**Федеральное государственное образовательное бюджетное учреждение высшего образования**
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**
**(ФИНАНСОВЫЙ УНИВЕРСИТЕТ)**

**Кафедра анализа данных и машинного обучения**
**Факультета информационных технологий и анализа больших данных**

# МЕТОДЫ ОПТИМИЗАЦИИ

**Учебное пособие**
**Часть I**

**Т.В. Золотова**

**Москва 2024**

**Федеральное государственное образовательное бюджетное учреждение высшего образования**
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ**
**ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**
**(Финансовый университет)**

**Кафедра анализа данных и машинного обучения**

**Факультета информационных технологий и анализа больших данных**

# МЕТОДЫ ОПТИМИЗАЦИИ

**Учебное пособие**
**Часть I**
*Издание переработанное и дополненное*

**Т.В. Золотова**

*Утверждено на заседании Кафедры анализа данных и машинного обучения*
*(протокол № 3 от 21.03.2024 г.)*

**Москва 2024**

**УДК 519.8(075.8)**
**ББК 65в6**
**З-81**

**Рецензент:** И.В. Трегуб, д.э.н., профессор кафедры математики Финансового университета при Правительстве РФ

**Т.В.Золотова. Методы оптимизации. Учебное пособие. Часть I.** – М.: Финансовый университет, кафедра анализа данных и машинного обучения, 2024. – 153 с.

Издание предназначено для студентов, обучающихся по направлению «Прикладная математика и информатика» и содержит наиболее широко используемые классы оптимизационных моделей, основные принципы оптимальности и методы нахождения оптимальных решений, приведены примеры постановки и решения финансово-экономических задач.

**УДК 519.8(075.8)**
**ББК 65в6**

***Учебное издание***

***Т.В.Золотова***

**Методы оптимизации**

Курс лекций (учебное пособие)

Компьютерный набор, верстка: ***Т.В.Золотова***

Формат 60x90/16. Гарнитура *Times New Roman*
Усл. п.л. 0,0. Изд. № 28.2 - 2024. Тираж - 0 экз.
Заказ № ______

**Отпечатано в Финансовом университете**

## СОДЕРЖАНИЕ

**Введение**…………………………………………………………….................... 3
**Глава 1. Классические оптимизационные задачи**…………………………... 5
1.1. Основные понятия теории экстремальных задач……………………… 5
1.2. Условия экстремума в задачах без ограничений…………………….. 7
1.3. Функция Лагранжа и условия экстремума в задачах с ограничениями типа равенств………………………………………… 10
1.4. Интерпретация множителей Лагранжа………………………………….. 15
1.5. Численные методы поиска экстремума без ограничений……………… 17
1.6. Примеры применения методов классической оптимизациии………… 22
Задачи и упражнения……………………………………………………….. 26
**Глава 2. Нелинейное программирование**……………………………………… 28
2.1. Общая задача математического программирования…………………… 28
2.2. Седловые точки и двойственность……………………………………… 30
2.3. Выпуклое программирование…………………………………………… 35
2.4. Интерпретация множителей Лагранжа в задаче математического программирования…………………………………….. 45
2.5. Численные методы нелинейного программирования………………….. 46
2.6. Использование методов нелинейного программирования в эколого-экономической сфере………………………………………… 54
Задачи и упражнения……………………………………………………….. 59
**Глава 3. Линейные задачи оптимизации**……………………………………... 62
3.1. Элементы линейного программирования……………………………….. 62
3.2. Дискретное программирование…………………………………………….. 68
3.3. Анализ чувствительности. Некорректные задачи...…………………….. 81
3.4. Несобственные задачи оптимизации. Задачи связанной коррекции данных в эколого-экономических моделях……………….... 86
3.5. Применение методов линейной оптимизации для решения матричных игр…………………………………………….... 95
Задачи и упражнения……………………………………………………….. 113
**Глава 4. Динамическая оптмизация**…….……………………………………… 116
4.1. Примеры задач оптимизации динамических систем…………………... 116
4.2. Задача динамического программирования. Функция Беллмана……… 122
4.3. Использование метода динамического программирования в решении экономических задач………………………………………… 128
4.4. Применение метода динамического программирования в сетевых задачах………………………………………………………… 138
4.5. Модели управления запасами…………………………………………… 142
Задачи и упражнения……………………………………………………….. 147
**Заключение**……………………………………………………………………….. 150
**Ответы к задачам и упражнениям**……………………………………………. 151
**Список рекомендуемой литературы к части I** ……………………………........... 153

# ВВЕДЕНИЕ

Практически любой вид человеческой деятельности связан с ситуациями, когда имеется несколько вариантов действий и человек может из этих действий выбрать любое, наиболее подходящее ему. Таким образом, приходится решать задачи наилучшего выбора, используя имеющуюся в наличии информацию и средства. Методы оптимизации являются математическим аппаратом для решения задачи наилучшего выбора.

Предложенные в пособии математические модели и соответственно методы оптимизации в различных сложных системах и процессах имеют разные аспекты. В оптимизационных моделях присутствуют такие характеристики, как многокритериальность, случайность, неопределенность. Оптимизационные модели имеют приложения в финансах, экономике, экологии.

Цель издания учебного пособия – полнота и доступность изложения предмета. При этом должно обеспечиваться получение необходимых математических знаний по методам оптимизации, используемых в различных ситуациях и областях профессиональной деятельности. Необходимо создать основу развития у студентов способности применять математические методы для исследования экономических и финансовых процессов.

Учебное пособие «Методы оптимизации» написано в соответствии с утвержденной программой дисциплины направления подготовки «Прикладная математика и информатика». Изучение дисциплины «Методы оптимизации» основывается на базе знаний, полученных студентами в ходе освоения дисциплин «Математический анализ», «Алгебра», «Теория вероятностей», «Математическая статистика».

В результате освоения дисциплины «Методы оптимизации» студент должен:

**знать**

– классификацию используемых оптимизационных моделей: статические и динамические, непрерывные и дискретные, линейные и нелинейные, детерминированные, случайные, неопределенные;

– методы оптимизации линейного, нелинейного, динамического, векторного, целевого программирования;

**уметь**

– ставить цели и задачи, связанные с реализацией решений, анализировать проблемные ситуации и разрабатывать решения по ним, формулировать критерии оптимизации и ограничения на управляемые переменные;

– определять альтернативы и давать им оценку исходя из наличия имеющихся ресурсов;

– применять методы оптимизации для нахождения оптимальных решений в экономических и финансовых областях;

– выбирать математический инструментарий для каждого этапа принятия решения;

**владеть**

– методиками анализа предметной области;

– методами математического моделирования экономических систем и процессов.

– навыками применения компьютерных технологий при решении задач;

– методикой построения, анализа и применения математических моделей для оценки состояния и прогноза развития экономических явлений и процессов.

Данное учебное пособие состоит из двух частей. Первая часть посвящена методам оптимизации в условиях полной информированности об условиях, в которых принимается решение. В первой главе рассматриваются классические задачи оптимизации. Вторая глава посвящена математическому программированию. Третья глава изучает линейные задачи оптимизации. Четвертая глава содержит динамические задачи оптимизации. В каждой главе имеются приложения математических методов оптимизации к решению различных экономических задач.

Вторая часть учебного пособия изучает методы оптимизации в условиях случайности и неопределенности.

Наряду со сведениями теоретического характера в каждой главе разбираются примеры и задачи, цель которых – уяснение основных понятий и математических методов. Кроме того, предложены задания для самостоятельного выполнения.

Дополнительные теоретические сведения для более глубокого изучения того или иного раздела можно получить из книг, приведенных в списке литературы.

Настоящее учебное пособие адресовано студентам, обучающимся по различным направлениям подготовки, учебные планы которых включают дисциплины «Методы оптимизации», «Методы оптимальных решений», «Теория принятия решений», «Методы принятия управленческих решений», «Исследование операций». Оно может найти применение в обучении студентов по направлениям «Прикладная математика и информатика», «Прикладная информатика», «Бизнес-информатика» и другим направлениям подготовки бакалавров, а также может оказаться полезным магистрантам и аспирантам, интересующимся проблемами оптимизации и новыми областями приложений.

# Глава 1. КЛАССИЧЕСКИЕ ОПТИМИЗАЦИОННЫЕ ЗАДАЧИ

## 1.1. Основные понятия теории экстремальных задач

В условиях полной информированности о состоянии системы задача поиска оптимальных решений сводится к экстремальной задаче. Если при этом на выбор стратегии не накладывается ограничений (что на практике встречается редко) или ограничения имеют вид только равенств, то применимы классические методы оптимизации. Эти методы и рассматриваются в данной главе.

Исходными данными при постановке задачи поиска экстремума является множество $X$ и определенная на нем функция $f(x)$. Мы будем рассматривать конечномерные задачи, поэтому $x$ является вектором произвольной размерности $n$, т.е. $x = (x_1, \ldots, x_n)$, а множество $X$ - подмножеством евклидова пространства $R^n$ (возможно совпадающим со всем пространством). Помимо задания $X$ (его называют допустимым множеством) и $f(x)$ (ее называют целевой функцией) необходимо определить, что понимается под решением задачи. Во-первых, речь может идти о нахождении точек максимума (одной или всех), минимума (одной или всех) или тех и других. Во-вторых, необходимо уточнить само понятие максимума (минимума), так как оно может пониматься в глобальном и локальном смысле.

**Определение 1.1.** Точка $x^* \in X$ называется точкой глобального (абсолютного) максимума функции $f(x)$ на множестве $X$, если

$$f(x^*) \geq f(x) \quad \forall x \in X. \tag{1.1}$$

**Определение 1.2.** Точка $x^* \in X$ называется точкой локального максимума функции $f(x)$ на множестве $X$, если $\exists\, \varepsilon > 0$ такое, что

$$f(x^*) \geq f(x) \ \forall x \in X \bigcap U_\varepsilon(x^*), \tag{1.2}$$

где $U_\varepsilon(x^*)$ - $\varepsilon$-окрестность точки $x^*$ (шар радиусом $\varepsilon$ с центром в $x^*$).

Определения глобального и локального минимумов получаются заменой в (1.1) и (1.2) знаков неравенств на противоположные.

Глобальные максимумы и минимумы называют глобальными экстремумами функции, локальные максимумы и минимумы – локальными экстремумами. Ясно, что глобальные экстремумы являются и локальными, но обратное, вообще говоря, не верно. Если глобальный максимум существует, т.е. точная верхняя грань функции на множестве достигается, то в принципе, можно найти все локальные максимумы и выбрать из них тот, в котором функция принимает наибольшее значение (аналогично для минимумов). Однако такой способ применим, если доказано существование глобального максимума и возможно найти все локальные максимумы.

Если $x^*$ есть точка глобального максимума $f(x)$ на $X$, то будем использовать запись

$$f(x^*) = \max_{x \in X} f(x), \tag{1.3}$$

т.е. под $\max\limits_{x \in X} f(x)$ понимается максимальное (абсолютное) значение $f(x)$ на $X$. Для точки глобального максимума принято обозначение

$$x^* = \arg\max_{x \in X} f(x). \tag{1.4}$$

Частным решением задачи поиска максимума называется нахождение одной точки $x^*$ и значения $f(x^*)$, определяемых (1.3), (1.4). Полным решением называется нахождение величины $\max\limits_{x \in X} f(x)$ и всех реализующих ее значений аргумента, множество которых принято обозначать

$$\operatorname{Arg}\max_{x \in X} f(x) = \{x^* \in X \mid f(x^*) = \max_{x \in X} f(x)\}. \tag{1.5}$$

Обычно, если это не оговаривается особо, под решением задачи максимизации будем понимать нахождение ее частного решения (глобального).

Аналогичные понятия и обозначения используются для задачи поиска глобального минимума (соответственно $\min\limits_{x \in X} f(x)$, $\arg\min\limits_{x \in X} f(x)$, $\operatorname{Arg}\min\limits_{x \in X} f(x)$).

Множество всех точек локальных максимумов включает в себя множество (1.5), но, вообще говоря, с ним не совпадает. Задачу нахождения всех локальных максимумов мы будем записывать в виде

$$f(x) \to \max, x \in X. \tag{1.6}$$

Аналогично для локальных минимумов

$$f(x) \to \min, x \in X, \tag{1.7}$$

а для задачи нахождения всех локальных экстремумов (максимумов и минимумов)

$$f(x) \to \text{extr}, x \in X. \tag{1.8}$$

Так как, очевидно, точки максимума функции $f(x)$ совпадают с точками минимума функции $-f(x)$, то все свойства и методы решения задачи (1.6) легко переносятся на (1.7), (1.8) (мы будем их рассматривать на примере задачи максимизации).

Экстремальные задачи принято делить на задачи поиска безусловного и условного экстремума. Если $X = R^n$, т.е. рассматриваются точки экстремума $f(x)$ на всем $n$-мерном пространстве, то имеем задачу на безусловный экстремум (или без ограничений). Если $X \subset R^n$ (собственное подмножество $n$-мерного пространства), то имеем задачу на условный экстремум (или с ограничениями). Терминология эта возникла в связи с тем, что множество $X$ в последнем случае обычно задается какими-то условиями (обычно ограничениями вида равенств и неравенств). В классическом анализе рассматриваются задачи без ограничений и с ограничениями типа равенств.

Задача поиска экстремума может, вообще говоря, не иметь решения (когда верхняя или нижняя грань функции на множестве не достигается). В курсе математического анализа доказывается следующий важный результат.

**Теорема 1.1.** (теорема Вейерштрасса) *Если $X$ – компакт в $R^n$ (замкнутое ограниченное множество), $f(x)$ – непрерывная на $X$ функция, то точки глобального максимума и минимума $f(x)$ на $X$ существуют.*

Из условий теоремы Вейерштрасса наиболее обременительным является ограниченность множества $X$, часто из постановки задачи она не вытекает, а для задачи на безусловный экстремум заведомо не выполняется. Поэтому полезной является следующая ее модификация.

**Следствие 1.1.** *Если $X$ - замкнутое непустое множество в $R^n$, $f(x)$ – непрерывная на $X$ функция и ограничено множество*

$$S = \{x \in X \mid f(x) \geq c\}, \tag{1.9}$$

*где $c$ - некоторая константа, меньшая верхней грани $f(x)$ на $X$, то точка глобального максимума $f(x)$ на $X$ существует.*

**Доказательство.** Множество $S$ является по определению непустым подмножеством множества $X$, оно ограничено, а в силу непрерывности $f(x)$, и замкнуто. Поэтому по теореме Вейерштрасса существует точка глобального максимума $f(x)$ на $S$ со значением не меньше $c$. Но на множестве $X \setminus S$ выполняется $f(x) < c$, следовательно, данная точка является и глобальным максимумом $f(x)$ на $X$, что и требовалось доказать.

Для существования глобального минимума в определении (1.9) множества $S$ знак неравенства следует заменить на противоположный, а константа $c$ должна быть больше нижней грани $f(x)$ на $X$.

Условие ограниченности множества $S$ можно заменить на следующее: если для любой последовательности векторов $x^{(k)} \in X$, стремящейся к бесконечности по модулю (евклидовой норме) векторов $\left\|x^{(k)}\right\| \to \infty$, последовательность $f(x^{(k)}) \to -\infty$, то точка глобального максимума $f(x)$ на $X$ существует, а глобального минимума, естественно, нет (для существования глобального минимума требуется, соответственно, $f(x^{(k)}) \to \infty$, при этом нет глобального максимума).

## 1.2. Условия экстремума в задачах без ограничений

Задача поиска безусловного экстремума (без ограничений) имеет вид

$$f(x) \to \text{extr}, x \in R^n. \tag{1.10}$$

Классические необходимые условия экстремума (локального или глобального, максимума или минимума) для этой задачи дает следующая теорема (Ферма).

**Теорема 1.2.** (теорема Ферма) *Пусть функция* $f(x)$ *дифференцируема в точке решения задачи* (1.10) $x^*$*, тогда*

$$\frac{\partial f(x^*)}{\partial x_j} = 0, \; j = 1,\dots,n \tag{1.11}$$

**Доказательство.** По определению дифференцируемой функции многих переменных для любого вектора $h \in R^n$ имеем

$$f(x^* + dh) = f(x^*) + d\,\langle f'(x^*), h\rangle + 0(d), \tag{1.12}$$

где $f'(x^*) = \left(\frac{\partial f(x^*)}{\partial x_1},\dots,\frac{\partial f(x^*)}{\partial x_n}\right)$ – градиент функции $f(x)$ в точке $x^*$, $d$ – достаточно малое положительное число, $0(d)$ – величина более высокого порядка малости, нежели $d$, $\langle f'(x^*), h\rangle$ – скалярное произведение векторов $f'(x^*)$ и $h$.

Предположим противное утверждению (1.11), т.е. что $f'(x^*) \neq 0$. Тогда из (1.12) при достаточно малых $d$ имеем $f(x^*+ dh) > f(x^*)$ при $h = f'(x^*)$ и $f(x^*+ dh) < f(x^*)$ при $h = -f'(x^*)$, что противоречит определению экстремумов.

Условие (1.11) или эквивалентное ему

$$f'(x^*) = 0 \tag{1.13}$$

называется условием первого порядка, т.к. использует первые частные производные. Оно является необходимым, но не достаточным, т.е. ему могут удовлетворять и точки, не являющиеся экстремумами (все эти точки называются стационарными).

Для того чтобы сформулировать необходимые и достаточные условия второго порядка (использующие вторые производные), напомним некоторые алгебраические понятия.

Пусть задана квадратная симметрическая матрица $A = (a_{ij})$ размера $n \times n$. Рассмотрим скалярное произведение $\langle Ah, h\rangle$, где первым вектором является произведение матрицы $A$ на $n$-мерный вектор $h$ (мы не будем делать различия в обозначениях между вектор-строкой и вектор-столбцом, считая всегда, что вектор задан в такой форме, которая позволяет производить соответствующую операцию умножения на него матрицы).

Это скалярное произведение равно $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}h_ih_j$, где $a_{ij}$ – элементы матрицы $A$, $h_i$ – компоненты вектора $h$, т.е. матрица $A$ задает квадратичную форму. Квадратичная форма $\langle Ah, h\rangle$ называется положительно определенной, если

$$\langle Ah, h\rangle > 0 \text{ при всех ненулевых } h \in R^n. \qquad (1.14)$$

Аналогично вводятся понятия неотрицательной, неположительной и отрицательной определенности (в (1.14) знак > меняется соответственно на ≥, ≤, <). При этом и матрица $A$ называется положительно (неотрицательно, неположительно, отрицательно) определенной.

Непосредственная проверка положительной определенности, т.е. неравенства (1.14), непроста. Поэтому желательно иметь для нее более конструктивные условия. Наиболее общим является *критерий Сильвестра.* Для положительной определенности матрицы $A$ необходимо и достаточно, чтобы были положительными ее главные миноры

$$a_{11}, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \dots, \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}.$$

Необходимым и достаточным условием отрицательной определенности матрицы $A$ является положительность главных миноров четного порядка и отрицательность главных миноров нечетного порядка, т.е.

$$a_{11} < 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} < 0 \text{ и т.д.}$$

Вернемся к задаче поиска безусловного экстремума. Так как условия второго порядка для максимумов и минимумов несколько отличаются, то рассмотрим сначала задачу

$$f(x) \to \max, \ x \in R^n. \qquad (1.15)$$

Будем считать $f(x)$ дважды дифференцируемой и введем матрицу ее вторых частных производных, называемую матрицей Гессе,

$$f''(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix}.$$

Это квадратная симметрическая матрица порядка $n$.

**Теорема 1.3.** (необходимые условия максимума второго порядка). *Пусть функция* $f(x)$ *дважды дифференцируема в точке решения задачи* (1.15) $x^*$, *тогда матрица ее вторых производных в данной точке* $f''(x^*)$ *неположительно определена, т.е.* $\langle f''(x^*)h, h\rangle \le 0$ *для любого* $h \in R^n$.

**Доказательство.** Используя разложение функции многих переменных в ряд Тейлора до квадратичных членов, имеем

$$f(x^*+dh)=f(x^*)+d\langle f'(x^*),h\rangle+\frac{1}{2}d^2\langle f''(x^*)h,h\rangle+0(d^2). \qquad (1.16)$$

Так как $x^*$ – локальный максимум, то по определению $f(x^*+dh)\le f(x^*)$ при достаточно малом $d$ и любом $h\in R^n$, а по теореме Ферма $f'(x^*)=0$, поэтому из (1.16) предельным переходом при $d\to 0$ получаем утверждение теоремы.

Условием минимума является неотрицательная определенность матрицы $f''(x^*)$ (доказывается аналогично).

**Теорема 1.4.**(достаточные условия максимума второго порядка). *Пусть функция* $f(x)$ *дважды дифференцируема в точке* $x^*$, $f'(x^*)=0$ *и матрица* $f''(x^*)$ *отрицательно определена, тогда* $x^*$ *– решение задачи* (1.15).

**Доказательство.** Используя разложение (1.16) и учитывая, что линейный член в этом разложении равен нулю, а квадратичный член отрицателен для любого ненулевого $h$, получаем утверждение теоремы.

Задаче на минимум соответствует условие положительной определенности матрицы $f''(x^*)$. Для функции одной переменной эти условия означают просто отрицательность (положительность) второй производной.

Для функции двух переменных с учетом критерия Сильвестра имеем условия максимума

$$\frac{\partial f(x_1^*,x_2^*)}{\partial x_1}=\frac{\partial f(x_1^*,x_2^*)}{\partial x_2}=0,\quad \frac{\partial^2 f(x_1^*,x_2^*)}{\partial x_1^2}<0,$$

$$\frac{\partial^2 f(x_1^*,x_2^*)}{\partial x_1^2}\cdot\frac{\partial^2 f(x_1^*,x_2^*)}{\partial x_2^2}-\left[\frac{\partial^2 f(x_1^*,x_2^*)}{\partial x_1\partial x_2}\right]^2>0.$$

Для задачи на минимум в этих условиях $\frac{\partial^2 f(x_1^*,x_2^*)}{\partial x_1^2}>0$, а остальные условия не меняются.

## 1.3. Функция Лагранжа и условия экстремума в задачах с ограничениями типа равенств

Задача поиска условного экстремума (с ограничениями), вообще говоря, сложнее задачи на безусловный экстремум. Полученные выше условия экстремума сохраняют силу для внутренних точек множества $X$, так как для такой точки $x^*$ при достаточно малом $d$ вектор $x^*+dh$ для любого $h$ принадлежит $X$ и все доказательства теорем о необходимых и достаточных условиях проходят без изменений. Однако часто экстремум оказывается на границе множества $X$, а здесь вектора $x^*+dh$ могут не принадлежать множеству $X$ и полученные результаты уже становятся неприменимыми.

Рассмотрим классическую задачу на условный экстремум, в которой множество $X$ задается конечным числом ограничений типа равенств

$$X=\{ x \in R^n \mid g_i(x) = 0,\ i=1,\dots,m \}. \tag{1.17}$$

Задачи с ограничениями-неравенствами будут рассмотрены в последующих главах.

Суть используемого подхода продемонстрируем сначала на частном случае задачи с двумя переменными и одним ограничением

$$f(x_1, x_2) \to \text{extr} \quad \text{на} \quad X = \{(x_1, x_2) \in R^2 \mid g(x_1, x_2) = 0\}. \tag{1.18}$$

Пусть из уравнения $g(x_1, x_2) = 0$ мы можем однозначно выразить одну переменную через другую, например $x_2=\varphi(x_1)$. Тогда, подставляя эту функциональную зависимость в функцию $f(x_1, x_2)$, получим задачу поиска безусловного экстремума функции одной переменной $\Psi(x_1) = f(x_1, \varphi(x_1))$. Если функции $f$ и $\varphi$ дифференцируемы, то по теореме Ферма, используя правило дифференцирования сложной функции, имеем для точки экстремума $x^*$

$$\frac{d\Psi(x_1^*)}{dx_1} = \frac{\partial f(x_1^*, \varphi(x_1^*))}{\partial x_1} + \frac{\partial f(x_1^*, \varphi(x_1^*))}{\partial x_2} \cdot \frac{d\varphi(x_1^*)}{dx_1} = 0. \tag{1.19}$$

Однако условие (1.19) неудобно, так как найти $\varphi(x_1)$ и ее производную в явном виде не всегда просто. Поэтому желательно выразить все через исходные функции. Так как $x_2=\varphi(x_1)$ эквивалентно $g(x_1, x_2)=0$, имеем $g(x_1, \varphi(x_1)) \equiv 0$, откуда при условии дифференцируемости функции $g$

$$\frac{\partial g(x_1, \varphi(x_1))}{\partial x_1} + \frac{\partial g(x_1, \varphi(x_1))}{\partial x_2} \cdot \frac{d\varphi(x_1)}{dx_1} \equiv 0. \tag{1.20}$$

Если $\dfrac{\partial g(x_1^*, \varphi(x_1^*))}{\partial x_2} \neq 0$, то из (1.20) получаем

$$\frac{d\varphi(x_1^*)}{dx_1} = -\left[\frac{\partial g(x_1^*, \varphi(x_1^*))}{\partial x_1}\right] \cdot \left[\frac{\partial g(x_1^*, \varphi(x_1^*))}{\partial x_2}\right]^{-1}.$$

Подставляя это выражение в (1.19), имеем в точке экстремума $(x_1^*, x_2^*)$, где $x_2^* = \varphi(x_1^*)$, условие

$$\frac{\partial f(x_1^*, x_2^*)}{\partial x_1} - \frac{\partial g(x_1^*, x_2^*)}{\partial x_1} \cdot \frac{\partial f(x_1^*, x_2^*)}{\partial x_2} \cdot \left[\frac{\partial g(x_1^*, x_2^*)}{\partial x_2}\right]^{-1} = 0.$$

Если ввести величину $y^* = -\frac{\partial f(x_1^*, x_2^*)}{\partial x_2}\left[\frac{\partial g(x_1^*, x_2^*)}{\partial x_2}\right]^{-1}$, то последнее условие можно записать в виде двух равенств

$$\frac{\partial f(x_1^*, x_2^*)}{\partial x_1} + y^* \frac{\partial g(x_1^*, x_2^*)}{\partial x_1} = 0, \qquad \frac{\partial f(x_1^*, x_2^*)}{\partial x_2} + y^* \frac{\partial g(x_1^*, x_2^*)}{\partial x_2} = 0. \tag{1.21}$$

Введем функцию $L(x_1, x_2, y) = f(x_1, x_2) + yg(x_1, x_2)$, тогда из (1.21) получаем

$$\frac{\partial L(x_1^*, x_2^*, y^*)}{\partial x_1} = \frac{\partial L(x_1^*, x_2^*, y^*)}{\partial x_2} = 0. \tag{1.22}$$

Условие (1.22) означает, что $(x_1^*, x_2^*)$ является стационарной точкой функции $L$. Правда оно дает два уравнения, а неизвестных три: $x_1$, $x_2$, $y$. Но у нас имеется еще одно условие $g(x_1^*, x_2^*) = 0$, т.е. получаем три уравнения с тремя неизвестными, которые в принципе позволяют найти решение задачи (1.18).

Рассмотрим теперь задачу с ограничениями типа равенств в общем виде (1.8), (1.17). Введем для нее аналогичную функцию

$$L(x, y) = f(x) + \sum_{i=1}^{m} y_i g_i(x), \tag{1.23}$$

которая называется функцией Лагранжа (здесь $y$ – $m$-мерный вектор, его компоненты называются множителями Лагранжа). Естественно считать, что неизвестных больше, чем связывающих их уравнений, т.е. $n > m$.

**Теорема 1.5** (принцип Лагранжа). *Пусть функции $f(x)$, $g_1(x)$, ..., $g_m(x)$ непрерывно дифференцируемы в некоторой окрестности точки $x^*$, являющейся решением задачи* (1.8), (1.17), *и градиенты $g'_1(x^*)$,..., $g'_m(x^*)$ линейно независимы. Тогда существует вектор $y^*$ такой, что*

$$\frac{\partial L(x^*, y^*)}{\partial x_j} = 0, \ j = 1, ..., n. \tag{1.24}$$

**Доказательство.** Так как градиенты $g'_1(x^*)$,…, $g'_m(x^*)$ линейно независимы, то ранг матрицы

$$\begin{pmatrix} \frac{\partial g_1(x^*)}{\partial x_1} & \dots & \frac{\partial g_1(x^*)}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial g_m(x^*)}{\partial x_1} & \dots & \frac{\partial g_m(x^*)}{\partial x_n} \end{pmatrix}$$

равен $m$. Значит, хотя бы один определитель $m$-го порядка, составленный из ее столбцов, отличен от нуля. Без ограничения общности можно считать, что это определитель из первых $m$ столбцов, т.е.

$$\begin{vmatrix} \frac{\partial g_1(x^*)}{\partial x_1} & \dots & \frac{\partial g_1(x^*)}{\partial x_m} \\ \dots & \dots & \dots \\ \frac{\partial g_m(x^*)}{\partial x_1} & \dots & \frac{\partial g_m(x^*)}{\partial x_m} \end{vmatrix} \neq 0. \tag{1.25}$$

Тогда по известной из анализа теореме о неявной функции в некоторой окрестности точки $x^*$ система равенств $g_i(x) = 0, i = 1,\dots,m$ эквивалентна системе вида

$$x_1 = \varphi_1(x_{m+1},\dots,x_n),\dots, x_m = \varphi_m(x_{m+1},\dots,x_n),$$

т.е. одна часть переменных (в количестве $m$) однозначно выражается через другую часть переменных (в количестве $n$-$m$), причем функции $\varphi_k$ имеют непрерывные частные производные. Подставляя эти функции в $f$, имеем задачу на безусловный экстремум функции $\Psi(x_{m+1},\dots,x_n) = f(\varphi_1(x_{m+1},\dots,x_n),\dots,\varphi_m(x_{m+1},\dots,x_n), x_{m+1},\dots,x_n)$.

Очевидно, точка $(x^*_{m+1},\dots,x^*_n)$ является экстремальной для $\Psi$, поэтому

$$\frac{\partial \Psi(x^*_{m+1},\dots,x^*_n)}{\partial x_j} = 0,\ j = m+1,\dots,n.$$

Вычисляя производные $\Psi$ и приравнивая их к нулю (в точке $x^*$), имеем

$$\frac{\partial \Psi}{\partial x_j} = \frac{\partial f}{\partial x_j} + \sum_{k=1}^{m} \frac{\partial f}{\partial x_k}\frac{\partial \varphi_k}{\partial x_j} = 0, \quad j = m+1,\dots,n. \tag{1.26}$$

Так как $g_i(\varphi_1(x_{m+1},\dots,x_n),\dots,\varphi_m(x_{m+1},\dots,x_n), x_{m+1},\dots,x_n) \equiv 0$, то

$$\frac{\partial g_i}{\partial x_j} + \sum_{k=1}^{m} \frac{\partial g_i}{\partial x_k}\frac{\partial \varphi_k}{\partial x_j} = 0, \quad i = 1,\dots,m, \quad j = m+1,\dots,n\,. \tag{1.27}$$

Умножим равенства (1.27) на $y_i$ и сложим с (1.26):

$$\left(\frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} y_i \frac{\partial g_i}{\partial x_j}\right) + \sum_{k=1}^{m}\left(\frac{\partial f}{\partial x_k} + \sum_{i=1}^{m} y_i \frac{\partial g_i}{\partial x_k}\right)\frac{\partial \varphi_k}{\partial x_j} = 0, \quad j = m+1,\dots,n\,. \tag{1.28}$$

Рассмотрим систему уравнений

$$\frac{\partial f(x^*)}{\partial x_k} + \sum_{i=1}^{m} y_i \frac{\partial g_i(x^*)}{\partial x_k} = 0, \quad k = 1,\dots,m, \tag{1.29}$$

относительно неизвестных $y_i$. Система (1.29) имеет отличный от нуля определитель (1.25), значит, существует ее единственное решение $y^*_1,\dots,y^*_m$. При данных $y^*_i$ из (1.28) и (1.29) следует

$$\frac{\partial f(x^*)}{\partial x_j} + \sum_{i=1}^{m} y_i^* \frac{\partial g_i(x^*)}{\partial x_j} = 0, \quad j = m+1,...,n. \tag{1.30}$$

Но (1.29) при $y_i = y_i^*$ и (1.30) с учетом определения функции Лагранжа эквивалентны (1.24), т.е. теорема доказана.

Условие линейной независимости градиентов $g'_i(x^*)$ называется условием регулярности. Без него принцип Лагранжа в сформулированном виде может не выполняться.

**Пример 1.1.** $f(x) = x_1 \to \text{extr}$, $X = \{x \in R^2 \mid x_1^3 - x_2^2 = 0\}$.

Из ограничения имеем $x_1 = x_2^{2/3}$, поэтому, очевидно, $f(x)$ имеет единственный локальный (и глобальный) минимум в (0, 0), а локальных (и глобальных) максимумов нет. Функция Лагранжа $L = x_1 + y(x_1^3 - x_2^2)$, поэтому $\frac{\partial L}{\partial x_1} = 1 + 3yx_1^2 \neq 0$ при $x_1 = 0$, т.е. принцип Лагранжа не выполняется. Здесь не выполнено условие регулярности, которое в случае одного ограничения превращается в требование неравенства нулю градиента ограничения.

Если градиенты $g'_i(x^*)$ линейно зависимы, то по определению существуют такие $y_i$, не все равные нулю, что

$$\sum_{i=1}^{m} y_i g'_i(x^*) = 0. \tag{1.31}$$

Условие (1.31) также выделяет точки, подозрительные на экстремум. Его можно объединить с (1.24), получив обобщенную форму принципа Лагранжа. Для этого введем обобщенную функцию Лагранжа

$$\tilde{L}(x, y) = y_0 f(x) + \sum_{i=1}^{m} y_i g_i(x).$$

Здесь $y$ уже $(m+1)$–мерный вектор. Условия стационарности для функции $\tilde{L}(x, y)$, т.е. $\frac{\partial \tilde{L}}{\partial x_j} = 0$, $j = 1,...,n$ при $y_0 = 1$ превращаются в (1.24), а при $y_0 = 0$ – в (1.31). Таким образом, можно утверждать, что если $x^*$ – решение задачи (1.8), (1.17), то существует ненулевой вектор $y$ такой, что $x^*$ является стационарной точкой обобщенной функции Лагранжа, причем достаточно ограничиться для $y_0$ значениями 0 и 1. В рассматриваемом выше примере (без условия регулярности) обобщенная функция Лагранжа

$$\tilde{L} = y_0 x_1 + y_1(x_1^3 - x_2^2)$$

и точка (0, 0) является уже для нее стационарной при $y_0 = 0$, $y_1 = 1$.

Принцип Лагранжа (обычный и обобщенный) представляет собой необходимое условие экстремума первого порядка в задаче с ограничениями типа равенств. Однако стационарные точки функции Лагранжа не обязаны быть решениями исходной задачи. Привлекая вторые производные, можно получить необходимые и достаточные условия второго порядка, аналогичные условиям для безусловного экстремума.

Сформулируем достаточные условия максимума в задаче с ограничениями: если функции $f(x)$, $g_1(x),\ldots, g_m(x)$ дважды дифференцируемы в точке $x^*$ и непрерывно дифференцируемы в некоторой ее окрестности, градиенты $g'_1(x^*),\ldots, g'_m(x^*)$ линейно независимы, в точке $x^*$ выполнены условия (1.24) при некоторых $y_i^*$ и, кроме того, для матрицы вторых производных по $x$ функции $L(x, y)$ справедливо $\langle L''_{xx}(x^*, y^*)h, h\rangle < 0$ для любого ненулевого $h\in R^n$, удовлетворяющего условиям $\langle g'_i(x^*), h\rangle = 0$, $i = 1,\ldots,m$, то $x^*$ – решение задачи (1.6), (1.17). Для задачи на минимум указанная квадратичная форма должна быть положительной.

Для проверки отрицательной (положительной) определенности матрицы вторых производных функции Лагранжа на подпространстве векторов, ортогональных всем градиентам ограничений в точке $x^*$, удобно использовать соответствующую форму критерия Сильвестра. Введем, так называемую, окаймленную матрицу Гессе размера $(m+n)\times(m+n)$, состоящую из 4 клеток-матриц вида $\begin{pmatrix} 0 & g' \\ g'^T & L'' \end{pmatrix}$, где левая верхняя матрица размера $m\times m$ нулевая, правая верхняя матрица $m\times n$ составлена из строк – векторов градиентов $g'_1(x^*),\ldots, g'_m(x^*)$, причем нумерация переменных выбрана так, что определитель из первых $m$ столбцов этой матрицы отличен от нуля, нижняя левая – транспонирование предыдущей, нижняя правая – матрица вторых производных функции Лагранжа. Последняя матрица отрицательно определена на соответствующем подпространстве, если знаки $n-m$ главных миноров окаймленной матрицы Гессе размеров от $2m+1$ до $m+n$ чередуются, причем знак «самого маленького» совпадает с $(-1)^{m+1}$, и положительно определена, если знаки этих миноров одинаковы и совпадают с $(-1)^m$.

### 1.4. Интерпретация множителей Лагранжа

Для рассмотрения данного вопроса необходимо в ограничениях $g_i(x)=0$, $i=1, \ldots, m$, исходной задачи выделить константы, т.е. представить ограничения в виде $g_i(x)=b_i$, $i=1, \ldots, m$.

Множители Лагранжа измеряют чувствительность оптимального значения целевой функции $f(x^*)$ к изменениям констант $b$ в ограничениях $g(x)=b$ (или $g_i(x)=b_i$, $i=1, \ldots, m$):

$$y^* = \frac{\partial f(x^*)}{\partial b}, \tag{1.32}$$

т.е. $y_i^* = \dfrac{\partial f(x^*)}{\partial b_i}, \quad i = 1,\ldots,m$. Для доказательства соотношений (1.32) запишем необходимые условия оптимальности в виде системы $m+n$ уравнений с $2m+n$ неизвестными $(b, y, x)$:

$$
\begin{aligned}
& b - g(x) = 0, \\
& \frac{\partial L}{\partial x} = \frac{\partial f(x)}{\partial x} - y\frac{\partial g(x)}{\partial x} = 0.
\end{aligned}
\tag{1.33}
$$

Матрица Якоби для этой системы уравнений имеет вид

$$
\begin{pmatrix}
I & 0 & -\dfrac{\partial g(x)}{\partial x} \\
0 & -\left(\dfrac{\partial g(x)}{\partial x}\right)^T & \dfrac{\partial^2 L}{\partial x^2}
\end{pmatrix},
$$

где $I$ – единичная матрица порядка $m$. Ранг матрицы Якоби равен $m$. Следовательно, по теореме о неявной функции, решая систему (1.33), переменные $x$ и множители Лагранжа $y$ можно представить как функции от констант $b$: $y=y(b)$, $x=x(b)$. Тогда функция Лагранжа зависит от констант ограничений: $L(b) = f(x(b)) + y(b)(b–g(x(b))$. Дифференцирование по $b$ дает

$$\frac{\partial L}{\partial b} = \frac{\partial f(x)}{\partial x}\frac{\partial x}{\partial b} + y(1 - \frac{\partial g(x)}{\partial x}\frac{\partial x}{\partial b}) + \frac{\partial y}{\partial b}(b - g(x)) = (\frac{\partial f(x)}{\partial x} - y\frac{\partial g(x)}{\partial x})\frac{\partial x}{\partial b} + (b - g(x))\frac{\partial y}{\partial b} + y.$$

Первый и второй член этого выражения в точке $(x^*, y^*)$ обращается в нуль согласно условиям (1.33), поэтому $\dfrac{\partial L(x^*, y^*)}{\partial b} = y^*$. Но значение функции Лагранжа в точке $(x^*, y^*)$ есть оптимальное значение целевой функции, т.е. $f(x^*)$. Следовательно, имеем

$$\frac{\partial L(x^*, y^*)}{\partial b} = \frac{\partial f(x^*)}{\partial b} = y^*,$$

что и требовалось доказать.

Таким образом, помимо того, что метод множителей Лагранжа дает решение классической задачи оптимизации, он позволяет проанализировать, насколько оптимальное значение целевой функции чувствительно к изменениям констант ограничений. Например, если какой-то множитель Лагранжа равен нулю, то малые изменения соответствующей константы ограничений не окажут никакого влияния на оптимальное значение целевой функции. Особенно важна интерпретация множителей Лагранжа в задачах экономической деятельности. В экономических задачах распределения ресурсов функция цели имеет размерность стоимости, т.е. цены, умноженной на объем продукции (например, прибыль, выручка, издержки), а с помощью ограничений устанавливается определенное значение некоторого количества (например, затрат). Поскольку в таких задачах с помощью множителя Лагранжа измеряют чувствительность величины, имеющей размерность стоимости, к изменениям некоторого количества, то он имеет размерность цены. По этой причине множитель Лагранжа часто называют *теневой ценой* (данного вида затрат).

### 1.5. Численные методы поиска экстремума без ограничений

Рассмотрим сначала задачу максимизации функции $f(x)$ без ограничений, т.е. в случае, когда $X$ совпадает со всем пространством $R^n$. Необходимые условия оптимальности в этом случае дает теорема 1.2, а достаточные – теорема 1.4. В простых случаях ими можно воспользоваться непосредственно.

**Пример 1.2.** $f(x_1, x_2) = x_1x_2 - x_1^2 - x_2^2 \to \max, x \in R^2$.

Условие первого порядка дает уравнения $x_2 - 2x_1 = 0$, $x_1 - 2x_2 = 0$, откуда $x^*_1 = 0$, $x^*_2 = 0$. Матрица вторых производных $\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}$ отрицательно определена, значит (0, 0) – локальный максимум. Проверяя поведение функции на бесконечности, убеждаемся, что у нее существует глобальный максимум, а глобального минимума нет, значит (0, 0) – глобальный максимум.

Непосредственное решение системы уравнений (1.11) может оказаться чересчур сложным, поэтому на практике используются итеративные градиентные методы. Для задачи на максимум, выбирая произвольную начальную точку $x^{(0)}$, строят итеративный процесс

$$x^{(k+1)} = x^{(k)} + \alpha_k f'(x^{(k)}),\ k=0,1,2,\ldots \tag{1.34}$$

Число $\alpha_k$ называют длиной шага или просто шагом. Если все $\alpha_k$ равны между собой, то имеем процесс с постоянным шагом.

Процесс (1.34), лежащий в основе градиентных методов, представляет собой движение в сторону возрастания функции $f(x)$, так как если градиент в текущей точке $f'(x^{(k)}) \neq 0$, то всегда можно выбрать $\alpha_k$, так, что $f(x^{(k+1)}) > f(x^{(k)})$. Для задачи на минимум, естественно, движение осуществляется в направлении антиградиента.

Существуют разные способы выбора $\alpha_k$. Вообще говоря, наилучшим является выбор такого $\alpha_k$, при котором обеспечивается максимальный рост функции $f(x)$. Такое $\alpha_k$ находится из условия

$$f(x^{(k)} + \alpha_k f'(x^{(k)})) = \max_{\alpha \geq 0} f(x^{(k)} + \alpha_k f'(x^{(k)})). \tag{1.35}$$

Градиентный метод поиска экстремума (1.34) с выбором шага по способу (1.35) называется методом скорейшего подъема (или спуска для задачи на минимум). Такой метод требует наименьшего числа итераций, но зато на каждом шаге приходится решать дополнительную задачу поиска экстремума (правда, в одномерном случае). На практике часто довольствуются нахождением любого $\alpha_k$, обеспечивающего рост функции. Для этого берут произвольное $\alpha_k$ и проверяют условие роста функции, если оно не выполняется, то дробят $\alpha_k$ до тех пор, пока это условие не будет выполнено (такое достаточно малое $\alpha_k$ при $f'(x^{(k)}) \neq 0$ существует всегда).

Возможен выбор постоянного шага $\alpha_k$, обеспечивающего *условие сходимости* метода, которое звучит так: если $f(x)$ принадлежит классу дважды непрерывно дифференцируемых функций на всем пространстве $R^n$, т.е. $f(x)$ непрерывно дифференцируемая функция на $R^n$ и градиент $f'(x)$ удовлетворяет условию Липшица $\| f'(u) - f'(v) \| \leq L \| u - v \|$ с константой $L$, то в (1.34) в качестве $\alpha_k$ может быть взято любое число, удовлетворяющее условию $0 < \alpha_k < \dfrac{2}{L}$ ($\| \cdot \|$ обозначает норму вектора).

Процесс (1.34), очевидно, останавливается, когда выполнено условие (1.13). Так это условие не обеспечивает наличия экстремума, необходимо провести дополнительное исследование функции $f(x)$ в окрестности найденной точки (например, с помощью достаточных условий). Однако даже если найдена точка максимума, определить локальный это максимум или глобальный не всегда возможно. Поэтому градиентные методы в общем случае дают лишь локальные экстремумы (при этом можно попытаться найти глобальный экстремум, применяя итеративный процесс многократно с разными начальными точками). Градиентные методы обеспечивают нахождение глобального экстремума только для вогнутых (выпуклых) функций; если функция $f(x)$ вогнута, то найденная стационарная точка будет глобальным максимумом, а если выпукла – глобальным минимумом (подробнее об этом в разделе «выпуклое программирование»).

Как правило, итеративный процесс (1.34) бесконечный, поэтому необходимо ввести правило останова. Для этого задается требуемая точность решения $\varepsilon$, но так как само решение неизвестно заранее, то процесс останавливают, когда либо расстояние между двумя последовательными точками становится меньше $\varepsilon$, либо градиент по модулю меньше $\varepsilon$.

**Пример 1.3.** Найти градиентным методом решение задачи из примера 1.2 с точностью $\varepsilon$=0,25 и начальной точкой $x^{(0)}$=(1,2).

Градиент функции равен $f'(x) = (x_2 - 2x_1,\ x_1 - 2x_2)$, а в начальной точке имеем $f'(x^{(0)})$=(0,–3). Согласно (1.34) следующая точка при произвольном шаге $\alpha$ есть (1, 2–3$\alpha$). Используем метод наискорейшего подъема. Находим максимум по $\alpha$ выражения $f(1, 2-3\alpha) = -9\alpha^2+9\alpha-3$, получаем $\alpha$=0.5 и следующая точка (1 , 0.5). В этой точке градиент функции равен (–1.5 , 0). Находим максимум по $\alpha$ выражения $f(1-1.5\alpha\ , 0.5)=$ $-2{,}25\alpha^2+2{,}25\alpha-0{,}75$ , получаем $\alpha$=0,5 и следующая точка (0.5 , 0.5). Расстояние между двумя последовательными точками становится меньше $\varepsilon$ на 4 шаге, поэтому если оно берется в качестве правила останова, то полученная на 4 шаге точка (0.0625 , 0.125) принимается за приближенное решение задачи.

Градиентные методы – это методы первого порядка, т.е. методы поиска экстремума, использующие первые производные. Фактически в этих методах производится линеаризация, связанная с заменой максимизируемой функции $f(x)$ линейным членом разложения ее в ряд Тейлора. Но если функция дважды непрерывно дифференцируема, то можно использовать для ее аппроксимации два члена ряда Тейлора, т.е. вторые производные. Использование такой аппроксимации в итеративном процессе может повысить скорость сходимости.

Наиболее известным из методов второго порядка является метод Ньютона, который состоит в следующем. На $k$-й итерации, имея приближение $x^{(k)}$, составим вспомогательную задачу максимизации функции

$$\varphi_k(x) = \langle f'(x^{(k)}), x - x^{(k)} \rangle + \frac{1}{2}\langle f''(x^{(k)})(x - x^{(k)}), x - x^{(k)} \rangle$$

на множестве $X$. Пусть

$$\varphi_k(\tilde{x}^{(k)}) = \max_{x \in X} \varphi_k(x),$$

тогда следующее приближение строится по формуле

$$x^{(k+1)} = x^{(k)} + \alpha_k(\tilde{x}^{(k)} - x^{(k)}), \quad 0 \le \alpha_k \le 1.$$

Шаг $\alpha_k$ можно выбирать разными способами, в частности, можно положить $\alpha_k$=1, тогда в качестве следующего приближения принимается просто решение вспомогательной задачи, т.е. $x^{(k+1)} = \tilde{x}^{(k)}$.

Так как в задаче без ограничений $X=R^n$, то имеем условие максимума $\varphi'_k(\tilde{x}^{(k)}) = 0$, откуда получается

$$f'(x^{(k)}) + f''(x^{(k)})(\tilde{x}^{(k)} - x^{(k)}) = 0. \qquad (1.36)$$

Значит, для нахождения $\tilde{x}^{(k)}$ надо решить систему линейных уравнений (1.36). Если матрица $f''(x^{(k)})$ невырождена, то при $\alpha_k = 1$ получаем следующий итеративный процесс

$$x^{(k+1)} = \tilde{x}^{(k)} = x^{(k)} - (f''(x^{(k)}))^{-1} \cdot f'(x^{(k)}). \qquad (1.37)$$

Достоинством метода Ньютона является высокая скорость сходимости, которая полностью компенсирует усложнение вычислений на каждой итерации. Недостатком метода является то, что теоретически он сходится при весьма жестких предположениях относительно начальной точки (достаточная ее близость к решению) и целевой функции (сильная выпуклость), хотя на практике, как правило, дело обстоит лучше.

**Пример 1.4.** Найти минимальное значение функции $f = (x_1-2)^4+(x_1-2x_2)^2$ с точностью $\varepsilon = 0{,}05$ и начальной точкой $x^{(0)}=(0;3)$.

Используем итерационную формулу метода Ньютона (1.37) и правило останова, когда градиент по модулю меньше $\varepsilon$, т.е. $\left\| f'(x^{(k)}) \right\| < \varepsilon$.

Вычислим $f(x^{(0)}) = 52$, $f'(x^{(0)}) = (-44{,}24)$, $f''(x^{(0)}) = \begin{pmatrix} 50 & -4 \\ -4 & 8 \end{pmatrix}$,

$(f''(x^{(0)}))^{-1} = \dfrac{1}{384}\begin{pmatrix} 8 & 4 \\ 4 & 50 \end{pmatrix}$, $(f''(x^{(0)}))^{-1} \cdot f'(x^{(0)}) = (-0.67,\, 2.67)$.

Теперь находим

$$x^{(1)} = x^{(0)} - (f''(x^{(0)}))^{-1} \cdot f'(x^{(0)}) = (0.67, 0.33).$$

Вычислим $f(x^{(1)}) = 3.13$, $f'(x^{(1)}) = (-9.4, -0.04)$, $f''(x^{(1)}) = \begin{pmatrix} 23{,}2 & -4 \\ -4 & 8 \end{pmatrix}$,

$(f''(x^{(1)}))^{-1} = \dfrac{1}{170}\begin{pmatrix} 8 & 4 \\ 4 & 23.2 \end{pmatrix}$, $x^{(2)} = x^{(1)} - (f''(x^{(1)}))^{-1} \cdot f'(x^{(1)}) = (1.11,\, 0.56)$ и так далее. Требуемая точность получается на 6-ом шаге:

$$\tilde{x} = x^{(6)} = (1.83,\, 0.91), \quad \left\| f'(x^{(6)}) \right\| = 0.04 < 0.05, \quad f(\tilde{x}) = 0.0009.$$

Таким образом, найдено $f_{min} = f(1/83,\, 0/91) = 0/0009$. Точное значение минимума равно 0, а точка минимума (2, 1). Как видим, начальное приближение было весьма далеким.

Для задачи с ограничениями типа равенств основным методом решения является метод множителей Лагранжа. Принцип Лагранжа дает нам $n+m$ уравнений ($n$ уравнений (1.24) и $m$ уравнений в определении (1.17) множества $X$) для нахождения $n+m$ неизвестных $x_1,\ldots, x_n, y_1,\ldots, y_m$.

**Пример 1.5.** $f(x) = x_1x_2x_3 \to \text{extr}$,

$$X = \{x \in R^3 \mid x_1 + x_2 + x_3 - 1 = 0,\ x_1^2 + x_2^2 + x_3^2 - 1 = 0\}.$$

Градиенты ограничений (1, 1, 1) и ($2x_1$, $2x_2$, $2x_3$) линейно независимы на $X$ (они зависимы при $x_1 = x_2 = x_3$, но точки такого вида не принадлежат $X$). Поэтому принцип Лагранжа справедлив. Функция Лагранжа имеет вид

$$L = x_1x_2x_3 + y_1(x_1 + x_2 + x_3 - 1) + y_2(x_1^2 + x_2^2 + x_3^2 - 1).$$

Имеем пять уравнений

$$x_2x_3 + y_1 + 2y_2x_1 = 0,\quad x_1x_3 + y_1 + 2y_2x_2 = 0,$$

$$x_1x_2 + y_1 + 2y_2x_3 = 0,\quad x_1 + x_2 + x_3 = 1,\quad x_1^2 + x_2^2 + x_3^2 = 1.$$

Эта система имеет две группы симметричных решений (по три точки):

$$x_i^* = x_j^* = \frac{2}{3},\ x_k^* = -\frac{1}{3},\ y_1^* = -\frac{2}{9},\ y_2^* = \frac{1}{3};$$

$$x_i^* = x_j^* = 0,\ x_k^* = 1,\ y_1^* = y_2^* = 0.$$

Для первой значение функции равно $-\frac{4}{27}$, для второй 0. Так как для данной задачи выполнены условия теоремы 1.1, то существуют глобальные максимум и минимум. Поскольку они находятся среди локальных, можно сделать вывод, что

$$\max_{x\in X} f(x) = 0,\ \min_{x\in X} f(x) = -\frac{4}{27},$$

а все три локальных максимума и три локальных минимума являются и глобальными.

**Пример 1.6.** $f(x) = x_1^2 + \frac{1}{2}x_2^2 \to \text{extr}$, $X = \{x \in R^2 \mid x_1^3 + x_2^3 - 1 = 0\}$.

Здесь условие регулярности выполняется, т.к. $g'(x) = (3x_1^2,\ 3x_2^2) \neq 0$ на $X$. Функция Лагранжа $L = x_1^2 + \frac{1}{2}x_2^2 + y(x_1^3 + x_2^3 - 1)$. Имеем систему из трех уравнений $2x_1 + 3yx_1^2 = 0$, $x_2 + 3yx_2^2 = 0$, $x_1^3 + x_2^3 = 1$. Она имеет три решения:

1) $x_1^* = 1$, $x_2^* = 0$, $y^* = -\frac{2}{3}$, 2) $x_1^* = 0$, $x_2^* = 1$, $y^* = -\frac{1}{3}$, 3) $x_1^* = \frac{2}{\sqrt[3]{9}}$, $x_2^* = \frac{1}{\sqrt[3]{9}}$, $y^* = -\frac{\sqrt[3]{9}}{3}$.

Найдем матрицу вторых производных функции Лагранжа

$$L''_{xx} = \begin{pmatrix} 2 + 6yx_1 & 0 \\ 0 & 1 + 6yx_2 \end{pmatrix}$$

и проверим достаточные условия для каждого решения.

1) $L''_{xx} = \begin{pmatrix} -2 & 0 \\ 0 & 1 \end{pmatrix}$, $g' = (3,0)$, если $\langle g', h\rangle = 0$, то $h_2$ – любое, $h_1 = 0$,

$\langle L''_{xx}h, h\rangle = h_2^2 > 0$, значит, (1, 0) – точка локального минимума со значением $f = 1$.

2) $L''_{xx} = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix}$, $g' = (0,3)$, если $\langle g', h\rangle = 0$, то $h_1$ –любое, $h_2 = 0$,

$\langle L''_{xx}h, h\rangle = 2h_1^2 > 0$, значит, (0,1) – точка локального минимума со значением $f = \frac{1}{2}$.

3) $L''_{xx} = \begin{pmatrix} -2 & 0 \\ 0 & -1 \end{pmatrix}$, $\langle L''_{xx}h, h\rangle = -2h_1^2 - h_2^2 < 0$ для любого $h \neq 0$, значит, $\left(\frac{2}{\sqrt[3]{9}}, \frac{1}{\sqrt[3]{9}}\right)$ – точка локального максимума со значением $f = \frac{\sqrt[3]{9}}{2}$.

Проверим теперь достаточность с использование окаймленной матрицы Гессе, которая имеет вид $\begin{pmatrix} 0 & 3x_1^2 & 3x_2^2 \\ 3x_1^2 & 2+6yx_1 & 0 \\ 3x_2^2 & 0 & 1+6yx_2 \end{pmatrix}$ и равна $\begin{pmatrix} 0 & 3 & 0 \\ 3 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ для первой точки. Здесь $n-m=1$, поэтому надо вычислить только определитель матрицы Гессе. В данной точке он равен $-9$, т.е. совпадает по знаку с $(-1)^m$, значит это точка локального минимума. Аналогично проверяются условия для остальных точек. Что касается поведения функции на бесконечности, то она ограничена снизу и не ограничена сверху, поэтому глобального максимума нет, а (0, 1) – точка глобального минимума

Конечно, непосредственное решение системы (1.17), (1.24) не всегда просто. Поэтому на практике используются различные численные реализации метода множителей Лагранжа, а также и другие численные методы решения экстремальных задач с ограничениями. Об этих методах пойдет речь в главе «Нелинейное программирование».

### 1.6. Примеры применения методов классической оптимизациии

В различных разделах исследования операций возникают проблемы, связанные с решением экстремальных задач. Естественно, что для этих целей могут быть использованы изложенные выше методы. Однако область их применения не ограничивается только такими задачами. Часто проблема в первоначальном виде не связана явно с какой-либо экстремальной задачей, но может быть переформулирована так, что для ее решения оказываются пригодными методы поиска экстремума. Рассмотрим некоторые примеры.

*Неравенство между средним арифметическим и средним квадратическим*

Средним арифметическим чисел $x_1, \ldots, x_n$ называется, как известно, величина $\frac{1}{n}\sum_{i=1}^{n} x_i$, а средним квадратическим – величина $(\frac{1}{n}\sum_{i=1}^{n} x_i^2)^{0.5}$. Для доказательства неравенства

$$\frac{1}{n}\sum_{i=1}^{n} x_i \le \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{0.5}$$

рассмотрим задачу поиска максимума функции $f(x) = \sum_{i=1}^{n} x_i$ при ограничении $\sum_{i=1}^{n} x_i^2 - B = 0$. Эта задача регулярна при $B \neq 0$ (градиент ограничения не равен нулю) и в соответствии с принципом Лагранжа ее решение удовлетворяет системе уравнений

$$1 + 2yx_i = 0, \quad i = 1,\ldots,n, \quad \sum_{i=1}^{n} x_i^2 = B.$$

Отсюда все $x_i^*$ равны между собой и $x_i^* = \pm\sqrt{\frac{B}{n}}$. Глобальный максимум и минимум функции $f(x)$ на сфере существуют по теореме Вейерштрасса. Точка максимума $x_i^* = \sqrt{\frac{B}{n}}$, $i = 1,\ldots,n,$ а другая точка – минимум. Значит

$$\frac{1}{n} f(x) = \frac{1}{n}\sum_{i=1}^{n} x_i \le \sqrt{\frac{B}{n}} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^{0.5},$$

что и требовалось доказать (заметим, что неравенство превращается в равенство только в случае, когда все $x_i$ равны между собой).

На этой задаче мы продемонстрировали общий прием для доказательства неравенств: одна часть неравенства приравнивается к произвольной величине $B$ и это равенство берется в качестве ограничения, а другая часть неравенства выступает в роли целевой функции, для которой в зависимости от знака неравенства ищется максимум или минимум, найденное экстремальное значение соответственно не больше или не меньше $B$. Таким стандартным приемом можно доказать все известные неравенства (например, Коши-Буняковского, Гельдера, со всевозможными средними), а также выводить новые неравенства.

*Метод наименьших квадратов*

Имеются результаты измерений значений параметров на входе $x$ и на выходе $y$ некоторого устройства:

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

По этим данным нужно восстановить функциональную зависимость $y=f(x)$, описывающую работу устройства. Уточним постановку задачи. Пусть априори известен класс возможных функциональных зависимостей $f(x,d_1,\ldots,d_k)$ с точностью до параметров $d_1,\ldots,d_k$. По методу наименьших квадратов конкретное значение вектора параметров $d = (d_1,\ldots,d_k)$ выбирается так, чтобы минимизировать сумму квадратов невязок в измерениях, т.е. величину

$$\delta(d) = \sum_{i=1}^{n}(y_i - f(x_i, d_1, ..., d_k))^2 .$$

Рассмотрим для простоты линейную зависимость $f(x, a, b) = ax + b$. Тогда

$$\delta(a, b) = \sum_{i=1}^{n}(y_i - ax_i - b)^2$$

и надо найти абсолютный минимум этой величины как функции аргументов $a$ и $b$. Условие экстремума

$$\frac{\partial\delta}{\partial a} = \frac{\partial\delta}{\partial b} = 0$$

дает систему из двух линейных уравнений

$$\begin{cases} a\left(\sum_{i=1}^{n} x_i^2\right) + b\left(\sum_{i=1}^{n} x_i\right) = \sum_{i=1}^{n} x_i y_i, \\ a\left(\sum_{i=1}^{n} x_i\right) + bn = \sum_{i=1}^{n} y_i. \end{cases}$$

Определитель этой системы

$$D = n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2.$$

Естественно считать, что не все $x_i$ равны между собой, поэтому из доказанного неравенства между средним арифметическим и средним квадратическим имеем $D > 0$. Значит, система имеет единственное решение

$$a^* = D^{-1}\left[n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)\right],$$

$$b^* = D^{-1}\left[\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)\right].$$

Матрица вторых производных $\delta$ есть

$$\delta'' = \begin{pmatrix} 2\sum_{i=1}^{n} x_i^2 & 2\sum_{i=1}^{n} x_i \\ 2\sum_{i=1}^{n} x_i & 2n \end{pmatrix}.$$

Ее главный минор первого порядка равен $2\sum_{i=1}^{n} x_i^2$, а главный минор второго порядка равен $4D$, т.е. они положительны (если не все $x_i$ равны между собой). Значит, по критерию Сильвестра матрица $\delta''$ положительно определена и в силу теоремы о достаточных условиях экстремума точка ($a^*$, $b^*$) доставляет минимум величине $\delta$ (по крайней мере, локальный).

Но так как выполнены условия следствия из теоремы Вейерштрасса, существует глобальный минимум $\delta$, а в силу единственности экстремума пара ($a^*$, $b^*$) является точкой глобального минимума $\delta$.

*Мера информации*

В теории информации количество (меру) информации в сообщении о значении величины, которая может принимать априори $n$ значений с вероятностями $p_1,\ldots,\ p_n$, принято определять величиной $H = -\sum_{i=1}^{n} p_i \log_2 p_i$.

Для этого имеется ряд оснований, связанных со свойствами функции $H$ (обычно эти свойства постулируются как аксиомы и однозначно определяют вид $H$). В частности, количество информации в сообщении о значении величины, принимающей априори два значения с равными вероятностями, равно единице (оно и принимается за единицу измерения информации, называемую битом). Для представления единицы информации достаточно одного разряда ячейки памяти машины, работающей с двоичными числами. Поставим вопрос: сколько двоичных разрядов достаточно для представления информации о величине, принимающей $n$ возможных значений с неизвестными вероятностями. Другими словами, какое максимальное количество информации может содержать сообщение о величине с $n$ возможными значениями. Для ответа на этот вопрос надо найти максимум функции $H$ при ограничении $\sum_{i=1}^{n} p_i - 1 = 0$ (вообще говоря, еще надо учитывать, что $p_i \ge 0,\quad i = 1,\ldots,n$). Функция $H$ не определена при $p_i = 0$, поэтому доопределим ее по непрерывности при $p_i \to 0$. Используя правило Лопиталя, нетрудно получить $\lim_{p_i \to 0}\left(p_i \log_2 p_i\right) = 0$, поэтому при $p_{i_0} = 1,\ p_i = 0,\ i \ne i_0$ полагаем $H = 0$ (так как $H \ge 0$, то это минимальное значение). Для нахождения максимума составим функцию Лагранжа

$$L = -\sum_{i=1}^{n} p_i \log_2 p_i + y\left(\sum_{i=1}^{n} p_i - 1\right).$$

Получим систему уравнений

$$\frac{\partial L}{\partial p_i} = -\log_2 p_i - \log_2 e + y = 0,\ i = 1,\ldots,n,\ \sum_{i=1}^{n} p_i = 1,$$

откуда видно, что все $p_i$ равны между собой, т.е. $p_i^* = \frac{1}{n},\quad i = 1,\ldots,n.$

Для нахождения глобального максимума, который существует в силу теоремы Вейерштрасса, надо это решение сравнить с граничными точками $p_i$=0, но на границах $H$=0.

Значит, очевидно, точка $\left(\frac{1}{n},...,\frac{1}{n}\right)$ доставляет глобальный максимум $H$ со значением $\log_2 n$. Это и есть искомое максимальное количество информации. Например, при $n$=8 требуется три разряда.

### Задачи и упражнения

**1**. Найти оптимальное распределение ограниченного ресурса в количестве $a$ между $n$ производственными процессами, если прибыль, получаемая при переработке $j$-м процессом $x_j$ единиц ресурса, вычисляется по формуле $c_j\sqrt{x_j}$ .

**2**. Завод А расположен на расстоянии $a$ км от железной дороги, идущей в город В, и на расстоянии $b$ км. от города В. Под каким углом к железной дороге следует провести шоссе с завода А, чтобы доставка грузов была наиболее дешевой, если стоимость перевозки одинакового количества груза на одинаковое расстояние по шоссе в $k$ раз дороже, чем по железной дороге?

**3**. Найти прямоугольник наименьшего периметра, ограничивающий заданную площадь, используя теорему о среднем.

**4**. Составляется электрическая цепь из двух параллельно соединенных сопротивлений. При каких величинах этих сопротивлений сопротивление всей цепи максимально, если при последовательном их соединении оно равно $R$?

**5**. Решить задачу $f(x)= 4x_1+6x_2-x_1^2-x_2^2 \to \max$ методом наискорейшего подъема, начиная процесс с точки $x^{(0)}$=(0, 0). Какова длина шага $\alpha$ на первой итерации.

**6.** По плану производства продукции предприятию необходимо изготовить 180 тонн продукции, которые могут быть изготовлены двумя технологическими способами. При производстве $x_1$ тонн продукции первым способом затраты составляют $4x_1+x_1^2$ тыс. у.е., а при изготовлении $x_2$ тонн продукции вторым способом они равны $8x_2+ x_2^2$ у.е. Определить сколько продукции каждым из способов следует изготовить, так чтобы общие затраты на производство продукции были минимальны.

**7.** Производственная функция фирмы имеет вид $f(x)=x_1^{0.25}x_2^{0.5}$, лимит на покупку ресурсов равен 18 денежных единиц, цена на ресурс $x_1$ равна 3, а на ресурс $x_2$ равна 4 денежным единицам. Решите задачу максимизации выпуска фирмы $f(x)$, используя условия первого и второго порядков для функции Лагранжа.

**8.** Точка $x^*\in X$ называется точкой глобального максимума функции $f(x)$ на множестве $X$, если

а. $f(x^*)\geq f(x)\ \ \forall\, x\in X$ б. $f(x^*) > f(x)\ \ \forall\, x\in X$ в. $\exists x\in X\ \ f(x^*)\geq f(x)$

**9.** Точки глобального максимума и минимума $f(x)$ на множестве $X$ существуют, если

а. $X$ – ограниченное множество в $R^n$, $f(x)$ – непрерывная на $X$ функция

б. $X$ – замкнутое ограниченное множество в $R^n$, $f(x)$ – непрерывная на $X$ функция

в. $X$ – замкнутое непустое множество в $R^n$, $f(x)$ – непрерывная на $X$ функция

**10**. Исследовать на знакоопределенность матрицу $\begin{pmatrix} 2 & -1 & 2 \\ 1 & 0 & 1 \\ -5 & 1 & 3 \end{pmatrix}$

а. отрицательно определена

б. положительно определена

в. не является знакоопределенной

**11.** Градиент функции $z(x,y)=y-\sin x$ равен

а. $(1,\cos x)$ б. $(1,-\cos x)$ в. $(-\cos x,1)$

**12.** Найти локальные экстремумы функции $f(x,y)=x^3+y^2+2xy$

а.$(0,0)$, $(\frac{2}{3},-\frac{2}{3})$ – точки локального минимума

б. $(0,0)$, $(\frac{2}{3},-\frac{2}{3})$ – точки локального максимума

в. $(\frac{2}{3},-\frac{2}{3})$ – точка локального минимума, в точке $(0,0)$ экстремума нет

**13.** Условием регулярности в задаче на экстремум функции $f(x)$ с ограничениями типа равенств является

а. линейная независимости градиентов функций ограничений в оптимальной точке

б. линейная зависимость градиентов функций ограничений в оптимальной точке

в. равенство нулю производных функции Лагранжа в оптимальной точке

**14.** Определить, выполняется ли условие регулярности для задачи оптимизации с ограничениями $x+2y+z=1$, $4y+2z=3$

а. выполняется в любой точке

б. нигде не выполняется

в. выполняется в некоторых точках

# Глава 2. НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

## 2.1. Общая задача математического программирования

Во многих задачах оптимизации, связанных с проблемами принятия решений в технике, экономике, экологии, финансах и т.д., связи являются нелинейными. Таковы, например, задачи увеличения масштабов производства, перехода на новую технологию, оптовой торговли и т.д. Поэтому ясна необходимость изучения нелинейных моделей принятия решений и методов их анализа. Здесь мы остановимся на случае ограничений типа неравенств, которые изучаются в рамках математического программирования.

**Общей задачей математического программирования** называется задача условной оптимизации вида

$$f(x) \to \max,$$

при ограничениях $$g_i(x) \geq 0, i = 1,\ldots,m, \tag{2.1}$$

$$x \in P, \tag{2.2}$$

где $P$ – некоторое непустое подмножество $n$-мерного евклидова пространства. В качестве $P$ обычно рассматривается множество простой структуры, например, координатный параллелепипед, частные случаи которого есть $P = R_+^n$ или $P = R^n$. Условия типа $g_i(x) \leq 0$ называют ограничениями-неравенствами; условие $x \in P$ называется прямым ограничением.

Если ввести множество

$$X = \{x \mid x \in P, g_i(x) \geq 0, i = 1,\ldots,m\}, \tag{2.3}$$

то кратко эту задачу можно записать как определение величины глобального максимума функции и точки его достижения (в предположении ее существования)

$$\bar{v} = \max_{x \in X} f(x) = f(x^0). \tag{2.4}$$

Разделение ограничений на (2.1) и (2.2) не носит принципиального характера, так как каждое из них может быть записано в одном из этих двух видов, но иногда оказывается полезным. Так в задаче линейного программирования записывается отдельно условие неотрицательности $x \geq 0$, хотя это просто частный вид линейного ограничения. Кстати, нетрудно видеть, что задача линейного программирования есть частный случай задачи (2.4). Действительно, если $f(x)$ и $g_i(x)$ линейные функции, $P$ – неотрицательный ортант, а (2.1) переписать в виде обратных неравенств, то получим задачу линейного программирования в стандартной форме.

Точку $x^0$ называют решением, или точкой глобального максимума, или оптимальным планом, множество $X$ – допустимым множеством или множеством допустимых планов (вообще говоря, оно может быть и пустым), его точки – допустимыми точками или допустимыми планами.

Вообще говоря, в задаче могут быть одновременно ограничения типа равенств и неравенств. Однако такую задачу можно преобразовать к указанному виду. Действительно, ограничение вида равенства $g(x)=0$ можно заменить двумя ограничениями вида неравенств $g(x)\geq 0,\ -g(x)\geq 0$, а ограничение вида $g(x)\leq 0$ можно заменить на $-g(x)\geq 0$, т.е. система ограничений-неравенств задачи (2.4) имеет общий вид. Практически при решении нелинейных задач равенства на неравенства не заменяют, а используют, например, известный из курса математического анализа метод множителей Лагранжа.

Для задачи (2.4) можно построить функцию Лагранжа (по аналогии со случаем ограничений типа равенств):

$$F(x,y)=f(x)+\sum_{i=1}^{m} y_i g_i(x), \tag{2.5}$$

где $y=(y_i,...,y_m)$ – вектор множителей Лагранжа. Однако применительно к задаче (2.4) метод множителей Лагранжа нуждается в некоторой модификации, для описания которой нам понадобятся новые важные понятия. Кроме того, метод множителей Лагранжа может быть усилен для некоторых частных случаев общей задачи математического программирования. Важнейшими такими случаями являются линейное программирование, для которого, как будет показано, метод Лагранжа уже фактически был обоснован и использован, и выпуклое программирование, которое будет далее рассматриваться. Кроме того, мы познакомимся с основными численными методами решения задач математического программирования.

Так как многие задачи принятия решений предполагают наличие условий неотрицательности на переменные, то рассмотрим метод множителей Лагранжа для задачи условной оптимизации (2.4), в которой $P=R_+^n$.

Метод множителей Лагранжа заключается в замене задачи условной максимизации функции $f(x)$ безусловной оптимизацией функции Лагранжа. Введем множество $I(x)=\{i\,|\,1\leq i\leq m, g_i(x)=0\}$, представляющее собой совокупность индексов активных в точке $x$ ограничений-неравенств; множество индексов активных простых ограничений $J(x)=\{j\,|\,1\leq j\leq n, x_j=0\}$; и $n$-мерные вектора $e_j=(0,\ldots,0,1,0,\ldots,0)$, $j=1,\ldots,n$, у которых $j$-я компонента – единица, а остальные – нули.

**Теорема 2.1.** *Если* $x^0$ *– локальное решение задачи математического программирования*

$$f(x) \to \max, \quad g_i(x) \ge 0, i = 1,...,m, \quad x \ge 0,$$

*функции* $f(x)$, $g_i(x)$, $i=1,\dots,m$, *непрерывно дифференцируемы в точке* $x^0$, *градиенты* $g_i'(x^0), i \in I(x^0)$ *и вектора* $e_j,\ j \in J(x^0)$ *линейно независимые, то найдется вектор* $y^0 \ge 0$ такой, что

$$\begin{aligned} &\frac{\partial F(x^0, y^0)}{\partial x_j} \le 0, \qquad x_j^0 \frac{\partial F(x^0, y^0)}{\partial x_j} = 0,\ j = 1,...,n, \\ &g_i(x^0) \ge 0, \qquad y_i^0 g_i(x^0) = 0,\ i = 1,...,m. \end{aligned} \tag{2.6}$$

***Замечание***. Если задача математического программирования сформулирована без условий неотрицательности, то условия оптимальности примут вид

$$\frac{\partial F(x^0, y^0)}{\partial x_j} = 0,\ j = 1,...,n, \quad g_i(x^0) \ge 0, \quad y_i^0 g_i(x^0) = 0, \quad y_i^0 \ge 0,\ i = 1,...,m. \tag{2.7}$$

Условия (2.6) называются **условиями Куна-Таккера**. Теорема 2.1. и Замечание к ней утверждают, что любое локально решение задачи является стационарной точкой, обратное гарантируется при некоторых предположениях только для задачи выпуклого программирования, к рассмотрению которой мы и переходим.

## 2.2. Седловые точки и двойственность

Важную роль в математическом программировании играет понятие седловой точки, которая определяется следующим образом.

**Определение 2.1.** Пара $(x^0, y^0)$ называется седловой точкой функции $F(x, y)$ на прямом произведении множеств $X \times Y$, если

$$F(x, y^0) \le F(x^0, y^0) \le F(x^0, y) \quad \forall x \in X, y \in Y$$

или эквивалентно

$$\max_{x \in X} F(x, y^0) = F(x^0, y^0) = \min_{y \in Y} F(x^0, y). \tag{2.8}$$

Основными свойствами седловых точек является взаимозаменяемость и эквивалентность. Если $(x^1, y^1)$ и $(x^2, y^2)$ – седловые точки, то $(x^1, y^2)$ и $(x^2, y^1)$ – также седловые точки (взаимозаменяемость), при этом

$$F(x^1, y^1) = F(x^1, y^2) = F(x^2, y^1) = F(x^2, y^2)$$

(эквивалентность). Доказательство этих фактов непосредственно вытекает из определения, и мы его оставляем в качестве упражнения.

Аргументы $x$ и $y$ функции $F(x, y)$ мы будем считать векторами соответственно $n$–мерного и $m$–мерного евклидовых пространств. Таким образом, из определения седловой точки следует, что в этой точке по одной группе переменных функция достигает максимума, а по другой – минимума. Хотя в этом состоит существенное отличие седловых точек от точек максимума или минимума функций, условия для их определения аналогичны условиям экстремума. Так, если $(x^0, y^0)$ является внутренней точкой произведения $X\times Y$ и функция $F(x, y)$ непрерывно дифференцируема по всем переменным, то в седловой точке $(x^0, y^0)$ необходимо выполнение условий

$$\frac{\partial F(x^0, y^0)}{\partial x_j} = \frac{\partial F(x^0, y^0)}{\partial y_i} = 0, \quad j = 1,..., n, \quad i = 1,..., m .$$

Среди решений этой системы уравнений могут быть и точки максимума, и точки минимума, и седловые точки, и другие критические точки (например, точки перегиба).

Возникает вопрос: если в седловой точке функция $F(x, y)$ достигает максимума по $x$ и минимума по $y$, то нельзя ли находить эти точки последовательным применением операций поиска максимума и минимума функций. Если мы берем минимум функции $F(x, y)$ по $y$, то получаем функцию от $x$: $\varphi(x) = \min\limits_{y\in Y} F(x, y).$ У этой функции можно брать максимум по $x$, в результате получается величина (в предположении достижимости верхней и нижней граней)

$$v_1 = \max_{x\in X} \varphi(x) = \max_{x\in X} \min_{y\in Y} F(x, y). \qquad (2.9)$$

Величина $v_1$ называется максимином функции $F(x, y)$, а задача ее определения называется максиминной задачей.

Аналогично определяется минимакс функции $F(x, y)$:

$$v_2 = \min_{y\in Y} \Psi(y) = \min_{y\in Y} \max_{x\in X} F(x, y), \qquad (2.10)$$

где $\Psi(y) = \max\limits_{x\in X} F(x, y).$

В общем случае, когда соответствующие верхняя и нижняя грани не обязательно достигаются, можно определить величины

$$v = \sup_{x\in X} \inf_{y\in Y} F(x, y), \qquad (2.11)$$

$$\tilde{v} = \inf_{y\in Y} \sup_{x\in X} F(x, y), \qquad (2.12)$$

которые превращаются в $v_1$ и $v_2$ при достижимости всех граней. Вообще говоря, существенным является достижимость только верхней грани в (2.11) и нижней грани в (2.12), эти грани называются наружными; если они достигаются, то соответствующие точки реализации верхней грани в (2.11) и нижней грани в (2.12) называются решениями максиминной и минимаксной задач. Теперь наш вопрос можно переформулировать так: имеют ли отношение решения максиминной и минимаксной задач к седловым точкам? Оказывается имеют и самое непосредственное. Но прежде чем показать это, докажем одну простую, но важную лемму.

**Лемма 2.1.** *Величины* $v$ *и* $\tilde{v}$*, определенные выражениями* (2.11), (2.12), *всегда связаны соотношением*

$$v \le \tilde{v},$$

*т.е. имеет место неравенство*

$$\sup_{x\in X}\inf_{y\in Y} F(x,y) \le \inf_{y\in Y}\sup_{x\in X} F(x,y). \tag{2.13}$$

**Доказательство**. Из определений верхней и нижней граней следуют следующие неравенства

$$\inf_{y\in Y} F(x,y) \le F(x,y') \quad \forall y' \in Y,$$

$$v = \sup_{x\in X}\inf_{y\in Y} F(x,y) \le \sup_{x\in X} F(x,y') \quad \forall y' \in Y,$$

откуда $v \le \inf_{y\in Y}\sup_{x\in X} F(x,y) = \tilde{v}$, что и требовалось доказать.

В (2.13) может быть как равенство, так и строгое неравенство. Это видно из следующих простых примеров.

**Пример 2.1.** $F(x,y) = (x-y)^2$, $0 \le x \le 1$, $0 \le y \le 1$.

Здесь $v = \max\limits_{0\le x\le 1}\min\limits_{0\le y\le 1}(x-y)^2 = 0$, $\quad \tilde{v} = \min\limits_{0\le y\le 1}\max\limits_{0\le x\le 1}(x-y)^2 = \min\{\min\limits_{0\le y\le \frac{1}{2}}(1-y)^2, \min\limits_{\frac{1}{2}\le y\le 1} y^2\} = \frac{1}{4}$,

т.е. имеет место неравенство $v < \tilde{v}$.

**Пример 2.2.** $F(x,y) = (x-y)$, $0 \le x \le 1$, $0 \le y \le 1$.

Здесь $v = \max\limits_{0\le x\le 1}\min\limits_{0\le y\le 1}(x-y) = \max\limits_{0\le x\le 1}(x-1) = 0$, $\quad \tilde{v} = \min\limits_{0\le y\le 1}\max\limits_{0\le x\le 1}(x-y) = \min\limits_{0\le y\le 1}(1-y) = 0$, т.е. имеет место равенство $v = \tilde{v}$.

Оказывается, что существование седловых точек связано с двумя моментами: достижимостью наружных граней в (2.11) и (2.12), т.е. существованием решений $x^0$ и $y^0$ максиминной и минимаксной задач, и выполнением равенства $v = \tilde{v}$. Если выполняется и то и другое, то пара $(x^0, y^0)$ образует седловую точку и любая седловая точка состоит из решений задач (2.11) и (2.12); естественно, их может быть несколько. Если хотя бы один из указанных моментов отсутствует, то у функции $F(x, y)$ седловых точек нет. Докажем эти важные факты.

**Теорема 2.2** (необходимые и достаточные условия существования седловой точки). *Для того чтобы функция* $F(x, y)$ *имела седловую точку на* $X\times Y$, *необходимо и достаточно выполнения равенства*

$$\max_{x\in X}\inf_{y\in Y} F(x,y) = \min_{y\in Y}\sup_{x\in X} F(x,y). \qquad (2.14)$$

**Доказательство.** *Необходимость.* Пусть существует точка $(x^0, y^0)$, удовлетворяющая (2.8), тогда

$$\sup_{x\in X}\inf_{y\in Y} F(x,y) \le \sup_{x\in X} F(x,y^0) = F(x^0,y^0) = \inf_{y\in Y} F(x^0,y),$$

откуда $$\sup_{x\in X}\inf_{y\in Y} F(x,y) = \max_{x\in X}\inf_{y\in Y} F(x,y) = \inf_{y\in Y} F(x^0,y) = F(x^0,y^0).$$

Аналогично

$$\inf_{y\in Y}\sup_{x\in X} F(x,y) \ge \inf_{y\in Y} F(x^0,y) = F(x^0,y^0) = \sup_{x\in X} F(x,y^0),$$

откуда $$\inf_{y\in Y}\sup_{x\in X} F(x,y) = \min_{y\in Y}\sup_{x\in X} F(x,y) = \sup_{x\in X} F(x,y^0) = F(x^0,y^0).$$

Тем самым доказано (2.14), причем $x^0$ и $y^0$ – решения задач (2.11), (2.12).

*Достаточность.* Так как равенство (2.14) подразумевает достижимость наружных верхней и нижней граней, то существуют $x^0$ и $y^0$ такие, что

$$\inf_{y\in Y} F(x^0,y) = \max_{x\in X}\inf_{y\in Y} F(x,y), \qquad \sup_{x\in X} F(x,y^0) = \min_{y\in Y}\sup_{x\in X} F(x,y),$$

причем $$\inf_{y\in Y} F(x^0,y) = \sup_{x\in X} F(x,y^0).$$

По определению верхней и нижней граней

$$\inf_{y\in Y} F(x^0,y) \le F(x^0,y^0) \le \sup_{x\in X} F(x,y^0) = \inf_{y\in Y} F(x^0,y),$$

т.е. справедливо (2.8) и $(x^0, y^0)$ – седловая точка.

Приведенные выше примеры показывают, что седловая точка может существовать и не существовать. В примере 2.2 седловая точка существует, это (1, 1); в примере 2.1 седловой точки нет.

В математическом программировании понятие седловой точки используют для построения теории двойственности. Вернемся к общей задаче математического программирования (2.3), (2.4). Введем функцию Лагранжа (2.5) и рассмотрим максиминную задачу

$$v = \max_{x\in P}\inf_{y\in Y} F(x,y), \qquad (2.15)$$

где $Y$ – неотрицательный ортант, т.е. $Y=\{y|\, y\ge 0\}$.

**Теорема 2.3.** *Задачи* (2.3), (2.4) и (2.5), (2.15) *эквивалентны, т.е.* $\bar{v} = v$ *и решение* $x^0$ *одной задачи* (*если оно существует*) *является решением другой*.

**Доказательство.** Рассмотрим функцию $\varphi(x) = \inf_{y \in Y} F(x, y)$, $x \in P$. Если $x \in X$, то $g_i(x) \geq 0$, $i = 1, \ldots, m$, и $F(x, y) \geq f(x)$. Так как $F(x,0)=f(x)$, $\varphi(x) \equiv f(x)$ на множестве $X$. Если $x \in P \backslash X$, то хотя бы для одного $i$ выполняется $g_i(x)<0$. Выбирая последовательность векторов $y^k \in Y$ с неограниченно возрастающей данной компонентой $y_i^k \to \infty$ и нулевыми остальными компонентами, получаем, что $\varphi(x) = -\infty$ на множестве $P \backslash X$. Значит, если $X=\varnothing$, то обе задачи не имеют решения. Если $X \neq \varnothing$, то $v = \sup_{x \in X} \varphi(x) = \sup_{x \in X} f(x) = \bar{v}$ и точки максимума функций $\varphi(x)$ и $f(x)$, если верхние грани достигаются, совпадают. Теорема доказана.

Рассмотрим теперь минимаксную задачу для функции Лагранжа:

$$\tilde{v} = \min_{y \in Y} \sup_{x \in P} F(x, y). \tag{2.16}$$

**Определение 2.2.** Задача (2.5), (2.16) называется двойственной к задаче (2.3), (2.4); соотношение $\tilde{v} = \bar{v}$, если оно выполняется, называется соотношением двойственности, а теоремы, устанавливающие при некоторых условиях справедливость этого соотношения, называются теоремами двойственности.

Так как выше было показано, что $\bar{v} = v$ и $v \leq \tilde{v}$ (последнее для любых функций $F(x, y)$), всегда справедливо $\bar{v} \leq \tilde{v}$, но соотношение двойственности не обязательно должно выполняться, и действительно иногда оно справедливо, а иногда и нет. Выполнение соотношения двойственности связано с существованием седловой точки у функции Лагранжа.

**Теорема 2.4.** *Если функция Лагранжа* (2.5) *имеет седловую точку на* $P \times Y$, *то выполняется соотношение двойственности; если* $(x^0, y^0)$ – *седловая точка функции Лагранжа, то* $x^0$ – *решение задачи* (2.3), (2.4), а $y^0$ – *решение двойственной задачи* (2.5), (2.16). *Обратно, если задачи* (2.3), (2.4) *и* (2.5), (2.16) *имеют решения* $x^0$ *и* $y^0$ *и выполнено соотношение двойственности, то* $(x^0, y^0)$ – *седловая точка функции Лагранжа* (2.5).

Доказательство данной теоремы непосредственно вытекает из теоремы об эквивалентности задач (2.3), (2.4) и (2.5), (2.15) и теоремы о необходимых и достаточных условиях существования седловой точки.

Факт существования седловой точки у функции Лагранжа представляет не просто теоретический интерес. На нем основываются методы решения задач математического программирования, в первую очередь, метод множителей Лагранжа. Действительно, если $(x^0, y^0)$ – седловая точка функции Лагранжа (2.5), то по определению $F(x^0, y^0) = \max_{x \in P} F(x\ , y^0)$, т.е.

$$f(x^0) + \sum_{i=1}^{m} y_i^0 g_i(x^0) = \max_{x \in P} (f(x) + \sum_{i=1}^{m} y_i^0 g_i(x)).$$

В простейшем случае, когда $x^0$ – внутренняя точка $P$, это означает, что существует такой вектор множителей Лагранжа $y^0$, что

$$\frac{\partial F(x^0, y^o)}{\partial x_j} = \frac{\partial f(x^0)}{\partial x_j} + \sum_{i=1}^{m} y_i^0 \frac{\partial g_i(x^0)}{\partial x_j} = 0, \quad j = 1,...,n. \tag{2.17}$$

Уравнения (2.17) представляют собой обычный метод множителей Лагранжа для нахождения экстремума функции с ограничениями на переменные. Однако тут заложены два предположения. Одно, менее существенное, что $x^0$ – внутренняя точка множества $P$. Если это не так, то условия оптимальности претерпевают некоторые изменения, мы это продемонстрируем в дальнейшем на примере задачи выпуклого программирования. Другое предположение, принципиальное, о существовании седловой точки у функции Лагранжа. Если она не существует (а такое действительно может быть), то метод Лагранжа уже не обоснован. Достаточно представительными случаями существования седловой точки у функции Лагранжа, да и то при дополнительных предположениях, являются задача линейого программирования и задача выпуклого программирования.

## 2.3. Выпуклое программирование

Для того чтобы сформулировать задачу выпуклого программирования, необходимо напомнить определения выпуклой и вогнутой функций.

Множество $X$ называется **выпуклым**, если все точки отрезка, соединяющие любые две точки этого множества, также принадлежат множеству $X$, т.е. $\forall x', x'' \in X$, $\alpha \in [0, 1]$ выполняется $x = \alpha x' + (1 - \alpha) x'' \in X$ .

**Определение 2.3.** Функция $f(x)$ называется выпуклой на выпуклом множестве $X$, если для любых точек $x', x'' \in X$ и произвольного числа $\alpha \in [0, 1]$ выполняется неравенство

$$f(\alpha x' + (1-\alpha) x'') \le \alpha f(x') + (1-\alpha) f(x'').$$

Функция $f(x)$ называется вогнутой на выпуклом множестве $X$, если для любых точек $x', x'' \in X$ и произвольного числа $\lambda \in [0, 1]$ выполняется обратное неравенство

$$f(\alpha x' + (1-\alpha) x'') \ge \alpha f(x') + (1-\alpha) f(x'').$$

Если в определениях неравенства выполняются строго, то функция $f(x)$ называется соответственно **строго выпуклой** или **строго вогнутой**. При этом определение выпуклой (или вогнутой) функции имеет смысл только на выпуклом множестве, т.к. в этом случае гарантируется, что точка $\alpha x' + (1-\alpha)x''$ принадлежит области определения функции $f(x)$.

***Критерий выпуклости (вогнутости) функции***: Пусть $X$ – выпуклое множество из $R^n$ и $\operatorname{int} X \neq \varnothing$, $f(x)$ – дважды непрерывно дифференцируемая функция. Тогда для вогнутости (выпуклости) $f(x)$ необходимо и достаточно, чтобы ее матрица вторых производных $f''(x)$ была неположительно (неотрицательно) определена на $X$, т.е. $\forall h \in R^n$ $\forall x \in X \langle f''(x)h, h\rangle \leq (\geq)\ 0$.

**Определение 2.4.** Задачей выпуклого программирования называется частный случай общей задачи математического программирования (2.3), (2.4), когда целевая функция $f(x)$ и функции ограничений $g_i(x)$ являются вогнутыми на выпуклом множестве $P$.

Так как задача максимизации функции эквивалентна задаче минимизации этой функции со знаком минус, ограничение $g_i \geq 0$ эквивалентно $-g_i \leq 0$, и из вогнутости функции $f(x)$ следует выпуклость $-f(x)$ и наоборот, то задачей выпуклого программирования называется также задача минимизации выпуклой функции $f(x)$ при ограничениях $g_i(x) \leq 0$, $i=1,\ldots,m$ , $x \in P$, где $g_i(x)$ – выпуклые функции, $P$ – выпуклое множество. Это просто другая форма той же задачи. Однако следует обратить внимание, что если задача выпуклого программирования формулируется как задача на максимум, то целевая функция обязательно вогнута, а если на минимум, то выпукла; если ограничения записываются в виде $g_i(x) \geq 0$, то функции ограничений вогнуты, если $g_i(x) \leq 0$, то выпуклы. Эта жесткая связь обусловлена наличием определенных свойств у задачи выпуклого программирования, которые в противном случае (при нарушении такой связи) могут и не выполняться. Два основных таких свойства сформулированы в следующей лемме.

**Лемма 2.2.** *Множество допустимых планов задачи выпуклого программирования* (2.3) *является выпуклым. Любой локальный максимум вогнутой функции* $f(x)$ *на* $X$ *является глобальным*.

**Доказательство.** Пусть $x'$, $x'' \in X$, т.е. $x'$, $x'' \in P$ и $g_i(x') \geq 0$, $g_i(x'') \geq 0$, $i=1,\ldots,m$, и $\alpha \in [0, 1]$.Тогда по определению выпуклого множества $x = \alpha x' + (1-\alpha)x'' \in P$
и по определению вогнутой функции

$$g_i(x) = g_i(\alpha x' + (1-\alpha)x'') \geq \alpha g_i(x') + (1-\alpha) g_i(x'') \geq 0,\ i=1,\ldots,m,$$

т.е. $x \in X$. Выпуклость множества $X$ доказана.

Далее, пусть $x^*$ – локальный максимум функции $f(x)$ на $X$, т.е. $f(x^*) \geq f(x)$ для любого $x$ из пересечения $X$ с некоторой достаточно малой окрестностью $\omega$ точки $x^*$. Нам надо показать, что $x^*$ является и глобальным максимумом, т.е. что $f(x^*) \geq f(x) \quad \forall x \in X$. Предположим противное, т.е. что существует $\bar{x} \in X$ такое, что $f(\bar{x}) > f(x^*)$. Возьмем $\tilde{x}_\alpha = \alpha x^* + (1-\alpha)\bar{x},$ где $0<\alpha<1$, тогда в силу вогнутости функции $f(x)$ имеем

$$f(\tilde{x}_\alpha) \geq \alpha f(x^*) + (1-\alpha) f(\bar{x}) > \alpha f(x^*) + (1-\alpha) f(x^*) = f(x^*).$$

При достаточно малом $\alpha>0$, очевидно, $\tilde{x}_\alpha \in X \cap \omega, \quad f(\tilde{x}_\alpha) > f(x^*),$ пришли к противоречию. Лемма доказана.

Если бы функции ограничений $g_i(x)$ были выпуклы, то для определяемого (2.3) множества $X$ выпуклость уже бы не следовала; в этом случае можно доказать выпуклость множества $X' = \{x \mid x \in P, g_i(x) \leq 0, i = 1,...,m\}$. Если бы целевая функция $f(x)$ была выпукла, то утверждение леммы относительно ее максимумов стало бы неверным, однако аналогичное утверждение можно было бы доказать для минимумов.

Бывают функции, которые выпуклы по одной группе переменных и вогнуты по другой. Такие функции представляют собой один из основных классов функций, у которых заведомо существует седловая точка. Приведем без доказательства следующую теорему.

**Теорема 2.6** (о существовании седловой точки у вогнуто-выпуклой функции). *Пусть $X$ и $Y$ – выпуклые замкнутые ограниченные подмножества конечномерных евклидовых пространств, а функция $F(x, y)$ непрерывна по $x$ и $y$, вогнута по $x \; \forall y$ и выпукла по $y \; \forall x$, тогда $F(x, y)$ имеет на $X \times Y$ седловую точку.*

Рассмотрим для задачи выпуклого программирования функцию Лагранжа (2.5), определенную на прямом произведении $P \times Y$, где $Y = \{y \mid y \geq 0\}$. Эта функция в силу вогнутости $f(x)$, $g_i(x)$ является вогнутой по $x \; \forall y$ и линейной, а следовательно, выпуклой по $y \; \forall x$. Однако нельзя утверждать на основании предыдущей теоремы, что она имеет седловую точку, так как множество $P$ не обязательно является замкнутым и ограниченным, а $Y$ заведомо неограничено. Тем не менее можно ожидать, что при некоторых условиях, заменяющих не выполняющиеся в данном случае условия этой теоремы, седловая точка все же будет существовать. Наиболее известной теоремой, относящейся к этому вопросу, является теорема Куна-Таккера, которая представляет собой центральный результат теории двойственности в выпуклом программировании. Эта теорема устанавливает связь между существованием решения задачи выпуклого программирования и седловой точки у функции Лагранжа при выполнении некоторого условия регулярности. Наиболее простым является следующее условие регулярности: задача выпуклого программирования удовлетворяет условию Слейтера, если существует такая точка $x \in P$, что $g_i(x)>0,\ i=1,\ldots,m$.

**Теорема 2.7 (Куна-Таккера).** *Если задача выпуклого программирования удовлетворяет условию Слейтера, то для оптимальности плана* $x^0 \in X$ *необходимо и достаточно существование такого* $y^0 \geq 0$*, что пара* $(x^0, y^0)$ *является седловой точкой функции Лагранжа* (2.5) *на* $P \times Y$.

**Доказательство.** *Достаточность* нами доказана для общей задачи математического программирования, следовательно, она справедлива и для задачи выпуклого программирования без всяких дополнительных условий (типа условия Слейтера).

Докажем *необходимость*. Для этого рассмотрим в $(m+1)$-мерном пространстве два множества

$$A=\left\{ z=\begin{pmatrix} z_1 \\ z_2 \\ \ldots \\ z_{m+1} \end{pmatrix} \middle| z_i \leq g_i(x), i=1,\ldots,m, z_{m+1} \leq f(x) \right\},$$

$$B=\left\{ z=\begin{pmatrix} z_1 \\ z_2 \\ \ldots \\ z_{m+1} \end{pmatrix} \middle| z_i > 0, i=1,\ldots,m, z_{m+1} > f(x^0) \right\},$$

где $x^0$ – решение задачи выпуклого программирования, $x$ - произвольная точка множества $P$. Покажем, что множество $A$ является выпуклым. Действительно, если $z', z'' \in A$, то $z_i' \leq g_i(x')$, $z_i'' \leq g_i(x'')$, $i=1,\ldots,m$, $z'_{m+1} \leq f(x')$, $z''_{m+1} \leq f(x'')$, $x', x'' \in P$. Рассмотрим $z^\alpha = \alpha z' + (1-\alpha) z''$, $0 \leq \alpha \leq 1$. Для него выполняются неравенства $z_i^\alpha \leq \alpha g_i(x')+$ $+(1-\alpha) g_i(x'') \leq g_i(\alpha x' + (1-\alpha) x'') = g_i(x^\alpha)$, где $x^\alpha = \alpha x' + (1-\alpha) x'' \in P$, $i=1,\ldots,m$; $z_{m+1}^\alpha \leq \alpha f(x') + (1-\alpha) f(x'') \leq f(\alpha x' + (1-\alpha) x'') = f(x^\alpha)$, т.е. по определению множества $A$ вектор $z^\alpha \in A$, значит $A$ – выпуклое множество.

Множество $B$ также является выпуклым и представляет собой внутренность ортанта с вершиной в точке $(0,\ldots,0, f(x^0))$.

В силу оптимальности $x^0$ множества $A$ и $B$ не имеют общих векторов, т.е. $A \bigcap B = \varnothing$. Действительно, если существует $z \in A \bigcap B$, то для некоторого $x \in P$ выполняются условия $g_i(x) \geq z_i$, $i = 1,...,m$, т.е. $x \in X$, и $f(x) \geq z_{m+1} > f(x^0)$, пришли к противоречию.

Значит, по теореме о разделяющей гиперплоскости существует такой ненулевой вектор $a = \begin{pmatrix} a_1 \\ a_2 \\ ... \\ a_{m+1} \end{pmatrix}$, что

$$\langle a, z^1 \rangle \leq \langle a, z^2 \rangle \qquad \forall z^1 \in A,\ z^2 \in B. \tag{2.18}$$

По определению множества $B$, очевидно, $a \geq 0$, так как если $a_i < 0$, то, выбирая последовательность векторов $z^k \in B$ с $z^k_i \to \infty$ и остальными компонентами равными нулю, получим $\langle a, z^k \rangle \to -\infty$, что противоречит неравенству (2.18). Далее, так как точка $(0,\ldots,0,\ f(x^0))$ является предельной для множества $B$, то имеем из (2.18) с учетом определения множества $A$

$$\sum_{i=1}^{m} a_i g_i(x) + a_{m+1} f(x) \leq a_{m+1} f(x^0) \tag{2.19}$$

для любой точки $x \in P$. Отсюда следует, что $a_{m+1} > 0$, поскольку, если $a_{m+1} = 0$, то из (2.19) вытекает $\sum_{i=1}^{m} a_i g_i(x) \leq 0\ \forall x \in P,$ что противоречит условию Слейтера, так как хотя бы одно $a_i > 0$ и при выполнении условия Слейтера существует $x \in P$ такое, что $g_i(x) > 0, i = 1,...,m$, т.е. $\sum_{i=1}^{m} a_i g_i(x) > 0$. Итак, $a \geq 0$ и $a_{m+1} > 0$. Положим $y_i^0 = \dfrac{a_i}{a_{m+1}}, i = 1,...,m$, тогда $y^0 = (y_1^0,\ldots, y_m^0) \geq 0$ и из (2.19) получаем

$$\sum_{i=1}^{m} y_i^0 g_i(x) + f(x) \leq f(x^0) \quad \forall x \in P. \tag{2.20}$$

Полагая в (2.20) $x = x^0$, получим $\sum_{i=1}^{m} y_i^0 g_i(x^0) \leq 0,$ но, с другой стороны, $\sum_{i=1}^{m} y_i^0 g_i(x^0) \geq 0$, так как $y_i^0 \geq 0$, $g_i(x^0) \geq 0$, $i = 1,...,m$, следовательно,

$$\sum_{i=1}^{m} y_i^0 g_i(x^0) = 0. \tag{2.21}$$

Из (2.20) и (2.21) имеем $F(x, y^0) \leq F(x^0, y^0) \quad \forall x \in P$

Учитывая, что $\sum_{i=1}^{m} y_i\, g_i(x^0) \ge 0\ \forall y \ge 0$, имеем $F(x^0,\ y^0) \le F(x^0,\ y)\quad \forall y \ge 0$. Значит $(x^0,\ y^0)$ – седловая точка функции $F(x, y)$ на $P\times Y$. Теорема доказана.

Теорема Куна-Таккера обосновывает сведение задачи выпуклого программирования к задаче поиска седловой точки функции Лагранжа. Задача поиска седловой точки аналогична по сложности задаче поиска экстремума функции, поэтому для того, чтобы такое сведение имело практический смысл, необходимо, чтобы получающаяся экстремальная задача была в чем-то проще исходной. Такое упрощение связано со следующим обстоятельством. Исходная задача представляет собой поиск максимума на множестве $X$, которое имеет вид (2.3). Разбиение ограничений, описывающих допустимое множество $X$, на $x\in P$ и $g_i(x)\ge 0$, как уже говорилось, является условным, так как любое ограничение можно записать и в том, и в другом виде. Поэтому обычно под $P$ понимается множество простого вида, это либо все пространство $R^n$, либо неотрицательный ортант, либо параллелепипед, т.е. границы множества $P$ задаются простыми ограничениями типа $a \le x \le b$.

Сложность же задачи выпуклого программирования определяется системой ограничений $g_i(x) \ge 0,\ \ i = 1,...,m$. Так как седловая точка функции Лагранжа ищется на произведении множеств $P\times Y$, где $Y$ также является множеством простого вида (неотрицательным ортантом), то смысл теоремы Куна-Таккера состоит в сведении задачи поиска экстремума функции со сложными ограничениями на переменные к задаче поиска экстремума новой функции с простыми ограничениями. Если множество $P$ совпадает со всем пространством $R^n$, то условия экстремума имеют вид (2.17), причем очень существенным моментом является то, что эти условия являются не только необходимыми для того, чтобы функция $F(x, y^0)$ достигала в точке $x^0$ максимума по $x$, но и достаточными. Это является важным свойством вогнутых функций (а $F(x, y)$ для задачи выпуклого программирования вогнута по $x$).

**Теорема 2.8.** *Для того чтобы непрерывно дифференцируемая вогнутая функция* $\varphi(x)$ *имела в точке* $x^0$ *максимум на* $R^n$, *необходимо и достаточно, чтобы* $\varphi'(x^0) = 0$ (докажите в качестве упражнения).

Таким образом, для нахождения седловой точки функции Лагранжа $F(x, y)$ на произведении $R^n\times Y$, а следовательно, и для нахождения решения $x^0$ задачи выпуклого программирования при $P=R^n$ надо решить систему уравнений (2.17). Но в этой системе $n$ уравнений, а неизвестных, вообще говоря, $n+m$, так как помимо $n$-мерного вектора $x^0$ нам неизвестен и $m$-мерный вектор соответствующих ему множителей Лагранжа $y^0$. Однако, если внимательно просмотреть доказательство теоремы Куна-Таккера, то мы увидим важное свойство, которое выполняется для седловой точки функции Лагранжа, а именно, условие (2.21). Так как $y_i^0 \geq 0$ и $g_i(x^0) \geq 0$, $i = 1,...m$, то из него вытекает, что либо $y_i^0 = 0$, либо $g_i(x^0) = 0$, либо то и другое одновременно. Это свойство аналогично второй теореме двойственности линейного программирования. Ограничения, которые выполняются в некоторой точке как равенства, называются активными. Таким образом, только те множители Лагранжа $y_i^0$ могут быть отличны от нуля, которые соответствуют активным в точке $x^0$ ограничениям. С учетом этого свойства для нахождения решения задачи выпуклого программирования получаем следующий способ. Введем множество $I(x) = \{i \mid 1 \leq i \leq m, g_i(x) = 0\}$. Множество $I(x)$ представляет собой совокупность индексов активных в точке $x$ ограничений. Тогда можно утверждать в силу сказанного выше, что $y_i^0 = 0 \ \ \forall i \notin I(x^0)$. Поэтому для определения $x^0$ и ненулевых $y_i^0$ имеем систему уравнений

$$\begin{cases} \dfrac{\partial f(x^0)}{\partial x_j} + \sum\limits_{i \in I(x^0)} y_i^0 \dfrac{\partial g_i(x^0)}{\partial x_j} = 0,\ j = 1,...,n, \\ g_i(x^0) = 0, i \in I(x^0). \end{cases} \tag{2.22}$$

Число неизвестных и число уравнений здесь совпадают, поэтому, если мы найдем решение системы уравнений (2.22), мы тем самым решим исходную задачу выпуклого программирования. Система (2.22) представляет собой необходимые и достаточные условия оптимальности для задачи выпуклого программирования в случае, когда $P = R^n$.

Рассмотрим, как трансформируются эти условия в случаях, когда $P$ не совпадает со всем пространством $R^n$.

Сначала остановимся на случае, когда $P$ является неотрицательным ортантом, т.е. $P = \{x \mid x \geq 0\}$. Тут для каждой компоненты $x_j^0$ решения задачи выпуклого программирования $x^0$ возможны две альтернативы: либо $x_j^0 > 0$, либо $x_j^0 = 0$. Для первой альтернативы условия оптимальности совпадают со случаем $P = R^n$, т.е. имеют вид (2.17) при данном $j$. Однако для второй альтернативы это не так. Если $x_j^0 = 0$ является компонентой точки максимума функции $F(x, y^0)$ по $x$ в области $x \geq 0$, то можно утверждать, что необходимо (а для вогнутой функции и достаточно) только выполнение неравенства $\dfrac{\partial F(x^0, y^0)}{\partial x_j} \leq 0$, т.е. функция $F(x, y^0)$ должна не возрастать по $x_j$ при движении внутрь области $x \geq 0$. Но так как $x_j^0 = 0$, то можно записать $\dfrac{\partial F(x^0, y^0)}{\partial x_j} x_j^0 = 0$. Объединяя обе альтернативы, а также условия дополняющей нежесткости, необходимые и достаточные условия оптимальности для задачи выпуклого программирования в случае $P=\{x|\ x\geq 0\}$ можно представить в виде

$$\begin{cases} \dfrac{\partial F(x^0, y^0)}{\partial x_j} \leq 0,\ j = 1,\dots,n, \\ \sum\limits_{j=1}^{n} \dfrac{\partial F(x^0, y^0)}{\partial x_j} x_j^0 = 0, \quad y_i^0 g_i(x^0) = 0, i = 1,\dots,m. \end{cases} \tag{2.23}$$

Условия оптимальности (2.23) представляют собой условия Куна-Таккера для задачи выпуклого программирования в случае $P=\{x|\ x\geq 0\}$. Для общей задачи математического программирования при выполнении условия регулярности (линейной независимости в точке решения векторов-градиентов активных ограничений, включая простейшие) они являются необходимыми (см. теорему 2.1). Это можно доказать, например, заменяя неравенства на равенства путем введения искусственных переменных, и применяя к полученной задаче с ограничениями типа равенств условия оптимальности Лагранжа.

Таким образом, конструктивный метод нахождения седловой точки функции Лагранжа задачи выпуклого программирования предполагает использование условий Куна-Таккера (2.6) или (2.7), которые для данной задачи являются необходимыми и достаточными условиями и могут быть получены непосредственно из определения седловой точки.

Следующий пример показывает, что условие Слейтера является существенным.

**Пример 2.3.** $f(x)= x$, $m=1$, $g(x)= -x^2$, $P=R^1$.

Очевидно, что $x=0$ есть решение задачи, но у функции $F(x, y)= x - yx^2$ седловой точки нет. Действительно, функция Лагранжа $F = x - yx^2$, поэтому $\dfrac{\partial F}{\partial x} = 1 - 2yx \neq 0$ при $x = 0$, т.е. условия Куна-Таккера (2.7) не выполняются. Здесь не выполнено условие регулярности (условие Слейтера), т.е. требование $-x^2>0$.

Если условие Слейтера не выполнено, то условия Куна-Таккера являются только достаточными условиями.

**Пример 2.4.** Найти максимум функции $x_1 + 2x_2$ при ограничении $x_1^2 + x_2^2 \leq 4$.

Здесь $f(x)= x_1 + 2x_2$, $m=1$, $n=2$, $g_1(x) = 4 - x_1^2 - x_2^2$. Условия Слейтера выполняются, поэтому используя условия (2.7), имеем либо систему из трех уравнений с тремя неизвестными

$$\begin{cases} 1-2x_1 y = 0, \\ 2-2x_2 y = 0, \\ x_1^2 + x_2^2 = 4, \end{cases}$$

откуда $x_1^0 = 2/\sqrt{5}, x_2^0 = 4/\sqrt{5}, y^0 = \sqrt{5}/4,$ либо вместо третьего уравнения $y$=0, этот вариант условий не дает решения.

Хотя в качестве множества $P$, как говорилось, обычно выбирается достаточно простое множество, рассмотрим общий случай, когда $P$ – произвольное выпуклое множество. Рассмотрение такой ситуации полезно еще и потому, что оно позволяет несколько по-другому подойти к задаче выпуклого программирования. Действительно, если в $P$ включить все ограничения $g_i(x)\geq 0$, то оно будет просто совпадать с множеством $X$ и задача выпуклого программирования состоит в максимизации вогнутой функции $f(x)$ на выпуклом множестве $P$ (или $X$). Поэтому, не конкретизируя вида ограничений и не вводя функцию Лагранжа, можно напрямую рассмотреть функцию $f(x)$ на $X$. При этом полезным является следующее понятие.

**Определение 2.5.** Направление $g$ является допустимым в точке $x$, принадлежащей выпуклому множеству $X$, если $\exists\ \varepsilon_0 > 0$ такое, что $\forall \varepsilon \leq \varepsilon_0$ выполняется $x+\varepsilon g \in X$.

Так как определение допустимого направления не зависит, очевидно, от длины вектора $g$, то будем его в дальнейшем считать единичным.

Допустимое направление – это такое направление, небольшое перемещение вдоль которого не выводит из допустимого множества, т.е. не нарушает ограничений задачи математического программирования. Но если точка $x^0$ является решением задачи математического программирования, т.е. доставляет максимум функции $f(x)$ на $X$, то при перемещении вдоль допустимого направления, по определению максимума, функция $f(x)$ не может возрастать. Изменение функции вдоль некоторого направления определяется ее производной по направлению.

**Определение 2.6.** Производной функции $f(x)$ по направлению, задаваемому единичным вектором $g$, называется величина

$$\frac{\partial f(x)}{\partial g} = \lim_{\varepsilon \to 0^+} \frac{f(x+\varepsilon g) - f(x)}{\varepsilon}, \tag{2.24}$$

если предел в (2.24) существует.

С помощью производных по направлению получаем другую форму условий оптимальности: для того чтобы $x^0$ было решением задачи математического программирования, необходимо, а для задачи выпуклого программирования и достаточно, чтобы для любого допустимого в $x^0$ направления $g$ выполнялось

$$\frac{\partial f(x^0)}{\partial g} \le 0. \qquad (2.25)$$

Применительно к функции Лагранжа на $P$ это условие имеет вид $\frac{\partial F(x^0, y^0)}{\partial g} \le 0,$ где $g$ – любое допустимое направление в точке $x^0 \in P$.

Важным частным случаем выпуклого программирования является задача квадратичного программирования, для которой, двойственную задачу можно выписать в явном виде. Эта задача имеет вид:

$$\langle Cx, x\rangle \to \max, \qquad Ax \le b, \qquad (2.26)$$

где $x \in R^n$, $b \in R^m$, $C$ – отрицательно определенная симметрическая матрица размера $n\times n$ (т.е. $C^T=C$ и $xCx \le 0$), $A$ – матрица размера $m\times n$ (для задачи на минимум $C$ положительно определена, иногда в целевую функцию включают дополнительно линейные слагаемые). Нетрудно видеть, что элементы матрицы $C$ представляют собой ее вторые частные производные $\partial f(x)/(\partial x_i \partial x_j)=d_{ij}$. Но, как было сказано выше, выпуклость (вогнутость) квадратичной функции определяется неотрицательной (неположительной) определенностью матрицы вторых частных производных, т.е. критерий выпуклости (вогнутости) установливает здесь вогнутость целевой функции в задаче (2.26).

В силу линейности ограничений задачи (2.26) условием регулярности является просто существование решения. Функция Лагранжа для задачи (2.26) имеет вид

$$F(x, y)=\langle Cx, x\rangle + \langle y, b-Ax\rangle, \text{ где } y \in R^m, y\ge 0.$$

Для построения двойственной задачи надо найти глобальный максимум $F(x, y)$ по $x \in R^n$. Для вогнутой квадратичной функции он существует и точка глобального максимума определяется из условия равенства нулю всех частных производных $F(x, y)$ по $x$, откуда имеем $2Cx= yA=A^Ty$.

Известно, что у отрицательно определенной матрицы существует обратная, поэтому

$$x=0.5C^{-1} A^Ty. \qquad (2.27)$$

Подставив это значение $x$ в $F(x, y)$, получим двойственную задачу к (2.26):

$$\langle y, b\rangle - 0.25\langle A^Ty, C^{-1} A^Ty\rangle \to \min, \quad y\ge 0. \qquad (2.28)$$

Задача (2.28), в принципе, проще (2.26), так как в ней имеется только простейшее условие неотрицательности переменных. Условия минимума в (2.28) – равенство нулю частных производных целевой функции для ненулевых компонент вектора $y$. Найдя решение задачи (2.28) $y^*$, по формуле (2.27) найдем и решение задачи (2.26) $x^*$.

**Пример 2.5.** Найти максимум $f(x_1, x_2) = x_1x_2 - x_1^2 x_2^2$ при $x_1+2x_2\ge 1$.

Здесь данные задачи (2.26) с учетом знака ограничения равны

$A$=(–1, –2), $b$= –1, $C=\begin{pmatrix} -1 & 1/2 \\ 1/2 & -1 \end{pmatrix}$, $C^{-1}=\begin{pmatrix} -4/3 & -2/3 \\ -2/3 & -4/3 \end{pmatrix}$, $A^T=\begin{pmatrix} -1 \\ -2 \end{pmatrix}$.

Подставляя эти данные в задачу (2.28), получаем двойственную задачу:

$$g(y) = -y + 7/3\, y^2 \to \min,\ y \ge 0.$$

Приравнивая производную $g(y)$ к нулю, получаем решение $y^*=3/14$, $g(y^*)=-3/28$. Подставляя $y^*$ в (2.27), получаем $x^*=(2/7,\ 5/14)$, $f(x^*)=-3/28$.

## 4.4. Интерпретация множителей Лагранжа в задаче математического программирования

Как и в предыдущей главе, множители Лагранжа можно истолковать как характеристики изменений оптимального значения целевой функции при изменениях констант ограничений $y^* = \dfrac{\partial f(x^*)}{\partial b}$.

Для доказательства следует сначала показать, что $x^*$ и $y^*$ можно представить в виде функций от констант $b$ в ограничениях $g(x)=b$. Затем продифференцировать функцию Лагранжа по этим константам аналогично тому, как это было сделано в предыдущей главе.

Условия Куна-Таккера можно было бы записать в виде равенств, если бы было известно, какие именно ограничения выполняются как равенства, а какие как неравенства, а также и то, какие именно переменные равны нулю и какие положительные. Допустим, что эти соотношения и переменные перенумерованы так, что в оптимальной точке первые $m_1$ из них выполняются как равенства, а остальные $m - m_1$ как неравенства ($0 \le m_1 \le m$) и что первые $n_1$ переменных положительны, а остальные $n - n_1$ равны нулю ($0 \le n_1 \le n$). Векторы можно расчленить следующим образом:

$$g(x)=\begin{pmatrix} g^1(x) \\ g^2(x) \end{pmatrix},\ b=\begin{pmatrix} b^1 \\ b^2 \end{pmatrix},\ x=\begin{pmatrix} x^1 \\ x^2 \end{pmatrix},\ y=\begin{pmatrix} y^1 & y^2 \end{pmatrix},$$

где $g^1(x)$, $b^1$, $y^1$ состоят из первых элементов векторов $g(x)$, $b$ и $y$ соответственно, а $x^1$ – из первых $n_1$ элементов $x$. Тогда условия Куна – Таккера можно записать как

$$\frac{\partial F}{\partial x^1} = \frac{\partial f(x)}{\partial x^1} - y^1 \frac{\partial g^1(x)}{\partial x^1} = 0, \quad x^2 = 0,$$
$$\frac{\partial F}{\partial y^1} = b^1 - g^1(x) = 0, \quad y^2 = 0.$$

Очевидно, что $y^* = \dfrac{\partial f(x^*)}{\partial b}$ выполняется для последних $m - m_1$ множителей Лагранжа, равных нулю. Следовательно, $y_i^* = \dfrac{\partial f(x^*)}{\partial b_i} = 0, \quad i = m_1 + 1, m_1 + 2, ..., m$.

Последние $m - m_1$ ограничений выполняются как неравенства, так что малые изменения соответствующих констант ограничений не могли бы изменить оптимального значения функции цели.

Что касается первых $m_1$ множителей Лагранжа, то заметим, что задача математического программирования имеет вид $\max\limits_{x^1} f(x^1, 0), \quad g^1(x^1, 0) = b^1$.

На том же основании, что в предыдущей главе, оказывается возможным представить что $x^1$ и $y^1$ как функции от $b^1$ и продифференцировать функцию Лагранжа по $b^1$. В результате получим, что $y_i^* = \dfrac{\partial f(x^*)}{\partial b_i} \geq 0, \quad i = 1,2,..., m_1$.

### 2.5. Численные методы нелинейного программирования

Хотя теорема Куна-Таккера и вытекающие из нее условия экстремума характеризуют решение задачи математического программирования, они еще не дают конструктивных методов нахождения этих решений для достаточно сложных случаев. Рассмотрим некоторые из наиболее распространенных численных методов, которые так или иначе используют полученные выше условия оптимальности.

1. *Градиентные методы*. Если рассматривается задача максимизации $f(x)$ при ограничениях, т.е. когда $X$ не совпадает с $R^n$, то непосредственное применение градиентного процесса (см. главу 2) может привести к нарушению ограничений, даже если начальная точка $x^{(0)} \in X$. Однако эту трудность можно преодолеть, например, если получаемую очередную точку проектировать на множество $X$. Если обозначить операцию проектирования точки $x$ на множестве $X$ через $\Pr_X(x)$, то соответствующий итеративный процесс имеет вид

$$x^{(k+1)} = \Pr_X(x^{(k)} + \alpha_k f'(x^{(k)})). \tag{2.29}$$

Полученный метод носит название метода проекции градиента. Шаг $\alpha_k$ в методе (2.29) может выбираться различными способами, например, как в методе скорейшего подъема или выбирая постоянный шаг, обеспечивающий условие сходимости (более подробно описано в главе 1). Стационарная точка этого процесса является решением задачи $\max_{x\in X} f(x)$ в случае вогнутой функции $f(x)$, а в общем случае требуется дополнительное исследование.

Недостатком метода проекции градиента является необходимость проведения операции проектирования, которая в общем случае эквивалентна некоторой задаче поиска экстремума. Однако, когда $X$ является шаром, параллелепипедом, гиперплоскостью, полупространством или ортантом, задача проектирования решается просто и в явном виде.

Еще одной разновидностью градиентных методов является метод условного градиента, который также предназначен для решения экстремальных задач с ограничениями и состоит в решении вспомогательной задачи максимизации на множестве $X$ линейной функции $\langle f'(x^{(k)}), x-x^{(k)}\rangle$, представляющей собой главную часть приращения функции $f(x)$ в точке $x^{(k)}$. Эта вспомогательная задача может быть непростой, но если $X$ задается линейными ограничениями, то она представляет собой задачу линейного программирования, которая решается за конечное число шагов стандартными методами (например, симплекс-методом). Если решение вспомогательной задачи $\tilde{x}^{(k)}$ найдено, то следующее приближение для исходной задачи строится по формуле

$$x^{(k+1)} = x^{(k)} + \alpha_k(\tilde{x}^{(k)} - x^{(k)}). \tag{2.30}$$

Если множество $X$ выпуклое, то для $\alpha_k \in [0,1]$ точка $x^{(k+1)}\in X$. Шаг $\alpha_k$ выбирается из условия максимального роста функции $f(x)$ или любым другим способом, обеспечивающим рост $f(x)$. На практике обычно решают вспомогательную задачу не точно, а приближенно. В процессе (2.30) направление движения не совпадает с градиентом функции $f(x)$ в точке $x^{(k)}$, но определяется им, так как его компоненты берутся в качестве коэффициентов линейной целевой функции вспомогательной задачи.

**Пример 2.6.** Найти минимальное значение функции $f=-x_1+x_2^2$ при условии $x_1^2+x_2^2\le 1$.

Решим данную задачу методом проекции градиента, завершая вычисления при выполнении условия $\| x^{(k-1)} - x^{(k)} \| \le 0.01$.

Используем (2.29) с постоянным шагом $\alpha_k\equiv\alpha\in(0, 1)$, $k$=0,1,2…, например, $\alpha_k = 0.75$. Нетрудно убедиться, что условие сходимости для данного шага выполняется (оставляем читателю в качестве упражнения). Так как $f'(x^{(0)})=(-1, 1)$, то

$$x^{(1)}=\mathrm{Pr}_X[x^{(0)}-\alpha f'(x^{(0)})] = \mathrm{Pr}_X[(0 , 0.5)-0.75(-1, 1)] = \mathrm{Pr}_X[(0.75, -0.25)].$$

Точка (0.75, -0.25) принадлежит допустимому множеству, так как $0.75^2+0.25^2=0.625 < 1$, поэтому $x^{(1)}=(0.75, -0.25)$. Требуемая точность не достигнута, так как $\| x^{(0)} - x^{(1)} \| = 1.6 > 0.01$. Далее находим $f'(x^{(1)})$=(-1 , -0.5), $x^{(2)}$ = $Pr_X$[(1.5 , 0.125)]. Точка (1.5 , 0.15) допустимому множеству не принадлежит, потому что $1.5^2+0.125^2=2.266>1$. Так как допустимое множество представляет собой замкнутый шар радиуса 1 с центром в точке (0, 0) в пространстве $R^2$, то

$$x^{(2)} = Pr_X[(1.5\,,\,0.125)] = \frac{(1.5\,,\,0.125)}{\sqrt{1.5^2 + 0.125^2}} = (0.9965\,,\,0.08304).$$

Требуемая точность снова не достигнута, так как $\| x^{(1)} - x^{(2)} \| = 0.298 > 0.01$.

Результаты остальных шагов приведены в табл. 2.1.

Таблица 2.1

| $k$ | $x^{(k)}$ | $\| x^{(k-1)} - x^{(k)} \|$ | $f'(x^{(k)})$ | $x^{(k)} - \alpha_k f'(x^{(k)})$ |
|---|---|---|---|---|
| 2 | (0.9965, 0.0830) | 0.298 | (-1, 0.1661) | (1.7465, -0.0415) |
| 3 | (0.9997, -0.0238) | 0.107 | (-1, -0.0476) | (1.7497, 0.0119) |
| 4 | (0.99998, 0.0077) | 0.031 | (-1, 0.0136) | (1.7500, -0.0034) |
| 5 | (0.99998, -0.00194 | 0.009 | точность достигнута | |

Итак, $x^* \approx x^{(5)}$=(0.99998, -0.00194), $f^* \approx f(x^{(5)}) \approx -1$. Отметим, что точное решение рассматриваемой задачи $f_{min}=f(1, 0)= -1$.

2. *Методы возможных направлений.* Идея методов возможных направлений состоит в следующем: на каждой итерации определяется допустимое направление на множестве $X$, вдоль которого функция $f(x)$ возрастает (такое направление называется возможным направлением возрастания функции $f(x)$), и по нему совершается шаг. Фактически в методе проекции градиента и в методе условного градиента мы находим такие направления. Однако там исходным было определение градиента, а допустимое направление определялось по нему однозначно. В методах же возможных направлений исходным пунктом является описание всех допустимых направлений и выбор из них такого, вдоль которого функция $f(x)$ возрастает и желательно скорейшим образом.

Рассмотрим вариант метода возможных направлений применительно к задаче максимизации $f(x)$ на множестве (2.3), где $P=R^n$. Пусть мы имеем $k$-е приближение $x^{(k)}$ к решению этой задачи и для построения следующего приближения поставим следующую вспомогательную задачу: максимизировать $u$ при ограничениях

$$\langle f'(x^{(k)}), a\rangle \ge u,\ \langle g'(x^{(k)}), a\rangle \ge u, \quad i \in I_k, \quad |a^j| \le 1, \quad j = 1,\dots,n,$$

где $I_k = \{i \,|\, 1 \le i \le m,\ g_i(x^{(k)}) = 0\}, \qquad a = (a^1,\dots,a^n).$

Эта задача представляет собой задачу линейного программирования в $(n+1)$-мерном пространстве векторов $(a, u)$. Множество допустимых планов замкнуто, ограничено и не пусто, так как $a$=0, $u$=0 является допустимым планом. Значит вспомогательная задача имеет решение ($a^k$, $u^k$), причем $u^k \geq 0$. Если $u^k>0$, то нетрудно показать, что направление $a^k$ является возможным направлением возрастания функции $f(x)$, т.е. точка $x^{(k+1)}=x^{(k)}+\alpha_k a^k$ при достаточно малом $\alpha_k$ принадлежит множеству $X$ и обеспечивает большее значение функции $f(x)$, чем $x^{(k)}$. Выбор пары ($a^k$, $u^k$) с возможно большим значением $u^k$ при этом означает выбор допустимого направления, наиболее близкого к градиенту функции $f(x)$, т.е. возможного направления с наибольшим ростом функции. Если $u^k=0$, то получается стационарная точка процесса, которая для задачи выпуклого программирования дает решение, а в общем случае требует дополнительного исследования. Для линейных функций $g_i(x)$ можно заменять в правых частях ограничений вспомогательных задач $u$ на нуль.

**Пример 2.7.** Найти минимум функции $f(x)=(x_1-7)^2+(x_2-5)^2$
при условиях $g_1(x) = x_1 - x_2 + 3 \geq 0, \quad g_2(x) = -2x_1 - x_2 + 15 \geq 0, \quad x_1, x_2 \geq 0.$

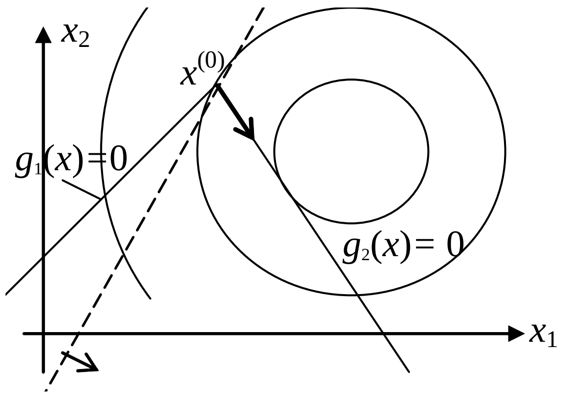


Рис. 2.1

Точку $x^{(0)}$ необходимо выбрать таким образом, чтобы она принадлежала допустимому множеству. Далее для определения возможного направления спуска ищем направление, удовлетворяющее активным ограничениям и составляющее минимальный угол с направлением антиградиента Возьмем $x^{(0)}$= (4,7), она принадлежит допустимому множеству, первые два ограничения активные, а антиградиент равен $f\,'(x^{(0)})$ = (6, -4) (см. рис. 2.1).

Для определения возможного направления спуска с учетом линейности ограничений вспомогательную задачу линейного программирования можно записать как

$$6a_1 - 4a_2 \to \max,$$

$$\begin{cases} a_1 - a_2 \geq 0, \\ -2a_1 - a_2 \geq 0, \\ -1 \leq a_1, a_2 \leq 1. \end{cases}$$

Решением этой задачи является $a^0$=(0.5, -1), а максимальное значение вспомогательной целевой функции равно 7. Значит исходная целевая функция по направлению $a^k$ убывает, причем если сделать шаг с α=2.8, то сразу получаем $x^{(1)}$ точку минимума целевой функции с координатами (5.4, 4.2) (иначе ее надо взять в качестве начальной точки для конструирования следующей итерации).

3. *Численная реализация метода множителей Лагранжа*. Эта группа методов основана на сведении решения задачи математического программирования к нахождению седловой точки функции Лагранжа. Ограничением здесь является то, что эта седловая точка существует далеко не всегда. По существу эквивалентность решения задачи математического программирования и нахождения седловой точки функции Лагранжа устанавливается только для выпуклого программирования, да и то при дополнительных условиях (теорема Куна-Таккера). Если такая эквивалентность есть, то можно использовать методы множителей Лагранжа, которые представляют собой различные процедуры поиска седловой точки. В качестве такой процедуры можно взять, например, итеративный процесс как в методе проекции градиента, применив его для подъема по переменной $x$ и спуска по переменной $y$. Получим процесс

$$x^{k+1} = \Pr_R(x^k + \alpha_k \frac{\partial F(x^k, y^k)}{\partial x}),$$

$$y^{k+1} = \Pr_Y(y^k - \alpha_k \frac{\partial F(x^k, y^k)}{\partial y}),$$

где $\dfrac{\partial F(x^k, y^k)}{\partial x} = f'(x^k) + y^k G'(x^k), \quad \dfrac{\partial F(x^k, y^k)}{\partial y} = g(x^k),$

$g(x) = (g_1(x), \ldots, g_m(x)), \quad G'$ – матрица градиентов функций $g_i(x), i = 1,...,m, Y = \{y \mid y \geq 0\}.$

Проекция на множество $Y$ определяется весьма просто: $\Pr_Y(y)=(z_1, \ldots, z_m)$, где $z_i=\max(y_i, 0), i = 1,...,m$. Если $R$ – неотрицательный ортант, то проекция на него определяется аналогичным образом.

**Пример 2.8.** Найти максимум функции $f(x) = 2x_1 + 4x_2 - x_1^2 - 2x_2^2$ при ограничениях $x_1 + 2x_2 \leq 2, \quad 2x_1 - x_2 \leq 1, \quad x_1, x_2 \geq 0.$

Функция $f(x)$ является вогнутой, система нетривиальных ограничений содержит только линейные неравенства: $g_1(x) = 2 - x_1 - 2x_2 \geq 0, \quad g_2(x) = 1 - 2x_1 + x_2 \geq 0$. Поэтому задача является задачей выпуклого программирования, для ее решения можно использовать теорему Куна-Таккера. Составим функцию Лагранжа:

$F(x, y) = 2x_1 + 4x_2 - x_1^2 - 2x_2^2 + y_1(2 - x_1 - 2x_2) + y_2(1 - 2x_1 + x_2)$.

Запишем необходимые и достаточные условия для седловой точки функции Лагранжа

$$\frac{\partial F}{\partial x_1} = 2 - 2x_1 - y_1 - 2y_2 \le 0, \quad x_1 \frac{\partial F}{\partial x_1} = x_1(2 - 2x_1 - y_1 - 2y_2) = 0,$$
$$\frac{\partial F}{\partial x_2} = 4 - 4x_2 - 2y_1 + y_2 \le 0, \quad x_2 \frac{\partial F}{\partial x_2} = x_2(4 - 4x_2 - 2y_1 + y_2) = 0,$$
$$\frac{\partial F}{\partial y_1} = 2 - x_1 - 2x_2 \ge 0, \quad y_1 \frac{\partial F}{\partial y_1} = y_1(2 - x_1 - 2x_2) = 0,$$
$$\frac{\partial F}{\partial y_2} = 1 - 2x_1 + x_2 \ge 0, \quad y_2 \frac{\partial F}{\partial y_2} = y_2(1 - 2x_1 + x_2) = 0,$$
$$x_i \ge 0, y_j \ge 0, i, j = 1,2.$$

Чтобы найти вручную решение задачи надо перебрать все варианты этих условий. В данном случае это несложно, решение есть точка $x^*$=(2/3, 2/3), $y^*$=(2/3, 0), которая находится из системы

$$\begin{cases} 2x_1 + y_1 = 2, \\ 4x_2 + 2y_1 = 4, \\ x_1 + 2x_2 = 2, \\ 2x_1 - x_2 < 1, y_2 = 0. \end{cases}$$

Рассмотрим для этой задачи метод проекции градиента. Выберем начальную допустимую точку, например, $x^{(0)} = (0.5, 0.5)$, $y^{(0)} = (0, 0)$. Точка ( $x^{(0)}$, $y^{(0)}$) не является седловой точкой функции Лагранжа (т.к. не удовлетворяет достаточным условиям).

Найдем седловую точку, используя итеративный процесс проекции градиента, применив его для подъема по переменной $x$ и спуска по переменной $y$ с постоянным шагом $\alpha$=0.1, обеспечивающего сходимость процесса. Для этого нам потребуются следующие величины:

$$f(x) = 2x_1 + 4x_2 - x_1^2 - 2x_2^2,\ f'(x) = (2 - 2x_1, 4 - 4x_2),$$
$$g(x) = (g_1(x), g_2(x)) = (2 - x_1 - 2x_2, 1 - 2x_1 + x_2),$$
$$G' = \begin{pmatrix} -1 & -2 \\ -2 & 1 \end{pmatrix}.$$

1 итерация. $f'(x^{(0)}) = (1, 2)$, $g(x^{(0)}) = (0.5\,, 0.5)$,

$$\frac{\partial F(x^{(0)}, y^{(0)})}{\partial x} = f'(x^{(0)}) + y^{(0)}G' = (1, 2),$$
$$\frac{\partial F(x^{(0)}, y^{(0)})}{\partial y} = g(x^{(0)}) = (0.5, 0.5),$$
$$x^{(1)} = \mathrm{Pr}_R(x^{(0)} + \alpha\, \frac{\partial F(x^{(0)}, y^{(0)})}{\partial x}) = \mathrm{Pr}_R((0.5, 0.5) + 0.1(1, 2)) = (0.6, 0.7).$$
$$y^{(1)} = \mathrm{Pr}_Y(y^{(0)} - \alpha\, \frac{\partial F(x^{(0)}, y^{(0)})}{\partial y}) = \mathrm{Pr}_Y((0, 0) - 0.1(0.5\,, 0.5)) = (0, 0).$$

2 итерация. $$\frac{\partial F(x^{(1)}, y^{(1)})}{\partial x} = (0.8,\ 1.2),\ \ x^{(2)} = (0.68,\ 0.82),$$

$$\frac{\partial F(x^{(1)}, y^{(1)})}{\partial y} = (-1.2,\ 0.6),\ \ y^{(2)} = (0.12,\ 0).$$

Продолжая процесс, получим табл. 2.2.

Как видно, процесс сходится довольно медленно (его можно ускорить выбором шага по методу наискорейшего подъема-спуска), однако с точки зрения численной реализации он значительно более удобен для алгоритмизации, нежели перебор систем уравнений в условиях Куна-Таккера.

Таблица 2.2

| $k$ | $x^{(k)}$ | $y^{(k)}$ | $\frac{\partial F(x^{(k)}, y^{(k)})}{\partial x}$ | $\frac{\partial F(x^{(k)}, y^{(k)})}{\partial y}$ |
|---|---|---|---|---|
| 2 | (0.68, 0.82) | (0.12,0) | (0.52, 0.48) | (-0.32, 0.46) |
| 3 | (0.73, 0.87) | (0.15,0) | (0.39, 0.12) | (-0.47, 0.41) |
| 4 | (0.72, 0.83) | (0.2,0) | (0.36, 0.28) | (-0.38, 0.39) |
| 5 | (0.76, 0.86) | (0.24,0) | (0.24, 0.08) | (-0.48, 0.34) |
| 6 | (0.78, 0.87) | (0.29,0) | (0.15, -0.06) | (-0.52, 0.31) |
| 7 | (0.79, 0.86) | (0.34,0) | (0.08, -0.12) | (-0.51, 0.28) |
| 8 | (0.80, 0.85) | (0.39,0) | (0.01, -0.18) | (-0.50, 0.25) |
| 9 | (0.80, 0.83) | (0.44,0) | (-0.04, -0.20) | (-0.46, 0.23) |
| 10 | (0.80, 0.81) | (0.49,0) | (-0.09, -0.32) | (-0.42, 0.21) |
| 11 | (0.79, 0.78) | (0.53,0) | … | … |

4. *Методы штрафных функций.* Как видно из предыдущих рассмотрений, наличие ограничений существенно усложняет задачу поиска экстремума функции. Почти все изложенные методы решения экстремальных задач включали сложные специальные процедуры, учитывающие наличие ограничений (например, проектирование). В этом смысле особую группу образуют методы множителей Лагранжа, которые состоят в сведении задачи поиска экстремума с ограничениями к задаче поиска экстремума (точнее седловой точки, что примерно то же самое) без ограничений или с ограничениями простого вида, учет которых не составляет труда (это видно на примере операции проектирования на неотрицательный ортант). Однако, как говорилось, эти методы применимы практически только в выпуклом случае. От этого недостатка избавлены методы штрафных функций, которые позволяют в весьма общем случае сводить экстремальные задачи со сложными ограничениями к задачам без ограничений или с

простыми ограничениями, а в сочетании с методами поиска безусловного экстремума служат универсальным средством решения задач математического программирования (конечно, не лишенным недостатков).

Рассмотрим общую задачу математического программирования (2.3), (2.4):

$$\max_{x \in X} f(x), \quad X = \{x \mid x \in P, g_i(x) \geq 0, i = 1,...,m\}.$$

Методы штрафных функций основаны на описании множества $X$ с помощью функций со специальными свойствами. Одна группа этих методов использует функции вида $\Phi(x, C)$, которые обладают следующими свойствами:

$$\lim_{C_k \to \infty} \Phi(x, C_k) = \begin{cases} 0, & x \in X, \\ -\infty, & x \notin X. \end{cases}$$

В качестве такой функции можно взять, например, квадратичную функцию вида

$$\Phi(x, C_k) = -C_k \sum_{i=1}^{m} \left[\min\left(0, g_i(x)\right)\right]^2,$$

где $\varphi(x)=\sum_{i=1}^{m} \left[\min\left(0, g_i(x)\right)\right]^2$ – функция штрафа, $C_k$ ($k$=1,2,…) – параметры штрафа. Функция $\Phi(x, C_k)$ является дифференцируемой, если дифференцируемы функции $g_i(x)$. Можно показать, что если $f(x)$ и $g_i(x), i = 1,...,m$, непрерывны на замкнутом, ограниченном множестве $P$ и $X \neq \varnothing$, то

$$\max_{x \in X} f(x) = \lim_{\{C_k\} \to \infty} \max_{x \in P} [f(x) + \Phi(x, C_k)]. \tag{2.31}$$

При этом, если $x^0(C_k)$ реализует максимум в правой части (2.31), то любая предельная точка последовательности $\{x^0(C_k)\}$ есть решение задачи (2.3), (2.4).

Таким образом, с помощью метода штрафных функций общую задачу математического программирования можно свести приближенно (при достаточно большом $C_k$) к задаче без ограничений или с простыми ограничениями (на простом множестве $P$). Недостатком методов штрафных функций является ухудшение свойств вспомогательных задач (в правой части (2.31)) при больших значениях $C_k$, в частности, чувствительность к ошибкам вычислений. Эти недостатки преодолены в некоторых вариантах методов штрафных функций, в которых можно ограничиться конечными значениями штрафных констант $C_k$.

**Пример 2.9.** Найти минимум функции $f(x)=2x_1^2+x_2^2$ при ограничениях $g_1(x) = -x_1 - x_2 + 2 = 0, \quad g_2(x) = x_1 - 2x_2 + 1 = 0$.

Так как целевая функция $f(x)$ является выпуклой функцией, а ограничения, определяющие допустимое множество задачи, линейны, то решение $x^{(k)}$ вспомогательной задачи $f_k(x)=f(x)+\varphi_k(x)\to \min$ $\forall k=1,2,\ldots$ , где $\varphi_k(x)=k\varphi(x)$ ( $k = C_k$ ), может быть найдено из условия $f_k'(x^{(k)})=0$. Для данной задачи вспомогательная функция $f_k(x)$ имеет вид $f_k(x)=f(k)+k[g_1^2(x)+g_2^2(x)]=(2+2k)x_1^2+(1+5k)x_2^2-2kx_1x_2-2kx_1-8kx_2+5k$.

Решив систему уравнений

$$\begin{cases} \dfrac{\partial f_k(x)}{\partial x_1} = (4+2k)x_1 - 2kx_2 - 2k = 0, \\ \dfrac{\partial f_k(x)}{\partial x_2} = -2kx_1 + (2+2k)x_2 - 8k = 0, \end{cases}$$

находим $x_1^{(k)} = \dfrac{9k^2+k}{9k^2+12k+2}, \qquad x_2^{(k)} = \dfrac{9k^2+8k}{9k^2+12k+2}.$

Последние два равенства определяют точку безусловного минимума $x^{(k)}$ вспомогательной функции $f_k(x)$. Окончательно находим

$$x^* = \lim_{k\to\infty} x^{(k)} = (1,1), \quad f_{\min}=f(1,\ 1)=3.$$

## 2.6. Использование методов нелинейного программирования в эколого-экономической сфере

1. *Модель поведения производителей*

Рассмотрим модели принятия решений в производственных фирмах. Основная цель фирмы заключается в максимизации прибыли путем рационального распределения затрачиваемых ресурсов.

Пусть производственная фирма выпускает $n$ видов продукции, тогда годовой выпуск фирмы в натурально-вещественной форме – это вектор единиц продукции каждого вида. Для производства продукции фирма использует настоящий труд $L$ (среднее число занятых в год либо отработанные за год человеко-часы) и прошлый труд в виде средств труда $K$ (основные производственные фонды) и предметов труда $M$ (затраченные за год топливо, энергия, материалы комплектующие и т.п.). Каждый из этих трех агрегированных видов ресурсов имеет определенное число разновидностей (труд разной квалификации, оборудование различного вида и т.п.). Обозначим $x=(x_1,\ldots,x_m)$ вектор объемов затрачиваемых видов ресурсов (факторов производства). Тогда технология фирмы определяется ее **производственной функцией** $F(x)$, выражающей связь между затратами ресурсов и выпуском продукции. Если $n > 1$, то $F(x)$ есть $n$–мерная вектор-функция.

Дважды непрерывно дифференцируемая производственная функция называется **неоклассической**, если выполняются условия:

$1^0$ $F(x_1,\ldots,x_m)=0$, если существует $i$, $x_i=0$, т. е. при отсутствии одного из ресурсов производство невозможно;

$2^0$ $\dfrac{\partial F(x)}{\partial x_i} > 0,\ i=1,\ldots,m$, т.е. с увеличением объемов ресурсов выпуск растет;

$3^0$ $\dfrac{\partial^2 F(x)}{\partial x_i^2} < 0,\ i=1,\ldots,m$, т.е. с увеличением объемов ресурсов скорость роста выпуска уменьшается;

$4^0$ если существует $i$, $x_i \to +\infty$, то $F(x_1,\ldots,x_n) \to +\infty$, т. е. при неограниченном увеличении объема одного из ресурсов выпуск неограниченно растет.

Условие $3^0$ обычно используется в расширенной трактовке: матрица вторых производных (матрица Гессе) производственной функции отрицательно определена.

Обозначим $Z(x)$ функцию издержек (затрат), являющуюся скалярной величиной. Предполагается, что $\dfrac{\partial Z(x)}{\partial x_i} > 0,\ i=1,\ldots,m$, т. е. с увеличением объемов ресурсов затраты растут, и $Z(0)=0$, т. е. издержки равны нулю, если никакой ресурс не используется в производстве. Пусть $c=(c_1,\ldots,c_n)$ – вектор цен выпускаемой продукции каждого вида, тогда прибыль $PR(x)$ представляет собой разность между полученным фирмой доходом и издержками: $PR(x)=cF(x) - Z(x)$.

Рассмотрим линейную функцию затрат $Z(x)=px$, где $p=(p_1,\ldots,p_m)$ – вектор цен на ресурсы. Цены на ресурсы имеют естественный и понятный смысл: если $x_i$ – среднегодовое число занятых определенной профессии, то $p_i$ – годовая заработная плата одного работника данной профессии; если $x_i$ – покупные материалы (топливо, энергия), то $p_i$ – покупная цена единицы данного материала; если $x_i$ – производственные фонды определенного вида, то $p_i$ – годовая арендная плата за единицу фондов или стоимость поддержания единицы фондов в исправности.

В случае долговременного промежутка времени фирма может свободно выбирать вектор $x$ ,т.е. на переменные $x_1,\ldots,x_m$ не налагается никаких дополнительных ограничений кроме естественного условия неотрицательности. Тогда получаем задачу принятия решений о рациональном использовании ресурсов, максимизирущем прибыль

$$PR(x) = cF(x) - px \to \max_{x \ge 0}. \tag{2.32}$$

Это задача нелинейного программирования. В силу отрицательной определенности матрицы вторых производных задача (2.32) – задача выпуклого программирования с $n$

условиями неотрицательности. Поэтому условия Куна-Таккера, являющиеся необходимыми и достаточными условиями оптимальности, есть $\frac{\partial PR(x^0)}{\partial x} = c\frac{\partial F(x^0)}{\partial x} - p \le 0,\ x^0\frac{\partial PR(x^0)}{\partial x} = x^0(c\frac{\partial F(x^0)}{\partial x} - p) = 0,\ x^0 \ge 0.$ Согласно условию $1^0$ в производстве будут использованы все виды ресурсов, т.е. $x^0 > 0$, поэтому условия Куна-Таккера принимают вид

$$c\frac{\partial F(x^0)}{\partial x} = p, \tag{2.33}$$

или $c\frac{\partial F(x^0)}{\partial x_i} = p_i,\ i = 1,...,m$.

В случае краткосрочного промежутка времени фирма должна учитывать ограничения на объемы затрачиваемых ресурсов. Эти ограничения могут быть записаны в виде нелинейных, вообще говоря, неравенств: $g_k(x) \le b_k$. Тогда задача (2.33) решается с учетом этих дополнительных ограничений.

Может возникнуть ситуация, при которой фирма будет иметь ограничения по затратам на приобретение ресурсов. Пусть $V$ – объем денег, расходуемых не приобретение ресурсов. Тогда задача принятия решений о рациональном использовании ресурсов может быть сформулирована в виде задачи на максимум дохода (выручки) при заданном объеме издержек:

$$cF(x) \to \max, \quad px \le V, \quad x \ge 0, \tag{2.34}$$

Задача (2.34) также является задачей выпуклого программирования с одним линейным ограничением и $n$ условиями неотрицательности. Функция Лагранжа есть $L(x, y) = cF(x) + y(V - px)$. Необходимые и достаточные условия Куна-Таккера имеют вид

$$\frac{\partial L(x^0, y^0)}{\partial x} = c\frac{\partial F(x^0)}{\partial x} - y^0 p \le 0, \quad x^0\frac{\partial L(x^0, y^0)}{\partial x} = x^0(c\frac{\partial F(x^0)}{\partial x} - y^0 p) = 0,$$
$$y^0(V - px^0) = 0, \quad px^0 \le V, \quad x^0 \ge 0.$$

Множитель Лагранжа $y^0 > 0$, иначе ограничение $px^0 \le V$ становится несущественным. Но тогда ограничение $px \le V$ в оптимальной точке выполняется как равенство: $px^0 = V$. Учитывая, что $x^0 > 0$, условия Куна-Таккера примут вид

$$c\frac{\partial F(x^0)}{\partial x} = y^0 p. \tag{2.35}$$

Можно показать, что если $V=V^0=px^0$, где $x^0$ – решение задачи (2.33), то решения задач (2.33) и (2.34) совпадают.

Если фирма выпускает один вид продукции, производственная функция которого $f(x)$ есть скалярная величина, то необходимые и достаточные условия (2.33) есть

$c\frac{\partial f(x^0)}{\partial x} = p$, где $c$ – скаляр. В данном случае можно рассмотреть задачу максимизации выпуска при заданном объеме издержек: $f(x) \to \max, \quad px \le V, \quad x \ge 0.$ Тогда условия (2.35) примут вид $\frac{\partial f(x^0)}{\partial x} = y^0 p$.

Возможна ситуация, связанная с ограничениями по объему выпускаемой продукции, например, фирма получила заказ на производство продукции каждого вида. Если вектор $R=(R_1,\ldots,R_n)$ есть портфель заказов, то задача принятия решений о рациональном использовании ресурсов может быть сформулирована в виде задачи на минимум издержек при фиксированном объеме выпускаемой продукции:

$$Z(x) = px \to \min, \quad F(x) \ge R, \quad x \ge 0. \qquad (2.36)$$

В задаче (2.36) целевая функция линейна, т.е. является одновременно выпуклой и вогнутой функцией, а правая часть ограничения-равенства – вогнутая функция. Значит, задача (2.36) – задача выпуклого программирования. Задача (2.36) эквивалентна задаче

$$Z(x) = -px \to \max, \quad F(x) \ge R, \quad x \ge 0.$$

Функция Лагранжа для задачи на максимум есть $L(x, y) = -px + \bar{y}(F(x) - R)$, где $\bar{y} = (y_1,..., y_n)$ – вектор множителей Лагранжа. Необходимые и достаточные условия Куна-Таккера имеют вид

$$\frac{\partial L(x^0, \bar{y}^0)}{\partial x} = -p + \bar{y}^0 \frac{\partial F(x^0)}{\partial x} \le 0, \; x^0 \frac{\partial L(x^0, \bar{y}^0)}{\partial x} = x^0(-p + \bar{y}^0 \frac{\partial F(x^0)}{\partial x}) = 0,$$
$$\bar{y}^0(F(x^0) - R) = 0, \quad F(x^0) \ge R, \quad x^0 \ge 0.$$

Множители Лагранжа $y_j > 0$, $j=1,\ldots,n$, иначе ограничения системы $F(x^0) \ge R$ становятся несущественными. Но тогда ограничения системы $F(x) \ge R$ в оптимальной точке выполняются как равенства: $F(x^0) = R$. Учитывая, что $x^0 > 0$, условия Куна-Таккера примут вид $\bar{y}^0 \frac{\partial F(x^0)}{\partial x} = p$.

2. *Задача планирования природоохранных мероприятий в условиях производственной деятельности*

Основная цель планирования природоохранной деятельности определяется как обеспечение наиболее благоприятных условий жизнедеятельности общества, заключающихся как в усилении ресурсовосстановительного потенциала природы, так и в неуклонном снижении отрицательного воздействия производственной деятельности человека на окружающую среду.

Важную роль при этом играет разработка нормативов качества окружающей среды с целью установления для окружающей природной среды предельно допустимых норм воздействия, гарантирующих экологическую безопасность населения и сохранение генетического фонда, обеспечивающих рациональное использование и воспроизводство природных ресурсов в условиях производственной деятельности. При нарушении требований нормативов качества окружающей природной среды выброс, сброс вредных веществ или иные виды воздействия на окружающую среду могут быть ограничены, приостановлены или прекращены. Такое распоряжение осуществляется по предписанию специально уполномоченных на то государственных органов Российской Федерации в области охраны окружающей природной среды.

Промышленное предприятие, используя, например, какой-то один природный ресурс, в результате своей производственной деятельности воздействует на различные природные среды (природоохранные объекты): водные ресурсы, земля, воздух, лесные угодья и др. Поэтому предприятия, планирующие свою производственную деятельность, должны регулировать качество природоохранных объектов того региона, в котором живет и проявляет себя человек.

Так как природоохранная деятельность любого предприятия связана с дополнительными затратами, то прежде чем руководство предприятия примет решение относительно природоохранных мероприятий, необходимо определить общий размер дополнительных затрат предприятия на улучшение качества природной среды. Такая задача может решаться на региональном уровне, например, совместно с органами, в ведении которых находится как распределение природных и дефицитных ресурсов региона, так и контроль за качеством окружающей природной среды.

Пусть критерий эффективности предприятия, отражающей его интересы, представляет собой функцию прибыли $\Pi(x, S)$, зависящую от стратегии предприятия $(x, S)$. Здесь $x=(x_1,\dots,x_m)$ – факторы производства, $S$ – средства, расходуемые на природоохранные мероприятия, $S\in[0, Y]$, где $Y$ – некоторый объем денег предприятия, часть которого может быть инвестирована на развитие самого предприятия. Существует некоторая система выплат за использование или загрязнение природных ресурсов в виде некоторой непрерывно дифференцируемой функции $\psi(S,\omega)$, зависящей от величины $S$ и значения параметра $\omega\in\Omega$, задаваемого региональным центром. Параметр $\omega$ может быть вектором или скалярной величиной и может рассматриваться как мера снижения выплат на одну единицу вложенных средств. Функция $\psi(S,\omega)$ обладает следующими свойствами $\forall\omega\in\Omega$:

$$\psi(S,\omega)>0,\quad \psi'_S(S,\omega)<0,\quad \psi''_S(S,\omega)>0,\quad \lim_{S\to\infty}\psi(S,\omega)=0,\quad \psi(0,\omega)=\psi_0,$$

где $\psi_0$ – плата при отсутствии дополнительных вложений средств $S$, $\psi_0 \in (0, \Psi]$, $\Psi$ – предельно допустимый размер платы.

Предположим, что для предприятия, выпускающего $n$ видов продукции, критерием эффективности является прибыль с учетом платы за использование или загрязнение природных ресурсов

$$\Pi(x,S) = \sum_{j=1}^{m} c_j F_j(x, Y-S) - C(x, Y-S) - \psi(S, \omega), \qquad (2.37)$$

где $c_j$ – цена на продукцию $j$-го вида $j=1,\ldots,n$. Выпуск $j$-го вида продукции предприятия $F_j(x, Y{-}S)$ будем считать неоклассической производственной функцией, т.е. такой функцией, для которой выполнены условия:

$$\forall x_i = 0, i = 1,\ldots,m, \quad F_j(x, Y-S) = 0, \quad \frac{\partial}{\partial x_i} F_j(x, Y-S) > 0; \quad \frac{\partial^2}{\partial x_i^2} X_j(x, Y-S) < 0,$$

$$\forall i = 1,\ldots,m, \quad \lim_{x_i \to \infty} F_j(x, Y-S) = \infty, \quad \frac{\partial}{\partial S} F_j(x, Y-S) < 0, \quad \frac{\partial^2}{\partial S^2} F_j(x, Y-S) > 0.$$

Функция издержек производства $C(x,Y{-}S)$ считается непрерывно дифференцируемой, выпуклой, возрастающей функцией ресурсов $x_i$, $i=1,\ldots,m$, и затрат на развитие производства $Y-S$. Тогда принятие управляющих решений предприятия определяется условием

$$\Pi(x,S) = \sum_{j=1}^{n} c_j F_j(x, Y-S) - C(x, Y-S) - \psi(S, \omega) \to \max_{x \ge 0,\, S \in [0,Y]}. \qquad (2.38)$$

Учитывая предположения относительно $F_j(x,Y{-}S)$, $j=1,\ldots,n$, $C(x,Y{-}S)$ и $\psi(S,\omega)$, функция прибыли (2.37), как сумма вогнутых функций, является вогнутой на выпуклом допустимом множестве. Значит задача (2.38) представляет собой задачу выпуклого программирования с простыми ограничениями на переменные. Необходимые и достаточные условия оптимальности в такой задаче имеют вид

$$\frac{\partial \Pi(x^0, S^0)}{\partial x_i} \le 0, \; x_i^0 \frac{\partial \Pi(x^0, S^0)}{\partial x_i} = 0, \; i = 1,\ldots,m, \quad \frac{\partial \Pi(x^0, S^0)}{\partial S}(Y - S^0) S^0 = 0.$$

### Задачи и упражнения

**1.** Найти глобальный максимум функции $f$ на множестве $X$:

$$f = 3x_1 + 4x_2, \; X: \begin{cases} x_1^2 + x_2^2 \le 25, \\ x_1 x_2 \ge 4, \\ x_1, x_2 \ge 0. \end{cases}$$

**2.** Найти глобальный максимум функции $f$ на множестве $X$:

$$f = x_1 x_2,\ X: \begin{cases} 6x_1 + 4x_2 \ge 12, \\ 2x_1 + 3x_2 \le 24, \\ -3x_1 + 4x_2 \le 12, \\ x_1, x_2 \ge 0. \end{cases}$$

**3.** Найти глобальные экстремумы функции *f* на множестве *X*:

$$f = x_1^2 + x_2^2,\ X: \begin{cases} x_1 + x_2 \ge 2, \\ 2x_1 + 3x_2 \le 6, \\ x_1, x_2 \ge 0. \end{cases}$$

**4.** На двух предприятиях отрасли необходимо изготовить 200 изделий некоторой продукции. Затраты, связанные с производством $x_1$ изделий на первом предприятии, равны $4x_1^2$ у.е., а затраты, обусловленные изготовлением $x_2$ изделий на втором предприятии, составляют $20x_2 + 6x_2^2$ у.е. Определить, сколько изделий на каждом из предприятий следует произвести, чтобы общие затраты, обусловленные изготовлением необходимой продукции, были минимальными.

**5.** Найти глобальные экстремумы функции

$f=x_1^2+x_2^2+x_2^3$ при условиях $x_1+x_2+x_3 \le 15$, $x_1, x_2, x_3 \ge 0$.

**6.** Найти глобальный максимум функции

$f=x_1+x_2-1/4\, x_1^2 - 1/8\, x_2^2 - 1/8\, x_1 x_2$ при условиях $x_1+x_2 \le 1$, $x_1, x_2 \ge 0$.

**7.** Найти глобальный максимум функции

$f= x_1^{0,25} x_2^{0,75}$ при условиях $x_1+2x_2 \le 4$, $x_1, x_2 \ge 0$.

**8.** Найти глобальный максимум функции

$f=x_1^2+(x_2-5)^2$ при условиях $\begin{cases} 2x_1 - 5x_2 \le -7, \\ 3x_1 + x_2 \le 15, \\ 4x_1 + 7x_2 \le 71, \\ x_1, x_2 \ge 0. \end{cases}$

**9.** Найти глобальные экстремумы функции

$f=(x_1-3)^2+(x_2-2)^2$ при условиях $\begin{cases} x_1^2 + x_2^2 \le 36, \\ x_1, x_2 \ge 0. \end{cases}$

В задачах 10-12 найти минимум методом штрафных функций:

**10.** $f= x_1^2+x_2^2-20x_1-30x_2$ при условиях $\begin{cases} 2x_1 + 3x_2 - 13 = 0, \\ 2x_1 + x_2 - 10 = 0. \end{cases}$

**11.** $f= x_1^2+x_2^2-5x_1-4x_2$ при условии $2x_1+3x_2+x_3=6$.

**12.** $f= x_1^2-2x_2^2-x_2$ при условиях $\begin{cases} 2x_1+3x_2-6=0, \\ 2x_1+x_2-4=0. \end{cases}$

В задачах 13-14 найти минимум методом возможных направлений:

**13.** $f= x_1^2+2x_2^2-16x_1-20x_2$ при условиях $\begin{cases} 2x_1+5x_2 \le 40, \\ 2x_1+x_2 \le 16, \\ x_1, x_2 \ge 0. \end{cases}$

**14.** $f= x_1^2+x_2^2-x_1 x_2-3x_2$ при условиях $\begin{cases} x_1+x_2 \le 4, \\ x_1, x_2 \ge 0. \end{cases}$

**15.** Задачей выпуклого программирования на максимум является частный случай задачи математического программирования $f(x) \to \max$ $x \in X=\{x \in P | \ g_i(x) \ge 0, \ i=1,\ldots,m\}$, когда

а) функция цели $f(x)$ и функции ограничений $g_i(x)$, $i=1,\ldots,m$ вогнуты на выпуклом множестве $P$

б) функция цели $f(x)$ и функции ограничений $g_i(x)$, $i=1,\ldots,m$ выпуклы на выпуклом множестве $P$

в) функция цели $f(x)$ и функции ограничений $g_i(x)$, $i=1,\ldots,m$ вогнуты на множестве $P$

**16.** Необходимые и достаточные условия существования седловой точки у непрерывной функции $F(x,y)$ на произведении замкнутых ограниченных множеств $X \times Y$

а) $\max\limits_{x\in X} \min\limits_{y\in Y} F(x,y) = \min\limits_{y\in Y} \max\limits_{x\in X} F(x,y)$

б) $\max\limits_{x\in X} \min\limits_{y\in Y} F(x,y) \le \min\limits_{y\in Y} \max\limits_{x\in X} F(x,y)$

в) $\max\limits_{x\in X} \min\limits_{y\in Y} F(x,y) \ge \min\limits_{y\in Y} \max\limits_{x\in X} F(x,y)$

**17.** Найти седловую точку функции $F(x,y)=x-2y$, $x\in[1, 2]$, $y\in[1, 2]$

а) (1, 1) – седловая точка

б) (2, 2) – седловая точка

в) седловой точки нет

# Глава 3. ЛИНЕЙНЫЕ ЗАДАЧИ ОПТИМИЗАЦИИ

## 3.1. Элементы линейного программирования

Если целевая функция линейна и допустимое множество задано линейными ограничениями, то задача математического программирования называется **задачей линейного программирования**. Задачей линейного программирования (ЛП) в стандартной форме называется задача

$$cx \to \max, \quad Ax \le b,\ x \ge 0, \tag{3.1}$$

где $x \in R^n$, $c \in R^n$, $b \in R^m$, $A$ – матрица размера $m \times n$.

**Пример 3.1** (задача планирования производства). Пусть имеется три вида ресурсов $R_1$, $R_2$, $R_3$, которые могут быть использованы для производства двух видов товаров $T_1$, $T_2$. Для производства единицы товара $T_j$ необходимо затратить $a_{ij}$ количества ресурса $R_i$, $i$=1,2,3, $j$=1,2 Матрица $A=(a_{ij})_{3\times 2}$ называется технологической матрицей. Известна цена $c_j$ на товар $T_j$ и запасы $b_i$ ресурса $R_i$. Векторы $b=(b_1,b_2,b_3)$, $c=(c_1,c_2)$ называются, соответственно, векторами ресурсов и цен. Требуется выбрать такой план производства товаров, который в условиях заданных ограничений на ресурсы максимизирует валовое производство (стоимость выпущенной продукции).

Построим математическую модель задачи. Пусть $x_j$ – производимое количество товара вида $T_j$, тогда вектор $x=(x_1, x_2)$ – производственный план предприятия. На производство всех видов товаров ресурс $R_1$ будет израсходован в количестве $a_{11}x_1+a_{12}x_2$, ресурс $R_2$ – в количестве $a_{21}x_1+a_{22}x_2$ и ресурс $R_3$ – в количестве $a_{31}x_1+a_{32}x_2$. Стоимость выпущенной продукции равен $c_1x_1+c_2x_2$. По смыслу задачи ясно, что $x_1$ и $x_2$ неотрицательны. С учетом ограничений по запасам ресурсов получаем задачу

$$c_1x_1+c_2x_2 \to \max, \qquad a_{i1}x_1 + a_{i2}x_2 \le b_i,\ i=1,2,3, \quad x_1 \ge 0,\ x_2 \ge 0.$$

Для задачи ЛП используют следующую терминологию. Стратегию – переменную $x$ – называют планом. План, удовлетворяющий всем ограничениям, называют **допустимым планом**, точку максимума или минимума – **оптимальным планом**.

Математические модели задач ЛП могут иметь различную форму: на максимум и минимум, ограничения в виде равенств, неравенства «больше или равно» или «меньше или равно», или смешанных типов. Все эти задачи обладают общими свойствами и сводятся одна к другой стандартными преобразованиями. Например, в задаче ЛП в стандартной форме для записи неравенства в форме равенства требуется введение новой переменной определенного знака. Применив этот прием ко всем ограничениям-неравенствам, получаем задачу ЛП в канонической форме записи:

$$cx \to \max, \quad Ax = b,\ x \ge 0.$$

Для любой конкретной задачи ЛП логически возможны три случая:

а) допустимых планов нет, т.е. множество допустимых планов пусто, тогда задача не имеет решения;

б) множество допустимых планов непусто, но на этом множестве целевая функция не ограничена сверху (для задачи на максимум) или снизу (для задачи на минимум), тогда задача не имеет решения;

в) множество допустимых планов непусто, и целевая функция ограничена (соответственно сверху или снизу), тогда задача имеет решение (существует оптимальный план).

**Пример 3.2.** Решить графическим методом задачу ЛП

$$x_1+x_2 \to \max, \quad x_1+2x_2 \leq 1,\ 2x_1+x_2 \leq 1, \quad x_1 \geq 0,\ x_2 \geq 0.$$

Множество допустимых планов $X$ представляет собой четырехугольник с вершинами (0, 0), (0, ½), (½, 0), (1/3, 1/3). На рис. 3.1 множество $X$ заштриховано. Линии уровня целевой функции $x_1+x_2$ – семейство прямых $x_1+x_2=C$. Одна из линий уровня (на которой она принимает одно и то же значение) изображена на рис. 3.1 пунктирной прямой. Направление возрастания функции $x_1+x_2$ указывает вектор $f'(x)=(1,1)$, изображенный на рис. 3.1.

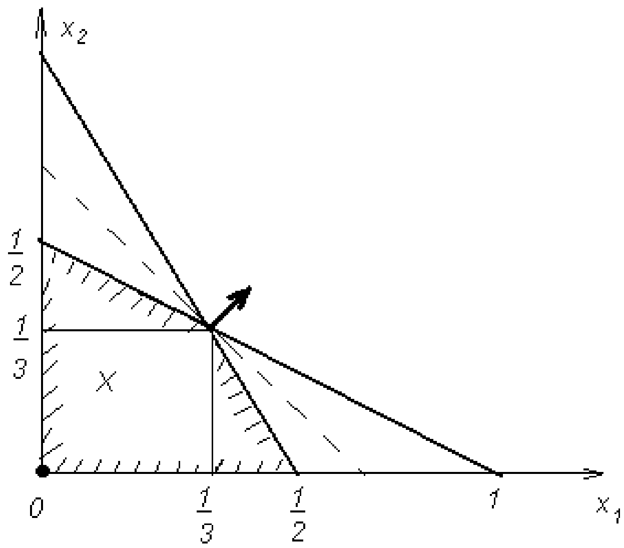


Рис. 3.1.

Так как множество $X$ замкнуто и ограничено, а целевая функция непрерывна, то по теореме Вейерштрасса задача имеет решение. Линия уровня функции $x_1+x_2$ на многоугольнике $X$ проходит через вершину с координатами $x_1^0 = x_2^0 = \frac{1}{3}$ и целевая функция в этой точке принимает максимальное значение.

Из геометрических представлений ясно, что если решение существует, то оно не может быть внутренне точкой, а должно принадлежать границе допустимого множества.

Множество решений системы линейных неравенств называется **выпуклой многогранной областью**. Если это множество ограничено, получим выпуклый многогранник.

Точка $x$ выпуклого множества $X$ называется **угловой**, или крайней, если не существует таких точек $x^1$, $x^2 \in X$, $x^1 \neq x^2$ таких, что $x=\alpha x^1+(1-\alpha)x^2$, $\alpha \in (0, 1)$. Например, для круга угловой является любая точка граничной окружности, для многоугольника – его вершины. Крайние точки выпуклой многогранной области называются **вершинами**. В задаче ЛП вершину допустимого множества называют также **опорным планом**.

**Теорема 3.1.** *Если задача линейного программирования имеет решение, то существует вершина допустимого множества, которая является оптимальным планом.*

В силу теоремы 3.1 для нахождения решения задачи ЛП достаточно ограничиться перебором угловых точек или вершин допустимого множества. На этом основан **симплекс-метод**, или метод последовательного улучшения плана. Алгоритм симплекс-метода осуществляет направленный перебор вершин, т.е. для задачи на максимум переходит от опорного плана с меньшим значением целевой функции к опорному плану с большим значением целевой функции. Этот процесс конечен, т.е. за конечное число шагов либо находится решение задачи ЛР, либо выясняется, что она не имеет решения.

*Алгоритм решения задачи ЛП симплекс-методом для задачи на максимум.*

Пусть система ограничений приведена к каноническому виду, в ней выделен допустимый базис (все правые части — неотрицательные числа), из целевой функции исключены базисные переменные и все слагаемые, кроме константы, перенесены в левую часть. Записав все данные в симплекс-таблицу, получим следующие результаты.

1. Если в последней строке *нет отрицательных оценок, то оптимальное решение достигнуто*.

2. Если в оценочной строке *есть хотя бы одна отрицательная оценка, то решение может быть улучшено*. Для этого выбирается разрешающий столбец (пусть он имеет номер $j$ ), содержащий отрицательную оценку, а в качестве разрешающего выбирается положительный элемент $a_{lj} > 0$, дающий минимум отношения элемента свободного члена $b_i$ к $a_{ij}$: $a_{lj} = \min\limits_{i, a_{ij} \geq 0} \left\{ \frac{b_i}{a_{ij}} \right\}$.

3. Если в симплекс-таблице *имеется отрицательная оценка, а в соответствующем столбце нет положительных элементов*, то исходная задача не имеет решения, т.е. $z_{\max} = cx = +\infty$.

4. Если оптимальное решение найдено, но при этом у одной (или нескольких) свободной переменной оценка равна 0, то задача имеет альтернативное решение, для получения

которого следует сделать шаг симплекс-метода, выбрав разрешающий элемент (по общему правилу) в свободном столбце с нулевой оценкой. При этом множество оптимальных решений является выпуклой комбинацией всех альтернативных решений.

В случае, когда рассматривается задача на *минимум*, то оптимальное решение достигнуто, если в последней строке *нет положительных оценок*. Если в строке оценок есть *хотя бы одна положительная оценка*, то решение может быть улучшено в соответствии с п. 2. Если же в симплекс-таблице имеется *положительная* оценка, а в соответствующем столбце нет положительных элементов, то исходная задача не имеет решения, т.е. $z_{\min} = -\infty$.

**Замечание.** Возможна ситуация, когда в заключительной симплекс-таблице в столбце свободной переменной, содержащем нулевую оценку, все элементы отрицательны. Это соответствует случаю, когда оптимальное множество неограниченно и не является выпуклой комбинацией угловых точек (например, луч).

Многие задачи ЛП можно сформулировать в виде так называемой *транспортной задачи*. Имеется $m$ пунктов производства некоторого продукта. В пункте $i$ продукт производится в количестве $a_i$, $i$=1,…,$m$. Имеется $n$ пунктов потребления продукта с потребностями $b_j$, $j$=1,…,$n$. Будем считать выполненным условие *баланса*: $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$.

Обозначим через $c_{ij}$ стоимость перевозки единицы продукции из $i$-го пункта производства в $j$-й пункт потребления. Пусть требуется выполнить перевозку продукта из пунктов производства в пункты потребления за минимальную стоимость. План перевозок можно задать матрицей $x=(x_{ij})_{m\times n}$, где $x_{ij}$ – количество перевозимого продукта из $i$-го в $j$-й пункт. Получаем задачу

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to \min,$$

$$\sum_{j=1}^{n} x_{ij} = a_i, \quad \sum_{i=1}^{mn} x_{ij} = b_j, \quad x_{ij} \ge 0, i = 1,...,m, \ j = 1,...,n.$$

Для решения транспортной задачи используется специальный метод: метод потенциалов, который является частным случаем симплекс-метода. Если все величины $a_i$ и $b_j$ целочисленны, то решение, определяемое методом потенциалов, также является целочисленным.

Построим двойственную задачу для задачи ЛП (3.1). Функция Лагранжа имеет вид $L(x, y) = cx - y(Ax - b)$. Найдем

$$\sup_{x\ge 0} L(x, y) = \sup_{x\ge 0}(cx - y(Ax - b)) = \sup_{x\ge 0}(cx - (A^T y)x + by) = by + \sup_{x\ge 0}(c - A^T y)x.$$

Если $(c - A^T y) \le 0$, то $\sup\limits_{x \ge 0}(c - A^T y)x = 0$. В противном случае существует положительная компонента вектора $(c - A^T y)$ и $\sup\limits_{x \ge 0}(c - A^T y)x = \infty$. Таким образом,

$$\sup_{x \ge 0}(cx - y(Ax - b)) = \begin{cases} by, \text{ если } (c - A^T y) \le 0, \\ \infty, \text{ если иначе.} \end{cases}$$

Поэтому двойственна задача $\min\limits_{y \in R_+^m} \sup\limits_{x \in P} L(x, y)$ эквивалентна задаче

$$by \to \min, \quad A^T y \ge c, \, y \ge 0. \tag{3.2}$$

Задача ЛП (3.2) является двойственной к задаче (3.1), которая является прямой задачей.

Двойственную задачу можно построить к задаче линейного программирования в любой форме. Общие правила построения двойственной задачи можно сформулировать следующим образом:

- если прямая задача – на максимум (минимум) критерия, то двойственная на минимум (максимум);
- если прямая задача – на максимум (минимум) критерия, то ограничения двойственной задачи имеют вид «больше либо равно» («меньше либо равно»);
- если прямая задача имеет ограничения неравенства (равенства), то в двойственной задаче соответствующие им переменные удовлетворяют условию неотрицательности (имеют произвольный знак);
- если в прямой задаче нет условия неотрицательности, то в двойственной задаче – ограничения равенства, иначе – ограничения неравенства.

Отметим, что задачи (3.1) и (3.2) **взаимодвойственные** или просто двойственные задачи ЛП, т.е. прямая задача является двойственной для двойственной задачи. Обозначим $X$ и $Y$ – допустимые множества взаимодвойственных задач и сформулируем для них теорему.

**Теорема 3.2** (теорема двойственности).

*1. Для любых допустимых планов $x \in X$ и $y \in Y$ выполняется $cx \ge by$.*

*2. Допустимые планы $x^0 \in X$ и $y^0 \in Y$ являются оптимальными тогда и только тогда, когда $cx^0 = by^0$.*

*3. (Свойство дополняющей нежесткости). Допустимые планы $x^0 \in X$ и $y^0 \in Y$ являются оптимальными тогда и только тогда, когда*

$$x_j^0(a^j y^0 - c_j) = 0, \; j = 1, \ldots, n, \quad y_i^0(a^i x^0 - b_i) = 0, \; i = 1, \ldots, m,$$

*где $a^i$ – строки матрицы $A$, $a^j$ – столбцы.*

*4. Прямая задача имеет решение тогда и только тогда, когда двойственная задача имеет решение.*

*5 Если $X \neq \varnothing$ и $Y \neq \varnothing$, то обе задачи имеют решение.*

*6. Если допустимое множество ровно одной из двух задач пусто: $X \neq \varnothing$, $Y = \varnothing$ (или $X = \varnothing$, $Y \neq \varnothing$), то целевая функция второй задачи не ограничена:* $\sup\limits_{x\in X} cx = \infty$ ($\inf\limits_{y\in Y} by = -\infty$).

Для задачи ЛП справедливо следующее утверждение.

**Теорема 3.3.** *Для того чтобы $x^0$ было решением задачи линейного программирования, необходимо и достаточно существования $y^0$ такого, что пара ($x^0$, $y^0$) является седловой точкой функции Лагранжа на прямом произведении неотрицательных ортантов.*

Достаточность доказана для общей задачи математического программирования, значит, в частности, и для линейного программирования. Необходимость следует из п. 2 теоремы двойственности (см. Теорему 3.2) в линейном программировании, так как если прямая задача имеет решение $x^0$, то и двойственная имеет решение $y^0$ и пара ($x^0$, $y^0$) является седловой точкой.

**Пример 3.3.** Найти решение задачи

$$f = 15x_1 + 6x_2 + 4x_3 \to \max,$$

$$3x_1 + 2x_2 - x_3 \le 9, \quad 5x_1 - 3x_2 + 4x_3 \le 25, \quad x_j \ge 0,\ j = 1,2,3.$$

Двойственная задача имеет вид

$$F = 9y_1 + 25y_2 \to \min,$$

$$3y_1 + 5y_2 \ge 15,\ 2y_1 - 3y_2 \ge 6, -y_1 + 4y_2 \ge 4,\ y_i \ge 0, \quad i = 1,2.$$

Графическое решение двойственной задачи дает следующий результат $y^0 = (7.2,\ 2{,}8)$, $F(y^0) = 134.8$. Подставляя полученное решение $y^0$ в систему ограничений двойственной задачи, находим $3\cdot 7.2 + 5\cdot 2.8 = 35.6 > 15$, $2\cdot 7.2 - 3\cdot 2.8 = 6$, $-7.2 + 4\cdot 2.8 = 4$. Первое ограничение выполняется строго, следовательно, переменная $x_1^0 = 0$. Так как $y_1^0 > 0$ и $y_2^0 > 0$, то оба ограничения прямой задачи выполняются как равенства в оптимальной точке. Подставляя $x_1^0 = 0$ в систему ограничений прямой задачи, получаем $2x_2 - x_3 = 9,\ -3x_2 + 4x_3 = 25$, откуда $x_2^0 = 12.2$, $x_3^0 = 15.4$. Решение прямой задачи $x^0 = (0, 12.2, 15.4)$, для которого значение целевой функции $f(x^0) = 134.8$.

Связь между двойственными задачами проявляется также при попытке одновременного решения взаимодвойственных задач симплекс-методом. При решении одной из задач с помощью симплекс-метода применение соответствующих преобразований к двойственной задаче называется *двойственным симплекс-методом*.

Двойственный симплекс-метод, как и обычный симплекс-метод, используется для решения задач линейного программирования. Но в отличие от обычного симплекс-метода его применяют в случае, когда среди свободных членов системы ограничений имеются отрицательные числа (при решении задачи обычным симплексным методом эти числа предполагаются неотрицательными), а оценки неотрицательные (для задачи на максимум). При этом $x=(b_1, b_2, \ldots, b_m, 0, \ldots, 0)$ есть *псевдоплан*.

*Алгоритм решения задачи ЛП двойственным симплекс-методом для задачи на максимум.*

Пусть система ограничений приведена к каноническому виду, в ней выделен базис, из целевой функции исключены базисные переменные и все слагаемые, кроме константы, перенесены в левую часть. Записав все данные в симплекс-таблицу, имеем следующие шаги алгоритма.

1. Если $b_i \geq 0$, $i = 1, \ldots, m$, то проверяем план на оптимальность. Если в последней строке *нет отрицательных оценок, то* план $x=(b_1, b_2, \ldots, b_m, 0, \ldots, 0)$ будет оптимальным решением исходной задачи.
2. Если же среди $b_i$, $i = 1, \ldots, m$, имеются отрицательные числа, то либо устанавливаем неразрешимость задачи – *если в псевдоплане* $x = (b_1, b_2, \ldots, b_m, 0, \ldots, 0)$ *есть хотя бы одно отрицательное число* $b_i < 0$ *такое, что все* $a_{ij} \geq 0$, *то задача не имеет решени,* – либо переходим к новому плану (правило перехода описано в п. 3).
3. Выбираем разрешающую строку как содержащую наибольшее по абсолютной величине отрицательное число в столбце свободных членов (пусть это строка со свободным членом $b_l$). Для выбора разрешающего столбца находим минимум модуля отношения элементов строки оценок к *отрицательным* элементам $l$-й строки. Пусть это минимальное значение принимается при $j = r$, тогда в базис вводят переменную $x_r$, а число $a_{lr}$ является разрешающим элементом. Переход к новой симплекс-таблице производят по обычным правилам симплексного метода.
4. Находим новый план и переходим к п. 1.

### 3.2. Дискретное программирование

Задачу оптимизации, в которой требуется, чтобы все или некоторые переменные принимали только целочисленные или дискретные значения, называют задачей целочисленного или дискретного программирования. Иначе говоря, задачей дискретного или **целочисленного программирования** называется задача $f(x) \to \max$, $x \in X$, $X$ – дискретное множество. Как правило, $X \subset \mathrm{Z}^n$ – множество точек с целочисленными координатами или $X \subset B^n$, $B=\{0, 1\}$ – множество точек, координаты которых принимают значение 0 или 1 (булевы переменные).

**Пример 3.4** (задача о назначениях)**.** Имеются $n$ сотрудников, которые могут выполнять $n$ работ, причем использование $i$-го сотрудника на $j$-й работе приносит доход $c_{ij}$ ($i,j=1,\ldots,n$). Предполагается, что каждый сотрудник выполняет только одну работу и каждую работу может выполнять только один сотрудник. Требуется поручить каждому сотруднику выполнение определенной работы так, чтобы максимизировать суммарный доход.

Введем $n^2$ переменных: $x_{ij}=1,$ если $i$-й сотрудник выполняет $j$-ю работу и $x_{ij}=0$ в противном случае. Условие, что каждый сотрудник выполняет только одну работу, можно записать в виде системы равенств $\sum_{j=1}^{n} x_{ij}=1, i=1,...,n$. Аналогично, каждую работу выполняет только один сотрудник тогда и и только тогда, когда выполняются равенства $\sum_{i=1}^{n} x_{ij}=1,\ j=1,...,n$. Получаем задачу

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} \to \max,$$
$$\sum_{j=1}^{n} x_{ij}=1, i=1,...,n, \quad \sum_{i=1}^{n} x_{ij}=1,\ j=1,...,n, \quad x_{ij}\in\{0,1\}. \tag{3.3}$$

**Пример 3.5** (задача коммивояжера)**.**

Имеется $n$ городов, занумерованных числами от 1 до $n$. Коммивояжер, выезжая из города 1, должен побывать в каждом городе ровно один раз и вернуться в исходный пункт. Известны расстояния между городами $c_{ij}$ ($i,j=1,\ldots,n,\ i\neq j$). Требуется найти самый короткий маршрут.

Введем переменные: $x_{ij}=1,$ если в маршрут входит переезд из города $i$ в город $j$, и $x_{ij}=0$ в противном случае. Сразу очевидны ограничения

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

которые отражают требования однократного въезда и выезда из каждого города. Однако эти ограничения еще полностью не описывают допустимые маршруты. Дело в том, что они не исключают возможности разрыва пути, т.е. появления нескольких не связанных между собой подмаршрутов для части городов. Чтобы исключить такую возможность, обычно вводят дополнительные переменные $u_i$, $i=1,\ldots,n,$ которые можно считать неотрицательными и целыми числами, и записывают еще $(n-1)^2-(n-1)$ ограничений

$$u_i - u_j + nx_{ij} \le n-1,\ i,\ j = 2,...,n,\ i \ne j\ . \tag{3.4}$$

Покажем, что ограничения (3.4) исключают возможность существования подмаршрутов. Действительно, если имеется подмаршрут, включающий $k$ городов и начинающийся и заканчивающийся в одном из них, то для соответствующих переменных, очевидно, выполняются соотношения $\sum_{i \in M}^{n} x_{ij} = 1, \quad j \in M$ , $\sum_{j \in M} x_{ij} = 1, \quad i \in M$ , где $M$ – множество номеров городов, образующих подмаршрут. Тогда, суммируя неравенства (3.4) для индексов $i,j \in M$ и учитывая, что все величины $u_i$ сокращаются вследствие связности и замкнутости подмаршрута, придем к противоречию: $kn \le k(n–1)$.

Необходимо еще показать, что ограничения (3.4) не исключают допустимый маршрут. Для этого положим число $u_i$ равным порядковому номеру города $i$ по маршруту следования коммивояжера. Тогда для тех пар ($i$, $j$), для которых $x_{ij}$=0, неравенства (3.4) вытекают просто из условий $1 \le u_i \le n$, откуда следует, что $u_i - u_j \le n–1$, а для тех пар ($i$, $j$), для которых $x_{ij}$=1, справедливы, очевидно, равенства $u_i - u_j = –1$ (город с номером $j$ следует за городом с номером $i$), и неравенства (3.4) выполняются как равенства.

Итак, задача коммивояжера состоит в минимизации целевой функции $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij}$ при ограничениях

$$\sum_{i=1}^{n} x_{ij} = 1,\ j = 1,...,n\ ,\ \sum_{j=1}^{n} x_{ij} = 1, i = 1,...,n\ ,\ (i \ne j)\ , \quad u_i - u_j + nx_{ij} \le n-1,\ i,\ j = 2,...,n,\ i \ne j\ ,$$

где переменные $x_{ij}$, $u_i$ – неотрицательные целые числа.

В приведенных примерах, а также во многих других прикладных задачах дискретной оптимизации критерий эффективности и функциональные ограничения являются линейными функциями. Отличие таких задач дискретной оптимизации от обычных задач линейного программирования состоит только в наличии дополнительных ограничений на значения переменных, целевая функция и основные ограничения по-прежнему линейны, поэтому имеются достаточные основания для включения их в линейное программирование, тем более что некоторые целочисленные задачи могут быть непосредственно сведены к задачам линейного программирования. Это целиком относится к задаче о назначениях. Действительно, если в этой задаче требование целочисленности переменных заменить условиями $x_{ij} \ge 0$, то она превращается, по существу, в транспортную задачу (только на максимум целевой функции), компоненты решения которой в силу имеющихся ограничений могут быть равны либо 0, либо 1.

Строго говоря, описанные задачи не относятся к задаче ЛП в их изначальном понимании, поэтому такие задачи часто называются просто задачами целочисленного (или дискретного) программирования. Однако их отличие от обычных задач ЛП состоит только в наличии дополнительных ограничений на значения переменных, целевая функция и основные ограничения по-прежнему линейны, поэтому имеются достаточные основания для включения их в линейное программирование, тем более что некоторые целочисленные задачи могут быть непосредственно сведены к задаче ЛП. Это целиком относится к задаче о назначениях. Действительно, если в этой задаче требование целочисленности переменных заменить условиями $x_{ij} \ge 0$, то она превращается, по существу, в транспортную задачу (на максимум целевой функции), найдя целочисленное решение которой описанными в предыдущем параграфе методами, решим задачу о назначениях (компоненты решения в силу имеющихся ограничений могут быть равны либо 0, либо 1).

**Венгерский метод**. Рассмотрим основные этапы венгерского метода на примере решения задачи о назначениях (3.3). Формально ее можно представить следующим образом: выбрать такую последовательность элементов $\{c_{1j_1}, c_{2j_2}, \ldots, c_{nj_n}\}$ из матрицы

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{pmatrix},$$

чтобы сумма $\sum_{k=1}^{n} c_{kj_k}$ была максимальна и при этом из каждой строки и столбца $C$ был выбран только один элемент. Введем следующие понятия.

**Определение 3.1.** Нулевые элементы $z_1$, $z_2, \ldots, z_k$ матрицы $C$ называются независимыми нулями, если для любого $i$, $1 \le i \le k$ строка и столбец, на пересечении которых расположен элемент $z_j$, не содержит другие такие элементы $z_j$.

**Определение 3.2.** Две прямоугольные матрицы $C$ и $D$ называются эквивалентными ($C \sim D$), если $c_{ij} = d_{ij} + \alpha_i + \beta_j$ для всех $i, j$.

Задачи о назначениях, определяемые эквивалентными матрицами, являются эквивалентными, т.е. оптимальный план одной из них будут оптимальным и для второй, и наоборот. Таким образом, под эквивалентностью здесь следует понимать то, что если ко всем элементам строки или столбца прибавить одно и то же число, то целевая функция изменится на это число, а оптимальный план не изменится. Действительно,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}x_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_j x_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}x_{ij} + \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n} x_{ij} + \sum_{j=1}^{n} \beta_j \sum_{i=1}^{n} x_{ij},$$

учитывая ограничения задачи (3.3) $\sum_{j=1}^{n} x_{ij} = 1,\ i = 1,\dots,n,\ \sum_{i=1}^{n} x_{ij} = 1,\ j = 1,\dots,n,$ имеем

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} x_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} x_{ij} + \sum_{i=1}^{n} \alpha_i + \sum_{j=1}^{n} \beta_j .$$

Алгоритм венгерского метода состоит из предварительного этапа и не более чем ($n$–2) последовательно проводимых итераций. Каждая итерация связана с эквивалентными преобразованиями матрицы, полученной в результате проведения предыдущей итерации, и с выбором максимального числа независимых нулей. Окончательным результатом итерации является увеличение числа независимых нулей на единицу. Как только количество независимых нулей станет равным $n$, проблема выбора оказывается решенной, а оптимальный вариант назначений определяется позициями независимых нулей в последней матрице. Приведем алгоритм для задачи на максимум.

Предварительный этап (нулевая итерация). Разыскивают максимальный элемент $\beta_j$ в $j$-м столбце и все элементы этого столбца последовательно вычитают из максимального. Эту операцию проделывают над всеми столбцами матрицы $C$. В результате образуется матрица с неотрицательными элементами, в каждом столбце которой имеется, по крайней мере, один нуль. Далее рассматривают $i$-ю строку полученной матрицы, разыскивают ее минимальный элемент $\alpha_i$ и из каждого элемента этой строки вычитают минимальный. Эту процедуру повторяют со всеми строками. На этом предварительный этап заканчивается. В результате получим матрицу $C_0$ ($C_0 \sim C$), в каждой строке и столбце которой имеется, по крайней мере один нуль. Описанный процесс преобразования $C$ в $C_0$ называется приведением матрицы.

I этап. Находим независимые нули $0^*$, начиная с произвольного нуля. Например, находим произвольный нуль в первом столбце и отмечаем его звездочкой “*”. Затем просматриваем второй столбец, и если в нем есть нуль, расположенный в строке, где нет нуля со звездочкой, то отмечаем его “*”. Аналогично просматриваем один за другим столбцы матрицы $C_0$ и отмечаем, если это возможно, следующие нули “*”. Очевидно, что нули матрицы $C_0$, отмеченные звездочкой, являются независимыми. Если в ней имеется ровно $n$ нулей со звездочкой, то процесс решения заканчивается. В противном случае переходим к этапу Ia или Iб. Перед началом следующего этапа знаком “+” выделяют столбцы матрицы, которые содержат нули со звездочками.

Ia этап. Невыделенный нуль (т.е. нуль, не попавший в столбец со знаком “+”) отмечаем штрихом, и выделяем знаком “+” строку, в которой он содержится.

Просматриваем эту строку, находим $0^*$ и убираем "+" выделения столбца, в котором содержится данный $0^*$. Если невыделенных нулей нет, то переходим к этапу II.

Iб этап. Невыделенный нуль (т.е. нуль, не попавший в столбец со знаком "+") отмечаем штрихом, и выделяем знаком "+" строку, в которой он содержится. Если строка, содержащая $0'$, не содержит $0^*$, то строим цепочку из нулей: от $0'$ к $0^*$ по столбцу и от $0^*$ к $0'$ по строке. Доказано, что этот процесс конечен, цепочка начинается и заканчивается $0'$. Далее $0'$ и $0^*$ меняем ролями, убираем все штрихи и знаки "+" и возвращаемся к этапу I.

II этап. К данному этапу переходят когда все нули выделены, т.е. содержатся в строке или столбце со знаком "+". Среди элементов матрицы, не принадлежащих выделенным стокам и столбцам, находим минимальный положительный элемент, вычтаем его из элементов, не принадлежащих выделенным стокам и столбцам и прибавляем его к элементам, расположенным на пересечении выделенных строк и столбцов. В результате появляются новые невыделенные нули. Переходим к этапу I.

**Пример 3.6.** Решить задачу о назначениях с матрицей

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

Предварительный этап. Разыскиваем максимальный элемент первого столбца, $\beta_1=7$. Вычитаем из него все элементы этого столбца. Аналогично для получения второго, третьего и четвертого столбцов новой матрицы вычитаем все элементы этих столбцов от $\beta_2=7$, $\beta_3=9$, $\beta_4=10$, соответственно. Получим матрицу $C'(C' \sim C)$: $C' = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 4 & 3 & 6 & 0 \\ 3 & 4 & 0 & 8 \\ 2 & 3 & 5 & 2 \end{pmatrix}$. В четвертой строке матрицы $C'$ нет нуля, отыскиваем минимальный элемент этой строки, $\alpha_4=2$. Вычитаем его из каждого элемента четвертой строки. Получим матрицу $C''(C'' \sim C)$: $C'' = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 4 & 3 & 6 & 0 \\ 3 & 4 & 0 & 8 \\ 0 & 1 & 3 & 0 \end{pmatrix}$. Так как в каждой строке $C''$ есть нуль, то $C'' = C_0$ и процесс приведения матрицы заканчивается. Далее ищем и отмечаем знаком “*” независимые нули в $C_0$, начиная с первого столбца: $C_0 = \begin{pmatrix} 0^* & 0 & 1 & 1 \\ 4 & 3 & 6 & 0^* \\ 3 & 4 & 0^* & 8 \\ 0 & 1 & 3 & 0 \end{pmatrix}$.

I этап. Выделяем знаком “+” первый, третий и четвертый столбцы матрицы $C_0$, которые содержат 0*: $C_0 = \begin{array}{c} \begin{array}{cccc} + & & + & + \end{array} \\ \begin{pmatrix} 0^* & 0 & 1 & 1 \\ 4 & 3 & 6 & 0^* \\ 3 & 4 & 0^* & 8 \\ 0 & 1 & 3 & 0 \end{pmatrix} \end{array}$.

На этапе Iа просматриваем невыделенный второй столбец, находим в нем невыделенный нуль $c_{12}$=0, отмечаем его штрихом и выделяем знаком “+” первую строку. В первой строке имеется нуль со звездочкой, расположенный в первом столбце. Поэтому уничтожаем знак “+” выделения этого столбца. Получаем матрицу

$$C_1 = \begin{array}{cc} \begin{array}{cccc} & & + & + \end{array} & \\ \begin{pmatrix} 0^* & 0' & 1 & 1 \\ 4 & 3 & 6 & 0^* \\ 3 & 4 & 0^* & 8 \\ 0 & 1 & 3 & 0 \end{pmatrix} & \begin{array}{c} + \\ \\ \\ \\ \end{array} \end{array}.$$

Просматриваем первый (ставший невыделенным) столбец и отыскиваем в нем невыделенный нуль. Этот нуль отмечаем штрихом и выделяем четвертую строку, содержащую такой нуль: $C_2 = \begin{array}{cc} \begin{array}{cccc} & & + & + \end{array} & \\ \begin{pmatrix} 0^* & 0' & 1 & 1 \\ 4 & 3 & 6 & 0^* \\ 3 & 4 & 0^* & 8 \\ 0' & 1 & 3 & 0 \end{pmatrix} & \begin{array}{c} + \\ \\ \\ + \end{array} \end{array}$.

В четвертой строке матрицы $C_2$ имеется невыделенный нуль, а нуля со звездочкой в этой строке нет. Отметим этот нуль штрихом: $C_2 = \begin{array}{cc} \begin{array}{cccc} & & + & + \end{array} & \\ \begin{pmatrix} 0^* & 0' & 1 & 1 \\ 4 & 3 & 6 & 0^* \\ 3 & 4 & 0^* & 8 \\ 0' & 1 & 3 & 0' \end{pmatrix} & \begin{array}{c} + \\ \\ \\ + \end{array} \end{array}$ и перейдем к этапу Iб.

На этапе Iб, начиная с элемента $c_{41}$ = 0′, строим цепочку, двигаясь от него по столбцу. Находим нуль со звездочкой $c_{11}$ = 0*, далее от него движемся вдоль первой строки и находим $c_{12}$=0′. Таким образом, цепочка построена: $0'_{41}$–$0^*_{11}$–$0'_{12}$. Заменяем штрих на звездочку и уничтожаем звездочки над четными элементами цепочки, а также все знаки выделения столбцов и строк: $C_3 = \begin{pmatrix} 0' & 0^* & 1 & 1 \\ 4 & 3 & 6 & 0^* \\ 3 & 4 & 0^* & 8 \\ 0^* & 1 & 3 & 0' \end{pmatrix}$.

В результате число независимых нулей (0*) увеличилось на единицу и стало равно 4, т.е. равно размерности матрицы $C$ и поэтому алгоритм заканчивает работу. Искомые элементы назначения соответствуют позициям независимых нулей матрицы $C_3$ (т.е. 0*). Соответствующее значение целевой функции $c_{12} + c_{24} + c_{33} + c_{41}$=7+10+9+5=31.

Однако в общем случае целочисленная задача не сводится к обычной задаче линейного программирования. Примером служит задача о коммивояжере. О возникающих тут сложностях говорит тот факт, что одной этой задаче посвящено множество книг.

Среди методов решения целочисленных задач можно выделить еще два направления: методы отсечения и комбинаторные методы.

В **методах отсечения** для решения целочисленных задач используются методы решения непрерывных задач. Рассмотрим это на примере задачи ЛП (3.1), в которой $x \in Z^n$. Сначала решается задача ЛП без условия целочисленности., которую назовем задачей 1′. Если найденное решение $x^1$ целочисленно, то оно является решением исходной задачи. Если нет, то к ограничениям задачи 1′ добавляется дополнительное ограничение, которое отсекает $x^1$ и сохраняет в допустимом множестве исходной задачи все его целочисленные точки. Такое ограничение называется правильным отсечением. На следующем этапе находится решение задачи с дополнительно введенным ограничением, которую назовем задачей 2′. Если решение $x^2$ не является целочисленным, то для него вводится правильное отсечение. И так далее до тех пор, пока решение очередной задачи ЛП не окажется целочисленным.

Конкретные способы построения правильных отсечений приводят к конкретным вычислительным алгоритмам. Наиболее известны алгоритмы Гомори.

Возьмем уравнение из симплексной таблицы, содержащей оптимальный план. Базисная переменная в этом уравнении имеет наибольшую дробную часть.

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \tag{3.5}$$

причем некоторые коэффициенты $a_{ij}$, $b_i$ являются дробными. Поскольку $x_j$ – неотрицательные целые числа, то к ограничениям исходной задачи можно добавить новое ограничение $\sum_{j=1}^{n}[a_{ij}]x_j \le [b_i]$. Знак "≤", так как $\sum[a_{ij}]x_j$ – целое и его всегда можно ограничить целой частью $b_i$. Откуда следует, что неравенство сохраняет все целые точки многограного множества $X$. Можно преобразовать последнее неравенство к виду

$$\sum_{j=1}^{n}[a_{ij}]x_j + x_{n+1} = [b_i], \tag{3.6}$$

добавив свободную переменную $x_{n+1}$ с условием ее неотрицательности и целочисленности. Условие (3.6) будет выполняться при целых $x_j$. Вычитая из (3.6) уравнение (3.5), получаем ограничение, содержащее дробные части:

$$\sum_{j=1}^{n}(-\{a_{ij}\})x_j + x_{n+1} = -\{b_i\}. \tag{3.7}$$

Уравнение (3.7) называется **отсечением Гомори**. Заметим, что (3.7) уже не содержит базисную переменную из (3.5), а новая базисная переменная $x_{n+1}$ будет выражена через прежние свободные.

Из (3.7) следует неравенство

$$\sum_{j=1}^{n}\{a_{ij}\}x_j \ge \{b_i\}. \tag{3.8}$$

Нецелочисленный план не удовлетворяет (3.8), т.к. $\forall i\ \{b_i\} > 0$, а левая часть (3.8) равна нулю ($\sum_{j=1}^{n}\{a_{ij}\}x_j = 0$), т.к. свободные переменные равны нулю. Любой целочисленный план удовлетворяет (3.8) как строгому равенству, т.к. в этом случае $\forall i\ \{b_i\} = 0$.

**Пример 3.7.** Найти оптимальное решение задачи

$$\max\ (x_1 + x_2),$$

$$x_1 + 3x_2 \le 10, \qquad 4x_1 + x_2 \le 12, \qquad x_1, x_2 \ge 0, \text{ целые.}$$

Шаг 1. Решаем задачу без условия целочисленности симплекс-методом.

Приведем задачу к каноническому виду:

$$\max\ (x_1 + x_2),$$

$$x_1 + 3x_2 + x_3 = 10,\ 4x_1 + x_2 + x_4 = 12,\ x_1 \ge 0,\ x_2 \ge 0.$$

Составим симплекс-таблицу:

| БП | $b$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|---|
| $x_3$ | 10 | 1 | 3 | 1 | 0 |
| $x_4$ | 12 | $\underline{4}$ | 2 | 0 | 1 |
| $f$ | 0 | -1 | -1 | 0 | 0 |

Согласно алгоритму симплекс-метода выбираем разрешающий элемент: $\min\{\frac{10}{1}, \frac{12}{4}\} = 3$, т.е. разрешающий элемент 4 (подчеркнут). Выполняем сиплекс-преобразования:

| БП | $b$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|---|
| $x_3$ | 7 | 0 | <u>11/4</u> | 1 | -1/4 |
| $x_1$ | 3 | 1 | 1/4 | 0 | 1/4 |
| $f$ | 3 | 0 | -3/4 | 0 | 1/4 |

Итоговая симплексная таблица имеет вид

| БП | $b$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|---|
| $x_2$ | 28/11 | 0 | 1 | 4/11 | -1/11 |
| $x_1$ | 26/11 | 1 | 0 | -1/11 | 3/11 |
| $f$ | 54/11 | 0 | 0 | 3/11 | 2/11 |

Оптимальный нецелочисленный план есть $x_1$ = 26/11, $x_2$ = 28/11. Первая строка таблицы содержит нецелую базисную переменную с наибольшей дробной частью.

$x_2 + \frac{4}{11}x_3 - \frac{1}{11}x_4 = \frac{28}{11}$. Преобразуем уравнение к виду (3.7):

$0 \cdot x_2 - \frac{4}{11}x_3 - \frac{10}{11}x_4 + x_5 = -\frac{6}{11}$, т.к. $-\frac{1}{11} = -1 + \frac{10}{11}$, $-1$ – целая часть числа $-\frac{1}{11}$;

$\frac{10}{11}$ – дробная часть числа $-\frac{1}{11}$. Получаем

$$-\frac{4}{11}x_3 - \frac{10}{11}x_4 + x_5 = -\frac{6}{11} \qquad (3.9)$$

Шаг 2: Решим задачу ЛП с дополнительным ограничением (3.9) двойственным симплекс-методом.

| БП | $b$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|---|---|
| $x_2$ | 28/11 | 0 | 1 | 4/11 | -1/11 | 0 |
| $x_1$ | 26/11 | 1 | 0 | -1/11 | 3/11 | 0 |
| $x_5$ | -6/11 | 0 | 0 | -4/11 | <u>-10/11</u> | 1 |
| $f$ | 54/11 | 0 | 0 | 3/11 | 2/11 | 0 |

Согласно алгоритму двойственного симплекс-метода определяем разрешающий элемент: $\min\{\frac{2/11}{10/11}, \frac{3/11}{4/11}\} = \frac{1}{5}$, т.е. разрешающий элемент $-\frac{10}{11}$. Делаем симплекс-пеобразования как в обычном симплекс-методе:

| БП | $b$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|---|---|
| $x_2$ | 13/5 | 0 | 1 | 2/10 | 0 | -1/5 |
| $x_1$ | 11/5 | 1 | 0 | -1/5 | 0 | 3/10 |
| $x_4$ | 3/5 | 0 | 0 | 2/5 | 1 | -11/10 |
| $f$ | 24/5 | 0 | 0 | 1/2 | 1/5 | 1/5 |

Полученный оптимальный план содержит нецелые компоненты. Построим новое отсечение.

Шаг 3. Новое отсечение построим из первого ограничения, т.к. базисная переменная $x_4$ имеет наибольшую дробную часть.

$x_4 + \frac{2}{5}x_3 - \frac{11}{10}x_5 = \frac{3}{5}$ , $0 \cdot x_4 - \frac{2}{5}x_3 - \frac{9}{10}x_5 + x_6 = -\frac{3}{5}$ , т.к. $-\frac{11}{6} = -2 + \frac{9}{10}$. Получаем

$$-\frac{2}{5}x_3 - \frac{9}{10}x_5 + x_6 = -\frac{3}{5}. \tag{3.10}$$

Добавляем (3.10) к ограничениям в последнюю симплексную таблицу:

| БП | $b$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|---|---|
| $x_2$ | 13/5 | 0 | 1 | 2/10 | 0 | -1/5 | 0 |
| $x_1$ | 11/5 | 1 | 0 | -1/5 | 0 | 3/10 | 0 |
| $x_4$ | 3/5 | 0 | 0 | 2/5 | 1 | -11/10 | 0 |
| $x_6$ | -3/5 | 0 | 0 | -9/10 | 0 | <u>-2/5</u> | 1 |
| $f$ | 24/5 | 0 | 0 | 1/2 | 1/5 | 1/5 | 0 |

Решаем задачу двойственным симплекс-методом. Разрешающий элемент определяется согласно $\min\{\frac{1/2}{9/10}, \frac{1/5}{2/5}\} = \frac{1}{2}$, т.е. разрешающий элемент $-\frac{2}{5}$. Делаем симплексное преобразование. Продолжаем процесс до тех пор, пока не получим полностью целочисленное решение. В нашем случае решение задачи имеет вид $x_1 = 1$, $x_2 = 3$, $x_3 = 0$, $x_4 = 5$, $x_5 = 4$, $x_6 = 3$, $x_7 = 2$, $x_8 = 0$. Так как все компоненты целые, то $x_1 = 1$, $x_2 = 3$ – решение исходной задачи целочисленного линейного программирования.

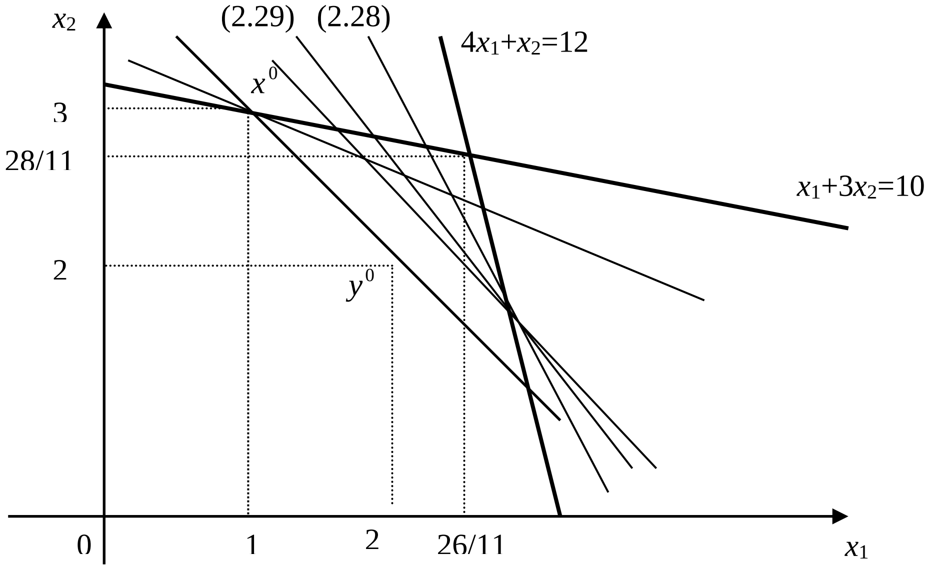


Рис. 3.2

На рис. 3.2 представлено графическое решение задачи, отмечены отсечения (3.9) и (3.10), полученные на шаге 2 и 3. Решение задачи $(3,1)$, полученное методом Гомори, обозначено $x^0$. Второе целочисленное решение $(2, 2)$ обозначено на том же рисунке $y^0$. Оба решения принадлежат одному отсечению.

Представление о комбинаторных методах дает наиболее известный и широко используемый на практике **метод ветвей и границ**.

Основная идея метода состоит в разбиении допустимого множества на подмножества (ветвление) и вычислении оценок (границ) значений целевой функции на этих подмножествах, позволяющие отбрасывать подмножества, заведомо не содержащие решений. Графически указанный процесс можно представить структурной схемой в виде дерева, у которого каждая вершина соответствует некоторому подмножеству планов (и, обычно, связанной с ним вспомогательной задаче ЛП), а выходящие из нее ветви отвечают разбиению этого подмножества на новые подмножества планов (и связанным с ними задачам).

Для решения задачи целочисленного программирования методом ветвей и границ можно использовать алгоритм Лэнда-Дойга. Прежде всего, решается исходная задача ЛП без условия целочисленности. Если в полученном решении все переменные целочисленные, то задача решена. Если некоторая компонента $x_k$ нецелочисленная, т.е. $x_k=[x_k]+\{x_k\}$, где $[x_k]$ – целая часть $x_k$, $\{x_k\}$ – дробная часть $x_k$, то исходная задача разбивается на две подзадачи ЛП. Одна подзадача решается дополнительным ограничением $x_k \leq [x_k]$, а другая – с дополнительным ограничением $x_k \geq [x_k]+1$. Вычисляется оценка (значение целевой функции). Если одна из этих двух подзадач, например, с ограничением $x_k \leq [x_k]$, дает нецелочисленное решение с компонентой $x_r=[x_r]+\{x_r\}$, то переходим к следующему шагу, на котором выбранная подзадача разветвляется на две новые, одна – с дополнительным ограничением $x_r \leq [x_r]$, а другая – с дополнительным ограничением $x_r \geq [x_r]+1$ соответственно. Вычисления заканчиваются, когда кандидатов на ветвление больше нет.

Описанный алгоритм применим и для задач, в которых содержится требование, чтобы переменная принимала только значения 0, 1.

**Пример 3.8.** Найти решение задачи из примера 3.7. методом ветвей и границ.

Оптимальное решение $x^0$ = (26/11, 28/11) является вектором с нецелыми компонентами (см. рис 3.2). Согласно схеме, решаем две задачи ЛП, одну с ограничением $x_1 \leq 2$, вторую с ограничением $x_1 \geq 3$.

Шаг 2а: Добавляем ограничение $x_1 \leq 2$ и получаем подзадачу

$$\max (x_1 + x_2),$$

$$x_1 + 3x_2 \leq 10,\ \ 4x_1 + x_2 \leq 12,\ x_1 \leq 2,\ x_1 \geq 0,\ x_2 \geq 0.$$

Получили решение: $x_1 = 2$, $x_2 = 8/3$.

Шаг 2б: Добавляем ограничение $x_1 \geq 3$ и получаем подзадачу

$$\max (x_1 + x_2),$$

$$x_1 + 3x_2 \leq 10,\ \ 4x_1 + x_2 \leq 12,\ x_1 \geq 3,\ x_2 \geq 0.$$

Здесь $x_1 = 3$, $x_2 = 0$, $f_{2б} = 3$.

Решение задачи шага 2а нецелочисленное, поэтому решаем две задачи: одну с ограничением $x_2 \leq 2$, другую с $x_2 \geq 3$ (при этом остается $x_1 \leq 2$).

Шаг 3а: $\max (x_1 + x_2)$,

$$x_1 + 3x_2 \leq 10,\ \ 4x_1 + x_2 \leq 12,\ x_1 \leq 2,\ x_2 \leq 2,\ x_1 \geq 0,\ x_2 \geq 0.$$

Получили решение: $x_1 = 2$, $x_2 = 2$, $f_{3а} = 4$.

Шаг 3б: $\max (x_1 + x_2)$,

$$x_1 + 3x_2 \leq 10,\ \ 4x_1 + x_2 \leq 12,\ x_1 \leq 2,\ x_2 \geq 3,\ x_1 \geq 0,\ x_2 \geq 0.$$

Эта подзадача решений не содержит, т.к. система ограничений не совместна.

Возвращаемся к шагу 1 и начинаем ветвление по нецелой компоненте $x_2$ = 28/11.

Необходимо решить две задачи: одну с ограничением $x_2 \leq 2$, вторую с $x_2 \geq 3$.

Шаг 2в: Добавляем ограничение $x_2 \leq 2$ и получаем подзадачу

$$\max (x_1 + x_2),$$

$$x_1 + 3x_2 \leq 10,\ \ 4x_1 + x_2 \leq 12,\ x_2 \leq 2,\ x_1 \geq 0,\ x_2 \geq 0.$$

Получили решение $x_1 = 5/2$, $x_2 = 2$.

Шаг 2г: Добавляем ограничение $x_2 \geq 3$ и получаем подзадачу

$\max (x_1 + x_2)$,

$x_1 + 3x_2 \le 10,\ \ 4x_1 + x_2 \le 12, x_2 \ge 3, x_1 \ge 0.$

Получили решение $x_1 = 1, x_2 \ge 3, f_{2д} = 4$.

Так как на шаге 2в получено нецелое решение, выполним шаг 3в и 3г.

Шаг 3в: $\max (x_1 + x_2)$,

$x_1 + 3x_2 \le 10, 4x_1 + x_2 \le 12, x_1 \le 2, x_2 \le 2, x_1 \ge 0, x_2 \ge 0.$

Получили решение: $x_1 = 2, x_2 = 2, f_{3в} = 4$.

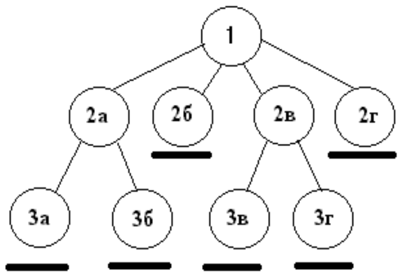


Рис. 3.3

Шаг 3г: $\max (x_1 + x_2)$,

$x_1 + 3x_2 \le 10,\ \ 4x_1 + x_2 \le 12, x_1 \ge 3, x_2 \le 2, x_1 \ge 0, x_2 \ge 0.$

Эта подзадача решений не содержит, т.к. система ограничений не совместна.

Алгоритм закончен, так как все ветви окончились целыми решениями (рис. 3.3).

Сравнивая подчеркнутые элементы схемы находим, что оптимальными являются решения шагов 3а, 2г, 3в, причем на 3а и 3в решения совпадают. На шаге 2б значение целевой функции не максимальное. Итак, (2, 2) и (1, 3) есть оптимальные решения задачи.

### 3.3. Анализ чувствительности. Некорректные задачи

При численном решении задачи ЛП (как и любых других) возникает проблема оценки точности найденного решения. Как вследствие неточности исходных данных (параметров модели), так и вследствие ошибок округления в процессе счета фактически происходит замена исходной задачи

$$\max cx \text{ при ограничениях } Ax \le b,\ x \ge 0 \tag{3.11}$$

возмущенной задачей

$$\max c(\delta)x \text{ при ограничениях } A(\delta)x \le b(\delta),\ x \ge 0, \tag{3.12}$$

где $A(\delta)$, $b(\delta)$, $c(\delta)$ отличаются от $A$, $b$, $c$ по норме на величину не более $\delta$. При этом совершенно не гарантировано, что для малого $\delta$ решение задачи (3.12) будет отличаться по норме от решения задачи (3.11) на малую величину $\varepsilon$ и даже вообще существовать.

Пусть задача (3.11) имеет решение, обозначим множество ее решений через $X^*$, экстремальное значение – $f^*$, а для задачи (3.12) – $X^*(\delta)$ и $f^*(\delta)$.

Задача (3.11) называется *устойчивой*, если существует такое $\delta_0$, что для всех $0\le\delta\le\delta_0$ задача (3.12) имеет решение.

Можно показать, что если задача устойчива, то для любого $\varepsilon > 0$ существует такое $\delta_0$, что $|f^* - f^*(\delta)| \le \varepsilon$ и $X^*(\delta)$ находится в $\varepsilon$-окрестности $X^*$ для любого $0\le\delta\le\delta_0$ , т.е. задача устойчива по функционалу и по решению. При этом задача в стандартной форме (3.11) устойчива тогда и только тогда, когда существуют такие вектора $x^*$, $y^*$, что $Ax^*<b$, $x^*>0$, $A^Ty^*>c$, $y^*>0$, т.е. в прямой и двойственной задачах существуют внутренние точки множеств допустимых планов (для задачи в канонической форме условия несколько иные).

Для устойчивой задачи можно провести анализ чувствительности, а именно, оценить изменение экстремального значения целевой функции при малых изменениях параметров. При достаточно малом изменении в задаче (3.11) вектора коэффициентов целевой функции, очевидно, не изменится хотя бы один оптимальный план $x^*$, поэтому $f^*-f^*(\delta)=\langle c-c(\delta),x^*\rangle$. Аналогично, при достаточно малом изменении в задаче (3.11) вектора правой части ограничений, который является вектором коэффициентов целевой функции двойственной задачи, не изменится хотя бы один оптимальный план двойственной задачи $y^*$. По теореме двойственности экстремальные значения двойственных задач равны, поэтому $f^*-f^*(\delta)=\langle b-b(\delta),y^*\rangle$. Малая вариация каждого коэффициента матрицы $A$ может изменить оптимальные планы, но при этом приблизительно $f^*-f^*(\delta)=-(a_{ij}-a_{ij}(\delta))\ x_j^* y_i^*$.

**Пример 3.9.** Исследовать на чувствительность задачу

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

с тремя перенумерованными ограничениями (не считая простейших)

$$x_1+2x_2 \le 4, \qquad (1)$$

$$5x_1+2x_2 \le 10, \qquad (2)$$

$$4x_1+3x_2 \le 12, \qquad (3)$$

$$x_1 \ge 0,\ x_2 \ge 0.$$

Решение данной задачи ЛП имеет вид $x^0=(1{,}5\ ,\ 1{,}25)$, $f^0=7$ (см. рис. 3.4). Двойственной к ней является задача

$$g(y)=4y_1+10y_2+12y_3 \to \min,$$

$$y_1+5y_2+4y_3 \ge 3,$$

$$2y_1+2y_2+3y_3 \ge 2,$$

$$y_1 \ge 0,\ y_2 \ge 0,\ y_3 \ge 0.$$

Так как ограничение (3) в прямой задаче на решении не является активным (выполняется как строгое неравенство), а обе компоненты решения $x^0$ отличны от нуля, то по теореме о дополняющей нежесткости в решении двойственной задачи $y_3=0$, а ограничения должны быть активными. Поэтому решение двойственной задачи есть $y^0=(0{,}5\ ,\ 0{,}5\ ,\ 0)$, $g^0=7$.

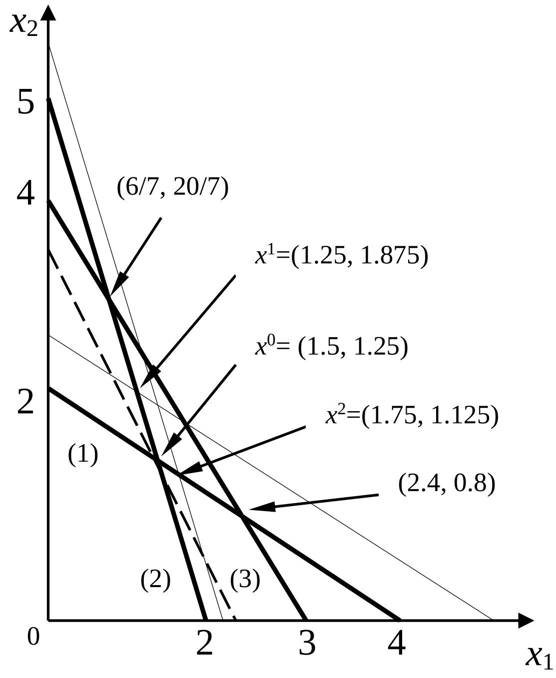


Рис. 3.4

Проведем анализ чувствительности решения исходной задачи при изменении коэффициентов функции цели $f(x)=c_1x_1 + c_2x_2$. Для этого найдем угловые коэффициенты целевой функции и функций ограничений: $k_0= -1{,}5$ , $k_1= -0{,}5$ , $k_2= -2{,}5$ , $k_3= -2$. Пусть изменению подвергается коэффициент $c_1$, а $c_2$ по условию равен 2. Тогда $k_0(c_1)= -c_1/2$. Для того чтобы оптимальный план оставался тем же, должно выполняться условие $0{,}5\le c_1/2 \le 2{,}5$, т.е. $c_1\in[1, 5]$. Пусть изменению подвергается коэффициент $c_2$, а $c_1$ по условию равен 3. Тогда $k_0(c_2)= -3/c_2$. Для того чтобы оптимальный план оставался тем же, должно выполняться условие $0{,}5 \le 3/c_2 \le 2{,}5$, т.е. $c_2\in[1{,}2 , 6]$. Значит, при изменении $c_1$ и $c_2$ в указанных пределах решение задачи не меняется.

Проведем анализ чувствительности решения исходной задачи при изменении свободных членов ограничений. Пусть $b_1=4$ в (1) увеличилось на 1, т.е. $\Delta b_1=1$. Тогда оптимальный план новой задачи ЛП получается из решения системы уравнений $x_1+2x_2=5$, $5x_1+2x_2=10$: $x^1=(1{,}25 , 1{,}875)$. При этом $f^{\,1}=7{,}5$ , $\Delta f^{\,1}=0{,}5=y_1^0$. Пусть $b_2=4$ в (2) увеличилось на 1, т.е. $\Delta b_2=1$. Тогда оптимальный план новой задачи ЛП получается из решения системы уравнений $x_1 + 2x_2 = 4$, $5x_1+2x_2=11$: $x^2=(1{,}75 , 1{,}125)$. При этом $f^{\,2}=7{,}5$ , $\Delta f^{\,2}=0{,}5=y_2^0$. Пусть $b_3=12$ в (3) увеличилось на 1, т.е. $\Delta b_3=1$. Тогда оптимальный план новой задачи ЛП совпадает с $x^0$ и $\Delta f^{\,3}=0=y_3^0$.

Посмотрим, как могут изменяться свободные члены ограничений, чтобы активные ограничения оставались теми же. Подставим (6/7, 20/7) (точка пересечения прямых, соответствующих (2) и (3)) в левую часть неравенства (1); получим $b_1=46/7$. Подставив точку (2, 0) в левую часть неравенства (1), получим $b_1$=2. Значит, при условии $b_1 \in [2, 46/7]$ ограничения (1) и (2) останутся активными. Аналогично, подставим (1,4 , 0.8) (точка пересечения прямых, соответствующих (1) и (3)) в левую часть неравенства (2); получим $b_2$=13,6. Подставив точку (0, 2) в левую часть неравенства (2), получим $b_2$=4. Значит, при условии $b_2 \in [4, 13{,}6]$ ограничения (1) и (2) останутся активными. Подставив точку (1,5 , 1.25) в левую часть неравенства (3), получим $b_3$=9,75. Значит, при условии $b_3 \in [9{,}75 , \infty]$ ограничения (1) и (2) останутся активными. Таким образом, структура решения не меняется.

Проведем анализ чувствительности решения исходной задачи при изменении коэффициента $a_{12}$=5 на 1, т.е. $\Delta a_{12}$=1. Пусть $a_{12}$=6, тогда имеем систему уравнений: $x_1+2x_2=4$, $6x_1+2x_2=10$. Откуда $\tilde{x}_1 = 1{,}2,\ \tilde{x}_2 = 1{,}4,\ \tilde{f} = 6{,}4,\ \Delta\tilde{f} = -0{,}6$. При этом теоретическая оценка изменения экстремального значения равна $-\Delta a_{12}\, x_2^0 y_1^0 = -0{,}625$, т.е. несколько отличается (что естественно, так как это приблизительная оценка при достаточно малом изменении параметра).

Неустойчивая задача ЛП с непустым множеством решений называется *некорректной*. Ее численное решение с любой (ненулевой) точностью может привести к результатам, значительно отличающимся от истинных. Для решения таких задач используют *методы регуляризации*, позволяющие найти решение с любой степенью точности. Наиболее простой метод регуляризации (3.11) заключается в аппроксимации ее задачей вида (3.12), где матрица ограничений не меняется, т.е. $A(\delta)=A$, вектор правых частей ограничений $b(\delta)$ получается из $b$ прибавлением к каждой компоненте $\delta$, а вектор коэффициентов $c(\delta)$ получается из $c$ вычитанием из каждой компоненты $\delta$. Можно показать, что тогда задача (3.12) является устойчивой при любом $\delta$ и $f^*(\delta)\to f^*$ при $\delta\to 0$.

Неточность исходных данных может приводить к тому, что множество допустимых планов задача ЛП является пустым. Такие задачи называются *несобственными*. Для несобственных задач разработаны *методы коррекции*, заключающиеся в их аппроксимации собственными задачами и различающиеся набором варьируемых параметров и критериями близости.

Рассмотрим метод коррекции задачи (3.11) с пустым множеством планов по минимаксному критерию при вариации всех элементов матрицы $A$, для которого задача коррекции (аппроксимации) является задача ЛП. В возмущенной задаче (3.12) положим $c(\delta)=c$, $b(\delta)=b\neq 0$, $A(\delta)=A-H$. Обозначим множество допустимых планов задачи (3.12) через $X(H)$, а строка матрицы $A$ через $a_i$. Поставим задачу минимальной коррекции: минимизировать максимум из модулей элементов матрицы $H$ при условии $X(H)\neq\varnothing$.

**Теорема 3.4.** *Задача коррекции несобственной задачи* (3.11)

$$F(H)=\max_{i,j}|h_{ij}|\to\min,\ \text{ при условии } X(H)=\{x \mid (A-H)x\le b,\ x\ge 0\}\ne\varnothing \quad (3.13)$$

*эквивалентна задаче линейного программирования*

$$\min_{(u,z,v)\ge 0} u, \qquad u\ge a_i v-b_i z,\ i=1,\dots,n,\ \sum_{j=1}^{n} v_j=1, \quad (3.14)$$

*при ограничениях* $u\ge a_i v-b_i z,\ i=1,\dots,n,\ \sum_{j=1}^{n} v_j=1,$

*а именно, если* $(u^*,\ z^*,v^*)$ – *решение задачи* (3.14) *и* $z^*\ne 0$, *то решение задачи* (3.13) *определяется по формулам*

$$F(H^*)=u^*,\ x^*=\frac{v^*}{z^*},\ h_{ij}=\max\{a_i v^*-b_i z^*,\ 0\},\ j=1,\dots,n,\ i=1,\dots,m. \quad (3.15)$$

**Доказательство.** Пусть система неравенств $Ax\le b$ не имеет неотрицательных решений. Зафиксируем произвольное $x\ne 0$ и рассмотрим некоторое нарушенное ограничение $a_i x>b_i$. При минимальной коррекции коэффициентов строки $a_i$ коэффициенты матрицы коррекции $H$ должны удовлетворять равенству $\sum_{j=1}^{n} h_{ij}x_j=\sum_{j=1}^{n} a_{ij}x_j-b_i$. При этом так как все $x_j\ge 0$, то, очевидно, минимальность максимального коэффициента $h_{ij}$ достигается при их равенстве:

$$h_{ij}=h_i=\left(\sum_{j=1}^{n} x_j\right)^{-1}\left(\sum_{j=1}^{n} a_{ij}x_j-b_i\right),\ j=1,\dots,n. \quad (3.16)$$

Сделаем в (3.16) замену переменных $z=\left(\sum_{j=1}^{n} x_j\right)^{-1}$, $v_j=zx_j$, тогда имеем

$$h_i=\sum_{j=1}^{n} a_{ij}v_j-b_i z,\ \sum_{j=1}^{n} v_j=1. \quad (3.17)$$

Минимизация максимальной из величин $h_i,\ i=1,..,m$, (3.17) эквивалентна минимизации величины $u$ при ограничениях $u\ge a_i v-b_i z,\ i=1,\dots,m$. Получили задачу ЛП (3.14) и формулы (3.15) с учетом того, что в каждой строке можно выбрать свое минимальное значение $h_i$ по невязке, а если неравенство выполняется, то можно положить $h_i=0$. Теорема доказана.

Аналогичные результаты имеют место для случая коррекции обеих частей ограничений задачи (3.11), а также для несобственной задачи ЛП в канонической форме. При использовании евклидовой нормы матрицы в качестве оценки величины коррекции получается задача квадратичного программирования. Учет исходного критерия $cx$ в задаче коррекции приводит к проблеме многокритериальности.

## 3.4. Несобственные задачи оптимизации. Задачи связанной коррекции данных в эколого-экономических моделях

Рассмотрим вопросы коррекции параметров моделей для несобственных задач принятия решений. Неточность исходных данных модели, жесткие условия (или дополнительными ограничениями), накладываемыми на переменные, противоречивость требований, предъявляемых к модели, может приводить к тому, что ограничения оказываются несовместными. Соответствующие задачи оптимизации не имеют решения и называются несобственными. Для таких задач предлагаются процедуры минимальной коррекции данных, в результате которых аппроксимирующие задачи уже имеют решение.

Появление дополнительных ограничений, например, по уровню загрязнения окружающей природной среды в задаче производственного планирования может привести к результату, не соответствующему требованиям к плановому заданию для конкретного предприятия, или к несовместности ограничений модели, т.е. привести к несобственной задаче. Для моделирования подобных ситуаций предлагается подход, связанный с минимальной коррекцией параметров модели, при которой обеспечивается выполнение планового задания предприятия с учетом дополнительных ограничений. Данные в модели эколого-экономической системы, как правило, взаимосвязаны (например, технологические параметры влияют уровень загрязнения окружающей среды). В связи с этим предлагается рассмотреть класс задач связанной коррекции данных на примере общей линейной модели планирования выпуска продукции с ограничениями по уровню загрязнения окружающей среды

$$cx \to \max, \quad Ax \le b, Hx \le Z, x \ge 0, \tag{3.18}$$

где $x = (x_1,...,x_n)$ – вектор валовой продукции предприятия, $c = (c_1,...,c_n)$ – вектор цен на продукцию, $A=(a_{ij})_{m\times n}$ – матрица технологических коэффициентов предприятия ($a_{ij}$ – количество ресурса вида $i$, затрачиваемое на производство единицы продукции вида $j$ на предприятии), $b = (b_1,...,b_m)$ – вектор производственных ресурсов предприятия, $H=(h_{sj})_{q\times n}$ – матрица коэффициентов загрязнения для предприятия ($h_{sj}$ – объем загрязнения по $s$-му показателю при производстве единицы продукции вида $j$); $Z = (Z_1,...,Z_q)$ – вектор предельно допустимых уровней загрязнения окружающей среды по каждому показателю.

Задан минимальный валовый доход $C_0$ для предприятия, обеспечивающий, например, условия его безубыточного функционирования. Рассмотрим проблему несобственности задачи (3.18), которая связана с тем, что при заданных ограничениях предприятие не в состоянии достигнуть значения целевой функции $C_0$. Будем анализировать параметры задачи (3.18), добиваясь тех значений параметров, при которых данная задача становится собственной. Пусть $x^0$ – решение задачи (3.2) и $cx^0 < C_0$. Рассмотрим задачи

$$cx \to \max, \quad Ax \le b,\ x \ge 0, \tag{3.19}$$

$$cx \to \max, \quad Hx \le Z,\ x \ge 0. \tag{3.20}$$

Будем полагать сначала, что задача (3.20) имеет такое решение $\tilde{x}^0$, что $c\tilde{x}^0 \ge C_0$. Необходимо так скорректировать технологию производства (матрицу $A$) и объемы затрачиваемых ресурсов (вектор $b$) в задаче (3.18), чтобы вывести производство на требуемый уровень $C_0$. Множество допустимых управлений предприятия представляет собой совокупность планов $x$ и матриц коррекций данных модели, для которых $C_0 \le cx^0$. При этом принцип оптимальности управления есть отображение в подмножество множества допустимых управлений, для которого выполняется требование минимальности матриц коррекции по некоторой норме. Используя в качестве критерия некоторую норму вектора или матрицы корректируемых параметров, получаем следующие постановки задач коррекции данных

$$\|\Delta A\| \to \min_{x,\Delta A}, \quad (A+\Delta A)x \le b,\ Hx \le Z,\ cx \ge C_0,\ x \ge 0, \tag{3.21}$$

$$\|\Delta b\| \to \min_{x,\Delta b}, \quad Ax \le b+\Delta b,\ Hx \le Z,\ cx \ge C_0,\ x \ge 0, \tag{3.22}$$

$$\|-\Delta b, \Delta A\| \to \min_{x,\Delta A,\Delta b}, \quad (A+\Delta A)x \le b+\Delta b,\ Hx \le Z,\ cx \ge C_0,\ x \ge 0. \tag{3.23}$$

Если задача (3.20) дает решение $\tilde{x}^0$ такое, что $c\tilde{x}^0 < C_0$, то коррекция ограничений $Ax \le b$ в задаче (3.18) оказывается недостаточной для того, чтобы достичь выполнения условия $cx^0 \ge C_0$. Приходим к тому, что ограничения $Hx \le Z$ так же требуют коррекции. Можно сделать это непосредственно, вводя корректирующий параметр $\Delta H$:

$$\|\Delta A, \Delta H\| \to \min_{x,\Delta A,\Delta H}, \quad (A+\Delta A)x \le b,\ (H+\Delta H)x \le Z,\ cx \ge C_0,\ x \ge 0. \tag{3.24}$$

Такая задача соответствует ситуации, когда предприятие осуществляет ряд мероприятий по защите окружающей среды: установка современных очистных сооружений, фильтров, золоуловителей и т.д. Приведем задачу (3.24) к виду

$$\|\Delta A'\| \to \min_{x,\Delta A'}, \quad (A'+\Delta A')x \le b',\ cx \ge C_0,\ x \ge 0, \tag{3.25}$$

где $A' = \begin{bmatrix} A \\ H \end{bmatrix}$, $b' = \begin{bmatrix} b \\ Z \end{bmatrix}$, $\Delta A' = \begin{bmatrix} \Delta A \\ \Delta H \end{bmatrix}$.

Постановки задач, аналогичные (3.21), (3.22), (3.23), (3.25) подробно исследованы в различных нормах $l_1, l_2, l_\infty$ и даже более общих [3, 5, 6]. Мы будем исходить из предположения: матрица коэффициентов загрязнения $H$ связана с технологической матрицей $A$ некоторой зависимостью, т. е. технология производства влияет на степень загрязнения окружающей среды. Пусть $H$ связана с $A$ линейной зависимостью вида

$$H_1(A) = D_1 + U_1 A, \tag{3.26}$$

где $D_1$ − неотрицательная матрица размером $q\times n$, $U_1$ − положительная матрица размером $q\times m$. Сформулируем такие постановки задач коррекции данных, в которых при коррекции тех или иных параметров задачи подвергаются изменению элементы матрицы $H$ согласно (3.26). Такие задачи называются связанной коррекцией. При этом будем использовать норму матрицы $l_\infty$.

В виду того, что $A$ и $b$ предполагаются положительными, имеем следующие задачи коррекции

$$\begin{gathered} \|\Delta A\|_\infty \to \min_{x,\Delta A}, \\ (A-\Delta A)x \le b, \quad (D_1 + U_1(A-\Delta A))x \le Z, \quad cx \ge C_0, \quad x \ge 0, \end{gathered} \tag{3.27}$$

$$\begin{gathered} \|-\Delta b, \Delta A\|_\infty \to \min_{x,\Delta A,\Delta b}, \\ (A-\Delta A)x \le b + \Delta b, \quad (D_1 + U_1(A-\Delta A))x \le Z, \quad cx \ge C_0, \quad x \ge 0. \end{gathered} \tag{3.28}$$

**Теорема 3.5.** *Нахождение минимальной корректирующей матрицы и соответствующего оптимального плана в задаче* (3.27) *сводится к нахождению решения задачи линейного программирования*

$$\begin{gathered} \min_{y,z,\delta_1} \delta_1, \\ Az - by \le \delta_1 e, \quad D_1 z + U_1 Az - Zy \le \delta_1 U_1 e, \\ cz \ge C_0 y, \quad ze_1 = 1, \quad \delta_1 \ge 0,\ y \ge 0,\ z \ge 0. \end{gathered} \tag{3.29}$$

**Доказательство.** Для задачи (3.27) имеем $Ax - b \le \Delta Ax$ или $\sum_{j=1}^{n} a_{ij}x_j - b_i \le \sum_{j=1}^{n} \Delta a_{ij}x_j$, $i = 1,...,m$, $D_1 x + U_1 Ax - Z \le U_1 \Delta Ax$ или

$\sum_{j=1}^{n} d_{1sj}x_j + \sum_{i=1}^{m} u_{1si} \sum_{j=1}^{n} a_{ij}x_j - Z_s \le \sum_{i=1}^{m} u_{1si} \sum_{j=1}^{n} \Delta a_{ij}x_j$, $s = 1,...,q$. Согласно норме $l_\infty$ задача состоит в $\max_{i,j} |\Delta a_{ij}| \to \min$. Введем переменную

$$\delta_1 = \max\{\max_{1\le i\le m} \frac{\sum_{j=1}^{n} a_{ij}x_j - b_i}{\sum_{j=1}^{n} x_j}, \max_{1\le s\le q} \frac{\sum_{j=1}^{n} d_{1sj}x_j + \sum_{i=1}^{m} u_{1si}\sum_{j=1}^{n} a_{ij}x_j - Z_s}{\sum_{i=1}^{m} u_{1si}\sum_{j=1}^{n} x_j}, 0\}.$$

Обозначим: $\sum_{j=1}^{n} x_j = y^{-1}$. Тогда задача (3.27) сводится к задаче

$$\min_{x,\delta_1} \delta_1,$$
$$Ayx - by \le \delta_1 e, \quad D_1 yx + U_1 Ayx - Zy \le \delta_1 U_1 e, \qquad (3.30)$$
$$cx \ge C_0, \quad \sum_{j=1}^{n} x_j = y^{-1}, \quad \delta_1 \ge 0, x \ge 0,$$

где $e = (\underbrace{1...1}_{m})^T$. Примем $yx_j = z_j$, тогда $ze_1 = 1$, где $e_1 = (\underbrace{1...1}_{n})$ и задача (3.30) сводится к задаче ЛП (3.29). Теорема доказана.

**Теорема 3.6.** *Нахождение минимальной корректирующей матрицы и соответствующего оптимального плана в задаче* (3.28) *сводится к нахождению решения задачи*

$$\min_{y,z,\delta_2} \delta_2,$$
$$Az - (b + \delta_2 e)y \le \delta_2 e, \quad D_1 z + U_1 Az - Zy \le \delta_2 U_1 e, \qquad (3.31)$$
$$cz \ge C_0 y, \quad ze_1 = 1, \quad \delta_2 \ge 0, y \ge 0, z \ge 0.$$

**Доказательство.** Для задачи (3.28) имеем $Ax - b \le \Delta Ax + \Delta b$ или $\sum_{j=1}^{n} a_{ij} x_j - b_i \le \sum_{j=1}^{n} \Delta a_{ij} x_j + \Delta b_i, i = 1,...,m$, $D_1 x + U_1 Ax - Z \le U_1 \Delta Ax$ или

$\sum_{j=1}^{n} d_{1sj} x_j + \sum_{i=1}^{m} u_{1si} \sum_{j=1}^{n} a_{ij} x_j - Z_s \le \sum_{i=1}^{m} u_{1si} \sum_{j=1}^{n} \Delta a_{ij} x_j, s = 1,...,q$. Согласно норме $l_\infty$ задача состоит в

$\max\{\max_{i,j} |\Delta a_{ij}|, \max_i |\Delta b_i|\} \to \min$. Введем переменную

$$\delta_2 = \max\{\max_{1\le i\le m} \frac{\sum_{j=1}^{n} a_{ij} x_j - b_i}{1 + \sum_{j=1}^{n} x_j}, \max_{1\le s\le q} \frac{\sum_{j=1}^{n} d_{1sj} x_j + \sum_{i=1}^{m} u_{1si} \sum_{j=1}^{n} a_{ij} x_j - Z_s}{\sum_{i=1}^{m} u_{1si} \sum_{j=1}^{n} x_j}, 0\}.$$

Тогда задача (3.28) эквивалентна задаче

$$\min_{x,\delta_2} \delta_2,$$
$$Ayx - (b + \delta_2 e)y \le \delta_2 e, \quad D_1 yx + U_1 Ayx - Zy \le \delta_2 U_1 e, \qquad (3.32)$$
$$cx \ge C_0, \quad \sum_{j=1}^{n} x_j = y^{-1}, \quad \delta_2 \ge 0, \quad x \ge 0.$$

Обозначив $yx_j = z_j$, приходим к задаче (3.31). Теорема доказана.

Задача (3.31) не является задачей ЛП, но, фиксируя значение $\overline{\delta}_2 > 0$, для которого задача (3.31) имеет решение, с помощью метода деления отрезка $[0; \overline{\delta}_2]$ сводим ее к последовательности задач ЛП.

**Пример 3.10.** (коррекция матрицы *A*).

Рассмотрим задачу (3.18), в которой $c=(7;8;9)$, $A=\begin{pmatrix}10 & 17 & 17\\ 18 & 19 & 16\end{pmatrix}$, $b=\begin{pmatrix}190\\ 195\end{pmatrix}$, $H=\begin{pmatrix}0{,}2 & 0{,}5 & 0{,}7\\ 0{,}5 & 0{,}7 & 0{,}6\\ 0{,}8 & 0{,}1 & 0{,}8\end{pmatrix}$, $Z=\begin{pmatrix}8\\ 6\\ 10\end{pmatrix}$. Задан минимальный валовый доход $C_0$=100. Решение задачи ЛП (3.18) есть $x^0=(0;0;10)$, $cx^0=90$, $cx^0<C_0$. При этом задача (3.19) имеет решение $x2^0=(1{,}884;0;10{,}068)$, $cx2^0=103{,}801$, $cx2^0>C_0$, а задача (3.20) – решение $x3^0=(0;0;10)$, $cx3^0=90$, $cx3^0<C_0$, т. е. дополнительные экологические ограничения приводят к результату, не соответствующему плановому заданию $C_0$. Скорректируем матрицу $A$ так, чтобы в задаче (3.18) функция цели имела значение в оптимальной точке не меньшее $C_0$. Зависимость (2.37) $H$ от $A$ задается с помощью матриц

$$D_1=\begin{pmatrix}0{,}136 & 0{,}426 & 0{,}635\\ 0{,}352 & 0{,}518 & 0{,}436\\ 0{,}754 & 0{,}045 & 0{,}751\end{pmatrix} \text{ и } U_1=\begin{pmatrix}0{,}001 & 0{,}003\\ 0{,}004 & 0{,}006\\ 0{,}001 & 0{,}002\end{pmatrix}.$$

Решение задачи (3.29) есть $\delta_1=6$, $y=0{,}09$, $z=(0;0;1)$. Тогда решение задачи (3.27) находим по формуле $x=z/y$, т. е. имеем $x0=(0;0;11{,}111)$, причем $cx0=100$. Ввиду того, что первые две компоненты вектора $x0$ равны нулю, то достаточно корректировать только третий столбец матрицы $A$. При этом в соответствии с (3.26) будут меняться компоненты третьего столбца матрицы $H_1$. Для матрицы коррекции $\Delta A=\begin{pmatrix}0 & 0 & \delta_1\\ 0 & 0 & \delta_1\end{pmatrix}$ ограничения $(A-\Delta A)x0\le b$ выполняются как строгие неравенства:

$\begin{pmatrix}10 & 17 & 17-6\\ 18 & 19 & 16-6\end{pmatrix}x0=\begin{pmatrix}122{,}222\\ 111{,}111\end{pmatrix}<\begin{pmatrix}190\\ 195\end{pmatrix}$, а одно из ограничений $H_1x0\le Z$ как равенство:

$$\left[\begin{pmatrix}0{,}136 & 0{,}426 & 0{,}635\\ 0{,}352 & 0{,}518 & 0{,}436\\ 0{,}754 & 0{,}045 & 0{,}751\end{pmatrix}+\begin{pmatrix}0{,}001 & 0{,}003\\ 0{,}004 & 0{,}006\\ 0{,}001 & 0{,}002\end{pmatrix}\begin{pmatrix}10 & 17 & 17-6\\ 18 & 19 & 16-6\end{pmatrix}\right]x0=\begin{pmatrix}7{,}511\\ 6\\ 8{,}689\end{pmatrix}\le\begin{pmatrix}8\\ 6\\ 10\end{pmatrix}.$$

Следовательно, если взять значение корректирующего параметра в какой-либо строке матрицы $\Delta A$ меньше максимального $\delta_1=6$ так, чтобы выполнялись ограничения $(A-\Delta A)x0\le b$, то это приведет к увеличению компонент 3-го столбца матрицы $H_1$. Тогда для выполнения ограничений $H_1x0\le Z$ потребуется уменьшение каких-то компонент оптимального плана, что в свою очередь дает уменьшение значения целевой функции в оптимальной точке. Значит, оптимальному плану задачи (3.27) соответствует минимальная матрица коррекции вида $\Delta A = \begin{pmatrix} 0 & 0 & \delta_1 \\ 0 & 0 & \delta_1 \end{pmatrix}$. Тогда скорректированная матрица технологических коэффициентов $A_1 = A - \Delta A$ имеет вид $A_1 = \begin{pmatrix} 10 & 17 & 11 \\ 18 & 19 & 10 \end{pmatrix}$, а соответствующая ей матрица коэффициентов загрязнения $H_1 = D_1 + U_1 A_1$ есть $H_1 = \begin{pmatrix} 0,2 & 0,5 & 0,676 \\ 0,5 & 0,7 & 0,54 \\ 0,8 & 0,1 & 0,782 \end{pmatrix}$.

**Пример 3.11.** (коррекция матрицы $A$ и вектора $b$ путем сведения к последовательности задач линейного программирования)

Найдем решение задачи (3.28) для исходных данных примера 6.7 методом деления отрезка. Зададим такое $\delta_2$, при котором задача (3.31) имеет решение. Пусть на первом шаге $\delta_2 = 10$. Тогда, решение задачи (3.31) имеет вид $y_1 = 0,077$, $z_1 = (0,504; 0,115; 0,381)$. При заданном $\delta_2$ задача (3.31) становится задачей ЛП. Далее, на втором шаге отрезок $[0; 10]$ делим пополам и решаем задачу (3.31) при $\delta_2 = 5$. При $\delta_2 = 5$ задача (3.31) не имеет решения, поэтому на третьем шаге отрезок $[5; 10]$ делим пополам и находим решение задачи (3.31) при $\delta_2 = 7,5$: $y_3 = 0,081$, $z_3 = (0,454; 0,023; 0,523)$. Далее, на четвертом шаге отрезок $[5; 7,5]$ делим пополам и находим решение задачи (3.31) при $\delta_2 = 6,25$: $y_4 = 0,09$, $z_4 = (0; 0,014; 0,986)$. На пятом шаге при $\delta_2 = 5,625$ задача (3.31) не имеет решения, поэтому следует делить пополам отрезок $[5,625; 6,25]$ и искать решение задачи (3.31) при $\delta_2 = 5,938$. Продолжая таким образом процесс деления все меньших отрезков, получаем решение задачи (3.31) с заданной точностью $|z_k - z_{k-1}| < \varepsilon$, где $k$ – последний шаг. Для $\varepsilon = 0,01$ имеем решение на восьмом шаге: $y_8 = 0,09$, $z_8 = (0; 0,0005466; 0,999)$, которое имеет место при $\delta_2 = 6,094$. Решение задачи (3.28), соответствующее решению задачи (3.31) на восьмом шаге, находим по формуле $x = z_8 / y_8$, т. е. имеем $x0 = (0; 0,006073; 11,106)$.

Использование прикладного программного обеспечения дает возможность получить решение задачи нелинейного программирования (3.31), однако в нелинейных задачах большой размерности такое случается далеко не всегда, при этом точность полученного решения остается неизвестной. Преимущество метода деления отрезка заключается в том, что он позволяет решать задачи линейного программирования большой размерности, а также задавать любую требуемую точность вычислений.

Предположим, что имеется цель максимально приблизиться к наилучшему в некотором смысле значению критерия $cx$ с минимальной корректировкой ограничений задачи (3.18). В качестве наилучшего значения критерия $cx$ можно брать согласно методу идеальной точки [4] такое значение $c^*$ или $c_0^*$ критерия $cx$, которое получается из решения задачи (3.19) или (3.20) соответственно. «Идеальные» пороговые значения в данном случае подлежат коррекции. Если задача (3.20) имеет такое решение $c_0^*$, что $c_0^* \geq C_0$, а для задачи (3.18) $cx^0 < C_0$, то $c_0^*$ является недостижимым и в качестве «идеального» порогового значения следует брать $c_0^*$.

Задачи коррекции с пороговым значением $c_0^*$, в которых $H_1(A)$ определена согласно (3.26), имеют вид

$$\begin{gathered}\max\{\|\Delta A\|_\infty, \Delta c^*\} \to \min_{x,\Delta A,\Delta c^*}, \\ (A-\Delta A)x \leq b, \quad H_1(A-\Delta A)x \leq Z, \quad cx = c_0^* - \Delta c^*, \Delta c^* \geq 0, x \geq 0.\end{gathered} \tag{3.33}$$

$$\begin{gathered}\max\{\|-\Delta b, \Delta A\|_\infty, \Delta c^*\} \to \min_{x,\Delta A,\Delta b,\Delta c^*}, \\ (A-\Delta A)x \leq b+\Delta b, H_1(A-\Delta A)x \leq Z, cx = c_0^* - \Delta c^*, \Delta c^* \geq 0, x \geq 0.\end{gathered} \tag{3.34}$$

Представим ограничение $cx = c_0^* - \Delta c^*$ в виде $\Delta c^* = c_0^* - cx$.

**Лемма 3.1.** *Нахождение минимальной корректирующей матрицы и соответствующего оптимального плана в задаче* (3.33) *сводится к нахождению решения задачи*

$$\begin{gathered}\min_{y,z,u_1} u_1, \\ Az - by \leq u_1 e, \quad D_1 z + U_1 A z - Zy \leq u_1 U_1 e, \\ u_1 y \geq c_0^* y - cz, \quad z e_1 = 1, \quad u_1 \geq 0, y \geq 0, z \geq 0.\end{gathered} \tag{3.35}$$

**Доказательство.** Введем обозначение $u_1 = \max\{\delta_1, c_0^* - cx\}$, где $\delta_1$ определено для задачи (3.20) в доказательстве теоремы 6.13. Тогда задача (3.33) эквивалентна задаче

$$\begin{gathered}\min_{x,u_1} u_1, \\ Ayx - by \leq u_1 e, \quad D_1 yx + U_1 Ayx - Zy \leq u_1 U_1 e, \\ u_1 \geq c_0^* - cx, \quad \sum_{j=1}^{l} x_j = y^{-1}, \quad u_1 \geq 0, x \geq 0.\end{gathered}$$

Обозначив $yx_j = z_j$, приходим к задаче (3.35). Лемма доказана

**Лемма 3.2.** *Нахождение минимальной корректирующей матрицы и соответствующего оптимального плана в задаче* (3.34) *сводится к нахождению решения задачи*

$$\min_{y,z,u_2} u_2,$$
$$Az-(b+u_2e)y\le u_2e,\quad D_1z+U_1Az-Zy\le u_2U_1e, \tag{3.36}$$
$$u_2y\ge c_0^*y-cz,\quad ze_1=1,\quad u_2\ge 0,\ y\ge 0,\ z\ge 0.$$

**Доказательство.** Введем обозначение $u_2=\max\{\delta_2, c_0^*-cx\}$, где $\delta_2$ определено для задачи (3.32) в доказательстве теоремы 6.14. Тогда задача (3.34) эквивалентна задаче

$$\min_{x,u_2} u_2,$$
$$Ayx-(b+u_2e)y\le u_2e,\quad D_1yx+U_1Ayx-Zy\le u_2U_1e,$$
$$u_2\ge c_0^*-cx,\quad \sum_{j=1}^{l}x_j=y^{-1},\quad u_2\ge 0,\ x\ge 0.$$

Делая подстановку $yx_j=z_j$, получаем задачу (3.36). Лемма доказана.

Задачи (3.35), (3.36) также не являются задачами ЛП, но, фиксируя значения $\bar{u}_1>0,\bar{u}_2>0$, для которых соответствующие задачи (3.35), (3.36) имеют решение, с помощью метода деления отрезка (соответственно $[0;\bar{u}_1]$, $[0;\bar{u}_2]$) сводим их к последовательности задач ЛП.

**Пример 3.12.** (коррекция матрицы *A* и порогового значения критерия *cx*)

Рассмотрим задачу (3.2) с исходными данными $c=(7;8;9)$, $A=\begin{pmatrix}4&6&5\\8&9&7\end{pmatrix}$,

$b=\begin{pmatrix}90\\70\end{pmatrix}$, $H=\begin{pmatrix}0{,}2&0{,}5&0{,}7\\0{,}5&0{,}7&0{,}6\\0{,}8&0{,}1&0{,}8\end{pmatrix}$, $Z=\begin{pmatrix}9\\10\\14\end{pmatrix}$, $C_0=100$. Решение задачи ЛП (3.18) есть $x^0=(0;0;10)$, $cx^0=90$. При этом задача (3.20) имеет решение $x3^0=(6{,}728;0{,}276;10{,}737)$, $cx3^0=145{,}945$, $cx3^0>cx^0$, т. е. производственная задача (3.18) дает результат меньший, чем задача (3.20). Скорректируем матрицу *A* так, чтобы в задаче (3.18) функция цели имела значение в оптимальной точке как можно более близкое к «идеальному» пороговому значению $c_0^*=cx3^0=145{,}945$. Зависимость (3.26) *H* от *A* задается с помощью матриц

$$D_1=\begin{pmatrix}0{,}172&0{,}467&0{,}674\\0{,}436&0{,}622&0{,}538\\0{,}78&0{,}076&0{,}781\end{pmatrix} \text{ и } U_1=\begin{pmatrix}0{,}001&0{,}003\\0{,}004&0{,}006\\0{,}001&0{,}002\end{pmatrix}.$$

Решение задачи (3.35) есть $u_1=3{,}22$, $y=0{,}059$, $z=(0{,}32;0;0{,}68)$. Тогда решение задачи (3.33) находим по формуле $x=z/y$, т. е. имеем $x0=(5{,}461;0;11{,}611)$, причем $cx0=142{,}726$. Вторая компонента вектора $x0$ равна нулю, поэтому достаточно корректировать первый и третий столбец матрицы *A*. При этом в соответствии с (3.26) будут меняться компоненты первого и третьего столбца матрицы $H_1$. Для матрицы коррекции $\Delta A = \begin{pmatrix} u_1 & 0 & u_1 \\ u_1 & 0 & u_1 \end{pmatrix}$ выполняется как равенство второе из ограничений $(A - \Delta A)x0 \le b$:

$\begin{pmatrix} 4-3{,}22 & 6 & 5-3{,}22 \\ 8-3{,}22 & 9 & 7-3{,}22 \end{pmatrix} x0 = \begin{pmatrix} 24{,}934 \\ 70 \end{pmatrix} \le \begin{pmatrix} 90 \\ 70 \end{pmatrix}$, и первое из ограничений $H_1 x0 \le Z$:

$$\left[ \begin{pmatrix} 0{,}172 & 0{,}467 & 0{,}674 \\ 0{,}436 & 0{,}622 & 0{,}538 \\ 0{,}78 & 0{,}076 & 0{,}781 \end{pmatrix} + \begin{pmatrix} 0{,}001 & 0{,}003 \\ 0{,}004 & 0{,}006 \\ 0{,}001 & 0{,}002 \end{pmatrix} \begin{pmatrix} 4-3{,}22 & 6 & 5-3{,}22 \\ 8-3{,}22 & 9 & 7-3{,}22 \end{pmatrix} \right] x0 = \begin{pmatrix} 9 \\ 9{,}147 \\ 13{,}493 \end{pmatrix} \le \begin{pmatrix} 9 \\ 10 \\ 14 \end{pmatrix}.$$

Следовательно, если взять значение корректирующего параметра в какой-либо строке матрицы $\Delta A$ меньше максимального $u_1 = 3{,}22$, то это приведет к увеличению компонент 1-го и 3-го столбца матрицы $H_1$. Тогда для выполнения ограничений $(A - \Delta A)x0 \le b$ и $H_1 x0 \le Z$ потребуется уменьшение каких-то компонент оптимального плана, что в свою очередь уменьшит значение целевой функции в оптимальной точке. Значит, найденному оптимальному плану задачи (3.33) соответствует минимальная матрица коррекции вида $\Delta A = \begin{pmatrix} u_1 & 0 & u_1 \\ u_1 & 0 & u_1 \end{pmatrix}$. Тогда скорректированная матрица технологических коэффициентов $A_1 = A - \Delta A$ и соответствующая ей матрица коэффициентов загрязнения $H_1 = D_1 + U_1 A_1$ имеют вид $A_1 = \begin{pmatrix} 0{,}78 & 6 & 1{,}78 \\ 4{,}78 & 9 & 3{,}78 \end{pmatrix}$,

$$H_1 = \begin{pmatrix} 0{,}187 & 0{,}5 & 0{,}687 \\ 0{,}468 & 0{,}7 & 0{,}568 \\ 0{,}79 & 0{,}1 & 0{,}79 \end{pmatrix}.$$

**Пример 3.13.** (коррекция матрицы *A*, вектора *b* и порогового значения критерия *cx*)

Рассмотрим задачу (3.18) с матрицами, вектором рыночных цен из примера 6.8. Скорректируем матрицу *A* и вектора *b* так, чтобы в задаче (3.18) функция цели имела значение в оптимальной точке как можно более близкое к «идеальному» пороговому значению $c_0^* = cx3^0 = 145{,}945$. Зависимость (3.26) *H* от *A* задается с помощью матриц $D_1$ и $U_1$ из примера 6.8. Решение задачи (3.36) есть $u_2 = 3{,}053$, $y = 0{,}058$, $z = (0{,}323;\ 0;\ 0{,}677)$. Тогда по формуле $x = z/y$ решение задачи (3.34) есть $x0 = (5{,}531;\ 0;\ 11{,}575)$, причем $cx0 = 142{,}893$. Вторая компонента вектора $x0$ равна нулю, поэтому достаточно корректировать 1-й и 3-й столбец матрицы $A$. Для матрицы коррекции $\Delta A=\begin{pmatrix} u_2 & 0 & u_2 \\ u_2 & 0 & u_2 \end{pmatrix}$

выполняется как равенство второе из ограничений $(A-\Delta A)x0 \le b+\Delta b$ :

$\begin{pmatrix} 4-3{,}053 & 6 & 5-3{,}053 \\ 8-3{,}053 & 9 & 7-3{,}053 \end{pmatrix} x0=\begin{pmatrix} 27{,}779 \\ 73{,}053 \end{pmatrix} \le \begin{pmatrix} 90+3{,}053 \\ 70+3{,}053 \end{pmatrix}$, и первое из ограничений $H_1 x0 \le Z$ :

$$\left[\begin{pmatrix} 0{,}172 & 0{,}467 & 0{,}674 \\ 0{,}436 & 0{,}622 & 0{,}538 \\ 0{,}78 & 0{,}076 & 0{,}781 \end{pmatrix}+\begin{pmatrix} 0{,}001 & 0{,}003 \\ 0{,}004 & 0{,}006 \\ 0{,}001 & 0{,}002 \end{pmatrix}\begin{pmatrix} 4-3{,}053 & 6 & 5-3{,}053 \\ 8-3{,}053 & 9 & 7-3{,}053 \end{pmatrix}\right] x0=\begin{pmatrix} 9 \\ 9{,}188 \\ 13{,}528 \end{pmatrix} \le \begin{pmatrix} 9 \\ 10 \\ 14 \end{pmatrix}.$$

Следовательно, если взять значение корректирующего параметра в какой-либо строке матрицы $\Delta A$ меньше максимального $u_2=3{,}053$, то это приведет в соответствии с (2.37) к увеличению компонент первого и третьего столбца матрицы $H_1$. Тогда для выполнения ограничений $(A-\Delta A)x0 \le b+\Delta b$ и $H_1 x0 \le Z$ потребуется уменьшение каких-то компонент оптимального плана, что в свою очередь уменьшит значение целевой функции в оптимальной точке. Значит, найденному оптимальному плану задачи (3.34) соответствует минимальная матрица коррекции вида $\Delta A=\begin{pmatrix} u_2 & 0 & u_2 \\ u_2 & 0 & u_2 \end{pmatrix}$. Нетрудно видеть, что в данном примере достаточно подвергнуть коррекции вторую компоненту вектора $b$, т.е минимальный корректирующий вектор имеет вид $\Delta b=\begin{pmatrix} 0 \\ u_2 \end{pmatrix}$. Тогда скорректированная матрица технологических коэффициентов $A_1=A-\Delta A$ и соответствующая ей матрица коэффициентов загрязнения $H_1=D_1+U_1A_1$, а также скорректированный вектор ресурсов $b_1=b+\Delta b$ имеют вид $A_1=\begin{pmatrix} 0{,}947 & 6 & 1{,}947 \\ 4{,}947 & 9 & 3{,}947 \end{pmatrix}$, $H_1=\begin{pmatrix} 0{,}188 & 0{,}5 & 0{,}688 \\ 0{,}469 & 0{,}7 & 0{,}569 \\ 0{,}791 & 0{,}1 & 0{,}791 \end{pmatrix}$, $b_1=\begin{pmatrix} 90 \\ 73{,}053 \end{pmatrix}$.

Отметим, что коррекция вектора $b$ вместе с матрицей $A$ приводит к большему значению целевой функции в оптимальной точке (т. е. к значению еще более близкому к «идеальному» пороговому значению критерия $cx$) по сравнению с результатом из примера 6.8; при этом матрица $A$ подвергается меньшей коррекции.

### 3.5. Применение методов линейной оптимизации для решения матричных игр

Теория игр занимается исследованием математических моделей принятия оптимальных решений в конфликтных ситуациях. Формальное описание принятия решений можно разбить на две части:

1) математическая модель конфликтной ситуации или игра – описание конфликтной ситуации, включающее в себя описание субъектов, принимающих решения, их возможностей и интересов;

2) принцип оптимальности – описание правил рационального поведения игроков.

Математическая модель конфликта и принцип оптимальности дают полное описание принятия решений в условиях конфликта. Именно в этом смысле они весьма тесно взаимосвязаны. Однако на начальной стадии удобнее их рассматривать отдельно.

В общем случае в описании игры можно выделить следующие элементы:

1) коалиции действий - совокупность действующих совместно в данной конфликтной ситуации субъектов;

2) коалиции интересов - множество одинаково заинтересованных в исходе конфликта сторон;

3) множества возможных выборов каждой из коалиций действия;

4) описание предпочтений каждой из коалиций интересов;

5) множество возможных исходов конфликта.

Появление термина *коалиция* указывает на тот факт, что участниками конфликта могут быть не только отдельные лица, но и большие, сложно организованные группы лиц. Коалиции действия и интересов могут не совпадать. Например, пассажиры самолета заинтересованы в его скорейшем прибытии к месту назначения и, таким образом, образуют коалицию интересов, однако они не могут предпринять никаких действий, направленных на достижение этой цели, т.е. не являются коалицией действия. Если коалиция действия совпадает с коалицией интересов, то такую монолитную коалицию можно считать одним лицом, поэтому ее называют игроком. Если все коалиции действия совпадают с соответствующими коалициями интересов, то игру называют *бескоалиционной*, а ее участников – *игроками.*

Множество возможных исходов конфликта определяет совместные ограничения на действия участников конфликта. Если такие ограничения задаются лишь формально (в виде прямого произведения множеств), то соответствующую игру называют *игрой без запрещенных ситуаций***,** если эти ограничения существенны, то - *игрой с запрещенными ситуациями.*

Для принятия решений необходимо обладать определенной информацией. При этом как сам выбор из множества стратегий, так и ожидаемый результат зависят от информации, которой обладает игрок к моменту принятия решения: о множествах выбора, мотивах поведения или принципах оптимальности игроков, о природных неопределенных факторах. Поэтому целесообразно дать общее определение **с**тратегии**.** Если множество $B_i$

описывает информацию, которую коалиция (игрок) $i$ использует при принятии решений, то под *стратегией* будем понимать отображение $u_i : B_i \to U_i$, где $U_i$ – множество выборов или управлений коалиции действия $i$. В результате выбора каждой коалиции действия элемента из множества управлений определяется *исход конфликта*.

Заинтересованность $j$-го субъекта формализуется, как правило, функцией выигрыша, которая определяется отображением $g_j : U \to R$, где $U$ – множество исходов. Выбор стратегии или управления определяется принципом оптимальности.

Таким образом, для описания конфликтной ситуации в общем виде необходимо задать систему

$$\Gamma = \{I, K_g, K_u, \{U_i\}_{i \in K_g}, \{g_j\}_{j \in K_u}, S\},$$

где $I$ – множество игроков, $K_g$ – множество коалиций действия, $K_и$ – множество коалиций интересов, $U_i$ – множество выборов коалиции $i \in K_g$, $g_j$ – функция выигрыша коалиции $j \in K_и$, $S$ – множество возможных исходов игры.

Протекание игры состоит в выборе каждым игроком своей стратегии и получении в сложившейся ситуации выигрыша. Основной целью теории игр является выработка рекомендаций для рационального поведения игроков в конфликте.

Игры, как и все задачи исследования операций, бывают статическими и динамическими. Фиксация параметров, а также различная их суперпозиция позволяют классифицировать игры. Рассмотрим основные классы теоретико-игровых моделей.

В качестве первого классификационного признака возьмем множество игроков. Различают игры 2-х лиц и игры $N$ лиц ($N$>2). Игры 2-х лиц называются *антагонистическими*, если игроки преследуют противоположные цели. Если в антагонистической игре игрок 1 стремится максимизировать свой выигрыш $g_1$, то целью игрока 2 является минимизация выигрыша игрока 1, так что $g_1 = -g_2$.

Другой важный принцип классификации связан с вопросом о допустимости образования тех или иных коалиций. Если в игре образование коалиций недопустимо, то такая игра называется *бескоалиционной.* Она определяется заданием множества игроков, пространств их стратегий и набором их функций выигрыша.

К бескоалиционным играм могут быть сведены также игры, в которых $K_g = K_и$. В истинно же коалиционных играх разрешены такие коалиции, что $K_g \neq K_и$. Среди подобных игр наиболее распространены *кооперативные игры,* в которых образуется одна коалиция. Целью этой коалиции является максимизация суммарного выигрыша, с тем, чтобы впоследствии разделить его между членами коалиции по соглашению.

Следующим признаком классификации являются стратегии. Если множество стратегий всех игроков конечно, то игра называется *конечной.* Если $N$=2 и игрок 1 имеет $n$ стратегий, а 2 – $m$, то возможны $n \cdot m$ исходов. Выигрыши игроков можно задать двумя матрицами, поэтому такая игра называется *биматричной*. В ней по умолчанию игрок 1 выбирает строки, 2 – столбцы, в результаты выбора пары $(i, j)$ игроки получают, соответственно, $a_{ij}$ и $b_{ij}$ (иногда обе матрицы сводят в одну, у которой элементами являются пары чисел, первое по умолчанию – выигрыш 1 игрока, второе – выигрыш 2 игрока). Конечная антагонистическая игра называется *матричной*.

Игры можно квалифицировать и в соответствии с формой их задания. При этом различают позиционные игры и игры в нормальной форме. Если процесс принятия решений описывается в виде динамического процесса, где игроки выбирают свои стратегии последовательно по шагам, обладая при этом определенной информацией при каждом шаге выбора стратегии, то такие игры называются *позиционными.* Классическим примером такой игры являются шахматы. Если же в игре стратегия представлена как одноактный выбор, который производится одновременно и независимо всеми игроками, то такая игра считается заданной *в нормальной форме.*

При проведении предварительного анализа конфликта использование исходных множеств управлений $U_i$ может привести к нежелательному для оперирующей стороны результату: либо решения не существует, либо (если это решение существует) количественная оценка этого решения неудовлетворительна. Тогда целесообразно расширить класс используемых управлений. Расширения класса стратегий можно добиться путем расширения либо области определения функции (информированности игроков), либо за счет расширения множества значений, т.е. физических возможностей. Достаточно традиционным способом расширения множества значений является использование выпуклой оболочки множества управлений путем перехода в пространство вероятностных мер. Исходные элементы множества $U_i$ называются *чистыми стратегиями*, а их произвольная выпуклая комбинация (мера) – *смешанной стратегией*. Существует два способа реализации смешанных стратегий:

1) введение искусственной рандомизации, т.е. использование функции распределения на исходном множестве управлений;

2) введение повторения, т.е. проведения конфликта многократно.

При этом с определенной частотой выбирают некоторые элементы исходного множества. В обоих способах соответствующим образом определяются функции выигрыша.

Наиболее интересная постановка проблемы расширения класса стратегий связана с увеличением информированности игроков. Действительно, чем больше неопределенность, тем больше разброс в ожидаемом результате при реализации выбранной стратегии. Таким образом, оказывается, очень выгодно делать ситуацию более определенной. Для этого необходимо четко фиксировать ожидаемую информацию, уметь ее структурировать с целью возможности ее математической записи, оценивать результаты реализации стратегий, построенных по заданной информации, и, наконец, видеть пути увеличения объема информации и оценивать, с одной стороны, ее результат и, с другой стороны, затраты на получение дополнительной информации.

При изолированном поведении игроков, действующих самостоятельно, не обмениваясь информацией, центральное место занимают игры в нормальной форме.

**Определение 3.3.** Игрой в нормальной форме называется совокупность

$$\Gamma = \langle I, \{U_i\}_{i\in I}, \{g_i\}_{i\in I} \rangle,$$

где $I$ – множество всех игроков; $U_i$ – множество стратегий $i$-го игрока; $g_i$ – функция выигрыша $i$-го игрока, которую он стремится максимизировать.

Обычно игроков нумеруют в произвольном порядке от 1 до $N$, поэтому $I = \{1,2, \ldots, N\}$. Стратегия $i$-го игрока для игры в нормальной форме сводится к одноактному выбору любой точки из множества $U_i$. Функция выигрыша ставит в соответствие каждому элементу $u = (u_1,...,u_N)$ из множества $U = \prod_{i=1}^{N} U_i$, называемому исходом или ситуацией игры, действительное число. Таким образом, $g_i$ есть однозначное отображение множества $U \to R$.

Исходная постановка игры в нормальной форме не предполагает никакой дополнительной информации у игроков о действиях друг друга. Поэтому можно считать, что все игроки одновременно и независимо осуществляют выбор своих стратегий, т.е. элементов $u_i \in U_i$. В результате складывается ситуация $u$, однозначно определяющая выигрыши всех игроков $g_1(u),..., g_N(u)$.

Для решения игры необходимо задать принцип оптимальности. В соответствии с общим определением понятия принципа оптимальности – это отображение (вообще говоря, неоднозначное) игры во множество ее исходов, в данном случае $\Gamma \to U$. Таких отображений можно задать сколько угодно, но желательно их связывать с некоторыми представлениями игроков о разумном поведении (выгодностью, устойчивостью, справедливостью и т.д.). Формулировка таких представлений является сложной проблемой, но без них в выборе понятия решения остается полный произвол.

*Антагонистические игры. Матричные игры.*

Антагонистическая игра представляет собой частный случай игры в нормальной форме $\Gamma$, когда имеется два игрока $(n = 2)$ и сумма функций выигрыша этих игроков тождественно равна нулю. Так как в этом случае в любом исходе по выигрышу одного игрока можно однозначно определить выигрыш другого, то достаточно задать функцию выигрыша одного игрока.

Таким образом, антагонистическая игра (или игра двух лиц с нулевой суммой) в нормальной форме – это тройка

$$\Gamma_a = \langle U_1, U_2, g(u_1, u_2)\rangle,$$

где $U_1$ – множество стратегий игрока 1; $U_2$ – множество стратегий игрока 2; $g(u_1,u_2)$ – вещественная функция, определенная на $U_1 \times U_2$ и представляющая собой выигрыш игрока 1 в ситуации $(u_1,u_2)$, когда он выбирает стратегию $u_1$, а игрок 2 – стратегию $u_2$; при этом выигрыш игрока 2 равен $-g(u_1,u_2)$.

Так как точки максимума функции совпадают с точками минимума этой функции, взятой со знаком минус, то в антагонистической игре игрок 1 стремится по возможности максимизировать функцию $g(u_1,u_2)$, а игрок 2 – минимизировать.

В антагонистической игре поведение на основе принципа гарантированного результата становится особенно уместным. Его применение с учетом того, что $\sup\limits_{u_2 \in U_2} \inf\limits_{u_1 \in U_1} (-g(u_1,u_2)) = -\inf\limits_{u_2 \in U_2} \sup\limits_{u_1 \in U_1} g(u_1,u_2)$, приводит к следующим понятиям.

*Нижней ценой игры* $\Gamma_a$ называется величина

$$\underline{v} = \sup_{u_1 \in U_1} \inf_{u_2 \in U_2} g(u_1,u_2), \tag{3.37}$$

*верхней ценой игры* $\Gamma_a$ – величина

$$\bar{v} = \inf_{u_2 \in U_2} \sup_{u_1 \in U_1} g(u_1,u_2). \tag{3.38}$$

Смысл величин $\underline{v}$ и $\bar{v}$ заключается в следующем: игрок 1 может гарантировать себе выигрыш не меньше $\underline{v} - \varepsilon$, независимо от действий игрока 2; а игрок 2 – выигрыш не больше $\bar{v} + \varepsilon$, независимо от действий игрока 1 ($\varepsilon$ – сколь угодно малое положительное число, если наружные грани в (3.37), (3.38) не достигаются; $\varepsilon$=0, если они достигаются). Как было доказано ранее, величины $\underline{v}$ и $\bar{v}$, определенные выражениями (3.37), (3.38), всегда связаны соотношением $\underline{v} \le \bar{v}$, т.е. имеет место неравенство

$$\sup_{u_1 \in U_1} \inf_{u_2 \in U_2} g(u_1,u_2) \le \inf_{u_2 \in U_2} \sup_{u_1 \in U_1} g(u_1,u_2). \tag{3.39}$$

Таким образом, если оба игрока разумны, то выигрыш игрока 1 должен лежать на отрезке $[\underline{v}, \bar{v}]$ (возможно с точностью до $\varepsilon$).

Величины $\underline{v}$ и $\bar{v}$ могут быть как положительными, так и отрицательными, поэтому в слово “выигрыш” здесь не вкладывается смысл успешного исхода. Так как выигрыш может быть и положительным и отрицательным, то фактически он может выступать как в виде дохода, так и в виде расхода.

Ранее приводились примеры, показывающие, что в (3.39) может иметь место как равенство, так и строгое неравенство.

Если $\underline{v} = \bar{v}$, то их общее значение называют *ценой игры* $\Gamma_a$; если $\underline{v} < \bar{v}$, то говорят, что игра не имеет цены.

Напомним, что пара $(u_1^0, u_2^0)$ называется седловой точкой функции $g(u_1, u_2)$ на прямом произведении множеств $U_1 \times U_2$, если

$$g(u_1, u_2^0) \le g(u_1^0, u_2^0) \le g(u_1^0, u_2) \quad \forall u_1 \in U_1,\ u_2 \in U_2$$

или эквивалентно

$$\max_{u_1 \in U_1} g(u_1, u_2^0) = g(u_1^0, u_2^0) = \min_{u_2 \in U_2} g(u_1^0, u_2)\,.$$

Основными свойствами седловых точек, как известно из предыдцщей главы, являются взаимозаменяемость и эквивалентность. Если $(u_1^*, u_2^*)$ и $(u_1^{**}, u_2^{**})$ – седловые точки, то $(u_1^*, u_2^{**})$ и $(u_1^{**}, u_2^*)$ – также седловые точки (свойство взаимозаменяемости), при этом $g(u_1^*, u_2^*) = g(u_1^*, u_2^{**}) = g(u_1^{**}, u_2^*) = g(u_1^{**}, u_2^{**})$ (свойство эквивалентности).

По доказанной ранее теореме о необходимых и достаточных условиях существования седловой точки для того чтобы функция $g(u_1, u_2)$ имела седловую точку на $U_1 \times U_2$, необходимо и достаточно выполнения равенства

$$\max_{u_1 \in U_1} \inf_{u_2 \in U_2} g(u_1, u_2) = \min_{u_2 \in U_2} \sup_{u_1 \in U_1} g(u_1, u_2)\,. \tag{3.40}$$

**Определение 3.4.** *Конечной антагонистической (или матричной) игрой называется тройка* $\langle U_1, U_2, g(u_1, u_2) \rangle$, *где множества стратегий* $U_1$ *и* $U_2$ *состоят из конечного числа точек, а* $g(u_1, u_2)$ *- функция дискретного аргумента.*

Если множество $U_1$ содержит $n$ точек, то выбор каждой стратегии игроком 1 можно представить в виде выбора натурального числа $i$ $(i = 1, \dots, n)$. Аналогично, если множество $U_2$ содержит $m$ точек, то выбор каждой стратегии игроком 2 можно представить в виде выбора натурального числа $j$ $(j = 1, \dots, m)$. Обозначим через $a_{ij}$

значение функции $g$ в ситуации, соответствующей выбору $i$-й чистой стратегии игрока 1 и $j$-й чистой стратегии игрока 2. Тогда конечную игру можно задавать матрицей $A=(a_{ij})$ размера $n\times m$ (отсюда и название – матричная игра). Теперь стратегией игрока 1 является выбор строки $i$ матрицы $A$, стратегией игрока 2 – выбор столбца $j$ матрицы $A$, выигрыш игрока 1 в ситуации $(i, j)$ равен $a_{ij}$, а игрока 2 соответственно $-a_{ij}$.

Нижняя и верхняя цены матричной игры

$$\underline{v}=\max_{1\le i\le n}\min_{1\le j\le m} a_{ij},$$

$$\bar{v}=\min_{1\le j\le m}\max_{1\le i\le n} a_{ij}.$$

Равенство $\underline{v}=\bar{v}$ имеет место тогда и только тогда, когда матрица $A$ содержит седловую точку $(i_0, j_0)$, т.е. $a_{ij_0}\le a_{i_0j_0}\le a_{i_0j}\quad \forall i, j$. В этом случае общее значение нижней и верхней цены игры называется *ценой игры,* а $i_0$ и $j_0$ – *оптимальными стратегиями* игроков.

**Пример 3.14.** В игре с матрицей $A=\begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}$ пара стратегий (1, 1) составляет седловую точку, цена игры равна 1. Выбором первой строки игрок 1 гарантирует себе выигрыш не меньше 1, выбором первого столбца игрок 2 гарантирует себе проигрыш не больше 1. Значит, при разумных игроках выигрыш игрока 1 равен 1 (игрок 2 получает –1).

Если $\underline{v}<\bar{v}$, то выигрыши игроков однозначно не определены ( можно указать только разумный диапазон).

**Пример3.15.** В игре с матрицей $A=\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}$ нижняя и верхняя цены, соответственно, $\underline{v}=1,\ \bar{v}=2$, т.е. игрок 1 может гарантировать себе выигрыш не меньше 1, а игрок 2 – проигрыш не более 2. У каждого игрока обе стратегии являются оптимальными гарантирующими, но без дополнительных предположений исход этой игры и выигрыш неоднозначны.

В общем случае антагонистическая игра (даже матричная) не имеет цены. В такой ситуации решающим становится вопрос информированности игроков о выборе противника. Без его конкретизации нельзя определить решение игры. Различные варианты уточнения информированности приводят к расширению множеств стратегий. Одним из наиболее традиционных подходов является введение случайности в выбор игроков. При этом считается, что конкретный выбор каждого игрока не может быть известным противнику.

Если множество чистых стратегий игрока $U$ состоит из конечного числа точек $u_1,...,u_n$, то *смешанной стратегией* называется вектор $p=(p_1,...,p_n)$, где $p_i$ – вероятность выбора чистой стратегии $u_i$, $p_i \ge 0$, $\sum_{i=1}^{n} p_i = 1$. Формально чистую стратегию можно рассматривать как смешанную стратегию, задаваемую вектором, $i$-я компонента которого равна 1, а все остальные – 0, поэтому множество чистых стратегий является подмножеством множества смешанных стратегий. Множество $S_n = \{ p=(p_1,..., p_n) \mid p_i \ge 0, i=1,...,n, \sum_{i=1}^{n} p_i = 1\}$ называется *n-мерным симплексом***.** Если ассоциировать $U$ с единичными векторами (вершинами симплекса), то $U \subset S_n$, т.е. множество смешанных стратегий представляет собой расширение множества чистых стратегий.

Рассмотрим общую матричную игру

$$A = \begin{pmatrix} a_{11} & a_{12} & ... & a_{1m} \\ a_{21} & a_{22} & ... & a_{2m} \\ ... & ... & ... & ... \\ a_{n1} & a_{n2} & ... & a_{nm} \end{pmatrix}. \tag{3.41}$$

Введем смешанные стратегии для обоих игроков. Смешанной стратегией игрока 1 в матричной игре (3.41) является вектор

$$p=(p_1,..., p_n),\ p_i \ge 0,\ i=1,...,n,\ \sum_{i=1}^{n} p_i = 1,$$

где $p_i$ – вероятность выбора $i$-й чистой стратегии ($i=1,...,n$);

смешанной стратегией игрока 2 – вектор

$$q=(q_1,...,q_m), q_j \ge 0, j=1,...,m,\ \sum_{j=1}^{m} q_j = 1,$$

где $q_j$ – вероятность выбора $j$-й чистой стратегии ($j=1,...,m$).

Функция выигрыша игрока 1 как математическое ожидание в соответствии с формулами теории вероятностей имеет вид $h(p,q) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} p_i q_j = \langle p, Aq\rangle$,

множество смешанных стратегий игрока 1

$$S_n = \{ p=(p_1,..., p_n) \mid p_i \ge 0, i=1,...,n,\ \sum_{i=1}^{n} p_i = 1\}\ ,$$

множество смешанных стратегий игрока 2

$$S_m = \{ q=(q_1,...,q_m) \mid q_j \ge 0, j=1,...,m,\ \sum_{j=1}^{m} q_j = 1\}.$$

Совокупность $\{S_n, S_m, h(p,q)\}$ называется *смешанным расширением* матричной игры. Чистые стратегии игроков являются подмножествами множеств смешанных стратегий (вершинами симплексов $S_n$ и $S_m$). Выбору $i$-й чистой стратегии соответствует вектор $p$ с компонентами $p_i = 1,\ p_k = 0,\ k \neq i$ (аналогично для игрока 2). Функция $h$ на множествах чистых стратегий принимает значения $a_{ij}$.

В смешанных стратегиях нижняя и верхняя цена равны, соответственно, $\max\limits_{p\in S_n}\min\limits_{q\in S_m} h(p,q),$ и $\min\limits_{q\in S_m}\max\limits_{p\in S_n} h(p,q).$ Следующая теорема утверждает, что нижняя и верхняя цены матричной игры в смешанных стратегиях всегда равны между собой (их общее значение $v$ называется *ценой матричной игры*).

**Теорема 3.7. (**основная теорема матричных игр фон Неймана**).** *Любая матричная игра имеет цену* (*в смешанных стратегиях*)*, а игроки имеют оптимальные смешанные стратегии, т.е.*

$$v = \max_{p\in S_n}\min_{q\in S_m} h(p,q) = \min_{q\in S_m}\max_{p\in S_n} h(p,q) = \min_{q\in S_m} h(p^0,q) = \max_{p\in S_n} h(p,q^0) = h(p^0,q^0).$$

**Доказательство.** Согласно теореме о необходимых и достаточных условиях существования седловой точки утверждение данной теоремы эквивалентно утверждению о существовании седловой точки у функции $h(p,q)$ на $S_n \times S_m$. Последнее вытекает из теоремы о существовании седловой точки у вогнуто-выпуклой функции. Действительно, множества $S_n$ и $S_m$ являются, очевидно, выпуклыми замкнутыми ограниченными подмножествами пространств $R^n$ и $R^m$, функция $h(p,q)$ непрерывна по $p,q$ и линейна по $p,q$, следовательно, является вогнуто-выпуклой. Таким образом, выполнены все условия теоремы о существовании седловой точки у вогнуто-выпуклой функции и существует такая точка $(p^0,q^0)$, что $h(p,q^0) \le h(p^0,q^0) \le h(p^0,q)\ \forall p,q.$ Из теоремы о необходимых и достаточных условиях существования седловой точки вытекает, что $p^0$ реализует $\max\limits_{p\in S_n}\min\limits_{q\in S_m} h(p,q),$ $q^0$ реализует $\min\limits_{q\in S_m}\max\limits_{p\in S_n} h(p,q)$ и нижняя и верхняя цены игры в смешанных стратегиях равны между собой, что и требовалось доказать.

Рассмотрим некоторые свойства оптимальных стратегий матричных игр.

В следующих теоремах формулируются необходимые и достаточные условия оптимальности стратегий.

**Теорема 3.8.** *Пусть $v$ – цена матричной игры, тогда для того, чтобы $p^0$ была оптимальной стратегией игрока* 1, *необходимо и достаточно, чтобы*

$$h(p^0,q) \ge v \qquad \forall q \in S_m. \tag{3.42}$$

*Аналогично, для того, чтобы* $q^0$ *была оптимальной стратегией игрока* 2, *необходимо и достаточно, чтобы*

$$h(p,q^0)\le v \qquad \forall p\in S_n. \tag{3.43}$$

**Доказательство.** *Необходимость.* Пусть $p^0$ – оптимальная стратегия, т.е. $\min\limits_{q\in S_m} h(p^0,q)=\max\limits_{p\in S_n}\min\limits_{q\in S_m} h(p,q)=v$, отсюда следует (3.42).

*Достаточность.* Пусть выполнено неравенство (3.42), тогда $\min\limits_{q\in S_m} h(p^0,q)\ge v=\max\limits_{p\in S_n}\min\limits_{q\in S_m} h(p,q),$ откуда $\min\limits_{q\in S_m} h(p^0,q)=\max\limits_{p\in S_n}\min\limits_{q\in S_m} h(p,q),$ т.е. $p^0$ – оптимальная стратегия.

Доказательство для игрока 2 аналогично.

Далее понадобятся обозначения

$$h(i,q)=\sum_{j=1}^{m} a_{ij}q_j, \tag{3.44}$$

$$h(p,j)=\sum_{i=1}^{n} a_{ij}p_i. \tag{3.45}$$

Выражение (3.44) представляет собой выигрыш игрока 1 при использовании им $i$-й чистой стратегии, тогда как игрок 2 использует смешанную стратегию $q$; выражение (3.45) – выигрыш игрока 1 при использовании им смешанной стратегии $p$, тогда как игрок 2 использует чистую $j$-ю стратегию.

**Следствие 3.1.** *Пусть* $v$ *– цена матричной игры, тогда для того чтобы* $p^0$ *была оптимальной стратегией игрока* 1, *необходимо и достаточно, чтобы*

$$h(p^0,j)\ge v, \quad j=1,\dots,m \tag{3.46}$$

*Аналогично, для того, чтобы* $q^0$ *была оптимальной стратегией игрока* 2, *необходимо и достаточно, чтобы*

$$h(i,q^0)\le v, \quad i=1,\dots,n \tag{3.47}$$

**Доказательство.** *Необходимость* следует непосредственно из теоремы 3.8, так как множество чистых стратегий является подмножеством множества смешанных стратегий.

*Достаточность*. Возьмем произвольную смешанную стратегию $q$ игрока 2, умножим неравенства (3.46) на соответствующие компоненты вектора $q$ и сложим полученные неравенства. Тогда $\sum\limits_{j=1}^{m} h(p^0,j)q_j\ge v\sum\limits_{j=1}^{m} q_j=v,$ но

$\sum_{j=1}^{m} h(p^0, j)q_j = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} p_i^0 q_j = h(p^0, q),$ значит, выполнено (3.42) и $p^0$ – оптимальная стратегия.

Доказательство для игрока 2 аналогично.

**Теорема 3.9.** *Для того чтобы* $v$ *было ценой игры, а* $p^0$ *и* $q^0$ *– оптимальными стратегиями игроков, необходимо и достаточно, чтобы выполнялись неравенства*

$$h(p, q^0) \le v \le h(p^0, q) \qquad \forall p, q. \tag{3.48}$$

**Доказательство.** *Необходимость.* Если $v$ – цена игры, а $p^0$ и $q^0$ – оптимальные стратегии, то ($p^0$, $q^0$) – седловая точка функции $h$ и по основной теореме матричных игр $h(p^0, q^0) = v$, откуда следует (3.48).

*Достаточность.* Положим в левой части неравенства (3.48) $p = p^0$, а в правой части $q = q^0$. Тогда $h(p^0, q^0) \le v \le h(p^0, q^0)$, откуда $h(p^0, q^0) = v$. Подставив в (3.48) вместо $v$ равное ему значение $h(p^0, q^0)$, получим, что ($p^0$, $q^0$) является седловой точкой функции $h$. Значит по основной теореме матричных игр $p^0$, $q^0$ – оптимальные стратегии, $v$ – цена игры.

**Следствие 3.2.** *Для того чтобы* $v$ *было ценой игры, а* $p^0$, $q^0$ *– оптимальными стратегиями игроков, необходимо и достаточно, чтобы*

$$h(i, q^0) \le v \le h(p^0, j) \qquad \forall i, j.$$

Доказательство аналогично доказательству следствия 3.1.

**Теорема 3.10** (о свойствах оптимальных стратегий). *Пусть* $p^0$ *и* $q^0$ *- произвольные оптимальные стратегии, а* $v$ *- цена игры. Тогда, если* $p_i^0 > 0$*, то* $h(i, q^0) = v$*; если* $q_j^0 > 0$*, то* $h(p^0, j) = v$*. Кроме того, имеют место равенства*

$$\min_{1\le j\le m} h(p^0, j) = \max_{p\in S_n} \min_{1\le j\le m} h(p, j) = \min_{q\in S_m} \max_{1\le i\le m} h(i, q) = \max_{1\le i\le n} h(i, q^0) = v. \tag{3.49}$$

**Доказательство.** Так как $q^0$ – оптимальная стратегия, $v$ – цена игры, то согласно (3.47) имеем $h(i, q^0) \le v,\ i = 1, ..., n$.

Предположим противное: существует такое $i_1$, что $p_{i_1}^0 > 0, h(i_1, q^0) < v$, тогда $p_{i_1}^0 h(i_1, q^0) < p_{i_1}^0 v;\ p_i^0\, h(i, q^0) \le p_i^0 v\ \ \forall i \ne i_1.$ Суммируя эти неравенства с учетом того, что одно из них строгое, получим $\sum_{i=1}^{n} h(i, q^0)\, p_i^0 = h(p^0, q^0) < \sum_{i=1}^{n} v p_i^0 = v.$

Пришли к противоречию, так как $(p^0, q^0)$ – седловая точка и $h(p^0, q^0) = v$. Аналогично доказывается утверждение $q_j^0 > 0 \Rightarrow h(p^0, j) = v$.

Далее имеем $h(p,q) = \sum_{j=1}^{m} h(p, j) q_j \geq \sum_{j=1}^{n} [\min_{1\leq j\leq m} h(p, j)] q_j = \min_{1\leq j\leq m} h(p, j) \sum_{j=1}^{m} q_j = \min_{1\leq j\leq m} h(p, j) \ \forall q$, поэтому $\min_{q\in S_m} h(p,q) \geq \min_{1\leq j\leq m} h(p, j)$.

С другой стороны, так как множество чистых стратегий является подмножеством множества смешанных стратегий, то $\min_{q\in S_m} h(p,q) \leq \min_{1\leq j\leq m} h(p, j)$. Следовательно, $p^0$ откуда

$$\max_{p\in S_n} \min_{1\leq j\leq m} h(p, j) = \max_{p\in S_n} \min_{q\in S_m} h(p,q) = v = \min_{q\in S_m} h(p^0, q) = \min_{1\leq j\leq m} h(p^0, j).$$

Аналогично доказываются и остальные равенства (3.49).

**Следствие 3.3.** *Если существует цена игры в чистых стратегиях, то она совпадает с ценой игры в смешанных стратегиях.*

**Доказательство.** Пусть $\underline{v} = \max_{1\leq i\leq n} \min_{1\leq j\leq m} a_{ij} = \bar{v} = \min_{1\leq j\leq m} \max_{1\leq i\leq n} a_{ij}$. Так как $\max_{1\leq i\leq n} \min_{1\leq j\leq m} a_{ij} \leq \max_{p\in S_n} \min_{1\leq j\leq m} h(p, j) = v$ и $\min_{1\leq j\leq m} \max_{1\leq i\leq n} a_{ij} \geq \min_{q\in S_m} \max_{1\leq i\leq n} h(i,q) = v$, то всегда $\underline{v} \leq v \leq \bar{v}$, где $v$ - цена игры в смешанных стратегиях. Из равенства $\underline{v} = \bar{v}$ следует, что $\underline{v} = \bar{v} = v$, что и требовалось доказать.

**Определение 3.5.** Некоторая линейная комбинация векторов $\alpha^1,...,\alpha^k$ доминирует (строго доминирует) вектор $\beta$, если существуют такие числа $\lambda_1,...,\lambda_k$, $\lambda_i \geq 0,\ i = 1,...,k,\ \sum_{i=1}^{k} \lambda_i = 1$, что вектор $\sum_{i=1}^{k} \lambda_i \alpha^i$ доминирует (строго доминирует) вектор $\beta$. При этом говорят, что вектор $\beta$ доминируется вектором $\alpha$ или линейной комбинацией.

**Теорема 3.11** (принцип доминирования). *Если $i_0$-я строка матрицы игры доминируется некоторой линейной комбинацией остальных строк, то существует такая оптимальная стратегия $p^0$ игрока* 1, *что $p_{i_0}^0 = 0$. Если $i_0$-я строка строго доминируется некоторой линейной комбинацией остальных строк, то для любой оптимальной стратегии $p^0$ игрока 1 выполняется $p_{i_0}^0 = 0$. Аналогично, если $j_0$-й столбец матрицы игры доминирует* (*строго доминирует*) *некоторую линейную комбинацию остальных столбцов, то существует такая* (*любая*) *оптимальная стратегия $q^0$ игрока* 2, *что $q_{j_0}^0 = 0$.*

**Доказательство.** Пусть сначала доминирование нестрогое, а $p^*$ – произвольная оптимальная стратегия игрока 1, тогда $h(p^*, j) \ge v \ \forall j$. По условию существуют такие $\lambda_i \ge 0,\ i = 1,...,n,\ \sum_{i=1}^{n} \lambda_i = 1$, что $a_{i_0 j} \le \sum_{i=1}^{n} \lambda_i a_{ij},\ j = 1,...,m\ (\lambda_{i_0} = 0)$. Введем вектор $p^0$ с компонентами $p_i^0 = p_i^* + \lambda_i p_{i_0}^*,\ i \ne i_0,\ p_{i_0}^0 = 0$. Тогда

$$h(p^0, j) = \sum_{i \ne i_0} a_{ij} p_i^* + p_{i_0}^* \sum_{i=1}^{n} a_{ij} \lambda_i \ge \sum_{i \ne i_0} a_{ij} p_i^* + p_{i_0}^* a_{i_0 j} = h(p^*, j) \ge v \quad \forall j.$$

Следовательно, $p^0$ – оптимальная стратегия, для которой $p_{i_0}^0 = 0$. Пусть теперь доминирование строгое, т.е. существуют такие $\lambda_i \ge 0,\ i = 1,...,n,\ \sum_{i=1}^{n} \lambda_i = 1,\ \lambda_{i_0} = 0$, что

$$a_{i_0 j} < \sum_{i=1}^{n} \lambda_i a_{ij},\ j = 1,...,m\ . \tag{3.50}$$

Пусть $q^0$ – произвольная оптимальная стратегия игрока 2. Умножим неравенства (3.50) на $q_j^0$ и сложим. Так как хотя бы одно $q_j^0 > 0,$ то

$$\sum_{j=1}^{m} a_{i_0 j} q_j^0 < \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \lambda_i q_j^0 \le \max_{p \in S_n} h(p, q^0) = v.$$

Значит $h(i_0, q^0) < v$, откуда по теореме о свойствах оптимальных стратегий для любой оптимальной стратегии $p^0$ игрока 1 выполняется равенство $p_{i_0}^0 = 0$.

Аналогично доказательство для игрока 2.

В соответствии с принципом доминирования при решении матричной игры (нахождении цены и оптимальных стратегий) можно предварительно вычеркивать доминируемые строки и доминирующие столбцы, сокращая размерность задачи.

*Связь матричных игр с линейным программированием*

Существует тесная связь между матричными играми и линейным программированием. Решение любой матричной игры можно свести к решению пары двойственных задач линейного программирования специального вида (соответствующий результат будет сформулирован для игр с положительными матрицами, но любую матричную игру, как уже было сказано, можно свести к игре с положительной матрицей прибавлением достаточно большой константы к каждому элементу матрицы). С другой стороны, любую задачу линейного программирования, имеющую решение, можно свести к матричной игре специального вида.

**Определение 3.6.** Матричная игра называется *симметричной,* если ее платежная матрица $(a_{ij})$ кососимметрическая, т.е. $a_{ij} = -a_{ji} \forall i, j$.

Так как кососимметрическая матрица квадратная, то размерности векторов в смешанных стратегиях обоих игроков одинаковы $(m = n)$, а множества всех смешанных стратегий совпадают. Свойство симметричных игр сформулировано в следующей теореме.

**Теорема 3.12.** *Цена симметричной матричной игры равна нулю, а множества оптимальных стратегий игроков совпадают.*

**Доказательство.** Справедлива цепочка равенств

$$v = \max_{p\in S_n} \min_{1\le j\le n} h(p, j) = \max_{p\in S_n} \min_{1\le j\le n} \sum_{i=1}^{n} a_{ij} p_i = -\min_{p\in S_n} \max_{1\le j\le n} \sum_{i=1}^{n} a_{ij} p_i =$$

$$= -\min_{q\in S_n} \max_{1\le i\le n} \sum_{j=1}^{n} a_{ij} q_j = -\min_{q\in S_n} \max_{1\le i\le n} h(i, q) = -v,$$

следовательно, $v = 0$ (при доказательстве равенств производится перемена обозначений $p$ на $q$ и $i$ на $j$).

Пусть $p^0$ – оптимальная стратегия игрока 1. Тогда

$$h(p^0, j) = \sum_{i=1}^{n} a_{ij} p_i^0 \ge 0,\ j = 1,\dots,n;$$

$$h(i, p^0) = \sum_{j=1}^{n} a_{ij} p_j^0 = -\sum_{j=1}^{n} a_{ji} p_j^0 = -\sum_{i=1}^{n} a_{ij} p_i^0 \le 0,\ j = 1,\dots,n$$

Следовательно, $p^0$ – оптимальная стратегия игрока 2. Аналогично, любая оптимальная стратегия игрока 2 является оптимальной стратегией игрока 1.

**Теорема 3.13.** *Решение матричной игры с матрицей*

$$A = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nm} \end{pmatrix}, \quad a_{ij} > 0 \quad \forall i, j,$$

*эквивалентно решению пары двойственных задач ЛП:*

$$\sum_{i=1}^{n} x_i \to \min, \quad x_i \ge 0,\ i = 1,\dots,n,\ \sum_{i=1}^{n} a_{ij} x_i \ge 1,\ j = 1,\dots,m; \tag{3.51}$$

$$\sum_{j=1}^{m} y_j \to \max, \quad y_j \ge 0,\ j = 1,\dots,m, \sum_{j=1}^{m} a_{ij} y_j \le 1,\ i = 1,\dots,n. \tag{3.52}$$

*Точнее, если* $x^0 = (x_1^0,\dots,x_n^0)$ *– решение задачи* (3.51) , $y^0 = (y_1^0,\dots,y_m^0)$ *– решение задачи* (3.52)*, то* $v = \dfrac{1}{\sum_{i=1}^{n} x_i^0} = \dfrac{1}{\sum_{j=1}^{m} y_j^0}$ *– цена игры с матрицей A,* $p^0 = vx^0$ *– оптимальная стратегия игрока* 1, $q^0 = vy^0$ *– оптимальная стратегия игрока* 2.

*Обратно, если* $p^0$ *и* $q^0$ *– оптимальные стратегии игроков,* $v$ *– цена игры, то* $x^0 = \frac{p^0}{v}$ *– решение задачи* (3.51)*, а* $y^0 = \frac{q^0}{v}$ *– решение задачи* (3.52).

**Доказательство.** Задачи (3.51) и (3.52) имеют хотя бы по одному допустимому вектору (для (3.52) нулевой вектор, а для (3.51) вследствие положительности матрицы $A$ вектор с достаточно большими компонентами), значит, они обе имеют решения.

Пусть $x^0$ – решение (3.51), $y^0$ – решение (3.52), тогда по теореме двойственности $\sum_{i=1}^{n} x_i^0 = \sum_{j=1}^{m} y_j^0 > 0$ (положительность этих сумм следует из того, что $x^0 \neq 0$). Положим

$$v = \frac{1}{\sum_{i=1}^{n} x_i^0} = \frac{1}{\sum_{j=1}^{m} y_j^0}, \quad p^0 = vx^0, \ q^0 = vy^0,$$

тогда

$$p_i^0 \geq 0, i = 1,...,n, \ \sum_{i=1}^{n} p_i^0 = 1,$$

$$q_j^0 \geq 0, \ j = 1,...,m, \ \sum_{j=1}^{m} q_j^0 = 1;$$

$$\sum_{i=1}^{n} a_{ij} p_i^0 \geq v \geq \sum_{j=1}^{m} a_{ij} q_j^0 \quad \forall i, j.$$

Значит, по следствию 3.2 получаем, что $v$ – цена игры, $p^0$ и $q^0$ – оптимальные стратегии игроков.

Пусть теперь $v$ – цена игры с матрицей $A$, $p^0$ и $q^0$ – оптимальные стратегии игроков. Так как матрица $A$ положительная, то $v > 0$. Положим $x^0 = \frac{p^0}{v}$, $y^0 = \frac{q^0}{v}$, тогда $x_i^0 \geq 0, i = 1,...,n,$ $q_j^0 \geq 0, j = 1,...,m,$ и из (3.46), (3.47) следуют неравенства $\sum_{i=1}^{n} a_{ij} x_i^0 \geq 1, \ j = 1,...,m,$ $\sum_{j=1}^{m} a_{ij} y_j^0 \leq 1, \ i = 1,...,n,$ т.е. $x^0$ – допустимый вектор задачи (3.51); $y^0$ – допустимый вектор задачи (3.52). При этом $\sum_{i=1}^{n} x_i^0 = \sum_{j=1}^{m} y_j^0 = 1,$ следовательно, $x^0$ – решение (3.51), $y^0$ – решение (3.52). Теорема доказана.

**Теорема 3.14.** *Решение пары двойственных задач линейного программирования*

$$\sum_{j=1}^{n} c_j x_j \to \max, \quad x_j \geq 0, \ j = 1,...,n, \ \sum_{j=1}^{n} a_{ij} x_j \leq b_i, \ i = 1,...,m; \tag{3.53}$$

$$\sum_{i=1}^{m} b_i y_i \to \min, \quad y_i \geq 0,\ i=1,...,m,\ \sum_{i=1}^{m} a_{ij} y_i \geq c_j,\ j=1,...,n, \tag{3.54}$$

*эквивалентно решению симметричной матричной игры с матрицей*

$$D = \begin{pmatrix} 0 & \dots & 0 & -a_{11} & \dots & -a_{m1} & c_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & -a_{1n} & \dots & -a_{mn} & c_n \\ a_{11} & \dots & a_{1n} & 0 & \dots & 0 & -b_1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m1} & \dots & a_{mn} & 0 & \dots & 0 & -b_m \\ -c_1 & \dots & -c_n & b_1 & \dots & b_m & 0 \end{pmatrix}$$

*Точнее, если* $z^0 = (\mu_1^0,...,\mu_n^0,\nu_1^0,...,\nu_n^0,\lambda^0)$ *– оптимальная стратегия любого игрока в игре с матрицей D и* $\lambda^0>0$, *то* $x^0 = (\frac{\mu_1^0}{\lambda^0},...,\frac{\mu_n^0}{\lambda^0})$ *– решение задачи* (3.53), $y^0 = (\frac{\nu_1^0}{\lambda^0},...,\frac{\nu_m^0}{\lambda^0})$ *– решение задачи* (3.54). *Обратно, если* $x^0 = (x_1^0,...,x_n^0)$ *– решение задачи* (3.53), $y^0 = (y_1^0,...,y_n^0)$ *– решение задачи* (3.54), *то*

$$z^0 = (\lambda^0 x_1^0,...,\lambda^0 x_n^0,\lambda^0 y_1^0,...,\lambda^0 y_m^0,\lambda^0),$$

*где* $\lambda^0 = \dfrac{1}{1+\sum_{j=1}^{n} x_j^0 + \sum_{i=1}^{m} y_i^0}$, *является оптимальной стратегией любого игрока в игре с матрицей D. Пара двойственных задач* (3.53), (3.54) *имеет решение тогда и только тогда, когда существует такая оптимальная стратегия в игре с матрицей D*

$$z^0 = (\mu_1^0,...,\mu_n^0,\nu_1^0,...,\nu_n^0,\lambda^0),$$

*для которой* $\lambda^0>0$.

**Доказательство.** Пусть $z^0 = (\mu_1^0,...,\mu_n^0,\nu_1^0,...,\nu_m^0,\lambda^0)$ – оптимальная стратегия, $\lambda^0>0$. Так как игра симметричная, то цена игры с матрицей $D$ равна нулю. Поэтому необходимые условия оптимальности (3.46) применительно к матрице $D$ дают следующие соотношения:

$$\sum_{i=1}^{m} a_{ij}\nu_i^0 - c_j\lambda^0 \geq 0,\ j=1,...,n; \tag{3.55}$$

$$-\sum_{j=1}^{n} a_{ij}\mu_j^0 + b_i\lambda^0 \geq 0, i=1,...,m; \tag{3.56}$$

$$\sum_{j=1}^{n} c_j\mu_j^0 - \sum_{i=1}^{m} b_i y_i^0 = 0 \tag{3.57}$$

(в (3.57) имеет место равенство по теореме о свойствах оптимальных стратегий, так как $\lambda^0>0$). Положим $x_j^0=\dfrac{\mu_j^0}{\lambda^0}, j=1,...,n,\quad y_i^0=\dfrac{\nu_i^0}{\lambda^0}, i=1,...,n.$ Тогда

$$x_j^0\geq 0,\ j=1,...,n,\quad \sum_{j=1}^{n}a_{ij}x_j^0\leq b_i,\ i=1,...,m,$$

$$y_i^0\geq 0,\ i=1,...,m,\quad \sum_{i=1}^{m}a_{ij}y_j^0\geq c_j,\ j=1,...,n,$$

$$\sum_{j=1}^{n}c_jx_j^0=\sum_{i=1}^{m}b_iy_i^0.$$

Следовательно, $x^0=(x_1^0,...,x_n^0)$ – решение задачи (3.53), $y^0=(y_1^0,...,y_n^0)$ – решение задачи (3.54).

Пусть теперь $x^0=(x_1^0,...,x_n^0)$ – решение задачи (3.53), $y^0=(y_1^0,...,y_n^0)$ – решение задачи (3.54). Положим $\lambda^0=\dfrac{1}{1+\sum_{j=1}^{n}x_j^0+\sum_{i=1}^{m}y_i^0}$, $\mu_j^0=\lambda^0x_j^0, j=1,...,n,\quad \nu_i^0=\lambda^0y_i^0, i=1,...,m.$

Тогда $\lambda^0>0, \mu_j^0\geq 0, j=1,...,n,\quad \nu_i^0\geq 0, i=1,...,m,\quad \sum_{j=1}^{n}\mu_j^0+\sum_{i=1}^{m}\nu_i^0+\lambda^0=1$, т.е. вектор $z^0=(\mu_1^0,...,\mu_n^0,\nu_1^0,...,\nu_n^0,\lambda^0)$ – смешанная стратегия в игре с матрицей $D$ размерности $(n+m+1)\times(n+m+1)$. Ограничения задач (3.53), (3.54) и соотношение двойственности $\sum_{j=1}^{n}c_jx_j^0=\sum_{i=1}^{m}b_iy_i^0$ дают для $z^0$ соотношения (3.55) – (3.57), которые являются достаточными условиями оптимальности (см. (3.46)) для игры с матрицей $D$. Следовательно, $z^0$ – оптимальная стратегия (любого игрока) в игре с матрицей $D$.

Наконец, последнее утверждение теоремы вытекает из уже доказанных утверждений. Действительно, в предположении существования решения задач (3.53), (3.54) построена оптимальная стратегия с последней компонентой, отличной от нуля. Поэтому, если нет оптимальных стратегий в игре с матрицей $D$, последняя компонента которых отлична от нуля, то задачи (3.53), (3.54) не могут иметь решения. С другой стороны, если существует оптимальная стратегия с отличной от нуля последней компонентой, то существование решения задач (3.53), (3.54) уже доказано конструктивно.

Полученные результаты позволяют свести решение игр к решению задач линейного программирования и применить к ним, например, симплекс-метод. Сведением задач линейного программирования к играм пользуются реже.

## Задачи и упражнения

**1.** Решите задачу линейного программирования симплекс-методом.

$$f(x) = x_1 + 9x_2 + 5x_3 + 3x_4 + 4x_5 + 14x_6 \to \min,$$

$$x_1 + x_4 = 20, \quad x_2 + x_5 = 50, \quad x_3 + x_6 = 30, \quad x_4 + x_5 + x_6 = 60, \quad x_j \geq 0, \ j = 1,...,6$$

**2.** Имеется три вида сырья А, В, С, которые используются для производства двух видов продуктов – 1 и 2. В распоряжении находятся 500 единиц сырья А, 750 единиц сырья В и 200 единиц сырья С. Для производства единицы продукта 1 требуется одна единица сырья А и две единицы сырья В. Для производства единицы продукта 2 требуется две единицы сырья А, одна единица сырья В и одна единица сырья С. Прибыль от производства одной единицы продукта 1 составляет 4 у.е., а от единицы продукта 2 – 5 у.е. Сколько нужно производить каждого продукта, чтобы максимизировать общую прибыль?

**3.** Необходимо составить наиболее дешевую смесь из трех веществ. В состав смеси должно входить не менее 6 единиц химического вещества А, не менее 8 единиц вещества В и не менее 12 единиц вещества С. Имеется три вида продукта (1, 2, 3), содержащих эти химические вещества в следующих пропорциях:

| | A | B | C |
|---|---|---|---|
| 1 | 2 | 1 | 3 |
| 2 | 1 | 2 | 4 |
| 3 | 3 | 1,5 | 2 |

Стоимость одной весовой единицы продукта 1 – 2 у.е., продукта 2 – 3 у.е., продукта 3 – 2,5 у.е.

**4.** Металлургический завод из металлов А, В, С может выпускать сплавы трех видов. В течение планируемого периода завод должен освоить не менее 640 т. металла А и 800 т. металла В, при этом металла С может быть израсходовано не более 860 т. Определить минимальные затраты, если нормы расхода и себестоимость следующие:

| Вид Металлов | Нормы расхода металла | | | $b$ |
|---|---|---|---|---|
| | Сплав 1 | Сплав 2 | Сплав 3 | |
| A | 1 | 4.3 | 2.6 | 640 |
| B | 5 | 1.5 | 3 | 800 |
| C | 3 | 3.9 | 4.3 | 860 |
| $c$ | 18 | 15 | 15 | |

**5.** Груз, находящийся в пунктах А и В, необходимо перебазировать в пункты С и D. В этих пунктах имеется соответственно груза на 6 и 4 машины. В пункты С и D надо отправить соответственно 3 и 7 машин груза. В таблице указаны расстояния между пунктами в километрах. Требуется спланировать перевозки так, чтобы суммарный пробег был наименьшим.

| | C | D |
|---|---|---|
| A | 80 | 30 |
| B | 60 | 90 |

**6.** На трех складах A, B, C находится сортовое зерно соответственно 10, 15, 25 т., которое надо доставить в четыре пункта: пункту 1 – 5 т., 2 – 10 т., 3 – 20 т., и 4 – 15 т. Стоимость доставки одной тонны со склада A в указанные пункты соответственно равны 80, 30, 50, 20 у.е.; со склада B – 40, 10, 60, 70 у.е. и со склада C – 10, 90, 40, 30 у.е. Составить оптимальный план перевозки зерна в четыре пункта, минимизирующий стоимость перевозок.

**7.** Найти максимальное значение функции $f = 3x_1 - 2x_2$ при ограничениях

$$7x_1 + 2x_2 \geq 14, \quad -x_1 + 2x_2 \geq 2, \quad 7x_1 + 10x_2 \leq 28, \quad x_i \geq 0, i = 1,2.$$

**8.** Найти максимальное значение функции $f = -16x_1 - x_2 + x_3 + 5x_4 + 5x_5$ при ограничениях $2x_1 + x_2 + x_3 = 10, \quad -2x_1 + 3x_2 + x_4 = 6, \quad 2x_1 + 4x_2 - x_5 = 8, \quad x_i \geq 0, i = 1,...,5$

**9.** При продаже двух видов товаров используется 3 типа ресурсов. Норма затрат ресурсов на реализацию единицы товара, общий объем каждого ресурса и прибыль от реализации единицы товара заданы в таблице.

| Ресурсы | $b$ | Норма затрат | |
|---|---|---|---|
| | | 1 товар | 2 товар |
| 1 тип | 12 | 2 | 2 |
| 2 тип | 8 | 1 | 2 |
| 3 тип | 16 | 4 | 0 |
| $c$ | | 2 | 3 |

Требуется найти оптимальный план реализации товаров, обеспечивающий торговому предприятию максимальную прибыль.

**10.** Построить двойственную задачу, решить ее графически и по ее решению найти оптимальное решение исходной:

$$f = 7x_1 + 8x_2 + 12x_3 \to \max,$$

$$2x_1 + x_2 + 3x_3 \leq 1{,}5, \quad x_1 + 2x_2 + 4x_3 \leq 2, \quad x_1 \geq 0, x_2 \geq 0.$$

**11.** Для производства двух видов продукции A и B используются три вида сырья. Исходные данные приведены в таблице:

| Сырье | $b$ | Продукция | |
|---|---|---|---|
| | | А | В |
| 1 тип | 35 | 1 | 5 |
| 2 тип | 16 | 2 | 1 |
| 3 тип | 6 | 1 | 0 |
| $c$ | | 2 | 3 |

Требуется 1) составить линейную оптимизационную модель, объяснить ее элементы и соотношения; 2) графическим методом найти оптимальный план и оптимальное значение целевой функции; 3) записать двойственную задачу; 4) найти решение двойственной задачи; дать интерпретацию полученным значениям $y_1^*, y_2^*, y_3^*$; 5) провести анализ чувствительности к изменению $c_1$, $c_2$; 6) провести анализ чувствительности к изменению $b_1$, $b_2$, $b_3$; 7) предприятие может приобрести дополнительно сырье одного из видов на сумму 1 у.е.; какое сырье следует приобрести, если стоимость единицы сырья 1, 2 и 3 видов равна 0.4 , 0,8 , 0,2 соответственно?

**12.** Какое из этих множеств не является выпуклым?

а. круг; б. кольцо; в. полупространство.

**13.** Какое из следующих утверждений верно?

а. если одна из пары двойственных задач ЛП имеет оптимальное решение, то оптимальное решение имеет и другая задача;

б. если одна из пары двойственных задач ЛП неразрешима ввиду неограниченности целевой функции, то другая задача имеет решения;

в. если $i$-я координата оптимального решения двойственной задачи ЛП отлична от нуля, то соответствующее ограничение исходной задачи ЛП является неравенством.

**14.** Признак оптимальности звучит следующим образом

а. допустимые планы двойственных ЗЛП являются оптимальными, если они совпадают;

б. допустимые планы двойственных ЗЛП являются оптимальными, если они опорные;

в. допустимые планы двойственных ЗЛП являются оптимальными, если значения целевые функций на них совпадают.

## Глава 4. ДИНАМИЧЕСКАЯ ОПТИМИЗАЦИЯ

### 4.1. Примеры задач оптимизации динамических систем

В предыдущих главах изучались статические задачи оптимизации, условия которых не зависели от времени. Однако практика ставит перед человеком проблемы управления динамическими системами. Состояние их есть функция времени, и управление ими представляет собой такой развернутый во времени план поведения управляющего органа, который обеспечивает достижение системой поставленной перед ней цели.

Такими системами могут являться технические объекты: поезд метро, крылатая ракета, ядерный реактор. Это хорошо известные управляемые объекты. Но во второй половине XX века внимание математиков, экономистов, социологов было обращено к задачам управления социально – экономическими системами, что привело к значительной «математизации» традиционно гуманитарных наук. Примером может служить следующая экономическая модель оптимального производства. Пусть некоторая фирма, ориентируясь на цену $p_{t-1}$, установленную на единицу продукта в период времени $t-1$, способна произвести $s(p_{t-1},u_t)$ единиц этого продукта. Здесь $u_t$ – некоторый управляющий параметр, выбираемый в момент времени $t$ из множества $U_t$. Пусть далее $d(p_t)$ функция спроса на продукт, продаваемый в период времени $t$ по цене $p_t$. Известно, что функции $s$ и $d$ являются монотонно возрастающей и монотонно убывающей по переменной $p$ соответственно. То есть фирма готова производить и продавать тем больше продукции, чем выше на нее цена, но спрос на рынке на продукт падает с увеличением его цены. Естественно, что фирме не имеет смысла производить в каждый период времени продукта больше, чем его может поглотить рынок. Равенство спроса и предложения, можно записать в виде $s(p_{t-1},u_t)=d(p_t)$. Разрешив это уравнение относительно $p_t$, получим, что цена $p_t$ является некоторой заданной функцией переменных $p_{t-1},u_t$, или

$$p_t = f(p_{t-1},u_t). \tag{4.1}$$

Задача фирмы состоит в том, чтобы, регулируя производство продукта с помощью управлений $u_t$, получить максимально возможный доход от продаж за период времени $T$. Иными словами, требуется максимизировать функцию $G(p_0,u_1,u_2,...,u_T)=\sum_{t=1}^{T} p_t\, s(p_{t-1},u_t)$. Здесь $p_0$ некоторое известное фирме заданное начальное значение цены, на которую она ориентируется, чтобы рассчитать план производства $s(p_0,u_1)$ на первый период. В этом примере состояние системы в момент времени $t$ можно характеризовать значением цены $p_t$, управлениями являются величины $u_t \in U_t,\ t=1,2,...,T$, а $G(p_0,u_1,u_2,...,u_T)$ – целевая функция. Изменение цены во времени описывается уравнением (4.1).

Рассмотрим частный случай приведенной задачи, так называемую *паутинную* модель рынка. В ней функции спроса и предложения линейны

$$s(p_{t-1},u_t)=S+u_t p_{t-1},\ d(p_t)=D-\delta p_t, \tag{4.2}$$

где $S<D,\delta$ положительные параметры, $0<a\le u_t\le b$. Параметр $u_t$ характеризует чувствительность функции производства продукта к его рыночной стоимости. Свое название «паутинная» модель получила из-за того, что траектория системы в плоскости с координатами цены – продажи напоминает паутину. Предположим, что параметр $u_t \equiv u$ не зависит от времени $t$. Графики функций спроса и предложения изображены на рис. 4.1. в виде прямых $s$ и $d$. Пусть в начальный момент времени фирма, ориентируясь на цену $p_0$, произведет продукцию в количестве $s_1=s(p_0,u)$ (точка $C_1$ на рис. 4.1.). Покупатель купит весь этот продукт, но уже по цене $p_1$ (точка $C_2$ на рис. 4.1.). При этом $s_1=d_1=d(p_1)$. В следующий период времени фирма, исходя из цены $p_1$, произведет продукции в количестве $s_2=s(p_1,u)$ (точка $C_3$ на рис. 4.1.). Покупатель вновь приобретет весь товар по цене $p_2$ (точка $C_4$ на рис. 4.1.). Причем $s_2=d_2=d(p_2)$. Продолжая этот процесс, мы получим последовательность точек $C_5,C_6...$ и т.д. На рисунке видно, что траектория сходится к точке $C$ с координатами $(p^*,s^*)$, которая является точкой устойчивого равновесия на рынке. Это значит, что любое отклонение от нее порождает траекторию, возвращающую систему в эту точку. Нетрудно видеть, что если бы прямая $s$ шла достаточно круто, то траектория уходила бы все дальше от точки $C$. И, наконец, если бы прямые $s$ и $d$ были наклонены к горизонтальной прямой под равными по модулю углами, то траектория представляла бы собой прямоугольный цикл, то есть рыночная ситуация периодически повторялась.

Исследуем эту модель аналитически. Итак, по–прежнему, $u_t \equiv u>0$. Приравняем правые части функций предложения и спроса (4.2). Тогда

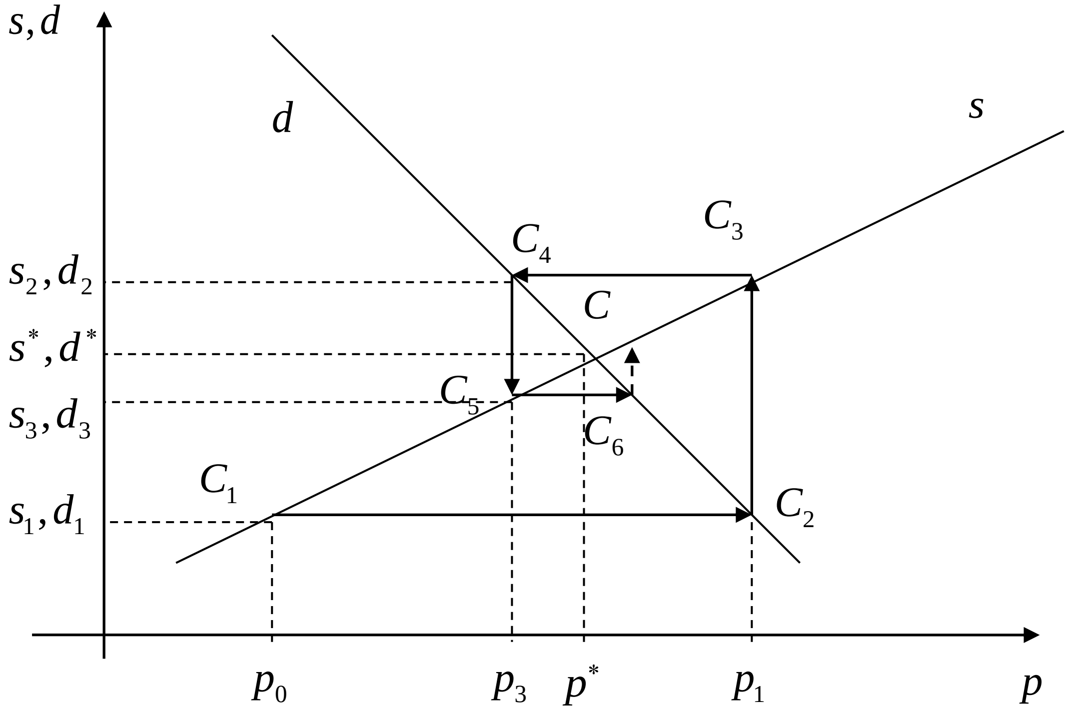


Рис. 4.1.

$$\delta p_t + u p_{t-1} = D - S . \tag{4.3}$$

Уравнение (4.3) является разностным уравнением первого порядка. Частное решение этого уравнения получим, если положим $p_t \equiv p^*$. Из (4.3) получаем выражение для равновесного значения цены

$$p^* = \frac{D - S}{\delta + u} .$$

При этом равновесное значение спроса и предложения равны

$$S + up^* = s^* = d^* = D - \delta p^* = \frac{uD + \delta S}{\delta + u} . \tag{4.4}$$

Непосредственной проверкой можно убедиться, что общее решение уравнения (4.3) можно записать в виде

$$p_t = p^* + A\left(-\frac{u}{\delta}\right)^t . \tag{4.5}$$

Значение постоянной $A$ найдем из начального условия $y \in Y$, откуда

$$A = p_0 - p^* .$$

Из (4.3) – (4.5) получим выражения для функций предложения и спроса как функций времени $t$

$$s_t = s^* + u\left[(p_0 - p^*)\left(-\frac{u}{\delta}\right)^{t-1}\right], \quad d_t = d^* - \delta\left[(p_0 - p^*)\left(-\frac{u}{\delta}\right)^{t}\right]. \qquad (4.6)$$

Решение (4.5) является колебательным, то есть цена последовательно принимает значения либо большие, либо меньшие равновесной величины $p^*$, а уровни продаж $s_t$ и покупок $d_t$ принимают согласно (4.6) последовательно значения большие или меньшие равновесных. При этом, из (4.4)–(4.6) следуют предельные соотношения $\lim\limits_{t\to\infty} p_t = p^*$, $\lim\limits_{t\to\infty} s_t = s^*$, $\lim\limits_{t\to\infty} d_t = d^*$, если $\frac{u}{\delta} < 1$, так как $\lim\limits_{t\to\infty}\left(-\frac{u}{\delta}\right)^t = 0$. Если же $\frac{u}{\delta} > 1$, то процесс будет расходящимся, и траектория будет удаляться от точки равновесия $C$. В случае равенства $\frac{u}{\delta} = 1$ система будет совершать периодические колебания

$$p_t = p^* + (p_0 - p^*)(-1)^t, s_t = s^* + u\left[(p_0 - p^*)(-1)^{t-1}\right],$$

$$d_t = d^* - \delta\left[(p_0 - p^*)(-1)^t\right].$$

Частным случаем приведенной выше оптимизационной задачи может служить следующая постановка. Требуется выбрать такое значение управляющего параметра $u$ из множества $0 < a \le u \le b$, чтобы функция дохода фирмы от продаж на отрезке времени $[0, T]$ была максимальной. Используя явную зависимость (4.5), (4.6) функций цены и производства от управления $u$ и времени $t$, можно записать целевую функцию дохода

$$G(p_0, u) = \sum_{t=1}^{T} p_t\, s_t =$$

$$= \sum_{t=1}^{T}\left\{\left[p^* + (p_0 - p^*)\left(-\frac{u}{\delta}\right)^t\right]\left[s^* + u(p_0 - p^*)\left(-\frac{u}{\delta}\right)^{t-1}\right]\right\}.$$

Ранее описанная динамическая оптимизационная задача, конечно, сложнее только что сформулированной, поскольку предполагает поиск управления $u_t$ как функции от времени. Обе эти задачи можно записать как задачи нелинейного программирования, изученные в первой главе. Об этом речь пойдет ниже. Однако, если считать время непрерывной величиной, то оптимизационная задача меняет свой тип, и ее решение требует специального исследования.

В качестве второго примера рассмотрим задачу управления самолетом. Везде ниже производную по времени от функции $f(t)$ будем обозначать $\dot{f} \equiv \frac{df}{dt}$.

В упрощенном виде движение самолета можно свести к движению его центра масс. В вертикальной плоскости в скоростной системе координат изменение горизонтальной координаты $x$ и вертикальной $h$ можно описать дифференциальными уравнениями

$$\dot{x} = v\cos\theta,$$

$$\dot{h} = v\sin\theta,$$

где $v$ – это модуль скорости, $\theta$ – угол наклона вектора скорости к горизонту или угол тангажа. В свою очередь скорость и тангаж меняются по закону

$$m\dot{v} = P - X(v,\alpha) - mg\sin\theta,$$

$$mv\dot{\theta} = (P\sin\alpha + Y(v,\alpha)) - mg\cos\theta,$$

$$\dot{m} = F(P)$$

Здесь $m, g$ масса самолета и ускорение свободного падения, $P, \alpha$ – модуль вектора тяги, и угол между вектором скорости и плоскостью крыла, (угол атаки), соответственно. Величины $X, Y$ представляют собой силу лобового сопротивления и подъемную силу. Считается, что они являются заданными функциями скорости и угла атаки. Функция расхода топлива $F(P)$ также считается известной. Таким образом, фазовое состояние самолета описывается в каждый момент времени пятью координатами $(x,h,v,\theta,m)$, изменяющимися во времени согласно описанной системе дифференциальных уравнений. Переменные $P, \alpha$ являются управляющими параметрами или функциями управления. Они не произвольны, и их значения лежат в определенных пределах $0 \le P \le P_{\max}$, $\alpha_1 \le \alpha \le \alpha_2$. Изменяя их, мы можем получать различные траектории движения самолета. Сформулируем некоторые оптимизационные задачи для изучаемого объекта. Пусть в начальный момент времени $t=0$ фазовые координаты самолета равны $x(0)=0, h(0)=0, v(0)=0, \theta(0)=0, m(0)=m_0$. Поставим две экстремальные задачи.

1.) Задача о скороподъемности. Требуется достичь высоты $H$ за минимально возможное время с заданным ограничением по расходу топлива. Формальная запись задачи следующая: $h(T)=H,\, m(T)\ge m_1,\, T\to\min$. Фазовые координаты $x(T), v(T), \theta(T)$ не задаются.

2.) Задача о дальности полета. За фиксированное время $T$ необходимо достигнуть максимально возможной горизонтальной координаты, израсходовав количество топлива не более заданного, или $m(T)\ge m_1,\, x(T)\to\max$.

Заметим, что в первой задаче координата $x$ не входит явно ни в одно уравнение, и ее можно исключить из рассмотрения. Во второй задаче ее можно так же не учитывать при интегрировании уравнений для $h, v, \theta, m$, а условие $x(T)\to\max$ заменить условием $\int\limits_0^T v(t)\cos(\theta(t))dt\to\max$.

Максимизируемые или минимизируемые величины в задачах оптимального управления называются целевыми *функционалами,* так как они зависят от всей траектории движения. Вообще, функционалами называются отображения функций из некоторого заданного множества в точки множества действительных чисел. Например, в задаче о скороподъемности значение $T$ зависит от того, как происходил набор высоты, т.е. от функций управления $P(t), \alpha(t)$, заданных на всем интервале $(0,T)$. В задаче о максимальной дальности целевой функционал $x(T)=\int\limits_0^T v(t)\cos(\theta(t))dt$ отображает функции $v(t)\cos(t)$ в точки действительной числовой прямой.

Как мы убедились на рассмотренных примерах, математические модели систем, состояние которых является функцией времени, делятся на два важных класса: дискретные, в которых время принимает дискретные значения, и непрерывные, в которых время непрерывно. В первом из них каждое последующее значение переменного фазового состояния однозначно определяется предыдущим состоянием и управлением. Такие системы называются *динамическими.* В непрерывных моделях изменение состояния описывается обыкновенными дифференциальными уравнениями с правой частью, зависящей от текущего значения состояния и управления.

В обоих рассмотренных примерах состояние системы определялось конечным набором величин или вектором. Существуют и более сложные системы, в которых состояние описывается вектором с бесконечным числом координат или функцией от нескольких переменных, такие системы называются системами с распределенными параметрами. Иногда поведение системы в данный момент времени зависит явно не только от текущего значения состояния и управления, но и от состояний и управлений, заданных в более ранние моменты. Такие системы называются системами с запаздыванием. Исследование распределенных систем и систем с запаздыванием зачастую представляет очень трудную задачу, и выходит за рамки данного учебника. Ниже будут изучаться лишь многошаговые задачи управления и системы, описываемые обыкновенными дифференциальными уравнениями.

Поставленные выше задачи оптимального управления можно в общем случае сформулировать следующим образом. В начальный момент времени система находится в состоянии, которое описывается конечномерным вектором $x_0$. Требуется перевести систему из этого состояния в конечное состояние $x_1$, которое не обязательно жестко зафиксировано, по траекториям системы так, чтобы добиться наилучшего результата. Соответствующее оптимальное управление, с помощью которого можно осуществить этот процесс, обозначим $u(t,x_0)$. Такое управление называется *программным*, т.е. оно представляет будущую программу действий. Это значит, что в каждый момент времени $t$ следует применять управление $u(t,x_0)$, чтобы добиться желаемого результата. Пусть $x(t,x_0)$ оптимальная траектория, соответствующая оптимальному управлению $u(t,x_0)$. Можно представить себе, что из-за каких-то помех в момент времени $\tau$ система оказалась не в точке $x(t,x_0)$ оптимальной траектории, а в некотором ином состоянии $y$. Это значит, что для достижения поставленной цели требуется скорректировать управление, т.е. найти оптимальное управление $u(t,y)$, которое наилучшим образом переведет систему из точки $y$ в состояние $x_1$. Если можно описать множество значений $Y$, в которые может попасть система из-за внешних помех, то имеет смысл решить задачу поиска управления $u(t,y)$ для каждого вектора состояний $y \in Y$. Задача построения оптимальных управлений $u(t,y)$ как функций не только переменной $t$, но и начального состояния $y \in Y$, называется задачей *синтеза* оптимального управления или задачей поиска управления с *обратной связью*. Понятно, что решение задачи оптимального управления представляет собой частный случай решения задачи синтеза. Для дискретных систем задачу синтеза часто можно решить с помощью метода динамического программирования, о котором речь пойдет в следующем пункте.

### 4.2. Задача динамического программирования. Функция Беллмана

Если задачи математического программирования допускают разбиение процесса решения на отдельные этапы (шаги), то их принято классифицировать как задачи динамического программирования. Задачи динамического программирования, как и любые другие задачи оптимизации, являются математическими моделями некоторых практических задач принятия решения (или управления).

В динамическом программировании рассматриваются методы, которые позволяют путем поэтапной оптимизации получить общий оптимум. Одним из таких методов является метод последовательной оптимизации (также называемый методом динамического программирования). Этот метод применяется в том случае, когда для оптимизации общего критерия качества системы надо многократно выбирать решение, причем система такова, что в ней можно выделить отдельные шаги, причем решения на последующих шагах не оказывают влияния на величину показателя качества, достигнутую на предыдущих шагах. Метод динамического программирования дает хорошие результаты при малом числе возможных решений на каждом шаге, и, прежде всего, когда динамический процесс описывается небольшим числом переменных. При этих условиях сложность решения задач при увеличении числа шагов возрастает умеренно.

Обратимся к многошаговым задачам оптимизации, т.е. задачам, оптимизацию в которых можно представить в виде ряда последовательных этапов. Рассмотрим некоторый процесс. Предположим, что состояние этого процесса описывается $n$-мерным вектором $x=(x_1,x_2,...,x_n)$, или, что то же самое, точкой $x$ пространства $R^n$, которое называют **фазовым пространством**. Будем считать, что процесс является $N$-шаговым, т.е. он происходит в $N$ этапов в соответствии со следующей схемой:

$$\underset{\text{начальное состояние}}{x^{(1)}} \to \ldots \to \underbrace{x^{(k)} \to x^{(k+1)}}_{k-\text{й шаг}} \to \ldots \to x^{(N)} \to \underset{\text{конечное состояние}}{x^{(N+1)}}.$$

Переход между состояниями на $k$-м шаге происходит в соответствии с уравнением состояний:

$$x^{(k+1)} = \varphi^{(k)}(x^{(k)}, u^{(k)}), \tag{4.7}$$

где $u^{(k)} \in R^m$ есть $m$-мерный вектор управления, выбираемый на $k$-м шаге, $\varphi^{(k)}(x,u)$ – заданная $n$-мерная вектор-функция аргументов $x \in R^n, u \in R^m$. Таким образом, предполагается, что в результате $k$-го шага процесс переходит в состояние $x^{(k+1)}$, которое определяется только начальным состоянием $x^{(k)}$ этого шага и выбранным на нем вектором управления $u^{(k)}$ и не зависит от “предыстории” процесса до $k$-го шага, т.е. от $x^{(1)},...,x^{(k-1)}$ и $u^{(1)},...,u^{(k-1)}$.

Показателем эффективности $k$-го шага является заданная числовая характеристика (целевая функция этого шага) $f_k = f_k(x^{(k)}, u^{(k)}),\ k = 1,2,...,N.$ Предположим, что эффективность всего процесса в целом характеризуется целевой функцией вида

$$F(\bar{x},\bar{u}) = \sum_{k=1}^{N} f_k(x^{(k)}, u^{(k)}) + f_{N+1}(x^{(N+1)}), \tag{4.8}$$

где $\bar{x} = \{x^{(1)}, x^{(2)},..., x^{(N+1)}\}$ – набор состояний, называемый фазовой **траекторией процесса**, а $\bar{u} = \{u^{(1)}, u^{(2)},..., u^{(N)}\}$ – набор векторов управления, который называется **управлением процессом**. Таким образом, рассматриваются только аддитивные (сепарабельные) целевые функции $F$, представимые в виде суммы целевых функций шагов $f_k$.

Предположим далее, что на фазовую траекторию и выбор управлений наложены ограничения

$$x^{(k)} \in X_k, \quad k = 1,..., N+1, \tag{4.9}$$

$$u^{(k)} \in U_k(x^{(k)}), \quad k = 1,..., N, \tag{4.10}$$

где $X_k$ и $U_k(x^{(k)})$ – заданные множества в пространствах $R^n$ и $R^m$ соответственно, причем множество $U_k$ зависит от начального состояния $x^{(k)}$ $k$-го шага. Ограничения (4.9) называются **фазовыми**. Ограничения на начальное и конечное состояния процесса $x^{(1)} \in X_1,\ x^{(N+1)} \in X_{N+1}$ называются начальными и конечными условиями. При этом множества $X_1$ и $X_{N+1} \subset R^n$ во многих случаях содержат по одной точке (начало и конец фазовой траектории).

Пусть $\bar{u} = \{u^{(1)}, u^{(2)},..., u^{(N)}\}$ – управление процессом, удовлетворяющее ограничениям (4.10) и переводящее его из некоторого начального состояния $x^{(1)} \in X_1$ в некоторое конечное состояние $x^{(N+1)} \in X_{N+1}$ в соответствии с уравнениями (4.7) с учетом ограничений (4.9). Обозначим множество всех таких управлений буквой $U$. **Многошаговая задача оптимизации** формулируется следующим образом: *среди всех управлений* $\bar{u} \in U$ *выбрать такое* $\bar{u}^* = \{u^{(1)*}, u^{(2)*},..., u^{(N)*}\}$, *для которого целевая функция* (4.8) *принимает минимальное или максимальное* (в зависимости от смысла задачи) *значение*. Управление $\bar{u}^*$ и соответствующая ему фазовая траектория $\bar{x}^*$ называются **оптимальными**. Таким образом, динамическая (многошаговая) задача оптимизации формулируется следующим образом:

$$F(\bar{x},\bar{u})=\sum_{k=1}^{N} f_k(x^{(k)},u^{(k)})+f_{N+1}(x^{(N+1)})\to \underset{\bar{u}}{extr}, \tag{4.11}$$

при условиях

$$x^{(k)}=\varphi^{(k)}(x^{(k)},u^{(k)}), k=1,...,N, \tag{4.12}$$

и ограничениях (4.9), (4.10).

К многошаговым задачам оптимизации сводятся многие прикладные задачи. Прежде чем решать такие задачи, необходимо научиться составлять условия.

Оптимальное уравнение в динамической задаче обладает тем свойством, что каковы бы ни были решения (управления) на начальном этапе (одном или нескольких шагах), последующие решения должны составлять оптимальное управление для критерия, представляющего собой сумму эффективностей на оставшихся шагах и полученного нового начального состояния. Это утверждение называется **принципом оптимальности**. Математическое выражение принципа оптимальности представляет рекуррентное соотношение, называемое **уравнением Беллмана**. Оптимальная траектория в задаче (4.9) – (4.12) обладает тем свойством, что любая ее завершающая часть, начинающаяся с $k$-го шага, $k=1,...,N-1,$ является оптимальной для остающихся шагов процесса.

Опишем метод динамического программирования, в основе которого лежит принцип оптимальности. Заметим, что в формулировке многошаговой задачи оптимизации (4.9) – (4.12) ограничения на фазовую траекторию (4.9) (включая конечное состояние) можно включить в ограничения на выбор управлений, заменив соотношения (4.9) и (4.10) следующим эквивалентным ограничением:

$$u^{(k)}\in\widetilde{U}_k(x^{(k)})=\{u^{(k)}\in U_k(x^{(k)})\,|\,\varphi^{(k)}(x^{(k)},u^{(k)})\in X_{k+1}\},\quad k=1,...,N. \tag{4.13}$$

С учетом этого перепишем формулировку задачи (4.9) – (4.12):

$$F(\bar{x},\bar{u})=\sum_{k=1}^{N} f_k(x^{(k)},u^{(k)})+f_{N+1}(x^{(N+1)})\to \underset{\bar{u}}{extr}, \tag{4.14}$$

$$x^{(k+1)}=\varphi^{(k)}(x^{(k)},u^{(k)}),\quad k=1,...,N, \tag{4.15}$$

$$u^{(k)}\in\widetilde{U}_k(x^{(k)}),\quad k=1,...,N, \tag{4.16}$$

$$x^{(1)}\in X_1. \tag{4.17}$$

Предположим, что в результате начальных $k-1$ шагов процесс перешел в состояние $x^{(k)}$. Рассмотрим задачу оптимизации оставшихся $N-k+1$ шагов при известном состоянии $x^{(k)}$. Пусть оптимальное управление $\bar{u}^*(k)=\{u^{(k)*},...,u^{(N)*}\}$ последних $N-k+1$ шагов и оптимальная траектория этих шагов $\bar{x}^*(k)=\{x^{(k)},x^{(k+1)*},...,x^{(N+1)*}\}$, начатая из состояния $x^{(k)}$, найдены. Целевая функция

$F^{(k)}(\bar{x}(k),\bar{u}(k))=\sum_{i=k}^{N} f_i(x^{(i)},u^{(i)})+f_{N+1}(x^{(N+1)})$ последних $N-k+1$ шагов при $\bar{x}(k)=\bar{x}^*(k),\quad \bar{u}(k)=\bar{u}^*(k)$ принимает оптимальное (т.е. минимальное или максимальное) значение, зависящее от начального состояния $x^{(k)}$ фазовой траектории этих шагов, т.е.

$$\underset{\bar{u}(k)}{extr}\, F^{(k)}(\bar{x}(k),\bar{u}(k))=F^{(k)}(\bar{x}(k),\bar{u}^*(k))=w_k(x^{(k)}). \qquad (4.18)$$

Функция $w_k(x^{(k)})$ из (4.18) называется **функцией Беллмана** последних $N-k+1$ шагов. Очевидно, для последнего шага имеем

$$\begin{aligned} w_N(x^{(N)}) &= \underset{u^{(N)}\in\tilde{U}_N(x^{(N)})}{extr}\left(f_N(x^{(N)},u^{(N)})+f_{N+1}(x^{(N+1)})\right)= \\ &= \underset{u^{(N)}\in\tilde{U}_N(x^{(N)})}{extr}\left(f_N(x^{(N)},u^{(N)})+f_{N+1}(\varphi^{(N)}(x^{(N)},u^{(N)}))\right). \end{aligned} \qquad (4.19)$$

Для всех остальных шагов функции Беллмана связаны между собой следующими соотношениями, вытекающими из принципа оптимальности:

$$\begin{aligned} w_k(x^{(k)}) &= \underset{u^{(k)}\in\tilde{U}_k(x^{(k)})}{extr}\left(f_k(x^{(k)},u^{(k)})+w_{k+1}(x^{(k+1)})\right)= \\ &= \underset{u^{(k)}\in\tilde{U}_k(x^{(k)})}{extr}\left(f_k(x^{(k)},u^{(k)})+w_{k+1}(\varphi^{(k)}(x^{(k)},u^{(k)}))\right),\ k=1,\dots,N-1. \end{aligned} \qquad (4.20)$$

Соотношения (4.19) и (4.20), позволяющие последовательно найти функции Беллмана $w_N(x^{(N)}), w_{N-1}(x^{(N-1)}),\dots,.w_1(x^{(1)}),$ называются **уравнениями Беллмана**.

Находя функции $w_k(x^{(k)}),\ k=N,\dots,1,$ из (4.19) и (4.20), мы одновременно определяем и управления $u^{(k)*}(x^{(k)}),$ которые реализуют экстремумы в правых частях, т.е. отвечают оптимальным значениям соответствующих величин из правых частей (4.19), (4.20). Управления $u^{(k)*}(x^{(k)}),\ k=1,\dots,N,$ называются **условно оптимальными управлениями**, а процесс их нахождения – **условной оптимизацией** (или обратной прогонкой). Отметим, что управления $u^{(k)*}(x^{(k)})$ удовлетворяют принципу оптимальности, т.е. в зависимости от начального состояния $x^{(k)}$ управление $u^{(k)*}(x^{(k)})$ учитывает оптимизацию не только $k$-го шага, но и следующих за ним $N-k$ шагов.

Итак, в результате условной оптимизации находятся функции Беллмана $w_k(x^{(k)})$ и условно оптимальные управления $u^{(k)*}(x^{(k)}),\ k=1,\dots,N$. После этого можно осуществить **безусловную оптимизацию** (прямую прогонку) в задаче (4.14) – (4.17), т.е. определить искомое оптимальное управление процессом $\bar{u}^*=\{u^{(1)*},\dots,u^{(N)*}\}$ и оптимальную фазовую траекторию $\bar{x}^*=\{x^{(2)*},\dots,x^{(N+1)*}\}$ следующим образом. Так как функция Беллмана $w_1(x^{(1)})$ для каждого начального состояния $x^{(1)}\in X_1$ равна оптимальному значению целевой функции за $N$ шагов, т.е. всего процесса, начатого из состояния $x^{(1)}$, то оптимальное начальное условие $x^{(1)*} \in X_0$ находим из соотношения

$$w_1(x^{(1)*}) = \underset{x^{(1)} \in X_1}{extr}\ w_1(x^{(1)}) \tag{4.21}$$

(если множество $X_1$ из (4.17) состоит из единственной точки $x^{(1)}$, то полагаем $x^{(1)*} = x^{(1)}$). Далее, используя найденные условные оптимальные управления, а также уравнения состояний (4.15), последовательно находим $u^{(1)*}, x^{(1)*}, u^{(2)*}, x^{(2)*}, ..., u^{(N)*}, x^{(N)*}$ из следующих соотношений:

$$\begin{aligned}
&u^{(1)*} = u^{(1)*}(x^{(1)*}), \quad x^{(2)*} = \varphi^{(1)}(x^{(1)*}, u^{(1)*}), \\
&u^{(2)*} = u^{(2)*}(x^{(2)*}), \quad x^{(3)*} = \varphi^{(2)}(x^{(2)*}, u^{(2)*}), \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&u^{(N)*} = u^{(N)*}(x^{(N)*}), \quad x^{(N+1)*} = \varphi^{(N)}(x^{(N)*}, u^{(N)*}).
\end{aligned} \tag{4.22}$$

Теперь рассмотрим **алгоритм** метода динамического программирования.

**Этап I** (***условная оптимизация***).

Шаг $N$. Находим условно оптимальное управление $u^{(N)*}(x^{(N)})$ и функцию Беллмана $w_N(x^{(N)})$ в соответствии с (4.19);

Шаг ($N$-1). Используя результаты $N$-го шага, находим $u^{(N-1)*}(x^{(N-1)})$ и $w_{N-1}(x^{(N-1)})$ с помощью равенств (4.20) при $k = N-1$;

…………………………………………………………………………

Шаг 1. Используя результаты предыдушего (второго) шага, определяем $u^{(1)*}(x^{(1)})$ и $w_1(x^{(1)})$ в соответствии с (4.20) при $k = 1$.

**Этап II** (***безусловная оптимизация***).

Шаг 1. Находим оптимальное начальное состояние $x^{(1)*}$ в соответствии с (4.21); Определяем оптимальное управление $u^{(1)*}$ и конечное состояние $x^{(2)*}$ первого шага процесса по формулам (4.22);

Шаг 2. Используя результаты предыдущего шага, находим $u^{(2)*}$ и $x^{(3)*}$ в соответствии с (4.22);

…………………………………………………………………………..

Шаг $N$. С помощью результатов ($N-1$)-го шага определяем $u^{(N)*}$ и $x^{(N+1)*}$ по формулам (4.22).

Окончательно имеем $\bar{u}^* = \{u^{(1)*}, ..., u^{(N)*}\}$, $\bar{x}^* = \{x^{(1)*}, ..., x^{(N+1)*}\}$.

Для многих экономических и производственных задач характерной является дискретная модель, предполагающая, что величины, описывающие процесс, могут принимать только дискретный ряд значений. Функциональные зависимости в таких задачах задаются, как правило, в виде таблиц, а не аналитически. Однако общая схема их решения методом динамического программирования остается без изменений. Более того, метод динамического программирования является одним из основных для решения дискретных и целочисленных задач.

### 4.3. Использование метода динамического программирования в решении экономических задач

Рассмотрим пример применения метода динамического программирования к задаче, которую можно интерпретировать как задачу распределения ресурсов между технологическими процессами.

**Пример 4.1.** Пусть имеется некоторое количество ресурсов $x$, которое необходимо распределить между $n$ различными предприятиями так, чтобы получить максимальную суммарную эффективность. Введем обозначения: $u_i$ – количество ресурсов, выделенных $i$-му предприятию (управление), $g_i(u_i)$ – величина дохода от использования $i$-м предприятием выделенного ресурса в количестве $u_i$. Математическая формулировка задачи:

$$\sum_{i=1}^{n} g_i(u_i) \to \max$$

при ограничениях $u_i \geq 0$, $i = 1,...,n$, $\sum_{i=1}^{n} u_i = x$.

Это задача нелинейного программирования, которая может быть решена методом Лагранжа, но мы будем решать ее здесь поэтапно, последовательно выделяя ресурсы предприятиям. Обозначим максимальный доход от распределения ресурсов в количестве $z$ между предприятиями с номерами от $k$ до $n$ через $f_k(z)$, тогда оптимальное значение задачи есть $w_1(x)$. Если мы дошли до последнего шага ($n$-го предприятия), то $w_n(z) = g_n(z)$. На предпоследнем шаге имеем

$$w_{n-1}(z) = \max_{0 \leq u_{n-1} \leq z} [g_{n-1}(u_{n-1}) + g_n(z - u_{n-1})] = \max_{0 \leq u_{n-1} \leq z} [g_{n-1}(u_{n-1}) + w_n(z - u_{n-1})].$$

Для произвольного шага имеем соотношение

$$w_k(z) = \max_{0 \leq u_k \leq z} [g_k(u_k) + w_{k+1}(z - u_k)], \quad k = 1,...,n-1,$$

которое является уравнением Беллмана данной задачи.

На первом шаге (при $k$=1) имеем соотношение $w_1(z)=\max\limits_{0\le u_1\le z}[g_1(u_1)+w_2(z-u_1)]$.

Последовательно определяя $w_n(z),\ldots,w_2(z),\ w_1(z),$ для произвольного $z$ и подставляя $z = x$ в $w_1(z),$ получаем оптимальный результат и оптимальное управление на первом шаге $u_1^*$. Оптимальное управление получается последовательной подстановкой в решения уравнения Беллмана, которые представляют собой условно оптимальными управления $u_k^*(z)$.

Для иллюстрации этой вычислительной процедуры рассмотрим конкретный пример с $g_i(u_i)=c_i\sqrt{u_i}$. Итак, требуется максимизировать $\sum\limits_{i=1}^{n}c_i\sqrt{u_i}$ при ограничении $\sum\limits_{i=1}^{n}u_i=x$. Имеем $w_1(x-\sum\limits_{i=2}^{n}u_i)=\max\limits_{0\le u_1\le x-\sum\limits_{i=2}^{n}u_i}c_1\sqrt{u_1}=c_1\sqrt{x-\sum\limits_{i=2}^{n}u_i}$, а условно оптимальное использование ресурса в первом процессе есть $u_1^*(x-\sum\limits_{i=2}^{n}u_i)=x-\sum\limits_{i=2}^{n}u_i$. Далее, $w_2(x-\sum\limits_{i=3}^{n}u_i)=\max\limits_{0\le u_2\le x-\sum\limits_{i=3}^{n}u_i}\left[c_2\sqrt{u_2}+c_1\sqrt{x-\sum\limits_{i=2}^{n}u_i}\right]$. Дифференцируя выражение в квадратных скобках по $u_2$, и приравнивая производную к нулю, получаем соотношение

$\frac{1}{2}c_2u_2^{-1/2}-\frac{1}{2}c_1(x-\sum\limits_{i=2}^{n}u_i)^{-1/2}=0$ ,откуда $u_2^*(x-\sum\limits_{i=3}^{n}u_i)=\dfrac{c_2^2}{c_1^2+c_2^2}(x-\sum\limits_{i=3}^{n}u_i)$,

$w_2(x-\sum\limits_{i=3}^{n}u_i)=\sqrt{(c_1^2+c_2^2)(x-\sum\limits_{i=3}^{n}u_i)}$. Используя вторую производную, легко убедиться, что $u_2^*$ дает именно максимум выражению в квадратных скобках.

На следующем шаге $w_3(x-\sum\limits_{i=4}^{n}u_i)=\max\limits_{0\le u_3\le x-\sum\limits_{i=4}^{n}u_i}\left[c_3\sqrt{u_3}+\sqrt{(c_1^2+c_2^2)(x-\sum\limits_{i=3}^{n}u_i)}\right]$. Решая задачу максимизации, как и на предыдущем шаге, получаем

$$u_3^*(x-\sum_{i=4}^{n}u_i)=\frac{c_3^2}{c_1^2+c_2^2+c_3^2}(x-\sum_{i=4}^{n}u_i);\qquad w_3(x-\sum_{i=4}^{n}u_i)=\sqrt{(c_1^2+c_2^2+c_3^2)(x-\sum_{i=4}^{n}u_i)}.$$

Продолжая вычисления по процедуре, получаем общие выражения для $u_k^*$ и $w_k$:

$u_k^*(x-\sum_{i=k+1}^{n}u_i)=\frac{c_k^2}{\sum_{i=1}^{k}c_i^2}(x-\sum_{i=k+1}^{n}u_i)$, $w_k(x-\sum_{i=k+1}^{n}u_i)=\sqrt{\sum_{i=1}^{k}c_i^2\left(x-\sum_{i=k+1}^{n}u_i\right)}$. Так как для последнего технологического процесса должно выполняться ограничение $0\le u_n\le x$, то на $n$-м шаге находим просто оптимальное значение задачи $w_n(x)=\sqrt{x\sum_{i=1}^{k}c_i^2}$ и оптимальное использование ресурса в $n$-ом процессе $u_n^*=\frac{c_n^2x}{\sum_{i=1}^{k}c_i^2}$. Подставляя $u_n^*$ в выражение для $u_{n-1}^*(x-u_n^*)$, затем в $u_{n-2}^*(x-u_n^*-u_{n-1}^*)$ и т.д., получаем решение задачи $u_k^0=\frac{c_k^2x}{\sum_{i=1}^{n}c_i^2}$.

Рассмотрим процесс решения уравнения Беллмана на следующем численном примере.

**Пример 4.2**. Имеются два предприятия *P* и *Q*. При выделении каждому из них на год средств в количестве *u* единиц предприятие *P* обеспечивает доход 5*u* единиц и остаток от выделенных средств 0.3*u* единиц, а предприятие *Q* – доход 4*v* единиц и остаток выделенных средств 0.5*v* единиц. Обоим предприятиям выделено на 4 года *a*=1000 единиц средств. Нужно распределить их в начале периода и затем остатки средств ежегодно между предприятиями, чтобы общий доход за указанный период был максимальным.

Разобьем весь период продолжительностью в 4 года на 4 этапа, приняв каждый год за один этап. Будем нумеровать этапы, начиная с первого года, и обозначим через $u_k, v_k$ средства, выделяемые соответственно предприятиям *P* и *Q* на *k*-м этапе; сумма $u_k+v_k=x_k$ представляет собой общее количество используемых на *k*-м этапе средств, оставшихся в конце предыдущего (*k*–1)-го этапа. Очевидно, $x_1=a=1000$. Доход, получаемый на *k*-м этапе, равен $f_k(u_k,v_k)=5u_k+4v_k$. Если обозначить через $w_k(x_k)$ максимальный доход, полученный на последних этапах, начиная с *k*-го, т.е. с распределения $x_k$ средств, то функциональное уравнение Беллмана, выражающее принцип оптимальности, имеет вид

$$w_k(x_k)=\max_{u_k+v_k=x_k}[5u_k+4v_k+w_{k+1}(x_{k+1})]. \qquad (4.23)$$

Так как $u_k+v_k=x_k$, то для каждого этапа надо найти значение величины $u_k$, при этом $v_k=x_k-u_k$. Доход на *k*-м этапе составит $5u_k+v_k=5u_k+4(x_k-u_k)=4x_k+u_k$. Количество средств, используемых на $k$-м этапе, выразится рекуррентной формулой $x_k = 0.3u_{k-1} + 0.5v_{k-1} = 0.3u_{k-1} + 0.5(x_{k-1} - u_{k-1})$ или

$$x_k = 0.1\cdot(5x_{k-1} - 2u_{k-1}), \tag{4.24}$$

а функциональное уравнение Беллмана (4.23) примет вид:

$$w_k(x_k) = \max_{0\le u_k \le x_k}[4x_k + u_k + w_{k+1}(x_{k+1})]. \tag{4.25}$$

Полагая последовательно $k = 4, 3, 2, 1$ (планирование начинается с последнего этапа), используя формулы (4.24) и (4.25) и учитывая, что $f_5(x_5) = 0$, получим:

$k = 4,\quad w_4(x_4) = \max\limits_{0\le u_4\le x_4}[4x_4 + u_4] = 5x_4,$ максимум функции $4x_4 + u_4$ достигается в точке $u_4 = x_4$ – правом конце отрезка $[0, x_4],\ v_4 = x_4 - u_4 = 0,\ x_4 = 0.1\cdot(5x_3 - 2u_3);$

$k = 3,\quad w_3(x_3) = \max\limits_{0\le u_3\le x_3}[4x_3 + u_3 + 5\cdot 0.1\cdot(5x_3 - 2u_3)] = \max\limits_{0\le u_3\le x_3}(6.5x_3 + 0u_3) = 6.5x_3$ при любом значении $u_3$, $v_3 = x_3 - u_3;\ x_3 = 0.1\cdot(5x_2 - 2u_2);$

$k = 2,\quad w_2(x_2) = \max\limits_{0\le u_2\le x_2}[4x_2 + u_2 + 6.5\cdot 0.1\cdot(5x_2 - 2u_2)] = \max\limits_{0\le u_2\le x_2}(7.25x_2 - 0.3u_2) = 7.25x_2$,

максимум линейной функции $7.25x_2 - 0.3u_2$ достигается в точке $u_2 = 0$ – левом конце отрезка $[0, x_2],\ v_2 = x_2 - u_2 = x_2,\ x_2 = 0.1\cdot(5x_1 - 2u_1)$;

$k = 1,\quad w_1(x_1) = \max\limits_{0\le u_1\le x_1}[4x_1 + u_1 + 7.25\cdot 0.1\cdot(5x_1 - 2u_1)] = \max\limits_{0\le u_1\le x_1}(7.625x_1 - 0.45u_1) = 7.625x_1,$

максимум линейной функции $7.625x_1 - 0.45u_1$ достигается в точке $u_1 = 0$ – левом конце отрезка $[0, x_1]$, $v_1 = x_1 - u_1 = x_1 = a = 1000$.

Максимальный доход за 4 года составит $w_1(x_1) = 7.625\cdot 1000 = 7625$ единиц. Для получения такого дохода надо в первый и второй годы все средства отдать предприятию $Q$ ($u_1 = 0, v_1 = x_1,\quad u_2 = 0, v_2 = x_2$); на третий год оставшиеся средства можно распределить между предприятиями $P$ и $Q$ произвольным образом ($0 \le u_3 \le x_3, v_3 = x_3 - u_3$); и, наконец, на четвертый год все оставшиеся средства надо передать предприятию $P$ ($u_4 = x_4, v_4 = 0$). Остаток средств, однако, при этом будет зависеть от распределения средств в третьем году. Действительно, $x_1 = a = 1000,\quad x_2 = 0.1\cdot(5x_1 - 2u_1) = 500,\quad a_3 = 0.1\cdot(5a_2 - 2x_2) = 250,$ $x_4 = 0.1\cdot(5x_3 - 2u_3) = 125 - 0.2u_3$; остаток же после четырех лет работы будет $x_5 = 0.3x_4 = 37.5 - 0.06u_3$. Величины $x_4$ и $x_5$ зависят от $u_3$. Например, если в третий год все оставшиеся средства передать предприятию $P$ ($u_3 = x_3 = 250, v_3 = 0$), то $x_5 = 37.5 - 0.06\cdot 250 = 22.5$ единиц, если же предприятию $Q$ ($u_3 = 0, v_3 = 250$), то $x_5 = 37.5$

единиц, а если поделить средства пополам ($u_3 = v_3 = 125$), то $x_5 = 37.5 - 0.06 \cdot 125 = 30$ единиц. Вообще в зависимости от $u_3$ величина $x_5$ будет меняться в пределах $22.5 \le x_5 \le 37.5$. Но в условии задачи нет требований к остатку. Речь идет только о максимальной прибыли, которая, независимо от остатка, останется равной 7625 единиц.

**Пример 4.3.** Общая сумма в 4 млн. у.е. распределяется между тремя предприятиями в количествах, кратных 1 млн. у.е. В результате выделения средств $k$-му предприятию в размере $u$ оно дает доход $f_k(u),\ k = 1,2,3,$ величина которого может быть найдена из табл. 4.1.

*Таблица 4.1*

**Доходы предприятий**

| *Доход* | *u, млн. у.е.* | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| $f_1(u)$ | 0 | 5 | 9 | 11 | 12 |
| $f_2(u)$ | 0 | 4 | 8 | 12 | 14 |
| $f_3(u)$ | 0 | 7 | 9 | 10 | 11 |

Требуется распределить средства между предприятиями так, чтобы их суммарный доход был максимальным.

Обозначив средства, выделенные $k$-му предприятию ($k = 1,2,3$), символом $u^{(k)}$, а сумму средств, выделенных предприятиям с номерами от 1 до $k$, символом $x^{(k)}$ сформулируем рассматриваемую задачу как многошаговую задачу (4.14) – (4.17):

$$F(\bar{x}, \bar{u}) = \sum_{k=1}^{3} f_k(u^{(k)}) \to \max,$$

$$x^{(k)} = x^{(k-1)} + u^{(k)},\ k = 1,2,3,$$

$$u^{(k)} \in [0; 4 - x^{(k-1)}] \cap \mathrm{N},\ k = 1,2,3,\quad \mathrm{N} - \text{ множество натуральных чисел,}$$

$$x^{(0)} = 0, \quad x^{(3)} = 4.$$

Для решения этой задачи применим метод динамического программирования.

Этап I (условная оптимизация).

Шаг 1. Найдем $w_3(x^{(2)}) = \max\limits_{u^{(3)} \in [0;4-x^{(2)}] \cap \mathrm{N}} f_3(u^{(3)}).$ Так как функция $f_3(u^{(3)})$ является возрастающей функцией аргумента $u^{(3)}$ (см. табл. 4.1), то ее максимум достигается при максимальном допустимом значении $u^{(3)}$, т.е. $u^{(3)*}(x^{(2)}) = 4 - x^{(2)}$. Отсюда $w_3(x^{(2)}) = f_3(4 - x^{(2)}).$ Значения $w_3(x^{(2)}),$ найденные с помощью табл. 4.1, представлены в табл. 4.2.

*Таблица 4.2*

**Зависимость максимального дохода третьего предприятия от выделенных средств второму предприятию**

| $x^{(2)}$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $w_3(x^{(2)})$ | 11 | 10 | 9 | 7 | 0 |

Шаг 2. Вычислим $w_2(x^{(1)}) = \max\limits_{u^{(2)} \in [0;4-x^{(1)}] \cap N} (f_2(u^{(2)}) + w_3(x^{(1)} + u^{(2)}))$. Для нахождения максимума функции $f_2(u^{(2)}) + w_3(x^{(1)} + u^{(2)})$ составляем табл. 4.3 значений этой функции, используя данные табл. 4.1 и 4.2.

*Таблица 4.3*

**Суммарный доход первого и второго предприятия**

| $x^{(1)}$ | $u^{(2)}$ | | | | |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 |
| 0 | 11 | 14 | 17 | (19) | 14 |
| 1 | 10 | 13 | (15) | 12 | - |
| 2 | 9 | (11) | 8 | - | - |
| 3 | (7) | 4 | - | - | - |
| 4 | (0) | - | - | - | - |

В табл. 4.3 скобками выделены максимальные по $u^{(2)}$ значения функции $w_2(x^{(1)})$ соответствующие различным значениям $x^{(1)}$. Используя табл. 4.3, находим функции $w_2(x^{(1)})$ и $u^{(2)*}(x^{(1)})$ представив их значения в табл. 4.4 и 4.5.

*Таблица 4.4*

**Зависимость максимального суммарного дохода первого и второго предприятия от выделенных средств первому предприятию**

| $x^{(1)}$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $w_2(x^{(1)})$ | 19 | 15 | 11 | 7 | 0 |

*Таблица 4.5*

**Зависимость условно-оптимального размера выделенных средств второму предприятию от выделенных средств первому предприятию**

| $x^{(1)}$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $u^{(2)*}(x^{(1)})$ | 3 | 2 | 1 | 0 | 0 |

Шаг 3. Так как множество $X_0$ состоит из единственной точки $x^{(0)} = 0$, то найдем только $w_1(0)$ и $u^{(1)*}(0)$: $w_1(0) = \max\limits_{u^{(1)} \in [0;4] \cap N} (f_1(u^{(1)}) + w_2(0 + u^{(1)}))$. Для определения максимума в правой части последнего равенства составим табл. 4.6 значений функции $f_1(u^{(1)}) + w_2(u^{(1)})$ которые найдем с помощью табл. 4.1 и 4.4.

*Таблица 4.6*

**Суммарный доход предприятий**

| $u^{(1)}$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $w_1(0)$ | 19 | (20) | (20) | 19 | 12 |

Из табл. 4.6 видно, что $w_1(0) = 20$, причем $u^{(1)*}(0) = 1$ или $u^{(1)*}(0) = 2,$ т.е. в данной задаче существует два оптимальных управления и две оптимальные траектории.

Этап II (безусловная оптимизация).

Шаг 1.

a) Пусть $u^{(1)*} = 1$, тогда $x^{(1)*} = x^{(0)*} + u^{(1)*} = 1.$

b) Пусть $u^{(1)*} = 2$, тогда $x^{(1)*} = x^{(0)*} + u^{(1)*} = 2.$

Шаг 2.

a) Для $x^{(1)*} = 1$ имеем $u^{(2)*} = u^{(2)*}(1) = 2$ (см. табл. 4.5), $x^{(2)*} = x^{(1)*} + u^{(2)*} = 3$.

b) Для $x^{(1)*} = 2$ получаем $u^{(2)*} = u^{(2)*}(2) = 1$ (см. табл. 4.5), $x^{(2)*} = x^{(1)*} + u^{(2)*} = 3.$

Шаг 3. Так как для обеих оптимальных фазовых траекторий $x^{(2)*} = 3$, то находим $u^{(3)*} = u^{(3)*}(3) = 1,$ $x^{(3)*} = x^{(2)*} + u^{(3)*} = 4$.

Окончательно получаем $\bar{u}^* = \{1, 2, 1\}$ или $\bar{u}^* = \{2, 1, 1\}$ и, соответственно, $\bar{x}^* = \{0, 1, 3, 4\}$ или $\bar{x}^* = \{0, 2, 3, 4\}$.

Таким образом, существует два оптимальных варианта распределения средств предприятиям:

1) первому предприятию выделяется 1 млн. у.е., второму – 2 млн. у.е. и третьему – 1 млн. у.е.;
2) первому – 2, второму – 1 и третьему – 1 млн. у.е.

В обоих случаях суммарный доход предприятий составит $w_1(0) = 20$ млн. у.е.

**Пример 4.4.** Для увеличения объемов выпуска пользующейся повышенным спросом продукции, изготовляемой предприятиями, выделены капиталовложения в объеме *S* = 700 тыс.у.е. Использование *i*-м предприятием $u_i$ тыс.у.е. из указанных средств обеспечивает прирост выпуска продукции, определяемый значением нелинейной функции $f_i(u_i)$. Найти распределение капиталовложений между предприятиями, обеспечивающее максимальное увеличение выпуска продукции, если значения $u_i$ и $f_i(u_i)$ приведены в табл. 4.7.

*Таблица 4.7*

**Зависимость прироста выпуска продукции от объема капитальвложений**

| Объем капиталовложений $u_i$ (тыс.у.е.) | Прирост выпуска продукции $f_i(u_i)$ в зависимости от объема капиталовложений (тыс.у.е.) | | |
|---|---|---|---|
| | Предприятие 1 | Предприятие 2 | Предприятие 3 |
| 0 | 0 | 0 | 0 |
| 100 | 30 | 50 | 40 |
| 200 | 50 | 80 | 50 |
| 300 | 90 | 90 | 110 |
| 400 | 110 | 150 | 120 |
| 500 | 170 | 190 | 180 |
| 600 | 180 | 210 | 220 |
| 700 | 210 | 220 | 240 |

Для решения данной задачи следует составить рекуррентное соотношение Беллмана. В рассматриваемом случае это соотношение приводит к следующим функциональным уравнениям:

$$
\begin{aligned}
&w_1(x) = \max_{0 \le u_1 \le x} f_1(u_1); \\
&w_2(x) = \max_{0 \le u_2 \le x} (f_2(u_2) + w_1(x - u_2); \\
&\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
&w_n(x) = \max_{0 \le u_n \le x} (f_n(u_n) + w_{n-1}(x - u_n)).
\end{aligned} \tag{4.26}
$$

Здесь функции $w_i(x)\,(i = 1,...,n)$ определяют максимальный прирост выпуска продукции $i$-го предприятия при соответствующих распределениях $x$ тыс.у.е. капиталовложений между предприятиями. Поэтому значение функции $w_n(x)$ вычисляется лишь для одного значения $x$=$S$, так как объем капиталовложений, выделяемых для всех $n$ предприятий, равен $S$ тыс.у.е.

Используя рекуррентное соотношение (4.26) и исходные данные табл. 4.7, решим задачу, т.е. определим сначала условно-оптимальные, а затем и оптимальные распределений капиталовложений между предприятиями. Начнем с определения условно оптимальных капиталовложений, выделяемых для развития первого предприятия. Для этого находим значения $w_1(x)$ для каждого $x$, принимающего значения 0, 100, 200, 300, 400, 500, 600 и 700.

Пусть $x = 0$, тогда $w_1(0) = 0$. Возьмем теперь $x$=100. Тогда, используя табл. 4.7, получаем

$$w_1(100) = \max\begin{Bmatrix} 0 \\ 30 \end{Bmatrix} = 30,\; u_1^* = 100.$$

Здесь первая строка соответствует решению $u_1 = 0$, а вторая строка – решению

$u_1 = 100$. Так как при первом решении прирост выпуска продукции не обеспечивается, а при втором равен 30 тыс.у.е., то условно-оптимальным решением является $u_1^0 = 100$.

Аналогично находим условно-оптимальные решения для других значений *x*. Результаты вычислений и полученные соответствующие условно оптимальные решения – в табл. 4.8.

*Таблица 4.8*

**Условно-оптимальные решения для первого предприятия**

| Объем капиталовложений *x*, выделяемых первому предприятию (тыс.у.е.) | Максимальный прирост $w_1(x)$ выпуска продукции (тыс.у.е.) | Условно оптимальный объем капиталовложений $u_1^*$, выделяемых первому предприятию (тыс.у.е.) |
|---|---|---|
| 0 | 0 | 0 |
| 100 | 30 | 100 |
| 200 | 50 | 200 |
| 300 | 90 | 300 |
| 400 | 110 | 400 |
| 500 | 170 | 500 |
| 600 | 180 | 600 |
| 700 | 210 | 700 |

Используя теперь данные табл. 4.7 и 4.8, определим условно оптимальные объемы капиталовложений, выделяемых второму предприятию. Найдем $w_2(x) = \max\limits_{0 \le u_2 \le x} (f_2(u_2) + w_1(x - u_2))$ для каждого из допустимых значений *x*, равных 0, 100, 200, 300, 400, 500, 600, 700: $w_2(0) = 0,\ u_2^* = 0;$

$$w_2(100) = \max \begin{Bmatrix} 0+30 \\ 50+0 \end{Bmatrix} = 50, \quad u_2^* = 100; \qquad w_2(200) = \max \begin{Bmatrix} 0+50 \\ 50+30 \\ 80+0 \end{Bmatrix} = 80, \quad u_2^* = 100;$$

$$w_2(300) = \max \begin{Bmatrix} 0+90 \\ 50+50 \\ 80+30 \\ 90+0 \end{Bmatrix} = 110, \quad u_2^* = 200; \qquad w_2(400) = \max \begin{Bmatrix} 0+110 \\ 50+90 \\ 80+50 \\ 90+30 \\ 150+0 \end{Bmatrix} = 150, \quad u_2^* = 400;$$

$$w_2(500) = \max \begin{Bmatrix} 0+170 \\ 50+110 \\ 80+90 \\ 90+50 \\ 150+30 \\ 190+0 \end{Bmatrix} = 190, \quad u_2^* = 500;$$

$$w_2(600) = \max \begin{Bmatrix} 0+180 \\ 50+170 \\ 80+110 \\ 90+90 \\ 150+50 \\ 190+30 \\ 210+0 \end{Bmatrix} = 220, \quad u_2^* = 100; \quad w_2(700) = \max \begin{Bmatrix} 0+210 \\ 50+180 \\ 80+170 \\ 90+110 \\ 150+90 \\ 190+50 \\ 210+30 \\ 220+0 \end{Bmatrix} = 250, \quad u_2^* = 200.$$

Полученные результаты и найденные условно оптимальные объемы капиталовложений, выделяемых второму предприятию, записываем в табл. 4.9.

*Таблица 4.9*

**Условно-оптимальные решения для второго предприятия**

| Объем капиталовложений $x$, выделяемых двум предприятиям (*тыс.у.е.*) | Максимальный прирост $w_2(x)$ выпуска продукции (*тыс.у.е.*) | Условно оптимальный объем капиталовложений $u_2^*$, выделяемых второму предприятию (*тыс.у.е.*) |
|---|---|---|
| 0 | 0 | 0 |
| 100 | 50 | 100 |
| 200 | 80 | 100 |
| 300 | 110 | 200 |
| 400 | 150 | 400 |
| 500 | 190 | 500 |
| 600 | 220 | 100 |
| 700 | 250 | 200 |

Переходим теперь к нахождению значений $w_3(x) = \max\limits_{0 \le u_3 \le x} (f_3(u_3) + w_2(x - u_3))$, используя для этого данные табл. 4.9 и 4.7. Так как в данном случае число предприятий равно трем, то проводим вычисление лишь для одного значения $x$=700:

$$w_3(700) = \max \begin{Bmatrix} 0+250 \\ 40+220 \\ 50+190 \\ 110+150 \\ 120+110 \\ 180+80 \\ 220+50 \\ 240+0 \end{Bmatrix} = 270, \quad u_3^* = 600.$$

Следовательно, максимальный прирост выпуска продукции составляет 270 тыс.у.е. Это имеет место тогда, когда третьему предприятию выделяется 600 тыс.у.е., а первому и второму предприятиям – 100 тыс.у.е. Итак, мы получили оптимальный план распределения капиталовложений между предприятиями, согласно которому обеспечивается максимальный прирост выпуска продукции.

## 4.4. Применение метода динамического программирования в сетевых задачах

При организации работы над различными проектами и выполнении комплексов операций широкое распространение получили методы сетевого планирования и управления. Начальная информация о проекте задается перечнем работ, их продолжительностью или стоимостью, последовательностью выполнения (для каждой работы должно быть указано, каким работам она предшествует и на какую опирается).

Реализацию проекта удобно представлять в виде **графа**, который задается множеством **вершин** или **узлов** и множеством **дуг** или **ребер**. Каждая дуга задается парой вершин. **Сетью** называется граф, каждой дуге которого поставлено в соответствие некоторое неотрицательное число. **Сетевым графиком проекта** называют наглядное представление проекта в виде сети. Дуги этой сети являются ориентированными, т. е. для каждой указаны начало и конец, и представляют **работы**; вершины (их называют **событиями**) представляют последовательность выполнения работ. Ни одна работа, представленная дугой, выходящей из некоторой вершины, не может начинаться прежде завершения всех работ, представленных дугами, входящими в данную вершину. Каждую пару событий должна соединять не более чем одна работа. В сетевом графике должно быть одно начальное и одно конечное событие. Для выполнения перечисленных требований допустимо введение **фиктивных работ** (работ нулевой продолжительности, не требующих затраты каких-либо ресурсов). Начальному событию присваивается номер 1. Чтобы пронумеровать остальные события нужно для каждого события найти величину максимального числа дуг в путях, соединяющих данное событие с начальным. События нумеруются по возрастанию этой величины. Если для некоторого числа событий указанная величина одинакова, то эти события нумеруются в произвольном порядке. После построения графа можно вычислить необходимые затраты времени (или ресурсов) на реализацию проекта. При этом удобно использовать метод динамического программирования, начиная с начального состояния и последовательно вычисляя наиболее ранние сроки наступления событий, вплоть до конечного. Цепочка событий, т.е последовательность работ (дуг), сдвиг начала которых приводит к увеличению общего времени выполнения проекта, называется **критическим путем**.

**Пример 4.5**. Перечень работ по строительству коттеджа и их продолжительности приведены в табл. 4.10. Построить сетевой график выполнения комплекса работ и определить время реализации проекта.

Приступая к построению сетевого графика (рис. 4.2), замечаем, что работа $a_1$ не опирается ни на какую работу, поэтому она изобразится дугой, выходящей из события 1, означающее исходный момент, с которого начинается выполнение рассматриваемого комплекса работ. На работу $a_1$ опираются работы $a_2, a_3, a_4$ поэтому дуги, соответствующие этим работам на сетевом графике будут следовать непосредственно за дугой $a_1$, от события 2, означающего момент окончания работы $a_1$, и начало работ $a_2, a_3, a_4$. На работу $a_4$ опирается работа $a_5$, а на $a_3$ и $a_5$ – работа $a_6$ , что и отражено на сетевом графике следующими друг за другом соответствующими дугами. Работа $a_7$ опирается на работы $a_2$ и $a_6$. поэтому дуга $a_7$ исходит из события номер 5, означающего момент, к которому завершены обе эти работы. Дуга $a_7$ соответствует последней работе, а конечное ее событие номер 6 означает момент завершения работ всего рассматриваемого комплекса.

*Таблица 4.10*

**Перечень работ по строительству коттеджа и их продолжительности**

| Работа | Содержание работы | Опирается на работу | Продолжи-тельность |
|---|---|---|---|
| $a_1$ | Подготовка проектной документации | | 6 |
| $a_2$ | Подводка коммуникаций | $a_1$ | 4 |
| $a_3$ | Возведение фундамента | $a_1$ | 3 |
| $a_4$ | Заводское изготовление строительного комплекта | $a_1$ | 5 |
| $a_5$ | Доставка комплекта | $a_4$ | 1 |
| $a_6$ | Сборка коттеджа | $a_3$, $a_5$ | 4 |
| $a_7$ | Регистрация собственности | $a_2$, $a_6$ | 5 |

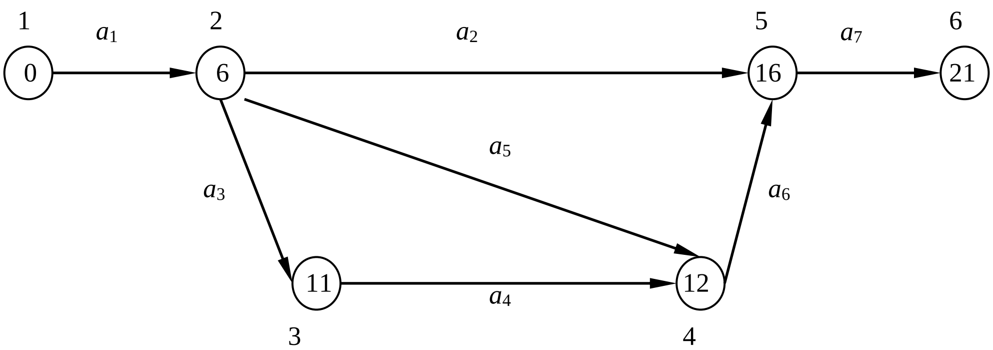


Рис. 4.2.

Вычислим методом динамического программирования ранние времена наступления событий $t_p(k)$, проставляя их последовательно на рис. 4.2 в кружках-вершинах графа.

1) $t_p(1)=0$,

2) $t_p(2)=t(a_1)=6$,

3) $t_p(3)=t(a_4)+6=11$,

4) $t_p(4)=\max(t(a_3)+6,\ t(a_5)+11)=12$,

5) $t_p(5)=\max(t(a_2)+6,\ t(a_6)+12)=16$,

6) $t_p(6)=t(a_7)+16=21$.

Учитывая, что максимум при вычислении $t_p(4)$ реализуется на дуге $a_5$, а максимум при вычислении $t_p(5)$ реализуется на дуге $a_6$, получаем, что критический путь образуют работы $a_1$, $a_4$, $a_5$, $a_6$, $a_7$. Если эти работы начнутся не сразу, например, работа $a_6$ начнется позже 12 дня, то время реализации проекта увеличится (станет больше 21 дня). Сдвиг же начала работ $a_2$ и $a_3$ допустим, они имеют резервы времени, соответственно, 6 и 3.

В данном примере в качестве функции Беллмана выступает $t_p(k)$, управления в явном виде нет, но расчеты позволяют выбрать время начала каждой работы и проекта в целом.

Другой широкий класс сетевых задач связан с транспортными потоками. Рассмотрим пример решение сетевой задачи о выборе кратчайшего маршрута методом динамического программирования.

**Пример 4.6**. Требуется перевезти груз из начального пункта в конечный пункт. На рис. 4.3 показана сеть дорог и стоимости перевозки единицы груза между отдельными пунктами сети (поставлены у соответствующих ребер). Необходимо определить маршрут доставки груза из пункта 1 в пункт 14, которому соответствуют наименьшие затраты.

Событиями здесь являются доставки груза в пункты, номера которых записаны в кружках (рис. 4.3), а работами – перевозки между отдельными пунктами.

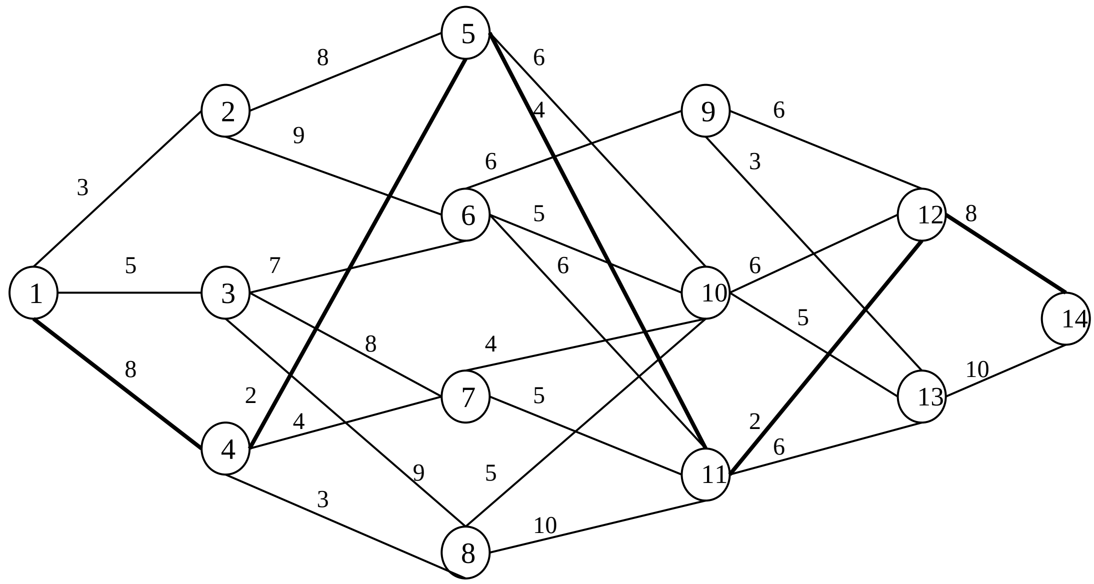

Рис. 4.3.

Для того, чтобы решить задачу методом динамического программирования, нужно весь процесс доставки груза из пункта 1 в пункт 14 разбить на этапы. На первом этапе транспорт с грузом из пункта 1 перемещается непосредственно в пункты 2, 3, 4; на втором этапе из пунктов 2, 3, 4 - в пункты 5, 6, 7, 8; на третьем этапе из пунктов 5, 6, 7, 8 – в пункты 9, 10, 11; на четвертом этапе из пунктов 9, 10, 11 – в пункты 12, 13 и на последнем этапе из пунктов 12, 13 – в пункт 14.

Введем обозначения: $k$ – номер этапа, $k=1,...,5$; $x_{k-1}$ – состояние (пункт) на $k$-м этапе; $u_{k-1}(x_{k-1})=(i,j)$ – выбор пути на $k$-м этапе; $f_{k-1}(x_{k-1},u_{k-1})=c_{i,j}$ – стоимость перевозки груза из пункта $i$ в пункт $j$; $w_{k-1}(x_{k-1})$ – минимальные затраты не перевозку груза из некоторого пункта на $k$-м этапе в последний (функция Беллмана).

Запишем функциональное уравнение для 5-го этапа:

$$w_4(x_4)=\min_{u_4}[f_4(x_4,u_4)+f_5(x_5)].$$

В пункт 14 груз может быть доставлен или из пункта 12 или из пункта 13, поэтому

$$w_4(12)=\min(c_{12,14}+c_{14,14})=8+0=8,\quad u_4^*(12)=(12,14) \text{ и}$$

$$w_4(13)=\min(c_{13,14}+c_{14,14})=10+0=10,\quad u_4^*(13)=(13,14).$$

Для 4-го этапа функциональное уравнение имеет вид

$$w_3(x_3)=\min_{u_3}\left[f_3(x_3,u_3)+w_4(x_4)\right].$$

Пусть груз на 4-м этапе оказался в пункте 9. Из этого пункта далее можно следовать либо через пункт 12, либо через пункт 13. Тогда

$w_3(9)=\min(c_{9,12}+w_4(12);\, c_{9,13}+w_4(13))=\min(6+8;\, 3+10)=13,\quad u_3^*(9)=(9,13,14)$, при этом минимальные затраты составляют 13.

Если груз оказался в пункте 10, то

$$w_3(10)=\min(c_{10,12}+w_4(12);\, c_{10,13}+w_4(13))=\min(6+8;\, 5+10)=14,$$

$u_3^*(10)=(10,12,14)$, при этом минимальные затраты составляют 14.

Для 11-го пункта получаем

$$w_3(11)=\min(c_{11,12}+w_4(12);\, c_{11,13}+w_4(13))=\min(2+8;\, 6+10)=10,$$

$u_3^*(11)=(11,12,14)$, при этом минимальные затраты составляют 10.

Переходим к 3-му этапу.

$$w_2(x_2)=\min_{u_2}\left[f_2(x_2,u_2)+w_3(x_3)\right].$$

$$w_2(5)=\min(c_{5,10}+w_3(10);\, c_{5,11}+w_3(11))=\min(6+14;\, 4+10)=14,$$

$u_2^*(5)=(5,11,12,14)$, при этом минимальные затраты составляют 14.

$$w_2(6)=\min(c_{6,9}+w_3(9);\, c_{6,10}+w_3(10);\, c_{6,11}+w_3(11))=\min(6+13;\, 5+14;\, 6+10)=16,$$

$u_2^*(6)=(6,11,12,14)$, при этом минимальные затраты составляют 16.

$$w_2(7)=\min(c_{7,10}+w_3(10);\, c_{7,11}+w_3(11))=\min(4+14;\, 5+10)=15,$$

$u_2^*(7) = (7,11,12,14)$, при этом минимальные затраты составляют 15.

$w_2(8) = \min(\ c_{8,10} + w_3(10);\ c_{8,11} + w_3(11)) = \min(\ 5+14;\ 10+10) = 19,$

$u_2^*(8) = (8,10,12,14)$, при этом минимальные затраты составляют 19.

Рассматриваем 2-й этап.

$$w_1(x_1) = \min_{u_1} \left[ f_1(x_1, u_1) + w_2(x_2) \right].$$

$w_1(2) = \min(\ c_{2,5} + w_2(5);\ c_{2,6} + w_2(6)) = \min(\ 8+14;\ 9+16) = 22,$

$u_1^*(2) = (2,5,11,12,14)$, при этом минимальные затраты составляют 22.

$w_1(3) = \min(\ c_{3,6} + w_2(6);\ c_{3,7} + w_2(7);\ c_{3,8} + w_2(8)) = \min(\ 7+16;\ 8+15;\ 2+19) = 21,$

$u_1^*(3) = (3,8,10,12,14)$, при этом минимальные затраты составляют 21.

$w_1(4) = \min(\ c_{4,5} + w_2(5);\ c_{4,7} + w_2(7);\ c_{4,8} + w_2(8)) = \min(\ 2+14;\ 4+15;\ 3+19) = 16,$

$u_1^*(4) = (4,5,11,12,14)$, при этом минимальные затраты составляют 16.

Наконец, переходим к 1-му этапу. На 1-м этапе состояние известно – это пункт 1. Поэтому, подставляя это начальное состояние в функцию $w_0(x_0)$, сразу получаем значение динамической задачи $w_0(1)$ и оптимальное управление $u_0^0(1)$.

$$w_0(x_0) = \min_{u_0} \left[ f_0(x_0, u_0) + w_1(x_1) \right].$$

$w_0(1) = \min(\ c_{1,2} + w_1(2);\ c_{1,3} + w_1(3);\ c_{1,4} + w_1(4)) = \min(\ 3+22;\ 5+21;\ 8+16) = 24,$

$u_0^0(1) = (1,4,5,11,12,14)$. Минимальные затраты при перевозке груза из 1 в 14 составляют 24 единицы, оптимальный маршрут 1-4-5-11-12-14.

### 4.5. Модели управления запасами

Предприятия, фирмы имеют различные запасы: сырье, комплектующие изделия, готовую продукцию, предназначенную для продажи, и т.д. Совокупность подобных материалов, представляющих временно не используемые экономические ресурсы, называют **запасами предприятия**. Запасы создаются по различным причинам. Одна из них состоит в том, что если в некоторый момент производства потребуется какой-то вид деталей, который поставляется другим предприятием, и он отсутствует на складе, то процесс производства может остановиться. Поэтому на складе всегда должно быть нужное количество деталей данного вида. Однако если запасы увеличить, то возрастет стоимость их хранения. Задача управления запасами состоит в выборе для предприятия целесообразного решения. Рассмотрим простейшие математические модели управления запасами $Q$.

Под $Q$ будем понимать изделия или материалы (товары) только одного вида. Если на изделие поступает заявка, то оно отпускается, и значение $Q$ падает. Предположим, что величина спроса непрерывна во времени. Если $Q$=0, то имеет место дефицит.

Любая математическая модель, которая применяется для изучения определенной ситуации в управлении запасами, должна учитывать факторы, связанные с издержками.

Различают **организационные издержки** – расходы, связанные с оформлением и доставкой товаров, издержки **содержания запасов** – затраты, связанные с хранением. Они возникают из-за амортизации в процессе хранения (изделия могут портиться, устаревать, их количество может уменьшаться и т.д.). Существуют издержки, связанные с **дефицитом**: если поставка со склада не может быть выполнена, то возникают дополнительные издержки, связанные с отказом. Это может быть денежный штраф или другой ущерб (например, ухудшение бизнеса в будущем и потеря потребителей). Количество товара, поставляемое на склад, называют **размером партии**. Введем обозначения величин, необходимых для составления модели (см. табл. 4.11).

*Таблица 4.11*

**Обозначения**

| Величина, обозначение | Единица измерения | Предположения |
|---|---|---|
| Интенсивность спроса, $g$ | Единиц товара в год | Спрос постоянен и непрерывен; весь спрос удовлетворяется |
| Организационные издержки, $b$ | Рублей за партию | Издержки постоянны, не зависят от размера партии |
| Стоимость товара, $s$ | Рублей за год | Цена единицы товара постоянна; рассматривается один вид товара |
| Издержки содержания запасов, $h$ | Рублей за единицу товара в год | Стоимость хранения единицы товара в течение года постоянна |
| Размер партии, $q$ | Единиц товара в одной партии | Размер партии постоянен; поступление товара происходит мгновенно, как только уровень запаса равен 0 |

Рассматриваемые далее задачи управления запасов относятся к динамическим, однако в силу периодичности и стационарности процессов могут быть решены без использования метода динамического программирования. Предполагается, что поставки производятся равными партиями, мгновенно. График изменения запасов представлен на рис. 4.4.

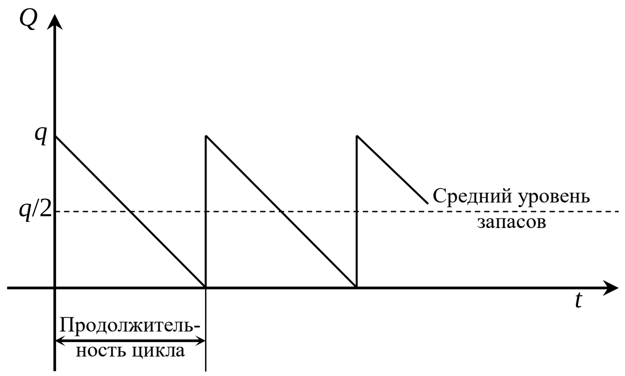


Рис. 4.4.

Чтобы полностью удовлетворить годовой спрос $g$ при размере поставки $q$, необходимо обеспечить $g/q$ поставок или партий за год. Средний уровень запасов составляет $q/2$. Уравнение издержек будет иметь вид $C = C_1 + C_2 + C_3 = bg/q + sg + hq/2$, где $C_1$ – общие организационные издержки; $C_2$ – стоимость товаров; $C_3$ – общие издержки содержания запасов. За исключением $q$ все величины в правой части уравнения постоянны и известны, т.е. $C = f(q)$. Для нахождения минимума $C$ найдем производную этой функции и приравняем ее к нулю, откуда получаем оптимальный размер партии $q_{opt}=\sqrt{2bg/h}$.

Иногда приходится заказывать размер партии товаров, не соответствующий оптимальному. Это приводит к увеличению издержек на содержание и организацию поставок. Действительно, предположим, что вместо оптимального размера была заказана партия товаров, равная $0.5q_{opt}$. Из основного уравнения издержек $C = C_1+ C_2 + C_3 = bg/q + sg + hq/2$ найдем при $q = q_{opt}$

$$C - C_2 = C - sg = bg\sqrt{h/2bg} + h\sqrt{2bg/4h} = \sqrt{2bgh}.$$

В случае заказа $0.5q_{opt}$ получим

$$C - C_2 = 2bg\sqrt{h/2bg} + h\sqrt{2bg/16h} = \sqrt{2bgh} + \sqrt{2bgh}/4 = 5\sqrt{2bgh}/4 = 1.25(C - C_2).$$

Таким образом, заказ партии товаров размером $0.5q_{opt}$ (вместо $q_{opt}$) приводит к увеличению общих издержек на содержание запасов и организацию поставок на 25%. Аналогичная картина наблюдается в случае заказа поставок больше чем $q_{opt}$. Изобразим графически изменение отдельных составляющих величины $C$.

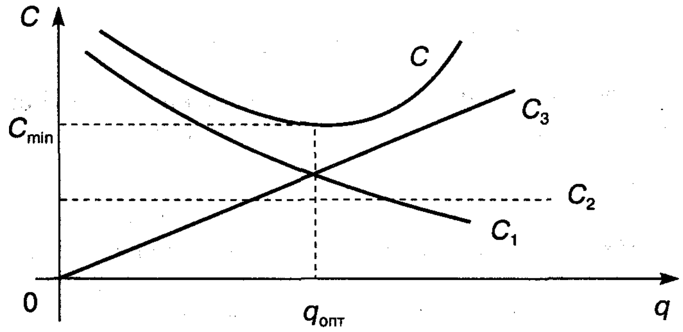


Рис. 4.5.

Из рис. 4.5 следует, что увеличение $q$ ведет к резкому снижению $C_1$, при этом $C_3$ увеличивается пропорционально $h/2$. При малых значениях $q$ величина $C$ падает до значения $C_{min}$ в точке $q_{opt.}$ При увеличении $q$ величина издержек $C$ приближается к $C_2+C_3$.

**Пример 4.5.** Интенсивность равномерного спроса составляет 2000 пылесосов в год. Организационные издержки для одной партии составляют 20 тыс.у.е. Цена единицы товара равна 3 тыс.у.е., а издержки содержания пылесосов составляют 0.1 тыс.у.е. за один пылесос в год. Найти оптимальный размер партии и продолжительность цикла.

По условию задачи $g = 2000$, $b = 20$, $s = 3$, $h = 0{,}1$. Общие издержки в течение года:

$$C = C_1 + C_2 + C_3 = 40000/q + 6000 + q/20, \qquad q_{opt} = \sqrt{800000} \approx 894 \text{ ед.},$$

Оптимальный размер партии – 894 пылесоса, продолжительность цикла – 163 дня.

В рассмотренной основной модели предполагается, что поступление товаров на склад происходит мгновенно, например, в течение одного дня. Рассмотрим случай, когда готовые товары поступают на склад непосредственно с производственной линии. Будем считать, что поступление партии товаров происходит непрерывно в течение некоторого времени с постоянной скоростью. Модель задачи в этом случае называют моделью **производственных поставок**. Обозначим через $p$ скорость поступающего на склад товара. Остальные обозначения и предположения те же, что и для основной модели управления запасами.

Определим оптимальный размер партии, минимизирующий общие затраты. График изменения модели производственных запасов представлен на рис. 4.6.

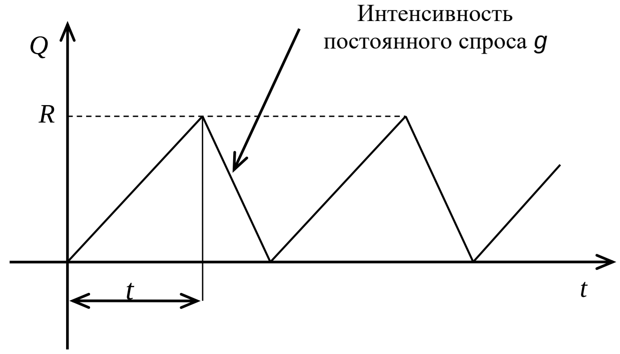


Рис. 4.6.

Общие издержки в течение года, как и для основной модели, составляют

$$C = C_1 + C_2 + C_3,$$

где $C_1 = bg/q$, $C_2 = sg$.

Для получения среднего уровня запасов следует учесть, что $R=(p–g)t$ – максимальный уровень запасов, где время поставки $t$ определяется из $q=pt$ – количество товаров в одной производственной поставке. Тогда средний уровень запасов составляет половину максимального и равен $(p–g)q/2p$. Поэтому $C_3=(p–g)hq/2p$. В итоге

$$C = bg/q + sg + hq(p - g)/2p.$$

Решая уравнение $dC/dq = 0$, найдем оптимальный размер партии в модели производственных поставок:

$$q_{\text{opt}} = \sqrt{2pbg \big/ (p-g)h}\,.$$

Рассмотрим модель, допускающую возможность существования периодов дефицита, который покрывается при последующих поставках, и штрафов за несвоевременную поставку. Пусть предприятие должно поставить $q$ ед. товара в течение каждого промежутка времени $L$, за единицу времени поставляется $g$ ед. товара ($q = Lg$). Предположим, что в начале каждого периода $L$ предприятие делает запас, равный $k$. Это означает, что в течение периода будет наблюдаться дефицит товара и некоторое время поставки не будут осуществляться. Невыполненные заявки будут накапливаться до максимальной величины $(q – k)$ и будут удовлетворены, как только поступит следующая партия товаров в количестве $q$. За то, что товары доставляются предприятием позже необходимого срока, на предприятие налагается штраф, который зависит от задержки поставки (иногда выгоднее заплатить штраф, чем расходовать дополнительные средства на хранение запасов). Задача управления состоит в том, чтобы выбрать такое значение $k$, которое ведет к минимизации всех затрат, включая затраты на хранение и штрафы. График изменения запасов модели представлен на рис. 4.7.

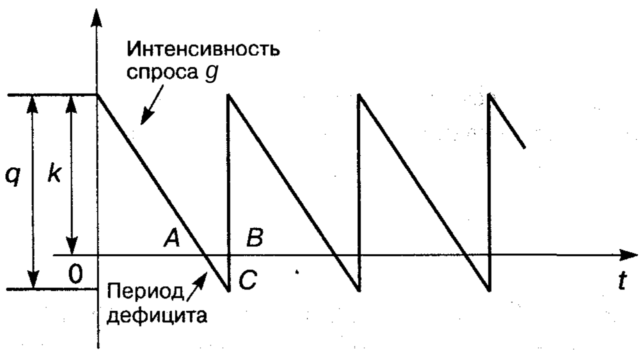


Рис. 4.7.

Пусть $d$ – штраф в расчете на единицу товара за один день просрочки. Издержки одного цикла определяются как $C=C_1+C_2$, где $C_1$ - общие издержки содержания запасов; $C_2$ - общие затраты на выплату штрафов. Так как товары находятся на складе в течение периода $OA$ (см. рис. 4.7), средний уровень запасов за этот период равен $k/2$. Если продолжительность периода $OA$ равна $k/g$, то $C_1 = h(k/2)k/g = hk^2/2g$.

Так как штраф выплачивается в течение периода $AB = (q–k)/g$, общее число "товаро-дней", на которые налагается штраф, равно площади треугольника $ABC$. Площадь составляет $(q – k)/g\cdot(q – k)/2$, откуда $C_2 = d(q – k)^2/2g$. Окончательно получаем $C= hk^2/2g + d\,(q–k)^2/2g$. Найдем $dC/dk$ и, решив уравнение $dC/dk = 0$, получим

$$k_{\text{opt}}= dq/(h+d).$$

Взяв $k_{\text{opt}}$ в качестве уровня запасов в начале каждого цикла при условии, что невыполненные заявки будут удовлетворены, сведем суммарные расходы к минимуму:

$$C_{min}=q^2hd/2g(h+ d).$$

## Задачи и упражнения

Сформулируйте задачи 1, 2 в виде многошаговых задач оптимизации.

**1.** Найдите $N$ неотрицательных чисел $u^{(k)}, k = 1,...N,$ сумма которых равна $S$, а произведение максимально.

**2.** Сумма средств $S$ выделяется предприятию в течение $N$ лет. Прибыль, получаемая предприятием в результате выделения ему средств $u$ в течение $k$-го года составляет $J_k(u), k = 1,..., N.$ Распределите выделяемые средства по годам таким образом, чтобы суммарная прибыль предприятия за $N$ лет была максимальной.

Задачи **3**, **4** решите с помощью уравнения Беллмана.

**3.** Найдите решение задачи **1**, если $N$=3, $S$=9.

**4.** Найдите решение задачи **2** об оптимальном распределении средств предприятию со следующими исходными данными: $S$=5 *млн.у.е.*, $N$=4. Средства распределяются в количествах, кратных 1 *млн.у.е.* Функции $J_k(u), k = 1,2,3,4,$ заданы следующей таблицей:

| $u$ (*млн.у.е.*) | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $J_1(u)$ | 0 | 1,5 | 2 | 3,5 | 5,5 | 9 |
| $J_2(u)$ | 0 | 3 | 4,5 | 5,5 | 6,5 | 7,5 |
| $J_3(u)$ | 0 | 4 | 5 | 5,5 | 6 | 9 |
| $J_4(u)$ | 0 | 2 | 3 | 4 | 6,5 | 8 |

**5.** Фирме по строительству судов требуется 20000 заклепок в год, расходуемых с постоянной интенсивностью. Организационные издержки составляют 0.5 тыс.р. за партию, цена заклепки - 20 р. Издержки на хранение одной заклепки оценены в 12.5% ее стоимости. Найдите оптимальный размер партии поставки, продолжительность цикла и оптимальное число поставок в год.

**6.** Система управления запасами некоторого товара подчиняется основной модели. Каждый год с постоянной интенсивностью спрос составляет 15000 ед. товара, издержки на организацию поставки составляют 10 ден.ед. на партию, цена единицы товара - 30 ден.ед., а издержки на ее хранение - 7,5 ден.ед. в год. Найдите оптимальный размер партии, число поставок, продолжительность цикла.

**7.** Интенсивность спроса в модели производственных поставок составляет четверть скорости производства, которая равна 20 тыс.ед. товара в год. Организационные издержки для одной партии равны 150 ден.ед., а издержки хранения единицы товара в течение года - 5 ден.ед. Определите оптимальный размер партии.

**8.** Решение задачи методом динамического программирования на некотором шаге при произвольном состоянии после предыдущих шагов дает

а) оптимальное управление на данном шаге

б) условно-оптимальное управление на данном шаге

в) последовательность условно-оптимальных управлений

**9.** Решение задачи методом динамического программирования на некотором шаге при произвольном состоянии после предыдущих шагов дает

а) оптимальный выигрыш на данном шаге

б) условно-оптимальный выигрыш на данном шаге

в) условно-оптимальный выигрыш за все предыдущие шаги

**10.** Решение задачи динамического программирования на первом шаге после завершения этапа условной оптимизации (обратной прогонки) дает

а) оптимальный выигрыш на данном шаге

б) условно-оптимальный выигрыш на данном шаге

в) оптимальный выигрыш в динамической задаче в целом

**11.** В динамической задаче последовательность фазовых состояний называется

а) траекторией процесса

б) оптимальным управлением

в) оптимальным поведением

**12.** Эффективность в задаче динамического программирования характеризуется целевой функцией а) аддитивного типа б) линейного типа в) кусочно-линейного типа

**13.** Математическое рекуррентное выражение принципа оптимальности называется

а) уравнением состояний б) уравнением Беллмана в) решением динамической задачи

**14.** Начальными условиями называются ограничения на

а) начальное состояние процесса б) фазовую траекторию в) выбор управления

**15.** Фазовые ограничения накладываются на

а) начальное состояние процесса

б) всю траекторию процесса

в) фазовую траекторию и управление

## Заключение

Математическая модель задачи выбора решений в условиях полной информации представляет собой задачу оптимизации. В задачах оптимизации одним из самых важных является вопрос существования решения. В теореме Вейерштрасса сформулированы условия существования глобального экстремума (максимума или минимума). Необходимое условие экстремума для задачи безусловной оптимизации дает теорема Ферма. Достаточное условие локального максимума (минимума) содержит требование отрицательной (положительной) определенности матрицы вторых производных целевой функции.

При изучении нелинейных моделей принятия решений в технике, экономике, экологии, финансах рассмотрен случай ограничений типа неравенств, которые изучаются в рамках математического программирования. Условия Куна-Таккера являются необходимыми условиями оптимальности для общей задачи математического программирования. В отличие от общей задачи математического программирования в задаче выпуклого программирования функция Лагранжа всегда имеет седловую точку (при выполнении условия регулярности Слейтера) и условия Куна-Таккера являются необходимыми и достаточными условиями оптимальности.

Если целевая функция линейна и допустимое множество задано линейными ограничениями, то приходим к задаче линейного программирования. Если задача ЛП имеет решение, то существует вершина допустимого множества, которая является оптимальным планом. На этом утверждении основан симплекс-метод, или метод последовательного улучшения плана. Среди методов решения целочисленных задач ЛП можно выделить два направления: методы отсечения, среди которых наиболее известны алгоритмы Гомори, и комбинаторные методы, например, метод ветвей и границ.

В динамическом программировании рассматриваются методы, которые позволяют путем поэтапной оптимизации получить общий оптимум в задаче принятия решений. Одним из таких методов является метод динамического программирования, в основе которого лежит принцип оптимальности. Математическое выражение принципа оптимальности представляет рекуррентное соотношение, называемое уравнением Беллмана. Метод динамического программирования может быть применен к сетевым задачам, связанных, например, с транспортными потоками, построением сетевых графиков. Задачи управления запасов относятся к динамическим задачам, однако в силу периодичности и стационарности процессов эти задачи могут быть решены без использования метода динамического программирования.

## Ответы к задачам и упражнениям

### К главе 1:

**1.** $x_j^* = \dfrac{c_j^2 a}{\sum\limits_{j=1}^{n} c_j^2}$ .

**2.** cosα = 1/*k*, если $k \geq \dfrac{b}{\sqrt{b^2 - a^2}}$ .

**3.** Минимальное значение периметра равно $4\sqrt{S}$ , оно достигается при $x_1^* = x_2^* = \sqrt{S}$ .

**4.** $x_1^* = x_2^* = R/2$.

**5.** $x^*=(2, 3)$, $\alpha_0=0,5$.

**6.** $x_1^*=144.4$, $x_2^*=35,6$.

**7.** $x_1^*=2$, $x_2^*=3$.

**8**. а. **9.** б. **10**. б. **11**. в. **12.** в. **13.** а. **14.** а.

### К главе 2:

**1.** $f_{\max}=f\,(4;3)=24$.

**2.** $f_{\max}=f\,(6;4)=24$.

**3.** $f_{\min}= f\,(1,1)=2$, $f_{\max}=f\,(3,0)=9$.

**4.** $f_{\min}=f(121;79)=66385$.

**5.** $f_{\min}=f(0,0,0)=0$, $f_{\max}=f\,(15,0,0)= f\,(0,15,0)= f\,(0,0,15)=225$.

**6.** $f_{\max}=f\,(0.25, 0/75)=57/64$.

**7.** $f_{\max}=f(1,1.5)=1,5^{0.75}$.

**8.** $f_{\max}=f(3,6)= 10$.

**9.** $f_{\min}=f(3,2)= 0$, $f_{\max}=f(0,6)=25$.

**10.** $f_{\min}=f(2,3)= -117$.

**11.** $f_{\min}=f(1.5,2,-5)=-10,25$.

**14.** $f_{\min}=f(1.5,2.5)= -2,75$.

**15.** а. **16.** а. **17. б.**

### К главе 3:

**1.** $f_{\min} = 390$ , $x^0 = (10;0;30;10;50;0)$

**2.** $f_{\max} = 1747$ , $x_1^0 = 333, x_2^0 = 83$ .

**3.** $f_{\min} = 12\dfrac{2}{9}$, $x^0 = (0;10/3;8/9)$ .

**4.** $x^0 = (115;110;20)$ , $f^0 = 4020$ ден.ед.

**5.** $f_{\min} = 450$ , $x_{AC} = 0, x_{AD} = 6, x_{BC} = 3, x_{BD} = 1$ .

**6.** $x_{14}^0 = x_{22}^0 = 10, x_{23}^0 = x_{31}^0 = x_{34}^0 = 5, x_{33}^0 = 15,$ остальные $x_{ij}^0 = 0,\ \ f^0 = 1400.$

**7**. $x^0 = (1,5;1,75)$ , $f^0 = 1$ .

**8**. $x^0 = (3;4;0;0;14)$ , $f^0 = 18$ .

**9.** $x^0=(4;2)$, $f_{\max}=14$ ден.ед.

**10**. $y^0=(2,3)$, $x^0=(1/3;5/6;0)$.

**11.** $x^0=(5,6)$, $y^0=(4/9;7/9;0)$, $f^0=28$, $c_1\in[0,6;6]$, $c_2\in[1,10]$, $b_1\in[26;80]$, $b_2\in[7;17,8]$, $b_3\in[5;\infty)$. Сырье первого вида.

**12.** б. **13.** а. **14.** в.

## К главе 4:

**1.** $J(\bar{x},\bar{u})=\sum_{k=1}^{N}J_k(x^{(k-1)},u^{(k)})\to\max,$

$$x_1^{(k)}=x^{(k-1)}+u^{(k)}, x_2^{(k)}=x^{(k-1)}\cdot u^{(k)},\ k=1,...,N,$$
$$x_1^{(k)}\in[0,S],\ k=1,...,N-1, \qquad u^{(k)}\in[0,S],$$
$$x_1^{(0)}=0, x_2^{(0)}=1, x_1^{(N)}=S,$$

где $x_1^{(k)}=\sum_{i=1}^{k}u^{(i)}, x_2^{(k)}=\prod_{i=1}^{k}u^{(i)}; J_k(x^{(k-1)},u^{(k)})=0, k=\overline{1,N},$ $J_N(x^{(N-1)},u^{(N)})=x_2^{(N-1)}\cdot u^{(N)}.$

**2.** $J(\bar{x},\bar{u})=\sum_{k=1}^{N}J_k(u^{(k)})\to\max,$

$$x^{(k)}=x^{(k-1)}+u^{(k)},\ k=1,...,N,$$
$$x^{(k)}\in[0,S],\ k=1,...,N-1, \qquad u^{(k)}\in[0,S],\ k=1,...,N,$$
$$x^{(0)}=0, x^{(N)}=S,$$

где $u^{(k)}$ – средства, выделенные предприятию в $k$-м году, $x^{(k)}$ – средства, выделенные ему за первые $k$ лет.

**3.** $\bar{u}^*=\{3,3,3\}, J^*=27.$

**4.** $\bar{u}^*=\{1,2,1,1\}, J^*=12$.

**5.** $q_{opt}=4000$; $n_{opt}=5$; $t_{opt}=73$ (дн.).

**6.** $q_{opt}=200$; $n_{opt}=75$; $t_{opt}=4.87$ (дн.).

**7.** $q_{opt}\approx 632$ ед.

**8.** б. **9.** в. **10.** в. **11.** а. **12**. а. **13.** б. **14.** а. **15.** б.

## Список рекомендуемой литературы к части I

1. Александрова И.А., Гончаренко В.М., Денежкина И.Е. и др. Методы оптимальных решений в экономике и финансах. Под редакцией В.М. Гончаренко и В.Ю. Попова. – М.: Кнорус, 2013. – 400 с.
2. Белолипецкий А.А., Горелик В.А. Экономико-математические методы. Университетский учебник. – М.: Издательский центр «Академия», 2010. – 368 с.
3. Васильев Ф.П. Методы оптимизации (в 2-х книгах). Кн. 1 — М.: МЦНМО, 2011, 620 с.
4. Горелик В.А. Матричная коррекция задачи линейного программирования с несовместной системой ограничений // Журн. вычисл. матем. и матем. физ. – 2001. – Т. 41. № 11. – С. 1607 – 1705.
5. Горелик В.А. Исследование операций и методы оптимизации. – Университетский учебник. – М.: Издательский центр «Академия», 2013. – 272 с.
6. Горелик В.А., Ерохин В.И. Оптимальная матричная коррекция несовместных систем линейных алгебраических уравнений по минимуму евклидовой нормы. – М.: ВЦ РАН, 2004. – 193 с.
7. Колемаев В.А. Математическая экономика: учебник для вузов. – М.: ЮНИТИ, 2005. – 399 с.
8. Сухарев А.Г., Тимохов А.В., Федоров В.В. Курс методов оптимизации. – М. Физматлит, 2008. – 367 с.
9. Таха Х.А. Введение в исследование операций. – М.: «Вильямс», 2007. – 901 с.