практически совпадают по количеству вычислений функции.

8.3 Полиномиальная аппроксимация и методы точечного оценивания

Применение методов исключения интервалов требует унимодальности целевой функции и основано на простом сравнении значений функции в двух пробных точках. Рассмотрим более эффективные методы, чем методы исключения интервалов. Но выигрыш в эффективности достигается ценой введения дополнительного требования – исследуемые функции должны быть гладкими.

Основная идея рассматриваемых ниже методов связана с возможностью аппроксимации гладкой функции полиномом и последующего использования этого полинома для оценивания координаты точки экстремума.

Необходимые условия эффективной реализации такого подхода – унимодальность и непрерывность исследуемой функции. Согласно теореме Вейерштрасса об аппроксимации, если функция непрерывна на некотором интервале, то ее с любой степенью точности можно аппроксимировать полиномом высокого порядка. Координату точки экстремума функции можно оценить путем вычисления координаты точки экстремума полинома.

Точность оценки точки экстремума можно повысить двумя способами:

- использование полинома более высокого порядка;
- уменьшение интервала аппроксимации.

Второй способ более прост и экономичен, если функция унимодальная.

8.3.1 Квадратичная аппроксимация

Простейший вариант полиномиальной аппроксимации основан на факте, что функция $f(x)$ имеет экстремум во внутренней точке интервала и должна быть по крайней мере квадратичной. Линейная функция достигает экстремума в двух граничных точках интервала.

Если задана последовательность точек x_1, x_2, x_3 и известны соответствующие этим последовательностям значения функции f_1, f_2, f_3 , то можно определить константы a_0, a_1, a_2 таким образом, что значения квадратичной функции:

$$q(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2)$$

совпадут со значениями функции в трех указанных точках. Перейдем к вычислению $q(x)$ в точках x_1, x_2, x_3 .

Т.к. $f_1 = f(x_1) = q(x_1) = a_0$, имеем $a_0 = f_1$.

Далее, поскольку $f_2 = f(x_2) = q(x_2) = f_1 + a_1(x_2 - x_1)$, получаем

$$a_1 = \frac{f_2 - f_1}{x_2 - x_1}$$

Наконец, при $x = x_3$ имеем:

$$f_3 = f(x_3) = q(x_3) = f_1 + [(f_2 - f_1)/(x_2 - x_1)](x_3 - x_1) + a_2(x_3 - x_1)(x_3 - x_2).$$

Разрешая последнее уравнение относительно a_2 , получим:

$$a_2 = \frac{1}{x_3 - x_2} \left(\frac{f_3 - f_1}{x_3 - x_1} - \frac{f_2 - f_1}{x_2 - x_1} \right).$$

Таким образом, по трем заданным точкам и соответствующим значениям функции можно оценить параметры a_0, a_1, a_2 аппроксимирующего квадратичного полинома.

Если точность аппроксимации исследуемой функции высока на интервале от x_1 до x_3 , то построенный полином можно использовать для оценивания координаты точки экстремума. В данном случае из уравнения

$$\frac{dq}{dx} = a_1 + a_2(x - x_2) + a_2(x - x_1) = 0$$

можно получить стационарную точку:

$$\bar{x} = \frac{x_2 + x_1}{2} - \frac{a_1}{2a_2}. \quad (8.9)$$

Так как $f(x)$ и аппроксимирующий полином являются унимодальными на рассматриваемом интервале, то \bar{x} является приемлемой оценкой координаты точки истинного экстремума x^* .

Пример 8.1: Оценить координаты минимума функции

$$f(x) = 2x^2 + \frac{16}{x}; \quad x \in [1; 5].$$

Пусть $x_1 = 1$, $x_3 = 5$, $x_2 = \frac{1+5}{2} = 3$.

Вычислим соответствующие значения функции: $f_1 = 18$; $f_2 = \frac{70}{3} = 23,33$; $f_3 = 52,3$.

Далее находим параметры a_0, a_1, a_2 . Имеем:

$$a_0 = 18; \quad a_1 = \frac{23,33 - 18}{3 - 1} = \frac{8}{3}; \quad a_2 = \frac{1}{5 - 3} \left[\frac{52,3 - 18}{5 - 1} - \frac{8}{3} \right] = \frac{46}{15}.$$

Подставляя эти значения в выражение (2.1), получим:

$$\bar{x} = \frac{3+1}{2} - \frac{8}{3} \cdot \left(2 \cdot \frac{46}{15} \right)^{-1} = 1,565.$$

Точный минимум достигается при $x^* = 1,5874$.

8.3.1 Метод Пауэлла

Этот метод основан на последовательном применении процедуры оценивания с использованием квадратичной аппроксимации.

Схема алгоритма Пауэлла: пусть x_1 - начальная точка; Δx - выбранная величина шага по оси x .

Шаг 1: Вычислить $x_2 = x_1 + \Delta x$.

Шаг 2: Вычислить $f(x_1)$ и $f(x_2)$.

Шаг 3: Если $f(x_1) > f(x_2)$, положить $x_3 = x_1 + 2\Delta x$, если $f(x_1) \leq f(x_2)$, то $x_3 = x_1 - \Delta x$. Если $x_3 < x_1$, то перенумеровать точки в естественном порядке: $x_1 = x_3$, $x_2 = x_1$, $x_3 = x_2$.

Шаг 4: Вычислить $f(x_3)$ и найти

$$F_{\min} = \min\{f_1, f_2, f_3\}.$$

X_{\min} равно точке x_i , которая соответствует F_{\min} .

Шаг 5: По трем точкам x_1, x_2, x_3 вычислить \bar{x} , используя формулу (2.1), т.е. используя квадратичную аппроксимацию.

Шаг 6: Проверка на окончание поиска:

а) является ли разность $|F_{\min} - f(\bar{x})| \leq \varepsilon$;

б) является ли разность $|X_{\min} - \bar{x}| \leq \delta$,

где $\varepsilon > 0$ и $\delta > 0$ - заданные точности.

Если условия а) и б) выполняются одновременно, то закончить поиск. Иначе переход на Шаг 7.

Шаг 7: Выбрать "наилучшую" точку (X_{\min} или \bar{x}) и две точки по обе стороны от нее. Обозначить эти точки в естественном порядке и перейти на Шаг 2.

Замечание: после пятого шага необходимо провести дополнительную проверку, т.к. точка \bar{x} может находиться за точкой x_3 . В этом случае точка \bar{x} заменяется точкой, координата которой вычисляется с учетом заранее установленной длины шага.

Пример 8.2: Найти $\min f(x) = 2x^2 + 16/x$.

Пусть начальная точка $x_1 = 1$ и длина шага $\Delta x = 1$.

Критерии останова:

$$\left| \frac{X_{\min} - \bar{x}}{x} \right| \leq 3 \cdot 10^{-2}; \quad \left| \frac{F_{\min} - f(\bar{x})}{f(x)} \right| \leq 3 \cdot 10^{-3}.$$

Замечание. Оба критерия должны выполняться одновременно.

Итерация 1:

Шаг 1: $x_2 = x_1 + \Delta x = 2$;

Шаг 2: $f(x_1) = 18$; $f(x_2) = 16$;

Шаг 3: $f(x_1) > f(x_2)$, следовательно $x_3 = 1 + 2 = 3$;

Шаг 4: $f(x_3) = 23,33$; $F_{\min} = 16$; $X_{\min} = x_2$.

Шаг 5:

$$a_1 = \frac{16-18}{2-1} = -2; \quad a_2 = \frac{1}{3-2} \cdot \left[\frac{23,33-18}{3-1} - (-2) \right] = \frac{5,33}{2} + 2 = 4,665.$$

В соответствии с (2.1) вычислим \bar{x} :

$$\bar{x} = \frac{1+2}{2} - \frac{(-2)}{2 \cdot 4,665} = 1,5 + \frac{1}{4,665} = 1,714; \quad f(\bar{x}) = 15,210.$$

Шаг 6: Проверка на окончание поиска:

$$a) \quad \left| \frac{16-15,210}{15,210} \right| = 0,0519 > 0,003. \quad \text{Критерий останова не выполняется,}$$

следовательно поиск продолжается. Переход на Шаг 7.

Шаг 7: Выбираем \bar{x} как "наилучшую" точку, а $x_1 = 1$, $x_2 = 2$ – как точки, которые ее окружают. Обозначим эти точки в естественном порядке и переходим к Шагу 2.

Итерация 2:

$$\text{Шаг 4:} \quad x_1 = 1, \quad f_1 = 18; \quad x_2 = 1,714, \quad f_2 = 15,210 = F_{\min};$$

$$X_{\min} = x_2 = 1,714; \quad x_3 = 2; \quad f_3 = 16.$$

Шаг 5:

$$a_1 = \frac{15,210-18}{1,714-1} = -3,908;$$

$$a_2 = \frac{1}{2-1,714} \cdot \left[\frac{16-18}{2-1} - (-3,908) \right] = \frac{1,908}{0,286} = 6,671;$$

$$\bar{x} = \frac{2,714}{2} - \frac{(-3,908)}{2 \cdot 6,671} = 1,357 + 0,293 = 1,650; \quad f(\bar{x}) = 15,142.$$

$$\text{Шаг 6: Проверка на окончание поиска: а) } \left| \frac{15,210-15,142}{15,142} \right| = 0,0045 > 0,003.$$

Второй критерий останова не выполняется. Переход на Шаг 7.

Шаг 7: Выбираем \bar{x} как "наилучшую" точку, а $x_1 = 1$, $x_2 = 1,714$ – как точки, которые ее окружают. Обозначим эти точки в естественном порядке и переходим к Шагу 2.

Итерация 3:

Шаг 4:

$$x_1 = 1; \quad f_1 = 18; \quad x_2 = 1,65; \quad f_2 = 15,142 = F_{\min}; \quad X_{\min} = x_2 = 1,65;$$

$$x_3 = 1,714; \quad f_3 = 15,210.$$

Шаг 5:

$$a_1 = \frac{15,142-18}{1,65-1} = -4,397;$$

$$a_2 = \frac{1}{1,714-1,650} \cdot \left[\frac{15,210-18}{1,714-1} - (-4,397) \right] = 7,647;$$

$$\bar{x} = \frac{2,65}{2} - \frac{(-4,397)}{2 \cdot 7,647} = 1,6125; \quad f(\bar{x}) = 15,123.$$

Шаг 6: Проверка на окончание поиска:

$$\text{а) } \left| \frac{15,142 - 15,123}{15,123} \right| = 0,0013 < 0,003; \quad \text{б) } \left| \frac{1,65 - 1,6125}{1,6125} \right| = 0,023 < 0,03.$$

Оба критерия останова выполняются, следовательно, поиск закончен.

Ответ: $\bar{x} = 1,6125$; (точное решение $x^* = 1,5874$).

8.4 Методы с использованием производных

Рассмотренные выше методы поиска основаны на предположениях об унимодальности и, в ряде случаев, о непрерывности исследуемой целевой функции. Если в дополнение к условию непрерывности ввести требование дифференцируемости, то эффективность поисковых процедур можно существенно повысить.

Напомним, что необходимым условием существования локального минимума функции в некоторой точке x^* является требование стационарности точки x^* , т.е.

$$f'(x^*) = \left. \frac{df}{dx} \right|_{x=x^*} = 0.$$

Рассматриваемые ниже методы поиска экстремумов функций основаны на методах решения нелинейных уравнений с одной переменной.

8.4.1 Метод Ньютона-Рафсона

Если $f(x)$ содержит члены, содержащие x в третьей и более высоких степенях, то непосредственное получение аналитического решения уравнения $f'(x) = 0$ может оказаться затруднительным. В таких случаях используются приближенные методы последовательного поиска стационарной точки $f(x)$. Классическая поисковая схема разработана Ньютоном и позднее уточнена Рафсоном (МНР).

Итерационная формула имеет вид:

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}. \quad (8.10)$$

Она легко получается, если разложить функцию $f'(x)$ в ряд Тейлора и ограничиться линейным приближением: $f'(x_n) \approx f'(x_n) + f''(x_n)(x_{n+1} - x_n) = 0$.

Геометрическая интерпретация приведена на рис. 8.1. Здесь начальной точкой является точка x_n , а искомым экстремумом исследуемой функции точка x^* .

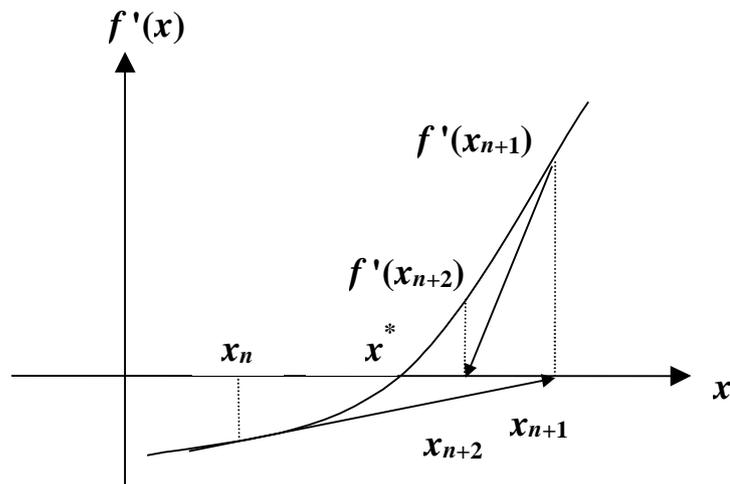


Рис. 8.1. Сходимость МНР

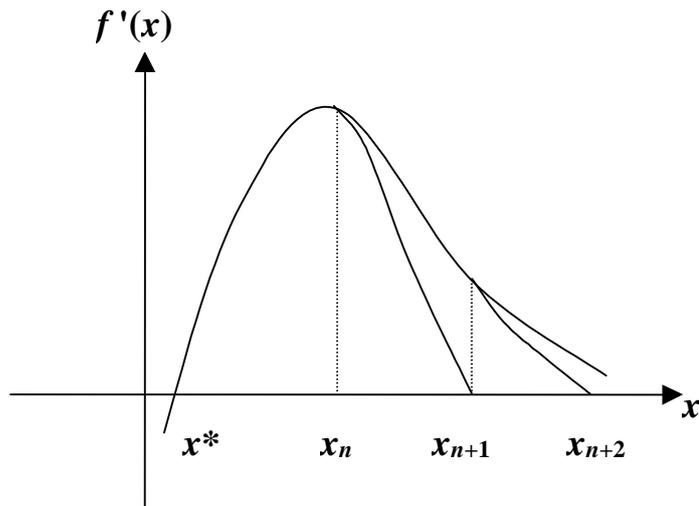


Рис. 8.2. Отсутствие сходимости МНР

Начальное приближение в МНР следует выбирать вблизи глобального экстремума, в противном случае метод будет расходящимся (рис. 8.2)

8.4.2 Метод средней точки (поиск Больцано)

Пусть $f(x) \in Q[a, b]$, т.е. функция унимодальна и дифференцируема на отрезке $[a, b]$.

Для нахождения корня уравнения $f'(x) = 0$ можно воспользоваться эффективным алгоритмом исключения интервалов, на каждой итерации которого рассматривается лишь одна пробная точка.

Если в точке z значение функции $f'(z) < 0$, то точка x^* не может быть левее точки z . В этом случае исключается интервал $[a, z]$.

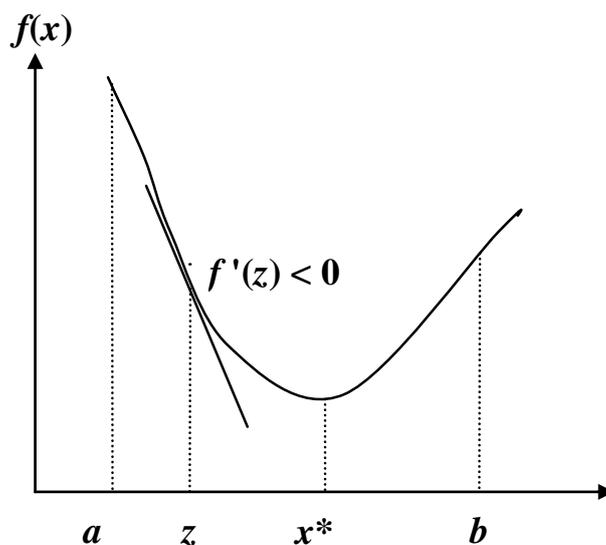


Рис. 8.3а. Поиск Больцано.

Исключается интервал $[a, z]$.

Если в точке z $f'(z) > 0$, то минимум x^* не может быть расположен правее z . Исключается интервал $[z, b]$.

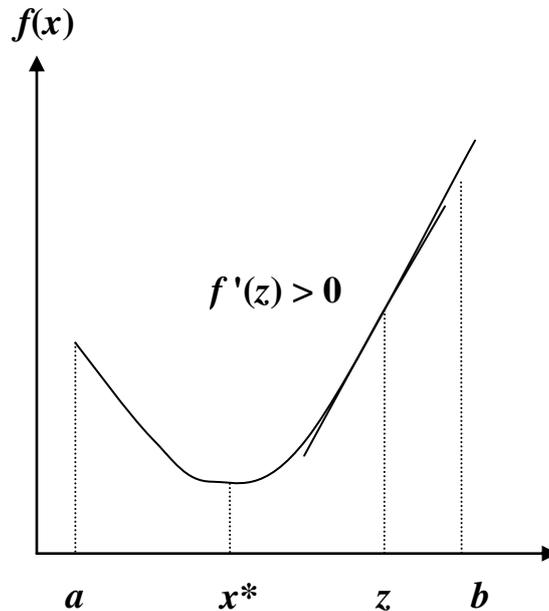


Рис. 8.3б. Поиск Больцано.

Исключается интервал $[z, b]$.

Определим две точки L и R таким образом, что $f'(L) < 0$ и $f'(R) > 0$. Стационарная точка $\bar{x} \in [L, R]$. Вычислим координату средней точки $z = (L + R)/2$ и найдем $f'(z)$.

$$\text{Если } \begin{cases} f'(z) > 0, & \text{то исключим } [z, R]; \\ f'(z) < 0, & \text{то исключим } [L, z]. \end{cases} \quad (8.11)$$

Формализованное описание алгоритма. Пусть задана функция $f(x)$, $x \in [a, b]$ и задана точность ε .

Шаг 1: Положить $L = a$, $R = b$ при этом $f'(a) < 0$; $f'(b) > 0$.

Шаг 2: Вычислить $z = \frac{R + L}{2}$; $f'(z)$.

Шаг 3: Если $|f'(z)| \leq \varepsilon$, закончить поиск. Иначе:

$$\text{Если } \begin{cases} f'(z) > 0, & \text{положить } R = z \rightarrow \text{Шаг 2}; \\ f'(z) < 0, & \text{положить } L = z \rightarrow \text{Шаг 2}. \end{cases}$$

Замечание: Здесь исследуется лишь знак производной независимо от ее значения.

Кроме рассмотренных выше интервальных методов на практике широкое распространение получили следующие методы поиска нулей функции одной переменной:

8.4.3 Другие методы поиска экстремума функций

Упрощенный метод Ньютона (сходимость линейная):

$$x_{n+1} = x_n - f'(x_n) / \alpha, \quad (8.12)$$

где α – константа, например, $\alpha = f''(x_0)$;

Метод ложного положения (сходимость линейная):

$$x_{n+1} = x_n - f'(x_n) \frac{x_n - z}{f'(x_n) - f'(z)}. \quad (8.13)$$

Значение абсциссы вспомогательной точки z необходимо выбирать как можно ближе к корню функции;

Метод Стефенсена (квадратичная сходимость):

$$x_{n+1} = x_n - f'(x_n) \frac{f'(x_n)}{f'[x_n + f'(x_n)] - f'(x_n)}; \quad (8.14)$$

Метод Уолла (кубическая сходимость):

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n) - \frac{f'(x_n) \cdot f'''(x_n)}{2f''(x_n)}}. \quad (8.15)$$

Метод секущих

$$x_{n+1} = x_n - \frac{x_{n-1} - x_n}{f'(x_{n-1}) - f'(x_n)} f'(x_n) \quad (8.16)$$

Метод хорд

$$x_{n+1} = \begin{cases} a - \frac{f'(a)}{f'(x_n) - f'(a)} (x_n - a), \\ x_n - \frac{f'(x_n)}{f'(b) - f'(x_n)} (b - x_n) \end{cases} \quad (8.17)$$

Комбинированный метод

$$\begin{cases} x_{n+1} = a_n - \frac{f'(a_n)(x_n - a_n)}{f'(x_n) - f'(a_n)}, & x_0 = b_0 = b; a_0 = a; n = 1, 2, \dots \\ a_n = a_{n-1} - \frac{f'(a_{n-1})}{f''(a_{n-1})} \end{cases} \quad (8.18a)$$

либо

$$\begin{cases} x_{n+1} = x_n - \frac{f'(x_n)(b_n - x_n)}{f'(b_n) - f'(x_n)}; & x_0 = a_0 = a; b_0 = b; n = 0, 1, 2, \dots \\ b_n = b_{n-1} - \frac{f'(b_{n-1})}{f''(b_{n-1})} \end{cases} \quad (8.18b)$$

8.4.4 Метод оптимизации с использованием кубичной аппроксимации

В этом методе также используются производные. Функция аппроксимируется полиномом третьего порядка. Схема метода аналогична схеме методов с использованием квадратичной аппроксимации, но построение полинома проводится на основе меньшего числа точек, т.к. в каждой точке можно вычислять значения и функции, и ее производной.

Схема метода. Произвольно выбирается точка x_1 , находится точка $x_2 = x_1 \pm \Delta_1$, такая, что $f'(x_1)f'(x_2) < 0$, т.е. производные имеют разные знаки. Иными словами,

стационарную точку \bar{x} , в которой производная $f'(\bar{x}) = 0$, необходимо заключить в интервал $[x_1, x_2]$.

Аппроксимирующий кубический полином записывается в виде:

$$F(x) = a_0 + a_1(x - x_1) + a_2(x - x_1)(x - x_2) + a_3(x - x_1)^2(x - x_2) \quad (10)$$

Параметры уравнения (10) подбираются таким образом, чтобы значения $F(x)$ и ее производной в точках x_1, x_2 совпадали со значениями $f(x), f'(x)$ в этих точках. Первая производная функции $F(x)$ равна:

$$\begin{aligned} \frac{dF(x)}{dx} = & a_1 + a_2(x - x_1) + a_2(x - x_2) + \\ & + a_3(x - x_1)^2 + 2a_3(x - x_1)(x - x_2). \end{aligned} \quad (8.19)$$

Коэффициенты a_0, a_1, a_2, a_3 уравнения (11) определяются по известным значениям $f(x_1), f(x_2), f'(x_1), f'(x_2)$ путем решения следующей системы уравнений:

$$\begin{aligned} f_1 = f(x_1) &= a_0; \\ f_2 = f(x_2) &= a_0 + a_1(x_2 - x_1); \\ f'_1 = f'(x_1) &= a_1 + a_2(x_1 - x_2); \\ f'_2 = f'(x_2) &= a_1 + a_2(x_2 - x_1) + a_3(x_2 - x_1)^2. \end{aligned}$$

Данная система легко решается рекурсивным методом. После того, как коэффициенты найдены, действуя по аналогии со случаем квадратичной аппроксимации, можно оценить координату стационарной точки $f(x)$ с помощью полинома (10). При этом приравнивание к нулю производной (11) приводит к квадратному уравнению. Используя формулу для вычисления корней квадратного уравнения, запишем решение, определяющее стационарную точку кубического полинома, в виде:

$$\bar{x} = \begin{cases} x_2, & \text{если } \mu < 0; \\ x_2 - \mu(x_2 - x_1), & \text{если } 0 \leq \mu \leq 1; \\ x_1, & \text{если } \mu > 1, \end{cases} \quad (8.20)$$

где:

$$\begin{aligned} \mu = \frac{f'_2 + w - z}{f'_2 - f'_1 + 2w}, \quad z = \left[\frac{3 \cdot (f_1 - f_2)}{x_2 - x_1} \right] + f'_1 + f'_2, \\ w = \begin{cases} \sqrt{z^2 - f'_1 f'_2}, & \text{если } x_1 < x_2; \\ -\sqrt{z^2 - f'_1 f'_2}, & \text{если } x_1 > x_2. \end{cases} \end{aligned}$$

Формула для вычисления значения w обеспечивает надлежащий выбор одного из двух корней квадратного уравнения.

Для значений $\mu \in [0, 1]$ формула (8.20) гарантирует, что получаемая точка $\bar{x} \in [x_1, x_2]$. Затем снова выбираются две точки: \bar{x} и одна из двух точек x_1 или x_2 , причем знаки производной функции в этих точках должны быть различны, и процедура повторяется.

Пример 8.3: Найти минимум функции:

$$f(x) = 2x^2 + 16/x; x_0 = 1; \Delta_1 = 1; \varepsilon_1 = 0,01; \varepsilon_2 = 0,03.$$

Ответ: $x = 1,5657$. Точное решение: $x^* = 1,5874$.

Напомним, что решение, полученное с помощью метода Пауэлла: $\bar{x} = 1,6125$.

Это указывает на преимущество полиномиальной аппроксимации более высокого порядка.

8.5 Сравнение методов одномерной оптимизации

С помощью теоретических выкладок можно показать, что такие методы точечного оценивания, как метод Пауэлла, метод поиска с использованием кубичной аппроксимации и производных существенно эффективнее методов исключения интервалов, среди которых выделяется метод золотого сечения. Данный вывод справедлив лишь в предположении, что интервалы сходимости сравнимы между собой, а исследуемая функция является достаточно гладкой и унимодальной. Выбор того или иного метода определяется из следующих соображений:

1. Если необходимо получить решение с очень высокой степенью точности, то лучшими оказываются методы на основе полиномиальной аппроксимации.

2. Если важно добиться надежной работы алгоритма, то целесообразно выбрать метод золотого сечения.

3. Поисковые методы типа метода Пауэлла следует использовать совместно с МЗС, когда возникают затруднения с реализацией соответствующих итераций на ЭВМ.

Для оценки эффективности методов обычно используются три характеристики:

- время, затраченное на получение решения;
- точность решения;
- чувствительность к изменениям параметра сходимости.

Например, метод средней точки, метод Пауэлла и поиск с использованием кубичной аппроксимации могут быть исследованы при поиске минимума функции

$f(x) = \sin^k(x)$, $x \in [0, 2\pi]$ для различных нечетных значений k :

$$k \in [1, 79], x^* = 4,71239, f(x^*) = -1, 0 \text{ для любого } k.$$

При возрастании значения k увеличивается время решения и уменьшается точность. Однако МЗС не чувствителен к росту k .

В заключение этого параграфа отметим, что МЗС обладает высокой вычислительной эффективностью и простотой реализации. Методы точечного оценивания позволяют определить точки экстремума с помощью квадратичной и кубичной аппроксимации целевой функции. Если интервалы сходимости сравнимы между собой, а целевая функция гладкая и унимодальная, то методы точечного оценивания сходятся значительно быстрее, чем методы исключения интервалов. Однако при исследовании мультимодальных или быстроизменяющихся функций (типа $f(x) = \sin^k x$, которая обладает узкими впадинами вблизи точки минимума) наиболее надежным оказывается метод "золотого сечения".

Вопросы для самопроверки

1. Метод равномерного поиска минимума функции.
2. Метод дихотомии поиска минимума функции.
3. Метод золотого сечения поиска минимума функции.
4. Метод Фибоначчи поиска минимума функции.

5. Показатели эффективности методов прямого поиска минимума функции.
6. Погрешности методов прямого поиска минимума функции.
7. Методы полиномиальной аппроксимации поиска минимума функции одной переменной.
8. Суть метода квадратичной аппроксимации.
9. Метод Пауэлла поиска минимума функции одной переменной.
10. Метод Ньютона-Рафсона. поиска минимума функции одной переменной.
11. Метод средней точки (поиск Больцано) поиска минимума функции одной переменной.
12. Какие методы поиска нулей функции используются для одномерной оптимизации функций?
13. Метод поиска минимума функции одной переменной с использованием кубичной аппроксимации.

Тема 9. Поиск экстремумов функции нескольких переменных (безусловная оптимизация)

9.1 Классификация методов безусловной оптимизации

Здесь исследуются методы и алгоритмы, позволяющие на итерационной основе получать оценки решения оптимизационной задачи x^* , которым соответствует минимальное значение целевой функции $f(x)$. Указанные методы применимы и к задачам максимизации, в которых ЦФ заменяется на обратную: $[-f(x)]$.

Задачи, которые здесь рассматриваются, имеют следующий вид: найти

$$\min f(x), \text{ где } f(x) = f(x_1, x_2, \dots, x_n), \quad x_i \in (-\infty, +\infty). \quad (9.1)$$

При определенных условиях многие задачи с ограничениями типа:

$$\min f(x), \quad g_i(x) \leq 0, \quad i = \overline{1, m},$$

можно свести к ЗО вида (9.1), т.е. к задаче безусловной оптимизации.

Напомним необходимые и достаточные условия существования локального минимума. Предполагается, что $f(x)$ непрерывна и имеет непрерывные 1-е и 2-е производные $\partial f(x)/\partial x_i$ и $\partial^2 f(x)/(\partial x_i \cdot \partial x_j)$ в любой точке $x \in R^n$:

- 1) $\nabla f(x^*) = 0$, (x^* - стационарная точка);
- 2) $\nabla^2 f(x^*)$ – матрица Гессе должна быть положительно определена.

Это условие означает, что квадратичная форма

$$y^T \nabla^2 f(x^*) y > 0, \quad \forall y \in R^n, \quad y \neq 0.$$

Для выпуклой непрерывно дифференцируемой функции необходимо и достаточно лишь выполнение равенства: $\nabla f(x^*) = 0$.

Методы, ориентированные на решение задач безусловной оптимизации, можно условно разделить на три больших класса:

- *Методы прямого поиска* (нулевого порядка), основанные на вычислении только значений ЦФ.
- *Градиентные методы* (методы 1-го порядка), в которых используются точные значения первых производных.
- *Методы 2-го порядка*, в которых используются также вторые производные целевой функции $f(x)$.

Ни один метод или класс методов не отличается высокой эффективностью при решении оптимизационных задач различных типов. В некоторых ситуациях вычисление значений ЦФ требует чрезмерных затрат времени. Иногда невозможно или очень трудно найти аналитическое выражение для производных целевой функции. Поэтому при использовании градиентных методов следует применять процедуру разностной аппроксимации производных. Таким образом, в каждом конкретном случае необходимо приспосабливать применяемый метод к конкретным характеристикам решаемой ЗО.

9.2. Методы прямого поиска

Рассмотрим два метода прямого поиска:

1. Поиск по симплексу или s^2 -метод.
2. Метод Хука-Дживса.

Достоинства данных методов: относительная простота вычислительных процедур, легкость реализации и корректировки.

Недостатки: значительные затраты времени по сравнению с градиентными методами.

9.2.1 Симплексный метод

Метод поиска по симплексу был предложен в 1962 г. Спендли, Хекстом и Химсвортом. Этот метод называют *последовательным симплекс-методом* (ПСМ). Следует отметить, что указанный метод и другие подобные методы не имеют отношения к симплекс-методу линейного программирования, а сходство названий носит чисто случайный характер.

Определение: В k -мерном евклидовом пространстве k -мерный симплекс представляет собой фигуру, образованную $k+1$ точками (вершинами), не принадлежащими одновременно ни одному пространству меньшей размерности.

В одномерном пространстве симплекс есть отрезок прямой; в двумерном – треугольник; в трехмерном – треугольная пирамида (тетраэдр) и т.д.

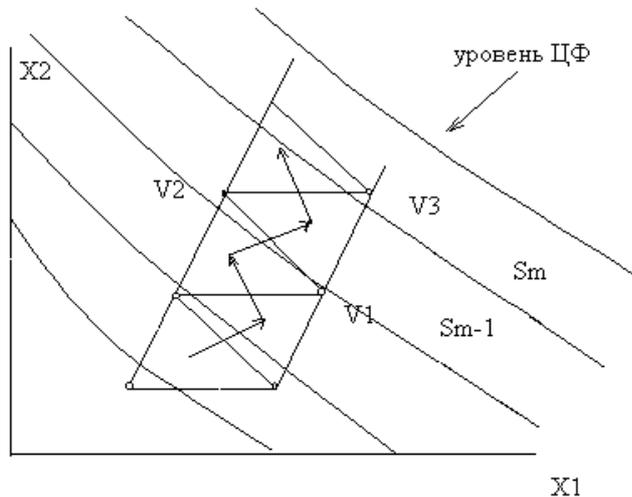


Рис. 9.1. Траектория движения по симплексу

Из любого симплекса, отбросив одну его вершину, можно получить новый симплекс, если к оставшимся добавит всего лишь одну точку.

Симплекс называется *регулярным*, если расстояния между вершинами равны. В ПСМ используются регулярные симплекс-планы.

Для оценки направления движения во всех вершинах симплекса V_j , $j = 1, 2, \dots, n+1$, где n - размерность вектора x , необходимо оценить значение ЦФ $f_j = f(V_j)$.

При поиске минимума наиболее целесообразно будет движение от вершины V_s с наибольшим значением f_s к противоположной грани симплекса. Шаг поиска выполняется переходом из некоторого симплекса S_{m-1} в новый симплекс S_m путем исключения вершины V_s и построения ее зеркального отображения относительно общей грани V_s (см. рис.1). Многократное отражение худших вершин приводит к шаговому движению центра симплекса к цели по траектории некоторой ломанной линии.

Алгоритм ПСМ

ШАГ 1.

Задается исходная вершина симплекса.

$$x^0 = (x_1^0, \dots, x_n^0)$$

Задается коэффициент сжатия $\gamma \in [0, 1]$ и размер симплекса L . Строится симплекс

$$(x_i^j) = \begin{pmatrix} x_1^0 & \dots & x_n^0 \\ x_1^1 & \dots & x_n^1 \\ \dots & \dots & \dots \\ x_1^{n-1} & \dots & x_n^{n-1} \\ x_1^n & \dots & x_n^n \end{pmatrix}. \quad (9.2)$$

Здесь j -я строка – это координаты j -ой вершины V_j . ($j = 1, \dots, n+1$), где n - размерность пространства (размерность вектора x), i – номер координаты $i = 1, \dots, n$.

Определение координат x_i^j , начиная со второй, производится по формуле

$$x_i^j = x_i^0 + \tilde{x}_i^j, \quad (j = 1, \dots, n; i = 1, \dots, n), \quad (9.3)$$

где \tilde{x}_i^j - матрица размерности $(n+1) \times n$

$$(\tilde{x}_i^j) = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ p_n & q_n & q_n & \dots & q_n \\ q_n & p_n & q_n & \dots & q_n \\ \dots & \dots & \dots & \dots & \dots \\ q_n & q_n & q_n & \dots & p_n \end{pmatrix}, \quad (9.4)$$

где

$$p_n = \frac{L}{n\sqrt{2}}(\sqrt{n+1} + n - 1), \quad q_n = p_n - \frac{\sqrt{2}}{2} = \frac{L}{n\sqrt{2}}(\sqrt{n+1} - 1).$$

Векторы соответствующие вершинам V_1, \dots, V_n , определяемые формулой (9.2), составят одинаковые углы с координатными осями x_1, \dots, x_n .

ШАГ 2.

В вершинах симплекса вычисляется ЦФ $f(x^j)$, $j = 0, \dots, n$

ШАГ 3.

Проверяем условия: $\|x^j - x^{j-1}\| \leq \varepsilon_1$, $|f(x^j) - f(x^{j+1})| \leq \varepsilon_2$

Если «да», то конец; если «нет», то на ШАГ 4.

ШАГ 4.

Находится «наихудшая» вершина симплекса (при поиске минимума «наихудшая» вершина – та, в которой значение функции максимально).

$$f(x^p) = \max_j \{f(x^j), j = \overline{1, n+1}\}$$

ШАГ 5.

Осуществляется расчет координат новой вершины (вершина отражения x^p):

$$\tilde{x}^p = \frac{2}{n} \left(\sum_{j=0}^n x^j - x^p \right) - x^p$$

ШАГ 6.

Если точка \tilde{x}^p оказывается «хуже» всех остальных точек симплекса, то осуществляется возврат к исходному симплексу с последующим его сжатием относительно «лучшей» из вершин x^k

$$f(x^k) = \min_j \{ f(x^j), j = \overline{1, n+1} \}$$

$$\tilde{x}^s = \gamma x^k + (1 - \gamma) x^s, \quad s = 0, 1, \dots, n; \quad s \neq k,$$

Переход на ШАГ 2.

Если \tilde{x}^p не является «худшей» в новом симплексе, то перейти на ШАГ 3.

9.2.2 Метод Хука-Дживса

Процедура Хука–Дживса представляет собой комбинацию двух поисков:

а) "*исследующий*" поиск (для выявления характера локального поведения ЦФ и определения направления движения вдоль "оврагов") с циклическим изменением переменных;

б) *ускоряющий поиск по образцу* с использованием определенных эвристических правил.

Исследующий поиск. Выбирается некоторая исходная точка x^0 . Задается величина шага Δ_i , которая может быть различной для разных координатных направлений и изменяться в процессе поиска.

Если значение ЦФ в пробной точке меньше значения ЦФ в исходной точке, то шаг поиска успешный. В противном случае из исходной точки делается шаг в противоположном направлении. После перебора всех n координат исследующий поиск завершается. Полученная точка называется *базовой*.

Поиск по образцу. Осуществляется шаг из полученной базовой точки вдоль прямой, соединяющей эту точку с предыдущей базовой. Новая точка образца определяется по формуле:

$$x_p^{k+1} = x^k + (x^k - x^{k-1}).$$

Как только движение по образцу не приводит к уменьшению ЦФ, точка x_p^{k+1} фиксируется в качестве временной базовой точки и вновь проводится исследующий поиск. Если в результате получается точка с меньшим значением ЦФ, чем в точке x^k , то она рассматривается как новая базовая точка x^{k+1} . Но если исследующий поиск неудачен, то следует вернуться в точку x^k и провести исследующий поиск с целью выявления нового направления минимизации. В конечном итоге возникает ситуация, когда такой поиск не приводит к успеху. В этом случае уменьшается шаг путем введения коэффициента α и возобновляется исследующий поиск.

Схема алгоритма Хука – Дживса

Введем следующие обозначения: x^k - текущая базовая точка; x^{k-1} - предыдущая базовая точка; x_p^{k+1} - точка, построенная при движении по образцу; x^{k+1} - следующая (новая) базовая точка.

Критерий останова: $\|\Delta x\| \leq \varepsilon$.

Шаг 1. Определить начальную точку x^0 ; приращения (шаги) Δ_i , $i = \overline{1, n}$; коэффициент уменьшения шага $\alpha > 1$; параметр окончания поиска .

Шаг 2. Провести исследующий поиск.

Шаг 3. Был ли исследующий поиск удачным (найдена ли точка с меньшим значением ЦФ)?

Да: переход на Шаг 5. Нет: продолжить, т.е. переход на Шаг 4.

Шаг 4. Проверка на окончание поиска. Выполняется ли неравенство $\|\Delta x\| \leq \varepsilon$? Да: окончание поиска, т.е. текущая точка аппроксимирует точку экстремума x^* .

Нет: уменьшить приращение Δ_i / α ; $i = 1, 2, \dots, n$. Переход на Шаг 2.

Шаг 5. Провести поиск по образцу: $x_p^{k+1} = x^k + (x^k - x^{k-1})$.

Шаг 6. Провести исследующий поиск, используя точку x_p^{k+1} в качестве временной базовой точки. Пусть в результате получена точка x^{k+1} .

Шаг 7. Выполняется ли неравенство: $f(x^{k+1}) < f(x^k)$?

Да: положить $x^{k-1} = x^k$; $x^k = x^{k+1}$. Переход на Шаг 5.

Нет: переход на Шаг 4.

Пример 9.1. Найти точку минимума ЦФ

$$f(x) = 8x_1^2 + 4x_1x_2 + 5x_2^2;$$

начальная точка: $x^0 = [-4; -4]^T$.

Решение. Зададим следующие величины:

$\Delta x = [1; 1]^T$ - векторная величина приращения шага;

$\alpha = 2$ - коэффициент уменьшения шага Δx ;

$\varepsilon = 10^{-4}$ - параметр окончания поиска.

Итерации начинаются с исследующего поиска вокруг точки x^0 , которой соответствует значение ЦФ $f(x^0) = 272$.

Фиксируя переменную $x_2 = -4$, дадим приращение x_1 :

$$x_1 = -4 + 1 \rightarrow f(-3; -4) = 200 < f(x^0) \rightarrow \text{успех.}$$

Далее, фиксируем $x_1 = -3$ и дадим приращение x_2 :

$$x_2 = -4 + 1 \rightarrow f(-3; -3) = 153 < 200 \rightarrow \text{успех.}$$

Таким образом, в результате исследующего поиска найдена точка $x^1 = [-3; -3]^T$, в которой значение ЦФ $f(x^1) = 153$.

Так как исследующий поиск был удачным, переходим к поиску по образцу:

$$x_p^2 = x^1 + (x^1 - x^0) = [-2; -2]^T; \quad f(x_p^2) = 68.$$

Далее проводится исследующий поиск вокруг точки x_p^2 . В результате получаем точку $x^2 = [-1; -1]^T$, в которой значение ЦФ $f(x^2) = 17$.

Поскольку $f(x^2) < f(x^1)$, поиск по образцу следует считать успешным, и x^2 становится новой базовой точкой. Итерации продолжаются до тех пор, пока уменьшение величины шага не укажет на окончание поиска в ε -окрестности точки минимума $x^* = [0; 0]^T$.

Достоинства метода Хука-Дживса: несложная стратегия поиска, простота вычислений, малый объем требуемой памяти (меньше чем в симплексе - методе).

Недостатки: при наличии значительных нелинейных эффектов процесс вырождается в последовательность исследующих поисков без перехода к ускоряющему поиску по образцу.

Возможные варианты модификации метода Хука - Дживса:

- а) если движение по образцу приводит к успеху, то желательно увеличить длину шага по образцу, чтобы полностью использовать возможность поиска вдоль прямой;
- б) введение дополнительных правил увеличения и уменьшения приращения переменных и другое.

9.3 Градиентные методы

Предполагается, что целевая функция $f(x)$ непрерывна и имеет, по крайней мере, непрерывные первые производные. Необходимым условием существования экстремума является наличие стационарной точки ЦФ. Таким образом, основная идея многих методов оптимизации без ограничений в пространстве R^n заключается в отыскании стационарной точки x^* , в которой градиент ЦФ $\nabla f(x^*) = 0$.

Эта задача эквивалентна решению нелинейной системы уравнений вида:

$$\begin{cases} \frac{\partial f(x)}{\partial x_i} = 0, & i = \overline{1, n}. \end{cases}$$

Можно отыскивать решение непосредственно этой системы, что приводит к методу Ньютона. При этом предполагается, что функция дважды дифференцируема.

В данном случае речь идет об итерационных процессах, порождающих последовательность точек x^0, x^1, \dots, x^k , сходящихся к локальному экстремуму функции f в точке x^* .

На каждом k -ом этапе значение x^{k+1} определяется выражением:

$$x^{k+1} = x^k + \lambda_k d_k,$$

где d_k - направление перемещения, которое может быть:

- либо градиентом функции в точке x^k , т.е. $d_k = -\nabla f(x^k)$;
- либо вычисленным, исходя из направления градиента $\nabla f(x^k)$;
- либо выбранным произвольно при условии, что это будет направление спуска. В

этом случае должно выполняться неравенство вида: $\nabla f^T(x^k) d_k < 0$.

Здесь λ_k - параметр, характеризующий длину шага.

Способ определения d_k и λ_k на каждой итерации связан с особенностями применяемого метода. Обычно выбор λ_k осуществляется путем решения задачи минимизации $f(x)$ в направлении d_k . Поэтому при реализации изучаемых методов необходимо использовать эффективные методы одномерной оптимизации.

В семействе градиентных методов следует выделить методы с заданным шагом, в которых заранее задаются значения λ_k .

Доказано, что построенная последовательность сходится к решению, т.е. $x^k \rightarrow x^*$, если выполняются два условия:

- 1) $\lambda_k \rightarrow 0$ при $k \rightarrow \infty$;
- 2) $\sum_{k=0}^{\infty} \lambda_k \rightarrow +\infty$ (например, $\lambda_k = \frac{1}{k}$).

Данная процедура ("*метод расходящегося ряда*") может оказаться медленной. *Критерии останова* (наиболее употребительные).

Пусть $\varepsilon > 0$ - заданная точность:

- 1) $\max_{i=1,n} \left| \frac{\partial f}{\partial x_i} \right| \leq \varepsilon$;
- 2) $\|\nabla f\|^2 = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2 \leq \varepsilon$;
- 3) $|f(x^{k+1}) - f(x^k)| \leq \varepsilon$;
- 4) $\|\nabla f\| = \sqrt{\sum_{i=1}^n \left(\frac{\partial f}{\partial x_i} \right)^2} \leq \varepsilon$;
- 5) $\|\Delta x\| = \|x^{k+1} - x^k\| \leq \varepsilon$.

9.3.1 Метод сопряженных направлений

Данный метод и его модификации являются наиболее эффективными алгоритмами прямого поиска.

Метод ориентирован на решение задач с квадратичными целевыми функциями и основывается на фундаментальных теоретических результатах.

Основная идея алгоритма: если квадратичная функция $f(x)$, $x = (x_1, \dots, x_n)$, приведена к виду суммы полных квадратов, то ее экстремум может быть найден в результате реализации n одномерных поисков по преобразованным координатным направлениям, т.е. не более чем за n шагов.

Рассмотрим общие принципы. Пусть имеем квадратичную функцию (в окрестности точки оптимума любую нелинейную функцию можно аппроксимировать квадратичной функцией) вида

$$f(x) = \frac{1}{2} x^T A x + b^T x + c,$$

где A – положительно определенная симметричная матрица порядка $n \times n$; b – n -

вектор $(b_1, \dots, b_n)^T$; c – константа.

Суть методов сопряженных направлений состоит в том, чтобы начиная от точки x^0 , минимизировать последовательно $f(x)$ по n линейно независимым направлениям d_0, d_1, \dots, d_{n-1} , обладающим свойством быть попарно сопряженными относительно квадратичной формы $f(x)$.

Это условие записывается в виде

$$\forall i, j \ (i, j = 0, \dots, n-1) \Rightarrow d_i^T A d_j = 0. \quad (9.3)$$

Итак, предположим, что точки $x^k, k = \overline{0, n-1}$ определены, начиная с x^{k-1} , формулой

$$x^k = x^{k-1} - \lambda_k d_{k-1},$$

где λ_k – значение величины λ , минимизирующее значение одномерной функции $f(x^{k-1} - \lambda d_{k-1})$.

Теорема. Точка

$$x^n = x^0 - \sum_{j=0}^{n-1} \lambda_{j+1} d_j \quad (9.4)$$

полученная на n -ом этапе, является точкой минимума функции $f(x)$, т.е. будет удовлетворять соотношению:

$$\nabla f(x^n) = A x^n + b = 0. \quad (9.5)$$

Таким образом, начиная с точки x^0 шаг за шагом строится последовательность x^1, x^2, \dots, x^n , определенная соотношениями:

$$\begin{aligned} f(x^1) &= f(x^0 - \lambda_1 d_0) = \min_{\lambda} f(x^0 - \lambda d_0); \\ f(x^2) &= f(x^1 - \lambda_2 d_1) = \min_{\lambda} f(x^1 - \lambda d_1); \\ &\dots\dots\dots \\ f(x^n) &= f(x^{n-1} - \lambda_n d_{n-1}) = \min_{\lambda} f(x^{n-1} - \lambda d_{n-1}). \end{aligned}$$

Точка x^n и является точкой минимума функции $f(x)$, т.е. $x^* = x^n$.

Займемся теперь вопросом вычисления сопряженных направлений d_0, d_1, \dots, d_{n-1} . Зададим начальную точку x^0 , начальное направление d_0 и вычислим точку x^1 по формуле

$$x^1 = x^0 - \lambda_1 d_0, \quad (9.10)$$

где λ_1 – скалярная величина, определяется из условия минимума функции $f(x)$ вдоль направления d_0 , т.е. из условия $\nabla f^T(x^1) d_0 = 0$ (см. формулу (9.7) при $k=1$).

Определим направление d_1 по формуле

$$d_1 = -\nabla f(x^1) + \mu_1 d_0 \quad (9.11)$$

так, чтобы оно было сопряжено с направлением d_0 , т.е.

$$d_0^T A d_1 = 0. \quad (9.12)$$

Здесь μ_1 – скалярный параметр, который мы будем определять из уравнения (9.12).

Разложим градиент $\nabla f(x)$ в окрестности точки x^0 и подставим $x = x^1$:

$\nabla f(x^1) - \nabla f(x^0) = A(x^1 - x^0)$. С учетом (10) получим

$\nabla f(x^1) - \nabla f(x^0) = -\lambda_1 A d_0$. Отсюда получим выражение для d_0^T :

$$d_0^T = -\frac{(\nabla f(x^1) - \nabla f(x^0))^T A^{-1}}{\lambda_1}. \quad (9.13)$$

Подставим выражения (9.11) и (9.13) в формулу (9.12), получим:

$(\nabla f(x^1) - \nabla f(x^0))^T (-\nabla f(x^1) + \mu_1 d_0) = 0$. Отсюда выражение для μ_1 :

$$\mu_1 = \frac{(\nabla f(x^1) - \nabla f(x^0))^T \nabla f(x^1)}{(\nabla f(x^1) - \nabla f(x^0))^T d_0} = \frac{(A(x^1 - x^0))^T (Ax^1 + b)}{(A(x^1 - x^0))^T d_0}. \quad (9.14)$$

Если потребовать выполнения дополнительного условия $(\nabla f(x^1))^T \nabla f(x^0) = 0$, то вместо (14) получим

$$\mu_1 = \frac{(\nabla f(x^1) - \nabla f(x^0))^T \nabla f(x^1)}{(\nabla f(x^1) - \nabla f(x^0))^T d_0} = \frac{(Ax^1 + b)^T (Ax^1 + b)}{(A(x^1 - x^0))^T d_0}. \quad (9.15)$$

Далее, направление d_2 находим по формуле $d_2 = -\nabla f(x^2) + \mu_2 d_1$, где $x^2 = x^1 - \lambda_2 d_1$. Проводя аналогичные рассуждения, получим для μ_2 :

$$\mu_2 = \frac{(\nabla f(x^2) - \nabla f(x^1))^T \nabla f(x^2)}{(\nabla f(x^2) - \nabla f(x^1))^T d_1} = \frac{(A(x^2 - x^1))^T (Ax^2 + b)}{(A(x^2 - x^1))^T d_1}.$$

Потребовав выполнения дополнительного условия $(\nabla f(x^2))^T \nabla f(x^1) = 0$, получим

$$\mu_2 = \frac{(\nabla f(x^2) - \nabla f(x^1))^T \nabla f(x^2)}{(\nabla f(x^2) - \nabla f(x^1))^T d_1} = \frac{(Ax^2 + b)^T (Ax^2 + b)}{(A(x^2 - x^1))^T d_1}$$

На k -ом шаге получим:

$$\mu_k = \frac{(\nabla f(x^k) - \nabla f(x^{k-1}))^T \nabla f(x^k)}{(\nabla f(x^k) - \nabla f(x^{k-1}))^T d_{k-1}} = \frac{(A(x^k - x^{k-1}))^T (Ax^k + b)}{(A(x^k - x^{k-1}))^T d_{k-1}}, \quad (9.16)$$

или с учетом условия $(\nabla f(x^k))^T \nabla f(x^{k-1}) = 0$

$$\mu_k = \frac{(\nabla f(x^k))^T \nabla f(x^k)}{(\nabla f(x^k) - \nabla f(x^{k-1}))^T d_{k-1}} = \frac{(Ax^k + b)^T (Ax^k + b)}{(A(x^k - x^{k-1}))^T d_{k-1}}, \quad (9.17)$$

$$d_k = -\nabla f(x^k) + \mu_k d_{k-1}, \quad (9.18)$$

где

$$x^k = x^{k-1} - \lambda_k d_{k-1}, \quad (9.19)$$

$$\lambda_k = \frac{d_{k-1}^T (Ax^{k-1} + b)}{d_{k-1}^T A d_{k-1}}, \quad k = 1, 2, \dots, n. \quad (9.20)$$

В качестве начального направления d_0 можно взять, например,

$$d_0 = -\nabla f(x^0) = -(Ax^0 + b).$$

Сформулируем теперь алгоритм поиска минимума функции $f(x)$ методом сопряженных направлений.

Шаг 0. Задаем начальную точку x^0 и вычисляем начальное направление d_0 по формуле (19).

Шаг k . По формуле (9.20) вычисляем λ_k и затем по формуле (9.19) — точку x^k . Затем по формулам (9.16) (либо (9.17)) и (9.18) вычисляем параметр μ_k и направление d_k . При $k = n$ процесс вычислений заканчиваем. В качестве точки минимума берем точку x^n .

Замечание. Формула для μ_k может быть получена непосредственно из (9.18) с учетом условия (9.12) $d_0^T A d_1 = 0$. В результате имеем

$$\mu_k = \frac{(\nabla f(x^k))^T A d_{k-1}}{d_{k-1}^T A d_{k-1}}. \quad (9.22)$$

К такому же результату мы придем, если в формуле (9.16) заменим $(x^k - x^{k-1})$ на $(-\lambda_k d_{k-1})$ (см. формулу (9.19)).

Пример 9.2. Найти минимум функции $f(x) = 8x_1^2 + 4x_1x_2 + 5x_2^2 + x_1x_3 + 2x_2x_3 + x_3^2$ методом сопряженных направлений. Начальная точка $x^0 = (5, 5, 5)^T$.

Решение. Вычислим матрицу Гессе $A = \nabla^2 f(x)$ и вектор коэффициентов линейной части

$$A = \begin{pmatrix} 16 & 4 & 1 \\ 4 & 10 & 2 \\ 1 & 2 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Тогда функцию $f(x)$ можно представить в матричной форме $f(x) = \frac{1}{2} x^T A x + b^T x + c$, где $c = 0$.

Шаг 0. Вычисляем начальное направление $d_0 = -(Ax^0 + b) = -(105, 80, 25)^T$.

Шаг 1. Вычисляем $\lambda_1 = \frac{d_0^T (Ax^0 + b)}{d_0^T A d_0} = -0.056$,

$$x^1 = x^0 - \lambda_1 d_0 = (-0.884, 0.516, 3.599)^T, \mu_1 = \frac{(A(x^1 - x^0))^T (Ax^1 + b)}{(A(x^1 - x^0))^T d_0} = 0.011,$$

$$d_1 = -\nabla f(x^1) + \mu_1 d_0 = -(Ax^1 + b) + \mu_1 d_0 = (7.292, -9.734, -7.630)^T.$$

Шаг 2. Вычисляем $\lambda_2 = \frac{d_1^T (Ax^1 + b)}{d_1^T A d_1} = -0.132,$

$$x^2 = x^1 - \lambda_2 d_1 = (0.085, -0.777, 2.584)^T, \mu_2 = \frac{(A(x^2 - x^1))^T (Ax^2 + b)}{(A(x^2 - x^1))^T d_1} = 0.095,$$

$$d_2 = -\nabla f(x^2) + \mu_2 d_1 = -(Ax^2 + b) + \mu_2 d_1 = (-0.147, 1.333, -4.429)^T.$$

Шаг 3. Вычисляем $\lambda_3 = \frac{d_2^T (Ax^2 + b)}{d_2^T A d_2} = -0.583,$

$$x^3 = x^2 - \lambda_3 d_2 = (0, 0, 0)^T. \text{ Точка минимума } x^* = x^3 = (0, 0, 0)^T.$$

9.3.2 Метод наискорейшего спуска (метод Коши)

Известный французский математик Коши Огюстен Луи (21.08.1789–23.05.1857) первым использовал аналогичный алгоритм для решения системы линейных уравнений.

В этом широко используемом методе (МК) λ_k выбираются так, чтобы минимизировать функцию по λ :

$$g(\lambda) = f[x^k - \lambda \nabla f(x^k)]$$

на множестве значений $\lambda \geq 0$ (одномерная минимизация).

Алгоритм Коши

Шаг 1. Выбрать начальную точку x^0 .

Шаг 2. На k -ой итерации, где $d_k = -\nabla f(x^k)$, найти такое λ_k , что $f(x^k + \lambda_k d_k) = \min_{\lambda \geq 0} f(x^k + \lambda d_k)$.

Положить $x^{k+1} = x^k + \lambda_k d_k$.

Шаг 3. Проверка критерия останова.

Да: окончание поиска \rightarrow конец. Нет: $k = k + 1, \rightarrow$ Ш. 2.

МК обладает устойчивостью, т.е. при достаточно малом значении λ_k обеспечивается выполнение неравенства:

$$f(x^{k+1}) \leq f(x^k).$$

МК позволяет существенно уменьшить значение ЦФ (быстро спуститься на "дно оврага") при движении из точек, расположенных на значительных расстояниях от точки x^* , но в дальнейшем может произойти "зацикливание". Поэтому МК часто используются совместно с другими градиентными методами (например, методом Ньютона) в качестве начальной процедуры.

Недостаток: для некоторых типов функций сходимость может оказаться медленной. В самом деле, если λ_k минимизирует функцию

$$g(\lambda) = f(x^k + \lambda d_k), \text{ то должно быть}$$

$$\frac{dg(\lambda_k)}{d\lambda} = d_k^T \nabla f(x^k + \lambda_k d_k) = d_k^T \nabla f(x^{k+1}) = 0,$$

откуда вытекает равенство:

$$d_k^T d_{k+1} = 0,$$

которое доказывает, что последовательные направления ортогональны. В случае

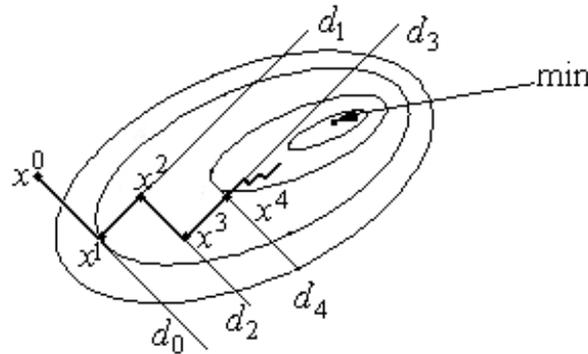


Рис. 9.2. Метод Коши-«зацикливание» траектории спуска

ярко выраженной нелинейности ("овражности") ЦФ происходит "зацикливание" (см. рис. 9.2).

Пример 9.3. Методом Коши найти

$$\min f(x) = 8x_1^2 + 4x_1x_2 + 5x_2^2; \quad x^0 = [10; 10]^T.$$

Критерии останова: $\|\nabla f(x^k)\| \leq \varepsilon$ или $\|x^{k+1} - x^k\| \leq \varepsilon$.

Решение. Вычислим компоненты градиента:

$$\frac{\partial f}{\partial x_1} = 16x_1 + 4x_2; \quad \frac{\partial f}{\partial x_2} = 10x_2 + 4x_1.$$

С помощью формулы $x^{k+1} = x^k - \lambda_k \nabla f(x^k)$ построим первое приближение:

$$x^1 = x^0 - \lambda_0 \nabla f(x^0).$$

Выберем λ_0 таким образом, чтобы минимизировать функцию по λ :

$$f(x^1) \rightarrow \min_{\lambda} \rightarrow \lambda_0 = 0,056.$$

Следовательно: $x^1 = [-1,18; 2,14]^T$. Далее найдем точку

$$x^2 = x^1 - \lambda_1 \nabla f(x^1),$$

вычислив градиент в точке x^1 и проведя поиск вдоль прямой.

Данные, полученные при проведении итерации на основе одномерного поиска по методу квадратичной интерполяции, приведены в табл. 9.1.

Таблица 9.1

| k | x_1^k | x_2^k | $f(x^k)$ |
|-----|---------|---------|----------|
| 1 | -1,2403 | 2,13 | 24,23 |
| 2 | 0,1441 | 0,1447 | 0,354 |
| 3 | -0,0181 | 0,0309 | 0,0052 |
| 4 | 0,0021 | 0,0021 | 0,0000 |

Несмотря на то, что МК не имеет большого практического значения, он реализует важнейшие шаги большинства градиентных методов.

9.3.3 Метод Ньютона (МН)

Итерационная формула МН имеет вид:

$$x^{k+1} = x^k - \left[\nabla^2 f(x^k) \right]^{-1} \nabla f(x^k).$$

Эта формула есть не что иное, как МН в применении к решению системы нелинейных уравнений:

$$\frac{\partial f(x)}{\partial x_i} = 0, \quad i = 1, \dots, n.$$

Замечание. В МН и направление, и шаг перемещения фиксированы.

Если $f(x)$ строго выпуклая функция, то МН сходится за одну итерацию.

Пример 9.4. Методом Ньютона найти

$$\min f(x) = 8x_1^2 + 4x_1x_2 + 5x_2^2, \text{ где } x^0 = [10; 10]^T.$$

Решение. Найдем градиент и матрицу Гессе в начальной точке

$$\nabla f(x) = (16x_1 + 4x_2; 10x_2 + 4x_1),$$

$$H_f(x) = \nabla^2 f(x^0) = \begin{pmatrix} 16 & 4 \\ 4 & 10 \end{pmatrix}.$$

Используя формулу МН, получаем:

$$x^1 = [10; 10]^T - \frac{1}{144} \cdot \begin{pmatrix} 10 & -4 \\ -4 & 16 \end{pmatrix} \cdot \begin{pmatrix} 200 \\ 140 \end{pmatrix} = [0; 0]^T,$$

что совпадает с точным решением.

Если функция $f(x)$ не квадратичная, то МН не отличается высоким быстродействием и надежностью.

9.3.4 Модифицированный метод Ньютона (МН)

В модифицированном МН последовательность итераций строится в соответствии с формулой:

$$x^{k+1} = x^k - \lambda_k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k).$$

Выбор λ_k осуществляется таким образом, чтобы минимизировать ЦФ по λ :

$$f(x^{k+1}) = f\left(x^k - \lambda [\nabla^2 f(x^k)]^{-1} \nabla f(x^k)\right) \rightarrow \min_{\lambda}.$$

Это гарантирует выполнение неравенства

$$f(x^{k+1}) \leq f(x^k).$$

Недостаток. При использовании модифицированного метода Ньютона на каждой итерации возникает необходимость построения и решения линейного уравнения, содержащего элементы матрицы Гессе $\nabla^2 f(x^k)$, т.е. направление спуска d_k находим из решения системы линейных уравнений

$[\nabla^2 f(x^k)] d_k = \nabla f(x^k)$ и затем вычисляем следующее приближение точки экстремума

$$x^{k+1} = x^k - \lambda_k d_k.$$

9.3.5 Метод Флетчера–Ривза (МФР)

Запишем формулу для μ_k (9.20)

$$\mu_k = \frac{(\nabla f(x^k) - \nabla f(x^{k-1}))^T \nabla f(x^k)}{(\nabla f(x^k) - \nabla f(x^{k-1}))^T d_{k-1}} \quad (9.23)$$

Потребуем выполнения следующего условия:

$$(\nabla f(x^k))^T \nabla f(x^{k-1}) = 0.$$

Кроме того, выбор длины шага λ_k (см. (20) производится из условия,

$$(\nabla f(x^k))^T d_{k-1} = 0, \text{ где } x^k = x^{k-1} - \lambda_k d_{k-1}.$$

Тогда (9.23) примет вид

$$\mu_k = - \frac{(\nabla f(x^k))^T \nabla f(x^k)}{(\nabla f(x^{k-1}))^T d_{k-1}} \quad (9.24)$$

Если в формуле (24) в знаменателе вместо d_{k-1} использовать антиградиент функции в точке x^{k-1} , т.е. $d_{k-1} = -\nabla f(x^{k-1})$, то получим метод Флетчера и Ривза.

Алгоритм Флетчера-Ривза

Шаг 1. Задаем начальную точку x^0 ; положить $d_0 = -\nabla f(x^0)$.

Шаг 2. Выбрать значение λ_k , минимизирующее значение функции $\Phi(\lambda) = f(x^{k-1} - \lambda d_k)$.

Положить:

$$x^k = x^{k-1} - \lambda_k d_{k-1}, \quad d_k = -\nabla f(x^k) + \mu_k d_{k-1}, \quad \mu_k = \frac{\|\nabla f(x^k)\|^2}{\|\nabla f(x^{k-1})\|^2}.$$

Шаг 3. Проверка критерия останова:

Да: \rightarrow конец. Нет: $k = k + 1 \rightarrow$ Ш. 2.

Достоинство МФР: малый объем памяти (необходимо хранить $3n$ -мерных вектора $\nabla f(x^k)$ и $\nabla f(x^{k-1})$, а также текущее направление d_k); скорость сходимости значительно выше классических градиентных методов.

Пример 9.5. Найти точку минимума целевой функции методом Флетчера-Ривза.

$$f(x) = 4x_1^2 + 3x_2^2 - 4x_1x_2 + x_1, \quad x^0 = [0; 0]^T.$$

Решение.

Шаг 1. $\nabla f(x) = (8x_1 - 4x_2 + 1, \quad 6x_2 - 4x_1), \quad d_0 = -\nabla f(x^0) = [-1; 0]^T.$

Шаг 2. Поиск вдоль прямой:

$$x^1 = x^0 - \lambda_1 \nabla f(x^0) \rightarrow \lambda_1 = 1/8;$$

$$x^1 = [0; 0] - (1/8)[1; 0] = [-1/8; 0].$$

Шаг 3. $k = 1: \nabla f(x^1) = [0; 1/2]; \mu_1 = 1/4;$

$$d_1 = -\nabla f(x^1) + \mu_1 d_0 = -[0; 1/2] + \frac{1}{4} [-1; 0] = [-1/4; -1/2].$$

Шаг 4. Поиск вдоль прямой:

$$x^2 = x^1 - \lambda_2 d_1 \rightarrow \lambda_2 = -1/4,$$

$$x^2 = [-1/8; 0] - \frac{1}{4} \times [1/4; 1/2] = [-3/16; -1/8],$$

$$\nabla f(x^2) = [0; 0]^T - \text{следовательно, имеем стационарную точку.}$$

Таким образом, $x^* = x^2$.

Решение получено в результате проведения двух одномерных поисков, а т.к. ЦФ - квадратичная функция, то ошибка округления отсутствует.

9.3.6 Вариант Полака-Рибьера (МПР)

Метод основан на точной процедуре проведения поиска вдоль прямой и на более общем предположении об аппроксимации ЦФ.

Метод реализуется с помощью следующих соотношений:

$$x^k = x^{k-1} - \lambda_k d_{k-1}, \quad d_k = -\nabla f_k + \mu_k d_{k-1},$$

$$\mu_k = \frac{\nabla f_k^T \Delta f_k}{\|\nabla f_{k-1}\|^2}, \quad (9.25)$$

где $\Delta f_k = \nabla f_k - \nabla f_{k-1}$; λ_k - параметр, значение которого определяется в результате поиска вдоль прямой.

Заметим, что формула (9.25) следует из (9.23). Действительно, знаменатель с учетом условия $(\nabla f(x^k))^T d_{k-1} = 0$ (из этого уравнения определяется длина шага λ_k)

равен $(-\nabla f_{k-1})^T d_{k-1}$. Если теперь положить $d_{k-1} = -\nabla f(x^{k-1})$, как мы это делали в методе Флетчера-Ривза, то придем к формуле (9.25)

Легко видеть, что единственное различие между методами Флетчера-Ривза и Полака-Рибьера заключается в способах выбора параметра μ .

9.4 Квазиньютоновские методы (КМ) (методы с переменной метрикой)

Квазиньютоновские методы также основаны на свойствах квадратичных функций. Данные методы обладают положительными чертами метода Ньютона, однако используют только первые производные. С чем это связано?

Итерационный поиск по методу Ньютона осуществлялся по формуле (обобщенный случай с λ):

$$x^{k+1} = x^k - \lambda_k [\nabla^2 f(x^k)]^{-1} \nabla f(x^k).$$

Трудность может возникнуть, когда матрица Гессе $\nabla^2 f(x^k) = H_f(x^k)$ не является положительно определенной. В этом случае направление перемещения

$$d_k = -[\nabla^2 f(x^k)]^{-1} \nabla f(x^k) = -(H_f)^{-1} \nabla f(x^k)$$

может не быть направлением спуска и глобальная сходимость метода не будет обеспечена.

В таких ситуациях обратную матрицу Гессе $[\nabla^2 f(x)]^{-1}$ заменяют положительно определенной матрицей A^k , дающей направление перемещения, исходя из градиента $\nabla f(x^k)$. Отсюда получаем итерационную формулу

$$x^{k+1} = x^k - \lambda_k A^k \nabla f(x^k),$$

где λ_k выбирается так, чтобы минимизировать функцию $f(x^k - \lambda d_k)$ в направлении

$$d_k = -A^k \nabla f(x^k).$$

Для аппроксимации матрицы $H_f^{-1} = [\nabla^2 f(x)]^{-1}$ пользуются следующим рекуррентным соотношением:

$$A^{k+1} = A^k + A_c^k,$$

где A_c^k - корректирующая матрица.

Для матрицы A_c^k получена следующая оценка

$$A_c^k = \frac{(\Delta x^k - A^k \Delta g^k)(\Delta x^k - A^k \Delta g^k)^T}{(\Delta g^k)^T (\Delta x^k - A^k \Delta g^k)},$$

где $\Delta x^k = x^{k+1} - x^k$, $\Delta g^k = g(x^{k+1}) - g(x^k) = \nabla f_{k+1} - \nabla f_k$.

Доказано, что для квадратичной функции $f(x)$ с положительно определенной матрицей A последовательность матриц A^k сходится не более чем за n этапов к обращению матрицы A^{-1} гессиана функции $f(x)$.

9.4.1 Метод Дэвидона–Флетчера–Пауэлла (ДФП)

Корректирующая матрица определяется по формуле:

$$\begin{cases} A_c^k = \frac{\Delta x^k (\Delta x^k)^T}{(\Delta x^k)^T \Delta g^k} - \frac{A^k \Delta g^k (\Delta g^k)^T (A^k)^T}{(\Delta g^k)^T (A^k)^T \Delta g^k}, \\ A^{k+1} = A^k + A_c^k. \end{cases} \quad (9.26)$$

Данная рекуррентная формула сохраняет свойства симметрии и положительной определенности матриц. Поэтому, если A^0 - симметричная положительно определенная матрица, то матрицы A^1, A^2, \dots будут так же симметричными положительно определенными матрицами.

Обычно удобно выбирать $A^0 = E$, где E - единичная матрица. Точка x^{k+1} получается из x^k перемещением в направлении $d_k = -A^k \nabla f(x^k)$.

В формуле (26): $\Delta x^k = x^{k+1} - x^k$; $\Delta g^k = \nabla f(x^{k+1}) - \nabla f(x^k)$.

Свойство сохранения положительной определенности матриц существенно, ибо оно гарантирует, что направления d_k , последовательно порождаемые алгоритмом, являются направлениями спуска. Действительно, первая вариация $f(x)$ равна $\Delta f(x) = \nabla f^T(x^k) \Delta x$. Используя формулы (15) и (16), получаем:

$$\Delta f(x^k) = -\lambda_k \nabla f^T(x^k) A^k \nabla f(x^k) < 0,$$

откуда $f(x^{k+1}) < f(x^k)$ при любых $\lambda_k > 0$, если A^k положительно определена.

Достоинства метода

1. Метод ДФП – наиболее широко используемый градиентный метод.
2. Устойчивость – при решении самых различных задач, возникающих на практике.

Недостаток. Необходимость хранения матрицы A^k порядка $n \times n$.

Пример 9.6. Найти минимум ЦФ $f(x)$ методом ДФП.

$$f(x) = 4x_1^2 + 3x_2^2 - 4x_1x_2 + x_1 \rightarrow \min, \quad x^0 = [0; 0].$$

Решение.

Шаг 1. Положить $d_0 = -\nabla f(x^0) = -[1; 0]$.

Шаг 2. Поиск вдоль d_0 приводит к результату: $\lambda_1 = 1/8$, $x^1 = [-1/8; 0]$.

$$\text{Шаг 3.} \quad k=1; \quad A^1 = A^0 + A_c^0;$$

$$A^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E;$$

$$A_c^0 = \frac{\Delta x^0 \Delta x^{0T}}{\Delta x^{0T} \Delta g^0} - \frac{A^0 \Delta g^0 \Delta g^{0T} A^0}{\Delta g^{0T} A^0 \Delta g^0};$$

$$\Delta x^0 = [-1/8; 0] - [0; 0] = [-1/8; 0];$$

$$\Delta g^0 = \nabla f(x^1) - \nabla f(x^0) = [0; 1/2] - [1; 0] = [-1; 1/2];$$

$$A_c^0 = \frac{[-1/8; 0]^T [-1/8; 0]}{[-1/8; 0] [-1; 1/2]^T} - \frac{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} [-1; 1/2]^T [-1; 1/2] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}{[-1; 1/2] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} [-1; \dots 1/2]^T};$$

$$A_c^0 = \begin{pmatrix} 1/8 & 0 \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} 4/5 & -2/5 \\ -2/5 & 1/5 \end{pmatrix} = \begin{pmatrix} -27/40 & 2/5 \\ 2/5 & -1/5 \end{pmatrix};$$

$$A^1 = A^0 + A_c^0 = \begin{pmatrix} 13/40 & 2/5 \\ 2/5 & 4/5 \end{pmatrix};$$

$$d_1 = -A^1 \nabla f(x^1) = -[1/5; 2/5].$$

Шаг 4. Поиск вдоль прямой

$$x^2 = x^1 - \lambda_2 d_1 \rightarrow \lambda_2 = 5/16;$$

$$x^2 = [-1/8; 0] - (5/16) \times [1/5; 2/5] = [-3/16; -1/8],$$

что совпадает с полученным ранее решением данного примера методом Флетчера-Ривза.

Вопросы для самопроверки

1. Классификация методов многомерной оптимизации.
2. Симплекс-метод поиска минимума функции многих переменных.
3. Алгоритм симплекс-метода поиска минимума функции многих переменных
4. Метод Хука-Дживса.
5. Градиентные методы поиска минимума функции многих переменных.
6. Метод сопряженных направлений.
7. Метод Коши.
8. Метод Ньютона.
9. Модифицированный метод Ньютона.
10. Метод Флетчера-Ривза.
11. Метод Поллака-Рибьера.
12. Квазиньютоновские методы с переменной метрикой.
13. Метод Дэвидона-Флетчера-Пауэлла.

Тема 10. Нелинейное программирование

10.1. Задачи с ограничениями в виде равенств

Рассмотрим общую задачу оптимизации с ограничениями – равенствами:

$$f(x) \rightarrow \min \quad x \in R^n \quad (10.1)$$

$$h_i(x) = 0 \quad i = \overline{1, m} \quad (10.2)$$

Функции $f(x)$ и $h_i(x)$ предполагаются непрерывными и дифференцируемыми.

10.1.1 Метод замены переменных (МЗП)

Задача нелинейного программирования в принципе может быть решена как задача безусловной оптимизации, путем исключения из целевой функции m независимых переменных с помощью заданных равенств (10.2). Наличие ограничений в виде равенств фактически позволяет уменьшить размерность исходной задачи с n до $(n - m)$. Для решения задачи безусловной минимизации можно использовать известные методы, изложенные ранее.

Пример 10.1.

$$f(x) = x_1^2 + x_2^2 \rightarrow \min$$

$$x_1 - x_2 = 1$$

Решение: $x_1 = 1 + x_2$ подставим в $f(x)$

$$f(x) = (1 + x_2)^2 + x_2^2 \rightarrow \min$$

$$f'(x) = 2(1 + x_2) + 2x_2 = 0 \Rightarrow x_1 = \frac{1}{2}; x_2 = -\frac{1}{2}$$

Пусть имеется m ограничений равенств, причем $m < n$. Если $m = n$, то задача оптимизации сводится к решению системы уравнений (2) и оптимизация здесь не нужна.

Таким образом, экстремальная точка функции $f(x)$ ищется лишь по тем значениям x , которые удовлетворяют независимой системе (2) при $m < n$, т. е. в некоторой части S пространства R^n ($S \subset R^n$).

Пусть первые m переменных x_1, \dots, x_m , относительно которых мы хотим разрешить систему (10.2), т. е. вектор x можно представить в виде:

$$x = \begin{pmatrix} y \\ u \end{pmatrix}; \quad y = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \equiv \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}; \quad u = \begin{pmatrix} x_{m+1} \\ \vdots \\ x_n \end{pmatrix} \equiv \begin{pmatrix} u_1 \\ \vdots \\ u_{n-m} \end{pmatrix} \quad (10.3)$$

Решаем систему (10.2):

$$y = \Psi(u) \quad u \in R^{n-m} \quad (10.4)$$

и подставляем u в целевую функцию $f(x) = f(y, u) = f(\Psi(u), u)$. Получим функцию $(n - m)$ переменных u_1, \dots, u_{n-m} , на которую не наложено никаких ограничений, т. е. получим задачу оптимизации без ограничений.

Для идентификации точки экстремума целевой функции можно использовать один из известных методов.

МЗП применим лишь в случаях, когда уравнения ограничения можно разрешить относительно некоторого конкретного набора независимых переменных. При наличии большого числа ограничений МЗП становится весьма трудоемкой процедурой. Кроме того, возможны ситуации, когда уравнение не удастся разрешить относительно переменной, например:

$$h_1(x) = x_1^2 x_3 + x_2 x_3^2 + x_2^{-1} x_1$$

В этом случае целесообразно использовать метод множителей Лагранжа.

10.1.2 Метод множителей Лагранжа (ММЛ)

С помощью ММЛ устанавливают необходимое условие, позволяющее идентифицировать точки экстремума в задаче оптимизации с ограничениями – равенствами. При этом задача с ограничениями преобразуется в эквивалентную задачу безусловной оптимизации, в которой фигурируют некоторые неизвестные параметры – множители Лагранжа.

Рассмотрим задачу оптимизации:

$$\begin{aligned} f(x_1, \dots, x_n) &\rightarrow \min, \\ h(x_1, \dots, x_n) &= 0. \end{aligned}$$

В соответствии с ММЛ эта задача преобразуется в задачу оптимизации без ограничений.

$$L(x, \lambda) = f(x) + \lambda h_1(x) \rightarrow \min.$$

Функция $L(x, \lambda)$ – функция Лагранжа. λ – *const* – множитель Лагранжа, на знак которого никаких требований не накладывается.

Проиллюстрируем ММЛ на конкретном примере:

Пример 10.2

$$\begin{aligned} \min f(x) &= x_1^2 + x_2^2 \\ h(x) &= 2x_1 + x_2 - 2 = 0 \end{aligned}$$

Соответствующая задача оптимизации без ограничений записывается в следующем виде:

$$L(x, \lambda) = x_1^2 + x_2^2 + \lambda(2x_1 + x_2 - 2) \rightarrow \min$$

Решение:

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 2x_1 + 2\lambda = 0 \rightarrow x_1^* = -\lambda \\ \frac{\partial L}{\partial x_2} &= 2x_2 + \lambda = 0 \rightarrow x_2^* = -\frac{\lambda}{2} \end{aligned}$$

Для того, чтобы проверить, соответствует ли стационарная точка x^* минимуму, вычислим матрицу Гессе функции $L(x, \lambda)$, рассматриваемой как функция x

$$H_L(x, \lambda) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

которая оказывается положительно определенной.

Это означает, что $L(x, \lambda)$ - выпуклая функция x .

Следовательно, координаты $x^* = (-\lambda, -\frac{\lambda}{2})$ определяют точку глобального

минимума. Оптимальное значение λ находится путем подстановки значений x_1^* и x_2^* в уравнение ограничений $2x_1 + x_2 - 2 = 0$, откуда:

$$2\lambda + \frac{\lambda}{2} = -2 \rightarrow \lambda^* = -\frac{4}{5}$$

Таким образом, минимум достигается при $x_1^* = \frac{4}{5}$ и $x_2^* = \frac{2}{5}$ и

$$f^* = \min f(x) = \frac{4}{5}.$$

Точка $(x_1^*, x_2^*, \lambda^*)$ - является седловой точкой функции Лагранжа $L(x, \lambda)$, если функцию Лагранжа рассматривать как функцию трех переменных x_1, x_2 и λ .

ММЛ можно распространить на случай, когда задача оптимизации имеет несколько ограничений - равенств. Рассмотрим общую задачу:

$$\begin{aligned} f(x_1, \dots, x_n) &\rightarrow \min, \\ h_i(x_1, \dots, x_n) &= 0, \quad i = \overline{1, m}. \end{aligned}$$

Функция Лагранжа принимает следующий вид:

$$L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i h_i(x),$$

где $\lambda_1, \dots, \lambda_m$ - множители Лагранжа, т. е. неизвестные параметры, значения которых необходимо определить.

Далее находят частные производные

$$\frac{\partial L}{\partial x_j}, \frac{\partial L}{\partial \lambda_i}, \quad j = \overline{1, n}; i = \overline{1, m}$$

и рассматривают систему $(n + m)$ уравнений с $(n + m)$ неизвестными x_j, λ_i :

$$\begin{cases} \frac{\partial L}{\partial x_j} = \frac{\partial f(x)}{\partial x_j} + \sum_{i=1}^m \lambda_i \frac{\partial h_i(x)}{\partial x_j} = 0 & j = \overline{1, n} \rightarrow \nabla_x L = 0 \\ \frac{\partial L}{\partial \lambda_i} = h_i(x_1, \dots, x_n) = 0 & i = \overline{1, m} \rightarrow \nabla_\lambda L = 0 \end{cases}$$

Это и есть необходимое условие экстремума функции Лагранжа.

Решение расширенной системы определяет стационарную точку функции L . Затем реализуется процедура проверки на минимум или максимум, которая проводится на основе вычисления элементов матрицы Гессе функции L , рассматриваемой как функция x .

Для некоторых задач расширенная система $(n + m)$ уравнений с $(n + m)$ неизвестными может не иметь решений, и ММЛ окажется неприемлемым. Однако на практике такие задачи встречаются редко.

10.2. Необходимые и достаточные условия оптимальности задач с ограничениями общего вида

Необходимые и достаточные условия оптимальности первого порядка.

Выше было установлено, что множители Лагранжа можно использовать при построении критериев оптимальности для задач оптимизации с ограничениями в виде равенств. Кун и Таккер обобщили этот подход на случай общей задачи нелинейного программирования с ограничениями как в виде равенств, так и в виде неравенств.

Рассмотрим общую задачу нелинейного программирования

$$\begin{aligned} f(x) &\rightarrow \min \\ g_i(x) &\begin{cases} \geq 0, & i = \overline{1, m}; \\ \leq 0, & i = \overline{m+1, m+p}; \end{cases} \\ h_l(x) &= 0, \quad l = \overline{1, p}. \end{aligned} \quad (10.1)$$

$f(x), g_i(x), h_l(x)$ – непрерывные и дифференцируемые функции.

Необходимые и достаточные условия первого порядка выражаются следующей теоремой:

Теорема 10.1: Для того, чтобы точка x^* была локальным оптимумом задачи (1), необходимо и достаточно, чтобы выполнялись следующие условия:

- а) $f(x)$ была выпукла;
- б) $g_i(x)$ – вогнута, если $g_i(x) \geq 0$,
– выпукла, если $g_i(x) \leq 0$;
- в) h_l – линейные;
- г) существовали такие числа $\mu_i^* \geq 0$ и λ_l^* (произвольного знака), что

$$\nabla f(x^*) + \sum_{i=1}^m \mu_i^* \nabla g_i(x^*) + \sum_{l=1}^p \lambda_l^* \nabla h_l(x^*) = 0; \quad (10.2)$$

$$\mu_i^* g_i(x^*) = 0; \quad i = \overline{1, m} \quad (10.3)$$

$$h_l(x^*) = 0; \quad l = \overline{1, p}. \quad (10.4)$$

В случаях задачи с ограничениями только типа равенств:

$$f(x) \rightarrow \min$$

$$h_l(x) = 0 \quad l = \overline{1, p}$$

Функцию Лагранжа имеет вид:

$$L(x, \lambda) = f(x) + \sum_{l=1}^p \lambda_l h_l(x)$$

Для этой функции условия оптимальности первого порядка выражаются следующими уравнениями:

$$\nabla_x L(x, \mu) = \nabla f(x) + \sum_{l=1}^p \lambda_l \nabla h_l(x)$$

$$\nabla_{\lambda} L(x, \mu) = h_l(x) = 0 \quad l = \overline{1, p}$$

Т. е. условия Куна – Таккера (2) и (4) совпадают с условиями оптимальности для задачи Лагранжа.

Пример 10.3.

$$f(x) = x_1^2 - x_2 \rightarrow \min$$

$$g_1(x) = x_1 - 1 \geq 0,$$

$$g_2(x) = 26 - x_1^2 - x_2^2 \geq 0,$$

$$h_1(x) = x_1 + x_2 - 6 = 0.$$

Проверим выполнение условий Куна-Таккера. Вычислим матрицы вторых производных целевой функции и функций ограничений.

$$\nabla f(x) = (2x_1, -1), \quad H_f(x) = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

H_f положительно полуопределена при всех x , следовательно $f(x)$ – выпуклая функция.

$g_1(x)$ – линейная функция, которая одновременно является выпуклой и вогнутой.

Матрица Гессе для Функции $g_2(x)$ имеет вид:

$$\nabla g_2(x) = (-2x_1, -2x_2); \quad H_{g_2} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix};$$

$$\Delta_1 = -2; \quad \Delta_2 = 4.$$

Так как матрица H_{g_2} отрицательно определена, то $g_2(x)$ является вогнутой.

Функция $h_1(x)$ входит в линейное ограничение в виде равенства. Следовательно, условия

(а), (б) и (в) теоремы (1) выполнены. Найдем теперь точку оптимума x^* . Имеем:

$$f(x) = x_1^2 - x_2 \quad \nabla f(x) = (2x_1, -1)$$

$$g_1(x) = x_1 - 1 \quad \nabla g_1(x) = (1, 0)$$

$$g_2(x) = 26 - x_1^2 - x_2^2 \quad \nabla g_2(x) = (-2x_1, -2x_2)$$

$$h_1(x) = x_1 + x_2 - 6 \quad \nabla h_1(x) = (1, 1)$$

Уравнение (2) принимает следующий вид:

$$\frac{\partial f}{\partial x_j} - \mu_1 \frac{\partial g_1}{\partial x_j} - \mu_2 \frac{\partial g_2}{\partial x_j} + \lambda_1 \frac{\partial h_1}{\partial x_j} = 0 \quad j = 1, 2$$

$$j = 1, \quad 2x_1 - \mu_1 + 2x_1\mu_2 + \lambda_1 = 0$$

откуда:

$$j = 2, \quad -1 + 2x_2\mu_2 + \lambda_1 = 0$$

Таким образом, условия Куна – Таккера (10.2) – (10.4) для нашего примера имеют следующий вид:

$$2x_1 - \mu_1 + 2x_1\mu_2 + \lambda_1 = 0,$$

$$-1 + 2x_2\mu_2 + \lambda_1 = 0,$$

$$x_1 + x_2 - 6 = 0,$$

$$\mu_1(x_1 - 1) = 0,$$

$$\mu_2(26 - x_1^2 - x_2^2) = 0,$$

$$\mu_{1,2} \geq 0$$

Из третьего уравнения находим: $x_1 = 6 - x_2$. Подставим x_1 в остальные уравнения, получим:

$$2(6 - x_2) - \mu_1 + 2\mu_2(6 - x_2) + \lambda_1 = 0$$

$$-1 + 2\mu_2x_2 + \lambda_1 = 0$$

$$\mu_1(6 - x_2 - 1) = 0 \Rightarrow x_2 = 5; \Rightarrow x_1 = 1$$

$$\mu_2(26 - (6 - x_2)^2 - x_2^2) = 0$$

Подставим $x_2 = 5$ в первое и второе уравнения:

$$\begin{cases} 2 - \mu_1 + 2\mu_2 + \lambda_1 = 0 \\ -1 + 10\mu_2 + \lambda_1 = 0 \end{cases}$$

Из второго уравнения выразим λ_1 и подставим в первое уравнение:

$$3 - \mu_1 - 8\mu_2 = 0$$

Пусть $\mu_1 = 0.1 > 0$, тогда $\mu_2 = 0.3625 > 0$. Таким образом, точка $x^* = (1, 5)$ является точкой минимума.

Вопросы для самопроверки

1. Записать задачу нелинейного программирования с ограничениями в виде равенств.
2. Метод замены переменных решения задачи нелинейного программирования с ограничениями в виде равенств.
3. Метод множителей Лагранжа решения задачи нелинейного программирования с ограничениями в виде равенств.
4. Решение задачи нелинейного программирования с ограничениями общего вида.

Тема 11. Методы штрафов

11.1. Общая схема метода штрафов

Здесь мы продолжим изучение задач оптимизации с ограничениями. Нас будут интересовать методы, существенно отличающиеся от рассмотренных выше. Их общий принцип заключается в замене исходной задачи на решение последовательности экстремальных задач без ограничений.

Рассмотрим задачу:

$$\begin{aligned} f(x) \rightarrow \min \quad x \in R^n & \quad (11.1) \\ g_i(x) \geq 0, \quad i = \overline{1, m}; \\ h_l(x) = 0, \quad l = \overline{1, p}. \end{aligned}$$

Предполагается, что точка x^* является решением этой задачи, известно некоторое начальное приближение x^0 , возможно недопустимое, т. е. не удовлетворяющее ограничениям задачи (11.1). С помощью рассматриваемых далее алгоритмов в пространстве R^n строится конечная последовательность точек $x^t, t = 0, 1, \dots, k$, которая начинается с заданной точки x^0 и заканчивается точкой x^k , дающей наилучшее приближение к x^* среди всех точек построенной последовательности. В качестве x^t берутся стационарные точки штрафной функции – ЦФ вспомогательной задачи безусловной минимизации. Конкретные методы, основанные на указанной общей схеме, определяются видом штрафной функции (ШФ), а также правилами, по которым производится пересчет штрафных параметров по окончании очередного цикла безусловной минимизации. Идея преобразования задачи с ограничениями в надлежащим образом построенную последовательность задач без ограничений представляется заманчивой главным образом в связи с наличием эффективных и надежных методов безусловной минимизации. При этом будем надеяться искать условный минимум с приемлемой точностью путем решения относительно небольшого числа не слишком сложных задач.

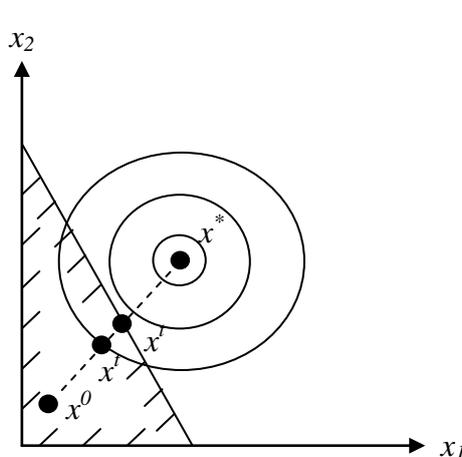


Рис. 11.1. Метод внутренних штрафов

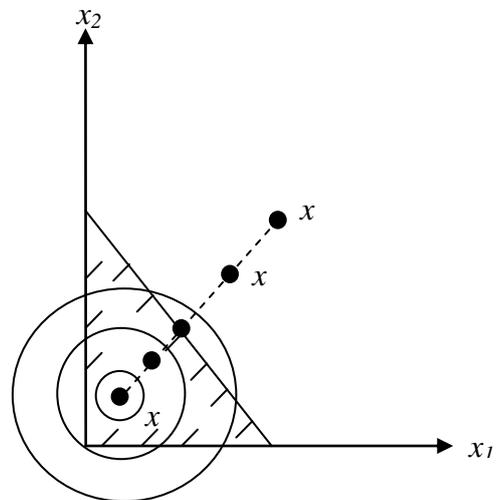


Рис. 11.2 Метод внешних штрафов

В зависимости от того, являются ли элементы последовательности x^t допустимыми или недопустимыми точками, говорят о методах **внутренней** или **внешней** точки соответственно. Иногда их называют методы **внутреннего** или **внешнего** штрафа. Методы внутренних штрафных функций называют также методами **барьерных функций**.

ШФ определяется выражением:

$$P(x, R) = f(x) + H(R, g(x), h(x)), \quad (11.2)$$

где R - набор штрафных параметров,

H - штраф – функция R и ограничений: $H(R, g(x), h(x)) = R\varphi(g(x), h(x))$.

Выражение для штрафа H определяется так, чтобы допустимые точки задачи имели преимущество перед недопустимыми в отношении безусловной минимизации ШФ.

Методы внутренней точки связаны с такими H , при которых стационарные точки функции $P(x, R)$ оказываются заведомо допустимыми. Эти методы называют также методами барьеров, поскольку здесь штраф как бы создает вдоль границы допустимой области S барьер из бесконечно больших значений функции P .

Методы ШФ представляют большой интерес лишь при выполнении следующих требований:

1. решения подзадач должны стремиться к решению исходной задачи нелинейного программирования:

$$\lim x^t \rightarrow x^*, \quad t \rightarrow k < \infty$$
2. сложность минимизации $P(x, R)$ должна быть того же порядка, что и для функции $f(x)$
3. правило пересчета $R^{t+1} = F(R^t)$ должно быть простым.

11.2. Основные типы штрафов

Рассмотрим широко используемые типы штрафов и различные процедуры учета ограничений при переходе к задачам безусловной минимизации.

11.2.1 Квадратичный штраф.

Этот штраф используется для учета ограничений – равенств.

$$H = R \cdot (h(x))^2 \quad (11.3)$$

При минимизации этот штраф препятствует отклонению величины $h(x)$ от нуля (как вправо, так и влево). Ниже будет видно, что при увеличении R стационарная точка соответствующей ШФ $P(x, R)$ приближается к точке x^* . т. к. в пределе $h(x^k) = 0$. Поскольку допустимая область S определяется ограничением $h(x) = 0$, то в допустимой области $H = 0$, а вне допустимой области $H > 0$ и тем больше, чем дальше точка x^t выйдет за пределы допустимой области и чем больше R .

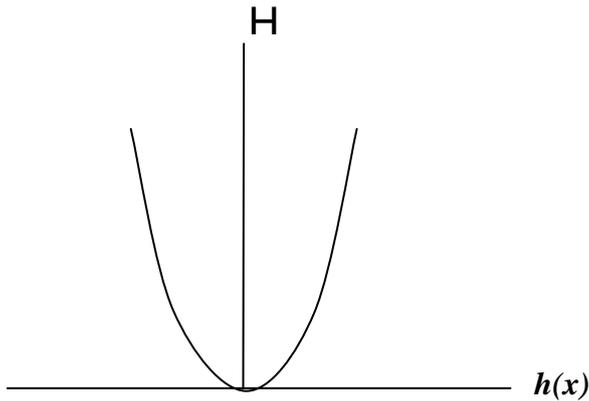


Рис. 11.3. Квадратичный штраф

Рассмотрим несколько типов штрафов для ограничений – неравенств.

11.2.1 Бесконечный барьер.

Задается в следующем виде

$$\begin{aligned} H &= \infty, \text{ а́ñëè } x \notin S(g_i(x) < 0), \\ H &= 0, \text{ а́ñëè } x \in S(g_i(x) \geq 0). \end{aligned} \quad (11.4)$$

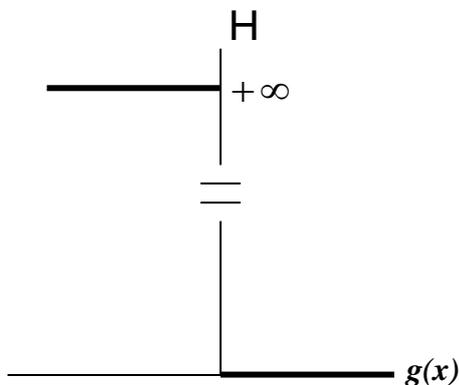


Рис. 11.4. Бесконечный барьер

Для этого штрафа функция $P(x, R)$ разрывна и недифференцируема на границе допустимой области.

Так как машинная реализация бесконечных штрафов невозможна, то используется большое положительное число, например, $H = 10^{20} \sum_{i \in I} |g_i(x)|$, где I - множество индексов нарушенных ограничений, т. е. $g_i(x) < 0$ при $i \in I$.

11.2.3 Логарифмический штраф

Логарифмический штраф задается в форме

$$H = -R \ln(g(x)) \quad (11.5)$$

Штраф положителен при всех x таких, что $0 < g(x) < 1$ и отрицателен при $g(x) > 1$. В данном случае вводится как бы искусственная дискриминация точек допустимой области: внутренним точкам отдается предпочтение перед граничными точками. Отрицательных значений штрафа можно было бы избежать, положив $H = 0$ для таких x , что $g(x) > 1$. Однако при этом у градиента ∇P появились бы разрывы.

Логарифмический штраф – это барьерная функция, не определенная в недопустимых точках (т. е. для таких x , для которых $g(x) < 0$).

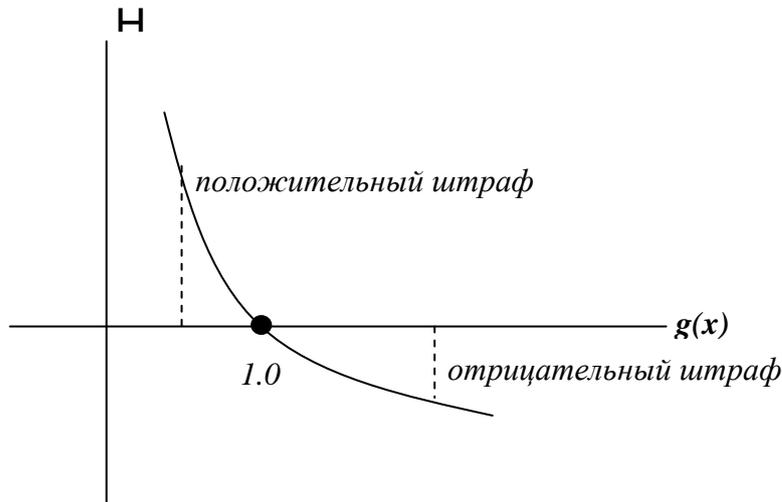


Рис. 11.5. Логарифмический штраф

Поэтому в тех случаях, когда приходится иметь дело с недопустимыми точками (например, когда точка $x^0 \notin S$) требуется процедура специального типа, обеспечивающая попадание в допустимую область S .

Итерационный процесс необходимо начинать из точки x^0 допустимой области при положительном начальном значении R (например, $R = 10$ или $R = 100$). После решения каждой подзадачи безусловной минимизации, параметр R уменьшается и в пределе стремится к нулю.

11.2.4 Штраф типа обратной функции

Задается в формулой

$$H = R \frac{1}{g(x)}. \quad (11.6).$$

Этот штраф внутренний, Итерационный процесс необходимо начинать из точки x^0 допустимой области.

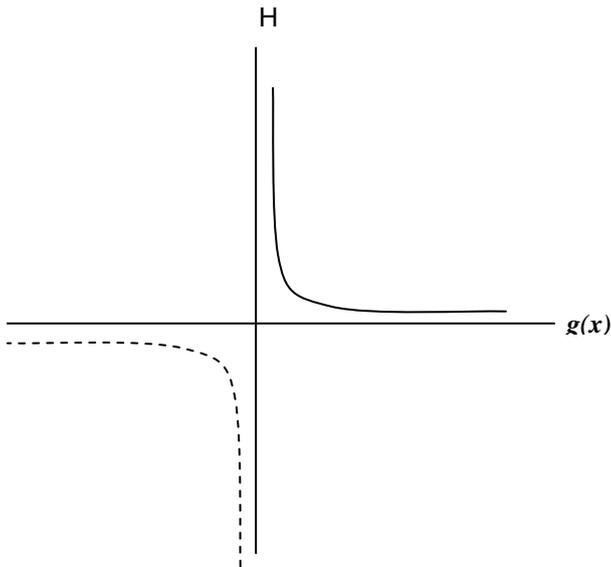


Рис. 11.6. Штраф типа обратной функции

H здесь в отличие от логарифмического штрафа не имеет отрицательных значений в допустимой области. Данный штраф называют также барьерной функцией.

Очевидна необходимость предотвращения появления недопустимых точек, в которых $H < 0$. В точках $x \in S$ вблизи границы значения штрафа положительны и быстро убывают при удалении от границы внутрь допустимой области. На самой границе $P(x, R)$ и ∇P не определены.

Итерации начинаются с точки x^0 – начальной допустимой точки при $R > 0$, причем в процессе вычисления $R \rightarrow 0$.

11.2.5 Штраф типа квадрата срезки

Штраф типа квадрата срезки является внешним штрафом. Этот штраф задается формулой

$$H = R \langle g(x) \rangle^2, \quad (11.7)$$

$$\langle g(x) \rangle = \begin{cases} g(x), & \text{если } g(x) \leq 0, \\ 0, & \text{если } g(x) > 0. \end{cases}$$

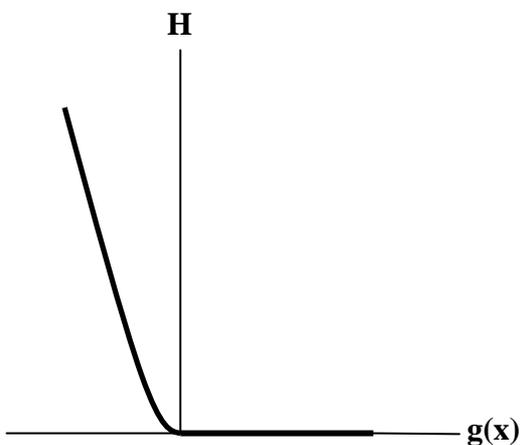


Рис. 11.7. Штраф типа квадрата срезки

Отметим, что этот штраф – внешний штраф, и стационарные точки функции $P(x, R)$ могут оказаться недопустимыми. С другой стороны, недопустимые точки не создают в данном случае дополнительных сложностей по сравнению с допустимыми. Различие между ними состоит лишь в том, что в допустимых и граничных точках $H = 0$.

Достоинства: функция $P(x, R)$ определена и непрерывна всюду. Вычисления проводятся с положительными малыми R ; после решения очередной задачи безусловной минимизации R увеличивается.

11.3. Примеры использования штрафов

Пример 11.1. С использованием квадратичного штрафа найти условный экстремум ЗО

$$f(x) = (x_1 - 4)^2 + (x_2 - 4)^2 \rightarrow \min$$

$$h(x) = x_1 + x_2 - 5 = 0$$

Решение. Используем квадратичный штраф. Введем ШФ вида

$$P(x, R) = (x_1 - 4)^2 + (x_2 - 4)^2 + R \cdot (x_1 + x_2 - 5)^2$$

В табл. 11.1 приведены координаты стационарных точек ШФ $P(x, R)$:

Таблица 11.1

| R | $x_1 = x_2$ | $P(x, R)$ | $f(x)$ |
|----------|-------------|-----------|--------|
| 0 | 4.0 | 0 | 0 |
| 0.1 | 3.75 | 0.75 | 0.2 |
| 1 | 3.0 | 3.0 | 2 |
| 10 | 2.57 | 4.29 | 4 |
| 100 | 2.51 | 4.447 | 4.44 |
| ∞ | 2.5 | 4.5 | 4.5 |

На рис. 11.8 изображены линии равных значений (линии уровня) ЦФ и вычисленный условный экстремум, находящийся на прямой, соответствующей ограничению-равенству $h(x)$. Линии уровня соответствуют следующим значениям целевой функции:

$A = 0.01$; $B = 3.0$; $C = 6.0$; $D = 9.0$; $J = 27$.

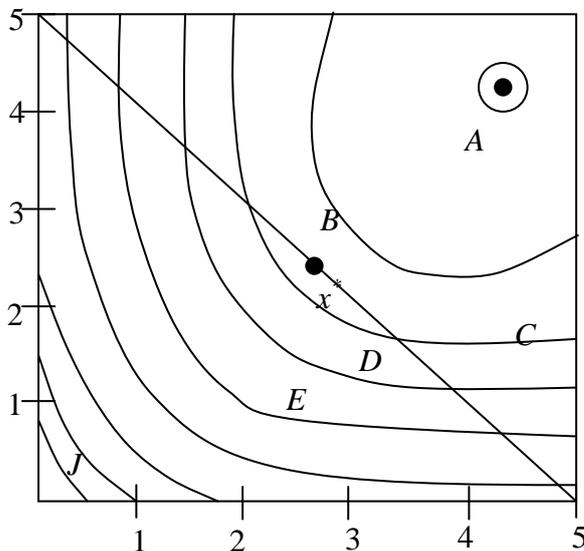


Рис. 11.8

Обсудим данный пример подробнее. Рассмотрим стационарную точку функции $P(x,1)$. Эта точка уже довольно близка к x^* , а именно $x = (3,3)$ и $f(x) = 2$, $h(x) = 1$. Во многих прикладных задачах такое приближенное решение может считаться вполне удовлетворительным, однако в общем случае величина невязки $h(x)$, как правило, неудовлетворительна. Возникает вопрос, как изменить ШФ, чтобы ее стационарная точка приблизилась к x^* .

Заметим, что в точно вычисленной стационарной точке функции $P(x,1)$ значение ЦФ $f(x^t) < P(x^t, 1)$ (оптимального) из-за нарушения условия $h(x) = 0$. При этом соотношение между значениями $f(x)$ и $P(x, R)$ связано с величиной R , в данном случае $R = 1$. По существу, R играет роль весового коэффициента, определяющего относительную значимость ограничения и ЦФ. Ясно, что при увеличении R стационарная точка функции $P(x, R)$ будет смещаться в сторону более точного выполнения ограничения $h(x) = 0$ и тем самым приближаться к x^* .

Если бы можно было положить $R = \infty$, то точное решение соответствующей подзадачи на безусловный минимум совпало бы с решением исходной задачи. Но выбрать $R = \infty$ (или очень большим) нельзя, т. к. при этом линии уровня ШФ становятся слишком вытянутыми – имеем овражную функцию. При этом сложность безусловной минимизации резко возрастает.

Предпочтительнее стратегия многоэтапного последовательного приближения к решению:

1. процесс начинают с относительно малого R , например, $R = 0.01$ и находят точку минимума функции $P(x, 0.01)$;
2. увеличивают R ($R = 0.1$) и находят минимум $P(x, 0.1)$, используя ранее полученную точку в качестве начальной;
3. R увеличивают до тех пор, пока не окажется, что элементы итерационных последовательностей $x^t, f(x^t), P(x^t, R)$ изменяются от шага к шагу достаточно мало. Однако, при этом увеличивается число итераций и время решения задачи. В этом основная трудность МШФ.

Пример 11.2:

$$f(x) = (x_1 - 4)^2 + (x_2 - 4)^2 \rightarrow \min$$

$$g(x) = 5 - x_1 - x_2 \geq 0$$

т. е. ограничение – равенство в примере 11.1 заменяется на неравенство. При этом решение x^* задачи останется прежним.

Используем штраф типа квадрата срезки.

$$P(x, R) = f(x) + R \langle g(x) \rangle^2,$$

$$\text{где } \langle \alpha \rangle = \begin{cases} \alpha, & \alpha \leq 0 \\ 0, & \alpha > 0 \end{cases}$$

Для нашего случая:

$$P(x, R) = (x_1 - 4)^2 + (x_2 - 4)^2 + R \langle 5 - x_1 - x_2 \rangle^2$$

В табл. 11.2 приведены координаты стационарных точек ШФ $P(x, R)$.

Таблица 11.2

| R | $x_1 = x_2$ | $f(x)$ | $g(x)$ | $H(x)$ | $P(x, R)$ |
|------|-------------|--------|--------|--------|-----------|
| 0 | 4 | 0 | -3 | 0 | 0 |
| 0.1 | 3.75 | 0.125 | -2.5 | 0.625 | 0.75 |
| 1 | 3.0 | 2 | -1 | 1.0 | 3.0 |
| 10 | 2.571 | 4.081 | -0.142 | 0.201 | 4.283 |
| 100 | 2.507 | 4.445 | -0.014 | 0.0196 | 4.474 |
| 1000 | 2.501 | 4.495 | -0.001 | 0.002 | 4.498 |

Таким образом, стационарная точка функции $P(x, 1)$ есть точка $\tilde{x} = (3, 3)$, причем $f(\tilde{x}) = 2$, $g(\tilde{x}) = -1$, $P(\tilde{x}, 1) = 3$. При изменении R от 0 до ∞ стационарная точка функции $P(x, R)$ перемещается вдоль прямой, соединяющей точку (4, 4) - безусловный минимум $f(x)$ с точкой (2.5, 2.5) ее условного минимума. При любом (конечном) R соответствующая стационарная точка недопустима, т. е. мы имеем дело с методом внешней точки.

Пример 11.3. Рассмотрим применительно к этой же задаче логарифмический штраф (барьерная функция).

$$P(x, R) = f(x) - R \ln(g(x))$$

В данном случае:

$$P(x, R) = (x_1 - 4)^2 + (x_2 - 4)^2 - R \ln(5 - x_1 - x_2).$$

Метод оказывается сходящимся в предположении, что безусловная минимизация проводится точно.

Таблица 11.3. Стационарные точки логарифмического штрафа

| R | $x_1 = x_2$ | $f(x)$ | $g(x)$ | H | $P(x, R)$ |
|-------|-------------|--------|--------|---------|-----------|
| 100 | -1.8059 | 67.417 | 8.6118 | -215.31 | -147.896 |
| 10 | 1.5 | 12.5 | 2.0 | -6.93 | 5.5686 |
| 1 | 2.3486 | 5.4542 | 0.3028 | 1.1947 | 6.6489 |
| 0.1 | 2.4835 | 4.5995 | 0.0034 | 0.3411 | 4.9406 |
| 0.01 | 2.4983 | 4.51 | 0.0034 | 0.0568 | 4.5668 |
| 0.001 | 2.5 | 4.5 | 0.00 | 0.00 | 4.5 |

Пример 11.4. Штраф – обратная функция (барьерная функция)

$$\text{Имеем } P(x, R) = f(x) + R \frac{1}{g(x)}.$$

Таблица 11.4. Стационарные точки штрафной функции для штрафа типа обратной функции

| R | $x_1 = x_2$ | $f(x)$ | $g(x)$ | $R/g(x)$ | $P(x, R)$ |
|-----|-------------|--------|--------|----------|-----------|
| 100 | 0.5864 | 3.3053 | 3.8272 | 26.1288 | 49.4341 |
| 10 | 1.7540 | 10.089 | 1.492 | 6.7024 | 16.7914 |

| | | | | | |
|-------|--------|--------|--------|--------|--------|
| 1 | 2.234 | 6.2375 | 0.532 | 1.8797 | 8.1172 |
| 0.1 | 2.4113 | 5.0479 | 0.1774 | 0.5637 | 5.6116 |
| 0.01 | 2.4714 | 4.6732 | 0.0572 | 0.1748 | 4.848 |
| 0.001 | 2.4909 | 4.5548 | 0.0182 | 0.0549 | 4.6097 |
| 0 | 2.5 | 4.5 | 0.0000 | 0.0000 | 4.5 |

Рассмотренные примеры демонстрируют с одной стороны, принципиальную возможность использования методов ШФ, а с другой стороны – их общие недостатки, а именно: сходимость метода связана с овражностью ШФ и здесь традиционные методы безусловной минимизации могут оказаться неэффективными.

Вопросы для самопроверки

1. Общая схема штрафов.
2. Методы внутреннего и внешнего штрафов.
3. Квадратичный штраф.
4. Штраф бесконечный барьер.
5. Логарифмический штраф.
6. Штраф типа обратной функции.
7. Штраф типа квадрата срезки.
8. Как строится последовательность штрафных параметров при использовании квадратичного штрафа?
9. Как строится последовательность штрафных параметров при использовании штрафа типа квадрата срезки?
10. Как строится последовательность штрафных параметров при использовании логарифмического штрафа?
11. Как строится последовательность штрафных параметров при использовании штрафа типа обратной функции?

Тема 12. Квадратичное программирование

12.1. Задача квадратичного программирования

Задача квадратичного программирования (ЗКП) представляет собой задачу с линейными ограничениями и квадратичной ЦФ. Общая ЗКП имеет вид:

$$f(x) = \sum_{j=1}^n c_j x_j + \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j \rightarrow \min$$

при ограничениях

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

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

или в матричном виде

$$\left. \begin{aligned} f(x) &= c^T x + x^T Q x \\ Ax &= b \\ x &\geq 0 \end{aligned} \right\}, \quad (12.1)$$

где A – матрица ограничений ($m \times n$);

b – вектор ограничений ($m \times 1$);

c – вектор ($1 \times n$);

Q – квадратная матрица ($n \times n$);

x – вектор переменных ($1 \times n$).

Проиллюстрируем на примерах некоторые применение этой модели на конкретном примере.

12.2 Задача выбора портфеля ценных бумаг

Определение *оптимального портфеля ценных бумаг* представляет собой одну из важнейших задач, с которыми сталкиваются инвестиционные фирмы – банки, фонды, страховые компании и частные лица.

Под портфелем понимаются размеры вложений в различные виды ценных бумаг: обычные облигации, облигации краткосрочных государственных займов, банковские депозитные сертификаты, акции и др.

Рассмотрим финансовую операцию, заключающуюся в покупке рискованных ценных бумаг по известной цене и в продаже их в будущем по цене, заранее не известной. Предполагается, что инвестор в настоящий момент времени инвестирует некоторую сумму денег в ценные бумаги. Эти деньги будут инвестированы на определённый промежуток времени, который называют *периодом владения*. В конце этого периода инвестор продает ценные бумаги, которые были куплены в начале периода. Таким образом, в момент $t = 0$ инвестор должен принять решение о покупке ценных бумаг, которые будут находиться в его портфеле до момента $t = 1$. Такую задачу называют *задачей выбора инвестиционного портфеля*.

Эффективность рискованной ценной бумаги зависит от трех факторов: цены покупки, которая точно известна; промежуточных выплат за период владения (дивидендов), которые

точно не известны; цены продажи, которая неизвестна. Таким образом, финансовая операция, заключающаяся в покупке ценной бумаги с целью получения определенного дохода в будущем, является рискованной. Основная гипотеза, которая позволяет анализировать такую операцию, состоит в следующем: предполагаем, что каждое конкретное значение эффективности такой финансовой операции является реализацией случайной величины

$$R = \frac{C_1 - C_0 + D}{C_0},$$

где C_0 – цена покупки, C_1 – цена продажи, D – дивиденды, выплаченные за период владения.

Формируя портфель ценных бумаг, инвестор хотел бы максимизировать ожидаемую доходность портфеля при минимальном риске. Как правило, эти две цели противоречат друг другу. Принимая решение, инвестор стремится сделать так, чтобы эти две цели были сбалансированы.

Доходность портфеля также является случайной величиной:

$$R_p = \frac{W_1 - W_0}{W_0},$$

где W_0 – совокупная цена покупки ценных бумаг, входящих в портфель в момент $t = 0$, W_1 – совокупная рыночная цена ценных бумаг в момент $t = 1$ и совокупный денежный доход от этих ценных бумаг (дивиденды), который владелец получит за период владения от момента $t = 0$ до $t = 1$.

Любая случайная величина может характеризоваться двумя параметрами: ожидаемое или среднее значение (математическое ожидание) и стандартное отклонение (среднеквадратичное отклонение).

Согласно рассматриваемой модели, предполагается, что инвестор основывает свое решение по выбору портфеля только на этих двух параметрах. Следовательно, инвестор должен оценить ожидаемую доходность и стандартное отклонение каждого возможного портфеля. Затем он должен выбрать лучший из портфелей, основываясь на соотношении этих двух параметров. При этом ожидаемая доходность рассматривается как мера потенциального вознаграждения, связанная с конкретным портфелем, а стандартное отклонение – как мера риска, связанная с данным портфелем.

Итак, рассмотрим финансовую операцию, которая заключается в покупке ценных бумаг в момент $t = 0$ по известной цене и в продаже их в момент $t = 1$ по цене, заранее не известной. При этом инвестор может рассчитывать на получение промежуточных выплат. Обозначим $m = M(R)$ – ожидаемое значение эффективности ценной бумаги – математическое ожидание случайной величины R (это среднее по всем реализациям (значениям) случайной величины, вычисленное с учетом частоты их возможного появления), $V = M\{(R - m)^2\}$ – дисперсия или вариация случайной величины – мера отклонения в среднем случайной величины R от ее ожидаемого значения. Часто вместо дисперсии используют среднеквадратичное или стандартное отклонение $\sigma = \sqrt{V}$.

Ковариация $V_{12} = M\{(R_1 - m_1)(R_2 - m_2)\}$ характеризует статистическую взаимосвязь двух случайных величин R_1 и R_2 .

Риск вложений в конкретные ценные бумаги связан с неопределённостью будущих доходов и, следовательно, с неопределённостью эффективности данной операции. Чем больше стандартное отклонение, тем больше в среднем случайная величина может отклониться от своего ожидаемого значения, тем больше неопределённость и выше риск. С другой стороны, если $\sigma = 0$, то эффективность не отклоняется от своего ожидаемого

значения, она принимает определённые не случайные значения, и риск отсутствует. Таким образом, стандартное отклонение характеризует уровень риска, связанный с конкретной ценной бумагой, и принимается в качестве *меры риска*.

Предположим, что инвестор вкладывает деньги не в один вид ценных бумаг, а несколько. В этом случае говорят, что инвестор диверсифицирует свой портфель. Рассмотрим эффект такой диверсификации.

Пусть x_j ($j=1,2,\dots,n$) – доля общего вложения, приходящаяся на j -й вид ценных бумаг; n – количество видов ценных бумаг, которые инвестор включает в портфель. Очевидно, должно выполняться равенство

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

Пусть R_p – эффективность портфеля, R_j – эффективность j -й ценной бумаги. Тогда

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

Ожидаемая эффективность портфеля:

$$m_p = M\{R_p\} = \sum_{j=1}^n x_j M\{R_j\} = \sum_{j=1}^n x_j m_j,$$

где $m_j = M\{R_j\}$ – ожидаемая эффективность j -й ценной бумаги. Отклонение от ожидаемой эффективности

$$R_p - m_p = \sum_{j=1}^n x_j (R_j - m_j).$$

Дисперсия эффективности портфеля:

$$V_p = M\{(R_p - m_p)^2\} = \sum_{i=1}^n \sum_{j=1}^n x_i x_j M\{(R_i - m_i)(R_j - m_j)\} = \sum_{i=1}^n \sum_{j=1}^n V_{ij} x_i x_j,$$

где $V_{ij} = M\{(R_i - m_i)(R_j - m_j)\}$ – ковариация случайных эффективностей R_i и R_j i -го и j -го видов ценных бумаг. Нетрудно заметить, что

$$V_{jj} = M\{(R_j - m_j)^2\} = \sigma_j^2.$$

Величина V_p (или σ_p) характеризует неопределённость портфеля в целом и называется *риском портфеля*.

При этих предположениях можно сформулировать следующую оптимизационную задачу: определить доли вложений x_j , $j=1,2,\dots,n$ минимизирующие вариацию (риск) портфеля

$$V_p = \sum_{i=1}^n \sum_{j=1}^n V_{ij} x_i x_j,$$

при условии, что обеспечивается заданное значение m_p ожидаемой эффективности портфеля

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

Кроме того, должны быть выполнены дополнительные ограничения вида

$$\sum_{j=1}^n x_j = 1, \quad x_j \geq 0$$

при всех $j = 1, 2, \dots, n$.

В этой модели целевая функция – квадратичная, ограничения – линейные.

12.2. Условие Куна-Таккера для ЗКП

Задача КП имеет вид

$$\begin{cases} f(x) = cx + x^T Qx \rightarrow \min, \\ g(x) = x \geq 0, \\ h(x) = Ax - b = 0. \end{cases} \quad (12.2)$$

Условия Куна-Таккера для ЗКП записываются следующим образом:

$$\begin{cases} c + x^T (Q + Q^T) - \mu - \lambda^T A = 0, \\ Ax = b, \quad x \geq 0, \\ \mu x = 0, \quad \mu \geq 0, \end{cases} \quad (12.3)$$

где λ – неограниченная по знаку переменная.

Пример 1.

$$\begin{aligned} f(x) &= -6x_1 + 2x_1^2 - 2x_1x_2 + 2x_2^2 \rightarrow \min, \\ x_1 + x_2 &= 2; \quad x_1, x_2 \geq 0. \end{aligned}$$

Для этой задачи:

$$c = (-6; 0); \quad Q = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}; \quad A = (1; 1); \quad b = (2);$$

$$Q = H_f \cdot \frac{1}{2} = \frac{1}{2} \nabla^2 f;$$

условия Куна-Таккера задаются следующим образом:

$$(-6; 0) + (x_1; x_2) \cdot \begin{pmatrix} 4 & -2 \\ -2 & 4 \end{pmatrix} - (\mu_1, \mu_2) - \lambda_1 \cdot (1; 1) = 0$$

$$\text{или} \quad -6 + 4x_1 - 2x_2 - \mu_1 - \lambda_1 = 0;$$

$$-2x_1 + 4x_2 - \mu_2 - \lambda_1 = 0;$$

$$x_1 + x_2 = 2 \quad x_1, x_2 \geq 0;$$

$$\mu_1 x_1 = 0; \quad \mu_2 x_2 = 0,$$

$\mu_1, \mu_2 \geq 0$; величина λ_1 неограничена по знаку.

Поскольку ограничение $h(x)$ – линейное, то применима теорема Куна-Таккера. Таким образом, условие Куна-Таккера является необходимым условием оптимальности решения ЗКП.

В случае положительной определенности или полуопределенности матрицы Q целевая функция является выпуклой и условия Куна-Таккера являются также и достаточными условиями оптимальности решения ЗКП. Т.о., в случае $H_f \geq 0$ достаточно найти точку x , удовлетворяющую условиям Куна-Таккера.

12.3 Решение ЗКП методом симплексного преобразования коэффициентов уравнений

Эту задачу можно решить гораздо проще, с помощью симплексного преобразования таблицы коэффициентов уравнений; так мы ищем начальное базисное решение. Отличием здесь является только то, что μ_i и x_i не могут быть одновременно базисными, поэтому мы делаем базисными только x_i и λ_j . При этом необходимо учесть, что λ_j могут быть любого знака, в то время как $x_j \geq 0$.

Для нашего примера имеем следующие таблицы.

Таблица 12.1

| x_1 | x_2 | μ_1 | μ_2 | λ | |
|-------|----------|---------|---------|-----------|---|
| 4 | -2 | -1 | 0 | -1 | 6 |
| -2 | 4 | 0 | -1 | -1 | 0 |
| 1 | 1 | 0 | 0 | 0 | 2 |

Таблица 12.2

| x_1 | x_2 | μ_1 | μ_2 | λ | |
|------------|-------|---------|---------|-----------|---|
| 3 | 0 | -1 | -1/2 | -3/2 | 6 |
| -1/2 | 1 | 0 | -1/4 | -1/4 | 0 |
| 3/2 | 0 | 0 | 1/4 | 1/4 | 2 |

Таблица 12.3

| x_1 | x_2 | μ_1 | μ_2 | λ | |
|-------|-------|---------|---------|-----------|-----|
| 0 | 0 | -1 | -1 | -2 | 2 |
| 0 | 1 | 0 | -1/6 | -1/6 | 2/3 |
| 1 | 0 | 0 | 1/6 | 1/6 | 4/3 |

Здесь мы получили базисные переменные x_1, x_2 . Теперь введем в базис переменную λ , не обращая внимания на ее знак.

Таблица 12.4

| x_1 | x_2 | μ_1 | μ_2 | λ | |
|-------|-------|---------|---------|-----------|-----|
| 0 | 0 | 1/2 | 1/2 | 1 | -1 |
| 0 | 1 | 1/12 | -1/12 | 0 | 1/2 |
| 1 | 0 | -1/12 | 1/12 | 0 | 3/2 |

Итак, получили базисные переменные:

$$x_1 = 3/2; x_2 = 1/2; \lambda = -1; \mu_1 = 0; \mu_2 = 0.$$

Заметим, в уравнении (12.3) мы взяли *минус* в слагаемом с λ . Если бы мы взяли *плюс*, то в таблице вместо (-1) был бы (+1) в первой и второй строке. В конечном решении мы бы получили $\lambda = +1$.

12.4 Метод решения ЗКП с помощью искусственного базиса

Одним из методов решения ЗКП может быть метод решения системы (3) как задачи ЛП с использованием искусственного базиса и первой операции симплекс-метода. При этом условие (4) неявно учитывается в процедуре симплекс-метода при помощи правила ограниченного ввода в базис. Это правило запрещает вводить в базис небазисную переменную μ_j , если значение соответствующей базисной переменной x_j положительное.

Введем искусственные переменные x_3, x_4, x_5 . В результате получим следующую целевую функцию:

$$z = x_3 + x_4 + x_5 = 8 - 3x_1 - 3x_2 + \mu_1 + \mu_2 + 2\lambda.$$

В табл. 12.5 – 12.8 жирным курсивом выделены опорные элементы.

Таблица 12.5

| | x_1 | x_2 | μ_1 | μ_2 | λ | |
|-------|-------|----------|---------|---------|-----------|----|
| x_3 | 4 | -2 | -1 | 0 | -1 | 6 |
| x_4 | -2 | 4 | 0 | -1 | -1 | 0 |
| x_5 | 1 | 1 | 0 | 0 | 0 | 2 |
| | -3 | -3 | 1 | 1 | 2 | -8 |

Таблица 12.6

| | x_1 | x_4 | μ_1 | μ_2 | λ | |
|-------|------------|------------|---------|---------|-----------|----|
| x_3 | 3 | 1/2 | -1 | -1/2 | -3/2 | 6 |
| x_2 | -1/2 | 1/4 | 0 | -1/4 | -1/4 | 0 |
| x_5 | 3/2 | -1/4 | 0 | 1/4 | 1/4 | 2 |
| | -9/2 | 3/4 | 1 | 1/4 | 5/4 | -8 |

Таблица 12.7

| | x_5 | x_4 | μ_1 | μ_2 | λ | |
|-------|-------|------------|---------|---------|-----------|-----|
| x_3 | -2 | 1 | -1 | -1 | -2 | 2 |
| x_2 | 1/3 | 1/6 | 0 | -1/6 | -1/6 | 2/3 |
| x_1 | 2/3 | -1/6 | 0 | 1/6 | 1/6 | 4/3 |
| | 3 | 0 | 1 | 1 | 2 | -2 |

Таблица 12.8

| | x_5 | x_4 | μ_1 | μ_2 | x_3 | |
|-----------|-------|-------|---------|---------|-------|-----|
| λ | 1 | -1/2 | 1/2 | 1/2 | -1/2 | -1 |
| x_2 | | | 1/12 | -1/12 | -1/12 | 1/2 |
| x_1 | | | -1/12 | 1/12 | 1/12 | 3/2 |
| | 1 | 1 | 0 | 0 | 1 | 0 |

К сожалению, в случае *положительно полуопределенной матрицы* Q (т. е. матрица Гессе $H_f(x) \geq 0$) этот метод может расходиться.

Вопросы для самопроверки

1. Запишите модель задачи квадратичного программирования.
2. Запишите условие Куна-Таккера для задачи квадратичного программирования.
3. Поясните суть метода искусственного базиса решения задачи квадратичного программирования.
4. Поясните суть метода симплексного преобразования таблицы коэффициентов уравнений решения задачи квадратичного программирования.

Тема 13. Модели динамического программирования

13.1. Общая постановка задачи динамического программирования

Динамическое программирование (ДП) — метод оптимизации, приспособленный к операциям, в которых процесс принятия решения может быть разбит на этапы (шаги). Такие операции называются *многошаговыми*. Начало развития ДП относится к 50-м годам XX в. Оно связано с именем американского математика Р.Беллмана¹.

Если модели линейного программирования можно использовать в экономике для принятия крупномасштабных плановых решений в сложных ситуациях, то модели ДП применяются при решении задач значительно меньшего масштаба, например, при разработке правил управления запасами, устанавливающими момент пополнения запасов и размер пополняющего заказа; при разработке принципов календарного планирования производства и выравнивания занятости в условиях колеблющегося спроса на продукцию; при распределении дефицитных капитальных вложений между возможными новыми направлениями их использования; при составлении календарных планов текущего и капитального ремонта сложного оборудования и его замены; при разработке долгосрочных правил замены выбывающих из эксплуатации основных фондов и т. п.

В реально функционирующих больших экономических системах еженедельно требуется принимать микроэкономические решения. Модели ДП ценны тем, что позволяют на основе стандартного подхода при минимальном вмешательстве человека принимать такие решения. И если каждое взятое в отдельности такое решение малосущественно, то в совокупности эти решения могут оказать большое влияние на прибыль.

Приведем *общую постановку задачи ДП*. Рассматривается управляемый процесс, например, экономический процесс распределения средств между предприятиями. Процесс описывается n -м вектором $x = (x_1, x_2, \dots, x_n)$, т.е. точкой в пространстве R^n , которое называют фазовым пространством. В результате управления система (объект управления) S переводится из начального состояния $x^{(0)}$ в состояние $x^{(N)}$. Предположим, что управление можно разбить на N шагов, т.е. решение принимается последовательно на каждом шаге, а управление, переводящее систему S из начального состояния в конечное, представляет собой совокупность N пошаговых управлений.

Обозначим через $u^{(k)}$ управление на k -м шаге ($k = 1, \dots, N$). Переменные $u^{(k)}$ удовлетворяют некоторым ограничениям и в этом смысле называются *допустимыми* ($u^{(k)}$ может быть числом, точкой в m -мерном пространстве, качественным признаком). В общем случае u представляет собой m -мерный вектор управления, т.е. $u \in R^m$.

В результате действия управления на каждом шаге получаем последовательность состояний $x^{(0)}, x^{(1)}, \dots, x^{(N)}$, которую изобразим кружками (рис. 13.1).

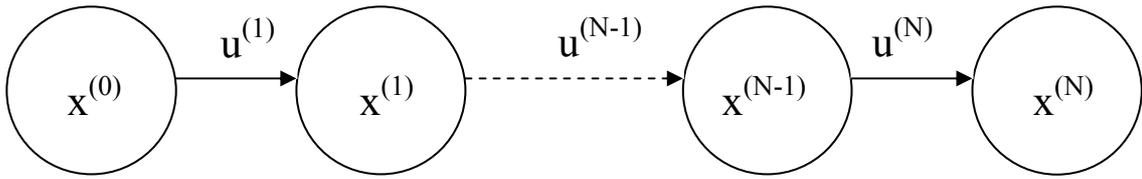


Рис. 13.1.

Показатель эффективности рассматриваемого управляемого процесса характеризуется целевой функцией, которая зависит от начального состояния и управления:

$$J = J(x^{(0)}, u). \quad (13.1)$$

Сделаем несколько предположений.

1. Состояние $x^{(k)}$ системы в конце k -го шага зависит только от предшествующего состояния $x^{(k-1)}$ и управления k -м шаге $u^{(k)}$ (и не зависит от предшествующих состояний и управлений). Это требование называется "отсутствием последействия". Сформулированное положение записывается в виде уравнений

$$x^{(k)} = f^{(k)}(x^{(k-1)}, u^{(k)}), \quad k = 1, 2, \dots, N, \quad (13.2)$$

которые называются *уравнениями состояний*.

2. Целевая функция (1) является аддитивной от показателя эффективности каждого шага. Обозначим показатель эффективности k -го шага через

$$J_k = J_k(x^{(k-1)}, u^{(k)}), \quad k = 1, 2, \dots, N, \quad (13.3)$$

тогда

$$J(x, u) = \sum_{k=1}^N J_k = \sum_{k=1}^N J_k(x^{(k-1)}, u^{(k)}). \quad (13.4)$$

Задача пошаговой оптимизации (задача ДП) формулируется так: *определить такое допустимое управление u , переводящее систему S из состояния $x^{(0)}$ в состояние $x^{(N)}$, при котором целевая функция (4) принимает наибольшее (наименьшее) значение.*

Предположим далее, что на фазовую траекторию и выбор управлений наложены ограничения

$$x^{(k)} \in X_k, \quad k = 1, \dots, N-1, \quad (13.5)$$

$$u^{(k)} \in U_k(x^{(k-1)}), \quad k = 1, \dots, N, \quad (13.6)$$

где X_k и $U_k(x^{(k-1)})$ - заданные множества в пространствах R^n и R^m соответственно, причем множество $U_k(x^{(k-1)})$ зависит, вообще говоря, от предыдущего состояния $x^{(k-1)}$ k -го шага.

Ограничения на начальное и конечное состояния процесса

$$x^{(0)} \in X_0, x^{(N)} \in X_N$$

называются *начальными и конечными условиями*.

Пусть $u = \{u^{(1)}, u^{(2)}, \dots, u^{(N)}\}$ - управление процессом, удовлетворяющее

ограничениям (13.6) и переводящее его из некоторого начального состояния $x^{(0)} \in X_0$ в некоторое конечное состояние $x^{(N)} \in X_N$ в соответствии с уравнением (2) с учетом ограничений (13.5). Обозначим множество всех таких управлений буквой U .

Многошаговая задача оптимизации формулируется следующим образом: *среди всех управлений $u \in U$ выбрать такое $\bar{u} = \{\bar{u}^{(1)}, \dots, \bar{u}^{(N)}\}$, для которого целевая функция (13.3) принимает минимальное или максимальное (в зависимости от смысла задачи) значение.*

Управление \bar{u} и соответствующая ему фазовая траектория \bar{x} называются *оптимальными*.

Условие многошаговой задачи оптимизации будем записывать следующим образом:

$$J(x, u) = \sum_{k=1}^N J_k = J_k(x^{(k-1)}, u^{(k)}) \rightarrow \underset{U}{extr}, \quad (13.7)$$

$$x^{(k)} = f^{(k)}(x^{(k-1)}, u^{(k)}), \quad k = 1, \dots, N, \quad (13.8)$$

$$u^{(k)} \in U_k(x^{(k-1)}), \quad k = 1, \dots, N, \quad (13.9)$$

$$x^{(0)} \in X_0, x^{(N)} \in X_N.$$

Выделим особенности модели ДП:

1. *Задача оптимизации интерпретируется как N -шаговый процесс управления,*

2. *Целевая функция равна сумме целевых функций каждого шага.*

3. *Выбор управления на k -м шаге зависит только от состояния системы к этому шагу, не влияет на предшествующие шаги (нет обратной связи).*

4. *Состояние $x^{(k)}$ после k -го шага управления зависит только от предшествующего состояния $x^{(k-1)}$ и управления $u^{(k)}$ (отсутствие последствий).*

5. *На каждом шаге управление $u^{(k)}$ зависит от конечного числа управляющих переменных, а состояние $x^{(k)}$ — от конечного числа параметров (смысл замечания 5 станет ясным из рассмотренных ниже примеров).*

Существуют различные способы решения подобных задач, применяемые в зависимости от вида функций, ограничений, размерности и т. п. Рассмотрим вычислительную схему ДП, которая окажется безразличной к способам задания функций и ограничений. Вычислительная схема связана с принципом оптимальности и использует рекуррентные соотношения.

13.2. Принцип оптимальности и уравнения Беллмана

Принцип оптимальности впервые был сформулирован Р. Беллманом в 1953 г. *Каково бы ни было состояние x системы в результате какого-либо числа шагов, на ближайшем шаге нужно выбирать управление так, чтобы оно в совокупности с оптимальным управлением на всех последующих шагах приводило к оптимальному выигрышу на всех оставшихся шагах,*

включая данный. Беллманом четко были сформулированы и условия, при которых принцип верен. Основное требование — процесс управления должен быть без обратной связи, т.е. управление на данном шаге не должно оказывать влияния на предшествующие шаги.

Принцип оптимальности утверждает, что для любого процесса без обратной связи оптимальное управление таково, что оно является оптимальным для любого подпроцесса по отношению к исходному состоянию этого подпроцесса. Поэтому решение на каждом шаге оказывается наилучшим с точки зрения управления в целом. Если изобразить геометрически оптимальную траекторию в виде ломаной линии, то любая часть этой ломаной будет являться оптимальной траекторией относительно начала и конца,

Уравнения Беллмана. Вместо исходной задачи ДП (см. разд. 6.1) с фиксированным числом шагов N и начальным состоянием $x^{(0)}$ рассмотрим последовательность задач, полагая последовательно $N=1, 2, \dots$ при различных x — одношаговую, двухшаговую и т.д., — используя принцип оптимальности.

Введем ряд новых обозначений. Обозначения в ДП несут большую информационную нагрузку, поэтому очень важно их четко усвоить.

На каждом шаге любого состояния системы $x^{(k-1)}$ решение $u^{(k)}$ нужно выбирать "с оглядкой", так как этот выбор влияет на последующее состояние $x^{(k)}$ и дальнейший процесс управления, зависящий от $x^{(k)}$. Это следует из принципа оптимальности.

Но есть один шаг, последний, который можно для любого состояния $x^{(N-1)}$ планировать локально-оптимально, исходя только из соображений этого шага.

Рассмотрим N -й шаг: $x^{(N-1)}$ — состояние системы к началу N -го шага, $x^{(N)}$ — конечное состояние, $u^{(N)}$ — управление на N -м шаге, а $J_k(x^{(N-1)}, u^{(N)})$ — целевая функция (выигрыш) N -го шага.

Согласно принципу оптимальности, $u^{(N)}$ нужно выбирать так, чтобы для любых состояний $x^{(N-1)}$ получить максимум (минимум) целевой функции на этом шаге.

Обозначим через $B_N(x^{(N-1)})$ максимум целевой функции — показателя эффективности N -го шага при условии, что к началу последнего шага система S была в произвольном состоянии $x^{(N-1)}$, а на последнем шаге управление было оптимальным.

$B_N(x^{(N-1)})$ называется *условным максимумом целевой функции на N -м шаге*. Очевидно, что

$$B_N(x^{(N-1)}) = \max_{u^{(N)} \in U_N(x^{(N-1)})} \left(J_N(x^{(N-1)}, u^{(N)}) \right). \quad (13.10)$$

Максимизация ведется по всем допустимым управлениям U_N .

Решение $u^{(N)}$, при котором достигается $B_N(x^{(N-1)})$, также зависит от $x^{(N-1)}$ и называется *условным оптимальным управлением на N -м шаге*. Оно обозначается через $u^{-(N)}(x^{(N-1)})$.

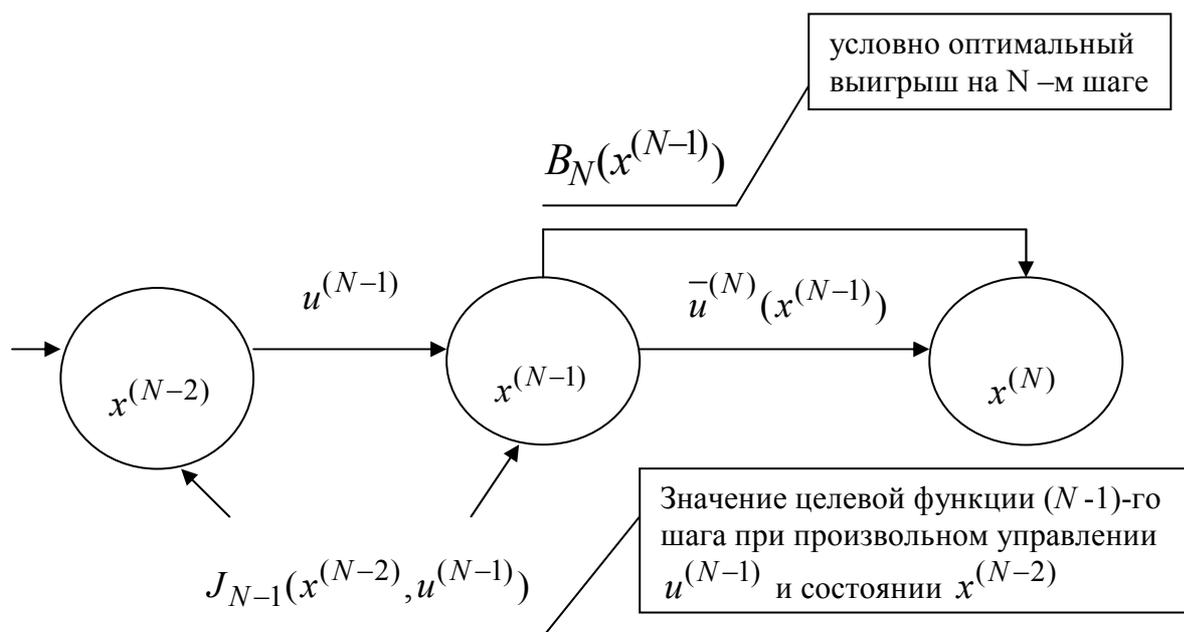


Рис. 13.2

Решив задачу локальной оптимизации (13.10), найдем для всех возможных состояний $x^{(N-1)}$ две функции: $B_N(x^{(N-1)})$ и $u^{(N)}(x^{(N-1)})$.

Рассмотрим теперь двухшаговую задачу: присоединим к N -му шагу $(N-1)$ -й шаг (рис. 13.2).

Для любых состояний $x^{(N-2)}$, произвольных управлений $u^{(N-1)}$ и оптимальном управлении на N -м шаге значение целевой функции на двух последних шагах равно:

$$J_{N-1}(x^{(N-2)}, u^{(N-1)}) + B_N(x^{(N-1)}). \quad (13.11)$$

Согласно принципу оптимальности для любых $x^{(N-2)}$ решение нужно выбирать так, чтобы оно вместе с оптимальным управлением на последнем (N -м) шаге приводило бы к максимуму целевой функции на двух последних шагах. Следовательно, нужно найти максимум выражения (13.11) по всем допустимым управлениям $u^{(N-1)}$. Максимум этой суммы зависит от $x^{(N-2)}$ и обозначается через $B_{N-1}(x^{(N-2)})$ и называется *условным максимумом целевой функции при оптимальном управлении на двух последних шагах*. Соответствующее управление $u^{(N-1)}$ на $(N-1)$ -м шаге обозначается через $u^{(N-1)}(x^{(N-2)})$ и называется *условным оптимальным управлением на $(N-1)$ -м шаге*.

$$B_{N-1}(x^{(N-2)}) = \max_{U_{N-1}} \left\{ J_{N-1}(x^{(N-2)}, u^{(N-1)}) + B_N(x^{(N-1)}) \right\}. \quad (13.12)$$

Следует обратить внимание на то, что выражение, стоящее в фигурных скобках (13.12), зависит только от $x^{(N-2)}$ и $u^{(N-1)}$, так как $x^{(N-1)}$ можно найти из уравнения состояний (13.2) при $k = N-1$

$$x^{(N-1)} = f^{(N-1)}(x^{(N-2)}, u^{(N-1)})$$

и подставить вместо $x^{(N-1)}$ в функцию $B_N(x^{(N-1)})$.

В результате максимизации только по одной переменной $u^{(N-1)}$ согласно уравнению (13.12) вновь получаются две функции:

$$B_{N-1}(x^{(N-2)}) \text{ и } \bar{u}^{(N-1)}(x^{(N-2)}).$$

Далее рассматривается трехшаговая задача: к двум последним шагам присоединяется $(N-2)$ -й шаг и т.д.

Обозначим через $B_k(x^{(k-1)})$ условный максимум целевой функции, полученный при оптимальном управлении на $N-(k-1)$ шагах, начиная с k -го до конца, при условии, что к началу k -го шага система находилась в состоянии $x^{(k-1)}$. Фактически эта функция равна

$$B_k(x^{(k-1)}) = \max_{\{U_k, \dots, U_N\}} \sum_{i=k}^N J_i(x^{(i-1)}, u^{(i)}).$$

Тогда

$$B_{k+1}(x^{(k)}) = \max_{\{U_k, \dots, U_N\}} \sum_{i=k+1}^N J_i(x^{(i-1)}, u^{(i)}).$$

Целевая функция на $(N-k)$ последних шагах (рис. 3) при произвольном управлении $u^{(k)}$ на k -м шаге и оптимальном управлении на последующих $(N-k)$ шагах равна

$$J_k(x^{(k-1)}, u^{(k)}) + B_{k+1}(x^{(k)}).$$

Согласно принципу оптимальности, $u^{(k)}$ выбирается из условия максимума этой суммы, т.е.

$$B_k(x^{(k-1)}) = \max_{U_k} \{J_k(x^{(k-1)}, u^{(k)}) + B_{k+1}(x^{(k)})\}, \quad k = N-1, N-2, \dots, 2, 1. \quad (13.13)$$

Управление $u^{(k)}$ на k -м шаге, при котором достигается максимум в (13.13), обозначается через $\bar{u}^{(k)}(x^{(k-1)})$ и называется *условным оптимальным управлением на k -м шаге* (также как и в (13.12) в правую часть уравнения (13.13) вместо $x^{(k)}$ следует подставить выражение $x^{(k)} = f^{(k)}(x^{(k-1)}, u^{(k)})$, найденное из уравнения состояния).

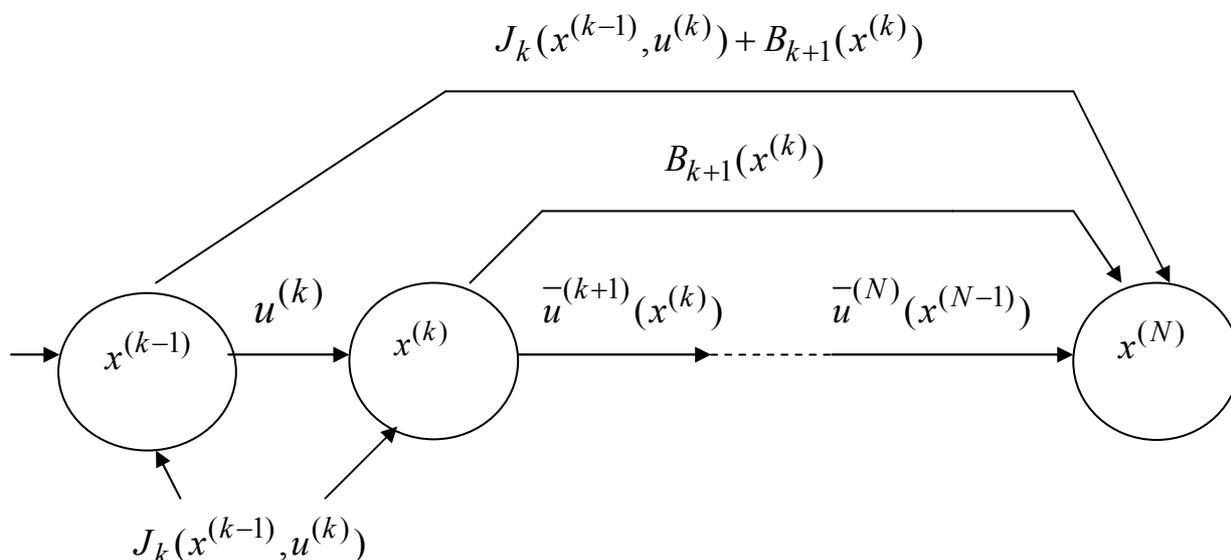


Рис. 13.3.

Функция $B_k(x^{(k-1)})$ называется *функцией Беллмана* последних $N - (k - 1)$ шагов.

Уравнения (13.13) называют *уравнениями Беллмана*. Это рекуррентные соотношения, позволяющие найти предыдущее значение функции, зная последующие. Если из (10) найти $B_N(x^{(N-1)})$, то при $k = N - 1$ из (13) можно определить, решив задачу максимизации для всех возможных значений $x^{(N-2)}$, выражение для $B_{N-1}(x^{(N-2)})$ и соответствующее управление $u^{(N-1)}(x^{(N-2)})$. Далее, зная $B_{N-1}(x^{(N-2)})$, находим $B_{N-2}(x^{(N-3)})$, используя (13.13) и уравнение состояния (13.2).

Процесс решения уравнений (13.10) и (13.13) называется *условной оптимизацией*.

В результате условной оптимизации получаются две последовательности:

1) условные максимумы целевой функции на последнем, двух последних, на ... N шагах

$$B_N(x^{(N-1)}), B_{N-1}(x^{(N-2)}), \dots, B_2(x^{(1)}), B_1(x^{(0)});$$

2) условные оптимальные управления на N -м, $(N - 1)$ -м, ..., 1-м шагах

$$\bar{u}^{(N)}(x^{(N-1)}), \bar{u}^{(N-1)}(x^{(N-2)}), \dots, \bar{u}^{(2)}(x^{(1)}), \bar{u}^{(1)}(x^{(0)}).$$

Используя эти последовательности, можно найти решение задачи ДП при данных N и $x^{(0)}$. По определению $B_1(x^{(0)})$ — условный максимум целевой функции за N шагов при условии, что к началу 1-го шага система была в состоянии $x^{(0)}$, т.е.

$$B_{\max} = B_1(x^{(0)}). \quad (13.14)$$

Далее, используя последовательность условных оптимальных управлений и уравнения состояний (8), последовательно находим $\bar{u}^{(1)}, \bar{x}^{(1)}, \bar{u}^{(2)}, \bar{x}^{(2)}, \dots, \bar{u}^{(N)}, \bar{x}^{(N)}$ из следующих соотношений:

$$\begin{aligned} \bar{u}^{(1)} &= \bar{u}^{(1)}(\bar{x}^{(0)}), & \bar{x}^{(1)} &= f^{(1)}(\bar{x}^{(0)}, \bar{u}^{(1)}), \\ \bar{u}^{(2)} &= \bar{u}^{(2)}(\bar{x}^{(1)}), & \bar{x}^{(2)} &= f^{(2)}(\bar{x}^{(1)}, \bar{u}^{(2)}), \\ & \dots, & & \\ \bar{u}^{(N)} &= \bar{u}^{(N)}(\bar{x}^{(N-1)}), & \bar{x}^{(N)} &= f^{(N)}(\bar{x}^{(N-1)}, \bar{u}^{(N)}). \end{aligned} \quad (13.15)$$

Этот этап называется *безусловной оптимизацией*.

13.3. Задача о распределении средств между предприятиями

Планируется деятельность четырех промышленных предприятий (системы) на очередной год. Начальные средства: $x^{(0)} = 5$ усл.ед. Размеры вложения в каждое предприятие кратны 1 усл.е. Средства u , выделенные k -му предприятию ($k = 1, 2, 3, 4$), приносят в конце года прибыль $J_k(u)$. Функции $J_k(u)$ заданы таблично (табл.1). Принято считать, что:

- прибыль $J_k(u)$ не зависит от вложения средств в другие предприятия;
- прибыль от каждого предприятия выражается в одних условных единицах;
- суммарная прибыль равна сумме прибылей, полученных от каждого предприятия.

Определить, какое количество средств нужно выделить каждому предприятию, чтобы суммарная прибыль была наибольшей.

Таблица 13.1.

| u | 1 | 2 | 3 | 4 | 5 |
|----------|---|----|----|----|----|
| $J_1(u)$ | 8 | 10 | 11 | 12 | 18 |
| $J_2(u)$ | 6 | 9 | 11 | 13 | 15 |
| $J_3(u)$ | 3 | 4 | 7 | 11 | 18 |
| $J_4(u)$ | 4 | 6 | 8 | 13 | 16 |

Решение. Обозначим через $u^{(k)}$ количество средств, выделенных k -му предприятию. Суммарная прибыль равна

$$J = \sum_{k=1}^4 J_k(u^{(k)}). \quad (13.16)$$

Переменные u удовлетворяют ограничениям:

$$\sum_{k=1}^4 u^{(k)} = 5, \quad u^{(k)} \geq 0, \quad k = 1, 2, 3, 4. \quad (13.17)$$

Требуется найти переменные $u^{(1)}, u^{(2)}, u^{(3)}, u^{(4)}$, удовлетворяющие системе ограничений (13.17) и обращающие в максимум функцию (13.16).

Схема решения задачи методом ДП имеет следующий вид: процесс решения распределения средств $x^{(0)} = 5$ можно рассматривать как 4-шаговый, номер шага совпадает с номером предприятия; выбор переменных $u^{(1)}, u^{(2)}, u^{(3)}, u^{(4)}$ — управление соответственно на I, II, III, IV шагах. Конечное состояние процесса распределения $x^{(4)} = 0$, так как все средства должны быть вложены в производство. Схема распределения показана на рис. 4.

Уравнения состояний (2) в данной задаче имеют вид:

$$x^{(k)} = x^{(k-1)} - u^{(k)}, \quad k = 1, 2, 3, 4, \quad (13.18)$$

где $x^{(k)}$ — параметр состояния — количество средств, оставшихся после k -го шага, т.е. средства, которые остается распределить между оставшимися $4 - k$ предприятиями.

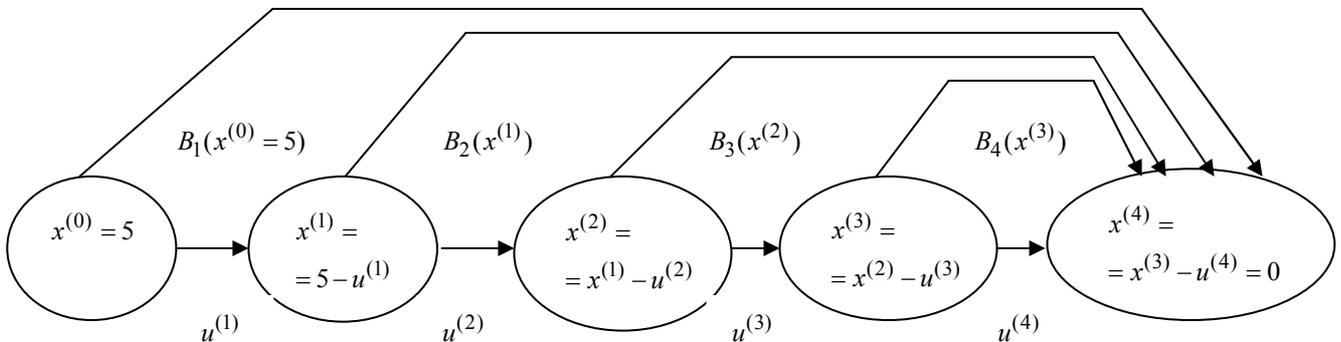


Рис. 13.4.

Введем в рассмотрение функцию $B_k(x^{(k-1)})$, ($k = 4, 3, 2, 1$) — условную оптимальную прибыль, полученную от 4-го предприятия ($k = 4$), 4-го и 3-го предприятий ($k = 3$), 4-го, 3-го и 2-го предприятий ($k = 2$) и всех четырех предприятий ($k = 1$). При этом

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|----------------------------|----------------------------|--|----|---|---|----|--------|--|----|----|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 1 | 1 0 | 0+4=4 3+0=3 | 4 | 0 | 0+4=4 6+0=6 | 6 | 1 | 0+6=6 8+0=8 | 8 | 1 |
| 2 | 0 1 2 | 2 1 0 | 0+6=6 3+4=7 4+0=4 | 7 | 1 | 0+7=7 6+4=10 9+0=9 | 10 | 1 | 0+10=10 8+6=14 10+0=10 | 14 | 1 |
| 3 | 0 1 2 3 | 3 2 1 0 | 0+8=8 3+6=9 4+4=8 7+0=7 | 9 | 1 | 0+9=9 6+7=13 9+4=13 11+0=11 | 13 | 1 2 | 0+13=13 8+10=18 10+6=16 11+0=11 | 18 | 1 |
| 4 | 0 1 2 3 4 | 4 3 2 1 0 | 0+13=13 3+8=11 4+6=10 7+4=11 11+0=11 | 13 | 0 | 0+13=13 6+9=15 9+7=16 11+4=15 13+0=13 | 16 | 2 | 0+16=16 8+13=21 10+10=20 11+6=17 12+0=12 | 21 | 1 |
| 5 | 0 1 2 3 4 5 | 5 4 3 2 1 0 | 0+16=16 3+13=16 4+8=12 7+6=13 11+4=15 18+0=18 | 18 | 5 | 0+18=18 6+13=19 9+9=18 11+7=18 13+4=17 15+0=15 | 19 | 1 | 0+19=19 8+16=24 10+13=23 11+10=21 12+6=18 18+0=18 | 24 | 1 |

I шаг. Условная оптимизация в соответствии с выражением (г) приведена в табл. 2 при $k = 1$ (здесь достаточно провести расчеты только для $x^{(0)} = 5$, так как начальное состояние задано и равно 5).

Дадим пояснение этим расчетам. Если $u^{(1)} = 0$, то $x^{(1)} = 5$, прибыль, полученная от четырех предприятий при условии, что $x^{(1)} = 5$ у.е. средств между оставшимися тремя предприятиями будут распределены оптимально, равна $J_1(0) + B_2(5) = 0 + 19 = 19$ (величина $B_2(5)$ взята из столбца 8 табл. 2 при $x^{(1)} = 5$). Если $u^{(1)} = 1$, то $x^{(1)} = 4$. Суммарная прибыль от четырех предприятий при условии, что $x^{(1)} = 4$ у.е. средств между оставшимися тремя предприятиями будут распределены оптимально, равна $J_1(1) + B_2(4) = 8 + 16 = 24$ ($J_1(1)$ взято из табл. 1, а $B_2(4)$ — из столбца 8 табл. 2). Аналогично при $u^{(1)} = 2$, $x^{(1)} = 3$, и $J_1(2) + B_2(3) = 10 + 13 = 23$. Продолжая этот процесс, мы получим:

$$\text{при } u^{(1)} = 3, x^{(1)} = 2 \text{ и } J_1(3) + B_2(2) = 11 + 10 = 21;$$

$$\text{при } u^{(1)} = 4, x^{(1)} = 1 \text{ и } J_1(4) + B_2(1) = 12 + 6 = 18;$$

$$\text{при } u^{(1)} = 5, x^{(1)} = 0 \text{ и } J_1(5) + B_2(0) = 18 + 0 = 18.$$

Сравнивая подчеркнутые числа, получим $B_{\max} = B_1(5) = 24$ при $u^{-(1)} = u^{-(1)}(5) = 1$.

Рассмотрим этап *безусловной оптимизации*. Используя уравнения (18), получим $x^{-(1)} = x^{-(0)} - u^{-(1)} = 5 - 1 = 4$. В табл. 2 в столбце 9 находим $u^{-(2)} = u^{-(2)}(4) = 2$. Продолжая этот процесс, имеем:

$$x^{-(2)} = x^{-(1)} - u^{-(2)} = 4 - 2 = 2, u^{-(3)} = u^{-(3)}(2) = 1 \text{ (см. столбец 6);}$$

$$x^{-(3)} = x^{-(2)} - u^{-(3)} = 2 - 1 = 1, \quad u^{-(4)} = u^{-(4)}(1) = 1, \quad \text{так как } \sum_{k=1}^5 u^{-(k)} = 5 \quad (\text{см. условие}$$

(17)).

Таким образом имеем: максимум суммарной прибыли равен 24 у.е. средств при условии, что 1-му предприятию выделено 1 у.е., 2-му предприятию — 2 у.е., 3-му предприятию — 1 у.е., 4-му предприятию — 1 у.е.

Замечание 1. Метод ДП безразличен к виду и способу задания функции прибыли $J_k(u)$.

Здесь они были заданы таблично, поэтому оптимальные $B_k(x)$ и $u^{-(k)}(x)$ принимали дискретные значения, представленные в табл. 13.2. Если бы $J_k(u)$ были заданы аналитически, то $B_k(x)$ и $u^{-(k)}(x)$ также были бы получены в аналитическом виде.

Замечание 2. В рассмотренных задачах состояния системы x и управления u были скалярными величинами, поэтому на каждом шаге у нас рассматривались одномерные задачи оптимизации. В общем случае, как мы отмечали ранее состояния системы $x \in R^n$ описываются n -мерным вектором, а управления — m -мерным вектором ($u \in R^m$). Поэтому с увеличением размерности векторов состояний и управлений объем вычислений резко возрастает, так как на каждом шаге возникает $n \cdot m$ -мерная задача оптимизации.

13.4. Задача об оптимальном распределении ресурсов между отраслями на N лет

Планируется деятельность двух отраслей производства на N лет. Начальные ресурсы $x^{(0)}$. Средства u , вложенные в I отрасль в начале года, дают в конце года прибыль $J_1(u)$ и возвращаются в размере $q_1(u) < u$; аналогично для II отрасли функция прибыли равна $J_2(u)$, а возврата — $q_2(u)$ ($q_2(u) < u$). В конце года все возвращенные средства заново перераспределяются между I и II отраслями, новые средства не поступают, прибыль в производство не вкладывается. (Последние условия определяют вид уравнений состояний; если поступают новые средства или часть прибыли вкладывается в производство, то это можно легко учесть, так как алгоритм метода ДП не изменится).

Требуется распределить имеющиеся средства $x^{(0)}$ между двумя отраслями производства на N лет так, чтобы суммарная прибыль от обеих отраслей за N лет оказалась максимальной.

Необходимо:

а) построить модель ДП для задачи и вычислительную схему;

б) решить задачу при условии, что $x^{(0)} = 10000$ у.е., $N = 4$, $J_1(u) = 0.6u$, $q_1(u) = 0.7u$, $J_2(u) = 0.5u$, $q_2(u) = 0.8u$.

Решение.

а) Процесс распределения средств между двумя отраслями производства разворачивается во времени, решения принимаются в начале каждого года, следовательно, осуществляется деление на шаги: номер шага — номер года. Управляемая система — две отрасли производства, а управление состоит в выделении средств каждой отрасли в очередном году. Параметры состояния к началу k -го года — $x^{(k-1)}$ ($k = 1, 2, \dots, N$) — количество средств, подлежащих распределению. Переменных

управления на каждом шаге две: $u_1^{(k)}$ — количество средств, выделенных I отрасли, и $u_2^{(k)}$ — II отрасли. Но так как все средства $x^{(k-1)}$ распределяются, то $u_2^{(k)} = x^{(k-1)} - u_1^{(k)}$, поэтому управление на k -м шаге зависит от одной переменной $u_1^{(k)}$, т.е. $U^{(k)}(u^{(k)}, x^{(k-1)} - u^{(k)})$.

Уравнения состояний

$$x^{(k)} = q_1(u^{(k)}) + q_2(x^{(k-1)} - u^{(k)}) \quad (13.19)$$

выражают остаток средств, возвращенных в конце k -го года.

Показатель эффективности k -го шага — прибыль, полученная в конце k -го года от обеих отраслей:

$$J_k = J_1(u^{(k)}) + J_2(x^{(k-1)} - u^{(k)}) \quad (13.20)$$

Суммарный показатель эффективности — целевая функция — прибыль за N лет:

$$J = \sum_{k=1}^N (J_1(u^{(k)}) + J_2(x^{(k-1)} - u^{(k)})). \quad (21)$$

Пусть $B_k(x^{(k-1)})$ — условная оптимальная прибыль за $N - k + 1$ лет, начиная с k -го года до N -го включительно, при условии что имеющиеся на начало k -го года средства $x^{(k-1)}$ в дальнейшем распределялись оптимально. Тогда оптимальная прибыль за N лет

$$J_{\max} = B_1(x^{(0)}).$$

Уравнения Беллмана приобретают вид:

$$B_N(x^{(N-1)}) = \max_{0 \leq u^{(N)} \leq x^{(N-1)}} \{J_1(u^{(N)}) + J_2(x^{(N-1)} - u^{(N)})\}, \quad (13.22)$$

$$B_k(x^{(k-1)}) = \max_{0 \leq u^{(k)} \leq x^{(k-1)}} \{J_1(u^{(k)}) + J_2(x^{(k-1)} - u^{(k)})\}, \quad (13.23)$$

$$k = N - 1, N - 2, \dots, 1$$

б) Выполним вычисления для конкретных данных.

Уравнение состояний (19) примет вид:

$$x^{(k)} = 0.7u^{(k)} + 0.8(x^{(k-1)} - u^{(k)}) = 0.8x^{(k-1)} - 0.1u^{(k)}. \quad (13.24)$$

Целевая функция k го шага (20)

$$J_k = 0.6u^{(k)} + 0.5(x^{(k-1)} - u^{(k)}) = 0.5x^{(k-1)} + 0.1u^{(k)}.$$

Целевая функция задачи

$$J = \sum_{k=1}^4 (0.5x^{(k-1)} + 0.1u^{(k)}). \quad (13.25)$$

Уравнения Беллмана

$$B_4(x^{(3)}) = \max_{0 \leq u^{(4)} \leq x^{(3)}} \{0.5x^{(3)} + 0.1u^{(4)}\}, \quad (13.26)$$

$$B_k(x^{(k-1)}) = \max_{0 \leq u^{(k)} \leq x^{(k-1)}} \{0.5x^{(k-1)} + 0.1u^{(k)} + B_{k+1}(x^{(k)})\}. \quad (13.27)$$

Проводим условную оптимизацию.

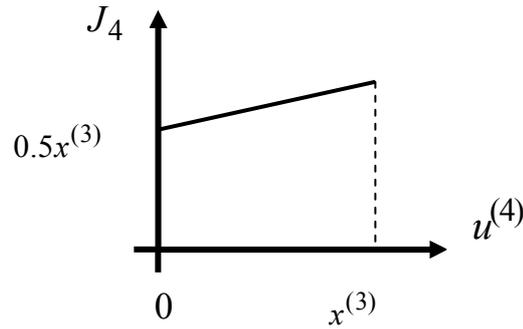


Рис. 13.5

IV шаг. Используем уравнение (13.26). Обозначим через J_4 функцию, стоящую в скобках, $J_4 = 0.5x^{(k-1)} + 0.1u^{(k)}$; функция J_4 — линейная, возрастающая. Поэтому максимум достигается на конце интервала $[0, x^{(3)}]$ (см. рис. 13.5) Поэтому, $B_4(x^{(3)}) = 0.6x^{(3)}$ при $\bar{u}^{(4)}(x^{(3)}) = x^{(3)}$.

III шаг. Запишем уравнение (13.27) для $k=3$

$$B_3(x^{(2)}) = \max_{0 \leq u^{(3)} \leq x^{(2)}} \{0.5x^{(2)} + 0.1u^{(3)} + 0.6x^{(3)}\}.$$

Найдем $x^{(3)}$ из уравнения состояния (24) $x^{(3)} = 0.8x^{(2)} - 0.1u^{(3)}$ и подставим его в правую часть, получим

$$B_3(x^{(2)}) = \max_{0 \leq u^{(3)} \leq x^{(2)}} \{0.5x^{(2)} + 0.1u^{(3)} + 0.6(0.8x^{(2)} - 0.1u^{(3)})\}, \text{ или}$$

$$B_3(x^{(2)}) = \max_{0 \leq u^{(3)} \leq x^{(2)}} \{0.98x^{(2)} + 0.04u^{(3)}\}.$$

Как и в предыдущем случае, максимум достигается при $u^{(3)} = x^{(2)}$, т.е.

$$B_3(x^{(2)}) = 1.02x^{(2)} \text{ при } \bar{u}^{(3)}(x^{(2)}) = x^{(2)}.$$

II шаг. Из уравнения состояния находим $x^{(2)}$: $x^{(2)} = 0.8x^{(1)} - 0.1u^{(2)}$. Уравнение (13.27) при $k=2$ примет вид:

$$B_2(x^{(1)}) = \max_{0 \leq u^{(2)} \leq x^{(1)}} \{0.5x^{(1)} + 0.1u^{(2)} + 1.02x^{(2)}\} =$$

$$= \max_{0 \leq u^{(2)} \leq x^{(1)}} \{0.5x^{(1)} + 0.1u^{(2)} + 1.02(0.8x^{(1)} - 0.1u^{(2)})\}. \text{ После преобразования получим:}$$

$$B_2(x^{(1)}) = \max_{0 \leq u^{(2)} \leq x^{(1)}} \{1.316x^{(1)} - 0.002u^{(2)}\}.$$

Имеем убывающую по $u^{(2)}$ функцию на отрезке $[0, x^{(1)}]$, поэтому максимум достигается при $u^{(2)} = 0$ (см. рис. 13.6) $B_2(x^{(1)}) = 1.316x^{(1)}$ при $\bar{u}^{(2)}(x^{(1)}) = 0$.

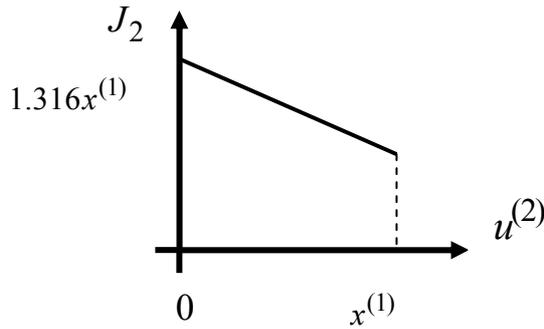


Рис. 13.6

I шаг. $x^{(1)} = 0.8x^{(0)} - 0.1u^{(1)}$. Уравнение (27) при $k=1$ имеет вид:

$$B_1(x^{(0)}) = \max_{0 \leq u^{(1)} \leq x^{(0)}} \{0.5x^{(0)} + 0.1u^{(1)} + 1.316(0.8x^{(0)} - 0.1x^{(1)})\}, \text{ или}$$

$$B_1(x^{(0)}) = \max_{0 \leq u^{(1)} \leq x^{(0)}} \{1.5528x^{(0)} - 0.0316u^{(1)}\}.$$

Максимум достигается в начале отрезка, т.е.

$$B_1(x^{(0)}) = 1.5528x^{(0)} \text{ при } u^{(1)}(x^{(0)}) = 0.$$

Переходим к безусловной оптимизации. $J_{\max} = B_1(10000) = 15528$.

$$u_1^{-(1)} = 0, u_2^{-(1)} = x^{(0)} - u_1^{-(1)} = 10000 \text{ (все средства выделяются II отрасли);}$$

$$\rightarrow x^{-(1)} = 0.8x^{-(0)} - 0.1u^{-(0)} = 0.8 \cdot 10000 - 0.1 \cdot 0 = 8000 \Rightarrow$$

$$u_1^{-(2)} = 0, u_2^{-(2)} = x^{-(1)} = 8000 \text{ (все средства выделяются II отрасли);}$$

$$\rightarrow x^{-(2)} = 0.8x^{-(1)} - 0.1u^{-(1)} = 0.8 \cdot 8000 - 0.1 \cdot 0 = 6400 \Rightarrow$$

$$u_1^{-(3)} = 6400, u_2^{-(3)} = x^{-(2)} - u_1^{-(3)} = 0 \text{ (все средства выделяются I отрасли);}$$

$$\rightarrow x^{-(3)} = 0.8x^{-(2)} - 0.1u^{-(2)} = 0.8 \cdot 6400 - 0.1 \cdot 6400 = 4480 \Rightarrow$$

$$u_1^{-(4)} = 4480, u_2^{-(4)} = x^{-(3)} - u_1^{-(4)} = 0 \text{ (все средства выделяются I отрасли).}$$

Оптимальная прибыль за 4 года, полученная от двух отраслей производства при начальных средствах 10000 у.е., равна 15528 у.е. при условии, что I отрасль получает по годам (0, 0, 6400, 4480), а II отрасль — соответственно (10000, 8000, 0, 0).

13.5. Задача о замене оборудования

Замена оборудования — важная экономическая проблема. Задача состоит в определении оптимальных сроков замены старого оборудования (станков, производственных зданий и т. п.). Старение оборудования включает его физический и моральный износ, в результате чего растут производственные затраты, затраты на ремонт и обслуживание, снижаются производительность труда, ликвидная стоимость. Критерием оптимальности являются, как правило, либо прибыль от эксплуатации оборудования (задача максимизации), либо суммарные затраты на эксплуатацию в течение планируемого периода (задача минимизации).

При построении модели задачи принято считать, что решение о замене выносится в начале каждого промежутка эксплуатации (например, в начале года) и что в принципе оборудование можно использовать неограниченно долго.

Основная характеристика оборудования — параметр состояния — его возраст t .

При составлении динамической модели замены процесс замены рассматривают как N -шаговый, разбивая весь период эксплуатации на N шагов. Возможное управление на каждом шаге характеризуется качественными признаками, например, U^c (сохранить оборудование), U^3 (заменить) и U^P (сделать ремонт).

Рассмотрим конкретный пример.

- **Пример.** Оборудование эксплуатируется в течение 5 лет, после этого продается. В начале каждого года можно принять решение сохранить оборудование или заменить его новым. Стоимость нового оборудования $p_0 = 4000$ ден. ед. После t лет эксплуатации ($1 \leq t \leq 5$) оборудование можно продать за $g(t) = p_0 2^{-t}$ ден. ед. (ликвидная стоимость). Затраты на содержание в течение года зависят от возраста t оборудования и равны $r(t) = 600(t+1)$. Определить оптимальную стратегию эксплуатации оборудования, чтобы суммарные затраты с учетом начальной покупки и заключительной продажи были минимальны.

Решение. Способ деления управления на шаги естественный, по годам, $N = 5$. Параметр состояния — возраст машины — $x^{(k-1)} = t$, $x^{(0)} = 0$ — машина новая в начале первого года эксплуатации. Управление на каждом шаге зависит от двух переменных U^c и U^3 .

Уравнения состояний зависят от управления:

$$x^{(k-1)} = \begin{cases} t+1, & \text{если } u^{(k)} = U^c, \\ 1, & \text{если } u^{(k)} = U^3, \end{cases} \quad k = 1, 2, 3, 4. \quad (13.28)$$

В самом деле, если к k -му шагу $x^{(k-1)} = t$, то при сохранении машины ($u^{(k)} = U^c$) через год возраст машины увеличится на 1. Если машина заменяется новой ($u^{(k)} = U^3$), то это означает, что к началу k -го шага ее возраст $t = 0$, а после года эксплуатации $t = 1$, т.е. $x^{(k)} = 1$.

Показатель эффективности k -го шага:

$$J_k(u^{(k)}, t) = \begin{cases} 600(t+1), & \text{если } u^{(k)} = U^c, \\ 4600 - 4000 \cdot 2^{-t}, & \text{если } u^{(k)} = U^3, \end{cases} \quad k = 1, \dots, 4. \quad (13.29)$$

(При U^c затраты только на эксплуатацию машины возраста t , при U^3 машина продается ($-4000 \cdot 2^{-t}$), покупается новая (4000) и эксплуатируется в течение первого года (600), общие затраты равны ($-4000 \cdot 2^{-t} - 4000 + 600$).

Пусть $B_k(t)$ — условные оптимальные затраты на эксплуатацию машины, начиная с k -го шага до конца, при условии, что к началу k -го шага машина имеет возраст t лет. Запишем для функций $B_k(t)$ уравнения Беллмана (10) и (13), заменив задачу максимизации на задачу минимизации:

$$B_5 = \min \begin{cases} 600(t+1) - 4000 \cdot 2^{-(t+1)}, & \text{если } u^{(5)} = U^c, \\ 4600 - 4000 \cdot 2^{-t} - 4000 \cdot 2^{-(t+1)}, & \text{если } u^{(5)} = U^3. \end{cases} \quad (13.30)$$

Величина $4000 \cdot 2^{-(t+1)}$ — стоимость машины возраста t лет (по условию машина после 5 лет эксплуатации продается).

$$B_k = \min \begin{cases} 600(t+1) + B_{k+1}(t+1), & \text{если } u^{(k)} = U^c, \\ 4600 - 4000 \cdot 2^{-t} + B_{k+1}(1), & \text{если } u^{(k)} = U^3. \end{cases} \quad (13.31)$$

$$k = 4, 3, 2, 1.$$

Из определения функций $B_k(t)$ следует

$$J_{\min} = B_1(0).$$

Дадим геометрическое решение этой задачи. На оси абсцисс будем откладывать номер шага k , на оси ординат — возраст t машины. Точка $(k-1, t)$ на плоскости соответствует началу k -го года эксплуатации машины возраста t лет. Перемещение на графике в зависимости от принятого управления на k -м шаге показано на рис. 7.

Состояние начала эксплуатации машины соответствует точке $x^{(0)}(0, 0)$, конец — точкам $\hat{x}(6, t)$. Любая траектория, переводящая точку $x(k-1, t)$ из $x^{(0)}$ в \hat{x} , состоит из отрезков — шагов, соответствующих годам эксплуатации. Надо выбрать такую траекторию, при которой затраты на эксплуатацию машины окажутся минимальными.

Над каждым отрезком, соединяющим точки $(k-1, t)$ и $(k, t+1)$, запишем соответствующие управлению U^c затраты, найденные из (13.29): $600(t+1)$, а

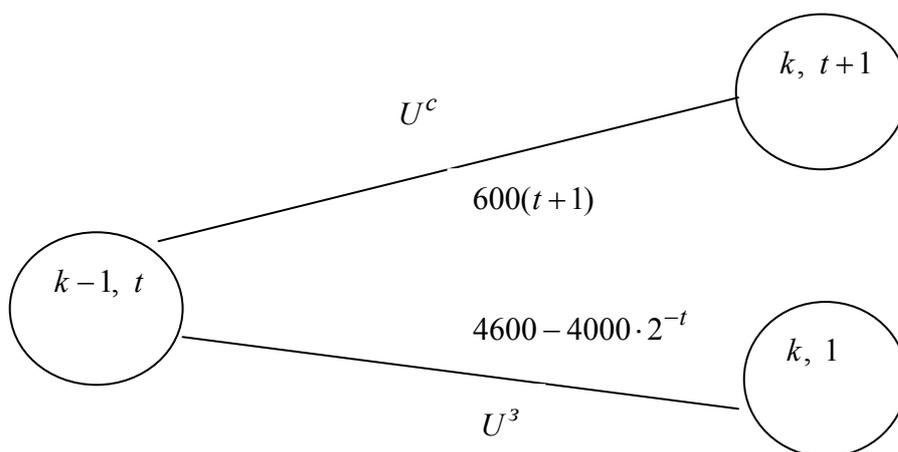


Рис. 13.7.

над отрезком, соединяющим $(k-1, t)$ и $(k, 1)$, запишем затраты, соответствующие управлению U^3 , т.е. $4600 - 4000 \cdot 2^{-t}$. Таким образом, мы разметим все отрезки, соединяющие точки на графике, соответствующие переходам из любого состояния $x^{(k-1)}$ в состояние $x^{(k)}$ (см. рис. 13.8). Например, над отрезками, соединяющими точки $(k, 2)$ и $(k+1, 3)$, стоит число 1800, что соответствует затратам на эксплуатацию в течение каждого года машины возраста $t = 2$ лет, а над отрезками,

соединяющими $(k, 2)$ и $(k+1, 1)$, стоит число 3600 — сумма затрат на покупку машины и эксплуатацию новой машины в течение года без "затрат" (выручки) за проданную машину возраста t лет. Следует учесть, что $0 \leq t \leq k$.

Проведем на размеченном графе состояний (см. рис. 13.8) условную оптимизацию.

V шаг. Начальные состояния — точки $(4; t)$, конечные $(5; t)$. В состояниях $(5; t)$ машина продается, условный оптимальный доход от продажи равен $4000 \cdot 2^{-t}$, но поскольку целевая функция связана с затратами, то в кружках точек $(5; t)$ поставим величину дохода со знаком минус.

Анализируем, как можно попасть из каждого начального состояния в конечное на V шаге.

Состояние $(4; 1)$. Из него можно попасть в состояние $(5; 2)$, затратив на эксплуатацию машины 1200 и выручив затем от продажи 1000, т.е. суммарные затраты равны 200, и в состояние $(5; 1)$ с затратами $2600 - 2000 = 600$. Значит, если к последнему шагу система находилась в точке $(4; 1)$, то следует идти в точку $(5; 2)$ (укажем это направление жирной стрелкой), а неизбежные минимальные затраты, соответствующие этому переходу, равны 200 (поместим эту величину $B_5(1) = 200$ в кружке точки $(4; 1)$).

Состояние $(4; 2)$. Из него можно попасть в точку $(5; 3)$ с затратами $1800 - 500 - 1300$ и в точку $(5; 1)$ с затратами $3600 - 2000 = 1600$. Выбираем первое управление, отмечаем его жирной стрелкой, а $B_5(2) = 1300$ проставляем в кружке точки $(4; 2)$.

Рассуждая таким же образом для каждой точки предпоследнего шага, мы найдем для любого исхода IV шага оптимальное управление на V шаге, отметим его на рис. 13.8 жирной стрелкой.

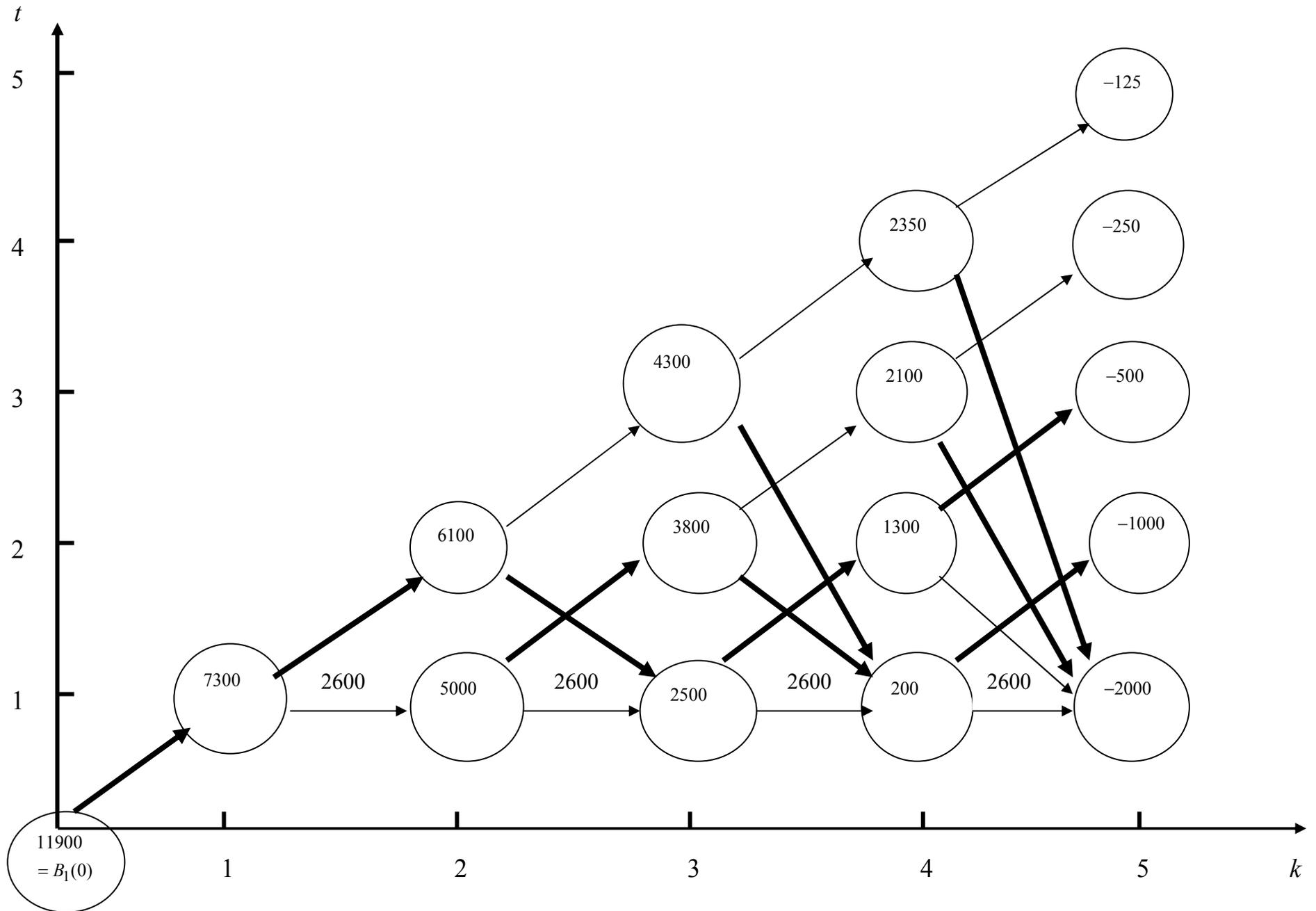
Далее планируем IV шаг, анализируя каждое состояние, в котором может оказаться система в конце III шага с учетом оптимального продолжения до конца процесса, т.е. решаем для всех $0 \leq t \leq 4$ при $k=4$ уравнения (13.28). Например, если начало IV шага соответствует состоянию $(3; 1)$, то при управлении U^c система переходит в точку $(4; 2)$, затраты на этом шаге 1200, а суммарные затраты за два последних шага равны $1200 + 1300 = 2500$. При управлении U^3 затраты за два шага равны $2600 + 200 = 2800$. Выбираем минимальные затраты 2500, ставим их в кружок точки $(3; 1)$, а соответствующее управление на этом шаге помечаем жирной стрелкой, ведущей из состояния $(3; 1)$ в состояние $(4; 2)$. Так поступаем для каждого состояния $(3; t)$ (см. рис. 13.8).

Продолжая условную оптимизацию III, II и I шагов, мы получим на рис. 8 следующую ситуацию: из каждой точки (состояния) выходит стрелка, указывающая, куда следует перемещаться в данном шаге, если система оказалась в этой точке, а в кружках записаны минимальные затраты на переход из этой точки в конечное состояние. На каждом шаге графически решались уравнения (13.28).

После проведения условной оптимизации получим в точке $(0; 0)$ минимальные затраты на эксплуатацию машины в течение 5 лет с последующей продажей: $J_{\min} = 11900$. Далее строим оптимальную траекторию, перемещаясь из точки $x^{(0)}(0; 0)$ по жирным стрелкам в \hat{x} . Получаем набор точек:

$$\{(0; 0), (1; 1), (2; 2), (3; 1), (4; 2), (5; 3)\},$$

который соответствует оптимальному управлению $\bar{u} = (U^c, U^c, U^3, U^c, U^c)$. Оптимальный режим эксплуатации состоит в том, чтобы заменить машину новой в начале 3-го года.



Таким образом, размеченный график (сеть) позволяет наглядно интерпретировать расчетную схему и решить задачу методом ДП.

Вопросы для самопроверки

1. Что понимают под динамическим программированием?
2. Запишите условие многошаговой задачи оптимизации
3. Перечислите особенности модели динамического программирования
4. В чем состоит принцип оптимальности управления при решении задачи динамического программирования?
5. Запишите уравнения Беллмана
6. Запишите модель задачи о распределении средств между предприятиями в виде модели динамического программирования
7. Запишите модель задачи об оптимальном распределении ресурсов между отраслями на N лет в виде модели динамического программирования
8. Запишите модель задачи о замене оборудования в виде модели динамического программирования

Литература

1. Кремер Н.Ш. и др. Исследование операций в экономике. Учебное пособие для вузов/ ред. : Н. Ш. Кремер. - М. : ЮНИТИ, 2006. - 407 с
2. Есипов Б.А. Методы исследования операций: Учебное пособие. – СПб.: Изд-во «Лань», 2010. – 256 с. (электр. ресурс). – Режим доступа: <http://e.lanbook.com/view/book/10250/>
3. Лесин В.В., Лисовец Ю.П. Основы методов оптимизации: Учебное пособие. – СПб.: Изд-во «Лань», 2011. – 352с. (электр. ресурс). – Режим доступа: <http://e.lanbook.com/view/book/1552>
4. Гладких Б. А. Методы оптимизации и исследование операций для бакалавров информатики Ч. 1.: учебное пособие. Томск: Изд-во НТЛ, 2009. – 198 с. /<http://sun.tsu.ru/mminfo/books/2010/000374996/000374996.djvu> (электронное издание djvu 1,0 Mb)
5. Гладких Б. А. Методы оптимизации и исследование операций для бакалавров информатики Ч. 2.: учебное пособие. Томск: Изд-во НТЛ, 2011. – 263 с./ <http://sun.tsu.ru/mminfo/books/2012/000416882/000416882.pdf> (электронное издание Adobe PDF 7,6 М)