Исследование операций в экономике

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ В ЭКОНОМИКЕ

Под редакцией профессора ***Н.Ш. Кремера***

Рекомендовано Министерством общего и профессионального образования Российской Федерации в качестве учебного пособия для студентов высших учебных заведений, обучающихся по экономическим специальностям

ЮНИТИ
UNITY
Москва • 2000

ББК 65в6я73
И87

Всероссийский заочный финансово-экономический институт
Ректор акад. *А.Н. Романов*
Председатель Научно-методического совета проф. *Д.М. Дайитбегов*

Коллектив авторов:
проф. *Н.Ш.Кремер* (предисловие, введение, гл. 1-3, 6, 8, 10-12,14-16), доц. *Б.А.Путко* (гл. 4, 13), доц. *И.М.Тришин* (гл. 7), доц. *М.Н.Фридман* (гл. 5,9). При подготовке гл. 10-12 использованы опубликованные материалы доц. *М.А. Войтенко* и доц. *Т.С. Фофановой.*

Рецензенты:
кафедра прикладной математики Московского государственного университета экономики, статистики и информатики
(зав. кафедрой проф. И.Н. Мастяева)
и д-р техн. наук проф. В.В. Подиновский

Главный редактор издательства *Н.Д. Эриашвили*

И87 **Исследование** операций в экономике: Учебн. пособие для вузов/ Н.Ш.Кремер, Б.А.Путко, И.М.Тришин, М.Н.Фридман; Под ред. проф. Н.Ш.Кремера. — М.: ЮНИТИ, 2000. — 407 с.
ISBN 5-85173-092-7

В учебном пособии представлены модели линейного и целочисленного программирования, классические методы оптимизации, задачи выпуклого и динамического программирования, модели управления запасами и сетевого планирования и управления, элементы теории игр и массового обслуживания. Рассмотрены некоторые вопросы применения ЭВМ для решения задач математического программирования. Приводится большое количество экономических задач с решениями и для самостоятельной работы.

Издательское объединение ЮНИТИ продолжает выпуск учебных пособий по математическим дисциплинам для студентов и абитуриентов экономических вузов. Вышли в свет "Математика для поступающих в экономические вузы"[1] и "Высшая математика для экономистов" [5].

Данное пособие написано в соответствии с требованиями государственных общеобразовательных стандартов в области математики для специалистов с высшим образованием по экономическим специальностям.

Исследование операций — комплексная научная дисциплина, имеющая важное методологическое значение в системе подготовки современного экономиста. В ней наиболее четко реализуется одна из основных идей изучения курса высшей математики в экономическом вузе — *идея математического моделирования экономических процессов.*

Круг проблем, изучаемых исследованием операций, еще недостаточно определен. В соответствии с государственными общеобразовательными стандартами для экономических специальностей комплекс вопросов, относящихся к исследованию операций в экономике, изучается в рамках математических дисциплин в отделе "Экономико-математические методы и модели". А примерная программа дисциплины "Математика" для экономических специальностей, утвержденная в 1996 г. Главным управлением образовательно-профессиональных программ и технологий Министерства общего и профессионального образования РФ, тот же комплекс вопросов относит к разделам "Исследование операций", "Методы оптимизации" и частично — к разделу "Основы дискретной математики и математической логики".

---

[1] Математика для поступающих в экономические вузы/Под. ред. Н.Ш.Кремера. — М.: Банки и биржи, ЮНИТИ, 1996.

Пособие содержит три раздела.

В разделе I (гл. 1—7) рассмотрены модели линейного программирования — постановка и примеры типовых задач, теоретические основы, теория двойственности, симплексный и распределительный методы решения задач. В гл. 8 представлены методы *целочисленного линейного программирования*, в частности, методы Гомори и ветвей и границ. Учитывая то, что к задачам линейного программирования могут быть сведены некоторые модели "конфликтных" ситуаций, авторы сочли возможным *элементы теории игр* (гл. 9) дать в этом же разделе.

Раздел II посвящен моделям нелинейного программирования. В гл. 10 описываются классические методы оптимизации: методы нахождения условного экстремума функции нескольких переменных и, в частности, метод множителей Лагранжа. В гл. 11 приводятся модели *выпуклого программирования* с использованием методов кусочно-линейной аппроксимации и градиентных методов, а в гл. 12 — модели *динамического программирования*, используемые при решении таких актуальных задач, как задача об оптимальном распределении средств между предприятиями, ресурсов между отраслями за ряд лет, задача о замене оборудования.

Учитывая, что приведенные в пособии задачи и наглядные способы их решения являются лишь условными примерами оптимизационных экономико-математических моделей, служащими для иллюстрации их сущности, в заключительной главе (гл. 13) раздела обсуждаются *вопросы применения ЭВМ* для решения реальных, многоразмерных задач математического программирования.

В разделе III рассматриваются специальные модели исследования операций: в гл. 14 — *модели сетевого планирования и управления*, в гл. 15 — *задачи массового обслуживания*, в гл. 16 — *модели управления запасами*. Эти модели, весьма отличные друг от друга по своей содержательной постановке, представляют основные, типичные классы задач исследования операций.

В пособии содержится большое количество задач. Задачи с решениями представлены на протяжении всего изложения учебного материала, а задачи для самостоятельной работы приведены в конце каждой главы в рубрике "Упражнения" (нумерация задач единая — начинается в основном тексте главы и продолжается в этой рубрике).

В книге знаком □ обозначается начало доказательства теоремы, ■ — ее окончание, знаком ▷ — начало условия задачи, ▶ — окончание ее решения.

ВВЕДЕНИЕ

Исследование операций — научная дисциплина, занимающаяся разработкой и практическим применением методов наиболее эффективного управления различными организационными системами.

Управление любой системой реализуется как процесс, подчиняющийся определенным закономерностям. Их знание помогает определить условия, необходимые и достаточные для осуществления данного процесса. Для этого все параметры, характеризующие процесс и внешние условия, должны быть количественно определены, измерены. Следовательно, ц е л ь и с с л е д о в а н и я о п е р а ц и й — *количественное обоснование принимаемых решений по организации управления.*

При решении конкретной задачи управления применение методов исследования операций предполагает:

- построение экономических и математических моделей для задач принятия решения в сложных ситуациях или в условиях неопределенности;
- изучение взаимосвязей, определяющих впоследствии принятие решений, и установление критериев эффективности, позволяющих оценивать преимущество того или иного варианта действия.

Примерами задач исследования операций, отражающих его специфику, могут служить следующие задачи.

З а д а ч а 1. Для обеспечения высокого качества выпускаемых изделий на заводе организуется система выборочного контроля. Требуется выбрать такие формы его организации — например, назначить размеры контрольных партий, указать последовательность контрольных операций, определить правила отбраковки, — чтобы обеспечить необходимое качество при минимальных расходах.

З а д а ч а 2. Для реализации определенной партии сезонных товаров создается сеть временных торговых точек. Требуется выбрать параметры сети — число точек, их размещение, количество персонала — так, чтобы обеспечить максимальную экономическую эффективность распродажи.

З а д а ч а 3. К заданному сроку необходимо провести массовое медицинское обследование группы населения с целью выявления определенных заболеваний. На обследование выделены материальные средства, оборудование, персонал. Требуется разработать такой план обследования — установить число медпунктов, их размещение, вид и количество анализов, чтобы выявить как можно больший процент из числа заболевших.

Необходимо отметить также задачи об использовании ресурсов (планирования производства), о смесях, об использовании мощностей (загрузке оборудования), о раскрое материалов (разд. 1.2), транспортную задачу (разд. 7.1) и др., в которых требуется найти решение, когда некоторый *критерий эффективности* (например, прибыль, выручка, затраты ресурсов и т.п.) принимает максимальное или минимальное значение.

Приведенные задачи относятся к разным областям практики, но в них есть общие черты: в каждом случае речь идет о каком-то *управляемом мероприятии* (*операции*), преследующем определенную *цель*. В задаче 1 — это организация выборочного контроля с целью обеспечить качество выпускаемой продукции; в задаче 2 — организация временных торговых точек с целью проведения сезонной распродажи; в задаче 3 — массовое медицинское обследование с целью определения процента заболевших. В каждой задаче заданы некоторые *условия* проведения этого мероприятия, в рамках которых следует принять *решение* — такое, чтобы мероприятие принесло определенную выгоду. Условиями проведения операции в каждой задаче оказываются средства, которыми мы располагаем, время, оборудование, технологии, а решение в задаче 1 заключается в выборе формы контроля — размера контрольных партий, правил отбраковки; в задаче 2 — в выборе числа точек размещения, количества персонала; в задаче 3 — в выборе числа медпунктов, вида и количества анализов.

Следует усвоить основные понятия и определения исследования операций.

Операция — любое управляемое мероприятие, направленное на достижение цели. Результат операции зависит от способа ее проведения, организации, иначе — от выбора некоторых параметров.

Всякий определенный выбор параметров называется *решением*. *Оптимальными* считают те решения, которые по тем или иным соображениям предпочтительнее других. Поэтому *основной задачей исследования операций является предварительное количественное обоснование оптимальных решений.*

Замечание 1. Следует обратить внимание на постановку проблемы: само *принятие решений* выходит за рамки исследования операций и относится к компетенции ответственного лица или группы лиц, которые могут учитывать и другие соображения, отличные от математически обоснованных.

Замечание 2. Если в одних задачах исследования операций (например, в разд. 1.2) оптимальным является решение, при котором некоторый критерий эффективности принимает максимальное или минимальное значение, то в других задачах это вовсе не обязательно. Так, в задаче 2 оптимальным можно считать такое количество торговых точек и персонала в них, при котором среднее время обслуживания покупателей не превысит, например, 5 мин, а длина очереди в среднем в любой момент окажется не более 3 человек.

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

Эффективность операции — степень ее приспособленности к выполнению задачи — количественно выражается в виде критерия эффективности — целевой функции. Например, в задаче об использовании ресурсов критерий эффективности — прибыль от реализации произведенной продукции, которую нужно максимизировать, в транспортной задаче — суммарные затраты на перевозку грузов от поставщиков к потребителям, которые нужно минимизировать. Выбор критерия эффективности определяет практическую ценность исследования. (Неправильно выбранный критерий может принести вред, ибо операции, организованные под углом зрения такого критерия эффективности, приводят порой к неоправданным затратам. Достаточно вспомнить пресловутый "вал" в качестве основного критерия хозяйственной деятельности предприятия.)

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

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

Например, в задаче об использовании ресурсов (*см.* разд. 1.2) к постоянным факторам следует отнести запасы ресурсов каждого вида, производственную матрицу, элементы которой определяют расход сырья каждого вида на единицу выпускаемой продукции каждого вида. Элементы решения — план выпуска продукции каждого вида.

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

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

Все модели исследования операций могут быть классифицированы в зависимости от природы и свойств операции, характера решаемых задач, особенностей применяемых математических методов.

Следует отметить прежде всего большой ***класс оптимизационных моделей.*** Такие задачи возникают при попытке оптимизировать планирование и управление сложными системами, в первую очередь экономическими системами. Оптимизационную задачу можно сформулировать в общем виде: ***найти переменные*** $x_1, x_2, ..., x_n$, ***удовлетворяющие системе неравенств (уравнений)***

$$\varphi_i(x_1, x_2, x_3, ..., x_n) \leq b_i, \quad i = 1, 2, ..., m \qquad \textbf{(0.1)}$$

и *обращающие в максимум* (*или минимум*) *целевую функцию, т.е.*

$$Z = f(x_1, x_2, ..., x_n) \rightarrow max\ (min). \qquad \textbf{(0.2)}$$

(Условия неотрицательности переменных, если они есть, входят в ограничения (0.1))

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

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

Пусть имеется n видов товаров и услуг, количества которых (в натуральных единицах) $x_1, x_2, ..., x_n$ по ценам соответственно $p_1, p_2, ..., p_n$ за единицу. Суммарная стоимость этих товаров и услуг составляет $\sum_{i=1}^{n} p_i x_i$.

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

$$Z = f(x_1, x_2, ..., x_n) \rightarrow \max \qquad \textbf{(0.3)}$$

при условии

$$\sum_{i=1}^{n} p_i x_i \leq I, \qquad \textbf{(0.4)}$$

$$x_i \geq 0\ (i = 1, 2, ..., n). \qquad \textbf{(0.5)}$$

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

Очевидно, что рассмотренная задача потребления (0.3)–(0.5), так же как и многие другие (*см.*, например, разд. 1.2), является частным случаем сформулированной выше общей задачи (0.1)–(0.2) на определение экстремума функции n переменных при некоторых ограничениях, т.е. задачей на *условный экстремум*.

В тех случаях, когда функции f и φ_i в задаче (0.1)–(0.2) хотя бы дважды дифференцируемы, можно применять *классические методы оптимизации*. Однако применение этих методов в исследовании операций весьма ограниченно, так как задача определения условного экстремума функции n переменных технически весьма трудна: метод дает возможность определить локальный экстремум, а из-за многомерности функции определение ее максимального (или минимального) значения (глобального экстремума) может оказаться весьма трудоемким — тем более, что этот экстремум возможен на границе области решений. Классические методы вовсе не работают, если множество допустимых значений аргумента дискретно или функция Z задана таблично. В этих случаях для решения задачи (0.1)—(0.2) применяются методы *математического программирования*.

Если критерий эффективности $Z = f(x_1, x_2, ..., x_n)$ (0.2) представляет линейную функцию, а функции $\varphi_i(x_1, x_2, x_3, ..., x_n)$ в системе ограничений (0.1) также линейны, то такая задача является задачей *линейного программирования*. Если, исходя из содержательного смысла, ее решения должны быть целыми числами, то эта задача *целочисленного линейного программирования*. Если критерий эффективности и (или) система ограничений задаются нелинейными функциями, то имеем задачу *нелинейного программирования*. В частности, если указанные функции обладают свойствами выпуклости, то полученная задача является задачей *выпуклого программирования*.

Если в задаче математического программирования имеется переменная времени и критерий эффективности (0.2) выражается не в явном виде как функция переменных, а косвенно — через уравнения, описывающие протекание операций во времени, то такая задача является задачей *динамического программирования*.

Если критерий эффективности (0.2) и система ограничений (0.1) задаются функциями вида $c x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$, то имеем задачу *геометрического программирования*. Если функции f и (или) φ_i в выражениях (0.2) и (0.1) зависят от параметров, то получаем задачу *параметрического программирования*, если эти функции носят случайный характер, — задачу *стохастического программирования*. Если точный оптимум найти алгоритмическим путем невозможно из-за чрезмерно большого числа вариантов решения, то прибегают к методам *эвристического программирования*, позволяющим существенно сократить просматриваемое число вариантов и найти если не оптимальное, то достаточно хорошее, удовлетворительное с точки зрения практики, решение.

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

По своей содержательной постановке множество других, типичных задач исследования операций может быть разбито на ряд классов.

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

Задачи массового обслуживания посвящены изучению и анализу систем обслуживания с очередями заявок или требований и состоят в определении показателей эффективности работы систем, их оптимальных характеристик, например, в определении числа каналов обслуживания, времени обслуживания и т.п.

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

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

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

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

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

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

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

В пособии нашли отражение основные из приведенных видов задач исследования операций.

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

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

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

$$f(x) = \omega_1 f_1(x)+\omega_2 f_2(x)+...+\omega_n f_n(x), \qquad (0.6)$$

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

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

В условиях определенности ω_i — числа, $f_i(x)$ — функции. В условиях неопределенности $f_i(x)$ могут оказаться случайными и вместо $f(x)$ в качестве целевой функции следует рассматривать математическое ожидание суммы (0.6).

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

Подробное обсуждение этих вопросов выходит за рамки нашего курса.

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

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

Методы исследования операций, как и любые математические методы, всегда в той или иной мере упрощают, огрубляют задачу, отражая порой нелинейные процессы линейными моделями, стохастические системы — детерминированными, динамические процессы — статическими моделями и т.д. Жизнь богаче любой схемы. Поэтому не следует ни преувеличивать значение количественных методов исследования операций, ни преуменьшать его, ссылаясь на примеры неудачных решений. Уместно привести в связи с этим шутливо-парадоксальное определение исследования операций, сделанное одним из его создателей Т. Саати, как "искусства давать плохие ответы на те практические вопросы, на которые даются еще худшие ответы другими методами" [3].

РАЗДЕЛ I. МОДЕЛИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ И ЕГО ПРИЛОЖЕНИЯ

Глава 1. ОБЩАЯ ПОСТАНОВКА ЗАДАЧИ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

1.1. Экономико-математическая модель

В настоящее время в литературе насчитывается несколько десятков определений понятия “модель”, отличающихся друг от друга. Тем не менее это понятие знакомо каждому: например, игрушечный самолет, бумажный голубь — модели самолета. Менее привычно представление о том, что фотоснимок пейзажа, географическая карта — это модель местности. И, наверное, новым для многих является то, что знакомая со школьных лет формула пути $s = vt$ — математическая модель. Под ***моделью*** будем понимать *условный образ какого-либо объекта, приближенно воссоздающий этот объект с помощью некоторого языка.* В экономико-математических моделях таким объектом является экономический процесс (например, использование ресурсов, распределение изделий между различными типами оборудования и т.п.), а языком — классические и специально разработанные математические методы.

Экономико-математическая модель — *математическое описание исследуемого экономического процесса или объекта.* Эта модель выражает закономерности экономического процесса в абстрактном виде с помощью математических соотношений. Использование математического моделирования в экономике позволяет углубить количественный экономический анализ, расширить область экономической информации, интенсифицировать экономические расчеты.

Можно выделить три основных этапа проведения экономико-математического моделирования. На первом этапе ставятся цели и задачи исследования, проводится качественное описание объекта в виде экономической модели. На втором этапе формируется математическая модель изучаемого объекта, осуществляется выбор (или разработка) методов исследования, проводится программирование модели на ЭВМ, подготавливается исходная информация. Далее проверяются пригодность машинной модели на основании правильности получаемых с ее помощью результатов и оценка их устойчивости. На третьем, основном, этапе экономико-математического моделирования осуществляются анализ математической модели, реализованной в виде программ для ЭВМ, проведение машинных расчетов, обработка и анализ полученных результатов.

Процедура экономико-математического моделирования заменяет дорогостоящие и трудоемкие натуральные эксперименты расчетами. Действительно, при использовании экономико-математических методов достаточно быстро и дешево производится на ЭВМ сравнение многочисленных вариантов планов и управленческих решений. В результате отбираются наиболее оптимальные варианты.

Ниже рассматриваются примеры экономико-математических моделей.

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

▷ **1.1**. Для изготовления двух видов продукции P_1 и P_2 используют четыре вида ресурсов S_1, S_2, S_3 и S_4. Запасы ресурсов, число единиц ресурсов, затрачиваемых на изготовление единицы продукции, приведены в табл. 1.1 (цифры условные).

Таблица 1.1

Вид ресурса	Запас ресурса	Число единиц ресурсов, затрачиваемых на изготовление единицы продукции	
		P_1	P_2
S_1	18	1	3
S_2	16	2	1
S_3	5	—	1
S_4	21	3	—

Прибыль, получаемая от единицы продукции P_1 и P_2, — соответственно 2 и 3 руб.

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

Решение. Составим экономико-математическую модель задачи.

Обозначим x_1, x_2 — число единиц продукции соответственно P_1 и P_2, запланированных к производству. Для их изготовления (см. табл. 1.1) потребуется $(1 \cdot x_1 + 3x_2)$ единиц ресурса S_1, $(2x_1 + 1 \cdot x_2)$ единиц ресурса S_2, $(1 \cdot x_2)$ единиц ресурса S_3 и $3x_1$ единиц ресурса S_4. Так как потребление ресурсов S_1, S_2, S_3 и S_4 не должно превышать их запасов, соответственно 18, 16, 5 и 21 единицы, то связь между потреблением ресурсов и их запасами выразится системой неравенств:

$$\begin{cases} x_1 + 3x_2 \le 18, \\ 2x_1 + x_2 \le 16, \\ x_2 \le 5, \\ 3x_1 \le 21. \end{cases} \qquad \mathbf{(1.1)}$$

По смыслу задачи переменные

$$x_1 \ge 0,\ x_2 \ge 0. \qquad \mathbf{(1.2)}$$

Суммарная прибыль F составит $2x_1$ руб. от реализации продукции P_1 и $3x_2$ руб. -- от реализации продукции P_2, т.е.

$$F = 2x_1 + 3x_2. \qquad \mathbf{(1.3)}$$

Итак, экономико-математическая модель задачи: *найти такой план выпуска продукции* $X = (x_1, x_2)$, *удовлетворяющий системе* (1.1) *и условию* (1.2), *при котором функция* (1.3) *принимает максимальное значение.*▶

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

Обозначим x_j $(j = 1, 2, \ldots, n)$ — число единиц продукции P_j, запланированной к производству; b_i $(i = 1, 2, \ldots, m)$ — запас ресурса S_i, a_{ij} — число единиц ресурса S_i, затрачиваемого на изготовление единицы продукции P_j (числа a_{ij} часто называют *технологическими коэффициентами*); c_j — прибыль от реализации единицы продукции P_j.

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

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m \end{cases} \qquad \textbf{(1.4)}$$

и условию

$$x_1 \ge 0,\ x_2 \ge 0, \ldots,\ x_n \ge 0, \qquad \textbf{(1.5)}$$

при котором функция

$$F = c_1x_1 + c_2x_2 + \ldots + c_nx_n \qquad \textbf{(1.6)}$$

принимает максимальное значение.

2. Задача составления рациона (задача о диете, задача о смесях).

> **1.2**. Имеется два вида корма I и II, содержащие питательные вещества (витамины) S_1, S_2 и S_3. Содержание числа единиц питательных веществ в 1 кг каждого вида корма и необходимый минимум питательных веществ приведены в табл. 1.2 (цифры условные).

Таблица 1.2

Питательное вещество (витамин)	Необходимый минимум питательных веществ	Число единиц питательных веществ в 1 кг корма	
		I	II
S_1	9	3	1
S_2	8	1	2
S_3	12	1	6

Стоимость 1 кг корма I и II соответственно равна 4 и 6 руб.

Необходимо составить дневной рацион, имеющий минимальную стоимость, в котором содержание каждого вида питательных веществ было бы не менее установленного предела.

Решение. Составим экономико-математическую модель задачи.

Обозначим x_1, x_2 — количество кормов I и II, входящих в дневной рацион. Тогда этот рацион (см. табл. 1.2) будет включать $(3x_1 + 1 \cdot x_2)$ единиц питательного вещества S_1, $(1 \cdot x_1 + 2x_2)$ единиц вещества S_2 и $(1 \cdot x_1 + 6x_2)$ единиц питательного вещества S_3. Так как содержание питательных веществ S_1, S_2 и S_3 в рационе должно быть не менее соответственно 9, 8 и 12 единиц, то получим систему неравенств:

$$\begin{cases} 3x_1 + x_2 \geq 9, \\ x_1 + 2x_2 \geq 8, \\ x_1 + 6x_2 \geq 12. \end{cases} \qquad \textbf{(1.7)}$$

Кроме того, переменные

$$x_1 \geq 0, x_2 \geq 0. \qquad \textbf{(1.8)}$$

Обшая стоимость рациона составит (в руб.)

$$F = 4x_1 + 6x_2. \qquad \textbf{(1.9)}$$

Итак, экономико-математическая модель задачи: *составить дневной рацион* $X = (x_1, x_2)$, *удовлетворяющий системе* (1.7) *и условию* (1.8), *при котором функция* (1.9) *принимает минимальное значение.*▶

Для формулировки задачи в общей постановке обозначим: x_j ($j = 1,2, ..., n$) — число единиц корма n-го вида; b_i ($i = 1, 2, ..., m$), — необходимый минимум содержания в рационе питательного вещества S_i; a_{ij} — число единиц питательного вещества S_i в единице корма j-го вида; c_j — стоимость единицы корма j-го вида. Тогда экономико-математическая модель задачи примет вид:

найти такой рацион $X = (x_1, x_2, ..., x_n)$, *удовлетворяющий системе*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \geq b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \geq b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \geq b_m \end{cases} \qquad \textbf{(1.10)}$$

и условию

$$x_1 \geq 0, x_2 \geq 0, ..., x_n \geq 0, \qquad \textbf{(1.11)}$$

при котором функция

$$F = c_1x_1 + c_2x_2 + ... + c_nx_n \qquad \textbf{(1.12)}$$

принимает максимальное значение.

3. Задача об использовании мощностей (задача о загрузке оборудования).

Предприятию задан план производства продукции по времени и номенклатуре: требуется за время T выпустить n_1, n_2,..., n_k единиц продукции P_1, P_2, ..., P_k. Продукция производится на станках S_1, S_2, ..., S_m. Для каждого станка известны производительность a_{ij} (т.е. число единиц продукции P_j, которое можно произвести на станке S_i) и затраты b_{ij} на изготовление продукции P_j на станке S_i в единицу времени.

Необходимо составить такой план работы станков (т.е. так распределить выпуск продукции между станками), чтобы затраты на производство всей продукции были минимальными.

Составим экономико-математическую модель задачи.

Обозначим x_{ij} — время, в течение которого станок S_i будет занят изготовлением продукции P_j ($i = 1, 2,..., m$; $j = 1, 2, ..., k$).

Так как время работы каждого станка ограничено и не превышает T, то справедливы неравенства:

$$\begin{cases} x_{11} + x_{12} + \ldots + x_{1k} \leq T, \\ x_{21} + x_{22} + \ldots + x_{2k} \leq T, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_{m1} + x_{m2} + \ldots + x_{mk} \leq T. \end{cases} \qquad (1.13)$$

Для выполнения плана выпуска по номенклатуре необходимо, чтобы выполнялись следующие равенства:

$$\begin{cases} a_{11}x_{11} + a_{21}x_{21} + \ldots + a_{m1}x_{m1} = n_1, \\ a_{12}x_{12} + a_{22}x_{22} + \ldots + a_{m2}x_{m2} = n_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1k}x_{1k} + a_{2k}x_{2k} + \ldots + a_{mk}x_{mk} = n_k. \end{cases} \qquad (1.14)$$

Кроме того,

$$x_{ij} \geq 0 \ (i = 1,2,..., m; j = 1, 2,...,k). \qquad (1.15)$$

Затраты на производство всей продукции выразятся функцией

$$F = b_{11}x_{11} + b_{12}x_{12} + \ldots + b_{mk}x_{mk}. \qquad (1.16)$$

Экономико-математическая модель задачи об использовании мощностей примет вид: *найти такое решение $X = (x_{11}, x_{12}, ..., x_{mk})$,*

удовлетворяющее системам (1.13) и (1.14) и условию (1.15), при котором функция (1.16) принимает минимальное значение.

4. Задача о раскрое материалов.

На раскрой (распил, обработку) поступает материал одного образца в количестве a единиц. Требуется изготовить из него l разных комплектующих изделий в количествах, пропорциональных числам b_1, b_2, $...b_l$ (условие комплектности). Каждая единица материала может быть раскроена n различными способами, причем использование i-го способа ($i = 1,2, ..., n$) дает a_{ik} единиц k-го изделия ($k = 1,2, ..., l$).

Необходимо найти план раскроя, обеспечивающий максимальное число комплектов.

Составим экономико-математическую модель задачи.

Обозначим x_i — число единиц материала, раскраиваемых i-м способом, и x — число изготавливаемых комплектов изделий.

Так как общее количество материала равно сумме его единиц, раскраиваемых различными способами, то

$$\sum_{i=1}^{n} x_i = a. \qquad (1.17)$$

Требование комплектности выразится уравнениями

$$\sum_{i=1}^{n} x_i a_{ik} = b_k x \ (k = 1,2, ..., l). \qquad (1.18)$$

Очевидно, что

$$x_i \geq 0 \ (i = 1,2,..., n). \qquad (1.19)$$

Экономико-математическая модель задачи: *найти такое решение* $X=(x_1, x_2,..., x_n)$, *удовлетворяющее системе уравнений* (1.17) — (1.18) *и условию* (1.19), *при котором функция* $F = x$ *принимает максимальное значение.*

▷ **1.3. Для изготовления брусьев длиной 1,2 м, 3 м и 5 м в соотношении 2:1:3 на распил поступают 195 бревен длиной 6 м. Определить план распила, обеспечивающий максимальное число комплектов. Составить экономико-математическую модель задачи.**

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

Таблица 1.3

Способ распила i	Число получаемых брусьев длиной, м		
	1,2	3,0	5,0
1	5	—	—
2	2	1	—
3	—	2	—
4	—	—	1

Обозначим: x_i — число бревен, распиленных i-м способом ($i = 1, 2, 3, 4$); x — число комплектов брусьев.

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

$$F = x \to \max$$

при ограничениях:

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 195, \\ 5x_1 + 2x_2 = 2x, \\ x_2 + 2x_3 = x, \\ x_4 = 3x, \end{cases}$$

$$x_i \geq 0 \ (i = 1, 2, 3, 4).$$▶

Задачу о раскрое легко обобщить на случай m раскраиваемых материалов.

Пусть каждая единица j-го материала ($j = 1, 2, ..., m$) может быть раскроена n различными способами, причем использование i-го способа ($i = 1, 2, ... , n$) дает a_{ijk} единиц k-го изделия ($k = 1, 2, ..., l$), а запас j-го материала равен a_j единиц.

Обозначим x_{ij} — число единиц j-го материала, раскрываемого i-м способом.

Экономико-математическая модель задачи о раскрое в общей постановке примет вид: ***найти такое решение*** $X = (x_{11}, x_{12}, ..., x_{nm})$, ***удовлетворяющее системе***

$$\begin{cases} \sum_{i=1}^{n} x_{ij} \leq a_j (j = 1,2,\ldots,m), \\ \sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij}a_{ijk} = b_k x (k = 1,2,\ldots,l) \end{cases}$$

и условию $x_{ij} \geq 0$, ***при котором функция*** $F = x$ ***принимает максимальное значение.***

5. Транспортная задача рассмотрена в гл. 7.

1.3. Общая задача линейного программирования

Рассмотренные выше примеры задач линейного программирования позволяют сформулировать ***общую задачу линейного программирования.***

Дана система m линейных уравнений и неравенств с n переменными

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le b_k, \\ a_{k+1,1}x_1 + a_{k+1,2}x_2 + \ldots + a_{k+1,n}x_n = b_{k+1}, \\ a_{k+2,1}x_1 + a_{k+2,2}x_2 + \ldots + a_{k+2,n}x_n = b_{k+2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases} \quad (1.20)$$

и линейная функция

$$F = c_1x_1 + c_2x_2 + \ldots + c_nx_n. \quad (1.21)$$

Необходимо найти такое решение системы $X = (x_1, x_2, \ldots, x_j, \ldots, x_n)$, где

$$x_j \ge 0 \ (j=1, 2, \ldots, l;\ l \le n), \quad (1.22)$$

при котором линейная функция F (1.21) *принимает оптимальное (т.е. максимальное или минимальное) значение.*

Система (1.20) называется *системой ограничений*, а функция F — *линейной функцией, линейной формой, целевой функцией* или *функцией цели.*

Более кратко общую задачу линейного программирования можно представить в виде:

$$F = \sum_{j=1}^{n} c_j x_j \to \max \ (\text{или} \to \min)$$

при ограничениях:

$$\begin{cases} \sum\limits_{j=1}^{n} a_{ij}x_j \le b_i (i = 1,2,\ldots,k), \\ \sum\limits_{j=1}^{n} a_{ij}x_j = b_i (i = k+1, k+2,\ldots,m), \end{cases}$$

$$x_j \ge 0 \ (j = 1,\ 2,\ \ldots,\ l;\ l \le n).$$

Оптимальным решением (или *оптимальным планом*) задачи линейного программирования называется решение $X = (x_1, x_2, ..., x_j, ..., x_n)$ системы ограничений (1.20), удовлетворяющее условию (1.22), при котором линейная функция (1.21) принимает оптимальное (максимальное или минимальное) значение.

Термины "решение" и "план" — синонимы, однако первый используется чаще, когда речь идет о формальной стороне задачи (ее математическом решении), а второй — о содержательной стороне (экономической интерпретации).

При условии, что все переменные неотрицательны ($x_j \ge 0, j = 1, 2, ..., n$), система ограничений (1.20) состоит лишь из одних неравенств, — такая задача линейного программирования называется *стандартной*; если система ограничений состоит из одних уравнений, то задача называется *канонической*[1]. Так, в приведенных выше примерах задач линейного программирования задачи 1 и 2 — стандартные, задача 4 — каноническая, а задача 3 — общая.

Любая задача линейного программирования может быть сведена к канонической, стандартной или общей задаче. Рассмотрим вначале вспомогательную теорему.

Теорема 1.1. *Всякому решению* $(\alpha_1, \alpha_2, ..., \alpha_n)$ *неравенства*

$$\alpha_{i1}x_1 + \alpha_{i2}x_2 + ... + \alpha_{in}x_n \le b_i \qquad (1.23)$$

соответствует определенное решение $(\alpha_1, \alpha_2, ..., \alpha_n; \alpha_{n+i})$ *уравнения*

$$\alpha_{i1}x_1 + \alpha_{i2}x_2 + ... + \alpha_{in}x_n + x_{n+i} = b_i, \qquad (1.24)$$

в котором

$$x_{n+i} \ge 0 \qquad (1.25)$$

и, наоборот, каждому решению $(\alpha_1, \alpha_2, ..., \alpha_n; \alpha_{n+i})$ *уравнения* (1.24) *и неравенства* (1.25) *соответствует определенное решение* $(\alpha_1, \alpha_2, ..., \alpha_n)$ *неравенства* (1.23).

---

1 В ряде работ по математическому программированию стандартную задачу называют *симметричной*, а каноническую — *основной*.

Используя эту теорему (которую принимаем без доказательства), представим в качестве примера стандартную задачу (1.4)—(1.6) в каноническом виде. С этой целью в каждое из m неравенств системы ограничений (1.4) введем дополнительные неотрицательные переменные x_{n+1}, x_{n+2}, ..., x_{n+m} и система ограничений (1.4) примет вид:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x_{n+2} = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + x_{n+m} = b_m. \end{cases} \quad \textbf{(1.26)}$$

Таким образом, стандартная задача (1.4)—(1.6) в канонической форме: *найти такое решение* $X = (x_1, x_2, ..., x_n, x_{n+1}, ..., x_{n+m})$, *удовлетворяющее системе* (1.26) *и условию* (1.5), *при котором функция* (1.6) *принимает максимальное значение.*

Замечание. В рассматриваемой задаче все неравенства вида "≤", поэтому дополнительные неотрицательные переменные вводились со знаком "+". В случае неравенства вида "≥", как, например, в задаче **(1.10) — (1.12)**, соответствующие дополнительные переменные следовало бы ввести со знаком "−".

УПРАЖНЕНИЯ

В задачах **1.4—1.7** составить экономико-математические модели.

1.4. Для производства двух видов изделий *A* и *B* предприятие использует три вида сырья. Другие условия задачи приведены в таблице.

Вид сырья	Нормы расхода сырья на одно изделие, кг		Общее количество сырья, кг
	A	*B*	
I	12	4	300
II	4	4	120
III	3	12	252
Прибыль от реализации одного изделия, ден.ед.	30	40	

Составить такой план выпуска продукции, при котором прибыль предприятия от реализации продукции будет максимальной при условии, что изделий *B* надо выпустить не менее, чем изделий *A*.

1.5. Рацион для питания животных на ферме состоит из двух видов кормов I и II. Один килограмм корма I стоит 80 ден. ед. и содержит: 1 ед. жиров, 3 ед. белков, 1 ед. углеводов, 2 ед. нитратов. Один килограмм корма II стоит 10 ден. ед. и содержит 3 ед. жиров, 1 ед. белков, 8 ед. углеводов, 4 ед. нитратов.

Составить наиболее дешевый рацион питания, обеспечивающий жиров не менее 6 ед., белков не менее 9 ед., углеводов не менее 8 ед., нитратов не более 16 ед.

1.6. На двух автоматических линиях выпускают аппараты трех типов. Другие условия задачи приведены в таблице.

Тип аппарата	Производительность работы линий, шт. в сутки		Затраты на работу линий, ден. ед. в сутки		План, шт.
	1	2	1	2	
A	4	3	400	300	50
B	6	5	100	200	40
C	8	2	300	400	50

Составить такой план загрузки станков, чтобы затраты были минимальными, а задание выполнено не более чем за 10 суток.

1.7. Необходимо распилить 20 бревен длиной по 5 м каждое на бруски по 2 м и 3 м; при этом должно получиться равное количество брусков каждого размера.

Составить такой план распила, при котором будет получено максимальное число комплектов и все бревна будут распилены (в один комплект входит по одному бруску каждого размера).

Глава 2. ЭЛЕМЕНТЫ ЛИНЕЙНОЙ АЛГЕБРЫ И ГЕОМЕТРИИ ВЫПУКЛЫХ МНОЖЕСТВ

2.1. Система *m* линейных уравнений с *n* переменными

Система m линейных уравнений с n переменными имеет вид

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1j}x_j + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2j}x_j + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{ij}x_j + \ldots + a_{in}x_n = b_i, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mj}x_j + \ldots + a_{mn}x_n = b_m, \end{cases} \tag{2.1}$$

или в краткой записи

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

В задачах линейного программирования представляют интерес системы, в которых ранг r матрицы системы $A = (a_{ij})$, i=1, 2, ..., m; j=1, 2, ..., n, или, что то же самое, максимальное число независимых уравнений системы r меньше числа переменных, т.е. $r < n$. Будем полагать, что в системе (2.1) все m уравнений системы независимы, т.е. $r = m$ и соответственно $m < n$.

Любые m переменных системы m линейных уравнений с n переменными ($m < n$) называются основными (или базисными), если определитель матрицы коэффициентов при них отличен от нуля[1].

---

[1] В литературе такой определитель часто называют *базисным минором* матрицы A.

Тогда остальные $n-m$ переменных называются ***неосновными*** *(или* ***свободными****).*

Основными могут быть разные группы из n переменных. Максимально возможное число групп основных переменных равно числу способов выбора m переменных из их общего числа n, т.е. числу сочетаний C_n^m. Но так как могут встретиться случаи, когда определитель матрицы коэффициентов при m переменных равен нулю, то общее число групп основных переменных не превосходит C_n^m.

▷ **2.1.** Найти все возможные группы основных переменных в системе

$$\begin{cases} x_1 - x_2 - 2x_3 + x_4 = 0, \\ 2x_1 + x_2 + 2x_3 - x_4 = 0. \end{cases}$$

Р е ш е н и е. Общее число групп основных переменных не более чем $C_4^2 = 4 \cdot 3 / 2 = 6$, т.е. возможные группы основных переменных: x_1, x_2; x_1, x_3; x_1, x_4; x_2, x_3; x_2, x_4; x_3, x_4.

Выясним, могут ли быть основными переменные x_1, x_2. Так как определитель матрицы из коэффициентов при этих переменных $\begin{vmatrix} 1 & -1 \\ 2 & 1 \end{vmatrix} = 1 \cdot 1 - 2(-1) = 3 \neq 0$, то x_1, x_2 могут быть основными переменными. Рассуждая аналогично, найдем, что могут быть основными переменные x_1, x_3; x_1, x_4, но не могут быть основными x_2, x_3; x_2, x_4; x_3, x_4, так как в трех последних группах переменных соответствующие определители равны нулю (например, для переменных x_3, x_4 $\begin{vmatrix} -2 & 1 \\ 2 & -1 \end{vmatrix} = 0$). ▶

Для решения системы (2.1) при условии $m < n$ докажем следующую теорему.

Теорема 2.1. *Если для системы m линейных уравнений с n переменными ($m < n$) ранг матрицы коэффициентов при переменных равен m, т.е. существует хотя бы одна группа основных переменных, то эта система является неопределенной, причем каждому произвольному набору значений неосновных переменных соответствует одно решение системы.*

□ Пусть, например, $x_1, x_2, ..., x_m$ — основные переменные (если это не так, то нумерацию соответствующих переменных можно изменить), т.е. определитель матрицы

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_m & a_{m2} & \dots & a_{mm} \end{vmatrix} \neq 0.$$

Оставим в левых частях уравнений системы (2.1) члены с переменными $x_1, x_2, ..., x_m$, а члены с переменными $x_{m+1}, x_{m+2}, ..., x_n$ перенесем в правые части. Получим

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m = b_1 - a_{1,m+1}x_{m+1} - \dots - a_{1n}x_n, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m = b_2 - a_{2,m+1}x_{m+1} - \dots - a_{2n}x_n, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mm}x_m = b_m - a_{m,m+1}x_{m+1} - \dots - a_{mn}x_n. \end{cases}$$

Задавая неосновным переменным $x_{m+1}, x_{m+2}, ..., x_n$ произвольные значения, каждый раз будем получать новую систему с новыми свободными членами $b_1', b_2', \dots, b_m'$. Каждая из полученных систем будет иметь один и тот же определитель $|A| \neq 0$, следовательно, в силу теоремы Крамера каждая из этих систем будет иметь единственное решение. Так как получаемых таким образом систем бесконечно много, то и система (2.1) будет иметь бесконечное множество решений ■.

▷ **2.2.** Решить систему уравнений

$$\begin{cases} x_1 - x_2 - 2x_3 + x_4 = 0, \\ 2x_1 + x_2 + 2x_3 - x_4 = 2. \end{cases}$$

Р е ш е н и е. В задаче **2.1** установлено, что основными могут быть переменные x_1, x_2; x_1, x_3; x_1, x_4. Возьмем в качестве основных переменные x_1, x_2, а переменные x_3, x_4 будем считать неосновными и перенесем их с соответствующими коэффициентами в правые части уравнений системы. Получим

$$\begin{cases} x_1 - x_2 = 2x_3 - x_4, \\ 2x_1 + x_2 = 2x_3 + x_4. \end{cases}$$

Решая данную систему любым способом, найдем $x_1 = 2/3$; $x_2 = 2/3 - 2x_3 + x_4$. Задавая неосновным переменным x_3 и x_4 произвольные значения, например, $x_3 = c_1$, $x_4 = c_2$, получим бесконечное множество решений этой системы ($x_1 = 2/3$; $x_2 = 2/3 - 2c_1 + c_2$; $x_3 = c_1$; $x_4 = c_2$). ▶

Решение $X = (x_1, x_2, ..., x_n)$ системы (2.1) называется *допустимым*[1], если оно содержит лишь неотрицательные компоненты, т.е. $x_j \geq 0$ для любых $j = 1, 2, ..., n$. В противном случае решение называется *недопустимым*. Так, в задаче **2.2** решение системы при $c_1 = 2$, $c_2 = 5$, т.е. $X_1 = (2/3; 5/3; 2; 5)$ является допустимым, а при $c_1 = 2$, $c_2 = 1$, т.е. $X_2 = (2/3; -7/3; 2; 1)$ — недопустимым.

Среди бесконечного множества решений системы выделяют так называемые базисные решения.

***Базисным** решением системы m линейных уравнений с n переменными называется решение, в котором все $n-m$ неосновных переменных равны нулю.*

В задачах линейного программирования особый интерес представляют *допустимые базисные решения*, или, как их еще называют, *опорные планы*. Число базисных решений является конечным, так как оно равно числу групп основных переменных, не превосходящему C_n^m. Базисное решение, в котором хотя бы одна из основных переменных равна нулю, называется *вырожденным*.

▷ **2.3**. Найти все базисные решения системы, приведенной в задаче **2.1**.

Р е ш е н и е. В задаче **2.1** было установлено, что существует три группы основных переменных x_1, x_2; x_1, x_3; x_1, x_4, т.е. число базисных решений равно 3.

Найдем первое базисное решение, взяв в качестве основных переменные x_1, x_2 и неосновных — переменные x_3, x_4. Приравняв неосновные переменные нулю, т.е. при $x_3 = x_4 = 0$, получим систему уравнений в виде

$$\begin{cases} x_1 - x_2 = 0, \\ 2x_1 + x_2 = 2, \end{cases}$$

откуда $x_1 = 2/3$; $x_2 = 2/3$. Следовательно, первое базисное решение системы $X_1 = (2/3; 2/3; 0; 0)$ — допустимое.

---

[1] **Именно такие решения представляют интерес в большинстве задач линейного программирования.**

Если взять за основные переменные x_1, x_3 и приравнять нулю соответствующие неосновные переменные $x_2 = x_4 = 0$, получим второе базисное решение $X_2 = (2/3; 0; 2/3; 0)$ — также допустимое. Аналогично можно найти и третье базисное решение $X_3 = (2/3; 0; 0; -2/3)$ — недопустимое.▶

Совместная система (2.1) имеет бесконечно много решений, из них базисных решений — конечное число, не превосходящее C_n^m.

В школьном курсе математики *выпуклыми* назывались многоугольники, целиком расположенные по одну сторону от прямых, на которых лежат их стороны.

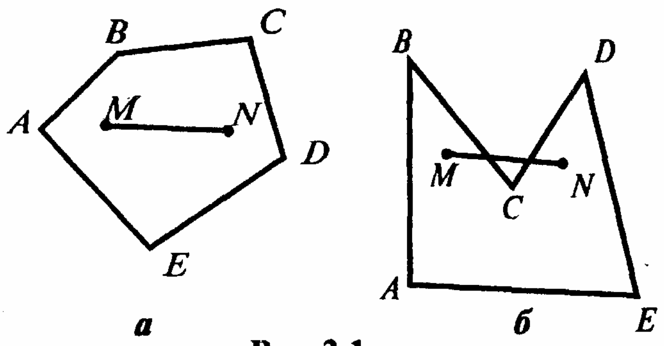


Рис. 2.1

Например, многоугольник на рис. 2.1, *а* — выпуклый, а многоугольник на рис. 2.1, *б* не является выпуклым (он расположен по обе стороны от прямой *BC*).

Общим определяющим свойством, которое отличает выпуклый многоугольник от невыпуклого, является то, что если взять любые две его точки и соединить их отрезком, то весь отрезок будет принадлежать этому многоугольнику. Это свойство может быть принято за определение выпуклого множества точек.

Множество точек называется ***выпуклым***, *если оно вместе с любыми двумя своими точками содержит весь отрезок, соединяющий эти точки.*

Согласно этому определению многоугольник на рис. 2.1, *а* является выпуклым множеством, а многоугольник на рис. 2.1, *б* таковым не является, ибо отрезок *MN* между двумя его точками *M* и *N* не полностью принадлежит этому многоугольнику.

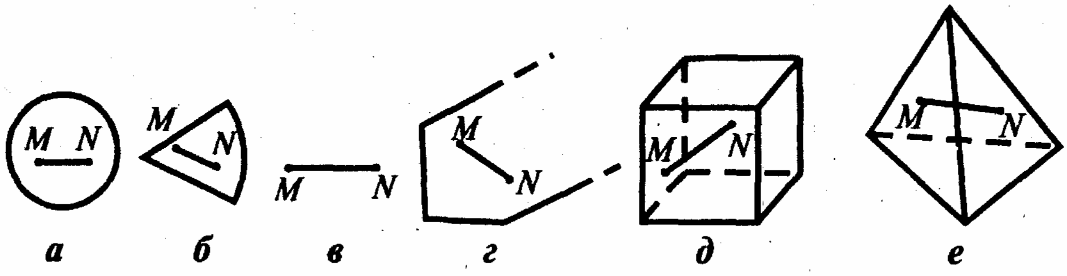


Рис. 2.2

Выпуклыми множествами могут быть не только многоугольники. Примерами выпуклых множеств являются круг, сектор, отрезок, многоугольная область, куб, пирамида (рис. 2.2, *а—е*), многогранная область, прямая, полуплоскость, полупространство и т.п.

Выпуклые множества обладают важным свойством, которое устанавливается следующей теоремой.

Теорема 2.2. *Пересечение (общая часть) любого числа выпуклых множеств есть выпуклое множество.*

□ Пусть M и N — любые две точки пересечения двух[1] множеств A и B (рис. 2.3). Так как точки M и N принадлежат пересечению множеств, т.е. одновременно и выпуклому множеству A, и выпуклому множеству B, то согласно определению выпуклого множества все точки отрезка MN будут принадлежать как множеству A, так и множеству B, т.е. пересечению этих множеств. А это и означает, что пересечение данных множеств есть выпуклое множество. ■

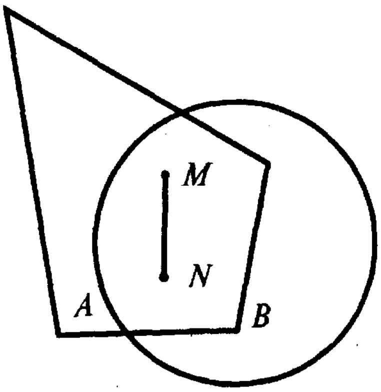


Рис. 2.3

Среди точек выпуклого множества можно выделить внутренние, граничные и угловые точки.

Точка множества называется *внутренней*, если в некоторой ее окрестности[2] содержатся точки только данного множества.

Точка множества называется *граничной*, если в любой ее окрестности содержатся как точки, принадлежащие данному множеству, так и точки, не принадлежащие ему.

---

[1] Для доказательства теоремы ограничимся случаем двух множеств.

[2] Под *окрестностью* точки плоскости (пространства) подразумевается круг (шар) с центром в этой точке.

Особый интерес в задачах линейного программирования представляют угловые точки.

Точка множества называется ***угловой*** *(или* ***крайней****), если она не является внутренней ни для какого отрезка, целиком принадлежащего данному множеству.*

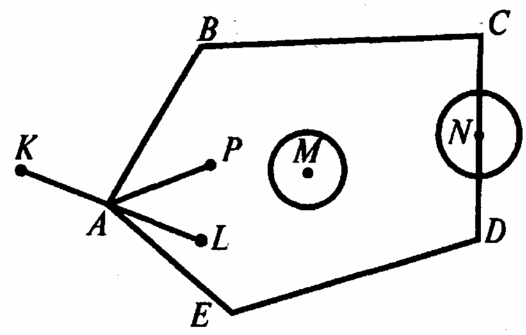


Рис. 2.4

На рис. 2.4 приведены примеры различных точек многоугольника: внутренней (точки M), граничной (точка N) и угловых (точки A, B, C, D, E). Точка A — угловая, так как для любого отрезка, целиком принадлежащего многоугольнику, например, отрезка AP, она не является внутренней; точка A — внутренняя для отрезка KL, но этот отрезок не принадлежит целиком многоугольнику.

Для выпуклого множества угловые точки всегда совпадают с вершинами многоугольника (многогранника), в то же время для невыпуклого множества это не обязательно. Так, на рис. 2.5 точка A является вершиной невыпуклого многоугольника, но не угловой (она является внутренней для отрезка KL, целиком принадлежащего этому многоугольнику).

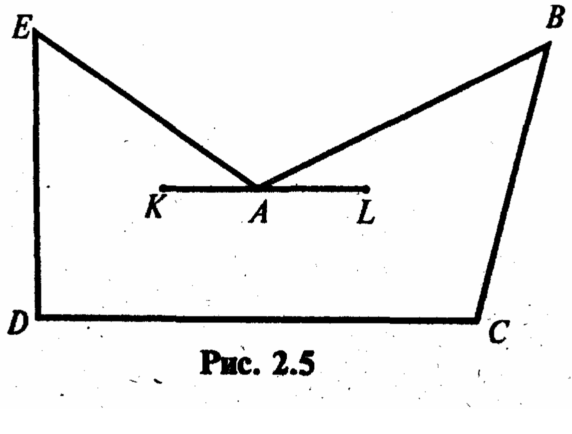


Рис. 2.5

Множество точек называется *замкнутым*, если включает все свои граничные точки. Множество точек называется *ограниченным*, если существует шар (круг) радиуса конечной длины с центром в любой точке множества, который полностью содержит в себе данное множество; в противном случае множество называется *неограниченным*.

Если фигура ограничена только прямыми или их отрезками, то число ее угловых точек конечно; в случае криволинейности границ фигура содержит бесконечно много угловых точек, что позволяет сделать следующее определение.

Выпуклое замкнутое множество точек пространства (плоскости), имеющее конечное число угловых точек, называется ***выпуклым многогранником (многоугольником)****, если оно ограниченное, и* ***выпуклой многогранной (многоугольной) областью****, если оно неограниченное.*

До сих пор рассматривались выпуклые множества точек на плоскости и в пространстве. Аналитически такие точки изображаются упорядоченной парой чисел (x_1, x_2) или упорядоченной тройкой чисел (x_1, x_2, x_3). Понятие точки можно обобщить, подразумевая под точкой (или вектором) упорядоченный набор n чисел $X = (x_1, x_2, ..., x_n)$, в котором числа $x_1, x_2, ..., x_n$ называются координатами точки (вектора). Такое обобщение имеет смысл, так как если взять какой-либо экономический объект, то для его характеристики двух-трех чисел обычно бывает недостаточно и необходимо взять n чисел, где $n > 3$.

Множество всех точек $X = (x_1, x_2,..., x_n)$ образует n-мерное точечное (векторное) пространство. При $n > 3$ точки и фигуры n-мерного пространства не имеют реального геометрического смысла и все исследования объектов этого пространства необходимо проводить в аналитической форме. Тем не менее оказывается целесообразным и в этом случае использовать геометрические понятия для облегчения представлений об объектах n-мерного пространства.

2.3. Геометрический смысл решений неравенств, уравнений и их систем

Рассмотрим **решения неравенств**.

Теорема 2.3. *Множество решений неравенства с двумя переменными*

$$a_{11}x_1 + a_{12}x_2 \le b_1 \tag{2.2}$$

является одной из двух полуплоскостей, на которые вся плоскость делится прямой $a_{11}x_1 + a_{12}x_2 = b_1$, включая и эту прямую, а другая полуплоскость с той же прямой есть множество решений неравенства

$$a_{11}x_1 + a_{12}x_2 \geq b_1. \tag{2.3}$$

□ Для произвольной абсциссы x_1 ордината точки M (рис. 2.6), лежащей на прямой $a_{11}x_1 + a_{12}x_2 = b_1$, при условии $a_{12} \neq 0$, есть $x_2 = -\frac{a_{11}}{a_{12}}x_1 + \frac{b_1}{a_{12}}$, т.е. координаты точки $M\left(x_1;\ -\frac{a_{11}}{a_{12}}x_1 + \frac{b_1}{a_{12}}\right)$.

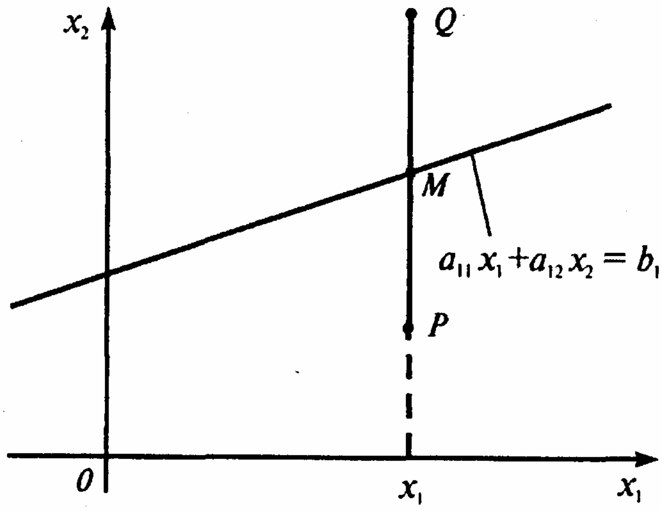


Рис. 2.6

Через точку M проведем прямую, параллельную оси $0x_2$. Тогда для любых точек P и Q этой прямой, расположенных выше и ниже точки M, т.е. в верхней и нижней полуплоскостях, будут верны неравенства $x_{2Q} \geq x_{2M}$ и $x_{2P} \leq x_{2Q}$ или $x_2 \geq -\frac{a_{11}}{a_{12}}x_1 + \frac{b_1}{a_{12}}$ и $x_2 \leq -\frac{a_{11}}{a_{12}}x_1 + \frac{b_1}{a_{12}}$. При условии $a_{12} > 0$ неравенства преобразуются соответственно к виду $a_{11}x_1 + a_{12}x_2 \geq b_1$ и $a_{11}x_1 + a_{12}x_2 \leq b_1$, т.е координаты всех точек верхней полуплоскости удовлетворяют неравенству (2.2), а нижней полуплоскости — неравенству (2.3). В случае $a_{12} < 0$, наоборот, координаты всех точек верхней полуплоскости удовлетворяют неравенству (2.3), а координаты нижней полуплоскости — неравенству (2.2).■

▷ **2.4**. Построить множество решений неравенства:

а) $3x_1 - 4x_2 + 12 \le 0$; б) $3x_1 - 2x_2 \ge 0$.

Р е ш е н и е. В соответствии с теоремой 2.3, множество решений неравенства есть полуплоскость.

а) Построим границу полуплоскости — прямую $3x_1 - 4x_2 + 12 = 0$, найдя точки ее пересечения с осями координат A (−4;0) и B (0;3) на рис. 2.7, *а*.

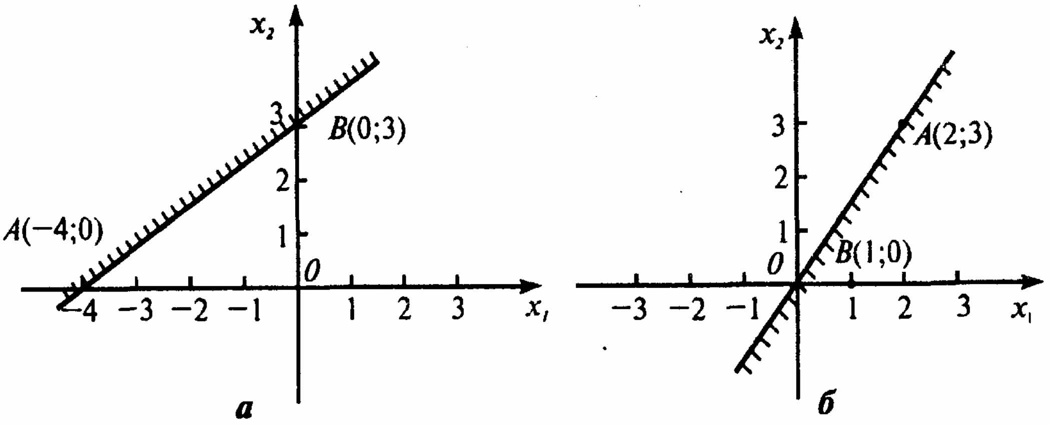


Рис. 2.7

Для определения искомой полуплоскости (верхней или нижней) рекомендуется задать произвольную контрольную точку, не лежащую на ее границе — построенной прямой. Если неравенство выполняется в контрольной точке, то оно выполняется и во всех точках полуплоскости, содержащей контрольную точку, и не выполняется во всех точках другой полуплоскости. И наоборот, в случае невыполнения неравенства в контрольной точке, оно не выполняется во всех точках полуплоскости, содержащей контрольную точку, и выполняется во всех точках другой полуплоскости.

В качестве контрольной точки удобно взять начало координат O (0;0), не лежащее на построенной прямой. Координаты точки O не удовлетворяют неравенству: $3 \cdot 0 - 4 \cdot 0 + 12 \le 0$, следовательно, решением данного неравенства является нижняя полуплоскость, не содержащая контрольную точку O. Искомая полуплоскость выделена штриховкой.

б) Построим границу полуплоскости — прямую $3x_1 - 4x_2 = 0$ по двум точкам. Одной из этих точек является начало координат на рис. 2.7, *б* (в уравнении прямой отсутствует свободный член), а другую точку берем на прямой произвольно, например, A (2; 3) на рис. 2.7, *б*. В качестве контрольной возьмем, например, точку B (1; 0). Самую "простую" точку O (0; 0) здесь в качестве контрольной брать не следует, ибо она лежит на построенной прямой. Так как координаты контрольной точки B (1; 0) удовлетворяют неравенству, т.е. $3 \cdot 1 - 2 \cdot 0 \geq 0$, то решением данного неравенства является нижняя (правая) полуплоскость, содержащая эту точку.▶

Учитывая, что множество точек, удовлетворяющих уравнению

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \quad \textbf{(2.4)}$$

при $n=3$, является *плоскостью*, а при $n>3$ — ее обобщением в n-мерном пространстве — *гиперплоскостью*, теорему 2.3 можно распространить на случай трех и более переменных.

Теорема 2.4. *Множество всех решений линейного неравенства с n переменными*

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1$$

является одним из полупространств, на которые все пространство делится плоскостью или гиперплоскостью (2.4), включая и эту плоскость (гиперплоскость).

Рассмотрим **множество решений систем неравенств.**

Теорема 2.5 *Множество решений совместной системы m линейных неравенств с двумя переменными*

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \leq b_1, \\ a_{21}x_1 + a_{22}x_2 \leq b_2, \\ \dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 \leq b_m \end{cases}$$

является выпуклым многоугольником (или выпуклой многоугольной областью).

◻ Каждое из неравенств в соответствии с теоремой 2.3 определяет одну из полуплоскостей, являющуюся выпуклым множеством точек. Множеством решений совместной системы линейных неравенств служат точки, которые принадлежат полуплоскостям решений всех неравенств, т.е. принадлежат их пересечению. Согласно теореме 2.2 о пересечении выпуклых множеств это множество является выпуклым и содержит конечное число угловых точек, т.е. является выпуклым многоугольником (выпуклой многоугольной областью). ■

▷ **2.5**. Построить множество решений системы неравенств

$$\begin{cases} -5x_1 + 4x_2 \le 20, & \text{(I)} \\ 2x_1 + 3x_2 \le 24, & \text{(II)} \\ x_1 - 3x_2 \le 3, & \text{(III)} \\ x_1 \ge 0, & \text{(IV)} \\ 0 \le x_2 \le 6. & \text{(V, VI)} \end{cases}$$

Р е ш е н и е. Для построения искомого множества решений системы неравенств строим последовательно множество решений каждого неравенства аналогично тому, как это делалось в задаче **2.4**. Рекомендуем после нахождения каждой полуплоскости и выделения ее соответствующей штриховкой находить последовательно их пересечение: сначала полуплоскостей решений первых двух неравенств (многоугольной области GFD на рис. 2.8), затем первых трех неравенств (треугольника GFD), потом — четырех неравенств (четырехугольника $HAFD$), далее — пяти неравенств (пятиугольника $OAFDE$) и, наконец, всех шести неравенств — выпуклого многоугольника $OABCDE$.

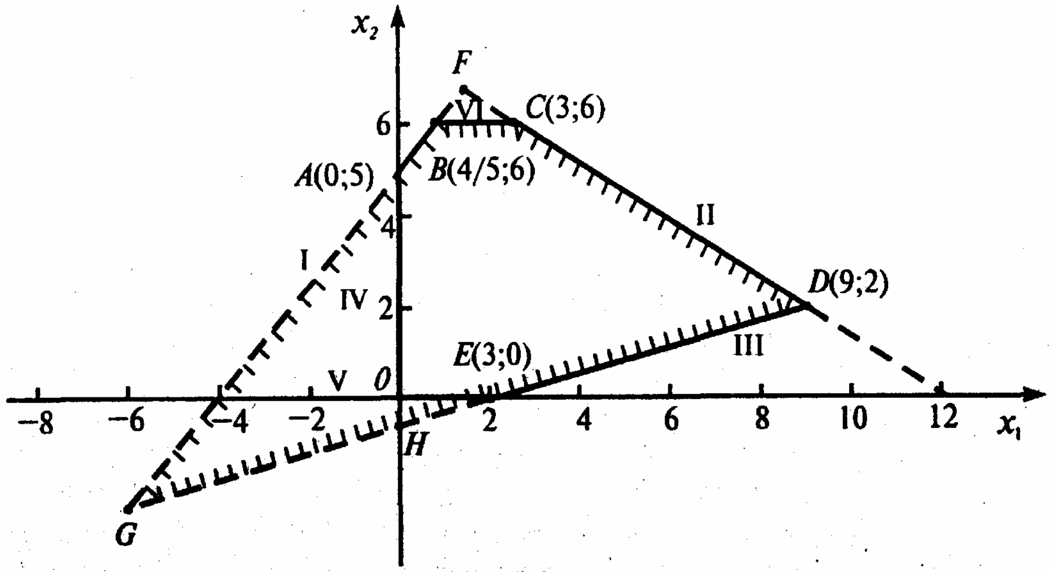


Рис. 2.8

Координаты угловых точек — вершин этого многоугольника найдем как координаты точек пересечения соответствующих прямых. Например, точка D является точкой пересечения прямых II и III, т.е. ее координаты являются решением системы

$$\begin{cases} 2x_1 + 3x_2 = 24, & \text{(II)} \\ x_1 - 3x_2 = 3, & \text{(III)} \end{cases}$$

откуда $x_1 = 9$, $x_2 = 2$, т.е. D (9;2). Аналогично находим координаты других угловых точек: O (0;0), A (0;5), B (4/5;6), C (3;6) E (3;0).▶

При построении областей решений систем неравенств могут встретиться и другие случаи: множество решений — выпуклая многоугольная область (рис. 2.9, *а*); одна точка (рис. 2.9, *б*); пустое множество, когда система неравенств несовместна (рис. 2.9, *в*).

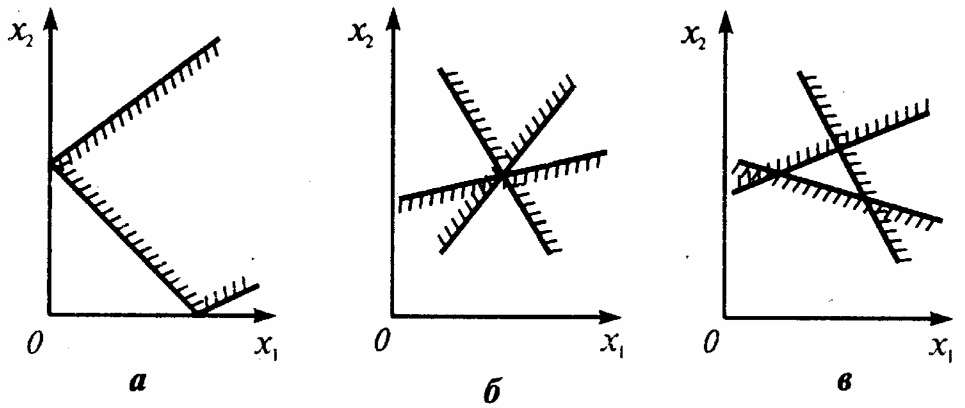


Рис. 2.9

Теорема 2.6. *Множество решений совместной системы m линейных неравенств с n переменными является выпуклым многогранником (выпуклой многогранной областью) в n-мерном пространстве.*

Рассмотрим **множество допустимых решений системы m линейных уравнений с n переменными.**

Теорема 2.7. *Множество всех допустимых решений совместной системы m линейных уравнений с n переменными ($m < n$) является выпуклым многогранником (выпуклой многогранной областью) в n-мерном пространстве.*

Доказательство этой теоремы приведено в гл. 3. Здесь же проиллюстрируем теорему на примерах.

▷ **2.6.** Построить множество допустимых решений:

а) уравнения $2x_1 + 3x_2 = 6$;

б) системы уравнений $\begin{cases} 2x_1 + 3x_2 + x_3 = 12, \\ x_1 + x_2 - x_4 = 1. \end{cases}$

Убедиться в справедливости теоремы 2.7.

Решение. а) Рассмотрим частный случай системы линейных уравнений ($m < n$), содержащей $n = 2$ переменных, т.е. состоящей из одного уравнения ($m = 1$). Множество всех решений данного уравнения есть прямая $2x_1 + 3x_2 = 6$, а множество допустимых решений (при дополнительном условии $x_1 \geq 0$, $x_2 \geq 0$) — точки отрезка AB (рис. 2.10, *а*), который можно рассматривать как частный случай выпуклого многогранника с двумя угловыми точками $A(3; 0)$ и $B(0; 2)$.

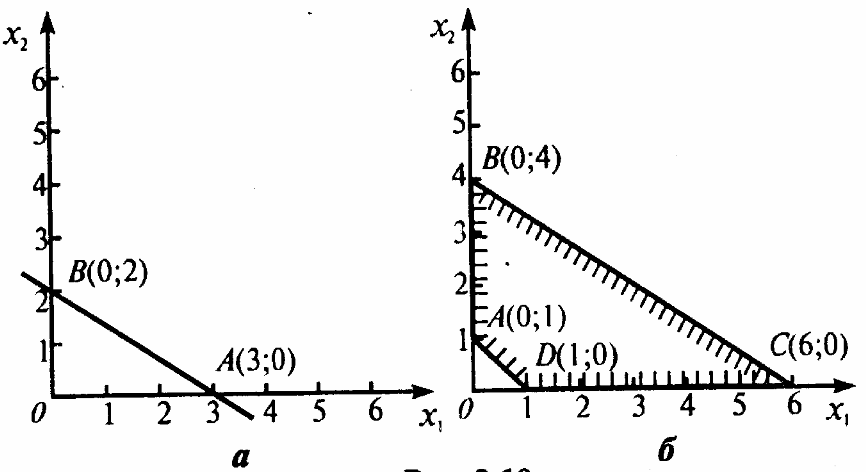


Рис. 2.10

б) Построить непосредственно множество решений системы уравнений с $n = 4$ ($n > 3$) переменными не представляется возможным. В данном случае (когда разность между числом переменных и уравнений $n - m = 2$) можно поступить так: разобьем все переменные на основные, например x_3 и x_4 (определитель из коэффициентов при них отличен от нуля), и неосновные (свободные) переменные x_1 и x_2, и вместо множества решений системы построим множество значений их неосновных переменных (выполнить это возможно, так как их всего две).

С этой целью выразим основные переменные через неосновные:

$$\begin{cases} x_3 = 12 - 2x_1 - 3x_2, \\ x_4 = -1 + x_1 + x_2. \end{cases}$$

Так как рассматриваются допустимые значения переменных, т.е. $x_1 \geq 0$, $x_2 \geq 0$, $x_3 \geq 0$, $x_4 \geq 0$, то

$$\begin{cases} 12 - 2x_1 - 3x_2 \geq 0, & \text{(I)} \\ -1 + x_1 + x_2 \geq 0, & \text{(II)} \\ x_1 \geq 0,\ x_2 \geq 0. & \text{(III,IV)} \end{cases}$$

Решениями полученной таким образом системы неравенств являются точки четырехугольника $ABCD$ на рис. 2.10, *б* с четырьмя угловыми точками $A(0;\ 1)$, $B(0;\ 4)$, $C(6;\ 0)$, $D(1;\ 0)$ (рекомендуем убедиться в этом самому читателю).

В данном примере графические построения проведены не в пространстве всех переменных, а в плоскости двух неосновных переменных x_1, x_2. Но так как любой паре неосновных переменных x_1, x_2 соответствуют определенные значения основных переменных x_3, x_4, а следовательно, одно и только одно решение данной системы уравнений, то каждой точке построенного четырехугольника $ABCD$ соответствует одна и только одна точка множества допустимых решений системы уравнений, представляющего в данном случае выпуклый многогранник в четырехмерном пространстве.▶

Между допустимыми базисными решениями и угловыми точками множества допустимых решений системы линейных уравнений существует взаимнооднозначное соответствие. Это утверждение будет доказано в гл. 3, здесь же ограничимся примером.

▷ **2.7.** Убедиться в том, что между базисными решениями систем, приведенных в задаче **2.6**, и угловыми точками множества их допустимых решений существует взаимнооднозначное соответствие.

Решение. а) Система, состоящая из одного уравнения, имеет два допустимых базисных решения. Первое базисное решение $X_1 = (3;\ 0)$ получается из уравнения, если в качестве основной взять переменную x_1, а неосновной — переменную $x_2 = 0$. Второе базисное решение $X_2 = (0;\ 2)$ получается, если основная переменная x_2, а неосновная — переменная $x_1 = 0$. Из рис. 2.10, *а* следует, что допустимым базисным решениям X_1 и X_2 однозначно соответствуют угловые точки отрезка AB — множества допустимых решений уравнения.

б) Для системы, приведенной в задаче **2.6**, *б*, можно получить четыре допустимых базисных решения (рекомендуем читателю найти их самостоятельно): $X_1 = (1;\ 0;\ 10;\ 0)$, $X_2 = (6;\ 0;\ 0;\ 5)$, $X_3 = (0;\ 1;\ 9;\ 0)$, $X_4 = (0;\ 4;\ 0;\ 3)$. Из рис. 2.10, *б* следует, что этим допустимым базисным решениям однозначно соответствуют точки $D(1;\ 0)$, $C(6;\ 0)$, $A(0;\ 1)$ и $B(0;\ 4)$ многоугольника $ABCD$ — множества допустимых решений системы уравнений.▶

В задачах **2.8** и **2.9** решить системы уравнений.

2.8. $\begin{cases} 2x_1 - 3x_2 + 4x_3 = 3, \\ 3x_1 - 4x_2 + 6x_3 = 5. \end{cases}$ **2.9** $\begin{cases} 3x_1 + x_2 - x_3 - 2x_4 = -4, \\ x_1 + x_2 - x_3 + 2x_4 = 1. \end{cases}$

В задачах **2.10** и **2.11** найти базисные решения.

2.10. $\begin{cases} x_1 + 2x_2 - x_3 = 5, \\ 2x_1 - x_2 - 3x_3 = -4. \end{cases}$ **2.11.** $\begin{cases} x_1 + x_2 + x_3 + x_4 = 2, \\ 2x_1 + 2x_2 - x_3 + 2x_4 = -2, \\ x_1 - x_2 - x_4 = 2. \end{cases}$

В задачах **2.12** и **2.13** построить множество решений неравенства.

2.12. $4x_1 - 5x_2 + 20 \le 0.$ **2.13.** $4x_1 - 3x_2 \ge 0.$

В задачах **2.14** и **2.15** построить множества решений системы неравенств и найти их угловые точки.

2.14. $\begin{cases} 5x_1 - 3x_2 + 15 \ge 0, \\ 0 \le x_2 \le 10, \\ x_1 + x_2 - 17 \le 0, \\ 0 \le x_1 \le 11. \end{cases}$ **2.15.** $\begin{cases} -x_1 + x_2 \le 1, \\ x_1 + x_2 \ge 1, \\ x_1 - x_2 \le 2, \\ 2x_1 - x_2 \le 0, \\ x_2 \ge 0. \end{cases}$

В задачах **2.16** и **2.17** построить множества допустимых решений уравнений.

2.16. $3x_1 + 5x_2 = 15.$ **2.17.** $2x_1 - 3x_2 = 0.$

Глава 3. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ МЕТОДОВ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Для рассмотрения теоретических основ методов линейного программирования целесообразно вновь вернуться к понятию выпуклого множества точек, дав ему более строгое определение в аналитической форме.

3.1. Выпуклые множества в n-мерном пространстве

В разд. 2.2 выпуклое множество точек определялось как множество, которое вместе с любыми своими двумя точками содержит весь отрезок, их соединяющий. Однако в случае n переменных не ясно, что следует понимать под "отрезком" в n-мерном пространстве. Очевидно, надо дать аналитическое определение этого понятия.

Начнем с $n = 2$ (двумерного пространства, плоскости). Пусть $X_1 = \left(x_1^{(1)}, x_2^{(1)}\right)$ и $X_2 = \left(x_1^{(2)}, x_2^{(2)}\right)$ – точки плоскости Ox_1x_2, а $X = (x_1, x_2)$ — любая точка отрезка X_1X_2 (рис. 3.1). Очевидно, что отношение α длин отрезков XX_2 и X_1X_2 удовлетворяет условию $0 \le \alpha \le 1$. Запишем это отношение α через координаты точек. Получим

$$\alpha = \frac{x_1^{(2)} - x_1}{x_1^{(2)} - x_1^{(1)}} = \frac{x_2^{(2)} - x_2}{x_2^{(2)} - x_2^{(1)}},$$

откуда

$$\begin{cases} x_1 = \alpha x_1^{(1)} + (1-\alpha)x_1^{(2)}, \\ x_2 = \alpha x_2^{(1)} + (1-\alpha)x_2^{(2)}, \end{cases} \tag{3.1}$$

где

$$0 \le \alpha \le 1. \tag{3.2}$$

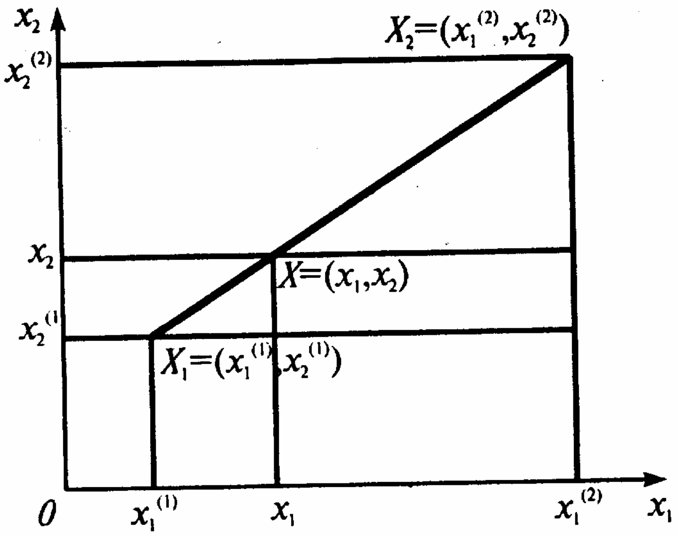


Рис. 3.1

Полагая $\alpha_1 = \alpha$ и $\alpha_2 = 1-\alpha$, условия (3.1), (3.2) примут вид

$$\begin{cases} x_1 = \alpha_1 x_1^{(1)} + \alpha_2 x_1^{(2)}, \\ x_2 = \alpha_1 x_2^{(1)} + \alpha_2 x_2^{(2)}, \end{cases} \tag{3.3}$$

$$\alpha_1 \ge 0, \;\; \alpha_2 \ge 0, \qquad \alpha_1 + \alpha_2 = 1. \tag{3.4}$$

Равенство (3.3) можно записать в виде

$$X = \alpha_1 X_1 + \alpha_2 X_2, \tag{3.5}$$

понимая, что в нем все операции выполняются покоординатно (т.е. отдельно по переменной x_1 и отдельно по переменной x_2).

Таким образом, ***отрезок*** X_1X_2 ***можно определить как множество точек (векторов), удовлетворяющих условиям (3.5) и (3.4).***

В случае n-мерного пространства определение отрезка будет таким же — множество точек, удовлетворяющих условиям (3.5) и

(3.4), если под X_1 *и* X_2 подразумевать точки (векторы) n-мерного пространства: $X_1 = (x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$ и $X_2 = (x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)})$.

Обобщением понятия отрезка для нескольких точек является их выпуклая линейная комбинация.

*Точка X называется **выпуклой линейной комбинацией** точек X_1, X_2, ..., X_n, если выполняются условия*

$$X = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n,$$

$$\alpha_j \ge 0 \ (j = 1, 2, \ldots, n), \quad \sum_{j=1}^{n} \alpha_j = 1.$$

Так, например, выражение $(1/6)X_1 + (1/2)X_2 + (1/3)X_3$ есть выпуклая линейная комбинация точек X_1, X_2, X_3, а выражения $(1/3)X_1 + (1/2)X_2 + (1/3)X_3$ или $(1/3)X_1 - (1/2)X_2 + (7/6)X_3$ являются линейными, но не выпуклыми комбинациями тех же точек (в первом $\frac{1}{3} + \frac{1}{2} + \frac{1}{3} \ne 1$, а во втором $\alpha_2 = -\frac{1}{2} < 0$).

Очевидно, что в частном случае при $n = 2$ выпуклой линейной комбинацией двух точек является соединяющий их отрезок. Поэтому *множество точек является **выпуклым**, если оно вместе с любыми своими двумя точками содержит их произвольную выпуклую линейную комбинацию.*

Рассмотрим теорему о представлении выпуклого многогранника.

Теорема 3.1. *Выпуклый n-мерный многогранник является выпуклой линейной комбинацией своих угловых точек.*

□ Возьмем для простоты $n = 2$, а в качестве многогранника — треугольник $X_1X_2X_3$ (рис.3.2). Через произвольную точку X треугольника проведем отрезок X_1X_4. Поскольку точка X лежит на этом отрезке, то

$$X = \alpha_1 X_1 + \alpha_4 X_4,$$

где

$$\alpha_1 \ge 0, \alpha_2 \ge 0, \qquad \alpha_1 + \alpha_4 = 1.$$

Точка X_4 лежит на отрезке X_2X_3, следовательно, $X_4 = \alpha_2 X_2 + \alpha_3 X_3$, где $\alpha_2 \ge 0, \alpha_3 \ge 0$, $\alpha_2 + \alpha_3 = 1$.

Подставив значение X_4 в выражение для X, получим

$$X = \alpha_1 X_1 + \alpha_4 (\alpha_2 X_2 + \alpha_3 X_3) = \alpha_1 X_1 + \alpha_2 \alpha_4 X_2 + \alpha_3 \alpha_4 X_3.$$

Обозначив $t_1 = \alpha_1$, $t_2 = \alpha_2\alpha_4$, $t_3 = \alpha_3\alpha_4$, получим окончательно

$$X = t_1X_1 + t_2X_2 + t_3X_3,$$

где

$$t_1 \geq 0, t_2 \geq 0, t_3 \geq 0 \text{ и } t_1 + t_2 + t_3 = 1.$$

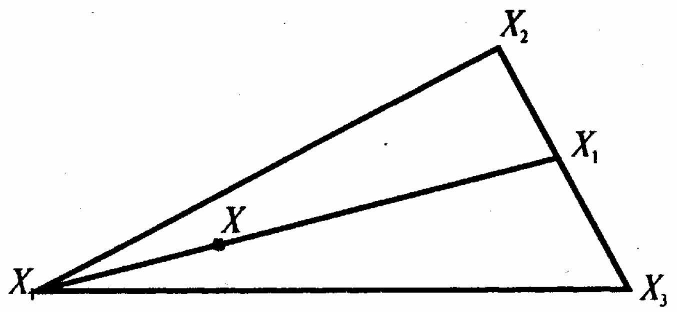


Рис. 3.2

Таким образом точка X есть выпуклая линейная комбинация угловых точек (вершин) треугольника $X_1X_2X_3$. ■

Из теоремы 3.1 следует, что выпуклый многогранник порождается своими угловыми точками или вершинами: отрезок — двумя точками, треугольник — тремя, тетраэдр — четырьмя точками и т.д. В то же время выпуклая многогранная область, являясь неограниченным множеством, не определяется однозначно своими угловыми точками: любую ее точку нельзя представить в виде выпуклой линейной комбинации угловых точек.

3.2. Свойства задачи линейного программирования

В разд. 1.3 были рассмотрены различные формы задачи линейного программирования и показано, что любая задача линейного программирования может быть представлена в виде общей, канонической или стандартной задачи.

В данном разделе будем рассматривать каноническую задачу (1.20) — (1.22), в которой система ограничений есть система уравнений (2.1).

Для обоснования свойств задачи линейного программирования и методов ее решения целесообразно рассмотреть еще два вида записи канонической задачи.

Матричная форма записи:

$$F = CX \to \max\ (\min) \tag{3.6}$$

при ограничениях

$$AX = B, \tag{3.7}$$

$$X \geq \mathbf{0}, \tag{3.8}$$

где

$$C = (c_1, c_2, \dots, c_n);\ A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix};\ X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix};\ B = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}.$$

Здесь C — матрица-строка, A — матрица системы, X — матрица-столбец переменных, B — матрица-столбец свободных членов.

Векторная форма записи:

$$F = CX \to \max\ (\min) \tag{3.9}$$

при ограничениях

$$P_1x_1 + P_2x_2 + \dots + P_nx_n = P, \tag{3.10}$$

$$X \geq \mathbf{0}, \tag{3.11}$$

где CX — скалярное произведение векторов[1] $C = (c_1, c_2, \dots, c_n)$ и $X = (x_1, x_2, \dots, x_n)$, векторы

$$P_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix},\ P_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{pmatrix},\ \dots,\ P_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{pmatrix},\ P = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}$$

состоят соответственно из коэффициентов при переменных и свободных членов.

Векторное неравенство $X \geq \mathbf{0}$ означает, что все компоненты вектора X неотрицательны, т.е. $x_j \geq 0$, $j = 1, 2, \dots, n$.

---

[1] ***Скалярным произведением** CX **двух векторов*** $C = (c_1, c_2, \dots, c_n)$ **и** $X = (x_1, x_2, \dots, x_n)$ **называется число, равное сумме произведений соответствующих координат этих векторов:** $CX = c_1x_1 + c_2x_2 + \dots + c_nx_n$.

В гл. 2 была сформулирована, но не доказана в общем виде следующая теорема.

Теорема 3.2. *Множество всех допустимых решений системы ограничений задачи линейного программирования является выпуклым.*

□ Пусть $X_1 = (x_1^{(1)}, x_2^{(1)}, \ldots, x_n^{(1)})$ и $X_2 = (x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)})$ — два допустимых решения задачи (3.6) — (3.8), заданной в матричной форме. Тогда $AX_1 = B$ и $AX_2 = B$. Рассмотрим выпуклую линейную комбинацию решений X_1 и X_2, т.е.

$$X = \alpha_1 X_1 + \alpha_2 X_2 \text{ при } \alpha_1 \geq 0, \alpha_2 \geq 0 \text{ и } \alpha_1 + \alpha_2 = 1,$$

и покажем, что она также является допустимым решением системы (3.7). В самом деле

$$AX = A(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 AX_1 + (1 - \alpha_1)AX_2 = \alpha_1 B + (1 - \alpha_1)B = B,$$

т.е. решение X удовлетворяет системе (3.7). Но так как $X_1 \geq \mathbf{0}$, $X_2 \geq \mathbf{0}$, $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, то и $X \geq \mathbf{0}$, т.е. решение X удовлетворяет и условию (3.8).■

Итак, доказано, что множество всех допустимых решений задачи линейного программирования является выпуклым, а точнее (если учесть изложенное в гл. 2), представляет выпуклый многогранник или выпуклую многогранную область, которые в дальнейшем будем называть одним термином — *многогранником решений.*

Ответ на вопрос, в какой точке многогранника решений возможно оптимальное решение задачи линейного программирования, дается в следующей фундаментальной теореме.

Теорема 3.3. *Если задача линейного программирования имеет оптимальное решение, то линейная функция принимает максимальное*[1] *значение в одной из угловых точек многогранника решений. Если линейная функция принимает максимальное значение более чем в одной угловой точке, то она принимает его в любой точке, являющейся выпуклой линейной комбинацией этих точек.*

□ Будем полагать, что многогранник решений является ограниченным. Обозначим его угловые точки через X_1, X_2, ..., X_p, а

[1] Формулировка теоремы остается такой же и при отыскании минимального значения линейной функции.

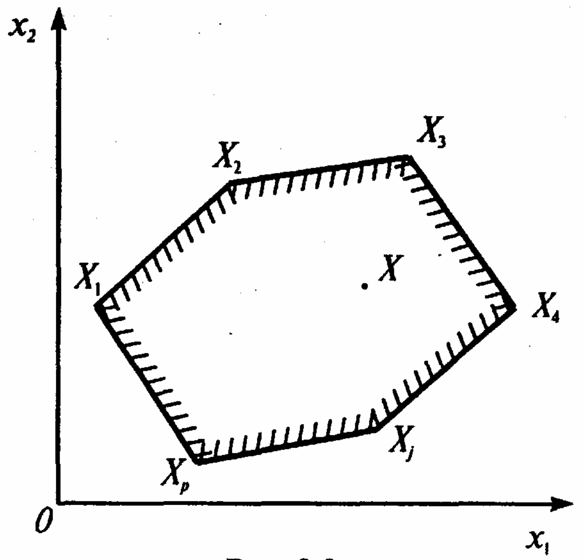


Рис. 3.3

оптимальное решение — через X^* (рис.3.3). Тогда $F(X^*) \geq F(X)$ для всех точек X многогранника решений. Если X^* — угловая точка, то первая часть теоремы доказана.

Предположим, что X^* не является угловой точкой, тогда на основании теоремы 3.1 X^* можно представить как выпуклую линейную комбинацию угловых точек многогранника решений, т.е.

$$X^* = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_p X_p,$$

$$\alpha_j \geq 0,\ j = 1,2,\ldots,p;\ \sum_{j=1}^{n} \alpha_j = 1.$$

Так как $F(X)$ — линейная функция, получаем

$$\begin{aligned} F(X^*) &= F(\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_p X_p) = \\ &= \alpha_1 F(X_1) + \alpha_2 F(X_2) + \ldots + \alpha_p F(X_p). \end{aligned} \qquad (3.12)$$

В этом разложении среди значений $F(X_j)$ ($j = 1, 2, \ldots, p$) выберем максимальное. Пусть оно соответствует угловой точке X_k $(1 \leq k \leq p)$; обозначим его через M, т.е. $F(X_k) = M$. Заменим в выражении (3.12) каждое значение этим максимальным значением M. Тогда, учитывая, что $\alpha_j \geq 0$, $\sum_{j=1}^{p} \alpha_j = 1$, найдем $F(X^*) \leq$ $\leq \alpha_1 M + \alpha_2 M + \ldots + \alpha_p M = M \sum_{j=1}^{p} \alpha_j = M$. По предположению X^* — оптимальное решение, поэтому, с одной стороны, $F(X^*) \geq F(X_k) = M$, но, с другой стороны, доказано, что $F(X^*) \leq M$, следовательно, $F(X^*) = M = F(X_k)$, где X_k — угловая точка. Итак, существует угловая точка X_k, в которой линейная функция принимает максимальное значение.

Для доказательства второй части теоремы допустим, что $F(X)$ принимает максимальное значение более чем в одной угловой точке, например, в точках X_1, X_2, ..., X_q. где $1 \le q \le p$; тогда $F(X_1) = F(X_2) = \ldots = F(X_q) = M$.

Пусть X — выпуклая линейная комбинация этих угловых точек, т.е.

$$X = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_q X_q,\ \alpha_j \ge 0\ (j = 1, 2, \ldots, q),\ \sum_{j=1}^{q} \alpha_j = 1.$$

В этом случае, учитывая, что функция $F(X)$ — линейная, получим

$$F(X) = F(\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_q X_q) = \alpha_1 F(X_1) + \alpha_2 F(X_2) + \ldots$$

$$\ldots + \alpha_q F(X_q) = \alpha_1 M + \alpha_2 M + \ldots + \alpha_q M = M \sum_{j=1}^{q} \alpha_j = M,$$

т.е. линейная функция F принимает максимальное значение в произвольной точке X, являющейся выпуклой линейной комбинацией угловых точек X_1, X_2, ..., X_q.■

Замечание. Требование ограниченности многогранника решений в теореме является существенным, так как в случае неограниченной многогранной области, как отмечалось в теореме 3.1, не каждую точку такой области можно представить выпуклой линейной комбинацией ее угловых точек.

Доказанная теорема является фундаментальной, так как она указывает принципиальный путь решения задач линейного программирования. Действительно, согласно этой теореме вместо исследования бесконечного множества допустимых решений для нахождения среди них искомого оптимального решения необходимо исследовать лишь конечное число угловых точек многогранника решений.

Следующая теорема посвящена аналитическому методу нахождения угловых точек.

Теорема 3.4. ***Каждому допустимому базисному решению задачи линейного программирования соответствует угловая точка много-***

гранника решений, и наоборот, каждой угловой точке многогранника решений соответствует допустимое базисное решение.

□ Пусть $X = (x_1, x_2, ..., x_m; 0, 0, ..., 0)$ — допустимое базисное решение системы ограничений (3.10) задачи, в котором первые m компонент — основные переменные, а остальные $n - m$ компонент — неосновные переменные, равные нулю в базисном решении (если это не так, то соответствующие переменные можно перенумеровать). Покажем, что X — угловая точка многогранника решений.

Предположим противное, т.е. что X не является угловой точкой. Тогда точку X можно представить внутренней точкой отрезка, соединяющего две различные, не совпадающие с X, точки

$$X_1 = (x_1^{(1)}, x_2^{(1)}, \ldots, x_m^{(1)};\ 0,\ 0, \ldots, 0) \text{ и}$$

$$X_2 = (x_1^{(2)}, x_2^{(2)}, \ldots, x_m^{(2)};\ 0,\ 0, \ldots, 0),$$

другими словами, — выпуклой линейной комбинацией точек X_1 и X_2 многогранника решений, т.е.

$$X = \alpha_1 X_1 + \alpha_2 X_2, \tag{3.13}$$

где $\alpha_1 > 0$, $\alpha_2 > 0$, $\alpha_1 + \alpha_2 = 1$ (полагаем, что $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, ибо в противном случае точка X совпадает с точкой X_1 или X_2).

Запишем векторное равенство (3.13) в координатной форме:

$$\begin{cases} x_1 = \alpha_1 x_1^{(1)} + \alpha_2 x_1^{(2)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_1 = \alpha_1 x_m^{(1)} + \alpha_2 x_m^{(2)}, \\ 0 = \alpha_1 x_{m+1}^{(1)} + \alpha_2 x_{m+1}^{(2)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ 0 = \alpha_1 x_n^{(1)} + \alpha_2 x_n^{(2)}. \end{cases}$$

Поскольку $x_j^{(1)} \geq 0$, $x_j^{(2)} \geq 0$ $(j = 1,2,\ldots,n)$, $\alpha_1 > 0$, $\alpha_2 > 0$, то из последних $n-m$ равенств следует, что $x_{m+1}^{(1)} = 0$, $x_{m+1}^{(2)} = 0$, ..., $x_n^{(1)} = 0$, $x_n^{(2)} = 0$, т.е. в решениях X_1, X_2 и X системы уравнений (3.10) значения $n-m$ компонент равны в данном случае нулю. Эти компоненты можно считать значениями неосновных переменных. Но значения неосновных переменных однозначно определяют значения основных, следовательно, $x_1^{(1)} = x_1^{(2)} = x_1, \ldots, x_m^{(1)} = x_m^{(2)} = x_2$. Таким образом, все n компонент в решениях X, X_1 и X_2 совпадают, и значит, точки X_1 и X_2 сливаются, что противоречит допущению. Следовательно, X — угловая точка многогранника решений.

Докажем обратное утверждение. Пусть $X = (x_1, x_2, \ldots, x_m; 0, 0, \ldots, 0)$ — угловая точка многогранника решений и первые ее m координат положительны. Покажем, что X — допустимое базисное решение.

Если векторы $P_1, P_2, \ldots, P_m$ линейно независимы[1], то ранг r матрицы A, составленной из компонент этих векторов, равен m, т.е. определитель $|A| \neq 0$, следовательно, переменные $x_1, x_2, \ldots, x_m$ являются основными, и решение $X = (x_1, x_2, \ldots, x_m; 0, 0, \ldots, 0)$ — базисное, допустимое, т.е. утверждение доказано.

Предположим противное, т.е. векторы $P_1, P_2, \ldots, P_m$ линейно зависимы; тогда в равенстве

$$\alpha_1 P_1 + \ldots + \alpha_m P_m + \ldots + \alpha_n P_n = \mathbf{0} \qquad \textbf{(3.14)}$$

хотя бы один из коэффициентов $\alpha_1, \ldots, \alpha_m, \ldots, \alpha_n$ отличен от нуля.

Умножим почленно равенство (3.14) на множитель $\mu > 0$:

$$\mu\alpha_1 P_1 + \ldots + \mu\alpha_m P_m + \ldots + \mu\alpha_n P_n = \mathbf{0}. \qquad \textbf{(3.15)}$$

Подставив координаты угловой точки X многогранника решений в систему ограничений (3.10), получим

$$P_1 x_1 + \ldots + P_m x_m + \ldots + P_n x_n = P. \qquad \textbf{(3.16)}$$

Равенство (3.15) почленно сложим с равенством (3.16), а затем вычтем его из равенства (3.16). Получим

$$P_1(x_1 + \mu\alpha_1) + \ldots + P_m(x_m + \mu\alpha_m) + \ldots + P_n(x_n + \mu\alpha_n) = P. \qquad \textbf{(3.17)}$$

$$P_1(x_1 - \mu\alpha_1) + \ldots + P_m(x_m - \mu\alpha_m) + \ldots + P_n(x_1 - \mu\alpha_n) = P. \qquad \textbf{(3.18)}$$

---

[1] Векторы $P_1, P_2, \ldots, P_n$ (они же столбцы матрицы A) называются *линейно зависимыми*, если можно подобрать такие числа $\alpha_1, \alpha_2, \ldots, \alpha_n$, не равные нулю одновременно, что $\alpha_1 P_1 + \alpha_2 P_2 + \ldots + \alpha_n P_n = \mathbf{0}$, где $\mathbf{0}$ — нулевой вектор (состоящий из нулей). В противном случае векторы $P_1, P_2, \ldots, P_n$ называются *линейно независимыми*.

Сравнивая полученные равенства (3.17), (3.18) с равенством (3.16), заключаем, что при любом μ системе ограничений (3.10) удовлетворяют решения $X_1 = (x_1 + \mu\alpha_1, \ ..., \ x_m + \mu\alpha_m; \ 0, 0, ..., 0)$ и $X_2 = (x_1 - \mu\alpha_1, \ \dots, \ x_m - \mu\alpha_m; 0, 0, ..., 0)$.

Поскольку $x_j \geq 0$ $(j = 1, 2, ..., n)$, то можно подобрать μ настолько малым, что все компоненты решений X_1 и X_2 будут неотрицательными. В результате X_1 и X_2 будут различными допустимыми решениями задачи (**3.9**) — (**3.11**). При этом, как легко видеть, решение $\frac{1}{2}\left(X_1 + X_2\right) = (x_1, x_2, \dots, x_m; \ 0, 0, ..., 0) = X$, т.е. точка X лежит на отрезке (в данном случае в его середине), расположенном в многограннике решений. Значит X не является угловой точкой, что противоречит условию. Следовательно, наше допущение неверно, т.е. векторы P_1, P_2, ..., P_m линейно независимы и X — допустимое базисное решение задачи (3.9) — (3.11).■

Из теорем 3.3 и 3.4 непосредственно вытекает важное следствие: *если задача линейного программирования имеет оптимальное решение, то оно совпадает, по крайней мере, с одним из ее допустимых базисных решений.*

Итак, *оптимум линейной функции задачи линейного программирования следует искать среди конечного числа ее допустимых базисных решений.*

Глава 4. ГЕОМЕТРИЧЕСКИЙ МЕТОД РЕШЕНИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

В гл. 2 и 3 было доказано, что множество допустимых решений (многогранник решений) задачи линейного программирования представляет собой выпуклый многогранник (или выпуклую многогранную область), а оптимальное решение задачи находится, по крайней мере, в одной из угловых точек многогранника решений.

Рассмотрим задачу в стандартной форме (1.4)—(1.6) с двумя переменными ($n = 2$). К такой форме может быть сведена и каноническая задача (с ограничениями в виде уравнений), когда число переменных n больше числа уравнений m на 2, т.е. $n - m = 2$.

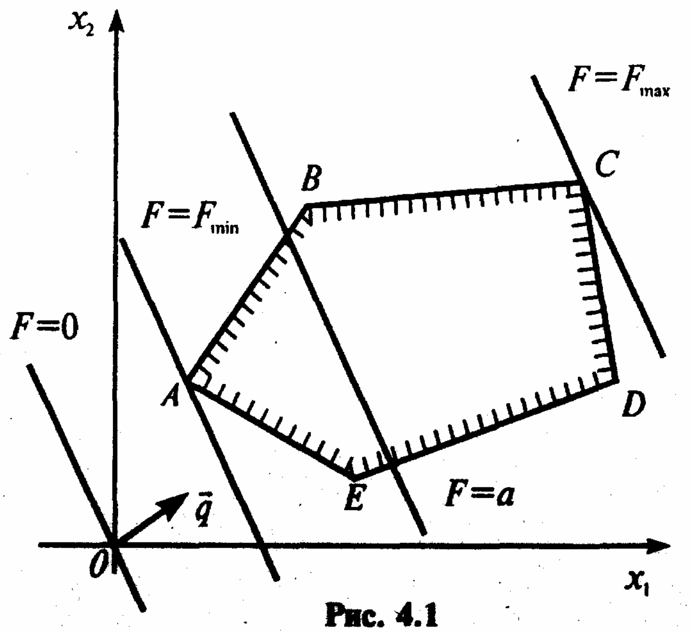


Рис. 4.1

Пусть геометрическим изображением системы ограничений является многоугольник $ABCDE$ (рис. 4.1). Необходимо среди точек этого многоугольника найти такую точку, в которой линейная функция $F = c_1x_1 + c_2x_2$ принимает максимальное (или минимальное) значение.

Рассмотрим так называемую *линию уровня* линейной функции F, т.е. линию, вдоль которой эта функция принимает одно и то же фиксированное значение a, т.е. $F = a$, или

$$c_1x_1 + c_2x_2 = a. \tag{4.1}$$

Линии уровня широко используются, например, на картах прогноза погоды, где извилистые линии — так называемые изотермы есть не что иное, как линии уровня температуры $T = c$. Еще более простым примером линий уровня являются параллели на географической карте. Это линии уровня широты.

Предположим, надо найти самую северную точку какой-либо области, например страны или материка. Это будет точка, имеющая наибольшую широту, т.е. точка, через которую проходит параллель (линия уровня) с самой большой широтой (уровнем).

Именно так и надо поступать при геометрическом решении задач линейного программирования. На многоугольнике решений следует найти точку, через которую проходит линия уровня функции F с наибольшим (если линейная функция максимизируется) или наименьшим (если она минимизируется) уровнем.

Уравнение линии уровня функции (4.1) есть уравнение прямой линии. При различных уровнях a линии уровня параллельны, так как их угловые коэффициенты определяются только соотношением между коэффициентами c_1 и c_2 и, следовательно, равны. Таким образом, линии уровня функции F — это своеобразные "параллели", расположенные обычно под углом к осям координат.

Важное свойство линии уровня линейной функции состоит в том, что при параллельном смещении линии в одну сторону уровень только возрастает, а при смещении в другую сторону — только убывает.

Пусть имеются три линии уровня

$$F = c_1x_1 + c_2x_2 = a_1, \quad \text{(I)}$$
$$F = c_1x_1 + c_2x_2 = a_2, \quad \text{(II)}$$
$$F = c_1x_1 + c_2x_2 = a_3, \quad \text{(III)}$$

причем линия II заключена между линиями I и III. Тогда $a_1 < a_2 < a_3$ или $a_1 > a_2 > a_3$.

В самом деле, на штриховой линии (перпендикулярной к линиям уровня на рис. 4.2) уровень является линейной функцией, а значит, при смещении в одном из направлений возрастает, а в другом — убывает.

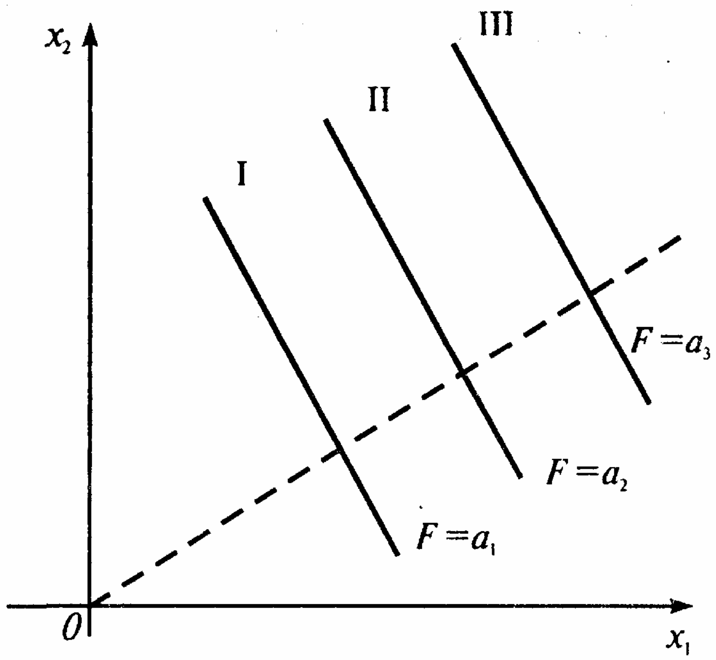


Рис. 4.2

Для определения направления возрастания рекомендуется изобразить две линии уровня и определить, на которой из них уровень больше. Например, одну из линий можно взять проходящей через начало координат (если линейная функция имеет вид $F = c_1x_1 + c_2x_2$, т.е. без свободного члена, то это соответствует нулевому уровню). Другую линию можно провести произвольно, так, например, чтобы она проходила через множество решений системы ограничений. Далее, определив направление возрастания линейной функции (обозначим его вектором $\vec{q}$), найдем точку, в которой функция принимает максимальное или минимальное значение, подобно тому как на карте находится самая северная или самая южная точка (на рис. 4.1 — это точка *C* или *A*).

▷ **4.1.** Решить геометрически 1-ю задачу из разд. 1.2:

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

при ограничениях:

$$\begin{cases} x_1 + 3x_2 & \le 18, (\mathrm{I}) \\ 2x_1 + x_2 & \le 16, (\mathrm{II}) \\ \quad\quad x_2 & \le 5, \ (\mathrm{III}) \\ 3x_1 & \le 21, (\mathrm{IV}) \\ x_1 \ge 0, \ x_2 \ge 0. & (\mathrm{V}, \mathrm{VI}) \end{cases}$$

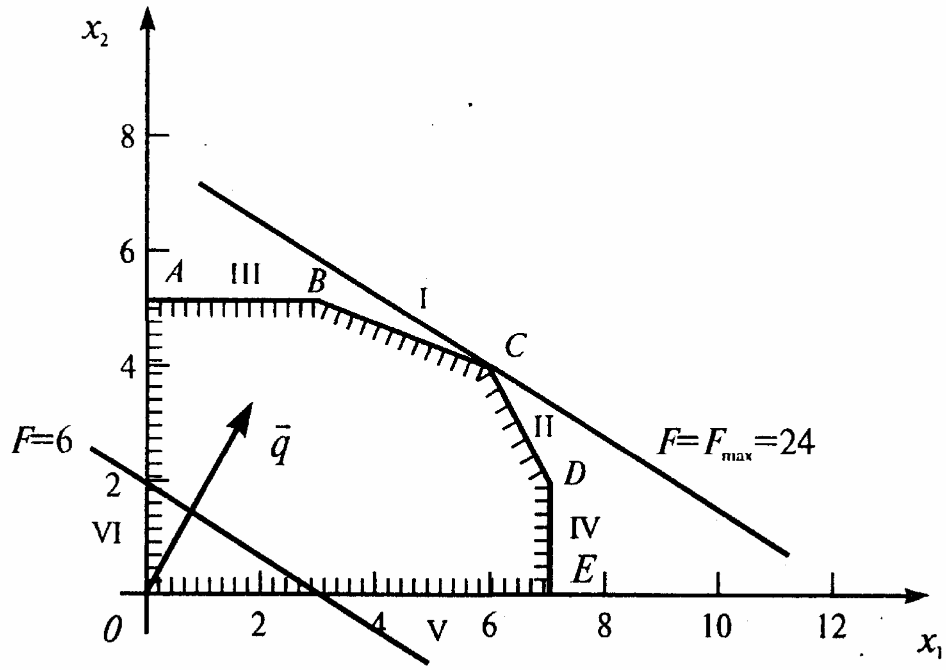


Рис. 4.3

Решение. Изобразим многоугольник решений (аналогично тому, как в задаче **2.5**) на рис. 4.3. Очевидно, что при $F = 0$ линия уровня $2x_1 + 3x_2 = 0$ проходит через начало координат (строить ее не обязательно). Зададим, например, $F = 6$ и построим линию уровня $2x_1 + 3x_2 = 6$. Ее расположение указывает на направление возрастания линейной функции (вектор $\vec{q}$). Так как рассматриваемая задача — на отыскание максимума, то оптимальное решение — в угловой точке C, находящейся на пересечении прямых I и II, т.е. координаты точки C определяются решением системы уравнений $\begin{cases} x_1 + 3x_2 = 18, \\ 2x_1 + x_2 = 16, \end{cases}$ откуда $x_1 = 6$, $x_2 = 4$, т.е. $C(6;4)$.

Максимум (максимальное значение) линейной функции равен $F_{\max} = 2 \cdot 6 + 3 \cdot 4 = 24$.

Итак, $F_{\max} = 24$ при оптимальном решении $x_1 = 6$, $x_2 = 4$, т.е. максимальная прибыль в 24 руб.[1] может быть достигнута при производстве 6 единиц продукции P_1 и 4 единиц продукции P_2.▶

▷ **4.2.** Решить геометрически 2-ю задачу из разд. 1.2:

$$F = 4x_1 + 6x_2 \to \min$$

при ограничениях:

$$\begin{cases} 3x_1 + x_2 \geq 9, & \text{(I)} \\ x_1 + 2x_2 \geq 8, & \text{(II)} \\ x_1 + 6x_2 \geq 12, & \text{(III)} \end{cases}$$

$$x_1 \geq 0,\ x_2 \geq 0. \quad \text{(IV, V)}$$

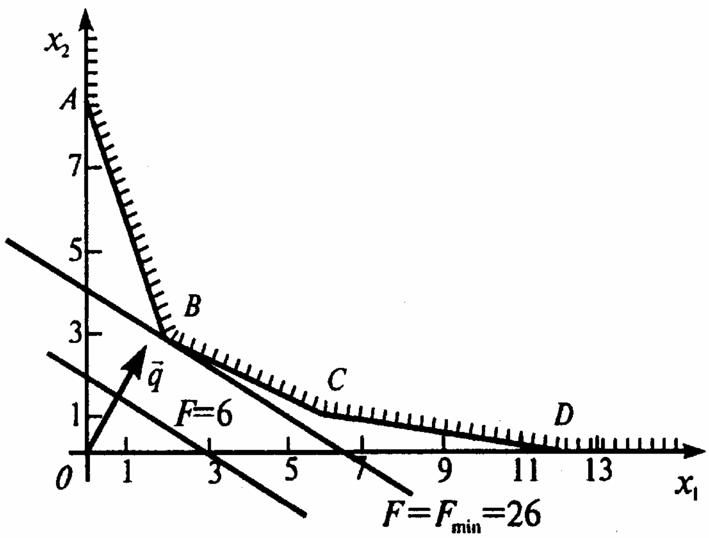


Рис. 4.4

Р е ш е н и е. Многоугольник решений представляет собой неограниченную многоугольную область (рис. 4.4). По расположению

---

[1] Напоминаем, что в задаче все цифры условные.

линии уровня, например, $F = 12$ или $4x_1 + 6x_2 = 12$, находим направление вектора $\vec{q}$ (этот вектор указывает на направление возрастания линейной функции). Очевидно, что точка минимума — это точка B "входа" в многоугольник решений, ибо при дальнейшем перемещении линии уровня в направлении вектора $\vec{q}$ значения линейной функции увеличиваются.

Находим координаты точки B (2;3), при этом $F_{\min} = 4 \cdot 2 + 6 \cdot 3 = 26$.

Итак, $F_{\min} = 26$ при оптимальном решении $x_1 = 2$, $x_2 = 3$, т.е. минимальная стоимость рациона 26 руб., если в него включить 2 единицы корма I и 3 единицы корма II. ▶

В рассмотренных задачах максимум и минимум линейной функции достигался в одной точке, так что задачи имели единственное оптимальное решение. На практике нередко встречаются задачи, которые этим условиям не удовлетворяют. В подобных случаях геометрический метод также позволяет получить ответ.

▷ **4.3.** Решить геометрически следующие задачи:

а) $F = 3x_1 + 3x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 + x_2 \le 8, & \text{(I)} \\ 2x_1 - x_2 \ge 1, & \text{(II)} \\ x_1 - 2x_2 \le 2, & \text{(III)} \end{cases}$$

$x_1 \ge 0,\ x_2 \ge 0;\ \text{(IV, V)}$

б) $F = 2x_1 - 3x_2 + 1 \to \min$

при ограничениях:

$$\begin{cases} x_1 + x_2 \ge 4, & \text{(I)} \\ 2x_1 - x_2 \ge 1, & \text{(II)} \\ x_1 - 2x_2 \le 1, & \text{(III)} \end{cases}$$

$x_1 \ge 0,\ x_2 \ge 0;\ \text{(IV, V)}$

Решение. **а)** Геометрическое решение задачи показано на рис. 4.5, *а*, из которого следует, что линия уровня с максимальным уровнем совпадает с граничной линией AB многоугольника решений $ABCD$, т.е. с линией $x_1 + x_2 = 8$. Следовательно, на всем отрезке AB линейная функция $F = 3x_1 + 3x_2$ принимает одно и то же максимальное значение, равное $3(x_1 + x_2) = 3 \cdot 8 = 24$. Это означает, что задача имеет бесконечно много оптимальных решений (их задают координаты точек отрезка AB), среди которых базисных оптимальных решений два — соответственно в угловых точках $A(3;\ 5)$ и $B(6;\ 2)$. Точки отрезка AB задаются уравнением $x_2 = 8 - x_1$, где $3 \le x_1 \le 6$.

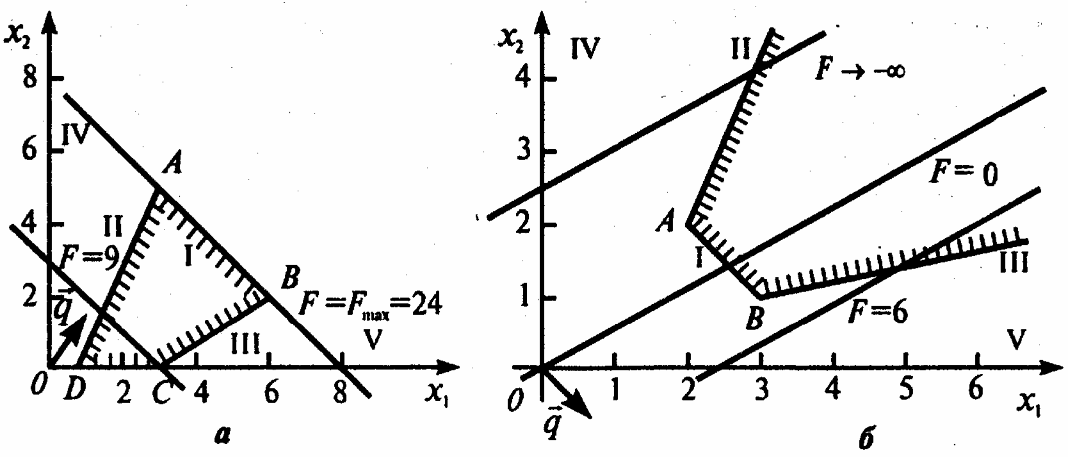


Рис. 4.5

Итак, $F_{max} = 24$ при бесконечном множестве оптимальных решений $x_1 = c$, $x_2 = 8 - c$, где $3 \le c \le 6$.

Замечание. При геометрическом решении подобных задач важно точно установить, действительно ли совпадает линия уровня с границей многоугольника решений или это связано с неточностью построений, мелким масштабом рисунка и т.п. Ответ на этот вопрос будет положительным, если линия уровня и граничная прямая параллельны, т.е. их коэффициенты при переменных пропорциональны. В рассматриваемом примере коэффициенты при переменных линии уровня $F = 3x_1 + 3x_2$ пропорциональны соответствующим коэффициентам граничной прямой $x_1 + x_2 = 6$.

б) Геометрическое решение задачи показано на рис. 4.5, *б*, из которого следует, что если линию уровня перемещать в направлении убывания линейной функции (т.е. в направлении, противоположном вектору $\vec{q}$), то она всегда будет пересекать многоугольник решений, следовательно, линейная функция неограниченно убывает.

Итак, конечного оптимума линейной функции нет[1], т.е. $F_{min} = -\infty$. ▶

---

[1] **Если задачу с той же линейной функцией и с той же системой ограничений решать на максимум ($F \to$ max), то линию уровня следует перемещать в направлении возрастания F (т.е. в направлении вектора $\vec{q}$), и в этой задаче отсутствует конечный оптимум (*см.* рис.4.5, *а*): $F_{max} = \infty$.**

При геометрическом решении задач линейного программирования возможны случаи, когда условия задач противоречивы, т.е. область допустимых решений системы ограничений представляет пустое множество (*см.* например, рис. 2.9, *в*). Очевидно, в таких задачах нет оптимальных решений и нет смысла строить линию уровня.

Рассмотренный в этой главе геометрический метод решения задач линейного программирования обладает рядом достоинств. Он прост, нагляден, позволяет быстро и легко получить ответ.

Однако только геометрический метод решения никак не может удовлетворить ни математиков, ни экономистов. Возможны "технические" погрешности, которые неизбежно возникают при приближенном построении графиков. Второй недостаток геометрического метода заключается в том, что многие величины, имеющие четкий экономический смысл (такие, как остатки ресурсов производства, избыток питательных веществ и т.п.), не выявляются при геометрическом решении задач. Но самое главное — геометрический метод неприемлем для решения практических задач. Его можно применить только в том случае, когда число переменных в стандартной задаче равно двум. Поэтому необходимы аналитические методы, позволяющие решать задачи линейного программирования с любым числом переменных и выявить экономический смысл входящих в них величин. Эти методы будут рассмотрены в следующих главах.

УПРАЖНЕНИЯ

Задачи **4.4–4.10** решить геометрически.

4.4. $F = 2x_1 - 6x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 + x_2 \geq 2, \\ -x_1 + 2x_2 \leq 4, \\ x_1 + 2x_2 \leq 8, \end{cases}$$

$x_1 \geq 0,\ x_2 \geq 0.$

4.5. $F = 2x_1 - x_2 \to \min$

при ограничениях:

$$\begin{cases} x_1 + x_2 \geq 4, \\ -x_1 + 2x_2 \leq 2, \\ x_1 + 2x_2 \leq 10, \end{cases}$$

$x_1 \geq 0,\ x_2 \geq 0.$

4.6. $F = x_1 + x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 - 4x_2 - 4 \le 0, \\ 3x_1 - x_2 \ge 0, \\ x_1 + x_2 - 4 \ge 0, \end{cases}$$

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

4.7. $F = 2x_1 - x_2 \to \min$

при ограничениях:

$$\begin{cases} x_1 + x_2 \ge 4, \\ 2x_1 - x_2 \ge 2, \\ -x_1 - 2x_2 \ge -10, \end{cases}$$

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

4.8. $F = x_1 - x_2 \to \max$

при ограничениях:

$$\begin{cases} -2x_1 + x_2 \le 2, \\ x_1 - 2x_2 \le -8, \\ x_1 + x_2 \le 5, \end{cases}$$

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

4.9. Текст условия приведен в задаче **1.4**.

4.10. Текст условия приведен в задаче **1.5**.

Задачи **4.11**, **4.12** решить геометрически, предварительно приведя их к стандартной форме.

4.11.

$F = -4x_1 - 2x_2 + x_3 - x_4 \to \max$

при ограничениях:

$$\begin{cases} x_1 - x_2 + 4x_3 - 2x_4 = 2, \\ 3x_1 + 2x_2 - x_3 + 4x_4 = 3, \end{cases}$$

$x_j \ge 0,\ j = 1,2,3,4.$

4.12.

$F = x_1 + 2x_2 + x_3 - x_4 - 6 \to \min$

при ограничениях:

$$\begin{cases} -x_1 + 5x_2 + x_3 + x_4 + x_5 = 10, \\ 2x_1 - x_2 + x_3 - 3x_4 = 6, \\ 10x_2 + x_3 + 2x_4 + 3x_5 = 25, \end{cases}$$

$x_j \ge 0,\ j = 1, 2, \ldots, 5.$

Глава 5. СИМПЛЕКСНЫЙ МЕТОД

5.1. Геометрическая интерпретация симплексного метода

В гл. 3 рассмотрены основные теоремы линейного программирования, из которых следует, что если задача линейного программирования имеет оптимальное решение, то оно соответствует хотя бы одной угловой точке многогранника решений и совпадает, по крайней мере, с одним из допустимых базисных решений системы ограничений (см. теоремы 3.3, 3.4). Там же был указан путь решения любой задачи линейного программирования: перебрать конечное число допустимых базисных решений системы ограничений и выбрать среди них то, на котором функция цели принимает оптимальное решение. Геометрически это соответствует перебору всех угловых точек многогранника решений. Такой перебор в конце концов приведет к оптимальному решению (если оно существует), однако его практическое осуществление связано с огромными трудностями, так как для реальных задач число допустимых базисных решений хотя и конечно, но может быть чрезвычайно велико.

Число перебираемых допустимых базисных решений можно сократить, если производить перебор не беспорядочно, а с учетом изменений линейной функции, т.е. добиваясь того, чтобы каждое следующее решение было "лучше" (или, по крайней мере, "не хуже"), чем предыдущее, по значениям линейной функции (увеличение ее при отыскании максимума $F \to \max$, уменьшение — при отыскании минимума $F \to \min$).

Такой перебор позволяет сократить число шагов при отыскании оптимума. Поясним это на графическом примере.

Пусть область допустимых решений изображается многоугольником $ABCDEGH$ (рис. 5.1). Предположим, что его угловая точка A соответствует исходному допустимому базисному решению. При беспорядочном переборе пришлось бы испытать семь допустимых базисных решений, соответствующих семи угловым точкам многоугольника. Однако из чертежа видно, что после вершины A выгодно перейти к соседней вершине B, а затем — к оптимальной точке C.

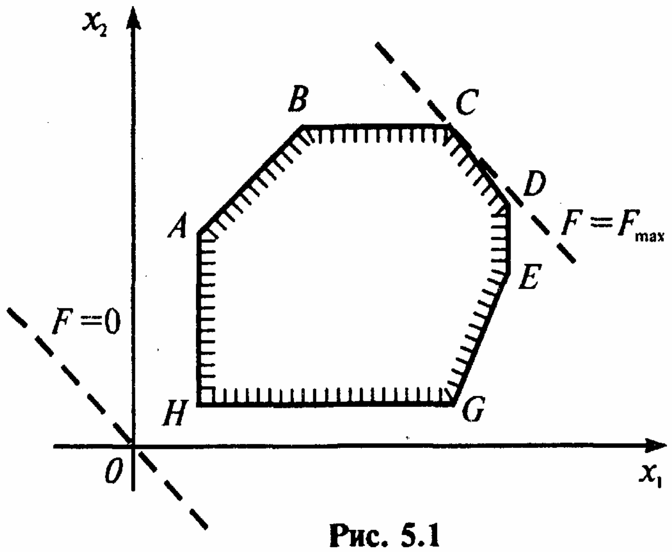


Рис. 5.1

Вместо семи перебрали только три вершины, последовательно улучшая линейную функцию.

Идея последовательного улучшения решения легла в основу универсального метода решения задач линейного программирования — ***симплексного***[1] ***метода.***

Геометрический смысл симплексного метода состоит в последовательном переходе от одной вершины многогранника ограничений (называемой первоначальной) к соседней, в которой линейная функция принимает лучшее (по крайней мере, не худшее) значение

---

[1] Симплекс (лат. simplex — простой) — простейший выпуклый многогранник в n-мерном пространстве с $n+1$ вершиной (например, тетраэдр в 3-мерном пространстве); симплексом является также область допустимых решений неравенства вида $\sum_{i=1}^{n} x_j \le 1$.

(по отношению к цели задачи) до тех пор, пока не будет найдено оптимальное решение — вершина, где достигается оптимальное значение функции цели (если задача имеет конечный оптимум).

Впервые симплексный метод был предложен американским ученым Дж. Данцигом в 1949 г., однако еще в 1939 г. идеи метода были разработаны российским ученым Л.В.Канторовичем.

Симплексный метод, позволяющий решить любую задачу линейного программирования, универсален. В настоящее время он используется для компьютерных расчетов, однако несложные примеры с применением симплексного метода можно решать и вручную.

Для реализации симплексного метода — последовательного улучшения решения — необходимо освоить *три основных элемента:*

- *способ определения какого-либо первоначального допустимого базисного решения задачи;*
- *правило перехода к лучшему (точнее, не худшему) решению;*
- *критерий проверки оптимальности найденного решения.*

Для использования симплексного метода задача линейного программирования должна быть приведена к каноническому виду, т.е. система ограничений должна быть представлена в виде уравнений. Алгоритм конкретной вычислительной реализации этих элементов рассмотрим на примерах.

5.2. Отыскание максимума линейной функции

В качестве первого примера рассмотрим задачу об использовании ресурсов, сформулированную в разд. 1.2 и уже решенную геометрически в задаче **4.1.**

▷ **5.1.** Решить симплексным методом задачу:

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

при ограничениях:

$$\begin{cases} x_1 + 3x_2 \le 18, \\ 2x_1 + x_2 \le 16, \\ x_2 \le 5, \\ 3x_1 \le 21, \end{cases} \qquad \textbf{(5.1)}$$

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

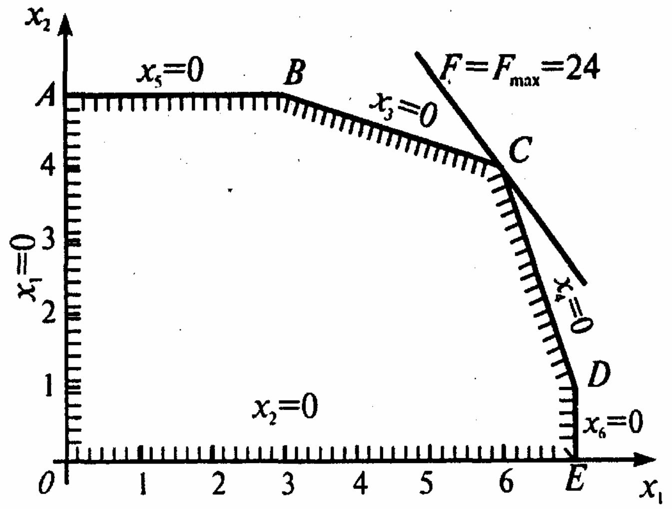


Рис. 5.2

Р е ш е н и е. С помощью дополнительных неотрицательных переменных перейдем к системе уравнений. В данном случае все дополнительные переменные вводятся со знаком "плюс", так как все неравенства имеют вид "$\leq$" [*см.* систему ограничений (1.26)].

Получим систему ограничений в виде:

$$\begin{cases} x_1 + 3x_2 + x_3 = 18, \\ 2x_1 + x_2 + x_4 = 16, \\ x_2 + x_5 = 5, \\ 3x_1 + x_6 = 21. \end{cases} \qquad (5.2)$$

Для нахождения первоначального базисного решения разобьем переменные на две группы — основные и неосновные. Так как определитель, составленный из коэффициентов при дополнительных переменных x_3, x_4, x_5, x_6, отличен от нуля, то (*см.* разд. 2.1) эти переменные можно взять в качестве основных на первом шаге решения задачи. При выборе основных переменных на первом шаге не обязательно составлять определитель из их коэффициентов и проверять, равен ли он нулю. Достаточно воспользоваться следующим правилом:

в качестве основных переменных на первом шаге следует выбрать (если возможно) такие m переменных, каждая из которых входит только в одно из m уравнений системы ограничений, при этом нет таких уравнений системы, в которые не входит ни одна из этих переменных.

Дополнительные переменные удовлетворяют этому правилу. Если выбранные по этому правилу переменные имеют те же знаки, что и соответствующие им свободные члены в правых частях уравнений, то полученное таким образом базисное решение будет допустимым. В данном случае так и получилось.

I шаг. Основные переменные: x_3, x_4, x_5, x_6.
Неосновные переменные: x_1, x_2.

Выразим основные переменные через неосновные:

$$\begin{cases} x_3 = 18 - x_1 - 3x_2, \\ x_4 = 16 - 2x_1 - x_2, \\ \boxed{x_5 = 5 - x_2,} \\ x_6 = 21 - 3x_1 \end{cases} \tag{5.3}$$

Положив неосновные переменные равными нулю, т.е. $x_1 = 0$, $x_2 = 0$, получим базисное решение $X_1 = (0;\ 0;\ 18;\ 16;\ 5;\ 21)$, которое является допустимым и соответствует вершине $O(0;0)$ многогранника $OABCDE$ на рис. 5.2. Поскольку это решение допустимое, нельзя отбросить возможность того, что оно оптимально. Выразим линейную функцию через неосновные переменные: $F = 2x_1 + 3x_2$. При решении X_1 значение функции равно $F(X_1)$. Легко понять, что функцию F можно увеличить за счет увеличения любой из неосновных переменных, входящих в выражение для F с положительным коэффициентом. Это можно осуществить, перейдя к такому новому допустимому базисному решению, в котором эта переменная будет неосновной, т.е. принимать не нулевое, а положительное значение (если новое решение будет вырождено, то функция цели сохранит свое значение). При таком переходе одна из основных переменных перейдет в неосновные, а геометрически произойдет переход к соседней вершине многоугольника, где значение линейной функции "лучше" (по крайней мере "не хуже"). В данном примере для увеличения F можно переводить в основные либо x_1, либо x_2, так как обе эти переменные входят в выражение для F со знаком "плюс". Для определенности в такой ситуации будем выбирать переменную, имеющую больший коэффициент, т.е. в данном случае x_2 (такое правило выбора не всегда дает наименее трудоемкое решение, иногда имеет смысл провести предварительные специальные оценки).

Система (5.3) накладывает ограничения на рост переменной x_2. Поскольку необходимо сохранять допустимость решений, т.е. все переменные должны оставаться неотрицательными, то должны выполняться следующие неравенства (при этом $x_1 = 0$ как неосновная переменная):

$$\begin{cases} x_3 = 18 - 3x_2 \geq 0, \\ x_4 = 16 - x_2 \geq 0, \\ x_5 = 5 - x_2 \geq 0, \\ x_6 = 21, \end{cases} \quad \text{откуда} \quad \begin{cases} x_2 \leq 18\,/\,3, \\ x_2 \leq 16\,/\,1, \\ x_2 \leq 5\,/\,1. \end{cases}$$

Каждое уравнение системы (5.3), кроме последнего, определяет оценочное отношение — границу роста переменной x_2, сохраняющую неотрицательность соответствующей переменной. Эта граница определяется абсолютной величиной отношения свободного члена к коэффициенту при x_2 при условии, что эти числа имеют разные знаки. Последнее уравнение системы не ограничивает рост переменной x_2, так как данная переменная в него не входит (или формально входит с нулевым коэффициентом). В этом случае условимся обозначать границу символом ∞. Такой же символ будем использовать, когда свободный член и коэффициент при переменной в уравнении имеют одинаковые знаки, так как и в этом случае нет ограничений на рост переменной.

Не накладывает ограничений на рост переменной, переводимой в основные, и такое уравнение, где свободный член отсутствует (т.е. равен 0), а переводимая переменная имеет положительный коэффициент. И в этом случае граница обозначается символом ∞. Обратите внимание, что при нулевом свободном члене и отрицательном коэффициенте при переводимой переменной уравнение ограничивает рост этой переменной нулем! (любое положительное ее значение вносит отрицательную компоненту в следующее базисное решение). Все возможные случаи, которые возникают при оценке переменной, переводимой в основные, перечислены в конце разд. 5.3.

Очевидно, что сохранение неотрицательности всех переменных (допустимость решения) возможно, если не нарушается ни одна из полученных во всех уравнениях границ. В данном примере наибольшее возможное значение для переменной x_2 определяется как $x_2 = \min\{18/3;\ 16/1;\ 5/1;\ \infty\} = 5$. При $x_2 = 5$ переменная x_5 обращается в нуль и переходит в неосновные.

Уравнение, где достигается наибольшее возможное значение переменной, переводимой в основные (т.е. где оценка минимальна), называется *разрешающим*. В данном случае — это третье уравнение. Разрешающее уравнение будем выделять рамкой в системе ограничений.

II шаг. Основные переменные: x_2, x_3, x_4, x_6.
Неосновные переменные: x_1, x_5.

Выразим новые основные переменные через новые неосновные, начиная с разрешающего уравнения (его используем при записи выражения для x_2):

$$\begin{cases} x_2 = 5 - x_5, \\ x_3 = 18 - x_1 - 3(5 - x_5), \\ x_4 = 16 - 2x_1 - (5 - x_5), \\ x_5 = 21 - 3x_1, \end{cases} \tag{5.4}$$

или

$$\begin{cases} x_2 = 5 - x_5, \\ \boxed{x_3 = 3 - x_1 + 3x_5,} \\ x_4 = 11 - 2x_1 + x_5, \\ x_5 = 21 - 3x_1. \end{cases}$$

Второе базисное решение $X_2 = (0;\ 5;\ 3;\ 11;\ 0;\ 21)$ является допустимым и соответствует вершине $A\ (0;5)$ на рис. 5.2. Геометрическая интерпретация перехода от X_1 к X_2 — переход от вершины O к соседней вершине A на многоугольнике решений $OABCDE$.

Выразив линейную функцию через неосновные переменные на этом шаге, получаем:

$$F = 2x_1 + 3x_2 = 2x_1 + 3(5 - x_5) = 15 + 2x_1 - 3x_5.$$

Значение линейной функции $F_2 = F(X_2) = 15$. Изменение значения линейной функции легко определить заранее как произведение наибольшего возможного значения переменной, переводимой в основные, на ее коэффициент в выражении для линейной функции; в данном случае $\Delta F_1 = 5 \cdot 3 = 15$, $F_2 = F_1 + \Delta F_1 = 0 + 15 = 15$.

Однако значение F_2 не является максимальным, так как повторяя рассуждения I шага, обнаруживаем возможность дальнейшего увеличения линейной функции за счет переменной x_1, входящей в выражение для F с положительным коэффициентом. Система уравнений (5.4) определяет наибольшее возможное значение для x_1: $x_1 = \min\{\infty; 3/1; 11/2; \infty\} = 3$. Второе уравнение является разрешающим, переменная x_3 переходит в неосновные, при этом $\Delta F_2 = 3 \cdot 2 = 6$.

III шаг. Основные переменные: x_1, x_2, x_4, x_6.
Неосновные переменные: x_3, x_5.

Как и на II шаге, выражаем новые основные переменные через новые неосновные, начиная с разрешающего уравнения (его используем при записи выражения для x_1). После преобразований получаем:

$$\begin{cases} x_1 = 3 - x_3 + 3x_5, \\ x_2 = 5 - x_5, \\ \boxed{x_4 = 5 + 2x_3 - 5x_5,} \\ x_6 = 12 + 3x_3 - 9x_5. \end{cases} \qquad (5.5)$$

Базисное решение $X_3 = (3;\ 5;\ 0;\ 5;\ 0;\ 12)$ соответствует вершине B (3;5). Выражаем линейную функцию через неосновные переменные: $F = 2x_1 + 3x_2 = 2(3 - x_3 + 3x_5) + 3(5 - x_5) = 21 - 2x_3 + 3x_5$, $F_3 = F(X_3) = 21$. Проверяем: $F_3 - F_2 = 21 - 15 = 6 = \Delta F_2$. Третье допустимое базисное решение тоже не является оптимальным, поскольку при неосновной переменной x_5 в выражении линейной функции через неосновные переменные содержится положительный коэффициент. Переводим x_5 в основную переменную. При определении наибольшего возможного значения для x_5 следует обратить внимание на первое уравнение в системе (5.5), которое не дает ограничений на рост x_5, так как свободный член и коэффициент при x_5 имеют одинаковые знаки. Поэтому $x_5 = \min\{\infty; 5; 1; 12/9\} = 1$. Третье уравнение является разрешающим, и переменная x_4 переходит в неосновные; $\Delta F_3 = 1 \cdot 3 = 3$.

IV шаг. Основные переменные: x_1, x_2, x_5, x_6.
Неосновные переменные: x_3, x_4.

После преобразований получим:

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

Базисное решение $X_4 = (6; 4; 0; 0; 1; 3)$ соответствует вершине $C(6; 4)$ на рис. 5.2.

Линейная функция, выраженная через неосновные переменные, имеет вид: $F = 24 - \frac{4}{5}x_3 - \frac{3}{5}x_4$. Это выражение не содержит положительных коэффициентов при неосновных переменных, поэтому значение $F_4 = F(X_4) = 24$ максимальное. Функцию F невозможно еще увеличить, переходя к другому допустимому базисному решению, т.е. решение X_4 оптимальное. Вспоминая экономический смысл всех переменных, можно сделать следующие выводы.

Прибыль предприятия принимает максимальное значение 24 руб. при реализации 6 единиц продукции $P_1(x_1 = 6)$ и 4 единиц продукции $P_2(x_2 = 4)$. Дополнительные переменные x_3, x_4, x_5, x_6 показывают разницу между запасами ресурсов каждого вида и их потреблением, т.е. остатки ресурсов. При оптимальном плане производства $x_3 = x_4 = 0$, т.е. остатки ресурсов S_1 и S_2 равны нулю, а остатки ресурсов S_3 и S_4 равны соответственно 1 и 3 единицам. ▶

Теперь можно в общем виде сформулировать ***критерий оптимальности решения при отыскании максимума линейной функции:*** ***если в выражении линейной функции через неосновные переменные отсутствуют положительные коэффициенты при неосновных переменных, то решение оптимально.***

5.3. Отыскание минимума линейной функции

При определении минимума линейной функции Z возможны два пути:

1) отыскать максимум функции F, полагая $F = -Z$ и учитывая, что $Z_{min} = -F_{max}$;

2) модифицировать симплексный метод: на каждом шаге уменьшать линейную функцию за счет той неосновной переменной, которая входит в выражение линейной функции с отрицательным коэффициентом.

Рассмотрим это на следующем примере.

▷ **5.2.** Решить симплексным методом задачу

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

при ограничениях:

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

Р е ш е н и е. Введем дополнительные неотрицательные переменные y_5 и y_6 со знаком "минус", так как неравенства имеют вид "$\geq$". Получим систему уравнений:

$$\begin{cases} y_1 + 2y_2 + 3y_4 - y_5 = 2, \\ 3y_1 + y_2 + y_3 - y_6 = 3. \end{cases}$$

Если на первом шаге в качестве основных взять дополнительные переменные, то получим недопустимое базисное решение: (0; 0; 0; 0; −2; −3). В данном случае в качестве основных удобно взять переменные y_3 и y_4 (это согласуется с правилом выбора основных переменных, сформулированным в разд. 5.2), коэффициенты при y_3 и y_4 положительны, поэтому в качестве первоначального получим допустимое базисное решение.

I шаг. Основные переменные: y_3, y_4.
Неосновные переменные: y_1, y_2, y_5, y_6.

Выражаем основные переменные через основные:

$$\begin{cases} y_3 = 3 - 3y_1 - y_2 + y_6, \\ y_4 = (2/3) - (1/3)y_1 - (2/3)y_2 + (1/3)y_5. \end{cases}$$

$Y_1 = (0;\ 0;\ 3;\ \frac{2}{3};\ 0;\ 0)$ — первое базисное решение, оно допустимое. Выражаем линейную функцию через неосновные переменные: $Z = 18y_1 + 16y_2 + 5y_3 + 21y_4 = 18y_1 + 16y_2 + 5(3 - 3y_1 - y_2 + y_6) + 21(2/3 - (1/3)y_1 - (2/3)y_2 + (1/3)y_5) = 29 - 4y_1 - 3y_2 + 7y_5 + 5y_6$.

$Z_1 = Z(Y_1) = 29$ — это значение не является минимальным, так как функцию Z можно уменьшить за счет перевода в основные любой из переменных y_1 или y_2, имеющих в выражении для Z отрицательные коэффициенты. Так как y_1 имеет бóльший по абсолютному значению коэффициент, то начнем с этой переменной. Для нее наибольшее возможное значение: $y_1 = \min\left\{\frac{3}{3}; \frac{2}{3}; \frac{1}{3}\right\} = 1$, т.е. первое уравнение является разрешающим; y_3 становится неосновной переменной, $\Delta Z_1 = -4 \cdot 1 = -4$.

II шаг. Основные переменные: y_1, y_4.
Неосновные переменные: y_2, y_3, y_5, y_6.

Получим после преобразований:

$$\begin{cases} y_1 = 1 - (1/3)y_2 - (1/3)y_3 + (1/3)y_6, \\ y_4 = 1/3 - (5/9)y_2 + (1/9)y_3 + (1/3)y_5 - (1/9)y_6, \end{cases}$$

$Z = 25 - (5/3)y_2 + (4/3)y_3 + 7y_5 + (11/3)y_6$ — линейная функция. При базисном решении $Y_2 = (1;\ 0;\ 0;\ 1/3;\ 0;\ 0)$ получаем $Z_2 = Z(Y_2) = 25$. $Z_2 - Z_1 = 25 - 29 = -4 = \Delta Z_1$. Переменную y_2 переводим в основные, так как в выражение для Z она входит с отрицательным коэффициентом. Наибольшее возможное значение $y_2 = \min\{3;\ 3/5\} = 3/5$, второе уравнение разрешающее и y_4 переходит в неосновные переменные; $\Delta Z_2 = 3/5(-5/3) = -1$.

III шаг. Основные переменные: y_1, y_2.
Неосновные переменные: y_3, y_4, y_5, y_6.

Получим после преобразований:

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

$Z = 24 + y_3 + 3y_4 + 6y_5 + 4y_6$. Базисное решение $Y_3 = (\frac{4}{5};\frac{3}{5};0;0;0;0)$ оптимальное, так как в выражении для Z нет неосновных переменных с отрицательными коэффициентами. Поэтому $Z_{\min} = Z_3 = Z(Y_3) = 24$. $Z_3 - Z_2 = 24 - 25 = -1 = \Delta Z_2$.▶

Сформулируем ***критерий оптимальности при отыскании минимума линейной функции:*** *если в выражении линейной функции через неосновные переменные отсутствуют отрицательные коэффициенты при неосновных переменных, то решение оптимально.*

Замечание. На каждом шаге симплексного метода какая-либо неосновная переменная переводится в основные, при этом каждое уравнение системы ограничений определяет конечное или бесконечное наибольшее возможное значение этой переменной — оценочное отношение. В задачах **5.1** и **5.2** встречались различные случаи оценки роста неосновной переменной, которые зависели от знаков и значений свободного члена и коэффициента при переводимой переменной. Сформулируем все возможные случаи. Обозначим: x_i — переводимая неосновная переменная, b_j — свободный член, a_{ij} — коэффициент при x_i. В общем виде уравнение $x_j = b_j + \ldots + a_{ij}x_i + \ldots$ определяет наибольшее возможное значение x_i по следующим правилам:

1) $x_i = |b_j/a_{ij}|$, если b_j и a_{ij} разного знака и $a_{ij} \neq 0$, $b_j \neq 0$; например: $x_3 = 8 - 2x_2 + \ldots$; $x_2 = 8/2 = 4$ или $x_3 = -8 + 2x_2 + \ldots$, $x_2 = 8/2 = 4$.

2) $x_i = \infty$, если b_j и a_{ij} одного знака и $a_{ij} \neq 0$, $b_j \neq 0$; например, $x_3 = 8 + 2x_2 + \ldots$; $x_2 = \infty$.

3) $x_i = 0$, если $b_j = 0$ и $a_{ij} < 0$, например, $x_3 = 0 - 2x_2 + \ldots$; $x_2 = 0$.

4) $x_i = \infty$, если $b_j = 0$ и $a_{ij} > 0$, например, $x_3 = 0 + 2x_2 + \ldots$; $x_2 = \infty$.

5) $x_i = \infty$, если $a_{ij} = 0$, например, $x_3 = 5 + 0 \cdot x_2 + \ldots$; $x_3 = -5 + 0 \cdot x_2 + \ldots$; $x_2 = \infty$.

5.4. Определение первоначального допустимого базисного решения

В рассмотренных выше примерах оптимальное решение получено путем последовательного перехода от первоначального допустимого базисного решения к следующему, "лучшему", и так — до достижения критерия оптимальности. Однако не всегда на первом же шаге получается допустимое базисное решение. В следующем примере рассмотрим один из возможных алгоритмов получения допустимого базисного решения. Другой, так называемый M-метод, будет изложен в разд. 5.7.

5.3. Решить симплексным методом задачу

$$F = x_1 + 2x_2 \to \max$$

при ограничениях:

$$\begin{cases} x_1 - x_2 \le -1, \\ x_1 - x_2 \ge -3, \\ x_1 \le 3, \end{cases}$$

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

Р е ш е н и е. Вводим дополнительные неотрицательные переменные x_3, x_4, x_5 с соответствующими знаками:

$$\begin{cases} x_1 - x_2 + x_3 = -1, \\ x_1 - x_2 - x_4 = -3, \\ x_1 + x_5 = 3. \end{cases}$$

В соответствии с правилом, сформулированным в разд. 5.2, на I шаге в качестве основных возьмем дополнительные переменные.

I шаг. Основные переменные: x_3, x_4, x_5.
Неосновные переменные: x_1, x_2.

Выражаем основные переменные через неосновные:

$$\begin{cases} x_3 = -1 - x_1 + x_2, \\ x_4 = 3 + x_1 - x_2, \\ x_5 = 3 - x_1. \end{cases}$$

$X_1 = (0;\ 0;\ \mathbf{-1};\ 3;\ 3)$ — первое базисное решение недопустимое (отрицательная компонента выделена), поэтому оно не может быть оптимальным. Линейную функцию на недопустимом решении не рассматриваем! В системе (5.6) выберем то уравнение, которое содержит отрицательный свободный член, т.е. первое уравнение (если таких уравнений несколько, выбираем любое из них).

Поскольку переменная x_3 принимает отрицательное значение при первом базисном решении, то ее необходимо увеличить. Это можно сделать за счет увеличения любой из неосновных переменных, входящих в первое уравнение с *положительным* коэффициентом, в данном случае — переменной x_2. Если перевести эту переменную в основные, то она, став положительной, увеличит переменную x_3. Как только x_2 достигнет уровня 1, то x_3 обратится в 0, т.е. исчезнет отрицательная компонента в решении. Можно считать, что переменная x_3 станет неосновной. Однако рост переменной x_2 ограничен условиями неотрицательности остальных переменных, которые определяют $x_2 = \min\{1; 3; \infty\} = 1$, т.е. первое уравнение — разрешающее. При $x_2 = 1$ переменная $x_3 = 0$ и переходит в неосновные переменные.

II шаг. Основные переменные: x_2, x_4, x_5.
Неосновные переменные: x_1, x_3.

Выражая новые основные переменные через новые неосновные, начиная с разрешающего уравнения, получаем

$$\begin{cases} x_2 = 1 + x_1 + x_3, \\ x_4 = 2 \qquad\quad - x_3, \\ x_5 = 3 - x_1 \end{cases}$$

и базисное решение $X_2 = (0;\ 1;\ 0;\ 2;\ 3)$, которое является допустимым. Поэтому выражаем через неосновные переменные линейную функцию $F = x_1 + x_2 = 1 + 2x_1 + x_3$ и продолжаем решение в соответствии с алгоритмом, изложенным в разд. 5.2 (читателю рекомендуется провести его самостоятельно).▶

Однако не всегда первый же шаг избавляет от недопустимого решения.

▷ **5.4.** Решить симплексным методом задачу

$$F = -2x_1 + 3x_2 \to \max$$

при ограничениях:

$$\begin{cases} 2x_1 + 3x_2 \geq 12, \\ -x_1 + x_2 \leq 7, \\ 2x_1 + x_2 \leq 10, \\ x_2 \geq 2, \end{cases}$$

$$x_1 \geq 0.$$

Решение. После введения дополнительных неотрицательных переменных с соответствующими знаками получим систему уравнений:

$$\begin{cases} 2x_1 + 3x_2 - x_3 = 12, \\ -x_1 + x_2 + x_4 = 7, \\ 2x_1 + x_2 + x_5 = 10, \\ x_2 - x_6 = 2. \end{cases}$$

На I шаге дополнительные переменные возьмем в качестве основных, так как они удовлетворяют правилу, изложенному в разд. 5.2: входят во все уравнения и только по одному разу.

I шаг. Основные переменные: x_3, x_4, x_5, x_6.
Неосновные переменные: x_1, x_2.

Выразим основные переменные через неосновные:

$$\begin{cases} x_3 = -12 + 2x_1 + 3x_2, \\ x_4 = 7 + x_1 - x_2, \\ x_5 = 10 - 2x_1 - x_2, \\ \boxed{x_6 = -2 + x_2.} \end{cases}$$

$X_1 = (0;\ 0;\ \mathbf{-12};\ 7;\ 10;\ \mathbf{-2})$ — первое базисное решение недопустимое, с двумя отрицательными компонентами (выделены).

Для получения допустимого базисного решения поступаем так, как в задаче **5.3**: выбираем любое уравнение, содержащее отрицательный свободный член (первое или четвертое), например первое, и в нем — любую неосновную переменную с положительным коэффициентом: x_1 или x_2, например x_1. Наибольшее возможное значение $x_1 = \min\{12/2;\ \infty;\ 10/2;\ \infty\} = 5$ достигается в третьем уравнении; оно разрешающее, и переменная x_5 переходит в неосновные переменные. Однако при этом ни одна из отрицательных компонент базисного решения не пропадает! Поэтому невыгодно переводить переменную x_1 в основные переменные. Переведем в основные x_2, тогда наибольшее возможное значение $x_2 = \min\{12/3;\ 7;\ 10;\ 2\} = 2$ достигается в четвертом уравнении; при этом переменная x_6 переходит в неосновные и исчезает одна отрицательная компонента в базисном решении.

II шаг. Основные переменные: x_2, x_3, x_4, x_5.
Неосновные переменные: x_1, x_6.

Выразим новые основные переменные через новые неосновные, начиная с четвертого уравнения:

$$\begin{cases} x_2 = 2 + x_6, \\ \boxed{x_3 = -6 + 2x_1 + 3x_6,} \\ x_4 = 5 + x_1 - x_6, \\ x_5 = 8 - 2x_1 - x_6. \end{cases}$$

$X_2 = (0;\ 2;\ \mathbf{-6};\ 5;\ 8;\ 0)$ — недопустимое базисное решение с одной отрицательной компонентой. Рассмотрим второе уравнение (с отрицательным свободным членом) и переведем в основные одну из неосновных переменных, x_1 или x_6, входящих в уравнение с положительными коэффициентами.

Получим из уравнений их наибольшие возможные значения: $x_1 = \min\{\infty;\ 3;\ \infty;\ 8/2\} = 3$ достигается во втором уравнении; $x_6 = \{\infty;\ 2;\ 5;\ 8\} = 2$ тоже определяет второе уравнение как разрешающее. Любой выбор устранит недопустимость решения, поэтому безразлично, какую переменную x_1 или x_6 выбрать. Переведем x_6 в основные.

III шаг. Основные переменные: x_2, x_4, x_5, x_6.
Неосновные переменные: x_1, x_3.

Выразим новые основные переменные через новые неосновные, начиная со второго уравнения:

$$\begin{cases} x_6 = 2 - (2/3)x_1 + (1/3)x_3, \\ x_2 = 4 - (2/3)x_1 + (1/3)x_3, \\ x_4 = 3 + (5/3)x_1 - (1/3)x_3, \\ x_5 = 6 - (4/3)x_1 - (1/3)x_3, \end{cases}$$

$X_3 = (0;\ 4;\ 0;\ 3;\ 4;\ 2)$ — допустимое базисное решение. Выражаем функцию цели через неосновные переменные $F = -2x_1 + 3x_2 = 12 - 4x_1 + x_3$. Дальнейшее решение предоставляем выполнить самостоятельно в соответствии с алгоритмом, изложенным в разд. 5.2.▶

Замечание 1. Если базисное решение недопустимое и для его улучшения есть возможность *выбора переменной*, переводимой из неосновных в основные, то рекомендуется выбрать такую неосновную переменную, которая определит в качестве разрешающего то уравнение системы, где содержится отрицательный свободный член. Только в этом случае новое базисное решение будет содержать, по крайней мере, на одну отрицательную компоненту меньше. Если в качестве разрешающего будет получено уравнение, не содержащее отрицательного свободного члена, то в новом базисном решении число отрицательных компонент не уменьшится.

Замечание 2. Из задачи **5.4** не следует делать вывод о том, что чем больше отрицательных компонент в первоначальном базисном решении, тем больше потребуется шагов, чтобы получить допустимое базисное решение. Оказывается, что в некоторых случаях невозможно получить допустимое базисное решение даже при одной отрицательной компоненте, а иногда его можно получить за один шаг, хотя все компоненты первоначального базисного решения отрицательны. Дальнейшие примеры пояснят это замечание.

Рассмотрим задачу о составлении рациона, приведенную в разд. 1.2 и решенную геометрически в задаче **4.2.**

▷ **5.5.** Решить симплексным методом задачу

$$F = 4x_1 + 6x_2 \to \min$$

при ограничениях:

$$\begin{cases} 3x_1 + x_2 \ge 9, \\ x_1 + 2x_2 \ge 8, \\ x_1 + 6x_2 \ge 12, \end{cases}$$

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

Решение. Введем дополнительные переменные x_3, x_4, x_5, x_6 (каждую со знаком "минус"). Их экономический смысл — это разность между содержанием и необходимым минимумом каждого из питательных веществ.

На I шаге берем дополнительные переменные в качестве основных.

I шаг. Основные переменные: x_3, x_4, x_5.
Неосновные переменные: x_1, x_2.

После преобразований получим:

$$\begin{cases} x_3 = -9 + 3x_1 + x_2, \\ x_4 = -8 + x_1 + 2x_2, \\ \boxed{x_5 = -12 + x_1 + 6x_2.} \end{cases}$$

$X_1 = (0;\ 0;\ -9;\ -8;\ -12)$ — первое базисное решение недопустимое, содержащее три отрицательные компоненты. Неосновная переменная x_2 входит в каждое уравнение с положительным коэффициентом, поэтому имеет смысл перевести ее в основные. В случае, когда все основные переменные принимают отрицательные значения, для ускорения решения можно в качестве значения для переменной x_2 взять максимальное оценочное отношение из полученных во всех уравнениях: $x_2 = \max\{9/3;\ 8;\ 12\} = 12$. Третье уравнение является разрешающим, при этом $x_5 = 0$ и переходит в основные, а остальные основные переменные принимают положительные значения.

II шаг. Основные переменные: x_2, x_4, x_5.
Неосновные переменные: x_1, x_3.

После преобразований получим:

$$\begin{cases} x_2 = 9 - 3x_1 + x_3, \\ x_4 = 10 - 5x_1 + 2x_3, \\ x_5 = 42 - 17x_1 + 6x_3. \end{cases}$$

$X_2 = (0;\ 9;\ 0;\ 10;\ 42)$ — допустимое базисное решение. Если действовать, как в предыдущих примерах, то для получения допустимого решения потребуется три шага!

Заканчивая решение задачи **5.5** симплексным методом (рекомендуем это сделать самостоятельно), на следующем шаге получаем оптимальное базисное решение $X_3 = (2;\ 3;\ 0;\ 0;\ 5)$, при котором минимальные затраты на рацион составляют $F_{\min} = 26$. Учитывая экономический смысл исходных и дополнительных переменных, получаем, что в оптимальном рационе используются 2 единицы корма I и 3 единицы корма II, при этом вещества S_1 и S_2 потребляются в необходимых минимальных количествах ($x_3 = x_4 = 0$), а питательное вещество S_3 оказывается в избытке на 5 единиц ($x_5 = 5$).▶

Итак, для ускорения отыскания допустимого базисного решения, когда все основные переменные отрицательны, рекомендуется выбрать, если возможно, неосновную переменную, входящую во все уравнения со знаком "плюс", и в качестве ее значения брать не минимум, а максимум оценочных отношений, получаемых из каждого уравнения.

▷ **5.6.** Решить симплексным методом задачу

$$F = x_1 - 3x_2 \to \max$$

при ограничениях:

$$\begin{cases} x_1 + x_2 \le 2, & \text{(I)} \\ 2x_1 + x_2 \ge 6, & \text{(II)} \end{cases}$$

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

Решение. Введем дополнительные переменные:

$$\begin{cases} x_1 + x_2 + x_3 = 2, \\ 2x_1 + x_2 - x_4 = 6. \end{cases}$$

I шаг. Основные переменные: x_3, x_4.
Неосновные переменные: x_1, x_2.

$$\begin{cases} x_3 = 2 - x_1 - x_2, \\ x_4 = -6 + 2x_1 + x_2. \end{cases}$$

$X_1 = (0;\ 0;\ 2;\ -1)$ — недопустимое базисное решение. Во втором уравнении выбираем любую переменную — x_1 или x_2, так как обе имеют знак "плюс", и переводим в основные. Для x_1: $\min\{2; 6/2\} = 2$, разрешающее первое уравнение; для x_2: $\min(2;6) = 2$, разрешающее тоже первое уравнение, поэтому в любом случае не удается сразу избавиться от отрицательной компоненты базисного решения (*см.* замечание 1 на с. 80).

II шаг. Основные переменные: x_1, x_4.
Неосновные переменные: x_2, x_3.

Получим после преобразований:

$$\begin{cases} x_1 = 2 - x_2 - x_3, \\ x_4 = -2 - x_1 - 2x_3. \end{cases}$$

$X_2 = (2;\ 0;\ 0;\ -2)$ — недопустимое базисное решение. Однако второе уравнение не содержит неосновной переменной с положительным коэффициентом, поэтому невозможно увеличить переменную x_4 и получить допустимое базисное решение. Задача противоречива (рис. 5.3).▶

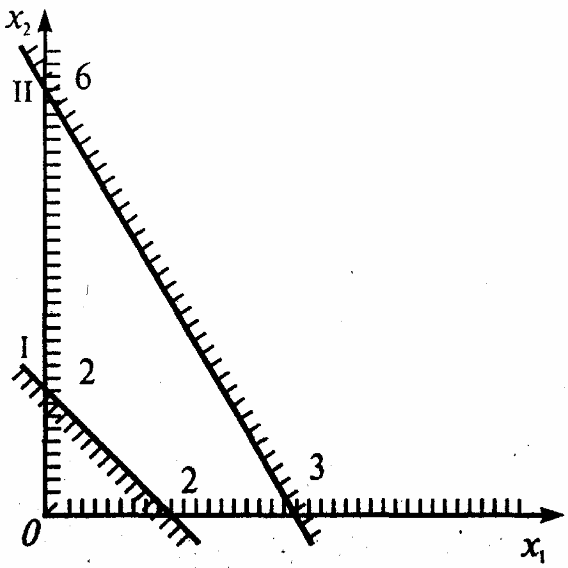


Рис. 5.3

После анализа результатов, полученных при решении задач **5.1–5.6**, сформулируем *алгоритм получения первоначального допустимого базисного решения:*

1. Если каждая дополнительная переменная входит в уравнение с тем же знаком, что и свободный

член, стоящий в правой части уравнения, то дополнительные переменные можно брать в качестве основных на I шаге. При этом получится допустимое базисное решение.

2. Если первое базисное решение получилось недопустимым (например, в качестве основных взяты дополнительные переменные, но хотя бы одна из них входила в уравнение со знаком, противоположным знаку свободного члена), то рассматриваем уравнение, содержащее отрицательный свободный член (любое, если их несколько), и переводим в основные ту неосновную переменную, которая в это уравнение входит с положительным коэффициентом (любую, если их несколько). Такие шаги повторяем до тех пор, пока не достигается допустимое базисное решение.

3. Если базисное решение недопустимое, а в уравнении, содержащем отрицательный свободный член, отсутствует неосновная переменная с положительным коэффициентом, то в этом случае допустимое базисное решение получить невозможно, т.е. условия задачи противоречивы.

5.5. Особые случаи симплексного метода

Рассмотрим особые случаи, которые могут возникнуть при решении задачи линейного программирования симплексным методом.

Неединственность оптимального решения (альтернативный оптимум)

▷ **5.7.** Решить симплексным методом задачу

$$F = 3x_1 + 3x_2 \to \max$$

при ограничениях:

$$\begin{cases} x_1 + x_2 \le 8, \\ 2x_2 - x_2 \ge 1, \\ x_1 - 2x_2 \le 2, \end{cases}$$

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

Решение. Геометрическое решение задачи приведено в задаче **4.3, *а*** (*см.* рис. 4.3, *а*): оптимум достигается в любой точке отрезка *AB*, так как линия уровня параллельна этому отрезку. Покажем, как проявляется наличие альтернативного оптимума при решении задачи симплексным методом.

На очередном шаге получаем:

основные переменные: x_1, x_2, x_5,

неосновные переменные: x_3, x_4.

Выражаем основные переменные через неосновные:

$$\begin{cases} x_1 = 5 - (2/3)x_3 - (1/3)x_4, \\ x_2 = 3 - (1/3)x_3 + (1/3)x_4, \\ x_5 = 9 - x_3 - x_4. \end{cases}$$

$X_1 = (3;\ 5;\ 0;\ 0;\ 9)$ — допустимое базисное решение, соответствующее угловой точке $A(3;\ 5)$. Линейная функция: $F = 24 - x_3$. В этом выражении отсутствуют положительные коэффициенты при неосновных переменных, значит критерий оптимальности выполнен, X_1 — оптимальное базисное решение, $F_{\max} = F(X_1) = 24$. Однако в последнем *выражении для F отсутствует неосновная переменная* x_4 (формально входит с нулевым коэффициентом), поэтому изменение этой переменной не повлечет за собой изменение линейной функции. Например, можно перевести в основную переменную x_4; $x_4 = \min\{15;\ \infty;\ 9\} = 9$. Переменная x_5 перейдет в неосновные, однако изменения линейной функции не произойдет: $\Delta F = 9 \cdot 0 = 0$. Действительно, на следующем шаге получим новое базисное решение $X_2 = (6;\ 2;\ 0;\ 9;\ 0)$, соответствующее угловой точке $B\ (6;\ 2)$, $F_{\max} = F\ (X_2) = 24$ (рекомендуем читателю убедиться в этом самостоятельно). Учитывая, что переменная $x_3 = 0$ (в базисном решении X_2 она осталась неосновной), а переменная x_4 удовлетворяет неравенству $0 \le x_4 \le 9$, из системы уравнений можно получить все множество оптимальных решений задачи. Положим для удобства $x_4 = t$, где $t \in [0;\ 9]$. Тогда множество оптимальных решений: $x_1 = 3 + (1/3)t$; $x_2 = 5 - (1/3)t$; $x_3 = 0$; $x_4 = t$; $x_5 = 9 + t\ (t \in [0;\ 9])$.

Замечание. В соответствии с теоремами 3.3 и 3.4 ***множество оптимальных решений можно представить как выпуклую линейную комбинацию X базисных решений*** **$X_1 = (3;\ 5;\ 0;\ 0;\ 9)$ и $X_2 = (6;\ 2;\ 0;\ 9;\ 0)$, т.е. в соответствии с выражениями (3.5) и (3.4) имеем: $X = \alpha X_1 + (1-\alpha)X_2$, где $0 \le \alpha \le 1$.▶**

▷ **5.8.** Решить симплексным методом задачу

$$F = 2x_1 - x_2 \to \max$$

при ограничениях:

$$\begin{cases} x_1 - x_2 \le 2, \\ 3x_1 - 2x_2 \le 6, \\ 6x_1 - 4x_2 \le 14, \end{cases}$$
$$x_1 \ge 0,\ x_2 \ge 0.$$

Решение. После введения дополнительных переменных, которые возьмем в качестве основных, получим:

I шаг. Основные переменные: x_3, x_4, x_5.
Неосновные переменные: x_1, x_2.

После преобразований получим:

$$\begin{cases} \boxed{x_3 = 2 - x_1 + x_2,} \\ x_4 = 6 - 3x_1 + 2x_2, \\ x_5 = 14 - 6x_1 + 4x_2. \end{cases}$$

$X_1 = (0;\ 0;\ 2;\ 6;\ 14)$ — допустимое базисное решение.

$F = 2x_1 - x_2$, $F_1 = F(X_1) = 0$. Критерий оптимальности на максимум не выполнен, поэтому переводим в основные переменную x_1, так как в выражение для F она входит с положительным коэффициентом: $x_1 = \min\{2; 6/3; 14/6\} = 2$. Оценочные отношения в двух первых уравнениях совпадают, поэтому в качестве разрешающего можно взять любое из них, например первое.

II шаг. Основные переменные: x_1, x_4, x_5.
Неосновные переменные: x_2, x_3.

Получим после преобразований:

$$\begin{cases} x_1 = 2 + x_2 - x_3, \\ x_4 = 0 - x_2 + 3x_3, \\ x_5 = 2 - 2x_2 + 6x_3. \end{cases}$$

$X_2 = (2;\ 0;\ 0;\ \mathbf{0};\ 2)$ — вырожденное базисное решение, основная компонента $x_4 = 0$.

Линейная функция цели, выраженная через неосновные переменные, имеет вид: $F = 4 + x_2 - 2x_3$. Переводя переменную x_2 в основные, получаем: $x_2 = \min\{\infty;\ 0;\ 1\} = 0$, поэтому на следующем шаге изменения функции цели не произойдет, $\Delta F = 0 \cdot 1 = 0$. Это нарушение принципа улучшения решения, который должен выполняться при использовании симплексного метода, в связи с чем уточним данный принцип: *каждый следующий шаг должен улучшить или, в крайнем случае, не ухудшить значение линейной функции.*

III шаг. Основные переменные: x_1, x_2, x_5.
Неосновные переменные: x_3, x_4.

После преобразований получим:

$$\begin{cases} x_1 = 2 + 2x_3 - x_4, \\ x_2 = 0 + 3x_3 - x_4, \\ x_5 = 2 + 2x_4. \end{cases}$$

$X_3 = (2;\ \mathbf{0};\ 0;\ 0;\ 2)$ — это базисное решение тоже вырождено. Покомпонентно оно совпадает с X_2, однако формально отличается набором основных переменных. Выражение линейной функции через неосновные переменные имеет вид: $F = 4 + x_3 - x_4$; $F(X_3) = 4$. (Решение предлагается продолжить читателю самостоятельно.)▶

Выполненный шаг хотя и не вызвал увеличения значения линейной функции, не является лишним, так как привел к новому базисному решению. Наличие "пустых" шагов ($\Delta F = 0$) может привести к так называемому "зацикливанию", т.е. возвращению к ранее найденному допустимому базисному решению (с тем же набором основных и неосновных переменных), и в этом случае процесс бесконечен. Избежать "зацикливания" можно с помощью определенных мер, которые в данном пособии не рассматриваются. Задачи с "зацикливанием" встречаются крайне редко, так как к нему приводит не только вырождение, но и сочетание его с другими специфическими условиями.

Вывод. *Если на каком-либо шаге наибольшее возможное значение переменной достигается в нескольких уравнениях одновременно*

(совпадают их оценочные отношения), то разрешающим является любое из них. На следующем шаге получаем вырожденное базисное решение, переход к очередному базисному решению может не изменить функцию цели ($\Delta F = 0$).

Замечание. Вырождение, полученное при оптимальном решении, может привести к альтернативному оптимуму даже при ненулевых коэффициентах при всех неосновных переменных в линейной функции (об этом упоминалось при рассмотрении случая альтернативного оптимума).

Отсутствие конечного оптимума

$$(F_{\max} = \infty \ \textbf{или} \ F_{\min} = -\infty)$$

▷ **5.9.** Решить симплексным методом задачу

$$F = 2x_1 - 3x_2 + 1 \to \min$$

при ограничениях:

$$\begin{cases} x_1 + x_2 \geq 4, & (\mathrm{I}) \\ 2x_1 - x_2 \geq 1, & (\mathrm{II}) \\ x_1 - 2x_2 \leq 1, & (\mathrm{III}) \end{cases}$$

$$x_1 \geq 0, \ x_2 \geq 0. \quad (\mathrm{IV, V})$$

Решение. Геометрическое решение этой задачи приведено в задаче **4.3, б** (*см.* рис. 4.5, *б*). На очередном шаге решения этой задачи симплексным методом получаем:

основные переменные x_1, x_2, x_5;

неосновные переменные x_3, x_4.

$$\begin{cases} x_1 = \dfrac{5}{3} + \dfrac{1}{3}x_3 + \dfrac{1}{3}x_4, \\ x_2 = \dfrac{7}{3} + \dfrac{2}{3}x_3 - \dfrac{1}{3}x_4, \\ x_5 = 4 + x_3 - x_4. \end{cases}$$

$X = \left(\dfrac{5}{3}; \ \dfrac{7}{3}; \ 0; \ 0; \ 4\right)$ **— базисное решение;** $F = -\dfrac{1}{3} - \dfrac{2}{3}x_3 + \dfrac{4}{5}x_4$;

минимум не достигнут, так как критерий оптимальности на это условие не выполнен: переменная x_3 имеет отрицательный коэф-

фициент в выражении для F. Определяем $x_3 = \min(\infty,\infty,\infty) = \infty$, так как в каждое из трех уравнений эта переменная входит с тем же знаком, что и свободный член. Уравнения не ограничивают рост x_3, поэтому и значение функции F неограниченно и отрицательно, т.е. $F_{\min} = -\infty$. ▶

Вывод. *Если на каком-либо шаге получаем, что во всех уравнениях системы бесконечны оценочные отношения той переменной, которая переводится в основные, то задача не имеет конечного оптимума* ($F_{\max} = \infty$, *если задача на отыскание максимума, и* $F_{\min} = -\infty$, *если задача на отыскание минимума*).

Подводя итоги, можно утверждать, что если система ограничений непротиворечива, то выполнение конечного числа последовательных шагов симплексного метода либо приводит к нахождению оптимального решения задачи (оно может быть неединственным), либо к установлению того факта, что линейная функция не имеет конечного оптимума.

5.6. Симплексные таблицы

Практические расчеты при решении реальных задач симплексным методом выполняются в настоящее время с помощью компьютеров. Однако если расчеты осуществляются без ЭВМ, то удобно использовать так называемые симплексные таблицы. Далее мы рассмотрим алгоритм их составления, не углубляясь в его подробное обоснование. Для определенности считаем, что решается задача на отыскание максимума.

I. После введения добавочных переменных систему уравнений и линейную функцию записываем в виде, который называется *расширенной системой*:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + x_{n+2} = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + x_{n+m} = b_m, \\ F - c_1x_1 - c_2x_2 - \ldots - c_nx_n = 0. \end{cases}$$

Предполагаем, что все добавочные переменные имеют тот же знак, что и свободные члены; в противном случае используем так называемый *М-метод*, который будет рассмотрен ниже.

II. Исходную расширенную систему заносим в первую симплексную таблицу. Последняя строка таблицы, в которой приведено уравнение для линейной функции цели, называется *оценочной.* В ней указываются коэффициенты функции цели с противоположным знаком: $b_i = - c_i$. В левом столбце таблицы записываем основные переменные (базис), в первой строке таблицы — все переменные (отмечая при этом основные), во втором столбце — свободные члены расширенной системы b_1, b_2, ..., b_m. Последний столбец подготовлен для оценочных отношений, необходимых при расчете наибольшего возможного значения переменной. В рабочую часть таблицы (начиная с третьего столбца и второй строки) занесены коэффициенты a_{ij} при переменных из расширенной системы. Далее таблица преобразуется по определенным правилам.

III. Проверяем выполнение критерия оптимальности при решении задачи на максимум — наличие в последней строке отрицательных коэффициентов $b_i < 0$ ($c_i > 0$). Если таких нет, то решение оптимально, достигнут max $F = c_0$ (в левом нижнем углу таблицы), основные переменные принимают значения a_{i0} (второй столбец), основные переменные равны 0, т.е. получаем оптимальное базисное решение.

IV. Если критерий оптимальности не выполнен, то наибольший по модулю отрицательный коэффициент $b_i < 0$ в последней строке определяет *разрешающий столбец s.*

Составляем оценочные ограничения каждой строки по следующим правилам:

1) ∞, если b_i и a_{is} имеют разные знаки;
2) ∞, если $b_i = 0$ и $a_{is} < 0$;
3) ∞, если $a_{is} = 0$;
4) 0, если $b_i = 0$ и $a_{is} > 0$;
5) $\left|\dfrac{b_i}{a_{is}}\right|$, если a_{i0} и a_{is} имеют одинаковые знаки.

Определяем $\min\limits_i \left\{\left|\dfrac{b_i}{a_{is}}\right|\right\}$. Если конечного минимума нет, то задача не имеет конечного оптимума ($F_{\max} = \infty$). Если минимум конечен, то выбираем строку q, на которой он достигается (любую, если их несколько), и называем ее ***разрешающей строкой.*** На пересечении разрешающих строки и столбца находится ***разрешающий элемент*** a_{qs}.

V. Переходим к следующей таблице по правилам:

а) в левом столбце записываем новый базис: вместо основной переменной x_q — переменную x_s;

б) в столбцах, соответствующих основным переменным, проставляем нули и единицы: 1 — против "своей" основной переменной, 0 — против "чужой" основной переменной, 0 — в последней строке для всех основных переменных;

в) новую строку с номером q получаем из старой делением на разрешающий элемент a_{qs};

г) все остальные элементы a'_{ij} вычисляем по *правилу прямоугольника*:

$$a'_{ij} = a_{ij} - \frac{a_{is}a_{qj}}{a_{qs}},$$

$$b'_i = b_i - \frac{a_{is}b_q}{a_{qs}}.$$

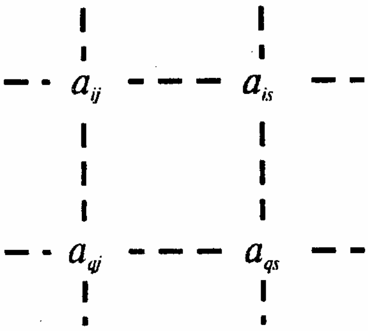


Рис. 5.4

Далее переходим к п. III алгоритма.

▷ **5.10.** Решить задачу об использовании ресурсов, приведенную в разд. 1.2, с помощью симплексных таблиц[1].

Р е ш е н и е . Расширенная система задачи имеет вид:

$$\begin{cases} x_1 + 3x_2 + x_3 = 18, \\ 2x_1 + x_2 + x_4 = 16, \\ x_2 + x_5 = 5, \\ 3x_1 + x_6 = 21. \end{cases}$$

Линейную функцию представим в виде $F - 2x_1 - 3x_2 = 0$.

Заполняем первую симплексную таблицу (табл. 5.1), в которой переменные x_3, x_4, x_5, x_6 основные. Последняя строка заполняется

---

[1] Эта задача уже решена геометрически в **4.1**, а в **5.1** — симплексным методом без использования таблиц.

коэффициентами линейной функции с противоположным знаком (см. п. II алгоритма).

Таблица 5.1

Базис	Свободный член	Переменные						Оценочное отношение
		x_1	x_2	x_3	x_4	x_5	x_6	
x_3	18	1	3	1	0	0	0	18/3
x_4	16	2	1	0	1	0	0	16
→ x_5	5	0	1	0	0	1	0	5 ←
x_6	21	3	0	0	0	0	1	∞
F	0	−2	−3	0	0	0	0	
			↑					

В соответствии с п. III алгоритма проверяем критерий оптимальности. В последней строке имеются отрицательные коэффициенты. Выбираем из них наибольший по модулю (−3); второй столбец разрешающий, переменная x_2 перейдет в основные (этот столбец отмечен стрелкой). В соответствии с п. IV находим оценочные отношения и $x_2 = \min\{18/3;\ 16;\ 5;\ \infty\} = 5$. Третья строка является разрешающей (отмечена горизонтальной стрелкой). На пересечении разрешающих строки и столбца стоит разрешающий элемент $a_{33} = 1$.

Строим табл. 5.2 по правилам п. V алгоритма:

а) в новом базисе основные переменные: x_3, x_4, x_2, x_6;

б) расставляем нули и единицы; например, в клетке, соответствующей основной переменной x_3 по столбцу и строке, ставим 1, а в клетке, соответствующей основной переменной x_3 по строке, а основной переменной x_2 — по столбцу, ставим 0 и т.п. В последней строке против всех основных переменных ставим 0. Третья строка получается делением на разрешающий элемент $a_{33} = 1$. Остальные клетки заполняем по правилу прямоугольника. Например: $b_1' = 18 - \frac{3 \cdot 5}{1} = 3$, $a_{11}' = 1 - \frac{3 \cdot 0}{1} = 1$ и т.д.

Получим вторую симплексную таблицу (табл. 5.2).

Таблица 5.2

	Базис	Свободный член	Переменные						Оценочное отношение
			x_1	x_2	x_3	x_4	x_5	x_6	
→	x_3	3	1	0	1	0	−3	0	3 ←
	x_4	11	2	0	0	1	−1	0	11/2
	x_2	5	0	1	0	0	1	0	∞
	x_6	21	3	0	0	0	0	1	7
	F	15	−2	0	0	0	3	0	
			↑						

Критерий оптимальности вновь не выполнен. Теперь первый столбец разрешающий; x_1 — переходит в основные, $\min\{3/1;\ 11/2;\ \infty;\ 7\} = 3$; первая строка разрешающая, a_{11} — разрешающий элемент.

Новая симплексная таблица примет вид табл. 5.3.

Таблица 5.3

	Базис	Свободный член	Переменные						Оценочное отношение
			x_1	x_2	x_3	x_4	x_5	x_6	
	x_1	3	1	0	1	0	−3	0	∞
→	x_4	5	0	0	−2	1	5	0	5/5 ←
	x_2	5	0	1	0	0	1	0	5/1
	x_6	12	0	0	−3	0	9	1	12/9
	F	21	0	0	2	0	−3	0	
							↑		

И на этот раз критерий оптимальности не выполнен; пятый столбец и вторая строка разрешающие, $a_{25} = 5$ — разрешающий элемент.

Переходим к табл. 5.4.

Таблица 5.4

Базис	Свободный член	Переменные						Оценочное отношение
		x_1	x_2	x_3	x_4	x_5	x_6	
x_1	6	1	0	−1/5	3/5	0	0	
x_5	1	0	0	−2/5	1/5	1	0	
x_2	4	0	1	2/5	−1/5	0	0	
x_6	3	0	0	3/5	−9/5	0	1	
F	24	0	0	4/5	3/5	0	0	

Критерий оптимальности выполнен, значит $F_{max} = 24$, оптимальное базисное решение (6; 4; 0; 0; 1; 3) совпадает с ранее полученным (*см.* решение задачи **5.1** симплексным методом).▶

5.7. Понятие об *М*-методе (методе искусственного базиса)

Выше был изложен алгоритм получения допустимого базисного решения в случае, когда первоначальное базисное решение недопустимо. Однако при расчете с помощью симплексных таблиц удобнее пользоваться так называемым *М-методом,* ***или*** *методом искусственного базиса.* Он заключается в следующем.

В каждое уравнение, дающее отрицательную компоненту в базисном решении, вводим свою новую неотрицательную искусственную переменную $y_1, y_2, ..., y_k$, которая имеет тот же знак, что и свободный член в правой части уравнения. В первой таблице включаем в число основных все искусственные переменные и те обычные добавочные переменные, которые определяют неотрицательные компоненты базисного решения. Составляем новую линейную функцию $T = F - M(y_1 + y_2 + ... + y_k)$, где M — произвольно большое число, и ищем ее максимум (T-задача). Назовем *М-функцией* выражение $M(y_1 + y_2 + ... + y_k)$. Справедлива ***теорема*** (доказательство здесь не приводится):

1. ***Если в оптимальном решении Т-задачи все искусственные переменные равны*** 0, ***то соответствующие значения остальных пере-***

менных дают оптимальное решение исходной задачи (т.е. $T_{max} = F_{max}$, если $y_1 = y_2 = \ldots = y_k = 0$, т.е. минимум М-функции равен 0).

2. *Если имеется оптимальное решение Т-задачи, в котором хотя бы одна из искусственных переменных отлична от 0, то система ограничений исходной задачи несовместна.*

3. *Если $T_{max} = \infty$, то исходная задача также неразрешима, причем либо $F_{max} = \infty$, либо условия задачи противоречивы.*

Из теоремы следует, что вначале следует найти минимум M-функции. Если он равен 0 и все искусственные переменные обращаются в 0, то далее можно отбросить эти переменные и решать исходную задачу, исходя из полученного допустимого базисного решения. На практике находят не минимум M-функции, а максимум $(-M)$-функции.

▷ **5.11.** Решить задачу **5.3** M-методом, используя симплексные таблицы.

Решение. Введем необходимое число искусственных переменных и столько же дополнительных строк в симплексной таблице.

Имеем $F = x_1 + 2x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 - x_2 + x_3 = -1, \\ x_1 - x_2 - x_4 = -3, \\ x_1 + x_5 = 3. \end{cases}$$

$X_1 = (0;\ 0;\ -1;\ 3;\ 5)$ — недопустимое базисное решение с одной отрицательной компонентой, поэтому в первое уравнение введем искусственную переменную y_1 с тем же знаком, что и свободный член:

$$\begin{cases} x_1 - x_2 + x_3 - y_1 = -1, \\ x_1 - x_2 - x_4 = -3, \\ x_1 + x_5 = 3 \end{cases}$$

или

$$\begin{cases} -x_1 + x_2 - x_3 + y_1 = 1, \\ -x_1 + x_2 + x_4 = 3, \\ x_1 + x_5 = 3, \end{cases}$$

$$T = x_1 + 2x_2 - My_1 \to \max.$$

Составляем первую симплексную таблицу (табл. 5.5).

Таблица 5.5

Базис	Свободный член	Переменные						Оценочное отношение
		x_1	x_2	x_3	x_4	x_5	y_1	
→ y_1	1	−1	1	−1	0	0	1	1 ←
x_4	3	−1	1	0	1	0	0	3
x_5	3	1	0	0	0	1	0	∞
F	0	−1	−2	0	0	0	0	max
$M_ф$	M	M	$-M$ ↑	M	0	0	M	max

Последняя строка — это $(-M)$-функция, т.е. $(-M_ф)y_1$. Заполняем ее, умножая строку y_1 на коэффициент $(-M)$. Проверяя выполнение критерия оптимальности при отыскании максимума $(-M)$-функции, определяем, что в последней строке имеется отрицательный элемент во втором столбце; значит он является разрешающим, переменная x_2 переходит в основные. Минимальное оценочное отношение в первой строке, она разрешающая. Переменная y_1 переходит в неосновные, обращается в 0 на следующем базисном решении и далее исключается из рассмотрения.

В соответствии с общим алгоритмом получаем табл. 5.6.

Таблица 5.6

Базис	Свободный член	Переменные					Оценочное отношение
		x_1	x_2	x_3	x_4	x_5	
x_2	1	−1	1	−1	0	0	
x_4	2	0	0	1	1	0	
x_5	3	1	0	0	0	1	
F	2	−3	0	−2	0	0	max
$-M_ф$	0	0	0	0	0	0	max

Последняя строка показывает, что критерий оптимальности выполнен; max $(-M_ф) = 0$, значит min $M_ф = 0$, далее эту строку можно не рассматривать. Получено допустимое базисное решение (0; 1; 0; 2; 3), начиная с которого решаем исходную задачу в соответствии с обычным алгоритмом. Читателям рекомендуется завершить ее самостоятельно, а также решить задачи **5.2**, **5.4**, **5.5**, **5.6** с помощью симплексных таблиц.▶

Замечание. При решении задачи на отыскание минимума линейной функции цели рекомендуется вместо Z_{min} находить $(-Z)_{max}$.

УПРАЖНЕНИЯ

Задачи **5.12–5.19** решить симплексным методом. Сравнить полученное решение с решением, найденным геометрически.

5.12. $F = 2x_1 - 6x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 + x_2 \ge 2, \\ -x_1 + 2x_2 \le 4, \\ x_1 + 2x_2 \le 8, \end{cases}$$

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

5.13. $F = 2x_1 - x_2 \to \min$

при ограничениях:

$$\begin{cases} x_1 + x_2 \ge 4, \\ -x_1 + 2x_2 \le 2, \\ x_1 + 2x_2 \le 10, \end{cases}$$

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

5.14. $F = x_1 + x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 - 4x_2 - 4 \le 0, \\ 3x_1 - x_2 \ge 0, \\ x_1 + x_2 - 4 \ge 0, \end{cases}$$

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

5.15. $F = 2x_1 - x_2 \to \min$

при ограничениях:

$$\begin{cases} x_1 + x_2 \ge 4, \\ 2x_1 - x_2 \ge 2, \\ -x_1 - 2x_2 \ge -10, \end{cases}$$

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

5.16. $F = x_1 - x_2 \to \max$

при ограничениях:

$$\begin{cases} -2x_1 + x_2 \le 2, \\ x_1 - 2x_2 \le -8, \\ x_1 + x_2 \le 5, \end{cases}$$

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

5.17. $F = 4x_1 - 2x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 - 2x_2 \le 0, \\ x_1 + 2x_2 \ge 2, \\ 2x_1 + x_2 \le 10, \end{cases}$$

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

5.18. Текст условия приведен в задаче **1.4.**

5.19. Текст условия приведен в задаче **1.5.**

Задачи **5.20—5.25** решить симплексным методом.

5.20. $F = 2x_1 - 13x_2 - 6x_3 \to \max$ при ограничениях:

$$\begin{cases} x_1 - x_2 + 3x_3 \ge 1, \\ x_1 - 2x_2 + x_3 \le -1, \end{cases}$$

$x_1 \ge 0,\ x_2 \ge 0,\ x_3 \ge 0.$

5.21. $F = -6x_1 + 10x_2 + 9x_3 + 8x_4 \to \min$ при ограничениях:

$$\begin{cases} -2x_1 + x_2 + x_3 \ge 2, \\ x_1 - x_2 - x_4 \le -1, \end{cases}$$

$x_j \ge 0,\ j = 1, 2, 3, 4.$

5.22. $F = x_1 + 3x_2 + 2x_3 \to \min$ при ограничениях:

$$\begin{cases} 3x_1 - 2x_2 + x_3 \ge 5, \\ x_1 + x_2 + 2x_3 \ge 10, \\ -2x_1 + 3x_2 - x_3 \ge 2, \end{cases}$$

$x_1 \ge 0,\ x_2 \ge 0,\ x_3 \ge 0.$

5.23. $F = x_1 + x_2 + x_3 + x_4 \to \min$ при ограничениях:

$$\begin{cases} x_1 - x_2 - 2x_3 + x_4 \ge 6, \\ -x_1 + x_3 \le 2, \\ 2x_2 - 3x_3 + 2x_4 \le 8, \end{cases}$$

$x_j \ge 0,\ j = 1, 2, 3, 4.$

5.24. $F = 4x_1 + 15x_2 + 12x_3 + 2x_4 \to \min$ при ограничениях:

$$\begin{cases} 2x_2 + 3x_3 + x_4 \ge 1, \\ x_1 + 3x_2 + x_3 - x_4 \ge 0, \end{cases}$$

$x_j \ge 0,\ j = 1, 2, 3, 4.$

5.25. $F = 14x_1 + 10x_2 + 14x_3 + 14x_4 \to \min$ при ограничениях:

$$\begin{cases} 4x_1 + 2x_2 + 2x_3 + x_4 \le 35, \\ x_1 + x_2 + 2x_3 + 3x_4 \le 30, \\ 3x_1 + x_2 + 2x_3 + x_4 \le 40, \end{cases}$$

$x_j \ge 0,\ j = 1, 2, 3, 4.$

Глава 6. ДВОЙСТВЕННЫЕ ЗАДАЧИ

Каждой задаче линейного программирования соответствует другая задача, называемая *двойственной* или *сопряженной* по отношению к исходной. Теория двойственности оказалась полезной для проведения качественных исследований задач линейного программирования.

6.1. Экономическая интерпретация задачи, двойственной задаче об использовании ресурсов

В гл. 1 рассмотрена задача об использовании ресурсов (экономико-математическая модель и содержательная интерпретация этой задачи I представлены в левой части табл. 6.1). В приведенной модели b_i ($i = 1, 2, ..., m$) обозначает запас ресурса S_i; a_{ij} — число единиц ресурса S_i, потребляемого при производстве единицы продукции P_j ($j = 1, 2, ..., n$); c_j — прибыль (выручка) от реализации единицы продукции P_j (или цена продукции P_j).

Предположим, что некоторая организация решила закупить ресурсы $S_1, S_2, ..., S_m$ предприятия и необходимо установить оптимальные цены на эти ресурсы $y_1, y_2, ..., y_m$.

Очевидно, что покупающая организация заинтересована в том, чтобы затраты на все ресурсы Z в количествах $b_1, b_2, ..., b_m$ по ценам соответственно $y_1, y_2, ..., y_m$ были минимальны, т.е

$$Z = b_1 y_1 + b_2 y_2 + \dots + b_m y_m \to \min .$$

С другой стороны, предприятие, продающее ресурсы, заинтересовано в том, чтобы полученная выручка была не менее той суммы, которую предприятие может получить при переработке ресурсов в готовую продукцию. На изготовление единицы продукции P_1 расходуется a_{11} единиц ресурса S_1, a_{21} единиц ресурса S_2, ..., a_{i1} единиц ресурса S_i, ..., a_{m1} единиц ресурса S_m по цене соответственно $y_1, y_2, ..., y_i, ..., y_m$. Поэтому для удовлетворения требований продавца затраты на ресурсы, потребляемые при изготовлении единицы продукции P_1, должны быть не менее ее цены c_1, т.е.

$$a_{11}y_1 + a_{21}y_2 + \dots + a_{m1}y_m \geq c_1.$$

Аналогично можно составить ограничения в виде неравенств по каждому виду продукции $P_1, P_2, ..., P_n$. Экономико-математическая модель и содержательная интерпретация полученной таким образом двойственной задачи II приведены в правой части табл. 6.1.

Таблица 6.1

Задача I (исходная)	Задача II (двойственная)
$F = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max$ **(6.1)** при ограничениях: $\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \leq b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \leq b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \leq b_m \end{cases}$ **(6.2)** и условии неотрицательности $x_1 \geq 0,\ x_2 \geq 0,\ \ldots,\ x_n \geq 0.$ **(6.3)** *Составить такой план выпуска продукции $X = (x_1, x_2, ..., x_n)$, при котором прибыль (выручка) от реализации продукции будет максимальной при условии, что потребление ресурсов по каждому виду продукции не превзойдет имеющихся запасов*	$Z = b_1y_1 + b_2y_2 + \ldots + b_my_m \to \min$ **(6.4)** при ограничениях: $\begin{cases} a_{11}y_1 + a_{21}y_2 + \ldots + a_{m1}y_1 \geq c_1, \\ a_{12}y_1 + a_{22}y_2 + \ldots + a_{m2}y_2 \geq c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}y_1 + a_{2n}y_2 + \ldots + a_{mn}y_m \geq c_n, \end{cases}$ **(6.5)** и условии неотрицательности $y_1 \geq 0,\ y_2 \geq 0,\ \ldots,\ y_m \geq 0.$ **(6.6)** *Найти такой набор цен (оценок) ресурсов $Y = (y_1, y_2, ..., y_m)$, при котором общие затраты на ресурсы будут минимальными при условии, что затраты на ресурсы при производстве каждого вида продукции будут не менее прибыли (выручки) от реализации этой продукции*

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

6.2. Взаимно двойственные задачи линейного программирования и их свойства

Рассмотрим формально две задачи I и II линейного программирования, представленные в табл. 6.1, абстрагируясь от содержательной интерпретации параметров, входящих в их экономико-математические модели. Обе задачи обладают следующими свойствами:

1. В одной задаче ищут максимум линейной функции, в другой — минимум.

2. Коэффициенты при переменных в линейной функции одной задачи являются свободными членами системы ограничений в другой.

3. Каждая из задач задана в стандартной форме, причем в задаче максимизации все неравенства вида "≤", а в задаче минимизации — все неравенства вида "≥".

4. Матрицы коэффициентов при переменных в системах ограничений обеих задач являются транспонированными друг к другу:

$$\text{для задачи I } A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

$$\text{для задачи II } A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{pmatrix}.$$

5. ***Число неравенств в системе ограничений одной задачи совпадает с числом переменных в другой задаче.***
6. ***Условия неотрицательности переменных имеются в обеих задачах.***

Две задачи I и II линейного программирования, обладающие указанными свойствами, называются ***симметричными взаимно двойственными задачами.*** В дальнейшем для простоты будем называть их просто *двойственными задачами.*

Исходя из определения, можно предложить следующий *алгоритм составления двойственной задачи.*

1. Привести все неравенства системы ограничений исходной задачи к одному смыслу: если в исходной задаче ищут максимум линейной функции, то все неравенства системы ограничений привести к виду "≤", а если минимум — к виду "≥". Для этого неравенства, в которых данное требование не выполняется, умножить на −1.
2. Составить расширенную матрицу системы A_1, в которую включить матрицу коэффициентов при переменных A, столбец свободных членов системы ограничений и строку коэффициентов при переменных в линейной функции.
3. Найти матрицу A_1', транспонированную к матрице A_1.
4. Сформулировать двойственную задачу на основании полученной матрицы A_1' и условия неотрицательности переменных.

▷ **6.1.** Составить задачу, двойственную следующей задаче:

$$F = -x_1 + 2x_2 \to \max$$

при ограничениях:

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

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

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

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

2. Составим расширенную матрицу системы

$$A_1 = \begin{pmatrix} -2 & 1 & -1 \\ -1 & 4 & 24 \\ 1 & -1 & 3 \\ -1 & -1 & -5 \\ -1 & 2 & F \end{pmatrix}.$$

3. Найдем матрицу A_1', транспонированную к A.

$$A_1' = \begin{pmatrix} -2 & -1 & 1 & -1 & -1 \\ 1 & 4 & -1 & -1 & 2 \\ -1 & 24 & 3 & -5 & Z \end{pmatrix}$$

4. Сформулируем двойственную задачу:

$$Z = -y_1 + 24y_2 + 3y_3 - 5y_4 \to \min$$

при ограничениях

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

$$y_1 \geq 0,\ y_2 \geq 0,\ \ldots,\ y_4 \geq 0. \blacktriangleright$$

6.3. Первая теорема двойственности

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

Основное неравенство теории двойственности. Пусть имеется пара двойственных задач I и II. Покажем, что *для любых допустимых решений* $X = (x_1, x_2, \ldots, x_n)$ *и* $Y = (y_1, y_2, \ldots, y_m)$ *исходной и двойственной задач справедливо неравенство*

$$F(X) \leq Z(Y) \text{ или } \sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i. \tag{6.7}$$

□ Умножив неравенства системы ограничений (6.2) исходной задачи $\sum_{j=1}^{n} a_{ij} x_j \leq b_i$ $(i = 1, 2, \ldots, m)$ соответственно на переменные y_1, y_2,

..., y_i, ..., y_m и сложив правые и левые части полученных неравенств, имеем

$$\sum_{i=1}^{m} y_i \sum_{j=1}^{n} a_{ij} x_j \le \sum_{i=1}^{m} b_i y_i. \qquad \textbf{(6.8)}$$

Аналогично преобразовав систему ограничений (6.5) двойственной задачи $\sum_{i=1}^{m} a_{ij} y_i \ge c_j$ ($j = 1, 2, ..., n$) путем умножения обеих частей ее неравенства на переменные x_1, x_2, ..., x_j, ..., x_n и последующего их сложения, получим

$$\sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} y_i \ge \sum_{j=1}^{n} c_j x_j. \qquad \textbf{(6.9)}$$

Так как левые части неравенств (6.8) и (6.9) представляют одно и то же выражение $\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_j y_i$, то в силу свойства транзитивности неравенств получим доказываемое неравенство (6.7).■

Теперь можно перейти к признакам оптимальности решений.

Достаточный признак оптимальности. Сформулируем теорему.

Теорема. *Если* $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ *и* $Y^* = (y_1^*, y_2^*, \dots, y_m^*)$ — *допустимые решения взаимно двойственных задач, для которых выполняется равенство*

$$F(X^*) = Z(Y^*), \qquad \textbf{(6.10)}$$

то X^* — *оптимальное решение исходной задачи I, а* Y^* — *двойственной задачи II.*

□ Пусть X_1 — любое допустимое решение исходной задачи I. Тогда на основании основного неравенства (6.7) получим $F(X_1) \le Z(Y^*)$. Однако X_1 — произвольное решение задачи I, отсюда в силу равенства (6.10) следует, что $F(X_1) \le F(X^*)$, т.е. X^* — оптимальное решение задачи I. Аналогично доказывается, что решение оптимально для задачи II. ■

Кроме достаточного признака оптимальности взаимно двойственных задач существуют и другие важные соотношения между их решениями. Прежде всего возникают вопросы: всегда ли для каждой пары двойственных задач есть одновременно оптимальные решения; возможна ли ситуация, когда одна из двойственных задач имеет решение, а другая нет. Ответ на эти вопросы дает следующая теорема.

Первая (основная) теорема двойственности. *Если одна из взаимно двойственных задач имеет оптимальное решение, то его имеет и другая, причем оптимальные значения их линейных функций равны:*

$$F_{\max} = Z_{\min} \text{ или } F(X^*) = Z(Y^*). \qquad \mathbf{(6.11)}$$

Если линейная функция одной из задач не ограничена, то условия другой задачи противоречивы.

Из первой части утверждения теоремы (которую мы принимаем без доказательства) следует, что равенство (6.10) является не только достаточным признаком оптимальности решений (доказанным выше), но и необходимым признаком оптимальности решений взаимно двойственных задач.

□ Утверждение второй части легко доказывается методом от противного. Предположим, что в исходной задаче линейная функция не ограничена, т.е. $F_{\max} = \infty$, а условия двойственной задачи не являются противоречивыми, т.е. существует хотя бы одно допустимое решение $Y = (y_1, y_2, ..., y_m)$. Тогда в силу основного неравенства теории двойственности (6.7) $F(X) \le Z(Y)$, что противоречит условию неограниченности $F(X)$. Следовательно, при $F_{\max} = \infty$ в исходной задаче допустимых решений в двойственной задаче быть не может. ■

Рассмотрим примеры, подтверждающие справедливость первой теоремы двойственности.

▷ **6.2.** Даны две взаимно двойственные задачи:

I. $F = 2x_1 + 3x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 + 3x_2 \le 18, \\ 2x_1 + x_2 \le 16, \\ x_2 \le 5, \\ 3x_1 \le 21, \end{cases}$$

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

II. $Z = 18y_1 + 16y_2 + 5y_3 + 21y_4 \to \min$

при ограничениях:

$$\begin{cases} y_1 + 2y_2 + 3y_4 \ge 2, \\ 3y_1 + y_2 + y_3 \ge 3, \end{cases}$$

$y_i \ge 0\ (i = 1, 2, 3, 4).$

Задача I об использовании ресурсов (*см.* разд. 1.2) и двойственная ей задача II были решены ранее (*см.* разд. 5.2, 5.3) и получены оптимумы линейных функций $F_{max} = 24$ для задачи I, и $Z_{min} = 24$ для задачи II, т.е заключение первой части основной теоремы двойственности (6.11) верно.▶

Экономический смысл первой (основной) теоремы двойственности. *План производства* $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ *и набор цен (оценок) ресурсов* $Y^* = (y_1^*, y_2^*, \dots, y_m^*)$ *оказываются оптимальными тогда и только тогда, когда прибыль (выручка) от продукции, найденная при "внешних" (известных заранее) ценах* $c_1, c_2, \dots, c_n$, *равна затратам на ресурсы по "внутренним" (определяемым только из решения задачи) ценам* $y_1, y_2, \dots, y_m$. Для всех же других планов X и Y обеих задач в соответствии с основным неравенством (6.7) теории двойственности прибыль (выручка) от продукции всегда меньше (или равна) затрат на ресурсы.

Так, в задаче **6.2** оптимумы прибыли от продукции F_{max} и затрат на ресурсы Z_{min} равны 24 руб.[1], для всех остальных планов $F(X) \le 24$, $Z(Y) \ge 24$.

Экономический смысл первой теоремы двойственности можно интерпретировать и так: предприятию безразлично, производить ли продукцию по оптимальному плану $X^* = (x_1^*, x_2^*, \dots, x_n^*)$ и получить максимальную прибыль (выручку) F_{max} либо продавать ресурсы по оптимальным ценам $Y^* = (y_1^*, y_2^*, \dots, y_m^*)$ и возместить от продажи равные ей минимальные затраты на ресурсы Z_{min}.

▷ **6.3.** Даны две взаимно двойственные задачи:

I. $Z = -8y_1 + 2y_2 \to \min$	II. $F = x_1 - x_2 \to \max$
при ограничениях:	при ограничениях:
$\begin{cases} -y_1 + 2y_2 \ge 1, \\ -2y_1 + y_2 \ge -1, \end{cases}$	$\begin{cases} x_1 + 2x_2 \ge 8, \\ 2x_1 + x_2 \le 2, \end{cases}$
$y_1 \ge 0,\ y_2 \ge 0.$	$x_1 \ge 0,\ x_2 \ge 0.$

Предлагаем читателю самостоятельно убедиться (симплексным методом или геометрически) в том, что в исходной задаче I ли-

[1] **Напоминаем, что в рассматриваемой задаче все цифры условные.**

нейная функция не ограничена ($Z_{\min} = -\infty$), а в двойственной задаче условия противоречивы, т.е. заключение второй части основной теоремы двойственности выполняется.▶

Замечание. Утверждение, обратное по отношению ко второй части основной теоремы двойственности, в общем случае неверно, т.е. из того, что условия исходной задачи противоречивы, не следует, что линейная функция двойственной задачи не ограничена.

▷ **6.4.** Даны две взаимно двойственные задачи:

I. $F = 3x_1 + 5x_2 \to \max$

при ограничениях:

$$\begin{cases} 3x_1 - 4x_2 \le 5, \\ -2x_1 \ge -7, \end{cases}$$

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

II. $Z = 5y_1 - 7y_2 \to \min$

при ограничениях:

$$\begin{cases} 3y_1 + 2y_2 \ge 3, \\ -4y_1 \ge 5, \end{cases}$$

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

Предлагаем читателю убедиться (симплексным методом или геометрически) в том, что в каждой из задач отсутствуют допустимые решения, т.е. условия обеих задач противоречивы.▶

6.4. Вторая теорема двойственности

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

Пусть даны две взаимно двойственные задачи I — (6.1)—(6.3) и задачи II — (6.4)—(6.6). Если каждую из этих задач решать симплексным методом, то необходимо привести их к каноническому виду, для чего в систему ограничений (6.2) задачи I (в краткой записи $\sum_{j=1}^{n} a_{ij}x_j \le b_i,$ $i = 1, 2, ..., m$) следует ввести m неотрицательных переменных $x_{n+1}, x_{n+2}, ..., x_{n+i}, ..., x_{n+m}$, а в систему ограничений (6.5) задачи II ($\sum_{i=1}^{m} a_{ij} y_i \geq c_j$, $j = 1, 2, ..., n$) — n неотрицательных переменных y_{m+1}, y_{m+2}, ..., y_{m+j}, ..., y_{n+m}, где $i(j)$ — номер неравенства, в которое введена дополнительная переменная $x_{n+i} \geq 0$ ($y_{m+j} \geq 0$).

Системы ограничений каждой из взаимно двойственных задач примут вид:

$$\sum_{j=1}^{n} a_{ij} x_j + x_{n+i} = b_i, \; i = 1,2,\dots,m. \qquad \textbf{(6.12)}$$

$$\sum_{i=1}^{m} a_{ij} y_i - y_{m+j} = c_j, \; j = 1,2,\dots,n. \qquad \textbf{(6.13)}$$

Установим соответствие между первоначальными переменными одной из двойственных задач и дополнительными переменными другой задачи (табл. 6.2).

Таблица 6.2

Переменные исходной задачи I									
Первоначальные					Дополнительные				
x_1	x_2	...	x_j	... x_n	x_{n+1}	x_{n+2}	...	x_{n+i}	... x_{n+m}
$\updownarrow$	$\updownarrow$		$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$		$\updownarrow$	$\updownarrow$ **(6.14)**
y_{m+1}	y_{m+2}	...	y_{m+j}	... y_{m+n}	y_1	y_2	...	y_j	... y_m
Дополнительные					Первоначальные				
Переменные двойственной задачи II									

Теорема. *Положительным (ненулевым) компонентам оптимального решения одной из взаимно двойственных задач соответствуют нулевые компоненты оптимального решения другой задачи, т.е. для любых $i = 1, 2, ..., m$ и $j = 1, 2, ..., n$: если $x_j^* > 0$, то $y_{m+j}^* = 0$; если $x_{n+i}^* > 0$, то $y_i^* = 0$, и аналогично, если $y_i^* > 0$, то $x_{n+i}^* = 0$; если $y_{m+j}^* > 0$, то $x_j^* = 0$.*

□ Выразим дополнительные переменные из систем ограничений (6.12) исходной задачи I и (6.13) двойственной задачи II, представленных в каноническом виде

$$x_{n+i} = b_i - \sum_{j=1}^{n} a_{ij}x_j, \; i = 1,2,\ldots,m, \tag{6.15}$$

$$y_{m+j} = \sum_{i=1}^{m} a_{ij}y_i - c_j, \; j = 1;2,\ldots,n. \tag{6.16}$$

Умножая каждое равенство системы (6.15) на соответствующие переменные $y_i \geq 0$ и складывая полученные равенства, найдем

$$\sum_{i=1}^{m} x_{n+j}y_i = \sum_{i=1}^{m} b_i y_i - \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_j y_i. \tag{6.17}$$

Аналогично, умножая каждое равенство системы (6.16) на соответствующие переменные $x_j \geq 0$ и складывая полученные равенства, найдем

$$\sum_{j=1}^{n} x_j y_{m+j} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_j y_i - \sum_{j=1}^{n} c_j x_j. \tag{6.18}$$

Равенства (6.17) и (6.18) будут справедливы для любых допустимых значений переменных, в том числе и для оптимальных значений $x_j^*, x_{n+i}^*, y_i^*, y_{m+j}^*$. В силу первой теоремы двойственности (6.11) $F(X^*) = Z(Y^*)$ или $\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$, поэтому из записи правых частей равенств (6.17) и (6.18) следует, что они должны отличаться только знаком. С другой стороны, из неотрицательности выражений $\sum_{j=1}^{m} x_{n+1}y_i$ и $\sum_{j=1}^{n} x_j y_{m+j}$, входящих в выражения (6.17) и (6.18), следует, что те же правые части этих равенств должны быть неотрицательны.

Эти условия могут выполняться одновременно только при равенстве этих правых частей для оптимальных значений переменных нулю:

$$\begin{cases} \sum_{i=1}^{m} x_{n+1}^{*} y_i^{*} = 0, \\ \sum_{j=1}^{n} x_j^{*} y_{m+j}^{*} = 0. \end{cases} \qquad \textbf{(6.19)}$$

В силу условия неотрицательности переменных каждое из слагаемых в равенствах (6.19) должно равняться нулю:

$$\begin{cases} x_{n+i}^{*} y_i^{*} = 0, \ i = 1,2,\dots,m, \\ x_j^{*} y_{m+j}^{*} = 0, \ j = 1,2,\dots,n, \end{cases}$$

откуда и вытекает заключение теоремы.■

Из доказанной теоремы следует важный вывод о том, что введенное ранее *соответствие (6.14) между переменными взаимно двойственных задач при достижении оптимума (т.е. на последнем шаге решения каждой задачи симплексным методом) представляет соответствие между* ***основными*** *(как правило, не равными нулю) переменными одной из двойственных задач и* ***неосновными*** *(равными нулю) переменными другой задачи, когда они образуют допустимые базисные решения.*

Рассмотренная теорема является следствием следующей теоремы.

Вторая теорема двойственности[1]. *Компоненты оптимального решения двойственной задачи равны абсолютным значениям коэффициентов при соответствующих переменных линейной функции исходной задачи, выраженной через неосновные переменные ее оптимального решения.*

▷ **6.5.** Убедиться в справедливости второй теоремы двойственности для взаимно двойственных задач I и II, приведенных в задаче **6.2**.

---

[1] В ряде работ по математическому программированию под второй теоремой двойственности фактически понимается предыдущая теорема (с. 108).

Решение. На основании выражения (6.14) установим следующее соответствие между переменными:

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

В гл. 5 обе задачи были решены симплексным методом. На последнем шаге решения каждой задачи (*см.* с. 72 и с. 75) получено:

в исходной задаче I

$$F = 24 - \frac{4}{5}x_3 - \frac{3}{5}x_4. \qquad \textbf{(6.20)}$$

$F(X^*) = F_{\max} = 24$ при оптимальном базисном решении $X^* = (6;\ 4;\ 0;\ 0;\ 1;\ 3)$.

в двойственной задаче II

$$Z = 24 + y_3 + 3y_4 + +6y_5 + 4y_6. \qquad \textbf{(6.21)}$$

$Z(Y^*) = Z_{\min} = 24$ при оптимальном базисном решении $Y^* = (\frac{4}{5};\ \frac{3}{5};\ 0;\ 0;\ 0;\ 0)$.

Компоненты оптимального решения двойственной задачи $y_1^* = \frac{4}{5}$, $y_2^* = \frac{3}{5}$, $y_3^* = 0$, $y_4^* = 0$, $y_5^* = 0$, $y_6^* = 0$ равны (по абсолютной величине) коэффициентам при *соответствующих* переменных линейной функции (6.20), которую можно представить в виде $F = 24 - \frac{4}{5}x_3 - \frac{3}{5}x_4 - 0 \cdot x_5 - 0 \cdot x_6 - 0 \cdot x_1 - 0 \cdot x_2$, а компоненты оптимального решения исходной задачи $x_1^* = 6$, $x_2^* = 4$, $x_3^* = 0$, $x_4^* = 0$, $x_5^* = 1$, $x_6^* = 3$ равны коэффициентам при *соответствующих* переменных линейной функции (6.21), которую можно представить в виде $Z = 24 + 6y_5 + 4y_6 + 0 \cdot y_1 + 0 \cdot y_2 + 1 \cdot y_3 + 3y_4$. ▶

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

С помощью теорем двойственности можно, решив симплексным методом исходную задачу, найти оптимум и оптимальное решение двойственной задачи.

▷ **6.6.** Решить симплексным методом задачу

$$F = -x_1 + 2x_2 \to \max$$

при ограничениях:

$$\begin{cases} -2x_1 + x_2 \le -1, \\ x_1 - 4x_2 \ge -24, \\ x_1 - x_2 \le 3, \\ x_1 + x_2 \ge 5, \end{cases}$$
$$x_1 \ge 0,\ x_2 \ge 0$$

и найти оптимум и оптимальное решение двойственной задачи, используя теоремы двойственности.

Решение. Решая задачу симплексным методом, получим на последнем шаге решения (рекомендуем читателю убедиться в этом самостоятельно): $F = 10 - \frac{2}{7}x_3 - \frac{3}{7}x_4$, т.е. $F_{\max} = 10$ при оптимальном базисном решении $X^* = (4;\ 7;\ 0;\ 0;\ 6;\ 6)$.

В данной задаче соответствие (6.14) между переменными примет вид:

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

На основании первой теоремы двойственности $Z_{\min} = F_{\max} = 10$. На основании второй теоремы двойственности оптимальное решение двойственной задачи $Y^* = (2/7;\ 3/7;\ 0;\ 0;\ 0;\ 0)$, так как $y_1^* = 2/7$, т.е. коэффициенту при соответствующей переменной x_3 в выражении линейной функции $F(x)$, $y_2^* = 3/7$, т.е. коэффициенту при переменной x_4, $y_3^* = y_4^* = y_5^* = y_6^* = 0$ (из-за отсутствия соответствующих переменных x_5, x_6, x_1, x_2 в выражении $F(x)$ их коэффициенты равны нулю). Итак, в двойственной задаче $Z_{\min} = 10$ при $Y^* = (2/7;\ 3/7;\ 0;\ 0;\ 0;\ 0)$.▶

Метод, при котором вначале симплексным методом решается двойственная задача, а затем оптимум и оптимальное решение исходной задачи находятся с помощью теорем двойственности, называется ***двойственным симплексным методом.*** Этот метод бывает выгодно применять, когда первое базисное решение исходной задачи недопустимое или, например, когда число ее ограничений m больше числа переменных n.

Компоненты оптимального решения двойственной задачи называются ***оптимальными (двойственными) оценками*** *исходной задачи.* Академик Л.В.Канторович назвал их ***объективно обусловленными оценками***[1].

Для выяснения смысла этих оценок вернемся к задаче I об использовании ресурсов и двойственной ей задаче II (*см.* задачу **6.2**). Компоненты оптимальных решений этих задач, приведенные в задаче **6.5**, даны в табл. 6.3.

Таблица 6.3

Компоненты оптимального решения исходной задачи I					
Число единиц продукции		Остатки ресурсов, единиц			
P_1	P_2	S_1	S_2	S_3	S_4
$x_1^* = 6$	$x_2^* = 4$	$x_3^* = 0$	$x_4^* = 0$	$x_5^* = 1$	$x_6^* = 3$
$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$
$y_5^* = 0$	$y_6^* = 0$	$y_1^* = 4/5$	$y_2^* = 3/5$	$y_3^* = 0$	$y_4^* = 0$
Превышение затрат на ресурсы над ценой реализации		**Объективно обусловленные оценки ресурсов (условные цены ресурсов)**			
Компоненты оптимального решения двойственной задачи II					

[1] **В литературе их еще называют** ***скрытыми доходами, маргинальными оценками, разрешающими множителями.***

В табл. 6.3 дополнительные переменные исходной задачи I x_3, x_4, x_5, x_6, представляющие согласно выражению (6.15) разность между запасами b_i ресурсов S_1, S_2, S_3, S_4 и их потреблением, выражают ***остатки ресурсов***, а дополнительные переменные двойственной задач II y_5, y_6, представляющие в соответствии с выражением (6.16) разность между затратами на ресурсы для производства из них единицы продукции и ценами c_j продукции P_1, P_2, выражают ***превышение затрат над ценой***.

Ресурсы S_1, S_2 по оптимальному плану полностью использованы ($x_3^* = 0$, $x_4^* = 0$), и объективно обусловленные оценки этих ресурсов ненулевые ($y_1^* = \frac{4}{5}$; $y_2^* = \frac{3}{5}$). Ресурсы S_3, S_4 не полностью используются в оптимальном плане ($x_5^* = 1$, $x_6^* = 3$) и объективно обусловленные оценки этих ресурсов нулевые ($y_3^* = 0$, $y_4^* = 0$).

Таким образом, *объективно обусловленные оценки ресурсов определяют степень дефицитности ресурсов: по оптимальному плану производства дефицитные (т.е. полностью используемые) ресурсы получают ненулевые оценки, а недефицитные — нулевые оценки.*

По оптимальному плану в исходной задаче следует производить оба вида продукции ($x_1^* = 6$, $x_2^* = 4$) и превышение затрат на ресурсы над ценой реализации равно нулю ($y_5^* = 0$, $y_6^* = 0$). Если бы затраты на ресурсы превышали цену изготавливаемой из них продукции, например, продукции P_2, т.е. если бы $y_6^* > 0$, то на основании теоремы на с. 108 оптимальное значение соответствующей переменной $x_2^* = 0$, и в этом случае по оптимальному плану производить продукцию P_2 не следовало.

Итак, ***в оптимальный план производства могут попасть только рентабельные, неубыточные виды продукции*** (правда, критерий рентабельности здесь своеобразный: цена продукции не превышает затраты на потребляемые при ее изготовлении ресурсы, а в точности равна им).

Для выяснения того, что показывают численные значения объективно обусловленных оценок ресурсов, докажем следующую теорему.

Третья теорема двойственности. *Компоненты оптимального решения двойственной задачи равны значениям частных производных линейной функции $F_{max}(b_1, b_2, ..., b_m)$ по соответствующим аргументам, т.е.*

$$\frac{\partial F_{max}}{\partial b_i} = y_i^*, \ (i = 1, 2, ..., m). \qquad \textbf{(6.22)}$$

□ Докажем эту теорему на примере задачи I (*см.* выражения (6.1)—(6.3) об использовании ресурсов и двойственной ей задачи II (*см.* выражения (6.4)—(6.6)).

В соответствии с выражением (6.4) оптимальное (минимальное) значение линейной функции

$$Z_{\min} = Z^* = \sum_{i=1}^{m} b_i y_i^*.$$

Изменим незначительно запас i-го ресурса b_i на величину Δb_i, при этом оптимальное решение двойственной задачи не изменится (на рис. 6.1 для простейшего случая двух переменных опти-

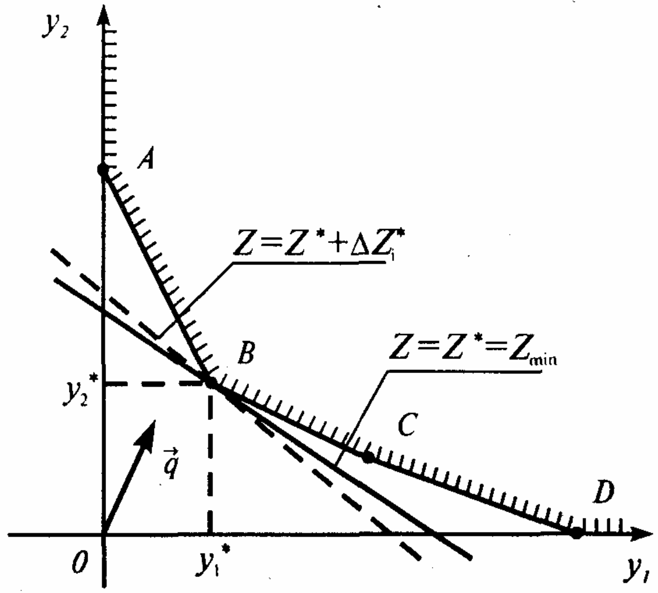


Рис. 6.1

мальное решение соответствует точке B), т.е. $y_1^*, y_2^*, \ldots, y_m^*$ прежние, а изменится только линейная функция Z (на рис. 6.1 для случая $Z = b_1 y_1 + b_2 y_2$ изменится наклон линии уровня).

Тогда изменение минимальных затрат на ресурсы

$$\Delta Z_i^* = Z_{2i}^* - Z_{1i}^* = [b_1 y_1^* + b_2 y_2^* + \ldots + (b_i + \Delta b_i) y_i^* + \ldots + b_m y_m^*] -$$
$$-(b_1 y_1^* + b_2 y_2^* + \ldots + b_i y_i^* + \ldots + b_m y_m^*) = \Delta b_i y_i^*,$$

т.е. $\Delta Z_i^* = \Delta b_i y_i^*$. А так как по первой теореме двойственности $F_{\max} = Z_{\min}$ или $F^* = Z^*$, то максимальная прибыль (выручка) увеличится на $\Delta F_{\max} = \Delta F_i^* = \Delta Z_i^* = \Delta b_i y_i^*$,
откуда

$$y_i^* = \frac{\Delta F_{\max}}{\Delta b_i}. \qquad \textbf{(6.23)}$$

Важно, что последнее соотношение верно при достаточно малом изменении Δb_i, ибо в противном случае могут существенно измениться наклон линии уровня (*см.* рис. 6.1) и оптимум, так как минимум линейной функции Z^* может перейти в другую угловую точку, соответственно с другим оптимальным решением $Y_1^* = (y_1^*, y_2^*, \ldots, y_m^*)$.

Понятие достаточной малости приращения запаса ресурса Δb_i можно формализовать, потребовав, чтобы Δb_i стремилось к нулю. При $\Delta b_i \to 0$ из соотношения (6.23) и получается доказываемое утверждение (6.22).■

Из соотношения (6.23) следует, что *объективно обусловленные оценки ресурсов показывают, на сколько денежных единиц изменится максимальная прибыль (выручка) от реализации продукции при изменении запаса соответствующего ресурса на одну единицу.*

▷ **6.7.** В задаче I из **6.2** пояснить смысл значений объективно обусловленных оценок ресурсов.

Решение. В задаче **6.5** получены объективно обусловленные оценки ресурсов рассматриваемой задачи: $y_1^* = 4/5$; $y_2^* = 3/5$; $y_3^* = y_4^* = 0$, т.е. при увеличении (уменьшении) запаса ресурсов S_1 или S_2 на 1 единицу максимальная прибыль (выручка) увеличится (уменьшится) соответственно на 4/5 и 3/5 руб., а при изменении запаса ресурсов S_3 или S_4 — не изменится.▶

Двойственные оценки могут служить инструментом анализа и принятия правильных решений в условиях постоянно меняющегося производства. Так, например, *с помощью объективно обусловленных оценок ресурсов возможно сопоставление оптимальных условных затрат и результатов производства.*

▷ **6.8.** В задаче I из **6.2**, в результате решения которой получен оптимальный план выпуска продукции P_1 и P_2, появилась возможность выпуска еще одного вида продукции P_3. Затраты на единицу продукции P_3 составляют: $a_{13} = 3$ единицы ресурса S_1, $a_{23} = 2$ единицы S_2, $a_{33} = 4$ единицы S_3 и $a_{43} = 1$ единица ресурса S_4; цена единицы продукции P_3 равна $c_3 = 3$ руб. Установить, даст ли прибыль включение в план выпуска дополнительно продукции P_3. Какой должна быть прибыль от единицы продукции P_3 (цена), чтобы ее производство было рентабельным?

Решение. Можно включить в условия задачи продукцию P_3 и заново решить задачу, однако это потребует новых затрат (трудовых, стоимостных, временных). Но необходимости в этом нет, так как известны объективно обусловленные оценки ресурсов. Действительно, сопоставим дополнительные затраты на ресурсы в расчете на единицу продукции P_3 с ценой ее реализации. Первая величина, равная $a_{13}y_1^* + a_{23}y_2^* + a_{33}y_3^* + a_{43}y_4^* = 3 \cdot 4/5 + $ $+ 2 \cdot 3/5 + 4 \cdot 0 + 1 \cdot 0 = 3{,}6$ руб., больше цены продукции $c_3 = 3$ руб., поэтому выпуск продукции P_3 не следует включать в план выпуска, и необходимость повторного решения задачи в изменившихся условиях отпадает. А чтобы производство продукции P_3 стало рентабельным, очевидно, ее цена должна составлять не менее 3,6 руб.▶

Объективно обусловленные оценки ресурсов позволяют судить об эффекте не любых, а лишь сравнительно небольших изменений ресурсов. При резких изменениях сами оценки могут стать другими, что приведет к невозможности их использования для анализа эффективности производства.

▷ **6.9.** **Для задачи I из 6.2 найти интервалы устойчивости (неизменности) двойственных оценок по отношению к изменениям запасов ресурсов каждого вида. Изменятся ли эти оценки, если увеличить запасы каждого из ресурсов на 10 единиц: а) в отдельности; б) одновременно? Найти соответствующее изменение максимальной прибыли (выручки) от реализации продукции.**

Решение. Предположим, что запасы ресурсов S_1, S_2, S_3, S_4, равные первоначально 18, 16, 5 и 21 единице, изменились соответственно на величины Δb_1, Δb_2, Δb_3 и Δb_4. Тогда затраты на ресурсы в соответствии с выражением (6.4) составят $Z = (18 + \Delta b_1)y_1 + (16 + \Delta b_2)y_2 + (5 + \Delta b_3)y_3 + (21 + \Delta b_4)y_4$.

Заменяя переменные y_1 и y_2 их выражениями через неосновные переменные оптимального решения $y_1 = \frac{4}{5} - \frac{2}{5}y_3 + \frac{3}{5}y_4 - \frac{1}{5}y_5 + \frac{2}{5}y_6$, $y_2 = \frac{3}{5} + \frac{1}{5}y_3 - \frac{9}{5}y_4 + \frac{3}{5}y_5 - \frac{1}{5}y_6$ — *см.* с. 74), получим после преобразований

$$Z = \left(24 + \frac{4}{5}\Delta b_1 + \frac{3}{5}\Delta b_2\right) + \left(1 - \frac{2}{5}\Delta b_1 + \frac{1}{5}\Delta b_2 + \Delta b_3\right)y_3 + \\ + \left(3 + \frac{3}{5}\Delta b_1 - \frac{9}{5}\Delta b_2 + \Delta b_4\right)y_4 + \left(6 - \frac{1}{5}\Delta b_1 + \frac{3}{5}\Delta b_2\right)y_5 + \\ + \left(4 + \frac{2}{5}\Delta b_1 - \frac{1}{5}\Delta b_2\right)y_6. \qquad (6.24)$$

В случае, если $\Delta b_1 = \Delta b_2 = \Delta b_3 = \Delta b_4 = 0$, т.е. запасы ресурсов равны первоначальным значениям, получилось бы знакомое выражение (6.21) линейной функции Z через неосновные переменные оптимального решения. Для того чтобы объективно обусловленные оценки ресурсов остались неизменными и при изменении запасов ресурсов, т.е. сохранилось оптимальное решение двойственной задачи $Y^* = (4/5; 3/5;\ 0;\ 0;\ 0;\ 0)$, достаточно, чтобы коэффициенты при неосновных переменных в выражении (6.24) оставались неотрицательными, т.е.

$$\begin{cases} 1 - (2/5)\Delta b_1 + (1/5)\Delta b_2 + \Delta b_3 \geq 0, \\ 3 + (3/5)\Delta b_1 - (9/5)\Delta b_2 + \Delta b_4 \geq 0, \\ 6 - (1/5)\Delta b_1 + (3/5)\Delta b_2 \geq 0, \\ 4 + (2/5)\Delta b_1 - (1/5)\Delta b_2 \geq 0. \end{cases} \qquad (6.25)$$

Предположим, что изменяется только запас ресурса S_1, а остальные запасы ресурсов остаются неизменными: $\Delta b_2 = \Delta b_3 = \Delta b_4 = 0$. Тогда из выражения (6.25) получим, что

$$\begin{cases} 1 - 0{,}4\Delta b_1 \geq 0, \\ 3 + 0{,}6\Delta b_1 \geq 0, \\ 6 - 0{,}2\Delta b_1 \geq 0, \\ 4 + 0{,}4\Delta b_1 \geq 0 \end{cases} \text{ и } \begin{cases} \Delta b_1 \leq 2{,}5, \\ \Delta b_1 \geq -5, \\ \Delta b_1 \leq 30, \\ \Delta b_1 \geq -10, \end{cases} \text{ откуда}$$

$-5 \leq \Delta b_1 \leq 2{,}5$ и $18 - 5 \leq b_1 + \Delta b_1 \leq 18 + 2{,}5$ или $13 \leq b_1 + \Delta b_1 \leq 20$, т.е. при неизменности объективно обусловленных оценок ресурсов запас ресурса S_1 может изменяться в пределах от 13 до 20,5 единиц. Аналогично можно получить, что $11 \leq b_2 + \Delta b_2 \leq 17\frac{2}{3}$, $4 \leq b_3 + \Delta b_3 < \infty$, $18 \leq b_4 + \Delta b_4 < \infty$. Таким образом, при изменении запаса только одного из ресурсов S_1 в пределах от 13 до 20,5 единиц или S_2 — в пределах от 11 до $17\frac{2}{3}$ единицы, или S_3 — в пределах не менее 4 единиц, или S_4 — в пределах не менее 18 единиц оптимальное решение двойственной задачи остается прежним, т.е. $Y^* = (4/5;\ 3/5;\ 0;\ 0;\ 0;\ 0)$.

На основании изложенного ясно, что увеличение на 10 единиц в отдельности запаса ресурса S_1 (равного 18 единицам) или S_2 (16 единиц) приведет к изменению их объективно обусловленных оценок, а запаса ресурса S_3 (5 единиц) или S_4 (21 единица) оставит оценки этих ресурсов прежними (равными нулю). В результате с помощью полученных оптимальных оценок ресурсов невозможно найти соответствующее изменение максимальной прибыли (выручки) $\Delta F_{\max}$.

Если запасы ресурсов изменяются одновременно, то исследование устойчивости объективно обусловленных оценок усложняется, поскольку в данном случае нужно найти многогранник решений системы неравенств (6.25). Однако всегда можно проверить, удовлетворяют ли конкретные изменения запасов ресурсов системе (6.25). Так, в нашей задаче при одновременном увеличении запасов всех ресурсов на 10 единиц, т.е. при $\Delta b_1 = \Delta b_2 = \Delta b_3 = \Delta b_4 = 10$ все неравенства системы (6.25) справедливы, следовательно, оптимальное решение двойственной задачи остается прежним, т.е. $Y^* = (4/5;\ 3/5;\ 0;\ 0;\ 0;\ 0)$. Поэтому изменение максимальной прибыли (выручки) с учетом выражения (6.23) составит

$$\Delta F_{\max} = y_1^* \Delta b_1 + y_2^* \Delta b_2 + y_3^* \Delta b_3 + y_4^* \Delta b_4 = \frac{4}{5} \cdot 10 + \frac{3}{5} \cdot 10 + $$
$$+\ 0 \cdot 10 + 0 \cdot 10 = 14 \text{ руб.} \ ▶$$

По соотношениям объективно обусловленных оценок могут быть определены расчетные нормы заменяемости ресурсов, при соблюдении которых проводимые замены в пределах устойчивости двойственных оценок не влияют на эффективность оптимального плана.

▷ **6.10.** По условию задачи I из **6.2** определить нормы заменяемости ресурсов S_1 и S_2, при осуществлении которых сохранится максимальная прибыль (выручка) от продукции.

Решение. Составим отношение объективно обусловленных оценок ресурсов S_1 и S_2, т.е. $y_1^* : y_2^* = (4/5) : (3/5) = 4:3$, т.е. для максимизации общей прибыли (выручки) каждые дополнительные 3 единицы ресурса S_1 эквивалентны дополнительным 4 единицам ресурса S_2, вывод верен в пределах устойчивости двойственных оценок, когда изменения запасов ресурсов Δb_1, Δb_2, а также Δb_3, Δb_4 удовлетворяют системе (6.25) ▶

▷ **6.11.** Решить задачу I из **6.2** при изменениях запасов ресурсов $\Delta b_1 = -2$, $\Delta b_2 = 1$, $\Delta b_3 = 5$, $\Delta b_4 = 19$, используя (если это возможно) двойственные оценки ресурсов.

Решение. Так как изменения запасов ресурсов $\Delta b_1 = -2$, $\Delta b_2 = 1$, $\Delta b_3 = 5$, $\Delta b_4 = 19$ находятся в пределах устойчивости двойственных оценок (они удовлетворяют системе (6.25)), то решение двойственной задачи останется тем же: $Y^* = (4/5;\ 3/5;\ 0;\ 0;\ 0;\ 0)$. На основании теоремы (с. 108) и соответствия (6.14) положительным значениям переменных $y_1 = 4/5$, $y_2 = 3/5$ оптимального решения двойственной задачи соответствуют нулевые значения переменных исходной задачи: $x_3 = 0$, $x_4 = 0$. Поэтому остальные компоненты оптимального решения исходной задачи можно найти непосредственно из ее системы ограничений, в которой $x_3 = 0$, $x_4 = 0$, а в правых частях указаны новые запасы ресурсов: $b_1' = 18 - 2 = 16$, $b_2' = 16 + 1 = 17$, $b_3' = 5 + 5 = 10$, $b_4' = 21 + 19 = 40$:

$$\begin{cases} x_1 + 3x_2 = 16, \\ 2x_1 + x_2 = 17, \\ x_2 + x_5 = 10, \\ 3x_1 + x_6 = 40, \end{cases}$$

откуда $x_1 = 7$, $x_2 = 3$, $x_3 = 7$, $x_4 = 19$. Соответствующее значение максимума линейной функции $F_{\max} = 2 \cdot 7 + 3 \cdot 3 = 23$.

Итак, максимальный размер прибыли (выручки) составит $F_{\max} = 23$ руб. при оптимальном плане производства $X^* = (7;\ 3;\ 0;\ 0;\ 7;\ 19)$.▶

УПРАЖНЕНИЯ

Задачи **6.12—6.15** решить симплексным методом, составить задачи, двойственные данным, и найти их решения, используя теоремы двойственности.

6.12. $F = x_1 + x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 - x_2 \le -2, \\ x_1 - 2x_2 \ge -13, \\ 3x_1 - x_2 \le 6, \end{cases}$$

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

6.13. $Z = 10y_2 - 3y_3 \to \min$

при ограничениях:

$$\begin{cases} -2y_1 + y_2 - y_3 \ge 1, \\ y_1 + 2y_2 - y_3 \ge 3, \end{cases}$$

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

6.14. Текст условия приведен в задаче **1.4**.

6.15. $F = x_1 + x_2 \to \max$

при ограничениях

$$\begin{cases} x_1 - 4x_2 - 4 \le 0, \\ 3x_1 - x_2 \ge 0, \\ x_1 + x_2 - 4 \ge 0, \end{cases}$$

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

6.16. Для изготовления четырех видов продукции (А, Б, В, Г) используются три вида ресурсов (I, II, III). Другие условия задачи представлены в таблице

Ресурсы	Запас ресурсов, ед.	Нормы расхода сырья на единицу продукции, ед.			
		А	Б	В	Г
I	3400	2	1	0,5	4
II	1200	1	5	3	0
III	3000	3	0	6	1
Прибыль от единицы продукции, ден. ед.		7,5	3	6	12

Необходимо: 1. Определить план выпуска продукции, при котором прибыль от ее реализации будет максимальной.

2. Сформулировать экономически, записать и решить двойственную задачу. Пояснить экономический смысл полученных объективно обусловленных оценок ресурсов.

3. Найти интервалы устойчивости двойственных оценок по отношению к изменению запаса ресурсов каждого вида.

4. Определить изменение максимальной прибыли от реализации продукции при увеличении запаса ресурса I на 40 ед., ресурса III — на 50 ед. и уменьшении запаса ресурса II на 30 ед. Оценить раздельное влияние этих изменений и суммарное влияние.

5. Определить нормы заменяемости ресурсов.

6. Сопоставить оценку затрат и прибыли по оптимальному плану и каждому виду продукции.

7. Оценить целесообразность введения в план пятого вида продукции Д, нормы расхода сырья на единицу которого соответственно равны 2, 4, 2 ед., а прибыль — 15 ден. ед.

6.17. Используя геометрическое решение двойственной задачи и теоремы двойственности, решить задачу линейного программирования

$$F = -4x_1 - 18x_2 - 30x_3 - 5x_4 \to \max$$

при ограничениях:

$$\begin{cases} 3x_1 + x_2 - 4x_3 - x_4 \le -3, \\ 2x_1 + 4x_2 + x_3 - x_4 \ge 3, \end{cases}$$

$$x_j \ge 0,\ j = 1,\ 2,\ 3,\ 4.$$

Глава 7. ТРАНСПОРТНАЯ ЗАДАЧА

7.1. Экономико-математическая модель транспортной задачи

Важным частным случаем задачи линейного программирования является так называемая транспортная задача.

▷ **7.1.** Построить экономико-математическую модель следующей задачи. Имеются три поставщика и четыре потребителя. Мощность поставщиков и спросы потребителей, а также затраты на перевозку единицы груза для каждой пары "поставщик — потребитель" сведены в таблицу поставок (табл. 7.1).

Таблица 7.1

Поставщики	Мощность поставщиков	Потребители и их спрос			
		1	2	3	4
		20	110	40	110
1	60	1 x_{11}	2 x_{12}	5 x_{13}	3 x_{14}
2	120	1 x_{21}	6 x_{22}	5 x_{23}	2 x_{24}
3	100	6 x_{31}	3 x_{32}	7 x_{33}	4 x_{34}

В левом верхнем углу произвольной (i, j)-клетки (i — номер строки, j — номер столбца) стоит так называемый *коэффициент затрат* — затраты на перевозку единицы груза от i-го поставщика

к j-му потребителю, например, в левом верхнем углу клетки (1,4) стоит число 3, следовательно, перевозка единицы груза от 1-го поставщика к 4-му потребителю обойдется в 3 условных денежных единицы и т. д.

Задача ставится следующим образом. *Найти объемы перевозок для каждой пары "поставщик — потребитель" так, чтобы:*

1) мощности всех поставщиков были реализованы;

2) спросы всех потребителей были удовлетворены;

3) суммарные затраты на перевозку были бы минимальны.

Решение. Построим экономико-математическую модель данной задачи. Искомый объем перевозки от i-го поставщика к j-му потребителю обозначим через x_{ij} и назовем *поставкой клетки* (i, j). Например, x_{12} — искомый объем перевозки от 1-го поставщика ко 2-му потребителю или поставка клетки (1,2) и т. д. Заданные мощности поставщиков и спросы потребителей накладывают ограничения на значения неизвестных x_{ij}. Так, например, объем груза, забираемого от 1-го поставщика, должен быть равен мощности этого поставщика — 60 единицам, т.е. $x_{11} + x_{12} + x_{13} + x_{14} = 60$ (уравнение баланса по первой строке). Таким образом, чтобы мощность каждого из поставщиков была реализована, необходимо составить уравнения баланса для каждой строки таблицы поставок, т. е.

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 60, \\ x_{21} + x_{22} + x_{23} + x_{24} = 120, \\ x_{31} + x_{32} + x_{33} + x_{34} = 100. \end{cases} \tag{7.1}$$

Аналогично, чтобы спрос каждого из потребителей был удовлетворен, подобные уравнения баланса составляем для каждого столбца таблицы поставок:

$$\begin{cases} x_{11} + x_{21} + x_{31} = 20, \\ x_{12} + x_{22} + x_{32} = 110, \\ x_{13} + x_{23} + x_{33} = 40, \\ x_{14} + x_{24} + x_{34} = 110. \end{cases} \tag{7.2}$$

Очевидно, что объем перевозимого груза не может быть отрицательным, поэтому следует дополнительно предположить, что

$$x_{ij} \geq 0 \quad (i = 1, 2, 3;\ j = 1, 2, 3, 4).$$

Суммарные затраты F на перевозку выражаются через коэффициенты затрат и поставки следующим образом:

$$F = 1 \cdot x_{11} + 2x_{12} + 5x_{13} + 3x_{14} + 1 \cdot x_{21} + 6x_{22} + 5x_{23} + \\ + 2x_{24} + 6x_{31} + 3x_{32} + 7x_{33} + 4x_{34}. \tag{7.3}$$

Теперь можно дать математическую формулировку задачи (без обращения к ее содержательному экономическому смыслу). *На множестве неотрицательных решений системы ограничений* (7.1) *и* (7.2) *найти такое решение* $X = (x_{11}, x_{12}, ..., x_{33}, x_{34})$, *при котором линейная функция* (7.3) *принимает минимальное значение.*

Особенности экономико-математической модели транспортной задачи:

- *система ограничений есть система уравнений (т.е. транспортная задача задана в канонической форме);*
- *коэффициенты при переменных системы ограничений равны единице или нулю;*
- *каждая переменная входит в систему ограничений два раза: один раз — в систему* (7.1) *и один раз — в систему* (7.2).

Для математической формулировки транспортной задачи в общей постановке обозначим через c_{ij} коэффициенты затрат, через M_i — мощности поставщиков, через N_j — мощности потребителей, где $i = 1, 2, ..., m$; $j = 1, 2, ..., n$; m — число поставщиков, n — число потребителей. Тогда система ограничений примет вид

$$\sum_{j=1}^{n} x_{ij} = M_i, \ i = 1, 2, ..., m, \tag{7.4}$$

$$\sum_{i=1}^{m} x_{ij} = N_j, \ j = 1, 2, ..., n. \tag{7.5}$$

Система (7.4) включает в себя уравнение баланса по строкам, а система (7.5) — по столбцам таблицы поставок. Линейная функция в данном случае

$$F = \sum_{j=1}^{n} \sum_{i=1}^{m} c_{ij} x_{ij}. \tag{7.6}$$

Математическая формулировка транспортной задачи в общей постановке будет следующей: ***на множестве неотрицательных (допустимых) решений системы ограничений*** (7.4), (7.5) ***найти такое*** *решение* $X = (x_{11}, x_{12}, ..., x_{ij}, ..., x_{mn})$, *при котором значение линейной функции* (7.6) *минимально.*

Произвольное допустимое решение $X = (x_{11}, x_{12}, ..., x_{ij}, ..., x_{mn})$ системы ограничений (7.4), (7.5) назовем *распределением поставок*. Такое решение задает заполнение таблицы поставок, поэтому в дальнейшем значение произвольной переменной x_{ij} и содержимое соответствующей клетки таблицы поставок будут отождествляться.

Транспортная задача, приведенная в примере **7.1**, обладает важной особенностью: суммарная мощность поставщиков равна суммарной мощности потребителей, т.е.

$$\sum_{i=1}^{m} M_i = \sum_{j=1}^{n} N_j . \tag{7.7}$$

Такие транспортные задачи называются *закрытыми* (говорят также, что транспортная задача в этом случае имеет закрытую модель). В противном случае транспортная задача называется *открытой* (открытая модель транспортной задачи).

Рассмотрим закрытую транспортную задачу. Являясь задачей линейного программирования, транспортная задача может быть решена симплексным методом. Однако специфичная форма системы ограничений данной задачи позволяет существенно упростить обычный симплексный метод. Модификация симплексного метода применительно к транспортной задаче называется *распределительным методом*. По аналогии с общим случаем решение в нем осуществляется по шагам, и каждому шагу соответствует разбиение переменных на основные (базисные) и неосновные (свободные).

Число r основных переменных транспортной задачи равно рангу системы линейных уравнений (максимальному числу линейно независимых уравнений в системе ограничений).

Теорема 7.1. *Ранг r системы уравнений* (7.4), (7.5) *при условии* (7.7) *равен* $m+n-1$.[1]

□ Прежде всего заметим, что уравнения системы (7.4), (7.5) при условии (7.7) линейно зависимы и, следовательно, ранг системы

---

[1] Все доказательства при первом чтении могут быть опущены.

не больше, чем $m + n$ — 1. Действительно, сравним сумму

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = \sum_{i=1}^{m} M_i \qquad \textbf{(7.8)}$$

первых m уравнений системы (сумму уравнений системы (7.4)) с суммой

$$\sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij} = \sum_{j=1}^{n} N_j \qquad \textbf{(7.9)}$$

оставшихся n уравнений (суммой уравнений системы (7.5)).

Согласно условию (7.7) правые части уравнений (7.8) и (7.9) совпадают. Левые части (7.8) и (7.9), являющиеся суммами всевозможных переменных x_{ij} данной задачи, также совпадают. Следовательно, совпадают уравнения (7.8) и (7.9), т.е. сумма первых m уравнений системы ограничений равна сумме оставшихся n уравнений системы ограничений: уравнения системы (7.4), (7.5) линейно зависимы.

Докажем, что ранг r системы *не меньше, чем* $m+n$—1. Из линейной алгебры известно, что если некоторые k переменных произвольной системы линейных уравнений можно линейно выразить через остальные переменные системы, то ранг этой системы не меньше, чем k.

Выразим, например, переменные x_{ij}, входящие в первый столбец и первую строку табл. 7.1, через остальные переменные x_{ij}, где $i = 2, 3, ..., m$, $j = 2, 3, ..., n$. Сначала найдем такие выражения для переменных, отличных от x_{11}. Для каждой переменной x_{1j} первой строки, где $j = 2, 3, ..., n$, воспользуемся уравнением баланса по соответствующему столбцу:

$$x_{1j} = N_j - \sum_{i=2}^{m} x_{ij}. \qquad \textbf{(7.10)}$$

Аналогично для каждой переменной x_{i1} первого столбца, где $i = 2, 3, ..., m$, воспользуемся уравнением баланса по соответствующей строке:

$$x_{i1} = M_i - \sum_{j=2}^{n} x_{ij}.$$

Для нахождения выражения для x_{11} воспользуемся, например, уравнением баланса по первой строке:

$$x_{11} = M_1 - \sum_{j=2}^{n} x_{1j}. \qquad \textbf{(7.11)}$$

Подставляя в правую часть уравнения (7.11) выражения для x_{1j}, где $j = 2, 3, ..., n$ из уравнения (7.10), получаем искомое выражение для x_{11}.

Таким образом, $m+n-1$ переменных этой задачи можно выразить через остальные $mn-m-n+1$ переменные, т.е. ранг r системы $r \geq m + n - 1$.

Сравнивая два полученных ограничения на ранг r системы (7.4) и (7.5) ($r \geq m + n - 1$ и $r \leq m + n - 1$), получаем, что $r = m+n-1$.■

Основное следствие теоремы 7.1 — *число r основных (базисных) переменных закрытой транспортной задачи равно $m+n-1$, где m — число поставщиков, n — число потребителей.* Каждому разбиению переменных x_{ij} задачи на основные (базисные) и неосновные (свободные) соответствует базисное решение (*см.* гл. 2) и, как следствие, заполнение таблицы поставок, которое также назовем *базисным.* Иными словами, *распределение поставок называется базисным, если переменные, соответствующие заполненным клеткам, можно принять за основные переменные.* Клетки, отвечающие базисным переменным, в дальнейшем будем называть *базисными*, а клетки, соответствующие свободным переменным, — *свободными* или *пустыми.* Поскольку в дальнейшем мы используем исключительно базисные распределения поставок, то термины "базисная клетка" и "заполненная клетка" будут считаться равнозначными.

Подобно тому, как это было в симплексном методе, в распределительном методе решения транспортной задачи будем переходить от одного базисного распределения поставок к другому в сторону невозрастания целевой функции вплоть до оптимального решения. Для начала такого движения потребуется исходное базисное распределение поставок — так называемый *опорный план.*

7.2. Нахождение первоначального базисного распределения поставок

Одним из возможных методов нахождения первоначального базисного распределения поставок является *метод "северо-западного" угла*, показанный в следующем примере.

▷ **7.2.** Найти первоначальное базисное распределение поставок для транспортной задачи **7.1**.

Р е ш е н и е. Дадим переменной x_{11} максимально возможное значение или, иными словами, максимально возможную поставку в клетку (1,1) — "северо-западный" угол таблицы поставок: $x_{11} = \min\{60, 20\} = 20$. После этого спрос 1-го потребителя будет полностью удовлетворен, в результате чего первый столбец таблицы поставок выпадет из последующего рассмотрения (заполненные клетки будем перечеркивать сплошной линией (*см.* табл. 7.2) клетки, выпавшие из последующего рассмотрения, перечеркнуты пунктирной линией. В таблице поставок найдем новый "северо-западный" угол — клетку (1,2) и дадим в нее максимально возможное значение. Учитывая, что 1-й поставщик уже отдал 20 единиц груза и у него осталось только 40 = 60−20 единиц груза, получаем, что $x_{12} = \min\{40, 110\} = 40$. После этого мощность 1-го поставщика полностью реализована и из рассмотрения выпадет первая строка таблицы поставок (перечеркиваем сплошной линией клетку (1,2) и пунктирной линией оставшиеся свободные клетки первой строки). В оставшейся таблице снова находим "северо-западный угол" и т. д. В результате получаем следующее исходное распределение поставок (*см.* табл.7.2).▶

Таблица 7.2

	20	110	40	110
60	1 / 20	2 / 40	5	3
120	1	6 / 70	5 / 40	2 / 10
100	6	3	7	4 / 100

Число заполненных клеток в полученном распределении оказалось равным $m+n-1 = 3+4-1 = 6$, т.е. числу основных (базисных) переменных. Это, конечно, не случайно. Действительно, на каждом шаге (кроме последнего) данного метода из рассмотрения выпадали либо строка, либо столбец, а на последнем шаге и столбец, и строка. Поэтому число заполненных клеток (число шагов) на единицу меньше, чем сумма числа строк и столбцов таблицы поставок, т.е. равно $m+n-1$. Оказывается (*см.* теорему 7.2), что эта особенность шагов метода "северо-западного" угла служит причиной того, что полученное распределение является базисным.

Существенный недостаток метода "северо-западного" угла состоит в том, что он построен без учета значений коэффициентов затрат задачи. С другой стороны, данный метод допускает модификацию, лишенную этого недостатка: на каждом шаге максимально возможную поставку следует давать не в "северо-западную" клетку оставшейся таблицы, а в клетку с наименьшим коэффициентом затрат. При этом распределение поставок оказывается, вообще говоря, ближе к оптимуму, чем распределение, полученное методом "северо-западного угла". Такой метод получения опорного плана называется *методом наименьших затрат*. Рассмотрим его на следующем примере.

▷ **7.3.** Найти методом наименьших затрат первоначальное распределение поставок в задаче **7.1.**

Решение. Находим в таблице поставок (*см.* табл. 7.1) клетки с наименьшим коэффициентом затрат. Таких клеток две — (1,1) и (2,1) с коэффициентами затрат, равными 1. Сравним максимально возможные поставки для этих клеток: для клетки (1,1) $x_{11} = \min\{60, 20\} = 20$, для клетки (2,1) $x_{21} = \min\{120, 20\} = 20$. Так как они совпадают, то максимально возможную поставку даем

Таблица 7.3

	20	110	40	110
60	1	2	5	3
120	1 20	6	5	2
100	6	3	7	4

Таблица 7.4

	20	110	40	110
60	1	2	5	3
120	1 20	6	5	2 100
100	6	3	7	4

в любую из них. Например, даем поставку, равную 20 единицам, в клетку (2,1). В результате спрос первого потребителя удовлетворен и первый столбец таблицы поставок выпадает из последующего рассмотрения (табл. 7.3).

Таблица 7.5

	20	110	40	110
60	1	2 / 60	5	3
120	1 / 20	6	5	2 / 100
100	6	3 / 50	7 / 40	4 / 10

В оставшейся таблице наименьшим коэффициентом затрат обладают две клетки: $c_{12} = c_{24} = 2$. Сравним максимально возможные поставки для этих клеток: для клетки (1,2) $x_{12} = \min\{60, 110\} = 60$; для клетки (2,4) $x_{24} = \min\{120-20, 110\} = 100$. Даем поставку в клетку (2,4), для которой максимально возможная поставка оказалась больше: $x_{24} = 100$. При этом из рассмотрения выпадает вторая строка таблицы поставок (табл. 7.4).

Аналогично, продолжая заполнение таблицы поставок шаг за шагом, получаем $\overline{x}_{12} = \min\{60, 110\} = 60$, $\overline{x}_{32} = \min\{100, 110-60\} = 50$, $x_{34} = \min\{100-50, 110-100\} = 10$, $\overline{x}_{33} = \min\{100-60, 40\} = 40$ (табл. 7.5).▶

Сравним найденное распределение поставок с распределением, полученным для той же задачи по методу “северо-западного” угла (*см.* задачу **7.2**, табл. 7.2). Вычислим для каждого из этих распределений суммарные затраты в денежных единицах:

в задаче **7.2**:

$$F_0 = 1 \cdot 20 + 2 \cdot 40 + 6 \cdot 70 + 5 \cdot 40 + 2 \cdot 10 + 4 \cdot 100 = 1140;$$

в задаче **7.3**:

$$F_0 = 1 \cdot 20 + 2 \cdot 60 + 3 \cdot 50 + 2 \cdot 100 + 7 \cdot 40 + 4 \cdot 10 = 810.$$

Как и ожидалось, при использовании метода “северо-западного” угла суммарные затраты больше, чем при применении метода наименьших затрат. Таким образом, во втором случае мы находимся ближе (по числу необходимых шагов) к оптимуму, чем в первом. Докажем, что распределения, получаемые с помощью указанных методов, являются базисными, и рассмотрим те особые случаи, которые могут встретиться при использовании этих методов.[1]

[1] **При первом чтении этот материал (с. 132) рекомендуется опустить.**

Теорема 7.2. *Пусть на каждом шаге заполнения таблицы поставок возникает одна заполненная клетка, причем из рассмотрения на каждом (кроме последнего) шаге выпадает либо одна строка, либо один столбец. Тогда переменные, соответствующие заполненным клеткам, можно принять за базисные.*

□ Из линейной алгебры известно, что если ранг системы линейных уравнений равен r и некоторые r переменных системы выражены через остальные переменные, то эти r переменных можно взять за основные (базисные). Из условия данной теоремы следует, что число заполненных клеток равно $m+n-1$, т. е. равно рангу системы (7.4), (7.5) (*см.* пояснения к методу "северо-западного" угла). Поэтому теорема будет доказана, если показать, что переменные, соответствующие заполненным клеткам, могут быть выражены через переменные, соответствующие свободным клеткам.

Предположим, что переменные заполненных клеток, возникшие на первых t шагах метода, где $t = 1, 2, ..., m+n-2$, можно выразить через переменные, соответствующие свободным клеткам тех строк и столбцов, которые были вычеркнуты (выпали из рассмотрения) на первых t шагах. Пусть на $(t+1)$-м шаге метода заполнена (p, q)-я клетка и из рассмотрения выпала, например, p-я строка. Выразим переменную x_{pq} из уравнения баланса по p-й строке:

$$x_{pq} = M_p - \sum_{\substack{j=1 \\ j \neq q}}^{n} x_{pj}.$$

Пусть среди переменных правой части последнего равенства есть переменные клеток, заполненных на одном из первых t шагов. Тогда по предположению их можно выразить через переменные свободных клеток тех строк и столбцов, которые были вычеркнуты на первых t шагах. В случае, если на $(t+1)$-м шаге из рассмотрения выпал q-й столбец, x_{pq} следует выразить из уравнения баланса по q-му столбцу. Подобные рассуждения следует последовательно провести для каждого из шагов заполнения таблицы поставок. ■

Из теоремы 7.2 следует, что методы "северо-западного" угла и наименьших затрат приводят к базисным распределениям поставок, если на каждом (кроме последнего) шаге из рассмотрения выпадают либо одна строка, либо один столбец.

Рассмотрим теперь те *особые случаи*, когда на некотором шаге заполнения из рассмотрения выпадают одновременно и строка и столбец. Укажем, как следует поступать, чтобы метод заполнения по-прежнему удовлетворял условиям теоремы 7.2 и получаемое распределение поставок было базисным.

▷ **7.4.** Найти первоначальное базисное распределение поставок для следующей транспортной задачи (табл. 7.6).

Таблица 7.6

	20	10	40
30	1	3	5
30	3	3	2
10	4	1	2

Решение. Воспользуемся методом "северо-западного" угла. На первом шаге следует дать поставку, равную 20 единицам, в клетку (1,1). В результате будет удовлетворен спрос 1-го потребителя и из рассмотрения выпадет первый столбец. На втором шаге поставку в 10 единиц следует дать в клетку (1,2). При этом из последующего рассмотрения выпадет и 1-й поставщик (который реализовал остатки своего груза), и 2-й потребитель, полностью удовлетворивший свой спрос. Продолжая использовать метод "северо-западного" угла, мы получим, конечно, заполнение таблицы поставок, но число заполненных клеток окажется меньше, чем число основных (базисных) переменных, равное $m+n-1=3+3-1=5$. Такое распределение не будет базисным, и для продолжения решения распределительный метод будет неприемлем. Избежать этого можно, используя следующий искусственный прием.

Разобьем второй шаг на два шага. Допустим, что после поставки в клетку (1,2) из рассмотрения выпадает, например, только первая строка. Для того чтобы вывести из рассмотрения второй столбец, делаем еще один шаг: даем нулевую (фиктивную) поставку в произвольную, но не вычеркнутую клетку второго столбца, например, в клетку (2,2). После таких трех шагов имеем табл. 7.7.

Аналогично можно было допустить, что после второго шага из рассмотрения выпал только второй столбец. Тогда на третьем шаге нулевую поставку следует дать в произвольную, но не вычеркнутую клетку первой строки.

При последующем заполнении таблицы поставок используем метод "северо-западного" угла обычным способом. В результате получаем распределение поставок (табл. 7.8).▶

Таблица 7.7

	20	10	40
30	1 20	3 10	3
30	3	3 0	2
10	4	1	2

Таблица 7.8

	20	10	40
30	1 20	3 10	3
30	3	3 0	2 30
10	4	1	2 10

Перечеркнутые сплошной чертой клетки, отвечающие базисным переменным, в дальнейшем будем называть заполненными, несмотря на то, что среди них возможны клетки с нулевыми поставками.

Рассмотренный искусственный прием применяется также при методе наименьших затрат, если при использовании этого метода на некотором шаге из рассмотрения выпадают одновременно и строка, и столбец.

7.3. Критерий оптимальности базисного распределения поставок

Подход к решению вопроса об оптимальности базисного решения был детально разобран в гл. 5, посвященной симплексному методу. Согласно ему вначале следует выразить линейную функцию задачи через неосновные (свободные) переменные. Транспортная задача — задача на минимум, поэтому оптимум достигнут тогда и только тогда, когда все коэффициенты при неосновных (свободных) переменных в выражении линейной функции неотрицательны. В транспортной задаче произвольная переменная x_{ij} отождествляется с содержимым соответствующей клетки (i, j) таблицы поставок. Коэффициент β_{ij} при свободной переменной x_{ij} в выражении линейной функции F через свободные переменные называется ***оценкой свободной клетки*** (i, j). Тогда ***критерий опти-***

мальности формулируется следующим образом: ***базисное распределение поставок оптимально тогда и только тогда, когда оценки всех свободных клеток неотрицательны.***

Таким образом, на первый план выходит задача о нахождении оценок свободных клеток для фиксированного базисного распределения поставок.

Пусть фиксировано некоторое базисное распределение поставок, при этом клетка (i, j) — свободная (переменная x_{ij} — свободная), β_{ij} — оценка клетки (i, j) или коэффициент при x_{ij} в выражении линейной функции F через свободные переменные, т.е.

$$F = F_0 + \beta_{ij} x_{ij} + \ldots, \qquad \textbf{(7.12)}$$

где многоточием обозначены слагаемые, отвечающие свободным переменным, отличным от x_{ij}, F_0 — суммарные затраты на перевозку данного распределения поставок. Тогда из выражения (7.12) следует, что *оценка β_{ij} свободной клетки (i, j) равна приращению ΔF суммарных затрат на перевозку при переводе в клетку (i, j) единичной поставки* (увеличение переменной x_{ij} от 0 до 1). Очевидно, что $\Delta F > 0$, если $\beta_{ij} > 0$; $\Delta F < 0$, если $\beta_{ij} < 0$. Последнее косвенное определение оценки свободной клетки обычно называют *экономическим смыслом оценки свободной клетки.*

Для нахождения оценок свободных клеток воспользуемся экономическим смыслом указанных оценок.

▷ **7.5.** Установить, является ли оптимальным базисное распределение поставок, найденное в задаче **7.3** (табл. 7.9).

Решение. Найдем, например, оценку свободной клетки (1,3). Для этого дадим в клетку (1,3) единичную поставку. При этом потребуется изменить поставки в заполненных клетках так, чтобы сохранился баланс по строкам и столбцам. (Будем полагать, что во всех свободных клетках, отличных от клетки (1,3), поставка останется нулевой.) Так, чтобы 3-й потребитель получил по-прежнему 40 единиц груза, поставку в клетке (3,3) следует уменьшить на 1. Для

Таблица 7.9

	20	110	40	110
60	1	2 / 60	5	3
120	1 / 20	6	5	2 / 100
100	6	3 / 50	7 / 40	4 / 10

Таблица 7.10

	20	110	40	110
60	1	2 / 59	5 / 1	3
120	1 / 20	6	5	2 / 100
100	6	3 / 51	7 / 39	4 / 10

того чтобы 3-й поставщик отправил по-прежнему 100 единиц груза, поставку в клетке (3,2) увеличиваем на 1. 2-му потребителю нужно только 110 единиц груза, поэтому поставку в клетке (1,2) придется уменьшить на 1. Существенно, что найденный вариант перераспределения поставок, затрагивающий заполненные клетки и увеличивающий на 1 поставку клетки (1,3), единственный. Полученное распределение поставок представлено в табл. 7.10.

Найдем изменение ΔF суммарных затрат при указанном перераспределении поставок.

Первоначально затраты на перевозку (*см.* табл. 7.9) составили

$$F_{\text{н}} = 2\cdot 60 + 1\cdot 20 + 5\cdot 0 + 3\cdot 50 + 7\cdot 40 + 2\cdot 100 + 4\cdot 10,$$

после перераспределения (*см.* табл. 7.10):

$$F_{\text{п}} = 2\cdot 59 + 1\cdot 20 + 5\cdot 1 + 3\cdot 51 + 7\cdot 39 + 2\cdot 100 + 4\cdot 10.$$

Тогда, учитывая экономический смысл оценки свободной клетки, получаем, что

$$\beta_{13} = \Delta F = F_{\text{п}} - F_{\text{н}} = 2\cdot(-1) + 5\cdot 1 + 3\cdot 1 + 7\cdot(-1) = -1.$$

Так как среди клеток табл. 7.9 есть клетка с отрицательной оценкой, то распределение поставок не оптимально.▶

Способ решения задачи **7.5** довольно громоздок (особенно учитывая, что часто в задачах приходится искать оценки всех свободных клеток заданного базисного распределения поставок). Проанализируем решение задачи **7.5** для упрощения вычислений.

При вычислении ΔF многие слагаемые из $F_{\text{п}}$ и $F_{\text{н}}$ взаимно уничтожаются, не влияя на значение ΔF: существенны лишь коэффициенты затрат тех клеток, в которых поставка при рассматриваемом перераспределении изменится. При этом в выражение для ΔF некоторые из них входят со знаком "+", а некоторые — со знаком "–". Для нахождения "правила знаков" удобен чертеж, представленный на рис. 7.1. На нем изображены клетки, в которых будет изменена поставка (слева от каждой клетки написан в скобках ее номер; клетки, соответствующие базисным переменным, перечеркнуты).

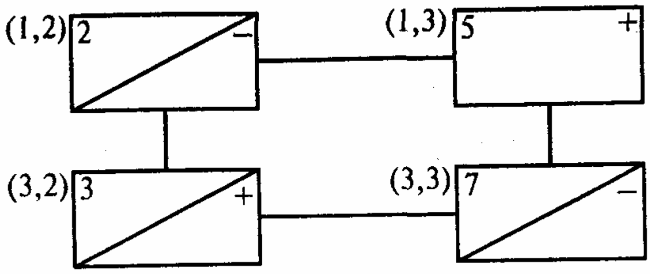


Рис. 7.1

При этом знаком "+" помечены те клетки, поставка в которых увеличится. Видно, что именно их коэффициенты затрат войдут в выражение для ΔF со знаком "+". В остальных клетках рис. 7.1 поставка уменьшится (в них вписан знак "–"), их коэффициенты затрат войдут в выражение для ΔF со знаком "–". Ломаную, соединяющую клетки с изменяемой поставкой, будем называть *означенным циклом пересчета* (*см.* рис. 7.1).

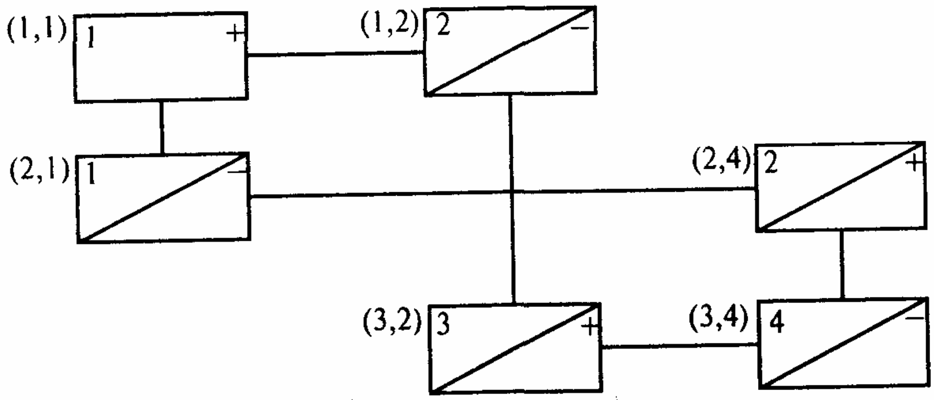


Рис. 7.2

Таким образом, можно сформулировать *правило 1 нахождения оценки свободной клетки:* для свободной клетки следует построить

цикл пересчета, в вершинах этого цикла расставить последовательно чередующие знаки, начиная со знака "+" в свободной клетке, тогда значение оценки свободной клетки равно алгебраической сумме коэффициентов затрат клеток цикла, взятых с соответствующими знаками.

Аналогично, составляя означенный цикл пересчета для каждой свободной клетки, можно найти ее оценку. При этом, конечно, цикл не всегда будет получаться таким простым, как в разобранном примере для клетки (1,3). Например, означенный цикл пересчета для клетки (1,1), показанный на рис. 7.2, более сложный.

Оценка клетки (1,1) в этом случае равна

$$\beta_{11} = (1 + 3 + 2) - (2 + 1 + 4) = -1.$$

Замечание. Иногда для произвольного означенного цикла вводится понятие ***оценка цикла*** *— алгебраическая сумма коэффициентов, стоящих в вершинах цикла, взятых с соответствующими знаками.* Приведенные выше рассуждения показывают, что оценка цикла равна оценке той единственной свободной клетке, которая входит в данный цикл.

Для облегчения нахождения цикла пересчета в конкретных задачах дадим его точное определение. ***Циклом*** *в матрице будем называть ломаную с вершинами в клетках и звеньями, лежащими вдоль строк и столбцов матрицы, удовлетворяющую условиям:*

- *ломаная должна быть* ***связной****, т.е. из любой ее вершины можно попасть в любую другую вершину по звеньям ломаной;*
- *в каждой вершине ломаной встречаются два звена, одно из которых располагается по строке, другое — по столбцу.*

Циклом пересчета называется такой цикл в таблице с базисным распределением поставок, при котором одна из его вершин лежит в свободной клетке, остальные — в заполненных. Цикл пересчета называется *означенным*, если в его вершинах расставлены знаки "+" и "−" так, что в свободной клетке стоит знак "+", а соседние вершины имеют противоположные знаки.

Для каждой свободной клетки базисного распределения поставок существует и притом единственный цикл пересчета, причем операция означивания цикла является корректной. Подробные доказательства можно найти, например, в [16].

Таким образом, получено правило, позволяющее найти оценку произвольной свободной клетки. Однако нахождение оценок свободных клеток можно существенно упростить. Рассмотрим следующую воображаемую ситуацию. Пусть коэффициенты затрат всех заполненных клеток равны нулю. Если теперь по рассмотренному правилу найти оценку свободных клеток, то окажется, что оценки свободных клеток равны их коэффициентам затрат, т.е. в этом случае значения оценок считываются с таблицы поставок и никаких циклов строить не надо.

С другой стороны, справедлива следующая теорема.

Теорема 7.3. *(о потенциалах). Оценка свободной клетки не изменится, если к коэффициентам затрат некоторой строки (столбца) таблицы поставок прибавить некоторое число. Это число, прибавляемое к коэффициентам затрат выделенной строки (столбца), будем называть* ***потенциалом*** *данной строки (столбца).*

□ Пусть для фиксированной свободной клетки построен цикл. Для каждого звена этого цикла, входящего в произвольную выделенную строку, имеем пару соседних вершин цикла. По определению означенного цикла пересчета одна из этих двух вершин имеет знак "+", другая — знак "−". Тогда вклад от этих двух вершин в значение искомой оценки определяется разностью коэффициентов затрат этих вершин. Очевидно, что если к каждому из этих коэффициентов затрат прибавить одно и то же число, то само значение разности не изменится.■

Рассмотрение воображаемого случая и теорема 7.3 приводят к *правилу 2 нахождения оценок свободных клеток:* к коэффициентам затрат таблицы поставок в каждой строке и столбце надо прибавить такие числа (потенциалы), чтобы коэффициенты затрат в *заполненных* клетках стали равными *нулю*. Полученные при этом коэффициенты затрат свободных клеток равны оценкам этих клеток.

▷ **7.6.** Найти оценки свободных клеток базисного распределения поставок, найденного в задаче **7.3**.

Р е ш е н и е. Найдем оценки свободных клеток, следуя изложенной выше последовательности действий. Изменение коэффициентов затрат можно начинать с любого столбца (строки). Потенциал столбца (строки), избранного для начала, может быть произвольным, но можно доказать, что после его фиксации потенциалы остальных столбцов и строк будут определены однозначно.

Начнем с первого столбца. Пусть потенциал этого столбца равен нулю (табл. 7.11). Рядом с потенциалом в скобках записываем номер шага (поставки опускаем).

Таблица 7.11

1	2	5	3	−2(7)
1	6	5	2	−1(2)
6	3	7	4	−3(4)
0(1)	0(6)	−4(5)	−1(3)	

$$\begin{pmatrix} -1 & 0 & -1 & 0 \\ 0 & 5 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{pmatrix}. \qquad \textbf{(7.13)}$$

После прибавления этого потенциала к коэффициентам затрат первого столбца коэффициент затрат заполненной клетки (2,1) не изменится; чтобы полученный после сложения коэффициент стал равен нулю, потенциал второй строки табл. 7.11 должен быть равен —1; для обнуления коэффициента затрат клетки (2,4) потенциал четвертого столбца должен быть равен −1 и т. д. Измененные коэффициенты затрат удобно выписать в виде отдельной *матрицы оценок* (7.13). *Элементы матрицы оценок, соответствующие свободным клеткам таблицы поставок, равны оценкам этих свободных клеток.*

Из предыдущих рассуждений вытекает, что для фиксированного базисного распределения поставок можно подобрать различные наборы потенциалов, удовлетворяющих правилу 2, однако матрица оценок во всех таких случаях будет одинаковой. Существование, по крайней мере, одного набора потенциалов, удовлетворяющих правилу 2, оставляем без доказательства.▶

Решение транспортной задачи в случае получения неоптимального распределения поставок рассмотрим в следующем разделе.

7.4. Распределительный метод решения транспортной задачи

▷ **7.7.** Найти оптимальное распределение поставок задачи из задачи 7.1

Решение. Начнем с базисного распределения поставок, полученного в задаче 7.3. Как было установлено ранее (*см.* задачу 7.5), данное распределение не оптимально. Оценка свободной клетки — это коэффициент при соответствующей свободной переменной в выражении линейной функции. Учитывая результат задачи **7.6**, имеем

$$F = 810 - x_{13} + 3x_{31} + 5x_{22} - x_{11}.$$

Значение $F_0 = 810$ для данного распределения поставок найдено в задаче **7.3**. Далее поступаем так, как поступили бы, решая задачу симплексным методом: переменную x_{13}, коэффициент при которой отрицателен, будем переводить в основные (базисные) переменные. Переменная x_{13} начинает возрастать от нуля. Как было показано в разд. 7.3, перевод поставки в свободную клетку вызывает перераспределение поставок (передвижение поставки по циклу). Означенный цикл пересчета для клетки (1,3) показан на рис. 7.3.

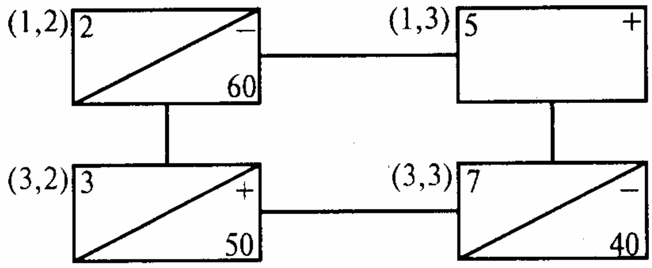


Рис. 7.3

Подобно тому, как это было в симплексном методе, увеличиваем поставку x_{13} в клетке (1,3) до тех пор, пока поставка в одной из заполненных клеток не станет равной нулю (дальнейшее увеличение x_{13} уводит в область недопустимых решений). Эта клетка принадлежит, конечно, циклу, построенному на рис. 7.3 для клетки (1,3). Найдем ее. Если в клетку (1,3) передать поставку, равную z, то поставка в клетках цикла со знаком "+" увеличится на z, а в клетках со знаком "–" — уменьшится на z. Поэтому искомая клетка находится среди клеток цикла, имеющих знак "–". Более того, она имеет минимальную поставку среди таких клеток. Так как (*см.* рис. 7.3) $\min\{60,40\} = 40$, то в нашем случае — это клетка (3,3), и для обнуления поставки в этой клетке по циклу следует передать 40 единиц груза, ***т.е. поставка, передаваемая по циклу, определяется как минимум среди поставок в клетках цикла со зна-***

ком "–". После этого клетка (1,3) считается заполненной, а клетка (3,3) — свободной.

В клетках со знаком "+" цикла поставка увеличивается на передаваемую поставку: поставка клетки (3,2) станет равной 90 единицам, поставка клетки (1,3) — 40 единицам. Аналогично в клетках со знаком "—" поставка уменьшится на передаваемую поставку, например, поставка клетки (1,2) станет равной 20 единицам, что видно из табл. 7.12. Нетрудно доказать, что вновь полученное распределение поставок — базисное.

И вновь возникает вопрос об оптимальности базисного распределения поставок — круг решения замкнулся.

Найдем оценки свободных клеток (матрицу оценок) распределения поставок. Для этого, как и прежде, подберем потенциалы так, чтобы коэффициенты затрат заполненных клеток стали равными нулю (*см.* табл. 7.12). Тогда матрица оценок примет вид (7.14).

Таблица 7.12

1	2 / 20	5 / 40	3	−2
1 / 20	6	5	2 / 100	−1
6	3 / 90	7	4 / 10	−3
0	0	−3	−1	

$$\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 5 & 1 & 0 \\ 3 & 0 & 1 & 0 \end{pmatrix}. \qquad \textbf{(7.14)}$$

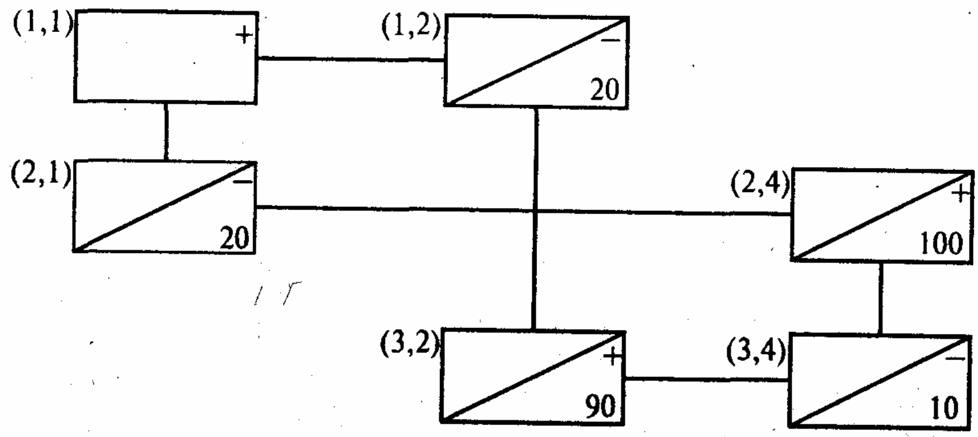


Рис. 7.4

Так как среди свободных клеток есть клетка (1,1) с отрицательной оценкой, то найденное распределение не оптимально и передача поставки в клетку (1,1) ведет к уменьшению суммарных затрат на перевозку. Означенный цикл пересчета для клетки (1,1) приведен на рис. 7.4. По правилу, сформулированному выше, поставка, передаваемая по циклу x_{11} = {20, 20, 10} = 10. Передвигая эту поставку по циклу (*см.* рис. 7.4), приходим к новому распределению поставок (табл. 7.13).

Найдя матрицу (7.15) оценок этого распределения, заключаем, что оно оптимально, так как среди оценок свободных клеток нет отрицательных.

Суммарные затраты на перевозку этого распределения поставок в денежных единицах составляют:

$$F_{\min} = 1 \cdot 10 + 2 \cdot 10 + 5 \cdot 40 + 1 \cdot 10 + 2 \cdot 110 + 3 \cdot 100 = 760.$$

Таблица 7.13

1 / 10	2 / 10	5 / 40	3
1 / 10	6	5	2 / 110
6	3 / 100	7	4

$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 4 & 0 & 0 \\ 4 & 0 & 1 & 1 \end{pmatrix} \qquad \textbf{(7.15)}$$

Экономия ΔF, достигнутая в результате применения метода перераспределения поставок, составляет в денежных единицах $\Delta F = F_{\min} - F_0 = 760 - 810 = -50$. Знак “−” в данном случае показывает, что при переходе к оптимальному распределению суммарные затраты на перевозку уменьшились.▶

Замечание 1. *Поставка, передаваемая по циклу, не может быть ни меньше, ни больше минимума поставок клеток цикла со знаком “—”.* Действительно, в первом случае ни одна из клеток цикла не будет иметь нулевой поставки, а потому общее число заполненных клеток таблицы будет равно $m+n$ и, следовательно, распределение не будет базисным. Во втором случае уходим в область недопустимых решений.

Замечание 2. ***Оптимальное распределение поставок,*** **найденное в задаче 7.7 (*см.* табл.7.13),** ***не единственное, так как среди оценок свободных клеток есть нулевые,*** **например клетка (2,3) в**

матрице (7.15). Аналогично при симплексном методе решение не единственное, если в выражении линейной функции оптимального решения через неосновные (свободные) переменные коэффициенты при некоторых свободных переменных равны нулю. В связи с данным замечанием можно предложить читателю следующее упражнение: проверить, не изменит ли перераспределение поставки в свободную клетку с нулевой оценкой оптимальное значение затрат $F_{\min}$.

Замечание 3. В некоторых случаях требуется определить изменение затрат ΔF_i на перевозку (экономию затрат) для некоторого i-го шага решения (или для каждого из шагов) транспортной задачи. Из экономического смысла оценки свободной клетки следует, что *экономия затрат ΔF_i, достигнутая на некотором i-м шаге, равна произведению оценки клетки, в которую передается поставка, на передаваемую поставку.* Например, при переходе от исходного распределения поставок (*см.* табл. 7.11) к распределению поставок по табл. 7.12 поставка 40 единиц передается в клетку, оценка которой равна —1. Тогда экономия затрат ΔF_1 на первом шаге задачи **7.7** составит

$$\Delta F_1 = (-1)40 = -40.$$

Последовательность действий по решению произвольной закрытой транспортной задачи теперь может быть изложена в виде следующего *алгоритма.*

1. Для данного базисного распределения поставок подбираем потенциалы строк и столбцов таблицы поставок так, чтобы коэффициенты затрат заполненных клеток стали равны нулю. Составляем матрицу оценок.

2. Если оценки всех свободных клеток неотрицательны, то найденное распределение оптимально — решение закончено. Если среди оценок свободных клеток есть отрицательные, то выбираем одну из них для передачи в нее поставки (для определенности можно брать, например, одну из клеток с наименьшей оценкой).

3. Для избранной свободной клетки строится означенный цикл пересчета. Поставка z, передаваемая по циклу, определяется как минимум среди поставок в клетках со знаком "—". Найденная поставка передвигается по циклу. При этом поставка в клетках цикла со знаком "+" увеличивается на z, а в клетках со

знаком “–” уменьшается на z. Клетка, поставка в которой при этом станет равной нулю, будет считаться свободной (далее рассмотрен случай, когда таких клеток несколько), остальные клетки цикла — заполненными. Таким образом, получено новое базисное распределение поставок.

4. Переходим к п. 1 алгоритма.

Рассмотрим *особые случаи*, которые могут возникнуть при решении транспортной задачи.

1. В некоторых случаях поставка, переводимая по циклу, может оказаться равной нулю. Это возможно тогда, когда клетка цикла со знаком “–” содержала нулевую поставку. В этом случае по циклу передается нулевая поставка. В результате та свободная клетка, для которой был построен цикл, становится заполненной (нулевой поставкой), а клетка с нулевой поставкой — свободной.

2. Если при переводе поставки по циклу поставка обращается в нуль сразу в нескольких заполненных клетках, то свободной из них следует считать только одну (любую), остальные клетки, поставка в которых стала равной нулю, следует считать заполненными нулевой поставкой. Разберем перечисленные особые случаи на примере.

▷ **7.8.** Завершить решение транспортной задачи из задачи **7.4**.

Р е ш е н и е . Установим сначала, оптимально ли распределение, полученное в указанном примере методом “северо-западного” угла (*см.* табл. 7.8). Подберем потенциалы строк и столбцов этой таблицы поставок так, чтобы коэффициенты затрат заполненных клеток стали равны нулю (табл. 7.14).

Т а б л и ц а 7.14

1 / 20	3 / 10	3	0
3	3 / 0	2 / 30	0
4	1	2 / 10	0
−1	−3	−2	

$$\begin{pmatrix} 0 & 0 & 1 \\ 2 & 0 & 0 \\ 3 & -2 & 0 \end{pmatrix}. \qquad \textbf{(7.16)}$$

Это приводит к матрице оценок (7.16). Так как среди свободных клеток таблицы есть клетка (3,2) с отрицательной оценкой, то данное базисное распределение поставок не оптимально. Переведем поставку в клетку (3,2) с отрицательной оценкой. Строя для клетки (3,2) означенный цикл пересчета (рис. 7.5), находим, что объем передаваемой поставки в данном случае равен $x_{32} = \min\{0,10\} = 0$. Передавая по построенному циклу нулевую поставку, приходим к новому базисному распределению (табл. 7.15).

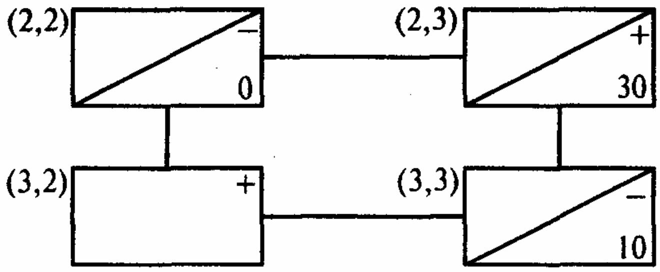


Рис. 7.5

Таблица 7.15

1 / 20	3 / 10	3	−2
3	3	2 / 30	0
4	1 / 0	2 / 10	0
1	−1	−2	

$$\begin{pmatrix} 0 & 0 & -1 \\ 4 & 2 & 0 \\ 5 & 0 & 0 \end{pmatrix} \qquad \textbf{(7.17)}$$

Подбирая потенциалы к строкам и столбцам табл. 7.15, находим матрицу (7.17) оценок данного распределения. Так как среди свободных клеток таблицы есть клетка (1,3) с отрицательной оценкой, то данное базисное распределение не оптимально. Найдем новое базисное распределение, передавая поставку в клетку (1,3) с отрицательной оценкой. Построим цикл для клетки (1,3), показанный на рис. 7.6. Поставка, передаваемая в клетку (1,3): $x_{13} = \min\{10,10\} = 10$. При передаче по циклу (*см.* рис. 7.6) 10 единиц груза станут равными нулю поставки в клетках (1,2) и (3,3). Полагаем, что только одна из них стала свободной, например клетка (3,3), а клетка (1,2) заполнена нулевой поставкой. Таким образом, будет получено базисное распределение поставок, представленное в табл. 7.16.

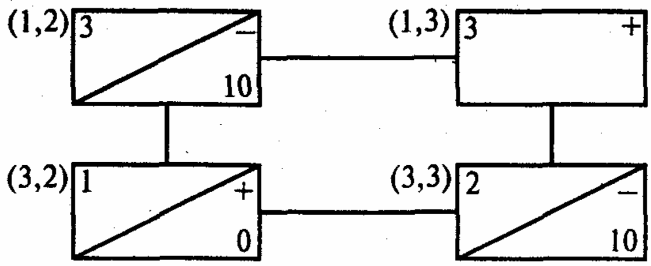


Рис. 7.6

Таблица 7.16

1 / 20	3 / 0	3 / 10	−1
3	3	2 / 30	0
4	1 / 10	2 / 10	1
0	−2	−2	

$$\begin{pmatrix} 0 & 0 & 0 \\ 3 & 1 & 0 \\ 5 & 0 & 1 \end{pmatrix}. \qquad \mathbf{(7.18)}$$

Определяя матрицу оценок (7.18), видим, что среди оценок свободных клеток найденного распределения нет отрицательных, т.е. найденное распределение (*см.* табл. 7.16) оптимально.▶

7.5. Открытая модель транспортной задачи

Открытая транспортная задача решается сведением ее к закрытой транспортной задаче.

▷ **7.9.** Найти оптимальное распределение поставок для транспортной задачи (табл. 7.17).

Решение. В данном случае суммарный спрос потребителей больше, чем суммарная мощность поставщиков (45+35+55+ +65 = 200 > 40+60+90 = 190). Введем "фиктивного поставщика" и в таблицу поставок добавим дополнительную строку (табл. 7.18) так, чтобы задача стала закрытой. Для этого мощность фиктивного поставщика следует принять равной 10 = 200—190. Коэффициенты затрат этой добавленной строки определяются издержками ввиду недогрузки мощностей потребителей. Если информация об этих издержках отсутствует, то их принимают равными одному и тому же числу (например, нулю, как в табл. 7.18). Согласно теореме 7.3, конкретное значение этого числа не влияет на оптимальное распределение поставок.

Таблица 7.17

	45	35	55	65
40	4	1	2	5
60	3	2	3	7
90	4	4	5	2

Таблица 7.18

	45	35	55	65
40	4	1	2	5
60	3	2	3	7
90	4	4	5	2
10	0	0	0	0

Первоначальное распределение поставок для сформулированной закрытой транспортной задачи найдем, например, по методу наименьших затрат. Для удобства укажем последовательность заполнения таблицы поставок: $x_{44} = 10$, $x_{12} = 35$, $x_{34} = 55$, $x_{13} = 5$, $x_{23} = 50$, $x_{21} = 10$, $x_{31} = 35$. В результате приходим к следующему базисному распределению поставок (табл. 7.19).

Таблица 7.19

4	1 / 35	2 / 5	5	−2
3 / 10	2	3 / 50	7	−3
4 / 35	4	5	2 / 55	−4
0	0	0	0 / 10	−2
0	1	0	2	

$$\begin{pmatrix} 2 & 0 & 0 & 5 \\ 0 & 0 & 0 & 6 \\ 0 & 1 & 1 & 0 \\ -2 & -1 & -2 & 0 \end{pmatrix} \qquad \textbf{(7.19)}$$

Установим, оптимально ли это распределение — найдем для него матрицу оценок (7.19). Так как среди оценок свободных клеток есть отрицательные, то найденное распределение не оптимально. Переведем поставку в одну из клеток с наименьшей отрицательной оценкой, например в клетку (4,3). Цикл для этой клетки изображен на рис 7.7. Поставка, передаваемая по циклу, равна $x_{43} = \min\{50, 35, 10\} = 10$. Передвигая по циклу поставку, равную 10 единицам, приходим к следующему распределению поставок (табл. 7.20). Найдем оценки свободных клеток данного распределения (*см.* матрицу оценок (7.20)). Так как оценки всех свободных клеток неотрицательны, то распределение поставок табл. 7.20 оптимально.▶

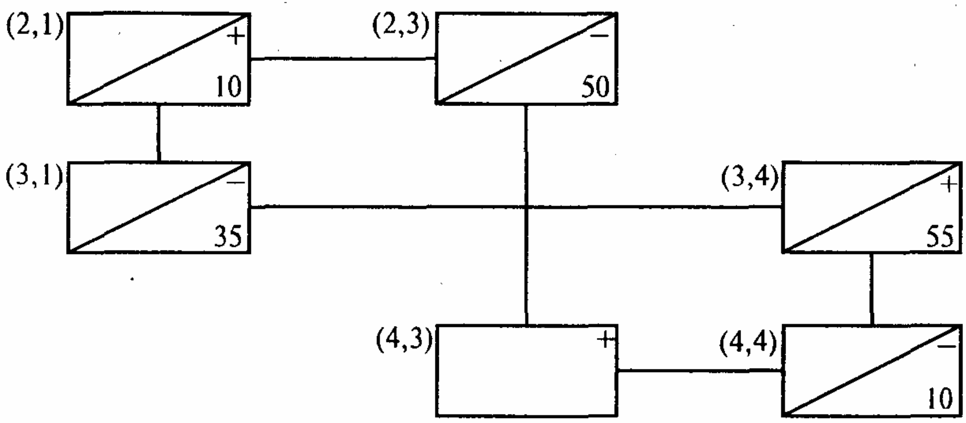


Рис. 7.7

Таблица 7.20

4	1 35	2 5	5
3 20	2	3 40	7
4 25	4	5	2 65
0	0	0 10	0

$$\begin{pmatrix} 2 & 0 & 0 & 5 \\ 0 & 0 & 0 & 6 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}. \qquad \textbf{(7.20)}$$

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

УПРАЖНЕНИЯ

В задачах **7.10, 7.11**: **а)** составить экономико-математическую модель задачи; **б)** найти оптимальное распределение поставок и минимальные затраты на перевозку, выполнив первоначальное распределение поставок методом наименьших затрат.

7.10.

Пос-тав-щики	Мощ-ность постав-щиков	Потребители и их спрос		
		1	2	3
		60	60	50
1	50	2	3	2
2	70	2	4	5
3	60	6	5	7

7.11.

Пос-тав-щики	Мощ-ность постав-щиков	Потребители и их спрос			
		1	2	3	4
		450	250	100	100
1	200	6	4	4	5
2	300	6	9	5	8
3	100	8	2	10	6

Решить транспортные задачи 7.10 (см. выше), **7.12, 7.13,** составив первоначальное распределение поставок методом "северо-западного" угла.

7.12.

Пос-тав-щики	Мощ-ность постав-щиков	Потребители и их спрос			
		1	2	3	4
		15	25	8	12
1	25	2	4	3	6
2	18	3	5	7	5
3	12	1	8	4	5
4	15	4	3	2	8

7.13.

Пос-тав-щики	Мощ-ность постав-щиков	Потребители и их спрос			
		1	2	3	4
		50	50	40	60
1	30	5	4	6	3
2	70	4	5	5	8
3	70	7	3	4	7

В задачах **7.14, 7.15** закончить решения транспортных задач начиная с заданных распределений поставок.

7.14.

Пос-тав-щики	Мощ-ность постав-щиков	Потребители и их спрос			
		1	2	3	4
		15	25	8	12
1	95	5 / 45	4 / 50	13	9
2	35	2	7	9 / 35	8
3	55	9	7 / 35	11	7 / 20
4	75	1	6	1	1 / 35

7.15.

Пос-тав-щики	Мощ-ность постав-щиков	Потребители и их спрос			
		1	2	3	4
		50	10	20	40
1	30	5	6	1 / 20	2 / 10
2	50	3 / 10	1 / 10	5	2 / 30
3	20	8 / 20	4	2	5
4	20	6 / 20	5	2	4

7.16. Решить транспортную задачу.

Пос-тав-щики	Мощ-ность постав-щиков	Потребители и их спрос					
		1	2	3	4	5	6
		10	35	15	25	55	10
1	30	5	7	1	5	4	9
2	5	7	5	8	6	3	4
3	45	6	4	8	3	2	5
4	50	3	1	7	4	2	3

Глава 8. МОДЕЛИ ЦЕЛОЧИСЛЕННОГО ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

8.1. Постановка задачи целочисленного программирования

По смыслу значительной части экономических задач, относящихся к задачам линейного программирования, компоненты решения должны выражаться в целых числах, т.е. быть целочисленными. К ним относятся, например, задачи, в которых переменные означают количество единиц неделимой продукции, число станков при загрузке оборудования, число судов при распределениях по линиям, число турбин в энергосистеме, число вычислительных машин в управляющем комплексе и многие другие.

Задача линейного целочисленного программирования формулируется следующим образом: *найти такое решение (план)* $X = (x_1, x_2, \dots, x_n)$, *при котором линейная функция*

$$Z = \sum_{j=1}^{n} c_j x_j \tag{8.1}$$

принимает максимальное или минимальное значение при ограничениях

$$\sum_{i=1}^{n} a_{ij} x_j = b_i, \quad i = 1, 2, \dots, m. \tag{8.2}$$

$$x_j \geq 0, \quad j = 1, 2, \dots, n. \tag{8.3}$$

$$x_j \text{ — целые числа.} \tag{8.4}$$

Следует отметить, что классическая транспортная задача и некоторые другие задачи транспортного типа “автоматически” обеспечивают решение задачи в целых числах (если, конечно, целочисленны параметры условий). Однако в общем случае условие целочисленности (8.4), добавляемое к обычным задачам линейного программирования, существенно усложняет ее решение.

Для решения задач линейного целочисленного программирования используется ряд методов. Самый простой из них — обычный метод линейного программирования. В случае если компоненты оптимального решения оказываются нецелочисленными, их округляют до ближайших целых чисел. Этот метод применяют тогда, когда отдельная единица совокупности составляет малую часть объема всей совокупности. В противном случае округление может привести к далекому от оптимального целочисленному решению, поэтому используют специально разработанные методы.

Методы целочисленной оптимизации можно разделить на три основные группы: а) методы отсечения; б) комбинаторные методы; в) приближенные методы. Остановимся подробнее на методах отсечения.

8.2. Методы отсечения. Метод Гомори

Сущность методов отсечения состоит в том, что сначала задача решается без условия целочисленности. Если полученный план целочисленный, задача решена. В противном случае к ограничениям задачи добавляется новое ограничение, обладающее следующими свойствами:

- *оно должно быть линейным;*
- *должно отсекать найденный оптимальный нецелочисленный план;*
- *не должно отсекать ни одного целочисленного плана.*

Дополнительное ограничение, обладающее указанными свойствами, называется ***правильным отсечением.***

Далее задача решается с учетом нового ограничения. После этого в случае необходимости добавляется еще одно ограничение и т. д.

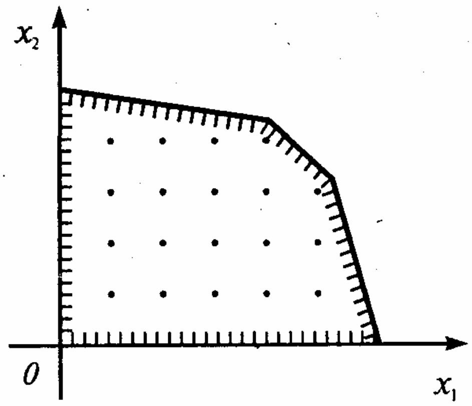


Рис. 8.1

Геометрически добавление каждого линейного ограничения отвечает проведению прямой (гиперплоскости), которая отсекает от многоугольника (многогранника) решений некоторую его часть вместе с оптимальной точкой с нецелыми координатами, но не затрагивает ни одной из целых точек этого многогранника. В результате новый многогранник решений содержит все целые точки, заключавшиеся в первоначальном многограннике решений и соответственно полученное при этом многограннике оптимальное решение будет целочисленным (рис. 8.1).

Один из алгоритмов решения задачи линейного целочисленного программирования (8.1)—(8.4), предложенный Гомори, основан на симплексном методе и использует достаточно простой способ построения правильного отсечения.

Пусть задача линейного программирования (8.1)—(8.3) имеет конечный оптимум и на последнем шаге ее решения симплексным методом получены следующие уравнения, выражающие основные переменные x_1, x_2, ..., x_j, ..., x_m через неосновные переменные x_{m+1}, x_{m+2}, ..., x_{m+i}, ..., x_n оптимального решения

$$\begin{cases} x_1 = \beta_1 - \alpha_{1\,m+1}x_{m+1} - \ldots - \alpha_{1n}x_n, \\ x_2 = \beta_2 - \alpha_{2\,m+1}x_{m+1} - \ldots - \alpha_{2n}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_i = \beta_i - \alpha_{i\,m+1}x_{m+1} - \ldots - \alpha_{in}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m = \beta_m - \alpha_{m\,m+1}x_{m+1} - \ldots - \alpha_{mn}x_n, \end{cases} \qquad \textbf{(8.5)}$$

так, что оптимальным решением задачи (8.1)—(8.3) является $X^* = (\beta_1, \beta_2, \ldots, \beta_i, \ldots, \beta_m, 0, 0, \ldots, 0)$, в котором, например β_i —

нецелая компонента. В этом случае можно доказать, что неравенство[1]

$$\{\beta_i\}-\{\alpha_{i\,m+1}\}x_{m+1}-\ldots-\{\alpha_{in}\}x_n\le 0, \qquad \textbf{(8.6)}$$

сформированное по i-му уравнению системы (8.5), обладает всеми свойствами правильного отсечения.

Для решения задачи целочисленного линейного программирования (8.1)—(8.4) методом Гомори используется следующий алгоритм:

1. Симплексным методом решить задачу (8.1)—(8.3) без учета условия целочисленности. Если все компоненты оптимального плана целые, то он является оптимальным и для задачи целочисленного программирования (8.1)—(8.4). Если первая задача (8.1)—(8.3) неразрешима (т.е. не имеет конечного оптимума или условия ее противоречивы), то и вторая задача (8.1)—(8.4) также неразрешима.

2. Если среди компонент оптимального решения есть нецелые, то выбрать компоненту с наибольшей целой частью и по соответствующему уравнению системы (8.5) сформировать правильное отсечение (8.6).

3. Неравенство (8.6) введением дополнительной неотрицательной целочисленной переменной преобразовать в равносильное уравнение

$$\{\beta_i\}-\{\alpha_{i\,m+1}\}x_{m+1}-\ldots-\{\alpha_{in}\}x_n+x_{n+1}, \qquad \textbf{(8.7)}$$

и включить его в систему ограничений (8.2).

4. Полученную расширенную задачу решить симплексным методом. Если найденный оптимальный план будет целочисленным,

---

[1] В неравенстве (8.6) присутствует символ { }, означающий дробную часть числа. *Целой частью числа a* называется наибольшее целое число $[a]$, не превосходящее a, а дробной частью числа — число $\{a\}$, равное разности между этим числом и его целой частью, т.е. $\{a\}=a-[a]$. Например, для $a=2\frac{1}{3}$ $[a]=2$, $\{a\}=2\frac{1}{3}-2=\frac{1}{3}$; для $a=-2\frac{1}{3}$ $[a]=-3$ (обратите внимание, именно −3, а не −2) и $\{a\}=-2\frac{1}{3}-(-3)=\frac{2}{3}$.

то задача целочисленного программирования (8.1)—(8.4) решена. В противном случае вернуться к п. 2 алгоритма.

Если задача разрешима в целых числах, то после конечного числа шагов (итераций) оптимальный целочисленный план будет найден.

Если в процессе решения появится уравнение (выражающее основную переменную через неосновные) с нецелым свободным членом и целыми остальными коэффициентами, то соответствующее уравнение не имеет решения в целых числах. В этом случае и данная задача не имеет целочисленного оптимального решения.

▷ **8.1.** Для приобретения оборудования по сортировке зерна фермер выделяет 34 ден. ед. Оборудование должно быть размещено на площади, не превышающей 60 кв. м. Фермер может заказать оборудование двух видов: менее мощные машины типа *A* стоимостью 3 ден. ед., требующие производственную площадь 3 кв. м (с учетом проходов) и обеспечивающие производительность за смену 2 т зерна, и более мощные машины типа *B* стоимостью 4 ден. ед., занимающие площадь 5 кв. м и обеспечивающие производительность за смену 3 т сортового зерна.

Требуется составить оптимальный план приобретения оборудования, обеспечивающий максимальную общую производительность при условии, что фермер может приобрести не более 8 машин типа *B*.

Решение. Обозначим через x_1, x_2 количество машин соответственно типа *A* и *B*, через Z — общую производительность. Тогда математическая модель задачи примет вид:

$$Z = 2x_1 + 3x_2 \to \max \qquad \textbf{(8.1')}$$

при ограничениях:

$$\begin{cases} 3x_1 + 5x_2 \le 60, & (1) \\ 3x_1 + 4x_2 \le 34, & (2) \\ x_2 \le 8, & (3) \end{cases} \qquad \textbf{(8.2')}$$

$$x_1 \ge 0,\ x_2 \ge 0, \qquad \textbf{(8.3')}$$

$$x_1, x_2 \text{ — целые числа.} \qquad \textbf{(8.4')}$$

Приведем задачу к каноническому виду, введя дополнительные неотрицательные переменные x_3, x_4, x_5. Получим систему ограничений:

$$\begin{cases} 3x_1 + 5x_2 + x_3 & = 60, \\ 3x_1 + 4x_2 \quad + x_4 & = 34, \\ \quad\;\; x_2 \qquad\quad + x_5 & = \;\; 8, \end{cases} \qquad \textbf{(8.5')}$$

$$x_j \ge 0,\ j = 1,2,\ldots,5.$$

Решаем задачу симплексным методом. Для наглядности решение иллюстрируем графически (рис. 8.2).

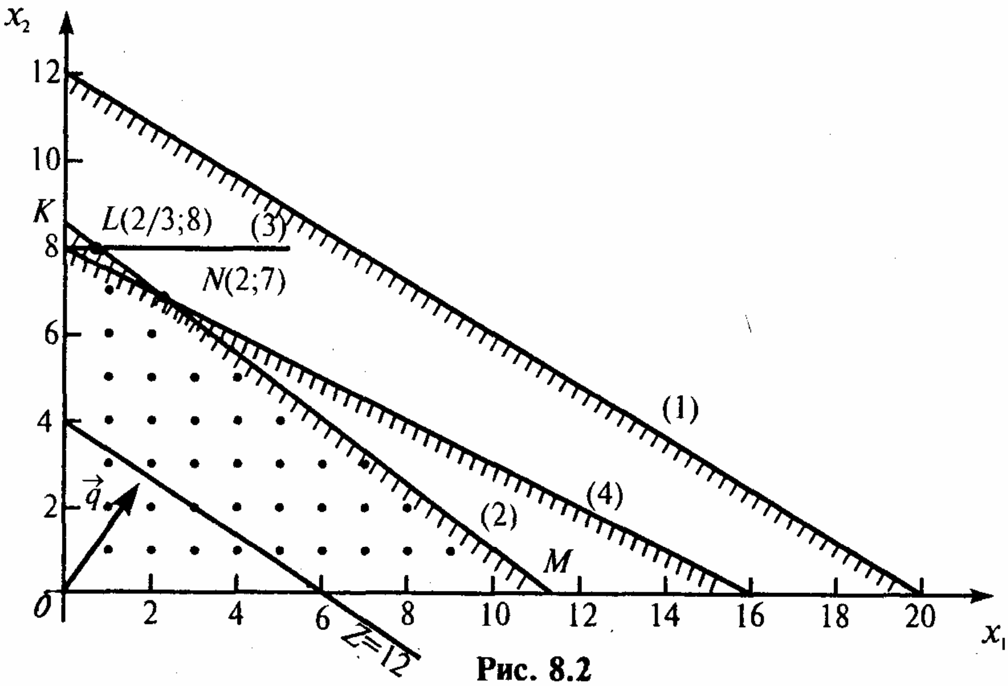


Рис. 8.2

На рис. 8.2 *OKLM* — область допустимых решений задачи (8.1′) – (8.3′), ограниченная прямыми (1), (2), (3) и осями координат; *L*(2/3; 8) — точка оптимального, но нецелочисленного решения задачи (8.1′) – (8.3′); (4) — прямая, отсекающая это нецелочисленное решение; *OKNM* — область допустимых решений расширенной задачи (8.1′) – (8.3′), (8.6′); *N*(2; 7) — точка оптимального целочисленного решения.

I шаг. Основные переменные x_3, x_4, x_5; неосновные переменные x_1, x_2.

$$\begin{cases} x_3 = 60 - 3x_1 - 5x_2, \\ x_4 = 34 - 3x_1 - 4x_2, \\ \boxed{x_5 = 8 \qquad\quad - x_2,} \end{cases}$$
$$Z = 2x_1 + 3x_2.$$

Первое базисное решение $X_1 = (0;\ 0;\ 60;\ 34;\ 8)$ — допустимое. Соответствующее значение линейной функции $Z_1 = 0$.

Переводим в основные переменные переменную x_2, которая входит в выражение линейной функции с наибольшим положительным коэффициентом. Находим максимально возможное значение переменной x_2, которое "позволяет" принять система ограничений, из условия минимума соответствующих отношений:

$$x_2 = \min\left\{\frac{60}{5}; \frac{34}{4}; \frac{8}{1}\right\} = 8,$$

т.е. разрешающим (выделенным) является третье уравнение. При $x_2 = 8$ в этом уравнении $x_5 = 0$, и в неосновные переходит переменная x_5.

II шаг. Основные переменные x_2, x_3, x_4; неосновные переменные x_1, x_5.

$$\begin{cases} x_2 = \ \ 8 \qquad\ \ - x_5, \\ x_3 = 20 - 3x_1 + 5x_5, \\ \boxed{x_4 = \ \ 2 - 3x_1 + 4x_5,} \end{cases}$$
$$Z = 24 + 2x_1 - 3x_5.$$

$X_2 = (0;\ 8;\ 20;\ 2;\ 0)$; $Z_2 = 24$. Переводим в основные переменную x_1, $x_1 = \min\left\{\infty; \frac{20}{3}; \frac{2}{3}\right\} = \frac{2}{3}$, а в неосновные x_4.

III шаг. Основные переменные x_1, x_2, x_3; неосновные переменные x_4, x_5.

После преобразований получим:

$$\begin{cases} x_1 = \frac{2}{3} - \frac{1}{3}x_4 + \frac{4}{3}x_5, \\ x_2 = 8 - x_5, \\ x_3 = 18 + x_4 + x_5, \end{cases}$$

$$Z = 25\frac{1}{3} - \frac{2}{3}x_4 - \frac{1}{3}x_5.$$

$$X_3 = (\frac{2}{3};\ 8;\ 18;\ 0;\ 0);\ Z_3 = 25\frac{1}{3}.$$

Базисное решение X_3 оптимально для задачи (8.1′) – (8.3′) ($Z_{\max} = Z_3 = 25\frac{1}{3}$), так как в выражении линейной функции отсутствуют неосновные переменные с положительными коэффициентами.

Однако решение X_3 не удовлетворяет условию целочисленности (8.4′). По первому уравнению с переменной x_1, получившей нецелочисленное значение в оптимальном решении (2/3), составляем дополнительное ограничение (8.6):

$$\left\{\frac{2}{3}\right\} - \left\{\frac{1}{3}\right\}x_4 - \left\{\frac{4}{3}\right\}x_5 \le 0.$$

Обращаем внимание на то, что согласно (8.5) и (8.6) берем дробную часть *свободного члена* с *тем же знаком*, который он имеет в уравнении, а дробные части *коэффициентов при неосновных переменных* x_4 и x_5 — с *противоположными знаками.*

Так как дробные части $\left\{\frac{2}{3}\right\} = \left\{0 + \frac{2}{3}\right\} = \frac{2}{3}$, $\left\{\frac{1}{3}\right\} = \left\{0 + \frac{1}{3}\right\} = \frac{1}{3}$, $\left\{-\frac{4}{3}\right\} = \left\{-2 + \frac{2}{3}\right\} = \frac{2}{3}$, то последнее неравенство запишем в виде

$$\frac{2}{3} - \frac{1}{3}x_4 - \frac{2}{3}x_5 \le 0. \qquad \textbf{(8.6′)}$$

Введя дополнительную целочисленную переменную $x_6 \geq 0$, получим равносильное неравенству (8.6′) уравнение

$$\frac{2}{3} - \frac{1}{3}x_4 - \frac{2}{3}x_5 + x_6 = 0. \qquad \textbf{(8.7′)}$$

Уравнение (8.7′) необходимо включить в систему ограничений (8.5′) исходной канонической задачи, после чего повторить процесс решения задачи симплексным методом применительно к расширенной задаче. При этом для сокращения числа шагов (итераций) рекомендуется вводить дополнительное уравнение (8.7′) в систему, полученную на последнем шаге решения задачи (без условия целочисленности).

IV шаг. Основные переменные x_1, x_2, x_3, x_6; неосновные переменные x_1, x_2.

$$\begin{cases} x_1 = \frac{2}{3} - \frac{1}{3}x_4 + \frac{4}{3}x_5, \\ x_2 = 8 - x_5, \\ x_3 = 18 + x_4 + x_5, \\ \boxed{x_6 = -\frac{2}{3} + \frac{1}{3}x_4 + \frac{2}{3}x_5.} \end{cases}$$

Базисное решение $X_4 = (\frac{2}{3};\ 8;\ 18;\ 0;\ 0;\ -\frac{2}{3})$ — недопустимое. (Заметим, что после включения в систему ограничений дополнительного уравнения, соответствующего правильному отсечению, всегда будет получаться недопустимое базисное решение).

Для получения допустимого базисного решения необходимо перевести в основные переменную, входящую с положительным коэффициентом в уравнение, в котором свободный член отрицательный, т.е. x_1 или x_5 (на этом этапе линейную функцию не рассматриваем). Переводим в основные, например, переменную x_5.[1]

V шаг. Основные переменные x_1, x_2, x_3, x_5; неосновные переменные x_1, x_2.

[1] Можно убедиться, что при этом решение задачи короче.

Получим после преобразований:

$$\begin{cases} x_1 = 2 - x_4 + 2x_6, \\ x_2 = 7 + \frac{1}{2}x_4 - \frac{3}{2}x_6, \\ x_3 = 19 + \frac{1}{2}x_4 + \frac{3}{2}x_6, \\ x_5 = 1 - \frac{1}{2}x_4 + \frac{3}{2}x_6, \end{cases}$$

$$Z = 25 - \frac{1}{2}x_4 - \frac{1}{2}x_6.$$

$$X_5 = (2;\ 7;\ 19;\ 0;\ 1;\ 0);\ Z_5 = 25.$$

Так как в выражении линейной функции нет основных переменных с положительными коэффициентами, то X_5 — оптимальное решение.

Итак, $Z_{max} = 25$ при оптимальном целочисленном решении $X^* = X_5 = (2;\ 7;\ 19;\ 0;\ 1;\ 0)$, т.е. максимальную производительность 25 т сортового зерна за смену можно получить приобретением 2 машин типа A и 7 машин типа B; при этом незанятая площадь помещения составит 19 кв. м, остатки денежных средств из выделенных равны 0, в резерве для покупки — 1 машина типа B (шестая компонента содержательного смысла не имеет).

Замечание. Для геометрической интерпретации на плоскости Ox_1x_2 (*см.* рис.8.2) отсечения (8.6′) необходимо входящие в него переменные x_4 и x_5 выразить через переменные x_1 и x_2. Получим (*см.* 2-е и 3-е уравнения системы ограничений (8.5′)):

$$\frac{2}{3} - \frac{1}{3}(34 - 3x_1 - 4x_2) - \frac{2}{3}(8 - x_2) \le 0 \text{ или } x_1 + 2x_2 \le 16.$$

(*см.* отсечение прямой (4) на рис 8.2).▶

▷ **8.2.** Имеется достаточно большое количество бревен длиной 3 м. Бревна следует распилить на заготовки двух видов: длиной 1,2 м и длиной 0,9 м, причем заготовок каждого вида должно быть получено не менее 50 шт. и 81 шт. соответственно. Каждое бревно можно распилить на указанные заготовки несколькими способами: 1) на 2 заготовки по 1,2 м; 2) на 1 заготовку по 1,2 м и 2 заготовки по 0,9 м; 3) на 3 заготовки по 0,9 м. Найти число бревен, распиливаемых каждым способом, с тем чтобы заготовок любого вида было получено из наименьшего числа бревен.

Решение. Обозначим через x_1, x_2, x_3 число бревен, распиливаемых соответственно 1-, 2-и 3-м способами. Из них можно получить $2x_1 + x_2$ заготовок по 1,2 м и $2x_1 + 3x_2$ заготовок по 0,9 м. Общее количество бревен обозначим Z. Тогда математическая модель задачи примет вид:

$$Z = x_1 + x_2 + x_3 \to \min \tag{8.1''}$$

при ограничениях

$$\begin{cases} 2x_1 + x_2 \geq 50, \\ 2x_2 + 3x_3 \geq 81, \end{cases} \tag{8.2''}$$

$$x_j \geq 0, \ \ j = 1, 2, 3, \tag{8.3''}$$

$$x_j \text{ — целые числа.} \tag{8.4''}$$

Введя дополнительные переменные $x_4 \geq 0$, $x_5 \geq 0$, приведем систему неравенств к равносильной системе уравнений

$$\begin{cases} 2x_1 + x_2 - x_4 = 50, \\ 2x_2 + 3x_3 - x_5 = 81, \end{cases} \tag{8.5''}$$
$$x_j \geq 0, \ j = 1,2,\ldots,5.$$

Решая полученную каноническую задачу (без условия целочисленности) симплексным методом, на последнем, III шаге, решения найдем следующие выражения основных переменных и линейной функции через неосновные переменные (рекомендуем студентам получить их самостоятельно).

III шаг. Основные переменные x_1, x_2; неосновные переменные x_3, x_4, x_5.

$$\begin{cases} x_1 = 4\frac{3}{4} + \frac{3}{4}x_3 + \frac{1}{2}x_4 - \frac{1}{4}x_5, \\ x_2 = 40\frac{1}{2} - \frac{3}{2}x_3 \qquad\qquad + \frac{1}{2}x_5, \end{cases}$$

$$Z = 45\frac{1}{4} + \frac{1}{4}x_3 + \frac{1}{2}x_4 + \frac{1}{4}x_5,$$

т.е. $Z_{\min} = 45\frac{1}{4}$ при оптимальном решении $X_3 = (4\frac{3}{4};\ 40\frac{1}{2};\ 0;\ 0;\ 0)$.

Получили, что две компоненты оптимального решения $x_1 = 4\frac{3}{4}$ и $x_2 = 40\frac{1}{2}$ не удовлетворяют условию целочисленности (8.4″), причем бо́льшую целую часть имеет компонента x_2. В соответствии с п.2 алгоритма решения задачи целочисленного программирования (*см.* с. 156) по второму уравнению, содержащему эту переменную x_2, составляем дополнительное ограничение (8.6):

$$\left\{40\frac{1}{2}\right\} - \left\{\frac{3}{2}\right\}x_3 - \left\{-\frac{1}{2}\right\}x_5 \le 0.$$

Найдем дробные части $\left\{40\frac{1}{2}\right\} = \left\{40 + \frac{1}{2}\right\} = \frac{1}{2}$, $\left\{\frac{3}{2}\right\} = \left\{1 + \frac{1}{2}\right\} = \left\{\frac{1}{2}\right\}$, $\left\{-\frac{1}{2}\right\} = \left\{-1 + \frac{1}{2}\right\} = \frac{1}{2}$ и запишем последнее неравенство в виде

$$\frac{1}{2} - \frac{1}{2}x_3 - \frac{1}{2}x_5 \le 0. \qquad \textbf{(8.6″)}$$

Введя дополнительную переменную $x_6 \ge 0$, получим равносильное неравенству (8.6″) уравнение

$$\frac{1}{2} - \frac{1}{2}x_3 - \frac{1}{2}x_5 + x_6 = 0. \qquad \textbf{(8.7″)}$$

Выразим из (8.7″) дополнительную переменную x_6 и полученное уравнение введем в систему ограничений, которую мы имели на последнем, III шаге, решения задачи (8.1″) – (8.3″) (без условия целочисленности).

IV шаг. Основные переменные x_1, x_2, x_6; неосновные переменные x_3, x_4, x_5.

$$\begin{cases} x_1 = 4\frac{3}{4} + \frac{3}{4}x_3 + \frac{1}{2}x_4 - \frac{1}{4}x_5, \\ x_2 = 40\frac{1}{2} - \frac{3}{2}x_3 \qquad + \frac{1}{2}x_5, \\ x_6 = -\frac{1}{2} + \frac{1}{2}x_3 \qquad + \frac{1}{2}x_5, \end{cases}$$

$$Z = 45\frac{1}{4} + \frac{1}{4}x_3 + \frac{1}{2}x_4 + \frac{1}{4}x_5.$$

Решая эту расширенную задачу симплексным методом (предлагаем студентам выполнить самостоятельно), получим следующее.

V шаг. Основные переменные x_1, x_2, x_3; неосновные переменные x_4, x_5, x_6.

$$\begin{cases} x_1 = 5\frac{1}{2} + \frac{1}{2}x_4 - x_5 + \frac{3}{2}x_6, \\ x_2 = 39 \qquad + 2x_5 - 3x_6, \\ x_3 = 1 \qquad - x_5 + 2x_6, \end{cases}$$

$$Z = 45\frac{1}{2} + \frac{1}{2}x_4 + \frac{1}{2}x_6,$$

т.е. $Z_{\min} = 45\frac{1}{2}$ при оптимальном решении $X_5 = (5\frac{1}{2};\ 39;\ 1;\ 0;\ 0;\ 0)$.

Полученное оптимальное решение расширенной задачи (8.1″) – (8.3″), (8.6″) вновь не удовлетворяет условию целочисленности (8.4″). По первому уравнению с переменной x_1, получившей нецелочисленное значение в оптимальном решении $(5\frac{1}{2})$, составляем второе дополнительное ограничение (8.6):

$$\left\{5\frac{1}{2}\right\} - \left\{-\frac{1}{2}\right\}x_4 - \{1\}x_5 - \left\{-\frac{3}{2}\right\}x_6 \le 0,$$

которое приводим к виду $\frac{1}{2}-\frac{1}{2}x_4-\frac{1}{2}x_6 \le 0.$

С помощью дополнительной переменной $x_7 \ge 0$ приводим это неравенство к равносильному уравнению, которое включаем в систему ограничений, полученную на последнем, V шаге, решения расширенной задачи (8.1″) – (8.3″), (8.6″) симплексным методом.

VI шаг. Основные переменные x_1, x_2, x_3, x_7; неосновные переменные x_4, x_5, x_6.

$$\begin{cases} x_1 = 5\frac{1}{2} + \frac{1}{2}x_4 - x_5 + \frac{3}{2}x_6, \\ x_2 = 39 + 2x_5 - 3x_6, \\ x_3 = 1 - x_5 + 2x_6, \\ x_7 = -\frac{1}{2} + \frac{1}{2}x_4 + \frac{1}{2}x_6, \end{cases}$$

$$Z = 45\frac{1}{2} + \frac{1}{2}x_4 + \frac{1}{2}x_6.$$

Опуская дальнейшее решение задачи симплексным методом (предлагаем сделать это самим студентам), получим на заключительном, VII шаге, следующее.

VII шаг. Основные переменные x_1, x_2, x_3, x_4; неосновные переменные x_5, x_6, x_7.

$$\begin{cases} x_1 = 6 - x_5 - x_6 - x_7, \\ x_2 = 39 + 2x_5 - 3x_6, \\ x_3 = 1 - x_5 + 2x_6, \\ x_4 = 1 - x_6 + 2x_7, \end{cases}$$

$$Z = 46 + x_7,$$

т.е. $Z_7 = 46$ при $X_7 = (6; 39; 1; 1; 0; 0; 0)$.

Так как в выражении линейной функции нет неосновных переменных с отрицательными коэффициентами, то X_7 — оптимальное целочисленное решение исходной задачи.

Следует обратить внимание на то, что в полученном выражении линейной функции Z отсутствуют неосновные переменные

x_5 и x_6. Это означает, что, вообще говоря, существует бесконечное множество оптимальных решений (любых, не обязательно целочисленных), при которых $Z^* = Z_{min} = 46$. Эти решения получаются при значении неосновной переменной x_7 (входящей в выражение для Z), равной нулю (т.е. при $x_7 = 0$), и при *любых* значениях неосновных переменных x_5 и x_6 (не входящих в выражение для Z), которые "позволяет" принять система ограничений: $0 \le x_5 \le \min\{6;\ 39/2;\ 1;\ \infty\} = 1$ и $0 \le x_6 \le \min\{\infty;\ 13;\ \infty;\ 1\} = 1$, т.е. при $0 \le x_5 \le 1$ и $0 \le x_6 \le 1$. Но в силу условия целочисленности переменные x_5 и x_6 могут принять только значения 0 или 1. Поэтому задача будет иметь четыре целочисленных оптимальных решения, когда x_5 и x_6 в любой комбинации принимают значения 0 или 1, а $x_7 = 0$. Подставляя эти значения в систему ограничений на VII шаге, найдем эти оптимальные решения:

$$X_7^{(1)} = (6;39;1;1;0;0;0),\ X_7^{(2)} = (7;36;3;0;0;1;0),$$

$$X_7^{(3)} = (5;41;0;1;1;0;0),\ X_7^{(4)} = (6;38;2;0;1;1;0).$$

Наличие альтернативных оптимальных целочисленных решений позволяет осуществить выбор одного из них, руководствуясь дополнительными критериями, не учитываемыми в математической модели задачи. Например, из условия данной задачи следует, что распиливание бревен не дает отходов лишь по третьему способу, поэтому естественно при выборе одного из четырех оптимальных решений отдать предпочтение решению $X_7^{(3)}$, при котором максимальное число бревен ($x_2 = 41$) распиливается без отходов.

Итак, $Z_{min} = 46$ при оптимальных целочисленных решениях **(5; 41; 0)**, (6; 39; 1), (7; 36; 3), (6; 38; 2). (При записи оптимальных решений мы оставили лишь первые три компоненты, выражающие число бревен, распиливаемых соответственно первым, вторым и третьим способами, и исключили последние четыре компоненты, не имеющие смыслового значения).▶

Недостатком метода Гомори является требование целочисленности для всех переменных — как основных (выражающих, например, в задаче об использовании ресурсов единицы продукции), так и дополнительных переменных (выражающих величину неиспользованных ресурсов, которые могут быть и дробными).

8.3. Понятие о методе ветвей и границ

Метод ветвей и границ — один из ***комбинаторных*** методов. Его суть заключается в упорядоченном переборе вариантов и рассмотрении лишь тех из них, которые оказываются по определенным признакам перспективными, и отбрасывании бесперспективных вариантов.

Метод ветвей и границ состоит в следующем: множество допустимых решений (планов) некоторым способом разбивается на подмножества, каждое из которых этим же способом снова разбивается на подмножества. Процесс продолжается до тех пор, пока не получено оптимальное целочисленное решение исходной задачи.

Пусть задача 1 (задача (8.1)—(8.3) максимизации линейной функции Z (без учета целочисленности переменных) решена симплексным методом и известны нижняя и верхняя границы для каждой целочисленной переменной x_j: $v_j \le x_j \le w_j$ (j = 1, 2, ..., n), а также нижняя граница линейной функции Z_0, т.е. при любом плане X $Z(X) \ge Z_0$. Предположим для определенности, что только первая компонента x_1^* оптимального плана X^* задачи 1 не удовлетворяет условию целочисленности. Тогда из области допустимых решений задачи 1 исключается область: $[x_1^*] < x_1^* < [x_1^*] + 1$, где $[x_1^*]$ — целая часть числа x_1^*. В результате из задачи 1 формируют две задачи: 2 и 3, отличающиеся друг от друга тем, что в задаче 2 кроме ограничений (8.2) задачи 1 добавлено ограничение $v_1 \le x_1^* \le [x_1^*] + 1$, а в задаче 3 кроме тех же ограничений (8.2) добавлено ограничение $[x_1^*] + 1 \le x_1^* \le w_1$. Получим список из двух задач: 2 и 3.

Решаем одну из них (в любом порядке). В зависимости от полученного решения список задач расширяется, либо уменьшается.

Если в результате решения одной из задач 2 или 3 получен нецелочисленный оптимальный план, для которого $Z(X^*) \le Z_0$, то данная задача исключается из списка. Если $Z(X^*) > Z_0$, то из данной задачи формируются новые две задачи.

Если полученное решение X^* удовлетворяет условию целочисленности и $Z(X^*) > Z_0$, то значение Z_0 исправляется и за величину Z_0 принимается оптимум линейной функции полученного оптимального целочисленного плана.

Процесс продолжается до тех пор, пока список задач не будет исчерпан, т.е. все задачи не будут решены.

Проиллюстрируем метод ветвей и границ на примере.

▷ **8.3.** Решить задачу

$$Z = 3x_1 + x_2 \to \max$$

при ограничениях:

$$\begin{cases} 4x_1 + 3x_2 \le 18, \\ x_1 + 2x_2 \le 6, \\ 0 \le x_1 \le 5, \\ 0 \le x_2 \le 4, \end{cases}$$

$$x_1, x_2 \text{ — целые числа.}$$

Решение. За нижнюю границу линейной функции примем, например, ее значение в точке (0,0), т.е. $Z_0 = Z\,(0;\,0) = 0$.

I этап. Решая задачу симплексным методом, получим $Z_{\max} = 13$ при $X_1^* = (4{,}5;\ 0;\ 0;\ 1{,}5;\ 0{,}5;\ 4)$; так как первая компонента x_1^* дробная, то из области решения исключается полоса, содержащая дробное оптимальное значение x_1^*, т.е. $4 < x_1 < 5$. Поэтому задача 1 разбивается на две задачи 2 и 3:

Задача 2

$$Z = 3x_1 + x_2 \to \max$$

при ограничениях:

$$\begin{cases} 4x_1 + 3x_2 \le 18, \\ x_1 + 2x_2 \le 6, \\ 0 \le x_1 \le 4, \\ 0 \le x_2 \le 4, \end{cases}$$

$$x_1, x_2 \text{ — целые числа.}$$

Задача 3

$$Z = 3x_1 + x_2 \to \max$$

при ограничениях:

$$\begin{cases} 4x_1 + 3x_2 \le 18, \\ x_1 + 2x_2 \le 6, \\ 5 \le x_1 \le 5, \\ 0 \le x_2 \le 4, \end{cases}$$

$$x_1, x_2 \text{ — целые числа.}$$

Список задач: 2 и 3. Нижняя граница линейной функции не изменилась: $Z_0 = 0$.

II этап. Решаем (по выбору) одну из задач списка, например задачу 3 симплексным методом.

Получим, что условия задачи 3 противоречивы.

III этап. Решаем задачу 2 симплексным методом. Получим $Z_{\max} = 4\frac{2}{3}$ при $X_3^* = \left(4; \frac{2}{3}; 0; \frac{2}{3}; 0; \frac{10}{3}\right)$. Хотя $Z(X_3^*) = 4\frac{2}{3} > Z_0 = 0$, по-прежнему сохраняется $Z_0 = 0$, ибо план X_3 нецелочисленный. Так как x_2^* — дробное число, из области решений исключаем полосу $0 < x_2 < 1$ и задачу 2 разбиваем на две задачи 4 и 5.

Задача 4

$Z = 3x_1 + x_2 \to \max$

при ограничениях:

$$\begin{cases} 4x_1 + 3x_2 \le 18, \\ x_1 + 2x_2 \le 6, \\ 0 \le x_1 \le 4, \\ 0 \le x_2 \le 0, \end{cases}$$

x_1, x_2 — целые числа.

Задача 5

$Z = 3x_1 + x_2 \to \max$

при ограничениях:

$$\begin{cases} 4x_1 + 3x_2 \le 18, \\ x_1 + 2x_2 \le 6, \\ 0 \le x_1 \le 4, \\ 1 \le x_2 \le 4, \end{cases}$$

x_1, x_2 — целые числа.

Список задач: 4 и 5. Значение $Z_0 = 0$.

IV этап. Решаем задачу 4 симплексным методом.

Получим $Z_{\max} = 12$ при $X_4^* = (4; 0; 2; 2; 0; 0)$. Задачу исключаем из списка, но при этом меняем Z_0; $Z_0 = Z(X_4^*) = 12$, ибо план X_4^* целочисленный.

V этап. Решаем задачу 5 симплексным методом.

Получим $Z_{\max} = 12{,}25$ при $X_5^* = (3{,}75;\ 1;\ 0;\ 0{,}25;\ 0{,}25;\ 0;\ 3)$. Z_0 не меняется, т.е. $Z_0 = 12$, ибо план X_5^* нецелочисленный. Так как x_1^* — дробное, из области решений исключаем полосу $3<x_1<4$, и задача 5 разбивается на две задачи: 6 и 7.

Задача 6	Задача 7
$Z = 3x_1 + x_2 \to \max$	$Z = 3x_1 + x_2 \to \max$
при ограничениях:	при ограничениях:
$\begin{cases} 4x_1 + 3x_2 \le 18, \\ x_1 + 2x_2 \le 6, \\ 0 \le x_1 \le 3, \\ 1 \le x_2 \le 4, \end{cases}$	$\begin{cases} 4x_1 + 3x_2 \le 18, \\ x_1 + 2x_2 \le 6, \\ 4 \le x_1 \le 4, \\ 1 \le x_2 \le 4, \end{cases}$
x_1, x_2 — целые числа.	x_1, x_2 — целые числа.

Список задач: 6 и 7. Значение $Z_0 = 12$.

VI этап. Решаем одну из задач списка, например задачу 7, симплексным методом.

Получим, что условия задачи 7 противоречивы.

VII этап. Решаем задачу 6 симплексным методом. Получим $Z_{\max} = 10{,}5$ при $X_6^* = (3;\ 1{,}5;\ 1{,}5;\ 0;\ 0;\ 0{,}5;\ 2{,}5)$. Так как $Z(X_6^*) = 10{,}5 < Z_0 = 12$, то задача исключается из списка.

Итак, список задач исчерпан и оптимальным целочисленным решением исходной задачи будет $X^* = X_4^* = (4;\ 0;\ 2;\ 2;\ 0;\ 0)$, а оптимум линейной функции $Z_{\max} = 12$.▶

Замечание 1. Нетрудно видеть, что каждая последующая задача, составляемая в процессе применения метода ветвей и границ, отличается от предыдущей лишь одним неравенством — ограничением. Поэтому при решении каждой последующей задачи нет смысла решать ее симплексным методом с самого начала (с I шага). А целесообразнее начать решение с *последнего шага* (итерации) предыдущей задачи, из системы ограничений которой исключить "старые" (одно или два) уравнения — ограничения и ввести в эту систему "новые" уравнения — ограничения.

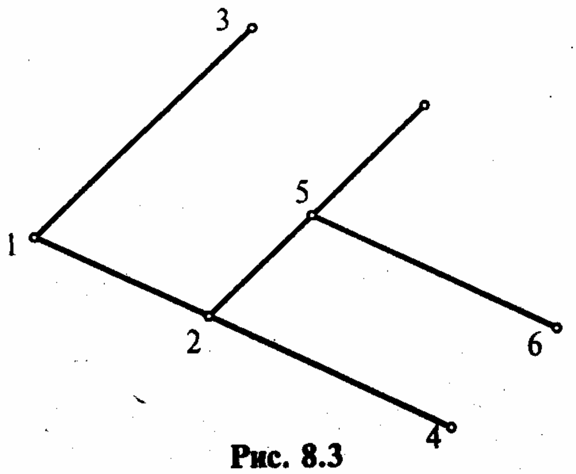


Рис. 8.3

Замечание 2. Название метода ветвей и границ объясняется тем, что в процессе решения задача последовательно "ветвится", заменяясь более простыми. Так, процесс решения данной задачи **8.3** можно представить в виде дерева (схемы), цифры в узлах (вершинах) которого обозначают номера задач (рис. 8.3).

УПРАЖНЕНИЯ

В задачах **8.4–8.7** методом Гомори (или методом ветвей и границ) найти оптимальные решения задач целочисленного линейного программирования. Дать геометрическую интерпретацию процесса решений задач.

8.4. $Z = 3x_1 + 2x_2 \to \max$

при ограничениях:

$$\begin{cases} x_1 + x_2 \le 13, \\ x_1 - x_2 \le 6, \\ -3x_1 + x_2 \le 9, \end{cases}$$

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

x_1, x_2 — целые числа

8.5. $Z = 5x_1 + 7x_2 \to \min$

при ограничениях:

$$\begin{cases} -3x_1 + 14x_2 \le 78, \\ 5x_1 - 6x_2 \le 26, \\ x_1 + 4x_2 \ge 25, \end{cases}$$

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

x_1, x_2 — целые числа

8.6. $Z = 2x_1 + x_2 \to \max$

при ограничениях:

$$\begin{cases} 6x_1 + 4x_2 \ge 24, \\ -3x_1 + 3x_2 \le 9, \\ -x_1 + 3x_2 \ge 3, \end{cases}$$

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

x_1, x_2 — целые числа

8.7. $Z = 6x_1 + x_2 \to \min$

при ограничениях:

$$\begin{cases} 3x_1 - x_2 \ge 9, \\ 2x_1 + 3x_2 \le 50, \\ -x_1 + 4x_2 \ge 18, \end{cases}$$

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

x_1, x_2 — целые числа

Глава 9. ЭЛЕМЕНТЫ ТЕОРИИ ИГР

9.1. Понятие об игровых моделях

На практике часто приходится сталкиваться с задачами, в которых необходимо принимать решения в условиях неопределенности, т.е. возникают ситуации, в которых две (или более) стороны преследуют различные цели, а результаты любого действия каждой из сторон зависят от мероприятий партнера. Такие ситуации, возникающие при игре в шахматы, шашки, домино и т.д., относятся к конфликтным: результат каждого хода игрока зависит от ответного хода противника, цель игры — выигрыш одного из партнеров. В экономике конфликтные ситуации встречаются очень часто и имеют многообразный характер. К ним относятся, например, взаимоотношения между поставщиком и потребителем, покупателем и продавцом, банком и клиентом. Во всех этих примерах конфликтная ситуация порождается различием интересов партнеров и стремлением каждого из них принимать оптимальные решения, которые реализуют поставленные цели в наибольшей степени. При этом каждому приходится считаться не только со своими целями, но и с целями партнера, и учитывать неизвестные заранее решения, которые эти партнеры будут принимать.

Для грамотного решения задач с конфликтными ситуациями необходимы научно обоснованные методы. Такие методы разработаны математической теорией конфликтных ситуаций, которая носит название ***теория игр***.

Ознакомимся с основными понятиями теории игр. Математическая модель конфликтной ситуации называется ***игрой***, стороны, участвующие в конфликте, — *игроками*, а исход конфликта — *выигрышем*. Для каждой формализованной игры вводятся *правила*, т.е. система условий, определяющая: 1) варианты действий игроков; 2) объем информации каждого игрока о поведении партнеров; 3) выигрыш, к которому приводит каждая совокупность действий. Как правило, выигрыш (или проигрыш) может быть задан количественно; например, можно оценить проигрыш нулем, выигрыш — единицей, а ничью — 1/2.

Игра называется *парной*, если в ней участвуют два игрока, и *множественной*, если число игроков больше двух. Мы будем рассматривать только парные игры. В них участвуют два игрока A и B, интересы которых противоположны, а под игрой будем понимать ряд действий со стороны A и B.

Игра называется *игрой с нулевой суммой*, или *антагонистической*, если выигрыш одного из игроков равен проигрышу другого, т.е. для полного задания игры достаточно указать величину одного из них. Если обозначить a — выигрыш одного из игроков, b — выигрыш другого, то для игры с нулевой суммой $b = -a$, поэтому достаточно рассматривать, например a.

Выбор и осуществление одного из предусмотренных правилами действий называется *ходом* игрока. Ходы могут быть личными и случайными. *Личный ход* — это сознательный выбор игроком одного из возможных действий (например, ход в шахматной игре). *Случайный ход* — это случайно выбранное действие (например, выбор карты из перетасованной колоды). В дальнейшем мы будем рассматривать только личные ходы игроков.

Стратегией игрока называется совокупность правил, определяющих выбор его действия при каждом личном ходе в зависимости от сложившейся ситуации. Обычно в процессе игры при каждом личном ходе игрок делает выбор в зависимости от конкретной ситуации. Однако в принципе возможно, что все решения приняты игроком заранее (в ответ на любую сложившуюся ситуацию). Это означает, что игрок выбрал определенную стратегию, которая может быть задана в виде списка правил или программы. (Так можно осуществить игру с помощью ЭВМ). Игра называется *конечной*, если у каждого игрока имеется конечное число стратегий, и *бесконечной* — в противном случае.

Для того чтобы *решить* игру, или найти р е ш е н и е и г р ы, следует для каждого игрока выбрать стратегию, которая удовлетворяет условию *оптимальности*, т.е. один из игроков должен получать *максимальный выигрыш*, когда второй придерживается своей стратегии. В то же время второй игрок должен иметь *минимальный проигрыш*, если первый придерживается своей стратегии. Такие *стратегии* называются *оптимальными*. Оптимальные стратегии должны также удовлетворять условию *устойчивости*, т.е. любому из игроков должно быть невыгодно отказаться от своей стратегии в этой игре.

Если игра повторяется достаточно много раз, то игроков может интересовать не выигрыш и проигрыш в каждой конкретной партии, а *средний выигрыш* (*проигрыш*) во всех партиях.

***Целью** теории игр является определение оптимальной стратегии для каждого игрока.* При выборе оптимальной стратегии естественно предполагать, что оба игрока ведут себя разумно с точки зрения своих интересов. Важнейшее ограничение теории игр — единственность выигрыша как показателя эффективности, в то время как в большинстве реальных экономических задач имеется более одного показателя эффективности. Кроме того, в экономике, как правило, возникают задачи, в которых интересы партнеров не обязательно антагонистические. Развитие аппарата теории игр для решения задач со многими участниками, имеющими непротиворечивые интересы, выходит за рамки настоящего пособия.

9.2. Платежная матрица. Нижняя и верхняя цена игры

Рассмотрим парную конечную игру. Пусть игрок A располагает m личными стратегиями, которые обозначим $A_1, A_2, ..., A_m$. Пусть у игрока B имеется n личных стратегий, обозначим их $B_1, B_2, ..., B_n$. Говорят, что игра имеет размерность $m \times n$. В результате выбора игроками любой пары стратегий

$$A_i \text{ и } B_j \; (i = 1, 2, ..., m; \; j = 1, 2, ..., n)$$

однозначно определяется исход игры, т.е. выигрыш a_{ij} игрока A (положительный или отрицательный) и проигрыш $(-a_{ij})$ игрока B. Предположим, что значения a_{ij} известны для любой пары стратегий (A_i, B_j). Матрица $P = (a_{ij})$, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$, элементами которой являются выигрыши, соответствующие стратегиям A_i и B_j, называется *платежной матрицей* или *матрицей игры*. Общий вид такой матрицы представлен в табл. 9.1.

Строки этой таблицы соответствуют стратегиям игрока A, а столбцы — стратегиям игрока B.

Таблица 9.1

B_j / A_i	B_1	B_2	...	B_n
A_1	a_{11}	a_{12}	...	a_{1n}
A_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mn}

Составим платежную матрицу для следующей игры.

▷ **9.1.** Игра "поиск".

Игрок A может спрятаться в одном из двух убежищ (I и II); игрок B ищет игрока A, и если найдет, то получает штраф 1 ден. ед. от A, в противном случае платит игроку A 1 ден. ед. Необходимо построить платежную матрицу игры.

Решение. Для составления платежной матрицы следует проанализировать поведение каждого из игроков. Игрок A может спрятаться в убежище I — обозначим эту стратегию через A_1 или в убежище II — стратегия A_2.

Игрок B может искать первого игрока в убежище I — стратегия B_1, либо в убежище II — стратегия B_2. Если игрок A находится в убежище I и там его обнаруживает игрок B, т.е. осуществляется пара стратегий (A_1, B_1), то игрок A платит штраф, т.е. $a_{11} = -1$. Аналогично получаем $a_{22} = -1$ (A_2, B_2). Очевидно, что стратегии (A_1, B_2) и (A_2, B_1) дают игроку A выигрыш 1, поэтому $a_{12} = a_{21} = 1$. Таким образом, для игры "поиск" размера 2×2 получаем платежную матрицу

$$P = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \blacktriangleright$$

Рассмотрим игру $m \times n$ с матрицей $P = (a_{ij})$, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$ и определим наилучшую среди стратегий $A_1, A_2, ..., A_m$. Выбирая стратегию A_i, игрок A должен рассчитывать, что игрок B ответит на нее той из стратегий B_j, для которой выигрыш для игрока A минимален (игрок B стремится "навредить" игроку A).

Обозначим через α_i наименьший выигрыш игрока A при выборе им стратегии A_i для всех возможных стратегий игрока B (наименьшее число в i-й строке платежной матрицы), т.е.

$$\min_{j=1,...,n} a_{ij} = \alpha_i. \qquad \textbf{(9.1)}$$

Среди всех чисел α_i ($i = 1, 2, ..., m$) выберем наибольшее: $\alpha = \max\limits_{i=1,2,...,m} \alpha_i$. Назовем α ***нижней ценой игры***, или ***максимальным выигрышем (максимином)***. Это *гарантированный выигрыш игрока А при любой стратегии игрока В*. Следовательно,

$$\alpha = \max_{i=1,...,m} \min_{j=1,...,n} a_{ij}. \tag{9.2}$$

Стратегия, соответствующая максимину, называется *максиминной стратегией*. Игрок B заинтересован в том, чтобы уменьшить выигрыш игрока A; выбирая стратегию B_j, он учитывает максимально возможный при этом выигрыш для A. Обозначим

$$\beta_j = \max_{i=1,...,m} a_{ij}. \tag{9.3}$$

Среди всех чисел β_j выберем наименьшее $\beta = \min\limits_{j=1,2,...,n} \beta_j$ и назовем β ***верхней ценой игры*** или ***минимаксным выигрышем (минимаксом)***. Это *гарантированный проигрыш игрока В*. Следовательно,

$$\beta = \min_{j=1,...,n} \max_{i=1,...,m} a_{ij}. \tag{9.4}$$

Стратегия, соответствующая минимаксу, называется *минимаксной стратегией*.

Принцип, диктующий игрокам выбор наиболее "осторожных" минимаксной и максиминной стратегий, называется *принципом минимакса*. Этот принцип следует из разумного предположения, что каждый игрок стремится достичь цели, противоположной цели противника. Определим нижнюю и верхнюю цены игры и соответствующие стратегии в задаче **9.1**. Рассмотрим платежную матрицу

$$P = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

из задачи **9.1**. При выборе стратегии A_1 (первая строка матрицы) минимальный выигрыш равен $\alpha_1 = \min(-1;1) = -1$ и соответствует стратегии β_1 игрока B. При выборе стратегии A_2 (вторая строка матрицы) минимальный выигрыш равен $\alpha_2 = \min(1;-1) = -1$, он достигается при стратегии B_2.

Гарантируя себе максимальный выигрыш при любой стратегии игрока B, т.е. нижнюю цену игры $\alpha = \max(\alpha_1, \alpha_2) = \max(-1;-1) = -1$, игрок A может выбирать любую стратегию: A_1 или A_2, т.е. любая его стратегия является максиминной.

Выбирая стратегию B_1 (столбец 1), игрок B понимает, что игрок A ответит стратегией A_2, чтобы максимизировать свой выигрыш (проигрыш B). Следовательно, максимальный проигрыш игрока B при выборе им стратегии B_1 равен $\beta_1 = \max(-1;\ 1) = 1$.

Аналогично максимальный проигрыш игрока B (выигрыш A) при выборе им стратегии B_2 (столбец 2) равен $\beta_2 = \max(1;\ -1) = 1$.

Таким образом, при любой стратегии игрока A гарантированный минимальный проигрыш игрока B равен $\beta = \min(\beta_1, \beta_2) = \min(1; 1) = 1$ — верхней цене игры.

Любая стратегия игрока B является минимаксной. Дополнив табл. 9.1 строкой β_j и столбцом α_i, получим табл. 9.2. На пересечении дополнительных строки и столбца будем записывать верхнюю и нижнюю цены игр.

Таблица 9.2

A_j \ B_i	B_1	B_2	α_i
A_1	-1	1	-1
A_2	1	-1	-1
β_j	1	1	$\alpha=-1$ / $\beta=1$

В задаче **9.1**, рассмотренной выше, верхняя и нижняя цены игры различны: $\alpha \neq \beta$.

Если верхняя и нижняя цены игры совпадают, то общее значение верхней и нижней цены игры $\alpha = \beta = \nu$ называется *чистой ценой игры*, или *ценой игры*. Минимаксные стратегии, соответствующие цене игры, являются *оптимальными стратегиями*, а их совокупность — *оптимальным решением,* или *решением* игры. В этом случае игрок A получает максимальный гарантированный (не зависящий от поведения игрока B) выигрыш ν, а игрок B добивается минимального гарантированного (вне зависимости от поведения игрока A) проигрыша ν. Говорят, что решение игры обладает *устойчивостью,* т.е. если один из игроков придерживается своей оптимальной стратегии, то для другого не может быть выгодным отклоняться от своей оптимальной стратегии.

Пара чистых стратегий A_i и B_j дает оптимальное решение игры тогда и только тогда, когда соответствующий ей элемент a_{ij} является одновременно наибольшим в своем столбце и наименьшим в своей строке. Такая ситуация, если она существует, называется *седловой точкой* (по аналогии с поверхностью седла, которая искривляется вверх в одном направлении и вниз — в другом).

Обозначим A^* и B^* — пару чистых стратегий, на которых достигается решение игры в задаче с седловой точкой. Введем функцию выигрыша первого игрока на каждой паре стратегий: $P(A_i, B_j) = a_{ij}$. Тогда из условия оптимальности в седловой точке выполняется двойное неравенство: $P(A_i, B^*) \le P(A^*, B^*) \le P(A^*, B_j)$, которое справедливо для всех $i = 1, ..., m$; $j = 1, ..., n$. Действительно, выбор стратегии A^* первым игроком при оптимальной стратегии B^* второго игрока максимизирует минимальный возможный выигрыш: $P(A^*, B^*) \ge P(A_i, B^*)$, а выбор стратегии B^* вторым игроком при оптимальной стратегии первого минимизирует максимальный проигрыш: $P(A^*, B^*) \le P(A^*, B)$.

▷ **9.2.** Определить нижнюю и верхнюю цену игры, заданной платежной матрицей

$$P = \begin{pmatrix} 0{,}5 & 0{,}6 & 0{,}8 \\ 0{,}9 & 0{,}7 & 0{,}8 \\ 0{,}7 & 0{,}6 & 0{,}6 \end{pmatrix}.$$

Имеет ли игра седловую точку?

Таблица 9.3

A_j \ B_j	B_1	B_2	B_3	α_i
A_1	0,5	0,6	0,8	0,5
A_2	0,9	0,7	0,8	$\boxed{0{,}7}$
A_3	0,7	0,6	0,6	0,6
β_j	0,9	$\boxed{0{,}7}$	0,8	$\alpha = \beta = 0{,}7$

Решение. Все расчеты удобно проводить в таблице, к которой, кроме матрицы P, введены столбец α_i и строка β_j (табл. 9.3). Анализируя строки матрицы (стратегии игрока A), заполняем столбец α_i: $\alpha_1 = 0{,}5$, $\alpha_2 = 0{,}7$, $\alpha_3 = 0{,}6$ — минимальные числа в строках 1, 2, 3. Аналогично $\beta_1 = 0{,}9$, $\beta_2 = 0{,}7$, $\beta_3 = 0{,}8$ — максимальные числа в столбцах 1, 2, 3 соответственно. Нижняя цена игры $\alpha = \max\limits_{i=1,2,3} \alpha_i = \max(0{,}5;\ 0{,}7;\ 0{,}6) = 0{,}7$ (наибольшее число в столбце α_i) и верхняя цена игры $\beta = \min\limits_{j=1,2,3} \beta_j = \min(0{,}9;\ 0{,}7;\ 0{,}8) = 0{,}7$ (наименьшее число в строке β_j). Эти значения равны, т.е. $\alpha = \beta$, и достигаются на одной и той же паре стратегий $(A_2,\ B_2)$. Следовательно, игра имеет седловую точку $(A_2,\ B_2)$ и цена игры $\nu = 0{,}7$. ▶

9.3. Решение игр в смешанных стратегиях

Если игра не имеет седловой точки, то применение чистых стратегий не дает оптимального решения игры. Так, в задаче **9.1** $\alpha \neq \beta$, седловая точка отсутствует. В таком случае можно получить оптимальное решение, случайным образом чередуя чистые стратегии.

Смешанной стратегией S_A игрока A называется применение чистых стратегий $A_1, A_2, \ldots, A_i, \ldots, A_m$ с вероятностями $p_1, p_2, \ldots, p_i, \ldots, p_m$, причем сумма вероятностей равна 1: $\sum\limits_{i=1}^{m} p_i = 1$. Смешанные стратегии игрока A записываются в виде матрицы

$$S_A = \begin{pmatrix} A_1\, A_2 \ldots A_i \ldots A_m \\ p_1\, p_2 \cdots p_i \cdots p_m \end{pmatrix},$$

или в виде строки $S_A = (p_1, p_2, \ldots p_i, \ldots, p_m)$. Аналогично смешанные стратегии игрока B обозначаются:

$$S_B = \begin{pmatrix} B_1\, B_2 \ldots B_j \ldots B_n \\ p_1\, p_2 \cdots p_j \cdots p_n \end{pmatrix}, \text{ или } S_B = (q_1, q_2, \ldots q_j, \ldots, q_n),$$

где сумма вероятностей появления стратегий равна 1: $\sum\limits_{j=1}^{n} q_j = 1$.

Чистые стратегии можно считать частным случаем смешанных и задавать строкой, в которой 1 соответствует чистой стратегии. На основании принципа минимакса определяется *оптимальное* *решение* (или *решение*) игры: это пара оптимальных стратегий S_A^*, S_B^*, в общем случае смешанных, обладающих следующим свойством: если один из игроков придерживается своей оптимальной стратегии, то другому не может быть выгодно отступать от своей. Выигрыш, соответствующий оптимальному решению, называется ***ценой игры*** ν. Цена игры удовлетворяет неравенству:

$$\alpha \le \nu \le \beta, \tag{9.5}$$

где α и β — нижняя и верхняя цены игры.

Справедлива следующая основная теорема теории игр — **теорема Неймана**[1]. *Каждая конечная игра имеет по крайней мере одно оптимальное решение, возможно, среди смешанных стратегий.*

Пусть $S_A^* = (p_1^*, p_2^*, \ldots, p_m^*)$ и $S_B^* = (q_1^*, q_2^*, \ldots, q_n^*)$ — пара оптимальных стратегий. Если чистая стратегия входит в оптимальную смешанную стратегию с отличной от нуля вероятностью, то она называется *активной.*

Справедлива **теорема об активных стратегиях:** *если один из игроков придерживается своей оптимальной смешанной стратегии, то выигрыш остается неизменным и равным цене игры* ν, *если второй игрок не выходит за пределы своих активных стратегий.*

Эта теорема имеет большое практическое значение — она дает конкретные модели нахождения оптимальных стратегий при отсутствии седловой точки.

Рассмотрим **игру размера 2 × 2**, которая является простейшим случаем конечной игры. Если такая игра имеет седловую точку, то оптимальное решение — это пара чистых стратегий, соответствующих этой точке.

Игра, в которой отсутствует седловая точка, в соответствии с основной теоремой теории игр *оптимальное решение существует и определяется парой смешанных стратегий* $S_A^* = (p_1^*, p_2^*)$ *и* $S_B^* = (q_1^*, q_2^*)$.

Для того чтобы их найти, воспользуемся теоремой об активных стратегиях. Если игрок *A* придерживается своей оптимальной стратегии S_A^*, то его средний выигрыш будет равен цене игры ν, какой бы активной стратегией ни пользовался игрок *B*. Для игры

[1] Джон фон Нейман (1903—1957) — американский математик.

2×2 любая чистая стратегия противника является активной, если отсутствует седловая точка. Выигрыш игрока A (проигрыш игрока B) — случайная величина, математическое ожидание (среднее значение) которой является ценой игры. Поэтому средний выигрыш игрока A (оптимальная стратегия) будет равен ν и для 1-й, и для 2-й стратегии противника.

Пусть игра задана платежной матрицей

$$P = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Средний выигрыш игрока A, если он использует оптимальную смешанную стратегию $S_A^* = \begin{pmatrix} A_1 & A_2 \\ p_1^* & p_2^* \end{pmatrix}$, а игрок B — чистую стратегию B_1 (это соответствует 1-му столбцу платежной матрицы P), равен цене игры ν:

$$a_{11}p_1^* + a_{21}p_2^* = \nu.$$

Тот же средний выигрыш получает игрок A, если 2-й игрок применяет стратегию B_2, т.е. $a_{12}p_1^* + a_{22}p_2^* = \nu$. Учитывая, что $p_1^* + p_2^* = 1$, получаем систему уравнений для определения оптимальной стратегии S_A^* и цены игры ν:

$$\begin{cases} a_{11}p_1^* + a_{21}p_2^* = \nu, \\ a_{12}p_1^* + a_{22}p_2^* = \nu, \\ p_1^* + p_2^* = 1. \end{cases} \qquad (9.6)$$

Решая эту систему, получим оптимальную стратегию

$$p_1^* = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}, \quad p_2^* = \frac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}} \qquad (9.7)$$

и цену игры

$$\nu = \frac{a_{22}a_{11} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}. \qquad (9.8)$$

Применяя теорему об активных стратегиях при отыскании S_B^* — оптимальной стратегии игрока B, получаем, что при любой чистой стратегии игрока A (A_1 или A_2) средний проигрыш игрока B равен цене игры ν, т.е.

$$\begin{cases} a_{11}q_1^* + a_{12}q_2^* = \nu, \\ a_{21}q_1^* + a_{22}q_2^* = \nu, \\ q_1^* + q_2^* = 1. \end{cases} \qquad \textbf{(9.9)}$$

Тогда оптимальная стратегия $S_B^*(q_1^*, q_2^*)$ определяется формулами:

$$q_1^* = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}, \quad q_2^* = \frac{a_{11} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}. \qquad \textbf{(9.10)}$$

Применим полученные результаты для отыскания оптимальных стратегий для игры, рассмотренной в задаче **9.1.**

▷ **9.3.** Найти оптимальные стратегии игры, приведенной в задаче **9.1.**

Решение. Игра “поиск” задана платежной матрицей без седловой точки:

$$P = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}, \quad \alpha = -1, \quad \beta = 1.$$

Поэтому ищем решение в смешанных стратегиях; для игрока A средний выигрыш равен цене игры ν (при B_1 и B_2); для игрока B средний проигрыш равен цене игры ν (при A_1 и A_2). Системы уравнений (9.6) и (9.9) в данном случае имеют вид:

$$\begin{cases} (-1)p_1^* + 1 \cdot p_2^* = \nu, \\ 1 \cdot p_1^* - 1 \cdot p_2^* = \nu, \\ p_1^* + p_2^* = 1, \end{cases} \qquad \begin{cases} (-1)q_1^* + 1 \cdot q_2^* = \nu, \\ 1 \cdot q_1^* - 1 \cdot q_2^* = \nu, \\ q_1^* + q_2^* = 1, \end{cases}$$

Решая эти системы, получаем $p_1^* = p_2^* = q_1^* = q_2^* = \frac{1}{2}, \quad \nu = 0.$

Это означает, что оптимальная стратегия каждого игрока состоит в том, чтобы чередовать свои чистые стратегии случайным образом, выбирая каждое из убежищ с вероятностью 1/2, при этом средний выигрыш равен 0.▶

9.4. Геометрическая интерпретация игры 2x2

Решение игры 2 × 2 допускает наглядную геометрическую интерпретацию. Пусть игра задана платежной матрицей $P = (a_{ij})$, $i, j = 1, 2$. По оси абсцисс (рис. 9.1) отложим единичный отрезок A_1A_2; точка $A_1(x=0)$ изображает стратегию A_1, а все промежуточные точки этого отрезка — смешанные стратегии S_A первого игрока, причем расстояние от S_A до правого конца отрезка — это вероятность p_1 стратегии A_1, расстояние до левого конца — вероятность p_2 стратегии A_2. На перпендикулярных осях I—I и II—II откладываем выигрыши при стратегиях A_1 и A_2 соответственно. Если 2-й игрок примет стратегию B_1, то она дает выигрыши a_{11} и a_{21} на осях I—I и II—II, соответствующие стратегиям A_1 и A_2. Обозначим эти точки на осях I—I и II—II буквой B_1. Средний выигрыш ν_1, соответствующий смешанной стратегии S_A, определяется по формуле математического ожидания $\nu_1 = a_{11}p_1 + a_{21}p_2$ и равен ординате точки M_1, которая лежит на отрезке B_1B_1 и имеет абсциссу S_A (рис. 9.1).

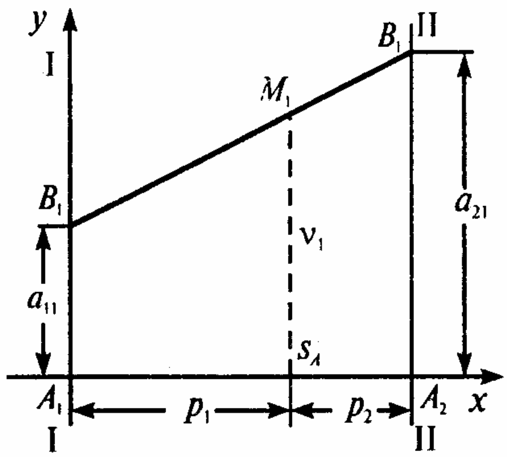


Рис. 9.1

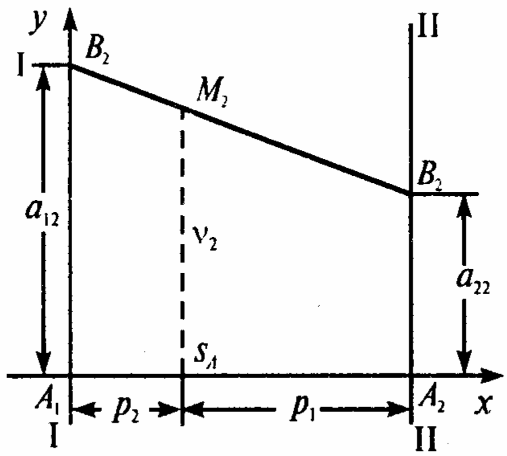


Рис. 9.2

Аналогично строим отрезок B_2B_2, соответствующий применению вторым игроком стратегии B_2 (рис. 9.2). При этом средний выигрыш $\nu_2 = a_{12}p_1 + a_{22}p_2$ — ордината точки M_2.

В соответствии с принципом минимакса оптимальная стратегия S_A^* такова, что минимальный выигрыш игрока A (при наихудшем поведении игрока B) обращается в максимум. Ординаты точек, лежащих на ломаной (рис. 9.3), показывают минимальный выигрыш игрока A при использовании им любой смешанной стратегии (на участке B_1N — против стратегии B_1, на участке NB_2 — против стратегии B_2). Оптимальную стратегию $S_A^* = (p_1^*, p_2^*)$ определяет точка N, в которой минимальный выигрыш достигает максимума; ее ордината равна цене игры ν. На рис. 9.3 обозначены также верхняя и нижняя цены игры α и β.

Применим геометрический метод для решения следующей задачи.

▷ **9.4.** Решить графически игру, заданную платежной матрицей

$$P = \begin{pmatrix} 1{,}5 & 3 \\ 2 & 1 \end{pmatrix}.$$

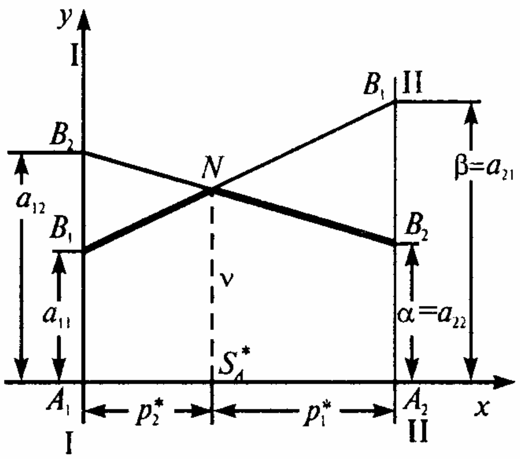


Рис. 9.3

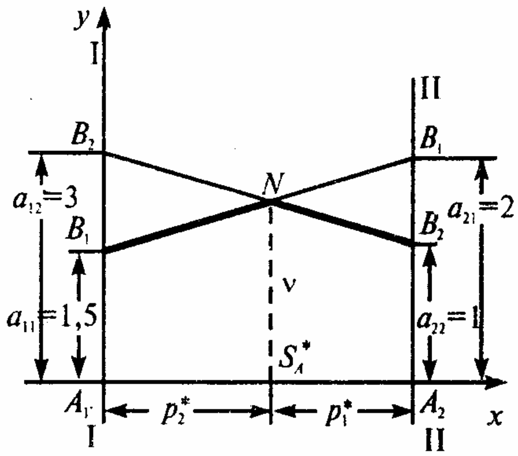


Рис. 9.4

Решение. Откладываем по оси абсцисс (рис. 9.4) единичный отрезок A_1A_2. На вертикальной оси I—I откладываем отрезки: $a_{11} = 1{,}5$, соответствующий стратегии B_1, и $a_{12} = 3$, соответствующий стратегии B_2. На вертикальной оси II—II отрезок $a_{21} = 2$ соответствует стратегии B_1, отрезок $a_{22} = 1$ соответствует стратегии B_2 (*см.* рис. 9.4). Нижняя цена игры $\alpha = a_{11} = 1{,}5$. Верхняя цена игры $\beta = a_{21} = 2$, седловая точка отсутствует. Из рис. 9.4 видно, что абсцисса точки N определяет оптимальную стратегию S_A^*, а ордината — цену игры ν. Точка N является точкой пересечения прямых B_1B_1 и B_2B_2. Уравнение прямой B_1B_1, проходящей через точки (0; 1,5) и (1; 2):

$$\frac{x-0}{1-0}=\frac{y-1{,}5}{2-1{,}5} \text{ или } y=0{,}5x+1{,}5.$$

Уравнение прямой B_2B_2, проходящей через точки (0; 3) и (1;1):

$$\frac{x-0}{1-0}=\frac{y-3}{1-3} \text{ или } y=2x+3.$$

Точка пересечения прямых является решением системы:

$$\begin{cases} y=0{,}5x+1{,}5, \\ y=-2x+3, \end{cases}$$

или $x=0{,}6;\ y=1{,}8$, т.е. $N\,(0{,}6;\ 1{,}8)$.

Таким образом, $p_1^*=0{,}6$, $p_2^*=1-0{,}6=0{,}4$; оптимальная стратегия $S_A^*=(0{,}6;0{,}4)$, цена игры $\nu=1{,}8$.

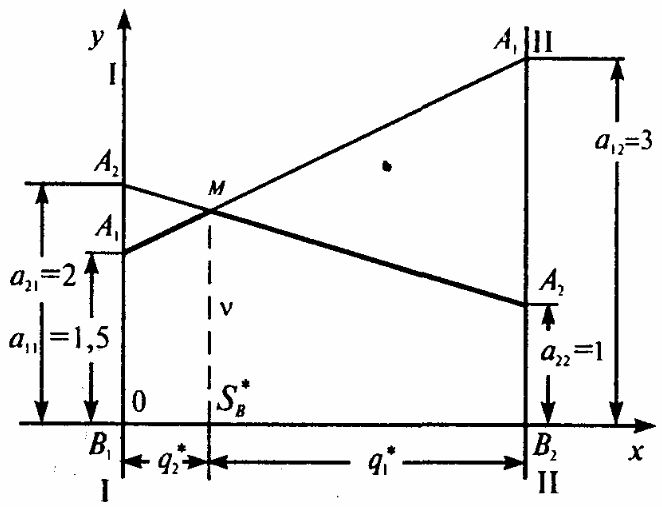


Рис. 9.5

Геометрически можно также определить оптимальную стратегию игрока B, если поменять местами игроков A и B и вместо максимума нижней границы A_2MA_1 в соответствии с принципом минимакса (рис. 9.5) рассмотреть минимум верхней границы.

Абсцисса точки M определяет q_2^* в оптимальной стратегии игрока B, ордината этой точки — цена игры. Прямая A_1A_1, проходящая через точки (0; 1,5) и (1; 3), удовлетворяет уравнению

$$y=1{,}5x+1{,}5.$$

Прямая A_2A_2, проходящая через точки (0; 2) и (1; 1), удовлетворяет уравнению $y=-x+2$.

Координаты их точки пересечения M — это решение системы уравнений:

$$\begin{cases} y = 1{,}5x + 1{,}5, \\ y = -x + 2, \end{cases}$$

откуда $x = 0{,}2$; $y = 1{,}8$, т.е. $q_2^* = 0{,}2$, $q_1^* = 1 - q_2^* = 0{,}8$, $x = y = 1{,}8$, $S_B^* = (0{,}8; 0{,}2)$.

Оптимальное решение игры найдено.▶

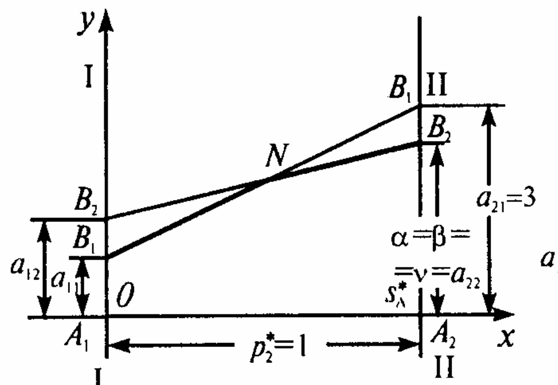


Рис. 9.6

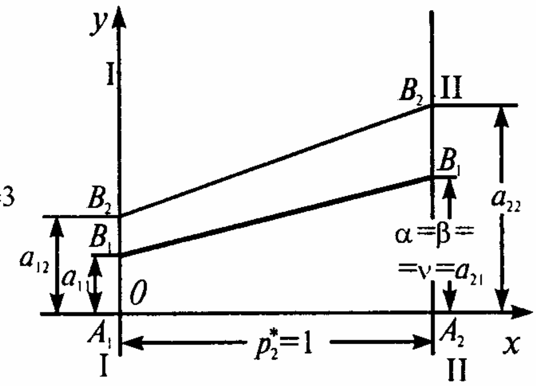


Рис. 9.7

Из решения задачи **9.4** следует, что геометрически можно определять оптимальную стратегию как игрока A, так и игрока B; в обоих случаях используется принцип минимакса, но во втором случае строится не нижняя, а верхняя граница выигрыша и на ней определяется не максимум, а минимум. Если платежная матрица содержит отрицательные числа, то для графического решения задачи лучше перейти к новой матрице с неотрицательными элементами; для этого к элементам исходной матрицы достаточно добавить соответствующее положительное число. Решение игры при этом не изменится, а цена игры увеличится на это число. В задаче **9.4** платежная матрица не имела седловой точки ($\alpha \neq \beta$).

При наличии седловой точки графическое решение дают варианты, изображенные на рис. 9.6 и 9.7. На рис. 9.6 наибольшей ординатой на ломаной B_1NB_2 обладает точка B_2, поэтому оптимальной является чистая стратегия A_2 для игрока A (B_2 — для игрока B), т.е. оптимальное решение: $S_A^* = (0; 1)$, $S_B^* = (0; 1)$. Игра имеет седловую точку $a_{22} = \nu$.

Чистая стратегия B_2 (рис. 9.7) не выгодна для игрока B, поскольку при любой стратегии игрока A она дает последнему больший выигрыш, чем чистая стратегия B_1. На основании принципа минимакса выделим прямую B_1B_1 и на ней точку B_1 с наибольшей ординатой на оси I—I. Чистая стратегия A_2 является оптимальной для игрока A, а чистая стратегия B_1 — для игрока B. Оптимальное решение: $S_A^* = (0;1)$, $S_B^* = (1;0)$, цена игры $\nu = a_{21} = \alpha = \beta$, т.е. имеется седловая точка.

Графический метод можно применять при решении игры $2 \times n$ и $m \times 2$.

9.5. Приведение матричной игры к задаче линейного программирования

Игра $m \times n$ в общем случае не имеет наглядной геометрической интерпретации. Ее решение достаточно трудоемко при больших m и n, однако принципиальных трудностей не имеет, поскольку может быть сведено к решению задачи линейного программирования. Покажем это. Пусть игра $m \times n$ задана платежной матрицей $p = (a_{ij})$, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$. Игрок A обладает стратегиями $A_1, A_2, ..., A_m$, игрок B — стратегиями $B_1, B_2, ..., B_n$. Необходимо определить оптимальные стратегии $S_A^* = (p_1^*, p_2^*, \dots, p_m^*)$ и $S_B^* = (q_1^*, q_2^*, \dots, q_n^*)$, где p_i^*, q_j^* — вероятности применения соответствующих чистых стратегий A_i, B_j;

$$p_1^* + p_2^* + \ldots + p_m^* = 1, \quad q_1^* + q_2^* + \ldots + q_n^* = 1.$$

Оптимальная стратегия S_A^* удовлетворяет следующему требованию. Она обеспечивает игроку A средний выигрыш, не меньший, чем цена игры ν, при любой стратегии игрока B и выигрыш, равный цене игры ν, при оптимальной стратегии игрока B. Без ограничения общности полагаем $\nu > 0$; этого можно добиться, сделав все элементы $a_{ij} \geq 0$. Если игрок A применяет смешанную стратегию $S_A^* = (p_1^*, p_2^*, \dots, p_m^*)$ против любой чистой стратегии B_j игрока B, то он получает *средний выигрыш*, или *математическое ожидание выигрыша* $a_j = a_{1j}p_1 + a_{2j}p_2 + \ldots + a_{mj}p_m$, $j = 1, 2, ..., n$ (т.е. элементы j-го столбца платежной матрицы почленно умножаются на соответствующие вероятности стратегий A_1, A_2, ..., A_m и результаты складываются).

Для оптимальной стратегии S_A^* все средние выигрыши не меньше цены игры ν, поэтому получаем систему неравенств:

$$\begin{cases} a_{11}p_1 + a_{21}p_2 + \ldots + a_{m1}p_m \geq \nu, \\ a_{12}p_1 + a_{22}p_2 + \ldots + a_{m2}p_m \geq \nu, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}p_1 + a_{2n}p_2 + \ldots + a_{mn}p_m \geq \nu. \end{cases} \quad \textbf{(9.11)}$$

Каждое из неравенств можно разделить на число $\nu > 0$. Введем новые переменные:

$$x_1 = p_1/\nu,\ x_2 = p_2/\nu, \ldots, x_m = p_m/\nu. \quad \textbf{(9.12)}$$

Тогда система (9.11) примет вид:

$$\begin{cases} a_{11}x_1 + a_{21}x_2 + \ldots + a_{m1}x_m \geq 1, \\ a_{12}x_1 + a_{22}x_2 + \ldots + a_{m2}x_m \geq 1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}x_1 + a_{2n}x_2 + \ldots + a_{mn}x_m \geq 1. \end{cases} \quad \textbf{(9.13)}$$

Цель игрока A — максимизировать свой гарантированный выигрыш, т.е. цену игры ν.

Разделив на $\nu \neq 0$ равенство $p_1 + p_2 + \ldots + p_m = 1$, получаем, что переменные x_i (i=1,2,...,m) удовлетворяют условию: $x_1 + x_2 + {} + \ldots + x_m = 1/\nu$. Максимизация цены игры ν эквивалентна минимизации величины $1/\nu$, поэтому задача может быть сформулирована следующим образом: *определить значения переменных* $x_i \geq 0$, $i = 1, 2, \ldots, m$, *так, чтобы они удовлетворяли линейным ограничениям* (9.13) *и при этом линейная функция*

$$Z = x_1 + x_2 + \ldots + x_m \quad \textbf{(9.14)}$$

обращалась в минимум. **Это задача линейного программирования. Решая задачу (9.13)—(9.14), получаем оптимальное решение** $p_1^*, p_2^*, \ldots, p_m^*$ **и оптимальную стратегию** S_A^*.

Для определения оптимальной стратегии $S_B^* = (q_1^*, q_2^*, \dots, q_n^*)$ следует учесть, что игрок B стремится минимизировать гарантированный выигрыш, т.е. найти $\max \frac{1}{\nu}$. Переменные $q_1, q_2, \dots, q_n$ удовлетворяют неравенствам

$$\begin{cases} a_{11}q_1 + a_{12}q_2 + \dots + a_{1n}q_n \le \nu, \\ a_{21}q_1 + a_{22}q_2 + \dots + a_{2n}q_n \le \nu, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}q_1 + a_{m2}q_2 + \dots + a_{mn}q_n \le \nu. \end{cases} \qquad \mathbf{(9.15)}$$

которые следуют из того, что средний проигрыш игрока B не превосходит цены игры, какую бы чистую стратегию не применял игрок A.

Если обозначить

$$y_j = q_j / \nu, \quad j = 1, 2, \dots, n, \qquad \mathbf{(9.16)}$$

то получим систему неравенств:

$$\begin{cases} a_{11}y_1 + a_{12}y_2 + \dots + a_{1n}y_n \le 1, \\ a_{21}y_1 + a_{22}y_2 + \dots + a_{2n}y_n \le 1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}y_1 + a_{m2}y_2 + \dots + a_{mn}y_n \le 1. \end{cases} \qquad \mathbf{(9.17)}$$

Переменные y_j ($j = 1, 2, \dots, n$) удовлетворяют условию $y_1 + y_2 + \dots + y_n = 1/\nu$.

Игра свелась к следующей задаче.

Определить значения переменных $y_j \ge 0$, $j = 1, 2, \dots, n$, *которые удовлетворяют системе неравенств* (9.17) *и максимизируют линейную функцию*

$$Z' = y_1 + y_2 + \dots y_n, \qquad \mathbf{(9.18)}$$

Решение задачи линейного программирования (9.16), (9.17) определяет оптимальную стратегию $S_B^* = (q_1^*, q_2^*, \dots, q_n^*)$. При этом цена игры

$$\nu = 1/\max Z' = 1/\min Z. \qquad \mathbf{(9.19)}$$

Составив расширенные матрицы для задач (9.13), (9.14) и (9.17), (9.18), убеждаемся, что одна матрица получилась из другой транспонированием:

$$\begin{array}{c} \geq \\ \left(\begin{array}{c|c} a_{11}\ a_{21} \ldots a_{m1} & 1 \\ a_{12}\ a_{22} \ldots a_{m2} & 1 \\ \ldots\ \ldots\ \ldots\ \ldots & \ldots \\ a_{1n}\ a_{2n} \ldots a_{mn} & 1 \\ \hline 1 \quad 1 \ \ldots \ 1 & \min Z \end{array}\right) \end{array} \leftrightarrow \begin{array}{c} \leq \\ \left(\begin{array}{c|c} a_{11}\ a_{12} \ldots a_{1n} & 1 \\ a_{21}\ a_{22} \ldots a_{2n} & 1 \\ \ldots\ \ldots\ \ldots\ \ldots & \ldots \\ a_{m1}\ a_{m2} \ldots a_{mn} & 1 \\ \hline 1 \quad 1 \ \ldots \ 1 & \max Z' \end{array}\right) \end{array}$$

Таким образом, задачи линейного программирования (9.13), (9.14) и (9.17), (9.18) являются взаимно-двойственными. Очевидно, при определении оптимальных стратегий в конкретных задачах следует выбрать ту из взаимно-двойственных задач, решение которой менее трудоемко, а решение другой задачи найти с помощью теорем двойственности.

Приведем примеры экономических задач, которые описываются игровыми моделями $m \times n$ и могут быть решены методами линейного программирования.

▷ **9.5.** Предприятие может выпускать три вида продукции (A_1, A_2 и A_3), получая при этом прибыль, зависящую от спроса, который может быть в одном из четырех состояний (B_1, B_2, B_3, B_4). Дана матрица (табл. 9.4), ее элементы a_{ij} характеризуют прибыль, которую получит предприятие при выпуске i-й продукции с j-м состоянием спроса.

Определить оптимальные пропорции в выпускаемой продукции, гарантирующие среднюю величину прибыли при любом состоянии спроса, считая его неопределенным.

Решение. Задача сводится к игровой модели, в которой игра предприятия A против спроса B задана платежной матрицей (*см.* табл. 9.4). Прежде чем решать задачу, можно попытаться упростить игру, проведя анализ платежной матрицы и отбросив стратегии, заведомо невыгодные или дублирующие. Так, вторая стратегия (второй столбец матрицы (*см.* табл. 9.4))

Таблица 9.4

	B_1	B_2	B_3	B_4
A_1	3	3	6	8
A_2	9	10	4	2
A_3	7	7	5	4

является явно невыгодной для игрока B по сравнению с первой (элементы второго столбца больше элементов первого столбца), так как цель игрока B — уменьшить выигрыш игрока A. Поэтому второй столбец можно отбросить. Получим матрицу P размера 3×3:

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

Таблица 9.5

Вид продукции \ Спрос	B_1	B_3	B_4	α_i
A_1	3	6	8	3
A_2	9	4	2	2
A_3	7	5	4	4
β_i	9	6	8	6 \ 4

Определим нижнюю и верхнюю цены игры в табл. 9.5.

Так как $\alpha \neq \beta$, то седловая точка отсутствует и оптимальное решение следует искать в смешанных стратегиях игроков:

$$S_A^* = (p_1, p_2, p_3) \text{ и } S_B^* = (q_1, q_2, q_3).$$

Обозначив $x_i = p_i/\nu$, $i = 1, 2, 3$ и $y_j = q_j/\nu$, $j = 1, 2, 3$, составим две взаимно-двойственные задачи линейного программирования (*см.* (9.13)—(9.14) и (9.17)—(9.18)).

Задача 1

$$\begin{cases} 3x_1 + 9x_2 + 7x_3 \geq 1, \\ 6x_1 + 4x_2 + 5x_3 \geq 1, \\ 8x_1 + 2x_2 + 4x_3 \geq 1, \end{cases}$$

$$x_i \geq 0, \ i = 1,2,3$$

$$Z = x_1 + x_2 + x_3 \to \min.$$

Задача 2

$$\begin{cases} 3y_1 + 6y_2 + 8y_3 \leq 1, \\ 9y_1 + 4y_2 + 2y_3 \leq 1, \\ 7y_1 + 5y_2 + 4y_3 \leq 1, \end{cases}$$

$$y_j \geq 0, \ j = 1,2,3,$$

$$Z = y_1 + y_2 + y_3 \to \max.$$

Решаем симплексным методом одну из задач, например, задачу 2, поскольку для нее первое базисное решение будет допустимым. Введем добавочные переменные и перейдем к уравнениям:

$$\begin{cases} 3y_1 + 6y_2 + 8y_3 + y_4 = 1, \\ 9y_1 + 4y_2 + 2y_3 + y_5 = 1, \\ 7y_1 + 5y_2 + 4y_3 + y_6 = 1, \end{cases}$$

$$y_j \geq 0,\ j = 1,2,\ldots,6.$$

I шаг. Основные переменные — y_4, y_5, y_6; неосновные переменные — y_1, y_2, y_3.

$$\begin{cases} \boxed{y_4 = 1 - 3y_1 - 6y_2 - 8y_3,} \\ y_5 = 1 - 9y_1 - 4y_2 - 2y_3, \\ y_6 = 1 - 7y_1 - 5y_2 - 4y_3, \end{cases}$$

$$Z' = y_1 + y_2 + y_3.$$

Базисное решение $Y_1 = (0;\ 0;\ 0;\ 1;\ 1;\ 1)$ допустимое; переводим y_2 в основные; $y_2 = \min\{1/6;\ 1/4;\ 1/5\} = 1/6$; переводим y_4 в неосновные переменные.

II шаг. Основные переменные — y_2, y_5, y_6; неосновные перменные — y_1, y_3, y_4.

Получим после преобразований:

$$\begin{cases} y_2 = \frac{1}{6} - \frac{3}{6}y_1 - \frac{8}{6}y_3 - \frac{1}{6}y_4, \\ y_5 = \frac{1}{3} - 7y_1 + \frac{10}{3}y_3 + \frac{2}{3}y_4, \\ \boxed{y_6 = \frac{1}{6} - \frac{27}{6}y_1 - \frac{16}{6}y_3 + \frac{5}{6}y_4,} \end{cases}$$

$$Z' = \frac{1}{6} + \frac{1}{2}y_1 - \frac{2}{6}y_3 - \frac{1}{6}y_4.$$

Базисное решение: $Y_2 = \left(0; \frac{1}{6};\ 0;\ 0; \frac{1}{3}; \frac{1}{6}\right)$. Переводим y_1 в основные; $y_1 = \min\left\{\frac{1}{3}; \frac{1}{21}; \frac{1}{27}\right\} = \frac{1}{27}$. Переводим y_6 в неосновные переменные.

III шаг. Основные переменные — y_1, y_2, y_5; неосновные переменные — y_3, y_4, y_6;

$$\begin{cases} y_1 = \frac{1}{27} - \frac{16}{27}y_3 + \frac{5}{27}y_4 - \frac{6}{27}y_6, \\ y_2 = \frac{4}{27} - \frac{28}{27}y_3 - \frac{7}{27}y_4 + \frac{1}{9}y_6, \\ y_5 = \frac{2}{27} + \frac{202}{27}y_3 - \frac{17}{27}y_4 + \frac{14}{9}y_6, \end{cases}$$

$$Z' = \frac{5}{27} - \frac{17}{27}y_3 - \frac{2}{27}y_4 - \frac{1}{9}y_6.$$

Базисное решение $Y_3 = \left(\frac{1}{27}; \frac{4}{27};\ 0;\ 0; \frac{2}{27};\ 0\right)$.

Так как отсутствуют положительные коэффициенты при неосновных переменных, то критерий оптимальности выполнен, $\max Z' = \frac{5}{27}$ и базисное решение $Y_3 = (1/27;\ 4/27;\ 0;\ 0;\ 2/27;\ 0)$ является оптимальным.

Установим соответствие между переменными взаимно-двойственных задач и определим оптимальное базисное решение задачи 1 с помощью теорем двойственности:

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow \\ y_4 & y_5 & y_6 & y_1 & y_2 & y_3 \\ \frac{2}{27} & 0 & \frac{1}{9} & \frac{1}{2} & 0 & \frac{17}{27} \end{array}$$

Оптимальное базисное решение задачи 1: (2/27; 0; 1/9; 1/2; 0; 17/27), причем $\min Z = \max Z' = 5/27$. Из соотношений (9.19) находим цену игры $v = \frac{1}{\max Z'} = \frac{1}{\min Z} = \frac{27}{5} = 5{,}4$. Оптимальную стратегию $S_A^* = (p_1^*; p_2^*; p_3^*)$ находим, используя (9.12):

$$p_i^* = x_i v,\ i = 1, 2, 3, \text{ т.е.}$$

$$p_1^* = 5{,}4\frac{2}{27} = 0{,}4,\quad p_2^* = 5{,}4 \cdot 0 = 0,$$

$$p_3^* = 5{,}4\frac{1}{9} = 0{,}6;\quad S_A^* = (0{,}4; 0; 0{,}6).$$

Следовательно, предприятие должно выпустить 40% продукции A_1 и 60% продукции A_3, а продукцию A_2 не выпускать.

Оптимальная стратегия спроса S_B^* определяется аналогично: $q_j^* = \nu y_j$, $j = 1, 2, 3$, т.е. $S_B^* = (0{,}2; 0; 0{,}8; 0)$ (здесь учтено, что второй столбец исходной матрицы был отброшен как невыгодный). Таким образом, оптимальный спрос в 20% находится в состоянии B_1 и в 80% — в состоянии B_3.▶

При решении произвольной конечной игры размера $m \times n$ рекомендуется придерживаться следующей схемы:

1. Исключить из платежной матрицы заведомо невыгодные стратегии по сравнению с другими стратегиями. Такими стратегиями для игрока A (игрока B) являются те, которым соответствуют строки (столбцы) с элементами, заведомо меньшими (большими) по сравнению с элементами других строк (столбцов).

2. Определить верхнюю и нижнюю цены игры и проверить, имеет ли игра седловую точку. Если седловая точка есть, то соответствующие ей стратегии игроков будут оптимальными, а цена совпадает с верхней (нижней) ценой.

3. Если седловая точка отсутствует, то решение следует искать в смешанных стратегиях. Для игр размера $m \times n$ рекомендуется симплексный метод, для игр размера 2×2, $2 \times n$, $n \times 2$ возможно геометрическое решение.

На практике реализация оптимального решения $S^* = \begin{pmatrix} A_1 \dots A_m \\ p_1 \dots p_2 \end{pmatrix}$ в смешанных стратегиях может происходить несколькими путями. Первый состоит в физическом смешении чистых стратегий A_i в пропорциях, заданных вероятностями p_i.

Другой путь — при многократном повторении игры — в каждой партии чистые стратегии применяются в виде случайной последовательности, причем каждая из них — с частотой, равной ее вероятности в оптимальном решении.

Рассмотрим еще одну экономическую задачу, сводящуюся к игровой модели.

▷ **9.6. Предприятие выпускает скоропортящуюся продукцию, которую может сразу отправить потребителю (стратегия A_1), отправить** на склад для хранения (стратегия A_2) или подвергнуть дополнительной обработке (стратегия A_3) для длительного хранения.

Потребитель может приобрести продукцию: немедленно (стратегия B_1), в течение небольшого времени (B_2), после длительного периода времени (B_3).

В случае стратегий A_2 и A_3 предприятие несет дополнительные затраты на хранение и обработку продукции, которые не требуются для A_1, однако при A_2 следует учесть возможные убытки из-за порчи продукции, если потребитель выберет стратегии B_2 или B_3.

Определить оптимальные пропорции продукции для применения стратегий A_1, A_2, A_3, руководствуясь "минимаксным критерием" (гарантированный средний уровень убытка) при матрице затрат, представленной табл. 9.6.

Таблица 9.6

A_i \ B_j	B_1	B_2	B_3
A_1	2	5	8
A_2	7	6	10
A_3	12	10	8

Решение. Получаем игру с платежной матрицей $P = \begin{pmatrix} 2 & 5 & 8 \\ 7 & 6 & 10 \\ 12 & 10 & 8 \end{pmatrix}$.

В этой матрице первую строку можно отбросить как невыгодную (ее элементы меньше соответствующих элементов второй строки). Матрица примет вид $P = \begin{pmatrix} 7 & 6 & 10 \\ 12 & 10 & 8 \end{pmatrix}$. Элементы первого столбца больше соответствующих элементов второго столбца, поэтому его можно отбросить.

Игра упростилась: $P = \begin{pmatrix} 6 & 10 \\ 10 & 8 \end{pmatrix}$.

По формулам (9.7) и (9.8) находим:

$$p_2^* = \frac{8-10}{6+8-10-10} = \frac{1}{3};\ p_3^* = 1 - \frac{1}{3} = \frac{2}{3};$$

$$\nu = \frac{6 \cdot 1}{3} + \frac{10 \cdot 2}{3} = \frac{26}{3} = 8\frac{2}{3}.$$

Вывод: оптимальная стратегия производителя продукции $S_A^* = \left(0; \frac{1}{3}; \frac{2}{3}\right)$, т.е. стратегия A_1 не применяется, 1/3 продукции отправляется на склад (стратегия A_2), 2/3 продукции дополнительно обрабатывается (стратегия A_3), при этом цена игры $\nu = 8\frac{2}{3}$. ▶

УПРАЖНЕНИЯ

9.7. Игрок A записывает одно из двух чисел: 1 или 2, игрок B — одно из трех чисел: 1, 2 или 3. Если оба числа одинаковой четности, то A выигрывает и выигрыш равен сумме этих чисел, если четности выбранных игроками чисел не совпадают, то B выигрывает, выигрыш равен сумме этих чисел. Построить платежную матрицу игры, определить нижнюю и верхнюю цены игры и проверить наличие седловой точки.

В задачах **9.8—9.13** для следующих платежных матриц определить нижнюю и верхнюю цены игры, минимаксные стратегии и оптимальные решения игры, если существует седловая точка.

9.8.
$$\begin{pmatrix} 0{,}3 & 0{,}6 & 0{,}8 \\ 0{,}9 & 0{,}4 & 0{,}2 \\ 0{,}7 & 0{,}5 & 0{,}4 \end{pmatrix}$$

9.9.
$$\begin{pmatrix} 4 & 5 & 3 \\ 6 & 7 & 4 \\ 5 & 2 & 3 \end{pmatrix}$$

9.10
$$\begin{pmatrix} 8 & 9 & 9 & 4 \\ 6 & 5 & 8 & 7 \\ 3 & 4 & 5 & 6 \end{pmatrix}$$

9.11.
$$\begin{pmatrix} 4 & 5 & 6 & 7 & 9 \\ 3 & 4 & 6 & 5 & 6 \\ 7 & 6 & 10 & 8 & 11 \\ 8 & 5 & 4 & 7 & 3 \end{pmatrix}$$

9.12.
$$\begin{pmatrix} 4 & 9 & 5 & 3 \\ 7 & 8 & 6 & 9 \\ 7 & 4 & 2 & 6 \\ 8 & 3 & 4 & 7 \end{pmatrix}$$

9.13.
$$\begin{pmatrix} 2 & 5 & 3 \\ 6 & 4 & 5 \\ 3 & 7 & 6 \\ 2 & 3 & 4 \end{pmatrix}$$

В задачах **9.14—9.16** решить и дать графическую интерпретацию для следующих игр 2×2.

9.14. $\begin{pmatrix} -2 & 2 \\ 1 & -1 \end{pmatrix}$ **9.15.** $\begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix}$ **9.16.** $\begin{pmatrix} 4 & -2 \\ 1 & 3 \end{pmatrix}$

9.17. Найти решения игр путем сведения их к задаче линейного программирования, используя платежные матрицы задач 9.9, 9.11, 9.12.

9.18. Найти решение игры, предварительно упростив ее:

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

9.19. Магазин может завезти в различных пропорциях товары трех типов (A_1, A_2, A_3); их реализация и прибыль магазина зависят от вида товара и состояния спроса.

Предполагается, что спрос может иметь три состояния (B_1, B_2, B_3) и не прогнозируется. Определить оптимальные пропорции в закупке товаров из условия максимизации средней гарантированной прибыли при следующей матрице прибыли (табл. 9.7).

Таблица 9.7

Тип товара	Спрос		
	B_1	B_2	B_3
A_1	20	15	10
A_2	16	12	14
A_3	13	18	15

РАЗДЕЛ II. МОДЕЛИ НЕЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Глава 10. КЛАССИЧЕСКИЕ МЕТОДЫ ОПТИМИЗАЦИИ

10.1. Классические методы определения экстремумов

Во многих экономических моделях исследования операций зависимости между постоянными и переменными факторами лишь в первом приближении можно считать линейными, более детальное рассмотрение позволяет обнаружить их нелинейность. Как правило, такие показатели, как прибыль, себестоимость, капитальные затраты на производство и др., в действительности зависят от объема производства, расхода ресурсов и т.п. нелинейно. В этом случае возникает задача нелинейного программирования, математическая модель которой (0.1), (0.2) приведена во введении.

Можно выделить класс нелинейных задач, которые относятся к *классическим методам оптимизации*. Допустим, что среди ограничений (0.1) нет неравенств, не обязательны условия неотрицательности, переменные не являются дискретными, $m < n$, а функции $\varphi_i(X)$ и $f(X)$ непрерывны и имеют частные производные по крайней мере второго порядка. В этом случае задачу оптимизации можно сформулировать так: ***найти переменные*** $x_1, x_2, ..., x_n$, ***удовлетворяющие системе уравнений***

$$\varphi_i(x_1, x_2, \ldots, x_n) = b_i,\ i = 1, 2, ..., m \tag{10.1}$$

и обращающие в максимум (минимум) целевую функцию

$$z = f(x_1, x_2, \ldots, x_n). \tag{10.2}$$

Такие задачи в принципе можно решать классическими методами дифференциального исчисления. Однако на этом пути встречаются такие вычислительные трудности, которые делают необходимым поиск других методов решения (например, *см.* гл. 11, 12). Поэтому классические методы часто используется не в качестве вычислительного средства, а как основа для теоретического анализа.

Примером типичной и простой нелинейной задачи является следующая: данное предприятие для производства какого-то продукта расходует два средства в количестве x_1 и x_2 соответственно. Это факторы производства, например, машины и труд, два различных вида сырья и т.п., а величины x_1 и x_2 — затраты факторов производства. Факторы производства впредь будем считать взаимозаменяемыми. Если это "труд" и "машины", то можно применять такие методы производства, при которых величина затрат машин в сопоставлении с величиной затрат труда оказывается больше или меньше (производство более или менее трудоемкое). В сельском хозяйстве взаимозаменяемыми факторами могут быть посевные площади или минеральные удобрения (экстенсивный или интенсивный метод производства).

Объем производства (выраженный в натуральных или стоимостных единицах) является функцией затрат производства $z = f(x_1, x_2)$. Эта зависимость называется ***производственной функцией***. Издержки зависят от расхода обоих факторов (x_1 и x_2) и от цен этих факторов (c_1 и c_2). Совокупные издержки выражаются формулой $b = c_1x_1 + c_2x_2$. Требуется при данных совокупных издержках определить такое количество факторов производства, которое максимизирует объем продукции z.

Математическая модель этой задачи имеет вид: определить такие переменные x_1 и x_2, удовлетворяющие условиям

$$\begin{gathered} c_1x_1 + c_2x_2 = b, \\ x_1 \ge 0,\ x_2 \ge 0, \end{gathered} \tag{10.3}$$

при которых функция

$$z = f(x_1, x_2) \tag{10.4}$$

достигает максимума.

Как правило, функция (10.4) может иметь произвольный нелинейный вид.

Используя классические методы оптимизации, следует четко представлять себе различие между *локальным* экстремумом функции, *глобальным* экстремумом и *условным* экстремумом. При этом полезно повторить определение локального и глобального экстремумов для функции одной переменной. Понятие условного экстремума вводится для случая, когда число переменных n не меньше 2 ($n \geq 2$).

Будем полагать, что функция $z = f(x_1, x_2, \ldots, x_n) = f(X)$ дважды дифференцируема в точке $X^* = (x_1^*, x_2^*, \ldots, x_n^*)$, $(X^* \in D(f))$ и в некоторой ее окрестности. Если для всех точек X этой окрестности $f(X^*) \geq f(X)$ или $f(X^*) \leq f(X)$, то говорят, что функция $f(X)$ имеет экстремум в X^* (соответственно максимум или минимум).

Точка X^*, в которой все частные производные функции $z = f(X)$ равны 0, называется *стационарной точкой.*

Необходимое условие экстремума. *Если в точке* X^* функция $z = f(X)$ *имеет экстремум, то частные производные функции в этой точке раны нулю:*

$$f'_{x_i}(X^*) = 0, \ i = 1,2,\ldots,n.$$

Следовательно, точки экстремума функции $z = f(X)$ удовлетворяют системе уравнений:

$$\begin{cases} f'_{x_1}(x_1, x_2, \ldots, x_n) = 0, \\ f'_{x_2}(x_1, x_2, \ldots, x_n) = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ f'_{x_n}(x_1, x_2, \ldots, x_n) = 0. \end{cases} \qquad (10.5)$$

Как и в случае одной переменной, необходимое условие не является достаточным для того, чтобы стационарная точка была точкой экстремума. Для получения достаточных условий следует определить в стационарной точке знак дифференциала второго порядка. Дифференциал второго порядка обозначается $d^2 f(x_1, x_2, \ldots, x_n)$ и равен сумме произведений частных производных второго порядка на соответствующие приращения аргументов. Если от частной производной $f'_{x_i}(X)$ найти частную производную по переменной x_j, то получим частную производную второго порядка по переменным x_i, x_j, которая обозначается $f''_{x_i x_j}(X)$. В этом случае

$$d^2 f(X) = \sum_{i=1}^{n} \sum_{j=1}^{n} f''_{x_i x_j}(X) \Delta x_i \Delta x_j.$$

Достаточные условия экстремума:

а) в стационарной точке X^0 функция $z = f(X)$ имеет максимум, если $d^2 f(X^0) < 0$, и минимум, если $d^2 f(X^0) > 0$, при любых Δx_i и Δx_j (в этих случаях $X^0 = X^*$), не обращающихся в нуль одновременно;

б) если $d^2 f(X^0)$ может принимать в зависимости от Δx_i и Δx_j и положительные, и отрицательные значения, то в точке X^0 экстремума нет;

в) если $d^2 f(X^0)$ может обращаться в нуль не только при нулевых приращениях Δx_i и Δx_j, то вопрос об экстремуме остается открытым.

Для функции двух переменных $z = f(x_1, x_2)$ достаточные условия еще не очень сложны. Существуют четыре частные производные второго порядка: $f''_{x_1^2}(X)$, $f''_{x_1 x_2}(X)$, $f''_{x_2 x_1}(X)$, $f''_{x_2^2}(X)$. Из них две смешанные производные $f''_{x_1 x_2}(X)$ и $f''_{x_2 x_1}(X)$, если непрерывны, то равны.

Найдем значение частных производных второго порядка в стационарной точке $X^0(x_1^0, x_2^0)$:

$$a_{11} = f''_{x_1^2}(X^0);\ a_{12} = f''_{x_1 x_2}(X^0);\ a_{21} = f''_{x_2 x_1}(X^0);\ a_{22} = f''_{x_2^2}(X^0)$$

(можно убедиться, что $a_{12} = a_{21}$). Обозначим через Δ определитель, составленный из a_{ij} для $i, j = 1, 2$:

$$\Delta = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21}.$$

Тогда достаточные условия экстремума функции двух переменных имеют вид:

а) если $\Delta > 0$ *и* $a_{11} < 0$ $(a_{22} < 0)$, *то в точке* X^0 *функция имеет максимум: если* $\Delta > 0$ *и* $a_{11} > 0$ $(a_{22} > 0)$, *то в точке* X^0 *— минимум (в этих случаях* $X^0 = X^*$*);*

б) *если* $\Delta < 0$, *то экстремума нет;*

в) *если* $\Delta = 0$, *то вопрос об экстремуме остается открытым.*

Схема определения экстремума функции n переменных совпадает с правилами определения локального экстремума функции одной переменной.

▷ **10.1.** Исследовать на экстремум функцию

$$z = x_1^4 + x_2^4 - x_1^2 - 2x_1x_2 - x_2^2.$$

Решение. Находим частные производные:

$$\begin{cases} z'_{x_1} = 4x_1^3 - 2x_1 - 2x_2, \\ z'_{x_2} = 4x_2^3 - 2x_1 - 2x_2. \end{cases} \qquad \textbf{(10.6)}$$

Приравниваем частные производные нулю:

$$\begin{cases} 4x_1^3 - 2x_1 - 2x_2 = 0, \\ 4x_2^3 - 2x_1 - 2x_2 = 0. \end{cases} \qquad \textbf{(10.7)}$$

Решаем систему уравнений (10.7). Вычитая из первого уравнения второе, получим $4x_1^3 - 4x_2^3 = 0$, поэтому $x_1 = x_2$, и из первого уравнения найдем $x_1^3 - x_1 = 0$, откуда $x_1 = 0$ или $x_1 = \pm 1$.

Имеем три стационарные точки: $X^1 = (0;0)$; $X^2 = (1;1)$; $X^3 = (-1;-1)$.

Найдем вторые частные производные, используя (10.6):

$$\begin{cases} z''_{x_1^2} = (4x_1^3 - 2x_1 - 2x_2)'_{x_1} = 12x_1^2 - 2; \\ z''_{x_1x_2} = (4x_1^3 - 2x_1 - 2x_2)'_{x_2} = -2; \\ z''_{x_2x_1} = (4x_2^3 - 2x_2 - 2x_1)'_{x_1} = -2; \\ z''_{x_2^2} = (4x_2^3 - 2x_2 - 2x_1)'_{x_2} = 12x_2^2 - 2. \end{cases}$$

Вычисляем значения вторых частных производных в каждой стационарной точке, составляем определитель Δ и применяем достаточные условия экстремума.

В точке $X^1 = (0; 0)$ $a_{11} = -2$; $a_{12} = a_{21} = -2$; $a_{22} = -2$;

$$\Delta = \begin{vmatrix} -2 & -2 \\ -2 & -2 \end{vmatrix} = 0.$$

Вопрос об экстремуме остается открытым (такая точка называется *седловой*).

В точке $X^2 = (1; 1)$ (а также и в точке $X^3 = (-1;-1)$):

$$a_{11} = 10;\ a_{12} = a_{21} = -2;\ a_{22} = 10;\quad \Delta = \begin{vmatrix} 10 & -2 \\ -2 & 10 \end{vmatrix} = 96.$$

Функция в этих точках имеет минимум, так как $\Delta > 0$, $a_{11} > 0$. $Z_{\min} = -2$[1].▶

Выше шла речь о локальном экстремуме функции n переменных. Как правило, в практических задачах необходимо определить наибольшее и наименьшее значения функции (глобальный экстремум) в некоторой области.

Говорят, что функция $z = f(X)$ имеет в точке X^0 заданной области D *глобальный максимум (наибольшее значение)* или *глобальный минимум (наименьшее значение)*, если неравенство $f(X) \le f(X^0)$ или $f(X) \ge f(X^0)$ соответственно выполняется для любой точки $X \in D$.

*Если область D замкнута и ограничена, то дифференцируемая функция $z = f(X)$ достигает в этой области своих наибольшего и наименьшего значений или в стационарной точке, или в граничной точке области (**теорема Вейерштрасса**)*[2].

---

[1] В точках $X^2 = (1; 1)$ и $X^3 = (-1;-1)$ $d^2 f(X^2) = d^2 f(X^3) =$ $= 10(\Delta x_1)^2 - 4\Delta x_1 \Delta x_2 + 10(\Delta x_2)^2 = 2(\Delta x_1 - \Delta x_2)^2 + 8(\Delta x_1)^2 + 8(\Delta x_2)^2$ и при любых Δx_1 и Δx_2, не равных нулю одновременно, $d^2 f(X^2) =$ $= d^2 f(X^3) > 0$.

[2] Вейерштрасс К. Т. (1815-1897) — немецкий математик.

Следовательно, чтобы найти наибольшее (наименьшее) значение функции $z = f(X)$ в области D, нужно:

1) найти все стационарные точки внутри области D и вычислить значения функции в них;

2) исследовать функцию на экстремум на границе области D;

3) сравнить значения функции, полученные в п. 1 и 2: наибольшее (наименьшее) из этих чисел и будет наибольшим (наименьшим) значением функции во всей области.

Граница области D аналитически может быть задана системой уравнений (условий) относительно переменных x_1, x_2, ..., x_n. Поэтому, исследуя экстремальные свойства функции на границе, необходимо решить задачу определения условного экстремума.

Условный экстремум. Пусть необходимо найти экстремум функции $z = f(x_1, x_2, \ldots, x_n)$ при условии, что переменные x_1, x_2, ..., x_n удовлетворяют уравнениям

$$\varphi_i(x_1, x_2, \ldots, x_n) = 0,\ i = 1, 2, \ldots, m,\ m < n. \qquad \textbf{(10.8)}$$

Предполагается, что функции f и φ_i имеют непрерывные частные производные по всем переменным. Уравнения (10.8) называют *уравнениями связи.*

Говорят, что в точке $X^0 = (x_1^0, x_2^0, \ldots, x_n^0)$, удовлетворяющей уравнениям связи (10.8), функция $z = f(X)$ имеет *условный максимум (минимум)*, если неравенство $f(X^0) \geq f(X)$ ($f(X^0) \leq f(X)$) имеет место для всех точек X, координаты которых удовлетворяют уравнениям связи.

Легко заметить, что задача определения условного экстремума совпадает с задачей нелинейного программирования (10.1), (10.2).

Один из способов определения условного экстремума применяется в том случае, если из уравнений связи (10.8) m переменных, например x_1, x_2, ..., x_m, можно явно выразить через оставшиеся $n-m$ переменных:

$$x_i = \psi_i(x_{m+1}, \ldots, x_n),\ i = 1, 2, \ldots, m. \qquad \textbf{(10.9)}$$

Подставив полученные выражения для x_i в функцию z, получим

$$z = f(\psi_1(x_{m+1}, \ldots, x_n), \ldots, \psi_m(x_{m+1}, \ldots, x_n), x_{m+1}, \ldots, x_n),$$

или

$$z = F(x_{m+1}, \ldots, x_n). \qquad \textbf{(10.10)}$$

Задача сведена к нахождению локального (глобального) экстремума для функции (10.10) от $n-m$ переменных. Если в точке $X^0 = (\tilde{x}^0_{m+1}, \ldots, \tilde{x}^0_n)$ функция (10.10) имеет экстремум, то в точке $X^0 = (\psi_1(\tilde{x}^0_{m+1}, \ldots, \tilde{x}^0_n), \ldots, \psi_m(\tilde{x}^0_{m+1}, \ldots, \tilde{x}^0_n), \tilde{x}^0_{m+1}, \ldots, \tilde{x}^0_n)$ функция $z = f(x_1, \ldots, x_n)$ имеет условный экстремум (сформулированный на с. 206).

▷ **10.2.** Решить задачу, сформулированную на с. 201 в предположении, что производственная функция равна z :

$$z = x_1^2 x_2 (4 - x_1 - x_2), \ c_1 = 1, \ c_2 = 2, \ b = 4.$$

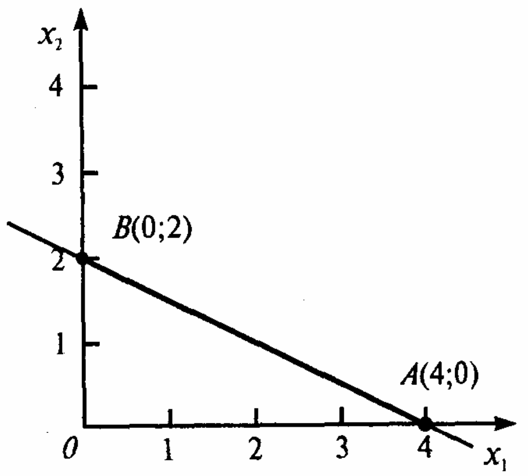


Рис. 10.1

Решение. Необходимо найти переменные x_1 и x_2, удовлетворяющие уравнению

$$x_1 + 2x_2 = 4, \qquad \textbf{(10.11)}$$

(уравнение связи), условию неотрицательности $x_1 \geq 0$, $x_2 \geq 0$ и обращающие в максимум функцию

$$z = x_1^2 x_2 (4 - x_1 - x_2). \qquad \textbf{(10.12)}$$

Ограничение (10.11) вместе с условиями неотрицательности определяют на плоскости $x_1 O x_2$ отрезок AB — замкнутую ограниченную область (рис. 10.1).

Согласно теореме Вейерштрасса максимум функции может достигаться либо внутри этого отрезка, либо в граничных точках: A (4;0) или B (0;2).

Следовательно, необходимо найти условный экстремум функции (10.12), если уравнение связи имеет вид (10.11).

Из уравнения связи найдем, например, x_1, и подставим в (10.12):

$$x_1 = 4 - 2x_2, \ z = (4 - 2x_2)^2 x_2 (4 - 4 + 2x_2 - x_2).$$

Упростив это выражение, получим

$$z = 4(2 - x_2)^2 x_2^2. \tag{10.13}$$

При этом $x_2 \in [0;2]$. Найдем глобальный экстремум функции (10.13) на отрезке [0;2]. Производная этой функции равна

$$z' = 16\,(2 - x_2)x_2(1 - x_2).$$

Стационарные точки: $x_2 = 0$, $x_2 = 1$ и $x_2 = 2$. Одна из них $x_2 = 1$, лежит внутри отрезка, две другие совпадают с концами. Найдем значения функции (10.13) в стационарной точке $x_2 = 1$ и на концах отрезка: $z\,(1) = 4$; $z\,(0) = 0$; $z\,(2) = 0$.

Следовательно, $z_{\max} = 4$ и достигается при $x_2 = 1$, $x_1 = 4 - 2x_2 = 2$, т.е. в точке (2; 1).

Максимальный объем производства, равный $z_{\max} = 4$ ед., достигается при условии, что затраты производственных факторов x_1 и x_2 равны соответственно 2 ед. и 1 ед.▶

10.2. Метод множителей Лагранжа

Другой способ определения условного экстремума начинается с построения вспомогательной функции Лагранжа[1], которая в области допустимых решений достигает максимума для тех же значений переменных $x_1, x_2, \ldots, x_n$, что и целевая функция z.

Пусть решается задача определения условного экстремума функции $z = f(X)$ при ограничениях (10.8).

Составим функцию

$$L(X) = f(X) + \sum_{i=1}^{m} \lambda_i \varphi_i(X), \tag{10.14}$$

которая называется *функцией Лагранжа*. λ_i — постоянные множители (*множители Лагранжа*). Отметим, что множителям Лагранжа можно придать экономический смысл. Если $f(x_1, x_2, \ldots, x_n)$ — доход, соответствующий плану $X = (x_1, x_2, \ldots, x_n)$, а функция $\varphi_i(x_1, x_2, \ldots, x_n)$ — издержки i-го ресурса, соответствующие этому плану, то λ_i — цена (оценка) i-го ресурса, характеризующая изменение экстремального значения целевой функции в зависимости от из-

---

[1] Лагранж Ж.Л. (1736-1813) — французский математик и механик.

менения размера i-го ресурса (маргинальная оценка). $L(X)$ — функция $n+m$ переменных ($x_1, x_2, ..., x_n, \lambda_1, \lambda_2, \ldots, \lambda_m$). Определение стационарных точек этой функции приводит к решению системы уравнений

$$\begin{cases} \dfrac{\partial L(X)}{\partial x_j} = 0, \ j = 1,2,\ldots,n, \\ \dfrac{\partial L(X)}{\partial \lambda_i} = 0, \ i = 1,2,\ldots,m. \end{cases} \tag{10.15}$$

Легко заметить, что $L'_{\lambda_i}(X) = \varphi_i(X)$, т.е. в (10.15) входят уравнения связи. Таким образом, задача нахождения условного экстремума функции $z = f(X)$ сводится к нахождению локального экстремума функции $L\ (X)$. Если стационарная точка найдена, то вопрос о существовании экстремума в простейших случаях решается на основании достаточных условий экстремума — исследования знака второго дифференциала $d^2L(X)$ в стационарной точке при условии, что переменные приращения Δx_j связаны соотношениями

$$\sum_{j=1}^{n} \frac{\partial \varphi_i}{\partial x_j} \Delta x_j = 0, \ i = 1, 2, ..., m, \tag{10.16}$$

полученными путем дифференцирования уравнений связи. Рассмотрим пример.

▷ **10.3.** Найти наибольшее и наименьшее значения функции

$$z = 9x_1^2 + 4x_2^2 + x_3^2 - (3x_1^2 + 2x_2^2 + x_3^2)$$

при условии, что x_1, x_2, x_3 удовлетворяют уравнению

$$x_1^2 + x_2^2 + x_3^2 = 1.$$

Решение. Уравнение связи определяет в пространстве сферу единичного радиуса с центром в начале координат (рис. 10.2). Так как сфера — замкнутое ограниченное множество, то согласно теоре-

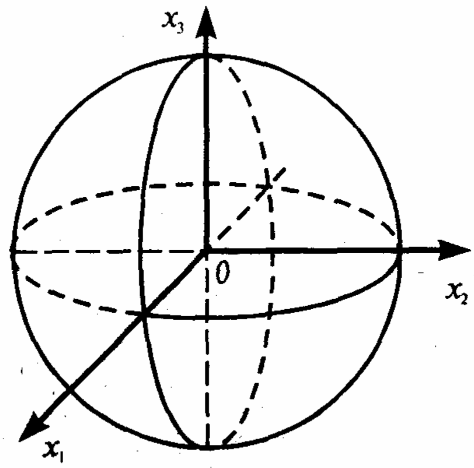


Рис. 10.2

ме Вейерштрасса функция достигает на ней своего наибольшего и наименьшего значений.

Необходимо найти условный глобальный экстремум. Запишем уравнение связи в виде: $x_1^2 + x_2^2 + x_3^2 - 1 = 0$.

Составим функцию Лагранжа:

$L(x_1, x_2, x_3) = 9x_1^2 + 4x_2^2 + x_3^2 - (3x_1^2 + 2x_2^2 + x_3^2)^2 + \lambda(x_1^2 + x_2^2 + x_3^2 - 1)$.

Найдем частные производные этой функции по x_1, x_2, x_3, λ.

$$L'_{x_1} = 18x_1 - 2(3x_1^2 + 2x_2^2 + x_3^2)6x_1 + 2\lambda x_1,$$
$$L'_{x_2} = 8x_2 - 2(3x_1^2 + 2x_2^2 + x_3^2)4x_2 + 2\lambda x_2,$$
$$L'_{x_3} = 2x_3 - 2(3x_1^2 + 2x_2^2 + x_3^2)2x_3 + 2\lambda x_3,$$
$$L'_{\lambda} = x_1^2 + x_2^2 + x_3^2 - 1.$$

Приравняв частные производные нулю, получим систему:

$$\begin{cases} x_1((9+\lambda) - 6(3x_1^2 + 2x_2^2 + x_3^2)) = 0, \\ x_2((4+\lambda) - 4(3x_1^2 + 2x_2^2 + x_3^2)) = 0, \\ x_3((1+\lambda) - 2(3x_1^2 + 2x_2^2 + x_3^2)) = 0, \\ x_1^2 + x_2^2 + x_3^2 = 1. \end{cases}$$

Решая систему, получим стационарные точки, в которых найдем значения функции z:

$$1.\ x_1 = x_2 = 0;\ x_3 = \pm 1 \Rightarrow z = 0.$$

$$2.\ x_1 = 0;\ x_2 = \pm 1;\ x_3 = 0 \Rightarrow z = 0.$$

$$3.\ x_1 = \pm 1;\ x_2 = x_3 = 0 \Rightarrow z = 0.$$

$$4.\ x_1 = 0;\ x_2 = x_3 = \pm\frac{1}{\sqrt{2}} \Rightarrow z = \frac{1}{4}.$$

$$5.\ x_1 = \pm\frac{1}{\sqrt{2}};\ x_2 = 0;\ x_3 = \pm\frac{1}{\sqrt{2}} \Rightarrow z = 1.$$

$$6.\ x_1 = x_2 = \pm\frac{1}{\sqrt{2}};\ x_3 = 0 \Rightarrow z = \frac{1}{4}.$$

Выберем из всех значений z наибольшее и наименьшее: $z_{\text{наиб.}} = 1$, а $z_{\text{наим.}} = 0$. Легко видеть, в каких точках сферы достигаются эти значения.▶

Если число переменных $n = 2$, нелинейные задачи можно решать геометрически. Ограничения должны быть записаны в виде неравенств

$$\varphi_i(x_1, x_2) \le b_i, \; i = 1,2,\dots,m, \tag{10.17}$$

а целевая функция иметь вид

$$z = f(x_1, x_2). \tag{10.18}$$

Как и в случае геометрического решения задач линейного программирования, сначала необходимо построить область допустимых решений (ОДР) — множество точек плоскости, удовлетворяющих неравенствам (10.17). Но в отличие от задач линейного программирования здесь ОДР не обязательно будет выпуклой и может быть даже разрывной. Экстремум функции может достигаться и внутри области, и на границе.

После построения ОДР следует записать уравнения линий уровня целевой функции — множество точек плоскости, в которых целевая функция (10.18) постоянна: $f(x_1, x_2) = C$, и определить направление возрастания (убывания) целевой функции, построив, например, линии уровня для разных значений C. Затем, перемещая линию уровня в нужном направлении в ОДР, найти точки области, в которых целевая функция принимает оптимальное значение.

▷ **10.4.** Найти наибольшие значения функции $z = 2x_1^2 - x_2$ при ограничениях

$$\begin{cases} x_1 - x_2 \le 2, \\ \quad\;\; x_2 \le 4, \\ x_1 + x_2 - x_1x_2 \ge 0, \\ x_1 \ge 0, \; x_2 \ge 0. \end{cases}$$

Решение. ОДР (рис.10.3) ограничена прямыми $x_1 - x_2 = 2$, $x_2 = 4$, осями координат $x_1 = 0$, $x_2 = 0$ и гиперболой $x_1 + x_2 - x_1x_2 = 0$, уравнение которой приводится к виду $x_2 = 1 + \dfrac{1}{x_1 - 1}$.

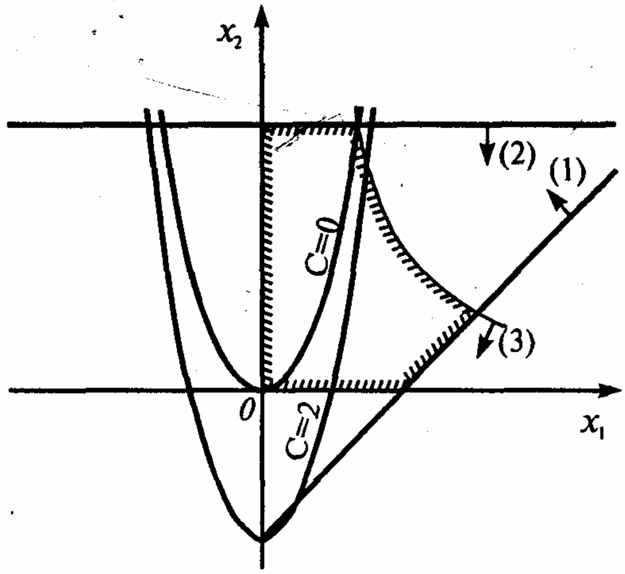


Рис. 10.3

Линии уровня целевой функции — $2x_1^2 - x_2 = C$. Для разных значений C графиком уравнения $x_2 = 2x_1^2 - C$ является парабола с осью симметрии, совпадающей с осью ординат.

При C=0 парабола проходит через начало координат. При $C > 0$ параболы сдвигаются вниз. Перемещая в направлении возрастания, получим, что линии уровня покидают ОДР через точку X^* пересечения гиперболы $x_2 = 1 + \dfrac{1}{x_1 - 1}$ и прямой $x_1 - x_2 = 2$.

Решая систему, составленную из этих двух уравнений, получим $x_1^* = 2 + \sqrt{2}$, $x_2^* = \sqrt{2}$, $X^* = (\sqrt{2} + 2; \sqrt{2})$. Поэтому $z_{\max} = 2(\sqrt{2} + 2)^2 - \sqrt{2}$ или $z_{\max} \approx 21{,}9$. ▶

Трудности применения классических методов оптимизации уже отмечались выше (с. 10). Поэтому разработаны приближенные методы решения нелинейных задач программирования, особенно плодотворные для некоторых классов функций, например, для выпуклых (вогнутых) функций, рассмотренных в следующей главе.

УПРАЖНЕНИЯ

В задачах **10.5—10.7** найти локальный экстремум следующих функций.

10.5. $z = x^3 + y^3 - 3xy$.

10.6. $z = x^3y^2(12 - x - y)$, $x > 0$, $y > 0$.

10.7. $z = x^2 + xy + y^2 + x - y + 1$.

В задачах **10.8—10.12** найти глобальный экстремум (наибольшее и наименьшее значения) функции z в области решений системы неравенств (или неравенства). Дать геометрическое решение.

10.8. $z = 3x_1 + x_2$

$$\begin{cases} x_1^2 + x_2^2 \le 40, \\ x_1^2 + x_2^2 \ge 4, \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

10.9. $z = x^2 + y^2 - 2x - 10y + 26$

$$\begin{cases} x_1 - 2x_2 \ge -4, \\ 5x_1 + 2x_2 \le 20, \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

10.10. $z = x_1^2 + 2x_2 - 3$

$$\begin{cases} x_1^2 + x_2^2 \le 10, \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

10.11. $z = e^{-x_1^2 - x_2^2}(2x_1^2 + 3x_2^2)$,

$$x_1^2 + x_2^2 \le 4.$$

10.12. $z = x_1 x_2$,

$$\begin{cases} 0 \le x_1 \le 3, \\ 2x_1 + x_2 \le 8. \end{cases}$$

В задачах **10.13—10.15** найти условный экстремум функции z с помощью метода исключения.

10.13. $z = x_1^2 + x_2^2$ при $x_1 + x_2 = 1$.

10.14. $z = x_1^2 - x_2^2$ в области $x_1^2 + x_2^2 \le 16$ при $x_1 - x_2 = 4$.

10.15. $z = 3x_1^2 + 2x_2^2 - 3x_1 + 1$ при $x_1^2 + x_2^2 = 4$.

В задачах **10.16—10.19** найти условный экстремум с помощью метода Лагранжа.

10.16. $z = x_1 x_2$ при $x_1^2 + x_2^2 = 2$.

10.17. $z = x_1 + x_2$ при $\dfrac{1}{x_1} + \dfrac{1}{x_2} = 1$.

10.18. $z = \dfrac{1}{x_1} + \dfrac{1}{x_2}$ при $\dfrac{1}{x_1^2} + \dfrac{1}{x_2^2} = 1$.

10.19. $z = x_1^3 + x_2^3$ при $x_1 + x_2 = 2$, $x_1 \ge 0$, $x_2 \ge 0$.

Глава 11. МОДЕЛИ ВЫПУКЛОГО ПРОГРАММИРОВАНИЯ

Рассмотрим задачу нелинейного программирования (10.1) и (10.2) при условии, что функции f и φ_i ($i = 1, 2, ..., m$) являются выпуклыми. Введем необходимые понятия.

11.1. Производная по направлению и градиент. Выпуклые функции

Производной $\dfrac{\partial F}{\partial l}$ *функции* $F(X) = F(x_1,..., x_n)$ *по направлению* l *в точке* X *называется предел*

$$\frac{\partial F}{\partial l} = \lim_{\lambda \to +0} \frac{F(X + \lambda l) - F(X)}{\lambda}.$$

Направление l обычно задается вектором $l = (l_1, ..., l_n)$.

Если функция F дифференцируема в точке X, то она имеет в этой точке производную по любому направлению l, которая выражается через частные производные по формуле

$$\frac{\partial F}{\partial l} = \frac{1}{|l|}\sum_{i=1}^{n} l_i \cdot \frac{\partial F}{\partial x_i}, \qquad \textbf{(11.1)}$$

где $|l|$ — длина вектора l, т.е. $|l| = \sqrt{l_1^2 + \ldots + l_n^2}$.

Абсолютная величина производной по направлению дает скорость изменения функции в этом направлении, а знак показывает характер изменения функции (возрастание или убывание).

Градиентом ∇F ***функции*** $F(X) = F(x_1, ..., x_n)$ ***называется вектор, проекциями которого на координатные оси служат соответствующие частные производные***, т.е.

$$\nabla F = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \dots, \frac{\partial F}{\partial x_n}\right).$$

Можно показать, что $\max\left(\frac{\partial F}{\partial l}\right)$ достигается тогда, когда направление l совпадает с направлением ∇F. По формуле (11.1) производная функции F по направлению градиента ∇F равна

$$\frac{\partial F}{\partial(\nabla F)} = \frac{1}{|\nabla F|} \cdot \sum_{i=1}^{n} \frac{\partial F}{\partial x_i} \cdot \frac{\partial F}{\partial x_i} = |\nabla F|.$$

Таким образом, в каждой точке X *направление градиента является направлением наибольшего возрастания функции*, а длина градиента равна наибольшей скорости возрастания функции в этой точке.

▷ **11.1.** Найти наибольшую скорость возрастания функции $F = x_1x_2x_3 + 2x_3$ в точке A (0; 1; 2) и определить характер изменения этой функции в точке A в направлении $l = (1; -2; 2)$.

Решение. Так как $\frac{\partial F}{\partial x_1} = x_2x_3$, $\frac{\partial F}{\partial x_2} = x_1x_3$, $\frac{\partial F}{\partial x_3} = x_1x_2 + 2$, то $\nabla F = (x_2x_3,\ x_1x_3,\ x_1x_2 + 2)$ и $\nabla F|_A = (2;\ 0;\ 2)$. Таким образом, наибольшая скорость возрастания функции в точке A равна $|\nabla F|_A = \sqrt{4+0+4} = 2\sqrt{2}$. Далее $|l| = \sqrt{1+4+4} = 3$ и $\left.\frac{\partial F}{\partial l}\right|_A =$ $= \frac{1}{3}(1\cdot 2 + (-2)\cdot 0 + 2\cdot 2) = \frac{6}{3} = 2$. Так как $\left.\frac{\partial F}{\partial l}\right|_A > 0$, функция F в точке A в направлении l возрастает.▶

Напомним (*см.* разд. 2.2), что множество точек называется ***выпуклым***, если оно вместе с любыми своими двумя точками содержит и весь отрезок, соединяющий эти точки. Если $[a,b]$ — отрезок на числовой прямой и $x \in [a,b]$, то, как показано в разд. 3.1, $x = \alpha a + (1-\alpha)b$, $0 \le \alpha \le 1$ или

$$x = \alpha_1 a + \alpha_2 b,\ \ \alpha_1 + \alpha_2 = 1,\ \ \alpha_1 \ge 0,\ \ \alpha_2 \ge 0. \qquad \textbf{(11.2)}$$

Нетрудно видеть и обратное: если выполняется (11.2), то $x \in [a,b]$. Таким образом, отрезок $[a,b]$ можно определить как множество всех точек x, удовлетворяющих условию (11.2). Тогда выпуклое множество — это множество, которое вместе с любой парой своих точек a, b содержит и все точки x, для которых выполняется (11.2). Эти определения отрезка и выпуклого множества сохраняются для случая, когда a, b, x — точки n-мерного пространства (где операции в равенстве (11.2) выполняются покоординатно).

Исходя из равенства (11.2), по индукции можно показать, что если M — выпуклое пространство, то $\sum_{i=1}^{r} t_i X_i \in M$ для любых точек $X_1, \dots, X_r \in M$ и любых действительных чисел $t_i \geq 0$, удовлетворяющих условию $\sum_{i=1}^{r} t_i = 1$.

Функция $F(X) = F(x_1, \dots, x_n)$, определенная на выпуклом множестве M n-мерного пространства, называется ***выпуклой*** *на этом множестве, если*

$$F(\alpha X_1 + (1-\alpha)X_2) \leq \alpha F(X_1) + (1-\alpha)F(X_2) \qquad \textbf{(11.3)}$$

для любых точек $X_1, X_2 \in M$ и любого числа $\alpha \in [0,1]$.

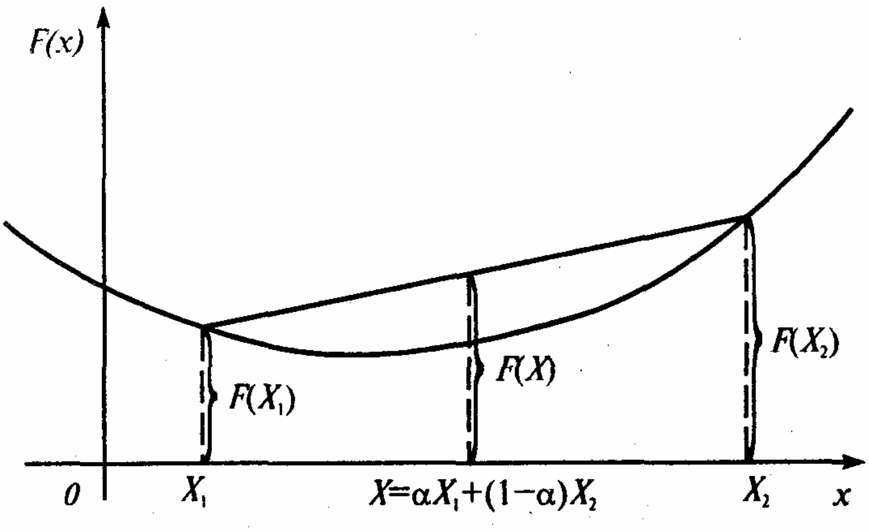


Рис. 11.1

Если в условии (11.2) изменить знак неравенства $\le$ на $\ge$, то получим определение ***вогнутой*** функции. Если же в условии (11.3) неравенство выполняется как строгое, то функция называется ***строго выпуклой*** (или ***строго вогнутой***).

На рис. 11.1 изображен график функции одной переменной, выпуклой на всей числовой прямой.

Для любой пары X_1, X_2 значений аргумента произвольную точку $X \in [X_1, X_2]$ (*см.* (11.2)) можно задать в виде $X = \alpha X_1 + (1-\alpha)X_2$. Как видно из рис. 11.1, неравенство (11.3) означает, что отрезок, соединяющий точки (X_1; $F(X_1)$) и (X_2; $F(X_2)$) расположен не ниже графика функции на этом участке (для строго выпуклой функции этот отрезок лежит выше графика). Таким образом, выпуклость или вогнутость функции одной переменной сразу же видна по ее графику.

▷ **11.2.** Установить выпуклость (вогнутость) функций, приведенных на рис. 11.2.

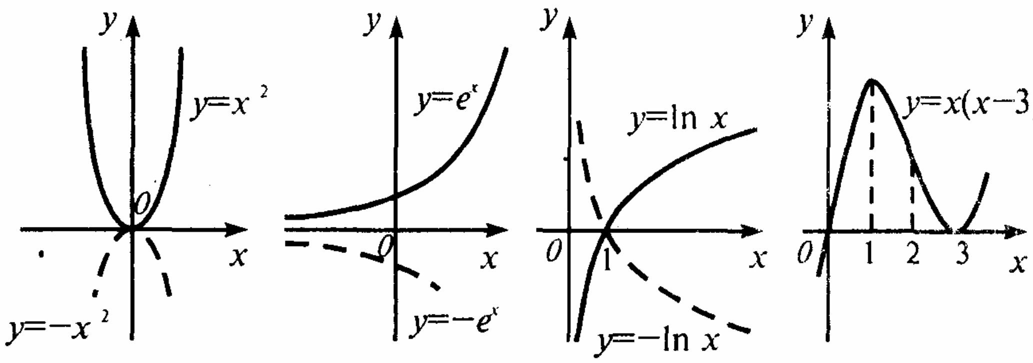


Рис. 11.2

Решение. Функции $y = x^2$ и $y = e^x$ являются всюду строго выпуклыми, а $y = \ln x$ строго вогнута на интервале $(0, \infty)$. Функция $y = x(x-3)^2$ не является ни выпуклой, ни вогнутой на всей числовой прямой, но точка $x = 2$ разбивает ее на два участка: слева от этой точки она вогнута, справа — выпукла (рис. 11.2).▶

Рассмотрим свойства выпуклых функций. Все рассматриваемые функции предполагаются определенными на некотором выпуклом множестве M. Свойства приводятся без доказательств, но многие из них легко проверяются по определению (11.3) и для функций одной переменной иллюстрируются графиками.

Алгебраические и аналитические свойства выпуклых функций:

1. Если функция $F(X)$ выпукла, то функция $-F(X)$ вогнута.

2. Функция $F(X) = C$ и линейная функция $F(X) = aX + b$ являются всюду выпуклыми и всюду вогнутыми.

3. Если функции $F_i(X)$, $i = 1, ..., m$, выпуклы, то при любых действительных числах $\alpha_i \geq 0$ функция $\sum_{i=1}^{m} \alpha_i F_i(X)$ также является выпуклой.

4. Если функция $F(X)$ выпукла, то для любого числа α область решений неравенства $F(X) < \alpha$ является либо выпуклым множеством, либо пустым.

Отметим следствия свойства 4 и теоремы о пересечении выпуклых множеств.

5. Если функции $\varphi_i(X)$ выпуклые при всех неотрицательных значениях переменных, то область решений системы неравенств $\varphi_i(X) \leq b_i$, $i = 1, ..., m$ является выпуклым множеством (если она не пуста).

6. Выпуклая (вогнутая) функция, определенная на выпуклом множестве M, непрерывна в каждой внутренней точке этого множества.

7. Всякая дифференцируемая строго выпуклая (вогнутая) функция имеет не более одной стационарной точки (т.е. точки, в которой равны нулю все частные производные). При этом для выпуклой (вогнутой) функции стационарная точка всегда является точкой локального и глобального минимума (максимума).

8. Дважды дифференцируемая функция $F(X) = F(x_1, ..., x_n)$ является выпуклой в том и только в том случае, когда

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 F(X)}{\partial x_i \partial x_j} \Delta x_i \Delta x_j \geq 0 \qquad \textbf{(11.4)}$$

для любых $X \in M$ и Δx_i, Δx_j, не обращающихся в нуль одновременно.

Чтобы использовать это условие для определения выпуклости конкретной функции, часто бывает полезен критерий Сильвестра[1]: условие (11.4) выполняется тогда и только тогда, когда не-

---

[1] Сильвестр Дж. (1814–1897) — английский математик.

отрицательны все главные миноры Δ_k матрицы вторых частных производных, т.е. определители

$$\Delta_k = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}, \quad a_{ij} = \frac{\partial^2 F(X)}{\partial x_i \partial x_j}, \quad k = 1, \ldots, n. \qquad \mathbf{(11.5)}$$

Если все $\Delta_k > 0$, то неравенство (11.4) выполняется как строгое, и тогда функция F является строго выпуклой.

▷ **11.3**. Показать, что следующие функции являются выпуклыми:

а) $F = 2x_1^2 + x_2^2 - x_1x_2 + 5x_1 - 6x_2 + 8$;

б) $G = -\sqrt{x_1x_2}$.

Решение. а) Находим частные производные: $\frac{\partial F}{\partial x_1} = 4x_1 - x_2 + 5$, $\frac{\partial F}{\partial x_2} = 2x_2 - x_1 - 6$, $\frac{\partial^2 F}{\partial x_1^2} = 4$, $\frac{\partial^2 F}{\partial x_1 \partial x_2} = -1$, $\frac{\partial^2 F}{\partial x_2^2} = 2$. Матрица вторых частных производных имеет вид $A = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$. Ее главные миноры $\Delta_1 = |4| = 4 > 0$, $\Delta_2 = |A| = 7 > 0$. Значит, на основании свойства 8, функция F является строго выпуклой при всех X.

б) Здесь $\frac{\partial G}{\partial x_1} = -\frac{x_2}{2\sqrt{x_1x_2}}$, $\frac{\partial G}{\partial x_2} = -\frac{x_1}{2\sqrt{x_1x_2}}$, $\frac{\partial^2 G}{\partial x_1 \partial x_2} = \frac{1}{4\sqrt{x_1x_2}}$,

$$\frac{\partial^2 G}{\partial x_1^2} = \frac{-x_2}{2}\left[(x_1x_2)^{-\frac{1}{2}}\right]'_{x_1} = \frac{x_2^2}{4(x_1x_2)^{\frac{3}{2}}} = \frac{1}{4x_1}\sqrt{\frac{x_2}{x_1}}, \qquad \frac{\partial^2 G}{\partial x_2^2} = \frac{1}{4x_2}\sqrt{\frac{x_1}{x_2}}.$$

Матрица вторых частных производных имеет вид:

$$A = \begin{pmatrix} \frac{1}{4x_1}\sqrt{\frac{x_2}{x_1}} & \frac{1}{4\sqrt{x_1x_2}} \\ \frac{1}{4\sqrt{x_1x_2}} & \frac{1}{4x_2}\sqrt{\frac{x_1}{x_2}} \end{pmatrix}; \quad \Delta_1 = \frac{1}{4x_1}\sqrt{\frac{x_2}{x_1}},$$

$$\Delta_2 = |A| = \frac{1}{16x_1x_2} - \frac{1}{16x_1x_2} = 0.$$

Таким образом, G является выпуклой, но не строго выпуклой функцией при $x_1 > 0$, $x_2 > 0$.▶

11.2. Задача выпуклого программирования

Пусть дана система неравенств вида

$$\varphi_i(x_1, x_2, \ldots, x_n) \leq b_i,\ i = 1, \ldots, m \qquad \textbf{(11.6)}$$

и функция

$$Z = f(x_1, x_2, \ldots, x_n), \qquad \textbf{(11.7)}$$

причем все функции $\varphi_i(X)$ является выпуклыми на некотором выпуклом множестве M, а функция Z либо выпукла на множестве M, либо вогнута. *Задача выпуклого программирования (ВП) состоит в отыскании такого решения системы ограничений* (11.6), *при котором целевая функция Z достигает минимального значения, или вогнутая функция Z достигает максимального значения.* (Условия неотрицательности переменных можно считать включенными в систему (11.6)).

Ввиду свойства 3 (разд. 11.1) всякая задача линейного программирования является частным случаем задачи ВП. В общем случае задачи ВП являются задачами нелинейного программирования. Выделение задач ВП в специальный класс объясняется экстремальными свойствами выпуклых функций: всякий локальный минимум выпуклой функции (локальный максимум вогнутой функции) является одновременно и глобальным; кроме того, ввиду свойства 2 выпуклая (вогнутая) функция, заданная на замкнутом ограниченном множестве, достигает на этом множестве глобального максимума и глобального минимума. Отсюда вытекает, что *если целевая функция Z является строго выпуклой (строго вогнутой) и если область решений системы ограничений не пуста и ограничена, то задача ВП всегда имеет единственное решение.* В этом случае минимум выпуклой (максимум вогнутой) функции достигается внутри области решений, если там имеется стационарная точка, или на границе этой области, если внутри нее нет стационарной точки. (В общем случае можно утверждать лишь, что множество оптимальных решений любой задачи ВП является выпуклым множеством).

11.4. Геометрически решить следующую задачу ВП: найти минимум функции $Z = 2 + (x_1 - 1)^2 + (x_2 - 1)^2$ при ограничениях:

$$\begin{cases} x_1^2 + x_2^2 \le 4, \\ x_1 \le 2x_2, \\ x_2 \le 2x_1, \\ x_1 \ge 0, x_2 \ge 0. \end{cases}$$

Решение. Строим область допустимых решений данной задачи:

а) $x_1^2 + x_2^2 = 4$ — окружность с центром в начале координат и радиусом $R = 2$. (рис. 11.3). Область решений неравенства $x_1^2 + x_2^2 \le 4$ состоит из точек, лежащих внутри этой окружности и на ней самой;

б) $x_1 = 2x_2$ — прямая, которую можно построить, например, по точкам (0; 0) и (2; 1). Область решений неравенства $x_1 \le 2x_2$ — полуплоскость, лежащая над этой прямой, включая и саму прямую;

в) $x_2 = 2x_1$ — прямая, которая строится, например, по точкам (0; 0) и (1; 2). Область решений неравенства $x_2 \le 2x_1$ — полуплоскость, лежащая под этой прямой, включая и саму прямую. Таким образом, с учетом условий неотрицательности переменных, областью допустимых решений данной задачи является замкнутый сектор OAB (рис. 11.3).

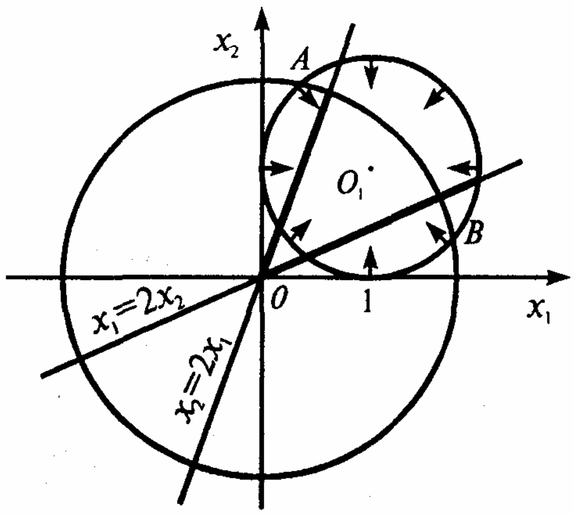


Рис. 11.3

Теперь построим линию уровня функции Z и определим направление убывания Z. Все линии уровня имеют уравнение $Z = C$, т.е. $(x_1 - 1)^2 + (x_2 - 1)^2 = C - 2$. При $C = 3$ получаем линию уровня $(x_1 - 1)^2 + (x_2 - 1)^2 = 1$ — это окружность с центром в точке $O_1(1; 1)$ и радиусом $R = 1$. Ясно, что в любой точке этой линии уровня при перемещении от центра окружности O_1 функция Z возрастает, а при перемещении к центру — убывает. Таким образом, минимум Z достигается в точке (1; 1), $Z_{\min} = 2$ (нетрудно убедиться, что точка (1; 1) является стационарной точкой функции Z).▶

11.3. Приближенное решение задач выпуклого программирования методом кусочно-линейной аппроксимации

Функция $F(X) = F(x_1,..., x_n)$ называется ***сепарабельной***, *если ее можно представить в виде суммы функций, каждая из которых зависит только от одной переменной*, т.е. если

$$F(X) = F_1(x_1) + F_2(x_2)+\ldots+F_n(x_n), \text{ или } F(X) = \sum_{i=1}^{n} F_i(x_i) \quad \textbf{(11.8)}$$

(не исключено, что $F_i(x_i) = 0$ при некоторых i).

Пусть в задаче ВП (11.6), (11.7) и функция цели Z, и все ограничения φ_i являются сепарабельными. Тогда задача имеет вид: *найти минимум выпуклой (максимум — вогнутой) функции* $Z = \sum_{j=1}^{n} f_j(x_j)$ *при ограничениях*

$$\sum_{j=1}^{n} \varphi_{ij}(x_j) \le b_i,\ i = 1, ..., m. \quad \textbf{(11.9)}$$

Идея метода кусочно-линейной аппроксимации состоит в том, что все f_i и все φ_{ij} заменяются ломаными линиями, состоящими из прямолинейных отрезков. При этом исходная задача ВП заменяется новой, приближенной задачей, которая является задачей линейного программирования. Эта задача решается обычно симплексным методом, и ее решение является приближенным решением исходной задачи ВП.

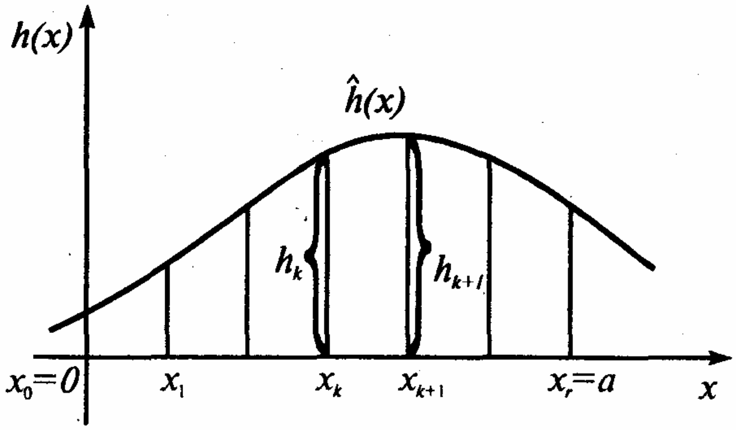


Рис. 11.4

Для построения приближенной задачи рассмотрим кусочно-линейную аппроксимацию функции одной переменной $h(x)$, заданной на отрезке $[0, a]$. Разобьем этот отрезок на r частей точками $x_0 < x_1 < \ldots < x_r$ так, чтобы $x_0 = 0$, $x_r = a$ (рис. 11.4). Вычислим значения функции $h_k(x)$ $(k = 0, \ldots, r)$ в этих точках. Соединим попарно точки $(x_k;\ h_k)$ и $(x_{k+1};\ h_{k+1})$ отрезками прямых. Состоящая из этих отрезков ломаная $\hat{h}(x)$ аппроксимирует функцию $h(x)$ на отрезке $[0, a]$. (Не рассматривая здесь оценку точности приближения, отметим только, что точность можно увеличить за счет более мелкого разбиения отрезка).

Уравнение участка ломаной $\hat{h}(x)$ между точками $(x_k;\ h_k)$ и $(x_{k+1};\ h_{k+1})$ имеет вид $\dfrac{\hat{h}(x) - h_k}{h_{k+1} - h_k} = \dfrac{x - x_k}{x_{k+1} - x_k}$ (уравнение прямой по двум точкам). Если каждое из отношений в этом равенстве обозначить через λ, то получим:

$$\begin{cases} x = \lambda x_{k+1} + (1-\lambda)x_k, \\ \hat{h}(x) = \lambda h_{k+1} + (1-\lambda)h_k, \end{cases} \qquad (11.10)$$

причем $0 \le \lambda \le 1$.

Отметим, что для каждого $x \in [x_k, x_{k+1}]$ существует единственное значение λ, удовлетворяющее условиям (11.10) (*см.* уравне-

ние отрезка (11.2)). Обозначив $1-\lambda = \lambda_k$, $\lambda = \lambda_{k+1}$, можно переписать (11.10) в виде:

$$\begin{cases} x = \lambda_k x_k + \lambda_{k+1} x_{k+1}, \\ \hat{h}(x) = \lambda_k h_k + \lambda_{k+1} h_{k+1}, \end{cases} \qquad \textbf{(11.11)}$$

где $\lambda_k + \lambda_{k+1} = 1$, $\lambda_k \geq 0$, $\lambda_{k+1} \geq 0$.

(Уравнения (11.11) называются *параметрическими уравнениями отрезка*. Если $h(x) \equiv 0$, то второе из этих уравнений обращается в тождество $0 = 0$, а первое принимает вид (11.2) — уравнение отрезка, лежащего на оси абсцисс.)

Таким образом, для любого $x \in [0, a]$ уравнение ломаной можно записать в виде:

$$\begin{cases} x = \sum\limits_{k=0}^{r} \lambda_k x_k, \\ \hat{h}(x) = \sum\limits_{k=0}^{r} \lambda_k h_k, \quad \sum\limits_{k=0}^{r} \lambda_k = 1 \text{ и } \lambda_k \geq 0 \ (k = 0, \ldots, r), \end{cases} \qquad \textbf{(11.12)}$$

причем всегда отличны от нуля только два значения λ_k (если x является внутренней точкой k-го отрезка разбиения) или одно (если x совпадает с концом отрезка).

Возвращаясь к задаче ВП с сепарабельными функциями, отметим, что прежде всего (в зависимости от системы ограничений) нужно определить интервал изменения каждой переменной x_j. Затем каждый этот интервал разбивается на части точками x_{jk} и с использованием формул (11.12) строится кусочно-линейная аппроксимация для функций f_j и φ_{ij}. После этого можно для исходной задачи **(11.9)** записать приближенную задачу:

$$\text{найти максимум функции } \hat{Z} = \sum_{j=1}^{n} \hat{f}_j(x_j)$$

$$\text{при ограничениях } \sum_{j=1}^{n} \hat{\varphi}_{ij}(x_j) \leq b_i, \ i = 1, \ldots, m. \qquad \textbf{(11.13)}$$

$$x_j \geq 0, \ j = 1, \ldots, n.$$

Поскольку приближенная задача (11.13) является задачей линейного программирования и мы обычно решаем ее симплексным методом, условия неотрицательности переменных записываются отдельно от остальных ограничений. Отличие полученной задачи (11.13) от обычной задачи линейного программирования состоит в том, что для каждого x_j имеется не более двух соседних ненулевых λ_{jk} и, значит, нельзя брать в качестве основных переменных два λ_{jk} с одинаковым j и несоседними k. Заметим также, что для условий неотрицательности переменных слагаемых $f_j(x_j)$ и $\varphi_{ij}(x_j)$ (если таковые окажутся) кусочно-линейную аппроксимацию проводить, конечно, не нужно.

▷ **11.5.** Найти минимум функции $Z = 2(x_1 - 1)^2 + (x_2 - 2)^2$ при ограничениях:

$$\begin{cases} x_1^2 + x_2 \le 4, \\ x_1 \ge 0, x_2 \ge 0. \end{cases}$$

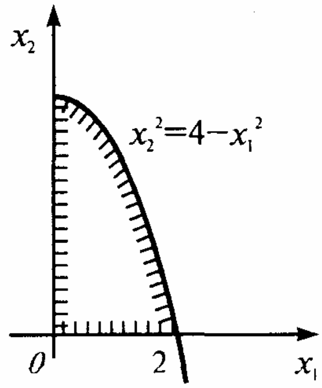


Рис. 11.5

Решить данную задачу методом кусочно-линейной аппроксимации.

Решение. Прежде всего рекомендуем самостоятельно убедиться в том, что данная задача является задачей ВП (используйте критерий Сильвестра, *см.* задачу (11.2) Далее, при условии неотрицательности переменных неравенство $x_1^2 + x_2 \le 4$ показывает, что x_1 может изменяться лишь от 0 до 2, а x_2 — от 0 до 4 (*см.* рис. 11.5).

Отрезок [0; 2] разобьем точками $x_{10} = 0$, $x_{11} = 1$, $x_{12} = 2$, а отрезок [0; 4] — точками $x_{20} = 0$, $x_{21} = 1$, $x_{22} = 2$, $x_{23} = 3$, $x_{24} = 4$. Сравнивая условие данной задачи с (11.9), видим, что

$$f_1(x_1) = 2(x_1 - 1)^2, \; f_2(x_2) = (x_2 - 2)^2.$$

Удобно сначала вычислить необходимые значения этих функций (так как имеем лишь одно ограничение, т.е. $m = 1$, будем писать φ_1 и φ_2 вместо φ_{11} и φ_{12}).

x_1	x_{10}	x_{11}	x_{12}
x_1	0	1	2
φ_1	0	1	4
f_1	2	0	2

x_2	x_{20}	x_{21}	x_{22}	x_{23}	x_{24}
x_2	0	1	2	3	4
φ_2	0	1	2	3	4
f_2	4	1	0	1	4

По формулам (11.12) имеем:

$$x_1 = \sum_{k=0}^{2} \lambda_{1k} x_{1k} = \lambda_{11} + 2\lambda_{12},$$

$$x_2 = \sum_{k=0}^{4} \lambda_{2k} x_{2k} = \lambda_{21} + 2\lambda_{22} + 3\lambda_{23} + 4\lambda_{24},$$

$$\varphi_1 = \sum_{k=0}^{2} \lambda_{1k} \varphi_{1k} = \lambda_{11} + 4\lambda_{12},$$

$$\varphi_2 = \sum_{k=0}^{4} \lambda_{2k} \varphi_{2k} = \lambda_{21} + 2\lambda_{22} + 3\lambda_{23} + 4\lambda_{24},$$

$$f_1 = \sum_{k=0}^{2} \lambda_{1k} f_{1k} = 2\lambda_{10} + 2\lambda_{12},$$

$$f_2 = \sum_{k=0}^{4} \lambda_{2k} f_{2k} = 4\lambda_{20} + \lambda_{21} + \lambda_{23} + 4\lambda_{24},$$

Таким образом, приближенная задача (11.13) для данной задачи ВП имеет вид: найти минимум функции

$$Z = 2\lambda_{10} + 2\lambda_{12} + 4\lambda_{20} + \lambda_{21} + \lambda_{23} + 4\lambda_{24}$$

при ограничениях

$$\begin{cases} \lambda_{11} + 4\lambda_{12} + \lambda_{21} + 2\lambda_{22} + 3\lambda_{23} + 4\lambda_{24} \le 4, \\ \lambda_{10} + \lambda_{11} + \lambda_{12} = 1, \\ \lambda_{20} + \lambda_{21} + \lambda_{22} + \lambda_{23} + \lambda_{24} = 1, \end{cases} \qquad \text{все } \lambda_{ij} \ge 0.$$

Это задача линейного программирования с 8 переменными λ_{ij}. Для решения ее симплексным методом первое ограничение-неравенство нужно преобразовать в уравнение путем введения дополнительной неотрицательной переменной, которую обозначим u:

$$\lambda_{11} + 4\lambda_{12} + \lambda_{21} + 2\lambda_{22} + 3\lambda_{23} + 4\lambda_{24} + u = 4.$$

Тогда система ограничений станет системой трех уравнений с 9 переменными, т.е. 3 переменные нужно взять за основные (берем u, λ_{10} и λ_{20}, так как они входят только в одно уравнение каждая), а остальные 6 являются свободными. Как обычно, на каждом шаге решения основные переменные и функция цели выражаются через свободные.

I шаг.

$$\begin{cases} u = 4 - \lambda_{11} - 4\lambda_{12} - \lambda_{21} - 2\lambda_{22} - 3\lambda_{23} - 4\lambda_{24}, \\ \lambda_{10} = 1 - \lambda_{11} - \lambda_{12}, \\ \lambda_{20} = 1 \qquad - \lambda_{21} - \lambda_{22} - \lambda_{23} - \lambda_{24}, \end{cases}$$

$$Z = 2(1 - \lambda_{11} - \lambda_{12}) + 2\lambda_{12} + 4(1 - \lambda_{21} - \lambda_{22} - \lambda_{23} - \lambda_{24}) +$$
$$+\lambda_{21} + \lambda_{23} + 4\lambda_{24} = 6 - 2\lambda_{11} - 3\lambda_{21} - 4\lambda_{22} - 3\lambda_{23}.$$

Так как в выражении Z имеются свободные переменные с отрицательными коэффициентами, то I базисное решение неоптимальное (хотя и допустимое), и согласно алгоритму симплексного метода λ_{22} следует перевести в основные переменные. Находим $\min\{4/2; \infty; 1/1\} = 1$, выделяем третье уравнение и переменную λ_{20} переводим в свободные переменные.

II шаг. Основные переменные λ_{22}, λ_{10}, u.

$$\begin{cases} \lambda_{22} = 1 - \lambda_{20} - \lambda_{21} - \lambda_{23} - \lambda_{24}, \\ \lambda_{10} = 1 - \lambda_{11} - \lambda_{12}, \\ u = 4 - 4 - \lambda_{11} - 4\lambda_{12} - \lambda_{21} - 2(1 - \lambda_{20} - \lambda_{21} - \lambda_{23} - \lambda_{24}) - 3\lambda_{23} - 4\lambda_{24} = \\ = 2 - \lambda_{11} - 4\lambda_{12} + 2\lambda_{20} + \lambda_{21} - \lambda_{23} - 2\lambda_{24}, \end{cases}$$

$$Z = 6 - 2\lambda_{11} - 3\lambda_{21} - 4(1 - \lambda_{20} - \lambda_{21} - \lambda_{23} - \lambda_{24}) - 3\lambda_{23} =$$
$$= 2 - 2\lambda_{11} + 4\lambda_{20} + \lambda_{21} + \lambda_{23} + 4\lambda_{24}.$$

III шаг. Основные переменные λ_{11}, λ_{22}, u.

$$\begin{cases} \lambda_{11} = 1 - \lambda_{10} - \lambda_{12}, \\ \lambda_{22} = 1 - \lambda_{20} - \lambda_{21} - \lambda_{23} - \lambda_{24}, \\ u = 1 + \lambda_{10} - 3\lambda_{12} + 2\lambda_{20} + \lambda_{21} - \lambda_{23} - 2\lambda_{24}, \end{cases}$$

$$Z = 2\lambda_{10} + 2\lambda_{12} + 4\lambda_{20} + \lambda_{21} + \lambda_{23} + 4\lambda_{24}.$$

Критерий оптимальности симплексного метода выполнен, значит на III шаге получено оптимальное базисное решение:

$$\lambda_{11} = \lambda_{22} = u = 1, \ \lambda_{10} = \lambda_{12} = \lambda_{20} = \lambda_{21} = \lambda_{23} = \lambda_{24} = 0.$$

Переходя к исходным переменным x_1, x_2, получаем:

$$x_1 = \lambda_{11} + 2\lambda_{12} = 1, \ \ x_2 = \lambda_{21} + 2\lambda_{22} + 3\lambda_{23} + 4\lambda_{24} = 2.$$

Таким образом, оптимальное решение приближенной задачи (1; 2) и $Z_{\min} = 0$.▶

Можно было бы геометрически решить исходную задачу ВП и убедиться в том, что оптимальное решение приближенной задачи в точности совпадает с оптимальным решением исходной задачи. Это совпадение случайное. В общем случае полученное решение является лишь некоторым приближением оптимального решения исходной задачи. Улучшить точность приближения можно, разбивая на более мелкие части уже не исходные отрезки изменения переменных, а другие, меньшие, взятые в окрестности полученного первого приближения. Недостатком метода является большое увеличение размерности задачи (т.е. числа переменных) при переходе к приближенной линейной модели.

11.4. Методы спуска. Приближенное решение задач выпуклого программирования градиентным методом

Общая схема решения задач математического программирования методами спуска состоит в построении последовательности

$$X_0, X_1, \ldots, X_k, \ldots \qquad \textbf{(11.14)}$$

решений системы ограничений данной задачи по следующему принципу: в качестве X_0 выбирается, вообще говоря, любая точка области решений и затем каждая последующая точка получается из предыдущей по формуле:

$$X_{k+1} = X_k + \lambda l, \qquad \textbf{(11.15)}$$

где $l = (l_1, \ldots, l_n)$ — некоторое направление (т.е. вектор), а λ — число. При этом направление l и “длина шага” λ выбираются так, чтобы обеспечить сходимость последовательности (11.13) к оптимальному решению X^*. В общем случае процесс получения последовательных приближений X_k бесконечен (и тогда некоторое X_{k_0} берется за приближенное значение оптимального решения X^*), однако иногда процесс может завершиться и за конечное число шагов, приводя к локальному, а в задачах ВП и глобальному оптимуму.

Находя производную по направлению $\frac{\partial Z}{\partial l}$, мы можем определять, является ли направление l "невыгодным" или "выгодным" в смысле приближения к оптимуму.

▷ **11.6.** В задаче ВП нужно найти минимум функции $Z = \frac{1}{x_1+1} + \frac{1}{x_2+1}$ при ограничениях: $x_1 + x_2 \le 5$, $x_1 \ge 0$, $x_2 \ge 0$.

Взяв за X_0 точку (1; 1), проверить, приблизимся ли мы к оптимуму по направлению: а) $l = (2; 1)$; б) $l_1 = (-2; 1)$.

Решение. По критерию Сильвестра (11.5) нетрудно убедиться, что функция Z является выпуклой при $x_1 \ge 0$, $x_2 \ge 0$. Находим $\frac{\partial Z}{\partial x_1} = -\frac{1}{(x_1+1)^2}$, $\frac{\partial Z}{\partial x_2} = -\frac{1}{(x_2+1)^2}$, значит, $\frac{\partial Z}{\partial x_1}(X_0) = -\frac{1}{4}$, $\frac{\partial Z}{\partial x_2}(X_0) = -\frac{1}{4}$.

Отсюда, учитывая, что $|l| = |l_1| = \sqrt{5}$, по формуле (11.1) получаем:

$$\frac{\partial Z}{\partial l}(X_0) = \frac{1}{\sqrt{5}}\left(2\cdot\left(-\frac{1}{4}\right) + 1\cdot\left(-\frac{1}{4}\right)\right) = -\frac{3}{4\sqrt{5}} < 0,$$

$$\frac{\partial Z}{\partial l_1}(X_0) = \frac{1}{\sqrt{5}}\left(-2\cdot\left(-\frac{1}{4}\right) + 1\cdot\left(-\frac{1}{4}\right)\right) = \frac{1}{4\sqrt{5}} > 0.$$

Таким образом, в направлении l функция Z убывает и по этому направлению мы приближаемся к оптимуму, а в направлении l_1 функция возрастает, т.е. мы удаляемся от оптимума.▶

Так как направление градиента ∇Z целевой функции является направлением ее наискорейшего роста, то при отыскании максимума вогнутой функции (минимума выпуклой функции) в качестве l часто берется $\nabla Z\,(-\nabla Z)$ и тогда формула (11.15) принимает вид

$$X_{k+1} = X_k + \lambda \cdot \nabla Z(X_k),\ \lambda > 0 \text{ (если ищется } Z_{\max}) \qquad \textbf{(11.16)}$$

или

$$X_{k+1} = X_k - \lambda \cdot \nabla Z(X_k),\ \lambda > 0 \text{ (если ищется } Z_{\min}) \qquad \textbf{(11.16')}$$

Методы спуска, в которых итерационная последовательность (11.14) находится по формуле (11.16) (или (11.16′)), называются *градиентными*. Друг от друга они отличаются способами выбора длины шага λ и алгоритмами нахождения точки X_{k+1}, если X_k находится на границе области решений и формула (11.16) выводит X_{k+1} за пределы этой области. (Выбор длины шага λ очень важен. Как видно из рис. 11.6, перемещаясь из точки X_0 в направлении ∇Z, мы в некоторый момент можем "проскочить" мимо точки X_1, в которой достигается искомый максимум). Если величина λ выбирается так, чтобы приращение функции ΔZ при перемещении из точки X_k в точку X_{k+1} было наибольшим (при отыскании $Z_{\max}$) или наименьшим (при отыскании $Z_{\min}$), то градиентный метод называется *методом скорейшего спуска*. Таким образом, по методу скорейшего спуска длина шага λ в формуле (11.16) (или (11.16′)) выбирается так, чтобы при этом λ достигался экстремум функции $\Delta Z = Z(X_{k+1}) - Z(X_k)$. (Обратите внимание на то, что при нахождении точки X_{k+1} предыдущая точка X_k считается уже известной, т.е. $Z(X_k)$ и $\nabla Z(X_k)$ являются постоянными величинами, а ΔZ — функцией одной переменной λ). Продифференцировав функцию ΔZ с учетом выражения X_{k+1} по формуле (11.16) и выражения градиента в точке X_k, $\nabla Z(X_k) = \left(\frac{\partial Z}{\partial x_1}(X_k), \frac{\partial Z}{\partial x_2}(X_k), \ldots, \frac{\partial Z}{\partial x_n}(X_k)\right)$, получим, что необходимое условие экстремума $\frac{d\Delta Z}{d\lambda} = 0$ примет вид:

$$\frac{\partial Z}{\partial x_1}(X_{k+1}) \cdot \frac{\partial Z}{\partial x_1}(X_k) + \frac{\partial Z}{\partial x_2}(X_{k+1}) \cdot \frac{\partial Z}{\partial x_2}(X_k) + \ldots$$

$$\ldots + \frac{\partial Z}{\partial x_n}(X_{k+1}) \cdot \frac{\partial Z}{\partial x_n}(X_k) = 0. \qquad \textbf{(11.17)}$$

Ему можно придать более компактную форму, если использовать скалярное произведение векторов:

$$\nabla Z(X_{k+1}) \cdot \nabla Z(X_k) = 0. \qquad \textbf{(11.17')}$$

(Напомним, что скалярное произведение векторов в прямоугольной системе координат равно сумме произведений их соответствующих координат). Например, если $l = (2, -1)$ и $l_1 = (3, 5)$, то $l \cdot l_1 = 2 \cdot 3 + (-1) \cdot 5 = 1$. Скалярное произведение векторов равно 0 тогда и только тогда, когда они ортогональны).

Если оптимум достигается внутри области решений системы ограничений данной задачи ВП, то нет опасности, что точка X_{k+1}, найденная по формуле (11.16) или (11.16′), выйдет за пределы этой области, и длину шага λ определяем по формуле (11.17) без каких-либо дополнительных ограничений. Рассмотрим примеры решения таких задач, а к случаю, когда оптимум достигается на границе области, обратимся позднее. Вместо $\nabla Z(X_0), \nabla Z(X_1)$ и т.д. будем писать $\nabla Z_0, \nabla Z_1$, ... Заметим также, что для различных k длина шага λ в формуле (11.16), вообще говоря, различна, и для строгости нужно было бы писать λ_k, но это сделало бы запись решения более громоздкой.

▷ **11.7.** Используя метод скорейшего спуска, найти максимум функции $Z = 5 - (x_1 - 4)^2 - (x_2 - 5)^2$ при ограничениях:

$$\begin{cases} x_1 + x_2 \le 10, \\ x_1 \cdot x_2 \ge 1, \\ x_1 \ge 0, x_2 \ge 0. \end{cases}$$

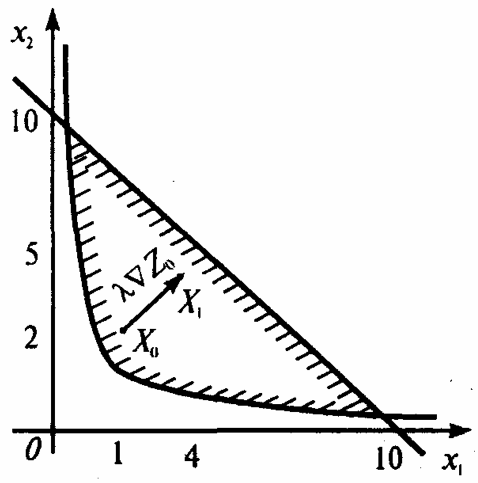


Рис. 11.6

Решение. Функция $-Z$ является выпуклой при любых x_1, x_2 (проверьте это самостоятельно, *см.* (11.4) и пример 11.2), значит, по свойству 1 выпуклых функций Z — вогнутая функция и поэтому ее локальный максимум совпадает с глобальным. Область решений системы ограничений данной задачи изображена на рис. 11.6 (хотя, вообще говоря, ее построение для решения задачи необязательно).

В качестве исходной точки возьмем точку X_0 (1;2). Найдем частные производные функции Z:

$$\frac{\partial Z}{\partial x_1} = 8 - 2x_1, \quad \frac{\partial Z}{\partial x_2} = 10 - 2x_2.$$

Запишем общее выражение градиента

$$\nabla Z = (8 - 2x_1); (10 - 2x_2). \qquad \textbf{(11.18)}$$

Подставляя в (11.18) координаты точки X_0, получим $\nabla Z_0 = (6; 6)$. По формуле (11.16) $X_1 = X_0 + \lambda \nabla Z = (1; 2) + \lambda(6; 6) = (1 + 6\lambda; 2 + 6\lambda)$. Подставляя координаты точки X_1 в (11.18), получаем

$$\nabla Z_1 = (8 - 2(1 + 6\lambda);\ 10 - 2(2 + 6\lambda)) = (6 - 12\lambda;\ 6 - 12\lambda).$$

Теперь можем записать уравнение (11.17′) для нахождения λ: $\nabla Z_0 \cdot \nabla Z_1 = 0$, т.е. $6(6 - 12\lambda) + 6(6 - 12\lambda) = 0$, откуда $1 - 2\lambda = 0$, т.е. $\lambda = \frac{1}{2}$. Подставив этот значение в выражения X_1 и ∇Z_1 через λ (*см.* выше), получаем $X_1 = \left(1 + 6 \cdot \frac{1}{2};\ 2 + 6 \cdot \frac{1}{2}\right) = (4;\ 5)$ и $\nabla Z_1 = \left(6 - 12 \cdot \frac{1}{2};\ 6 - 12 \cdot \frac{1}{2}\right) = (0;\ 0)$. Так как градиент в точке (4;5) равен нулю, то X_1 является точкой максимума, $Z_{max} = 5$. (За один шаг мы достигли точного оптимума. Это не является общей закономерностью; есть задачи, в которых оптимальное решение получается за большее число шагов, и задачи, в которых за любое конечное число шагов мы получаем лишь приближенное решение).▶

Для случая функции двух переменных метод скорейшего спуска имеет простую геометрическую интерпретацию: для любого k луч, идущий от точки X_k к точке X_{k+1}, перпендикулярен к линии уровня функции Z, проходящей через точку X_k (так как направлен по градиенту), и касается линии уровня, проходящей через точку X_{k+1} (так как ввиду условия (11.16′) он перпендикулярен к следующему лучу, который в свою очередь перпендикулярен к этой линии уровня). Таким образом, на плоскости скорейший спуск происходит по двум взаимно перпендикулярным направлениям так, как показано на рис. 11.7.

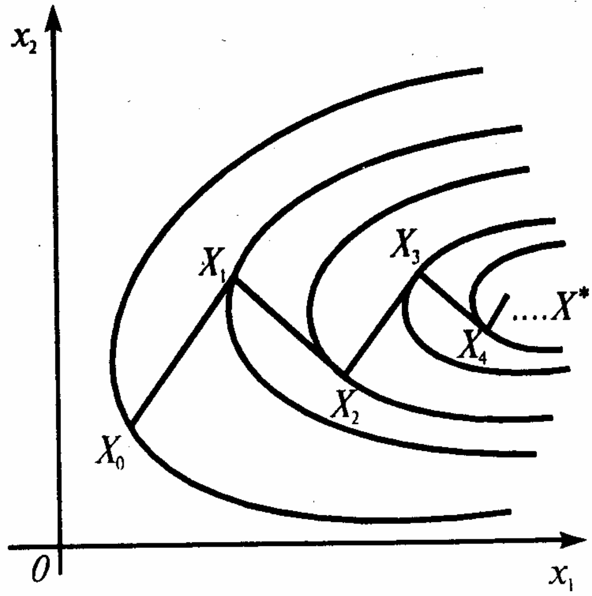


Рис. 11.7

Замечание. Для упрощения счета можно в формулах (11.16) и (11.17′) брать вместо ∇Z_k любой вектор с тем же направлением, т.е. координаты ∇Z_k можно умножать или делить на положительное число.

▷ **11.8.** Найти методом скорейшего спуска с точностью до 0,01 минимум функции $Z = 2x_1^2 - x_1x_2 + x_2^2 - x_1 - x_2 + 1$ при ограничениях $x_1^2 + x_2^2 \le 4$, $x_1 \ge 0$, $x_2 \ge 0$.

Решение. Найдем частные производные $\dfrac{\partial Z}{\partial x_1} = 4x_1 - x_2 - 1$, $\dfrac{\partial Z}{\partial x_2} = -x_1 + 2x_2 - 1$ и общее выражение градиента целевой функции:

$$\nabla Z = (4x_1 - x_2 - 1; -x_1 + 2x_2 - 1).$$

В качестве исходной берем точку (1;1), которая лежит в области решений. Так как Z является выпуклой функцией, то для нахождения точек X_{k+1} будем пользоваться формулой (11.16′), в которой, возможно, вместо ∇Z_k будем брать вектор l_k с тем же направлением, но более простыми координатами (*см.* замечание). Длина шага λ находится по формуле (11.17′), на каждом шаге

производятся аналогичные вычисления, которые подробно объяснены в предыдущем примере.

I шаг.

$X_0 = (1; 1),\ \nabla Z_0 = (4\cdot 1 - 1 - 1; -1 + 2\cdot 1 - 1) = (2; 0).$

$X_1 = X - \lambda\cdot\nabla Z_0 = (1; 1) - \lambda(2; 0) = (1 - 2\lambda; 1).$

$\nabla Z_1 = (4(1 - 2\lambda) - 1 - 1; -1 + 2\lambda + 2\cdot 1 - 1) = (2 - 8\lambda; 2\lambda).$

$$\nabla Z_0\cdot\nabla Z_1 = (2; 0)\cdot(2 - 8\lambda; 2\lambda) = 4 - 16\lambda + 0\cdot 2\lambda = 4 - 16\lambda = 0 \Rightarrow \lambda = \frac{1}{4}.$$

$$X_1 = \left(1 - 2\cdot\frac{1}{4}; 1\right) = \left(\frac{1}{2}; 1\right),\quad \nabla Z_1 = \left(0; \frac{1}{2}\right).$$

II шаг. Вместо ∇Z_1 возьмем $l_1 = (0;1)$. Тогда

$$X_2 = X_1 - \lambda l_1 = \left(\frac{1}{2}; 1\right) - \lambda(0; 1) = \left(\frac{1}{2}; 1 - \lambda\right).$$

$$\nabla Z_2 = \left(4\cdot\frac{1}{2} - (1 - \lambda) - 1; -\frac{1}{2} + 2(1 - \lambda) - 1\right) = \left(\lambda; \frac{1}{2} - 2\lambda\right).$$

$$l_1\cdot\nabla Z_2 = (0; 1)\left(\lambda; \frac{1}{2} - 2\lambda\right) = 0\cdot\lambda + 1\cdot\left(\frac{1}{2} - 2\lambda\right) = \frac{1}{2} - 2\lambda = 0 \Rightarrow \lambda = \frac{1}{4}.$$

$$X_2 = \left(\frac{1}{2}; 1 - \frac{1}{4}\right) = \left(\frac{1}{2}; \frac{3}{4}\right),$$

$$\nabla Z_2 = \left(4\cdot\frac{1}{2} - \frac{3}{4} - 1; -\frac{1}{2} + 2\cdot\frac{3}{4} - 1\right) = \left(\frac{1}{4}; 0\right).$$

III шаг. Вместо ∇Z_2 возьмем $l_2 = (1; 0)$. Тогда

$$X_3 = X_2 - \lambda l_2 = \left(\frac{1}{2}; \frac{3}{4}\right) - \lambda(1; 0) = \left(\frac{1}{2} - \lambda; \frac{3}{4}\right).$$

$$\nabla Z_3 = \left(4\left(\frac{1}{2} - \lambda\right) - \frac{3}{4} - 1; -\frac{1}{2} + \lambda + 2\cdot\frac{3}{4} - 1\right) = \left(\frac{1}{4} - 4\lambda; \lambda\right).$$

$$l_2\cdot\nabla Z_3 = 1\cdot\left(\frac{1}{4} - 4\lambda\right) + 0\cdot\lambda = \frac{1}{4} - 4\lambda = 0,\ \text{откуда}\ \lambda = \frac{1}{16}.$$

$$X_3 = \left(\frac{1}{2} - \lambda; \frac{3}{4}\right) = \left(\frac{7}{16}; \frac{3}{4}\right) = (0{,}4375; 0{,}75)\ \text{и}$$

$$\nabla Z_3 = \left(\frac{1}{4} - 4\lambda; \lambda\right) = \left(0; \frac{1}{16}\right).$$

IV шаг. Берем $l_3 = (0;\ 1)$. Тогда

$$X_4 = X_3 - \lambda l_3 = \left(\frac{7}{16};\ \frac{3}{4}\right) - \lambda(0;\ 1) = \left(\frac{7}{16};\ \frac{3}{4} - \lambda\right).$$

$$\nabla Z_4 = \left(\frac{7}{4} - \frac{3}{4} + \lambda - 1;\ -\frac{7}{16} + \frac{3}{2} - 2\lambda - 1\right) = \left(\lambda;\ \frac{1}{16} - 2\lambda\right).$$

$$l_3 \cdot \nabla Z_4 = 0 \cdot \lambda + 1 \cdot \left(\frac{1}{16} - 2\lambda\right) = 0, \text{ откуда } \lambda = \frac{1}{32}.$$

$$X_4 = \left(\frac{7}{16};\ \frac{3}{4} - \lambda\right) = \left(\frac{7}{16};\ \frac{23}{32}\right) = (0{,}4375;\ 0{,}71875).$$

V шаг. Берем $l_4 = (1;\ 0)$. Тогда

$$X_5 = X_4 - \lambda l_4 = \left(\frac{7}{16};\ \frac{23}{32}\right) - \lambda(1;\ 0) = \left(\frac{7}{16} - \lambda;\ \frac{23}{32}\right).$$

$$\nabla Z_5 = \left(\frac{7}{4} - 4\lambda - \frac{23}{32} - 1;\ -\frac{7}{16} + \lambda + \frac{23}{16} - 1\right) = \left(\frac{1}{32} - 4\lambda;\ \lambda\right).$$

$$l_4 \cdot \nabla Z_5 = 1 \cdot \left(\frac{1}{32} - 4\lambda\right) + 0 \cdot \lambda = \frac{1}{32} - 4\lambda = 0,$$

откуда $\lambda = \dfrac{1}{128}$ и $X_5 = \left(\dfrac{55}{128};\ \dfrac{23}{32}\right)$.

Таким образом, $X_5 = (0{,}4296875;\ 0{,}71875)$. Сравнение X_4 и X_5 показывает, что координаты этих точек отличаются меньше, чем на 0,01, и поэтому (правда, не совсем строго) можно положить $X^* \approx (0{,}43;\ 0{,}72)$. Нетрудно убедиться, что все точки $X_0, X_1, \ldots, X_5, X^*$ лежат в области решений системы ограничений.▶

Рассмотрим теперь задачу ВП, когда *оптимум целевой функции достигается на границе области решений системы ограничений.* В этом случае, взяв, как и ранее, в качестве исходной точки X_0 любую точку из области решений и находя последующие точки по формуле (11.16) или (11.16′), мы на некотором шаге получим, что X_k уже не лежит в области решений (*см.* рис. 11.8 *а*). Тогда вместо X_k берем точку X'_k, которая лежит на пересечении направления спуска с границей области решений, а все последующие точки находятся путем проектирования на эту границу точек, получаемых обычным методом скорейшего спуска. Поскольку общий оператор проектирования не изучается в нашем курсе, ограничимся случаем, когда система ограничений линейная, т.е. область решений задачи для случая двух переменных ограничена отрезками прямых (рис. 11.8, *б*).

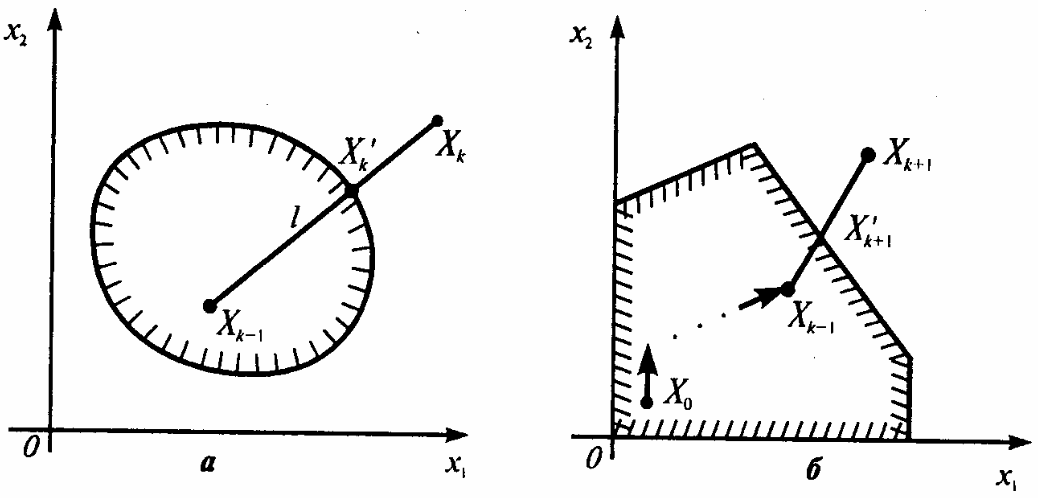


Рис. 11.8

В этом случае система ограничений данной задачи примет вид

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i, \quad i = 1, \dots, m. \tag{11.19}$$

Пусть по методу скорейшего спуска мы построили точки X_0, ..., X_k, X_{k+1} и убедились (подставляя в (11.19) координаты этих точек), что X_1, ..., X_k лежат в области решений, а точка X_{k+1} уже не лежит в ней. Значит, координаты точки X_{k+1} не удовлетворяют хотя бы одному неравенству системы (11.19), например неравенству $\sum_{j=1}^{n} a_{ij}x_j \le b_i$ (здесь i — номер конкретного неравенства).

Пусть $X_k = (x_1^{(k)}, \dots, x_n^{(k)})$, $X_{k+1} = (x_1^{(k+1)}, \dots, x_n^{(k+1)})$ и $X_{k+1} = X_k + \lambda l$, т.е.

$$x_j^{(k+1)} = x_j^{(k)} + \lambda l_j, \; j = 1, \dots, n. \tag{11.20}$$

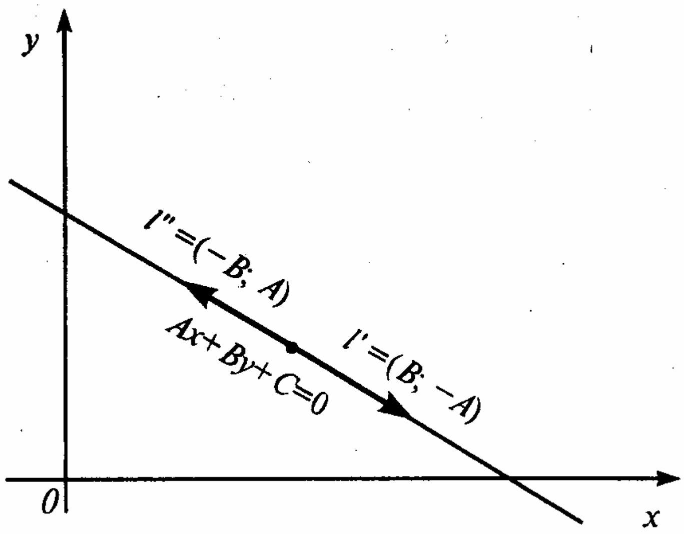


Рис. 11.9

Здесь $l = (l_1, \ldots, l_n)$ — направление спуска, которое совпадает с направлением $\nabla Z(X_n)$ при отыскании максимума или с направлением $-\nabla Z(X_n)$ при отыскании минимума. Чтобы остаться в пределах области решений, следует вместо точки X_{k+1} взять точку, лежащую на том же направлении спуска. т.е. с координатами. удовлетворяющими условию (11.20), но с меньшей длиной шага λ. Значение λ нужно выбирать так, чтобы точка оказалась на границе области решений, т.е. чтобы выполнялось равенство $\sum_{j=1}^{n} a_{ij} x_j^{(k+1)} = b_j$. Учитывая (11.20), получаем $\sum_{j=1}^{n} a_{ij}(x_j^{(k)} + \lambda l_j) = b_j$ и $\sum_{j=1}^{n} a_{ij} x_j^{(k)} - b_i + \lambda \sum_{j=1}^{n} a_{ij} l_j = 0$. Отсюда

$$\lambda = -\frac{\sum_{j=1}^{n} a_{ij} x_j^{(k)} - b_i}{\sum_{j} a_{ij} l_j}. \qquad \textbf{(11.21)}$$

Формула (11.21) дает значение λ, при котором направление спуска пересекает границу области решений. Если координаты точки X_{k+1} не удовлетворяют нескольким неравенствам системы (11.19), то нужно найти λ по формуле (11.21) для каждого из этих неравенств и взять наименьшее из найденных значений. Подставляя его в (11.20), найдем координаты точки X'_{k+1}, которая будет исходной для следующего шага решения. Так как проекция градиента $\nabla Z(X'_{k+1})$ задачи с линейными ограничениями лежит на границе области решений, а для случая двух переменных — просто на прямой $a_{i1}x_1 + a_{i2}x_2 = b_i$, то направление спуска l на следующем шаге берется на этой границе. Оптимум достигается в той точке, в которой градиент перпендикулярен границе области решений.

Замечание. Если прямая на плоскости имеет уравнение $Ax + By + C = 0$, то ее направление задается вектором $l' = (B; -A)$ или вектором $l'' = (-B; A)$ (рис. 11.9).

▷ **11.9.** Используя метод скорейшего спуска, найти максимум функции $Z = 5x_1 - \frac{1}{2}x_1^2 + x_1x_2 - x_2^2$ при ограничениях:

$$\begin{cases} x_1 + 2x_2 \le 10, \\ x_1 + x_2 \le 6, \\ x_1 \ge 0, x_2 \ge 0. \end{cases}$$

Решение. Отметим, что функция Z вогнутая, т.е. данная задача является задачей ВП. Находим частные производные $\frac{\partial Z}{\partial x_1}, \frac{\partial Z}{\partial x_2}$ функции Z и записываем общее выражение ее градиента:

$$\nabla Z = (5 - x_1 + x_2;\ x_1 - 2x_2).$$

I шаг. В качестве исходной возьмем точку $X_0 = (1; 1)$ (ее координаты удовлетворяют системе ограничений). Далее, используя формулы (11.15) и (11.16′), так же, как и в задачах **11.6** и **11.7**, находим точку X_1: $\nabla Z_0 = l_0 = (5; -1)$. Тогда

$$X_1 = X_0 + \lambda l_0 = (1; 1) + \lambda(5; -1) = (1 + 5\lambda; 1 - \lambda),$$

$$\nabla Z_1 = (5 - 1 - 5\lambda + 1 - \lambda; 1 + 5\lambda - 2 + 2\lambda) = (5 - 6\lambda; -1 + 7\lambda),$$

$$l_0 \cdot \nabla Z_1 = 5(5 - 6\lambda) + (-1)(-1 + 7\lambda) = 26 - 37\lambda = 0.$$

Отсюда $\lambda = \frac{26}{37}$, т.е. $X_1 = \left(1 + 5 \cdot \frac{26}{37}; 1 - \frac{26}{37}\right) = \left(\frac{167}{37}; \frac{11}{37}\right) \approx (4{,}5; 0{,}3)$.

Так как 4,5 +0,6 < 10 и 4,5 + 0,3 < 6, то X_1 находится внутри области решений (рис. 11.10).

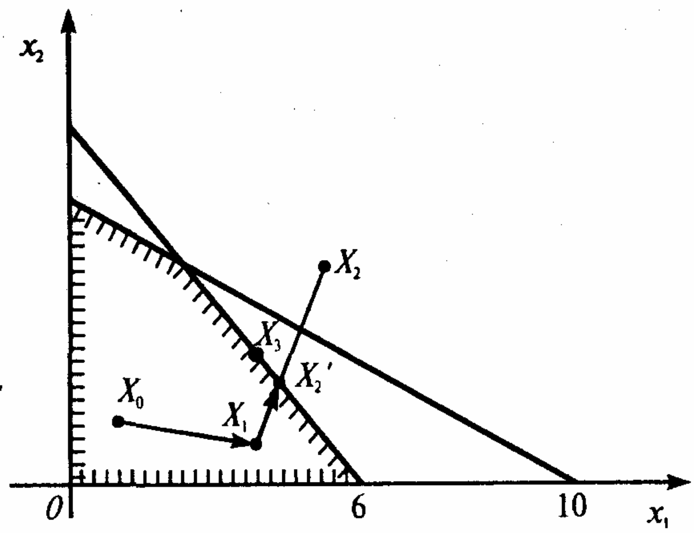


Рис. 11.10

II шаг: Находим $\nabla Z_1 = \left(5 - \frac{167}{37} + \frac{11}{37}; \frac{167}{37} - 2 \cdot \frac{11}{37}\right)$. Можно не производить этих вычислений. Достаточно убедиться, что первая компонента положительна, и тогда взять $l_1 = (1;\ 5)$. (В самом деле, $l_1 \perp l_0$, так как $l_1 \cdot l_0 = 1 \cdot 5 + 5 \cdot (-1) = 0$, и l_1 имеет то же направление, что ∇Z_1).

Теперь

$$X_2 = X_1 + \lambda l_0 = \left(\frac{167}{37}; \frac{11}{37}\right) + \lambda(1;5) = \left(\frac{167}{37} + \lambda; \frac{11}{37} + 5\lambda\right),$$

$$\nabla Z_2 = \left(5 - \frac{167}{37} - \lambda + \frac{11}{37} + 5\lambda; \frac{167}{37} + \lambda - \frac{22}{37} - 10\lambda\right) =$$

$$= \left(\frac{29}{37} + 4\lambda; \frac{145}{37} - 9\lambda\right).$$

$l_1 \cdot \nabla Z_2 = 1 \cdot \left(\frac{29}{37} + 4\lambda\right) + 5 \cdot \left(\frac{145}{37} - 9\lambda\right) = 0$, откуда $41\lambda = \frac{754}{37}$,

т.е. $\lambda \approx 0{,}50$ и $X_2 \approx (5{,}01;\ 2{,}78)$. Мы видим, что X_2 лежит уже вне области решений, причем не выполняются первое и вто-

рое неравенства. По формуле (11.21) находим λ для обоих неравенств:

$$\lambda' = -\frac{167/37 + 2\cdot 11/37 - 10}{1+2\cdot 5} \approx 0{,}445\,,$$

$$\lambda'' = -\frac{167/37 + 11/37 - 6}{1+5} = \frac{22}{111} \approx 0{,}2\,.$$

Берем $\lambda = \min\{\lambda', \lambda''\} = \dfrac{22}{111}$ и теперь получаем $X_2' = X_1 + \lambda l_1 = = \left(\dfrac{167}{37}; \dfrac{11}{37}\right) + \lambda(1;\, 5) = \left(\dfrac{523}{111}; \dfrac{143}{111}\right)$. Нетрудно убедиться, что эта точка лежит на границе области $x_1 + x_2 = 6$.

III шаг. Попав на границу области, мы должны двигаться по ней в сторону увеличения функции Z. Ввиду замечания, сделанного перед этим примером, направление прямой $x_1 + x_2 = 16$ задается вектором $l' = (1;\, -1)$ или $l'' = (-1;\, 1)$. По формуле (11.1) находим производную функции Z по направлению l' в точке X_2:

$$\frac{\partial Z}{\partial l'}(X_2') = \frac{\partial Z}{\partial x_1}(X_2')\cdot 1 + \frac{\partial Z}{\partial x_2}(X_2')\cdot(-1) =$$

$$= \left(5 - \frac{523}{111} + \frac{143}{111}\right) - \left(\frac{523}{111} - 2\cdot\frac{143}{111}\right) < 0.$$

Значит, в этом направлении функция Z убывает и в качестве направления спуска следует взять $l'' = (-1;\, 1)$.

Итак, $l_2 = (-1;\, 1)$, $X_3 = X_2' + \lambda l_2 = \left(\dfrac{523}{111}; \dfrac{143}{111}\right) + \lambda(-1;\, 1) = = \left(\dfrac{523}{111} - \lambda;\, \dfrac{143}{111} + \lambda\right)$. Далее поступаем, как и на предыдущих шагах:

$$\nabla Z_3 = \left(5 - \frac{523}{111} + \lambda + \frac{143}{111} + \lambda;\, \frac{523}{111} - \lambda - \frac{286}{111} - 2\lambda\right) =$$

$$= \left(\frac{175}{111} + 2\lambda;\, \frac{237}{111} - 3\lambda\right).$$

$$l_2 \cdot \nabla Z_3 = 0 \to \left(\frac{175}{111} + 2\lambda\right) \cdot (-1) + \left(\frac{237}{111} - 3\lambda\right) \cdot 1 = \frac{62}{111} - 5\lambda = 0.$$

Отсюда $\lambda = \dfrac{62}{555}$ и $X_3 = \left(\dfrac{523}{111} - \dfrac{62}{555}; \dfrac{143}{111} + \dfrac{62}{155}\right) = (4{,}6;\ 1{,}4)$. Так как $(4{,}6 + 2{,}8) < 10$, то точка X_3 является решением данной задачи.

Поскольку $\dfrac{\partial Z}{\partial l_2}(X_3) = (5 - 4{,}6 + 1{,}4) \cdot (-1) + (4{,}6 - 2{,}8) \cdot 1 = 0$, проекция градиента на это направление равна 0, значит, мы достигли оптимального значения функции при данных ограничениях.

Таким образом, $X^* = (4{,}6;\ 1{,}4)$ и $Z_{\max} = 5 \cdot 4{,}6 - \dfrac{1}{2} \cdot (4{,}6)^2 + {}$ $+ 4{,}6 \cdot 1{,}4 - (1{,}4)^2 = 16{,}9.$ ▶

11.5. Понятие о параметрическом и стохастическом программировании

Параметрическое программирование. В экономической практике нередко возникают задачи, в математических моделях которых коэффициенты линейной формы или системы ограничений (или те и другие) не являются постоянными числами, а меняются в зависимости от некоторых параметров. Например, в задаче об оптимальном использовании ресурсов (оптимальном планировании производства) прибыль от реализации (или цена) продукции может носить сезонный характер и являться функцией времени, а запасы ресурсов и технологические коэффициенты (выражающие размеры их потребления на единицу продукции каждого вида) могут изменяться в зависимости от времени, технологии производства, вместимости складских помещений и т. п.

Параметрическое программирование рассматривает экстремальные задачи с целевыми функциями и ограничениями, зависящими от параметров, разрабатывает методы нахождения оптимальных решений для совокупностей значений параметров и изучает поведение оптимальных планов этих задач при изменении параметров.

Наиболее простой и хорошо изученной является *задача линейного параметрического программирования* с одним параметром, от которого зависят только коэффициенты целевой функции. Эта задача состоит в максимизации линейной функции $F_t = \sum_{j=1}^{n}(c_j + d_j t)x_j$ при ограничениях $\sum_{j=1}^{n} a_{ij}x_j = b_i \quad (i = \overline{1,m})$ и условии неотрицательности переменных $x_j \geq 0 \quad (j = \overline{1,n})$, где c_j, d_j, a_{ij}, b_i — заданные постоянные, а t — параметр, изменяющийся в пределах от α до β: $t \in [\alpha;\beta]$.

В результате решения задачи (если таковое существует) отрезок $[\alpha;\beta]$ разбивается на конечное число отрезков значений параметра t таким образом, чтобы для каждого из них максимальное значение линейной функции F_t достигалось в одной и той же вершине многогранника решений. Тем самым для каждого промежутка значений параметра находится оптимум и оптимальное решение.

Стохастическое программирование представляет собой совокупность методов решения оптимизационных задач вероятностного (стохастического) характера. Задача об оптимальном использовании ресурсов, транспортная задача и т. п. становятся задачами стохастического программирования, если параметры целевой функции либо системы ограничений (или и те и другие) рассматривать как случайные величины. В стохастической постановке эти задачи будут полнее отображать экономическую действительность.

При решении стохастических задач проще всего найти средние значения всех случайных параметров и свести такие задачи к обычным, детерминированным задачам математического программирования. Однако такой подход не всегда эффективен, так как при некоторых реализациях случайных величин (параметров) можно прийти к решению, далекому от оптимального, или даже к отсутствию решений задачи.

Другой подход состоит в том, что на первом этапе устанавливается предварительный оптимальный план на основе решения детерминированной задачи, который и реализуется на этом этапе. Затем на втором (последующих) этапе этот план корректируется в соответствии с реальными статистическими характеристиками параметров. Так поступают, например, при решении задачи об оптимальном использовании ресурсов, транспортной задачи при неопределенном спросе на продукцию.

УПРАЖНЕНИЯ

11.10. Исследовать выпуклость следующих функций:

а) $y = \dfrac{2}{x}$ при $x > 0$;

б) $Z = e^{2x_1 - x_2}$;

в) $Z = 5 - x_1^2 - x_2^2$;

г) $Z = -x_1 x_2 + 2x_1$;

д) $Z = -\dfrac{1}{x_1} - \dfrac{1}{x_2}$ при $x_1 > 0$, $x_2 > 0$.

11.11. Решить геометрически задачу выпуклого программирования:

$$Z = 2 - x_1^2 - x_2^2 \to \max$$

$$\begin{cases} x_2 \le 4 - x_1^2, \\ x_1 + x_2 \ge 1, \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

Задачи выпуклого программирования **11.12**, **11.13** решить методом кусочно-линейной аппроксимации.

11.12. $Z = (x_1 - 3)^2 + 2(x_2 - 2)^2 \to \min$

$$\begin{cases} x_1 + 4x_2 \le 16, \\ 3x_1 + x_2 \le 15, \\ x_1 \ge 0, x_2 \ge 0. \end{cases}$$

Указание: Отрезок изменения переменной x_1 [0; 5] разбить на 5 частей, а отрезок переменной x_2 [0; 4] разбить на 4 части.

11.13. $Z = x_2 - x_1^2 \to \max$

$$\begin{cases} 2x_1 + 3x_2^2 \le 3, \\ 0 \le x_1 \le 2/3, \\ x_2 \ge 0. \end{cases}$$

Указание: Отрезок изменения каждой из переменных разбить на 4 части.

11.14. Для данной функции $Z = x_1 x_2$ найти:

а) общее выражение градиента функции ∇Z;

б) линию уровня, проходящую через точку $A(4; 1)$ и градиент $\nabla Z(A)$, и изобразить их на чертеже;

в) производную по направлению $l = (-1; -1)$.

11.15. Решить задачу **11.14** при условии, что $Z = x_1^2 - x_2 + 2$.

Глава 12. МОДЕЛИ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

12.1. Общая постановка задачи динамического программирования

Динамическое программирование (ДП) — метод оптимизации, приспособленный к операциям, в которых процесс принятия решения может быть разбит на этапы (шаги). Такие операции называются *многошаговыми.* Начало развития ДП относится к 50-м годам XX в. Оно связано с именем Р.Беллмана[1].

Если модели линейного программирования можно использовать в экономике для принятия крупномасштабных плановых решений в сложных ситуациях, то модели ДП применяются при решении задач значительно меньшего масштаба, например, при разработке правил управления запасами, устанавливающими момент пополнения запасов и размер пополняющего заказа; при разработке принципов календарного планирования производства и выравнивания занятости в условиях колеблющегося спроса на продукцию; при распределении дефицитных капитальных вложений между возможными новыми направлениями их использования; при составлении календарных планов текущего и капитального ремонта сложного оборудования и его замены; при разработке долгосрочных правил замены выбывающих из эксплуатации основных фондов и т. п.

В реально функционирующих больших экономических системах еженедельно требуется принимать микроэкономические ре-

---

[1] Беллман Р.Э. (р.1920 г.) — американский математик.

шения. Модели ДП ценны тем, что позволяют на основе стандартного подхода с использованием при минимальном вмешательстве человека принимать такие решения. И если каждое взятое в отдельности такое решение малосущественно, то в совокупности эти решения могут оказать большое влияние на прибыль.

Приведем *общую постановку задачи ДП.* Рассматривается управляемый процесс, например, экономический процесс распределения средств между предприятиями, использования ресурсов в течение ряда лет, замены оборудования, пополнения запасов и т. п. В результате управления система (объект управления) S переводится из начального состояния s_0 в состояние $\hat{s}$. Предположим, что управление можно разбить на n шагов, т.е. решение принимается последовательно на каждом шаге, а управление, переводящее систему S из начального состояния в конечное, представляет собой совокупность n пошаговых управлений.

Обозначим через X_k управление на k-м шаге (k=1, 2, ..., n). Переменные X_k удовлетворяют некоторым ограничениям и в этом смысле называются *допустимыми* (X_k может быть числом, точкой в n-мерном пространстве, качественным признаком).

Пусть $X(X_1, X_2, ..., X_n)$ — управление, переводящее систему S из состояния s_0 в состояние $\hat{s}$. Обозначим через s_k состояние системы после k-го шага управления. Получаем последовательность состояний s_0, s_1, ..., s_{k-1}, s_k,..., s_{n-1}, $s_n = \hat{s}$, которую изобразим кружками (рис. 12.1).

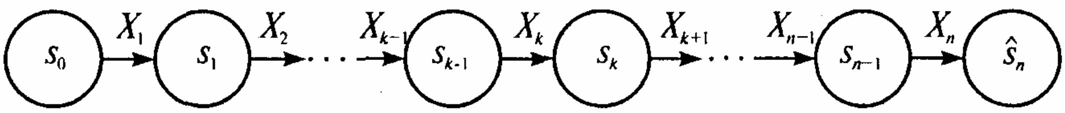


Рис. 12.1

Показатель эффективности рассматриваемой управляемой операции — целевая функция — зависит от начального состояния и управления:

$$Z = F(s_0, X). \qquad \textbf{(12.1)}$$

Сделаем несколько предположений.

1. Состояние s_k системы в конце k-го шага зависит только от предшествующего состояния s_{k-1} и управления на k-м шаге X_k

(и не зависит от предшествующих состояний и управлений). Это требование называется “отсутствием последействия”. Сформулированное положение записывается в виде уравнений

$$s_k = \varphi_k(s_{k-1}, X_k),\ k = 1, 2, ..., n, \qquad \textbf{(12.2)}$$

которые называются *уравнениями состояний.*

2. Целевая функция (12.1) является аддитивной от показателя эффективности каждого шага [3]. Обозначим показатель эффективности k-го шага через

$$Z_k = f_k\,(s_{k-1}, X_k),\ k=1, 2, ..., n; \qquad \textbf{(12.3)}$$

тогда

$$Z = \sum_{k=1}^{n} f_k\ (s_{k-1}, X_k\,). \qquad \textbf{(12.4)}$$

Задача пошаговой оптимизации (задача ДП) формулируется так: *определить такое допустимое управление X, переводящее систему S из состояния s_0 в состояние $\hat{s}$, при котором целевая функция* (12.4) *принимает наибольшее* (*наименьшее*) *значение.*

Выделим особенности модели ДП:

1. *Задача оптимизации интерпретируется как n-шаговый процесс управления.*

2. *Целевая функция равна сумме целевых функций каждого шага.*

3. *Выбор управления на k-м шаге зависит только от состояния системы к этому шагу, не влияет на предшествующие шаги* (*нет обратной связи*).

4. *Состояние s_k после k-го шага управления зависит только от предшествующего состояния s_{k-1} и управления X_k* (*отсутствие последействия*).

5. *На каждом шаге управление X_k зависит от конечного числа управляющих переменных, а состояние s_k — от конечного числа параметров* (смысл замечания 5 станет ясным из рассмотренных ниже примеров).

Следует вспомнить, что существуют различные способы решения подобных задач, применяемые в зависимости от вида функций, ограничений, размерности и т. п. Рассмотрим вычислительную схему ДП, которая окажется безразличной к способам задания функций и ограничений. Вычислительная схема связана с принципом оптимальности и использует рекуррентные соотношения.

12.2. Принцип оптимальности и уравнения Беллмана

Принцип оптимальности впервые был сформулирован Р. Беллманом в 1953 г. *Каково бы ни было состояние s системы в результате какого-либо числа шагов, на ближайшем шаге нужно выбирать управление так, чтобы оно в совокупности с оптимальным управлением на всех последующих шагах приводило к оптимальному выигрышу на всех оставшихся шагах, включая данный*[1]. Беллманом четко были сформулированы и условия, при которых принцип верен. Основное требование — процесс управления должен быть без обратной связи, т.е. управление на данном шаге не должно оказывать влияния на предшествующие шаги.

Принцип оптимальности утверждает, что для любого процесса без обратной связи оптимальное управление таково, что оно является оптимальным для любого подпроцесса по отношению к исходному состоянию этого подпроцесса. Поэтому решение на каждом шаге оказывается наилучшим с точки зрения управления в целом. Если изобразить геометрически оптимальную траекторию в виде ломаной линии, то любая часть этой ломаной будет являться оптимальной траекторией относительно начала и конца.

Уравнения Беллмана. Вместо исходной задачи ДП (см. разд. 12.1) с фиксированным числом шагов n и начальным состоянием s_0 рассмотрим последовательность задач, полагая последовательно n=1, 2, ... при различных s — одношаговую, двухшаговую и т.д., — используя принцип оптимальности.

Введем ряд новых обозначений. Обозначения в ДП несут большую информационную нагрузку, поэтому очень важно их четко усвоить.

На каждом шаге любого состояния системы s_{k-1} решение X_k нужно выбирать "с оглядкой", так как этот выбор влияет на последующее состояние s_k и дальнейший процесс управления, зависящий от s_k. Это следует из принципа оптимальности.

Но есть один шаг, последний, который можно для любого состояния s_{n-1} планировать локально-оптимально, исходя только из соображений этого шага.

---

[1] **Формулировка принадлежит Е.С.Вентцель [3] и немного отличается от предложенной Беллманом.**

Рассмотрим n-й шаг: s_{n-1} — состояние системы к началу n-го шага, $s_n = \hat{s}$ — конечное состояние, X_n — управление на n-м шаге, а $f_n(s_{n-1}, X_n)$ — целевая функция (выигрыш) n-го шага.

Согласно принципу оптимальности, X_n нужно выбирать так, чтобы для любых состояний s_{n-1} получить максимум[1] целевой функции на этом шаге.

Обозначим через $Z_n^*(s_{n-1})$ максимум целевой функции — показателя эффективности n-го шага при условии, что к началу последнего шага система S была в произвольном состоянии s_{n-1}, а на последнем шаге управление было оптимальным.

$Z_n^*(s_{n-1})$ *называется условным максимумом целевой функции на n-м шаге*. Очевидно, что

$$Z_n^*(s_{n-1}) = \max_{\{X_n\}} f_n(s_{n-1}, X_n). \qquad \textbf{(12.5)}$$

Максимизация ведется по всем допустимым управлениям X_n.

Решение X_n, при котором достигается $Z_n^*(s_{n-1})$, также зависит от s_{n-1} и называется *условным оптимальным управлением* на *n-м шаге*. Оно обозначается через $X_n^*(s_{n-1})$.

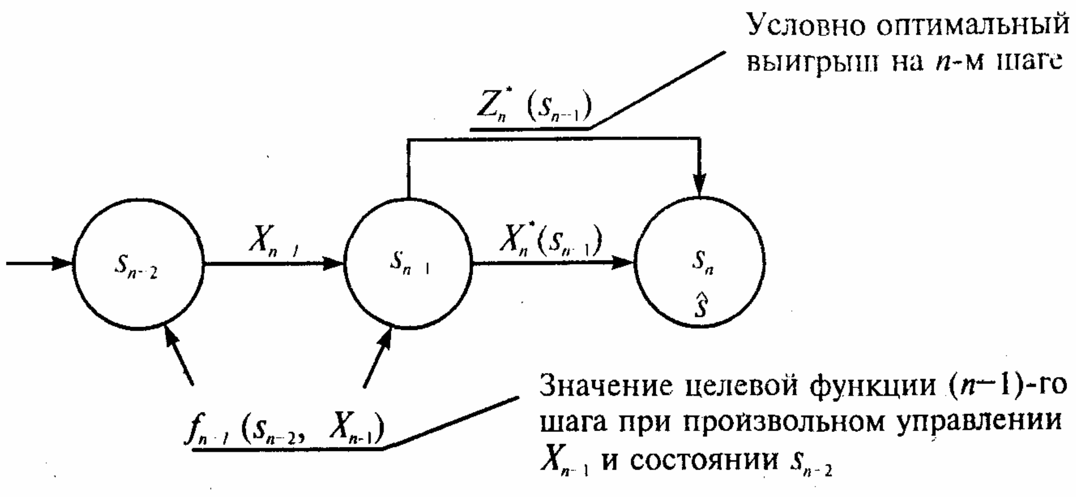


Рис. 12.2

---

[1] Ограничимся здесь задачей максимизации целевой функции.

Решив одномерную задачу локальной оптимизации по уравнению (12.5), найдем для всех возможных состояний s_{n-1} две функции: $Z_n^*(s_{n-1})$ и $X_n^*(s_{n-1})$.

Рассмотрим теперь двухшаговую задачу: присоединим к n-му шагу $(n-1)$-й (рис. 12.2).

Для любых состояний s_{n-2}, произвольных управлений X_{n-1} и оптимальном управлении на n-м шаге значение целевой функции на двух последних шагах равно:

$$f_{n-1}(s_{n-2}, X_{n-1}) + Z_n^*(s_{n-1}). \qquad \textbf{(12.6)}$$

Согласно принципу оптимальности для любых s_{n-2} решение нужно выбирать так, чтобы оно вместе с оптимальным управлением на последнем (n-м) шаге приводило бы к максимуму целевой функции на двух последних шагах. Следовательно, нужно найти максимум выражения (12.6) по всем допустимым управлениям X_{n-1}. Максимум этой суммы зависит от s_{n-2}, обозначается через $Z_{n-1}^*(s_{n-2})$ и называется *условным максимумом целевой функции при оптимальном управлении на двух последних шагах*. Соответствующее управление X_{n-1} на $(n-1)$-м шаге обозначается через $X_{n-1}^*(s_{n-2})$ и называется *условным оптимальным управлением* на $(n-1)$-м шаге.

$$Z_{n-1}^*(s_{n-2}) = \max_{\{X_{n-1}\}} \left\{ f_{n-1}\left(s_{n-2}, X_{n-1}\right) + Z_n^*\left(s_{n-1}\right) \right\}. \qquad \textbf{(12.7)}$$

Следует обратить внимание на то, что выражение, стоящее в фигурных скобках (12.7), зависит только от s_{n-2} и X_{n-1}, так как s_{n-1} можно найти из уравнения состояний (12.2) при $k=n-1$

$$s_{n-1} = \varphi_{n-1}(s_{n-2}, X_{n-1})$$

и подставить вместо s_{n-1} в функцию $Z_n^*(s_{n-1})$.

В результате максимизации только по одной переменной X_{n-1} согласно уравнению (12.7) вновь получаются две функции:

$$Z_{n-1}^*(s_{n-2}) \text{ и } X_{n-1}^*(s_{n-2}).$$

Далее рассматривается трехшаговая задача: к двум последним шагам присоединяется $(n-2)$-й и т. д.

Обозначим через $Z_k^*(s_{k-1})$ ***условный максимум целевой функции, полученный при оптимальном управлении на $n-k+1$ шагах, начиная с***

k-го до конца, при условии, что к началу k-го шага система находилась в состоянии s_{k-1}. Фактически эта функция равна

$$Z_k^*(s_{k-1}) = \max_{\{(x_k,\dots,x_n)\}} \sum_{i=k}^{n} f_i(s_{i-1}, X_i).$$

Тогда

$$Z_{k+1}^*(s_k) = \max_{\{(x_{k+1},\dots,x_n)\}} \sum_{i=k+1}^{n} f_i(s_{i-1}, X_i).$$

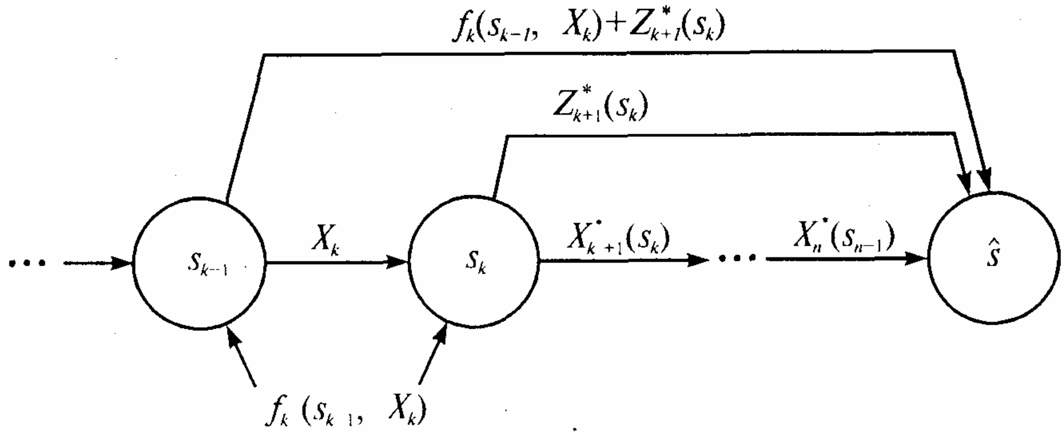


Рис. 12.3

Целевая функция на $n-k$ последних шагах (рис. 12.3) при произвольном управлении X_k на k-м шаге и оптимальном управлении на последующих $n-k$ шагах равна

$$f_k\,(s_{k-1}, X_k) + Z_{k+1}^*\,(s_k).$$

Согласно принципу оптимальности, X_k выбирается из условия максимума этой суммы, т.е.

$$Z_k^*(s_{k-1}) = \max_{\{X_k\}}\left\{f_k(s_{k-1}, X_k) + Z_{k+1}^*(s_k)\right\}, \qquad \textbf{(12.8)}$$

$$k = n-1,\ n-2,\ \dots,\ 2,\ 1.$$

Управление X_k на k-м шаге, при котором достигается максимум в (12.8), обозначается через $X_k^*(s_{k-1})$ и называется ***условным оптимальным управлением на k-м шаге*** (в правую часть уравнения

(12.8) следует вместо s_k подставить выражение $s_k=\varphi_k(s_{k-1}, X_k)$, найденное из уравнений состояния).

Уравнения (12.8) называют *уравнениями Беллмана*. Это рекуррентные соотношения, позволяющие найти предыдущее значение функции, зная последующие. Если из (12.5) найти $Z_n^*(s_{n-1})$, то при $k=n-1$ из (12.8) можно определить, решив задачу максимизации для всех возможных значений s_{n-2}, выражения для $Z_{n-1}^*(s_{n-2})$ и соответствующее $X_{n-1}^*(s_{n-2})$. Далее, зная $Z_{n-1}^*(s_{n-2})$, находим, используя (12.8) и (12.2), уравнения состояний.

Процесс решения уравнений (12.5) и (12.8) называется *условной оптимизацией*[1].

В результате условной оптимизации получаются две последовательности:

$$Z_n^*(s_{n-1}),\ Z_{n-1}^*(s_{n-2}), \ldots,\ Z_2^*(s_1),\ Z_1^*(s_0) \ -$$

условные максимумы целевой функции на последнем, на двух последних, на ...n шагах и

$$X_n^*(s_{n-1}),\ X_{n-1}^*(s_{n-2}), \ldots,\ X_2^*(s_1),\ X_1^*(s_0) \ -$$

условные оптимальные управления на n-м, $(n-1)$-м, ..., 1-м шагах.

Используя эти последовательности, можно найти решение задачи ДП при данных n и s_0. По определению (*см.* разд. 12.1) $Z_1^*(s_0)$ — условный максимум целевой функции за n шагов при условии, что к началу 1-го шага система была в состоянии s_0, т.е.

$$Z_{\max} = Z_1^*(s_0). \qquad \textbf{(12.9)}$$

Далее следует использовать последовательность условных оптимальных управлений и уравнения состояний (12.2).

При фиксированном s_0 получаем $X_1^* = X_1^*(s_0)$. Далее из уравнений (12.2) находим $s_1^* = \varphi_1(s_0, X_1^*)$ и подставляем это выраже-

---

[1] Здесь описан способ решения задачи ДП, начинающийся с последнего шага ("обратная схема"). Можно n-й и 1-й шаги поменять местами ("прямая схема").

ние в последовательность условных оптимальных управлений:

$$X_2^* = X_2^*(s_1^*) \text{ и т.д. по цепочке}^1:$$

$$X_1^* = X_1^*(s_0) \to s_1^* = \varphi_1(s_0, X_1^*) \to X_2^* = X_2^*(s_1^*) \to$$

$$\to s_2^* = \varphi_2(s_1^*, X_2^*) \Rightarrow X_3^* = X_3^*(s_2^*) \to \ldots \to$$

$$\to s_{n-1}^* = \varphi_{n-1}(s_{n-2}^*, X_{n-1}^*) \Rightarrow X_n^* = X_n^*(s_{n-1}^*).$$

Получаем оптимальное решение задачи ДП:

$$X^* = \left(X_1^*, X_2^*, \ldots, X_n^*\right).$$

(Стрелка $\to$ означает использование уравнений состояния, а стрелка $\Rightarrow$ — последовательности условных оптимальных управлений).

12.3. Задача о распределении средств между предприятиями

Рассмотрим предложенную выше схему на конкретной задаче о распределении средств между предприятиями.

▷ **12.1.** Планируется деятельность четырех промышленных предприятий (системы) на очередной год. Начальные средства: $s_0 = 5$ усл. ед. Размеры вложения в каждое предприятие кратны 1 усл. ед. Средства x, выделенные k-му предприятию (k=1, 2, 3, 4), приносят в конце года прибыль $f_k(x)$. Функции $f_k(x)$ заданы таблично (табл. 12.1). Принято считать, что:

а) прибыль $f_k(x)$ не зависит от вложения средств в другие предприятия;

б) прибыль от каждого предприятия выражается в одних условных единицах;

в) суммарная прибыль равна сумме прибылей, полученных от каждого предприятия.

Определить, какое количество средств нужно выделить каждому предприятию, чтобы суммарная прибыль была наибольшей.

---

[1] Через s_k^* здесь обозначено состояние системы после k-го шага при условии, что на k-м шаге выбрано оптимальное управление.

Таблица 12.1

x	$f_1(x)$	$f_2(x)$	$f_3(x)$	$f_1(x)$
1	8	6	3	4
2	10	9	4	6
3	11	11	7	8
4	12	13	11	13
5	18	15	18	16

Решение. Обозначим через x_k количество средств, выделенных k-му предприятию. (Нумерацию предприятий 1, 2, 3, 4 сохраняем в процессе решения неизменной.)

Суммарная прибыль равна

$$Z = \sum_{k=1}^{4} f_k(x_k). \tag{12.10}$$

Переменные x удовлетворяют ограничениям:

$$\begin{gathered}\sum_{k=1}^{4} x_k = 5, \\ x_k \geq 0,\ k = 1,2,3,4.\end{gathered} \tag{12.11}$$

Требуется найти переменные x_1, x_2, x_3, x_4, удовлетворяющие системе ограничений (12.11) и обращающие в максимум функцию (12.10).

Особенности модели. Ограничения линейные, но переменные целочисленные, а функции $f_k(x_k)$ заданы таблично, поэтому нельзя применить методы целочисленного линейного программирования (*см.* гл. 8).

Схема решения задачи методом ДП имеет следующий вид: процесс решения распределения средств s_0=5 можно рассматривать как 4-шаговый, номер шага совпадет с номером предприятия; выбор переменных x_1, x_2, x_3, x_4 — управление соответственно на I, II, III, IV шагах. $\hat{s}$ — конечное состояние процесса распределения — равно нулю, так как все средства должны быть вложены в производство, $\hat{s}$ =0. Схема распределения показана на рис. 12.4.

Уравнения состояний (12.2) в данной задаче имеют вид:

$$s_k = s_{k-1} - x_k,\quad k=1, 2, 3, 4, \tag{12.12}$$

где s_k — параметр состояния — количество средств, оставшихся после k-го шага, т.е. средства, которые остается распределить между оставшимися $4-k$ предприятиями.

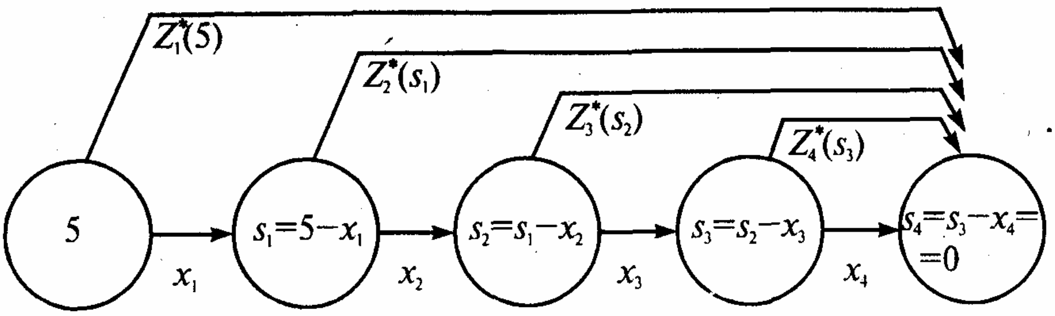


Рис. 12.4

Введем в рассмотрение функцию $Z_k^*(s_{k-1})$ — условную оптимальную прибыль, полученную от k-го, $(k+1)$-го, ..., 4-го предприятий, если между ними распределялись оптимальным образом средства $s_{k-1}(0 \le s_{k-1} \le 5)$. Допустимые управления на k-м шаге удовлетворяют условию: $0 \le x_k \le s_{k-1}$ (либо k-му предприятию ничего не выделяем, $x_k=0$, либо не больше того, что имеем к k-му шагу, $x_k \le s_{k-1}$).

Уравнения (12.5) и (12.8) имеют вид:

$$k=4, \ s_4 = 0 \Rightarrow Z_4^*(s_3) = \max_{0 \le x_4 \le s_3} f_4(x_4), \qquad \textbf{(а)}$$

$$Z_3^*(s_2) = \max_{0 \le x_3 \le s_2} \{f_3(x_3) + Z_4^*(s_3)\}, \qquad \textbf{(б)}$$

$$Z_2^*(s_1) = \max_{0 \le x_2 \le s_1} \{f_2(x_2) + Z_3^*(s_2)\}, \qquad \textbf{(в)}$$

$$Z_1^*(5) = \max_{0 \le x_1 \le 5} \{f_1(x_1) + Z_2^*(s_1)\}. \qquad \textbf{(г)}$$

Последовательно решаем записанные уравнения, проводя условную оптимизацию (*см.* рис. 12.4) каждого шага.

IV шаг. В табл. 12.1 $f_4(x)$ прибыли монотонно возрастают, поэтому все средства, оставшиеся к IV шагу, следует вложить в 4-е предприятие. При этом для возможных значений $s_3=0, 1, ..., 5$ получим:

$$Z_4^*(s_3) = f_4(s_3) \text{ и } x_4^*(s_3) = s_3.$$

I I I шаг. Делаем все предположения относительно остатка средств s_2 к III шагу (т.е. после выбора x_1 и x_2). s_2 может принимать значения 0, 1, 2, 3, 4, 5 (например, $s_2 = 0$, если все средства отданы 1-му и 2-му предприятиям, $s_2=5$, если 1-е и 2-е предприятия ничего не получили, и т.д.). В зависимости от этого выбираем $0 \le x_3 \le s_2$, находим $s_3=s_2-x_3$ и сравниваем для разных x_3 при фиксированном s_2 значения суммы $f_3(x_3)+ Z_4^*(s_3)$. Для каждого s_2 наибольшее из этих значений есть $Z_3^*(s_2)$ — условная оптимальная прибыль, полученная при оптимальном распределении средств s_2 между 3-м и 4-м предприятиями. Оптимизация дана в табл. 12.2 при $k=3$. Для каждого значения s_2 $Z_3^*(s_2)$ и $X_3^*(s_2)$ помещены в графах 5 и 6 соответственно.

I I шаг. Условная оптимизация, согласно уравнению (в), проведена в табл. 12.2 при $k=2$. Для всех возможных значений s_1 значения $Z_2^*(s_1)$ и $X_2^*(s_1)$ находятся в столбцах 8 и 9 соответственно; первые слагаемые в столбце 7 — значения $f_2(x_2)$, взяты из табл. 12.1, а вторые слагаемые взяты из столбца 5 табл. 12.2 при $s_2 = s_1 - x_2$.

I шаг. Условная оптимизация (уравнение (г)) проведена в табл. 12.2 при $k=1$ для $s_0=5$[1]. Поясним решение подробно: если $x_1=0$, то $s_1=5$, прибыль, полученная от четырех предприятий при условии, что $s_1=5$ ед. средств между оставшимися тремя предприятиями будут распределены оптимально, равна $f_1(0)+ Z_2^*(5) = 0+19=19$ ($Z_2^*(5)$ взято из столбца 9 табл. 12.2 при $s_1=5$). Если $x_1=1$, то $s_2=4$. Суммарная прибыль при условии, что $s_2=4$ ед. средств между оставшимися тремя предприятиями будут распределены оптимально, равна $f_1(1)+ Z_2^*(4)=8+16=\underline{24}$ ($f_1(1)$ взято из табл. 12.1, а $Z_2^*(4)$ — из столбца 9 табл. 12.2.) Аналогично при $x_1=2$, $s_2=3$ и $f_1(2)+ Z_2^*(3)=10+13=\underline{23}$;

при $x_1=3$, $s_2=2$ и $f_1(3)+ Z_2^*(3)=11+10=\underline{21}$;

при $x_1=4$, $s_2=1$ и $f_1(4)+ Z_2^*(1)=12+16=\underline{28}$;

при $x_1=5$, $s_2=0$ и $f_1(5)+ Z_2^*(0)=18+0=\underline{18}$.

---

[1] На I шаге условной оптимизации достаточно заполнить раздел таблицы, соответствующий $s_0=5$.

Сравнивая подчеркнутые числа, получим $Z_1^*(5)=24$ усл. ед. $= Z_{\max}$ при $x_1^* = x_1^*(5) = 1$.

Используя уравнения (12.12), получим $s_1^*=5-1=4$, а по табл. 12.2 в столбце 9 находим $x_2^* = x_2^*(4) = 2$. Далее находим $s_2^*=4-2=2$, а по табл. 12.2 в столбце 6 — $x_3^* = x_3^*(2) = 1$. Наконец, $s_3^*=2-1=1$ и $x_4^* = x_4^*(1) = 1$, т.е. $X^*(1;\ 2;\ 1;\ 1)$.

Максимум суммарной прибыли равен 24 усл. ед. средств при условии, что 1-му предприятию выделено 1 усл. ед.; 2-му предприятию — 2 усл. ед.; 3-му предприятию — 1 усл. ед.; 4-му предприятию — 1 усл.ед.▶

Замечание 1. Решение четырехмерной задачи **12.1** на определение условного экстремума сведено фактически к решению четырех одномерных задач: на каждом шаге определялась одна переменная x.

Замечание 2. На разобранной задаче **12.1** видно, что метод ДП безразличен к виду и способу задания функции: $f_k(x)$ были заданы таблично, поэтому и $Z_k^*(s)$ и $X_k^*(s)$ принимали дискретные значения, представленные в табл. 12.2.

Таблица 12.2

			$k=3$			$k=2$			$k=1$		
s_{k-1}	x_k	s_k	$f_3(x_3)+$ $+Z_4^*(s_3)$	Z_3^* (s_2)	x_3^* (s_2)	$f_2(x_2)+$ $+Z_3^*(s_2)$	Z_2^* (s_1)	x_2^* (s_1)	$f_1(x_1)+$ $+Z_2^*(s_1)$	Z_1^* (s_0)	x_1^* (s_0)
1	2	3	4	5	6	7	8	9	10	11	12
0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	0+1=1			0+4=4			0+6=6	8	1
	1	0	3+0=3	4	0	6+0=6	6	1	8+0=8		
2	0	2	0+6=6			0+7=7			0+10=10		
	1	1	3+4=7	7	1	6+4=10	10	1	8+6=14	14	1
	2	0	4+0=4			9+0=9			10+0=10		
	0	3	0+8=8			0+9=9			0+13=13		
3	1	2	3+6=9	9	1	6+7=13	13	1	8+10=18	18	1
	2	1	4+4=8			9+4=13		2	10+6=16		

s_{k-1}	x_k	s_k	k=3			k=2			k=1		
			$f_3(x_3)+$ $+Z_4^*(s_3)$	Z_3^* (s_2)	x_3^* (s_2)	$f_2(x_2)+$ $+Z_3^*(s_2)$	Z_2^* (s_1)	x_2^* (s_1)	$f_1(x_1)+$ $+Z_2^*(s_1)$	Z_1^* (s_0)	x_1^* (s_0)
1	2	3	4	5	6	7	8	9	10	11	12
	3	0	7+0=7			11+0=11			11+0=11		
	0	4	0+13=13			0+13=13			0+16=16		
	1	3	3+8=11			6+9=15			8+13=21		
4	2	2	4+6=10	13	0	9+7=16	16	2	10+10=20	21	1
	3	1	7+4=11			11+4=15			11+6=17		
	4	0	11+0=11			13+0=13			12+0=12		
	0	5	0+16=16			0+18=18			0+19=19		
	1	4	3+13=16			6+13=19			8+16=24		
5	2	3	4+8=12	18	5	9+9=18	19	1	10+13=23	24	1
	3	2	7+6=13			11+7=18			11+10=21		
	4	1	11+4=15			13+4=17			12+6=18		
	5	0	18+0=18			15+0=15			18+0=18		

Замечание 3. Альтернативой между ДП для подобной дискретной задачи является метод перебора. Метод ДП предпочтительнее, так как на этапе условной оптимизации отбрасываются заведомо негодные варианты.

Замечание 4. Достоинством метода является возможность анализа решения на чувствительность к изменению s_0 и n. Проведенные расчеты можно использовать для изменившихся начального состояния s_0 и числа шагов n. Например, пусть в задаче **12.1** произошло уменьшение начальных средств на 1 усл. ед. Для s_0=4 достаточно в таблицу добавить расчеты при k=1 (это сделано в той же табл. 12.2). Получаем в этом случае Z_{max}=21 усл. ед. при распределении:

$$x_1^*=1 \to s_1^*=4-1=3 \Rightarrow x_2^*=1,\text{ или } x_2^*=2 \to$$

$$\to s_2^*=3-1=2,\text{ или } s_2^*=3-2=1 \Rightarrow x_3^*=1,\text{ или}$$

$$x_3^*=0 \to s_3^*=2-1=1,\text{ или } s_3^*=1-0=1 \Rightarrow x_4^*=1.$$

В результате найдены два оптимальных решения: $X^{(1)*}(1;\ 1;\ 1;\ 1)$ и $X^{(2)*}\ (1;\ 2;\ 0;\ 1)$. Если начальные средства увеличились, например, на 1 усл. ед., $s_0=6$, а функции прибыли $f_k(x)$ остались прежними, то в табл. 12.2 достаточно добавить раздел для $s_0=6$ при $k=3$, 2, 1; этот фрагмент расчетов помещен в табл. 12.3.

Таблица 12.3

1	2	3	4	5	6	7	8	9	10	11	12
6	0	6	0+16=16	22	5	0+22=22	24	1	0+24=24	27	1
	1	5	3+16=19			6+18=24			8+19=27		
	2	4	4+13=17			9+13=22			10+16=26		
	3	3	7+8=15			11+9=20			11+13=24		
	4	2	11+6=17			13+7=20			12+10=22		
	5	1	18+4=22			15+4=19			18+6=24		

Получаем $Z_{\max}=27$, $x_1^*=1 \to s_1^*=6-1=5 \Rightarrow x_2^*=1 \to s_2^*=5-1= =4 \Rightarrow x_3^*=0 \to s_3^*=4-0=4 \Rightarrow x_4^*=4$.

Оптимальное решение $X^*(1;\ 1;\ 0;\ 4)$.

Если принято решение распределить средства $s_0=5$ между 2-, 3- и 4-м предприятиями, то задача уже решена в табл. 12.2. В разделе $k=2$ таблицы находим $Z_{\max}=\ Z_2^*(5)=19$ при условии, что $x_2^*=1$, $x_3^*=0$, $x_4^*=4$.

Наконец, если увеличилось количество предприятий (число шагов), то схему можно дополнить, присоединяя шаги с номерами $k=0$, -1, ... и т. д. Например, пусть средства $\xi_0 = 6$ распределяются между пятью предприятиями. Функция прибыли для пятого предприятия задана формулой $f(x)=3x+1$, если $x\neq 0$ и $f(0)=0$. Присвоим 5-му предприятию номер $k=0$, тогда x_0 — средства, выделенные этому предприятию. Обозначим через $Z_0^*(6)$ оптимальную прибыль, полученную от пяти предприятий:

$$Z_0^*(6) = \max_{0\le x_0\le 6}\left\{f_0(x_0) + Z_1^*(s_1)\right\},$$

а $s_1=6-x_0$. Условная оптимизация 0-го шага дана в табл. 12.4.

Таблица 12.4

x_0	0	1	2	3	4	5	6
$s_1=6-x_0$	6	5	4	3	2	1	0
$f(0)=0$	0	4	7	10	13	16	19
$Z_1^*(s_1)$ (взята из табл. 12.2 и 12.3 при $k=1$)	27	24	21	18	14	8	0
$f(x_0)+Z_1(s_1)$	27	28	28	28	27	24	19

Следовательно, $Z_{max}=28$, а оптимальных решений четыре: $X_1^*(1;\ 1;\ 2;\ 1;\ 1)$, $X_2^*(2;\ 1;\ 1;\ 1;\ 1)$, $X_3^*(2;\ 1;\ 2;\ 0;\ 1)$, $X_4^*(3;\ 1;\ 1;\ 0;\ 1)$.

Замечание 5. К недостаткам метода по-прежнему следует отнести возникновение технических трудностей при вычислениях в случае увеличения размерности. Если каждое управление X_k^* будет зависеть от r переменных, а состояние s_k^* — от p параметров, то на каждом шаге возникает rp-мерная задача оптимизации. (В задаче **12.1** $r=1$, $p=1$, т.е. решалась одномерная задача). Даже при реализации метода ДП на ЭВМ практически можно решать задачи для небольших r, p, n.

12.4. Общая схема применения метода ДП. Задача об оптимальном распределении ресурсов между отраслями на *n* лет

Прежде чем перейти к конкретным задачам, следует усвоить *общую схему применения метода ДП.*

Предположим, что все требования, предъявляемые к задаче методом ДП, выполнены. (Эти требования сформулированы в разд. 12.1). Построение модели ДП и применение метода ДП для решения сводится к следующим моментам:

1. Выбирают способ деления процесса управления на шаги.
2. Определяют параметры состояния s_k и переменные управления X_k на каждом шаге.
3. Записывают уравнения состояний.
4. Вводят целевые функции k-го шага и суммарную целевую функцию.

5. Вводят в рассмотрение условные максимумы (минимумы) $Z_k^*(s_{k-1})$ и условное оптимальное управление на k-м шаге: $X_k^*(s_{k-1})$, $k = n, n-1, ..., 2, 1$.

6. Записывают основные для вычислительной схемы ДП уравнения Беллмана для $Z_n^*(s_{n-1})$ и $Z_k^*(s_{k-1})$, $k=n-1, ..., 1$.

7. Решают последовательно уравнения Беллмана (условная оптимизация) и получают две последовательности функций:

$$\{Z_k^*(s_{k-1})\} \text{ и } \{X_k^*(s_{k-1})\}.$$

8. После выполнения условной оптимизации получают оптимальное решение для конкретного начального состояния s_0:

а) $Z_{\max} = Z_1^*(s_0)$ и

б) по цепочке $s_0 \Rightarrow X_1^* \to s_1^* \Rightarrow X_2^* \to s_2^* \Rightarrow ... \Rightarrow X_{n-1}^* \to s_{n-1}^* \Rightarrow$ $\Rightarrow X_n^* \to s_n^*$ оптимальное управление: $X^*\left(X_1^*, X_2^*, \ldots, X_n^*\right)$.

Решая задачи, следует по возможности придерживаться этой схемы по крайней мере в начале изучения темы. Рассмотрим, как работает схема на примере задачи об оптимальном распределении ресурсов между двумя отраслями на n лет.

▷ **12.2.** Планируется деятельность двух отраслей производства на n лет. Начальные ресурсы s_0. Средства x, вложенные в I отрасль в начале года, дают в конце года прибыль $f_1(x)$ и возвращаются в размере $q_1(x)<x$; аналогично для II отрасли функция прибыли равна $f_2(x)$, а возврата — $q_2(x)$ ($q_2(x)<x$). В конце года все возвращенные средства заново перераспределяются между I и II отраслями, новые средства не поступают, прибыль в производство не вкладывается[1].

Требуется распределить имеющиеся средства s_0 между двумя отраслями производства на n лет так, чтобы суммарная прибыль от обеих отраслей за n лет оказалась максимальной.

Необходимо:

а) построить модель ДП для задачи и вычислительную схему;

б) решить задачу при условии, что $s_0 = 10000$ ед., $n = 4$, $f_1(x) = 0{,}6x$, $q_1(x) = 0{,}7x$, $f_2(x) = 0{,}5x$, $q_2(x) = 0{,}8x$.

---

[1] Последние условия определяют вид уравнений состояний; если поступают новые средства или часть прибыли вкладывается в производство, это можно легко учесть, так как алгоритм метода ДП не изменяется.

Решение. а) Процесс распределения средств между двумя отраслями производства разворачивается во времени, решения принимаются в начале каждого года, следовательно, осуществляется деление на шаги: номер шага — номер года. Управляемая система — две отрасли производства, а управление состоит в выделении средств каждой отрасли в очередном году. Параметры состояния к началу k-го года — s_{k-1} (k=1, 2,..., n) — количество средств, подлежащих распределению. Переменных управления на каждом шаге две: x_k — количество средств, выделенных I отрасли, и y_k — II отрасли. Но так как все средства s_{k-1} распределяются, то $y_k=s_{k-1}-x_k$, и поэтому управление на k-м шаге зависит от одной переменной x_n, т.е. $X_k(x_k, s_{k-1}-x_k)$.

Уравнения состояний

$$s_k=q_1(x_k)+q_2(s_{k-1}-x_k) \qquad \textbf{(12.13)}$$

выражают остаток средств, возвращенных в конце k-го года.

Показатель эффективности k-го шага — прибыль, полученная в конце k-го года от обеих отраслей:

$$f_1(x_k)+f_2(s_{k-1}-x_k). \qquad \textbf{(12.14)}$$

Суммарный показатель эффективности — целевая функция задачи — прибыль за n лет:

$$Z=\sum_{k=1}^{n} f_1(x_k)+f_2(s_{k-1}-x_k). \qquad \textbf{(12.15)}$$

Пусть $Z_k^*(s_{k-1})$ — условная оптимальная прибыль за $n-k+1$ лет, начиная с k-го года до n-го года включительно, при условии, что имеющиеся на начало k-го года средства s_{k-1} в дальнейшем распределялись оптимально. Тогда оптимальная прибыль за n лет $Z_{\max} = Z_1^*(s_0)$.

Уравнения Беллмана имеют вид:

$$Z_n^*(s_{n-1}) = \max_{0\le x_k \le s_{k-1}} f_1(x_n)+f_2(s_{n-1}-x_n)\}, \qquad \textbf{(12.16)}$$

$$Z_k^*(s_{k-1}) = \max_{0\le x_k \le s_{k-1}} \{f_1(x_k)+f_2(s_{k-1}-x_k)+ Z_{k+1}^*(s_k)\}, \qquad \textbf{(12.17)}$$

$$(k=n-1,\ n-2,\ ...,\ 2).$$

б) Используем конкретные данные.

Уравнение состояний (12.13) примет вид:

$$s_k=0{,}7x_k+0{,}8(s_{k-1}-x_k),\ \text{или}\ s_k=0{,}8s_{k-1}-0{,}1x_k. \qquad \textbf{(12.18)}$$

Целевая функция k-го шага (12.14)

$$0{,}6x_k+0{,}5(s_{k-1}-x_k)=0{,}1x_k+0{,}5s_{k-1}.$$

Целевая функция задачи

$$Z=\sum_{k=1}^{4}0{,}5s_{k-1}+0{,}1x_k\ . \qquad \textbf{(12.19)}$$

Функциональные уравнения

$$Z_4^*(s_3)=\max_{0\le x_4\le s_3}\{0{,}5s_3+0{,}1x_4\}, \qquad \textbf{(12.20)}$$

$$\text{и}\ Z_k^*(s_{k-1})=\max_{0\le x_k\le s_{k-1}}\{0{,}1x_k+0{,}5s_{k-1}+Z_{k+1}^*(s_k)\}. \qquad \textbf{(12.21)}$$

Проводим условную оптимизацию.

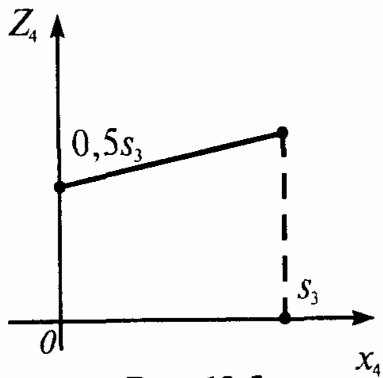


Рис. 12.5

IV шаг. Используем уравнение (12.20). Обозначим через Z_4 функцию, стоящую в скобках, $Z_4=0{,}1x_4+0{,}5s_3$; функция Z_4 — линейная, возрастающая, так как угловой коэффициент 0,1 больше нуля. Поэтому максимум достигается на конце интервала $[0;\ s_3]$ (рис. 12.5). Следовательно, $Z_4^*(s_3)=0{,}6s_3$ при $x_4^*(s_3)=s_3$.

III шаг. Уравнение:

$$Z_3^*(s_2)=\max_{0\le x_3\le s_2}\{0{,}1x_3+0{,}5s_2+0{,}6s_3\}.$$

Найдем s_3 из уравнений состояний (12.18): $s_3=0{,}8s_2-0{,}1x_3$ и, подставив его выражение в правую часть уравнения, получим

$$Z_3^*(s_2)=\max_{0\le x_2\le s_2}\{0{,}1x_3+0{,}5s_2+0{,}6(0{,}8s_2-0{,}1x_3)\},$$

$$Z_3^*(s_2)=\max_{0\le x_3\le s_2}\{0{,}04x_3+0{,}98s_2\}.$$

Как и в предыдущем случае, максимум достигается при $x_3=s_2$; т.е. $Z_3^*(s_2) = 1{,}02s_2$ при $x_3^*(s_2)=s_2$.

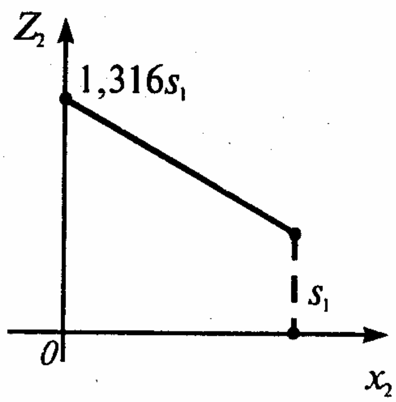


Рис. 12.6

II шаг. Из уравнения состояния: $s_2=0{,}8s_1-0{,}1x_2$. Поэтому уравнение (12.20) при $k=2$ примет вид:

$$Z_2^*(s_1)=\max_{0\le x_2\le s_1}\{1{,}316s_1-0{,}002x_2\}.$$

Линейная относительно x_2 функция $Z_2^* = 1{,}316s_1-0{,}002x_2$ убывает на отрезке $[0; s_1]$, и поэтому ее максимум достигается при $x_2=0$ (рис.12.6):

$$Z_2^*(s_1)=1{,}316s_1 \text{ при } x_2^*(s_1)=0.$$

I шаг. $s_1=0{,}8s_0-0{,}1x_1$. Уравнение (12.20) при $k=1$ имеет вид:

$$Z_1^*(s_0)=\max_{0\le x_1\le s_0}\{1{,}5528s_0-0{,}0316x_1\}.$$

Как и в предыдущем случае, максимум достигается в начале отрезка, т.е.

$$Z_1^*(s_0) = 1{,}5528s_0 \text{ при } x_1^*(s_0)=0.$$

На этом условная оптимизация заканчивается. Используя ее результат и исходные данные, получим $Z_{\max} = Z_1^*(10000)$, $Z_{\max} = 15528$.

$$x_1^* = 0, \quad y_1^* = s_0 = 10000$$

(все средства выделяются II отрасли) $\to$

$$s_1^*=0{,}8\cdot 10000-0{,}1\cdot 0=8000 \Rightarrow x_2^*=0,\ y_2^*=s_1=8000$$

(все средства выделяются II отрасли) $\to$

$$\to s_2^*=0{,}8\cdot 8000-0{,}1\cdot 0=6400 \Rightarrow x_3^*=6400,\ y_3^*=0 \to$$

(все средства выделяются I отрасли) $\to$

$$\to s_3^*=0{,}8\cdot 6400-0{,}1\cdot 6400=4480 \Rightarrow x_4^*=4480,\ y_4^*=0$$

(все средства выделяются I отрасли).

Оптимальная прибыль за 4 года, полученная от двух отраслей производства при начальных средствах 10000 ед., равна 15528 ед.

при условии, что I отрасль получает по годам (0; 0; 6400; 4480), а II отрасль — соответственно (10000; 8000; 0; 0).▶

12.5. Задача о замене оборудования

Замена оборудования — важная экономическая проблема. Задача состоит в определении оптимальных сроков замены старого оборудования (станков, производственных зданий и т. п.). Старение оборудования включает его физический и моральный износ, в результате чего растут производственные затраты, затраты на ремонт и обслуживание, снижаются производительность труда, ликвидная стоимость. Критерием оптимальности являются, как правило, либо прибыль от эксплуатации оборудования (задача максимизации), либо суммарные затраты на эксплуатацию в течение планируемого периода (задача минимизации).

При построении модели задачи принято считать, что решение о замене выносится в начале каждого промежутка эксплуатации (например, в начале года) и что в принципе оборудование можно использовать неограниченно долго.

Основная характеристика оборудования — параметр состояния — его возраст t.

При составлении динамической модели замены процесс замены рассматривают как n-шаговый, разбивая весь период эксплуатации на n шагов. Возможное управление на каждом шаге характеризуется качественными признаками, например, $X^{с}$ (сохранить оборудование), $X^{з}$ (заменить) и $X^{р}$ (сделать ремонт).

Рассмотрим конкретный пример.

▷ **12.3.** Оборудование эксплуатируется в течение 5 лет, после этого продается. В начале каждого года можно принять решение сохранить оборудование или заменить его новым. Стоимость нового оборудования p_0=4000 руб.[1] После t лет эксплуатации ($1 \leq t \leq 5$) оборудование можно продать за $g(t)=p_0 2^{-t}$ рублей (ликвидная стоимость). Затраты на содержание в течение года зависят от возраста t оборудования и равны $r(t)=600(t+1)$. Определить оптимальную стратегию эксплуатации оборудования, чтобы суммарные затраты с учетом начальной покупки и заключительной продажи были минимальны.

---

[1] Все цены условные.

Решение. Способ деления управления на шаги естественный, по годам, $n = 5$. Параметр состояния — возраст машины — $s_{k-1}=t$, $s_0=0$ — машина новая в начале первого года эксплуатации. Управление на каждом шаге зависит от двух переменных $X^с$ и $X^з$.

Уравнения состояний зависят от управления:

$$s_k=\begin{cases} t+1, & \text{если} \quad X_k = X^с, \\ 1, & \text{если} \quad X_k = X^з,\ k=1,2,3,4. \end{cases} \qquad \textbf{(12.22)}$$

В самом деле, если к k-му шагу $s_{k-1}=t$, то при сохранении машины ($X_k=X^с$) через год возраст машины увеличится на 1. Если машина заменяется новой ($X_k=X^з$), то это означает, что к началу k-го шага ее возраст $t=0$, а после года эксплуатации $t=1$, т.е. $s_k=1$.

Показатель эффективности k-го шага:

$$f_k=(X_k,\ t)=\begin{cases} 600(t+1), & \text{если} \quad X_k = X^с, \\ 4600-4000\cdot 2^{-t}, & \text{если}\ X_k = X^з, \end{cases} \quad k=1,\ 2,\ 3,\ 4. \qquad \textbf{(12.23)}$$

(При $X^с$ затраты только на эксплуатацию машины возраста t, при $X^з$ машина продается ($-4000\cdot 2^{-t}$), покупается новая (4000) и эксплуатируется в течение первого года (600), общие затраты равны ($-4000\cdot 2^{-t}+4000+600$)).

Пусть $Z_k^*(t)$ — условные оптимальные затраты на эксплуатацию машины, начиная с k-го шага до конца, при условии, что к началу k-го шага машина имеет возраст t лет. Запишем для функций $Z_k^*(t)$ уравнения Беллмана (12.5) и (12.8), заменив задачу максимизации на задачу минимизации:

$$Z_5^* = \min\begin{cases} 600(t+1)-4000\cdot 2^{-(t+1)}, & \text{если} \quad X_5 = X^с, \\ 4600-4000\cdot 2^{-t}-4000\cdot 2^{-(t+1)}, & \text{если} \quad X_5 = X^з. \end{cases} \qquad \textbf{(12.24)}$$

Величина $4000\cdot 2^{-(t+1)}$ — стоимость машины возраста t лет (по условию машина после 5 лет эксплуатации продается).

$$Z_k^* = \min\begin{cases} 600(t+1)+Z_{k+1}^*(t+1), & \text{если} \quad X_k = X^с, \\ 4600-4000\cdot 2^{-t}+Z_{k+1}^*(1), & \text{если} \quad X_k = X^з, \end{cases} \qquad \textbf{(12.25)}$$

$$k=4,\ 3,\ 2,\ 1.$$

Из определения функций $Z_k^*(t)$ следует

$$Z_{\min} = Z_1^*(0).$$

Дадим геометрическое решение этой задачи. На оси абсцисс будем откладывать номер шага k, на оси ординат — возраст t машины. Точка $(k-1,\ t)$ на плоскости соответствует началу k-го года эксплуатации машины возраста t лет. Перемещение на графике в зависимости от принятого управления на k-м шаге показано на рис. 12.7.

Состояние начала эксплуатации машины соответствует точке $s_0^*(0;\ 0)$, конец — точкам $\hat{s}\,(6;\ t)$. Любая траектория, переводящая точку $s(k-1;\ t)$ из s_0^* в $\hat{s}$, состоит из отрезков — шагов, соответствующих годам эксплуатации. Надо выбрать такую траекторию, при которой затраты на эксплуатацию машины окажутся минимальными.

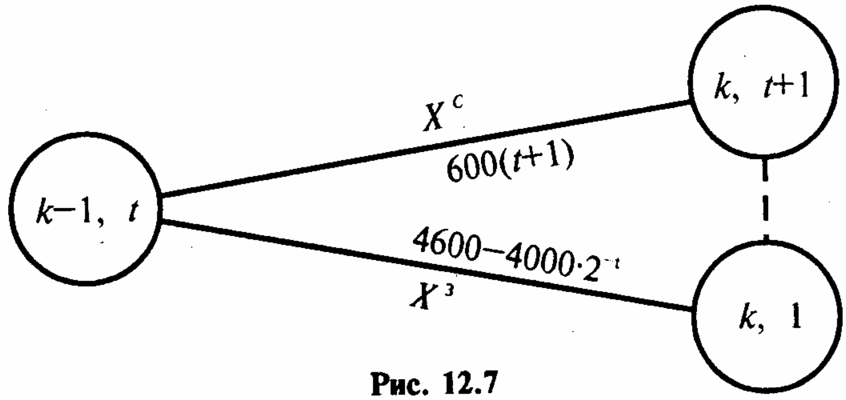
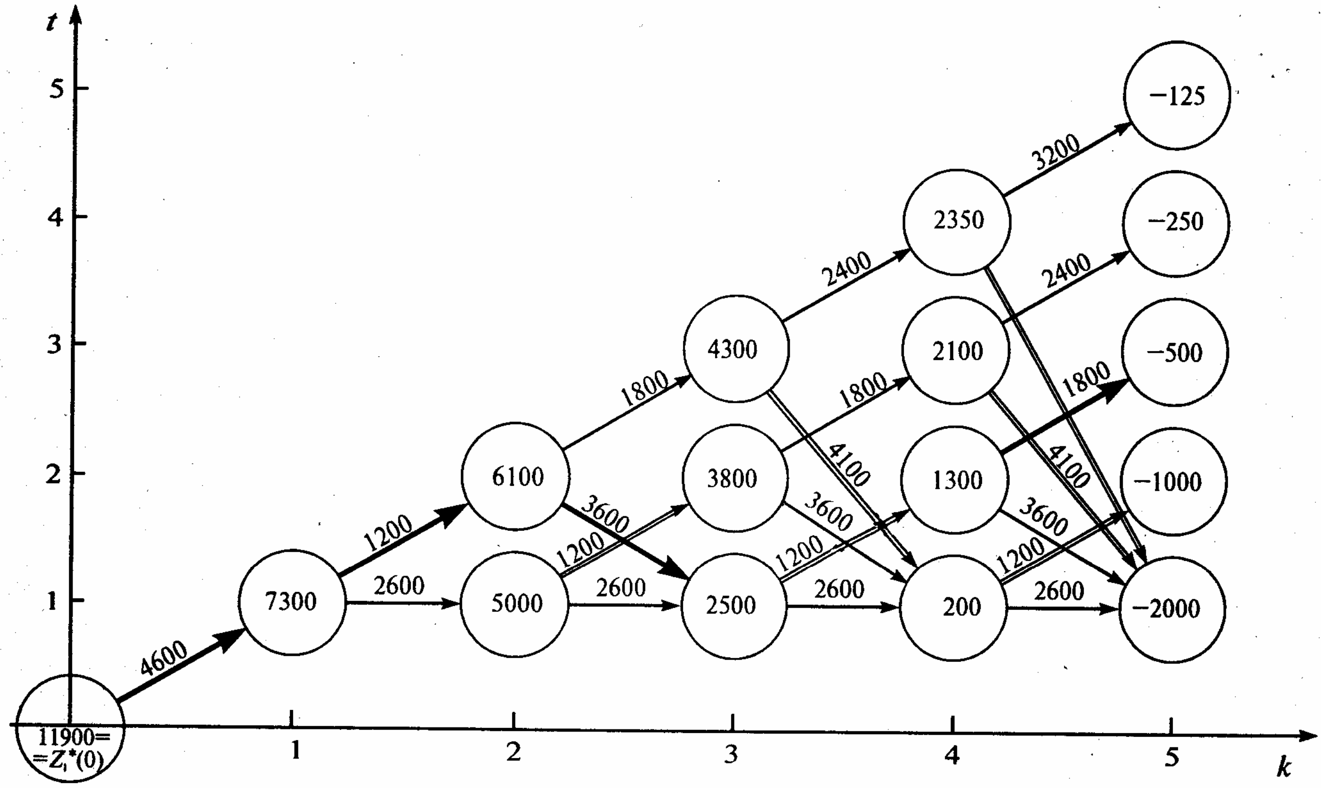


Рис. 12.7

Над каждым отрезком, соединяющим точки $(k-1;\ t)$ и $(k;\ t+1)$, запишем соответствующие управлению X^c затраты, найденные из (12.23): $600(t+1)$, а над отрезком, соединяющим точки $(k-1;\ t)$ и $(k;\ t)$, запишем затраты, соответствующие управлению $X^з$, т.е. $4600-4000 \cdot 2^{-t}$. Таким образом мы разметим все отрезки, соединяющие точки на графике, соответствующие переходам из любого состояния s_{k-1} в состояние s_k (рис. 12.8). Например, над отрезками, соединяющими точки $(k;\ 2)$ и $(k+1;\ 3)$, стоит число 1800[1], что соответствует затратам на эксплуатацию в течение каждого года

[1] Напоминаем, что все затраты выражены в условных рублях.



t
5
4
3
2
1
−125
3200
2350
−250
2400
2400
4300
2100
−500
1800
1800
1800
6100
3800
4100
1300
4100
−1000
3600
3600
3600
1200
1200
1200
1200
7300
2600
5000
2600
2500
2600
200
2600
−2000
4600
11900=
$=Z_1^*(0)$
1
2
3
4
5
k

Рис. 12.8

машины возраста $t = 2$ лет, а над отрезками, соединяющими (k; 2) и (k+1; 1), стоит число 3600 — это сумма затрат на покупку машины и эксплуатацию новой машины в течение года без "затрат" (выручки) за проданную машину возраста t лет. Следует учесть, что $0 \le t \le k$.

Проведем на размеченном графе состояний (*см.* рис. 12.8) условную оптимизацию.

V шаг. Начальные состояния — точки (4; t), конечные (5; t). В состояниях (5; t) машина продается, условный оптимальный доход от продажи равен $4000 \cdot 2^{-t}$, но поскольку целевая функция связана с затратами, то в кружках точек (5; t) поставим величину дохода со знаком минус.

Анализируем, как можно попасть из каждого начального состояния в конечное на V шаге.

Состояние (4; 1). Из него можно попасть в состояние (5; 2), затратив на эксплуатацию машины 1200 и выручив затем от продажи 1000, т.е. суммарные затраты равны 200, и в состояние (5; 1) с затратами 2600−2000=600. Значит, если к последнему шагу система находилась в точке (4; 1), то следует идти в точку (5; 2) (укажем это направление двойной стрелкой), а неизбежные минимальные затраты, соответствующие этому переходу, равны 200 (поместим эту величину $Z_5^*(1)=200$ в кружке точки (4; 1).

Состояние (4; 2). Из него можно попасть в точку (5; 3) с затратами 1800−500=1300 и в точку (5; 1) с затратами 3600−2000=1600. Выбираем первое управление, отмечаем его двойной стрелкой, а $Z_5^*(2)=1300$ проставляем в кружке точки (4; 2).

Рассуждая таким же образом для каждой точки предпоследнего шага, мы найдем для любого исхода IV шага оптимальное управление на V шаге, отметим его на рис. 12.8 двойной стрелкой. Далее планируем IV шаг, анализируя каждое состояние, в котором может оказаться система в конце III шага с учетом оптимального продолжения до конца процесса, т.е. решаем для всех $0 \le t \le 4$ при k=4 уравнения (12.22). Например, если начало IV шага соответствует состоянию (3; 1), то при управлении X^c система переходит в точку (4; 2), затраты на этом шаге 1200, а суммарные затраты за два последних шага равны 1200+1300=2500. При управлении $X^з$ затраты за два шага равны 2600+200=2800. Выбираем минимальные затраты 2500, ставим их в кружок точки (3; 1), а соответствующие управления на этом шаге помечаем двойной стрелкой,

ведущей из состояния (3; 1), в состояние (4; 2). Так поступаем для каждого состояния (3; t) (*см.* рис. 12.8).

Продолжая условную оптимизацию III, II и I шагов, мы получим на рис. 12.8 следующую ситуацию: из каждой точки (состояния) выходит стрелка, указывающая, куда следует перемещаться в данном шаге, если система оказалась в этой точке, а в кружках записаны минимальные затраты на переход из этой точки в конечное состояние. На каждом шаге графически решались уравнения (12.22).

После проведения условной оптимизации получим в точке (0; 0) минимальные затраты на эксплуатацию машины в течение 5 лет с последующей продажей: $Z_{\min} = 11900$. Далее строим оптимальную траекторию, перемещаясь из точки $s_0(0;\ 0)$ по двойным стрелкам в $\hat{s}$. Получаем набор точек:

$$\{(0;\ 0),\ (1;\ 1),\ (2;\ 2),\ (3;\ 1),\ (4;\ 2),\ (5;\ 3)\},$$

который соответствует оптимальному управлению X^* ($X^с$, $X^с$, $X^з$, $X^с$, $X^с$). Оптимальный режим эксплуатации состоит в том, чтобы заменить машину новой в начале 3-го года.▶

Таким образом, размеченный график (сеть) позволяет наглядно интерпретировать расчетную схему и решить задачу методом ДП.

Как уже отмечалось, модели и вычислительная схема ДП очень гибки в смысле возможностей включения в модель различных модификаций задачи. Например, аналогичная задача может быть рассмотрена для большого числа вариантов управления, "ремонт", "капитальный ремонт" и т. д. Можно рассматривать замену оборудования новым с учетом технического прогресса, можно учесть изменения в затратах на эксплуатацию оборудования после его ремонта, в зависимости от года эксплуатации (дороже, дешевле). Все эти факторы можно учитывать вычислительной схемой ДП.

УПРАЖНЕНИЯ

В задачах 12.4–12.5 найти оптимальное распределение средств между n предприятиями при условии, что прибыль $f(x)$, полученная от каждого предприятия, является функцией от вложенных в него средств x. Вложения кратны Δx, а функции $f(x)$ заданы таблично.

12.4.

x	1	2	3	4	5	6	7	8	9
$f_1(x)$	5	9	12	14	15	18	20	24	27
$f_2(x)$	7	9	11	13	16	19	21	22	25
$f_3(x)$	6	10	13	15	16	18	21	22	25

$s_0=9$, $n=3$, $\Delta x=1$.

12.5.

x	1	2	3	4	5
$f_1(x)$	0,2	0,9	1,0	1,2	2,0
$f_2(x)$	1,0	1,1	1,3	1,4	1,8
$f_3(x)$	2,1	2,5	2,9	3,9	4,9
$f_1(x)$	0	2,0	2,5	3,0	4,0

$s_0=5$, $n=4$, $\Delta x=1$.

12.6. В условиях задачи **12.4** найти оптимальное распределение средств $s_0=8$.

12.7. В условиях задачи **12.4** найти оптимальное распределение средств $s_0=9$ между четырьмя предприятиями, если функция прибыли для четвертого предприятия задана следующей таблицей:

x	1	2	3	4	5	6	7	8	9
$f_4(x)$	3	5	7	11	13	15	20	22	24

12.8. В условиях задачи **12.5** найти оптимальное распределение средств $s_0=6$ между четырьмя предприятиями.

12.9. В условиях задачи **12.6** найти оптимальное распределение средств между 2-, 3- и 4-м предприятиями (1-е предприятие исключить).

В задачах **12.10–12.11** найти оптимальное распределение ресурсов s_0 между двумя отраслями производств I и II в течение n лет, если даны функции доходов $f_1(x)$ и $f_2(x)$ для каждой отрасли, функции возврата $\varphi_1(x)$ и $\varphi_2(x)$. По истечении года только все возвращенные средства перераспределяются, доход в производство не вкладывается.

12.10. $s_0=40000$ ед.; $n=4$; $f_1(x)=0,4x$; $f_2(x)=0,3x$; $\varphi_1(x)=0,5x$; $\varphi_2(x)=0,8x$.

12.11. $s_0=10000$ ед.; $n=4$; $f_1(x)=0,1x^2$; $f_2(x)=0,5x$; $\varphi_1(x)=0,75x$; $\varphi_2(x)=0,3x$.

12.12. Решить задачу **12.10** при условии, что в начале каждого года дополнительно поступают средства в размерах $\Delta s = 10000$.

В задачах **12.13–12.15** составить математическую модель, записать уравнения Беллмана и решить графически следующие задачи на определение оптимальных сроков замены оборудования. Даны: первоначальная стоимость оборудования p_0, его ликвидная стоимость $\varphi(t)$, стоимость содержания $r(t)$ в течение года оборудования возраста t лет, n — срок эксплуатации, в конце которого оборудование продается. Критерий оптимальности — суммарные затраты на эксплуатацию оборудования в течение n лет с учетом первоначальной покупки и последующей продажи.

12.13. $p_0=8000$; $\varphi(t)=p_0 2^{-t}$; $r(t)=0{,}1p_0=(t+1)$; $n=5$.

12.14. $n=5$. Стоимость нового оборудования зависит от года покупки $p_k=5000+500(k-1)$; $(k=1, 2, ..., 5)$; $\varphi(t)=p_k 2^{-t}$; $r_k(t)=0{,}1p_k(t+1)$.

12.15. $p_0=8000$; $n=5$, $\varphi(t)$ и $r(t)$ заданы таблично:

t	0	1	2	3	4	5
$\varphi(t)$	–	6000	5000	3000	1000	500
$r(t)$	600	800	1100	1500	2000	–

Глава 13. ПРИМЕНЕНИЕ ЭВМ ДЛЯ РЕШЕНИЯ ЗАДАЧ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Математические методы, изложенные в предыдущих главах рассматривались на примерах условных малоразмерных задач. При этом использовались лишь наглядные способы их решения, необходимые для иллюстрации сущности этих методов. На практике, как правило, специалисты имеют дело с многоразмерными задачами математического программирования. И здесь возникает еще одна проблема — техническая. В этом случае, чтобы найти оптимальное решение, следует проделать большое количество вычислений, которые невозможно осуществить без применения ЭВМ.

13.1. Алгоритмы решения задач

Для практического решения задачи математического программирования математический метод должен быть преобразован в алгоритм. *Алгоритм* — это такое предписание, задающее вычислительный процесс, который начинается с некоторых исходных данных и направлен на получение полностью определенного этими исходными данными результата. Для оптимизационных задач алгоритм в общих чертах описывается следующим образом. Пусть вектор X^* является оптимальным решением, т.е. X^* задает минимум (или максимум) целевой функции на множестве решений системы ограничений. К этому вектору X^* мы приближаемся последовательно, на каждом (n-м) шаге алгоритма получая вектор X_n

так, что $X_n \to X^*$. Таким образом, алгоритм решения оптимизационных задач — это последовательность формул:

$$
\begin{aligned}
&X_1=\psi_1(X_0),\\
&X_2=\psi_2(X_0, X_1),\\
&\dots\dots\dots\dots\\
&X_n=\psi_n(X_0, X_1, \dots, X_{n-1}).
\end{aligned}
\tag{13.1}
$$

Такая последовательность может быть и бесконечной; в этом случае работа программы заканчивается, когда получается решение, достаточно близкое к точному. Обычно на практике это выглядит следующим образом. Компьютер распечатывает не только величины X_n, но и значения целевой функции $F(X_n)$. В какой-то момент значения стабилизируются — каждое следующее значение с большой точностью повторяет предыдущее. Это значит, что оптимальное решение X^* практически достигнуто, и вычислительный процесс можно остановить.

Для того чтобы формулы (13.1) представляли собой действительно алгоритм, необходимо, чтобы действия, выражающие результат $(n+1)$-го шага — X_{n+1} — через результаты предыдущих шагов (X_0, X_1, ..., X_n) были доступны компьютеру. Это условие выполняется, например, если ψ_n — элементарная функция от X_0, X_1, ..., X_n, т.е. получается из основных элементарных функций с помощью конечного числа алгебраических операций и операции взятия функции от функции. В данном случае любой квалифицированный математик-программист способен по алгоритму написать текст программы, реализуемой на компьютере.

Наиболее полные и законченные результаты получены для *линейных* задач. Развитие теории линейного программирования совпало по времени с развитием ЭВМ. Такое совпадение не случайно: без ЭВМ эта теория имела бы весьма узкую область применения.

Математическим методом решения задачи линейного программирования служит симплексный метод, описанный в гл. 5. Напомним, что суть его заключается в том, что на каждом шаге одно допустимое базисное решение заменяется другим, улучшающим значение целевой функции. Таким образом, формулы (13.1) имеют вид

$$X_{n+1}=\psi_n(X_n),$$

где X_n и X_{n+1} — допустимые базисные решения.

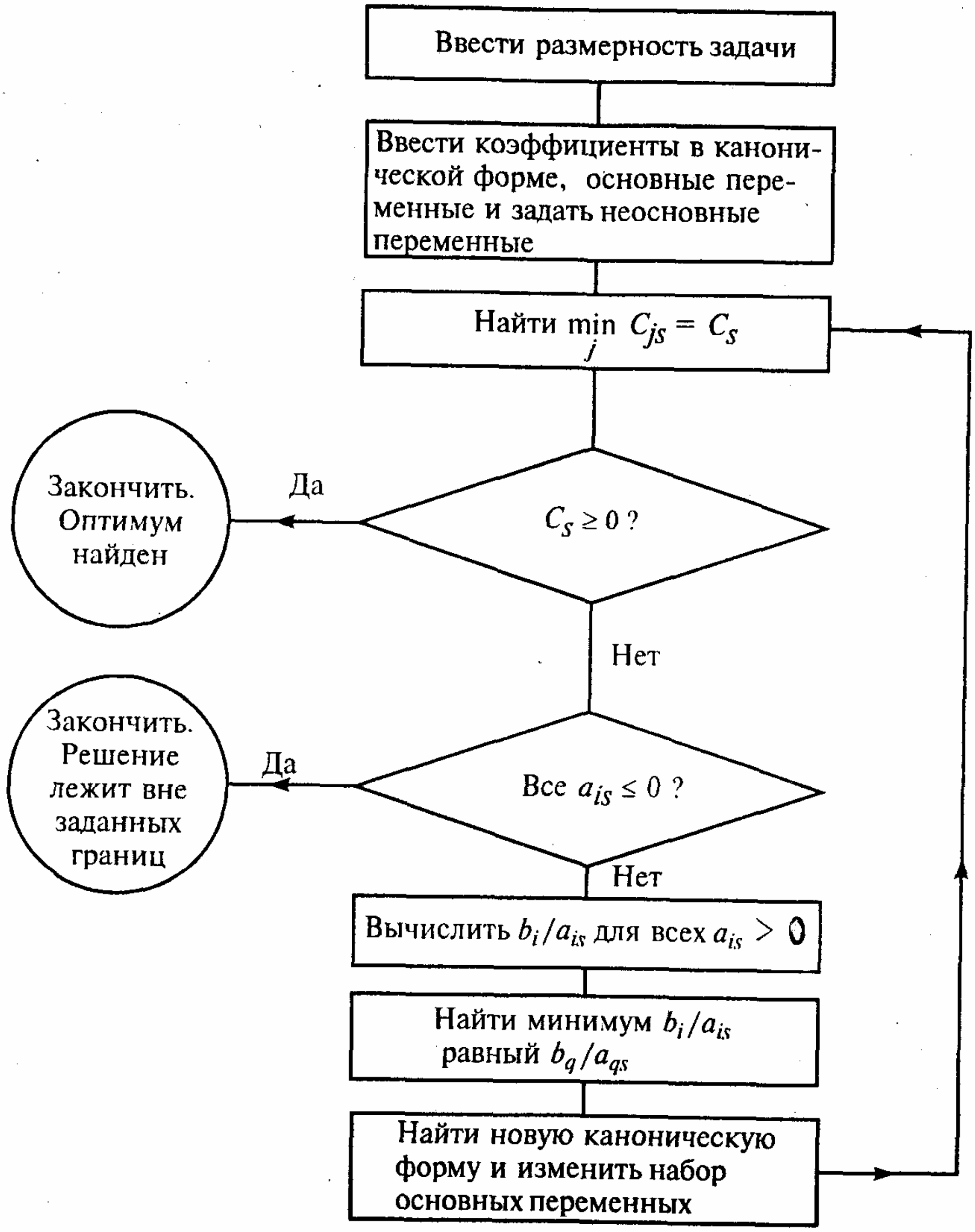

Ввести размерность задачи
Ввести коэффициенты в канонической форме, основные переменные и задать неосновные переменные
Найти $\min_j C_{js} = C_s$
$C_s \geq 0$?
Да
Закончить. Оптимум найден
Нет
Все $a_{is} \leq 0$?
Да
Закончить. Решение лежит вне заданных границ
Нет
Вычислить b_i/a_{is} для всех $a_{is} > 0$
Найти минимум b_i/a_{is} равный b_q/a_{qs}
Найти новую каноническую форму и изменить набор основных переменных

Рис. 13.1

Фактически явное описание функций ψ_n приведено в разд. 5.6 — алгоритм составления симплексных таблиц и есть по сути пошаговый алгоритм симплексного метода. Текст программы, реализующей приведенный в разд. 5.6 алгоритм, можно найти в [2]. Соответствующая программа написана на языке Бейсик и является достаточно простой: вместе с комментариями она содержит 120 строк.

На рис. 13.1 представлена схема этой программы (рассматривается задача минимизации целевой функции).

Нетрудно написать также алгоритмы отыскания допустимого базисного решения, а также алгоритмы решения транспортной задачи и других задач линейного программирования. Решения этих задач давно получены в законченном виде (текст программ можно найти, например, в [2]).

Задачи нелинейного программирования в принципе значительно более сложны: единого метода их решения не существует. Однако разработано множество *численных методов*, позволяющих получить решение с достаточной точностью (абсолютно точного решения, как правило, получить все же не удается).

Приведем некоторые примеры алгоритмов решения задач математического программирования.

Сначала рассмотрим *задачу безусловной оптимизации*, т.е. нахождение минимума нелинейной функции в отсутствие каких-либо ограничений: $F(X)\to\min$, где X — n-мерный вектор.

Основная идея алгоритма состоит в использовании следующего свойства функций нескольких переменных. Пусть X не является точкой экстремума функции $F(X)$. Тогда наибольшая скорость изменения функции достигается в направлении градиента (*см.* разд. 11.4), т.е. чтобы "быстрее" продвинуться от значения X к оптимальному X^*, следует двигаться в направлении градиента.

Алгоритм решения задачи [9] имеет следующий вид:

а) выбрать произвольный вектор X_0; $P_0=-\nabla f(X_0)$;

б) если $\nabla f(X_0)=0$, то процесс останавливается.

В противном случае

$$X_{n+1} = X_n+\lambda_n P_n,$$

$$P_{n+1}=-\nabla f(X_{n+1})+\frac{\left\|\nabla f(X_{n+1})\right\|^2}{\left\|\nabla f(X_n)\right\|^2} P_n. \qquad \textbf{(13.2)}$$

Здесь λ_n — первый неотрицательный корень уравнения

$$<\nabla f(X_n+\lambda P_n),\ P_n>=0. \qquad \textbf{(13.3)}$$

Формулы (13.2) задают простые вычислительные процедуры. Кроме того, на каждом шаге алгоритма требуется получить корень уравнения (13.3). Такая процедура сама требует алгоритмизации, однако во многих случаях уравнение (13.3) оказывается достаточно простым.

Теперь рассмотрим *задачу математического программирования* в общей постановке: найти минимум целевой функции $F(X)$, т.е.

$$F(X)\to \min \qquad \textbf{(13.4)}$$

при ограничениях

$$\begin{cases} f_i(X)\geq 0 & i=1,2,\ldots,m, \qquad \textbf{(13.5)} \\ q_j(X)=0 & j=1,2,\ldots,l. \qquad \textbf{(13.6)} \end{cases}$$

Наиболее простая идея ее решения — свести такую задачу к последовательности задач безусловной оптимизации. Эта идея осуществляется с помощью *метода штрафных функций*. Пусть $\mathcal{X}$ — множество решений системы (13.5), (13.6). Функция $\varphi(X)$ называется *штрафной функцией*, если

$$\varphi(X)\begin{cases} =0 & \text{при } X\in\mathcal{X}, \\ >0 & \text{при } X\notin\mathcal{X}. \end{cases}$$

Например, если система ограничений состоит только из одних неравенств (уравнения (13.6) отсутствуют), в качестве штрафной функции может быть выбрана функция вида $\varphi(X)=\sum_{i=1}^{n}\left[\min(f_i(X),0)\right]^2$. Рассмотрим теперь целевую функцию $\widetilde{F}(X)$, зависящую также от параметра α, и задачу безусловной оптимизации:

$$\widetilde{F}(X)=F(X)+\frac{1}{\alpha}\varphi(X)\to\min. \qquad \textbf{(13.7)}$$

Можно показать, что во многих случаях решение задачи (13.7) при $\alpha\to 0$ сходится к решению задачи (13.4)—(13.6). Поэтому для решения рассматриваемой задачи (13.4)—(13.6) целесообразно задать последовательность $\alpha_n\to 0$ и рассмотреть последовательность задач безусловной оптимизации (13.7). Соответствующие решения можно получать с помощью алгоритма (13.2) и (13.3).

Очевидно, что алгоритм решения задач нелинейного программирования существенно сложнее, чем линейного (и в частности, алгоритма симплексного метода). Более сложными оказываются, как правило, и сами программы. Кроме того, остаются и принципиальные проблемы — такие, как поиск новых алгоритмических методов. В частности, это относится к задачам выпуклого и динамического программирования.

13.2. Некоторые проблемы решения оптимизационных задач на ЭВМ

При решении задач на ЭВМ возникает проблема анализа *чувствительности алгоритма* к незначительным изменениям параметров задачи, в частности, возможная опасность так называемой *ошибки округления*. В процессе решения задач на компьютере числа округляются: простые дроби при вычислениях заменяются на десятичные. Это означает, что при каждой операции допускается какая-то ошибка. Если операций очень много, ошибка может накапливаться и в результате стать весьма значительной.

Как это получается, можно увидеть на следующем примере. Предположим, надо вычислить выражение $[(1/3+2/9)\cdot 9]^2$. Допустим, вычисления производятся с точностью до второго знака, т.е. числа округляются до второго знака после запятой. Тогда получится: $1/3=0{,}33$; $2/9=0{,}22$; $1/3+2/9=0{,}55$; $0{,}55\cdot 9=4{,}95$, $4{,}95^2=24{,}50$, в то время, как точное значение равно 25. Ошибка произошла на пол-единицы, т.е. на 2%.

Разумеется, ЭВМ вычисляет значения с гораздо бóльшей точностью — например, до 16-го знака. И все же результирующая ошибка при большом числе операций, выполняемых современными компьютерами (иногда миллиарды операций в секунду), может сказаться не только на точных значениях оптимального решения, но и на качественном результате!

Рис. 13.2 иллюстрирует, как при совсем незначительных изменениях чисел, входящих в условие задачи, может качественно измениться результат.

На рис. 13.2 область допустимых решений системы ограничений конечна и имеет конечный оптимум. Но при незначительном изменении параметров (на тысячные доли) область становится неограниченной, и целевая функция может неограниченно возрастать (рис. 13.3).

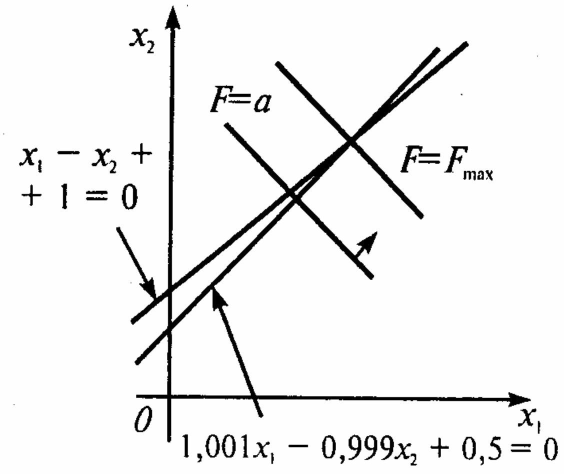


Рис. 13.2

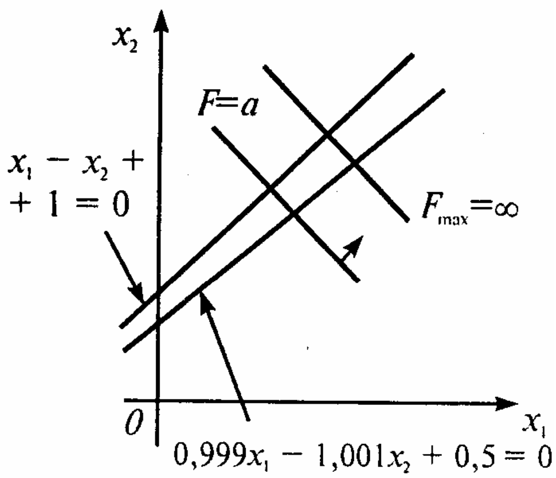


Рис. 13.3

Аналогичные (и более разнообразные) ситуации возникают при рассмотрении нелинейных задач. Таким образом, при разработке алгоритма решения задач математического программирования требуется оценить ошибку округления и в случае необходимости модифицировать алгоритм так, чтобы уменьшить ее. С этой целью создан, например, *модифицированный симплексный метод* (*МСМ*). Основное внимание в вычислительных процедурах МСМ сосредоточено именно на минимизации ошибок округления.

Другие проблемы связаны с *большой размерностью задач*. Например, если задача линейного программирования содержит до 1500 строк в матрице системы ограничений, то современный компьютер позволяет реализовать симплексный метод. Однако порой встречаются задачи от 8000 до 16000 строк. В этом случае требуются модификации известных алгоритмов. Такие модификации рассматриваются, например в [9].

Задачи модификации алгоритмов решения нелинейных задач также чрезвычайно актуальны, и в последние годы в этой области получены важные результаты.

13.3. Стандартные пакеты прикладных программ

Рассмотрим некоторые технические вопросы решения задач математического программирования на ЭВМ. Если специалисту часто приходится решать однотипные задачи (например, задачи линейного программирования), он может приобрести и установить на своем компьютере готовую и апробированную совокупность программ их решения. Такая совокупность программ называется *пакетом прикладных программ*. Основу пакета составляет выполняемый файл (в каталоге — exe-файл), который, собственно, и является программой решения задачи. Кроме него, в пакет входят также файлы, содержащие описание программы, методические указания (help-файлы) и другие файлы.

Один из стандартных пакетов прикладных программ для решения задач линейного программирования — *пакет* LP88. Этот пакет широко используется специалистами в конкретных предметных областях. Программа LP88 позволяет организовывать и решать линейные задачи, содержащие до 255 ограничений и до 2 255 переменных.

Программа LP88 функционирует на IBM PC под управлением операционной системы PC DOS версий 2.00 и выше.

Подробное описание программы LP88 дано в [7]. Там же приведены методические указания по работе с этой программой. Здесь мы отметим только некоторые ее особенности.

1. Программа работает в диалоговом режиме; вопросы пользователю высвечиваются на экране. Ввод данных осуществляется с помощью редактора LP88 (Display Editor), который легко осваивается пользователем.

2. По завершении работы программы на экране появляются все ненулевые значения компонент оптимального решения, а также значения двойственных переменных.

3. Программа позволяет вносить изменения во входные данные (коэффициенты системы ограничений и целевой функции) с помощью редактора LP88. При этом не требуется вновь вводить неизмененные входные данные.

В качестве примера рассмотрим задачу об использовании ресурсов.

Фабрика выпускает три вида тканей, причем плановое задание составляет: не менее 90 м ткани 1-го вида, 70 м — 2-го и 60 м —

3-го вида. Технологические характеристики производства приведены в табл. 13.1.

Таблица 13.1

Ресурсы	Расход ресурсов на 1 м ткани вида, единиц			Суточные запасы ресурсов, единиц
	1-го	2-го	3-го	
Оборудование	2	3	4	780
Сырье	1	4	5	850
Электроэнергия	3	4	2	790
Цена 1 м ткани, ден. единиц	80	70	60	

Поиск оптимального плана на компьютере занимает несколько секунд. После завершения работы программы экран будет выглядеть так:

```
tkani       SOLUTION is OPTIMAL                                DATE...
MAXIMUM                  ENTERS            BASIS X:  3
PIVOTS:     5            LEAVES            BASIS S:  3
LAST INV:   0            DELTA:   0 .      стоим     19075

BASIS       S.6          S.2      S.4      тк1       тк2       тк3
PRIMAL      26.25        26.25    22.25    112.5     70        8.25
DUAL        2.5          0        25       0         -37.5     0
```

В строке BASIS указаны основные переменные оптимального решения (тк1, тк2, тк3 — оптимальный план выпуска ткани 1-го, 2-го, 3-го видов, S.1, S.2, S.3 — остаток 1-го, 2-го, 3-го ресурса, S.4, S.5, S.6 — перевыполнение задания по тканям соответственно 1-го, 2-го и 3-го видов). В строке PRIMAL указаны значения основных переменных оптимального решения, в строке DUAL — значения двойственных переменных. Указатель "стоим" задает оптимальное значение целевой функции.

Подробное описание решения этой задачи приведено в [7].

С помощью функциональных клавиш <F1>—<F5> можно получать дополнительную информацию о решении, а также проводить анализ зависимости его структуры от входных данных.

Например, клавиша <F3> дает возможность получать допустимые интервалы изменения коэффициентов целевой функции:

tkani	SOLUTION IS MAXIMUM OBJECTIVE ROW RANGES		стоим 19 075	DATE... TIME...	
VARIABLE	STATUS	VALUE	стоим/UNIT	MINIMUM	MAXIMUM
тк1	BASIS	112.5	80	50	90
тк2	BASIS	70	70	NONE	107.5
тк1	BASIS	86.25	60	53.33333	160

Эта таблица указывает границы значений коэффициентов целевой функции (столбцы MINIMUM и MAXIMUM), которые обеспечивают сохранение типа переменных в оптимальном решении. Так, если цена 1 м ткани 1-го типа станет меньше 50 или больше 90 денежных единиц, набор основных переменных в оптимальном решении изменится (т.е. целевая функция будет достигать своего максимума в другой угловой точке системы ограничений).

Для задачи об использовании ресурсов при отсутствии ограничений снизу такие данные интерпретируются как пределы рентабельности производства продукции. Если переменная x_i, которая задает выпуск i-го вида продукции, перестает быть основной в оптимальном решении, это значит, что продукцию i-го вида становится невыгодно производить и ее не следует включать в оптимальный план производства.

Использование клавиши <F4> позволяет также получить интервалы устойчивости двойственных оценок, при нажатии <F4> на экране появляется запись следующего вида:

tkani	SOLUTION IS MAXIMUM RIGHT HAND SIDE RANGES		стоим 19 075	DATE... TIME...	
ROW ID	STATUS	DUAL VALUE	RHS VALUE	MINIMUM	MAXIMUM
оборуд	BINDING	2.5	780	710	796.1539
сырье	NONBINDING	0	850	823.75	NONE
эл-эн	BINDING	25	790	755	895
тк1-пл	NONBINDING	0	90	NONE	112.5
тк2-пл	BINDING	–37.5	70	0	82.35294
тк3-пл	NONBINDING	0	60	NONE	86.25

В этом случае в столбцах MINIMUM и MAXIMUM указаны граница коэффициентов правой части системы ограничений, при которых не изменяется набор основных переменных в оптимальном решении двойственной задачи.

Таким образом программа LP88 позволяет не только найти решение задачи линейного программирования, но и провести его *экономико-математический анализ* (оценку дефицитности ресурсов, рентабельности продукции, эффективности отдельных планов, сопоставление оптимальных условных затрат и результатов производства, определение норм заменяемости ресурсов и т.п.). *Возможность проведения экономико-математического анализа результатов существенно увеличивает преимущество решения задач на ЭВМ* даже для относительно простых задач линейного программирования (*см.* для сравнения, например, разд. 6.5, где элементы такого анализа проведены вручную).

Количественный анализ задачи производится с помощью пробного изменения ее исходных данных. Сначала следует выявить факторы, ограничивающие производство, т.е. имеющие ненулевые объективно обусловленные оценки. Это можно сделать с помощью клавиши <F2>. Ее использование приведет к следующей записи на экране:

tkani	SOLUTION IS MAXIMUM DUAL PROBLEM SOLUTION			стоим 19 075	
ROW ID	STATUS	DUAL VALUE	RHS VALUE	USAGE	SLACK
оборуд	BINDING	2.5	780	780	0
сырье	NONBINDING	0	850	823.75	26.25
эл-эн	BINDING	25	790	790	0
тк1-пл	NONBINDING	0	90	112.5	−22.5
тк2-пл	BINDING	−37.5	70	70	0
тк3-пл	NONBINDING	0	60	86.25	−26.5

Здесь RHS VALUE — правые части системы ограничений.

USAGE — фактическое использование ресурсов в ограничениях или фактическое выполнение плановых заданий, SLACK — разность между значениями этих столбцов. Отсюда видно (*см.* также разд. 6.5), что оборудование и электроэнергия являются дефицитными ресурсами и их увеличение приводит к повышению максимального дохода. Продукция тк2 — нерентабельная — уменьшение плана по этой продукции также увеличит максимальное значение целевой функции.

Рассмотрим, например, два вида изменений исходных данных: а) уменьшение задания по ткани 2-го вида на 10 единиц и б) увеличение запаса ресурса электроэнергии на 12 единиц.

В результате первого изменения на экране появится решение следующего вида:

MAXIMUM		ENTERS:		BASIS	X:	3
PIVOTS:	5	LEAVES:		BASIS	S:	3
LAST INV:	0	DELTA 0		стоим	19 450	
BASIS	S.6	S.2	S.4	тк1	тк2	тк3
PRIMAL	27.5	47.5	35	125	60	87.5
DUAL	2.5	0	25	0	−37.5	0

Уменьшение плана по нерентабельному виду продукции позволило перераспределить ресурсы в пользу более рентабельных — 1-го и 3-го видов. В результате целевая функция увеличилась на 375 ден. единиц.

Второе изменение приведет к следующему решению:

MAXIMUM		ENTERS:		BASIS	X:	3
PIVOTS:	5	LEAVES:		BASIS	S:	3
LAST INV:	0	DELTA 0		стоим 19 750		
BASIS	S.6	S.2	S.4	тк1	тк2	тк3
PRIMAL	24.5	56.5	41	131	60	84.5
DUAL	2.5	0	25	0	−37.5	0

Целевая функция возросла еще на 300 ден. единиц.

Программа LP88 позволяет также получать таблицу значений базисных решений и значений целевой функции задачи, что дает возможность следить за пошаговым его изменением. Программа LP88 может быть использована и в учебных целях.

Составление и отладка программ, подобной LP88, требуют немалых затрат и имеют смысл только в том случае, если спрос на готовые пакеты достаточно высок. Так, например, существуют стандартные программы решения оптимизационных задач с квадратичной целевой функцией. Если требуется решить частную задачу, то, как правило, приходится составлять программу самостоятельно, используя разработанный алгоритм. Стандартных текстовых программ, задающих решение частных математических задач, обычно не существует. Программистам бывает проще самим написать текст программы, чем приспособить чужой текст под свой компьютер.

Эффективность решения реальных многоразмерных оптимизационных задач определяется, как правило, тесным взаимодействием специалиста в предметной области (экономиста, менеджера и т.п.), математика и программиста.

РАЗДЕЛ III. СПЕЦИАЛЬНЫЕ МОДЕЛИ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ

Глава 14. МОДЕЛИ СЕТЕВОГО ПЛАНИРОВАНИЯ И УПРАВЛЕНИЯ

14.1. Назначение и области применения сетевого планирования и управления

Поиски более эффективных способов планирования сложных процессов привели к созданию принципиально новых методов сетевого планирования и управления (СПУ).

Система методов СПУ — система методов планирования и управления разработкой крупных народнохозяйственных комплексов, научными исследованиями, конструкторской и технологической подготовкой производства, новых видов изделий, строительством и реконструкцией, капитальным ремонтом основных фондов путем применения сетевых графиков.

Первые системы, использующие сетевые графики, были применены в США в конце 50-х годов и получили названия *CPM* (английская аббревиатура, означающая *метод критического пути*) и *PERT* (*метод оценки и обзора программы*). Система *CPM* была впервые применена при управлении строительными работами, система *PERT* — при разработке систем "Поларис".

В России работы по сетевому планированию начались в 60-х годах. Тогда методы СПУ нашли применение в строительстве и научных разработках. В дальнейшем сетевые методы стали широко применяться и в других областях народного хозяйства.

СПУ основано на моделировании процесса с помощью сетевого графика и представляет собой совокупность расчетных методов, организационных и контрольных мероприятий по планированию и управлению комплексом работ.

Система СПУ позволяет:

• формировать календарный план реализации некоторого комплекса работ;

• выявлять и мобилизовывать резервы времени, трудовые, материальные и денежные ресурсы;

• осуществлять управление комплексом работ по принципу "ведущего звена" с прогнозированием и предупреждением возможных срывов в ходе работ;

• повышать эффективность управления в целом при четком распределении ответственности между руководителями разных уровней и исполнителями работ.

Диапазон применения СПУ весьма широк: от задач, касающихся деятельности отдельных лиц, до проектов, в которых участвуют сотни организаций и десятки тысяч людей (например, разработка и создание крупного территориально-промышленного комплекса).

Под *комплексом работ* (*комплексом операций,* или *проектом*) мы будем понимать всякую задачу, для выполнения которой необходимо осуществить достаточно большое количество разнообразных работ. Это может быть и строительство некоторого здания, корабля, самолета или любого другого сложного объекта, и разработка проекта этого сооружения, и даже процесс построения планов реализации проекта.

Для того чтобы составить план работ по осуществлению больших и сложных проектов, состоящих из тысяч отдельных исследований и операций, необходимо описать его с помощью некоторой математической модели. Таким средством описания проектов (комплексов) является *сетевая модель.*

14.2. Сетевая модель и ее основные элементы

Сетевая модель представляет собой план выполнения некоторого комплекса взаимосвязанных работ (операций), заданного в специфической форме сети, графическое изображение которой называется *сетевым графиком*. Отличительной особенностью сетевой модели является четкое определение всех временных взаимосвязей предстоящих работ.

Главными элементами сетевой модели являются *события* и *работы.*

Термин ***работа*** используется в СПУ в широком смысле. Во-первых, это *действительная работа* — протяженный во времени процесс, требующий затрат ресурсов (например, сборка изделия, испытание прибора и т.п.). Каждая действительная работа должна быть конкретной, четко описанной и иметь ответственного исполнителя.

Во-вторых, это *ожидание* — протяженный во времени процесс, не требующий затрат труда (например, процесс сушки после покраски, старения металла, твердения бетона и т.п.).

В-третьих, это *зависимость*, или *фиктивная работа* — логическая связь между двумя или несколькими работами (событиями), не требующими затрат труда, материальных ресурсов или времени. Она указывает, что возможность одной работы непосредственно зависит от результатов другой. Естественно, что продолжительность фиктивной работы принимается равной нулю.

Событие *— это момент завершения какого-либо процесса, отражающий отдельный этап выполнения проекта.* Событие может являться частным результатом отдельной работы или суммарным результатом нескольких работ. Событие может свершиться только тогда, когда закончатся все работы, ему предшествующие. Последующие работы могут начаться только тогда, когда событие свершится. Отсюда *двойственный* характер события: для всех непосредственно предшествующих ему работ оно является конечным, а для всех непосредственно следующих за ним — начальным. При этом *предполагается, что событие не имеет продолжительности и свершается как бы мгновенно.* Поэтому каждое событие, включаемое в сетевую модель, должно быть полно, точно и всесторонне определено, его формулировка должна включать в себя результат всех непосредственно предшествующих ему работ.

Среди событий сетевой модели выделяют *исходное* и *завершающее* события. Исходное событие не имеет предшествующих работ и событий, относящихся к представленному в модели комплексу работ. Завершающее событие не имеет последующих работ и событий.

События на сетевом графике (или, как еще говорят, *на графе*) изображаются кружками (вершинами графа), а работы — стрелками (ориентированными дугами), показывающими связь между работами. Пример фрагмента сетевого графика представлен на рис. 14.1.

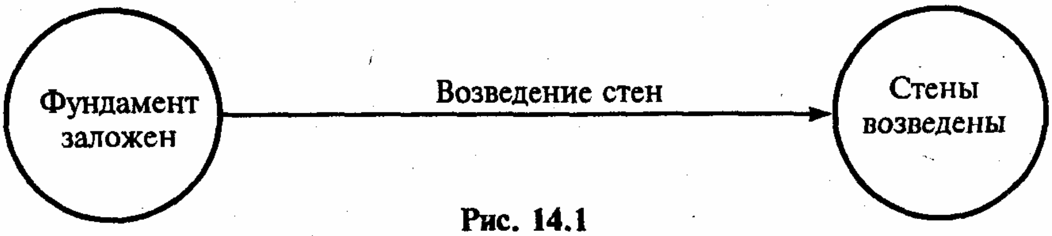


Рис. 14.1

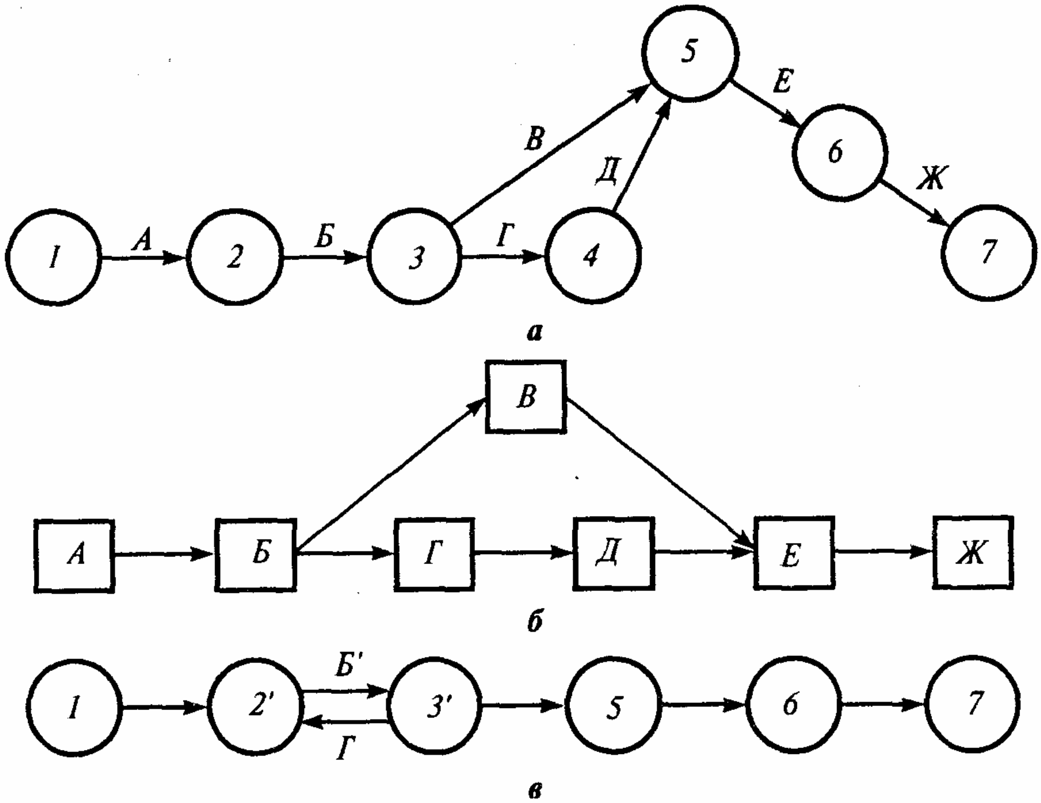


Рис. 14.2

На рис. 14.2, *а* приведен сетевой график задачи моделирования и построения оптимального плана некоторого экономического объекта. Чтобы решить эту задачу, необходимо провести следующие работы: *А* — сформулировать проблему исследования; *Б* — построить математическую модель изучаемого объекта; *В* — собрать информацию; *Г* — выбрать метод решения задачи; *Д* — построить и отладить программу для ЭВМ; *Е* — рассчитать оптимальный план; *Ж* — передать результаты расчета заказчику. Цифрами на графике обозначены номера событий, к которым приводит выполнение соответствующих работ.

Из графика, например, следует, что работы *В* и *Г* можно начать выполнять независимо одна от другой только после свершения события *3*, т.е. когда выполнены работы *А* и *Б*; работу *Д* — после свершения события *4*, когда выполнены работы *А*, *Б* и *Г*; а работу *Е* можно выполнить только после наступления события *5*, т.е при выполнении всех предшествующих ему работ *А, Б, В, Г* и *Д*.

В сетевой модели, представленной на рис. 14.2 *а*, нет числовых оценок. Такая сеть называется *структурной*. Однако на практике чаще всего используются сети, в которых заданы оценки продолжительности работ (указываемые в часах, неделях, декадах, месяцах и т.д. над соответствующими стрелками), а также оценки других параметров, например трудоемкости, стоимости и т.п. Именно такие сети мы будем рассматривать в дальнейшем.

Но прежде сделаем следующее замечание. В рассмотренных примерах сетевые графики состояли из работ и событий. Однако может быть и иной принцип построения сетей — без событий. В такой сети вершины графа (например, изображенные прямоугольниками) означают определенные работы, а стрелки — зависимости между этими работами, определяющие порядок их выполнения. В качестве примера сетевой график "события — работы" задачи моделирования и построения оптимального плана некоторого экономического объекта, приведенный на рис. 14.2 *а*, представлен в виде сети "работы — связи" на рис. 14.2 *б*. А сетевой график "события — работы" той же задачи, но с неудачно составленным перечнем работ, представлен на рис. 14.2 *в* (*см.* правило 3 в разд. 14.3).

Следует отметить, что сетевой график "работы — связи" в отличие от графика "события — работы" обладает известными преимуществами: не содержит фиктивных работ, имеет более простую технику построения и перестройки, включает только хорошо знакомое исполнителям понятие работы без менее привычного понятия события. Вместе с тем сети без событий оказываются значительно более громоздкими, так как событий обычно значительно меньше, чем работ (*показатель сложности сети*, равный отношению числа работ к числу событий, как правило, существенно больше единицы). Поэтому эти сети менее эффективны с точки зрения управления комплексом. Этим и объясняется тот факт, что (при отсутствии в целом принципиальных различий между двумя формами представления сети) в настоящее время наибольшее распространение получили сетевые графики "события — работы".

14.3. Порядок и правила построения сетевых графиков

Сетевые графики составляются на начальном этапе планирования. Вначале планируемый процесс разбивается на отдельные работы, составляется перечень работ и событий, продумываются их логические связи и последовательность выполнения, работы закрепляются за ответственными исполнителями. С их помощью оценивается длительность каждой работы. Затем составляется (*сшивается*) сетевой график. После упорядочения сетевого графика рассчитываются параметры событий и работ, определяются резервы времени и *критический путь.* Наконец, проводятся анализ и оптимизация сетевого графика, который при необходимости вычерчивается заново с пересчетом параметров событий и работ.

При построении сетевого графика необходимо соблюдать ряд правил.

1. *В сетевой модели не должно быть "тупиковых" событий, т.е. событий, из которых не выходит ни одна работа, за исключением завершающего события* (рис. 14.3 *а*). Здесь либо работа (*2*, *3*) не нужна и ее необходимо аннулировать, либо не замечена необходимость определенной работы, следующей за событием *3* для свершения какого-либо последующего события. В таких случаях необходимо тщательное изучение взаимосвязей событий и работ для исправления возникшего недоразумения.

2. *В сетевом графике не должно быть "хвостовых" событий (кроме исходного), которым не предшествует хотя бы одна работа* (событие *3* — на рис. 14.3 *б*). Здесь работы, предшествующие событию *3*, не предусмотрены. Поэтому событие *3* не может свершиться, а следовательно, не может быть выполнена и следующая за ним работа (*3*, *5*). Обнаружив в сети такие события, необходимо определить исполнителей предшествующих им работ и включить эти работы в сеть.

3. ***В сети не должно быть замкнутых контуров и петель, т.е. путей, соединяющих некоторые события с ними же самими*** (рис. 14.3 *в*, *г*).

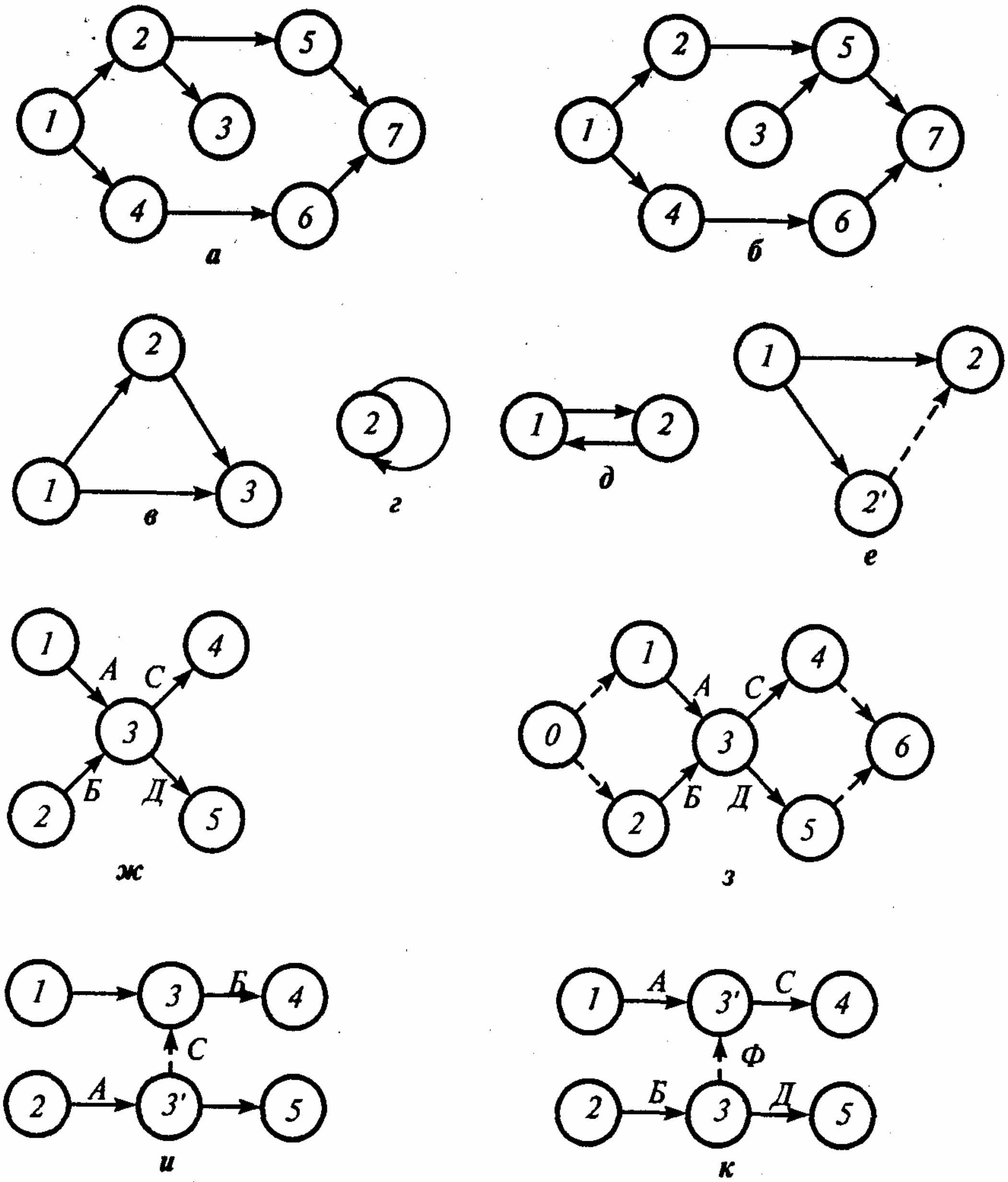


Рис. 14.3

Представим себе, что в сетевом графике, изображенном на рис. 14.2 *а*, работы *Б* (построение математической модели) и *Д* (построение и отладка программы для ЭВМ) при формулировании первоначального списка работ мы объединили бы в одну работу *Б'*. Тогда получили бы сетевой график, представленный на рис. 14.2 *в*. Событие *2'* означает, что можно переходить к работе *Б'*, которую нельзя выполнить до выбора метода расчета (работа *Г*), а выбор метода расчета нельзя начинать до окончания построения модели (событие *3'*). Другими словами, в сети образовался простейший контур: *2'→3'→2'*.

При возникновении контура (а в сложных сетях, т.е. в сетях с высоким показателем сложности, это встречается довольно часто и обнаруживается лишь при помощи ЭВМ) необходимо вернуться к исходным данным и путем пересмотра состава работ добиться его устранения. Так, в нашем примере потребовалось бы разделение работы *Б'* на *Б* и *Д*.

4. ***Любые два события должны быть непосредственно связаны не более чем одной работой-стрелкой.***

Нарушение этого условия происходит при изображении параллельно выполняемых работ (рис. 14.3 *д*). Если эти работы так и оставить, то произойдет путаница из-за того, что две различные работы будут иметь одно и то же обозначение (*1, 2*); обычно принято под (*i, j*) понимать работу, связывающую *i*-е событие с *j*-м событием. Однако содержание этих работ, состав привлекаемых исполнителей и количество затрачиваемых на работы ресурсов могут существенно отличаться.

В этом случае рекомендуется ввести *фиктивное событие* (событие *2'* на рис. 14.3 *е*) и *фиктивную работу* (работа *2', 2*), при этом одна из параллельных работ (*1, 2'*) замыкается на это фиктивное событие. Фиктивные работы изображаются на графике пунктирными линиями.

5. ***В сети рекомендуется иметь одно исходное и одно завершающее событие.*** Если в составленной сети это не так (*см.* рис. 14.3 *ж*), то добиться желаемого можно путем введения фиктивных событий и работ, как это показано на рис. 14.3 *з*.

Фиктивные работы и события необходимо вводить и в ряде других случаев. Один из них — отражение зависимости событий, не связанных с реальными работами. Например, работы *А* и *Б* (рис. 14.3 *и*) могут выполняться независимо друг от друга, но по условиям производства работа *Б* не может начаться раньше, чем окончится работа *А*. Это обстоятельство требует введения фиктивной работы *С*.

Другой случай — неполная зависимость работ. Например, работа *С* требует для своего начала завершения работ *А* и *Б*, но работа *Д* связана только с работой *Б*, а от работы *А* не зависит. Тогда требуется введение фиктивной работы *Ф* и фиктивного события *3'*, как показано на рис. 14.3 *к*.

Кроме того, фиктивные работы могут вводиться для отражения реальных отсрочек и ожидания. В отличие от предыдущих случаев здесь фиктивная работа характеризуется протяженностью во времени.

14.4. Упорядочение сетевого графика. Понятие о пути

Предположим, что при составлении некоторого проекта выделено 12 событий: *0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11* и 24 связывающие их работы: (*0, 1*), (*0, 2*), (*0, 3*), (*1, 2*), (*1, 4*), (*1, 5*), (*2, 3*), (*2, 5*), (*2, 7*), (*3, 6*), (*3, 7*), (*3, 10*), (*4, 8*), (*5, 8*), (*5, 7*), (*6, 10*), (*7, 6*), (*7, 8*), (*7, 9*), (*7, 10*), (*8, 9*), (*9, 11*), (*10, 9*), (*10, 11*). Необходимо составить и упорядочить сетевой график.

Как следует из перечня работ, исходным событием сетевого графика является событие *0* (ему не предшествуют никакие работы), а завершающим — событие *11* (за ним не следует ни одна работа). Полагая на сетевых графиках изменение времени слева направо, поместим событие *0* в левую часть графика, а событие 11 — в правую часть, разместив между ними промежуточные события в некотором порядке, соответствующем их номерам (рис. 14.4). События свяжем работами-стрелками в соответствии с перечнем работ.

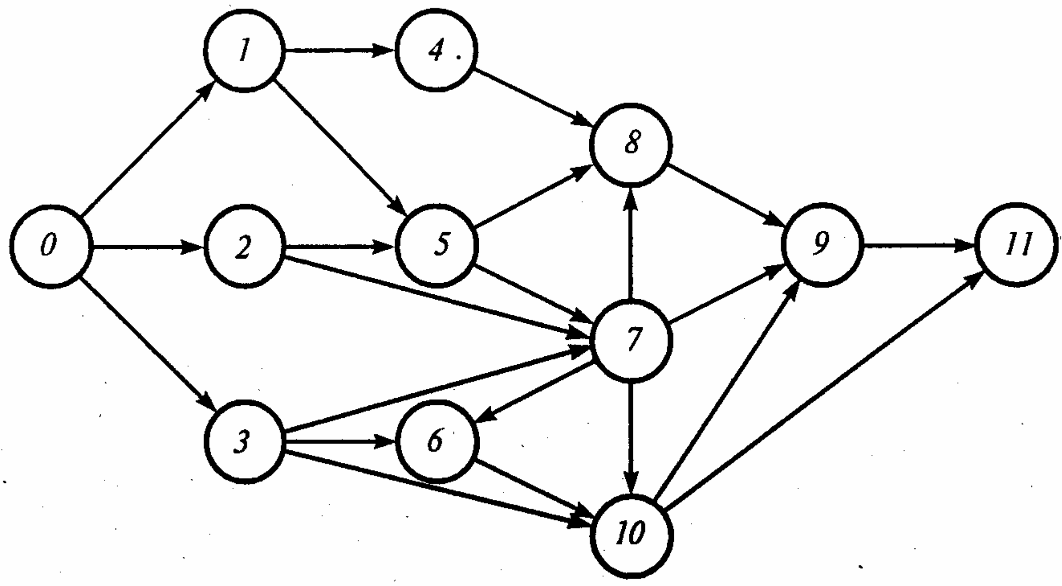


Рис. 14.4

Построенный сетевой график удовлетворяет сформулированным в разд. 14.3 правилам, предъявляемым к его построению. Однако этот график не полностью упорядочен.

Упорядочение сетевого графика *заключается в таком расположении событий и работ, при котором для любой работы предшествующее ей событие расположено левее и имеет меньший номер по сравнению с завершающим эту работу событием.* Другими словами, в упорядоченном сетевом графике все работы-стрелки направлены слева направо: от событий с меньшими номерами к событиям с бóльшими номерами.

Разобьем условно сетевой график на несколько вертикальных слоев (обводим их пунктирными линиями и обозначаем римскими цифрами).

Поместив в I слое начальное событие *0* (рис. 14.5), мысленно вычеркнем из графика (*см.* рис. 14.4) это событие и все выходящие из него работы-стрелки. Тогда без входящих стрелок останется событие *1*, образующее II слой. Вычеркнув мысленно событие *1* и все выходящие из него работы, увидим, что без входящих стрелок остаются события *4* и *2*, которые образуют III слой. Продолжая указанную процедуру вычеркивания, получим IV слой с событиями-

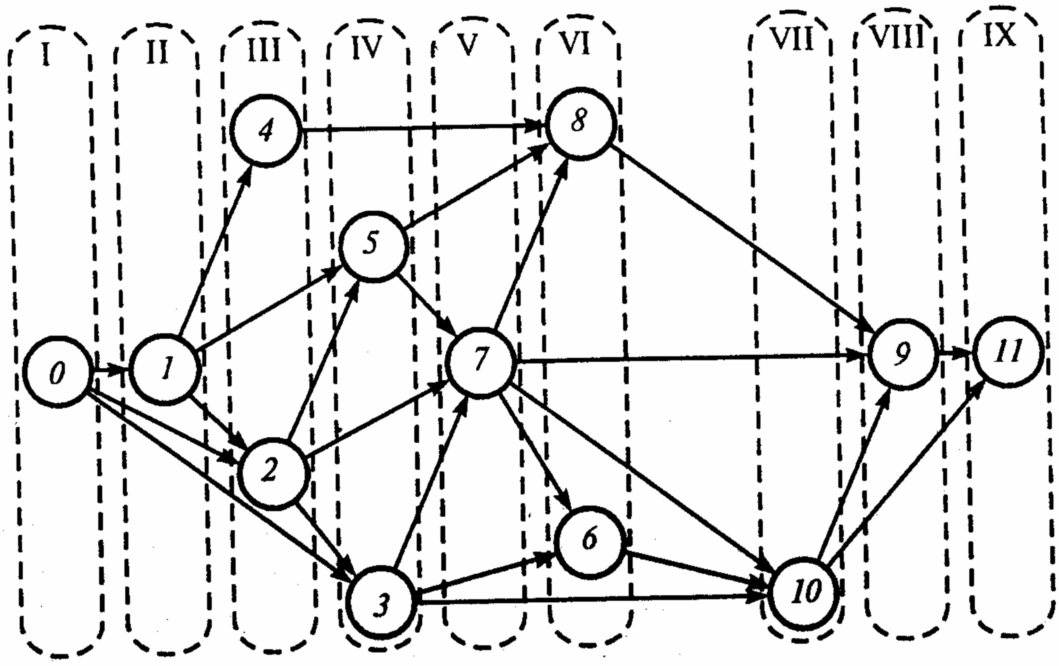


Рис. 14.5

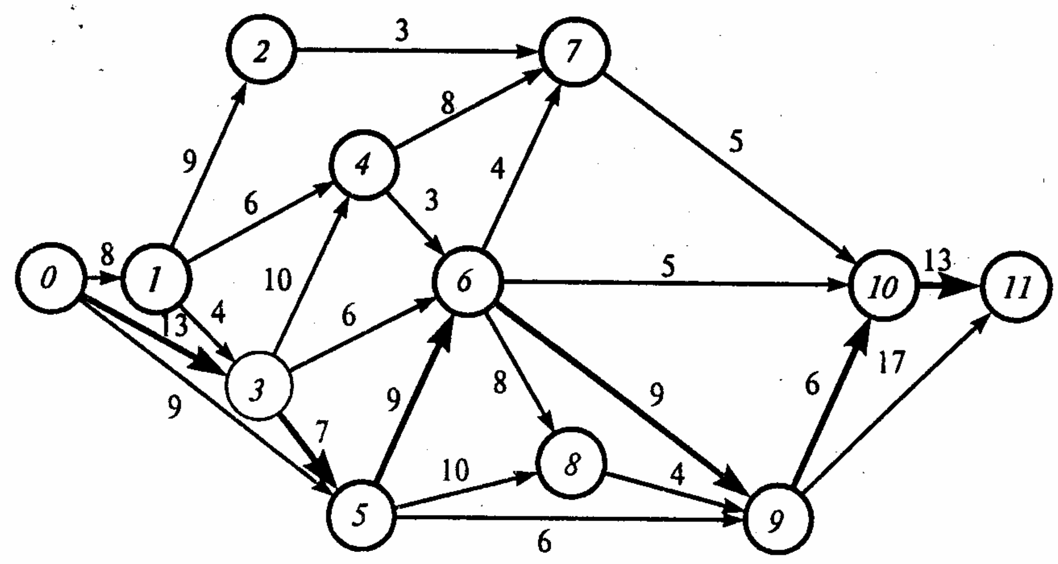


Рис. 14.6

ми *5* и *3*, V слой — с событием *7*, VI слой—с событиями *8* и *6*, VII слой — с событием *10*, VIII слой—с событием *9* и, наконец, IX слой — с событием *11*.

Теперь видим, что первоначальная нумерация событий не совсем правильная: так, событие *6* лежит в VI слое и имеет номер, меньший, чем событие *7* из предыдущего слоя. То же можно сказать о событиях *9* и *10*.

Изменим нумерацию событий в соответствии с их расположением на графике (*см.* рис. 14.5) и получим упорядоченный сетевой график[1] (рис. 14.6), в котором над стрелками указана продолжительность соответствующих работ (в сутках).

Одно из важнейших понятий сетевого графика — понятие пути. ***Путь** — любая последовательность работ, в которой конечное событие каждой работы совпадает с начальным событием следующей за ней работы.* Среди различных путей сетевого графика наибольший интерес представляет *полный путь L* — любой путь, начало которого совпадает с исходным событием сети, а конец — с завершающим.

Наиболее продолжительный полный путь в сетевом графике называется критическим. Критическими называются также работы и события, расположенные на этом пути.

---

[1] **Порядок нумерации событий, расположенных в одном вертикальном слое, принципиального значения не имеет, так что нумерация одного и того же сетевого графика может быть неоднозначной.**

Например, для рассматриваемого нами сетевого графика (*см.* рис. 14.6) полными путями будут: путь *0→1→2→7→10→11* продолжительностью 8+9+3+5+13=38 суток, путь *0→1→3→4→6→→10→11* продолжительностью 8+4+10+3+5+13=43 суток, путь *0→5→8→9→11* продолжительностью 9+10+4+17=40 суток, путь *0→3→5→6→9→10→11* продолжительностью 13+7+9+13+6+13=61 сутки и т.д.

Можно убедиться в том, что последний путь имеет наибольшую продолжительность (не только среди приведенных четырех полных путей, но и среди всех полных путей, которых в данном случае насчитывается 64), поэтому он и является критическим (способ определения критического пути, не основанного на переборе всех полных путей сетевого графика, приводится в разд. 14.5). Продолжительность критического пути составляет 61 сутки, т.е. для проведения комплекса работ понадобятся 61 сутки. Быстрее комплекс выполнить нельзя, так как для достижения завершающего события критический путь надо пройти обязательно.

Действительно, для достижения события *11* надо выполнить работу (*10, 11*), т.е. достичь события *10;* для достижения события *10* надо провести работу (*9, 10*), т.е. достичь события *9;* для достижения события *9* надо провести работу (*6, 9*), т.е. достичь события *6*, и т.д.

Определив критический путь, мы тем самым установили критические события сети *0, 3, 5, 6, 9, 10 и 11* и критические работы (*0, 3*), (*3, 5*), (*5, 6*), (*6, 9*), (*9, 10*), (*10, 11*).

Критический путь имеет особое значение в системе СПУ, так как работы этого пути определяют общий цикл завершения всего комплекса работ, планируемых при помощи сетевого графика. И для сокращения продолжительности проекта необходимо в первую очередь сокращать продолжительность работ, лежащих на критическом пути.

Следует отметить, что классический вид сетевого графика — это сеть, вычерченная без масштаба времени. Поэтому сетевой график, хотя и дает четкое представление о порядке следования работ, но недостаточно нагляден для определения тех работ, которые должны выполняться в каждый данный момент времени. В связи с этим небольшой проект после упорядочения сетевого графика рекомендуется дополнить *линейной диаграммой проекта*. Такая линейная диаграмма для рассматриваемой сети показана на рис. 14.7.

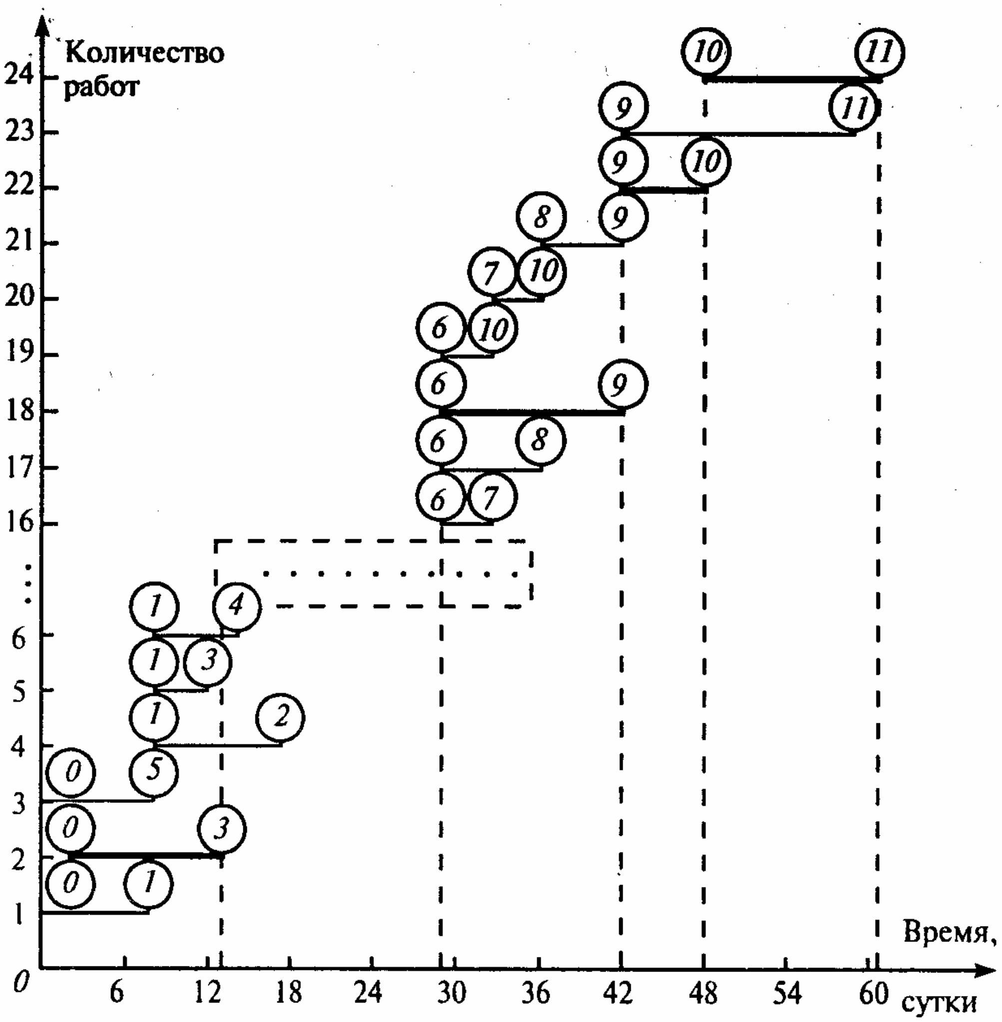


Рис. 14.7

При построении линейной диаграммы каждая работа изображается параллельным оси времени отрезком, длина которого равна продолжительности этой работы. При наличии фиктивной работы нулевой продолжительности (в рассматриваемой сети ее нет) она изображается точкой. События *i* и *j*, начало и конец работы (*i*, *j*) помещают соответственно в начале и конце отрезка. Отрезки располагают один над другим, снизу вверх в порядке возрастания индекса *i*, а при одном и том же *i* — в порядке возрастания индекса *j* (на рис. 14.7 вследствие ограниченности места не показаны работы-отрезки, выходящие из *2*-, *3*-, *4*- и *5*-го событий).

По линейной диаграмме проекта можно определить критическое время, критический путь, а также резервы времени всех работ (*см.* об этом в разд. 14.5).

Так, *критическое время комплекса работ равно координате на оси времени самого правого конца всех отрезков диаграммы:*

$$t_{кр} = t\,(11) = 61 \text{ (сутки)}.$$

Для определения критического пути рассматриваем работы-отрезки, конечные события которых совпадают с завершающим событием сети (в нашем примере (*9, 11*) и (*10, 11*). Затем находим отрезок (*9, 10*), правый конец которого лежит на одной вертикали $t\,(10)$ с левым концом одного из рассматриваемых ранее отрезков (*10, 11*). Аналогично определяем и другие работы-отрезки критического пути: (*6, 9*), ..., (*0, 3*) (на рис. 14.7 все они выделены жирным шрифтом).

14.5. Временные параметры сетевых графиков

В табл. 14.1 приведены основные временные параметры сетевых графиков.

Таблица 14.1

Элемент сети, характеризуемый параметром	Наименование параметра	Условное обозначение параметра
Событие i	Ранний срок свершения события Поздний срок свершения события Резерв времени события	$t_р(i)$ $t_п(i)$ $R(i)$
Работа (i, j)	Продолжительность работы Ранний срок начала работы Ранний срок окончания работы Поздний срок начала работы Поздний срок окончания работы Полный резерв времени работы Частный резерв времени работы первого вида Частный резерв времени работы второго вида или свободный резерв времени работы Независимый резерв времени работы	$t\,(i, j)$ $t_{рн}(i, j)$ $t_{ро}(i, j)$ $t_{пн}(i, j)$ $t_{по}(i, j)$ $R_п(i, j)$ $R_1(i, j)$ $R_с(i, j)$ $R_н(i, j)$
Путь L	**Продолжительность пути** **Продолжительность критического пути** **Резерв времени пути**	$t\,(L)$ $t_{кр}$ $R(L)$

Рассмотрим содержание и расчет указанных параметров.

Начнем с **параметров событий**. Как уже отмечалось, событие не может наступить прежде, чем свершатся все предшествующие работы. Поэтому *ранний* (*или ожидаемый*) *срок* $t_р(i)$ *свершения i-го события определяется продолжительностью максимального пути, предшествующего этому событию:*

$$t_р(i)=\max_{L_{пi}} t(L_{пi}),$$

где $L_{пi}$ — любой путь, предшествующий i-му событию, т.е. путь от исходного до i-го события сети.

Если событие j имеет несколько предшествующих путей, а следовательно, несколько предшествующих событий i, то ранний срок свершения события j удобно находить по формуле

$$t_р(j)=\max_{i,j} [t_р(i)+t(i,j)]. \quad \textbf{(14.2)}$$

Задержка свершения события i по отношению к своему раннему сроку не отразится на сроке свершения завершающего события (а значит, и на сроке выполнения комплекса работ) до тех пор, пока сумма срока свершения этого события и продолжительности (длины) максимального из последующих за ним путей не превысит длины критического пути.

Поэтому *поздний* (*или предельный*) *срок* $t_п(i)$ *свершения i-го события равен*

$$t_п(i)= t_{кр}-\max_{L_{ci}} t(L_{ci}), \quad \textbf{(14.3)}$$

где L_{ci} — любой путь, следующий за i-м событием, т.е. путь от i-го до завершающего события сети.

Если событие i имеет несколько последующий путей, а следовательно, несколько последующих событий j, то поздний срок свершения события i удобно находить по формуле

$$t_п(i)=\min_{i,j} [t_п(j) - t(i,j)]. \quad \textbf{(14.4)}$$

Резерв времени $R(i)$ ***i-го события определяется как разность между поздним и ранним сроками его свершения:***

$$R(i)=t_п(i) - t_р(i). \quad \textbf{(14.5)}$$

Резерв времени события показывает, на какой допустимый период времени можно задержать наступление этого события, не вызывая при этом увеличения срока выполнения комплекса работ.

Критические события резервов времени не имеют, так как любая задержка в свершении события, лежащего на критическом пути, вызовет такую же задержку в свершении завершающего события.

Из этого следует, что для того чтобы определить длину и топологию критического пути, вовсе не обязательно перебирать все полные пути сетевого графика и определять их длины. *Определив ранний срок наступления завершающего события сети, мы тем самым определяем длину критического пути, а выявив события с нулевыми резервами времени, определяем его топологию*[1].

▷ **14.1.** Определить временные параметры событий и критический путь для сетевого графика, изображенного на рис. 14.6.

Решение. Найденные параметры сведем в табл. 14.2.

При определении ранних сроков свершения событий $t_р(i)$ двигаемся по сетевому графику слева направо и используем формулы (14.1) и (14.2).

Для $i = 0$ (нулевого события), очевидно, что $t_р(0) = 0$. Для $i = 1$ $t_р(1) = t_р(0)+t(0,1) = 0+8 = 8$ (суток), так как для события 1 существует только один предшествующий путь $L_{п1}$ $0 \to 1$. Для $i = 2$ $t_р(2) = t_р(1)+t(1,2) = 8+9 = 17$ (суток), так как для события 2 существует только один предшествующий путь $L_{п2}$ $0 \to 1 \to 2$. Для $i = 3$ $t_р(3) = \max\{t_р(0)+t(0,3);\ t_р(1)+t(1,3)\} = \max\{0+13;\ 8+4\} = \max\{13;\ 12\} = 13$ (суток), так как для события 3 существуют два предшествующих пути $L_{п3}$ $0 \to 1 \to 3$ и $0 \to 3$ и два предшествующих события 0 и 1.

---

[1] Если сетевой график имеет единственный критический путь, то этот путь проходит через все критические события, т.е. события с нулевыми резервами времени. Если критических путей несколько, то выявление их с помощью критических событий может быть затруднено, так как через часть критических событий могут проходить как критические, так и некритические пути. В этом случае для определения критических путей рекомендуется использовать *критические работы*.

Таблица 14.2

Номер события	Сроки свершения события, сутки		Резерв времени $R(i)$, сутки
	ранний $t_р(i)$	поздний $t_п(i)$	
0	0	0	0
1	8	9	1
2	17	40	23
3	13	13	0
4	23	26	3
5	20	20	0
6	29	29	0
7	33	43	10
8	37	38	1
9	42	42	0
10	48	48	0
11	61	61	0

Аналогично:

$t_р(4) = \max\{t_р(1)+t(1,4);\ t_р(3)+t(3,4)\} = \max\{8+6;\ 13+10\} = \max\{14;\ 23\} = 23$ (суткам);

$t_р(5) = \max\{t_р(3)+t(3,5);\ t_р(0)+t(0,5)\} = \max\{13+7;\ 0+9\} = \max\{20;\ 9\} = 20$ (суткам);

$t_р(6) = \max\{t_р(4)+t(3,4);\ t_р(3)+t(3,6);\ t_р(5)+t(5,6)\} = \max\{23+3;\ 13+6;\ 20+9\} = \max\{26;\ 19;\ 29\} = 29$ (суткам) и т.д.

Длина критического пути равна раннему сроку свершения завершающего собьггия *11* (*см*. табл. 14.2):

$$t_{кр}=t_р(11)=61 \text{ (суткам)}.$$

При определении поздних сроков свершения событий $t_п(i)$ двигаемся по сети в обратном направлении, т.е. справа налево, и используем формулы (14.3) и (14.4).

Для $i=11$ (завершающего события) поздний срок свершения события должен равняться его раннему сроку (иначе изменится длина критического пути): $t_п(11)=t_р(11)=61$ (сутки).

Для $i = 10$ $t_п(10) = t_р(11) - t(10, 11) = 61-13 = 48$ (суток), так как для события 10 существует только один последующий путь L_{c10}: $10 \to 11$.

Для $i = 9$ $t_п(9) = \min\{t_п(10) - t(9,10); t_п(11) - t(9,10)\} = \min\{48-6; 61-17\} = \min\{42; 44\} = 42$ (суткам), так как для события 9 существуют два последующих пути L_{c9}: $9 \to 10 \to 11$ и $9 \to 11$ и два последующих события 10 и 11.

Аналогично:

$t_п(8) = t_п(9) - t(8, 9) = 42-4 = 38$ (суток);

$t_п(7) = t_п(10) - t(7, 10) = 48-5 = 43$ (суткам);

$t_п(6) = \min\{t_п(7) - t(6,7); t_п(10) - t(6,10); t_п(9) - t(6, 9); t_п(8) - t(6,8)\} = \min\{43-4; 48-5; 42-13; 38-8\} = \min\{39; 43; 29; 30\} = 29$ (суткам) и т.д.

По формуле (14.5) определяем резервы времени i-го события:

$R(0) = 0$; $R(1) = 9-8 = 1$; $R(2) = 40-17 = 23$ и т.д.

Резерв времени, например, события 2 — $R(2) = 23$ — означает, что время свершения события 2 может быть задержано на 23 суток без увеличения общего срока выполнения проекта. Анализируя табл. 14.2, видим, что не имеют резервов времени события 0, 3, 5, 6, 9, 11. Эти события и образуют критический путь (на рис. 14.6 он выделен жирным шрифтом).▶

Теперь перейдем к **параметрам работ.**

Отдельная работа может начаться (и окончиться) в ранние, поздние или другие промежуточные сроки. В дальнейшем при оптимизации графика возможно любое размещение работы в заданном интервале.

Очевидно, что ранний срок $t_{рн}(i, j)$ начала работы (i, j) совпадает с ранним сроком наступления начального (предшествующего) события i, т.е.

$$t_{рн}(i, j) = t_р(i). \tag{14.6}$$

Тогда ранний срок $t_{ро}(i, j)$ окончания работы (i, j) определяется по формуле

$$t_{ро}(i, j) = t_р(i) + t(i, j). \tag{14.7}$$

Ни одна работа не может окончиться позже допустимого позднего срока своего конечного события i. Поэтому поздний срок $t_{по}(i, j)$ окончания работы (i, j) определяется соотношением

$$t_{по}(i, j) = t_н(j), \tag{14.8}$$

а поздний срок $t_{пн}(i, j)$ начала этой работы — соотношением

$$t_{пн}(i, j) = t_н(j) - t(i, j). \qquad \textbf{(14.9)}$$

Таким образом, в рамках сетевой модели моменты начала и окончания работы тесно связаны с соседними событиями ограничениями (14.6)—(14.9).

Прежде чем рассматривать резервы времени работ, обратимся к резерву времени пути. Такие резервы имеют все некритические пути. ***Резерв времени пути*** *$R(L)$ определяется как разность между длиной критического и рассматриваемого пути*

$$R(L) = t_{кр} - t(L). \qquad \textbf{(14.10)}$$

Он показывает, на сколько в сумме могут быть увеличены продолжительности всех работ, принадлежащих этому пути. Если затянуть выполнение работ, лежащих на этом пути, на время большее чем $R(L)$, то критический путь переместится на путь L.

Отсюда можно сделать вывод, что *любая из работ пути L на его участке, не совпадающем с критическим путем (замкнутым между двумя событиями критического пути), обладает резервом времени.*

Среди резервов времени работ выделяют четыре разновидности.

Полный резерв времени *$R_п(i, j)$ работы (i, j) показывает, на сколько можно увеличить время выполнения данной работы при условии, что срок выполнения комплекса работ не изменится. Полный резерв $R_п(i, j)$ определяется по формуле*

$$R_п(i, j) = t_п(j) - t_р(i) - t(i, j). \qquad \textbf{(14.11)}$$

Полный резерв времени работы равен резерву максимального из путей, проходящего через данную работу. Этим резервом можно располагать при выполнении данной работы, если ее начальное событие свершится в самый ранний срок, и можно допустить свершение конечного события в его самый поздний срок (рис. 14.8 *а*).

Важным свойством полного резерва времени работы является то, что он принадлежит не только этой работе, но и всем полным путям, проходящим через нее. При использовании полного резерва времени только для одной работы резервы времени остальных работ, лежащих на максимальном пути, проходящем через нее, будут полностью исчерпаны. Резервы времени работ, лежащих на других (немаксимальных по длительности) путях, проходящих через эту работу, сократятся соответственно на величину использованного резерва.

Остальные резервы времени работы являются частями ее полного резерва.

Частный резерв времени первого вида R_1 *работы* (i, j) *есть часть полного резерва времени, на которую можно увеличить продолжительность работы, не изменив при этом позднего срока ее начального события.* Этим резервом можно располагать при выполнении данной работы в предположении, что ее начальное и конечное события свершаются в свои самые поздние сроки (*см.* рис. 14.8 *б*).

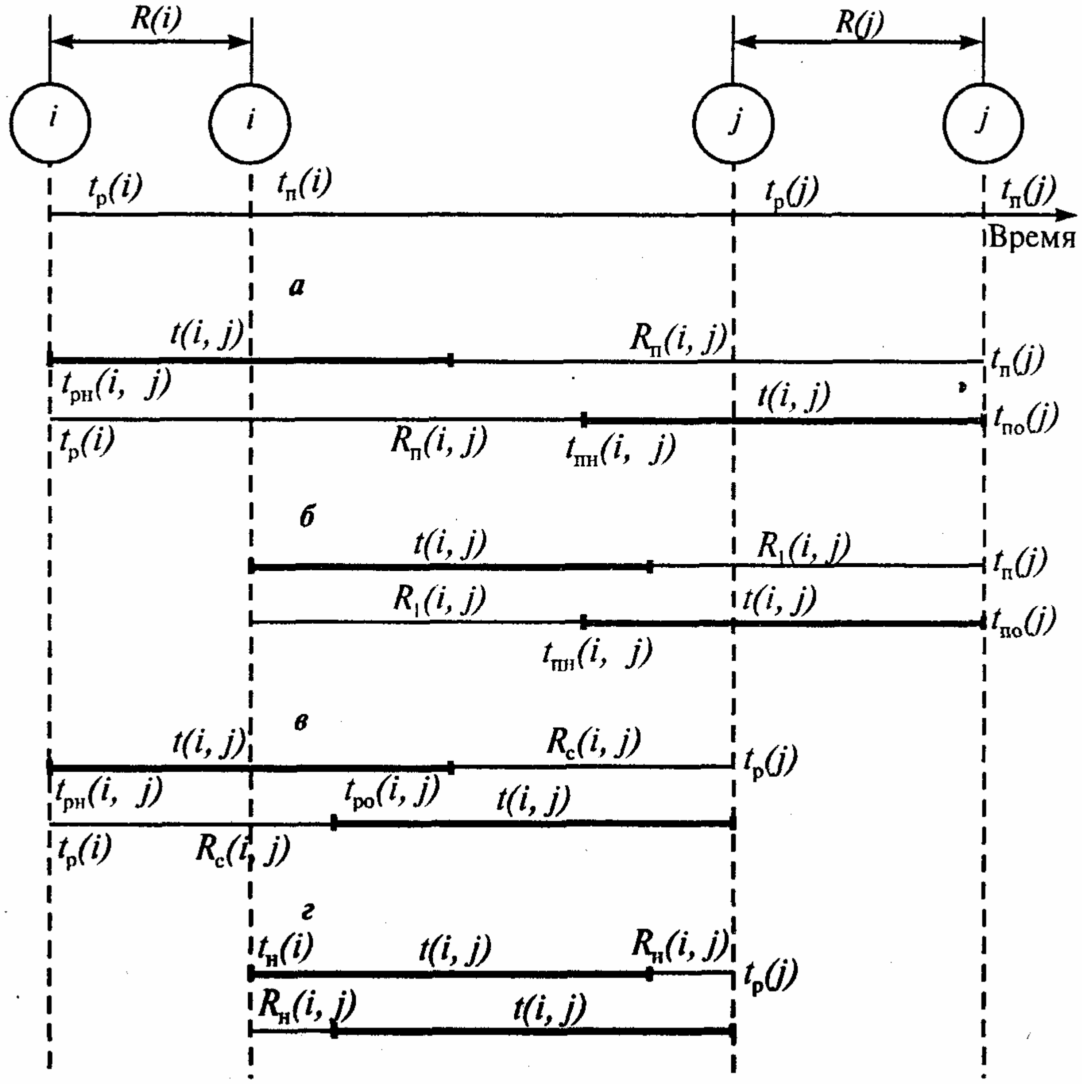


Рис. 14.8

R_1 находится по формуле

$$R_1(i,j)=t_п(j)-t_п(i)-t(i,j), \tag{14.12}$$

$$R_1(i,j)=R_п(i,j)-R(i). \tag{14.13}$$

Частный резерв времени второго вида, или ***свободный резерв времени*** $R_с$ *работы* (i, j) *представляет часть полного резерва времени, на которую можно увеличить продолжительность работы, не изменив при этом раннего срока ее конечного события.* Этим резервом можно располагать при выполнении данной работы в предположении, что ее начальное и конечное события свершатся в свои самые ранние сроки (*см.* рис. 14.8 *в*). $R_с$ находится по формуле

$$R_с(i,j)=t_р(j)-t_р(i)-t(i,j), \tag{14.14}$$

$$R_с(i,j)=R_п(i,j)-R(j). \tag{14.15}$$

Свободным резервом времени можно пользоваться для предотвращения случайностей, которые могут возникнуть в ходе выполнения работ. Если планировать выполнение работ по ранним срокам их начала и окончания, то всегда будет возможность при необходимости перейти на поздние сроки начала и окончания работ.

Независимый резерв времени $R_н$ *работы* (i, j)[1] *— часть полного резерва времени, получаемая для случая, когда все предшествующие работы заканчиваются в поздние сроки, а все последующие работы начинаются в ранние сроки* (*см.* рис. 14.8 *г*)

$$R_н(i,j)=t_р(j)-t_п(i)-t(i,j), \tag{14.16}$$

или

$$R_н(i,j)=R_п(i,j)-R(i). \tag{14.17}$$

Использование независимого резерва времени не влияет на величину резервов времени других работ. Независимые резервы стремятся использовать тогда, когда окончание предыдущей работы произошло в поздний допустимый срок, а последующие работы хотят выполнить в ранние сроки. Если величина независимого резерва, определяемая по формуле (14.16) или (14.17), равна нулю или положительна, то такая возможность есть. Если же величина

---

[1] В ряде работ по сетевому планированию резерв времени $R_н(i, j)$ называют *свободным*, а резерв $R_с(i, j)$ специального названия не имеет.

$R_н(i, j)$ отрицательна, то этой возможности нет, так как предыдущая работа еще не оканчивается, а последующая уже должна начаться. Поэтому отрицательное значение $R_н(i, j)$ не имеет реального смысла. А фактически независимый резерв имеют лишь те работы, которые не лежат на максимальных путях, проходящих через их начальные и конечные события.

Следует отметить, что резервы времени работы (i, j), показанные на рис. 14.8, могут состоять из двух временны́х отрезков, если интервал продолжительности работы $t(i, j)$ занимает промежуточную позицию между двумя его крайними положениями, изображенными на графиках.

Таким образом, *если частный резерв времени первого вида может быть использован на увеличение продолжительности данной и последующих работ без затрат резерва времени предшествующих работ, а свободный резерв времени — на увеличение продолжительности данной и предшествующих работ без нарушения резерва времени последующих работ, то независимый резерв времени может быть использован для увеличения продолжительности только данной работы.*

Работы, лежащие на критическом пути, так же как и критические события, резервов времени не имеют[1].

Если на критическом пути лежит начальное событие i, то

$$R_п(i, j)=R_1(i, j). \qquad \textbf{(14.18)}$$

Если на критическом пути лежит конечное событие j, то

$$R_п(i, j)=R_с(i, j). \qquad \textbf{(14.19)}$$

Если на критическом пути лежат начальное и конечное события i и j, но сама работа не принадлежит этому пути, то

$$R_п(i, j)=R_1(i, j)=R_с(i, j)=R_н(i, j). \qquad \textbf{(14.20)}$$

Соотношения (14.18)—(14.20) можно использовать при проверке правильности расчетов резервов времени отдельных работ.

---

[1] С помощью критических работ, т.е. работ, не имеющих резервов времени, может быть определен критический путь сетевого графика. Этот способ определения критического пути целесообразно использовать тогда, когда сеть содержит несколько критических путей.

▷ **14.2.** Вычислить временные параметры работ для сетевого графика, изображенного на рис. 14.6.

Результаты расчетов сведем в табл. 14.3.

Вычисление временных параметров работы (i, j) покажем на примере работы $(1, 4)$:

ранний срок начала работы (по формуле (14.6)): $t_{рн}(1, 4) = t_р(1) = 8$ (суток),

ранний срок окончания работы (по формуле (14.7)): $t_{ро}(1, 4) = t_р(1)+t(1,4) = 8+6 = 14$ (суток);

поздний срок начала работы (по формуле (14.9)): $t_{пн}(1, 4) = t_п(4)-t(1,4) = 26-6 = 20$ (суток), где $t_п(4) = 26$ (*см.* табл. 14.2);

поздний срок окончания работы (по формуле (14.8)): $t_{по}(1, 4) = t_п(4) = 26$ (суток).

Таким образом, работа $(1, 4)$ должна начаться в интервале [8; 28] (суток) и окончиться в интервале [12; 26] (суток) от начала выполнения проекта.

Полный резерв работы $(1, 4)$ (по формуле (14.11)): $R_п(1, 4) = t_п(4)-t_р(1)-t(1,4) = 26-8-6=12$ (суток), т.е. срок выполнения данной работы можно увеличить на 12 суток, при этом срок выполнения комплекса работ не изменится.

Таблица 14.3

№ п/п	Работа (i, j)	Продолжительность работы $t(i, j)$	Сроки начала и окончания работы				Резервы времени работы			
			$t_{рн}(i,j)$	$t_{ро}(i,j)$	$t_{пн}(i,j)$	$t_{по}(i,j)$	$R_п(i,j)$	$R_1(i, j)$	$R_с(i, j)$	$R_н(i,j)$
1	(0,1)	8	0	8	1	9	1	1	0	0
2	(0,3)	13	0	13	0	13	0	0	0	0
3	(0,5)	9	0	9	11	20	11	11	11	11
4	(1,2)	9	8	17	31	40	23	22	0	-
5	(1,4)	6	8	14	20	26	12	11	9	8
6	(1,3)	4	8	12	9	13	1	0	1	0
7	(2,7)	3	17	20	40	43	23	0	13	-
8	(3,4)	10	13	23	16	26	3	3	0	0
9	(3,5)	7	13	20	13	20	0	0	0	0
10	(3,6)	6	13	19	23	29	10	10	10	10
11	(4,7)	8	23	31	35	43	12	9	2	-
12	(4,6)	3	23	26	26	29	3	0	3	0
13	(5,6)	9	20	29	20	29	0	0	0	0
14	(5,8)	10	20	30	28	38	8	8	7	7

№ п/п	Работа (i, j)	Продолжительность работы $t(i, j)$	Сроки начала и окончания работы				Резервы времени работы			
			$t_{рн}(i,j)$	$t_{ро}(i,j)$	$t_{пн}(i,j)$	$t_{по}(i,j)$	$R_п(i,j)$	$R_1(i, j)$	$R_c(i, j)$	$R_н(i,j)$
15	(*5,9*)	6	20	26	36	42	16	16	16	16
16	(*6,7*)	4	29	33	39	43	10	10	0	0
17	(*6,10*)	5	29	34	43	48	14	14	14	14
18	(*6,9*)	13	29	42	29	42	0	0	0	0
19	(*6,8*)	8	29	37	30	38	1	1	0	0
20	(*7,10*)	5	33	38	43	48	10	0	10	0
21	(*8,9*)	4	37	41	38	42	1	0	1	0
22	(*9,10*)	6	42	48	42	48	0	0	0	0
23	(*9,11*)	17	42	59	44	61	2	2	2	2
24	(*10,11*)	13	48	61	48	61	0	0	0	0

Покажем на примере работы (*1*, *4*), что полный резерв времени работы равен продолжительности максимального из путей, проходящих через данную работу.

Через работу (*1*, *4*) проходят семь полных путей (*см.* рис. 14.6):

	Путь	*Продолжительность, сутки*
L_1	$0 \to 1 \to 4 \to 6 \to 7 \to 10 \to 11$	39
L_2	$0 \to 1 \to 4 \to 6 \to 8 \to 9 \to 10 \to 11$	48
L_3	$0 \to 1 \to 4 \to 6 \to 8 \to 9 \to 11$	46
L_4	$0 \to 1 \to 4 \to 6 \to 9 \to 10 \to 11$	49
L_5	$0 \to 1 \to 4 \to 6 \to 9 \to 11$	47
L_6	$0 \to 1 \to 4 \to 6 \to 10 \to 11$	35
L_7	$0 \to 1 \to 4 \to 7 \to 10 \to 11$	40

Отсюда максимальным из путей, проходящих через работу (*1*, *4*), является путь L_4 продолжительностью 49 (суток), резерв времени которого (по формуле (14.10)) $R(L_4)=61-49=12$ (суток).

Как видим, полный резерв времени работы (*1*, *4*) оказался равным резерву пути L_4 — максимального из путей, проходящих через эту работу. Если увеличить продолжительность выполнения работы $t(1, 4)$ на 12 суток, т.е. с 6 до 18 суток, то полностью будет исчерпан резерв времени пути L_4, т.е. этот путь станет также критическим, а резервы времени других путей уменьшатся соответственно на 12 суток.

Частный резерв времени работы $(1, 4)$ первого вида определим по формуле (14.12) (или по формуле 14.13): $R_1(1, 4) = t_п(4)-t_п(1)-t(1, 4) = 26-9-6 = 11$ (суток) (или $R_1(1, 4) = R_п(1, 4)-R(1) = 12-1 = 11$ (суток), т.е. при сохранении общего срока выполнения проекта на 11 суток может быть задержано выполнение работы $(1, 4)$ и последующих работ (по любому из путей $L_1, L_2, ..., L_7$) без затрат резерва времени предшествующих ей работ (в данном случае без затрат резерва времени одной предшествующей работы $(0, 1)$).

Частный резерв времени второго вида, или свободный резерв времени, работы $(1, 4)$ найдем по формуле (14.14) (или 14.15)): $R_c(1, 4) = t_р(4)-t_р(1)-t(1, 4) = 23-8-6 = 9$ (суток) (или $R_c(1, 4) = R_п(1, 4)-R(1) = 12-3 = 9$ (суток), т.е. при сохранении общего срока выполнения проекта на 9 суток может быть задержано выполнение работы $(1, 4)$ и предшествуюших ей работ (в данном случае работы $(0, 1)$) без нарушения резерва времени последующих работ.

Независимый резерв времени работы $(1, 4)$ определим по формуле (14.16) (или (14.17)): $R_н(1, 4) = t_р(4)-t_п(1)-t(1, 4) = 23-9-6 = 8$ (суток) (или $R_н(1,4) = R_п(1,4)-R(1)-R(4) = 12-1-3 = 8$ (суток), т.е. на 8 суток может быть увеличена продолжительность работы $(1, 4)$ без изменения резервов времени всех остальных работ.

Обратим внимание на то, что независимые резервы работ $(1, 2)$, $(2, 7)$ и $(4, 7)$ отрицательны (в табл. 14.3 они обозначены прочерком). Например $R_н(2, 7) = t_р(7)-t_п(2)-t(2, 7) = 33-40-3 = -10$. Это означает, что работа $(2, 7)$ продолжительностью 3 (суток) должна закончиться на 33-и сутки после начала комплекса работ, а начаться на 40-е сутки, что, естественно, невозможно.

Подчеркнем, что резервы критических работ $(0, 3)$, $(3, 5)$, $(5, 6)$, $(6, 9)$, $(9, 10)$, $(10, 11)$, так же как и резервы критических событий, равны нулю.▶

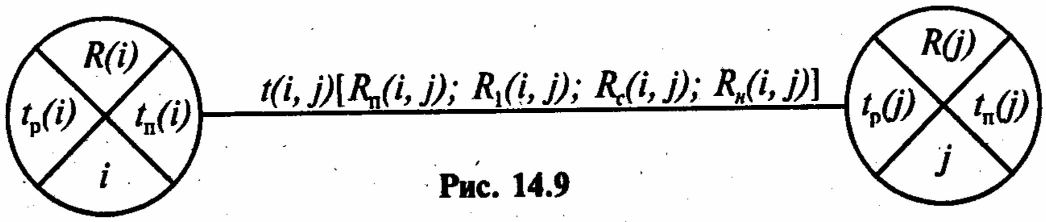


Рис. 14.9

Следует отметить, что в случае достаточно простых сетевых графиков результаты расчета их временных параметров можно фиксировать прямо на графике. Параметры событий записываются в кружках, разделенных на четыре части, а параметры работ — над соответствующими стрелками (рис. 14.9). При этом отпадает необходимость составления таблиц.

14.6. Сетевое планирование в условиях неопределенности

При определении временных параметров сетевого графика до сих пор предполагалось, что время выполнения каждой работы точно известно. Такое предположение в действительности выполняется редко: напомним, система СПУ обычно применяется для планирования сложных разработок, не имевших в прошлом никаких аналогов. Чаще всего продолжительность работы по сетевому графику заранее не известна и может принимать лишь одно из ряда возможных значений. Другими словами, продолжительность работы $t(i, j)$ является случайной величиной, характеризующейся своим законом распределения, а значит, своими числовыми характеристиками — *средним значением*, или *математическим ожиданием*, $\bar{t}(i, j)$ и *дисперсией* $\sigma^2(i, j)$.

Практически во всех системах СПУ априори принимается, что распределение продолжительности работ обладает тремя свойствами: а) непрерывностью; б) унимодальностью, т.е. наличием единственного максимума у кривой распределения; в) двумя точками пересечения кривой распределения с осью Ox, имеющими неотрицательные абсциссы.

Кроме того, установлено, что распределение продолжительности работ обладает *положительной асимметрией*, т.е. максимум кривой смещен влево относительно медианы (линии, делящей площадь под кривой на две равные части). Распределение, как правило, более круто поднимается при удалении от минимального значения t и полого опускается при приближении к максимальному значению t (рис.14.10).

Простейшим распределением с подобными свойствами является известное в математической статистике β-*распределение*. Анализ большого количества статистических данных (хронометражи времени реализации отдельных работ, нормативные данные и т.д.)

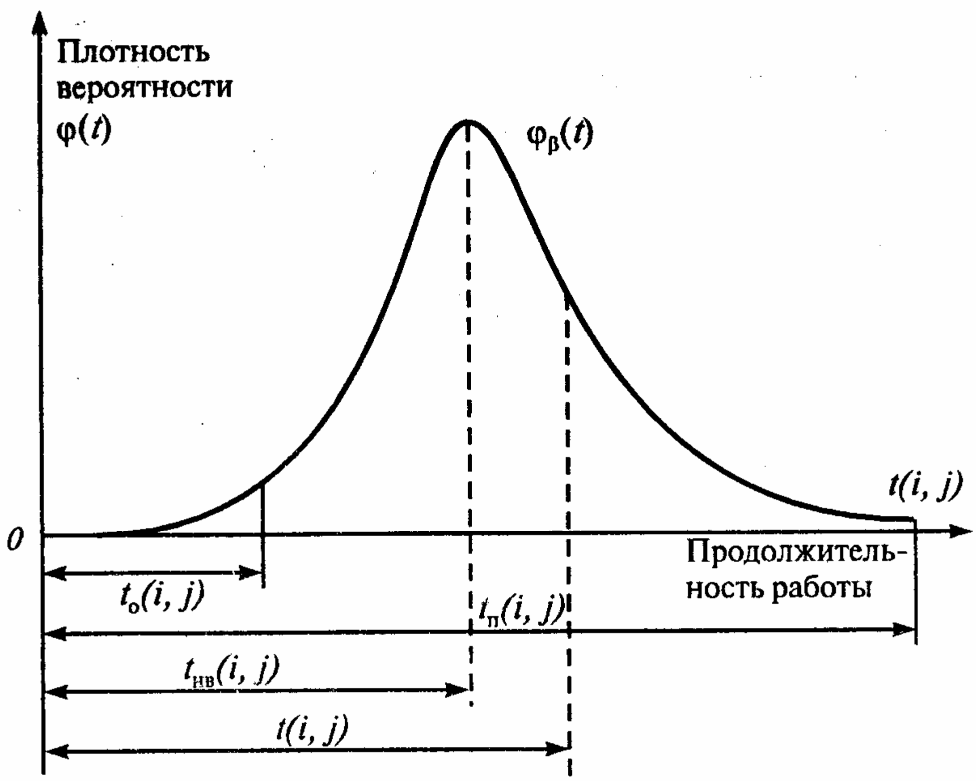


Рис. 14.10

показывает, что β-*распределение* можно использовать в качестве априорного для всех работ.

Для определения числовых характеристик $\bar{t}\,(i,\ j)$, и $\sigma^2(i,\ j)$ этого распределения для работы (i, j) на основании опроса ответственных исполнителей проекта и экспертов определяют три временны́е оценки (рис. 14.10):

а) *оптимистическую оценку* $t_о(i,\ j)$, т.е. продолжительность работы (i, j) при самых благоприятных условиях;

б) *пессимистическую оценку* $t_п(i,\ j)$, т.е. продолжительность работы (i, j) при самых неблагоприятных условиях;

в) *наиболее вероятную оценку* $t_{нв}(i,\ j)$, т.е. продолжительность работы (i, j) при нормальных условиях.

Предположение о β-распределении продолжительности работы (i, j) позволяет получить следующие оценки ее числовых характеристик:

$$\bar{t}\,(i, j) = \frac{t_о(i, j) + 4t_{нв}(i, j) + t_п(i, j)}{6}; \qquad (14.21)$$

$$\sigma^2(i, j)=\left[\frac{t_п(i,j)-t_о(i,j)}{6}\right]^2. \qquad \textbf{(14.22)}$$

Следует отметить, что обычно специалистам сложно оценить наиболее вероятное время выполнения работы $t_{нв}(i, j)$. Поэтому в реальных проектах используется упрощенная (и менее точная) оценка средней продолжительности работы (i, j) на основании лишь двух задаваемых временны́х оценок $t_о(i, j)$ и $t_п(i, j)$:

$$\bar{t}\,(i,j)=\frac{2t_о(i,j)+3t_п(i,j)}{5}. \qquad \textbf{(14.23)}$$

Зная $\bar{t}\,(i, j)$ и $\sigma^2(i, j)$, можно определять временны́е параметры сетевого графика и оценивать их надежность.

Так, при достаточно большом количестве работ, принадлежащих пути L, и выполнении некоторых весьма общих условий можно применить центральную предельную теорему Ляпунова, на основании которой можно утверждать, что общая продолжительность пути L имеет нормальный закон распределения со средним значением $\bar{t}\,(L)$, равным сумме средних значений продолжительности составляющих его работ $\bar{t}\,(i, j)$ и дисперсией $\sigma^2(L)$, равной сумме соответствующих дисперсий $\sigma^2(i, j)$:

$$\bar{t}\,(L)=\sum_{i,j}\bar{t}\,(i,j); \qquad \textbf{(14.24)}$$

$$\sigma^2(L)=\sum_{i,j}\sigma^2(i,j). \qquad \textbf{(14.25)}$$

Предположим, что сетевой график на рис. 14.6 представляет сеть не с детерминированными (фиксированными), а со случайными продолжительностями работ и цифры над работами-стрелками указывают средние значения $\bar{t}\,(i, j)$ продолжительности соответствующих операций, найденные по формуле (14.21) или (14.23), и известны все дисперсии $\sigma^2(i, j)$, вычисленные по формуле (14.22).

Следует отметить, что и в этом случае временны́е параметры сетевого графика — длина критического пути, ранние и поздние сроки свершения событий, резервы времени событий и работ и т.д. — будут такие же, как и найденные в разд. 14.5. Но при этом необходимо учесть, что эти параметры, представленные в табл. 14.2 и 14.3, теперь будут являться *средними* значениями соответствующих случайных величин: средней длиной критического пути $\bar{t}_{кр}$, средним значением раннего срока наступления события $\bar{t}_р(i)$, средним значением полного резерва времени работы $\overline{R}_п(i, j)$ и т.п.

Так, $\bar{t}_{кр}$=61 будет означать, что длина критического пути лишь *в среднем* составляет 61 сутки, а в каждом конкретном проекте возможны заметные отклонения длины критического пути от ее среднего значения (причем, чем больше суммарная дисперсия продолжительности работ критического пути, тем более вероятны значительные по абсолютной величине отклонения).

Поэтому предварительный анализ сетей со случайными продолжительностями работ, как правило, не ограничивается расчетами временных параметров сети. Весьма важным моментом анализа становится оценка вероятности того, что срок выполнения проекта $t_{кр}$ не превзойдет заданного директивного срока T.

Полагая $t_{кр}$ случайной величиной, имеющей нормальный закон распределения, получим

$$P(t_{кр} \le T) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{T - \bar{t}_{кр}}{\sigma_{кр}}\right), \qquad (14.26)$$

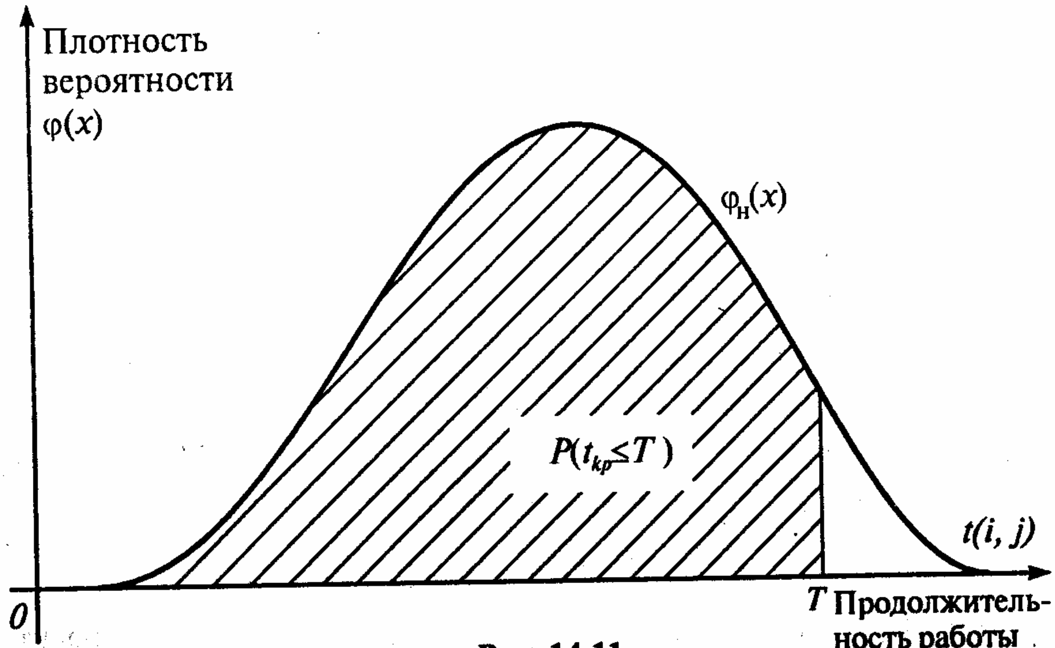


Рис. 14.11

(на рис. 14.11 это площадь заштрихованной фигуры), где $\Phi(z)$ — значение интеграла вероятностей Лапласа, где $z = (T - \bar{t}_{кр})/\sigma_{кр}$; $\sigma_{кр}$ — среднее квадратическое отклонение длины критического пути:

$$\sigma_{кр}=\sqrt{\sigma^2_{кр}}\,, \qquad (14.27)$$

а $\bar{t}_{кр}$, $\sigma^2_{кр}$ определяются по формулам (14.24) и (14.25).

Если $P(t_{кр} \le T)$ мала (например, меньше 0,3), то опасность срыва заданного срока выполнения комплекса велика, необходимо принятие дополнительных мер (перераспределение ресурсов по сети, пересмотр состава работ и событий и т.п. — об этом речь пойдет дальше). Если $P(t_{кр} \le T)$ значительна (например, более 0,8), то, очевидно, с достаточной степенью надежности можно прогнозировать выполнение проекта в установленный срок.

В некоторых случаях представляет интерес и решение обратной задачи: определение максимального срока выполнения проекта T, который возможен с заданной надежностью (вероятностью) β. В этом случае

$$T=\bar{t}_{кр}+z_\beta\sigma^2_{кр}, \qquad (14.28)$$

где z_β — нормированное отклонение случайной величины, определяемое с помощью функции Лапласа $\Phi(z_\beta)=\beta$.

▷ **14.3.** Пусть, например, для сети (рис.14.6) дисперсии продолжительности работ критического пути равны: $\sigma^2(0,3)=2{,}5$; $\sigma^2(3,5)=2{,}1$; $\sigma^2(5,6)=3{,}2$; $\sigma^2(6,9)=4{,}0$; $\sigma^2(9,10)=1{,}5$; $\sigma^2(10,11)=3{,}5$. Оценить вероятность выполнения проекта в срок $T = 63$ суткам.

Решение: Найдем $\sigma_{кр}$, используя формулы (14.25) и (14.27):

$$\sigma_{кр}=\sqrt{\sigma^2(0,3)+\sigma^2(3,5)+\sigma^2(5,6)+\sigma^2(6,9)+\sigma^2(9,10)+\sigma^2(10,11)}=$$

$$=\sqrt{2{,}5+2{,}1+3{,}2+4{,}0+1{,}5+3{,}5}=\sqrt{16{,}8}\approx 4{,}1.$$

Теперь искомая вероятность

$$P(t_{кр}\le 63)=\frac{1}{2}+\frac{1}{2}\Phi\left(\frac{63-61}{4{,}1}\right)=\frac{1}{2}+\frac{1}{2}\Phi\left(\frac{2}{4{,}1}\right)=$$

$$=\frac{1}{2}+\frac{1}{2}\Phi(0{,}49)=0{,}5+0{,}5\cdot 0{,}376=0{,}688\approx 0{,}69,$$

т.е. можно с известным риском предполагать выполнение проекта в срок.

Рассмотрим и пример решения обратной задачи: оценить максимально возможный срок T выполнения проекта с надежностью $\beta=0,95$.

По формуле (14.28)[1] $T=61+z_{0,95}\cdot 4,1=61+1,96\cdot 4,1\approx 69$, т.е. с надежностью 0,95 срок выполнения проекта не превысит 69 суток.▶

Следует отметить, что для данной сети мы можем найти лишь весьма приближенные оценки $P(t_{кр} \le T)$ и T, ибо на основании теоремы Ляпунова вывод о нормальном законе распределения случайной величины $t_{кр}$ правомерен лишь для достаточно большого числа критических работ, а в рассматриваемой сети их всего 6.

Однако приведенный метод расчета имеет принципиальные недостатки оценки параметров даже сложных сетей с большим количеством работ. Дело в том, что на практике нередки случаи, когда дисперсии $\sigma^2(L)$ длин некритических (но близких к критическому) путей существенно больше, чем $\sigma^2_{кр}$. Поэтому при изменении ряда условий в данном конкретном комплексе работ возможен переход к новым критическим путям, которые в расчете не учитываются.

Различия между событиями с детерминированными случайными продолжительностями работ не следует путать с различием детерминированных и стохастических сетей. Последнее различие связано со структурой самой сети.

Рассмотренные до сих пор сети являлись *детерминированными*, хотя работы в них могли характеризоваться не только детерминированными, но и случайными продолжительностями. Вместе с тем встречаются проекты, когда на некоторых этапах тот или иной комплекс последующих работ зависит от неизвестного заранее результата. Какой из этих комплексов работ будет фактически выполняться, заранее не известно, а может быть предсказано лишь с некоторой вероятностью. Например, может быть предусмотрено несколько вариантов продолжения исследования в зависимости от полученных опытных данных или несколько вариантов строительства предприятий различной мощности по обработке сырья в зависимости от результатов разведки запасов этого сырья. Такие сети называются *стохастическими*.

В свою очередь стохастические сети, так же как и детерминированные, могут характеризоваться детерминированными либо случайными продолжительностями.

---

[1] $z_{0,95}=1,96$ определяем по таблице значений функции Лапласа.

14.7. Коэффициент напряженности работы. Анализ и оптимизация сетевого графика

После нахождения критического пути и резервов времени работ и оценки вероятности выполнения проекта в заданный срок должен быть проведен всесторонний анализ сетевого графика и приняты меры по его оптимизации. Этот весьма важный этап в разработке сетевых графиков раскрывает основную идею СПУ. Он заключается в приведении сетевого графика в соответствие с заданными сроками и возможностями организации, разрабатывающей проект.

Вначале рассмотрим анализ и оптимизацию *календарных сетей*, в которых заданы только оценки продолжительности работ.

Анализ сетевого графика начинается с анализа топологии сети, включающего контроль построения сетевого графика, установление целесообразности выбора работ, степени их расчленения.

Затем проводятся классификация и группировка работ по величинам резервов. Следует отметить, что величина полного резерва времени далеко не всегда может достаточно точно характеризовать, насколько напряженным является выполнение той или иной работы некритического пути. Все зависит от того, на какую последовательность работ распространяется вычисленный резерв, какова продолжительность этой последовательности.

Определить степень трудности выполнения в срок каждой группы работ некритического пути можно с помощью коэффициента напряженности работ.

Коэффициентом напряженности $K_н$ работы (i, j) *называется отношение продолжительности несовпадающих (заключенных между одними и теми же событиями) отрезков пути, одним из которых является путь максимальной продолжительности, проходящий через данную работу, а другим — критический путь:*

$$K_{н}(i,j)=\frac{t(L_{\max})-t'_{кр}}{t_{кр}-t'_{кр}}, \qquad (14.29)$$

где $t(L_{\max})$ — продолжительность максимального пути, проходящего через работу (i, j);

$t_{кр}$ — продолжительность (длина) критического пути;

$t'_{кр}$ — продолжительность отрезка рассматриваемого пути, совпадающего с критическим путем.

Формулу (14.29) можно легко привести к виду

$$K_{\text{н}}(i, j)=1-\frac{R_{\text{п}}(i, j)}{t_{\text{кр}}-t'_{\text{кр}}}, \tag{14.30}$$

где $R_{\text{п}}(i, j)$ — полный резерв времени работы (i, j).

Коэффициент напряженности $K_{\text{н}}(i, j)$ может изменяться в пределах от 0 (для работ, у которых отрезки максимального из путей, не совпадающие с критическим путем, состоят из фиктивных работ нулевой продолжительности) до 1 (для работ критического пути).

▷ **14.4.** Найти коэффициент напряженности работы $(1, 4)$ для сетевого графика (рис.14.6).

Р е ш е н и е. В разд. 14.5 мы установили, что длина критического пути $t_{\text{кр}}=61$ (сутки), а максимальный путь, проходящий через работу $(1, 4)$ — путь L_4 $0\to 1\to 4\to 6\to 9\to 10\to 11$ — имеет продолжительность $t(L_{\max})=t(L_4)=49$ (суток). Максимальный путь L_4 совпадает с критическим (*см.* рис.14.6) на отрезке $6\to 9\to 10\to 11$ продолжительностью $t_{\text{кр}}=13+6+13=32$ (сутки). Используя формулу (14.29), найдем

$$K_{\text{н}}(1, 4)=\frac{49-32}{61-32}=\frac{17}{29}\approx 0{,}59.$$

Или иначе: зная полный резерв работы $R_{\text{п}}(1, 4)=12$ (*см.* рис. 14.3), по формуле (14.30) находим

$$K_{\text{н}}(1,4)=1-\frac{12}{61-32}=\frac{17}{29}\approx 0{,}59. \blacktriangleright$$

Чем ближе к 1 коэффициент напряженности $K_{\text{н}}(i, j)$, тем сложнее выполнить данную работу в установленные сроки. Чем ближе $K_{\text{н}}(i, j)$ к нулю, тем бо́льшим относительным резервом обладает максимальный путь, проходящий через данную работу.

Работы могут обладать одинаковыми полными резервами, но степень напряженности сроков их выполнения, выражаемая коэффициентом напряженности $K_{\text{н}}(i, j)$, может быть различна. И наоборот, различным полным резервам могут соответствовать одинаковые коэффициенты напряженности.

Так, полные резервы работ (*3*, *6*) и (*6*, *7*) для сетевого графика равны: $R_п(3, 6)=R_п(6, 7)=10$ (суток) — *см.* табл. 14.3, а их коэффициенты напряженности различны:

$$K_н(3, 6)=\frac{6}{16}\approx 0{,}38,\ K_н(6, 7)=\frac{9}{19}\approx 0{,}47.$$

Обратим внимание на то, что бóльший полный резерв одной работы (по сравнению с другой) не обязательно свидетельствует о меньшей степени напряженности ее выполнения. Так, в рассматриваемой сети (*см.* рис. 14.6), хотя работа (*2*, *7*) обладает бóльшим полным резервом по сравнению с работой (*6*, *10*): $R_п(2, 7)= 23>R_п(6, 10) = 14$, но имеет вдвое бóльший коэффициент напряженности: $K_н(2,\ 7)=\frac{25}{48}\approx 0{,}52$ против $K_н(6,\ 10)=\frac{5}{19}\approx 0{,}26$. Это объясняется разным удельным весом полных резервов работ в продолжительности отрезков максимальных путей, не совпадающих с критическим путем.

Вычисленные коэффициенты напряженности позволяют дополнительно классифицировать работы по зонам. В зависимости от величины $K_н(i, j)$ выделяют три зоны: *критическую* ($K_н(i, j)>0{,}8$); *подкритическую* ($0{,}6\le K_н(i, j)\le 0{,}8$); *резервную* ($K_н(i, j)<0{,}6$).

Оптимизация *сетевого графика* представляет процесс улучшения организации выполнения комплекса работ с учетом срока его выполнения. Оптимизация проводится с целью сокращения длины критического пути, выравнивания коэффициентов напряженности работ, рационального использования ресурсов.

В первую очередь принимаются меры по сокращению продолжительности работ, находящихся на критическом пути. Это достигается:

- перераспределением всех видов ресурсов, как временных (использование резервов времени некритических путей), так и трудовых, материальных, энергетических (например, перевод части исполнителей, оборудования с некритических путей на работы критического пути); при этом перераспределение ресурсов должно идти, как правило, из зон, менее напряженных, в зоны, объединяющие наиболее напряженные работы;
- сокращением трудоемкости критических работ за счет передачи части работ на другие пути, имеющие резервы времени;
- параллельным выполнением работ критического пути;

• пересмотром топологии сети, изменением состава работ и структуры сети.

В процессе сокращения продолжительности работ критический путь может измениться, и в дальнейшем процесс оптимизации будет направлен на сокращение продолжительности работ нового критического пути и так будет продолжаться до получения удовлетворительного результата. В идеале длина любого из полных путей может стать равной длине критического пути или по крайней мере пути критической зоны. Тогда все работы будут вестись с равным напряжением, а срок завершения проекта существенно сократится.

Весьма эффективным является использование *метода статистического моделирования*, основанного на многократных последовательных изменениях продолжительности работ (в заданных пределах) и "проигрывании" на компьютере различных вариантов сетевого графика с расчетами всех его временных параметров и коэффициентов напряженности работ. Процесс "проигрывания" продолжается до тех пор, пока не будет получен приемлемый вариант плана или пока не будет установлено, что все имеющиеся возможности улучшения плана исчерпаны и поставленные перед разработчиком проекта условия невыполнимы.

До сих пор мы говорили лишь о соблюдении директивных сроков выполнения комплекса работ и не затрагивали непосредственно вопросов стоимости разработки проектов. Однако на практике при попытках эффективного улучшения составленного плана неизбежно введение дополнительно к оценкам сроков фактора стоимости работ.

14.8. Оптимизация сетевого графика методом "время — стоимость"

Оптимизация сетевого графика в зависимости от полноты решаемых задач может быть условно разделена на частную и комплексную. Видами *частной оптимизации* сетевого графика являются: минимизация времени выполнения комплекса работ при заданной его стоимости; минимизация стоимости комплекса работ при заданном времени выполнения проекта.

Комплексная оптимизация представляет собой нахождение оптимального соотношения величин стоимости и сроков выполнения проекта в зависимости от конкретных целей, ставящихся при его реализации.

При использовании метода "время — стоимость" предполагают, что уменьшение продолжительности работы пропорционально возрастанию ее стоимости. Каждая работа (i, j) характеризуется продолжительностью $t(i, j)$, которая может находиться в пределах

$$a(i, j) \le t(i, j) \le b(i, j), \tag{14.31}$$

где $a(i, j)$ — минимально возможная (экстренная) продолжительность работы (i, j), которую только можно осуществить в условиях разработки;

$b(i, j)$ — нормальная продолжительность выполнения работы (i, j).

При этом стоимость $c(i, j)$ работы (i, j) заключена в границах от $c_{\min}(i, j)$ (при нормальной продолжительности работы) до $c_{\max}(i, j)$ (при экстренной продолжительности работы).

Используя аппроксимацию по прямой (*см.* рис. 14.12), можно легко найти изменение стоимости работы $\Delta c(i, j)$ при сокращении ее продолжительности на величину

$$\Delta c(i, j)=[b(i, j)-t(i, j)]h(i, j). \tag{14.32}$$

Величина $h(i, j)$, равная тангенсу угла α наклона аппроксимирующей прямой (*см.* рис. 14.12), показывает *затраты на ускорение*

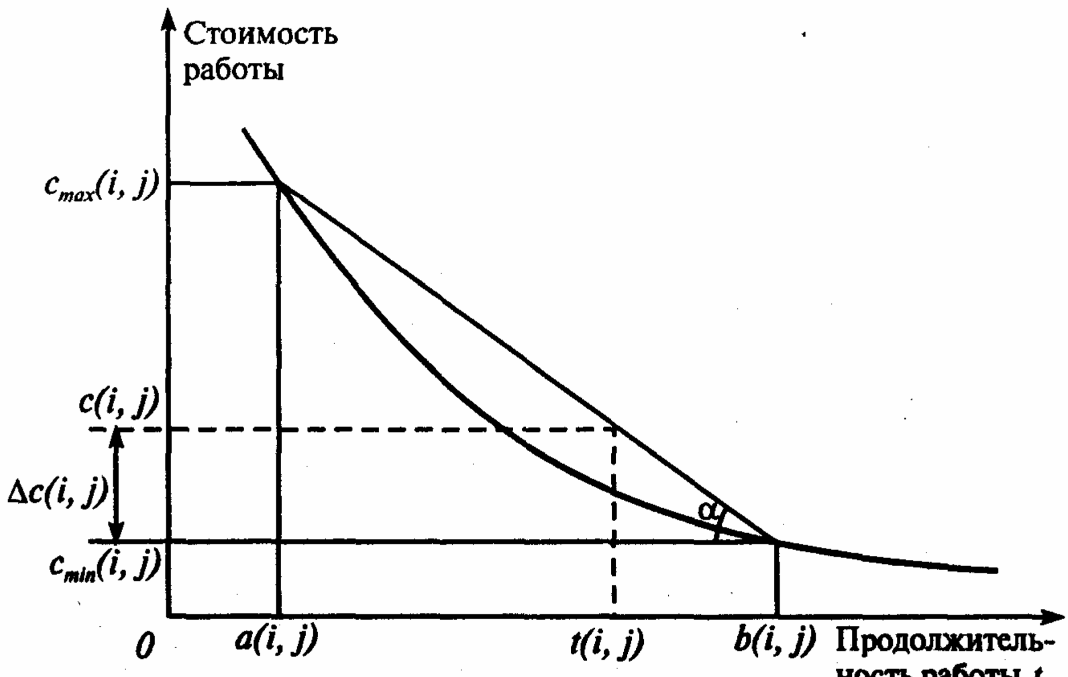


Рис. 14.12

работы (i, j) (по сравнению с нормальной продолжительностью) на единицу времени:

$$h(i,j) = \operatorname{tg} \alpha = \frac{c_{\max}(i,j) - c_{\min}(i,j)}{b(i,j) - a(i,j)}. \qquad \textbf{(14.33)}$$

Самый очевидный вариант частной оптимизации сетевого графика с учетом стоимости предполагает использование резервов времени работ. Продолжительность каждой работы, имеющей резерв времени, увеличивают до тех пор, пока не будет исчерпан этот резерв или пока не будет достигнуто верхнее значение продолжительности $b\,(i, j)$. При этом стоимость выполнения проекта, равная до оптимизации

$$C = \sum_{i,j} c(i,j)\,, \qquad \textbf{(14.34)}$$

уменьшится на величину

$$C = \sum_{i,j} \Delta c(i,j) = \sum_{i,j} \left[b(i,j) - t(i,j) \right] h(i,j). \qquad \textbf{(14.35)}$$

Для проведения частной оптимизации сетевого графика кроме продолжительности работ $t\,(i, j)$, необходимо знать их граничные значения $a\,(i, j)$ и $b\,(i, j)$, а также показатели затрат на ускорение работ $h\,(i, j)$, вычисляемые по формуле (14.33). Продолжительность каждой работы $t\,(i, j)$ целесообразно увеличить на величину такого резерва, чтобы не изменить ранние (ожидаемые) сроки наступления всех событий сети, т.е. на величину свободного резерва времени $R_c\,(i, j)$.

▷ **14.15.** Провести частную оптимизацию сетевого графика (рис. 14.6). Граничные значения продолжительностей работ $a\,(i, j)$ и $b\,(i, j)$, их стоимости $c\,(i, j)$, коэффициенты затрат на ускорение работ $h\,(i, j)$ приведены в табл. 14.4.

Р е ш е н и е. Свободные резервы времени работ $R_c\,(i, j)$ были вычислены нами ранее (*см.* табл. 14.3). Их ненулевые значения даны в табл. 14.4. Там же представлены результаты частной оптимизации рассматриваемой сети.

Стоимость первоначального варианта сетевого графика или плана по формуле (14.34) равна сумме стоимостей всех работ (включая работы, не имеющие резервов и не включенные в табл. 14.4):

$$C=694+50+45+...+35+10=1216 \text{ (усл.руб.)}.$$

Стоимость нового плана равна $C-\Delta C$=1216−293=923 (усл.руб.), т.е. уменьшилась почти на 25%. Новый оптизированный сетевой график представлен на рис. 14.13. Нетрудно убедиться в том, что появились новые критические пути длиной $t_{кр}$=61 (сутки), например: $0\to1\to3\to4\to7\to10\to11$; $0\to3\to5\to8\to9\to11$; $0\to1\to3\to4\to\to6\to7\to10\to11$; $0\to3\to5\to6\to8\to9\to11$ и т.д.

Таблица 14.4

№ п/п	Работа	Продолжительность работы, в сутки			Свободный резерв времени работы, в сутки	Стоимость работы	Коэффициент затрат на ускорение работы, руб./сутки	Уменьшение стоимости проекта, усл. руб.
	(i, j)	$a(i,j)$	$t(i,j)$	$b(i,j)$	$R_c(i,j)$	$c(i,j)$	$h(i,j)$	$\Delta C(i,j)$
1	(0,5)	5	9	14	11	60	8	5·8=40
2	(1,4)	4	6	10	9	28	4	4·4=16
3	<u>(1,3)</u>	3	4	6	1	37	12	1·12=12
4	(2,7)	2	3	7	13	86	6	4·6=24
5	(3,6)	4	6	9	10	92	10	3·10=30
6	<u>(4,7)</u>	3	8	14	2	48	5	2·5=10
7	<u>(4,6)</u>	1	3	6	3	64	12	3·12=36
8	<u>(5,8)</u>	5	10	18	7	15	1	7·1=7
9	(5,9)	3	6	12	16	86	7	6·7=42
10	(6,10)	2	5	10	14	44	5	5·5=25
11	<u>(7,10)</u>	1	5	15	10	74	4	10·4=40
12	<u>(8,9)</u>	2	4	8	1	20	3	1·3=3
13	<u>(9,11)</u>	11	17	23	2	40	4	2·4=8
	Итого					694	–	293

Примечания: 1. В таблице представлены параметры лишь тех работ, которые имеют свободный резерв времени.

2. Стоимости $c(i, j)$ остальных работ: $c(0,1)$=50; $c(0,3)$=45; $c(1,2)$=82; $c(3,4)$=55; $c(3,5)$=72; $c(5,6)$=30; $c(6,7)$=26; $c(6,9)$=75; $c(6,8)$=42; $c(9,10)$=35; $c(10,11)$=10 (усл.руб.).

3. Подчеркнуты те работы, свободные резервы времени которых полностью использованы на увеличение их продолжительности.

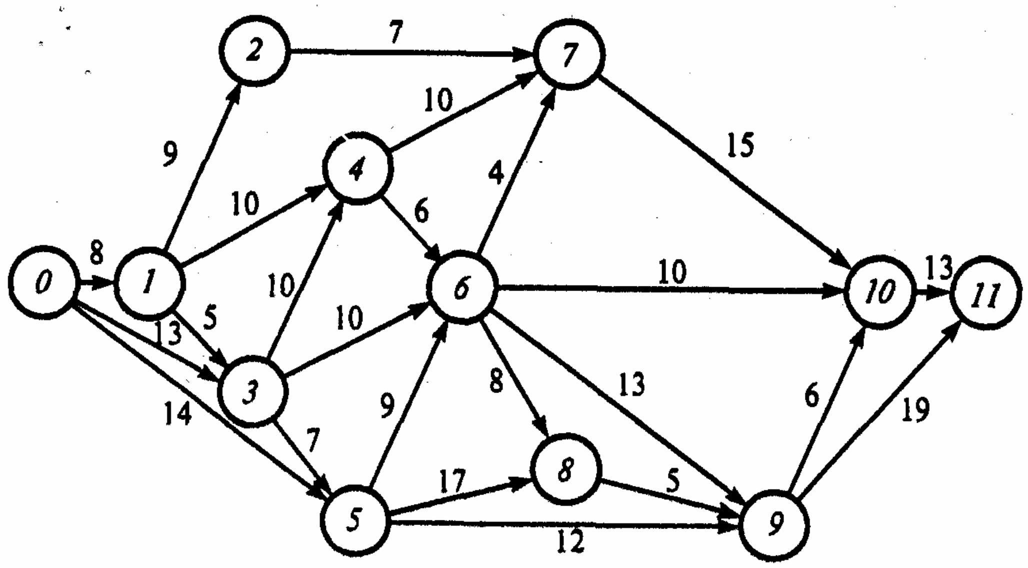


Рис. 14.13

Можно показать, что в этом варианте сетевого графика из 64 полных путей 28 — критические. Если бы верхние границы продолжительностей работ дали возможность полностью использовать резерв времени всех работ, представленных в табл. 14.4, то в новом плане все полные пути были бы критические.

Итак, в результате оптимизации сети мы пришли к плану, позволяющему выполнить комплекс работ в срок $t_{кр}$=61 (сутки) при минимальной его стоимости C = 923 (усл.руб.).

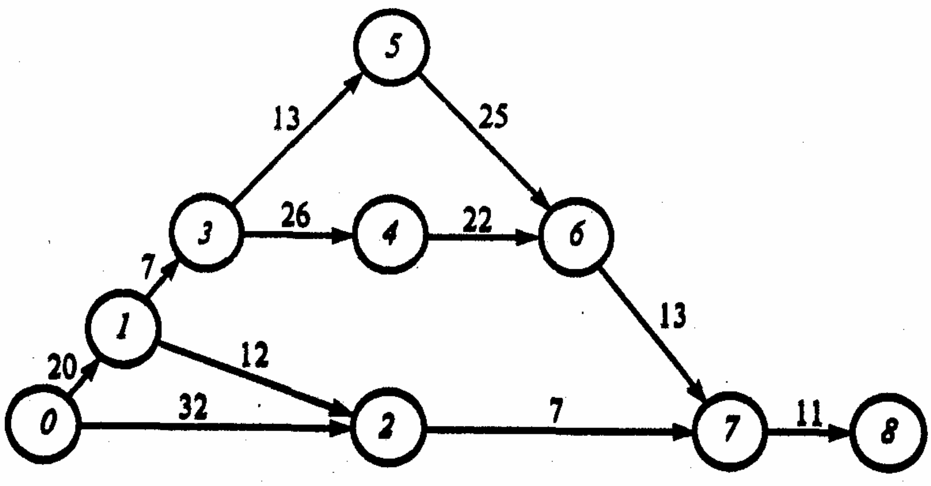


Рис. 14.14

В реальных условиях выполнения проекта может потребоваться ускорение его выполнения, что, естественно, отразится на стоимости проекта: она увеличится. Поэтому необходимо определить оптимальное соотношение между стоимостью проекта C и продолжительностью его выполнения $t = t_{кр}$, представленное, например, в виде функции $C = C(t)$.

Для оптимизации сетей и, в частности, для нахождения функции $C(t)$ могут быть использованы *эвристические методы*, т.е. методы, учитывающие индивидуальные особенности сетевых графиков.

▷ **14.6.** Оптимизировать сетевой график, изображенный на рис. 14.14, в котором указаны максимально возможные продолжительности работ (в сутках). Необходимые для оптимизации исходные данные представлены в табл. 14.5.

Таблица 14.5

№ п/п	Работа (i, j)	Продолжительность работы, сутки		Коэффициент затрат на ускорение работы $h(i, j)$	Стоимость работы, усл. руб. $C = (i, j)$ при $t(i, j) = b(i, j)$
		минимальная $a(i, j)$	максимальная $b(i, j)$		
1	*(0,1)*	10	20	6	35
2	*(0,2)*	12	32	3	50
3	*(1,2)*	2	12	3	15
4	*(1,3)*	2	7	8	10
5	*(2,7)*	2	7	3	10
6	*(3,4)*	16	26	2	50
7	*(3,5)*	8	13	6	15
8	*(4,6)*	12	22	4	40
9	*(5,6)*	20	25	4	30
10	*(6,7)*	8	13	5	25
11	*(7,8)*	6	11	9	20
	Итого:				300

Р е ш е н и е. Исходный для оптимизации план (*см.* рис.14.14) имеет максимальную продолжительность работ $t(i, j)=b(i, j)$ и соответственно минимальную стоимость C=300 (усл.руб.). Найдем все полные пути сетевого графика.

Их четыре:

L_1 $0 \to 1 \to 3 \to 5 \to 6 \to 7 \to 8$ продолжительностью $t(L_1)=89$ (суток);

L_2 $0 \to 1 \to 3 \to 4 \to 6 \to 7 \to 8$ продолжительностью $t_{кр}=t(L_2)=99$ (суток);

L_3 $0 \to 1 \to 2 \to 7 \to 8$ продолжительностью $t(L_3)=50$(суток);

L_4 $0 \to 2 \to 7 \to 8$ продолжительностью $t(L_4)=50$ (суток).

Для удобства дальнейших расчетов представим эти пути графически в виде цепочек работ (рис. 14.15), в которых цифры над стрелками показывают коэффициенты затрат на ускорение работ $h(i, j)$, а под стрелками — максимально возможные величины уменьшения продолжительности работ $\Delta t(i, j) = b(i, j) - a(i, j)$.

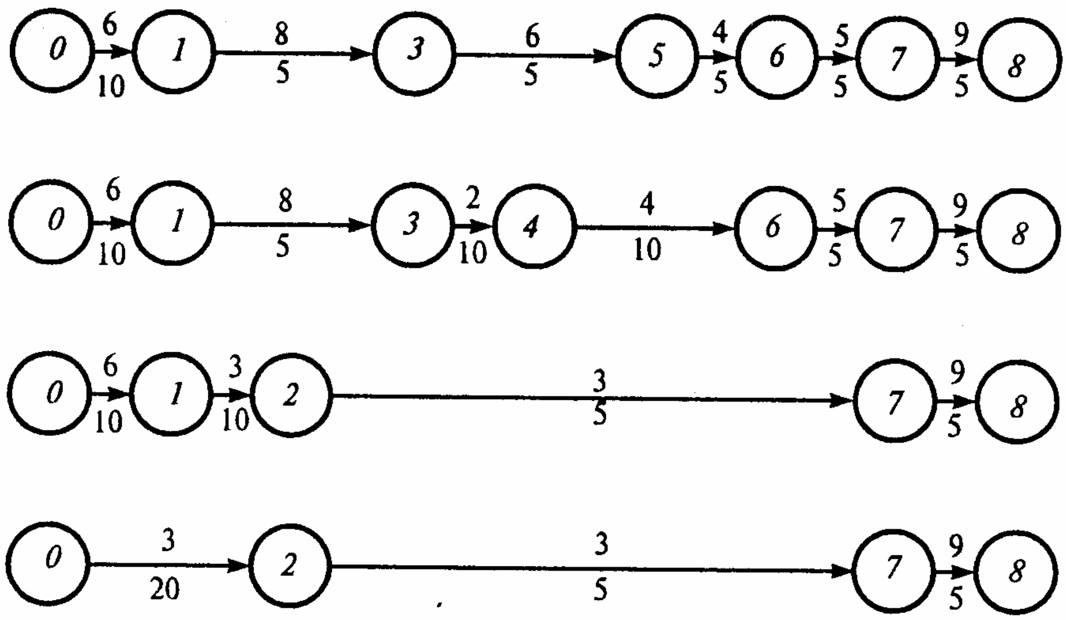


Рис. 14.15

I ш а г. Уменьшить продолжительность выполнения комплекса можно, как известно, только за счет сокращения продолжительности работ критического пути $t_{кр}^{\prime}= t(L_2)$. Из работ критического пути L_2 наименьший коэффициент затрат на ускорение $h(i, j)$ имеет работа $(3,4)$: $h_{\min}(i, j)=\min\{h(0, 1); h(1, 3); h(3, 4); h(4, 6); h(6, 7); h(7, 8)\}=\min\{6; 8; 2; 4; 5; 9\}=2$, т.е. $h_{\min}(i, j)=h(3, 4)=2$. Продолжительность работы $t(3,4)$ можно сокращать не более чем на 10 суток. При этом изменится длина только критического пути (с 99 до 89 суток) L_2 — единственного из четырех путей, проходящего через работу $(3,4)$. А стоимость проекта за счет ускорения работы $(3,4)$ с учетом формул (14.34) и (14.35) возрастет до $300+2\cdot10=320$ (усл.руб.). Итак, на I шаге:

$$C = 300 + 2\cdot(99-t), \text{ где } 89 \le t \le 99;$$

новые длины путей равны $t(L_1)=t(L_2)=89$; $t(L_3)=t(L_4)=50$.

II шаг. Теперь мы имеем два критических пути L_1 и L_2 и сократить срок выполнения проекта можно за счет одновременного сокращения их продолжительности. Сократить одновременно $t(L_1)$ и $t(L_2)$ можно, уменьшив продолжительность работ, лежащих на этих путях (*см.* рис. 14.15): либо $t(0, 1)$, либо $t(6, 7)$, либо $t(7, 8)$. Останавливаемся на $t(6, 7)$, поскольку при этом обеспечивается минимум затрат на ускорение работы: $h_{\min}(i, j)=\min\{h(0, 1)$; $h(1, 3)$; $h(6, 7)$; $h(7, 8)\}=\min\{6; 8; 5; 9\}=5$, т.е. $h_{\min}(i, j)=h(6, 7)=5$.

Продолжительность работы $t(6, 7)$, можно уменьшить не более чем на 5 суток. На эту величину уменьшатся длины критических путей $t(L_1)$ и $t(L_2)$, а следовательно, и срок выполнения проекта $t=t(L_1)=t(L_2)$. При этом стоимость проекта увеличится с 320 до $320+5\cdot5=345$ (усл.руб.). Итак, на II шаге:

$$C=320+5\cdot(89-t), \text{ где } 84 \le t \le 89;$$

$$t(L_1)=t(L_2)=84, \; t(L_3)=t(L_4)=50.$$

Продолжая аналогичным образом сокращать продолжительность работ, получим

III шаг. $h_{\min}(i, j) = \min\{h(0, 1); h(1, 3); h(7, 8)\} = \min\{6; 8; 9\} = 6$, т.е. $h_{\min}(i, j)=h(0, 1)=6$. Сокращая продолжительность работы $t(0, 1)$ до 10 суток, найдем

$$C = 345+6\cdot(84-t), \text{ где } 74 \le t \le 84;$$

$$t(L_1)=t(L_2)=74, \; t(L_3)=40; \; t(L_4)=50.$$

IV шаг. $h_{\min}(i, j) = \min\{h(1, 3); h(7, 8)\} = \min\{8; 9\} = 8$, т.е. $h_{\min}(i, j) = h(1, 3) = 8$. Сокращая продолжительность работы $t(1, 3)$ до 5 суток, найдем

$$C = 405+8\cdot(74-t), \text{ где } 69 \le t \le 74;$$

$$t(L_1)=t(L_2)=69, \; t(L_3)=40; \; t(L_4)=50.$$

V. ш а г. Сокращая продолжительность работы $t(7, 8)$ до 5 суток, найдем (учитывая, что $h(7, 8)=9$)

$$C = 445+9\cdot(69-t), \text{ где } 64 \le t \le 69;$$

$$t(L_1)=t(L_2)=64;\ t(L_3)=35;\ t(L_4)=45.$$

VI ш а г. Теперь несокращенными остались продолжительности трех критических работ: $t(3, 5)$ и $t(5, 6)$ критического пути L_1, каждую из которых можно сократить до 5 суток, и $t(4, 6)$ критического пути L_2, которую можно сократить до 10 суток. Сокращение какой-либо одной из названных величин не приведет к сокращению продолжительности выполнения проекта, ибо при этом сократится лишь один из двух путей, а длина несокращенного пути, который станет единственным критическим путем, не изменится. Поэтому последовательно сокращая $t(4, 6)$ и $t(5, 6)$ до 5 суток (с учетом времени сокращения продолжительности работ), найдем (теперь коэффициент затрат на ускорение работ равен $h(4, 6)+h(5, 6) = 4+4 = 8$):

$$C = 490+8\cdot(64-t), \text{ где } 59 \le t \le 64;$$

$$t(L_1)=t(L_2)=59;\ t(L_3)=35;\ t(L_4)=45.$$

VII ш а г. Продолжительность работы $t(4, 6)$ можно сократить еще до 5 суток и на тот же срок можно сократить $t(3, 5)$ (иначе срок выполнения проекта не изменится). Полагая, что $h(4, 6)+h(3, 5)=4+6=10$, найдем

$$C = 530+10\cdot(59-t), \text{ где } 54 \le t \le 59.$$

График оптимальной зависимости стоимости проекта $C(t)$ от продолжительности его выполнения показан на рис. 14.16. С помощью этого графика можно, с одной стороны, оценить минимальную стоимость проекта при любом возможном сроке его выполнения, а с другой стороны — найти предельную продолжительность выполнения проекта при заданной его стоимости. Например, при продолжительности проекта $t=79$ (суток) минимальная стоимость выполнения рассматриваемого комплекса составит 375 (усл. руб.), а при стоимости выполнения комплекса, например, 540 (усл. руб.) предельная продолжительность проекта составит 55 (суток). С помощью функции $C(t)$ можно оценить дополнительные затраты, связанные с сокращением сроков завершения комплекса. Так, со-

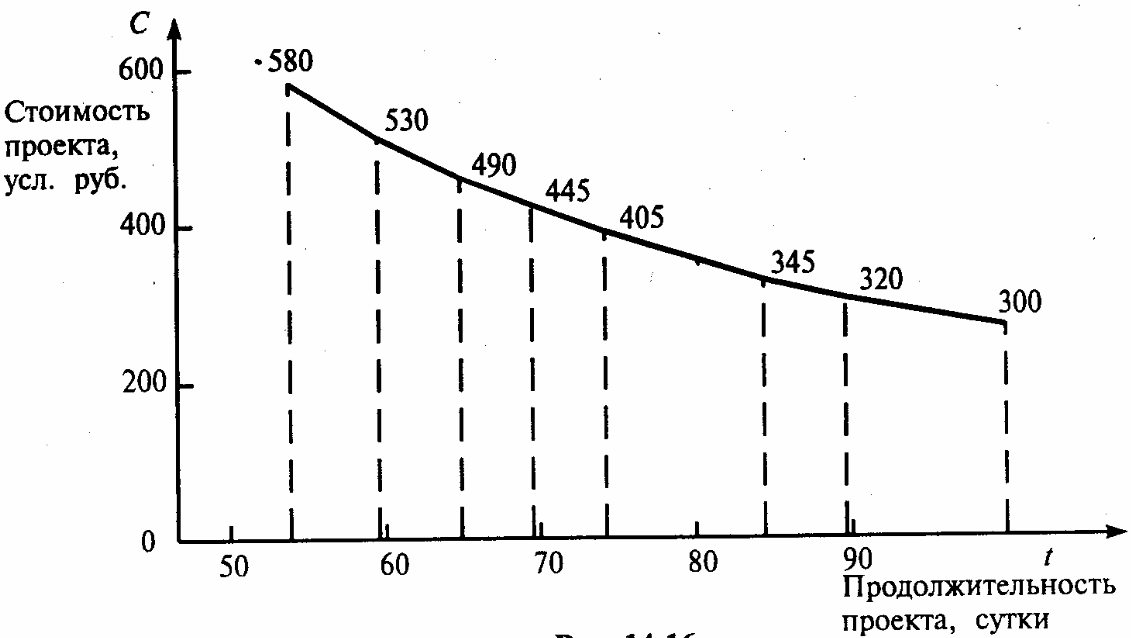


Рис. 14.16

кращение продолжительности проекта с 79 до 55 суток потребует дополнительных затрат 540–375=165 (усл. руб.).▶

Итак, мы рассмотрели один из возможных эвристических алгоритмов оптимизации сетевого графика (*см.* рис. 14.14). Можно было использовать и другие алгоритмы. Например, взять в качестве первоначального план, имеющий не максимальные, а минимальные значения продолжительности работ $t(i, j) = a(i, j)$ и соответственно максимальную стоимость проекта. А затем последовательно увеличивать продолжительность выполнения комплекса работ путем увеличения продолжительности работ, расположенных на некритических, а затем и на критическом(ских) пути в порядке убывания коэффициентов затрат $h(i, j)$.

Следует заметить, что при линейной зависимости стоимости работ от их продолжительности задача построения оптимального сетевого графика может быть сформулирована как задача линейного программирования, в которой необходимо минимизировать стоимость выполнения проекта при двух группах ограничений. Первая группа ограничений показывает, что продолжительность каждой работы должна находиться в пределах, установленных неравенством (14.31). Вторая группа ограничений требует, чтобы продолжительность любого полного пути сетевого графика не превышала установленного директивного срока выполнения проекта. Однако решать такие задачи классическими методами линейного программирования, как правило, неэффективно, в связи с чем используются специально разработанные методы.

УПРАЖНЕНИЯ

В задачах **14.7**, **14.8** построить сетевой график. Найти продолжительность выполнения комплекса работ, временные характеристики событий и работ. В скобках указана продолжительность работ.

14.7. Сделать деревянный ящик (работу выполняет один человек). Разместить доски в соответствии с размерами ящика (15 мин); разрезать доски (12 мин); склеить части ящика (40 мин); прибить к крышке ящика петли (8 мин); подождать, пока ящик высохнет, и вытереть его (15 мин); петли (с крышкой) прибить к ящику (10 мин).

14.8. Заменить колесо машины (работу выполняют два человека). Достать из багажника домкрат и инструменты (40 с); снять диск с колеса (30 с); освободить колесо (50 с); поставить домкрат под машину (26 с); поднять машину (20 с); из багажника взять запасное колесо (25 с); снять гайки и колесо (20 с); установить запасное колесо на ось (10 с); завинтить (не сильно) гайки на оси (15 с); опустить машину и собрать домкрат (25 с); поставить домкрат обратно в багажник (10 с); завинтить гайки на оси до конца (12 с); положить плохое колесо и инструменты в багажник (40 с); поставить на место диск колеса (10 с).

14.9. Для сетевого графика (рис. 14.17) найти все полные пути, критический путь; рассчитать ранние и поздние сроки свершения событий, начала и окончания работ; определить резервы времени полных путей и событий, резервы времени (полные, частные резервы первого вида, свободные и независимые) работ и коэффициенты напряженности работ.

14.10. Как изменится срок выполнения проекта (*см.* рис. 14.17), резервы времени работ и событий, коэффициенты напряженности работ, если увеличить продолжительность работы $t(9, 10)$ на величину: а) $R_п(9, 10)$; б) $R_1(9, 10)$; в) $R_с(9, 10)$; г) $R_н(9, 10)$?

14.11. Как изменится срок выполнения проекта (*см.* рис. 14.17), резервы времени работ и событий, коэффициенты напряженности работ, если продолжительность каждой работы $t(i, j)$ увеличить на величину: а) $R_п(i, j)$; б) $R_1(i, j)$; в) $R_с(i, j)$; г) $R_н(i, j)$? Для случаев б), в) и г) найти все критические пути сетевого графика.

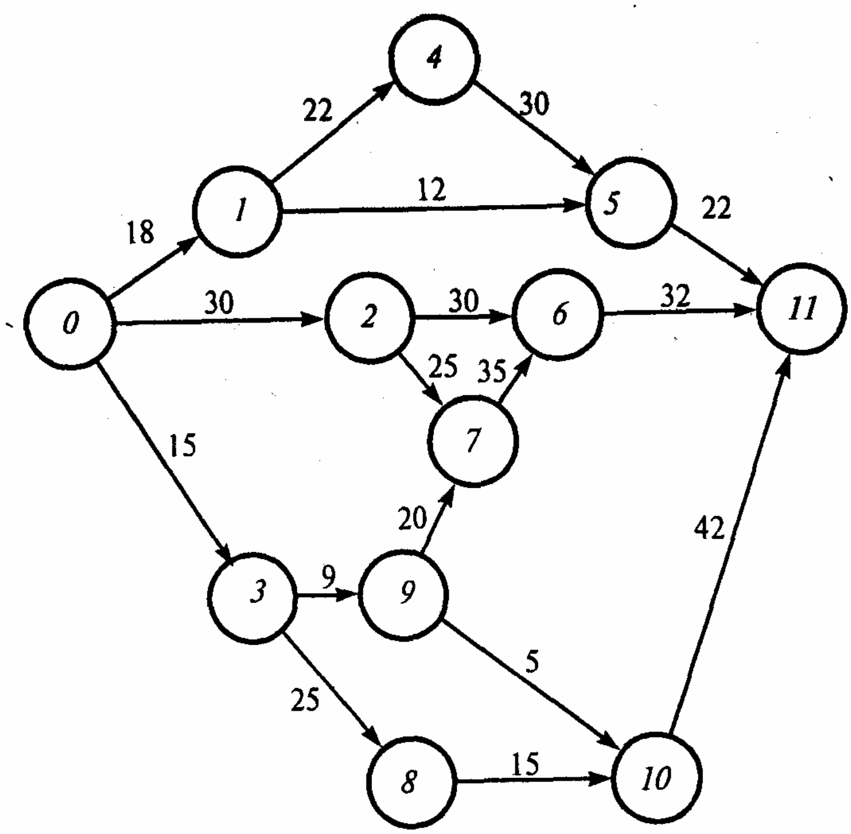


Рис. 14.17

14.12. В табл. 14.6 указаны оценки времени выполнения работ сетевого графика, данные ответственными исполнителями и экспертами.

Таблица 14.6

№ п/п	Работа (i, j)	Оценки времени выполнения работы, сутки		
		оптимистическая $t_о(i, j)$	пессимистическая $t_п(i, j)$	наиболее вероятная $t_{нв}(i, j)$
1	*(1,2)*	5	9	6
2	*(1,3)*	2	7	5
3	*(1,4)*	4	10	8
4	*(3,4)*	9	14	11
5	*(2,5)*	7	13	10
6	*(4,5)*	1	4	3

Необходимо: а) построить сетевой график; б) определить средние (ожидаемые) значения продолжительности работ; в) определить критический путь и его длину. Полагая, что продолжительность критического пути распределена по нормальному закону, найти: а) вероятность того, что срок выполнения комплекса работ не превысит 17 суток; б) максимальное значение продолжительности выполнения проекта, которое можно гарантировать с надежностью 0,95.

14.13. По данным табл. 14.7 необходимо: 1) построить сетевой график; 2) определить критический путь и стоимость проекта при минимально возможных значениях продолжительности всех работ; 3) найти минимальную стоимость проекта при том же сроке его завершения; 4) рассчитать и построить оптимальную зависимость стоимости проекта от продолжительности его выполнения, используя в качестве первоначального варианта сетевого графика: а) план с максимальными значениями продолжительности всех работ и соответственно минимальной стоимостью проекта; б) план, полученный в результате выполнения п.3.

Таблица 14.7

Работа	Нормальный план выполнения работы, сутки		Срочный план выполнения работы, сутки		Коэффициент затрат на ускорение работы
	min	max	min	max	
(1,2)	4	5	2	15	5
(1,3)	4	3	2	11	4
(1,4)	12	150	9	180	10
(2,3)	6	11	5	30	19
(2,4)	7	18	6	30	12
(3,4)	10	10	8	20	5
(3,5)	24	147	19	212	13
(4,5)	10	4	7	25	7
(5,6)	3	2	2	5	3

Глава 15. ЭЛЕМЕНТЫ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

15.1. Основные понятия. Классификация СМО

При исследовании операций часто приходится сталкиваться с системами, предназначенными для многоразового использования при решении однотипных задач. Возникающие при этом процессы получили название *процессов обслуживания*, а системы — *систем массового обслуживания* (*СМО*). Примерами таких систем являются телефонные системы, ремонтные мастерские, вычислительные комплексы, билетные кассы, магазины, парикмахерские и т.п.

Каждая СМО состоит из определенного числа обслуживающих единиц (приборов, устройств, пунктов, станций), которые будем называть *каналами* обслуживания. Каналами могут быть линии связи, рабочие точки, вычислительные машины, продавцы и др. По числу каналов СМО подразделяют на *одноканальные* и *многоканальные*.

Заявки поступают в СМО обычно не регулярно, а случайно, образуя так называемый *случайный поток заявок (требований)*. Обслуживание заявок, вообще говоря, также продолжается какое-то случайное время. Случайный характер потока заявок и времени обслуживания приводит к тому, что СМО оказывается загруженной неравномерно: в какие-то периоды времени скапливается очень большое количество заявок (они либо становятся в очередь, либо покидают СМО необслуженными), в другие же периоды СМО работает с недогрузкой или простаивает.

Предметом теории массового обслуживания является построение математических моделей, связывающих заданные условия работы СМО (число каналов, их производительность, характер потока заявок и т.п.) с показателями эффективности СМО, описывающими ее способность справляться с потоком заявок.

В качестве *показателей эффективности* СМО используются: среднее[1] число заявок, обслуживаемых в единицу времени; среднее число заявок в очереди; среднее время ожидания обслуживания; вероятность отказа в обслуживании без ожидания; вероятность того, что число заявок в очереди превысит определенное значение и т.п.

СМО делят на два основных типа (класса): СМО *с отказами* и СМО *с ожиданием* (*очередью*). В СМО с отказами заявка, поступившая в момент, когда все каналы заняты, получает отказ, покидает СМО и в дальнейшем процессе обслуживания не участвует (например, заявка на телефонный разговор в момент, когда все каналы заняты, получает отказ и покидает СМО необслуженной). В СМО с ожиданием заявка, пришедшая в момент, когда все каналы заняты, не уходит, а становится в очередь на обслуживание.

СМО с ожиданием подразделяются на разные виды в зависимости от того, как организована очередь: с ограниченной или неограниченной длиной очереди, с ограниченным временем ожидания и т.п.

Для классификации СМО важное значение имеет *дисциплина обслуживания*, определяющая порядок выбора заявок из числа поступивших и порядок распределения их между свободными каналами. По этому признаку обслуживание заявки может быть организовано по принципу *"первая пришла — первая обслужена"*, *"последняя пришла — первая обслужена"* (такой порядок может применяться, например, при извлечении для обслуживания изделий со склада, ибо последние из них оказываются часто более доступными) или *обслуживание с приоритетом* (когда в первую очередь обслуживаются наиболее важные заявки). Приоритет может быть как *абсолютным,* когда более важная заявка"вытесняет" из-под обслуживания обычную заявку (например, в случае аварийной ситуации плановые работы ремонтных бригад прерываются до ликвидации аварии), так и *относительным*, когда более важная заявка получает лишь "лучшее" место в очереди.

---

[1] Здесь и в дальнейшем средние величины понимаются как математические ожидания соответствующих случайных величин.

15.2. Понятие марковского случайного процесса

Процесс работы СМО представляет собой *случайный процесс.*

Под *случайным (вероятностным или стохастическим) процессом* понимается процесс изменения во времени состояния какой-либо системы в соответствии с вероятностными закономерностями.

Процесс называется *процессом с дискретными состояниями*, если его возможные состояния S_1, S_2, S_3... можно заранее перечислить, а переход системы из состояния в состояние происходит мгновенно (скачком). Процесс называется *процессом с непрерывным временем*, если моменты возможных переходов системы из состояния в состояние не фиксированы заранее, а случайны.

Процесс работы СМО представляет собой случайный процесс с дискретными состояниями и непрерывным временем. Это означает, что состояние СМО меняется скачком в случайные моменты появления каких-то событий (например, прихода новой заявки, окончания обслуживания и т.п.).

Математический анализ работы СМО существенно упрощается, если процесс этой работы — марковский. Случайный процесс называется *марковским* или *случайным процессом без последствия,* если для любого момента времени t_0 вероятностные характеристики процесса в будущем зависят только от его состояния в данный момент t_0 и не зависят от того, когда и как система пришла в это состояние.

Пример марковского процесса: система S — счетчик в такси. Состояние системы в момент t характеризуется числом километров (десятых долей километров), пройденных автомобилем до данного момента. Пусть в момент t_0 счетчик показывает S_0. Вероятность того, что в момент $t > t_0$ счетчик покажет то или иное число километров (точнее, соответствующее число рублей) S_1, зависит от S_0, но не зависит от того, в какие моменты времени изменялись показания счетчика до момента t_0.

Многие процессы можно приближенно считать марковскими. Например, процесс игры в шахматы; система S — группа шахматных фигур. Состояние системы характеризуется числом фигур противника, сохранившихся на доске в момент t_0. Вероятность того, что в момент $t > t_0$ материальный перевес будет на стороне одного из противников, зависит в первую очередь от того, в каком состоянии находится система в данный момент t_0, а не от того, когда и в какой последовательности исчезли фигуры с доски до момента t_0.

В ряде случаев предысторией рассматриваемых процессов можно просто пренебречь и применять для их изучения марковские модели.

При анализе случайных процессов с дискретными состояниями удобно пользоваться геометрической схемой — так называемым *графом состояний*. Обычно состояния системы изображаются прямоугольниками (кружками), а возможные переходы из состояния в состояние — стрелками (ориентированными дугами), соединяющими состояния.

▷ **15.1.** Построить граф состояний следующего случайного процесса: устройство S состоит из двух узлов, каждый из которых в случайный момент времени может выйти из строя, после чего мгновенно начинается ремонт узла, продолжающийся заранее неизвестное случайное время.

Решение. Возможные состояния системы: S_0 — оба узла исправны; S_1 — первый узел ремонтируется, второй исправен; S_2 — второй узел ремонтируется, первый исправен; S_3 — оба узла ремонтируются. Граф системы приведен на рис. 15.1.

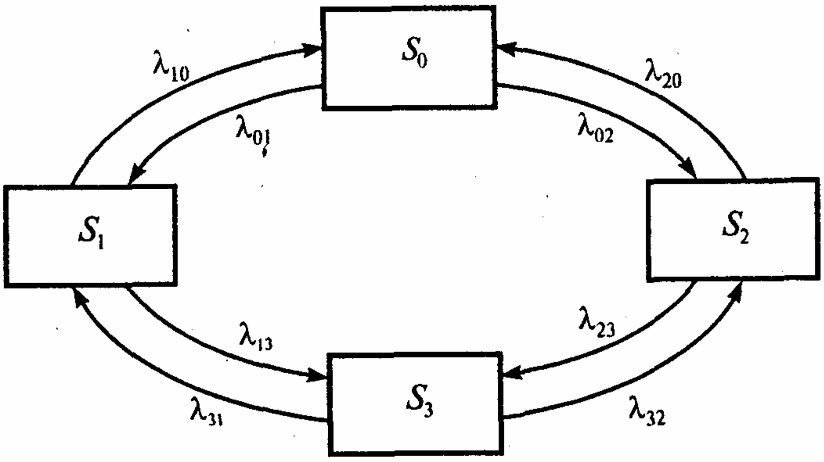


Рис. 15.1

Стрелка, направленная, например, из S_0 в S_1, означает переход системы в момент отказа первого узла, из S_1 в S_0 — переход в момент окончания ремонта этого узла.

На графе отсутствуют стрелки из S_0 в S_3 и из S_1 в S_2. Это объясняется тем, что выходы узлов из строя предполагаются независимыми друг от друга и, например, вероятностью одновременного выхода из строя двух узлов (переход из S_0 в S_3) или одновременного окончания ремонтов двух узлов (переход из S_3 в S_0) можно пренебречь.▶

Для математического описания марковского случайного процесса с дискретными состояниями и непрерывным временем, протекающего в СМО, познакомимся с одним из важных понятий теории вероятностей — понятием потока событий.

15.3. Потоки событий

Под *потоком событий* понимается последовательность однородных событий, следующих одно за другим в какие-то случайные моменты времени (например, поток вызовов на телефонной станции, поток отказов ЭВМ, поток покупателей и т.п.).

Поток характеризуется *интенсивностью* λ — частотой появления событий или средним числом событий, поступающих в СМО в единицу времени.

Поток событий называется *регулярным*, если события следуют одно за другим через определенные равные промежутки времени. Например, поток изделий на конвейере сборочного цеха (с постоянной скоростью движения) является регулярным.

Поток событий называется *стационарным*, если его вероятностные характеристики не зависят от времени. В частности, интенсивность стационарного потока есть величина постоянная: $\lambda(t)=\lambda$. Например, поток автомобилей на городском проспекте не является стационарным в течение суток, но этот поток можно считать стационарным в течение суток, скажем, в часы пик. Обращаем внимание на то, что в последнем случае фактическое число проходящих автомобилей в единицу времени (например, в каждую минуту) может заметно отличаться друг от друга, но среднее их число будет постоянно и не будет зависеть от времени.

Поток событий называется *потоком без последействия*, если для любых двух непересекающихся участков времени τ_1 и τ_2 — число событий, попадающих на один из них, не зависит от числа событий, попадающих на другие. Например, поток пассажиров, входящих в метро, практически не имеет последействия. А, скажем, поток покупателей, отходящих с покупками от прилавка, уже имеет последействие (хотя бы потому, что интервал времени между отдельными покупателями не может быть меньше, чем минимальное время обслуживания каждого из них).

Поток событий называется *ординарным*, если вероятность попадания на малый (элементарный) участок времени Δt двух и более событий пренебрежимо мала по сравнению с вероятностью попадания одного события. Другими словами, поток событий ординарен, если события появляются в нем поодиночке, а не группами. Например, поток поездов, подходящих к станции, ординарен, а поток вагонов не ординарен.

Поток событий называется ***простейшим*** *(или* ***стационарным пуассоновским****), если он одновременно стационарен, ординарен и не имеет последействия.* Название "простейший" объясняется тем, что СМО с простейшими потоками имеет наиболее простое математическое описание. Заметим, что регулярный поток не является "простейшим", так как он обладает последействием: моменты появления событий в таком потоке жестко зафиксированы.

Простейший поток в качестве предельного возникает в теории случайных процессов столь же естественно, как в теории вероятностей нормальное распределение получается в качестве предельного для суммы случайных величин: *при наложении (суперпозиции) достаточно большого числа n независимых, стационарных и ординарных потоков (сравнимых между собой по интенсивностям λ_i (i=1,2, ..., n) получается поток, близкий к простейшему с интенсивностью λ, равной сумме интенсивностей входящих потоков,* т.е.

$$\lambda=\sum_{i=1}^{n}\lambda_i .$$

Рассмотрим на оси времени Ot (рис. 15.2) простейший поток событий как неограниченную последовательность случайных точек.

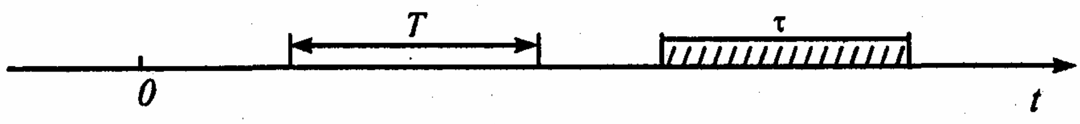


Рис. 15.2

Можно показать (*см.*, например, [3]), что для простейшего потока число m событий (точек), попадающих на произвольный участок времени τ, распределено по *закону Пуассона*

$$P_m(\tau) = \frac{(\lambda\tau)^m}{m!} e^{-\lambda\tau}, \tag{15.1}$$

для которого математическое ожидание случайной величины равно ее дисперсии: $a = \sigma^2 = \lambda\tau$.

В частности, вероятность того, что за время τ не произойдет ни одного события (m=0), равна

$$P_0(\tau) = e^{-\lambda\tau}. \tag{15.2}$$

Найдем распределение интервала времени T между произвольными двумя соседними событиями простейшего потока.

В соответствии с (15.2) вероятность того, что на участке времени длиной t не появится ни одного из последующих событий, равна

$$P(T \geq t) = e^{-\lambda t}, \tag{15.3}$$

а вероятность противоположного события, т.е. функция распределения случайной величины T, есть

$$F(t) = P(T < t) = 1 - e^{-\lambda t}. \tag{15.4}$$

Плотность вероятности случайной величины есть производная ее функции распределения (рис. 15.3), т.е.

$$\varphi(t) = F'(t) = \lambda e^{-\lambda t}. \tag{15.5}$$

Распределение, задаваемое плотностью вероятности (15.5) или функцией распределения (15.4), называется *показательным* (или *экспоненциальным*). Таким образом, интервал времени между двумя соседними произвольными событиями имеет показательное распределение, для которого матема-

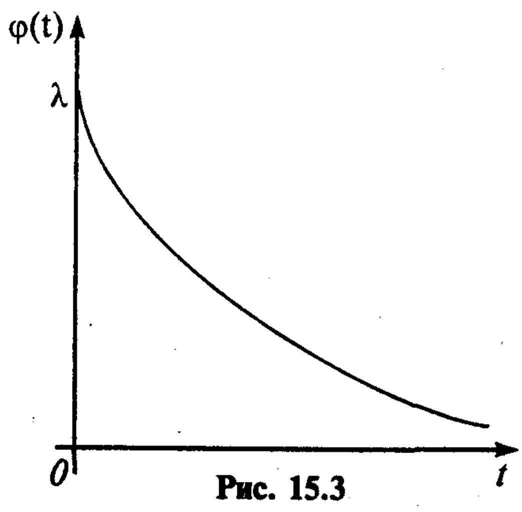


Рис. 15.3

тическое ожидание равно среднему квадратическому отклонению случайной величины

$$a = \sigma = \frac{1}{\lambda} \qquad (15.6)$$

и обратно по величине интенсивности потока λ.

Важнейшее свойство показательного распределения (присущее только показательному распределению) состоит в следующем: если промежуток времени, распределенный по показательному закону, уже длился некоторое время τ, то это никак не влияет на закон распределения оставшейся части промежутка ($T-\tau$): он будет таким же, как и закон распределения всего промежутка T.

Другими словами, для интервала времени T между двумя последовательными соседними событиями потока, имеющего показательное распределение, любые сведения о том, сколько времени протекал этот интервал, не влияют на закон распределения оставшейся части. Это свойство показательного закона представляет собой, в сущности, другую формулировку для "отсутствия последействия" — основного свойства простейшего потока.

Для простейшего потока с интенсивностью λ вероятность попадания на *элементарный* (*малый*) отрезок времени Δt хотя бы одного события потока равна согласно (15.4)

$$P_{\Delta t} = P(T < \Delta t) = 1 - e^{-\lambda \Delta t} \approx \lambda \Delta t\,. \qquad (15.7)$$

(Заметим, что эта приближенная формула, получаемая заменой функции $e^{-\lambda \Delta t}$ лишь двумя первыми членами ее разложения в ряд по степеням Δt, тем точнее, чем меньше Δt).

15.4. Уравнения Колмогорова. Предельные вероятности состояний

Рассмотрим математическое описание марковского процесса с дискретными состояниями и непрерывным временем на примере случайного процесса из задачи **15.1**, граф которого изображен на рис. 15.1. Будем полагать, что все переходы системы из состояния S_i в S_j происходят под воздействием простейших потоков событий с интенсивностями λ_{ij} (i, j=0, 1, 2, 3); так, переход системы из состояния S_0 в S_1 будет происходить под воздействием потока отказов первого узла, а обратный переход из состояния S_1 в S_0 — под воздействием потока "окончаний ремонтов" первого узла и т.п.

Граф состояний системы с проставленными у стрелок интенсивностями будем называть *размеченным* (*см.* рис. 15.1). Рассматриваемая система S имеет четыре возможных состояния: S_0, S_1, S_2, S_3.

Вероятностью i-го состояния называется вероятность $p_i(t)$ того, что в момент t система будет находиться в состоянии S_i. Очевидно, что для любого момента t сумма вероятностей всех состояний равна единице:

$$\sum_{i=0}^{3} p_i(t) = 1. \qquad (15.8)$$

Рассмотрим систему в момент t и, задав малый промежуток Δt, найдем вероятность $p_0(t+\Delta t)$ того, что система в момент $t+\Delta t$ будет находиться в состоянии S_0. Это достигается разными способами.

1. Система в момент t с вероятностью $p_0(t)$ находилась в состоянии S_0, а за время Δt не вышла из него.

Вывести систему из этого состояния (*см.* граф на рис. 15.1) можно суммарным простейшим потоком с интенсивностью $(\lambda_{01}+\lambda_{02})$, т.е. в соответствии с (15.7), с вероятностью, приближенно равной $(\lambda_{01}+\lambda_{02})\Delta t$. А вероятность того, что система не выйдет из состояния S_0, равна $[1-(\lambda_{01}+\lambda_{02})\Delta t]$. Вероятность того, что система будет находиться в состоянии S_0 по первому способу (т.е. того, что находилась в состоянии S_0 и не выйдет из него за время Δt), равна по теореме умножения вероятностей:

$$p_0(t)\left[1-(\lambda_{01}+\lambda_{02})\Delta t\right].$$

2. Система в момент t с вероятностями $p_1(t)$ (или $p_2(t)$) находилась в состоянии S_1 или S_2 и за время Δt перешла в состояние S_0.

Потоком интенсивностью λ_{10} (или λ_{20} — *см.* рис. 15.1) система перейдет в состояние S_0 с вероятностью, приближенно равной $\lambda_{10}\Delta t$ (или $\lambda_{20}\Delta t$). Вероятность того, что система будет находиться в состоянии S_0 по этому способу, равна $p_1(t)\lambda_{10}\Delta t$ (или $p_2(t)\lambda_{20}\Delta t$).

Применяя теорему сложения вероятностей, получим

$$p_0(t+\Delta t) = p_1(t)\lambda_{10}\Delta t + p_2(t)\lambda_{20}\Delta t + p_0(t)\left[1-(\lambda_{01}+\lambda_{02})\Delta t\right],$$

откуда

$$\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t}=p_1(t)\lambda_{10}+p_2(t)\lambda_{20}-\left(\lambda_{01}+\lambda_{02}\right)p_0(t).$$

Переходя к пределу при $\Delta t \to 0$ (приближенные равенства, связанные с применением формулы (15.7), перейдут в точные), получим в левой части уравнения производную $p_0'(t)$ (обозначим ее для простоты p_0'):

$$p_0'=\lambda_{10}p_1+\lambda_{20}p_2-\left(\lambda_{01}+\lambda_{02}\right)p_0.$$

Получили дифференциальное уравнение первого порядка, т.е. уравнение, содержащее как саму неизвестную функцию, так и ее производную первого порядка.

Рассуждая аналогично для других состояний системы S, можно получить систему дифференциальных уравнений Колмогорова для вероятностей состояний:

$$\begin{cases} p_0'=\lambda_{10}p_2+\lambda_{20}p_2-(\lambda_{01}+\lambda_{02})p_0, \\ p_1'=\lambda_{01}p_0+\lambda_{31}p_3-(\lambda_{10}+\lambda_{13})p_1, \\ p_2'=\lambda_{02}p_0+\lambda_{32}p_3-(\lambda_{20}+\lambda_{23})p_2, \\ p_3'=\lambda_{13}p_1+\lambda_{23}p_2-(\lambda_{31}+\lambda_{32})p_3. \end{cases} \quad (15.9)$$

Сформулируем **правило составления уравнений Колмогорова**. *В левой части каждого из них стоит производная вероятности i-го состояния. В правой части — сумма произведений вероятностей всех состояний (из которых идут стрелки в данное состояние) на интенсивности соответствующих потоков событий, минус суммарная интенсивность всех потоков, выводящих систему из данного состояния, умноженная на вероятность данного (i-го состояния).*

В системе (15.9) независимых уравнений на единицу меньше общего числа уравнений. Поэтому для решения системы необходимо добавить уравнение (15.8).

Особенность решения дифференциальных уравнений вообще состоит в том, что требуется задать так называемые начальные условия, т.е. в данном случае вероятности состояний системы в начальный момент $t = 0$. Так, например, систему уравнений (15.9) естественно решать при условии, что в начальный момент оба узла исправны и система находилась в состоянии S_0, т.е. при начальных условиях $p_0(0)=1$, $p_1(0)=p_2(0)=p_3(0)=0$.

Уравнения Колмогорова дают возможность найти все вероятности состояний как *функции времени*. Особый интерес представляют вероятности системы $p_i(t)$ в *предельном стационарном режиме*, т.е. при $t \to \infty$, которые называются *предельными* (или *финальными*) вероятностями состояний.

В теории случайных процессов доказывается, что *если число состояний системы конечно и из каждого из них можно (за конечное число шагов) перейти в любое другое состояние, то предельные вероятности существуют.*

Предельная вероятность состояния S_i имеет четкий смысл: она показывает *среднее относительное время пребывания системы в этом состоянии*. Например, если предельная вероятность состояния S_0, т.е. $p_0=0{,}5$, то это означает, что в среднем половину времени система находится в состоянии S_0.

Так как предельные вероятности постоянны, то, заменяя в уравнениях Колмогорова их производные нулевыми значениями, получим систему линейных алгебраических уравнений, описывающих стационарный режим. Для системы S с графом состояний, изображенном на рис. 15.1), такая система уравнений имеет вид:

$$\begin{cases} (\lambda_{01}+\lambda_{02})p_0 = \lambda_{10}p_1+\lambda_{20}p_2, \\ (\lambda_{10}+\lambda_{13})p_1 = \lambda_{01}p_0+\lambda_{31}p_3, \\ (\lambda_{20}+\lambda_{23})p_2 = \lambda_{02}p_0+\lambda_{32}p_3, \\ (\lambda_{31}+\lambda_{32})p_3 = \lambda_{13}p_1+\lambda_{23}p_2. \end{cases} \qquad \textbf{(15.10)}$$

Систему (15.10) можно составить непосредственно по размеченному графу состояний, если руководствоваться п р а в и л о м , согласно которому *слева в уравнениях стоит предельная вероятность данного состояния p_i, умноженная на суммарную интенсивность всех потоков, ведущих из данного состояния, а справа — сумма произведений интенсивностей всех потоков, входящих в i-е состояние, на вероятности тех состояний, из которых эти потоки исходят.*

▷ 15.2. Найти предельные вероятности для системы S из задачи 15.1, граф состояний которой приведен на рис. 15.1, при $\lambda_{01}=1$, $\lambda_{02}=2$, $\lambda_{10}=2$, $\lambda_{13}=2$, $\lambda_{20}=3$, $\lambda_{23}=1$, $\lambda_{31}=3$, $\lambda_{32}=2$.

Р е ш е н и е. Система алгебраических уравнений, описывающих стационарный режим для данной системы, имеет вид (15.10) или

$$\begin{cases} 3p_0 = 2p_1 + 3p_2, \\ 4p_1 = p_0 + 3p_3, \\ 4p_2 = 2p_0 + 2p_3, \\ p_0 + p_1 + p_2 + p_3 = 1. \end{cases} \qquad \textbf{(15.11)}$$

(Здесь мы вместо одного "лишнего" уравнения системы (15.10) записали нормировочное условие (15.8)).

Решив систему (15.11), получим p_0=0,40, p_1=0,20, p_2=0,27, p_3=0,13, т.е. в предельном, стационарном режиме система S в среднем 40% времени будет находиться в состоянии S_0 (оба узла исправны), 20% — в состоянии S_1 (первый узел ремонтируется, второй работает), 27% — в состоянии S_2 (второй узел ремонтируется, первый работает) и 13% времени — в состоянии S_3 (оба узла ремонтируются).▶

▷ **15.3.** Найти средний чистый доход от эксплуатации в стационарном режиме системы S в условиях задач **15.1** и **15.2**, если известно, что в единицу времени исправная работа первого и второго узлов приносит доход соответственно в 10 и 6 ден.ед., а их ремонт требует затрат соответственно в 4 и 2 ден.ед. Оценить экономическую эффективность имеющейся возможности уменьшения вдвое среднего времени ремонта каждого из двух узлов, если при этом придется вдвое увеличить затраты на ремонт каждого узла (в единицу времени).

Р е ш е н и е. Из задачи **15.2** следует, что в среднем первый узел исправно работает долю времени, равную p_0+p_3=0,40+0,27=0,67, а второй узел — p_0+p_1=0,40+0,20=0,60. В то же время первый узел находится в ремонте в среднем долю времени, равную p_1+p_3= =0,20+0,13=0,33, а второй узел — p_2+p_3=0,27+0,13=0,40. Поэтому средний чистый доход в единицу времени от эксплуатации системы, т.е. разность между доходами и затратами, равен

$$Д=0,67\cdot 10+0,60\cdot 6-0,33\cdot 4-0,40\cdot 2=8,18 \text{ ден.ед.}$$

Уменьшение вдвое среднего времени ремонта каждого из узлов в соответствии с (15.6) будет означать увеличение вдвое интенсивностей потока "окончаний ремонтов" каждого узла, т.е. теперь λ_{10}=4, λ_{20}=6, λ_{31}=6, λ_{32}=4 и система линейных алгебраических уравнений (15.10), описывающая стационарный режим системы S, вместе с нормировочным условием (15.8) примет вид[1]:

$$\begin{cases} 3p_0 = 4p_1 + 6p_2, \\ 6p_1 = p_0 + 6p_3, \\ 7p_2 = 2p_0 + 4p_3, \\ p_0 + p_1 + p_2 + p_3 = 1. \end{cases}$$

Решив систему, получим p_0=0,60, p_1=0,15, p_2=0,20, p_3=0,05.

Учитывая, что p_0+p_2=0,60+0,20=0,80, p_0+p_1=0,60+0,15=0,75, p_1+p_3=0,15+0,05=0,20, p_2+p_3=0,20+0,05=0,25, а затраты на ремонт первого и второго узла составляют теперь соответственно 8 и 4 ден. ед., вычислим средний чистый доход в единицу времени:

$$Д_1=0{,}80\cdot 10+0{,}75\cdot 6-0{,}20\cdot 8-0{,}25\cdot 4=9{,}9 \text{ ден.ед.}$$

Так как $Д_1$ больше Д (примерно на 20%), то экономическая целесообразность ускорения ремонтов узлов очевидна.▶

15.5. Процесс гибели и размножения

В теории массового обслуживания широкое распространение имеет специальный класс случайных процессов — так называемый *процесс гибели и размножения*. Название этого процесса связано с рядом биологических задач, где он является математической моделью изменения численности биологических популяций.

Граф состояний процесса гибели и размножения имеет вид, показанный на рис. 15.4.

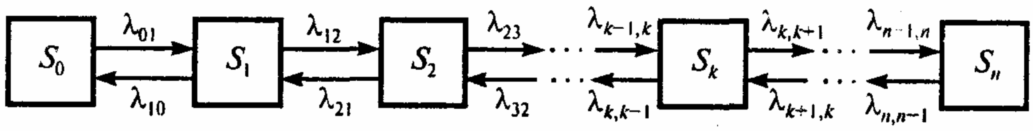


Рис. 15.4

Рассмотрим упорядоченное множество состояний системы S_0, S_1, S_2, ..., S_k. Переходы могут осуществляться из любого со-

[1] При записи системы (15.10) одно "лишнее" уравнение мы исключили.

стояния только в состояния с соседними номерами, т.е. из состояния S_k возможны переходы только либо в состояние S_{k-1}, либо в состояние S_{k+1}[1].

Предположим, что все потоки событий, переводящие систему по стрелкам графа, простейшие с соответствующими интенсивностями $\lambda_{k,k+1}$ или $\lambda_{k+1,k}$.

По графу, представленному на рис. 15.4, составим и решим алгебраические уравнения для предельных вероятностей состояний (их существование вытекает из возможности перехода из каждого состояния в каждое другое и конечности числа состояний).

В соответствии с правилом составления таких уравнений (*см.* 15.13) получим: для состояния S_0

$$\lambda_{01} p_0 = \lambda_{10} p_1, \tag{15.12}$$

для состояния S_1 – $(\lambda_{12}+\lambda_{10})p_1 = \lambda_{01}p_0+\lambda_{21}p_2$, которое с учетом (15.12) приводится к виду

$$\lambda_{12} p_1 = \lambda_{21} p_2. \tag{15.13}$$

Аналогично, записывая уравнения для предельных вероятностей других состояний, можно получить следующую систему уравнений:

$$\begin{cases} \lambda_{01} p_0 = \lambda_{10} p_1, \\ \lambda_{12} p_1 = \lambda_{21} p_2, \\ \dots\dots\dots\dots\dots\dots \\ \lambda_{k-1,k} p_{k-1} = \lambda_{k,k+1} p_k, \\ \dots\dots\dots\dots\dots\dots \\ \lambda_{n-1,n} p_{n-1} = \lambda_{n,n+1} p_n, \end{cases} \tag{15.14}$$

к которой добавляется нормировочное условие

$$p_0 + p_1 + p_2 + \ldots + p_n = 1. \tag{15.15}$$

---

[1] При анализе численности популяций считают, что состояние S_k соответствует численности популяции, равной k, и переход системы из состояния S_k в состояние S_{k+1} происходит при рождении одного члена популяции, а переход в состояние S_{k-1} — при гибели одного члена популяции.

Решая систему (15.14), (15.15), можно получить

$$p_0=\left(1+\frac{\lambda_{01}}{\lambda_{10}}+\frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}+\ldots+\frac{\lambda_{n-1,n}\cdots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\cdots\lambda_{21}\lambda_{10}}\right)^{-1}, \qquad (15.16)$$

$$p_1=\frac{\lambda_{01}}{\lambda_{10}}p_0,\ p_2=\frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}}p_0,\ \ldots,\ p_n=\frac{\lambda_{n-1,n}\cdots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\cdots\lambda_{21}\lambda_{10}}p_0. \qquad (15.17)$$

Легко заметить, что в формулах (15.17) для p_1, p_2, ..., p_n коэффициенты при p_0 есть слагаемые, стоящие после единицы в формуле (15.16). Числители этих коэффициентов представляют произведение всех интенсивностей, стоящих у стрелок, ведущих слева направо до данного состояния S_k (k=1, 2, ..., n), а знаменатели — произведение всех интенсивностей, стоящих у стрелок, ведущих справа налево до состояния S_k.

▷ **15.4.** Процесс гибели и размножения представлен графом (рис. 15.5). Найти предельные вероятности состояний.

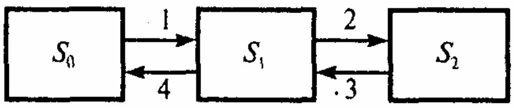


Рис. 15.5

Решение. По формуле (15.16) найдем

$$p_0=\left(1+\frac{1}{4}+\frac{2\cdot 1}{3\cdot 4}\right)^{-1}=0{,}706,$$

по (15.17) — $p_1=\frac{1}{4}0{,}706=0{,}176$, $p_2=\frac{2\cdot 1}{3\cdot 4}0{,}706=0{,}118$, т.е. в установившемся, стационарном режиме в среднем 70,6% времени система будет находиться в состоянии S_0, 17,6% — в состоянии S_1 и 11,8% — в состоянии S_2.▶

15.6. СМО с отказами

В качестве показателей эффективности СМО с отказами будем рассматривать:

А — абсолютную пропускную способность СМО, т.е. среднее число заявок, обслуживаемых в единицу времени;

Q — *относительную пропускную способность*, т.е. среднюю долю пришедших заявок, обслуживаемых системой;

$P_{отк.}$ — *вероятность отказа*, т.е. того, что заявка покинет СМО необслуженной;

$\bar{k}$ — *среднее число занятых каналов* (для многоканальной системы).

Одноканальная система с отказами. Рассмотрим задачу.

Имеется один канал, на который поступает поток заявок с интенсивностью λ. Поток обслуживаний имеет интенсивность μ[1]. Найти предельные вероятности состояний системы и показатели ее эффективности.

Система S (СМО) имеет два состояния: S_0 — канал свободен, S_1 — канал занят. Размеченный граф состояний представлен на рис. 15.6.

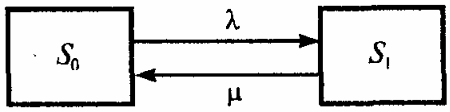


Рис. 15.6

В предельном, стационарном режиме система алгебраических уравнений для вероятностей состояний имеет вид (*см.* правило составления таких уравнений на с. 343).

$$\begin{cases} \lambda p_0 = \mu p_1, \\ \mu p_1 = \lambda p_0, \end{cases} \qquad \textbf{(15.18)}$$

т.е. система вырождается в одно уравнение. Учитывая нормировочное условие $p_0 + p_1 = 1$, найдем из (15.18) предельные вероятности состояний

$$p_0 = \frac{\mu}{\lambda + \mu}, \; p_1 = \frac{\lambda}{\lambda + \mu}, \qquad \textbf{(15.19)}$$

которые выражают среднее относительное время пребывания системы в состоянии S_0 (когда канал свободен) и S_1 (когда канал

---

[1] **Здесь и в дальнейшем предполагается, что все потоки событий, переводящие СМО из состояния в состояние, будут простейшими. К ним относится и поток обслуживаний — поток заявок, обслуживаемых одним непрерывно занятым каналом. Среднее время обслуживания $\bar{t}_{об.}$ обратно по величине интенсивности μ, т.е. $\bar{t}_{об.} = 1/\mu$.**

занят), т.е. определяют соответственно относительную пропускную способность Q системы и вероятность отказа $P_{отк}$:

$$Q = \frac{\mu}{\lambda + \mu}, \tag{15.20}$$

$$P_{отк.} = \frac{\lambda}{\lambda + \mu}. \tag{15.21}$$

Абсолютную пропускную способность найдем, умножив относительную пропускную способность Q на интенсивность потока отказов

$$A = \frac{\lambda\mu}{\lambda + \mu}. \tag{15.22}$$

▷ **15.5.** Известно, что заявки на телефонные переговоры в телевизионном ателье поступают с интенсивностью λ, равной 90 заявок в час, а средняя продолжительность разговора по телефону $\bar{t}_{об.}=2$ мин. Определить показатели эффективности работы СМО (телефонной связи) при наличии одного телефонного номера.

Решение. Имеем $\lambda=90$ (1/ч), $\bar{t}_{об.}=2$ мин. Интенсивность потока обслуживаний $\mu=1/\bar{t}_{об.}=1/2=0,5$ (1/мин)=30 (1/ч). По (15.20) относительная пропускная способность СМО $Q=30/(90+30)=0,25$, т.е. в среднем только 25% поступающих заявок осуществят переговоры по телефону. Соответственно вероятность отказа в обслуживании составит $P_{отк.}=0,75$ (*см.* (15.21)). Абсолютная пропускная способность СМО по (15.29) $A=90\cdot0,25=22,5$, т.е. в среднем в час будут обслужены 22,5 заявки на переговоры. Очевидно, что при наличии только одного телефонного номера СМО будет плохо справляться с потоком заявок.▶

Многоканальная система с отказами. Рассмотрим классическую *задачу Эрланга.*

Имеется n каналов, на которые поступает поток заявок с интенсивностью λ. Поток обслуживаний имеет интенсивность μ. Найти предельные вероятности состояний системы и показатели ее эффективности.

Система S (СМО) имеет следующие состояния (нумеруем их по числу заявок, находящихся в системе): S_0, S_1, S_2, ..., S_k, ... S_n, где S_k — состояние системы, когда в ней находится k заявок, т.е. занято k каналов.

Граф состояний СМО соответствует процессу гибели и размножения и показан на рис. 15.7.

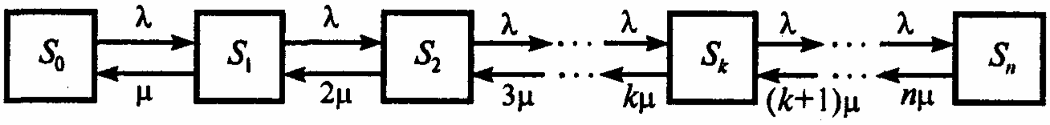


Рис. 15.7

Поток заявок последовательно переводит систему из любого левого состояния в соседнее правое с одной и той же интенсивностью λ. Интенсивность же потока обслуживаний, переводящих систему из любого правого состояния в соседнее левое состояние, постоянно меняется в зависимости от состояния. Действительно, если СМО находится в состоянии S_2 (два канала заняты), то она может перейти в состояние S_1 (один канал занят), когда закончит обслуживание либо первый, либо второй канал, т.е. суммарная интенсивность их потоков обслуживаний будет 2μ. Аналогично суммарный поток обслуживаний, переводящий СМО из состояния S_3 (три канала заняты) в S_2, будет иметь интенсивность 3μ, т.е. может освободиться любой из трех каналов и т.д.

В формуле (15.16) для схемы гибели и размножения получим для предельной вероятности состояния

$$p_0=\left(1+\frac{\lambda}{\mu}+\frac{\lambda^2}{2!\mu^2}+\ldots+\frac{\lambda^k}{k!\mu^k}+\ldots+\frac{\lambda^n}{n!\mu^n}\right)^{-1}, \qquad (15.23)$$

где члены разложения $\frac{\lambda}{\mu}$, $\frac{\lambda^2}{2!\mu^2}$, ..., $\frac{\lambda^n}{n!\mu^n}$ будут представлять собой коэффициенты при p_0 в выражениях для предельных вероятностей $p_1, p_2, \ldots, p_k, \ldots, p_n$. Величина

$$\rho=\frac{\lambda}{\mu} \qquad (15.24)$$

называется ***приведенной интенсивностью потока заявок*** или ***интенсивностью нагрузки канала***. Она выражает среднее число заявок, приходящее за среднее время обслуживания одной заявки. Теперь

$$p_0=\left(1+\rho+\frac{\rho^2}{2!}+\ldots+\frac{\rho^k}{k!}+\ldots+\frac{\rho^n}{n!}\right)^{-1}, \qquad (15.25)$$

$$p_1=\rho p_0,\ p_2=\frac{\rho^2}{2!}p_0,\ \ldots,\ p_k=\frac{\rho^k}{k!}p_0,\ \ldots,\ p_n=\frac{\rho^n}{n!}p_0. \qquad (15.26)$$

Формулы (15.25) и (15.26) для предельных вероятностей получили названия *формул Эрланга*[1] в честь основателя теории массового обслуживания.

Вероятность отказа СМО есть предельная вероятность того, что все n каналов системы будут заняты, т.е.

$$P_{\text{отк.}}=\frac{\rho^n}{n!}p_0. \qquad (15.27)$$

Относительная пропускная способность — вероятность того, что заявка будет обслужена:

$$Q=1-P_{\text{отк}}=1-\frac{\rho^n}{n!}p_0. \qquad (15.28)$$

Абсолютная пропускная способность:

$$A=\lambda Q=\lambda\left(1-\frac{\rho^n}{n!}p_0\right). \qquad (15.29)$$

Среднее число занятых каналов $\bar{k}$ есть математическое ожидание числа занятых каналов:

$$\bar{k}=\sum_{k=0}^{n}kp_k,$$

где p_k — предельные вероятности состояний, определяемых по формулам (15.25), (15.26).

Однако среднее число занятых каналов можно найти проще, если учесть, что абсолютная пропускная способность системы A есть не что иное, как интенсивность потока *обслуженных* системой заявок (в единицу времени). Так как каждый занятый канал обслуживает в среднем μ заявок (в единицу времени), то среднее число занятых каналов

$$\bar{k}=\frac{A}{\mu} \qquad (15.30)$$

---

[1] **Эрланг А.К. (конец XIX в. — начало XX в.) — датский инженер, математик.**

или, учитывая (15.29), (15.24):

$$\bar{k} = \rho\left(1 - \frac{\rho^n}{n!} p_0\right). \qquad \textbf{(15.31)}$$

▷ **15.6.** В условиях задачи **15.5** определить оптимальное число телефонных номеров в телевизионном ателье, если условием оптимальности считать удовлетворение в среднем из каждых 100 заявок не менее 90 заявок на переговоры.

Решение. Интенсивность нагрузки канала по формуле (15.25) $\rho=90/30=3$, т.е. за время среднего (по продолжительности) телефонного разговора $\bar{t}_{\text{об.}}=2$ мин. поступает в среднем 3 заявки на переговоры.

Будем постепенно увеличивать число каналов (телефонных номеров) $n=2$, 3, 4, ... и определим по формулам (15.25), (15.28), (15.29) для получаемой n-канальной СМО характеристики обслуживания. Например, при $n = 2$ $p_0 = \left(1 + 3 + 3^2/2!\right)^{-1} = 0{,}118 \approx 0{,}12$; $Q = 1 - \left(3^2/2!\right) \cdot 0{,}118 = 0.471 \approx 0{,}47$; $A=90 \cdot 0{,}471=42{,}4$ и т.д. Значение характеристик СМО сведем в табл. 15.1.

Таблица 15.1

Характеристика обслуживания	Число каналов (телефонных номеров)					
	1	2	3	4	5	6
Относительная пропускная способность Q	0,25	0,47	0,65	0,79	0,90	0,95
Абсолютная пропускная способность A	22,5	42,4	58,8	71,5	80,1	85,3

По условию оптимальности $Q \geq 0{,}9$, следовательно, в телевизионном ателье необходимо установить 5 телефонных номеров (в этом случае $Q = 0{,}90$ — *см.* табл. 15.1). При этом в час будут обслуживаться в среднем 80 заявок ($A = 80{,}1$), а среднее число занятых телефонных номеров (каналов) по формуле (15.30) $\bar{k} = 80{,}1/30 = 2{,}67$.▶

▷ **15.7.** В вычислительный центр коллективного пользования с тремя ЭВМ поступают заказы от предприятий на вычислительные работы. Если работают все три ЭВМ, то вновь поступающий заказ не принимается, и предприятие вынуждено обратиться в другой вычислительный центр. Среднее время работы с одним заказом составляет 3 ч. Интенсивность потока заявок 0,25 (1/ч). Найти предельные вероятности состояний и показатели эффективности работы вычислительного центра.

Решение. По условию n=3, λ=0,25 (1/ч), $\bar{t}_{об.}$=3 (ч). Интенсивность потока обслуживаний $\mu=1/\bar{t}_{об.}=1/3=0,33$. Интенсивность нагрузки ЭВМ по формуле (15.24) $\rho=0,25/0,33=0,75$. Найдем предельные вероятности состояний:

по формуле (15.25) $p_0=(1+0,75+0,75^2/2!+0,75^3/3!)^{-1}=0,476$;

по формуле (15.26) $p_1=0,75\cdot 0,476=0,357$; $p_2=(0,75^2/2!)\cdot 0,476=$ $=0,134$; $p_3=(0,75^3/3!)\cdot 0,476=0,033$, т.е. в стационарном режиме работы вычислительного центра в среднем 47,6% времени нет ни одной заявки, 35,7% — имеется одна заявка (занята одна ЭВМ), 13,4% — две заявки (две ЭВМ), 3,3% времени — три заявки (заняты три ЭВМ).

Вероятность отказа (когда заняты все три ЭВМ), таким образом, $P_{отк.}=p_3=0,033$.

По формуле (15.28) относительная пропускная способность центра Q = 1−0,033 = 0,967, т.е. в среднем из каждых 100 заявок вычислительный центр обслуживает 96,7 заявок.

По формуле (15.29) абсолютная пропускная способность центра $A = 0,25\cdot 0,967 = 0,242$, т.е. в один час в среднем обслуживается 0,242 заявки.

По формуле (15.30) среднее число занятых ЭВМ $\bar{k}$ = = 0,242/0,33 = 0,725, т.е. каждая из трех ЭВМ будет занята обслуживанием заявок в среднем лишь на 72,5/3 = 24,2%.

При оценке эффективности работы вычислительного центра необходимо сопоставить доходы от выполнения заявок с потерями от простоя дорогостоящих ЭВМ (с одной стороны, у нас высокая пропускная способность СМО, а с другой стороны — значительный простой каналов обслуживания) и выбрать компромиссное решение.▶

15.7. СМО с ожиданием (очередью)

В качестве показателей эффективности СМО с ожиданием, кроме уже известных показателей — абсолютной A и относительной Q пропускной способности, вероятности отказа $P_{отк.}$, среднего числа занятых каналов $\bar{k}$ (для многоканальной системы) будем рассматривать также следующие: $L_{сист.}$ — *среднее число заявок в системе*; $T_{сист.}$ — *среднее время пребывания заявки в системе*; $L_{оч.}$ — *среднее число заявок в очереди* (*длина очереди*); $T_{оч.}$ — *среднее время пребывания заявки в очереди*; $P_{зан.}$ — *вероятность того, что канал занят* (*степень загрузки канала*).

Одноканальная система с неограниченной очередью. На практике часто встречаются одноканальные СМО с неограниченной очередью (например, телефон-автомат с одной будкой). Рассмотрим задачу.

Имеется одноканальная СМО с очередью, на которую не наложены никакие ограничения (ни по длине очереди, ни по времени ожидания). Поток заявок, поступающих в СМО, имеет интенсивность λ, а поток обслуживаний — интенсивность μ. Необходимо найти предельные вероятности состояний и показатели эффективности СМО.

Система может находиться в одном из состояний $S_0, S_1, S_2, ..., S_k$, по числу заявок, находящихся в СМО: S_0 — канал свободен; S_1 — канал занят (обслуживает заявку), очереди нет; S_2 — канал занят, одна заявка стоит в очереди; ... S_k — канал занят, $(k-1)$ заявок стоят в очереди и т.д.

Граф состояний СМО представлен на рис. 15.8.

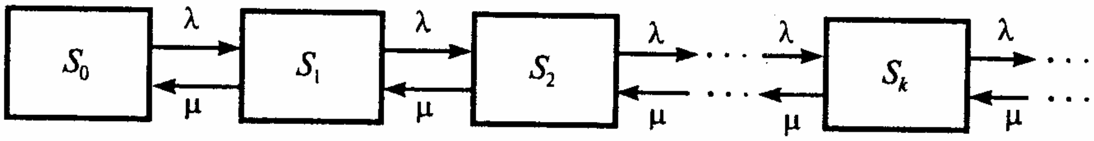


Рис. 15.8

Это процесс гибели и размножения, но с бесконечным числом состояний, в котором интенсивность потока заявок равна λ, а интенсивность потока обслуживаний μ.

Прежде чем записать формулы предельных вероятностей, необходимо быть уверенным в их существовании, ведь в случае, когда время $t \to \infty$, очередь может неограниченно возрастать. Доказано, что *если* $\rho < 1$, *т.е. среднее число приходящих заявок меньше среднего числа обслуженных заявок (в единицу времени), то предельные вероятности существуют. Если* $\rho \geq 1$, *очередь растет до бесконечности.*

Для определения предельных вероятностей состояний воспользуемся формулами (15.16), (15.17) для процесса гибели и размножения (здесь мы допускаем известную нестрогость, так как ранее эти формулы были получены для случая конечного числа состояний системы). Получим:

$$p_0=\left[1+\frac{\lambda}{\mu}+\left(\frac{\lambda}{\mu}\right)^2+\ldots+\left(\frac{\lambda}{\mu}\right)^k+\ldots\right]^{-1}=$$

$$=\left(1+\rho+\rho^2+\ldots+\rho^k+\ldots\right)^{-1}. \qquad \textbf{(15.32)}$$

Так как предельные вероятности существуют лишь при $\rho < 1$, то геометрический ряд со знаменателем $\rho < 1$, записанный в скобках в формуле (15.32), сходится к сумме, равной $\frac{1}{1-\rho}$. Поэтому

$$p_0=1-\rho, \qquad \textbf{(15.33)}$$

и с учетом соотношений (15.17)

$$p_1=\rho p_0,\ \ p_2=\rho^2 p_0,\ \ \ldots,\ \ p_k=\rho^k p_0,\ \ \ldots$$

найдем предельные вероятности других состояний

$$p_1=\rho(1-\rho),\ p_2=\rho^2(1-\rho),\ \ldots,\ p_k=\rho^k(1-\rho),\ \ \ldots\ . \qquad \textbf{(15.34)}$$

Предельные вероятности p_0, p_1, p_2, ..., p_k, ... образуют убывающую геометрическую прогрессию со знаменателем $\rho < 1$, следовательно, вероятность p_0 — наибольшая. Это означает, что если СМО справляется с потоком заявок (при $\rho < 1$), то наиболее вероятным будет отсутствие заявок в системе.

Среднее число заявок в системе $L_{\text{сист.}}$ определим по формуле математического ожидания, которая с учетом (15.34) примет вид

$$L_{\text{сист.}}=\sum_{k=1}^{\infty} kp_k=(1-\rho)\sum_{k=1}^{\infty} k\rho^k \qquad \textbf{(15.35)}$$

(суммирование от 1 до ∞, так как нулевой член $0p_0=0$).

Можно показать, что формула (15.35) преобразуется (при $\rho < 1$) к виду

$$L_{\text{сист.}} = \frac{\rho}{1-\rho}. \qquad \textbf{(15.36)}$$

Найдем среднее число заявок в очереди $L_{\text{оч}}$. Очевидно, что

$$L_{\text{оч.}} = L_{\text{сист.}} - L_{\text{об.}}, \qquad \textbf{(15.37)}$$

где $L_{\text{об.}}$ — среднее число заявок, находящихся под обслуживанием.

Среднее число заявок под обслуживанием определим по формуле математического ожидания числа заявок под обслуживанием, принимающего значения 0 (если канал свободен) либо 1 (если канал занят):

$$L_{\text{об.}} = 0 \cdot p_0 + 1(1 - p_0),$$

т.е. среднее число заявок под обслуживанием равно вероятности того, что канал занят:

$$L_{\text{об.}} = P_{\text{зан.}} = 1 - p_0. \qquad \textbf{(15.38)}$$

В силу (15.33)

$$L_{\text{об.}} = P_{\text{зан.}} = \rho. \qquad \textbf{(15.39)}$$

Теперь по формуле (15.37) с учетом (15.36) и (15.39)

$$L_{\text{оч.}} = \frac{\rho^2}{1-\rho}. \qquad \textbf{(15.40)}$$

Доказано, что *при любом характере потока заявок, при любом распределении времени обслуживания, при любой дисциплине обслуживания среднее время пребывания заявки в системе (очереди) равна среднему числу заявок в системе (в очереди), деленному на интенсивность потока заявок*, т.е.

$$T_{\text{сист.}} = \frac{1}{\lambda} L_{\text{сист.}}, \qquad \textbf{(15.41)}$$

$$T_{\text{оч.}} = \frac{1}{\lambda} L_{\text{оч.}}. \qquad \textbf{(15.42)}$$

Формулы (15.41) и (15.42) называются *формулами Литтла.* Они вытекают из того, что ***в предельном, стационарном режиме среднее число заявок, прибывающих в систему, равно среднему числу заявок, покидающих ее***: оба потока заявок имеют одну и ту же интенсивность λ.

На основании формул (15.41) и (15.42) с учетом (15.36) и (15.40) среднее время пребывания заявки в системе определится по формуле:

$$T_{\text{сист.}} = \frac{\rho}{\lambda(1-\rho)}, \quad \textbf{(15.43)}$$

а среднее время пребывания заявки в очереди —

$$T_{\text{оч.}} = \frac{\rho^2}{\lambda(1-\rho)}. \quad \textbf{(15.44)}$$

▷ **15.8.** В порту имеется один причал для разгрузки судов. Интенсивность потока судов равна 0,4 (судов в сутки). Среднее время разгрузки одного судна составляет 2 суток. Предполагается, что очередь может быть неограниченной длины. Найти показатели эффективности работы причала, а также вероятность того, что ожидают разгрузки не более чем 2 судна.

Решение. Имеем $\rho = \lambda/\mu = \lambda \bar{t}_{\text{об.}} = 0{,}4 \cdot 2 = 0{,}8$. Так как $\rho = 0{,}8 < 1$, то очередь на разгрузку не может бесконечно возрастать и предельные вероятности существуют. Найдем их.

Вероятность того, что причал свободен, по (15.33) $p_0 = 1 - 0{,}8 = 0{,}2$, а вероятность того, что он занят, $P_{\text{зан.}} = 1 - 0{,}2 = 0{,}8$. По формуле (15.34) вероятности того, что у причала находятся 1, 2, 3 судна (т.е. ожидают разгрузки 0, 1, 2 судна), равны $p_1 = 0{,}8(1-0{,}8) = 0{,}16$; $p_2 = 0{,}8^2 \cdot (1-0{,}8) = 0{,}128$; $p_3 = 0{,}8^3 \cdot (1-0{,}8) = 0{,}1024$.

Вероятность того, что ожидают разгрузку не более чем 2 судна, равна

$$P = p_1 + p_2 + p_3 = 0{,}16 + 0{,}128 + 0{,}1024 = 0{,}3904.$$

По формуле (15.40) среднее число судов, ожидающих разгрузки,

$$L_{\text{оч.}} = 0{,}8^2/(1-0{,}8) = 3{,}2,$$

а среднее время ожидания разгрузки по формуле (15.42)

$$T_{\text{оч.}} = 3{,}2/0{,}8 = 4 \text{ (сутки)}.$$

По формуле (15.36) среднее число судов, находящихся у причала, $L_{сист.} = 0{,}8/(1-0{,}8) = 4$ (сутки) (или проще по (15.37) $L_{сист.} = 3{,}2+0{,}8 = 4$ (сутки), а среднее время пребывания судна у причала по формуле (15.41) $T_{сист.} = 4/0{,}8 = 5$ (сутки).

Очевидно, что эффективность разгрузки судов невысокая. Для ее повышения необходимо уменьшение среднего времени разгрузки судна $\bar{t}_{об.}$ либо увеличение числа причалов n.▶

Многоканальная СМО с неограниченной очередью. Рассмотрим задачу. Имеется n-канальная СМО с неограниченной очередью. Поток заявок, поступающих в СМО, имеет интенсивность λ, а поток обслуживаний — интенсивность μ. Необходимо найти предельные вероятности состояний СМО и показатели ее эффективности.

Система может находиться в одном из состояний S_0, S_1, S_2, ..., S_k, ..., S_n, ..., нумеруемых по числу заявок, находящихся в СМО: S_0 — в системе нет заявок (все каналы свободны); S_1 — занят один канал, остальные свободны; S_2 — заняты два канала, остальные свободны; ..., S_k — занято k каналов, остальные свободны; ..., S_n — заняты все n каналов (очереди нет); S_{n+1} — заняты все n каналов, в очереди одна заявка; ..., S_{n+r} — заняты все n каналов, r заявок стоит в очереди, ...

Граф состояний системы показан на рис. 15.9. Обратим внимание на то, что в отличие от предыдущей СМО, интенсивность потока обслуживаний (переводящего систему из одного состояния в другое справа налево) не остается постоянной, а по мере увеличения числа заявок в СМО от 0 до n увеличивается от величины μ до $n\mu$, так как соответственно увеличивается число каналов обслуживания. При числе заявок в СМО большем, чем n, интенсивность потока обслуживаний сохраняется равной $n\mu$.

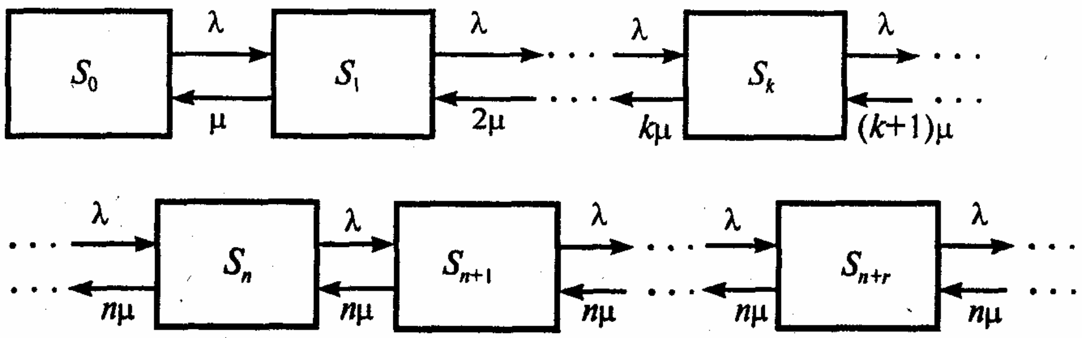


Рис. 15.9

Можно показать, что при $\rho/n < 1$ предельные вероятности существуют. Если $\rho/n \geq 1$, очередь растет до бесконечности. Используя формулы (15.16) и (15.17) для процесса гибели и размножения, можно получить следующие формулы для предельных вероятностей состояний n-канальной СМО с неограниченной очередью

$$p_0 = \left(1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \ldots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n!(n-\rho)}\right)^{-1}, \tag{15.45}$$

$$p_1 = \frac{\rho}{1!} p_0, \ldots, p_k = \frac{\rho^k}{k!} p_0, \ldots, p_n = \frac{\rho^n}{n!} \rho_0, \tag{15.46}$$

$$p_{n+1} = \frac{\rho^{n+1}}{n \cdot n!} p_0, \ldots, p_{n+r} = \frac{\rho^{n+r}}{n^r \cdot n!} p_0, \ldots \ . \tag{15.47}$$

Вероятность того, что заявка окажется в очереди,

$$P_{\text{оч.}} = \frac{\rho^{n+1}}{n!(n-\rho)} p_0 \,. \tag{15.48}$$

Для n-канальной СМО с неограниченной очередью, используя прежние приемы, можно найти:

среднее число занятых каналов

$$\bar{k} = \frac{\lambda}{\mu} = \rho, \tag{15.49}$$

среднее число заявок в очереди

$$L_{\text{оч.}} = \frac{\rho^{n+1} p_0}{n \cdot n! \left(1 - \frac{\rho}{n}\right)^2}, \tag{15.50}$$

среднее число заявок в системе

$$L_{\text{сист.}} = L_{\text{оч.}} + \rho. \tag{15.51}$$

Среднее время пребывания заявки в очереди и среднее время пребывания заявки в системе, как и ранее, находятся по формулам Литтла (15.42) и (15.41).

Замечание. Для СМО с неограниченной очередью при $\rho < 1$ любая заявка, пришедшая в систему, будет обслужена, т.е. вероятность отказа $P_{\text{отк.}} = 0$, относительная пропускная способность $Q = 1$, а абсолютная пропускная способность равна интенсивности входящего потока заявок, т.е. $A = \lambda$.

▷ **15.9.** В универсаме к узлу расчета поступает поток покупателей с интенсивностью $\lambda = 81$ чел. в час. Средняя продолжительность обслуживания контролером-кассиром одного покупателя $\bar{t}_{\text{об.}} = 2$ мин. Определить:

а. Минимальное количество контролеров-кассиров $n_{\min}$, при котором очередь не будет расти до бесконечности, и соответствующие характеристики обслуживания при $n=n_{\min}$.

б. Оптимальное количество $n_{\text{опт.}}$ контролеров-кассиров, при котором относительная величина затрат $C_{\text{отн.}}$, связанная с издержками на содержание каналов обслуживания и с пребыванием в очереди покупателей, задаваемая, например, как $C_{\text{отн.}} = \frac{1}{\lambda}n + 3T_{\text{оч.}}$, будет минимальна, и сравнить характеристики обслуживания при $n=n_{\min}$ и $n=n_{\text{опт.}}$.

в. Вероятность того, что в очереди будет не более трех покупателей.

Решение. **а.** По условию $\lambda = 81(1/\text{ч}) = 81/60 = 1{,}35$ (1/мин.). По формуле (15.24) $\rho = \lambda/\mu = \lambda\bar{t}_{\text{об.}} = 1{,}35 \cdot 2 = 2{,}7$. Очередь не будет возрастать до бесконечности при условии $\rho/n < 1$, т.е. при $n > \rho = 2{,}7$. Таким образом, минимальное количество контролеров-кассиров $n_{\min} = 3$.

Найдем характеристики обслуживания СМО при $n = 3$.

Вероятность того, что в узле расчета отсутствуют покупатели, по формуле (15.45) $p_0 = (1+2{,}7+2{,}7^2/2!+2{,}7^3/3!+2{,}7^4/3!(3-2{,}7))^{-1} = 0{,}025$, т.е. в среднем 2,5% времени контролеры-кассиры будут простаивать.

Вероятность того, что в узле расчета будет очередь, по (15.48)

$$P_{\text{оч.}} = (2{,}7^4/3!(3-2{,}7))0{,}025 = 0{,}735.$$

Среднее число покупателей, находящихся в очереди, по (15.50)

$$L_{\text{оч.}} = (2{,}7^4/3 \cdot 3!(1-2{,}7/3)^2)0{,}025 = 7{,}35.$$

Среднее время ожидания в очереди по (15.42)

$$T_{\text{оч.}} = 7{,}35/1{,}35 = 5{,}44 \text{ (мин)}.$$

Среднее число покупателей в узле расчета по (15.51)

$$L_{сист.} = 7{,}35+2{,}7 = 10{,}05.$$

Среднее время нахождения покупателей в узле расчета по (15.41)

$$T_{сист.} = 10{,}05/1{,}35 = 7{,}44 \text{ (мин)}.$$

Среднее число контролеров-кассиров, занятых обслуживанием покупателей, по (15.49) $\bar{k} = 2{,}7$.

Коэффициент (доля) занятых обслуживанием контролеров-кассиров

$$\bar{k} = \rho/n = 2{,}7/3 = 0{,}9.$$

Абсолютная пропускная способность узла расчета $A = 1{,}35$ (1/мин), или 81 (1/ч), т.е. 81 покупатель в час.

Анализ характеристик обслуживания свидетельствует о значительной перегрузке узла расчета при наличии трех контролеров-кассиров.

б. Относительная величина затрат при $n = 3$

$$C_{отн.} = \frac{1}{\lambda}n + 3T_{оч.} = 3/1{,}35+3\cdot5{,}44 = 18{,}54.$$

Рассчитаем относительную величину затрат при других значениях n (табл. 15.2).

Таблица 15.2

Характеристика обслуживания	Число контролеров-кассиров				
	3	4	5	6	7
Вероятность простоя контролеров-кассиров p_0	0,025	0,057	0,065	0,067	0,067
Среднее число покупателей в очереди $T_{оч.}$	5,44	0,60	0,15	0,03	0,01
Относительная величина затрат $C_{отн.}$	18,54	4,77	4,14	4,53	5,22

Как видно из табл. 15.2, минимальные затраты получены при $n = n_{опт.} = 5$ контролерах-кассирах.

Определим характеристики обслуживания узла расчета при $n = n_{\text{опт.}} = 5$. Получим $P_{\text{оч.}} = 0{,}091$; $L_{\text{оч.}} = 0{,}198$; $T_{\text{оч.}} = 0{,}146$ (мин); $L_{\text{сист.}} = 2{,}90$; $T_{\text{сист.}} = 2{,}15$ (мин); $\bar{k} = 2{,}7$; $k_3 = 0{,}54$.

Как видим, при $n = 5$ по сравнению с $n = 3$ существенно уменьшились вероятность возникновения очереди $P_{\text{оч.}}$, длина очереди $L_{\text{оч.}}$ и среднее время пребывания в очереди $T_{\text{оч.}}$ и соответственно среднее число покупателей $L_{\text{сист.}}$ и среднее время нахождения в узле расчета $T_{\text{сист.}}$, а также доля занятых обслуживанием контролеров k_3. Но среднее число занятых обслуживанием контролеров-кассиров $\bar{k}$ и абсолютная пропускная способность узла расчета A естественно не изменились.

в. Вероятность того, что в очереди будет не более 3 покупателей, определится как

$$P(r \le 3) = \underbrace{p_1+p_2+p_3+p_4+p_5}_{\substack{\text{(когда заняты от 1}\\ \text{до 5 контролеров-кассиров)}}} + \underbrace{p_{5+1}+p_{5+2}+p_{5+3}}_{\substack{\text{(когда в очереди}\\ \text{стоят от 1 до 3 покупателей)}}} =$$

$= 1-P_{\text{оч.}}+p_{5+1}+p_{5+2}+p_{5+3}$, где каждое слагаемое найдем по формулам (15.45)—(15.48). Получим при $n = 5$:

$$P(r \le 3) = 1-\frac{2{,}7^6}{5!(5-2{,}3)}0{,}065+\frac{2{,}7^6}{5\cdot 5!}0{,}065+$$

$$+\frac{2{,}7^7}{5^2\cdot 5!}0{,}065+\frac{2{,}7^8}{5^3\cdot 5!}0{,}065 = 0{,}986.$$

(Заметим, что в случае n=3 контролеров-кассиров та же вероятность существенно меньше: $P(r \le 3)$=0,464).▶

▷ **15.10.** Железнодорожная касса с двумя окошками продает билеты в два пункта A и B. Интенсивность потока пассажиров, желающих купить билеты, для обоих пунктов одинакова: $\lambda_A = \lambda_B = 0{,}45$ (пассажиров в минуту). На обслуживание пассажиров кассир тратит в среднем 2 мин. Рассматриваются два варианта продажи билетов: первый — билеты продаются в одной кассе с двумя окошками одновременно в оба пункта A и B; второй — билеты продаются в двух специализированных кассах (по одному окошку в каждой), одна только в пункт A, другая — только в пункт B. Необходимо:

а. Сравнить два варианта продажи билетов по основным характеристикам обслуживания.

б. Определить, как надо изменить среднее время обслуживания одного пассажира, чтобы по второму варианту продажи пассажиры затрачивали на приобретение билетов в среднем меньше времени, чем по первому варианту.

Р е ш е н и е. **а.** По **первому варианту** имеем двухканальную СМО, на которую поступает поток заявок интенсивностью $\lambda = 0{,}45+0{,}45 = 0{,}9$; интенсивность потока обслуживаний $\mu = 1/2 = 0{,}5$; $\rho = \lambda/\mu = 1{,}8$. Так как $\rho/n = 1{,}8/2 = 0{,}9 < 1$, то предельные вероятности существуют.

Вероятность простоя двух кассиров по (15.45)

$$p_0 = \left(1 + \frac{1{,}8}{1!} + \frac{1{,}8^2}{2!} + \frac{1{,}8^3}{2!(2-1{,}8)}\right)^{-1} = 0{,}0526.$$

Среднее число пассажиров в очереди по (15.50)

$$L_{\text{оч.}} = 1{,}8^3/2 \cdot 2!(1-1{,}8/2) \cdot 0{,}0526 = 7{,}67.$$

Среднее число пассажиров у кассы по (15.51)

$$L_{\text{сист.}} = 7{,}67+1{,}8 = 9{,}47.$$

Среднее время на ожидание в очереди и покупку билетов равно соответственно (по формулам (15.42) и (15.41)): $T_{\text{оч.}} = 7{,}67/0{,}9 = 8{,}52$ (мин) и $T_{\text{сист.}} = 9{,}47/0{,}9 = 10{,}5$ (мин).

По **второму варианту** имеем две одноканальные СМО (два специализированных окошка); на каждую поступает поток заявок с интенсивностью $\lambda = 0{,}45$. По-прежнему $\mu = 0{,}5$; $\rho = \lambda/\mu = 0{,}9 < 1$, предельные вероятности существуют. По формулам (15.40), (15.36), (15.42), (15.41)

$$L_{\text{оч.}} = 0{,}9^2/(1-0{,}9) = 8{,}1; \; L_{\text{сист.}} = 0{,}9/(1-0{,}9) = 9{,}0;$$

$$T_{\text{оч.}} = 8{,}1/0{,}45 = 18{,}0 \text{ (мин)}, \; T_{\text{сист.}} = 9{,}0/0{,}45 = 20{,}0 \text{ (мин)}.$$

Итак, по второму варианту увеличились и длина очереди, и среднее время ожидания в ней и в целом на покупку билетов. Такое различие объясняется тем, что в первом варианте (двухканальная СМО) меньше средняя доля времени, которую простаивает каждый из двух кассиров: если он не занят обслуживанием пассажира, покупающего билет в пункт *A*, он может заняться обслуживанием пассажира, покупающего билет в пункт *B*, и наоборот. Во втором варианте такой взаимозаменяемости нет.

Можно заметить, что среднее время на покупку билетов по второму варианту увеличилось более чем в 2 раза. Такое значительное увеличение связано с тем, что СМО работает на пределе своих возможностей ($\rho = 0{,}9$): достаточно незначительно увеличить среднее время обслуживания $\bar{t}_{об.}$, т.е. уменьшить μ, и ρ превзойдет 1, т.е. очередь начнет неограниченно возрастать.

б. Выше было получено, что по первому варианту продажи билетов при среднем времени обслуживания одного пассажира $\bar{t}_{об.} = 2$ (мин) среднее время на покупку билетов составит $T_{сист.1} = 10{,}5$ (мин). По условию для второго варианта продажи

$T_{сист.2} < T_{сист.1}$ или с учетом (15.36) и (15.41): $\frac{1}{\lambda}\frac{\rho}{1-\rho} < T_{сист.1}$.

Полагая $\rho = \lambda/\mu = \lambda \bar{t}_{об.}$, получим $\frac{\bar{t}_{об.}}{1 - \lambda\bar{t}_{об.}} < T_{сист.1}$, откуда

найдем $\bar{t}_{об.} < \frac{T_{сист.1}}{1 + \lambda T_{сист.1}}$ или $\bar{t}_{об} < \frac{10{,}5}{1 + 0{,}45 \cdot 10{,}5} = 1{,}83$ (мин).

Итак, средние затраты времени на покупку билетов по второму варианту продажи уменьшатся, если среднее время обслуживания одного пассажира уменьшится более чем на 0,17 мин, или более чем на 8,5%.▶

СМО с ограниченной очередью. СМО с ограниченной очередью отличаются от рассмотренных выше задач лишь тем, что число заявок в очереди ограничено (не может превосходить некоторого заданного m). Если новая заявка поступает в момент, когда все места в очереди заняты, она покидает СМО необслуженной, т.е. получает отказ.

Очевидно: для вычисления предельных вероятностей состояний и показателей эффективности таких СМО может быть использован тот же подход, что и выше, с той разницей, что суммировать надо не бесконечную прогрессию (как, например, мы делали при выводе формулы (15.33)), а конечную. Соответствующие формулы сведем в табл. 15.3.

Среднее время пребывания заявки в очереди и в системе, как и ранее, определяем по формулам Литтла (15.44) и (15.43).

▷ **15.11. По условию задачи 15.8 найти показатели эффективности работы причала. Известно, что приходящее судно покидает причал (без разгрузки), если в очереди на разгрузку стоит более 3 судов.**

Р е ш е н и е. По условию $m = 3$. Используем формулы, приведенные во второй графе табл. 15.3.

Вероятность того, что причал свободен:

$$p_0 = \frac{1-0{,}6}{1-0{,}8^{3+2}} = 0{,}297.$$

Вероятность того, что приходящее судно покинет причал без разгрузки:

$$P_{\text{отк.}}=0{,}8^{3+1}\cdot 0{,}297=0{,}122.$$

Относительная пропускная способность причала:

$$Q = 1-0{,}122 = 0{,}878.$$

Абсолютная пропускная способность причала $A = 0{,}4 \cdot 0{,}878 = 0{,}351$, т.е. в среднем в сутки разгружается 0,35 судна.

Среднее число судов, ожидающих разгрузку

$$L_{\text{оч.}} = \frac{0{,}8^2\left[1-0{,}8^3(3+1-3\cdot 0{,}8)\right]}{(1-0{,}8^{3+2})(1-0{,}8)} = 0{,}861,$$

а среднее время ожидания разгрузки по (15.42)

$$T_{\text{оч.}} = \frac{0{,}861}{0{,}8} = 1{,}076 \text{ (сутки)}.$$

Среднее число судов, находящихся у причала

$$L_{\text{сист.}} = 0{,}861+(1-0{,}297) = 1{,}564,$$

а среднее время пребывания судна у причала по (15.41):

$$T_{\text{сист.}} = \frac{1{,}564}{0{,}8} = 1{,}955 \text{ (сутки)}.\blacktriangleright$$

СМО с ограниченным временем ожидания. На практике часто встречаются СМО с так называемыми “нетерпеливыми” заявками. Такие заявки могут уйти из очереди, если время ожидания превышает некоторую величину. В частности, такого рода заявки возникают в различных технологических системах, в которых задержка с началом обслуживания может привести к потере качества продукции, в системах оперативного управления, когда срочные сообщения теряют ценность (или даже смысл), если они не поступают на обслуживание в течение определенного времени.

Таблица 15.3

Показатели	Одноканальная СМО с ограниченной очередью	Многоканальная СМО с ограниченной очередью
Предельные вероятности	$p_0 = \dfrac{1-\rho}{1-\rho^{m+2}}$ $p_1=\rho p_0,\ p_2=\rho^2 p_0,\ ...,\ p_k=\rho^k p_0$	$p_0=\left[1+\dfrac{\rho}{1!}+\ldots+\dfrac{\rho^n}{n!}+\ldots+\dfrac{\rho^{n+1}(1-(\rho/n)^m)}{n\cdot n!(1-\rho/n)}\right]^{-1}$ $p_1=\dfrac{\rho}{1!}p_0,\ldots,p_k=\dfrac{\rho^k}{k!}p_0,\ldots,p_n=\dfrac{\rho^n}{n!}p_0,$ $p_{n+1}=\dfrac{\rho^{n+1}}{n\cdot n!}p_0,\ldots,p_{n+r}=\dfrac{\rho^{n+r}}{n^r\cdot n!}p_0\ (r=1,\ ...,\ m)$
Вероятность отказа	$P_{\text{отк.}}=P_{m+1}=\rho^{m+1}p_0$	$P_{\text{отк.}}=P_{n+m}=\dfrac{\rho^{n+m}}{n^m\cdot n!}p_0$
Абсолютная пропускная способность	$A=\lambda Q=\lambda(1-\rho^{m+1}p_0)$	$A=\lambda Q=\lambda(1-\dfrac{\rho^{n+m}}{n^m\cdot n!}p_0)$
Относительная пропускная способность	$Q=1-P_{\text{отк.}}=1-\rho^{m+1}p_0$	$Q=1-P_{\text{отк.}}=1-\dfrac{\rho^{n+m}}{n^m\cdot n!}p_0$
Среднее число заявок в очереди	$L_{\text{оч.}}=\rho^2\dfrac{[1-\rho^m(m+1-m\rho)]}{(1-\rho^{m+2})(1-\rho)}$	$L_{\text{оч.}}=\dfrac{\rho^{n+1}p_0\left[1-\left(m+1-m\dfrac{\rho}{n}\right)\left(\dfrac{\rho}{n}\right)^m\right]}{n\cdot n!\left(1-\dfrac{\rho}{n}\right)^2}$
Среднее число заявок под обслуживанием (среднее число занятых каналов)	$L_{\text{об.}}=1-p_0$	$\bar{k}=\rho\left(1-\dfrac{\rho^{n+m}}{n^m\cdot n!}p_0\right)$
Среднее число заявок в системе	$L_{\text{сист.}}=L_{\text{оч.}}+L_{\text{об.}}$	$L_{\text{сист.}}=L_{\text{оч.}}+\bar{k}$

В простейших математических моделях таких систем предполагается, что заявка может находиться в очереди случайное время, распределенное по показательному закону с некоторым параметром ν, т.е. можно условно считать, что каждая заявка, стоящая в очереди на обслуживание, может покинуть систему с интенсивностью ν.

Соответствующие показатели эффективности СМО с ограниченным временем получаются на базе результатов, полученных для процесса гибели и размножения.

В заключение отметим, что на практике часто встречаются *замкнутые* системы обслуживания, у которых входящий поток заявок существенным образом зависит от состояния самой СМО. В качестве примера можно привести ситуацию, когда на ремонтную базу поступают с мест эксплуатации некоторые машины: понятно, что чем больше машин находится в состоянии ремонта, тем меньше их продолжает эксплуатироваться и тем меньше интенсивность потока вновь поступающих на ремонт машин. Для замкнутых СМО характерным является ограниченное число источников заявок, причем каждый источник "блокируется" на время обслуживания его заявки (т.е. он не выдает новых заявок). В подобных системах при конечном числе состояний СМО предельные вероятности будут существовать при любых значениях интенсивностей потоков заявок и обслуживаний. Они могут быть вычислены, если вновь обратиться к процессу гибели и размножения.

15.8. Понятие о статистическом моделировании СМО (методе Монте-Карло)

Основное допущение, при котором анализировались рассмотренные выше СМО, состоит в том, что все потоки событий, переводящие их из состояния в состояние, были простейшими. При нарушении этого требования общих аналитических методов для таких систем не существует. Имеются лишь отдельные результаты, позволяющие выразить в аналитическом виде характеристики СМО через параметры задачи.

В случаях, когда для анализа работы СМО аналитические методы не применимы (или же требуется проверить их точность), используют универсальный *метод статистического моделирования*, или, как его называют, *метод Монте-Карло*.

Идея метода Монте-Карло состоит в том, что вместо аналитического описания СМО производится "розыгрыш" случайного процесса, проходящего в СМО, с помощью специально организованной процедуры. В результате такого "розыгрыша" получается каждый раз новая, отличная от других реализация случайного процесса. Это множество реализаций можно использовать как некий искусственно полученный статистический материал, который обрабатывается обычными методами математической статистики. После такой обработки могут быть получены приближенно любые характеристики обслуживания.

Например, необходимо проанализировать очереди, возникающие в магазине, для решения вопроса о расширении магазина. Время подхода покупателей и время их обслуживания носят случайный характер, и их распределения могут быть установлены по имеющейся информации. В результате взаимодействия этих случайных процессов создается очередь.

Согласно методу Монте-Карло перебирают (с помощью ЭВМ) все возможные состояния системы с различным числом покупателей в час, временем их обслуживания и т.п., сохраняя те же характеристики распределения. В результате многократного искусственного воссоздания работы магазина рассчитывают характеристики обслуживания, как если бы они были получены при наблюдении над реальным потоком покупателей.

При моделировании случайных явлений методом Монте-Карло мы пользуемся самой случайностью как аппаратом исследования. Заметим, что для сложных систем обслуживания с немарковским случайным процессом метод статистического моделирования, как правило, оказывается проще аналитического.

УПРАЖНЕНИЯ

15.12. Построить граф состояний следующего случайного процесса: система состоит из двух автоматов по продаже газированной воды, каждый из которых в случайный момент времени может быть либо занятым, либо свободным.

15.13. Построить граф состояний системы S, представляющей электрическую лампочку, которая в случайный момент времени может быть либо включена, либо выключена, либо выведена из строя.

15.14. Найти предельные вероятности для систем S, граф которых изображен на рис. 15.10 и 15.11.

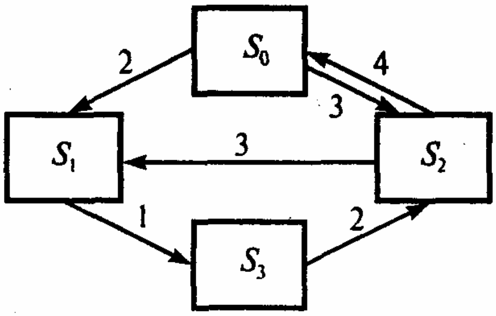


Рис. 15.10

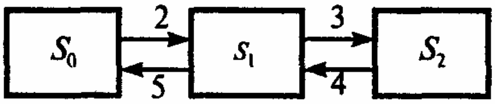


Рис. 15.11

15.15. Рассматривается круглосуточная работа пункта проведения профилактического осмотра автомашин с одним каналом (одной группой проведения осмотра). На осмотр и выявление дефектов каждой машины затрачивается в среднем 0,5 ч. На осмотр поступает в среднем 36 машин в сутки. Потоки заявок и обслуживаний — простейшие. Если машина, прибывшая в пункт осмотра, не застает ни одного канала свободным, она покидает пункт осмотра необслуженной. Определить вероятности состояний и характеристики обслуживания профилактического пункта осмотра.

15.16. Решить задачу **15.15** для случая $n = 4$ канала (групп проведения осмотра). Найти число каналов, при котором относительная пропускная способность пункта осмотра будет не менее 0,9.

15.17. Анализируется работа междугородного переговорного пункта в небольшом городке. Пункт имеет один телефонный аппарат для переговоров. В среднем за сутки поступает 240 заявок на переговоры. Средняя длительность переговоров (с учетом вызова абонентов в другом городе) составляет 5 мин. Никаких ограничений на длину очереди нет. Потоки заявок и обслуживаний простейшие. Определить предельные вероятности состояний и характеристики обслуживания переговорного пункта в стационарном режиме.

15.18. Решить задачу **15.17** для случая $n = 3$ телефонных аппаратов.

15.19—15.20. Решить задачи **15.15**, **15.16** при условии, что машина, прибывшая на пункт осмотра, покидает этот пункт лишь в случае, если в очереди на осмотр стоят более 5 машин.

15.21—15.22. Решить задачи **15.17**, **15.18** при условии, что длина очереди не должна превышать 60 чел.

Глава 16. МОДЕЛИ УПРАВЛЕНИЯ ЗАПАСАМИ

16.1. Основные понятия

Задачи управления запасами составляют один из наиболее многочисленных классов экономических задач исследования операций, решение которых имеет важное народнохозяйственное значение. Правильное и своевременное определение оптимальной стратегии управления запасами, а также нормативного уровня запасов позволяет высвободить значительные оборотные средства, замороженные в виде запасов, что в конечном счете повышает эффективность используемых ресурсов.

Рассмотрим основные характеристики моделей управления запасами.

Спрос. Спрос на запасаемый продукт может быть *детерминированным* (в простейшем случае — постоянным во времени) или *случайным*. Случайность спроса описывается либо случайным моментом спроса, либо случайным объемом спроса в детерминированные или случайные моменты времени.

Пополнение склада. Пополнение склада может осуществляться либо периодически через определенные интервалы времени, либо по мере исчерпания запасов, т. е. снижения их до некоторого уровня.

Объем заказа. При периодическом пополнении и случайном исчерпании запасов объем заказа может зависеть от того состояния, которое наблюдается в момент подачи заказа. Заказ обычно подается на одну и ту же величину при достижении запасом заданного уровня — так называемой *точки заказа.*

Время доставки. В идеализированных моделях управления запасами предполагается, что заказанное пополнение доставляется на склад мгновенно. В других моделях рассматривается задержка поставок на фиксированный или случайный интервал времени.

Стоимость поставки. Как правило, предполагается, что стоимость каждой поставки слагается из двух компонент — разовых затрат, не зависящих от объема заказываемой партии, и затрат, зависящих (чаще всего — линейно) от объема партии.

Издержки хранения. В большинстве моделей управления запасами считают объем склада практически неограниченным, а в качестве контролирующей величины служит объем хранимых запасов. При этом полагают, что за хранение каждой единицы запаса в единицу времени взимается определенная плата.

Штраф за дефицит. Любой склад создается для того, чтобы предотвратить дефицит определенного типа изделий в обслуживаемой системе. Отсутствие запаса в нужный момент приводит к убыткам, связанным с простоем оборудования, неритмичностью производства и т. п. Эти убытки в дальнейшем будем называть *штрафом за дефицит*.

Номенклатура запаса. В простейших случаях предполагается, что на складе хранится запас однотипных изделий или однородного продукта. В более сложных случаях рассматривается *многономенклатурный запас*.

Структура складской системы. Наиболее полно разработаны математические модели одиночного склада. Однако на практике встречаются и более сложные структуры: иерархические системы складов с различными периодами пополнения и временем доставки заказов, с возможностью обмена запасами между складами одного уровня иерархии и т. п.

В качестве критерия эффективности принятой стратегии управления запасами выступает *функция затрат (издержек)*, представляющая суммарные затраты на хранение и поставку запасаемого продукта (в том числе потери от порчи продукта при хранении и его морального старения, потери прибыли от омертвления капитала и т. п.) и затраты на штрафы.

Управление запасами состоит в отыскании такой стратегии пополнения и расхода запасами, при котором функция затрат принимает минимальное значение.

Ниже рассматриваются простейшие модели управления запасами.

Пусть функции $A(t)$, $B(t)$ и $R(t)$ выражают соответственно пополнение запасов, их расход и спрос на запасаемый продукт за промежуток времени $[0, t]$. В моделях управления запасами обычно используются производные этих функций по времени $a(t)$, $b(t)$, $r(t)$, называемые соответственно *интенсивностями пополнения, расхода и спроса.*

Если функции $a(t)$, $b(t)$, $r(t)$ — не случайные величины, то модель управления запасами считается *детерминированной*, если хотя бы одна из них носит случайный характер — *стохастической*. Если все параметры модели не меняются во времени, она называется *статической*, в противном случае — *динамической*. Статические модели используются, когда принимается разовое решение об уровне запасов на определенный период, а динамические — в случае принятия последовательных решений об уровнях запаса или корректировке ранее принятых решении с учетом происходящих изменений.

Уровень запаса в момент t определяется основным уравнением запасов

$$J(t) = J_0 + A(t) - B(t), \qquad (16.1)$$

где J_0 — начальный запас в момент $t = 0$.

Уравнение (16.1) чаще используется в интегральной форме:

$$J(t) = J_0 + \int_0^t a(t)dt - \int_0^t b(t)dt. \qquad (16.2)$$

▷ **16.1**. Интенсивность поступления деталей на склад готовой продукции цеха составляет в начале смены 5 дет./мин, в течение первого часа линейно возрастает, достигая к концу его 10 дет./мин, и затем остается постоянной. Полагая, что поступление деталей на склад происходит непрерывно в течение всех семи часов смены, а вывоз деталей со склада производится только в конце работы, записать выражение для уровня запаса в произвольный момент времени и, используя его, найти количество деталей на складе: а) через 30 мин после начала работы; б) в конце смены.

Решение. По условию в течение смены не происходит выдачи деталей со склада, т. е. $b(t) = 0$. Интенсивность пополнения запаса в течение первого часа линейно возрастает, т. е. $a(t) = kt + b$. Учитывая, что $a(0) = 5$, получаем $b = 5$. Так как в конце часа, т. е. при $t = 60$ $a(60) = 10$, то $10 = k \cdot 60 + 5$, откуда $k = 1/12$. Таким образом, для первого часа смены $a(t) = (1/12)t + 5$, а затем $a(t) = 10$.

Учитывая продолжительность смены (7 ч = 420 мин) и соотношение (16.2), получаем:

$$J(t)=\int_0^t (t/12+5)dt = t^2/24+5t\,,$$

если $0 \le t \le 60$, и

$$J(t)=\int_0^{60}(t/12+5)dt+\int_{60}^{t}10dt=\left(t^2/24+5t\right)\Big|_0^{60}+10t\Big|_{60}^{t}=$$

$$=450+10t-600=10t-150,$$

если $60 \le t \le 420$.

Количество деталей на складе через 30 мин после начала работы: $J(30)$ = 900/24 + 5·30 = 187,5, а в конце смены: $J(420)$ = = 10·420 − 150 = 4050.▶

16.2. Статическая детерминированная модель без дефицита

Предположение о том, что дефицит не допускается, означает полное удовлетворение спроса на запасаемый продукт, т.е. совпадение функций $r(t)$ и $b(t)$. Пусть общее потребление запасаемого продукта за рассматриваемый интервал времени θ равно N. Рассмотрим простейшую модель, в которой предполагается, что расходование запаса происходит непрерывно с постоянной интенсивностью, т.е. $b(t) = b$. Эту интенсивность можно найти, разделив общее потребление продукта на время, в течение которого он расходуется:

$$b=\frac{N}{\theta}. \qquad (16.3)$$

Пополнение заказа происходит партиями одинакового объема, т.е. функция $a(t)$ не является непрерывной: $a(t) = 0$ при всех t, кроме моментов поставки продукта, когда $a(t) = n$, где n — объем партии. Так как интенсивность расхода равна b, то вся партия будет использована за время

$$T=\frac{n}{b}.. \qquad (16.4)$$

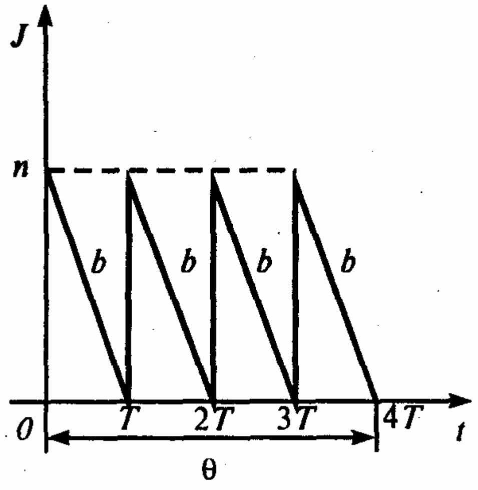


Рис. 16.1

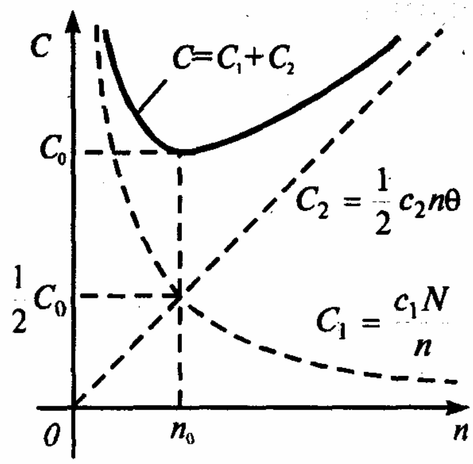


Рис. 16.2

Если отсчет времени начать с момента поступления первой партии, то уровень запаса в начальный момент равен объему этой партии n, т.е. $J(0) = n$. Графически уровень запаса в зависимости от времени представлен на рис 16.1.

На временно́м интервале $[0, T]$ уровень запаса уменьшается по прямой $J(t) = n - bt$ от значения n до нуля. Так как дефицит не допускается, то в момент T уровень запаса мгновенно пополняется до прежнего значения n за счет поступления партии заказа. И так процесс изменения $J(t)$ повторяется на каждом временном интервале продолжительностью T (*см.* рис. 16.1).

Задача управления запасами состоит в определении такого объема партии n, при котором суммарные затраты на создание и хранение запаса были бы минимальными.

Обозначим суммарные затраты через C, затраты на создание запаса — через C_1, затраты на хранение запаса — через C_2 и найдем эти величины за весь промежуток времени T.

Пусть затраты на доставку одной партии продукта, не зависимые от объема партии, равны c_1, а затраты на хранение одной единицы продукта в единицу времени — c_2. Так как за время θ необходимо запастись N единицами продукта, который доставляется партиями объема n, то число таких партий k равно:

$$k = \frac{N}{n} = \frac{\theta}{T}. \qquad (16.5)$$

Отсюда получаем

$$C_1 = c_1 k = c_1 \frac{N}{n}. \qquad \textbf{(16.6)}$$

Мгновенные затраты хранения запаса в момент времени t равны $c_2 J(t)$. Значит, за промежуток времени $[0, T]$ они составят

$$c_2 \int_0^T J(t)\,dt = c_2 \int_0^T (n - bt)\,dt$$

или, учитывая (16.4):

$$c_2 \int_0^T J(t)\,dt = c_2 \int_0^T \left(n - \frac{n}{T}t\right) dt = c_2 \left(nt - \frac{nt^2}{2T}\right)\Bigg|_0^T = \frac{c_2 nT}{2}.$$

Средний запас за промежуток $[0, T]$ равен $nT/2$, т.е. *затраты на хранение всего запаса при линейном (по времени) его расходе равны затратам на хранение среднего запаса.*

Учитывая периодичность функции $J(t)$ (всего за промежуток времени θ будет $k = \frac{N}{n}$ "зубцов", аналогичных рассмотренному на отрезке $[0, T]$), и формулу (16.5), получаем, что затраты хранения запаса за промежуток времени θ равны:

$$C_2 = \frac{c_2 nT}{2} k = \frac{c_2 nT}{2} \cdot \frac{N}{n} = \frac{c_2 TN}{2} = \frac{c_2 \theta n}{2}. \qquad \textbf{(16.7)}$$

Нетрудно заметить, что затраты C_1 обратно пропорциональны, а затраты C_2 прямо пропорциональны объему партии n. Графики функций $C_1(n)$ и $C_2(n)$, а также функции суммарных затрат

$$C = \frac{c_1 N}{n} + \frac{c_2 \theta}{2} n \qquad \textbf{(16.8)}$$

приведены на рис. 16.2.

В точке минимума функции $C(n)$ ее производная $C'(n) = -\frac{c_1 N}{n^2} + \frac{c_2 \theta}{2} = 0$, откуда

$$n = n_0 = \sqrt{\frac{2c_1 N}{c_2 \theta}} \qquad \textbf{(16.9)}$$

или, учитывая (16.3):

$$n_0 = \sqrt{\frac{2c_1 b}{c_2}}. \tag{16.10}$$

Формула (16.10), называемая *формулой Уилсона* или *формулой наиболее экономичного объема партии*, широко используется в экономике. Эта формула может быть получена и другим способом, если учесть, что произведение $C_1C_2 = 0{,}5c_1c_2N\theta$ есть величина постоянная, не зависящая от n. В этом случае, как известно, сумма двух величин принимает наименьшее значение, когда они равны, т. е. $C_1 = C_2$ или

$$\frac{c_1 N}{n} = \frac{c_2 n\theta}{2}, \tag{16.11}$$

откуда получаем (16.9).

Из (16.11) следует, что *минимум общих затрат задачи управления запасами достигается тогда, когда затраты на создание запаса равны затратам на хранение запаса.* При этом минимальные суммарные затраты

$$C_0 = C(n_0) = \frac{2c_1 N}{n}, \tag{16.12}$$

откуда, учитывая (16.9) и (16.3), получим $C_0 = \sqrt{2c_1c_2\theta N}$

или

$$C_0 = \theta\sqrt{2c_1c_2 b}\ . \tag{16.13}$$

Число оптимальных партий за время θ с учетом (16.5), (16.9) и (16.3) равно:

$$k_0 = \frac{N}{n_0} = \sqrt{\frac{c_2 N\theta}{2c_1}} = \theta\sqrt{\frac{c_2 b}{2c_1}}.$$

Время расхода оптимальной партии на основании (16.4) с учетом (16.9) и (16.3) равно

$$T_0 = \frac{n_0}{b} = n_0\frac{\theta}{N} \tag{16.14}$$

или

$$T_0 = \sqrt{\frac{2c_1\theta}{c_2 N}} = \sqrt{\frac{2c_1}{c_2 b}}. \tag{16.15}$$

▷ **16.2**. Потребность сборочного предприятия в деталях некоторого типа составляет 120 000 деталей в год, причем эти детали расходуются в процессе производства равномерно и непрерывно. Детали заказываются раз в год и поставляются партиями одинакового объема, указанного в заказе. Хранение детали на складе стоит 0,35 ден. ед. в сутки, а поставка партии — 10 000 ден. ед. Задержка производства из-за отсутствия деталей недопустима. Определить наиболее экономичный объем партии и интервал между поставками, которые нужно указать в заказе (предполагается, что поставщик не допускает задержки поставок).

Решение. По условию затраты на одну партию составляют $c_1 = 10\ 000$ ден. ед., затраты хранения единицы запаса в сутки $c_1 = 0{,}35$ ден. ед. Общий промежуток времени $\theta = 1$ год $= 365$ дней, а общий объем запаса за этот период $N = 120\ 000$ деталей. По формуле (16.9) $n_0 = \sqrt{\dfrac{2 \cdot 10\ 000 \cdot 120\ 000}{0{,}35 \cdot 365}} \approx 4335$ дет., а по (16.14)

$T_0 = n_0 \dfrac{\theta}{N} = 13{,}2 \approx 13$ дней.

Итак, наиболее экономичный объем партии равен 4335 деталей, а интервал между поставками ≈ 13 дней.▶

На практике, естественно, объем партии может отличаться от оптимального n_0, вычисленного по (16.9). Так, в предыдущей задаче может оказаться удобным заказывать партии по 4 500 или даже по 5 000 деталей и возникает вопрос, как при этом изменятся суммарные затраты.

Для ответа на этот вопрос разложим функцию $C(n)$ в ряд Тейлора в окрестности точки n_0, ограничившись первыми тремя членами ряда при достаточно малых изменениях объема партии Δn:

$$C(n) = C(n_0) + C'(n_0)\Delta n + \frac{c''(n_0)}{2!}\Delta n^2 + \ldots .$$

Учитывая, что при $n = n_0$ $C'(n_0) = 0$, $C''(n_0) = \dfrac{2c_1 N}{n_0^3}$, а $C_0 = C(n_0)$ определяется по формуле (16.12), найдем:

$$\frac{\Delta C}{C_0} = \frac{C(n) - C(n_0)}{C(n_0)} \approx \frac{C''(n_0)\Delta n^2}{2C(n_0)} = \frac{2c_1 N \Delta n^2}{n_0^3 (2c_1 N / n_0)}$$

или

$$\frac{\Delta C}{C_0} \approx \frac{1}{2}\left(\frac{\Delta n}{n_0}\right)^2. \qquad \textbf{(16.16)}$$

Формула (16.16) свидетельствует об определенной устойчивости суммарных затрат по отношению к наиболее экономичному объему партии, ибо при малых Δn относительное изменение затрат примерно на порядок меньше относительного изменения объема партии по сравнению с оптимальным.

▷ **16.3**. По условию задачи **16.2** определить, на сколько процентов увеличатся затраты на создание и хранение запаса по сравнению с минимальными затратами при объеме заказываемых партий 5 000 деталей.

Решение. Относительное изменение объема партии по сравнению с оптимальным n_0 = 4335 составляет $\Delta n/n_0$ = (5000 − − 4335)/4335=0,153. В соответствии с (16.16) относительное изменение суммарных затрат составит $\Delta C/C_0 = 0{,}153^2/2 \approx 0{,}012$, или лишь 1,2%.▶

▷ **16.4**. В условиях задачи **16.3** предположим, что заказываются не все партии сразу, а каждая отдельно, причем срок выполнения заказа равен 16 дней. Определить точки заказа, т. е. при каком уровне запаса следует заказывать следующую партию.

Решение. Так как по результатам решения задачи **16.2** длина интервала между поставками равна 13,2 дней, то заказ в условиях налаженного производства следует возобновить, когда уровень запаса достаточен для удовлетворения потребности на 16 − 13,2 = = 2,8 дня. Так как ежедневная потребность (интенсивность расхода запаса) равна по формуле (16.3) b = 120 000/365 = 329 деталей, то заказы должны делаться регулярно при достижении уровня запаса $329 \cdot 2{,}8 \approx 922$ деталей.▶

16.3. Статическая детерминированная модель с дефицитом

В рассматриваемой модели будем полагать наличие *дефицита*. Это означает, что при отсутствии запасаемого продукта, т.е. при $J(t) = 0$ спрос сохраняется с той же интенсивностью $r(t) = b$, но потребление запаса отсутствует — $b(t) = 0$, вследствие чего накапливается дефицит со скоростью b. График изменения уровня запаса в этом случае представлен на рис. 16.3. Убывание графика ниже оси абсцисс в область отрицательных значений в отличие от графика на рис. 16.2 характеризует накопление дефицита.

Из рис. 16.3 видно, что каждый период "пилы" $T = \frac{n}{b}$ разбивается на два временных интервала, т. е. $T = T_1 + T_2$, где T_1 — время, в течение которого производится потребление запаса, T_2 — время, когда запас отсутствует и накапливается дефицит, который будет перекрыт в момент поступления следующей партии.

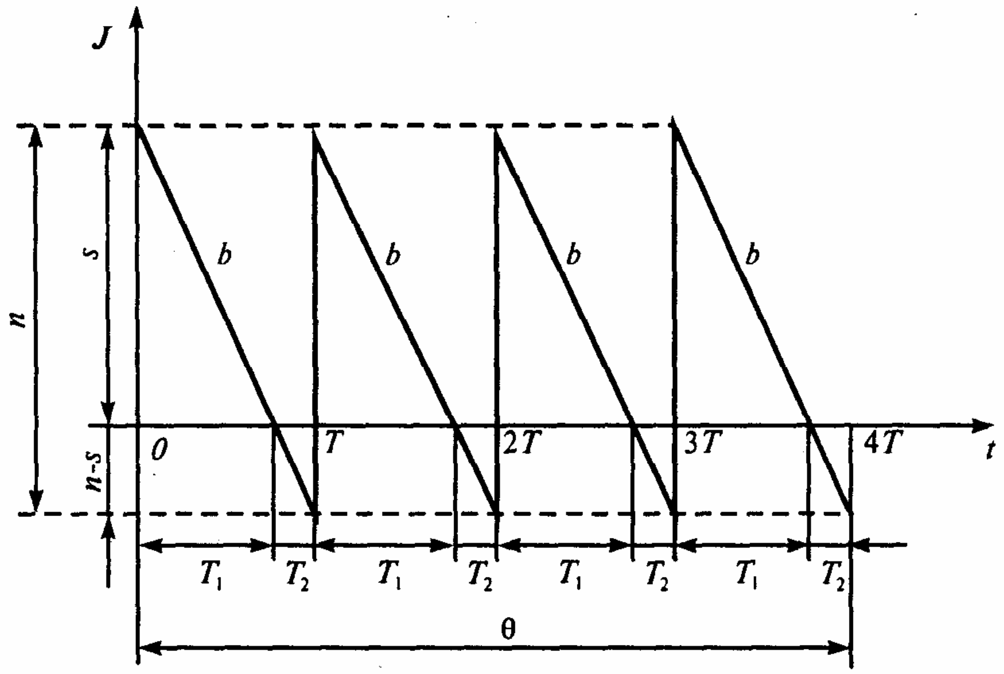


Рис. 16.3

Необходимость покрытия дефицита приводит к тому, что максимальный уровень запаса s в момент поступления каждой партии теперь не равен ее объему n, а меньше его на величину дефицита $n - s$, накопившегося за время T_2 (*см.* рис. 16.3).

Из геометрических соображений легко установить, что

$$T_1 = \frac{s}{n}T, \quad T_2 = \frac{n-s}{n}T. \qquad \textbf{(16.17)}$$

В данной модели в функцию суммарных затрат C наряду с затратами C_1 (на пополнение запаса) и C_2 (на хранение запаса) необходимо ввести затраты C_3 — на штраф из-за дефицита, т.е.

$$C = C_1 + C_2 + C_3.$$

Затраты C_1, как и ранее, находим по формуле (16.11). В разд. 16.2 было показано, что затраты C_2 при линейном расходе запаса равны затратам на хранение среднего запаса, который за время потребления T_1 равен $sT_1/2$; поэтому с учетом (16.7) и (16.5) эти затраты составят

$$C_2 = \frac{c_2 s T_1}{2} k = \frac{c_2 s \cdot sT}{2} \cdot \frac{\theta}{T} = \frac{c_2 s^2 \theta}{2n}. \qquad \textbf{(16.18)}$$

При расчете затрат C_3 будем считать, что штраф за дефицит составляет в единицу времени c_3 на каждую единицу продукта. Так как средний уровень дефицита за период T_2 равен $(n - s)\, T_2/2$, то штраф за этот период T_2 составит $\frac{1}{2} c_3 (n-s) T_2$, а за весь период θ с учетом (16.7) и (16.19) —

$$C_3 = \frac{1}{2} c_3 (n-s) T_2 k = \frac{1}{2} c_3 (n-s) \frac{n-s}{n} T \frac{\theta}{T} = \frac{c_3 \theta (n-s)^2}{2n}. \qquad \textbf{(16.19)}$$

Теперь, учитывая (16.12), (16.18) и (16.19), суммарные затраты равны

$$C = c_1 \frac{N}{n} + \frac{c_2 \theta s^2}{2n} + \frac{c_3 \theta (n-s)^2}{2n}. \qquad \textbf{(16.20)}$$

Нетрудно заметить, что при $n = s$ формула (16.19) совпадает с ранее полученной (16.8) в модели без дефицита.

Рассматриваемая *задача управления запасами сводится к отысканию такого объема партии n и максимального уровня запаса s, при которых функция C* (16.19) *принимает минимальное значение.* Другими словами, необходимо исследовать функцию двух переменных $C(n, s)$ на экстремум. Приравнивая частные производные $\partial C/\partial n$, $\partial C/\partial s$ к нулю, получим после преобразований систему уравнений:

$$\begin{cases} n^2 c_3 - (c_2 + c_3) s^2 = 2 c_1 N/\theta, \\ s = n \dfrac{c_3}{c_2 + c_3}. \end{cases} \qquad \textbf{(16.21)}$$

Решая систему, получаем формулы наиболее экономичного объема партии $\tilde{n}_0$ и максимального уровня запаса $\tilde{s}_0$ для модели с дефицитом[1]**:**

$$\tilde{n}_0 = \sqrt{\frac{2c_1N}{c_2\theta}}\sqrt{\frac{c_2+c_3}{c_3}} = \sqrt{\frac{2c_1b}{c_2}}\sqrt{\frac{c_2+c_3}{c_3}}, \qquad \textbf{(16.22)}$$

$$\tilde{s}_0 = \sqrt{\frac{2c_1N}{c_2\theta}}\sqrt{\frac{c_3}{c_2+c_3}} = \tilde{n}_0\frac{c_3}{c_2+c_3}. \qquad \textbf{(16.23)}$$

Величина

$$\rho = \frac{c_3}{c_2+c_3} \qquad \textbf{(16.24)}$$

называется *плотностью убытков из-за неудовлетворенного спроса* и играет важную роль в управлении запасами. Заметим, что $0 \le \rho \le 1$. Если значение c_3 мало по сравнению с c_2, то величина ρ близка к нулю: когда c_3 значительно превосходит c_2, то ρ близка к 1. Недопустимость дефицита равносильна предположению, что $c_3 = \infty$ или $\rho = 1$.

Используя (16.24), основные формулы (16.22) и (16.23) можно записать компактнее:

$$\tilde{n}_0 = \sqrt{\frac{2c_1b}{c_2\rho}}, \qquad \textbf{(16.25)}$$

$$\tilde{s}_0 = \tilde{n}_0\rho. \qquad \textbf{(16.26)}$$

Следует учесть, что в силу (16.17) и (16.26) $T_1/T = \tilde{s}_0/\tilde{n}_0 = \rho$ и $T_2/T = (\tilde{n}_0 - \tilde{s}_0)/\tilde{n}_0 = 1-\rho$. Поэтому утверждение о том, что плотность убытков из-за неудовлетворенного спроса равна ρ, означает, что в течение $(1-\rho)100\%$ времени от полного периода T запас продукта будет отсутствовать.

[1] С помощью достаточного условия экстремума можно убедиться в том, что действительно при $n = n_0$, $s = s_0$ функция $C(n, s)$ достигает минимума.

Из сравнения формул (16.25) и (16.10) следует, что оптимальные объемы партий для задач с дефицитом и без дефицита при одинаковых параметрах связаны соотношением

$$\tilde{n}_0 = \frac{n_0}{\sqrt{\rho}}, \qquad (16.27)$$

откуда вытекает, что *оптимальный объем партии в задаче с дефицитом всегда больше (в $1/\sqrt{\rho}$ раз), чем в задаче без дефицита.*

▷ **16.5**. Найти наиболее экономичный объем партии и интервал между поставками, сохраняя условия задачи **16.2**, кроме недопустимости дефицита, если известно, что отсутствие на сборке каждой детали приносит в сутки убытки в размере 3,5 ден. ед.

Решение. По условию $c_3 = 3{,}5$. Ранее было получено по формуле (16.9) $n_0 = 4335$ и по (16.15) $T_0 = 13{,}2$. Найдем плотность убытков из-за неудовлетворенного спроса по формуле (16.24): $\rho = 3{,}5/(0{,}35 + 3{,}5) = 0{,}909$, т.е. $100(1-0{,}909) = 9{,}1\%$ времени между поставками детали на сборке будут отсутствовать.

Теперь оптимальный размер партии по формуле (16.27) $\tilde{n}_0 = 4335/\sqrt{0{,}909} = 4547$. В силу (16.15) пропорционально увеличению $\tilde{n}_0$ должен увеличиться интервал между поставками, т.е. $\tilde{T}_0 = T_0/\sqrt{\rho} = 13{,}2/\sqrt{0{,}909} = 13{,}8 \approx 14$ дней.▶

16.4. Стохастические модели управления запасами

Рассмотрим *стохастические модели управления запасами*, у которых спрос является *случайным*. Этот факт существенным образом сказывается на характере соответствующих моделей и значительно усложняет их анализ, в связи с чем в рамках данной книги ограничимся рассмотрением наиболее простых моделей.

Предположим, что спрос r за интервал времени T является случайным и задан его закон (ряд) распределения $p(r)$ или плотность вероятностей $\varphi(r)$ (обычно функции $p(r)$ и $\varphi(r)$ оцениваются на основании опытных или статистических данных). Если спрос r ниже уровня запаса s, то приобретение (хранение, продажа) излишка продукта требует дополнительных затрат c_2 на единицу продукта; наоборот, если спрос r выше уровня запаса s, то это приводит к штрафу за дефицит c_3 на единицу продукции.

В качестве функции суммарных затрат, являющейся в стохастических моделях случайной величиной, рассматривают ее среднее значение или математическое ожидание.

В рассматриваемой модели при дискретном случайном спросе r, имеющем закон распределения $p(r)$, математическое ожидание суммарных затрат имеет вид[1]:

$$C(s) = c_2 \sum_{r=0}^{s} (s-r)p(r) + c_3 \sum_{r=s+1}^{\infty} (r-s)p(r). \qquad \textbf{(16.28)}$$

В выражении (16.28) первое слагаемое учитывает затраты на приобретение (хранение) излишка $s - r$ единиц продукта (при $r \le s$), а второе слагаемое — штраф за дефицит на $r - s$ единиц продукта (при $r > s$).

В случае непрерывного случайного спроса, задаваемого плотностью вероятностей $\varphi(r)$, выражение $C(s)$ принимает вид:

$$C(s) = c_2 \int_0^s (s-r)\varphi(r)dr + c_3 \int_0^s (r-s)\varphi(r)dr. \qquad \textbf{(16.29)}$$

Задача управления запасами состоит в отыскании такого запаса s, при котором математическое ожидание суммарных затрат (16.28) или (16.29) принимает минимальное значение.

Доказано, например в [17] или [24], что при дискретном случайном спросе r выражение (16.28) минимально при запасе s_0, удовлетворяющем неравенствам

$$F(s_0) < \rho < F(s_0 + 1), \qquad \textbf{(16.30)}$$

а при непрерывном случайном спросе r выражение (16.29) минимально при значении s_0, определяемом из уравнения

$$F(s_0) = \rho, \qquad \textbf{(16.31)}$$

где

$$F(s) = p(r < s) \qquad \textbf{(16.32)}$$

есть функция распределения спроса r, $F(s_0)$ и $F(s_0 + 1)$ — ее значения; ρ — плотность убытков из-за неудовлетворенного спроса, определяемая по (16.24).

---

[1] Учитываем только расходы на неиспользованные единицы продукта.

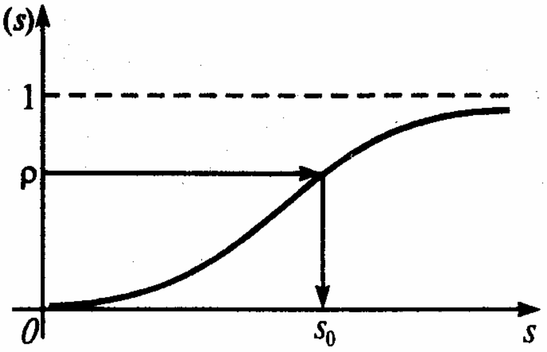


Рис. 16.4

Оптимальный запас s_0 при непрерывном спросе по данному значению ρ может быть найден и графически (рис. 16.4).

▷ **16.6**. Предприятие закупает агрегат с запасными блоками к нему. Стоимость одного блока равна 5 ден. ед. В случае выхода агрегата из строя из-за поломки блока, отсутствующего в запасе, простой агрегата и срочный заказ нового блока к нему обойдется в 100 ден. ед. Опытное распределение агрегатов по числу блоков, потребовавших замену, представлено в табл. 16.1.

Таблица 16.1

Число замененных блоков r	0	1	2	3	4	5	6
Статистическая вероятность (доля) агрегатов $p(r)$, которым потребовалась замена r блоков	0,90	0,05	0,02	0,01	0,01	0,01	0,00

Необходимо определить оптимальное число запасных блоков, которое следует приобрести вместе с агрегатом.

Решение. По условию $c_2 = 5$, $c_3 = 100$. Вычислим плотность убытков из-за нехватки запасных блоков по формуле (16.24) $\rho =$ $= 100/(5+100) = 0{,}952$.

Учитывая (16.32), найдем значения функции распределения спроса (табл. 16.2).

Таблица 16.2

s	0	1	2	3	4	5	6	>6
r	0	1	2	3	4	5	6	>6
$F(s)$	0,00	0,00	0,90	0,95	0,97	0,98	0,99	1,00

Очевидно (*см.* табл. 16.2), что оптимальный запас составит $s_0 = 3$, ибо он удовлетворяет неравенству (16.30): $F(3) < 0{,}952 < F(4)$. ▶

▷ **16.7.** Решить задачу **16.6** при условии непрерывного случайного спроса r, распределенного по показательному закону с функцией распределения $F(r) = 1 - e^{-\lambda r}$ при $\lambda = 0{,}98$.

Решение. Оптимальное число запасных блоков s_0 найдем из уравнения (16.31): $1 - e^{-\lambda s_0} = \rho$, откуда $e^{-\lambda s_0} = 1 - \rho$ и $s_0 = -\dfrac{1}{\lambda}\ln(1-\rho)$. При $\lambda = 0{,}98$ $s_0 = -(1/0{,}98)\ln 0{,}02 \approx 4$ (блока).▶

В условиях рассматриваемой модели предположим, что расходование запаса происходит *непрерывно* с одинаковой интенсивностью. Такую ситуацию можно представить графически (рис. 16.5).

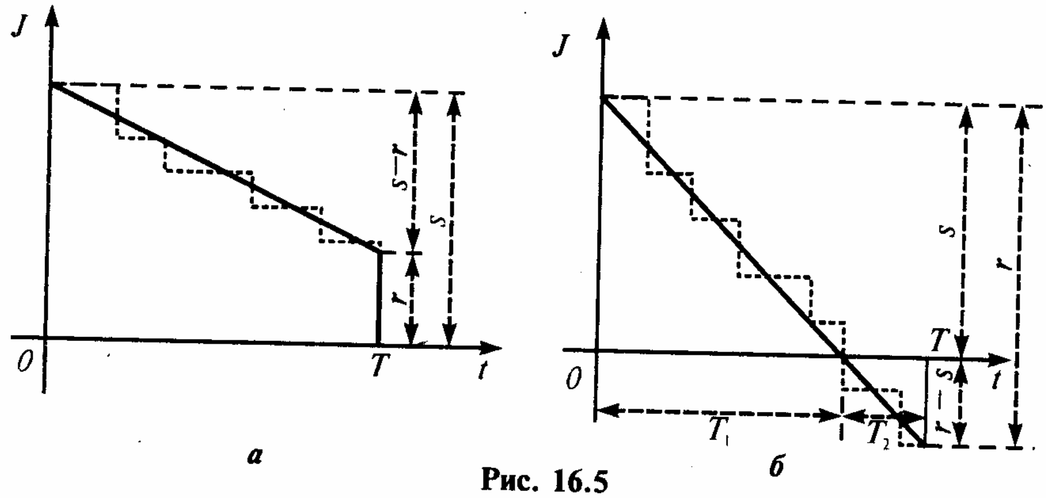


Рис. 16.5

Рис. 16.5, *а* соответствует случаю $r \le s$, когда спрос не превосходит запаса, а рис. 16.5, *б* — случаю, когда спрос превышает запас, т.е. $r > s$. Следует отметить, что на самом деле график $J(t)$ представляет ступенчатую ломаную, показанную на рис. 16.5 пунктиром, но для исследования модели нам проще рассматривать $J(t)$ в виде прямой, сглаживающей эту ломаную.

Средний запас, соответствующий рис. 16.5, *а*, равен

$$\bar{s}_1 = \frac{1}{2}(s + (s - r)) = s - \frac{1}{2}r. \qquad \textbf{(16.33)}$$

Средний запас, соответствующий рис. 16.6, *б* с учетом формулы (16.17), в которой полагаем $n = r$, составляет

$$\bar{s}_2 = \frac{1}{2} s \frac{T_1}{T} = \frac{1}{2} \frac{s^2}{r}. \qquad \mathbf{(16.34)}$$

Средний дефицит продукта за период T_2 для случая, соответствующего рис. 16.5, *б* с учетом (16.17), где $n = r$, равен

$$\bar{s}_3 = \frac{1}{2}(r - s)\frac{T_2}{T_1} = \frac{1}{2}\frac{(r - s)^2}{r}. \qquad \mathbf{(16.35)}$$

Математическое ожидание суммарных затрат составит:

$$C(s) = c_2 \sum_{r=0}^{s} \left(s - \frac{r}{2}\right) p(r) + c_2 \sum_{r=s+1}^{\infty} \frac{1}{2} \frac{s^2}{r} p(r) + c_3 \sum_{r=0}^{\infty} \frac{1}{2} \frac{(r - s)^2}{r} p(r).$$

(16.36)

Доказано [17, 24] что в этом случае математическое ожидание (16.36) минимально при запасе s_0, удовлетворяющем неравенству

$$L(s_0) < \rho < L(s_0 + 1), \qquad \mathbf{(16.37)}$$

где ρ по-прежнему определяется по формуле (16.24):

$$L(s) = F(s) + \left(s - \frac{1}{2}\right) \sum_{r=s}^{\infty} \frac{p(r)}{r}, \qquad \mathbf{(16.38)}$$

$L(s_0)$ и $L(s_0 + 1)$ — значения функции (16.38), а $F(s)$ находится в соответствии с определением (16.32).

▷ **16.8.** Имеющиеся на складе изделия равномерно расходуются в течение месяца. Затраты на хранение одного изделия составляют 5 ден. ед., а штраф за дефицит одного изделия обходится в 100 ден. ед. Изучение спроса дало распределение числа потребляемых за месяц изделий, представленное в табл. 16.3.

Таблица 16.3

Спрос r	0	1	2	3	4	5	≥6
Статистическая вероятность $p(r)$	0,1	0,2	0,2	0,3	0,1	0,1	0,0

Необходимо, определить оптимальный месячный запас склада.

Решение. Так же как в задаче **16.6**, $c_2 = 5$, $c_3 = 100$, $\rho = 0{,}952$. Значения функции $L(r)$ определим с помощью табл. 16.4.

Таблица 16.4

s	r	$p(r)$	$\frac{p(r)}{r}$	$\sum_{r=s}^{\infty}\frac{p(r)}{r}$	$\left(s-\frac{1}{2}\right)\sum_{r=s}^{\infty}\frac{p(r)}{r}$	$F(r)$	$L(r)$
0	0	0,1	—	—	—	0,0	—
1	1	0,2	0,200	0,445	0,2225	0,1	0,3225
2	2	0,2	0,100	0,245	0,3675	0,3	0,6675
3	3	0,3	0,100	0,145	0,3625	0,5	0,8625
4	4	0,1	0,025	0,045	0,1575	0,8	0,9575
5	5	0,1	0,020	0,020	0,0900	0,9	0,9900
≥ 6	≥ 6	0,0	0,000	0,000	0,0000	1,0	1,0000

Очевидно, что оптимальный запас изделий $s_0 = 3$, ибо он удовлетворяет условию (16.37): $L(3) < 0{,}952 < L(4)$.▶

16.5. Стохастические модели управления запасами с фиксированным временем задержки поставок

В рассмотренных выше идеализированных моделях управления запасами предполагалось, что пополнение запаса происходит практически мгновенно. Однако в ряде задач *время задержки* поставок может оказаться настолько значительным, что его необходимо учитывать в модели.

Пусть за время задержек поставок θ уже заказаны n партий по одной в каждый из n периодов продолжительностью $T = \theta/n$. Обозначим:

$s_{нз}$ — первоначальный уровень запаса (к началу первого периода);

s_i — запас за i-й период;

r_i — спрос за i-й период;

q_i — пополнение запаса за i-й период.

q_i — пополнение запаса за i-й период.

Тогда к концу n-го периода на склад поступит $\sum_{i=1}^{n} q_i$ единиц продукта, а будет израсходовано $\sum_{i=1}^{n} r_i$ единиц, т.е.

$$s_n = s_{\text{н з}} + \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} r_i, \tag{16.39}$$

или

$$s_n = s - r, \tag{16.40}$$

где

$$s = s_{\text{н з}} + \sum_{i=1}^{n} q_i, \tag{16.41}$$

$$r = \sum_{i=1}^{n} r_i . \tag{16.42}$$

Требуется найти оптимальный объем партии заказа, который необходимо сделать за последний n-й период, предшествующий поступлению сделанного ранее заказа.

Математическое ожидание суммарных затрат в этом случае определяется по формуле (16.28), а оптимальный запас s находится по формуле (16.30), т.е.

$$F(s_0) < \rho < F(s_0 + 1). \tag{16.43}$$

Найдя оптимальный запас s_0 и зная $q_1, q_2, \ldots, q_{n-1}$, можно вычислить q_n по формуле (16.41), т.е.

$$q_n = s_0 - \left(s_{\text{н з}} + \sum_{i=1}^{n-1} q_i \right). \tag{16.44}$$

▷ **16.9. Ежедневно заказываемый скоропортящийся товар поступает в магазин спустя 7 дней после заказа. В момент очередного заказа запас товара составил в стоимостном выражении 10 ден. ед. В** данное в день изготовления, приносит прибыль 0,95 ден. ед., а не проданный в этот день товар может быть затем реализован с убытком 0,10 ден. ед.

На основании опытных данных получено следующее распределение спроса на данный товар (табл. 16.5).

Таблица 16.5

r	0	10	20	30	40	50	60	70	80	90	100
$p(r)$	0,00	0,00	0,01	0,02	0,05	0,08	0,11	0,12	0,14	0,13	0,10
r	110	120	130	140	150	160	170	180	190	200	>200
$p(r)$	0,08	0,05	0,03	0,02	0,02	0,01	0,01	0,01	0,01	0,00	0,00

Необходимо определить оптимальный объем заказанного товара q_7 на седьмой день после заказа.

Решение. Плотность убытков из-за дефицита товара по формуле (16.24) равна $\rho = 0{,}95/(0{,}10 + 0{,}95) = 0{,}905$. Учитывая условия (16.32), найдем значения функции распределения спроса (табл. 16.6).

Таблица 16.6

s	r	$F(s)$	s	r	$F(s)$	s	r	$F(s)$	s	r	$F(s)$
0	0	0,00	50	50	0,16	100	100	0,86	150	150	0,96
10	10	0,00	60	60	0,27	110	110	0,84	160	160	0,97
20	20	0,01	70	70	0,39	120	120	0,89	170	170	0,98
30	30	0,03	80	80	0,53	130	130	0,92	180	180	0,99
40	40	0,08	90	90	0,76	140	140	0,94	≥ 190	≥ 190	1,00

Условию (16.43) удовлетворяет $s_0 = 120$, ибо $F(120) < 0{,}905 < F(130)$.

Таким образом, оптимальный запас товара за 7 дней должен быть на сумму 120 ден. ед., откуда оптимальный объем заказанного товара на седьмой день по (16.44) составит: $q_7 = 120 - (10 + (10 + 20 + 10 + 10 + 20 + 10)) = 30$ ден. ед.▶

В заключении главы отметим, что найти аналитически оптимальные значения точки запаса s_0 и объема партии n удается только в относительно простых случаях. Если же система хранения запасов имеет сложную структуру (много видов хранимой продукции, иерархическая система складов), используемые стохастические модели сложны, а их параметры меняются во времени, то единственным средством анализа такой системы становится *имитационное моделирование*, позволяющее имитировать ("проигрывать") на ЭВМ функционирование системы, исследуя ее поведение при различных условиях, значениях параметров, отражая их случайный характер, изменение во времени и т. п.

УПРАЖНЕНИЯ

16.10. В условиях задачи **16.1** интенсивность поступления деталей на склад в течение первых 50 мин растет по закону $a(t) = 0{,}2t + 5$, а затем до конца смены остается постоянной. Найти количество деталей на складе: а) через 10 мин после начала работы; б) в конце работы.

16.11. По условию задачи **16.3** найти изменение затрат на создание и хранение запаса при изменении объема партии на 10%.

16.12. Ежедневный спрос на некоторый продукт составляет 100 ед. Затраты на приобретение каждой партии этого продукта, не зависимые от объема партии, равны 100 ден. ед., а затраты на хранение единицы продукта — 0,02 ден. ед. в сутки. Определить наиболее экономичный объем партии и интервал между поставками партий такого объема.

16.13. Срок выполнения заказа на партию продукта в условиях задачи **16.9** равен 12 дням. При каком уровне запаса следует заказывать очередную партию продукта?

16.14. По условию задачи **16.5** определить максимальный уровень запаса и интервал между поставками.

16.15. Решить задачу **16.12** в предположении, что возможен дефицит, который приносит 0,03 ден. ед. убытка в день на единицу продукта.

16.16. Кондитерское предприятие торгует вразвес своими тортами. Каждый килограмм торта приносит 2 ден. ед. прибыли. Все торты можно продать на следующий день со скидкой 0,2 ден. ед. На основании опыта получено распределение спроса на торты, представленное в табл. 16.3. Найти оптимальную дневную выработку тортов.

16.17. Решить задачу **16.16** при условии, что спрос на торты есть случайная величина, распределенная по показательному закону с параметром $\lambda = 0{,}9$.

16.18. Склад пополняется каждый месяц некоторыми изделиями. В течение первых пяти месяцев года объемы пополнения равны соответственно 10, 20, 20, 20 и 30 изделиям. Начальный запас к началу первого месяца равен 10 изделиям. На основании опыта получено распределение спроса на товар, представленное в табл. 16.5. Сдвиг по времени между заказом на пополнение и доставкой на склад равен 6 мес. Издержки в расчете на одно изделие из-за излишка изделий равны 10 ден. ед, а от их нехватки — 120 ден. ед. Найти оптимальное пополнение склада на шестой месяц.

ЛИТЕРАТУРА

1. *Акулич И.Л.* Математическое программирование в примерах и задачах. — М.: Высшая школа, 1986.

2. *Банди Б.* Основы линейного программирования. — М.: Радио и связь, 1989.

3. *Вентцель Е.С.* Исследование операций. — М.: Сов. радио, 1972.

4. *Вентцель Е.С.* Исследование операций. Задачи, принципы, методология. — М.: Наука, 1980.

5. Высшая математика для экономистов/Под ред. Н.Ш. Кремера. — М.: Банки и биржи, ЮНИТИ, 1997.

6. *Горелик В.А., Ушаков И.А.* Исследование операций. — М.: Машиностроение, 1986.

7. *Горчаков А.А., Орлова И.В.* Компьютерные экономико-математические модели. — М.: Компьютер, ЮНИТИ, 1995.

8. *Зайченко Ю.П.* Исследование операций. — Киев: Выща школа, 1986.

9. Исследование операций. — В 2-х т./Под ред. Дж. Моудера и С. Элмаграби. Пер. с англ. — М.: Мир, 1981.

10. Исследование операций/Под ред. М.А. Войтенко и Н.Ш. Кремера. — М.: Экономическое образование, 1992.

11. *Калихман И.Л.* Линейная алгебра и программирование. — М.: Высшая школа, 1967.

12. *Калихман И.Л., Войтенко М.А.* Динамическое программирование. — М.: Высшая школа, 1979.

13. *Калихман И.Л.* Сборник задач по математическому программированию. — М.: Высшая школа, 1975.

14. *Карасев А.И., Аксютина З.М., Савельева Т.И.* Курс высшей математики для экономических вузов. — Ч.2. — М.: Высшая школа, 1982.

15. *Карасев А.И., Кремер Н.Ш., Савельева Т.И.* Математические методы и модели в планировании. — М.: Экономика, 1987.

16. *Карпелевич Ф.И., Садовский Л.Е.* Элементы линейной алгебры и линейного программирования. — М.: Наука, 1967.

17. *Кофман А.* Методы и модели исследования операций/Пер. с франц. — М.: Мир, 1966.

18. *Крушевский А.В., Швецов К.И.* Математическое программирование и моделирование в экономике. — Киев: Выща школа, 1979.

19. *Кузнецов А.В., Сакович В.А., Холод Н.И.* Высшая математика. Математическое программирование. — Минск: Вышэйшая школа, 1994.

20. *Кузнецов Ю.Н., Кузубов В.И., Волощенко А.Б.* Математическое программирование. — М.: Высшая школа, 1986.

21. Математическое программирование /Под ред. Н.Ш. Кремера. — М.: Финстатинформ, 1995.

22. *Нит И.В.* Линейное программирование. — М.: Изд-во МГУ, 1978.

23. Руководство к решению задач с экономическим содержанием по курсу высшей математики/Под ред. А.И. Карасева и Н.Ш. Кремера. — М.: ВЗФЭИ, 1989.

24. *Черчмен У., Акоф Р., Арноф Л.* Введение в исследование операций/Пер. с англ. — М.: Наука, 1968.

Содержание

Учебное пособие

Наум Шевелевич Кремер,
Борис Александрович Путко,
Иван Михайлович Тришин,
Мира Нисоновна Фридман

ИССЛЕДОВАНИЕ ОПЕРАЦИЙ В ЭКОНОМИКЕ

Редактор ***Л.А. Артемова***
Корректоры *Л.И. Ганина, К.В. Федорова*
Оформление художника *А.В. Лебедева*

Компьютерный набор и оригинал-макет
выполнены АО "ВИОЛАНТА"

Лицензия № 071252 от 04.01.96
Подписано в печать 01.07.97. Формат 60x88 1/16
Усл. печ. л. 25,5. Доп. тираж 5000 экз. Заказ 1729

Издательское объединение “ЮНИТИ”
Генеральный директор *В.Н. Закаидзе*

123298, Москва, Тепличный пер., 6
Тел. (095) 194-00-15. Тел/факс (095) 194-00-14
E.mail: unity@tech.ru

Отпечатано в ГУП ИПК “Ульяновский Дом печати”
432601, г. Ульяновск, ул. Гончарова, 14