ШЕСТОЕ
ИЗДАНИЕ

Введение в ИССЛЕДОВАНИЕ ОПЕРАЦИЙ

ХЭМДИ А. ТАХА
Университет Арканзаса, Фейетвилл

Издательский дом "Вильямс"
Москва ♦ Санкт-Петербург ♦ Киев
2001

ББК 32.973.26-018.2.75
Т24
УДК 681.3.07

Издательский дом “Вильямс”

Зав. редакцией *С.Н. Тригуб*

Перевод с английского *канд.физ.-мат.наук В.И. Тюпти (главы 9 – 21), канд физ.-мат.наук А.А. Минько (все остальное)*

По общим вопросам обращайтесь в Издательский дом “Вильямс” по адресу: info@williamspublishing.com, http://www.williamspublishing.com

Таха, Хэмди, А.

Т24 Введение в исследование операций. 6-е издание. : Пер. с англ. — М. : Издательский дом “Вильямс”, 2001. — 912 с. : ил. — Парал. тит. англ.

ISBN 5-8459-0180-4 (рус.)

Исследование операций, как научная дисциплина и практические методы, ориентировано на решение практических задач, которые можно корректно описать с помощью той или иной математической модели с целью получения оптимального решения. Данная книга может служить учебным пособием по теории и практическому применению методов исследования операций. Каждая тема начинается с вводного материала, доступного студентам начальных курсов, далее уровень изложения постепенно повышается и рассчитан уже на студентов старших курсов. В конце каждой главы приводится набор комплексных задач, связанных излагаемой с темой, которые значительно углубляют и расширяют ее.

Написанная без излишнего академизма, но достаточно строго, книга будет интересна широкому кругу читателей: студентам, аспирантам и преподавателям высших учебных заведений, экономистам, инженерам, разработчикам программного обеспечения и др.

ББК 32.973.26-018.2.75

ISBN 5-8459-0180-4 (рус.)
ISBN 0-13-272915-6 (англ.)

Оглавление

Содержание

Предисловие

Поскольку первое издание этой книги вышло в далеком 1971 году[1], я вынужден был внести многочисленные изменения как в стиль изложения, так и в содержание шестого издания данной книги. Я пришел к выводу, что внесение отдельных изменений и исправлений может привести лишь к непреднамеренным искажениям и ошибкам. Я посчитал необходимым добавить новые упражнения и изменить многие "старые" упражнения. Таким образом, я пришел к заключению, что в этом издании надо существенно изменить как основной материал, так и упражнения.

В данной книге первые 18 глав переписаны полностью. Оставшиеся три главы пересмотрены и исправлены. Добавлено много нового материала, старый текст сокращен или вовсе удален. В этом издании существенно изменен уровень излагаемого материала, в частности о линейном программировании. Теперь каждая тема начинается с вводного материала, доступного студентам первых курсов, далее уровень изложения постепенно повышается, предлагая материал, доступный для студентов старших курсов.

Я использовал многочисленные примеры и упражнения как средство для представления основных идей и принципов, лежащих в основе различных методов теории исследования операций. Каждый представленный в книге решенный пример состоит из ряда подзадач, которые охватывают в определенных пропорциях (надеюсь, сбалансированно) вопросы создания и формализации моделей, вычислительные аспекты и теоретические основы. В конце каждой главы приводится набор комплексных задач, связанных с темой, излагаемой в главе, и значительно углубляющих и расширяющих ее. Шестое издание содержит более 1000 упражнений (60% из них появились только в этом издании).

Книга разбита на три части. Часть *Детерминированные модели* включает темы линейного программирования, сетевых моделей, многокритериальной оптимизации (целевого программирования), динамического (детерминированного) программирования и моделей управления запасами. Тема линейного программирования изложена так, что начинающий студент получит здесь основы практического применения методов, включая теорию двойственности и анализ чувствительности. Глава, посвященная сетевым моделям, содержит обобщенные модели и алгоритмы, включая алгоритм нахождения кратчайших путей (алгоритм Флойда), а также исследование потоков в сетях с помощью методов линейного программирования и показывает их связь с транспортными моделями. В отдельной главе собран углубленный материал по теории линейного программирования. Отдельные главы посвящены целевому и целочисленному программированию. В главе о целочисленном программировании основной упор сделан на применении многообещающего метода ветвей и границ. Новые приложения также включены в главу, посвященную динамическому программированию. Детерминированные модели управления запасами вынесены в отдельную главу.

Часть *Вероятностные модели* начинается с глав, содержащих основы теории вероятностей и математической статистики. Материал о теории принятия решений охватывает аналитический иерархический подход и раскрывает роль функции полезности. В теории

[1] Ранее на русский язык было переведено 3-е издание этой книги: Таха Х. *Введение в исследование операций*: В 2 кн. — М.: Мир, 1985. — *Прим. ред.*

игр метод, основанный на линейном программировании, в настоящем издании, с одной стороны, упрощен, с другой стороны — усилен. Стохастическое динамическое программирование представлено новой главой, за которой следует глава о вероятностных моделях управления запасами. Новое изложение теории массового обслуживания позволяет студентам изучить как ее практическое применение, так и саму теорию или, при желании, сосредоточиться только на практических аспектах темы. В отдельную главу вынесены основы и принципы дискретного имитационного моделирования, а введение в язык моделирования SIMMET II перенесено в Приложение Б. Материал о марковских процессах принятия решений переписан в соответствии с новой концепцией книги.

Часть *Нелинейные модели* повторяет материал пятого издания, но она также подверглась значительным изменениям.

Программное обеспечение, сопровождающее эту книгу, включает программу TORA и "студенческую" версию языка SIMMET II.[1] Программа TORA реализует различные алгоритмы, описанные в книге, и может помочь в их изучении либо может просто использоваться для решения соответствующих задач. SIMMET II имеет все возможности и средства, присущие ее коммерческой версии, — различие заключается только в том, что данная версия имеет ограничения на размер решаемых задач.

Книга имеет пять приложений. Приложение А содержит обзор теории матриц. Приложение Б предлагает введение в язык имитационного моделирования SIMMET II. Материал, посвященный инсталляции и использованию программного обеспечения TORA и SIMMET II, представлен в Приложении В. Приложение Г содержит таблицы нормального, Стьюдента и χ^2 распределений. Ответы к половине упражнений представлены в Приложении Д.

Благодарности

Многие мои коллеги поддержали меня в работе над этой книгой своими советами и критическими замечаниями. Я глубоко благодарен им всем и выражаю надежду на наше дальнейшее взаимовыгодное сотрудничество. Особо хочу поблагодарить профессоров Гая Карри (Guy Curry) из Техасского университета, Дона Э. Дела (Don E. Deal) из университета Хьюстона, Ричарда Френсиса (Richard Francis) из университета Флориды, Ясcера Хосни (Yasser Hosni) из Флоридского центрального университета, Аллена С. Шермана (Allen C. Schuermann) из университета шт. Оклахома и Эвангелоса Триантафиллу (Evangelos Triantaphyllou) из университета шт. Луизиана.

Я также благодарен своему новому издателю Prentice Hall за мягкий и гладкий переход под его покровительство. Выражаю особую благодарность моим редакторам Бейни М. де Леон (Bayani M. de Leon), Алисе Дворкин (Alice Dworkin) и Редоре Пифиаренда (Rhodora Pefiaranda) за их профессиональную работу по подготовке шестого издания книги.

Хэмди А. Таха

[1] Упомянутое программное обеспечение можно найти на Web-узле Издательского дома "Вильямс" по адресу: www.williamspublishing.com. — *Прим ред.*

Исследование операций: обзор

1.1. Математические модели исследования операций

Предположим, что в соответствии с деловыми обязательствами вам необходимо в течение пяти недель пять раз посетить город В (постоянное ваше пребывание — город А). Вы должны быть в городе В в понедельник первой недели и окончательно возвратиться в город А в среду пятой недели. Заказной билет из города А в город В и обратно стоит \$400, однако вы можете получить 20% скидки от стоимости билетов, если вылет придется на конец недели. Кроме того, следует учесть, что стоимость билета только в одну сторону равна 75% от стоимости заказного билета. Вы, естественно, хотите минимизировать стоимость перелетов. Как это сделать?

Описанную ситуацию можно рассматривать как задачу принятия решений, где для нахождения оптимального решения требуется определить три основных компонента.

1. Что в данном случае считать **альтернативными** решениями?
2. Каким **ограничениям** должно удовлетворять возможное решение?
3. По какому **критерию** должны отбираться альтернативные решения?

В нашей задаче возможны следующие альтернативы.

1. Покупка пяти заказных билетов А-В-А (т.е. из города А в город В и обратно).
2. Покупка одного билета в одну сторону А-В, четырех билетов А-В-А, захватывающих конец недели, и одного “однонаправленного” билета В-А.
3. Покупка билета А-В-А для первой недели, причем между датами вылетов должен быть понедельник; для последней недели покупка билета А-В-А, между датами которого должна быть среда, причем первый и последний билеты должны захватывать последние дни недели; четыре билета А-В-А, между датами которых также есть последние дни недели.

Ограничением в данной задаче являются дни прибытия: понедельник первой недели и среда пятой.

В данном случае естественным критерием для оценивания возможных альтернатив является цена билетов. Альтернатива, обеспечивающая наименьшую стоимость билетов, будет наилучшей. В данном случае имеем следующие альтернативы.

Альтернатива 1: стоимость билетов = 5×400 = \$2000.

Альтернатива 2: стоимость билетов = $0.75 \times 400 + 4 \times 0.8 \times 400 + 0.75 \times 400 = \$1800$.

Альтернатива 3: стоимость билетов = $5 \times (0.8 \times 400) = \mathbf{\$1600}$.

Очевидно, что наилучшей является третья альтернатива.

Приведенный пример показывает основные принципиальные составляющие модели исследования операций (ИО), а именно альтернативы, ограничения и критерий отбора альтернатив. В общем случае в задачах принятия решений альтернативы зависят от определенного набора переменных, которые затем могут использоваться при формализации ограничений и критерия в виде подходящих математических функций. В результате формализации получаем **математическую модель**, содержащую изменяемые переменные, ограничения и функцию критерия, которая также называется **целевой функцией**. Решением математической модели будет такой набор значений переменных, который **оптимизирует** (**максимизирует** или **минимизирует**) функцию критерия и удовлетворяет всем ограничениям. Такой набор переменных называется **оптимальным допустимым решением**.

Типичную математическую модель ИО схематически можно представить следующим образом.

Максимизация или минимизация *целевой функции* при условии выполнения *ограничений.*

Пример 1.1–1

Рассмотрим следующую задачу. Среди всех прямоугольников с периметром фиксированной длины L необходимо найти прямоугольник максимальной площади. Какую длину и ширину будет иметь такой прямоугольник?

В этой задаче переменными будут длина l и ширина w прямоугольника. Площадь прямоугольника A вычисляется по формуле $A = lw$. Таким образом, надо максимизировать величину A при условии, что длина периметра прямоугольника, вычисляемая по формуле $2(l + w)$, равна заданной величине L. Математическая модель будет записана следующим образом.

$$\text{Максимизировать } A = lw$$

при ограничении

$$l + w = \frac{L}{2}.$$

Эту задачу можно решить, выразив из равенства $l + w = L/2$ одну переменную (например, l) через другую (w) и подставив ее в формулу целевой функции. В результате получим следующее.

$$A = \left(\frac{L}{2} - w\right)w = \frac{Lw}{2} - w^2.$$

Как известно, функция A будет иметь экстремум при том значении w, при котором производная по w этой функции будет обращаться в нуль. Другими словами, надо решить уравнение

$$\frac{dA}{dw} = \frac{L}{2} - 2w = 0.$$

В результате получим решение $w = L/4$. Отрицательность второй производной при данном значении w доказывает, что эта точка действительно является точкой максимума функции A. Из равенства ограничения получаем, что $l = L/4$. Следовательно, оптимальным решением данной задачи будет $w = L/4$ и $l = L/4$, т.е. среди прямоугольников с фиксированным периметром максимальную площадь будет иметь квадрат.

Хотя математические модели являются краеугольным камнем в изучении ИО, задача принятия решений не ограничивается построением и решением математических моделей. По большому счету, задача принятия решений содержит "нематериальные" факторы, которые трудно поддаются количественному определению, но, безусловно, значительно влияют на качество получаемого решения. Среди них, конечно же, имеется и человеческий фактор. Часто эффект от человеческого поведения может свести на "нет", самое оптимальное решение, полученное на основе любой математической модели. Иллюстрацией этого может служить широко известная *проблема лифта*. По многочисленным жалобам жильцов высотного дома на длительное ожидание лифта было оптимизировано время ожидания, рассчитанное на основе модели массового обслуживания. Но предложенное решение не уменьшило поток жалоб. Дальнейшее изучение ситуации показало, что жильцам просто скучно ждать лифт. Проблема была решена, когда в холле возле лифтов повесили большие зеркала. Жалобы на длительное ожидание лифта прекратились: теперь жильцы коротают время возле лифтов, разглядывая себя и других в зеркале, что, согласитесь, почти не надоедает.

Математический аспект исследования операций обязательно должен рассматриваться в широком контексте всего процесса принятия решений. Это было осознано британскими учеными, которые стали "пионерами" в области ИО еще во время Второй мировой войны. Хотя их работы, в основном, были сосредоточены на оптимизации размещения ограниченных военных ресурсов, в команду разработчиков ИО входили также специалисты по социологии, психологии и поведенческим наукам, что подчеркивало важность человеческого фактора в процессе принятия решений.

1.2. Методы исследования операций

В моделях исследования операций переменные, от которых зависят ограничения и целевая функция, могут быть дискретными (чаще всего целочисленными) и континуальными (непрерывными). В свою очередь, ограничения и целевая функция делятся на линейные и нелинейные. Задачи оптимизации, представленные этими моделями, дали толчок к разработке различных методов решения, которые должны учитывать соответствующие математические свойства этих моделей. Наиболее известными и эффективными из них являются методы **линейного программирования**, когда целевая функция и все ограничения будут линейными. Для решения математических моделей других типов предназначены методы **динамического программирования**, **целочисленного программирования**, **нелинейного программирования**, **многокритериальной оптимизации** и методы **сетевых моделей**.

Практически все методы исследования операций порождают вычислительные алгоритмы, которые являются итерационными по своей природе. Это подразумевает, что задача решается **последовательно** (итерационно), когда на каждом шаге (итерации) получаем решения, постепенно сходящиеся к оптимальному решению. Итерационная природа алгоритмов обычно приводит к объемным однотипным вычислениям. В этом и заключается причина того, что эти алгоритмы разрабатываются, в основном, для реализации с помощью вычислительной техники.

Использование компьютера как неотъемлемого средства решения задач ИО порождает определенные вычислительные сложности, а именно ошибки машинного округления. Такие ошибки особенно заметны при увеличении числа итераций. Проблема ошибок округления усугубляется, если переменные модели ИО должны принимать только целочисленные значения. Поскольку компьютер все вычисления выполняет в арифметике с плавающей запятой, точное представление некоторых целочисленных значений становится невозможным; в таком случае о решении можно только сказать, что оно принадлежит определенной области значений.

Некоторые математические модели могут быть такими сложными, что их невозможно решить никакими доступными методами оптимизации. В этом случае остается только **эвристический подход**: поиск *подходящего* "хорошего" решения вместо *оптимального*. Эвристический подход предполагает наличие *эмпирических правил*, в соответствии с которыми ведется поиск подходящего решения. Обычно эвристические алгоритмы выполняются значительно быстрее, чем алгоритмы нахождения точного решения.

1.3. Имитационное моделирование

Несмотря на впечатляющие достижения математического моделирования, многие реальные ситуации невозможно адекватно представить с помощью соответствующих математических моделей. В одних случаях в этом "виновата" определенная "жесткость" математики как языка описания и представления событий и явлений. Кроме того, даже если есть возможность формализовать рассматриваемую жизненную ситуацию посредством построения математической модели, полученная на ее основе задача оптимизации может быть слишком сложной для современных алгоритмов решения задач этого класса.

Альтернативой математическому моделированию сложных систем может служить **имитационное моделирование**. Этот вид моделирования часто является наилучшим (если не единственным) способом исследования реальных систем. Различие между математической и имитационной моделями заключается в том, что в последней отношение между "входом" и "выходом" модели может быть явно не задано. Вместо явного математического описания взаимоотношения между входными и выходными переменными математической модели, при имитационном моделировании реальная система разбивается на ряд достаточно малых (в функциональном отношении) элементов или модулей. Затем поведение исходной системы имитируется как поведение совокупности этих элементов, определенным образом связанных (путем установления соответствующих взаимосвязей между ними) в единое целое. Вычислительная реализация такой модели начинается с входного элемента, далее проходит по всем элементам, пока не будет достигнут выходной элемент модели.

Вычислительные аспекты имитационных моделей обычно сравнительно несложные, но, как правило, очень трудоемкие. Поэтому реализация таких моделей подразумевает использование вычислительной техники.

Имитационные модели значительно гибче в представлении реальных систем, чем их математические "конкуренты". Причина такой гибкости заключается в том, что при имитационном моделировании исходная система рассматривается на элементном уровне, в то время как математические модели стремятся описать системы на глобальном, как можно более общем уровне.

Но за гибкость имитационных моделей приходится платить высокими требованиями к потребляемым временным и вычислительным ресурсам. Поэтому реализация некоторых имитационных моделей даже на современных быстрых и высокопроизводительных компьютерах может быть очень медленной.

1.4. Искусство моделирования

Решения реальных задач исследования операций должны быть плодом *коллективной работы*, когда бок о бок работают аналитики ИО и клиент-заказчик задачи принятия решений. Аналитикам ИО с их знаниями возможностей математического моделирования необходимы опыт и знание реальной ситуации, исходящие от клиента, для которого, собственно, и решается задача ИО.

Исследование операций, как инструмент задачи принятия решения, можно рассматривать и как науку, и как искусство. Наука здесь представлена всей мощью математических методов, а искусство — тем обстоятельством, что успех на всех этапах, предшествующих получению оптимального решения математической модели, в большей степени зависит от творчества и опыта всей команды, занимающейся решением задачи ИО. Виллимейн (Willemain, [4]) утверждает, что "эффективная практика [ОИ] требует нечто большего, чем только знания и компетентность. Она также требует, среди прочего, "технической" мудрости (т.е. понимание того, когда и как применять тот или иной метод или алгоритм) и определенного уровня коммуникабельности и организационных способностей".

Из-за "неуловимого" человеческого фактора трудно дать точные предписания для реализации теории исследования операций на практике. Можно попытаться показать только общую направленность такой реализации.

На практике реализация методов ИО должна включать следующие этапы.

1. Формализация исходной проблемы.
2. Построение математической модели.
3. Решение модели.
4. Проверка адекватности модели.
5. Реализация решения.

Из всех пяти приведенных этапов только третий, *решение модели*, достаточно точно определен и наиболее прост для реализации в рамках методики ИО, поскольку действия на этом этапе основываются на точной математической теории. Выполнение остальных этапов в значительной мере является искусством, а не наукой. Поэтому мы не можем точно описать процедуры выполнения этих этапов.

Формализация проблемы требует исследования той предметной области, где возникла рассматриваемая проблема. Это начальный этап работы любой команды аналитиков ИО. В результате такого исследования должны быть получены следующие три принципиальных элемента решаемой задачи: 1) описание возможных альтернативных решений, 2) определение целевой функции, 3) построение системы ограничений, накладываемых на возможные решения.

Построение математической модели означает перевод формализованной задачи, описание которой получено на предыдущем этапе, на четкий язык математических соотношений. Если полученная модель является одной из стандартных математических моделей, таких как модель линейного программирования, то решение обычно достигается путем использования соответствующих существующих алгоритмов. Если же результирующая модель очень сложная и не приводится к какому-либо стандартному типу моделей, то команда ИО может либо упростить модель, либо применить эвристический подход, либо использовать имитационное моделирование. В некоторых случаях комбинация математической, имитационной и эвристической моделей может привести к решению исходной проблемы.

Решение модели, как уже упоминалось, — наиболее простой из всех этапов реализации методов исследования операций, так как здесь используются известные алгоритмы оптимизации. Важным аспектом этого этапа является *анализ чувствительности* полученного решения. Это подразумевает получение дополнительной информации о поведении "оптимального" решения при изменении некоторых параметров модели. Анализ чувствительности особенно необходим, когда невозможно точно оценить параметры модели. В этом случае важно изучить поведение оптимального решения в окрестности первоначальных оценок значений параметров модели.

Проверка адекватности модели предполагает проверку правильности модели, т.е. определения того, соответствует ли поведение модели в определенных ситуациях поведению исходной реальной системы. Но сначала команда аналитиков ИО должна удостовериться, что модель не содержит "сюрпризов". Другими словами, надо убедиться, что решение, полученное в рамках построенной модели, имеет смысл и интуитивно приемлемо. Формальным общепринятым методом проверки адекватности модели является сравнение полученного решения (поведение модели) с известными ранее решениями или поведением реальной системы. Модель считается адекватной, если при определенных начальных условиях ее поведение совпадает с поведением исходной системы при тех же начальных условиях. Конечно, это не гарантирует, что при других начальных условиях поведение модели будет совпадать с поведением реальной системы. В некоторых случаях в силу разных причин невозможно прямое сравнение модели с реальной системой или сравнение решений, полученных в рамках этой модели, с известными решениями (например, из-за отсутствия таких данных). В такой ситуации для проверки адекватности математической модели можно использовать имитационное моделирование, т.е. сравнивать поведение математической и имитационной моделей.

Реализация решения подразумевает перевод результатов решения модели в рекомендации, представленные в форме, понятной для лиц, принимающих решения, т.е. заказчиков решения исходной проблемы. Бремя этой непростой задачи ложится непосредственно на плечи команды аналитиков ИО.

1.5. Об этой книге

Моррис (Morris, [3]) утверждает, что "изучение моделей не эквивалентно изучению моделирования". Автор постоянно держал эту важную мысль в голове во время подготовки шестого издания данной книги. Он сознательно старался привнести искусство моделирования в теорию исследования операций. Эта книга, кроме описания математических моделей, содержит большое количество упражнений и задач, которые позволяют проникнуть в суть анализа практических ситуаций.

Автор надеется, что эта книга даст студентам не только фундаментальную основу для понимания математических методов исследования операций, но и понимание возможностей их применения. Такое понимание должно показать, что недостаточно сосредоточиться только на философских и "художественных" аспектах ИО. Необходимы фундаментальные знания математических методов исследования операций. Только на этой основе студенты могут "взращивать" свой "художественный" потенциал в искусстве моделирования ИО. Хорошим подспорьем здесь может служить изучение публикаций и статей в различных журналах. Автор настоятельно рекомендует журнал *Interfaces* (издательство INFORMS Института управленческих наук и исследования операций) как богатый источник интересных приложений теории ИО.

Литература

1. Evans J. *Creative Thinking in the Decision and Management Sciences*, South-Western Publishing, Cincinnati, Ohio, 1991.
2. Gass S. *Model World: Danger, Beware the User as a Modeler*, Interfaces, Vol. 20, No. 3, pp. 60–64, 1990.
3. Morris W. *On the Art of Modeling*, Management Science, Vol. 13, pp. B707–B717, 1967.
4. Willemain T.R. *Insights on Modeling from a Dozen Experts*, Operations Research, Vol. 42, No. 2, pp. 213–222, 1994.

Литература, добавленная при переводе[1]

Вагнер Г. *Основы исследования операций.* — М.: Мир, 1972.

Вентцель Е.С. *Исследование операций.* — М.: Советское радио, 1972.

Вилкас Э.Й., Майминас Е.З. *Решения: теория, информация, моделирование.* — М.: Радио и связь, 1981.

Гермейер Ю.Б. *Введение в теорию исследования операций.* — М.: Наука, 1971.

Ларичев О.И. *Наука и искусство принятия решений.* — М.: Наука, 1979.

Ларичев О.И. *Объективные модели и субъективные решения.* — М.: Наука, 1987.

Кофман А. *Методы и модели исследования операций.* — М.: Мир, 1966.

Краснощеков П.С., Петров А.А. *Принципы построения моделей.* — М.: Изд-во МГУ, 1983.

Шеннон Р. *Имитационное моделирование систем — искусство и наука.* — М.: Мир, 1978.

[1] Литература по исследованию операций на русском языке очень обширна, но, к сожалению, в силу известных причин в последнее десятилетие издание новых книг по этой тематике практически прекратилось (впрочем, как и другой научной литературы). Поэтому пусть извинит нас читатель, если добавленная литература покажется ему устаревшей (утешением может служить то, что математика, как вечная наука и наука о вечном (с определенным допущением), не может устареть). Мы будем приводить, в основном, монографии, "устоявшиеся" в качестве учебных пособий для вузов. — *Прим. ред.*

ЧАСТЬ I

ДЕТЕРМИНИРОВАННЫЕ МОДЕЛИ

Введение в линейное программирование

2.1. Введение

Линейное программирование (ЛП) — это метод математического моделирования, разработанный для *оптимизации* использования *ограниченных* ресурсов. ЛП успешно применяется в военной области, индустрии, сельском хозяйстве, транспортной отрасли, экономике, системе здравохранения и даже в социальных науках. Широкое использование этого метода также подкрепляется высокоэффективными компьютерными алгоритмами, реализующими данный метод. На алгоритмах линейного программирования (учитывая их компьютерную эффективность) базируются оптимизационные алгоритмы для других, более сложных типов моделей и задач исследования операций (ИО), включая целочисленное, нелинейное и стохастическое программирование.

Вычисления в методе ЛП, как и во многих задачах ИО, как правило, очень трудоемкие и поэтому требуют применения вычислительной техники. Свободно распространяемая программа TORA, сопровождающая изложение материала данной книги, значительно облегчает эту вычислительную ношу.[1]

2.2. Ограничения в модели линейного программирования

В этом разделе на простом примере с двумя переменными показаны основные элементы модели ЛП. Далее этот пример будет обобщен в общую задачу линейного программирования.

Пример 2.2–1. (Компания Reddy Mikks[2])

Компания Reddy Mikks производит краску для внутренних и наружных работ из сырья двух типов: M1 и M2. Следующая таблица представляет основные данные для задачи.

[1] Напомним еще раз, что программу TORA можно загрузить с Web-страницы данной книги по адресу: www.williamspublishing.com (Web-узел Издательского дома "Вильямс"). — *Прим. ред.*

[2] Автор часто использует в примерах шуточные названия компаний, которые адекватно трудно перевести на русский язык. Например, в данном случае Reddy Mikks дословно не переводится, но по звучанию можно перевести и как "Охряные смеси" (намек на производимые краски), и как "Краснощекие бездельники". Эти названия, как правило, не несут смысловой нагрузки. Поэтому в большинстве случаев мы будем оставлять их без перевода. — *Прим. перев.*

	Расход сырья (в тоннах) на тонну краски		Максимально возможный ежедневный расход сырья
	для наружных работ	для внутренних работ	
Сырье М1	6	4	24
Сырье М2	1	2	6
Доход (в $1000) на тонну краски	5	4	

Отдел маркетинга компании ограничил ежедневное производство краски для внутренних работ до 2 т (из-за отсутствия надлежащего спроса), а также поставил условие, чтобы ежедневное производство краски для внутренних работ не превышало более чем на тонну аналогичный показатель производства краски для внешних работ. Компания хочет определить оптимальное (наилучшее) соотношение между видами выпускаемой продукции для максимизации общего ежедневного дохода.

Задача (модель) линейного программирования, как и любая задача исследования операций, включает три основных элемента.

1. **Переменные**, которые следует определить.
2. **Целевая функция**, подлежащая оптимизации.
3. **Ограничения**, которым должны удовлетворять переменные.

Определение переменных — первый шаг в создании модели. После определения переменных построение ограничений и целевой функции обычно не вызывает трудностей.

В нашем примере необходимо определить ежедневные объемы производства краски для внутренних и наружных работ. Обозначим эти объемы как переменные модели:

x_1 — ежедневный объем производства краски для наружных работ;
x_2 — ежедневный объем производства краски для внутренних работ.

Используя эти переменные, далее строим целевую функцию. Логично предположить, что целевая функция, как суммарный ежедневный доход, должна возрастать при увеличении ежедневных объемов производства красок. Обозначим эту функцию через z (она измеряется в тысячах долларов) и положим, что

$$z = 5x_1 + 4x_2.$$

В соответствии с целями компании получаем задачу:

$$\text{Максимизировать } z = 5x_1 + 4x_2.$$

Итак, остался не определенным последний элемент модели — условия (ограничения), которые должны учитывать ограниченные возможности ежедневного потребления сырья и ограниченность спроса на готовую продукцию. Другими словами, ограничения на сырье можно записать следующим образом.

$$\begin{pmatrix}\text{Используемый объем}\\ \text{сырья для производства}\\ \text{обоих видов краски}\end{pmatrix} \le \begin{pmatrix}\text{Максимально возможный}\\ \text{ежедневный расход сырья}\end{pmatrix}$$

Из таблицы с данными имеем следующее.

Используемый объем сырья М1 = $6x_1 + 4x_2$ (т)
Используемый объем сырья М2 = $1x_1 + 2x_2$ (т)

Так как ежедневный расход сырья М1 и М2 ограничен соответственно 24 и 6 тоннами, получаем следующие ограничения.

$$6x_1 + 4x_2 \le 24 \text{ (сырье M1)}$$
$$1x_1 + 2x_2 \le 6 \text{ (сырье M2)}$$

Существует еще два ограничения по спросу на готовую продукцию: (1) максимальный ежедневный объем производства краски для внутренних работ не должен превышать 2 т и (2) ежедневный объем производства краски для внутренних работ не должен превышать ежедневный объем производства краски для наружных работ более чем на одну тонну. Первое ограничение простое и записывается как $x_2 \le 2$. Второе можно сформулировать так: разность между ежедневными объемами производства красок для внутренних и наружных работ не должна превышать одной тонны, т.е. $x_2 - x_1 \le 1$.

Еще одно неявное ограничение состоит в том, что переменные x_1 и x_2 должны быть неотрицательными. Таким образом, к сформулированным выше ограничениям необходимо добавить **условие неотрицательности переменных**: $x_1 \ge 0$, $x_2 \ge 0$. (В главе 3 мы увидим, что это условие весьма важно для разработки алгоритмов решения задачи ЛП.)

Окончательно задача будет записана следующим образом:

$$\text{Максимизировать } z = 5x_1 + 4x_2$$

при выполнении ограничений

$$\begin{aligned} &6x_1 + 4x_2 \le 24, \\ &x_1 + 2x_2 \le 6, \\ &-x_1 + x_2 \le 1, \\ &x_2 \le 2, \\ &x_1 \ge 0, x_2 \ge 0. \end{aligned}$$

Любое решение, удовлетворяющее ограничениям модели, является **допустимым**. Например, решение $x_1 = 3$ и $x_2 = 1$ будет допустимым, так как не нарушает ни одного ограничения, включая условие неотрицательности. Чтобы удостовериться в этом, подставьте значения $x_1 = 3$ и $x_2 = 1$ в левые части неравенств системы ограничений и убедитесь, что ни одно неравенство не нарушается. Значение целевой функции при этом решении будет равно $z = 5\times 3 + 4\times 1 = 19$ (тысяч долларов).

Итак, задача сформулирована, теперь встает вопрос о нахождении **оптимального допустимого решения**, доставляющего максимум целевой функции. После некоторых раздумий приходим к выводу, что задача имеет много (фактически, бесконечно много) допустимых решений. По этой причине невозможна подстановка значений переменных для поиска оптимума, т.е. нельзя применить простой перебор всех допустимых решений. Следовательно, необходима эффективная процедура отбора допустимых решений для поиска оптимального. В разделе 2.3 показан графический метод нахождения оптимального допустимого решения, а в главе 3 описано его алгебраическое обобщение.

В предыдущем примере целевая функция и все ограничения были линейными. Свойство *линейности* функций предполагает следующее.

1. Значения левых частей неравенств ограничений и значение целевой функции *прямо пропорциональны* значениям переменных.
2. *Аддитивность переменных* означает, что общий вклад всех переменных в значения целевой функции и левых частей неравенств ограничений является прямой суммой вкладов каждой отдельной переменной.

Упражнения 2.2,а

1. В модели для компании Reddy Mikks постройте новые ограничения, исходя из следующих условий.

   a) Ежедневный объем производства краски для внутренних работ должен *не менее чем* на одну тонну превышать ежедневный объем производства краски для наружных работ.

   b) Ежедневное потребление сырья M2 должно быть *не менее* 3 т и *не более* 6 т.

   c) Ежедневный объем производства краски для внутренних работ не может быть меньше ежедневного объема производства краски для наружных работ.

   d) Минимальный ежедневный общий объем производства краски обоих типов составляет 3 т.

   e) Отношение ежедневного объема производства краски для внутренних работ к общему объему производства краски обоих типов не должно превышать 1/2.

2. Для компании Reddy Mikks найдите *оптимальное допустимое решение* модели среди следующих решений.

   a) $x_1 = 1, x_2 = 4$.

   b) $x_1 = 2, x_2 = 2$.

   c) $x_1 = 3, x_2 = 1.5$.

   d) $x_1 = 2, x_2 = 1$.

   e) $x_1 = 2, x_2 = -1$.

3. Для допустимого решения $x_1 = 2$, $x_2 = 2$ в модели компании Reddy Mikks определите следующее.

   a) объем используемого сырья M1

   b) объем используемого сырья M2

4. Предположим, что компания Reddy Mikks продает свою краску для наружных работ оптовому покупателю со скидкой, зависящей от объема поставок. В результате доход на тонну продукции составляет $5000, если оптовик покупает не более 2 т краски в день, и $4500 — в противном случае. Можно ли для этой ситуации построить модель линейного программирования?

2.3. Графическое решение задачи линейного программирования

В этом разделе будет показано, как в задаче ЛП с двумя переменными можно получить решение графическим способом. Хотя такая задача редко встречается на практике (типовая задача ЛП обычно содержит тысячи переменных), идеи, вытекающие из графического способа нахождения оптимального решения, будут положены в основу построения общего метода решения задачи ЛП (называемого симплекс-методом), описанного в главе 3.

Графический способ решения задачи ЛП состоит из двух этапов.

1. Построение пространства допустимых решений, удовлетворяющих всем ограничениям модели.
2. Нахождение оптимального решения среди всех точек пространства допустимых решений.

Далее графический способ решения описан в двух вариантах: для максимизации и минимизации целевой функции.

2.3.1. Нахождение максимума целевой функции

Пример 2.3–1

Мы используем модель, построенную для компании Reddy Mikks, чтобы показать оба этапа графического решения задачи ЛП.

Этап 1. Построение пространства допустимых решений.

Сначала проведем оси: на горизонтальной будут указываться значения переменной x_1, а на вертикальной — x_2 (рис. 2.1). Далее рассмотрим условие неотрицательности переменных: $x_1 \geq 0$ и $x_2 \geq 0$. Эти два ограничения показывают, что пространство допустимых решений будет лежать в первом квадранте (т.е. выше оси x_1 и правее оси x_2).

Чтобы учесть оставшиеся ограничения, проще всего заменить неравенства на равенства, в результате чего получим уравнения прямых, а затем на плоскости провести эти прямые. Например, неравенство $6x_1 + 4x_2 \leq 24$ заменяется уравнением прямой $6x_1 + 4x_2 = 24$. Чтобы провести эту линию, надо найти две различные точки, лежащие на этой прямой. Можно положить $x_1 = 0$, тогда $x_2 = 24/4 = 6$. Аналогично для $x_2 = 0$ находим $x_1 = 24/6 = 4$. Итак, наша прямая проходит через две точки (0, 6) и (4, 0). Эта прямая обозначена на рис. 2.1 как линия (1).

Теперь рассмотрим, как графически интерпретируются неравенства. Каждое неравенство делит плоскость (x_1, x_2) на два полупространства, которые располагаются по обе стороны прямой, которая, как показано выше, соответствует данному неравенству. Точки плоскости, расположенные по одну сторону прямой, удовлетворяют неравенству (допустимое полупространство), а точки, лежащие по другую сторону, — нет. "Тестовой" точкой, проверяющей, точки какого полупространства удовлетворяют неравенству, а какого — нет, может служить точка (0, 0). Например, эта точка удовлетворяет первому неравенству $6x_1 + 4x_2 \leq 24$ (здесь $6 \times 0 + 4 \times 0 = 0 \leq 24$). Это означает, что точки полупространства, содержащего начальную точку (0, 0), удовлетворяют этому неравенству. На рис. 2.1 допустимые полупространства показаны стрелочками.

В том случае, когда точка (0, 0) не удовлетворяет неравенству, допустимым полупространством будет то, которое не содержит эту точку. Если же прямая проходит через эту точку, следует в качестве "тестовой" взять какую-либо другую точку.

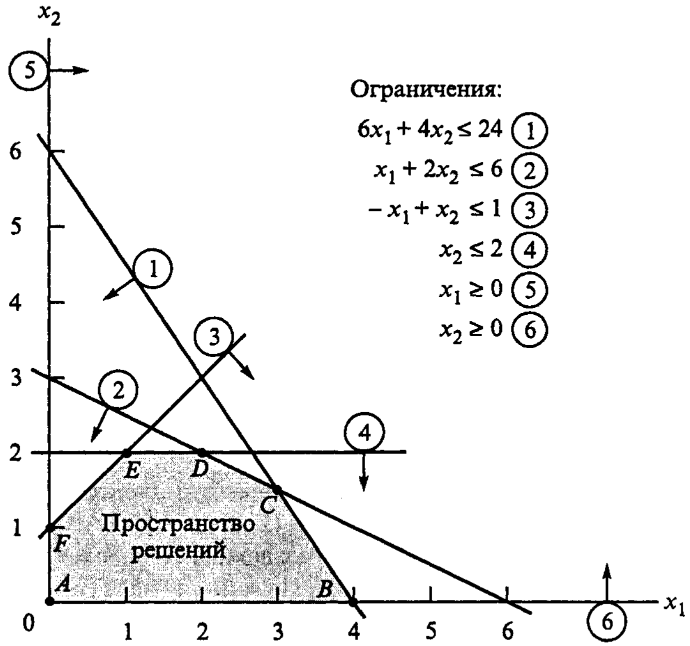


Рис. 2.1

Этап 2. Нахождение оптимального решения.

Точки пространства допустимых решений, показанного на рис. 2.1, удовлетворяют одновременно всем ограничениям. Это пространство ограничено отрезками прямых, которые соединяются в угловых точках *A*, *B*, *C*, *D*, *E* и *F*. Любая точка, расположенная внутри или на границе области, ограниченной ломаной *ABCDEF*, является допустимым решением, т.е. удовлетворяет всем ограничениям. Поскольку пространство допустимых решений содержит бесконечное число точек, необходима некая процедура поиска оптимального решения.

Нахождение оптимального решения требует определения направления возрастания целевой функции $z = 5x_1 + 4x_2$ (напомним, что мы *максимизируем* функцию z). Мы можем приравнять z к нескольким возрастающим значениям, например 10 и 15. Эти значения, подставленные вместо z в выражение целевой функции, порождают уравнения прямых; для значений 10 и 15 получаем уравнения прямых $5x_1 + 4x_2 = 10$ и $5x_1 + 4x_2 = 15$. На рис. 2.2 эти прямые показаны штриховыми линиями, а направление возрастания целевой функции — толстой стрелкой. Целевая функция может возрастать до тех пор, пока прямые, соответствующие возрастающим значениям этой функции, пересекают область допустимых решений. Точка пересечения области допустимых решений и прямой, соответствующей максимально возможному значению целевой функции, и будет точкой оптимума.

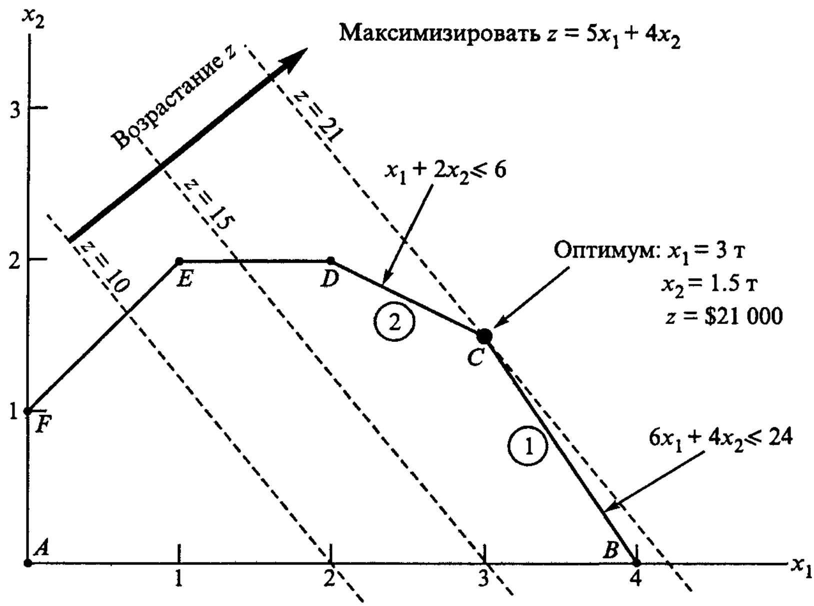


Рис. 2.2

На рис. 2.2 видно, что оптимальное решение соответствует точке C. Эта точка является местом пересечения прямых (1) и (2), поэтому ее координаты x_1 и x_2 находятся как решение системы уравнений, задающих эти прямые:

$$6x_1 + 4x_2 = 24,$$
$$x_1 + 2x_2 = 6.$$

Решением этой системы будет $x_1 = 3$ и $x_2 = 1.5$, при этом значение целевой функции равно $z = 5 \times 3 + 4 \times 1.5 = 21$. Полученное решение означает, что для компании Reddy Mikks оптимальным выбором будет ежедневное производство 3 т краски для наружных работ и 1.5 т — для внутренних работ с ежедневным доходом в \$21 000.

Не случайно, что оптимальное решение расположено в **угловой точке** пространства допустимых решений, где пересекаются две прямые. Если мы изменим наклон функции z (путем изменения ее коэффициентов), то обнаружим, что в любом случае решение достигается в одной из угловых точек (или одновременно в нескольких угловых точках). В этом и состоит основная идея построения общего *симплексного алгоритма*, который будет рассмотрен в главе 3.

Упражнения 2.3,а

1. Для каждого из следующих неравенств определите допустимое полупространство, предполагая, что $x_1, x_2 \geq 0$.

   а) $-3x_1 + x_2 \leq 7$.

b) $x_1 - 2x_2 \geq 5$.
c) $2x_1 - 3x_2 \leq 8$.
d) $x_1 - x_2 \leq 0$.
e) $-x_1 + x_2 \geq 0$.

2. Определите направление возрастания целевой функции z в следующих случаях.

a) Максимизировать $z = x_1 - x_2$.
b) Максимизировать $z = -5x_1 - 6x_2$.
c) Максимизировать $z = -x_1 + 2x_2$.
d) Максимизировать $z = -3x_1 + x_2$.

3. В рамках модели компании Reddy Mikks постройте пространство допустимых решений и найдите оптимальное решение, учитывая (независимо) следующие условия.

a) Ежедневный объем производства краски для наружных работ не должен превышать 2.5 т.
b) Ежедневный объем производства краски для внутренних работ должен быть не менее 2 т.
c) Ежедневный объем производства краски для внутренних работ должен превышать ежедневный объем производства краски для наружных работ в точности на одну тонну.
d) Ежедневный расход сырья M1 должен быть не менее 24 т.
e) Ежедневный расход сырья M1 должен быть не менее 24 т и ежедневный объем производства краски для внутренних работ должен не менее чем на одну тонну превышать ежедневный объем производства краски для наружных работ.

4. Для исходной задачи компании Reddy Mikks определите *угловые точки* области допустимых решений, где достигается оптимальное решение для следующих целевых функций.

a) $z = 3x_1 + x_2$.
b) $z = x_1 + 3x_2$.
c) $z = 6x_1 + 4x_2$.

В чем решение для целевой функции п. c) отличается от решений для целевых функций пп. a) и b)?

5. Джек — студент-первокурсник. Он пришел к выводу, что одна только учеба, без ежедневной игры в баскетбол, плохо влияет на его умственное, нравственное и физическое развитие. Поэтому он решил распределить свое дневное время (примерно 10 часов) для учебы и игры в баскетбол. Привлекательность игрового времени он оценивает в два раза выше, чем привлекательность времени, затраченного на учебу. Но имея совесть и чувство долга, Джек решил, что время для игры не должно превышать время учебы. Кроме того, он заметил, что если выполнять все учебные задания, на игру останется не более 4 часов в день. Помогите Джеку распределить его дневное время так, чтобы он получал максимальное удовольствие и от работы, и от игры.

Пример 2.3–2. (Задача "диеты")

Фармацевтическая фирма Ozark ежедневно производит не менее 800 фунтов[1] некой пищевой добавки, которая состоит из смеси кукурузной и соевой муки, состав которой представлен в следующей таблице.

Мука	Белок (в фунтах на фунт муки)	Клетчатка (в фунтах на фунт муки)	Стоимость (в $ за фунт)
Кукурузная	0.09	0.02	0.30
Соевая	0.60	0.06	0.90

Диетологи требуют, чтобы в пищевой добавке было не менее 30% белка и не более 5% клетчатки. Фирма Ozark хочет определить рецептуру смеси наименьшей стоимости с учетом требований диетологов.

Поскольку пищевая добавка состоит только из кукурузной и соевой муки, переменными для этой задачи, очевидно, будут

x_1 — количество (в фунтах) кукурузной муки, используемой в дневном производстве пищевой добавки;

x_2 — количество (в фунтах) соевой муки, используемой в дневном производстве пищевой добавки.

Целевая функция равна общей стоимости пищевой добавки, производимой за один день, и должна быть минимальной. В данном случае это можно записать следующим образом:

$$\text{минимизировать } z = 0.3x_1 + 0.9x_2.$$

Ограничения модели должны отражать производственные требования и рекомендации диетологов. Фирма должна выпускать не менее 800 фунтов смеси в день; соответствующее ограничение будет записано следующим образом:

$$x_1 + x_2 \geq 800.$$

Рассмотрим ограничение, связанное с количеством белка в пищевой добавке. Общее количество белка в смеси, состоящей из x_1 фунтов кукурузной муки и x_2 фунтов соевой муки, равно $0.09x_1 + 0.6x_2$ (фунтов). Это количество должно составлять не менее 30% от общего объема смеси $x_1 + x_2$. Отсюда получаем следующее неравенство

$$0.09x_1 + 0.6x_2 \geq 0.3(x_1 + x_2).$$

Аналогично строится ограничение для клетчатки:

$$0.02x_1 + 0.06x_2 \leq 0.05(x_1 + x_2).$$

[1] Практически во всех примерах при описании реальных ситуаций автор пользуется системой мер, принятой в США. Мы не стали переводить эти единицы измерения в метрическую систему единиц, так как названия единиц никак не влияют ни на описание примеров, ни на понимание методов, иллюстрируемых этими примерами. — *Прим. ред.*

В последних двух неравенствах переменные x_1 и x_2 надо перенести из правых частей неравенств в левые. Окончательно модель примет следующий вид:

$$\text{Минимизировать } z = 0.3x_1 + 0.9x_2$$

при ограничениях

$$\begin{aligned}
&x_1 + x_2 \geq 800,\\
&0.21x_1 - 0.30x_2 \leq 0,\\
&0.03x_1 - 0.01x_2 \geq 0,\\
&x_1, x_2 \geq 0.
\end{aligned}$$

На рис. 2.3 показано графическое решение этой задачи. В отличие от модели примера 2.3–1, здесь две прямые, соответствующие неравенствам ограничений, проходят через начальную точку (0, 0). Для того чтобы провести на графике такую прямую, необходима еще одна точка. Координаты этой точки можно найти, подставив в уравнение прямой любое значение для одной переменной и затем из этого уравнения найти значение для другой. Например, для второго неравенства из системы ограничений положим $x_1 = 200$, тогда для второй переменной получаем уравнение $0.21\times200 - 0.3x_2 = 0$; отсюда имеем $x_2 = 140$. Таким образом, прямая $0.21x_1 - 0.30x_2 = 0$ проходит через точки (0, 0) и (200, 140). Заметим также, в данном случае для определения допустимого полупространства нельзя использовать в качестве "тестовой" точку (0, 0), здесь следует взять какую-либо другую, например (100, 0) или (0, 100).

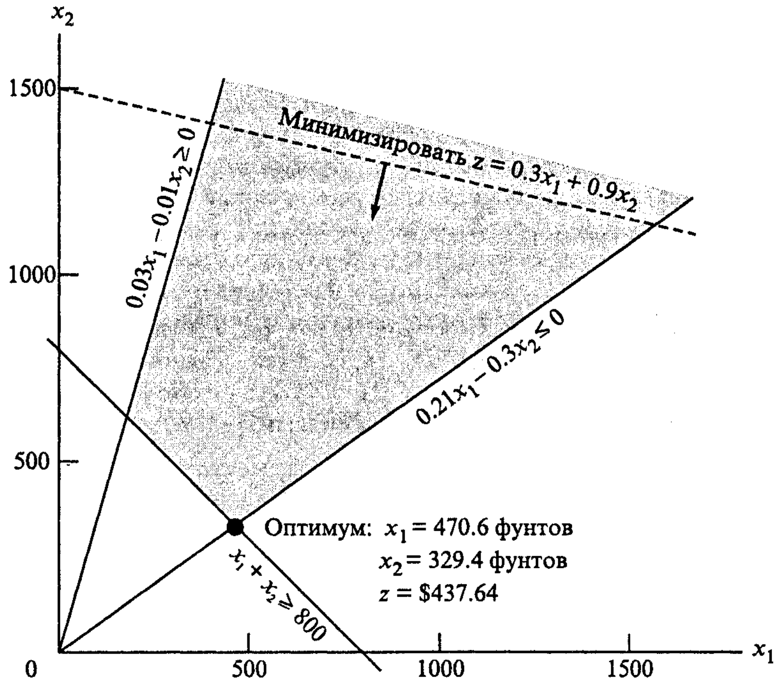


Рис. 2.3

Поскольку в данной модели следует минимизировать целевую функцию, поэтому нужно идти в направлении уменьшения ее значений (это направление на рис. 2.3 показано стрелкой). Оптимальное решение находится на пересечении прямых $x_1 + x_2 = 800$ и $0.21x_1 - 0.30x_2 = 0$, откуда получаем $x_1 = 470.59$ (фунтов) и $x_2 = 329.41$ (фунтов). При этих значениях переменных минимальная стоимость производимой ежедневно пищевой добавки составляет $z = 0.3 \times 470.59 + 0.9 \times 329.41 = \$437.65$.

Упражнения 2.3,b

1. Определите направление убывания следующих целевых функций.
   a) Минимизировать $z = 4x_1 - 2x_2$.
   b) Минимизировать $z = -3x_1 + x_2$.
   c) Минимизировать $z = -x_1 - 2x_2$.
2. В задачу "диеты" добавлено еще одно ограничение: ежедневный расход кукурузной муки ограничен 450 фунтами. Постройте новое пространство допустимых решений и найдите новое оптимальное решение.
3. Найдите оптимальное решение в задаче "диеты" при условии, что ежедневное производство пищевой добавки не должно превышать 800 фунтов. Имеет ли такое решение смысл?
4. Джон, помимо занятий в школе, для поддержания надлежащего финансового уровня должен подрабатывать не менее 20 часов в неделю. Для этого у него есть прекрасная возможность работать в двух розничных магазинчиках. В первом он может работать от 5 до 12 часов в неделю, а во втором — от 6 до 10 часов. Оба магазина предлагают одинаковую почасовую оплату. Джон должен определиться, в каком магазине и сколько ему работать, исходя из фактора "напряженности" работы. Основываясь на сведениях, полученных при общении с работниками этих магазинов, он оценил этот фактор по 10-балльной шкале: для первого и второго магазинов соответственно 8 и 6 баллов. Понятно, что суммарная "напряженность" работы за неделю пропорциональна количеству отработанных часов. Сколько часов Джон должен работать в каждом магазине, чтобы минимизировать общую суммарную "напряженность" работы?

2.3.3. Дополнительные переменные

В обоих примерах этой главы мы использовали неравенства типа "меньше или равно" (знак неравенства $\leq$) и "больше или равно" (знак неравенства $\geq$). В этих примерах также предполагалась неотрицательность всех переменных. В данном разделе мы введем два типа *дополнительных* неотрицательных переменных (назовем их *остаточными* и *избыточными*)[1], которые связаны с неравенствами типа "$\leq$" и "$\geq$" соответственно. Введем

---

[1] Отметим, что в русской математической литературе для этих типов переменных не используются какие-либо специальные названия — они известны просто как *дополнительные* переменные (хотя иногда их также называют *балансными*) В неравенствах они различаются тем, что перед остаточной переменной всегда стоит знак "плюс", а перед избыточной — "минус". — *Прим ред*

также понятие *свободной переменной*, которая может принимать как положительные, так и отрицательные значения (и, конечно, значение 0).

Остаточная переменная. Неравенства типа "$\leq$" обычно можно интерпретировать как ограничения на использование некоторых ресурсов (представленных в левой части неравенств переменными модели). В такой интерпретации *остаточная переменная* показывает количество неиспользованных ресурсов. В примере с компанией Reddy Mikks неравенство $6x_1 + 4x_2 \leq 24$ связано с использованием сырья M1. Это неравенство эквивалентно равенству $6x_1 + 4x_2 + s_1 = 24$, где $s_1 \geq 0$. Здесь остаточная переменная s_1 ($= 24 - 6x_1 - 4x_2$) равна неиспользуемому количеству сырья M1.

Избыточная переменная. Неравенство типа "$\geq$" показывает, что "что-то" должно быть не меньше определенной величины. *Избыточная переменная* определяет превышение значения левой части неравенства над этой величиной. Например, в модели "диеты" неравенство $x_1 + x_2 \geq 800$ показывает, что суточное производство пищевой добавки не должно быть меньше 800 фунтов. Математически это неравенство эквивалентно равенству $x_1 + x_2 - S_1 = 800$, где $S_1 \geq 0$. Положительное значение избыточной переменной S_1 показывает превышение суточного производства добавки над минимальным значением в 800 фунтов.

Свободная переменная. В приведенных выше примерах условие неотрицательности переменных является естественным. Но, конечно, возможны ситуации, когда переменные могут принимать любые действительные значения. Такая ситуация показана в следующем примере.

Пример 2.3–3

Ресторан быстрого обслуживания McBurger торгует порционными мясными пирогами и чизбургерами. На порцию мясного пирога идет четверть фунта мяса, а на чизбургер — только 0.2 фунта. В начале рабочего дня в ресторане имеется 200 фунтов мяса, можно еще прикупить мясо в течение дня, но уже с наценкой в 25 центов. Мясо, оставшееся в конце рабочего дня, жертвуется благотворительной организации "Горячий суп". Ресторан имеет доход 20 центов от одной порции мясного пирога и 15 центов — от одного чизбургера. Как и многие другие, этот ресторан не может продать в день более 900 бутербродов. Какова должна быть доля каждого из бутербродов (т.е. сколько порций мясного пирога и сколько чизбургеров) в ежедневном производстве ресторана, чтобы максимизировать его доход?

Сначала рассмотрим ограничения. Обозначим через x_1 и x_2 соответственно количество порций мясного пирога и чизбургеров, производимых рестораном. Для их производства ресторан может ограничиться 200 фунтами мяса или может прикупить еще. В первом случае получаем ограничение в виде неравенства $0.25x_1 + 0.2x_2 \leq 200$, а во втором — $0.25x_1 + 0.2x_2 \geq 200$. Естественно, выбор одного из этих неравенств будет существенно влиять на возможное оптимальное решение. Так как мы не знаем, какое из них необходимо, логично заменить их одним равенством $0.25x_1 + 0.2x_2 + x_3 = 200$, где x_3 — свободная переменная. Фактически свободная переменная x_3 в данной ситуации одновременно играет роли как остаточной, так и избыточной переменных.

Далее построим целевую функцию. Ресторан хочет максимизировать свой доход. Очевидно, что для максимизации дохода желательно как можно больше продавать своей продукции, но для этого необходимы дополнительные закупки мяса.

В этом случае переменная x_3 должна быть отрицательной, т.е. должна играть роль избыточной переменной.

Для того чтобы раскрыть "двойственную" природу переменной x_3, используем стандартный математический прием, а именно представим ее в следующем виде.

$$x_3 = x_3^+ - x_3^-, \text{ где } x_3^+,\ x_3^- \geq 0$$

Если $x_3^+ > 0$ и $x_3^- = 0$, тогда переменная x_3 играет роль остаточной. Если, напротив, $x_3^- > 0$ и $x_3^+ = 0$, тогда переменная x_3 выступает в роли избыточной. (В главе 3 будет показана ситуация, когда оптимальное решение задачи линейного программирования достигается при положительных значениях как x_3^+, так и x_3^-.) Итак, теперь ограничение можно записать в виде равенства

$$0.25x_1 + 0.2x_2 + x_3^+ - x_3^- = 200.$$

Целевая функция получает следующее выражение.

$$\text{Максимизировать } z = 0.20 + 0.15x_2 - 0.25\,x_3^-$$

Упражнения 2.3,c

1. В модели компании Reddy Mikks рассмотрите допустимое решение $x_1 = 3$ т и $x_2 = 1$ т. Для этого решения найдите недоиспользование сырья M1 и M2.
2. В модели "диеты" определите превышение над минимальным допустимым объемом производства пищевой добавки, на которую расходуется 500 фунтов кукурузной муки и 600 фунтов — соевой.
3. В некотором машинном центре производятся два изделия, причем на производство одной единицы первого изделия затрачивается 10 минут рабочего времени, а на единицу второго изделия — 12 минут. Рабочее время машинного центра ограничено величиной в 2500 минут в день (некоторые операции центр может выполнять параллельно); возможно превышение этой величины, но каждая дополнительная минута работы машинного центра стоит 50 центов. В рабочий день допустимо производить от 150 до 200 единиц первого изделия, но не более 45 единиц второго изделия.
   a) Предполагая, что доход от единицы первого изделия составляет $6.00, а второго — $7.50, постройте модель и найдите оптимальное соотношение между объемами производства изделий, максимизирующее общий доход, а также дополнительное время работы машинного центра.
   b) Если стоимость дополнительного времени работы машинного центра увеличится до $1.50, будет ли компания использовать это время?

2.4. Графический анализ чувствительности

Модель линейного программирования является как бы "моментальным снимком" реальной ситуации, когда параметры модели (коэффициенты целевой функции и неравенств ограничений) предполагаются неизменными. Естественно изучить влияние изменения параметров модели на полученное оптимальное решение задачи ЛП. Такое исследование называется **анализом чувствительности**.

В этом разделе анализ чувствительности основывается на графическом решении задачи ЛП. Рассмотрим два случая: (1) изменение коэффициентов целевой функции и (2) изменение значений констант в правой части неравенств ограничений. Хотя проведенное здесь исследование будет элементарным и ограниченным, оно покажет основные идеи методов анализа чувствительности. Подробно методы этого анализа описаны в главе 4.

2.4.1. Изменение коэффициентов целевой функции

В общем виде целевую функцию задачи ЛП с двумя переменными можно записать следующим образом.

$$\text{Максимизировать или минимизировать } z = c_1x_1 + c_2x_2$$

Изменение значений коэффициентов c_1 и c_2 приводит к изменению угла наклона прямой z. Графический способ решения задачи ЛП, описанный в разделе 2.3, показывает, что это может привести к изменению оптимального решения: оно будет достигаться в другой *угловой точке* пространства решений. Вместе с тем, очевидно, существуют интервалы изменения коэффициентов c_1 и c_2, когда текущее оптимальное решение сохраняется. Задача анализа чувствительности и состоит в получении такой информации. В частности, представляет интерес определение *интервала оптимальности* для отношения c_1/c_2 (или, что то же самое, для c_2/c_1); если значение отношения c_1/c_2 не выходит за пределы этого интервала, то оптимальное решение в данной модели сохраняется неизменным. Следующий пример показывает, как можно получить необходимый результат с помощью анализа графического представления модели ЛП.

Пример 2.4–1

Применим процедуру анализа чувствительности к задаче примера 2.2–1 (модель для компании Reddy Mikks). На рис. 2.4 видно, что функция $z = 5x_1 + 4x_2$ достигает максимального значения в угловой точке C. При изменении коэффициентов целевой функции $z = c_1x_1 + c_2x_2$ точка C останется точкой оптимального решения до тех пор, пока угол наклона линии z будет лежать между углами наклона двух прямых, пересечением которых является точка C. Этими прямыми являются $6x_1 + 4x_2 = 24$ (ограничение на сырье M1) и $x_1 + 2x_2 = 6$ (ограничение на сырье M2). Алгебраически это можно записать следующим образом:

$$\frac{4}{6} \le \frac{c_2}{c_1} \le \frac{2}{1}, \quad c_1 \ne 0,$$

или

$$\frac{1}{2} \le \frac{c_1}{c_2} \le \frac{6}{4}, \quad c_2 \ne 0.$$

В первой системе неравенств условие $c_1 \ne 0$ означает, что прямая, соответствующая целевой функции, не может быть горизонтальной. Аналогичное условие в следующей системе неравенств означает, что эта же прямая не может быть вертикальной. Из рис. 2.4 видно, что интервал оптимальности данной задачи (он определяется двумя прямыми, пересекающимися в точке C) не разрешает целевой функции быть ни горизонтальной, ни вертикальной прямой. Таким образом, мы получили две системы неравенств, определяющих интервал оптимальности в нашем примере. (Когда c_1 и c_2 могут принимать нулевые значения, интервал оптимальности для отношения c_1/c_2 (или c_2/c_1) необходимо разбить на два множества, где знаменатели не обращались бы в нуль. Эта ситуация представлена в упр. 2.4,а(1).)

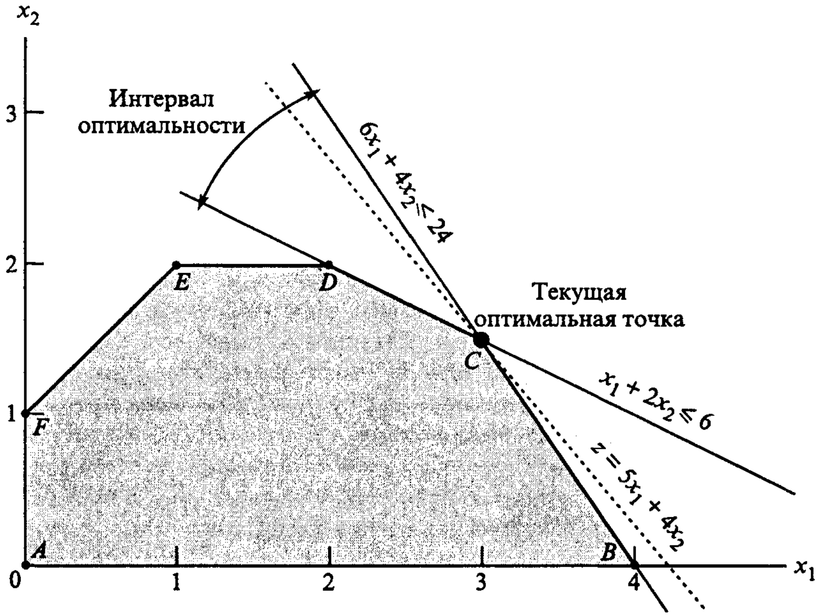


Рис. 2.4

Итак, если коэффициенты c_1 и c_2 удовлетворяют приведенным выше неравенствам, оптимальное решение будет достигаться в точке C. Отметим, если прямая $z = c_1x_1 + c_2x_2$ совпадет с прямой $x_1 + 2x_2 = 6$, то оптимальным решением будет любая точка отрезка CD. Аналогично, если прямая, соответствующая целевой функции, совпадет с прямой $6x_1 + 4x_2 = 24$, тогда любая точка отрезка BC будет оптимальным решением. Однако заметим, что в обоих случаях точка C остается точкой оптимального решения.

Приведенные выше неравенства можно использовать при определении интервала оптимальности для какого-либо одного коэффициента целевой функции, если предположить, что другой коэффициент остается неизменным. Например, если в нашей модели зафиксировано значение коэффициента c_2 (пусть $c_2 = 4$), тогда интервал оптимальности для коэффициента c_1 получаем из неравенств $\frac{1}{2} \leq \frac{c_1}{c_2} \leq \frac{6}{4}$ путем подстановки туда значения $c_2 = 4$. После выполнения элементарных арифметических операций получаем неравенства для коэффициента c_1: $2 \leq c_1 \leq 6$. Аналогично, если зафиксировать значение коэффициента c_1 (например, $c_1 = 5$), тогда из неравенств $\frac{4}{6} \leq \frac{c_2}{c_1} \leq \frac{2}{1}$ получаем интервал оптимальности для коэффициента c_2: $10/3 \leq c_2 \leq 10$.

Упражнения 2.4,а

1. Определите графически интервал оптимальности для отношения c_1/c_2 в следующих задачах; отдельно рассмотрите случаи, когда коэффициенты c_1 и c_2 могут обращаться в нуль.

   a) Максимизировать $z = 2x_1 + 3x_2$ при выполнении условий
   $$3x_1 + 2x_2 \le 5,$$
   $$-x_1 + x_2 \le 0,$$
   $$x_1, x_2 \le 0.$$
   b) Максимизировать $z = 4x_1 + 3x_2$ при выполнении условий
   $$3x_1 + 5x_2 \le 5,$$
   $$x_1 - x_2 \le 0,$$
   $$x_1, x_2 \le 0.$$
   c) Максимизировать $z = x_1 + x_2$ при выполнении условий
   $$-x_1 + x_2 \le 0,$$
   $$3x_1 - x_2 \le 3,$$
   $$x_1, x_2 \ge 0.$$

2. В задаче "диеты" из примера 2.3–2

   a) определите интервал оптимальности для отношения стоимости фунта кукурузной муки к стоимости фунта соевой муки;
   b) если стоимость фунта кукурузной муки увеличится на 20%, а стоимость фунта соевой уменьшится на 5%, то будет ли в этом случае оптимальным ранее найденное решение?
   c) если стоимость фунта кукурузной муки останется фиксированной (30 центов), а стоимость фунта соевой муки возрастет до $1.10, то останется ли оптимальным ранее найденное решение?

3. Магазин B&K продает два вида безалкогольных напитков: колу A1 известного производителя и колу B&K собственного производства. Доход от одной банки колы A1 составляет 5 центов, тогда как доход от одной банки собственной колы — 7 центов. В среднем магазин за день продает не более 500 банок обоих напитков. Несмотря на то что A1 известная торговая марка, покупатели предпочитают колу B&K, поскольку она значительно дешевле. Подсчитано, что объемы продаж колы B&K и A1 (в натуральном исчислении) должны соотноситься не менее 2:1. Кроме того, известно, что магазин продает не менее 100 банок колы A1 в день.

   a) Сколько банок каждого напитка должен иметь магазин в начале рабочего дня для максимизации дохода?
   b) Определите соотношение доходов от напитков A1 и B&K, при котором сохраняется оптимальное решение, найденное на предыдущем шаге.

4. Мебельная фабрика для сборки столов и стульев привлекает к работе на 10 дней четырех столяров. Каждый столяр затрачивает 2 часа на сборку стола и 30 минут — на сборку стула. Покупатели обычно приобретают вместе со столом от че-

тырех до шести стульев. Доход от одного стола составляет \$135 и \$50 — от одного стула. На фабрике установлен 8-часовой рабочий день.

a) Определите графически структуру производства (на 10 рабочих дней), которая максимизировала бы суммарный доход.
b) Определите для найденного решения интервал оптимальности для отношения доходов на единицу продукции.
c) Изменится ли найденное выше оптимальное решение, если доходность столов и стульев уменьшится на 10%?
d) Изменится ли найденное выше оптимальное решение, если доходность столов и стульев составит соответственно \$120 и \$25 на единицу продукции?

5. Банк Elkins в течение нескольких месяцев планирует вложить до \$200 000 в кредитование частных лиц (клиентов) и покупок автомобилей. Банковские комиссионные составляют 14% при кредитовании частных лиц и 12% при кредитовании покупок автомобилей. Оба типа кредитов возвращаются в конце годичного периода кредитования. Известно, что около 3% клиентских и 2% автомобильных кредитов никогда не возвращаются. В этом банке объемы кредитов на покупку автомобилей обычно более чем в два раза превышают объемы других кредитов для частных лиц.

   a) Найдите оптимальное размещение средств по двум описанным видам кредитования и определите коэффициент возврата по всем кредитам.
   b) Определите интервал оптимальности для отношения процентных ставок по двум видам кредитов для найденного на предыдущем шаге оптимального решения.
   c) Предположим, что невозврат кредитов составит 4% и 3% для кредитов частных лиц и кредитов на покупку автомобилей соответственно. Изменится ли при этом оптимальное решение, полученное выше?

6. Завод Electra производит два типа электрических двигателей, каждый на отдельной сборочной линии. Производительность этих линий составляет 600 и 750 двигателей в день. Двигатель первого типа использует 10 единиц некоего комплектующего, а двигатель второго типа — 8 единиц этого же комплектующего. Поставщик может обеспечить на день 8000 единиц этого комплектующего. Доходность двигателя первого типа составляет \$60, второго — \$40.

   a) Определите оптимальную структуру ежедневного производства двигателей.
   b) Найдите интервал оптимальности для отношения доходности двигателей.

7. Консервный завод Popeye перерабатывает за смену 60 000 фунтов спелых помидоров (7 пенсов за фунт) в томатный сок и пасту. Готовая продукция пакетируется в упаковки по 24 банки. Производство одной банки сока требует одного фунта спелых помидоров, а одной банки пасты — трети фунта. Заводской склад может принять за смену только 2000 упаковок сока и 6000 упаковок пасты. Оптовая цена одной упаковки томатного сока составляет \$18, одной упаковки томатной пасты — \$9.

   a) Найдите оптимальную структуру производства консервного завода.
   b) Найдите отношение оптовых цен на продукцию завода, при котором заводу будет выгоднее производить больше томатной пасты, чем сока.

8. Мебельная фабрика собирает из готовых комплектующих два вида кухонных шкафов: обычные и дорогие. Обычный шкаф покрывается белой краской, а дорогой — лаком. Покраска и покрытие лаком производятся на одном производственном покрасочном участке. Сборочная линия фабрики ежедневно может собирать не более 200 обычных шкафов и 150 дорогих. Лакирование одного дорогого шкафа требует вдвое больше времени, чем покраска одного простого шкафа. Если покрасочный участок занят только лакированием дорогих шкафов, то за день здесь можно подготовить 180 таких шкафов. Фабрика оценивает доход от обычных и дорогих кухонных шкафов в \$100 и \$140 соответственно.

   a) Сформулируйте задачу линейного программирования и составьте оптимальное ежедневное расписание работы покрасочного участка.
   b) Предположим, что в результате конкуренции доход от производства одного обычного шкафа снизился до \$80, а дорогого — до \$110. Используя анализ чувствительности, определите, будет ли в этой ситуации решение, полученное на предыдущем шаге, оптимальным.

2.4.2. Стоимость ресурсов

Во многих моделях линейного программирования ограничения трактуются как условия ограниченности ресурсов. В таких ограничениях правая часть неравенств является верхней границей количества доступных ресурсов. В этом разделе мы изучим чувствительность оптимального решения к изменению ограничений, накладываемых на ресурсы. Такой анализ задачи ЛП предлагает простую меру чувствительности решения, называемую **стоимостью единицы ресурса**; при изменении количества доступных ресурсов (на единицу) значение целевой функции в оптимальном решении изменится на стоимость единицы ресурса. Проиллюстрируем этот вид анализа задачи ЛП на следующем примере.

Пример 2.4–2

В модели для компании Reddy Mikks (пример 2.2–1) первые два неравенства представляют собой ограничения на использование сырья M1 и M2 соответственно. Определим стоимость единиц этих ресурсов.

Начнем с ограничения для сырья M1. Напомним, что в данной задаче оптимальное решение достигается в угловой точке C, являющейся точкой пересечения прямых, соответствующих ограничениям на сырье M1 и M2 (рис. 2.5). При изменении уровня доступности материала M1 (увеличение или уменьшение текущего уровня, равного 24 т) точка C оптимального решения "плывет" вдоль отрезка DG. Любое изменение уровня доступности материала M1, приводящее к выходу точки пересечения C из этого отрезка, ведет к неосуществимости оптимального решения в точке C. Поэтому можно сказать, что концевые точки $D = (2, 2)$ и $G = (6, 0)$ отрезка DG определяют *интервал осуществимости* для ресурса M1. Количество сырья M1, соответствующего точке $D = (2, 2)$, равно $6x_1 + 4x_2 = 6 \times 2 + 4 \times 2 = 20$ т. Аналогично количество сырья, соответствующего точке $G = (6, 0)$, равно $6 \times 6 + 4 \times 0 = 36$ т. Таким образом, интервал осуществимости для ресурса M1 составляет $20 \le M_1 \le 36$ (здесь через M_1 обозначено количество материала M1). Если мы определим M_1 как $M_1 = 24 + D_1$, где D_1 — отклонение количества материала M1 от текущего уровня в 24 т, тогда последние неравенства можно переписать как

$20 \le 24 + D_1 \le 36$ или $-4 \le D_1 \le 12$. Это означает, что текущий уровень ресурса M1 может быть уменьшен не более чем на 4 т и увеличен не более чем на 12 т. В этом случае гарантируется, что оптимальное решение будет достигаться в точке C — точке пересечения прямых, соответствующих ограничениям на ресурсы M1 и M2.

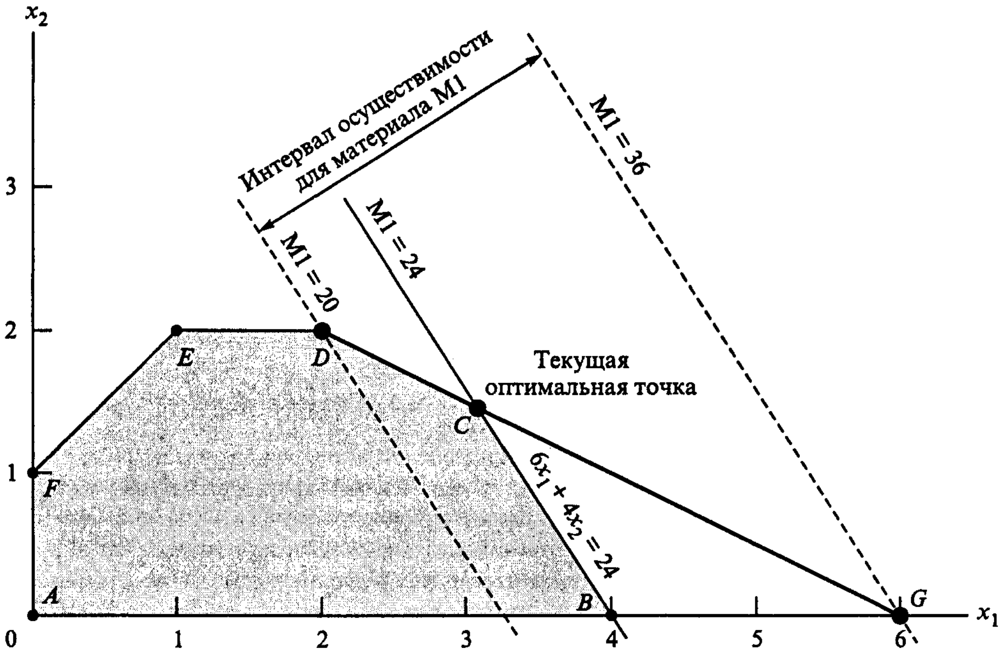


Рис. 2.5

Теперь вычислим стоимость единицы материала M1. При изменении количества сырья M1 от 20 до 36 тонн, значения целевой функции z будут соответствовать положению точки C на отрезке DG. Обозначив через y_1 стоимость единицы ресурса M1, получим следующую формулу.

$$y_1 = \frac{\text{изменение значения } z \text{ при перемещении т. } C \text{ от } D \text{ до } G}{\text{изменение количества } M_1 \text{ при перемещении т. } C \text{ от } D \text{ до } G}$$

Если точка C совпадает с точкой $D = (2, 2)$, то $z = 5 \times 2 + 4 \times 2 = 18$ (тысяч долларов), если же точка C совпадает с точкой $G = (6, 0)$, тогда $z = 5 \times 6 + 4 \times 0 = 30$ (тысяч долларов). Отсюда следует, что

$$y_1 = \frac{30-18}{36-20} = \frac{3}{4} \text{ (тысяч долларов на тонну материала M1).}$$

Этот результат показывает, что изменение количества ресурса M1 на одну тонну (если общее количество этого ресурса не меньше 20 и не больше 36 тонн) приводит к изменению в оптимальном решении значения целевой функции на \$750.

Теперь рассмотрим ресурс M2. На рис. 2.6 видно, что интервал осуществимости для ресурса M2 определяется концевыми точками B и H отрезка BH, где $B = (4, 0)$ и $H = (8/3, 2)$. Точка H находится на пересечении прямых ED и BC. Находим, что количество сырья M2, соответствующего точке B, равно $x_1 + 2x_2 = 4 + 2 \times 0 =$ 4 т, а точке H — $8/3 + 2 \times 2 = 20/3$ т. Значение целевой функции в точке B равно $5x_1 + 4x_2 = 5 \times 4 + 4 \times 0 = 20$ (тысяч долларов), а в точке H — $5 \times 8/3 + 4 \times 2 = 64/3$ (тысяч долларов). Отсюда следует, что количество сырья M2 может изменяться от 4 до 20/3 тонн, а стоимость единицы ресурса M2, обозначенная как y_2, равна

$$y_2 = \frac{64/3 - 20}{20/3 - 4} = \frac{1}{2} \text{ (тысяч долларов на тонну материала M2).}$$

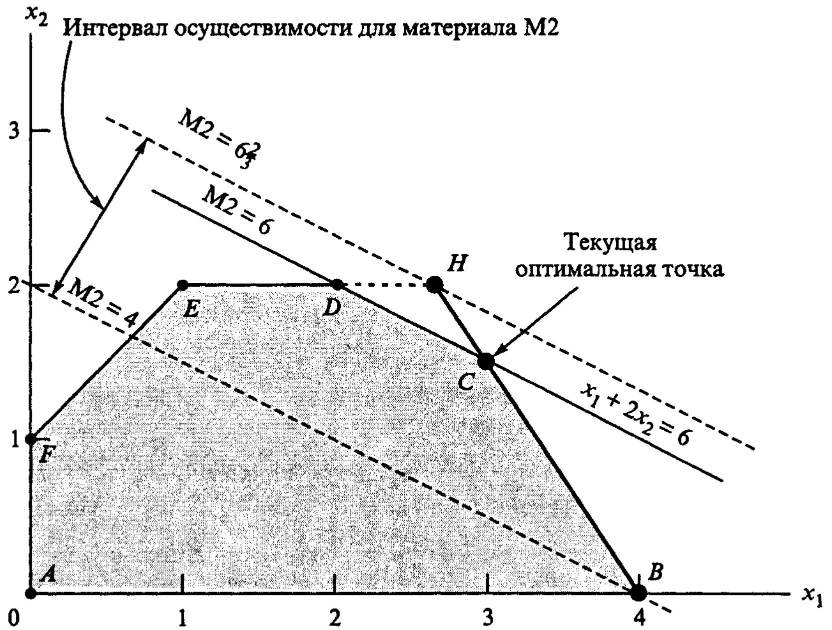


Рис. 2.6

Упражнения 2.4,b

1. Фабрика Wild West производит два типа ковбойских шляп. Производство шляпы первого типа требует в два раза больше временных ресурсов, чем производство шляпы второго типа. Если фабрика будет производить только шляпы второго типа, то в день она сможет изготовить 400 таких шляп. Рынок накладывает ограничения на производство шляп: не более 150 шляп первого типа и 200 шляп второго типа. Доход от производства шляп составляет \$8 на единицу первого типа и \$5 — второго типа.
   a) Примените графический метод для определения ежедневного оптимального производства шляп обоих типов.
   b) Определите стоимость возрастания производства на одну шляпу второго типа и интервал значений числа ежедневного производства этих шляп, для которого данная стоимость была бы применима.

c) Используя метод стоимости единицы ресурса, определите, на сколько изменится максимальный доход фабрики, если ежедневное производство шляп первого типа не будет превышать 120 единиц.

d) Чему равна стоимость возрастания предельного спроса на одну шляпу второго типа?

2. Компания производит два вида продукции, A и B. Объем продаж продукта A составляет не менее 80% от общего объема продаж продуктов A и B. Вместе с тем компания не может производить более 100 единиц продукта A в день. Для производства этих продуктов используется одно и то же сырье, поступление которого ограничено 240 фунтами в день. На изготовление единицы продукта A расходуется 2 фунта сырья, а единицы продукта B — 4 фунта. Цена одной единицы продуктов A и B составляет \$20 и \$50 соответственно.

a) Найдите оптимальную структуру производства этой компании.

b) Определите стоимость единицы сырья и интервал изменения потребляемого сырья, при котором справедлива данная стоимость.

c) С помощью графического анализа чувствительности определите, как изменится значение целевой функции при изменении максимального уровня производства продукта A на ±10 единиц.

3. Компания на производство двух продуктов в день затрачивает 10 часов. Производство каждого продукта состоит из последовательного выполнения трех процессов. Данные по этим продуктам и процессам приведены в следующей таблице.

Продукт	Процесс 1	Процесс 2	Процесс 3	Доход (на единицу продукта)
	(количество минут выполнения процесса на единицу продукта)			
1	10	6	8	\$2
2	5	20	10	\$3

a) Найдите структуру оптимального производства.

b) Предположим, что появилась возможность увеличить время на выполнение одного из трех процессов. Для выбора процесса, время которого будет увеличено, создайте логически обоснованные приоритеты процессов.

4. Компания Show&Sell имеет возможность рекламировать свою продукцию по местному радио и телевидению. Бюджет на рекламу ограничен суммой \$10 000 в месяц. Одна минута рекламного времени на радио стоит \$15, а на телевидении — \$300. Компания предполагает, что реклама на радио по времени должна превышать рекламу на телевидении не менее чем в два раза. Вместе с тем известно, что нерационально использовать более 400 минут рекламы на радио в месяц. Последние исследования показали, что реклама на телевидении в 25 раз эффективнее рекламы на радио.

a) Разработайте оптимальный бюджет для рекламы на радио и телевидении.

b) Определите стоимость единицы месячного лимита на рекламу по радио.

c) Примените метод нахождения стоимости единицы ресурса для определения возможной эффективности рекламной кампании при увеличении ежемесячного бюджета на рекламу до \$15 000.

5. Корпорация Wyoming Electric является собственником электрогенерирующей станции. Поскольку эта корпорация имеет богатые запасы угля, на электростанции для генерации электрического тока используется уголь. Агентство по защите окружающей среды установило следующие ограничения: концентрация выбрасываемого в воздух сернистого газа не должна превышать 0.002, количество выбрасываемых аэрозольных частиц не должно превышать 20 фунтов в час. Корпорация для генерации электрического тока использует пылевидный уголь двух сортов, C1 и C2. Перед сжиганием эти сорта угля обычно смешиваются. Для простоты предположим, что сернистая составляющая в смеси углей определяется как средневзвешенное от доли угля каждого сорта в смеси. Характеристики используемых сортов угля приведены в следующей таблице.

Сорт угля	Концентрация серы	Количество выделяемых аэрозольных частиц (фунт/час)	Генерируемая мощность (фунт/час)
C1	0.0018	2.1	12 000
C2	0.0021	0.9	9 000

a) Найдите оптимальную смесь углей обоих сортов.

b) На сколько изменится количество генерируемой энергии (в час), если ослабить на 1 фунт в час ограничение на количество выбрасываемых аэрозольных частиц?

6. Факультет послевузовского обучения местного колледжа города Озарк предлагает в общей сложности до 30 курсов каждый семестр. Все курсы условно можно разбить на два типа: практические, такие как деревообработка, обучение работе на компьютере, ремонт и поддержка автомобилей и т.п.; и гуманитарные, например история, музыка и изобразительное искусство. Чтобы удовлетворить запросы обучающихся, в каждом семестре должно предлагаться не менее 10 курсов каждого типа. Факультет оценивает доход от одного практического курса в $1500, а гуманитарного — в $1000.

a) Какова оптимальная структура курсов для факультета?

b) Определите, какой будет иметь доход факультет при увеличении минимального количества практических курсов на единицу.

c) Определите доход факультета при увеличении минимального количества гуманитарных курсов на единицу.

7. Швейная фабрика Burroughs производит мужские сорочки и женские блузки для магазина Walmark. Этот магазин принимает всю продукцию, вырабатываемую фабрикой Burroughs. Производство швейного изделия состоит из раскроя, пошива и пакетирования готового изделия. На участке раскроя работают 25 человек, непосредственно на пошиве изделий — 35 человек и пакетируют готовые изделия 5 человек. Швейная фабрика Burroughs работает в одну смену (8 часов) пять дней в неделю. Трудозатраты на выпускаемые фабрикой изделия и доход от них показаны в следующей таблице.

Изделие	Раскрой	Пошив	Пакетирование	Доход
		(минуты на изделие)		(в $ на изделие)
Рубашка	20	70	12	2.50
Блузка	60	60	4	3.20

a) Определите оптимальную структуру еженедельного производства для этой швейной фабрики.

b) Если магазину Walmark потребуется не менее 2000 рубашек и 3000 блузок в неделю, то сможет ли швейная фабрика выполнить этот заказ при 5-дневной рабочей неделе? Если нет, то какой выход из этой ситуации вы можете предложить и какое оптимальное решение возможно в этом случае?

c) Определите стоимость одного часа рабочего времени, затрачиваемого отдельно на раскрой, пошив и пакетирование.

d) Предположим, что можно организовать сверхурочную работу на участках раскроя и пошива. Какую максимальную почасовую добавку за сверхурочные может предложить швейная фабрика?

8. Завод бытовой химии производит два вида чистящих средств, A и B, используя при этом сырье I и II. Обработка одной единицы сырья I стоит \$8, в результате производится 0.5 единицы средства A и столько же средства B. Обработка одной единицы сырья II стоит \$5, в результате получается 0.6 единицы средства A и 0.4 единицы средства B. Ежедневное производство средства A должно быть не менее 10 и не более 15 единиц. Для производства средства B аналогичные ограничения составляют 12 и 20 единиц.

a) Найдите оптимальную структуру выпуска чистящих средств.

b) Определите стоимость единицы изменения граничных значений ежедневного выпуска средств A и B.

9. Конвейер состоит из трех последовательных линий для сборки двух видов радиоприемников: HiFi-1 и HiFi-2. Время, необходимое для сборки одного радиоприемника на каждой линии, приведено в следующей таблице.

	Количество минут, затрачиваемых на сборку одного изделия	
Сборочная линия	HiFi-1	HiFi-2
1	6	4
2	5	5
3	4	6

Ежедневные профилактические работы на соответствующих линиях составляют 10%, 14% и 12% от всего рабочего времени, которое для любой линии не превышает 480 минут в смену.

a) Компания желает иметь такую структуру выпускаемой продукции, чтобы минимизировать время простоя всех трех линий.

b) Определите стоимость одного процента уменьшения времени профилактических работ для каждой линии.

2.5. Компьютерное решение задач ЛП

Графический метод, описанный в разделе 2.3, основывается на некоторых фундаментальных свойствах оптимального решения задач ЛП. Но на практике, где типичная задача линейного программирования содержит сотни и даже тысячи переменных и ограничений, графический метод неприменим. Здесь требуется использование компьютерных программ.

В этом разделе мы рассмотрим решение задач ЛП с помощью программы TORA. Хотя эта программа также имеет некоторые ограничения на размер решаемых задач, ее возможностей вполне достаточно для решения большинства практических задач.

Пример 2.5–1

На рис. 2.7 показана выходная распечатка решения задачи ЛП для компании Reddy Mikks (пример 2.2–1), выполненная программой TORA. На примере этой неоднократно исследованной задачи проиллюстрируем решение, полученное с помощью программы TORA.

Выходные результаты программы разбиты на два основных раздела: OPTIMUM SOLUTION SUMMARY (Итоговая сводка оптимального решения) и SENSITIVITY ANALYSIS (Анализ чувствительности). В первом разделе представлены оптимальные значения переменных (первая таблица) и значение целевой функции (Objective value). В рассматриваемом примере оптимальное значение переменной x_1 (количество выпускаемой краски для наружных работ) равно 3 т, а переменной x_2 (количество выпускаемой краски для внутренних работ) — 1.5 т.[1] Соответственно, доход составляет $21 000. В следующей таблице данного раздела представлены ограничения (Constraint). В этой таблице показаны значения дополнительных переменных, остаточных (Slack(-)) и избыточных (Surplus(+)), и значения правых частей неравенств (столбец RHS[2]). В данной задаче нет ограничений типа "≥", поэтому все дополнительные переменные являются остаточными. Из таблицы видно, что значения дополнительных переменных для первых двух неравенств равны нулю. Это означает, что сырье M1 и M2 потребляется полностью, без остатка. В третьем и четвертом ограничениях значения дополнительных переменных отличны от нуля, т.е. неравенства этих ограничений выполняются строго.

В верхней части раздела SENSITIVITY ANALYSIS в двух первых таблицах показаны результаты анализа чувствительности при изменении *по отдельности* (single changes) коэффициентов целевой функции (таблица Objective coefficients) и значений правых частей неравенств ограничений (таблица Right-hand Side). Например, текущее оптимальное решение сохраняется до тех пор, пока доход на тонну краски для внешних работ составляет от $2000 до $6000. Текущее оптимальное решение также будет сохранено, если ежедневные поступления сырья M1 будут находиться в пределах от 20 до 36 тонн. Такой же результат получен графическим способом в примере 2.4–2.

Вариации коэффициентов целевой функции в определенных пределах (точнее, в тех пределах, которые не приводят к перемещению точки оптимального решения) не вызывают изменения оптимальных значений переменных x_1 и x_2, а изменяют только значение самой целевой функции (см. раздел 2.4.1). Но изменение значений правых частей неравенств ограничений приводит к изменению как оптимальных значений переменных x_1 и x_2, так и значения целевой функции (более подробно об этом сказано в разделе 2.4.2). Эффективные методы вычисления нового оптимального решения будут описаны в разделе 4.7.

---

[1] В итоговой таблице слова EXT и INT, стоящие рядом с обозначениями переменных, являются просто описанием этих переменных: EXT обозначает "exterior paint" (краска для наружных работ), а INT — "interior paint" (краска для внутренних работ). — *Прим. перев.*

[2] RHS — сокращение от Right Hand Side, т.е. "правая сторона". — *Прим. перев.*

```
*** OPTIMUM SOLUTION SUMMARY ***
------------------------------------------------------------------------
Title  Reddy Mikks model
Final iteration No. 3
Objective value (max) =   21.0000
--------------------------------------------------------------------
Variable          Value      Obj Coeff   Obj Val Contrib
--------------------------------------------------------------------
x1 EXT            3.0000     5.0000        15.0000
x2 INT            1.5000     4.0000         6.0000
--------------------------------------------------------------------
Constraint         RHS          Slack(-)/Surplus(+)
--------------------------------------------------------------------
1 (<)           24.0000         0.0000-
2 (<)            6.0000         0.0000-
3 (<)            1.0000         2.5000-
4 (<)            2.0000         0.5000-

                   *** SENSITIVITY ANALYSIS ***
Objective coefficients -- Single Changes.
--------------------------------------------------------------------------------
Variable    Current Coeff   Min Coeff     Max Coeff    Reduced Cost
--------------------------------------------------------------------------------
x1 EXT          5.0000        2.0000         6.0000          0.0000
x2 INT          4.0000        3.3333        10.0000          0.0000

Right-hand Side -- Single Changes:
--------------------------------------------------------------------------------
Constraint  Current RHS     Min RHS       Max RHS      Dual Price
--------------------------------------------------------------------------------
1 (<)         24.0000       20.0000      36.0000        0.7500
2 (<)          6.0000        4.0000       6.6667        0.5000
3 (<)          1.0000       -1.5000      infinity       0.0000
4 (<)          2.0000        1.5000      infinity       0.0000

Objective Coefficients -- Simultaneous Changes d.
----------------------------------------------------------------------------------
Nonbasic Var     Optimality Condition
----------------------------------------------------------------------------------
sx3              0.7500 +     0.2500 d1 +    -0.1250 d2 >= 0
sx4              0.5000 +    -0.5000 d1 +     0.7500 d2 >= 0

Right-hand Side Ranging -- Simultaneous Changes D:
------------------------------------------------------------------------------------
Basic Var    Value/Feasibilty Condition
------------------------------------------------------------------------------------
x1 EXT     3.0000 +     0.2500 D1 +    -0.5000 D2 >= 0
x2 INT     1.5000 +    -0.1250 D1 +     0.7500 D2 >= 0
sx5        2.5000 +     0.3750 D1 +    -1.2500 D2 +     1.0000 D3  >= 0
sx6        0.5000 +     0.1250 D1 +    -0.7500 D2 +     1.0000 D4  >= 0
```

Рис. 2.7

Теперь рассмотрим информацию, приведенную в столбцах Reduced Cost (Приведенная стоимость) и Dual Price (Двойственная цена) первых двух таблиц раздела SENSITIVITY ANALYSIS. Начнем с приведенной стоимости.

С точки зрения экономиста переменные в модели ЛП можно рассматривать как числовые характеристики интенсивности определенных видов деятельности (или процессов), в результате которых потребляются ресурсы ("вход" модели) в целях получения прибыли ("выход" модели). Из этой интерпретации вытекает следующее определение.

$$\begin{pmatrix}\text{Приведенная стоимость на}\\ \text{единицу интенсивности}\\ \text{процесса } j\end{pmatrix} = \begin{pmatrix}\text{Стоимость потребленных}\\ \text{ресурсов на единицу}\\ \text{интенсивности процесса } j\end{pmatrix} - \begin{pmatrix}\text{Доход на единицу}\\ \text{интенсивности}\\ \text{процесса } j\end{pmatrix}$$

Если приведенная стоимость какого-либо процесса положительна, то отсюда следует, что стоимость потребленных ресурсов больше возможного дохода (все на единицу интенсивности процесса), поэтому такой процесс экономически не приемлем. Это обуславливает нулевое значение соответствующей переменной в оптимальном решении. Если же экономически привлекательный процесс имеет нулевую приведенную стоимость, то это означает, что достигнута точка равновесия, когда "выход" (единичный доход) равен "входу" (единичной стоимости ресурсов). В оптимальном решении, показанном на рис. 2.7, обе переменные x_1 и x_2 имеют положительные значения и нулевые приведенные стоимости.

Теперь рассмотрим определение **двойственной цены**. Двойственная цена — это просто еще одно название стоимости единицы ресурсов, определенной в разделе 2.4.2. Она равна вкладу, который привносит в значение целевой функции изменение на одну единицу лимита, определяющего доступность ресурса. Термин "двойственная цена" произошел от названия "проблема двойственности" из теории линейного программирования (см. главу 4). Для обозначения стоимости единицы ресурсов такие термины, как *теневая цена* (shadow price) и *симплексный мультипликатор* (simplex multiplier), применяются значительно реже. Хотя название "стоимость единицы ресурсов", введенное в разделе 2.4.2, лучше отображает смысл, вкладываемый в это понятие, мы все же будем использовать термин "двойственная цена" (dual price), так как он применяется во всех коммерческих программах решения задач ЛП.

На рис. 2.7 видно, что двойственные цены для сырья M1 и M2 равны соответственно 0.75 и 0.5 (т.е. \$750 и \$500) за тонну. Такие же результаты получены графическим способом в примере 2.4–2 и справедливы только при условии, что $20 \le M_1 \le 36$ и $4 \le M_2 \le 20/3$ (здесь через M_1 и M_2 обозначено количество материалов M1 и M2 соответственно). Отсюда следует, что если, например, доступное количество сырья M1 возрастет от текущего уровня в 24 т до 28, тогда оптимальное значение целевой функции увеличится на $\$750 \times (28 - 24) = \$3000$.

Двойственные цены для третьего и четвертого ограничений в данной модели равны нулю при изменении значений правых частей этих неравенств от -1.5 до ∞ и от 1.5 до ∞ соответственно. Это означает, что ограничения, накладываемые рынком на структуру производства (ограничение на производство краски для внутренних работ до 2 т и т.п.), в данной ситуации не оказывают влияния на оптимальное решение.

В оставшейся части раздела SENSITIVITY ANALYSIS приведены результаты анализа влияния на оптимальное решение *одновременного* изменения (simultaneous changes) всех коэффициентов целевой функции (таблица Objective Coefficients – Simultaneous Changes) и *одновременного* изменения значений правых частей всех неравенств (таблица Right-hand Side Ranging – Simultaneous Changes). Чтобы отследить результат одновременного изменения коэффициентов целевой функции, в данной модели представим эту функцию как $z = (5 + d_1)x_1 + (4 + d_2)x_2$. Тогда условие неотрицательности, накладываемое на дополнительные переменные sx_3 и sx_4 первого и второго ограничений, будет записано в виде следующих неравенств.

$$0.75 + 0.25d_1 - 0.125d_2 \geq 0$$
$$0.50 - 0.50d_1 + 0.75d_2 \geq 0$$

Эти неравенства эквивалентны системе неравенств $\frac{1}{2} \leq \frac{c_1}{c_2} \leq \frac{6}{4}$, которую мы получили в разделе 2.4.1 графическим способом. Для того чтобы убедиться в эквивалентности этих неравенств, подставьте в последние из них выражения $c_1 = 5 + d_1$ и $c_2 = 4 + d_2$ и выполните необходимые действия (проверьте!). Отметим, что при вариации d_1 и d_2 в пределах, не нарушающих приведенных неравенств, изменяется только оптимальное значение целевой функции; значения переменных остаются неизменными. Если измененные значения d_1 и d_2 не удовлетворяют неравенствам, то необходимо искать новое оптимальное решение (используя процедуры, описанные в разделе 4.7).

Теперь рассмотрим одновременное изменение правых частей неравенств ограничений. В данной модели представим правые части неравенств как $24 + D_1$, $6 + D_2$, $1 + D_3$ и $2 + D_4$. Тогда оптимальные значения переменных можно выразить через новые переменные D_1, D_2, D_3 и D_4 (соответствующие формулы показаны в таблице Right-hand Side Ranging — Simultaneous Changes на рис. 2.7). Для того чтобы решение было осуществимым, необходимо выполнение условия неотрицательности для значений переменных в оптимальном решении — тем самым накладываются ограничения на новые переменные D_1, D_2, D_3 и D_4. Например, положим $D_1 = 5$, $D_2 = -1$, $D_3 = 1$ и $D_4 = 2$. Тогда

$$x_1 = 3 + 0.25 \times 5 - 0.5 \times (-1) = 4.75,$$
$$x_2 = 1.5 - 0.125 \times 5 + 0.75 \times (-1) = 0.125,$$
$$sx_5 = 2.5 + 0.375 \times 5 - 1.25 \times (-1) + 1 \times 1 = 6.625,$$
$$sx_6 = 0.5 + 0.125 \times 5 - 0.75 \times (-1) + 1 \times 2 = 3.875.$$

Здесь значения всех переменных положительные. Следовательно, имеем оптимальное решение $x_1 = 4.75$ т, $x_2 = 0.125$ т и $z = 5 \times 4.75 + 4 \times 0.125 = 24.25$ (т.е. \$24 250). Полученное значение z больше текущего оптимального на величину $\Delta z = \$24\,250 - \$21\,000 = \$3\,250$. Эту разность можно подсчитать также по формуле

$$\Delta z = D_1y_1 + D_2y_2 + D_3y_3 + D_4y_4,$$

где y_1, y_2, y_3 и y_4 — двойственные цены соответствующих ресурсов. На основании последней формулы получаем

$$\Delta z = 5 \times \$750 + (-1) \times \$500 + 1 \times 0 + 2 \times 0 = \$3\,250.$$

Этот способ вычислений применим только тогда, когда изменения D_i, $i = 1, 2, 3, 4$, приводят к допустимому решению. Если в результате изменений правых частей ограничений значение какой-нибудь переменной будет отрицательным, то это означает, что необходимо искать новую точку оптимального решения, как показано в разделе 4.7.

Упражнения 2.5,а

1. На рис. 2.8 показаны выходные результаты, полученные с помощью программы TORA для задачи "диеты" из примера 2.3–2.

   a) Интерпретируйте двойственную цену первого ограничения.

   b) Используя результаты, показанные на рис. 2.8, найдите новое оптимальное решение, если необходимо ежедневно производить не менее 900 фунтов пищевой добавки.

   c) Если стоимость кукурузной и соевой муки изменится соответственно до \$0.40 и \$1.05 за один фунт, то будет ли текущее решение оптимальным?

   d) На основе графического решения данной задачи получите условия оптимальности решения, приведенные на рис. 2.8, для одновременного изменения коэффициентов целевой функции.

2. Задача ЛП из упр. 2.4,а(7) формулируется следующим образом.

   $$\text{Максимизировать } z = 18x_1 + 9x_2$$

   при выполнении условий

   $$24x_1 + 8x_2 \le 60\,000,$$
   $$x_1 \le 2000,\ x_2 \le 6000,$$
   $$x_1, x_2 \ge 0,$$

   где x_1 и x_2 — количество производимых консервным заводом упаковок томатного сока и томатной пасты соответственно. Решение этой задачи приведено на рис. 2.9.

   a) Определите единичную стоимость дополнительного фунта помидоров.

   b) Изменится ли оптимальное решение, если склад сможет принимать больше упаковок томатного сока? А при увеличении емкости склада для упаковок томатной пасты?

   c) Найдите новое оптимальное решение, если компания уменьшит объем перерабатываемых помидоров до 50 000 фунтов за смену.

   d) Останется ли текущее решение оптимальным, если цена одной упаковки томатного сока уменьшится до \$15, а цена одной упаковки томатной пасты возрастет до \$10?

   e) Найдите новое оптимальное решение, если завод будет перерабатывать за смену 80 000 фунтов помидоров, а склад сможет принять до 7 000 упаковок томатной пасты.

   f) Если количество перерабатываемых заводом помидоров уменьшится до 30 000 фунтов, а емкость складских помещений увеличится до 8 000 упаковок томатной пасты, то можно ли будет найти новое оптимальное решение на основании данных, приведенных на рис. 2.9?

*** OPTIMUM SOLUTION SUMMARY ***

```
--------------------------------------------------------------------------------
Title: Diet problem
Final iteration No: 3
Objective value (min) = 437.6471
----------------------------------------------------------------
Variable          Value      Obj Coeff   Obj Val Contrib
----------------------------------------------------------------
x1 Corn          470.5882     0.3000       141.1765
x2 Soy           329.4118     0.9000       296.4706
----------------------------------------------------------------
Constraint         RHS          Slack(-)/Surplus(+)
----------------------------------------------------------------
1 (>)           800.0000         0.0000+
2 (<)             0.0000         0.0000-
3 (>)             0.0000        10.823+

                  *** SENSITIVITY ANALYSIS ***
Objective coefficients -- Single Changes:
--------------------------------------------------------------------------------
Variable    Current Coeff   Min Coeff     Max Coeff    Reduced Cost
--------------------------------------------------------------------------------
x1 Corn        0.3000       -0.6300         0.9000        0.0000
x2 Soy         0.9000        0.3000         infinity      0.0000

Right-hand Side -- Single Changes:
--------------------------------------------------------------------------------
Constraint    Current RHS       Min RHS      Max RHS      Dual Price
--------------------------------------------------------------------------------
1 (>)          800.0000          0.0000      infinity       0.5471
2 (<)            0.0000       -138.0000      168.0000      -1.1765
3 (>)            0.0000        -infinity      10.8235       0.0000

Objective Coefficients -- Simultaneous Changes d:
--------------------------------------------------------------------------------
Nonbasic Var  Optimality Condition
--------------------------------------------------------------------------------
Sx3        -0.5471 +    -0.5882 d1 +    -0.4118 d2 <= 0
sx4        -1.1765 +     1.9608 d1 +    -1.9608 d2 <= 0

Right-hand Side Ranging -- Simultaneous Changes D:
--------------------------------------------------------------------------------
Basic Var   Value/Feasibilty Condition
--------------------------------------------------------------------------------
x1 Corn    470.5882 +    0.5882 D1 +     1.9608 D2 >= 0
x2 Soy     329.4118 +    0.4118 D1 +    -1.9608 D2 >= 0
Sx5         10.8235 +    0.0135 D1 +     0.0784 D2 +    -1.0000 D3  >= 0
```

Рис. 2.8

*** OPTIMUM SOLUTION SUMMARY ***

Title: Problem 2.5a-2
Final iteration No: 4
Objective value (max) =63000.0000

Variable	Value	Obj Coeff	Obj Val Contrib
x1 Juice	500.0000	18.0000	9000.0000
x2 Paste	6000.0000	9.0000	54000.0000

Constraint	RHS	Slack(-)/Surplus(+)
1 (<)	60000.0000	0.0000-
2 (<)	2000.0000	1500.0000-
3 (<)	6000.0000	0.0000-

*** SENSITIVITY ANALYSIS ***

Objective coefficients -- Single Changes:

Variable	Current Coeff	Min Coeff	Max Coeff	Reduced Cost
x1 Juice	18.0000	0.0000	27.0000	0.0000
x2 Paste	9.0000	6.0000	infinity	0.0000

Right-hand Side -- Single Changes:

Constraint	Current RHS	Min RHS	Max RHS	Dual Price
1 (<)	60000.0000	48000.0000	96000.0000	0.7500
2 (<)	2000.0000	500.0000	infinity	0.0000
3 (<)	6000.0000	1500.0000	7500.0000	3.0000

Objective Coefficients -- Simultaneous Changes d:

Nonbasic Var	Optimality Condition
sx3	0.7500 + 0.0417 d1 >= 0
sx5	3.0000 + 1.0000 d2 + -0.3333 d1 >= 0

Right-hand Side Ranging -- Simultaneous Changes D:

Basic Var	Value/Feasibilty Condition
x2 Paste	6000.0000 + 1.0000 D3 >= 0
x1 Juice	500.0000 + 0.0417 D1 + -0.3333 D3 >= 0
sx4	1500.0000 + -0.0417 D1 + 1.0000 D2 + 0.3333 D3 >= 0

Рис. 2.9

2.6. Примеры моделей ЛП

В этом разделе исследовано несколько реалистических моделей ЛП, в которых определение переменных и построение целевой функции и ограничений не так однозначно, как в ранее рассмотренных моделях с двумя переменными. Использование программы TORA позволит интерпретировать результаты решения этих задач ЛП.

Пример 2.6–1. (Кредитная политика банка)

Банк Thriftem, предоставляющий полный набор банковских услуг, находится в процессе формирования портфеля кредитов объемом 12 миллионов долларов. В следующей таблице представлены возможные типы банковских кредитов.

Тип кредита	Ставка процента	Вероятность безнадежных долгов
Кредиты физическим лицам	0.140	0.10
Кредиты на покупку автомобилей	0.130	0.07
Кредиты на покупку жилья	0.120	0.03
Сельскохозяйственные	0.125	0.05
Коммерческие	0.100	0.02

Безнадежные долги считаются невозвратимыми, поэтому они должны вычитаться из возможного дохода.

Конкурентная борьба с другими финансовыми институтами вынуждает банк не менее 40% капитала помещать в сельскохозяйственные и коммерческие кредиты. Для содействия строительной индустрии своего региона банк планирует вложить в кредиты на покупку жилья не менее 50% от общей суммы кредитов физических лиц, на покупку автомобилей и жилья. Банк также поддерживает государственную политику, указывающую, что отношение безнадежных долгов ко всей сумме кредитов не должно превышать 0.04.

Математическая модель

Переменные для создаваемой модели можно определить следующим образом:

x_1 — кредиты физическим лицам,
x_2 — кредиты на покупку автомобилей,
x_3 — кредиты на покупку жилья,
x_4 — сельскохозяйственные кредиты,
x_5 — коммерческие кредиты.

Банк Thriftem, естественно, желает максимизировать чистую прибыль, т.е. разность между доходом от инвестируемых сумм и суммой невозвращенных кредитов. Поскольку безнадежные долги считаются невозвратимыми, они вычитаются как из инвестируемых сумм, так и общей прибыли. Исходя из этих соображений, целевую функцию можно записать следующим образом.

Максимизировать $z = 0.14\times(0.9x_1) + 0.13\times(0.93x_2) + 0.12\times(0.97x_3) + 0.125\times(0.95x_4) + 0.1\times(0.98x_5) - 0.1x_1 - 0.07x_2 - 0.03x_3 - 0.05x_4 - 0.02x_5$.

После приведения подобных членов получаем

Максимизировать $z = 0.026x_1 + 0.0509x_2 + 0.0864x_3 + 0.06875x_4 + 0.078x_5$.

Задача имеет пять ограничений.

1. *Ограничение общей суммы кредитов*

$$x_1 + x_2 + x_3 + x_4 + x_5 \leq 12.$$

2. *Ограничение на сельскохозяйственные и коммерческие кредиты*

$$x_4 + x_5 \geq 0.4 \times 12$$

или

$$x_4 + x_5 \geq 4.8.$$

3. *Ограничения кредитов на покупку жилья*

$$x_3 \geq 0.5(x_1 + x_2 + x_3).$$

4. *Ограничения на невозвращенные кредиты*

$$\frac{0.1x_1 + 0.07x_2 + 0.03x_3 + 0.05x_4 + 0.02x_5}{x_1 + x_2 + x_3 + x_4 + x_5} \leq 0.04$$

или

$$0.06x_1 + 0.03x_2 - 0.01x_3 + 0.01x_4 - 0.02x_5 \leq 0.$$

5. *Условия неотрицательности*

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0, x_5 \geq 0.$$

Необходимо сделать еще одно "тонкое" замечание, что все кредиты выделяются примерно в одно и то же время. Это позволит игнорировать временной фактор в процессе размещения капитала в различные кредиты.

Оптимальное решение сформулированной задачи линейного программирования показано на рис. 2.10. Оно рекомендует использовать только кредиты на покупку жилья и коммерческие кредиты. Среди неиспользованных типов кредитов наименее привлекательны кредиты физическим лицам и не только из-за того, что коэффициент при переменной x_1 в целевой функции наименьший (равен 0.026), но и потому, что приведенная стоимость таких кредитов наибольшая (равна 0.0604) среди всех остальных. Приведенная стоимость показывает, что "рентабельность" этой переменной должна возрасти до 0.0604, чтобы кредиты для физических лиц стали привлекательными для инвестиций. Двойственная цена первого ограничения показывает, что увеличение суммы всех кредитов на единицу (миллион долларов) приводит к увеличению чистой прибыли на 0.0864 (миллиона долларов). Это эквивалентно 8.64% годовых от суммы инвестиций. Поскольку соответствующий интервал для значения правой части этого ограничения простирается от 4.8 до бесконечности, отсюда следует, что указанный процент годовых гарантирован для любой общей суммы кредитов, превышающих 12 миллионов долларов. Но этот процент годовых, конечно, очень мал, поскольку наименьший процент по банковским вложениям составляет 10% (для коммерческих кредитов). Разность в величинах этих процентов обуславливают невозвращенные кредиты, которые вычитаются и из общей суммы кредитов, и чистой прибыли. В формуле целевой функции наибольший коэффициент (0.0864) стоит перед переменной, соответствующей объему кредитов на покупку жилья. Интересно, что в данном решении ему оказалась равной двойственная цена первого ограничения. Отсюда можно заключить, что любые новые вложения следует размещать в виде кредитов на покупку жилья.

*** OPTIMUM SOLUTION SUMMARY ***

Title. Bank Model
Final iteration No: 5
Objective value (max) = 0.9965

Variable	Value	Obj Coeff	Obj Val Contrib
x1 pers'nl	0.0000	0.0260	0.0000
x2 car	0.0000	0.0509	0.0000
x3 home	7 2000	0.0864	0.6221
x4 farm	0.0000	0.0688	0.0000
x5 com'l	4.8000	0.0780	0.3744

Constraint	RHS	Slack(-)/Surplus(+)
1 (<)	12.0000	0.0000-
2 (>)	4.8000	0.0000+
3 (<)	0.0000	3.6000-
4 (<)	0.0000	0.1680-

*** SENSITIVITY ANALYSIS ***

Objective coefficients -- Single Changes:

Variable	Current Coeff	Min Coeff	Max Coeff	Reduced Cost
x1 pers'nl	0.0260	-infinity	0.0864	0.0604
x2 car	0.0509	-infinity	0.0864	0.0355
x3 home	0.0864	0.0780	infinity	0.0000
x4 farm	0.0688	-infinity	0.0780	0.0092
x5 com'l	0.0780	0.0688	0.0864	0.0000

Right-hand Side -- Single Changes:

Constraint	Current RHS	Min RHS	Max RHS	Dual Price
1 (<)	12.0000	4.8000	infinity	0.0864
2 (>)	4.8000	0.0000	12.0000	-0.0084
3 (<)	0.0000	-3.6000	infinity	0.0000
4 (<)	0.0000	-0.1680	infinity	0.0000

Objective Coefficients -- Simultaneous Changes d:

Nonbasic Var	Optimality Condition
x1 pers'nl	0.0604 + 1.0000 d3 - d1 >= 0
x2 car	0.0355 + 1.0000 d3 - d2 >= 0
x4 farm	0.0092 + 1.0000 d5 - d4 >= 0
Sx6	0.0084 + 1.0000 d3 + -1.0000 d5 >= 0
sx7	0.0864 + 1.0000 d3 >= 0

Right-hand Side Ranging -- Simultaneous Changes D:

Basic Var	Value/Feasibilty Condition
x3 home	7.2000 + 1.0000 D1 + -1.0000 D2 >= 0
x5 com'l	4.8000 + 1.0000 D2 >= 0
sx9	3.6000 + 0.5000 D1 + -0.5000 D2 + 1.0000 D3 >= 0
sx10	0.1680 + 0.0100 D1 + 0.0100 D2 + 1.0000 D4 >= 0

Рис. 2.10

Второе ограничение данной задачи указывает нижнюю границу суммы сельскохозяйственных и коммерческих кредитов. Отрицательная двойственная цена (–0.0084) показывает, что увеличение этой границы приведет к уменьшению чистой прибыли банка. Другими словами, экономически не выгодно устанавливать нижнюю границу для суммы сельскохозяйственных и коммерческих кредитов. Это еще раз подтверждает, как было видно при рассмотрении первого ограничения, что любые дополнительные вложения следует помещать в кредиты на покупку жилья, а не в сельскохозяйственные и коммерческие кредиты. Если мы вообще удалим это ограничение, тогда все инвестиции переместятся в кредиты на покупку жилья. Проверьте это утверждение, удалив второе ограничение и решив новую задачу с помощью программы TORA (для этого достаточно использовать опцию MODIFY (Изменить) при решении текущей задачи).

Пример 2.6–2. (Освоение и использование земли)

Компания Birdeyes[1] владеет 800 акрами необработанной земли с живописным озером в центре горного массива Озарк. В прошлом освоение территории вокруг озера происходило нерегулярно и понемногу. Сейчас вокруг озера построено много разбросанных домиков для отдыхающих. Поскольку канализация отсутствует, интенсивно используются септические танки (большие емкости для хранения бытовых отходов). В настоящее время многие септические танки приходят в негодность (текут), поэтому появились проблемы с загрязнением воды в озере.

Чтобы предотвратить дальнейшее загрязнение озера, официальные власти округа разработали строгие ограничительные требования для возможного дальнейшего освоения этой территории.

1. Разрешается возводить домики только на одну, две или три семьи, причем доля домиков на одну семью должна составлять не менее 50% от всех построенных.
2. Ограничивается количество септических танков: минимальная территория, отведенная для размещения одного танка, обслуживающего один домик на одну, две или три семьи, составляет соответственно 2, 3 и 4 акра.
3. Рекреационная зона рассчитывается исходя из нормы — не менее одного акра на 200 семей.
4. Для предотвращения загрязнения озера запрещается использовать поверхностные воды для полива и бытовых нужд.

Правление компании Birdeyes решило изучить возможности дальнейшего освоения тех 800 акров земли, которые ей принадлежат. Дальнейшее освоение этих земель предусматривает строительство домиков для отдыха на одну, две и три семьи. Подсчитано, что необходимо отвести не менее 15% всей территории на прокладку дорог и строительство вспомогательных сооружений. Компания планирует сдавать домики отдыхающим и рассчитывает получить от одного домика следующую прибыль.

---

[1] Название компании можно перевести как "Птичьи глазки". — *Прим. перев.*

Тип домика	на одну семью	на две семьи	на три семьи
Чистая прибыль ($)	10 000	12 000	15 000

Стоимость планируемой водопроводной сети пропорциональна количеству домиков, подключенных к водопроводу. Однако власти округа настаивают на том, что необходимо собрать не менее $100 000, чтобы этот проект был экономически осуществимым. Кроме того, известно, что даже в пиковый период водопроводная сеть не сможет поставить больше 200 000 галлонов воды в день. В следующей таблице приведена стоимость подключения к водопроводной сети одного домика и указана ежедневная потребность в воде одной средней семьи.

Тип домика	на одну семью	на две семьи	на три семьи	Рекреационная зона
Стоимость подключения к водопроводу ($)	1000	1200	1400	800
Потребность в воде (галлон/день)	400	600	840	450

Математическая модель

Компания должна определить, сколько строить домиков различного типа и какова должна быть рекреационная зона, чтобы удовлетворить требованиям властей округа. Положим

x_1 — количество домиков на одну семью,
x_2 — количество домиков на две семьи,
x_3 — количество домиков на три семьи,
x_4 — количество акров земли, отводимых под рекреационную зону.

Очевидно, что компания должна максимизировать свой доход, поэтому целевую функцию можно записать так:

$$\text{Максимизировать } z = 1000x_1 + 12000x_2 + 15000x_3.$$

Задача имеет следующие ограничения.

1. Ограничение на общую площадь используемой земли.
2. Ограничение, связанное со строительством домиков преимущественно на одну семью.
3. Ограничение, учитывающее требования к рекреационной зоне.
4. Ограничение, связанное со строительством водопроводной сети.
5. Ограничение на потребление воды в пиковый период.

Эти ограничения математически будут записаны следующим образом.

1. *Ограничение на общую площадь используемой земли*:

$$2x_1 + 3x_2 + 4x_3 + 1x_4 \leq 680 \ (= 0.85 \times 800).$$

2. *Ограничение, связанное со строительством домиков преимущественно на одну семью*:

$$\frac{x_1}{x_1+x_2+x_3} \geq 0.5$$

или

$$0.5x_1 - 0.5x_2 - 0.5x_3 \geq 0.$$

3. *Ограничение, учитывающее требования к рекреационной зоне*:

$$x_4 \geq \frac{x_1+2x_2+3x_3}{200}$$

или

$$200x_4 - x_1 - 2x_2 - 3x_3 \geq 0.$$

4. *Ограничение, связанное со строительством водопроводной сети*:

$$1000x_1 + 1200x_2 + 1400x_3 + 800x_4 \geq 100\,000.$$

5. *Ограничение на потребление воды в пиковый период*:

$$400x_1 + 600x_2 + 840x_3 + 450x_4 \leq 200\,000.$$

6. *Условия неотрицательности*:

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0.$$

Еще на этапе формулирования математической модели следует обращать внимание на возможные проблемы, которые могут возникнуть из-за ошибок машинного округления. В нашей модели коэффициенты в четвертом и пятом ограничениях значительно больше по величине, чем все остальные. Эта несоразмерность (относительная) величин коэффициентов может привести к непредвиденным ошибкам машинного округления и, следовательно, к неверному результату. В нашем случае мы можем предотвратить эту потенциальную опасность путем масштабирования коэффициентов, для чего все коэффициенты в четвертом и пятом неравенствах надо разделить на 1000. После этого неравенства будут записаны следующим образом.

$$x_1 + 1.2x_2 + 1.4x_3 + 0.8x_4 \geq 100,$$
$$0.4x_1 + 0.6x_2 + 0.84x_3 + 0.45x_4 \leq 200.$$

К подобным вычислительным проблемам может привести и наличие в неравенствах слишком малых по величине коэффициентов. В этом случае также применяется масштабирование коэффициентов, но уже в сторону их увеличения. Большинство программного обеспечения, предназначенного для решения задач ЛП (включая программу TORA), пытается согласовать величины коэффициентов еще до начала расчетов. Однако это желательно сделать еще на этапе формализации модели.

На рис. 2.11 представлено оптимальное решение для построенной модели. Отметим, что обычная задача линейного программирования не предполагает целочисленного решения. Поэтому мы получили значения $x_1 = 339.152$, $x_4 = 1.696$ и $x_2 = x_3 = 0$. Мы можем округлить полученные значения, тогда $x_1 = 339$ и $x_4 = 2$ (эти значения в данном случае совпадут с оптимальным решением задачи целочисленного линейного программирования).

*** OPTIMUM SOLUTION SUMMARY ***

Title: Land Development
Final iteration No: 6
Objective value (max) =3391521.0000

Variable	Value	Obj Coeff	Obj Val Contrib
x1 SINGLE	339.1521	10000.0000	3391521.0000
x2 DOUBLE	0.0000	12000.0000	0.0000
x3 TRIPLE	0.0000	15000.0000	0.0000
x4 RECR'N	1.6958	0.0000	0.0000

Constraint	RHS	Slack(-)/Surplus(+)
1 (<)	680.0000	0.0000-
2 (>)	0.0000	169.5760+
3 (>)	0.0000	0.0000+
4 (>)	100.0000	252.7182+
5 (<)	200.0000	63.5761-

*** SENSITIVITY ANALYSIS ***

Objective coefficients -- Single Changes:

Variable	Current Coeff	Min Coeff	Max Coeff	Reduced Cost
x1 SINGLE	10000.0000	7993.3551	infinity	0.0000
x2 DOUBLE	12000.0000	-infinity	15012.4690	3012.4690
x3 TRIPLE	15000.0000	-infinity	20024.9355	5024.9355
x4 RECR'N	0.0000	-2000000.0000	5000.0044	0.0000

Right-hand Side -- Single Changes:

Constraint	Current RHS	Min RHS	Max RHS	Dual Price
1 (<)	680.0000	192.7885	996.8926	4987.5308
2 (>)	0.0000	-infinity	169.5760	0.0000
3 (>)	0.0000	-340.0000	50988.0938	-24.9377
4 (>)	100.0000	-infinity	352.7182	0.0000
5 (<)	200.0000	136.4239	infinity	0.0000

Objective Coefficients -- Simultaneous Changes d:

Nonbasic Var	Optimality Condition		
x2 DOUBLE	3012.4690 +	1.5012 d1 +	-0.0025 d4 - d2 >= 0
x3 TRIPLE	5024.9355 +	2.0025 d1 +	-0.0050 d4 - d3 >= 0
Sx6	24.9377 +	0.0025 d1 +	-0.0050 d4 >= 0
sx8	4987.5308 +	0.4988 d1 +	0.0025 d4 >= 0

Right-hand Side Ranging -- Simultaneous Changes D:

Basic Var	Value/Feasibilty Condition			
Sx7	252.7182 +	0.5187 D1 +	0.0374 D3 +	-1.0000 D4 >= 0
x1 SINGLE	339.1521 +	0.4988 D1 +	-0.0025 D3 >= 0	
x4 RECR'N	1.6958 +	0.0025 D1 +	0.0050 D3 >= 0	
Sx5	169.5760 +	0.2494 D1 +	-1.0000 D2 +	-0.0012 D3 >= 0
sx12	63.5761 +	-0.2006 D1 +	-0.0012 D3 +	1.0000 D5 >= 0

Рис. 2.11

Оптимальное решение не рекомендует строить домики на две и три семьи, несмотря на то что доходность этих домиков ($12 000 и $15 000) выше, чем домиков, рассчитанных на одну семью. Этот результат показывает, что только по коэффициентам целевой функции нельзя судить о "рентабельности" отдельных видов экономической деятельности, которым соответствуют данные коэффициенты. Следует учитывать также стоимость ресурсов, необходимых для осуществления такой деятельности. Это тот фактор, который выражается *приведенной стоимостью*. Приведенные стоимости для домиков на две и три семьи составляют $3012.47 и $5024.94 соответственно. Это означает, для того чтобы такие домики были рентабельными, необходимо либо на величины приведенных стоимостей уменьшить стоимость ресурсов, необходимых для их строительства и эксплуатации, либо на такую же величину увеличить их доходность.

В ограничениях 2, 4 и 5 дополнительные переменные имеют положительные значения, а это указывает, что соответствующие "ресурсы" использованы не полностью. В результате *двойственные цены*, соответствующие этим ограничениям, равны нулю. Двойственная цена первого ограничения (ограничение на общую площадь используемой земли) равна $4987.53; это говорит о том, что увеличение общей площади на один акр должно принести $4987.53 чистой прибыли. Эту информацию можно использовать для определения цены при покупке дополнительного земельного участка.

В третьем ограничении двойственная цена равна –$24.938, поэтому любое увеличение этого "ресурса" приведет к снижению общего дохода. Но почему так происходит? Чтобы ответить на этот вопрос, нужно знать, сколько единиц этого "ресурса" содержится в данном ограничении. Рассмотрим подробнее данное ограничение.

$$200x_4 - x_1 - 2x_2 - 3x_3 \geq 0.$$

Это неравенство показывает, что минимальная площадь рекреационной зоны зависит от количества домиков. В такой записи единицы измерения левой части неравенства не видны. Поэтому мы разделим все неравенство на 200, в результате получим

$$x_4 - (0.005x_1 + 0.01x_2 + 0.015x_3) \geq 0.$$

Поскольку площадь рекреационной зоны измеряется в акрах, выражение в скобках также должно измеряться в акрах. Поэтому увеличение на одну единицу правой части неравенства (т.е. возрастание от 0 до 1) можно интерпретировать как увеличение на один акр площади рекреационной зоны. На основании такого представления ограничения можно сказать, что двойственная цена соответствует стоимости увеличения на акр рекреационной зоны. Но новая запись ограничения показывает, что двойственная цена должна быть равной $200 \times (-24.937) = -\$4987.53$. (Если повторить расчеты программы TORA при такой записи третьего ограничения, то должны получить именно такую величину двойственной цены — проверьте это!)

Новая двойственная цена свидетельствует о том, что увеличение площади рекреационной зоны на один акр приведет к уменьшению общего дохода на $4987.53. Интересно, что это число в точности равно двойственной цене первого ограничения, но с противоположным знаком. Такой результат имеет экономический смысл, поскольку перевод акра земли в рекреационную зону означает исключение этого акра земли из той площади, на которой можно построить домики, приносящие прибыль. Поэтому не стоит удивляться, что эти величины совпадают.

Пример 2.6–3. (Задача расписания движения автобусов)

Городская транспортная компания изучает возможность ввести такую систему движения городских автобусов, которая снизила бы проблему загазованности в городе путем уменьшения количества используемых автобусов. Вначале нужно было определить минимальное количество автобусов, необходимое для удовлетворения транспортных потребностей горожан. Оказалось, что в различное время суток требуется разное количество автобусов. Дальнейшее изучение этого вопроса позволило аппроксимировать суточную потребность в автобусах кусочно-постоянной функцией с 4-часовыми интервалами постоянных значений. Эта функция показана на рис. 2.12. При составлении расписания движения автобусов следует учитывать, что каждый автобус должен находиться на линии непрерывно в течение 8 часов (одна рабочая смена).

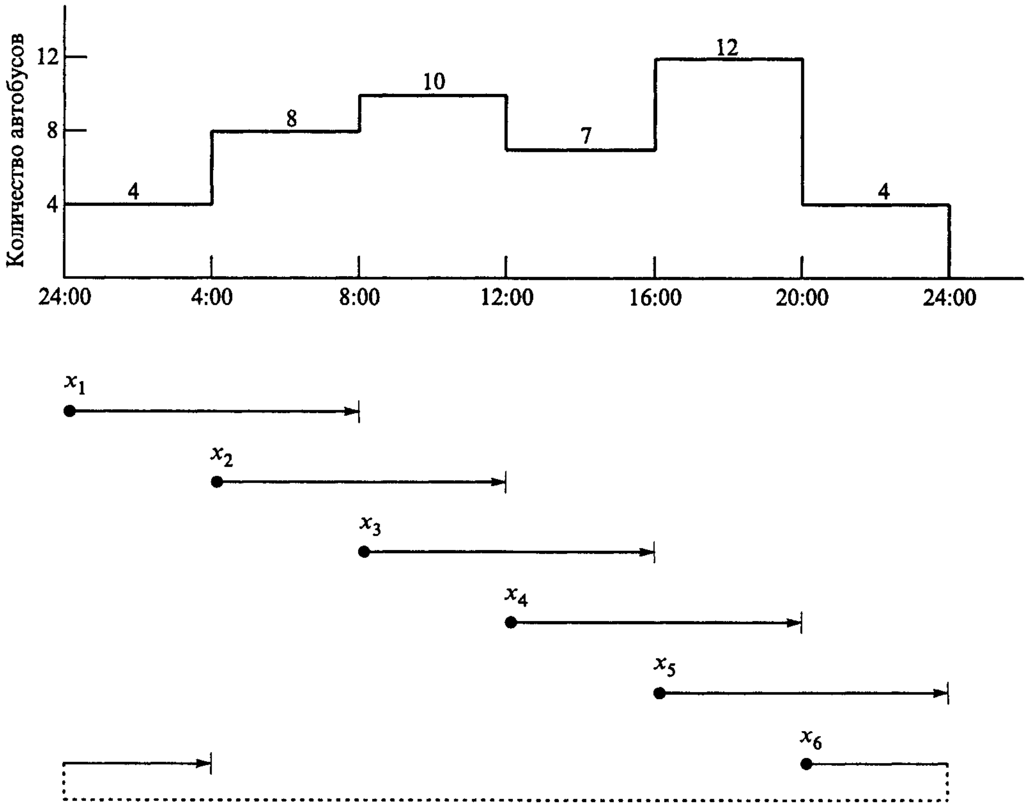


Рис. 2.12

Математическая интерпретация задачи

Итак, требуется определить *число автобусов, выходящих на линию в определенную смену*, (переменные) так, чтобы *удовлетворить минимальные потребности в транспортных услугах* (ограничения) и по возможности *минимизировать общее количество автобусов, выходящих на линию в течение суток* (целевая функция).

Нетрудно заметить, что такое определение переменных неоднозначно. Мы знаем, что каждый автобус должен отработать 8-часовую смену, но мы не знаем, когда эта смена должна начинаться. Если следовать схеме обычной 3-сменной работы (1-я смена с 8:01 до 16:00, 2-я — с 16:01 до 24:00 и 3-я — с 00:01 до 8:00) и обозначить через x_1, x_2 и x_3 количество автобусов, работающих на линии в эти смены, тогда, исходя из транспортных потребностей (рис. 2.12), получаем $x_1 \geq 10$, $x_2 \geq 12$ и $x_3 \geq 8$, а общее количество ежедневно используемых автобусов составляет $x_1 + x_2 + x_3 = 10 + 12 + 8 = 30$.

Это решение приемлемо, если смены *должны* начинаться так, как при обычной организации 3-сменной работы. Однако можно оптимизировать расписание движения автобусов, если поискать другое, "лучшее" время начала рабочих смен. Предположим, что между началом "соседних" смен может быть 4 часа, а не 8, как в обычной схеме 3-сменной работы. В нижней части рис. 2.12 схематично показана такая схема организации перекрывающихся рабочих смен, когда они должны начинаться в 00:01, 4:01, 8:01, 12:01, 16:01 и 20:01, при этом каждая смена продолжается 8 часов. Теперь мы готовы определить переменные:

x_1 — количество автобусов, начинающих работу в 00:01,
x_2 — количество автобусов, начинающих работу в 4:01,
x_3 — количество автобусов, начинающих работу в 8:01,
x_4 — количество автобусов, начинающих работу в 12:01,
x_5 — количество автобусов, начинающих работу в 16:01,
x_6 — количество автобусов, начинающих работу в 20:01.

Задача линейного программирования будет записана следующим образом.

$$\text{Минимизировать } z = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

при выполнении условий

$$\begin{aligned}
&x_1 + x_6 \geq 4 \text{ (время от 00:01 до 4:00)},\\
&x_1 + x_2 \geq 8 \text{ (время от 4:01 до 8:00)},\\
&x_2 + x_3 \geq 10 \text{ (время от 8:01 до 12:00)},\\
&x_3 + x_4 \geq 7 \text{ (время от 12:01 до 16:00)},\\
&x_4 + x_5 \geq 12 \text{ (время от 16:01 до 20:00)},\\
&x_5 + x_6 \geq 4 \text{ (время от 20:01 до 24:00)},\\
&x_j \geq 0, j = 1, 2, ..., 6.
\end{aligned}$$

Результаты решения этой задачи ЛП представлены на рис. 2.13. Отметим, что в оптимальном решении общее количество автобусов равно 26, при этом $x_1 = 4$, $x_2 = 10$, $x_3 = 0$, $x_4 = 8$, $x_5 = 4$ и $x_6 = 0$. Приведенные стоимости всех переменных равны нулю; это указывает на альтернативные оптимальные решения (но с тем же значением целевой функции). Интерес представляет информация о рассчитанных двойственных ценах. Двойственная цена, равная 1, показывает, что увеличение правой части соответствующего неравенства на единицу приводит к возрастанию общего количества автобусов также на единицу. Если двойственная цена равна нулю, то увеличение значения правой части соответствующего неравенства не приводит к возрастанию общего числа используемых автобусов. Однако эти утверждения справедливы только тогда, когда значения правых частей неравенств не выходят за пределы соответствующих интервалов (значения в столбцах Min RHS и Max RHS таблицы Right-hand Side – Single Changes). Например, правая часть второго неравенства может возрасти от 8 до 14, без увеличения общего количества используемых автобусов. А увеличение на какое-либо значение правой части третьего неравенства (начиная со значения 10 и выше) приводит к такому же увеличению общего количества автобусов. Эта информация очень важна для анализа оптимального решения.

Анализ чувствительности коэффициентов целевой функции в данном случае не имеет особого смысла, поскольку в этой модели они естественным образом имеют значение 1 и не могут иметь другого значения. Если бы целевая функция отражала какую-нибудь другую "цель" (например, минимизировала бы стоимость эксплуатации автобусов), тогда ситуация могла быть иной и анализ этих коэффициентов имел бы смысл.

Пример 2.6–4. (Задача минимизации потерь при разрезании рулонов бумаги)

Тихоокеанская бумажная фабрика производит стандартные рулоны бумаги шириной в 20 футов. Специальные заказы клиентов требуют разрезания стандартных рулонов. Типовой заказ (такие заказы могут меняться каждый день) приведен в следующей таблице.

Позиции заказа	Требуемая ширина рулона (футы)	Требуемое количество рулонов (шт.)
1	5	150
2	7	200
3	9	300

На фабрике заказы выполняются путем разрезания стандартных рулонов на требуемые специальными ножами. Существует несколько вариантов разрезки стандартного рулона, три из которых показаны на рис. 2.14. Конечно, существуют и другие варианты, не показанные на этом рисунке (они будут описаны ниже), но сейчас мы ограничимся только представленными — *A*, *B* и *C*. Для выполнения заказа можно совместно использовать несколько вариантов разрезки стандартных рулонов. Например, для выполнения заказа, приведенного в таблице, можно применить следующие комбинации вариантов *A*, *B* и *C*.

1. Разрезать 300 стандартных рулонов, используя вариант *A*, и 75 рулонов с помощью варианта *B*.
2. Разрезать 200 стандартных рулонов, используя вариант *A*, и 100 рулонов с помощью варианта *C*.

*** OPTIMUM SOLUTION SUMMARY ***

Title: Bus Model
Final iteration No: 5
Objective value (min) = 26.0000, ALTERNATIVE solution detected at x3

Variable	Value	Obj Coeff	Obj Val Contrib
x1 00:01	4.0000	1.0000	4.0000
x2 4:00	10.0000	1.0000	10.0000
x3 8:00	0.0000	1.0000	0.0000
x4 12:00	8.0000	1.0000	8.0000
x5 16:00	4.0000	1.0000	4.0000
x6 20:00	0.0000	1.0000	0.0000

Constraint	RHS	Slack(-)/Surplus(+)
1 (>)	4.0000	0.0000+
2 (>)	8.0000	6.0000+
3 (>)	10.0000	0.0000+
4 (>)	7.0000	1.0000+
5 (>)	12.0000	0.0000+
6 (>)	4.0000	0.0000+

*** SENSITIVITY ANALYSIS ***

DEGENERATE or ALTERNATE optimum. Ranges may not be unique
Objective coefficients -- Single Changes:

Variable	Current Coeff	Min Coeff	Max Coeff	Reduced Cost
x1 00:01	1.0000	0.0000	1.0000	0.0000
x2 4:00	1.0000	0.0000	1.0000	0.0000
x3 8:00	1.0000	1.0000	infinity	0.0000
x4 12:00	1.0000	1.0000	1.0000	0.0000
x5 16:00	1.0000	1.0000	1.0000	0.0000
x6 20:00	1.0000	1.0000	infinity	0.0000

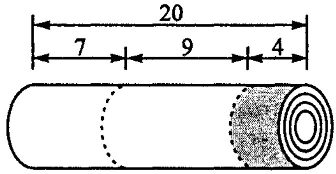
Right-hand Side -- Single Changes:

Constraint	Current RHS	Min RHS	Max RHS	Dual Price
1 (>)	4.0000	0.0000	infinity	1.0000
2 (>)	8.0000	-infinity	14.0000	0.0000
3 (>)	10.0000	4.0000	infinity	1.0000
4 (>)	7.0000	-infinity	8.0000	0.0000
5 (>)	12.0000	11.0000	infinity	1.0000
6 (>)	4.0000	0.0000	5.0000	0.0000

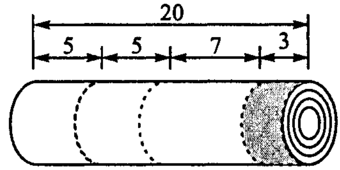
Right-hand Side Ranging -- Simultaneous Changes D:

Basic Var	Value/Feasibilty Condition
x1 00:01	4.0000 + 1.0000 D1 >= 0
Sx8	6.0000 + 1.0000 D1 + -1.0000 D2 + 1.0000 D3 >= 0
x2 4:00	10.0000 + 1.0000 D3 >= 0
Sx10	1.0000 + -1.0000 D4 + 1.0000 D5 + -1.0000 D6 >= 0
x4 12:00	8.0000 + 1.0000 D5 + -1.0000 D6 >= 0
x5 16:00	4.0000 + 1.0000 D6 >= 0

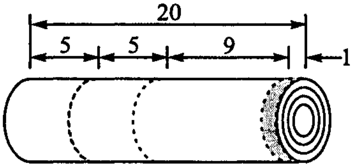
Рис. 2.13



Вариант *A*



Вариант *B*



Вариант *C*

Рис. 2.14. (Размеры даны в футах)

Какая из этих комбинаций лучше? Чтобы ответить на этот вопрос, надо рассмотреть потери от каждой из комбинаций. На рис. 2.14 серым цветом показаны отходы бумаги после каждого варианта разрезки. Мы можем оценить преимущества каждой комбинации, если подсчитаем суммарные отходы, полученные после их применения. Но поскольку отрезки рулонов, идущие в отходы, имеют разную длину, нам надо подсчитать *объем* этих отходов, а не просто количество отрезков. Предполагая, что стандартный рулон бумаги имеет площадь поперечного сечения, равную L квадратным футам, подсчитываем потери для комбинаций 1 и 2 (рис. 2.14).

Комбинация 1: $300\times(4 \times L) + 75\times(3 \times L) = 1425L$ куб. футов.
Комбинация 2: $200\times(4 \times L) + 100\times(1 \times L) = 900L$ куб. футов.

Если число рулонов шириной 5, 7 или 9 футов, которые были получены в результате применения какой-либо комбинации вариантов разрезки, превышает количество, необходимое для выполнения заказа, то разность между ними также следует отнести к потерям. В первой комбинации при использовании варианта *A* получено $300 - 200 = 100$ лишних рулонов шириной 7 футов; применение варианта *B* добавляет еще 75 лишних рулонов такой же ширины. Таким образом, дополнительные потери составляют $175 \times (7 \times L) = 1225L$ куб. футов. Комбинация 2 не производит лишних рулонов шириной 7 и 9 футов, но применение варианта *C* приводит к появлению $200 - 150 = 50$ лишних рулонов шириной 5 футов, что составляет $50 \times (5 \times L) = 250L$ куб. футов дополнительных отходов бумаги. В результате этих выкладок имеем следующее.

Объем отходов от комбинации $1 = 1425L + 1225L = 2650L$ куб. футов.
Объем отходов от комбинации $2 = 900L + 250L = 1150L$ куб. футов.

Очевидно, что комбинация 2 лучше, так как имеет меньше отходов.

Для получения оптимального решения данной задачи необходимо определить все допустимые варианты разрезки стандартных рулонов и затем получить *все возможные* комбинации этих вариантов. Хотя определить все допустимые варианты разрезки несложно, перебор всех комбинаций этих вариантов уже является нетривиальной задачей. Здесь необходим систематический подход к организации такого перебора. В данном случае это может выполнить задача линейного программирования.

Математическая интерпретация задачи

Мы должны найти комбинацию *вариантов разрезки* (переменные), с помощью которой *можно было бы выполнить заказ* (ограничения) с *наименьшими отходами бумаги* (целевая функция).

Переменные надо определить таким образом, чтобы их значения можно было бы интерпретировать в способ разрезки стандартных рулонов бумаги. Поэтому определим переменные как *количество стандартных рулонов, разрезанных с помощью конкретных вариантов разрезки*. Это определение требует описания всех возможных вариантов разрезки. Три варианта показаны на рис. 2.14, другие возможные варианты приведены в следующей таблице. Убедитесь, что не пропущены еще какие-нибудь варианты разрезки, при этом помните, что в допустимом варианте разрезки ширина остатка стандартного рулона должна быть меньше 5 футов.

Требуемая ширина (футы)	Варианты						Требуемое количество рулонов
	1	2	3	4	5	6	
5	0	2	2	4	1	0	150
7	1	1	0	0	2	0	200
9	1	0	1	0	0	2	300
Остаток (футы)	4	3	1	0	1	2	

Теперь можно определить переменные: x_j — количество стандартных рулонов, разрезанных вариантом j, $j = 1, 2, \ldots, 6$.

Ограничением в этой модели является требование того, что произведенных рулонов заданных размеров (5, 7 и 9 футов) будет достаточно для выполнения заказа. Если использовать все варианты разрезки, приведенные в таблице, то получим, что

количество рулонов шириной 5 футов $= 2x_2 + 2x_3 + 4x_4 + x_5$,
количество рулонов шириной 7 футов $= x_1 + x_2 + 2x_5$,
количество рулонов шириной 9 футов $= x_1 + x_3 + 2x_6$.

Эти числа должны быть не меньше 150, 200 и 300 соответственно.

Для построения целевой функции заметим, что общий объем отходов можно подсчитать как разность между объемом всех использованных стандартных рулонов и объемом рулонов, необходимых для выполнения заказа. Запишем это следующим образом:

объем использованных стандартных рулонов $= 20L(x_1 + x_2 + x_3 + x_4 + x_5 + x_6)$,
объем заказных рулонов $= L(150 \times 5 + 200 \times 7 + 300 \times 9) = 4750L$.

Поскольку объем рулонов, необходимых для выполнения заказа, постоянен, а также постоянна величина поперечного сечения L, целевую функцию можно записать просто как сумму всех переменных: $z = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$.

Таким образом, получаем следующую задачу линейного программирования.

$$\text{Минимизировать } z = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

при выполнении условий

$$2x_2 + 2x_3 + 4x_4 + x_5 \geq 150 \text{ (ограничение на рулоны шириной 5 футов)},$$
$$x_1 + x_2 + 2x_5 \geq 200 \text{ (ограничение на рулоны шириной 7 футов)},$$
$$x_1 + x_3 + 2x_6 \geq 300 \text{ (ограничение на рулоны шириной 9 футов)},$$
$$x_j \geq 0, j = 1, 2, \ldots, 6.$$

Оптимальное решение этой задачи, представленное на рис. 2.15, показывает, что задача имеет и другие оптимальные решения, где для выполнения того же заказа используется такое же количество рулонов стандартной длины, но применяются другие варианты разрезки. В приведенном решении 25 стандартных рулонов разрезаются в соответствии с вариантом 3, 100 стандартных рулонов — в соответствии с вариантом 5, а 137.5 стандартных рулонов — с вариантом 6. В таком виде это решение нельзя реализовать, так как значение переменной x_6 нецелое. В этой ситуации можно применить к данной задаче алгоритм целочисленного программирования (см. главу 9) или округлить значение переменной x_6 до 138.

При интерпретации результатов, полученных с помощью программы TORA, следует учитывать требование целочисленности значений переменных, которое неявно присутствует в данной задаче. Например, двойственная цена 0.25, соответствующая первому ограничению, показывает, что увеличение на 1 количества заказных рулонов шириной 5 футов потребует дополнительно еще четверти стандартного рулона (шириной 20 футов). Но эта информация в данном случае не имеет практического смысла. Ее нужно переформулировать следующим образом (исходя из условия целочисленности): потребуется дополнительный стандартный рулон при увеличении на 4 единицы количества заказных рулонов шириной 5 футов. Подобные переформулировки следует сделать при интерпретации других двойственных цен.

Упражнения 2.6,а

1. Вернитесь к задаче из примера 2.6–1 (модель банковских инвестиций) и ее решению, приведенному на рис. 2.10.
   a) Рассмотрим таблицу, в которой приведены результаты анализа чувствительности правых частей неравенств ограничений. Объясните, почему минимальное и максимальное значения интервала изменений величины правой части первого неравенства равны соответственно 4.8 и 12 (миллиона долларов).
   b) Предположим, что банк решил вложить все 12 миллионов долларов в сельскохозяйственные и коммерческие кредиты. Вычислите двумя способами чистую прибыль банка.
   c) Предположим, что объем капитала, предназначенного для инвестиций, возрос до 20 миллионов долларов, а лимит на сельскохозяйственные и коммерческие кредиты увеличился до 9 миллионов долларов. Будет ли новое оптимальное решение включать только кредиты на покупку жилья и коммерческие кредиты? Найдите это новое оптимальное решение.

*** OPTIMUM SOLUTION SUMMARY ***

Title: Trim Loss Model
Final iteration No: 5
Objective value (min) = 262.5000
==> ALTERNATIVE solution detected at x4

Variable	Value	Obj Coeff	Obj Val Contrib
x1	0.0000	1.0000	0.0000
x2	0.0000	1.0000	0.0000
x3	25.0000	1.0000	25.0000
x4	0.0000	1.0000	0.0000
x5	100.0000	1.0000	100.0000
x6	137.5000	1.0000	137.5000

Constraint	RHS	Slack(-)/Surplus(+)
1 (>)	150.0000	0.0000+
2 (>)	200.0000	0.0000+
3 (>)	300.0000	0.0000+

*** SENSITIVITY ANALYSIS ***

Objective coefficients -- Single Changes:
==> DEGENERATE or ALTERNATE optimum. Ranges may not be unique

Variable	Current Coeff	Min Coeff	Max Coeff	Reduced Cost
x1	1.0000	0.8750	infinity	-0.1250
x2	1.0000	0.8750	infinity	-0.1250
x3	1.0000	0.5000	1.0000	0.0000
x4	1.0000	1.0000	infinity	0.0000
x5	1.0000	0.2500	1.2500	0.0000
x6	1.0000	1.0000	1.2000	0.0000

Right-hand Side -- Single Changes:
==> DEGENERATE or ALTERNATE optimum. Ranges may not be unique

Constraint	Current RHS	Min RHS	Max RHS	Dual Price
1 (>)	150.0000	100.0000	700.0000	0.2500
2 (>)	200.0000	0.0000	300.0000	0.3750
3 (>)	300.0000	25.0000	infinity	0.5000

Рис. 2.15

2. Вернитесь к примеру 2.6–2 (модель освоения и использования земли). Предположим, что компания Birdeyes может купить дополнительные 100 акров земли за $450 000. Используя результаты анализа чувствительности по этой задаче (см. рис. 2.11), подскажите компании, стоит ли покупать эту землю.
3. Вернитесь к задаче из примера 2.6–3 (составление расписания движения автобусов).
   a) На основе оптимального решения, представленного на рис. 2.13, определите оптимальное количество используемых автобусов в предположении, что минимальные потребности в автобусах для шести 4-часовых периодов (см. рис. 2.12) составляют (i) (4, 12, 10, 7, 12, 4) и (ii) (4, 8, 7, 7, 12, 4).
   b) Предположим, что минимальные потребности в автобусах для шести 4-часовых периодов изменились и теперь составляют (6, 9, 12, 7, 15, 6). Используя результаты анализа чувствительности, определите, будет ли в этой ситуации возможным применить текущее решение. При необходимости найдите новое оптимальное решение.
4. Вернитесь к примеру 2.6–4 (задача минимизации потерь при разрезании рулонов бумаги) и ее оптимальному решению, показанному на рис. 2.15.
   a) Определите потери бумаги при разрезании 200 стандартных рулонов по варианту 1 и 100 стандартных рулонов — по варианту 2.
   b) Предположим, что ширина стандартного рулона равна 15 футам. Определите возможные варианты разрезки на рулоны шириной 5, 7 и 9 футов и укажите потери бумаги при использовании каждого варианта.
   c) В исходной задаче (со стандартными рулонами шириной 20 футов) поступил новый заказ, где потребность в рулонах шириной 7 футов уменьшилась до 80 штук, а необходимое число рулонов другой ширины (5 и 9 футов) осталось неизменным. Сколько стандартных рулонов необходимо для выполнения нового заказа?
   d) В исходной задаче требуемое количество рулонов шириной 9 футов возросло до 400 штук. Сколько дополнительных стандартных рулонов необходимо для выполнения такого заказа?
5. Нефтедобывающая компания, расположенная на острове Аруба, добывает 600 000 баррелей сырой нефти в день. Нефтеперерабатывающий завод производит два вида неэтилированного бензина: рядовой и высококачественный. Процесс нефтепереработки включает три стадии: 1) перегонка сырой нефти на перегонной колонне — на выходе бензиновый полуфабрикат, 2) часть полуфабриката поступает на крекинг-установку, где производится бензиновый дистиллят, 3) смесительная установка смешивает полуфабрикат, полученный на выходе перегонной колонны, и бензиновый дистиллят. Как рядовой, так и высококачественный бензин можно получить на основе либо бензинового полуфабриката, либо бензинового дистиллята (это зависит от того, что является основой смеси в смесительной установке), но стоимость таких видов бензина будет разной. Компания подсчитала, что чистая прибыль от одного барреля рядового бензина составляет $7.70 и $5.20, в зависимости от того, будет ли основой бензина полуфабрикат или дистиллят. Аналогичная чистая прибыль от одного барреля высококачественного бензина составляет соответственно $12.30 и $10.40.

Далее, на производство одного барреля бензинового полуфабриката (получаемого на выходе перегонной колонны) идет 5 баррелей сырой нефти. Крекинг-установка за день не может переработать более 40 000 баррелей полуфабриката. Весь остальной полуфабрикат идет на изготовление чистого бензина через смесительную установку. Ежедневно требуется производить не более 80 000 баррелей рядового бензина и 50 000 баррелей высококачественного бензина.

a) Разработайте математическую модель для нахождения оптимального производственного плана нефтеперерабатывающего завода.

b) Предположим, что появилась возможность увеличить производительность перегонной колонны до 650 000 баррелей в день, для чего необходимо одноразовое вложение \$3 500 000, а после этого \$15 000 ежедневно для поддержания такой производительности. Порекомендуете ли вы реализовать такую возможность? Обоснуйте свою рекомендацию.

6. Сахарный завод из сиропа сахарного тростника производит желтый сахар, обычный белый, сахарную пудру и мелассу (черную патоку). Компания еженедельно закупает 4000 т сиропа и планирует производить не менее 25 т каждого сахарного продукта в неделю. Процесс производства начинается с переработки сахарного сиропа в желтый сахар и мелассу. Из одной тонны сиропа получается 0.3 т желтого сахара и 0.1 т мелассы. Далее из желтого сахара вырабатывается белый: из тонны желтого сахара получается 0.8 т белого. Наконец, сахарная пудра получается из белого сахара путем размельчения на специальной мельнице. Производительность этой мельницы равна 95%, т.е. из тонны белого сахара получается 0.95 т сахарной пудры. Доход от одной тонны желтого и белого сахара, сахарной пудры и мелассы составляет \$150, \$200, \$230 и \$35 соответственно.

a) Сформулируйте задачу линейного программирования и найдите ее оптимальное решение.

b) Определите экономическую целесообразность расширения производства сахарного завода для переработки более 4000 т сахарного сиропа еженедельно.

7. Некая компания рассматривает возможность реализации шести проектов в течение 4 лет. Ожидаемые затраты на реализацию каждого проекта и доход от них приведены в следующей таблице. Компания может выполнить любой проект частично или полностью. При частичном выполнении проекта доход и затраты считаются пропорционально реализованной доле проекта.

	Затраты (на \$1000)				
Проект	1-й год	2-й год	3-й год	4-й год	Доход (на \$1000)
1	10.5	14.4	2.2	2.4	32.40
2	8.3	12.6	9.5	3.1	35.80
3	10.2	14.2	5.6	4.2	17.75
4	7.2	10.5	7.5	5.0	14.80
5	12.3	10.1	8.3	6.3	18.20
6	9.2	7.8	6.9	5.1	12.35
Возможное вложение (в \$1000)	60.0	70.0	35.0	20.0	

a) Сформулируйте задачу линейного программирования и найдите решение (состоящее из набора выполняемых частей проектов), максимизирующее общую прибыль.

b) Предположим, что никакая часть второго проекта не может быть выполнена без выполнения такой же или большей части шестого проекта. Измените формулировку задачи и найдите новое оптимальное решение.

c) Какой эффект от денег, вложенных на 4-м году?

d) Предположим, что деньги, оставшиеся в конце года, можно использовать в следующем году. Найдите новое оптимальное решение и определите, какую сумму каждый год может "занять" у предыдущего.

e) Предположим, что суммы, вкладываемые ежегодно в течение первых трех лет, при необходимости можно получить в виде займа у внешних заемщиков. Переформулируйте задачи и найдите оптимальное решение. Будет ли новое решение требовать займов каждый год? Если "Да", то каков процент прибыли можно получить на заемные деньги?

8. Производственная компания "Прогресс" получила заказ на производство оконных блоков, рассчитанный на 6 месяцев. В течение этого срока надо поставить 100, 250, 190, 140, 220 и 110 оконных блоков ежемесячно. Стоимость оконных блоков в разные месяцы может быть разной, в зависимости от стоимости трудовых ресурсов, материалов и оконной фурнитуры. Компания подсчитала, что стоимость ее продукции на следующие 6 месяцев будет равна \$50, \$45, \$55, \$48, \$52 и \$50 за один оконный блок. Учитывая изменения стоимости, компания может производить больше оконных блоков, чем необходимо, и использовать ранее произведенную продукцию для покрытия потребности следующих месяцев. Однако хранение одного оконного блока стоит \$8 в месяц, причем начисления за хранение происходят при инвентаризации продукции в конце каждого месяца.

a) Сформулируйте задачу линейного программирования для определения оптимальной временн*о*й схемы производства.

b) Решите задачу в предположении, что компания имеет в начале первого месяца в запасе 25 оконных блоков.

c) Объясните, почему двойственные цены для 1-го, 2-го, 4-го и 5-го месяцев в точности равны стоимостям производства единицы продукции в эти месяцы, тогда как в 3-й месяц этого не наблюдается.

d) Если стоимость хранения оконного блока возрастет до \$9, изменится ли полученное ранее оптимальное решение?

9. Некий инвестор имеет четыре проекта инвестирования суммы в размере \$100 000. В следующей таблице показаны денежные потоки для каждого инвестиционного проекта.

	Денежные потоки (в \$1000) на начало года				
Проект	1-й год	2-й год	3-й год	4-й год	5-й год
1	−1.00	0.50	0.30	1.80	1.20
2	−1.00	0.60	0.20	1.50	1.30
3	0.00	−1.00	0.80	1.90	0.80
4	−1.00	0.40	0.60	1.80	0.95

Рассмотрим подробнее эту таблицу. Например, для проекта 1 вложение $1.00 в начале первого года принесет $0.50 в начале второго года, $0.30 — в начале третьего года, $1.80 — в начале четвертого и $1.20 — в начале пятого. Для остальных проектов данные интерпретируются таким же образом. Значение 0.00 показывает, что поступления денег в этом году нет. Инвестор также может положить деньги в банк под 6.5% годовых. Деньги, полученные по итогам года, можно реинвестировать в последующие годы.

a) Сформулируйте задачу линейного программирования и найдите решение, оптимизирующее размещение инвестиций.
b) Используя двойственные цены, определите доходность инвестиций.
c) Если в конце первого года вы планируете истратить $1000 на удовольствия, то как это отразится на сумме, получаемой в начале пятого года?
d) Если в конце каждого из первых четырех годов вы тратите $1000 на удовольствия, то как это отразится на сумме, получаемой в начале пятого года?

10. Производственная компания Toolco заключила контракт с сетью магазинов автотоваров AutoMate на поставку 1500 гаечных ключей и 1200 специальных отверток еженедельно. Работая в одну смену, компания Toolco не может выполнить этот контракт, поэтому вынуждена ввести сверхурочные и воспользоваться услугами субподрядчиков, в результате чего возрастет себестоимость ее инструментов, как показано в следующей таблице. Отметим также, что рыночная цена гаечных ключей более чем в два раза выше рыночной цены отверток.

Инструмент	Тип производства	Еженедельные производственные возможности (шт.)	Себестоимость единицы продукции ($)
Гаечный ключ	Обычный	$0-550$	2.00
	Использование сверхурочных	$551-800$	2.80
	Использование субподрядчиков	$801-\infty$	3.00
Отвертка	Обычный	$0-620$	2.10
	Использование сверхурочных	$621-900$	3.20
	Использование субподрядчиков	$901-\infty$	4.20

a) Сформулируйте задачу линейного программирования и найдите оптимальную схему производства каждого инструмента.
b) Свяжите двойственные цены в анализе чувствительности оптимального решения с себестоимостью продукции, приведенной в таблице.
c) Как скажется на общей стоимости продукции увеличение на единицу производственных возможностей обычной рабочей смены и сверхурочной работы?

11. Четыре изделия последовательно обрабатываются на двух станках. Данные, описывающие этот технологический процесс, приведены в следующей таблице.

Станок	Стоимость часа работы ($)	Время обработки (часы) Изделие 1	Изделие 2	Изделие 3	Изделие 4	Максимальная нагрузка (часы)
1	10	2	3	4	2	500
2	5	3	2	1	2	380
Цена единицы изделия		65	70	55	45	

a) Сформулируйте задачу линейного программирования и найдите ее оптимальное решение.

b) Предположим, что увеличить нагрузку на станки можно только за счет сверхурочной работы. Какую максимальную стоимость часа сверхурочной работы каждого станка может позволить компания?

c) Какую максимальную стоимость машинного часа можно применить при обработке третьего изделия, чтобы его продажа приносила хоть какую-то прибыль.

d) Предположим, что в исходной модели цены на изделия 1, 3 и 4 возросли до $80, $65 и $60 соответственно. Определите границы на стоимость изделия 2, в пределах которых сохраняются текущие значения переменных (т.е. найденного ранее оптимального решения).

e) В исходной модели нагрузка на первый станок возросла до 550 часов. Определите границы для нагрузки второго станка, сохраняющие текущее решение.

12. Завод производит три типа (I, II и III) некоторого изделия, используя для этого материал A и B. В следующей таблице приведены необходимые данные для этой задачи.

Материал	Расход материалов на единицу изделия I	II	III	Доступно
A	2	3	5	4000
B	4	2	7	6000
Надо произвести не менее (шт.)	200	200	150	
Доход на единицу изделия ($)	30	20	50	

На изготовление единицы изделия типа I затрачивается в два раза больше рабочего времени, чем на изготовление единицы изделия типа II, и в три раза больше, чем на изготовление единицы изделия типа III. Рабочие ресурсы завода эквивалентны ресурсам, необходимым для изготовления 1500 шт. изделия типа I.

a) Сформулируйте задачу линейного программирования и найдите ее оптимальное решение.

b) Предположим, что завод имеет возможность приобрести дополнительный объем материала A по цене $12 за единицу. Целесообразно ли делать такую покупку?

c) Порекомендуете ли вы заводу купить дополнительный объем материала B по цене $5 за единицу?

d) Целесообразно ли заводу увеличить текущее отношение (3:2) цены на изделия типа I к цене на изделия типа II?

13. Компания HiRise получила предложение участвовать в двух одногодичных проектах. Ежеквартальные денежные потоки для этих проектов показаны в следующей таблице.

Проект	Объем денежных потоков (млн долл.) на указанную дату				
	1.01.99	1.04.99	1.07.99	1.10.99	31.12.99
I	−1.0	−3.1	−1.5	1.8	5.0
II	−3.0	−2.5	1.5	1.8	2.8

Компания в начале каждого квартала может инвестировать $1 000 000, а также взять заем на сумму, не превышающую 10% совокупного годового дохода. Все займы должны быть возвращены в конце квартала. Прибавочные суммы могут ежеквартально приносить прибыль, равную 10% годовых. Все суммы, аккумулированные в конце квартала, можно инвестировать в следующем квартале.

a) Предположим, что компания для реализации проектов может привлекать к долевому участию внешних партнеров. Определите уровень их долевого участия для максимизации чистой прибыли, аккумулированной на 31.12.99.

b) Объясните, можно ли в каком-нибудь квартале использовать внешний заем и получить в конце этого же квартала чистый доход?

c) Дайте экономическую интерпретацию двойственным ценам в этой модели.

d) Покажите, что двойственную цену, соответствующую верхней границе займа в начале третьего квартала, можно выразить через двойственную цену, соответствующую балансным уравнениям денежных потоков на каждую дату, приведенную в таблице.

14. В предчувствии больших расходов на обучение своего ребенка в колледже, семейная пара решила ежегодно откладывать определенную сумму в течение 10 лет, начиная с 8-летнего возраста ребенка. Они распределили эти суммы следующим образом.

Год	1	2	3	4	5	6	7	8	9	10
Сумма ($)	2000	2000	2500	2500	3000	3500	3500	4000	4000	5000

Для предотвращения нежелательных финансовых сюрпризов, семейная пара решила вложить деньги в (1) страховой полис с 7.5% годовых; (2) шестилетние правительственные ценные бумаги с 7.9% годовых (их текущая рыночная цена равна 98% номинальной стоимости); (3) девятилетние муниципальные ценные бумаги с доходностью 8.5% и текущей рыночной стоимостью 1.02 от их номинальной стоимости.

a) Помогите этой семейной паре вложить свои деньги наиболее оптимальным образом.

b) Вычислите их ежегодные доходы.

15. Бизнесмен имеет возможность вложить деньги в два инвестиционных проекта: проект A гарантирует \$0.70 на каждый вложенный доллар ежегодно, проект B — \$2 на вложенный доллар по истечении двух лет. В проекте A вложения можно делать ежегодно, а в проекте B только в периоды, кратные двум годам.

a) Как инвестировать \$100 000 для получения максимального дохода в конце третьего года инвестирования.

b) Стоит ли вкладывать еще деньги в эти проекты, т.е. можно ли получить больший процент прибыли при бóльших вложениях?

16. Рассмотрим задачу назначения трех типов самолетов на четыре маршрута в соответствии со следующими данными.

Тип самолета	Вместимость самолета (чел.)	Количество самолетов	Число ежедневных полетов по маршруту 1	2	3	4
1	50	5	3	2	2	1
2	30	8	4	3	3	2
3	20	10	5	5	4	2
Ежедневное число пассажиров			1000	2000	900	1200

Стоимость полетов, включая неустойку за "потерю" пассажиров вследствие нелетной погоды, приведена в следующей таблице.

Тип самолета	Стоимость (в \$) одного полета по маршруту 1	2	3	4
1	1000	1100	1200	1500
2	800	900	1000	1000
3	600	800	800	900
Неустойка (\$) за одного "потерянного" пассажира	40	50	45	70

a) Определите оптимальное распределение самолетов по маршрутам.

b) Целесообразно ли увеличивать количество самолетов какого-либо типа?

c) Интерпретируйте двойственные цены, соответствующие ограничениям на количество пассажиров того или иного маршрута.

17. Сплавы A и B состоят из металлов I, II, III и IV согласно следующей спецификации.

Сплав	Спецификация	Рыночная цена (\$)
A	Не более 80% металла I	200
	Не более 30% металла II	
	Не менее 50% металла IV	
B	От 40% до 60% металла II	300
	Не менее 30% металла III	
	Не более 70% металла IV	

Металлы, используемые для сплавов, получают из руды трех типов, как показано в следующей таблице.

Руда	Максимум добычи (т)	Содержание металлов (%) I	II	III	IV	Другое	Стоимость 1 т ($)
1	1000	20	10	30	30	10	30
2	2000	10	20	30	30	10	40
3	3000	5	5	70	20	0	50

a) Сколько сплавов каждого типа можно произвести? (*Совет.* Обозначьте через x_{ik} количество руды i, расходуемой на производство сплава k, и через w_k — количество произведенного сплава k.)

b) Сколько руды каждого типа расходуется на производство сплава A и сколько — на производство сплава B?

c) Какое из ограничений, связанных со спецификацией сплавов, самое неблагоприятное для оптимального решения?

d) Какую максимальную цену можно позволить при покупке руды 1? А руды 2 и 3?

18. Игрок участвует в игре, где требуется разделить ставку по четырем полям. Игра имеет три исхода. В следующей таблице показаны прибыль и потери для каждого поля в зависимости от исхода игры.

Исход игры	Возврат на $1, поставленный на поле 1	2	3	4
1	–3	4	–7	15
2	5	–3	9	4
3	3	–9	10	–8

Игрок имеет $500, которые он может поставить только один раз. Шансы какого-либо исхода игры неизвестны. В условиях этой неопределенности найдите стратегию, которая максимизировала бы *минимальный* возврат сделанной ставки при всех возможных исходах игры.

a) Как игрок должен разложить $500 по четырем полям? (*Подсказка.* Чистая прибыль игрока может быть положительной, нулевой или отрицательной.)

b) Ваш совет игроку о том, как сделать ставки, если появится дополнительная сумма.

2.7. Заключение

Техника графического решения задач линейного программирования позволяет выявить многие свойства оптимального решения ЛП. Основное свойство заключается в том, что оптимальное решение задачи ЛП достигается в одной из угловых точек пространства решений. На основе этого был разработан симплекс-метод для решения задач линейного программирования, который описан в следующей главе. В данной главе также дана экономическая интерпретация оптимального решения и представлены результаты анализа чувствительности, примененного к оптимальному решению, полученному графическим способом.

Литература

1. Bazaraa M., Jarvis J., Sherall M. *Linear Programming and Network Flows*, 2nd ed., Wiley, New York, 1990.
2. Schrage L. *LINDO: An Optimization Modeling System, Text and Software*, 4th ed., Boyd and Fraser, Danvers, Mass, 1991.
3. William H. *Model Building in Mathematical Programming*, 3rd ed., Wiley, New York, 1990.

Литература, добавленная при переводе

Гольштейн Е.Г., Юдин Д.Б. *Линейное программирование: Теория, методы и приложения.* — М.: Наука, 1969.

Данциг Д. *Линейное программирование, его обобщение и применение.* — М.: Прогресс, 1966.

Кофман А. *Методы и модели исследования операций.* — М.: Мир, 1966.

Комплексные задачи

■ **2-1.**[1] Компания Hi-C производит три апельсиновых продукта: концентрат, обычный сок и джем, которые расфасовываются в 5-галонные банки. Для джема расходуются апельсины только первого сорта, а для других продуктов — апельсины второго сорта. В следующей таблице показано, сколько апельсинов идет на производство продуктов, а также максимальный возможный объем их производства на следующий год.

Продукт	Сорт апельсинов	К-во апельсинов (фунты), необходимых для изготовления 5-галонной банки продукта	Возможности производства (к-во банок)
Джем	I	5	10 000
Концентрат	II	30	12 000
Сок	II	15	40 000

Маркетинговые исследования показывают, что рыночные потребности в обычном соке более чем в два раза превосходят потребности в апельсиновом концентрате.

В прошлом году компания закупала апельсины первого и второго сортов по отдельности по цене соответственно 25 и 20 центов за фунт. В этом году в силу ряда причин поставщики поставляют апельсины без сортировки. Подсчитано, что урожай текущего года (всего собрано 3 миллиона фунтов апельсинов) на 30% состоит из апельсинов первого сорта и на 60% — из апельсинов второго сорта. Оптовая цена неотсортированных апельсинов опустилась до 19 центов за фунт. Компания Hi-C подсчитала, что сортировка апельсинов обойдется ей в 2.15 цента за фунт. Несортовые апельсины (10% от поставок) идут в отходы.

Для определения себестоимости продукции экономический отдел компании использует следующий способ вычисления стоимости фунта апельсинов первого и второго

[1] Материал для этой задачи взят из статьи “Red Brand Canners”, *Stanford Business Cases 1965*, Graduate School of Business, Stanford University (Высшая школа бизнеса Станфордского университета).

сортов. Поскольку 10% апельсинов идет в отходы, средняя стоимость сортовых апельсинов равна (19 + 2.15)/0.9 = 23.5 цента. Так как отношение объема апельсинов первого сорта к объему апельсинов второго сорта составляет 1:2, средняя цена (на основе прошлогодних цен) должна быть равной (20 × 2 + 25 × 1)/3 = 21.67 цента. Итак, в этом году средняя цена апельсинов возросла на 1.83 цента (= 23.5 – 21.67). Эту разность надо "разбросать" на стоимость апельсинов первого и второго сортов, учитывая соотношения их объемов 1:2. В результате, стоимость апельсинов первого сорта равна 25 + 1.83×(1/3) = 25.61 цента за фунт, а апельсинов второго сорта — 20 + 1.83×(2/3) = 21.22 цента за фунт. На основе этой информации экономический отдел вычислил доходность всех трех производимых компанией продуктов.

	Джем (для 5-галонных банок)	Концентрат	Сок
Отпускная цена	\$15.50	\$30.25	\$20.75
Стоимость сырья	9.85	21.05	13.28
Другие расходы	1.05	2.15	1.96
Себестоимость	\$10.90	\$23.20	\$15.24
Чистый доход	\$4.60	\$7.05	\$5.51

Составьте оптимальный производственный план для компании Hi-C.

■ **2-2.**[1] Сталелитейная компания имеет литейный цех и два прокатных стана. Литейный цех производит три типа стальных заготовок, которые, прежде чем попасть на прокатные станы, обрабатываются в механическом цехе.

В начале каждого квартала прокатные станы определяют помесячно свои потребности в стальных заготовках и делают заказ литейному цеху. Руководство литейного цеха на основе этого заказа строит свой производственный план, на который влияют ограниченные мощности механического цеха. Возможная недостача стальных заготовок для прокатных станов покрывается покупкой аналогичных заготовок у сторонних производителей. Стоимость собственных заготовок и закупаемых у сторонних производителей показана в следующей таблице. Руководство считает, что такие внешние закупки происходят не часто и их можно позволить при условии, что они не превышают 5% от требуемого количества заготовок.

Тип заготовки	Длина заготовки (футы)	Внутренняя цена (в \$ на одну заготовку)	Цена на рынке (в \$ на одну заготовку)
1	800	90	108
2	1200	130	145
3	1650	180	194

В механическом цехе заготовки обрабатываются на станках четырех различных типов, имеющих следующую производительность.

[1] Взято из S. Jain, K. Stott, E. Vasold. "Orderbook Balancing Using a Combination of Linear Programming and Heuristic Techniques", *Interfaces*, Vol. 9, No. 1, November 1978, pp. 55–67.

Тип станка	Время обработки одной заготовки типа			Количество станков	Общее время работы станка в месяц (часы)
	1	2	3		
1	1	5	7	10	320
2	0	4	6	8	310
3	6	3	0	9	300
4	3	6	9	5	310

Потребности прокатных станов в стальных заготовках на следующие три месяца.

Месяц	Потребность в заготовках					
	Первый прокатный стан			Второй прокатный стан		
	Заготовки типа 1	Заготовки типа 2	Заготовки типа 3	Заготовк и типа 1	Заготовки типа 2	Заготовки типа 3
1	5000	2000	4000	2000	1000	0
2	0	3000	5000	3000	2000	2000
3	1000	0	3000	0	4000	2000

Составьте оптимальный производственный график для механического цеха.

■ **2-3.** Фирма ArkTec собирает персональные компьютеры для частных клиентов. На следующие четыре квартала имеются заказы на 400, 700, 500 и 200 компьютеров соответственно. Фирма может производить больше компьютеров, чем указано в заказах, но в таком случае приходится платить $100 за хранение одного компьютера в течение квартала. Увеличение производства в следующем квартале, по сравнению с предыдущим, приводит к дополнительному набору работников, что повышает себестоимость одного компьютера на $60. При уменьшении производства в следующем квартале, по сравнению с предыдущим, придется прибегнуть к сокращению персонала, что также увеличивает себестоимость одного компьютера на $50. Как организовать сборку компьютеров, чтобы удовлетворить все заказы?

■ **2-4.** Мебельная фабрика осуществляет производство и сборку стульев, столов и книжных полок. Заготовительный цех производит получистовую продукцию, которая затем собирается в сборочном цехе фабрики.

Ежемесячная производительность заготовительного цеха составляет 3000 стульев, 1000 столов и 580 книжных полок (несобранных). В сборочном цехе работают 150 рабочих в две 8-часовые смены 5 дней в неделю. Среднее время сборки одной единицы продукции равно 20, 40 и 15 минут соответственно для стульев, столов и книжных полок.

Количество рабочих в сборочном цехе колеблется в зависимости от того, сколько рабочих в настоящее время находится в ежегодном отпуске. В мае, июне и июле планируют уйти в отпуск соответственно 20, 25 и 45 человек.

Отдел маркетинга фабрики оценил потребность рынка в их продукции на май, июнь и июль следующим образом.

Продукт	Май	Июнь	Июль	Остаток на конец апреля
Стул	2800	2300	3350	30
Стол	500	800	1400	100
Книжная полка	320	300	600	50

Себестоимость продукции и ее отпускная цена показаны в следующей таблице.

Продукт	Себестоимость ($)	Отпускная цена ($)
Стул	150	250
Стол	400	750
Книжная полка	60	120

Если какая-либо продукция не продана в течение того месяца, в каком она произведена, то она может быть продана в следующем месяце. Хранение одной единицы стоит 2% от стоимости этой продукции.

Помогите фабрике разработать план ежегодных отпусков для работников сборочного цеха.

Симплекс-метод

3.1. Введение

Графический способ решения задачи ЛП из главы 2 показывает, что оптимальное решение этой задачи всегда ассоциируется с *угловой точкой* пространства решений (в математике она также называется **крайней точкой** множества). Это является ключевой идеей при разработке общего алгебраического *симплекс-метода* для решения любой задачи линейного программирования.

Переход от геометрического способа решения задачи ЛП к симплекс-методу лежит через алгебраическое описание крайних точек пространства решений. Для реализации этого перехода сначала надо привести задачу ЛП к **стандартной форме**, преобразовав неравенства ограничений в равенства путем введения дополнительных переменных.

Стандартная форма задачи ЛП необходима, потому что она позволяет получить **базисное решение** (используя систему уравнений, порожденную ограничениями). Это (алгебраическое) базисное решение полностью определяет все (геометрические) крайние точки пространства решений (доказательство этого утверждения приведено в разделе 7.2). Симплекс-метод позволяет эффективно найти оптимальное решение среди всех базисных.

3.2. Стандартная форма задачи ЛП и ее базисные решения

В этом разделе мы сначала рассмотрим стандартную форму записи задачи линейного программирования, а затем покажем, как определить базисное решение.

3.2.1. Стандартная форма задачи ЛП

Стандартная форма записи задачи ЛП предполагает выполнение следующих требований.

1. Все ограничения (включая ограничения неотрицательности переменных) преобразуются в равенства с неотрицательной правой частью.
2. Все переменные неотрицательные.
3. Целевую функцию следует или максимизировать, или минимизировать.

1. ***Преобразование неравенств в равенства.*** Неравенства любого типа (со знаками неравенств $\leq$ или $\geq$) можно преобразовать в равенства путем добавления в левую часть неравенств дополнительных переменных — *остаточных* или *избыточных* (см. определения этих переменных в разделе 2.3.3).

Пример неравенства типа "≤". Неравенство $x_1 + 2x_2 \le 3$ эквивалентно равенству $x_1 + 2x_2 + s_1 = 3$, где s_1 — остаточная переменная и $s_1 \ge 0$.

Пример неравенства типа "≥". Неравенство $3x_1 + x_2 \ge 5$ эквивалентно равенству $3x_1 + x_2 + S_1 = 5$, где S_1 — избыточная переменная и $S_1 \ge 0$.

Правую часть равенства всегда можно сделать неотрицательной путем умножения всего равенства на –1. Кроме того, заметим, что неравенство типа "≤" также преобразуется в неравенство типа "≥" посредством умножения обеих частей неравенства на –1. Например, неравенство $2 < 4$ после умножения на –1 становится неравенством $-2 > -4$.

2. *ПРЕОБРАЗОВАНИЕ СВОБОДНЫХ ПЕРЕМЕННЫХ В НЕОТРИЦАТЕЛЬНЫЕ.* *Свободную* переменную x_j (т.е. переменную, которая может принимать как отрицательные, так и положительные значения) можно представить как разность двух *неотрицательных* переменных следующим образом.

$$x_j = x_j^+ - x_j^-, \quad x_j^+,\ x_j^- \ge 0.$$

Например, для $x_j = -5$ положим $x_j^+ = 0$ и $x_j^- = 5$. Если же $x_j = +5$, тогда $x_j^+ = 5$ и $x_j^- = 0$. В обоих случаях переменные x_j^+ и x_j^- неотрицательны.

Такое преобразование свободных переменных следует выполнить во всех неравенствах и в целевой функции. После решения задачи с переменными x_j^+ и x_j^- значения исходных переменных восстанавливаются с помощью обратной подстановки.

3. *ПРЕОБРАЗОВАНИЕ ЗАДАЧИ МАКСИМИЗАЦИИ В ЗАДАЧУ МИНИМИЗАЦИИ.* Задача максимизации функции $f(x_1, x_2, ..., x_n)$ эквивалентна задаче минимизации функции $-f(x_1, x_2, ..., x_n)$, поскольку при решении обеих задач предоставляется один и тот же набор значений переменных $x_1, x_2, ..., x_n$.

Пример 3.2–1

Преобразуем следующую задачу ЛП в стандартную форму.

$$\text{Максимизировать } z = 2x_1 + 3x_2 + 5x_3$$

при выполнении условий

$$\begin{aligned}
&x_1 + x_2 - x_3 \ge -5,\\
&-6x_1 + 7x_2 - 9x_3 \le 4,\\
&x_1 + x_2 + 4x_3 = 10,\\
&x_1, x_2 \ge 0,\\
&x_3 \text{ — свободная переменная.}
\end{aligned}$$

Для преобразования задачи в стандартную форму выполним следующие действия.

1. Вычтем из левой части первого неравенства дополнительную (избыточную) переменную S_1 и затем умножим все неравенство на –1, для того чтобы правая часть неравенства стала положительной. (Другой путь преобразования неравенства: сначала умножим его на –1 — неравенство примет вид "≤" вместо "≥", далее к левой части неравенства прибавим дополнительную (остаточную) переменную s_1.)

2. Добавим дополнительную (остаточную) переменную s_2 к левой части второго неравенства.
3. Так как третье ограничение изначально записано в виде равенства, поэтому оставляем его без изменения.
4. Выполняем замену $x_3 = x_3^+ - x_3^-$, где x_3^+, $x_3^- \geq 0$, во всех ограничениях и целевой функции.

Получаем следующую стандартную задачу линейного программирования.

$$\text{Максимизировать } z = 2x_1 + 3x_2 + 5x_3^+ - 5x_3^-$$

при выполнении следующих условий

$$\begin{aligned}
&-x_1 - x_2 + x_3^+ - x_3^- + x_4 = 5,\\
&-6x_1 + 7x_2 - 9x_3^+ + 9x_3^- + x_5 = 4,\\
&x_1 + x_2 + 4x_3^+ - 4x_3^- = 10,\\
&x_1, x_2, x_3^+, x_3^-, x_4, x_5 \geq 0.
\end{aligned}$$

Упражнения 3.2,а

1. Запишите задачу примера 3.2–1 в стандартной форме, предполагая, что первое ограничение является равенством, второе — неравенством типа "≥", а третье — неравенством типа "≤". Также предполагается, что x_1 является свободной переменной, а $x_3 \geq 0$.
2. Два изделия, P1 и P2, можно произвести на двух различных станках M1 и M2. Время изготовления любого изделия на любом станке одинаково. Производительность станка M1 составляет 200 изделий за смену, а производительность станка M2 — 250 изделий. Мастер планирует сбалансировать рабочее время таким образом, чтобы общее количество изделий, произведенных на одном станке, не превышало, более чем на 5 единиц, общего количества изделий, изготовленных на другом станке. Доход от одного изделия P1 составляет \$10, а от второго — \$15. Запишите эту задачу в стандартной форме линейного программирования.
3. Покажите, как следующую целевую функцию можно записать в стандартной форме задачи ЛП.

$$\text{Минимизировать } z = \max\{|x_1 - x_2 + 3x_3|, |-x_1 + 3x_2 - x_3|\},$$

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

4. Покажите, что m равенств вида

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

эквивалентны следующим $m + 1$ неравенствам.

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i, \quad i = 1, 2, \ldots, m,$$

$$\sum_{j=1}^{n}\left(\sum_{i=1}^{m} a_{ij}\right)x_j \ge \sum_{i=1}^{m} b_i.$$

3.2.2. Определение базисных решений

Задача линейного программирования, записанная в стандартной форме, содержит m линейных равенств с n неизвестными переменными ($m < n$). Разделим n переменных на два множества: (1) $n - m$ переменные, которые положим равными нулю, и (2) оставшиеся m переменные, значения которых определяются как решение системы из m линейных уравнений. Если это решение *единственное*, тогда соответствующие m переменные называются **базисными**, а остальные $n - m$ нулевые переменные — **небазисными**. В этом случае результирующие значения переменных составляют **базисное решение**. Если все переменные принимают неотрицательные значения, то такое базисное решение является *допустимым*.[1] В противном случае — *недопустимым*.

Основываясь на этих определениях, нетрудно подсчитать, что количество всех *положительных* базисных решений для m уравнений с n неизвестными не превосходит[2]

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

Пример 3.2–2

Рассмотрим следующую систему двух уравнений с пятью неизвестными ($m = 2$, $n = 5$)

$$x_1 + x_2 + 4x_3 + 2x_4 + 3x_5 = 8,$$
$$4x_1 + 2x_2 + 2x_3 + x_4 + 6x_5 = 4.$$

Определим различные решения этой системы. Количество положительных базисных решений равно $\frac{5!}{3!2!} = 10$. Ниже мы покажем, что некоторые из этих решений на самом деле не будут базисными.

По определению базисное решение включает только две ($= m$) переменные, предполагая, что небазисных *нулевых* переменных три ($= n - m$).

Случай 1. Допустимое базисное решение.

Нулевые (небазисные) переменные: x_2, x_4 и x_5.

---

[1] В русской математической литературе также применяются термины "план", "опорный план" и "оптимальный опорный план" как эквиваленты терминам "базисное решение", "допустимое базисное решение" и "оптимальное базисное решение". — *Прим. ред.*

[2] Через $\binom{n}{m}$ здесь и далее обозначаются биномиальные коэффициенты C_n^m. — *Прим. ред.*

Уравнения: $x_1 + 4x_3 = 8,$
$4x_1 + 2x_3 = 4.$

Решение: *единственное* решение $x_1 = 0, x_3 = 2.$

Заключение: базисное решение допустимо, так как $x_1, x_3 \geq 0.$

Случай 2. Недопустимое базисное решение.

Нулевые (небазисные) переменные: x_3, x_4 и x_5.

Уравнения: $x_1 + x_2 = 8,$
$4x_1 + 2x_2 = 4.$

Решение: *единственное* решение $x_1 = -6, x_3 = 14.$

Заключение: базисное решение недопустимо, так как $x_1 < 0.$

Случай 3. Решение не единственное.

Нулевые (небазисные) переменные: x_1, x_2 и x_5.

Уравнения: $4x_3 + x_4 = 8,$
$2x_3 + x_4 = 4.$

Решение: единственного решения не существует, так как уравнения *зависимы* (если первое уравнение разделить на 2, то получим второе уравнение).

Заключение: бесконечное количество решений.

Случай 4. Решения не существует.

Нулевые (небазисные) переменные: x_1, x_3 и x_4.

Уравнения: $x_2 + 3x_5 = 8,$
$2x_2 + 6x_5 = 4.$

Решение: решения не существует, так как уравнения *несовместны*.

Заключение: решения не существует.

Упражнения 3.2,b

1. В примере 3.2–2 мы проверили только четыре комбинации базисных переменных. Проверьте комбинации переменных (x_1, x_4), (x_1, x_5) и (x_2, x_3) и найдите решения системы уравнений.
2. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 2x_1 + 3x_2$$

при ограничениях

$$x_1 + 3x_2 \leq 6,$$
$$3x_1 + 2x_2 \leq 6,$$
$$x_1, x_2 \geq 0.$$

a) Запишите задачу в стандартной форме.

b) Найдите все базисные решения этой задачи и определите, какие из них допустимые, а какие — нет.

c) Путем непосредственной подстановки решений в целевую функцию определите наилучшее *допустимое базисное* решение.

d) Проверьте графическим способом, что решение, полученное в предыдущем пункте, является оптимальным. Отсюда следует, что оптимальное решение можно получить алгебраически путем перебора множества допустимых базисных решений.

e) Определите, чему на рисунке, где представлено пространство допустимых решений, соответствует *недопустимое* базисное решение.

3. Путем перебора всех базисных решений найдите оптимальные решения следующих задач линейного программирования.

a) Максимизировать $z = 2x_1 - 4x_2 + 5x_3 - 6x_4$
при ограничениях

$$x_1 + 4x_2 - 2x_3 + 8x_4 \le 2,$$
$$-x_1 + 2x_2 + 3x_3 + 4x_4 \le 1,$$
$$x_1, x_2, x_3, x_4 \ge 0.$$

b) Минимизировать $z = x_1 + 2x_2 - 3x_3 - 2x_4$
при ограничениях

$$x_1 + 2x_2 - 3x_3 + x_4 = 4,$$
$$x_1 + 2x_2 + x_3 + 2x_4 = 4,$$
$$x_1, x_2, x_3, x_4 \ge 0.$$

4. Покажите, что все базисные решения следующей задачи ЛП недопустимы.

Максимизировать $z = x_1 + x_2$

при ограничениях

$$x_1 + 2x_2 \le 6,$$
$$2x_1 + x_2 \ge 16,$$
$$x_1, x_2 \ge 0.$$

3.2.3. Свободные переменные и базисные решения

Свободные переменные мы определили в разделе 2.3.2. Затем в разделе 3.2.1 показали, что в стандартной форме записи задачи ЛП свободная переменная x_j должна быть представлена как разность двух неотрицательных переменных:

$$x_j = x_j^+ - x_j^-, \quad x_j^+,\ x_j^- \ge 0.$$

Основываясь на определении базисного решения из раздела 3.2.2, нетрудно показать, что невозможна ситуация, когда x_j^+ и x_j^- одновременно будут базисными переменными, что вытекает из их зависимости. В свою очередь, их зависимость следует из того, что в ограничении коэффициент при x_j^+ имеет знак, противоположный знаку коэффициента при x_j^-. Это означает, что в любом базисном решении по крайней мере одна из переменных x_j^+ и x_j^- должна быть небазисной, т.е. нулевой (см. упр. 3.2,c(1)).

Упражнения 3.2,c

1. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 2x_1 + 3x_2 + 5x_3$$

при ограничениях

$$\begin{aligned} &-6x_1 + 7x_2 - 9x_3 \geq 4,\\ &-x_1 + x_2 + 4x_3 = 10,\\ &x_1, x_3 \geq 0,\\ &x_2 \text{ — свободная переменная.}\end{aligned}$$

Приведите задачу к стандартной форме, используя для этого подстановку $x_2 = x_2^+ - x_2^-$. Покажите, что в этой задаче не существует базисных решений, в которые входили бы одновременно x_2^+ и x_2^-.

2. Рассмотрите следующую задачу ЛП.

$$\text{Максимизировать } z = x_1 + 3x_2$$

при ограничениях

$$\begin{aligned} &x_1 + x_2 \leq 2,\\ &-x_1 + x_2 \leq 4,\\ &x_1 \text{ — свободная переменная,}\\ &x_2 \geq 0.\end{aligned}$$

a) Найдите все допустимые базисные решения этой задачи.
b) С помощью подстановки решений в целевую функцию определите наилучшее базисное решение.
c) Решите эту задачу графическим способом и проверьте, что решение, найденное в предыдущем пункте, является оптимальным.

3. Фирма производит три вида изделий, прибыль от которых составляет соответственно \$2, \$5 и \$3 на единицу изделия. Для производства этих изделий фирма располагает 80 рабочими часами ручного труда и 65 часами машинного времени. Для производства одной единицы изделия каждого из трех видов требуется 2, 1 и 2 часа ручного труда и 1, 1 и 0.5 часов машинного времени соответственно. При необходимости фирма может увеличить количество рабочих часов ручного труда и количество часов машинного времени, но каждый дополнительный час ручного труда будет стоить \$15, а машинного — \$10.

a) Сформулируйте задачу линейного программирования и найдите все ее допустимые базисные решения.
b) Найдите оптимальное решение этой задачи.

4. В задачах ЛП, где есть несколько свободных переменных, преобразование типа $x_j = x_j^+ - x_j^-$, $x_j^+, x_j^- \geq 0$, удваивает соответствующее число неотрицательных переменных. Но при использовании подстановок $x_j = x_j' - w$, $x_j', w \geq 0$, можно k свободных переменных заменить на $k + 1$ неотрицательных. Используя программу TORA, покажите, что эти два метода приводят к одному и тому же решению следующей задачи ЛП.

$$\text{Максимизировать } z = -2x_1 + 3x_2 - 2x_3$$

при ограничениях

$$4x_1 - x_2 - 5x_3 = 10,$$
$$2x_1 + 3x_2 + 2x_3 = 12,$$
$$x_1 \geq 0,$$
$$x_2, x_3 \text{ — свободные переменные.}$$

3.3. Алгоритм симплекс-метода

Основываясь на определениях из раздела 3.2, мы можем найти оптимальное решение задачи линейного программирования, записанной в стандартной форме, путем простого перебора всех базисных (допустимых) решений. Но, конечно, такая процедура не эффективна. Алгоритм симплекс-метода находит оптимальное решение, рассматривая ограниченное количество допустимых базисных решений.

Алгоритм симплекс-метода всегда начинается с некоторого допустимого базисного решения и затем пытается найти другое допустимое базисное решение, "улучшающее" значение целевой функции. Это возможно только в том случае, если *возрастание* какой-либо нулевой (небазисной) переменной ведет к улучшению значения целевой функции. Но для того, чтобы небазисная переменная стала положительной, надо одну из текущих базисных переменных сделать нулевой, т.е. перевести в небазисные. Это необходимо, чтобы новое решение содержало в точности m базисных переменных. (Напомним, что нас интересуют только *базисные* решения, содержащие в точности m базисных переменных.) В соответствии с терминологией симплекс-метода выбранная нулевая переменная называется **вводимой** (в базис), а удаляемая базисная переменная — **исключаемой** (из базиса).

Пример 3.3–1

Используем задачу о компании Reddy Mikks (пример 2.2–1) для рассмотрения деталей выполнения симплекс-метода. Для удобства изложения напомним основные элементы этой задачи. Переменные задачи:

x_1 = объем ежедневного производства краски для наружных работ (тонны),
x_2 = объем ежедневного производства краски для внутренних работ (тонны).

Краска обоих видов производится из сырья M1 и M2. Первые два ограничения задачи порождены ограниченными ежедневными запасами сырья. Другие два отображают ограниченный рыночный спрос на краску. Прибыль от одной тонны краски для наружных работ составляет \$5000, а краски для внутренних работ — \$4000. Целевая функция задачи должна максимизировать общую прибыль. Для удобства вычислений доходность производства красок масштабирована в тысячах долларов.

Эта задача в стандартной форме записывается так:

максимизировать $z = 5x_1 + 4x_2 + 0s_1 + 0s_2 + 0s_3 + 0s_4$

при ограничениях

$$\begin{aligned} 6x_1 + 4x_2 + s_1 &= 24 && \text{(Ограничение на сырье M1)},\\ x_1 + 2x_2 + s_2 &= 6 && \text{(Ограничение на сырье M2)},\\ -x_1 + x_2 + s_3 &= 1 && \text{(Ограничение на спрос)},\\ x_2 + s_4 &= 2 && \text{(Ограничение на спрос)},\\ x_1, x_2, s_1, s_2, s_3, s_4 &\geq 0. \end{aligned}$$

Здесь s_1, s_2, s_3, s_4 — дополнительные (остаточные) переменные, добавленные в неравенства для преобразования их в равенства.

Задачу ЛП в стандартной форме можно представить в виде следующей компактной таблицы.

Базис	z	x_1	x_2	s_1	s_2	s_3	s_4	Решение	
z	1	–5	–4	0	0	0	0	0	z-строка
s_1	0	6	4	1	0	0	0	24	s_1-строка
s_2	0	1	2	0	1	0	0	6	s_2-строка
s_3	0	–1	1	0	0	1	0	1	s_3-строка
s_4	0	0	1	0	0	0	1	2	s_4-строка

Нижние четыре строки этой таблицы представляют равенства ограничений; значения правых частей этих равенств даны в столбце "Решение". Строка z получена из равенства $z - 5x_1 - 4x_2 = 0$.

Дополнительные переменные s_1, s_2, s_3 и s_4 составляют очевидное начальное *допустимое базисное* решение, при этом, поскольку небазисные переменные x_1 и x_2 равны нулю, их значения автоматически отображаются в столбце "Решение": $s_1 = 24$, $s_2 = 6$, $s_3 = 1$ и $s_4 = 2$. Значение целевой функции при этом решении равно нулю.

Будет ли это начальное решение оптимальным? Конечно, нет, поскольку переменные x_1 и x_2 здесь равны нулю, а возрастание этих переменных даже на единицу приводит к увеличению значения целевой функции $z = 5x_1 + 4x_2$ на 5 (при увеличении x_1) или 4 единицы (при увеличении x_2). Поскольку коэффициент при переменной x_1 в формуле целевой функции больше, чем коэффициент при x_2, переменную x_1 следует ввести в число базисных (в этом случае она станет *вводимой*). Если обратиться к приведенной выше таблице, то вводимая переменная определяется среди множества небазисных как переменная, имеющая наибольший отрицательный коэффициент в z-строке.

Включаемая переменная x_1 должна принять положительное значение. На рис. 3.1 видно, что, исходя из начальной точки A ($x_1 = 0$, $x_2 = 0$), наибольшее значение, которое можно присвоить переменной x_1 (не выходя из пространства допустимых решений), равно 4, что соответствует точке B ($x_1 = 4$, $x_2 = 0$). Это означает, что решение переместилось от крайней точки A в крайнюю точку B.

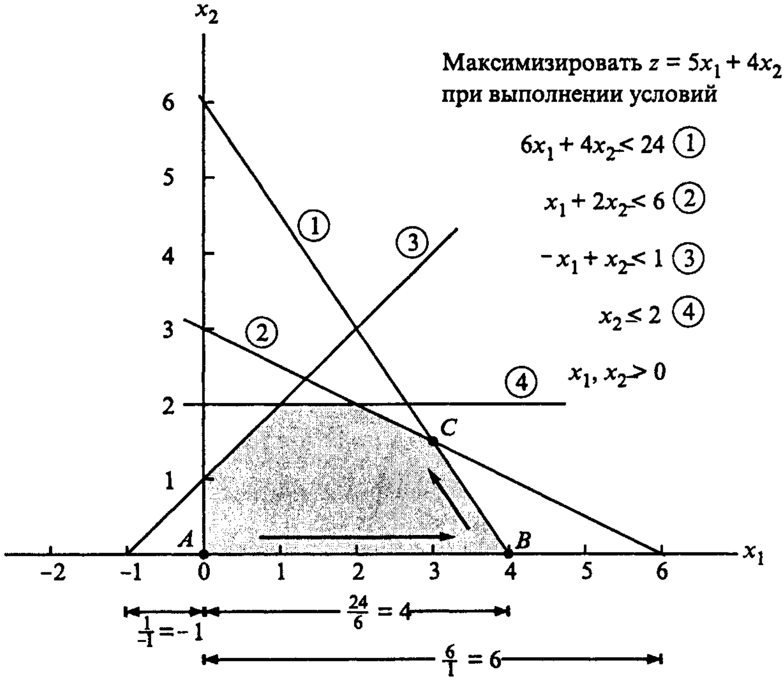


Рис. 3.1

Поскольку симплекс-метод не должен основываться на графическом представлении задачи ЛП, необходимо определить, как выбрать точку B алгебраически. Из рис. 3.1 видно, что B является одной из точек пересечения прямых, соответствующих ограничениям, с осью x_1. Алгебраически точку пересечения можно найти как **отношение** правой части равенства (значение в столбце "Решение") к коэффициенту при переменной x_1 в этом равенстве, как показано в следующей таблице.

Базис	x_1	Решение	Отношение (точка пересечения)
s_1	6	24	24/6 = 4 (Минимум)
s_2	1	6	6/1 = 6
s_3	–1	1	1/(–1) = –1 (Не подходит)
s_4	0	2	2/0 = ∞ (Не подходит)

Нас интересуют только *неотрицательные* отношения (или точки пересечения на положительной полуоси x_1), так как они соответствуют направлению возрастания переменной x_1. Отметим, что все значения в столбце "Решение" неотрицательные, поэтому вычисляемое отношение будет неотрицательным и конечным только в том случае, если знаменатель отношения строго положителен. Именно поэтому мы не рассматриваем отношения, соответствующие третьему и четвертому равенствам, так как там знаменатели или отрицательные, или равны нулю.

Минимальное неотрицательное отношение равно значению вводимой переменной x_1 в новом решении, а именно $x_1 = 4$ (сравните с точкой B на рис. 3.1). Значение целевой функции при этом значении x_1 возрастет до 20 (= 5×4).

Теперь среди текущих базисных переменных s_1, s_2, s_3 и s_4 следует определить *исключаемую переменную*, которая примет нулевое значение после введения в базис переменной x_1. (Напомним, что в этом примере должно быть в точности 4 базисные переменные.) Поскольку наименьшее (неотрицательное) отношение соответствует переменной s_1, в точке B именно эта переменная обращается в нуль. Таким образом, переменная s_1 будет *исключаемой*, в этом случае переменная x_1 автоматически получает значение 4. Замена исключаемой переменной s_1 на вводимую переменную x_1 приводит к новому базисному решению (x_1, s_2, s_3, s_4).

Вычисление нового базисного решения основывается на методе исключения переменных (методе Гаусса–Жордана). В следующей таблице, которая пока совпадает с начальной таблицей задачи ЛП, определим **ведущий столбец**, ассоциируемый с вводимой переменной, и **ведущую строку**, ассоциируемую с исключаемой переменной. Элемент, находящийся на пересечении ведущего столбца и ведущей строки, назовем **ведущим**.

Базис	z	x_1	x_2	s_1	s_2	s_3	s_4	Решение	
z	1	–5	–4	0	0	0	0	0	
s_1	0	6	4	1	0	0	0	24	Ведущая строка
s_2	0	1	2	0	1	0	0	6	
s_3	0	–1	1	0	0	1	0	1	
s_4	0	0	1	0	0	0	1	2	
		Ведущий столбец							

Процесс вычисления нового базисного решения состоит из двух этапов.

1. Вычисление элементов новой ведущей строки.

   Новая ведущая строка = Текущая ведущая строка / Ведущий элемент.

2. Вычисление элементов остальных строк, включая z-строку.

   Новая строка = Текущая строка – Ее коэффициент в ведущем столбце × Новая ведущая строка.

В нашем примере ведущая строка (s_1-строка) делится на ведущий элемент (= 6). В следующей таблице показана новая ведущая строка, где вводимая (небазисная) переменная x_1 заменила исключаемую переменную s_1. В столбце "Решение" получаем новое значение переменной x_1 (= 4).

Базис	z	x_1	x_2	s_1	s_2	s_3	s_4	Решение	
z									
x_1	0	1	4/6	1/6	0	0	0	24/6 = 4	Новая ведущая строка
s_2									
s_3									
s_4									

Вычисляем элементы остальных строк таблицы.

1. z-строка.

Текущая z-строка: $(1\ \ -5\ \ -4\ \ 0\ \ 0\ \ 0\ \ 0\ \ |\ \ 0)$

$-(-5)\times$ Новая ведущая строка: $(0\ \ 5\ \ 10/3\ \ 5/6\ \ 0\ \ 0\ \ 0\ \ |\ \ 20)$

= Новая z-строка: $(1\ \ 0\ \ -2/3\ \ 5/6\ \ 0\ \ 0\ \ 0\ \ |\ \ 20)$

2. s_2-строка.

Текущая s_2-строка: $(0\ \ 1\ \ 2\ \ 0\ \ 1\ \ 0\ \ 0\ \ |\ \ 6)$

$-(1)\times$ Новая ведущая строка: $(0\ \ -1\ \ -2/3\ \ -1/6\ \ 0\ \ 0\ \ 0\ \ |\ \ -4)$

= Новая s_2-строка: $(0\ \ 0\ \ 4/3\ \ -1/6\ \ 1\ \ 0\ \ 0\ \ |\ \ 2)$

3. s_3-строка.

Текущая s_3-строка: $(0\ \ -1\ \ 1\ \ 0\ \ 0\ \ 1\ \ 0\ \ |\ \ 1)$

$-(1)\times$ Новая ведущая строка: $(0\ \ 1\ \ 2/3\ \ 1/6\ \ 0\ \ 0\ \ 0\ \ |\ \ 4)$

= Новая s_3-строка: $(0\ \ 0\ \ 5/3\ \ 1/6\ \ 0\ \ 1\ \ 0\ \ |\ \ 5)$

4. s_4-строка. Новая s_4-строка повторяет текущую s_4-строку, поскольку ее коэффициент в *ведущем столбце* равен нулю.

Новая симплекс-таблица, соответствующая новому базисному решению (x_1, s_2, s_3, s_4), имеет следующий вид.

Базис	z	x_1	x_2	s_1	s_2	s_3	s_4	Решение
z	1	0	–2/3	5/6	0	0	0	20
x_1	0	1	2/3	1/6	0	0	0	4
s_2	0	0	4/3	–1/6	1	0	0	2
s_3	0	0	5/3	1/6	0	1	0	5
s_4	0	0	1	0	0	0	1	2

Отметим, что новая таблица обладает теми же свойствами, что и начальная: только небазисные переменные x_2 и s_1 равны нулю, в столбце "Решение" представлено новое базисное решение ($x_1 = 4$, $s_2 = 2$, $s_3 = 5$, $s_4 = 2$) вместе с новым значением целевой функции z (= 20). Это результат применения метода Гаусса–Жордана.

Из последней таблицы видно, что полученное базисное решение не является оптимальным, поскольку в z-строке переменная x_2 имеет отрицательный коэффициент. Так же, как и в начальной таблице, строку z можно интерпретировать как уравнение

$$z = \frac{2}{3}x_2 - \frac{5}{6}s_1 + 20.$$

Из последнего уравнения следует, что увеличение значения переменной x_2 (ее текущее значение равно нулю) приведет к увеличению значения целевой функции. Таким образом, переменная x_2 должна стать вводимой в базис.

Далее определим исключаемую переменную. Для этого вычислим отношения правых частей равенств, соответствующих ограничениям, к коэффициентам, стоящим при x_2 в этих равенствах.

Базис	x_1	Решение	Отношение
x_1	2/3	4	4/(2/3) = 6
s_2	4/3	2	2/(4/3) = 3/2
s_3	5/3	5	5/(5/3) = 3
s_4	1	2	2/1 = 2

Вычисления показывают, что минимальное (неотрицательное) отношение соответствует переменной s_2, которая становится исключаемой; при этом значение отношения (= 3/2) равно новому значению переменной x_2. Соответствующее увеличение значения целевой функции составит $\frac{2}{3}\times\frac{3}{2} = 1$ и $z = 20 + 1 = 21$.

В этой ситуации ведущей строкой будет s_2-строка, а ведущим столбцом — столбец, соответствующий переменной x_2. Ведущий элемент равен 4/3.

Вычисляем элементы новой симплекс-таблицы.

1. Новая ведущая (s_2-) строка = Текущая s_2-строка / $\frac{4}{3}$.
2. Новая строка z-строка = Текущая z-строка – $(-\frac{2}{3})$ × Новая ведущая строка.
3. Новая x_1-строка = Текущая x_1-строка – $\frac{2}{3}$ × Новая ведущая строка.
4. Новая s_3-строка = Текущая s_3-строка – $\frac{5}{3}$ × Новая ведущая строка.
5. Новая s_4-строка = Текущая s_4-строка – 1 × Новая ведущая строка.

Эти вычисления приводят к следующей таблице.

Базис	z	x_1	x_2	s_1	s_2	s_3	s_4	Решение
z	1	0	0	3/4	1/2	0	0	21
x_1	0	1	0	1/4	–1/2	0	0	3
x_2	0	0	1	–1/8	3/4	0	0	3/2
s_3	0	0	0	3/8	–5/4	1	0	5/2
s_4	0	0	0	1/8	–3/4	0	1	1/2

Поскольку z-строка не имеет отрицательных коэффициентов, соответствующих небазисным переменным s_1 и s_2, полученное решение оптимально.

Оптимальное решение задачи ЛП можно "считать" из симплекс-таблицы следующим образом. Неотрицательные (базисные) переменные представлены в столбце "Базис", а их значения — в столбце "Решение". В данном примере имеем следующее.

Переменные задачи	Оптимальные значения	Интерпретация
x_1	3	Ежедневно следует производить 3 т краски для наружных работ
x_2	3/2	Ежедневно следует производить 1.5 т краски для внутренних работ
z	21	Ежедневный доход составляет $21 000

С помощью симплекс-таблицы можно получить много дополнительной информации (кроме непосредственно оптимального решения).

1. Состояние ресурсов.
2. Цена единицы ресурсов (двойственные цены).
3. Все данные, необходимые для проведения анализа чувствительности оптимального решения.

Здесь мы покажем, как определить состояние (статус) ресурсов. Нахождение двойственных цен и использование данных симплекс-таблицы для проведения анализа чувствительности описано в главе 4.

Статус ресурса определяется как дефицитный или недефицитный, в зависимости от того, будет он использован полностью или нет. Эту информацию можно получить из результирующей симплекс-таблицы путем проверки значений дополнительных (остаточных) переменных, ассоциируемых с соответствующими ограничениями, накладываемыми на ресурсы. Если дополнительная переменная равно нулю, значит, ресурс использован полностью, и он получает статус дефицитного. Положительное значение дополнительной переменной указывает на недефицитность соответствующего ресурса.

В задаче о компании Reddy Mikks четыре ограничения классифицируются следующим образом.

Ресурс	Остаточная переменная	Статус
Сырье M1	$s_1 = 0$	Дефицитный
Сырье M2	$s_2 = 0$	Дефицитный
Ограничение на спрос 1	$s_3 = 2.5$	Недефицитный
Ограничение на спрос 2	$s_4 = 0.5$	Недефицитный

Статус ресурсов показывает, что, поскольку M1 и M2 дефицитны, их увеличение может привести к улучшению оптимального решения (т.е. увеличению значения целевой функции). Процедура анализа чувствительности может помочь определить, на сколько можно увеличить объемы ресурсов. Этот вопрос исследовался в разделах 2.4 и 2.5, полное его решение представлено в главе 4.

В примере 3.3–1 велся поиск максимума целевой функции. В случае минимизации целевой функции исключаемые переменные определяются точно так же, как и при максимизации целевой функции. Как показано в разделе 3.2.1, задача минимизации сводится к задаче максимизации простым соотношением $\max z = -\min(-z)$. Поэтому в случае минимизации целевой функции вводимая переменная выбирается как небазисная с наибольшим *положительным* коэффициентом в z-строке симплекс-таблицы, а минимум целевой функции будет достигнут тогда, когда все коэффициенты в z-строке будут неположительные.

Два правила выбора вводимых и исключающих переменных в симплекс-методе назовем **условием оптимальности** и **условием допустимости**. Сформулируем эти правила, а также рассмотрим последовательность действий, выполняемых при реализации симплекс-метода.

Условие оптимальности. Вводимой переменной в задаче максимизации (минимизации) является *небазисная* переменная, имеющая наибольший отрицательный (положительный) коэффициент в z-строке. Если в z-строке есть несколько таких коэффициентов, то выбор вводимой переменной делается произвольно. Оптимальное решение достигнуто тогда, когда в z-строке все коэффициенты при небазисных переменных будут неотрицательными (неположительными).

Условие допустимости. Как в задаче максимизации, так и в задаче минимизации в качестве исключаемой выбирается *базисная* переменная, для которой отношение значения правой части ограничения к положительному коэффициенту ведущего столбца минимально. Если базисных переменных с таким свойством несколько, то выбор исключаемой переменной выполняется произвольно.

Теперь приведем последовательность действий, выполняемых в симплекс-методе.

Шаг 0. Находится начальное допустимое базисное решение.

Шаг 1. На основе условия оптимальности определяется *вводимая* переменная. Если вводимых переменных нет, вычисления заканчиваются.

Шаг 2. На основе условия допустимости выбирается исключаемая переменная.

Шаг 3. Методом Гаусса–Жордана вычисляется новое базисное решение. Переход к шагу 1.

Вычисления в симплекс-методе выполняются итерационно в том смысле, что условия оптимальности и допустимости, а также вычисления применяются к текущей симплекс-таблице, в результате чего получается следующая таблица. Мы будем называть последовательные симплекс-таблицы **итерациями**.

Отслеживание последовательных базисных решений из примера 3.3–1 по графическому представлению пространства решений (рис. 3.1) показывает, что итерационный процесс симплекс-метода начался в крайней (угловой) точке A, затем переместился в крайнюю точку B и закончился в крайней точке C — точке оптимума. Эти итерации соответствуют **смежным** крайним точкам границы пространства решений. Алгоритм симплекс-метода никогда не сможет переместиться сразу из точки A в точку C. (См. раздел 7.8, где описан алгоритм, использующий внутренние точки пространства решений.)

Упражнения 3.3,а

1. Это упражнение должно показать значимую роль условия допустимости в симплекс-методе. В примере 3.3–1 в первой симплекс-таблице мы использовали минимальное неотрицательное значение из множества вычисленных отношений для определения

исключаемой переменной. Такой выбор исключаемой переменной гарантирует, что новое значение базисной переменной не будет отрицательным. Для иллюстрации этого утверждения примените программу TORA к примеру 3.3–1 и на первом шаге с помощью программной опции User-guided вместо переменной s_1 исключите из базисного решения переменную s_2. Далее посмотрите на результирующую симплекс-таблицу, вы должны увидеть, что переменная s_1 приняла отрицательное значение (= –12). Это показывает, что полученное решение недопустимо.

2. Рассмотрим следующее множество ограничений.

$$\begin{gathered} x_1 + 2x_2 + 2x_3 + 4x_4 \le 40, \\ 2x_1 - x_2 + x_3 + 2x_4 \le 8, \\ 4x_1 - 2x_2 + x_3 - x_4 \le 10, \\ x_1, x_2, x_3, x_4 \ge 0. \end{gathered}$$

С учетом этих ограничений с помощью программы TORA решите следующие задачи ЛП.

a) Максимизировать $z = 2x_1 + x_2 - 3x_3 + 5x_4$.
b) Максимизировать $z = 8x_1 + 6x_2 + 3x_3 - 2x_4$.
c) Максимизировать $z = 3x_1 - x_2 + 3x_3 + 4x_4$.
d) Минимизировать $z = 5x_1 - 4x_2 + 6x_3 - 8x_4$.
e) Минимизировать $z = -4x_1 + 6x_2 - 2x_3 + 4x_4$.

3. Следующая таблица представляет отдельную итерацию симплекс-метода. Все переменные неотрицательные.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	Решение
z	0	–5	0	4	–1	–10	0	0	620
x_8	0	3	0	–2	–3	–1	5	1	12
x_3	0	1	1	3	1	0	3	0	6
x_1	1	–1	0	0	6	–4	0	0	0

a) Определите исключаемую переменную, если вводимой является (1) x_2, (2) x_4, (3) x_5, (4) x_6, (5) x_7.
b) Для каждой из вводимых переменных, перечисленных в п. а), без использования метода Гаусса–Жордана найдите ее значение и вычислите значение целевой функции.

4. Дана следующая система уравнений:

$$\begin{aligned} &x_1 + 2x_2 - 3x_3 + 5x_4 + x_5 = 4, \\ &5x_1 - 2x_2 + 6x_4 + x_6 = 8, \\ &2x_1 + 3x_2 - 2x_3 + 3x_4 + x_7 = 3, \\ &-x_1 + x_3 + 2x_4 + x_8 = 0, \\ &x_1, x_2, \ldots, x_8 \ge 0. \end{aligned}$$

Пусть переменные x_5, x_6, x_7 и x_8 составляют начальное допустимое базисное решение. Предположим, что в базис вводится переменная x_1. Определите, какую из переменных текущего базисного решения следует исключить из базиса так, чтобы все переменные остались неотрицательными. Найдите значение переменной x_1 в этом новом базисе. Повторите это упражнение для переменных x_2, x_3 и x_4.

5. a) Решите следующую задачу ЛП, обосновывая и проверяя ход решения; исходите из предположений, на которых основывается симплекс-метод.

$$\text{Максимизировать } z = x_1$$

при ограничениях

$$5x_1 + x_2 = 4,$$
$$6x_1 + x_3 = 8;$$
$$3x_1 + x_4 = 3,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

b) Повторите решения задачи ЛП при тех же ограничениях, что и в п. a), но для минимизации целевой функции $z = x_1$.

6. Решите следующую задачу ЛП, обосновывая и проверяя ход решения; исходите из предположений, на которых основывается симплекс-метод.

$$\text{Максимизировать } z = 5x_1 - 6x_1 + 3x_3 - 5x_4 + 12x_5$$

при ограничениях

$$x_1 + 3x_2 + 5x_3 + 6x_4 + 3x_5 \leq 90,$$
$$x_1, x_2, x_3, x_4, x_5 \geq 0.$$

(*Подсказка*. Базисное решение содержит только одну переменную.)

7. Рассмотрите двухмерное пространство решений на рис. 3.2.

a) Определите графическим способом точку оптимума в предположении, что целевая функция имеет следующий вид:

$$\text{Максимизировать } z = 3x_1 + 6x_2.$$

b) Предполагая, что реализация симплекс-метода начинается в точке A, определите последовательность угловых точек, приводящих к точке оптимума.

c) Определите вводимую переменную, значения соответствующих отношений в условии допустимости и значение целевой функции, полагая, что реализация симплекс-метода начинается в точке A и целевая функция имеет вид

$$\text{Максимизировать } z = 4x_1 + x_2.$$

d) Повторите п. c) применительно к целевой функции

$$\text{Максимизировать } z = x_1 + 4x_2.$$

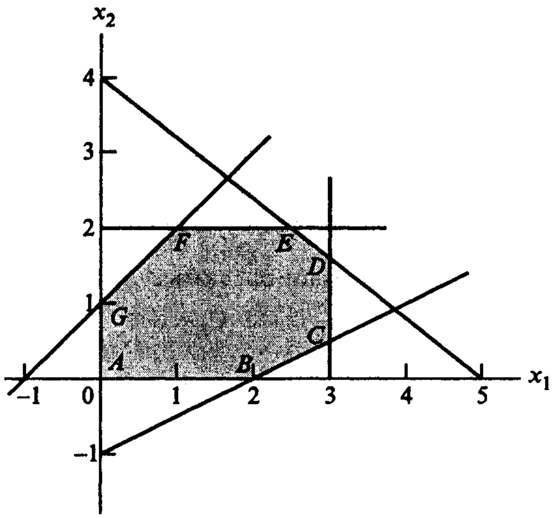


Рис. 3.2

8. На рис. 3.3 показано пространство допустимых решений трехмерной задачи ЛП с крайними точками $A, B, C, \ldots, J$.

   a) Являются ли следующие пары крайних точек смежными: (A, B), (B, D), (E, H), (A, I)?

   b) Предположим, что реализация симплекс-метода начинается в точке A и заканчивается в точке оптимума H. Определите, какие из следующих последовательностей крайних точек могут привести к точке оптимума. Обоснуйте свой вывод.

      i) $A \to B \to G \to H$.

      ii) $A \to E \to I \to H$.

      iii) $A \to C \to E \to B \to A \to D \to H$.

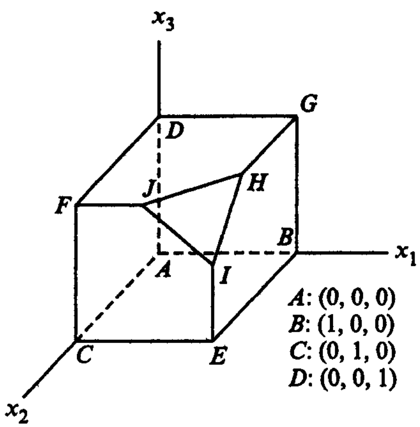


Рис. 3.3

9. Пусть в задаче ЛП, которой соответствует пространство допустимых решений, показанное на рис. 3.3, все ограничения являются неравенствами типа "≤", а все переменные задачи (т.е. x_1, x_2 и x_3) неотрицательны. Обозначим через s_1, s_2, s_3 и s_4 дополнительные (неотрицательные) переменные, ассоциируемые с ограничениями, представленными плоскостями *CEIJF*, *BEIHG*, *DFJHG* и *IJH* соответственно. Определите базисные и небазисные переменные для каждой крайней точки пространства допустимых решений.

10. Пусть в задаче ЛП, для которой пространство допустимых решений представлено на рис. 3.3, реализация симплекс-метода начинается в точке *A*. Определите вводимую переменную на *первой* итерации симплекс-метода, а также ее значение и значение целевой функции, если целевая функция имеет следующий вид.

    a) Максимизировать $z = x_1 - 2x_2 + 3x_3$.
    b) Максимизировать $z = 5x_1 + 2x_2 + 4x_3$.
    c) Максимизировать $z = -2x_1 + 7x_2 + 2x_3$.
    d) Максимизировать $z = x_1 + x_2 + x_3$.

11. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 16x_1 + 15x_2$$

при ограничениях

$$40x_1 + 31x_2 \le 124,$$
$$-x_1 + x_2 \le 1,$$
$$x_1 \le 3,$$
$$x_1, x_2 \ge 0.$$

Используя программу TORA, выполните следующее.

    a) Решите поставленную задачу симплекс-методом, выбирая в качестве вводимой небазисную переменную, имеющую *наибольший по абсолютной величине* отрицательный коэффициент в z-строке симплекс-таблицы.
    b) Решите поставленную задачу симплекс-методом, выбирая в качестве вводимой небазисную переменную, имеющую *наименьший по абсолютной величине* отрицательный коэффициент в z-строке симплекс-таблицы.
    c) Подсчитайте количество итераций, используемых для решения задачи в пп. а) и b). Действительно ли выбор вводимой переменной среди небазисных, имеющих *наибольший по абсолютной величине* отрицательный коэффициент в z-строке, ведет к уменьшению числа итераций, выполняемых при реализации симплекс-метода? Обоснуйте это наблюдение.
    d) Предположим, что задача максимизации целевой функции z заменена на задачу минимизации путем умножения функции z на -1. Как это скажется на количестве итераций симплекс-метода?

12. Компания Gutchi производит дорожные сумки, чемоданы и рюкзаки. Для трех видов изделий используется натуральная кожа и синтетические материалы, причем кожа считается ограниченным ресурсом. Производство этих изделий также требу-

ет выполнения двух ручных операций: прошивка и окончательная отделка изделия. В следующей таблице приведены ограничения на используемые в производстве ресурсы, а также отпускная цена каждого вида изделия.

Ресурс	Расход ресурса на производство одного изделия			Ограничение на ресурс (ежедневно)
	Сумка	Чемодан	Рюкзак	
Кожа (кв. футы)	2	1	3	42
Прошивка (часы)	2	1	2	40
Отделка (часы)	1	0.5	1	45
Отпускная цена ($)	24	22	45	

Сформулируйте задачу линейного программирования, с помощью программы TORA найдите ее оптимальное решение и определите статус ресурсов.

3.4. Искусственное начальное решение

В примере 3.3–1 при начальном допустимом базисном решении гарантировалось, что все последующие базисные решения, получаемые при выполнении симплекс-метода, также будут допустимыми. В задачах линейного программирования, где все ограничения являются неравенствами типа "≤" (с неотрицательной правой частью), дополнительные (остаточные) переменные позволяют сформировать начальное допустимое базисное решение. Естественно, возникает вопрос: как найти начальное допустимое базисное решение в задачах ЛП, где есть ограничения в виде равенств или неравенств типа "≥"?

Наиболее общим способом построения начального допустимого базисного решения задачи ЛП является использование **искусственных переменных**. Эти переменные в первой итерации играют роль дополнительных остаточных переменных, но на последующих итерациях от них освобождаются. Разработано два тесно связанных между собой метода нахождения начального решения, которые используют искусственные переменные: *M*-метод[1] и двухэтапный метод.

3.4.1. *M*-метод

Пусть задача ЛП записана в стандартной форме. Для любого равенства i, в котором не содержится дополнительная остаточная переменная, введем искусственную переменную R_i, которая далее войдет в начальное базисное решение. Но поскольку эта переменная искусственна (другими словами, не имеет никакого "физического смысла" в данной задаче), необходимо сделать так, чтобы на последующих итерациях она обратилась в нуль. Для этого в выражение целевой функции вводят **штраф**.

Переменная R_i, с помощью достаточно большого положительного числа M, штрафуется путем ввода в целевую функцию выражения $-MR_i$ в случае максимизации целевой функции и выражения $+MR_i$ — в случае минимизации. Вследствие этого штрафа естественно предположить, что процесс оптимизации симплекс-метода приведет к нулевому значению переменной R_i. Следующий пример проясняет детали этого метода.

[1] *M*-метод также называют методом больших штрафов. — *Прим. ред.*

Пример 3.4–1

$$\text{Минимизировать } z = 4x_1 + x_2$$

при выполнении условий

$$3x_1 + x_2 = 3,$$
$$4x_1 + 3x_2 \geq 6,$$
$$x_1 + 2x_2 \leq 4,$$
$$x_1, x_2 \geq 0.$$

Стандартная форма этой задачи получается в результате добавления дополнительной (избыточной) переменной x_3 во второе неравенство и дополнительной (остаточной) переменной x_4 в третье неравенство. Эта задача в стандартной форме будет записана следующим образом.

$$\text{Минимизировать } z = 4x_1 + x_2$$

при выполнении условий

$$3x_1 + x_2 = 3,$$
$$4x_1 + 3x_2 - x_3 = 6,$$
$$x_1 + 2x_2 + x_4 = 4,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

В полученной задаче первое и второе уравнения не имеют дополнительных (остаточных) переменных, которые можно ввести в базисное решение. Поэтому введем в эти уравнения искусственные переменные R_1 и R_2, а в целевую функцию добавим штраф $MR_1 + MR_2$. В результате получим следующую задачу ЛП.

$$\text{Минимизировать } z = 4x_1 + x_2 + MR_1 + MR_2$$

при выполнении условий

$$3x_1 + x_2 + R_1 = 3,$$
$$4x_1 + 3x_2 - x_3 + R_2 = 6,$$
$$x_1 + 2x_2 + x_4 = 4,$$
$$x_1, x_2, x_3, x_4, R_1, R_2, \geq 0.$$

В этой модифицированной задаче переменные R_1, R_2 и x_4 можно использовать в качестве начального допустимого базисного решения. В результате получим следующую симплекс-таблицу.

Базис	x_1	x_2	x_3	R_1	R_2	x_4	Решение
z	–4	–1	0	$-M$	$-M$	0	0
R_1	3	1	0	1	0	0	3
R_2	4	3	–1	0	1	0	6
x_4	1	2	0	0	0	1	4

Прежде чем применять симплекс-метод, надо согласовать значения в z-строке с остальной частью таблицы. В частности, значение функции z, соответствующее начальному базисному решению $R_1 = 3$, $R_2 = 6$ и $x_4 = 4$, должно равняться $3M + 6M + 0 = 9M$, а не 0, как показано в таблице. Это несоответствие связано с тем, что переменным R_1 и R_2 соответствуют ненулевые коэффициенты ($-M$, $-M$) в строке z (сравните с начальным решением в примере 3.3–1, где дополнительным переменным соответствуют нулевые коэффициенты в z-строке). Чтобы сделать эти коэффициенты нулевыми, следует умножить элементы строк R_1 и R_2 на величину M, и затем сложить эти строки с z-строкой. (Обратите внимание на "подсвеченные" единицы в этих строках — если бы эти коэффициенты были отличны от единиц, то необходимо было бы сначала разделить все элементы этих строк на данные коэффициенты.) Кратко это действие можно записать следующим образом.

$$\text{Новая } z\text{-строка} = \text{Старая } z\text{-строка} + M \times R_1\text{-строка} + M \times R_2\text{-строка}$$

Выполним эту операцию в данном примере.

$$\begin{array}{rcccccc|c}
\text{Старая } z\text{-строка: } (-4 & -1 & 0 & -M & -M & 0 & & 0) \\
+ M \times R_1\text{-строка: } (3M & M & 0 & M & 0 & 0 & & 3M) \\
+ M \times R_2\text{-строка: } (4M & 3M & -M & 0 & M & 0 & & 6M) \\
= \text{Новая } z\text{-строка: } (-1 + 7M & -1 + 4M & -M & 0 & 0 & 0 & & 9M)
\end{array}$$

Измененная симплекс-таблица примет следующий вид.

Базис	x_1	x_2	x_3	R_1	R_2	x_4	Решение
z	$-4 + 7M$	$-1 + 4M$	$-M$	0	0	0	$9M$
R_1	3	1	0	1	0	0	3
R_2	4	3	-1	0	1	0	6
x_4	1	2	0	0	0	1	4

Отметим, что теперь $z = 9M$, что соответствует базисному решению $R_1 = 3$, $R_2 = 6$ и $x_4 = 4$.

Последняя таблица готова к применению симплекс-метода с использованием условий оптимальности и допустимости. Поскольку мы минимизируем целевую функцию, находим наибольший *положительный* коэффициент в z-строке. Наибольший коэффициент $-4 + 7M$ соответствует переменной x_1, которая и будет вводимой. Условие допустимости указывает на переменную R_1 в качестве исключаемой.

Поскольку вводимая и исключаемая переменные определены, новую симплекс-таблицу можно вычислить с помощью метода Гаусса–Жордана. Заметим, что вычисления в z-строке, где присутствует M, следует проводить алгебраически. Так, для получения новой z-строки надо новую ведущую строку умножить на $-(-4 + 7M)$ и сложить с текущей z-строкой. В результате получим следующую таблицу.

Базис	x_1	x_2	x_3	R_1	R_2	x_4	Решение
z	0	$(1 + 5M)/3$	$-M$	$(4 - 7M)/3$	0	0	$4 + 2M$
x_1	1	1/3	0	1/3	0	0	1
R_2	0	5/3	-1	$-4/3$	1	0	2
x_4	0	5/3	0	$-1/3$	0	1	3

Отметим, что уже первая итерация исключила из базисного решения искусственную переменную R_1, что является результатом включения штрафа в целевую функцию.

Последняя таблица показывает, что следующими, вводимой и исключаемой, переменными будут x_2 и R_2 соответственно. Конечно, для получения оптимального решения может потребоваться больше двух итераций. В данной задаче оптимальным решением будет $x_1 = 2/5$, $x_2 = 9/5$, $x_3 = 1$ и $z = 17/5$ (см. упр. 3.4,а(1)).

При использовании M-метода следует обратить внимание на следующие два обстоятельства.

1. Использование штрафа M может и не привести к исключению искусственной переменной в конечной симплекс-итерации. Если исходная задача линейного программирования не имеет допустимого решения (например, система ограничений несовместна), тогда в конечной симплекс-итерации, по крайней мере, одна искусственная переменная будет иметь положительное значение. Это "индикатор" того, что задача не имеет допустимого решения.
2. Теоретически применение M-метода требует, чтобы $M \to \infty$. Однако с точки зрения компьютерных вычислений величина M должна быть конечной и, вместе с тем, *достаточно большой*. Как понимать термин "достаточно большая" — это открытый вопрос. Величина M должна быть настолько большой, чтобы выполнять роль "штрафа", но не слишком большой, чтобы не уменьшить точность вычислений. На практике вы должны помнить о возможных ошибках машинного округления при выполнении вычислений, в которых совместно участвуют как большие, так и малые числа. Эту проблему иллюстрирует следующий пример.

Пример 3.4–2

Рассмотрим следующую задачу.

$$\text{Максимизировать } z = 0.2x_1 + 0.5x_2$$

при выполнении условий

$$\begin{aligned} 3x_1 + 2x_2 &\geq 6, \\ x_1 + 2x_2 &\leq 4, \\ x_1, x_2 &\geq 0. \end{aligned}$$

Проведем два эксперимента для демонстрации влияния величины M на результат вычислений симплекс-метода.

Эксперимент 1

С помощью программы TORA найдите решение симплекс-методом поставленной задачи при $M = 10$ и повторите вычисления при $M = 999\,999$. При первом значении M получим правильное решение $x_1 = 1$ и $x_2 = 1.5$, тогда как во втором случае получим явно неверное решение $x_1 = 4$, $x_2 = 0$.

Эксперимент 2

Умножив целевую функцию на 1000, получим новую целевую функцию $z = 200x_1 + 500x_2$. Решим эту задачу при $M = 10$ и $M = 999\,999$. Теперь именно второе значение M обеспечивает нахождение правильного решения.

Из этих экспериментов можно сделать вывод, что правильный выбор значения M зависит от данных исходной задачи. Бездумное следование теоретическому требованию, что значение M должно быть "очень большим", может привести к значительным ошибкам округления. Именно поэтому M-метод никогда не применяется в коммерческих программах, реализующих симплекс-метод. Вместо него используется *двухэтапный метод*, который будет описан в следующем разделе.

Упражнения 3.4,а

1. Завершите решение задачи из примера 3.4–1 и получите оптимальное решение.
2. В примере 3.4–1 найдите начальную симплекс-таблицу для каждого из следующих (независимых) случаев и подсчитайте коэффициенты z-строки после определения всех искусственных переменных.

   a) Третье ограничение имеет вид $x_1 + 2x_2 \geq 4$.
   b) Второе ограничение имеет вид $4x_1 + 3x_2 \leq 6$.
   c) Второе ограничение имеет вид $4x_1 + 3x_2 = 6$.
   d) Целевая функция имеет вид: максимизировать $z = 4x_1 + x_2$.

3. Существует следующее множество ограничений.

$$\begin{aligned} -2x_1 + 3x_2 &= 3, && (1)\\ 4x_1 + 5x_2 &\geq 10, && (2)\\ x_1 + 2x_2 &\leq 5, && (3)\\ 6x_1 + 7x_2 &\leq 3, && (4)\\ 4x_1 + 8x_2 &\geq 5, && (5)\\ x_1, x_2 &\geq 0. \end{aligned}$$

   Для каждой из следующих задач найдите коэффициенты z-строки симплекс-таблицы после введения искусственных переменных.

   a) Максимизировать $z = 5x_1 + 6x_2$ при ограничениях (1), (3) и (4).
   b) Максимизировать $z = 2x_1 - 7x_2$ при ограничениях (1), (2), (4) и (5).
   c) Минимизировать $z = 3x_1 + 6x_2$ при ограничениях (3), (4) и (5).
   d) Минимизировать $z = 4x_1 + 6x_2$ при ограничениях (1), (2) и (5).
   e) Минимизировать $z = 3x_1 + 2x_2$ при ограничениях (1) и (2).

4. Дано следующее множество ограничений:

$$\begin{aligned} &x_1 + x_2 + x_3 = 7,\\ &2x_1 - 5x_2 + x_3 \geq 10,\\ &x_1, x_2, x_3 \geq 0. \end{aligned}$$

   При этих ограничениях решите задачи ЛП для следующих целевых функций.

   a) Максимизировать $z = 2x_1 + 3x_2 - 5x_3$.
   b) Минимизировать $z = 2x_1 + 3x_2 - 5x_3$.
   c) Максимизировать $z = x_1 + 2x_2 + x_3$.
   d) Минимизировать $z = 4x_1 - 8x_2 + 3x_3$.

5. Дана следующая задача.

$$\text{Максимизировать } z = 2x_1 + 4x_2 + 4x_3 - 3x_4$$

при ограничениях

$$x_1 + x_2 + x_3 = 4,$$
$$x_1 + 4x_2 + x_4 = 8,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

Решите эту задачу без искусственных переменных, используя в начальном базисном решении переменные x_3 и x_4.

6. Без применения искусственных переменных решите следующую задачу, используя в начальном базисном решении переменные x_3 и x_4.

$$\text{Минимизировать } z = 3x_1 + 2x_2 + 3x_3$$

при ограничениях

$$x_1 + 4x_2 + x_3 \geq 7,$$
$$2x_1 + x_2 + x_4 \geq 10,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

7. Дана следующая задача.

$$\text{Максимизировать } z = x_1 + 5x_2 + 3x_3$$

при ограничениях

$$x_1 + 2x_2 + x_3 = 3,$$
$$2x_1 - x_2 = 4,$$
$$x_1, x_2, x_3 \geq 0.$$

В первом равенстве переменная x_3 может войти в базисное решение вместо искусственной переменной. Однако во втором равенстве искусственная переменная R_2 необходима. Используя начальное базисное решение, состоящее из переменных x_3 и R_2, найдите оптимальное решение этой задачи.

8. Покажите, как с помощью M-метода можно определить, что следующая задача не имеет допустимого решения.

$$\text{Максимизировать } z = 2x_1 + 5x_2$$

при ограничениях

$$3x_1 + 2x_2 \geq 6,$$
$$2x_1 + x_2 \leq 2,$$
$$x_1, x_2 \geq 0.$$

3.4.2. Двухэтапный метод

Пример 3.4–2 демонстрирует проблемы, которые могут возникнуть при M-методе вследствие ошибок округления. Двухэтапный метод полностью лишен тех недостатков, которые присущи M-методу. Как следует из названия этого метода, процесс решения задачи ЛП разбивается на два этапа. На первом этапе ведется поиск начального допустимого базисного решения. Если такое решение найдено, то на втором этапе решается исходная задача.

Этап1. Задача ЛП записывается в стандартной форме, а в ограничения добавляются необходимые искусственные переменные (как и в M-методе) для получения начального базисного решения. Решается задача ЛП *минимизации* суммы искусственных переменных с исходными ограничениями. Если минимальное значение этой новой целевой функции больше нуля, значит, исходная задача не имеет допустимого решения, и процесс вычислений заканчивается. (Напомним, что положительные значения искусственных переменных указывают на то, что исходная система ограничений несовместна.) Если новая целевая функция равна нулю, переходим ко второму этапу.

Этап2. Оптимальное базисное решение, полученное на первом этапе, используется как начальное допустимое базисное решение *исходной* задачи.

Пример 3.4–3

К задаче из примера 3.4–1 применим двухэтапный метод.

Этап 1

$$\text{Минимизировать } r = R_1 + R_2$$

с ограничениями

$$\begin{aligned}
&3x_1 + x_2 + R_1 = 3,\\
&4x_1 + 3x_2 - x_3 \ + R_2 = 6,\\
&x_1 + 2x_2 + x_4 = 4,\\
&x_1, x_2, x_3, x_4, R_1, R_2, \geq 0.
\end{aligned}$$

Соответствующая таблица имеет следующий вид.

Базис	x_1	x_2	x_3	R_1	R_2	x_4	Решение
r	0	0	0	–1	–1	0	0
R_1	3	1	0	1	0	0	3
R_2	4	3	–1	0	1	0	6
x_4	1	2	0	0	0	1	4

Как и в M-методе, сначала вычисляется новая r-строка.

$$\begin{aligned}
\text{Старая } r\text{-строка: } &(0 \quad 0 \quad 0 \quad -1 \quad -1 \quad 0 \mid 0)\\
+1 \times R_1\text{-строка: } &(3 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \mid 3)\\
+1 \times R_2\text{-строка: } &(4 \quad 3 \quad -1 \quad 0 \quad 1 \quad 0 \mid 6)\\
= \text{Новая } r\text{-строка: } &(7 \quad 4 \quad -1 \quad 0 \quad 0 \quad 0 \mid 9)
\end{aligned}$$

Новая r-строка

$$r + 7x_1 + 4x_2 - x_3 + 0R_1 + 0R_2 + 0x_4 = 9$$

используется для решения задачи первого этапа, что приведет к следующему оптимальному решению (проверьте!).

Базис	x_1	x_2	x_3	R_1	R_2	x_4	Решение
r	0	0	0	–1	–1	0	0
x_1	1	0	1/5	3/5	–1/5	0	3/5
x_2	0	1	–3/5	–4/5	3/5	0	6/5
x_4	0	0	1	1	–1	1	1

Поскольку достигнут минимум $r = 0$, значит, на первом этапе получено допустимое базисное решение $x_1 = 3/5$, $x_2 = 6/5$ и $x_4 = 1$. Искусственные переменные полностью выполнили свою "миссию", поэтому из последней таблицы можно удалить их столбцы. Переходим ко второму этапу.

Этап 2

После удаления искусственных переменных *исходная* задача будет записана следующим образом.

$$\text{Минимизировать } z = 4x_1 + x_2$$

с ограничениями

$$x_1 + \frac{1}{5}x_3 = \frac{3}{5},$$

$$x_2 - \frac{3}{5}x_3 = \frac{6}{5},$$

$$x_3 + x_4 = 1,$$

$$x_1, x_2, x_3, x_4 \geq 0.$$

Обратите внимание, что после первого этапа исходная задача претерпела некоторые изменения, которые учитывают полученное базисное решение. Этой трансформированной задаче соответствует следующая таблица.

Базис	x_1	x_2	x_3	x_4	Решение
z	–4	–1	0	0	0
x_1	1	0	1/5	0	3/5
x_2	0	1	–3/5	0	6/5
x_4	0	0	1	1	1

Поскольку базисные переменные x_1 и x_2 имеют ненулевые коэффициенты в z-строке, эту строку следует преобразовать.

Старая z-строка:	(–4	–1	0	0	0)
+ 4 × x_1-строка:	(4	0	4/5	0	12/5)
+ 1 × x_2-строка:	(0	1	–3/5	0	6/5)
= Новая z-строка:	(0	0	1/5	0	18/5)

Начальная таблица второго этапа примет следующий вид.

Базис	x_1	x_2	x_3	x_4	Решение
z	0	0	1/5	0	18/5
x_1	1	0	1/5	0	3/5
x_2	0	1	–3/5	0	6/5
x_4	0	0	1	1	1

Так как решается задача минимизации, следует ввести переменную x_3 в базис. Применение алгоритма симплекс-метода уже на следующей итерации приведет к оптимальному решению (проверьте!).

Удаление искусственных переменных в конце первого этапа имеет смысл только тогда, когда все они являются *небазисными* (как в примере 3.4–2). Однако возможна ситуация, когда в конце первого этапа искусственные переменные *останутся в базисе*, но будут иметь *нулевые значения*. В этом случае такие переменные при необходимости будут формировать часть начального базисного решения для второго этапа. При этом необходимо так изменить вычисления, выполняемые на втором этапе, чтобы искусственные переменные никогда не смогли принять положительные значения ни в каких итерациях симплекс-метода.

Существует простое правило, которое гарантирует, что *нулевая* базисная искусственная переменная на втором этапе никогда не станет положительной. Если в ведущем столбце коэффициент, соответствующий нулевой базисной искусственной переменной, положителен, тогда *ведущий элемент* определяется автоматически (поскольку ему соответствует минимальное отношение, равное нулю) и искусственная переменная на следующей итерации становится небазисной. Если ведущий элемент равен нулю, следующая итерация оставляет искусственную переменную нулевой. И наконец, рассмотрим отрицательный ведущий элемент. В этом случае минимальное отношение не ассоциируется с базисной (нулевой) искусственной переменной. Если минимальное отношение будет положительным, то на следующей итерации искусственная переменная примет положительное значение (обоснуйте это утверждение). Чтобы исключить эту возможность, мы *принуждаем* искусственную переменную всегда оставаться в базисном решении. Поскольку искусственная переменная равна нулю, ее удаление из базисного решения не влияет на то, будет ли допустимым решение из оставшихся в базисе переменных. (Было бы полезно для читателя рассмотреть все описанные случаи с помощью симплекс-таблиц.)

Итак, правило для второго этапа заключается в том, чтобы искусственные переменные оставлять в базисе всегда, когда коэффициент в ведущем столбце положительный или отрицательный. Фактически это правило применяется в конце первого этапа, когда удаляем нулевые искусственные переменные из базисного решения, перед тем как приступить ко второму этапу (см. упр. 3.4,b(5)).

Упражнения 3.4,b

1. Ответьте на следующие вопросы.
   a) Почему на первом этапе двухэтапного метода всегда минимизируется сумма искусственных переменных?

b) Если в задаче ЛП требуется найти максимум целевой функции, то следует ли на первом этапе максимизировать сумму искусственных переменных?

2. Для каждой задачи из упр. 3.4,a(3) запишите соответствующую целевую функцию для первого этапа.

3. Решите задачи из упр. 3.4,a(4) двухэтапным методом.

4. Покажите, что следующая задача не имеет допустимого решения. Запишите задачу ЛП для первого этапа и затем примените программу TORA для поиска ее решения.

$$\text{Максимизировать } z = 2x_1 + 5x_2$$

при выполнении условий

$$\begin{aligned} &3x_1 + 2x_2 \ge 6, \\ &2x_1 + x_2 \le 2, \\ &x_1, x_2 \ge 0. \end{aligned}$$

5. Дана следующая задача.

$$\text{Максимизировать } z = 3x_1 + 2x_2 + 3x_3$$

при выполнении условий

$$\begin{aligned} &2x_1 + x_2 + x_3 \le 2, \\ &3x_1 + 4x_2 + 2x_3 \ge 8, \\ &x_1, x_2, x_3 \ge 0. \end{aligned}$$

a) Используя программу TORA, покажите, что первый этап закончится с нулевой искусственной базисной переменной.

b) Выполните вручную вычисления второго этапа с нулевой искусственной переменной, являющейся частью начального базисного решения. Убедитесь, что искусственные переменные никогда не принимают положительных значений.

c) Покажите, что нулевые искусственные переменные можно удалить из базисного решения на первом этапе (до начала второго) путем выбора вводимой переменной с помощью *ненулевого* ведущего элемента в строке искусственной переменной.

6. Рассмотрите следующую задачу.

$$\text{Максимизировать } z = 3x_1 + 2x_2 + 3x_3$$

при выполнении условий

$$\begin{aligned} &2x_1 + x_2 + x_3 = 2, \\ &x_1 + 3x_2 + x_3 = 6, \\ &3x_1 + 4x_2 + 2x_3 = 8, \\ &x_1, x_2, x_3 \ge 0. \end{aligned}$$

a) С помощью программы TORA покажите, что первый этап закончится с двумя нулевыми искусственными переменными в базисном решении.

b) Покажите, что описанная в упр. 5(c) процедура, применяемая в конце первого этапа, может только одну из двух нулевых искусственных переменных сделать небазисной.

c) Покажите, что исходное ограничение, соответствующее нулевой искусственной переменной, которую нельзя сделать небазисной в п. b), является избыточным. Следовательно, строку этого ограничения, как и его искусственную переменную, можно удалить перед началом второго этапа.

7. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = 3x_1 + 2x_2 + 3x_3$$

при выполнении условий

$$2x_1 + x_2 + x_3 \le 2,$$
$$3x_1 + 4x_2 + 2x_3 \ge 8,$$
$$x_1, x_2, x_3 \ge 0.$$

Оптимальная симплекс-таблица, полученная в конце первого этапа, имеет следующий вид

Базис	x_1	x_2	x_3	x_4	x_5	R	Решение
z	–5	0	–2	–1	–4	0	0
x_2	2	1	1	0	1	0	2
R	–5	0	–2	–1	–4	1	0

Покажите, что небазисные переменные x_1, x_3, x_3 и x_4 никогда не примут положительные значения в конце второго этапа. Следовательно, столбцы этих переменных можно удалить из симплекс-таблицы еще до начала второго этапа. В сущности, удаление этих переменных сводит ограничительные равенства к одному: $x_2 = 2$. Это означает, что не было необходимости выполнять второй этап вовсе, так как пространство допустимых решений состоит только из одной точки.

Из этого упражнения можно сделать общий вывод, что любые небазисные переменные, которые имеют *строго отрицательные* коэффициенты в z-строке в конце вычислений первого этапа, можно удалить из симплекс-таблицы, так как они никогда не примут положительные значения в результате выполнения второго этапа. В связи с этим отметим, что отрицательные коэффициенты в z-строке для "естественных" (не искусственных) переменных могут появиться только в том случае, если в конце первого этапа в базисе присутствуют искусственные (нулевые) переменные.

8. Дана задача линейного программирования.

$$\text{Минимизировать } z = 2x_1 - 4x_2 + 3x_3$$

при выполнении условий

$$5x_1 - 6x_2 + 2x_3 \ge 5,$$
$$-x_1 + 3x_2 + 5x_3 \ge 8,$$

$$2x_1 + 5x_2 - 4x_3 \geq 4,$$
$$x_1, x_2, x_3 \geq 0.$$

Покажите, что неравенства можно свести к множеству равенств, которое потребует введения только одной искусственной переменной (вместо возможных двух искусственных).

3.5. Особые случаи применения симплекс-метода

В этом разделе рассмотрим четыре особых случая, встречающихся при использовании симплекс-метода.

1. Вырожденность.
2. Альтернативные оптимальные решения.
3. Неограниченные решения.
4. Отсутствие допустимых решений.

При изучении этих случаев основное внимание мы уделим (1) *теоретическому* обоснованию причин, приводящих к таким ситуациям, и (2) их *практическим* интерпретациям применительно к реальным задачам.

3.5.1. Вырожденность

В ходе выполнения симплекс-метода проверка условия допустимости может привести к неоднозначному выбору исключаемой переменной. В этом случае на следующей итерации одна или более *базисных* переменных примут нулевое значение. Тогда новое решение будет **вырожденным**.

В вырожденном решении нет никакой опасности, за исключением небольших теоретических неудобств, которые мы далее кратко обсудим. С практической точки зрения вырожденность объясняется тем, что в исходной задаче присутствует по крайней мере одно *избыточное* ограничение. Для того чтобы лучше понять практические и теоретические аспекты явления вырожденности, рассмотрим численный пример. Графическая интерпретация задачи поможет наглядно разобраться в этом явлении.

Пример 3.5–1. (Вырожденное оптимальное решение)

Рассмотрим следующую задачу ЛП.

$$\text{Максимизировать } z = 3x_1 + 9x_2$$

при выполнении условий

$$x_1 + 4x_2 \leq 8,$$
$$x_1 + 2x_2 \leq 4,$$
$$x_1, x_2 \geq 0.$$

Обозначим через x_3 и x_4 дополнительные переменные. Результаты применения симплекс-метода представлены в следующей таблице.

Итерация	Базис	x_1	x_2	x_3	x_4	Решение
Начальная	z	–3	–9	0	0	0
Вводится x_2	x_3	1	4	1	0	8
Исключается x_3	x_4	1	2	0	1	4
Первая	z	–3/4	0	9/4	0	18
Вводится x_1	x_2	1/4	1	1/4	0	2
Исключается x_4	x_4	1/2	0	–1/2	1	0
Вторая	z	0	0	3/2	3/2	18
Оптимум	x_2	0	1	1/2	–1/2	2
	x_1	1	0	–1	2	0

На начальной итерации в качестве исключаемой можно выбрать как переменную x_3, так и x_4. Если оставить в базисе переменную x_4, на следующей итерации она примет значение 0 (как показано в таблице), т.е. получим вырожденное базисное решение. Оптимальное решение получается на следующей итерации.

Что же практически приводит к вырожденности решения? Рассмотрим рис. 3.4, графически представляющий решение этой задачи. Точка оптимума $x_1 = 0$, $x_2 = 2$ является пересечением трех прямых. Поскольку данная задача двухмерна, эта точка переопределена (на плоскости для определения точки достаточно двух прямых), и, следовательно, одно из ограничений избыточно. На практике информация о том, что некоторые ресурсы недефицитны, может быть полезной при интерпретации результатов решения задачи. Эти сведения также могут помочь выявить неточности и ошибки в постановке исходной задачи. К сожалению, не существует способов определить избыточное ограничение непосредственно из данных симплекс-таблиц.

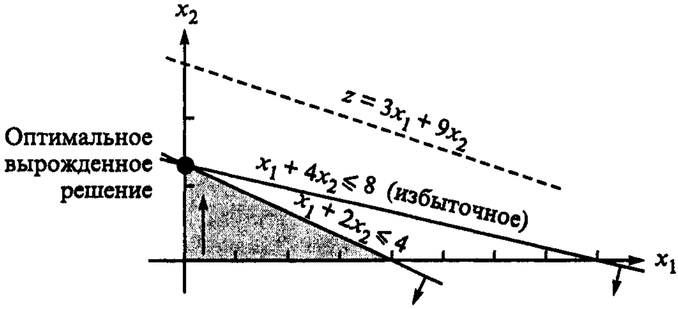


Рис. 3.4

С вычислительной и теоретической точек зрения вырожденность может привести к двум последствиям. Во-первых, в процессе вычислений может возникнуть **зацикливание**. Если в приведенной выше таблице сравнить первую и вторую итерации, то можно заметить, что значение целевой функции не изменилось ($z = 18$). Поэтому может возникнуть ситуация, когда при реализации симплекс-метода некоторая последовательность будет повторяться, не изменяя значения целевой функции и не приводя к завершению вычислительного процесса. Существуют методы, предотвращающие зацикливание, однако они значительно замедляют процесс вычислений. Поэтому в большинстве программ, реализующих симплекс-метод, отсутствуют специальные средства защиты от зацикливания, тем более, что вероятность зацикливания очень мала.

Во-вторых, последствие вырожденности решения можно обнаружить, сравнивая первую и вторую итерации в приведенной выше таблице. Хотя в этих итерациях состав базисных и небазисных переменных различен, значения всех переменных и значение целевой функции не изменяются:

$$x_1 = 0, x_2 = 2, x_3 = 0, x_4 = 0, z = 18.$$

Можно ли, несмотря на то что оптимальное решение не достигнуто, остановить вычисления на первой итерации (когда впервые обнаруживается вырожденность)? Ответ отрицательный, так как решение может быть только *временно* вырожденным (упр. 3.5,а(2)).

Упражнения 3.5,а

1. Рассмотрите графическое представление области допустимых решений (рис. 3.5). Предположим, что выполнение симплекс-метода начинается из точки A, оптимальное решение соответствует точке D, а целевая функция такова, что в точке A вводимой переменной будет x_1.
   a) Покажите (на рисунке) крайние точки, которые соответствуют последовательности симплекс-итераций, приводящих к точке оптимума.
   b) Определите максимально возможное число симплекс-итераций, необходимых для достижения оптимального решения.

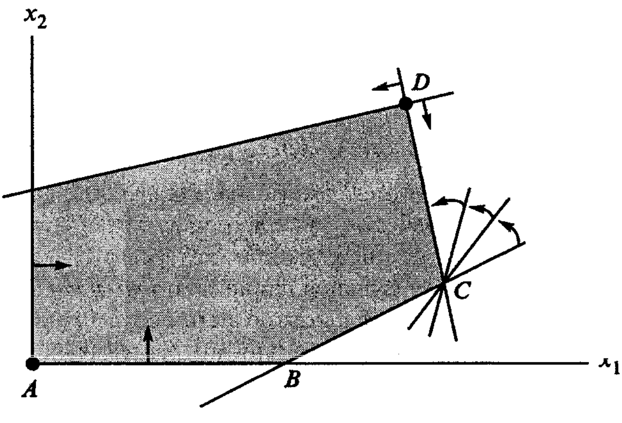


Рис. 3.5

2. Покажите (графически и путем выполнения симплекс-метода), что следующая задача имеет *временно* вырожденные решения.

$$\text{Максимизировать } z = 3x_1 + 2x_2$$

при выполнении условий

$$4x_1 + 3x_2 \leq 12,$$

$$4x_1 + x_2 \le 8,$$
$$4x_1 - x_2 \le 8,$$
$$x_1, x_2 \ge 0.$$

3. Используя в программе TORA опции, позволяющие управлять процессом вычислений, выполните последовательность симплекс-итераций в следующей задаче ЛП (задача придумана E.M. Beale). Начальное допустимое базисное решение, состоящее из дополнительных переменных, повторится на седьмой итерации. Этот пример иллюстрирует явление зацикливания при выполнении симплекс-метода; в этом случае оптимальное решение никогда не будет достигнуто.

$$\text{Максимизировать } z = \frac{3}{4}x_1 - 20x_2 + \frac{1}{2}x_3 - 6x_4$$

при выполнении условий

$$\frac{1}{4}x_1 - 8x_2 - x_3 + 9x_4 \le 0,$$
$$\frac{1}{2}x_1 - 12x_2 - \frac{1}{2}x_3 + 3x_4 \le 0,$$
$$x_3 \le 1,$$
$$x_1, x_2, x_3, x_4 \ge 0.$$

Интересно отметить, что если сделать все коэффициенты в этой задаче целыми (путем умножения на соответствующие множители), то алгоритм симплекс-метода достигнет оптимального решения за конечное число итераций (проверьте это!).

(*Предостережение*. Не используйте программу TORA в автоматическом режиме, так как в этом случае циклические вычисления будут выполняться бесконечно.)

3.5.2. Альтернативные оптимальные решения

Когда прямая (если рассматривается двухмерная задача ЛП, в общем случае — гиперплоскость), представляющая целевую функцию, параллельна прямой (гиперплоскости), соответствующей *связывающему* неравенству (которое в точке оптимума выполняется как точное равенство), целевая функция принимает *одно и то же оптимальное значение* на некотором множестве точек границы области допустимых решений. Эти решения называются **альтернативными оптимальными решениями**. Следующий пример показывает, что таких решений (если они существуют) бесконечное множество. Этот пример также проиллюстрирует практическую значимость альтернативных практических решений.

Пример 3.5–2. (Бесконечное множество решений)

Рассмотрим следующую задачу ЛП.

$$\text{Максимизировать } z = 2x_1 + 4x_2$$

при ограничениях

$$x_1 + 2x_2 \le 5,$$
$$x_1 + x_2 \le 4,$$
$$x_1, x_2 \ge 0.$$

На рис. 3.6 показано множество альтернативных оптимальных решений, которые являются следствием того, что прямая, представляющая целевую функцию, параллельна прямой, соответствующей связывающему ограничению. Каждая точка *отрезка BC* соответствует оптимальному решению со значением целевой функции $z = 10$.

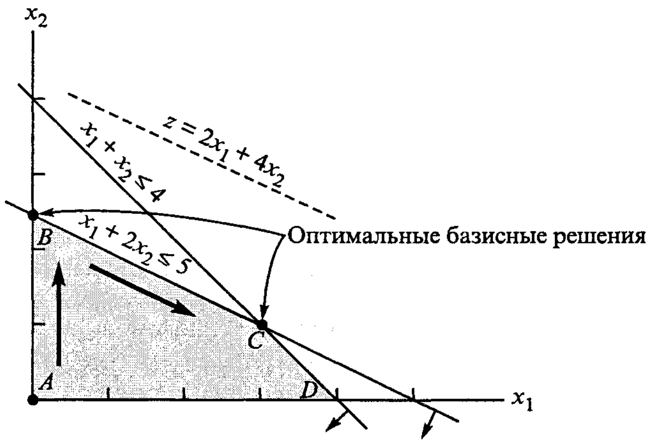


Рис. 3.6

Последовательные итерации выполнения симплекс-метода представлены в следующей таблице.

Итерация	Базис	x_1	x_2	x_3	x_4	Решение
Начальная	z	–2	–4	0	0	0
Вводится x_2	x_3	1	2	1	0	5
Исключается x_3	x_4	1	1	0	1	4
Первая (оптимум)	z	0	0	2	0	10
Вводится x_1	x_2	1/2	1	1/2	0	5/2
Исключается x_4	x_4	1/2	0	–1/2	1	3/2
Вторая	z	0	0	2	0	10
(альтернативный	x_2	0	1	1	–1	1
оптимум)	x_1	1	0	–1	2	3

На первой итерации получаем оптимальное решение $x_1 = 0$, $x_2 = 5/2$ и $z = 10$, которое соответствует точке B на рис. 3.6. Как узнать из симплекс-таблицы, что существует альтернативное оптимальное решение? Посмотрите на коэффициенты *небазисных* переменных в z-строке первой итерации. Коэффициент небазисной переменной x_1 равен нулю, это означает, что данную переменную можно ввести в базис без изменения значения целевой функции, но значение самой переменной x_1 изменится. Введение переменной x_1 в базисное решение выполнено на второй итерации, при этом из базиса исключена переменная x_4. Получено новое решение $x_1 = 3$, $x_2 = 1$, $z = 10$, которое соответствует точке C на рис. 3.6.

Симплекс-метод может определить только две угловые точки B и C. Математически мы можем найти все точки (x_1', x_2') отрезка BC как взвешенное среднее (с неотрицательными весами) точек B и C. Полагая $0 \le \alpha \le 1$ и

$$B\colon x_1 = 0, x_2 = 5/2,$$
$$C\colon x_1 = 3, x_2 = 1,$$

координаты любой точки отрезка BC можно записать следующим образом:

$$x_1' = \alpha \times 0 + (1-\alpha) \times 3 = 3 - 3\alpha,$$

$$x_2' = \alpha \times \frac{5}{2} + (1-\alpha) \times 1 = 1 + \frac{3}{2}\alpha.$$

При $\alpha = 0$ $(x_1', x_2') = (3, 1)$, что соответствует точке C. При $\alpha = 1$ получаем $(x_1', x_2') = (0, 5/2)$ — это точка B. Если значение α лежит строго между 0 и 1, получаем внутренние точки отрезка BC.

На практике альтернативные оптимальные решения весьма полезны, поскольку позволяют сделать выбор среди множества решений без ухудшения значения целевой функции. Например, в рассмотренной выше задаче переменная x_2 принимает нулевое значение в точке B, тогда как в других альтернативных оптимальных решениях ее значение положительно. Если интерпретировать задачу как задачу организации производства двух видов товара (которые соответствуют переменным x_1 и x_2), то, с учетом конкуренции на рынке, более рационально производить оба вида товара, а не один. В этом случае решение, соответствующее точке C, предпочтительнее.

Упражнения 3.5,b

1. Для следующей задачи ЛП найдите не менее трех альтернативных оптимальных базисных решений. Запишите также общее выражение для всех небазисных альтернативных оптимальных решений, частными случаями которых будут найденные вами решения.

   $$\text{Максимизировать } z = x_1 + 2x_2 + 3x_3$$

   при выполнении условий

   $$x_1 + 2x_2 + 3x_3 \le 10,$$
   $$x_1 + x_2 \le 5,$$
   $$x_1 \le 1,$$
   $$x_1, x_2, x_3 \ge 0.$$

2. Покажите, что следующая задача ЛП имеет альтернативные оптимальные решения и все они небазисные. Для этой задачи представьте на плоскости область допустимых решений и прямую, соответствующую целевой функции.

$$\text{Максимизировать } z = 2x_1 - x_2 + 3x_3$$

при выполнении условий

$$x_1 - x_2 + 5x_3 \le 10,$$
$$2x_1 - x_2 + 3x_3 \le 40,$$
$$x_1, x_2, x_3 \ge 0.$$

3. Покажите, что в следующей задаче ЛП оптимальное решение вырождено и все существующие альтернативные решения являются небазисными.

$$\text{Максимизировать } z = 3x_1 + x_2$$

при выполнении условий

$$x_1 + 2x_2 \le 5,$$
$$x_1 + x_2 - x_3 \le 2,$$
$$7x_1 + 3x_2 - 5x_3 \le 20,$$
$$x_1, x_2, x_3 \ge 0.$$

3.5.3. Неограниченные решения

В некоторых задачах ЛП значения переменных могут неограниченно возрастать без нарушения ограничений. Это говорит о том, что пространство допустимых решений **не ограничено** по крайней мере по одному направлению. В результате этого целевая функция может возрастать (задача максимизации) или убывать (задача минимизации) неограниченно.

Неограниченность решения задачи свидетельствует только об одном: модель разработана не достаточно корректно. Типичные ошибки, приводящие к построению таких моделей, заключаются в том, что не учитываются ограничения, не являющиеся избыточными, и не точно оцениваются параметры (коэффициенты) ограничений.

В следующем примере показано, как на основе данных, приведенных в симплекс-таблице, можно определить, когда не ограничено пространство решений и значения целевой функции.

Пример 3.5–3. (Неограниченная целевая функция)

Рассмотрим задачу

$$\text{Максимизировать } z = 2x_1 + x_2$$

при выполнении условий

$$x_1 - x_2 \le 10,$$
$$2x_1 \le 40,$$
$$x_1, x_2 \ge 0.$$

Симплекс-таблица начальной итерации этой задачи имеет следующий вид.

Базис	x_1	x_2	x_3	x_4	Решение
z	–2	–1	0	0	0
x_3	1	–1	1	0	10
x_4	2	0	0	1	40

Из этой таблицы видно, что в качестве вводимой переменной можно взять как x_1, так и x_2. Поскольку переменная x_1 имеет максимальный (по абсолютной величине) отрицательный коэффициент в z-строке, именно ее следует ввести в базисное решение. Однако заметим, что *во всех ограничениях* коэффициенты, стоящие перед переменной x_2, *отрицательны* или *равны нулю*. Это означает, что значение переменной x_2 может возрастать до бесконечности, и при этом не нарушается ни одно ограничение. Поскольку увеличение на 1 значения переменной x_2 приводит к увеличению на 1 значения целевой функции, значит, неограниченное увеличение значения переменной x_2 ведет к неограниченному увеличению значения целевой функции. Эта ситуация проиллюстрирована на рис. 3.7. На этом рисунке видно, что пространство допустимых решений не ограничено в направлении оси x_2 и значение целевой функции может быть каким угодно большим.

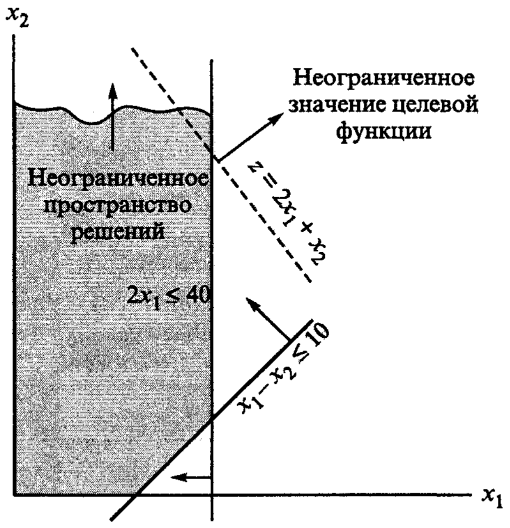


Рис. 3.7

Правило выявления неограниченности решения следующее. Если на какой-либо симплекс-итерации коэффициенты в ограничениях для какой-нибудь *небазисной* переменной будут неположительными, значит, *пространство решений* не ограничено в направлении возрастания этой переменной. Кроме того, если коэффициент этой переменной в z-строке отрицателен, когда рассматривается задача максимизации, или положителен в задаче минимизации, *целевая функция* также не ограничена.

Упражнения 3.5,с

1. В задаче из примера 3.5–3 покажите, что применение симплекс-метода, когда, согласно условию оптимальности, вычисления начинаются с переменной x_1 в качестве вводимой, обязательно приведет к неограниченному решению.
2. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 20x_1 + 10x_2 + x_3$$

при выполнении условий

$$3x_1 - 3x_2 + 5x_3 \le 50,$$
$$x_1 + x_3 \le 10,$$
$$x_1 - x_2 + 4x_3 \le 20,$$
$$x_1, x_2, x_3 \ge 0.$$

   a) Рассмотрев ограничения, определите направление (по оси x_1, x_2 или x_3), в котором пространство допустимых решений не ограничено.
   b) Без дополнительных вычислений сделайте заключение относительно оптимального значения целевой функции.

3. В некоторых плохо построенных моделях ЛП пространство допустимых решений может быть неограниченным даже тогда, когда задача имеет конечное (ограниченное) оптимальное решение. Такое может случиться, если при построении ограничений допущены ошибки. В больших задачах иногда трудно визуально определить, является ли пространство допустимых решений ограниченным. Разработайте процедуру определения неограниченности пространства решений и примените ее к следующей задаче.

$$\text{Максимизировать } z = 40x_1 + 20x_2 + 2x_3$$

при ограничениях

$$3x_1 - 3x_2 + 5x_3 \le 50,$$
$$x_1 + x_3 \le 10,$$
$$x_1 - x_2 + 4x_3 \le 2,$$
$$x_1, x_2, x_3 \ge 0.$$

3.5.4. Отсутствие допустимых решений

Если ограничения задачи ЛП несовместны (т.е. они не могут выполняться одновременно), то задача не имеет допустимых решений. Такая ситуация не может возникнуть, если *все* неравенства, составляющие систему ограничений, имеют тип "≤" с неотрицательными правыми частями, так как в этом случае дополнительные переменные могут составить допустимое решение. Для других типов ограничений мы используем искусственные переменные. И хотя в оптимальном решении все искусственные переменные в силу штрафов равны нулю, такой исход возможен только тогда, когда задача имеет непустое пространство допустимых решений. В противном случае в оптимальном решении будет присутствовать хотя бы одна *положительная* искусственная переменная.

С практической точки зрения отсутствие допустимых решений свидетельствует о том, что задача плохо сформулирована.

Пример 3.5–4. (Отсутствие допустимых решений)

Рассмотрим следующую задачу.

$$\text{Максимизировать } z = 3x_1 + 2x_2$$

при выполнении условий

$$2x_1 + x_2 \le 2,$$
$$3x_1 + 4x_3 \ge 12,$$
$$x_1, x_2 \ge 0.$$

Результат применения симплекс-метода представлен в следующей таблице.

Итерация	Базис	x_1	x_2	x_4	x_3	R	Решение
Начальная	z	$-3-3M$	$-2-4M$	M	0	0	$-12M$
Вводится x_2	x_3	2	1	0	1	0	2
Исключается x_3	R	3	4	-1	0	1	12
Первая	z	$1+5M$	0	M	$2+4M$	0	$4-4M$
(псевдооптимум)	x_2	2	1	0	1	0	2
	R	-5	0	-1	-4	1	4

Данные из этой таблицы показывают, что в точке оптимума искусственная переменная R имеет положительное значение (= 4), что свидетельствует об отсутствии допустимого решения. На рис. 3.8 графически представлена ситуация данной задачи. Алгоритм симплекс-метода, допуская положительные значения искусственной переменной, по существу, превращает неравенство $3x_1 + 4x_3 \ge 12$ в $3x_1 + 4x_3 \le 12$. (Объясните, почему так происходит.) В результате получаем то, что можно назвать **псевдооптимальным решением**.

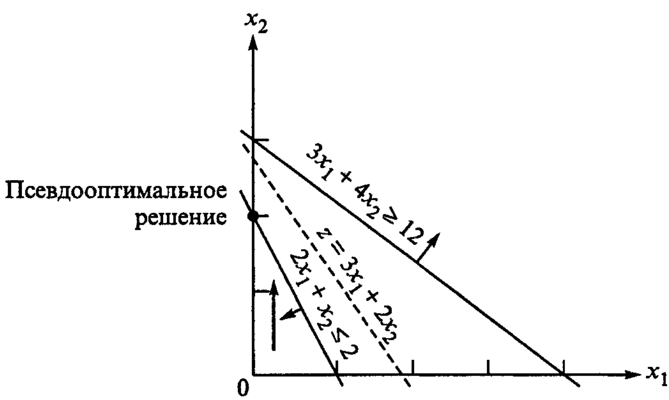


Рис. 3.8

Упражнения 3.5,d

1. Компания производит изделия трех типов T1, T2 и T3. На их изготовление расходуется материал M1 и M2 согласно данным следующей таблицы.

Материал	Расход материалов на единицу изделия		
	T1	T2	T3
M1	3	5	6
M2	5	3	4

   Ежедневно можно использовать не более 1000 единиц материала M1 и 1200 единиц материала M2. Отдел маркетинга настаивает, чтобы суммарное ежедневное производство изделий составляло не менее 500 единиц. Может ли производство удовлетворить это требование? Если ответ отрицательный, помогите компании составить наиболее выгодную структуру производства изделий.

2. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 3x_1 + 2x_2 + 3x_3$$

   при ограничениях

$$2x_1 + x_2 + x_3 \le 2,$$
$$3x_1 + 4x_2 + 2x_3 \ge 8,$$
$$x_1, x_2, x_3 \ge 0.$$

   Используя M-метод, покажите, что оптимальное решение этой задачи содержит искусственную базисную переменную. Но поскольку значение этой переменной равно нулю, задача имеет *допустимое* оптимальное решение.

3.6. Заключение

Из теоретических положений, лежащих в основе симплекс-метода, следует, что оптимальное решение задачи линейного программирования соответствует крайней точке пространства допустимых решений задачи. В свою очередь, крайние точки пространства допустимых решений полностью определяются базисными решениями задачи ЛП, представленной в стандартной форме. Для компьютерной реализации симплекс-метода разработан способ использования искусственных переменных, что позволяет найти начальное базисное решение задачи. В этой главе также рассмотрены теоретические и практические аспекты особых случаев реализации симплекс-метода: вырожденность, альтернативные оптимальные решения, неограниченность и отсутствие допустимых решений.

Литература

1. Bazaraa M., Jarvis J., Sherali M. *Linear Programming and Network Flows*, 2nd ed., Wiley, New York, 1990.
2. Dantzig G. *Linear Programming and Extensions*, Princeton University Press, Princeton, N.J., 1963. (Русский перевод: Данциг Дж. Линейное программирование, его применение и обобщение. — М.: Прогресс, 1966.)

3. Nering E., Tucker A. *Linear Programming and Related Problems*, Academic Press, Boston, 1992.
4. Taha H. *Linear Programming*, Chapter II–1 in Handbook of Operations Research, J. Moder and S. Elmaghraby (eds.), Van Nostrand Reinhold, New York, 1978. (Русский перевод: *Исследование операций*: в 2-х томах. Под ред. Дж. Моудера, С. Элмаграби. — М.: Мир, 1981.)

Литература, добавленная при переводе

Ашманов С.А. *Линейное программирование*. — М.: Наука, 1981.
Гасс С. *Линейное программирование*. — М.: Физматгиз, 1961.
Гольштейн Е.Г., Юдин Д.Б. *Линейное программирование: Теория, методы и приложения*. — М.: Наука, 1969.

Комплексные задачи

■ **3-1.** Небольшой консервный завод производит пять типов консервов на основе трех видов свежих фруктов. Завод имеет два производственных цеха, которые при необходимости могут наращивать выпуск продукции. В настоящее время производственные цеха работают в одну смену, но без труда могут перейти на двух- или трехсменную работу. Реальным ограничением наращивания выпуска продукции является только объем поставляемых на переработку свежих фруктов. Поскольку объемы холодильных камер ограничены, желательно, чтобы все фрукты перерабатывались в течение рабочего дня.

Молодой аналитик, специалист по исследованию операций, горит желанием увеличить уровень производства этого завода. После анализа производственной ситуации он сформулировал задачу линейного программирования. Задача включает пять переменных (для пяти видов продукции) и три ограничения (для трех видов сырья). Теория ЛП говорит о том, что в задаче с пятью переменными и тремя ограничениями оптимальное решение не может содержать более трех базисных переменных. Следовательно, оптимальное решение предполагает производство не более трех видов продукции. "Ага, — подумал аналитик, — производство этой компании явно не оптимально."

Он добился встречи с менеджером по производству для обсуждения построенной модели ЛП. Менеджер, подробно ознакомившись с задачей, нашел, что модель адекватно отображает сложившуюся производственную ситуацию.

Аналитик объяснил, что, согласно теории ЛП, поскольку задача имеет только три ограничения, оптимальный план производства не должен содержать более трех видов продукции и поэтому завод должен прекратить выпуск двух бесприбыльных видов продукции. Менеджер, внимательно выслушав аналитика, сказал, что его компания не может прекратить выпуск двух видов продукции, так как это не выгодно предприятию. Но чтобы согласовать теорию и практику, он предложил добавить в модель еще два ограничения, тогда будет пять ограничений, и в этом случае оптимальное решение должно разрешить производство всех пяти видов продукции.

Это предложение немного смутило аналитика, так как очевидно, что добавление новых ограничений не может улучшить оптимальное решение, однако он смог честно ответить менеджеру, что данное предложение согласуется с теорией ЛП.

Как разрешить этот "парадокс"?

■ **3-2.**[1] Транспортная компания, специализирующаяся на перевозках грузов, имеет множество терминалов, расположенных в "стратегических" точках по всей территории США. Когда грузы поступают на терминал, они сортируются: часть груза, предназначенная местному потребителю, ему же и поступает, остальной груз отправляется к следующему терминалу. Терминальные доки обслуживают как *постоянные*, так *временные работники*, набираемые по найму. Постоянным работникам гарантирована 40-часовая рабочая неделя. Они работают в одну из трех стандартных смен непрерывно в течение пяти дней, но их рабочая неделя может начаться в любой день недели. Временные работники нанимаются на любое количество рабочих часов при пиковых поступлениях грузов, превышающих возможности их обработки постоянными работниками.

Грузы могут поступать на терминалы в любое время суток, причем неравномерно в течение суток. Изучение статистических данных показывает, что распределение поступления грузов примерно одинаково каждую неделю, и пик поступления грузов обычно приходится на конец недели (пятница—воскресение). Политика компании требует, чтобы грузы задерживались на терминалах не более чем на 16 часов.

Разработайте модель, с помощью которой можно было бы назначать в рабочие смены постоянных работников и нанимать временных.

---

[1] Эта задача основана на исследованиях, проводимых автором для национальной транспортной компании.

Двойственность и анализ чувствительности

4.1. Введение

Оптимальное решение задачи линейного программирования определяется теми условиями, которые нашли отражение в модели в момент ее формирования. В реальной жизни условия, формирующие модель, не остаются неизменными. В связи с этим особое значение приобретают средства, позволяющие оценить изменения в оптимальном решении, вызванные изменениями в параметрах исходной модели. Таким средством является *анализ чувствительности.* Он предлагает эффективные вычислительные методы, позволяющие изучить динамическое поведение оптимального решения.

Мы уже встречались с анализом чувствительности (на элементарном уровне) в разделе 2.4. В этой главе мы подробнее рассмотрим методы анализа чувствительности, основанные на *теории двойственности.*

Анализ чувствительности предполагает, что изменение параметров модели происходит *линейно.* Обобщая, можно предположить, что параметры модели изменяются согласно заранее определенным *непрерывным* функциям. Соответствующая этому обобщению совокупность методов, называемая *параметрическим программированием*, приведена в разделе 7.7.

4.2. Определение двойственной задачи

Исходную задачу линейного программирования будем называть **прямой**. **Двойственная задача** — это задача, формулируемая с помощью определенных правил непосредственно из прямой задачи. В этом разделе рассмотрены правила построения двойственных задач.

При изложении теории двойственности часто рассматривают формулировки двойственной задачи в зависимости от различных видов прямой задачи, которые определяются типами ограничений, знаками переменных (неотрицательные или свободные, т.е. без ограничения в знаке) и типом оптимизации (максимизация или минимизация целевой функции). Такая "привязка" двойственной задачи к исходной не всегда оправдана (см. упр. 4.2,a(7)). В этой книге приводится *единая* формулировка двойственной задачи, применимая ко *всем* видам прямой задачи. В основу такой формулировки положена *стандартная форма прямой задачи* (см. раздел 3.2). Напомним, что задача ЛП в стандартной форме записывается следующим образом.

Максимизировать или минимизировать целевую функцию $z = \sum_{j=1}^{n} c_j x_j$

при ограничениях

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

$$x_j \geq 0, j = 1, 2, ..., n.$$

В состав n переменных x_j входят также дополнительные переменные. Стандартная форма задачи ЛП предполагает выполнение следующих условий.

1. Все ограничения записаны в виде равенств (с неотрицательной правой частью).
2. Все переменные неотрицательны.
3. Оптимизация определяется как максимизация или минимизация целевой функции.

Стандартная форма задачи ЛП порождает стандартную таблицу симплекс-метода. Поэтому, как будет показано в разделе 4.3, решение двойственной задачи можно получить непосредственно из симплекс-таблицы, соответствующей оптимальному решению прямой задачи. Таким образом, определив двойственную задачу на основе стандартной формы прямой задачи, после вычислений симплекс-метода мы автоматически получаем решение двойственной задачи.

Переменные и ограничения двойственной задачи формируются путем симметричных структурных преобразований прямой задачи по следующим правилам.

1. Каждому из m ограничений прямой задачи соответствует переменная двойственной задачи.
2. Каждой из n переменных прямой задачи соответствует ограничение двойственной задачи.
3. Коэффициенты при какой-либо переменной в ограничениях прямой задачи становятся коэффициентами ограничения двойственной задачи, соответствующей этой переменной, а правая часть формируемого ограничения равна коэффициенту при этой переменной в выражении целевой функции.
4. Коэффициенты целевой функции двойственной задачи равны правым частям ограничений прямой задачи.

Графически эти правила представлены в табл. 4.1.

ТАБЛИЦА 4.1

Переменные двойственной задачи	Переменные прямой задачи						
	x_1	x_2	...	x_j	...	x_n	
	c_1	c_2	...	c_j	...	c_n	
y_1	a_{11}	a_{12}	...	a_{1j}	...	a_{1n}	b_1
y_2	a_{21}	a_{22}	...	a_{2j}	...	a_{2n}	b_2
.	.	.		.		.	.
.	.	.	...	.	...	.	.
.	.	.		.		.	.
y_n	a_{m1}	a_{m2}	...	a_{mj}	...	a_{mn}	b_m
				j-е ограничение двойственной задачи			Коэффициенты целевой функции двойственной задачи

Правила, определяющие тип оптимизации и ограничений, а также знак переменных двойственной задачи, приведены в табл. 4.2.

Таблица 4.2

Целевая функция прямой задачи	Двойственная задача		
	Целевая функция	Тип ограничений	Переменные
Максимизация	Минимизация	$\geq$	Свободные
Минимизация	Максимизация	$\leq$	Свободные

Следующие примеры иллюстрируют правила построения двойственной задачи.

Пример 4.2–1

Прямая задача	Прямая задача в стандартной форме	Двойственные переменные
Максимизировать $z = 5x_1 + 12x_2 + 4x_3$ при ограничениях $x_1 + 2x_2 + x_3 \leq 10,$ $2x_1 - x_2 + 3x_3 = 8,$ $x_1, x_2, x_3 \geq 0.$	Максимизировать $z = 5x_1 + 12x_2 + 4x_3 + 0x_4$ при ограничениях $x_1 + 2x_2 + x_3 + x_4 = 10,$ $2x_1 - x_2 + 3x_3 + 0x_4 = 8,$ $x_1, x_2, x_3, x_4 \geq 0.$	 y_1 y_2

Двойственная задача

Минимизировать $w = 10y_1 + 8y_2$

при ограничениях

$$y_1 + 2y_2 \geq 5,$$
$$2y_1 - y_2 \geq 12,$$
$$y_1 + 3y_3 \geq 4,$$
$$\left.\begin{array}{r} y_1 + 0y_2 \geq 0, \\ y_1, y_2 \ \text{— свободные} \end{array}\right\} \Rightarrow (y_1 \geq 0,\ y_2 \ \text{— свободная переменная}).$$

Пример 4.2–2

Прямая задача	Прямая задача в стандартной форме	Двойственные переменные
Минимизировать $z = 15x_1 + 12x_2$ при ограничениях $x_1 + 2x_2 \geq 3,$ $2x_1 - 4x_2 \leq 5,$ $x_1, x_2 \geq 0.$	Минимизировать $z = 15x_1 + 12x_2 + 0x_3 + 0x_4$ при ограничениях $x_1 + 2x_2 - x_3 = 3,$ $2x_1 - 4x_2 + x_4 = 5,$ $x_1, x_2, x_3, x_4 \geq 0.$	 y_1 y_2

Двойственная задача

Максимизировать $w = 3y_1 + 5y_2$

при ограничениях

$$y_1 + 2y_2 \leq 15,$$

$$2y_1 - 4y_2 \leq 12,$$
$$-y_1 \leq 0 \text{ (или } y_1 \geq 0),$$
$$y_2 \leq 0,$$
$$y_1, y_2 \text{ — свободные переменные (избыточное условие).}$$

Пример 4.2–3

Прямая задача	Прямая задача в стандартной форме	Двойственные переменные
Максимизировать $z = 5x_1 + 6x_2$ при ограничениях	Подстановка $x_1 = x_1^+ - x_1^-$ приводит к задаче: Максимизировать $z = 5x_1^+ - 5x_1^- + 6x_2$ при ограничениях	
$x_1 + 2x_2 = 5,$	$x_1^+ - x_1^- + 2x_2 = 5,$	y_1
$-x_1 + 5x_2 \geq 3,$	$-x_1^+ + x_1^- + 5x_2 - x_3 = 3,$	y_2
$4x_1 + 7x_2 \leq 8,$	$4x_1^+ - 4x_1^- + 7x_2 + x_4 = 8,$	y_3
x_1 — свободная, $x_2 \geq 0.$	$x_1^+, x_1^-, x_2 \geq 0.$	

Двойственная задача

$$\text{Минимизировать } w = 5y_1 + 3y_2 + 8y_3$$

при ограничениях

$$\left.\begin{array}{r} y_1 - y_2 + 4y_3 \geq 5 \\ -y_1 + y_2 - 4y_3 \geq -5 \end{array}\right\} \Rightarrow y_1 - y_2 + 4y_3 = 5,$$
$$2y_1 + 5y_2 + 7y_3 \geq 6,$$
$$-y_2 \geq 0 \Rightarrow y_2 \leq 0,$$
$$y_3 \geq 0,$$
$$y_1 \text{ — свободная переменная,}$$
$$y_2, y_3 \text{ — свободные переменные (избыточное условие).}$$

Первое и второе ограничения двойственной задачи заменены одним ограничением в виде равенства. Здесь действует следующее правило: свободной переменной прямой задачи соответствует ограничение в виде равенства двойственной задачи и, наоборот, ограничению в виде равенства прямой задачи соответствует свободная переменная двойственной задачи.

Упражнения 4.2,а

1. Для задачи из примера 4.2–1 запишите двойственную задачу в предположении, что находится не максимум целевой функции, а ее минимум.
2. Вернитесь к задаче из примера 4.2–1. Применение симплекс-метода к прямой задаче ЛП, записанной в стандартной форме, требует введения искусственной переменной во второе ограничение для получения начального базисного решения. По-

кажите, что введение искусственной переменной не влияет на построение двойственной задачи, поскольку искусственная переменная прямой задачи порождает избыточное ограничение двойственной задачи.

3. В примере 4.2–2 сформулируйте двойственную задачу в предположении, что в прямой задаче этого примера добавлено третье ограничение $3x_1 + x_2 = 4$.
4. На основе задачи из примера 4.2–3 покажите, что даже если изменить тип оптимизации (с максимизации на минимизацию целевой функции), то все равно свободным переменным прямой задачи будут соответствовать в двойственной задаче ограничения в виде равенств.
5. Запишите двойственные задачи для следующих прямых задач ЛП.

a) Максимизировать $z = -5x_1 + 2x_2$
при ограничениях

$$-x_1 + x_2 \le -2,$$
$$2x_1 + 3x_2 \le 5,$$
$$x_1, x_2 \ge 0.$$

b) Минимизировать $z = 6x_1 + 3x_2$
при ограничениях

$$6x_1 - 3x_2 + x_3 \ge 2,$$
$$3x_1 + 4x_2 + x_3 \ge 5,$$
$$x_1, x_2, x_3 \ge 0.$$

c) Максимизировать $z = 5x_1 + 6x_2$
при ограничениях

$$x_1 + 2x_2 = 5,$$
$$-x_1 + 5x_2 \ge 3,$$
x_1 — свободная переменная,
$$x_2 \le 0.$$

d) Минимизировать $z = 3x_1 + 4x_2 + 6x_3$
при ограничениях

$$x_1 + x_2 \ge 10,$$
$$x_1 \ge 0,$$
$$x_2 \le 0,$$
$$x_3 \ge 0.$$

e) Максимизировать $z = x_1 + x_2$
при ограничениях

$$2x_1 + x_2 = 5,$$
$$3x_1 - x_2 = 6,$$
x_1, x_2 — свободные переменные.

6. Истинны или ложны следующие утверждения?

a) Задача, двойственная к двойственной, совпадает с исходной прямой задачей.
b) Если в прямой задаче есть ограничение в виде равенства, то соответствующая переменная двойственной задачи обязательно будет свободной.

c) Если в прямой задаче ограничение имеет вид неравенства типа "≤", тогда соответствующая переменная в двойственной задаче будет неотрицательной (неположительной), в зависимости от того, следует ли в прямой задаче максимизировать или минимизировать целевую функцию.

d) Если в прямой задаче ограничение является неравенством типа "≥", тогда соответствующая переменная в двойственной задаче будет неотрицательной (неположительной), в зависимости от того, следует ли в прямой задаче минимизировать или максимизировать целевую функцию.

e) Свободная переменная прямой задачи порождает в двойственной задаче ограничение в виде равенства.

7. В следующей таблице приведены правила построения двойственных задач, которые часто приводятся в литературе по исследованию операций и линейному программированию. Покажите, что эти правила являются частными случаями общих правил, приведенных в табл. 4.2.

Задача максимизации		Задача минимизации
Ограничения		Переменные
≥	⇔	≤ 0
≤	⇔	≥ 0
=	⇔	Свободная
Переменные		Ограничения
≥ 0	⇔	≥
≤ 0	⇔	≤
Свободная	⇔	=

4.3. Соотношения между оптимальными решениями прямой и двойственной задач

Прямая и двойственная задачи так тесно взаимосвязаны, что оптимальное решение одной задачи можно получить непосредственно (без дополнительных вычислений) из симплекс-таблицы, представляющей оптимальное решение другой. Это утверждение основывается на следующем соотношении.

Соотношение 1. *Для любой симплекс-итерации прямой или двойственной задачи.*

$$\begin{pmatrix}\text{Коэффициент при } j\text{-й} \\ \text{переменной в } z\text{-строке} \\ \text{одной задачи}\end{pmatrix} = \begin{pmatrix}\text{Разность между левой и правой} \\ \text{частями } j\text{-го неравенства} \\ \text{другой задачи}\end{pmatrix}$$

Это соотношение симметрично относительно прямой и двойственной задач. Его можно использовать для определения оптимального решения одной задачи непосредственно из симплекс-таблицы, содержащей оптимальное решение другой. Данное обстоятельство обуславливает возможность проведения вычислений именно по той задаче (прямой или двойственной), которая требует меньших вычислительных ресурсов. Например, если прямая задача имеет 100 переменных и 500 ограничений, то предпочтительнее нахождение оптимального решения двойственной задачи, так как она будет содержать только 100 ограничений.[1]

Пример 4.3–1

Рассмотрим прямую и двойственную задачи из примера 4.2–1, которые для удобства сведены в следующую таблицу.

Прямая задача	Двойственная задача
Максимизировать $z = 5x_1 + 12x_2 + 4x_3$ при ограничениях $x_1 + 2x_2 + x_3 \le 10,$ $2x_1 - x_2 + 3x_3 = 8,$ $x_1, x_2, x_3 \ge 0.$	Минимизировать $w = 10y_1 + 8y_2$ при ограничениях $y_1 + 2y_2 \ge 5,$ $2y_1 - y_2 \ge 12,$ $y_1 + 3y_2 \ge 4,$ $y_1 \ge 0$, y_2 — свободная переменная.

В следующей таблице представлены симплекс-итерации решения прямой задачи.

Итерация	Базис	x_1	x_2	x_3	x_4	R	Решение
0	z	$-5-2M$	$-12+M$	$-4-3M$	0	0	$-8M$
	x_4	1	2	1	1	0	10
	R	2	-1	3	0	1	8
1	z	$-7/3$	$-40/3$	0	0	$4/3+M$	$32/3$
	x_4	$1/3$	$7/3$	0	1	$-1/3$	$22/3$
	x_3	$2/3$	$-1/3$	1	0	$1/3$	$8/3$
2	z	$-3/7$	0	0	$40/7$	$-4/3+M$	$368/7$
	x_2	$1/7$	1	0	$3/7$	$-1/7$	$22/7$
	x_3	$5/7$	0	1	$1/7$	$2/7$	$26/7$
3	z	0	0	$3/5$	$29/5$	$-2/5+M$	$274/5$
	x_2	0	1	$-1/5$	$2/5$	$-1/5$	$12/5$
	x_1	1	0	$7/5$	$1/5$	$2/5$	$26/5$

Применяя на третьей симплекс-итерации соотношение 1 для переменных x_4 и R из начального базисного решения, получим следующие данные.

Начальные базисные переменные прямой задачи	x_4	R
Коэффициенты в z-строке 3-й итерации	$29/5$	$-2/5+M$
Соответствующие ограничения двойственной задачи	$y_1 \ge 0$	$y_2 \ge -M$
Уравнения, полученные на основе соотношения 1	$y_1 - 0 = 29/5$	$y_2 - (-M) = -2/5 + M$

[1] Трудоемкость вычислений задачи линейного программирования в большей степени зависит от количества ограничений, чем переменных. — *Прим. ред.*

Решениями полученных уравнений будут $y_1 = 29/5$ и $y_2 = -2/5$. Если бы решения двойственной задачи искались независимо от решения прямой задачи, то было бы получено такое же решение. Аналогично, вследствие симметричности соотношения 1, из данного решения двойственной задачи и ее начальной симплекс-таблицы получаем оптимальное решение прямой задачи: $x_1 = 26/5$, $x_2 = 12/5$ и $x_3 = 0$. (С помощью программы TORA решите двойственную задачу и убедитесь, что ее решение действительно совпадает с приведенным выше.)

Применение соотношения 1 к переменным начального решения всегда приводит к легко решаемым уравнениям, так как каждое такое уравнение содержит только одну переменную. Конечно же, ничего не мешает использовать другие переменные (такие, как x_1, x_2 и x_3 в рассматриваемой задаче) при построении уравнений, необходимых для определения значений переменных двойственной задачи. Например, соотношение 1 порождает следующие уравнения, ассоциируемые с переменными x_1 и x_3:

$$y_1 + 2y_2 - 5 = 0,$$
$$y_1 + 3y_2 - 4 = \frac{3}{5}.$$

Решение этой системы уравнений также приводит к оптимальным значениям двойственной задачи $y_1 = 29/5$ и $y_2 = -2/5$. Однако аналогичные уравнения, ассоциируемые с переменными x_4 и R, будут уже не так просты. (Убедитесь самостоятельно, что уравнения, ассоциированные с любыми двумя переменными из множества x_1, x_2, x_3, x_4, R, дают одни и те же значения переменных двойственной задачи.)

Представим еще одно соотношение между прямой и двойственной задачами, которое совместно с соотношением 1 предлагает интересную экономическую интерпретацию задачи линейного программирования.

Соотношение 2. *Для любой пары допустимых решений прямой и двойственной задач*

$$\begin{pmatrix}\text{Значение целевой функции}\\ \text{в задаче } \textit{максимизации}\end{pmatrix} \le \begin{pmatrix}\text{Значение целевой функции}\\ \text{в задаче } \textit{минимизации}\end{pmatrix}.$$

В точке оптимума это соотношение принимает вид строгого равенства.

Пример 4.3–2

В примере 4.3–1 нетрудно показать (путем подстановки в ограничения), что прямая и двойственная задачи имеют допустимые решения $x_1 = 0$, $x_2 = 0$, $x_3 = 8/3$ и $y_1 = 6$, $y_2 = 0$. Для этих решений значения целевых функций соответственно равны $z = 32/3$ и $w = 60$. Для оптимальных решений $x_1 = 26/5$, $x_2 = 12/5$, $x_3 = 0$ и $y_1 = 29/5$, $y_2 = -2/5$ имеем $z = w = 54.8$. Таким образом, приведенные значения целевых функций подтверждают соотношение 2.

Из соотношения 2 следует, что для всех допустимых решений прямой и двойственной задач значения целевой функции задачи минимизации всегда будут верхним пределом значений целевой функции задачи максимизации. Таким образом, итерационное решение задачи максимизации ведет к возрастанию значений целевой функции, а итерационное решение задачи минимизации — к ее убыванию. В итоге, при успешном завершении процессов вычисления прямой и двойственной задач приходим к точке "равновесия", где значения целевых функций задач максимизации и минимизации становятся равными.

Упражнения 4.3,а

1. В примере 4.3–1, используя ограничения двойственной задачи, ассоциированные с переменными x_2 и x_3, найдите оптимальные значения переменных y_1 и y_2.
2. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = 5x_1 + 2x_2 + 3x_3$$

при ограничениях

$$x_1 + 5x_2 + 2x_3 = 30,$$
$$x_1 - 5x_2 - 6x_3 \le 40,$$
$$x_1, x_2, x_3 \ge 0.$$

Оптимальное решение этой задачи удовлетворяет уравнению (получено из z-строки симплекс-таблицы)

$$z + 0x_1 + 23x_2 + 7x_3 + (5 + M)x_4 + 0x_5 = 150,$$

где искусственная переменная x_4 и дополнительная переменная x_5 входили в начальное базисное решение. Запишите соответственную двойственную задачу и найдите ее оптимальное решение, исходя из приведенного уравнения.

3. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = x_1 + 5x_2 + 3x_3$$

при ограничениях

$$x_1 + 2x_2 + x_3 = 3,$$
$$2x_1 - x_2 = 4,$$
$$x_1, x_2, x_3 \ge 0.$$

a) Запишите соответствующую двойственную задачу.

b) Используя информацию, что оптимальное базисное решение этой задачи содержит переменные x_1 и x_3, найдите оптимальное решение двойственной задачи.

4. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = 2x_1 + 4x_2 + 4x_3 - 3x_4$$

при ограничениях

$$x_1 + x_2 + x_3 = 4,$$
$$x_1 + 4x_2 + x_4 = 8,$$
$$x_1, x_2, x_3, x_4 \ge 0.$$

Используя двойственную задачу, убедитесь, что базисное решение (x_1, x_2) не является оптимальным.

5. В предыдущем упражнении уравнение, полученное из z-строки оптимальной симплекс-таблицы, имеет следующий вид

$$z + 2x_1 + 0x_2 + 0x_3 + 3x_4 = 16.$$

Найдите оптимальное решение соответствующей двойственной задачи.

6. С помощью двойственной задачи найдите допустимое решение следующей системы неравенств:

$$2x_1 + 3x_2 \le 12,$$
$$-3x_1 + 2x_2 \le -4,$$
$$3x_1 - 5x_2 \le 2,$$
$$x_1 \text{ — свободная переменная},$$
$$x_2 \ge 0.$$

(*Совет.* Добавьте к этой системе неравенств тривиальную целевую функцию *максимизировать* $z = 0x_1 + 0x_2 + 0x_3$ и решите двойственную задачу.)

7. Найдите оптимальное значение целевой функции следующей задачи, основываясь только на свойствах ее двойственной задачи (т.е. не применяя симплекс-метод к двойственной задаче).

$$\text{Минимизировать } z = 10x_1 + 4x_2 + 5x_3$$

при ограничениях

$$5x_1 - 7x_2 + 3x_3 \ge 50,$$
$$x_1, x_2, x_3 \ge 0.$$

8. Определите интервалы изменения значений целевой функции в следующих задачах ЛП.

a) Минимизировать $z = 5x_1 + 2x_2$
при ограничениях

$$x_1 - x_2 \ge 3,$$
$$2x_1 + 3x_2 \ge 5,$$
$$x_1, x_2 \ge 0.$$

b) Максимизировать $z = x_1 + 5x_2 + 3x_3$
при ограничениях

$$x_1 + 2x_2 + x_3 = 3,$$
$$2x_1 - x_2 = 4,$$
$$x_1, x_2, x_3 \ge 0.$$

c) Максимизировать $z = 2x_1 + x_2$

при ограничениях

$$x_1 - x_2 \le 10,$$
$$2x_1 \le 40,$$
$$x_1, x_2 \ge 0.$$

d) Максимизировать $z = 3x_1 + 2x_2$
при ограничениях

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

9. В упр. 8(а) обозначим через y_1 и y_2 переменные двойственной задачи. Определите, какая из следующих пар решений прямой и двойственной задач является оптимальной.

a) $x_1 = 3, x_2 = 1; y_1 = 4, y_2 = 1.$
b) $x_1 = 4, x_2 = 1; y_1 = 1, y_2 = 0.$
c) $x_1 = 3, x_2 = 0; y_1 = 5, y_2 = 0.$

4.4. Экономическая интерпретация двойственности

Задачу линейного программирования можно рассматривать как модель распределения ограниченных ресурсов, в которой целевая функция, отображающая прибыль или доход от производственной деятельности, подлежит максимизации. Если рассматривать задачу ЛП с этой точки зрения, соответствующая ей двойственная задача получает интересную экономическую интерпретацию.

Чтобы формализовать рассматриваемый вопрос, приведем еще раз общее представление прямой и двойственной задач, причем прямая задача будет играть роль модели распределения ресурсов.

Прямая задача	Двойственная задача
Максимизировать $z = \sum_{j=1}^{n} c_j x_j$	Минимизировать $w = \sum_{i=1}^{m} b_i y_i$
при ограничениях	при ограничениях
$\sum_{j=1}^{n} a_{ij} x_j \le b_i,\ i = 1, 2, \ldots, m,$	$\sum_{i=1}^{m} a_{ij} y_i \ge c_j,\ j = 1, 2, \ldots, n,$
$x_j \ge 0, j = 1, 2, \ldots, n.$	$y_i \ge 0, i = 1, 2, \ldots, m.$

Исходя из модели распределения ресурсов, прямая задача отображает n видов экономической (производственной) деятельности и возможности получения m ресурсов. В прямой задаче коэффициент c_j представляет собой прибыль на единицу продукции j-го вида экономической деятельности, причем на единицу продукции этого вида деятельности расходуется a_{ij} единиц ресурса i, максимальные запасы которого ограничены величиной b_i.

4.4.1. Экономическая интерпретация переменных двойственной задачи

Соотношение 2 из раздела 4.3 устанавливает, что для любой пары решений прямой и двойственной задач значения (конечные) их целевых функций удовлетворяют неравенству

$$z = \sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i = w.$$

Строгое равенство здесь достигается только тогда, когда решения прямой и двойственной задач оптимальны.

Рассмотрим сначала случай оптимума, т.е. когда $z = w$. Исходя из представления прямой задачи как модели распределения ресурсов, можно считать, что величина z соответствует величине дохода (в долларах[1]). Поскольку b_i — общее доступное количество ресурса i, равенство $z = w$ можно переписать следующим образом.

$$\text{Доход (\$)} = \sum_i \text{(количество ресурса } i) \times \text{(доход (\$) на единицу ресурса } i).$$

Это означает, что переменная y_i двойственной задачи должна представлять *стоимость единицы ресурса i*. (Данное понятие уже вводилось в разделе 2.4.2, исходя из графического представления задачи ЛП.) В литературе по исследованию операций переменные y_i двойственной задачи часто называют **двойственными ценами**. Кроме того, иногда их именуют **теневыми ценами** и **симплексными мультипликаторами**.

Аналогично для любой пары допустимых решений прямой и двойственной задач неравенство $z < w$ можно интерпретировать следующим образом.

Доход < Общая стоимость ресурсов

Это соотношение показывает, что до тех пор, пока суммарный доход от всех видов деятельности строго меньше суммарной стоимости всех используемых ресурсов, решение как прямой, так и двойственной задачи не может быть оптимальным. Оптимум (максимальный доход) может быть достигнут только тогда, когда все потребляемые ресурсы использованы полностью. Если модель ЛП рассматривать более обще как модель некой системы, имеющую "вход" и "выход", то потребляемые ресурсы характеризуют "вход" этой системы, а получаемый доход — ее "выход". Система будет находиться в *нестабильном* (неоптимальном) состоянии, пока вход превышает выход. Устойчивое состояние системы характеризуется равенством входа и выхода.

Пример 4.4–1

Приведем формулировки прямой и двойственной задач, описывающие модель производства компании Reddy Mikks из примера 2.2–1.

Напомним вкратце, что в этой модели описывается производство двух видов краски (для внутренних и наружных работ) на основе двух видов сырья, M1 и M2 (ресурсы 1 и 2), с учетом рыночных условий, выражаемых третьим и четвертым ограничениями. Задача состоит в определении объемов производства красок каждого вида (в тоннах), при которых будет получен максимальный доход (в тысячах долларов).

[1] Читатель вместо долларов может, конечно, подставить любую другую денежную единицу — это не принципиально для понимания изложенного материала. — *Прим. ред.*

Прямая задача	Двойственная задача
Максимизировать $z = 5x_1 + 4x_2$ при ограничениях $6x_1 + 4x_2 \le 24$, (ресурс 1, сырье M1) $x_1 + 2x_2 \le 6$, (ресурс 2, сырье M2) $-x_1 + x_2 \le 1$, (ресурс 3) $x_2 \le 2$, (ресурс 3) $x_1, x_2 \ge 0$.	Минимизировать $w = 24y_1 + 6y_2 + y_3 + 2y_4$ при ограничениях $6y_1 + y_2 - y_3 \ge 5$, $4y_1 + 2y_2 + y_3 + y_4 \ge 4$, $y_1, y_2, y_3, y_4 \ge 0$.
Оптимальное решение $x_1 = 3, x_2 = 1.5, z = 21$	Оптимальное решение $y_1 = 0.75, y_2 = 0.5, y_3 = y_4 = 0, w = 21$

Оптимальное решение двойственной задачи показывает, что стоимость единицы первого ресурса (сырье M1) составляет $y_1 = 0.75$ (или \$750 за тонну), а второго (сырье M2) — $y_2 = 0.5$ (или \$500 за тонну). В разделе 2.4.2 мы графически показали, что приведенные значения стоимостей справедливы, если значение первого ресурса не выходит из интервала (20, 36), а второго — из интервала (4, 6.67) (эти же результаты алгебраически будут получены в разделе 4.7.1). Таким образом, расход сырья M1 может возрасти с 24 до 36 тонн, что приведет к соответствующему увеличению дохода на величину $12 \times \$750 = \$9000$. Аналогично количество второго ресурса (сырье M2) можно увеличить с 6 до 6.67 тонн с увеличением дохода на величину $0.67 \times \$500 = \$335$. Но еще раз напомним, что подобные расчеты применимы только тогда, когда увеличение числа используемых ресурсов не выходит за приведенные выше интервалы значений. Конечно, это не означает, что количество используемых ресурсов в принципе не может выходить за указанные пределы. Однако приведенные выше стоимости ресурсов определены только для случая, когда количество этих ресурсов не выходит за указанные пределы.

Для третьего и четвертого ресурсов двойственные цены (оптимальное решение двойственной задачи) равны нулю. Это указывает на то, что данные ресурсы недефицитны. Поэтому их стоимость равна нулю.

Упражнения 4.4,а

1. В задаче из примера 4.4–1 подсчитайте оптимальный доход при выполнении следующих условий.

   a) Ограничение для первого ресурса: $6x_1 + 4x_2 \le 22$.

   b) Ограничение для второго ресурса: $x_1 + 2x_2 \le 4.5$.

   c) Четвертое ограничение: $x_2 \le 10$.

2. Электронная компания NWAC производит четыре типа кабеля для оборонного ведомства. Каждый тип кабеля подвергается четырем последовательным операциям: разделка, пайка, оплетка и проверка. В следующей таблице приведены данные, характеризующие производство кабелей.

Тип кабеля	Затраты времени на изделие (в минутах)				Доход ($)
	Разделка	Пайка	Оплетка	Проверка	
SC320	10.5	20.4	3.2	5.0	9.40
SC325	9.3	24.6	2.5	5.0	10.80
SC340	11.6	17.7	3.6	5.0	8.75
SC370	8.2	26.5	5.5	5.0	7.80
Ежедневный фонд рабочего времени (в минутах)	4800.0	9600.0	4700.0	4500.0	

Оборонное ведомство гарантирует для компании минимальный уровень производства в 100 единиц каждого типа кабеля.

a) Сформулируйте задачу линейного программирования и с помощью программы TORA найдите ее оптимальное решение.

b) Основываясь на двойственных ценах, приведенных программой TORA, определите возможное увеличение ежедневного фонда времени по каждой технологической операции.

c) Выгодно ли компании выполнение требования заданного минимального уровня производства? Обоснуйте ответ, основываясь на величинах двойственных цен.

d) Возможно ли увеличение на 10% временного фонда операции пайки с сохранением величины ее вклада в суммарный доход, определяемый текущей двойственной ценой?

3. Компания производит кожаные чехлы и сумки. На производство одного чехла требуется 8 м^2 кожи и 12 часов рабочего времени, на производство сумки — 2 м^2 кожи и 5 часов рабочего времени. Текущие еженедельные ресурсы производства ограничены 1200 м^2 кожи и 1850 часами рабочего времени. Компания продает чехлы и сумки по цене $350 и $130 соответственно. Определите для этой компании схему производства, максимизирующую чистую прибыль. Допустим, компания желает расширить свое производство. Какова максимальная цена, по которой компании имеет смысл закупать дополнительную кожу? А какова допустимая максимальная цена дополнительных трудовых ресурсов?

4.4.2. Экономическая интерпретация ограничений двойственной задачи

Для интерпретации ограничений двойственной задачи используем соотношение 1 из раздела 4.3. В соответствии с этим соотношением на любой итерации решения прямой задачи справедливо равенство

$$\text{Коэффициент при } x_j \text{ в } z\text{-строке} = \sum_{i=1}^{m} a_{ij} y_i - c_j \ .$$

Для интерпретации этого равенства воспользуемся анализом размерностей входящих в него величин. Коэффициент c_j — доход (в долларах) на единицу "выхода" j-го вида производственной (экономической) деятельности. Следовательно, для согласования размерностей величина $\sum_{i=1}^{m} a_{ij} y_i$ также должна иметь размерность доллар/единица. Далее, поскольку c_j представляет доход, сумма $\sum_{i=1}^{m} a_{ij} y_i$, которая входит в равенство с противоположным знаком, должна соответствовать затратам или издержкам. В то же время коэффициент a_{ij} равен количеству ресурса i, используемого на поддержание j-го вида деятельности; переменная y_i представляет **вменяемые издержки** (вменяемую стоимость) единицы ресурса i. Таким образом, величина $\sum_{i=1}^{m} a_{ij} y_i$ соответствует суммарной стоимости всех ресурсов, необходимых для производства единицы продукции j-го вида деятельности.

Условие оптимальности симплекс-метода в задаче максимизации говорит о том, что j-й вид деятельности (переменная x_j), не представленный в текущем базисном решении, можно ввести в базис для увеличения дохода только тогда, когда коэффициент при x_j в z-строке (равный $\sum_{i=1}^{m} a_{ij} y_i - c_j$) будет неотрицательным. В рамках предлагаемой экономической интерпретации это означает, что j-й вид деятельности должен быть представлен в базисном решении, если выполняется следующее неравенство.

$$\begin{pmatrix} \text{Стоимость всех ресурсов, используемых} \\ \text{для производства единицы продукции} \\ j\text{-го вида деятельности} \end{pmatrix} < \begin{pmatrix} \text{Доход от реализации} \\ \text{единицы продукции} \\ j\text{-го вида деятельности} \end{pmatrix}$$

Таким образом, условие оптимальности (в задаче максимизации) говорит о том, что деятельность любого вида следует наращивать до тех пор, пока доход от нее превышает возможные издержки (стоимость ресурсов), затрачиваемые на ее поддержку.

Приведем стандартные определения, используемые в литературе по линейному программированию. Введем обозначение $z_j = \sum_{i=1}^{m} a_{ij} y_i$. Величина z_j представляет суммарную стоимость ресурсов, используемых на производство единицы продукции j-го вида деятельности. Величина $z_j - c_j$ равна коэффициенту при x_j в z-строке симплекс-таблицы и часто называется **приведенной стоимостью** (приведенными издержками) j-го вида деятельности. В некоторых случаях разности $z_j - c_j = \sum_{i=1}^{m} a_{ij} y_i - c_j$ используются непосредственно для вычисления коэффициентов в z-строке симплекс-таблицы (вместо метода Гаусса-Жордана). Такие вычисления используются в *модифицированном симплекс-методе* (этот метод описан в главе 7).

Пример 4.4–2

Фабрика игрушек TOYCO собирает три вида игрушек: модели поездов, грузовиков и легковых автомобилей; при сборке каждого вида используется три типа операций. Ежедневный фонд рабочего времени на каждую операцию ограничен предельными величинами 430, 460 и 420 минут. Доход на одну игрушку каждого вида составляет соответственно \$3, \$2 и \$5. На каждой из трех операций для сборки модели поезда требуется 1, 2 и 1 минуты рабочего времени. Соответствующее время для сборки моделей грузовиков и легковых автомобилей составляет $(2, 0, 4)$ и $(1, 2, 0)$ минут (нуль указывает на то, что соответствующая операция не выполняется).

Обозначив через x_1, x_2 и x_3 количество собираемых ежедневно моделей трех видов, получаем прямую и двойственную задачи ЛП.

Прямая задача	Двойственная задача
Максимизировать $z = 3x_1 + 2x_2 + 5x_3$ при ограничениях $x_1 + 2x_2 + x_3 \le 430$, (операция 1) $3x_1 + 2x_3 \le 460$, (операция 2) $x_1 + 4x_2 \le 420$, (операция 3) $x_1, x_2, x_3 \ge 0.$	Минимизировать $w = 430y_1 + 460y_2 + 420y_3$ при ограничениях $y_1 + 3y_2 + y_3 \ge 3,$ $2y_1 + 4y_3 \ge 2,$ $y_1 + 2y_2 \ge 5,$ $y_1, y_2, y_3 \ge 0.$
Оптимальное решение $x_1 = 0, x_2 = 100, x_3 = 230, z = \$1350$	Оптимальное решение $y_1 = 1, y_2 = 2, y_3 = 0$

Оптимальное решение предусматривает производство моделей грузовых ($x_2 = 100$) и легковых ($x_3 = 230$) автомобилей и требует отказа от производства моделей поездов ($x_1 = 0$). Это означает, что в текущей экономической ситуации производство моделей поездов не рентабельно. Вместе с тем рынок игрушек требует выпуска этого вида моделей. Как сделать их производство доходным? В соответствии с экономической интерпретацией задач ЛП, приведенной в этом разделе, производство моделей поездов будет выгодным только тогда, когда будет выполняться неравенство $z_1 < c_1$. Для выполнения этого неравенства нужно либо повысить коэффициент c_1 (доход от продажи одной модели поезда), например путем увеличения цены модели, либо снизить стоимость ресурсов z_1 ($= y_1 + 3y_2 + y_3$), необходимых для производства этих игрушек. Увеличение цены игрушек не желательно, так как это снизит их конкурентоспособность на рынке игрушек. Уменьшение величины коэффициента z_1 более привлекательно, поскольку для этого надо просто сократить время выполнения операций, необходимых для производства моделей поездов. Обозначим через r_1, r_2 и r_3 величины, пропорциональные долям сокращения времени соответствующих операций. Эти величины находим из условия, чтобы новая стоимость производственных операций не превышала дохода от одной модели поезда. Это условие записывается следующим образом.

$$1(1 - r_1)y_1 + 3(1 - r_2)y_2 + 1(1 - r_3)y_3 < 3$$

После подстановки значений $y_1 = 1$, $y_2 = 2$ и $y_3 = 0$ получим следующее неравенство (проверьте!): $r_1 + 6r_2 > 4$.

Таким образом, любые значения величин r_1 и r_2, от 0 до 1, удовлетворяющие неравенству $r_1 + 6r_2 > 4$, приведут к доходности производства моделей поездов. Например, для значений $r_1 = 0.6$ и $r_2 = 0.6$ получаем $z_1 - c_1 = 4 - 0.6 - 6\times0.6 = -0.2$. Вместе с тем отметим, что сокращение времени выполнения второй операции в 6 раз эффективнее сокращения времени выполнения первой операции.

Упражнения 4.4,b

1. В задаче из примера 4.4–2 предположим, что время выполнения второй операции при сборке модели поезда сокращено с 3 до 1.25 минуты. На сколько должно быть сокращено время выполнения первой операции, чтобы производство этой игрушки стало доходным?

2. В задаче из примера 4.4–2 предположим, что фабрика игрушек рассматривает возможность производства еще одного вида игрушки, а именно модели пожарной машины. При сборке этой модели первая операция не используется, а вторая и третья требуют соответственно 1 и 3 минут для сборки одной модели. Доход от одной модели пожарной машины составляет $4. Посоветуете ли вы фабрике производить эти игрушки?

3. Компания использует токарные и сверлильные станки для производства четырех типов деталей: PP1, PP2, PP3 и PP4. В следующей таблице представлены технологические данные, характеризующие производство этих деталей.

	Время обработки одного изделия (минуты)				Фонд машинного времени (минуты)
Станок	PP1	PP2	PP3	PP4	
Токарный	2	5	3	4	5300
Сверлильный	3	4	6	4	5300
Доход от одного изделия ($)	3	6	5	4	

Для тех изделий, которые не войдут в оптимальное базисное решение, определите степень уменьшения оптимального дохода при увеличении их производства на единицу.

4. Рассмотрите оптимальное решение задачи из предыдущего упражнения. Компания подсчитала, что с помощью специальных мероприятий можно уменьшить общее время производства изделий, не вошедших в оптимальное базисное решение, на 20%. Будет ли после этого производство таких изделий рентабельно? Если нет, то на сколько следует сократить время производства данных изделий?

4.5. Двойственный симплекс-метод

В этом разделе рассмотрим симплексный алгоритм, который основан на соотношениях между прямой и двойственной задачами. Этот алгоритм эффективно решает определенный класс задач ЛП. Кроме того, он позволяет довести до конца анализ чувствительности модели ЛП (см. раздел 4.7).

В двойственном симплекс-методе решение задачи ЛП начинается с недопустимого, но лучшего, чем оптимальное решение. Последовательные итерации этого метода приближают решение к области допустимости без нарушения оптимальности (точнее, "супероптимальности") промежуточных решений. Когда будет достигнута область допустимых решений, процесс вычислений заканчивается, так как последнее решение будет оптимальным. Уже из этого краткого описания двойственного симплекс-метода видно, что он значительно отличается от обычного (прямого) симплекс-метода, рассмотренного в главе 3, где начальное решение всегда допустимое, но не оптимальное, и промежуточные решения никогда не выходят из пространства допустимых решений.

В двойственном симплекс-методе начальная симплекс-таблица обязательно должна иметь в базисном решении недопустимую (т.е. отрицательную) переменную. Для реализации двойственного симплекс-метода разработаны следующие два условия, выполнение которых гарантирует оптимальность последовательных промежуточных решений и приближение их к области допустимых решений.

Двойственное условие допустимости. В качестве исключаемой переменной x_r выбирается базисная переменная, имеющая наибольшее по абсолютной величине отрицательное значение. Если таких переменных несколько, то выбор произволен. Если все базисные переменные неотрицательные, процесс вычислений заканчивается.

Двойственное условие оптимальности. Вводимая в базис переменная определяется как переменная, на которой достигается следующий минимум:

$$\min_{\text{небазисные } x_j} \left\{ \left| \frac{z_j - c_j}{\alpha_{rj}} \right|, \quad \alpha_{rj} < 0 \right\},$$

где α_{rj} — коэффициент из симплекс-таблицы, расположенный на пересечении ведущей строки (соответствующей исключаемой переменной x_r) и столбца, соответствующего переменной x_j. При наличии нескольких альтернативных переменных, выбор делается произвольно.

После некоторых раздумий вы должны заметить, что двойственное условие оптимальности основывается на той же основной идее, что и условие оптимальности прямого симплекс-метода из главы 3.

Пример 4.5–1

Дана следующая задача ЛП.

Минимизировать $z = 3x_1 + 2x_2$

при ограничениях

$$3x_1 + x_2 \geq 3,$$
$$4x_1 + 3x_2 \geq 6,$$
$$x_1 + x_2 \leq 3,$$
$$x_1, x_2 \geq 0.$$

Начальная симплекс-таблица этой задачи имеет следующий вид.

Базис	x_1	x_2	x_3	x_4	x_5	Решение
z	–3	–2	0	0	0	0
x_3	–3	–1	1	0	0	–3
x_4	–4	–3	0	1	0	–6
x_5	1	1	0	0	1	3

Среди дополнительных переменных этой задачи, x_3 и x_4 являются избыточными, а x_5 — остаточной. Мы умножили каждое равенство, соответствующее избыточным дополнительным переменным, на –1; в результате правые части этих равенств непосредственно указывают на базисные переменные, которые являются недопустимыми ($x_3 = -3$, $x_4 = -6$, $x_5 = 3$). Этот подход всегда используется при реализации двойственного симплекс-метода. Поскольку $z_j - c_j \leq 0$ для всех $j = 1, ..., 5$, начальное базисное решение является оптимальным (но не допустимым). Таким образом, приведенная таблица удовлетворяет требованиям начальной таблицы двойственного симплекс-метода, а именно — оптимальности и недопустимости.

Двойственное условие допустимости указывает на переменную x_4 (= –6) как на вводимую в базис. Теперь применим двойственное условие оптимальности для определения исключаемой переменной. Для этого используем следующую таблицу.

Переменные	x_1	x_2	x_3	x_4	x_5
z-строка ($z_j - c_j$)	–3	–2	0	0	0
x_4-строка, α_{4j}	–4	–3	0	1	0
Отношение $\left\|\frac{z_j - c_j}{\alpha_{4j}}\right\|$	$\frac{3}{4}$	$\frac{2}{3}$	—	—	—

Приведенные отношения показывают, что исключаемой переменной будет x_2. Отметим, что переменные x_j будут кандидатами на исключение из базисного решения только тогда, когда коэффициент α_{4j} будет строго отрицательным. По этому критерию переменные x_3, x_4 и x_5 не рассматриваются как кандидаты на исключение из базиса.

Следующая таблица получена с помощью известных операций над строками, применяемых в прямом симплекс-методе.

Базис	↓ x_1	x_2	x_3	x_4	x_5	Решение
z	–1/3	0	0	–2/3	0	4
← x_3	–5/3	0	1	–1/3	0	–1
x_2	4/3	1	0	–1/3	0	2
x_5	–1/3	0	0	1/3	1	1
Отношение	1/5	—	—	2	—	

Последняя таблица показывает, что из базиса исключается переменная x_3 и вводится x_1. В результате получаем следующую симплекс-таблицу.

Базис	x_1	x_2	x_3	x_4	x_5	Решение
z	0	0	–1/5	–3/5	0	21/5
x_1	1	0	–3/5	1/5	0	3/5
x_2	0	1	4/5	–3/5	0	6/5
x_5	0	0	–1/5	2/5	1	6/5

Решение, представленное в последней таблице, допустимо (и оптимально), поэтому вычисления заканчиваются. Это решение имеет вид $x_1 = 3/5$, $x_2 = 6/5$ и $z = 21/5$.

На рис. 4.1 показана последовательность шагов двойственного симплекс-метода при решении задачи из примера 4.5–1. Алгоритм начинается в крайней точке A (которой соответствует недопустимое, но "лучше, чем оптимальное" решение), затем он переходит к точке B (которой также соответствует недопустимое, но "лучше, чем оптимальное" решение) и заканчивается в точке C, уже принадлежащей области допустимых решений.

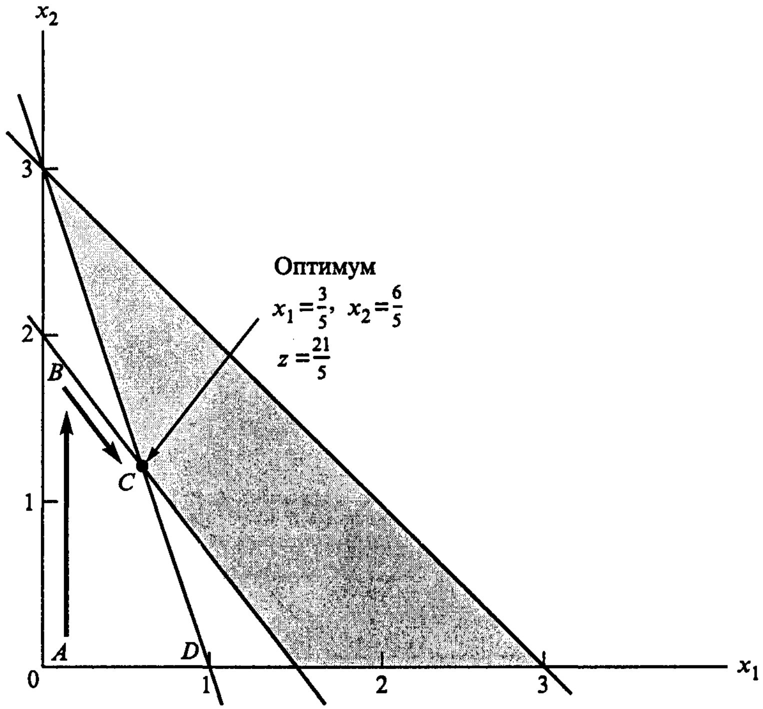


Рис. 4.1

Упражнения 4.5,а

1. На рис. 4.2 показано пространство решений, соответствующее задаче минимизации целевой функции $z = 2x_1 + x_2$. Предполагается, что поиск решения выполняется двойственным симплекс-методом; оптимальное решение соответствует точке $F = (0.5, 1.5)$.

   a) Если начальное базисное (недопустимое) решение соответствует точке G, будет ли алгоритм двойственного симплекс-метода проходить через точки G, E, F? Обоснуйте это.

   b) Если начальное базисное (недопустимое) решение соответствует точке L, определите на рис. 4.2 возможную последовательность точек, через которые будет проходить алгоритм двойственного симплекс-метода для достижения оптимального решения в точке F.

2. Решите следующие задачи с помощью двойственного симплекс-метода и определите на графически представленном пространстве решений этих задач последовательность точек прохождения алгоритма двойственного симплекс-метода для достижения оптимального решения.

   a) Минимизировать $z = 2x_1 + 3x_2$
   при ограничениях

$$2x_1 + 2x_2 \leq 30,$$
$$x_1 + 2x_2 \geq 10,$$
$$x_1, x_2 \geq 0.$$

b) Минимизировать $z = 5x_1 + 6x_2$
при ограничениях

$$x_1 + x_2 \geq 2,$$
$$4x_1 + x_2 \geq 4,$$
$$x_1, x_2 \geq 0.$$

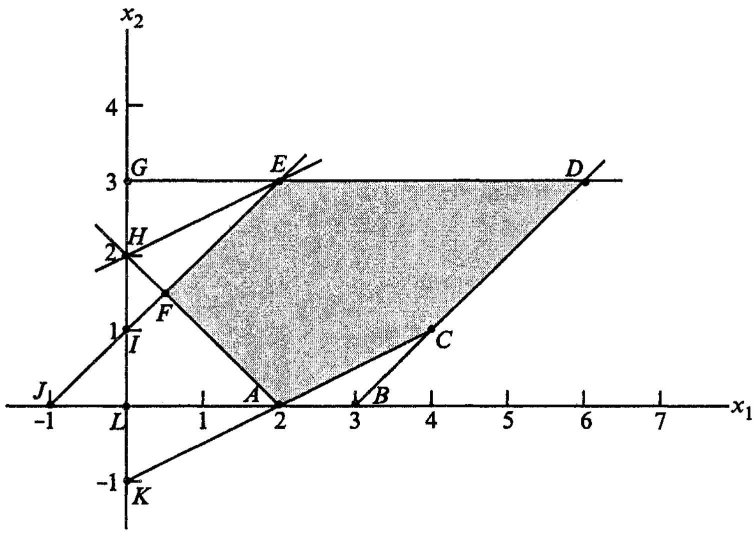


Рис. 4.2

3. Решите следующие задачи двойственным симплекс-методом и определите на графически представленном пространстве решений этих задач последовательность точек прохождения алгоритма двойственного симплекс-метода для достижения оптимального решения.

a) Минимизировать $z = 4x_1 + 2x_2$
при ограничениях

$$x_1 + x_2 = 1,$$
$$3x_1 - x_2 \geq 2,$$
$$x_1, x_2 \geq 0.$$

b) Минимизировать $z = 2x_1 + 3x_2$
при ограничениях

$$2x_1 + x_2 \geq 3,$$
$$x_1 + x_2 = 2,$$
$$x_1, x_2 \geq 0.$$

(*Совет.* Замените ограничение в виде равенства двумя неравенствами.)

4. *Двойственный симплекс-метод с искусственными ограничениями.* Дана следующая задача ЛП.

$$\text{Максимизировать } z = 2x_1 - x_2 + x_3$$

при ограничениях

$$2x_1 + 3x_2 - 5x_3 \geq 4,$$
$$-x_1 + 9x_2 - x_3 \geq 3,$$
$$4x_1 + 6x_2 + 3x_3 \leq 8,$$
$$x_1, x_2, x_3 \geq 0.$$

Начальное базисное решение содержит дополнительные переменные x_4, x_5 и x_6 и является недопустимым, поскольку $x_4 = -4$ и $x_5 = -3$. Но непосредственное применение двойственного симплекс-метода невозможно, так как переменные x_1 и x_3 не удовлетворяют условию оптимальности для задачи максимизации. Покажите, что введение искусственного ограничения $x_1 + x_3 \leq M$ (где M — достаточно большое положительное число и такое, что данное неравенство не сужает область допустимых решений исходной задачи) и последующее использование его как ведущей строки симплекс-таблицы позволяют получить оптимальную z-строку путем выбора переменной x_1 в качестве вводимой. Таким образом становится возможным применение двойственного симплекс-метода к модифицированной задаче с искусственным ограничением.

5. Используя процедуру введения искусственного ограничения, описанную в предыдущем упражнении, решите двойственным симплекс-методом следующие задачи ЛП. Во всех задачах укажите, будет ли окончательное решение допустимым, недопустимым или неограниченным.

a) Максимизировать $z = 2x_3$
при ограничениях

$$-x_1 + 2x_2 - 2x_3 \geq 8,$$
$$-x_1 + x_2 + x_3 \leq 4,$$
$$2x_1 - x_2 + 4x_3 \leq 10,$$
$$x_1, x_2, x_3 \geq 0.$$

b) Максимизировать $z = x_1 - 3x_2$
при ограничениях

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

c) Минимизировать $z = -x_1 + x_2$
при ограничениях

$$x_1 - 4x_2 \geq 5,$$
$$x_1 - 3x_2 \leq 1,$$
$$2x_1 - 5x_2 \geq 1,$$
$$x_1, x_2 \geq 0.$$

d) Максимизировать $z = 2x_3$
при ограничениях

$$-x_1 + 3x_2 - 7x_3 \geq 5,$$
$$-x_1 + x_2 - x_3 \leq 1,$$
$$3x_1 + x_2 - 10x_3 \leq 8,$$
$$x_1, x_2, x_3 \geq 0.$$

6. Решите следующую задачу ЛП тремя различными методами (используя в качестве инструмента программу TORA). Определите, какой метод будет наиболее эффективен при вычислениях.

$$\text{Минимизировать } z = 6x_1 + 7x_2 + 3x_3 + 5x_4$$

при ограничениях

$$5x_1 + 6x_2 - 3x_3 + 4x_4 \geq 12,$$
$$x_2 - 5x_3 - 6x_4 \geq 10,$$
$$2x_1 + 5x_2 + x_3 + x_4 \geq 8,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

7. *Обобщение симплексного алгоритма.* В прямом симплекс-методе (описан в главе 3) начальное решение допустимо, но не оптимально. В двойственном симплекс-методе данное решение оптимально (точнее, "супероптимально"), но не допустимо. Возникает естественный вопрос: можно ли начать решение задачи ЛП с неоптимального и недопустимого решений? Мы видели, что в прямом симплекс-методе при отсутствии допустимого начального решения используются искусственные переменные. В двойственном симплекс-методе при отсутствии оптимального начального решения применяются искусственные ограничения. Хотя задача этих процедур и состоит в обеспечении *автоматического* выполнения вычислений, необходимо не терять из виду основную идею симплексных алгоритмов, а именно то, что оптимальное решение задачи ЛП достигается в одной из крайних (угловых) точек пространства допустимых решений. С учетом этих замечаний разработайте собственный симплексный алгоритм решения задач ЛП, где начальное решение будет и неоптимальным, и недопустимым. Для проверки своего алгоритма решите задачу из упр. 5(а). В качестве начальной таблицы можно принять следующую симплекс-таблицу, где представлено начальное решение, которое не оптимально (из-за переменной x_3) и не допустимо (так как $x_4 = -8$). (Заметим, что в этой таблице первое равенство умножено на -1 для того, чтобы показать недопустимость решения непосредственно в столбце "Решение".)

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	0	0	-2	0	0	0	0
x_4	1	-2	2	1	0	0	-8
x_5	-1	1	1	0	1	0	4
x_6	2	-1	4	0	0	1	10

Решение задачи ЛП без использования каких-либо искусственных переменных или ограничений может быть следующим. Сначала освобождаемся от свойства недопустимости базисного решения путем применения версии двойственного условия допустимости. В нашем примере это приведет к выбору переменной x_4 в качестве исключаемой из базиса. Для определения вводимой переменной надо найти в x_4-строке строго отрицательный коэффициент, соответствующий небазисной переменной. Выбор вводимой переменной можно осуществить без удовлетворения требования оптимальности решения, так как в данном случае это не существенно (сравните с двойственным условием оптимальности). Описанная процедура повторяется до тех пор, пока не будет получено допустимое решение. Далее основное внимание уделяется оптимальности решения путем применения условия оптимальности прямого симплекс-метода.

Данный пример симплексного метода характеризуется гибкостью. В литературе описано большое количество вариаций симплекс-метода (например, метод одновременного решения прямой и двойственной задач, симметричный метод, перекрестный метод и мультиплексный метод), причем создается впечатление, что каждый из них существенно отличается от других, тогда как все они просматривают экстремальные точки пространства решений с различной степенью автоматизации вычислений и вычислительной эффективности.

8. Задача ЛП из упр. 5(c) не имеет допустимого решения. Покажите, что это свойство задачи ЛП можно определить с помощью *обобщенного симплексного алгоритма*, описанного в предыдущем упражнении.
9. Задача ЛП из упр. 5(d) не имеет ограниченного решения. Покажите, что это свойство задачи ЛП можно определить с помощью *обобщенного симплексного алгоритма*, описанного в упр. 7.

4.6. Матричное представление симплексных вычислений

В анализе чувствительности решения задачи ЛП нас интересует один принципиальный вопрос: как изменение коэффициентов целевой функции и ограничений влияет на оптимальность и допустимость текущего оптимального решения? Если эти изменения ведут к новому решению, то как найти это новое оптимальное решение (если оно существует)? Конечно, мы всегда можем ответить на этот вопрос путем решения задачи заново. Но для типовой практической задачи, имеющей тысячи переменных и ограничений, такой путь будет не эффективным. Поэтому необходимо выяснить, как вычисления в симплекс-методе зависят от изменений коэффициентов исходной задачи. В частности, следует выяснить, как от этих изменений зависят оптимальность и допустимость решения, представленного в виде симплекс-таблицы.

Наиболее компактный способ записи вычислений, производимых при симплекс-методе, заключается в использовании матриц. Для этого не потребуется знаний из теории матриц, нам необходимы только три элементарные матричные операции умножения: вектор-строка на матрицу, матрица на вектор-столбец и скаляр (число) на матрицу. Для удобства мы сформулируем определение этих операций, а также некоторые другие основополагающие определения (в Приложении А приведен более полный обзор теории матриц).

1. *Матрица* **A** порядка $m \times n$ — это прямоугольный массив элементов с m строками и n столбцами.
2. *Вектор-строка* **V** размерности n — матрица порядка $1 \times n$.
3. *Вектор-столбец* **P** размерности m — матрица порядка $m \times 1$.

Эти определения математически можно представить следующим образом.

$$\mathbf{V} = (v_1, v_2, \dots, v_n), \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} p_1 \\ p_2 \\ . \\ . \\ . \\ p_m \end{bmatrix}.$$

Теперь определим три матричные операции умножения.

1. *Умножение вектор-строки* **V** *на матрицу* **A**: $\mathbf{V} \times \mathbf{A}$. Эта операция определена только тогда, когда количество элементов в вектор-строке **V** равно числу строк в матрице **A**.

$$\mathbf{V} \times \mathbf{A} = \left(\sum_{i=1}^{m} v_i a_{i1}, \sum_{i=1}^{m} v_i a_{i2}, \dots, \sum_{i=1}^{m} v_i a_{in} \right).$$

   Например,

$$(11, 22, 33) \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = (1 \times 11 + 3 \times 22 + 5 \times 33,\ 2 \times 11 + 4 \times 22 + 6 \times 33) = (242, 308).$$

2. *Умножение матрицы* **A** *на вектор-столбец* **P**: $\mathbf{A} \times \mathbf{P}$. Эта операция определена только тогда, когда количество столбцов матрицы **A** равно количеству элементов вектор-столбца **P**.

$$\mathbf{A} \times \mathbf{P} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} p_j \\ \sum_{j=1}^{n} a_{2j} p_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj} p_j \end{bmatrix}.$$

   Например,

$$\begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} 11 \\ 22 \\ 33 \end{bmatrix} = \begin{bmatrix} 11 \times 1 + 22 \times 3 + 33 \times 5 \\ 11 \times 2 + 22 \times 4 + 33 \times 6 \end{bmatrix} = \begin{bmatrix} 242 \\ 308 \end{bmatrix}.$$

3. *Умножение скаляра на матрицу.* При умножении скаляра (константы) α на матрицу **A**, $\alpha\mathbf{A}$, получаем матрицу порядка $m \times n$, у которой (i, j)-й элемент равен αa_{ij}. Например, для $\alpha = 10$ имеем

$$(10)\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{bmatrix}.$$

В общем случае $\alpha\mathbf{A} = \mathbf{A}\alpha$. Этим же свойством обладает операция умножения вектора на скаляр: $\alpha\mathbf{V} = \mathbf{V}\alpha$ и $\alpha\mathbf{P} = \mathbf{P}\alpha$.

С применением матричных обозначений общую задачу ЛП можно записать следующим образом.

$$\text{Максимизировать } z = \sum_{j=1}^{n} c_j x_j$$

при ограничениях

$$\sum_{j=1}^{n} \mathbf{P}_j x_j = \mathbf{b},$$

$$\text{для всех } x_j \ge 0,$$

где $\mathbf{P}_j$ и $\mathbf{b}$ — вектор-столбцы размерности m.

При анализе чувствительности, как указывалось выше, нас интересует влияние (на оптимальное решение задачи ЛП) изменений коэффициентов c_j целевой функции и правых частей ограничений, представленных вектором **b**. Для того чтобы разобраться в этом, используем обобщенное матричное представление симплекс-таблиц. (Математическое обоснование такого представления дано в разделе 7.4.)

Базис	$\ldots x_j \ldots$	Начальный базис $\mathbf{X}_Б$	Решение
z	$z_j - c_j$		
$\mathbf{X}_Б$	$\mathbf{B}^{-1}\mathbf{P}_j$	$\mathbf{B}^{-1}$	$\mathbf{B}^{-1}\mathbf{b}$

Эта таблица имеет точно такой же формат, как и обычная симплекс-таблица, описанная в главе 3. Здесь вектор $\mathbf{X}_Б$ содержит базисные переменные текущего решения. Матрица $\mathbf{B}^{-1}$, расположенная под начальными базисными переменными, называется **обратной**. Элементы обратной матрицы будут изменяться при переходе от одной симплекс-таблицы к другой как функции вектора $\mathbf{X}_Б$.

Вычисление элементов ограничений, соответствующих новой симплекс-таблице, выполняется непосредственно. Например, столбцы $\mathbf{P}_j$ и **b** исходной задачи (начальной симплекс-таблицы) преобразуются в столбцы $\mathbf{B}^{-1}\mathbf{P}_j$ и $\mathbf{B}^{-1}\mathbf{b}$ следующей симплекс-таблицы. Осталось рассмотреть вычисление элементов z-строки. Обозначим через $\mathbf{Y} = (y_1, y_2, ..., y_m)$ m-мерный вектор переменных двойственной задачи. В разделе 4.3 было показано, что значения переменных двойственной задачи можно найти как решение системы линейных уравнений. В компактной матричной форме это можно записать так: $\mathbf{Y} = \mathbf{C}_Б\mathbf{B}^{-1}$, где $\mathbf{C}_Б$ — m-мерный вектор, состоящий из коэффициентов c_j *исходной целевой функции*, соответствующих базисному вектору $\mathbf{X}_Б$ (эта формула будет доказана в разделе 7.4). Таким образом, разности $z_j - c_j$ можно вычислить следующим образом.

$$z_j - c_j = \sum_{i=1}^{m} a_{ij} y_i - c_j = \mathbf{Y}\mathbf{P}_j - c_j.$$

Последнее выражение показывает, что величина $z_j - c_j$ равна разности между левой и правой частями ограничения двойственной задачи (как и показано в разделе 4.4.2).

Таким образом, все элементы текущей симплекс-таблицы вычисляются на основе текущей обратной матрицы $\mathbf{B}^{-1}$ и *исходных данных* задачи. Этот вывод будет использован в следующем разделе при выполнении анализа чувствительности задач ЛП.

Пример 4.6–1

Используем задачу из примера 4.3–1 для демонстрации матричных вычислений. Прямая и двойственная задачи этого примера с помощью матриц будут записаны следующим образом.

Прямая задача	Двойственная задача
Максимизировать $z = (5, 12, 4)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$	Минимизировать $w = (y_1, y_2)\begin{bmatrix} 10 \\ 8 \end{bmatrix}$
при ограничениях	при ограничениях
$\begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 8 \end{bmatrix},$	$(y_1, y_2)\begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 3 \end{bmatrix} \geq (5, 12, 4),$
$x_1, x_2, x_3 \geq 0.$	$y_1 \geq 0, y_2$ — свободная переменная.

Описываемые вычисления применимы на любом шаге выполнения симплекс-метода. Продемонстрируем их на первой итерации симплекс-метода, основываясь на начальном решении, приведенном в примере 4.3–1. Из этого решения имеем

$$\mathbf{X}_Б = \begin{bmatrix} x_4 \\ x_3 \end{bmatrix}, \mathbf{C}_Б = (0, 4)$$

(Отметим, что порядок элементов в векторе $\mathbf{C}_Б$ должен быть такой же, как и в базисном векторе $\mathbf{X}_Б$, т.е. сначала c_4, затем c_3.) Обратная матрица имеет следующий вид

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & \frac{1}{3} \end{bmatrix}$$

Теперь вычисляем вектор переменных двойственной задачи.

$$(y_1, y_2) = \mathbf{Y} = \mathbf{C}_Б\mathbf{B}^{-1} = (0, 4)\begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & \frac{1}{3} \end{bmatrix} = \left(0, \frac{4}{3}\right).$$

Далее находим коэффициенты z-строки.

$$z_1 - c_1 = \mathbf{YP}_1 - c_1 = \left(0, \frac{4}{3}\right)\begin{bmatrix} 1 \\ 2 \end{bmatrix} - 5 = -\frac{7}{3}.$$

Аналогично вычисляются следующие разности: $z_2 - c_2 = -40/3$, $z_3 - c_3 = 0$, $z_4 - c_4 = 0$ и $z_5 - c_5 = 4/3 + M$ (проверьте!).

Коэффициенты равенств, соответствующих ограничениям, вычисляются с использованием обратной матрицы $\mathbf{B}^{-1}$ и вектор-столбцов $\mathbf{P}_j$ и $\mathbf{b}$ коэффициентов *исходных* ограничений. Столбец правых частей ограничений на первой итерации симплекс-метода вычисляется следующим образом.

$$\mathbf{X}_Б = \mathbf{B}^{-1}\mathbf{b} = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 10 \\ 8 \end{bmatrix} = \begin{bmatrix} \frac{22}{3} \\ \frac{8}{3} \end{bmatrix}.$$

Вычисление столбцов коэффициентов левых частей ограничений покажем на примере вычисления столбца коэффициентов, соответствующего переменной x_1.

$$\mathbf{B}^{-1}\mathbf{P}_1 = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Столбцы коэффициентов, соответствующие переменным x_2, x_3, x_4, x_5 и R, вычисляются аналогично (вычислите их!).

Основной вывод, следующий из примера 4.6–1, заключается в том, что все коэффициенты симплекс-таблицы на любой итерации можно вычислить только на основании соответствующей обратной матрицы $\mathbf{B}^{-1}$ и *исходных данных* задачи ЛП. Таким образом, при анализе чувствительности оптимального решения конкретной задачи ЛП, где известна обратная матрица $\mathbf{B}^{-1}$ этого решения, можно исследовать эффект от изменения коэффициентов целевой функции и значений правых частей неравенств ограничений посредством новых вычислений всех разностей $z_j - c_j$ (коэффициенты z-строки) и произведения $\mathbf{B}^{-1}\mathbf{b}$ (значения правых частей ограничений, т.е. коэффициенты в столбце "Решение" симплекс-таблицы). Если результаты этих вычислений покажут, что текущее базисное решение (т.е. вектор $\mathbf{X}_Б$) остается допустимым и оптимальным, вычисления анализа чувствительности заканчиваются. В противном случае необходимы дополнительные вычисления, возвращающие оптимальность и допустимость исследуемому решению. В следующем разделе эти вопросы будут рассмотрены подробно.

Упражнения 4.6,а

1. В задаче из примера 4.3–1 определите элементы обратной матрицы
   a) на второй итерации симплекс-метода,
   b) на третьей итерации симплекс-метода.

2. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 3x_1 + 2x_2 + 5x_3$$

при ограничениях

$$\begin{aligned} &x_1 + 2x_2 + x_3 + x_4 = 30, \\ &3x_1 + 2x_2 + x_5 = 60, \\ &x_1 + 4x_2 + x_6 = 20, \\ &x_1, x_2, x_3, x_4, x_5, x_6 \geq 0. \end{aligned}$$

Проверьте оптимальность и допустимость следующих базисных решений.

a) $\mathbf{X}_Б = \begin{bmatrix} x_4 \\ x_3 \\ x_6 \end{bmatrix}, \quad \mathbf{B}^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$

b) $\mathbf{X}_Б = \begin{bmatrix} x_2 \\ x_3 \\ x_1 \end{bmatrix}, \quad \mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{4} & -\frac{1}{8} & \frac{1}{8} \\ \frac{3}{2} & -\frac{1}{4} & -\frac{3}{4} \\ -1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$

c) $\mathbf{X}_Б = \begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix}, \quad \mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix}.$

3. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 4x_1 + 14x_2$$

при ограничениях

$$\begin{aligned} &2x_1 + 7x_2 + x_3 = 21, \\ &7x_1 + 2x_2 + x_4 = 21, \\ &x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

Проверьте оптимальность и допустимость следующих базисных решений.

a) $\mathbf{X}_Б = \begin{bmatrix} x_2 \\ x_4 \end{bmatrix}, \quad \mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{7} & 0 \\ -\frac{2}{7} & 1 \end{bmatrix}.$

b) $\mathbf{X}_Б = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}$, $\mathbf{B}^{-1} = \begin{bmatrix} 0 & \frac{1}{2} \\ 1 & -\frac{7}{2} \end{bmatrix}$.

c) $\mathbf{X}_Б = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$, $\mathbf{B}^{-1} = \begin{bmatrix} \frac{7}{45} & -\frac{2}{45} \\ -\frac{2}{45} & \frac{7}{45} \end{bmatrix}$.

d) $\mathbf{X}_Б = \begin{bmatrix} x_1 \\ x_4 \end{bmatrix}$, $\mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 \\ -\frac{7}{2} & 1 \end{bmatrix}$.

4. Дана следующая задача ЛП.

$$\text{Минимизировать } z = 2x_1 + x_2$$

при ограничениях

$$\begin{aligned} 3x_1 + x_2 - x_3 &= 3, \\ 4x_1 + 3x_2 - x_4 &= 6, \\ x_1 + 2x_2 + x_5 &= 3, \\ x_1, x_2, x_3, x_4, x_5 &\geq 0. \end{aligned}$$

Вычислите симплекс-таблицу, соответствующую следующему базисному решению, и проверьте его оптимальность и допустимость.

$$\mathbf{X}_Б = \begin{bmatrix} x_1 \\ x_2 \\ x_5 \end{bmatrix}, \quad \mathbf{B}^{-1} = \begin{bmatrix} \frac{3}{5} & -\frac{1}{5} & 0 \\ -\frac{4}{5} & \frac{1}{2} & 0 \\ 1 & -1 & 1 \end{bmatrix}.$$

5. Дана следующая задача ЛП.

$$\text{Максимизировать } z = 5x_1 + 12x_2 + 4x_3$$

при ограничениях

$$\begin{aligned} x_1 + 2x_2 + x_3 + x_4 &= 10, \\ 2x_1 - x_2 + 3x_3 &= 2, \\ x_1, x_2, x_3, x_4 &\geq 0. \end{aligned}$$

a) Найдите наилучшее решение среди следующих базисных допустимых решений.

i) $\mathbf{X}_Б = \begin{bmatrix} x_4 \\ x_3 \end{bmatrix}$, $\mathbf{B}^{-1} = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & \frac{1}{3} \end{bmatrix}$.

ii) $\mathbf{X}_Б = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$, $\mathbf{B}^{-1} = \begin{bmatrix} \frac{2}{5} & -\frac{1}{5} \\ \frac{1}{5} & \frac{2}{5} \end{bmatrix}$.

iii) $\mathbf{X}_Б = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix}$, $\mathbf{B}^{-1} = \begin{bmatrix} \frac{3}{7} & -\frac{1}{7} \\ \frac{1}{7} & \frac{2}{7} \end{bmatrix}$.

b) Присутствует ли среди них оптимальное решение?

6. В следующей таблице представлено оптимальное решение задачи максимизации с тремя ограничениями типа "$\leq$" и неотрицательными переменными x_1 и x_2. Переменные x_3, x_4 и x_5 являются дополнительными (остаточными) переменными, соответствующими ограничениям задачи. Двумя различными способами, используя целевые функции прямой и двойственной задач, найдите оптимальное значение целевой функции исходной задачи.

Базис	x_1	x_2	x_3	x_4	x_5	Решение
z	0	0	0	3	2	?
x_3	0	0	1	1	–1	2
x_2	0	1	0	1	0	6
x_1	1	0	0	–1	1	2

7. Рассмотрите следующую задачу ЛП.

$$\text{Максимизировать } z = 5x_1 + 2x_2 + 3x_3$$

при ограничениях

$$x_1 + 5x_2 + 2x_3 \leq b_1,$$
$$x_1 - 5x_2 - 6x_3 \leq b_2,$$
$$x_1, x_2, x_3 \geq 0.$$

Определите значения констант b_1 и b_2, при которых симплекс-таблица с оптимальным решением имеет следующий вид.

Базис	x_1	x_2	x_3	x_4	x_5	Решение
z	0	a	7	d	e	150
x_1	1	b	2	1	0	30
x_5	0	c	–8	–1	1	10

Константы a, b, c, d и e можно найти на основе данных исходной задачи и условий оптимальности и допустимости решения, представленного в симплекс-таблице.

a) Найдите значения правых частей неравенств исходной задачи, т.е. константы b_1 и b_2.
b) Найдите оптимальное решение двойственной задачи.
c) Найдите значения констант a, b и c, соответствующие переменной x_2.

4.7. Анализ чувствительности оптимального решения

Анализ чувствительности оптимальных решений задач ЛП на элементарном уровне рассмотрен в разделе 2.4. В этом разделе, используя соотношения двойственности и матричное представление симплексных вычислений, мы проведем анализ чувствительности значительно глубже.

Анализ чувствительности выполняется уже после получения оптимального решения задачи ЛП. Его цель — определить, приведет ли изменение коэффициентов исходной задачи к изменению текущего оптимального решения, и если да, то как эффективно найти новое оптимальное решение (если оно существует).

В общем случае изменение коэффициентов исходной задачи может привести к одной из следующих четырех ситуаций.

1. Текущее базисное решение остается неизменным.
2. Текущее решение становится недопустимым.
3. Текущее решение становится неоптимальным.
4. Текущее решение становится неоптимальным и недопустимым.

Во второй ситуации можно использовать двойственный симплекс-метод для восстановления допустимости решения. В третьей ситуации мы используем прямой симплекс-метод для получения нового оптимального решения. В четвертой для получения нового оптимального и допустимого решения следует воспользоваться как прямым, так и двойственным симплекс-методом. Четвертая ситуация, как комбинация второй и третьей, рассмотрена в комплексной задаче 4-3.

Для объяснения различных процедур анализа чувствительности используем модель фабрики игрушек TOYCO из примера 4.4–2. Напомним, что фабрика TOYCO собирает три вида детских игрушек: модели поездов, грузовиков и легковых автомобилей. Сборка модели каждого вида требует последовательного применения трех операций. В задаче необходимо определить объемы производства каждого вида игрушек, максимизирующие общий доход. Для удобства изложения материала повторим формулировки прямой и двойственной задач.

Прямая задача	Двойственная задача
Максимизировать $z = 3x_1 + 2x_2 + 5x_3$ при ограничениях $x_1 + 2x_2 + x_3 \le 430$, (операция 1) $3x_1 + 2x_3 \le 460$, (операция 2)	Минимизировать $w = 430y_1 + 460y_2 + 420y_3$ при ограничениях $y_1 + 3y_2 + y_3 \ge 3$, $2y_1 + 4y_3 \ge 2$,

Прямая задача	Двойственная задача
$x_1 + 4x_2 \le 420$, (операция 3) $x_1, x_2, x_3 \ge 0$.	$y_1 + 2y_2 \ge 5$, $y_1, y_2, y_3 \ge 0$.
Оптимальное решение $x_1 = 0, x_2 = 100, x_3 = 230, z = \$1350$	Оптимальное решение $y_1 = 1, y_2 = 2, y_3 = 0$

Приведем симплекс-таблицу, содержащую оптимальное решение прямой задачи.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	4	0	0	1	2	0	1350
x_2	–1/4	1	0	1/2	–1/4	0	100
x_3	3/2	0	1	0	1/2	0	230
x_6	2	0	0	–2	1	1	20

4.7.1. Изменения, влияющие на допустимость решения

К недопустимости текущего оптимального решения может привести (1) изменение правых частей ограничений (т.е. изменение элементов вектора **b**) и (2) введение в множество ограничений задачи нового ограничения. В любом случае недопустимость решения проявится в том, что, по крайней мере, один элемент в векторе $\mathbf{B}^{-1}\mathbf{b}$ станет отрицательным, т.е. одна или несколько базисных переменных примут отрицательные значения.

Изменение элементов вектора b правых частей ограничений. В следующем примере проиллюстрирован подход к исследованию ситуации, когда изменяется несколько элементов вектора **b**, содержащего значения правых частей ограничений.

Пример 4.7–1

Предположим, что фабрика игрушек TOYCO планирует расширить производство своей продукции путем увеличения возможностей сборочных линий на 40%, что даст следующий фонд рабочего времени для каждого вида сборочной операции: 602, 644 и 588 минут соответственно. Эти изменения влияют только на правые части неравенств ограничений. По формуле $\mathbf{X}_Б = \mathbf{B}^{-1}\mathbf{b}$ найдем новое решение задачи.

$$\begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 602 \\ 644 \\ 588 \end{bmatrix} = \begin{bmatrix} 140 \\ 322 \\ 28 \end{bmatrix}.$$

Таким образом, текущие базисные переменные x_2, x_3 и x_6 с новыми значениями 140, 322 и 28 по прежнему составляют допустимое решение. Соответствующее этому решению оптимальное значение целевой функции (максимальный доход) равно \$1890.

Хотя новое решение и приводит к увеличению дохода фабрики, реализация мероприятий, необходимых для такого наращивания производства, требует определенного времени. Временной альтернативой такой модернизации производства может служить “перенос” неиспользуемого фонда рабочего времени третьей операции ($x_6 = 20$ минут) в фонд первой. Тогда фонд рабочего времени трех сборочных операций будет равен 450, 460 и 400 минут соответственно. С учетом новых ограничений получаем следующее решение.

$$\begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 450 \\ 460 \\ 400 \end{bmatrix} = \begin{bmatrix} 110 \\ 230 \\ -40 \end{bmatrix}.$$

Полученное решение не является допустимым, поскольку теперь $x_6 = -40$. Для возврата в область допустимых решений применим двойственный симплекс-метод. Сначала изменим значения в столбце “Решение” симплекс-таблицы (эти новые значения выделены в следующей симплекс-таблице). Отметим, что соответствующее значение целевой функции равно $z = 3 \times 0 + 2 \times 100 + 5 \times 230 = \$1370$.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	4	0	0	1	2	0	1370
x_2	–1/4	1	0	1/2	–1/4	0	110
x_3	3/2	0	1	0	1/2	0	230
x_6	2	0	0	–2	1	1	–40

В соответствии с двойственным симплекс-методом исключаемой переменной будет x_6, а вводимой — x_4. В результате получим следующую симплекс-таблицу с оптимальным допустимым решением. (В общем случае для получения допустимого решения может потребоваться несколько итераций двойственного симплекс-метода.)

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	5	0	0	0	5/2	1/2	1350
x_2	1/4	1	0	0	0	1/4	100
x_3	3/2	0	1	0	1/2	0	230
x_4	–1	0	0	1	–1/2	–1/2	20

По существу, оптимальное решение осталось неизменным. Это означает, что в данном случае “перенос” части фонда рабочего времени третьей операции в фонд рабочего времени первой операции не приводит к улучшению целевой функции.

Упражнения 4.7,а

1. В модели для фабрики TOYCO 20-минутная часть фонда рабочего времени третьей операции перенесена в фонд рабочего времени второй операции. Улучшит ли это оптимальное решение?

2. Предположим, что фабрика TOYCO планирует изменить фонды рабочего времени сборочных операций следующим образом.

$$\text{a)}\ \begin{bmatrix}460\\500\\400\end{bmatrix},\quad \text{b)}\ \begin{bmatrix}500\\400\\600\end{bmatrix},\quad \text{c)}\ \begin{bmatrix}300\\800\\200\end{bmatrix},\quad \text{d)}\ \begin{bmatrix}450\\700\\350\end{bmatrix}.$$

Воспользуйтесь возможностями анализа чувствительности для нахождения оптимального решения.

3. Вернитесь к модели предприятия Reddy Mikks из примера 2.2–1. Ее симплекс-таблица с оптимальным решением приведена в примере 3.3–1. Используя анализ чувствительности, найдите новое оптимальное решение этой задачи в предположении, что ограничения на сырье M1 и M2 составляют 28 и 8 тонн соответственно.

4. Птицефабрика Ozark содержит 20 000 цыплят, которых выращивают до 8-недельного возраста и затем отправляют на рынок. В следующей таблице представлен недельный расход корма на одного цыпленка, в зависимости от его возраста.

Неделя	1	2	3	4	5	6	7	8
Расход корма (фунты)	0.26	0.48	0.75	1.00	1.30	1.60	1.90	2.10

Для того чтобы цыплята к 8-й неделе могли достичь определенного веса, их рацион должен удовлетворять определенным требованиям к калорийности. Хотя обычно список кормов очень большой, здесь для упрощения мы ограничимся только тремя ингредиентами: известняк, зерно и соевая мука крупного помола. Требования к качественному составу рациона также ограничим только тремя показателями: кальций, белок и клетчатка. В следующей таблице приведены обобщенные данные по их содержанию в кормовых ингредиентах.

	Содержание веществ (фунт/фунт ингредиента)			Стоимость
Ингредиент	Кальций	Белок	Клетчатка	($/фунт)
Известняк	0.380	0.00	0.00	0.12
Зерно	0.001	0.09	0.02	0.45
Соевая мука	0.002	0.50	0.08	1.60

Кормовой рацион должен содержать:

a) кальция — не менее 8% и не более 12%,

b) белка — не менее 22%,

c) клетчатки — не более 5%.

Составьте оптимальный кормовой рацион для каждой недели.

Интервалы допустимых изменений для элементов вектора b. Другой способ исследования влияния изменения доступности ресурсов (т.е. элементов вектора **b** правых частей неравенств ограничений) заключается в определении интервалов допустимости для этих элементов, сохраняющих текущее решение допустимым. Следующий пример иллюстрирует данный метод анализа чувствительности.

Пример 4.7–2

Пусть в задаче о фабрике игрушек TOYCO нас интересует интервал допустимости для значения фонда рабочего времени первой операции. Заменим вектор **b** вектором

$$\mathbf{b}_1 = \begin{bmatrix} 430 + D_1 \\ 460 \\ 420 \end{bmatrix}.$$

Переменная D_1 представляет изменение фонда рабочего времени первой операции по сравнению с текущим уровнем в 430 минут. Для того чтобы текущее базисное решение осталось допустимым, необходимо выполнение неравенства $\mathbf{X}_Б = \mathbf{B}^{-1}\mathbf{b}_1 \geq 0$. Отсюда получаем следующую систему неравенств.

$$\begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 430 + D_1 \\ 460 \\ 420 \end{bmatrix} = \begin{bmatrix} 100 + \frac{D_1}{2} \\ 230 \\ 20 - 2D_1 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Первое неравенство $x_2 \geq 0$ порождает $D_1 \geq -200$, второе неравенство $x_3 \geq 0$ не зависит от D_1, третье $x_6 \geq 0$ дает условие $D_1 \leq 10$. Таким образом, текущее базисное решение останется допустимым при выполнении неравенств $-200 \leq D_1 \leq 10$. Это эквивалентно следующему интервалу допустимости для фонда рабочего времени первой операции.

$$430 - 200 \leq \text{Фонд рабочего времени операции } 1 \leq 430 + 10$$

или

$$230 \leq \text{Фонд рабочего времени операции } 1 \leq 440$$

Изменение значения целевой функции, соответствующее изменению D_1, равно $D_1 y_1$, где y_1 — стоимость (в долларах) одной минуты фонда рабочего времени первой операции (т.е. двойственная цена этого ресурса).

Чтобы проиллюстрировать использование данного интервала допустимости, предположим, что фонд рабочего времени первой операции изменился от 430 до 400 минут. Текущее базисное решение остается допустимым, поскольку новое значение фонда рабочего времени первой операции принадлежит интервалу допустимости. Для вычисления новых значений переменных воспользуемся значением $D_1 = 400 - 430 = -30$. Далее получим следующее.

$$\begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix} = \begin{bmatrix} 100 + \frac{1}{2}(-30) \\ 230 \\ 20 - 2(-30) \end{bmatrix} = \begin{bmatrix} 85 \\ 230 \\ 80 \end{bmatrix}.$$

Для вычисления нового значения целевой функции сначала найдем значения двойственных цен.

$$(y_1, y_2, y_3) = \mathbf{C}_Б \mathbf{B}^{-1} = (2,5,0)\begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} = (1,2,0).$$

Таким образом, стоимость одной минуты фонда рабочего времени первой операции равна $y_1 = \$1$. Тогда изменение оптимального дохода составит $D_1 y_1 = -30 \times 1 = -\$30$. Следует помнить, что данная стоимость минуты фонда рабочего времени первой операции, равная $y_1 = \$1$, справедлива только для указанного выше интервала изменения D_1. Любое изменение, выходящее за этот интервал, приводит к недопустимому решению. В таком случае следует использовать двойственный симплекс-метод для нахождения нового решения, если оно существует.

Аналогичную процедуру можно использовать при определении интервалов допустимости для переменных D_2 и D_3, равных изменению фондов рабочего времени второй и третьей сборочных операций (см. упр. 4.7,b(1)). Определение интервалов допустимости для D_1, D_2 и D_3, как описано выше, и их соотношения с переменными y_1, y_1 и y_3 двойственной задачи корректны только тогда, когда эти ресурсы рассматриваются независимо друг от друга. Далее мы рассмотрим возможность *одновременного* изменения всех трех ресурсов, в этом случае текущий вектор $\mathbf{b}$ необходимо заменить на $\mathbf{b}^*$ с элементами $430 + D_1$, $460 + D_2$ и $420 + D_3$ (упр. 4.7,b(2)).

Упражнения 4.7,b

1. Пусть в задаче о фабрике игрушек TOYCO переменные D_2 и D_3 представляют изменения фондов рабочего времени второй и третьей операций.

   a) Определите интервалы для D_2 и D_3, гарантирующие допустимость текущего решения. Предполагается, что изменения фондов рабочего времени каждой операции выполняются по отдельности.

   b) Определите стоимость одной минуты фондов рабочего времени второй и третьей операций.

   c) Пусть фонд рабочего времени второй операции изменен от текущего значения 460 минут до 500 минут. Найдите новое оптимальное решение и определите соответствующее изменение значения целевой функции.

   d) Пусть фонд рабочего времени третьей операции изменен от текущего значения 420 минут до 450 минут. Найдите новое оптимальное решение и определите соответствующее изменение значения целевой функции.

   e) Пусть фонд рабочего времени третьей операции изменен от текущего значения 420 минут до 380 минут. Найдите новое оптимальное решение и определите соответствующее изменение значения целевой функции.

2. Пусть в задаче о фабрике игрушек TOYCO изменения D_3, D_2 и D_3 фондов рабочего времени всех операций производятся одновременно.

   a) Наложите условия на переменные D_3, D_2 и D_3, гарантирующие допустимость текущего оптимального решения.

   b) Пусть фонды рабочего времени всех трех операций изменены до 438, 500 и 410 минут соответственно. На основании условия, найденного в предыдущем пункте, покажите, что текущее базисное решение останется допустимым. С помощью двойственных цен найдите изменение значения целевой функции.

   c) Пусть фонды рабочего времени всех трех операций изменены до 460, 440 и 380 минут соответственно. На основании условия, найденного в п. а), покажите, что текущее базисное решение будет недопустимым. С помощью двойственного симплекс-метода найдите новое оптимальное решение.

3. Вернитесь к модели фабрики игрушек TOYCO.

   a) Предположим, что стоимость дополнительного времени, выделяемого для первой сборочной операции сверх текущего фонда времени в 430 минут, равна \$50 за час. В эту стоимость входит оплата сверхурочных работ персонала и стоимость машинного времени. Будет ли экономически целесообразным использовать дополнительное время для первой операции?

   b) Пусть на второй сборочной операции оператор может ежедневно работать два часа сверхурочно с оплатой \$45 за каждый час. Стоимость дополнительного машинного времени составляет \$10 за час. Будет ли экономически целесообразным использовать дополнительное время для второй операции?

   c) На каких условиях экономически целесообразно использовать дополнительное время для третьей операции?

   d) Предположим, что фонд рабочего времени первой операции увеличен до 440 минут, но любое превышение текущего фонда этой операции (430 минут) стоит \$40 за час. Найдите новое оптимальное решение, включая значение целевой функции.

   e) Предположим, что фонд рабочего времени второй операции уменьшен на 15 минут. Стоимость одного часа этой операции равна \$30 (в течение обычной рабочей смены). Будет ли экономически целесообразным уменьшение фонда рабочего времени для второй операции?

4. Компания производит бумажники, кошельки и небольшие рюкзаки. Конструкция всех трех видов изделий предусматривает использование кожи и синтетических материалов, причем кожа является дефицитным материалом. В производственном процессе используется два вида ручных работ: прошивка и зачистка. В следующей таблице приведены данные, характеризующие производственный процесс, потребность в ресурсах и доход на единицу производимого изделия.

	Ресурсы, необходимые для изготовления одного изделия			Ежедневный лимит
Ресурс	Бумажник	Кошелек	Рюкзак	ресурса
Кожа (кв. футы)	2	1	3	42
Прошивка (часы)	2	1	2	40
Зачистка (часы)	1	0.5	1	45
Отпускная цена (\$)	24	22	45	

Сформулируйте задачу линейного программирования и найдите ее оптимальное решение с помощью программы TORA. Для приведенных ниже изменений в предельных значениях доступных ресурсов определите, какие из них сохраняют допустимость текущего решения. В случае сохранения допустимости решения найдите новое оптимальное решение (т.е. значения переменных задачи и значение целевой функции).

a) Ежедневный лимит кожи возрос до 45 кв. футов.
b) Ежедневный лимит кожи уменьшился на 1 кв. фут.
c) Фонд рабочего времени операции прошивки изменился до 38 часов.
d) Фонд рабочего времени операции прошивки изменился до 46 часов.
e) Фонд рабочего времени операции зачистки уменьшился до 15 часов.
f) Фонд рабочего времени операции зачистки увеличился до 50 часов.
g) Следует ли рекомендовать компании набор временных рабочих на операцию прошивки с оплатой $15 в час?

5. Компания производит две модели электронных устройств, при изготовлении которых используются резисторы, конденсаторы и микросхемы. В следующей таблице приведены данные, характеризующие производство этих моделей.

Ресурс	Количество комплектующих на одно изделие		Лимит комплектующих (шт.)
	Модель 1	Модель 2	
Резистор	2	3	1200
Конденсатор	2	1	1000
Микросхема	0	4	800
Доход на одно изделие ($)	3	4	

Обозначим через x_1 и x_2 количество производимых устройств моделей 1 и 2 соответственно. Ниже приведена сформулированная задача ЛП и соответствующая симплекс-таблица с ее оптимальным решением.

$$\text{Максимизировать } z = 3x_1 + 4x_2$$

при ограничениях

$$2x_1 + 3x_2 \le 1200, \text{ (ограничение на резисторы)}$$

$$2x_1 + x_2 \le 1000, \text{ (ограничение на конденсаторы)}$$

$$4x_2 \le 800, \text{ (ограничение на микросхемы)}$$

$$x_1, x_2 \ge 0.$$

Базис	x_1	x_2	s_1	s_2	s_3	Решение
z	0	0	5/4	1/4	0	1750
x_1	1	0	–1/4	3/4	0	450
s_3	0	0	–2	2	1	400
x_2	0	1	1/2	–1/2	0	100

a) Определите статус каждого ресурса (комплектующего).

b) В терминах оптимального дохода определите стоимость одного резистора, одного конденсатора и одной микросхемы.

c) Найдите интервал применимости двойственных цен для каждого ресурса.

d) Найдите новое оптимальное решение при условии возрастании числа доступных резисторов до 1300.

e) Если количество доступных микросхем будет уменьшено до 350, можно ли будет найти новое оптимальное решение непосредственно из приведенной выше информации? Обоснуйте свой ответ.

f) В п. с) был определен интервал допустимости для доступного количества используемых конденсаторов. На основе этих данных определите соответствующий интервал изменения оптимального дохода и соответствующие интервалы изменения количества производимых изделий первой и второй моделей.

g) Новый контракт позволяет компании закупить дополнительное число резисторов по 40 центов за единицу, но только при условии, что закупочная партия составит не менее 500 единиц. Выгоден ли компании такой контракт?

6. Компания для производства двух видов продукции имеет ежедневный фонд рабочего времени 320 часов и 350 единиц расходных материалов (сырья). При необходимости компания может позволить 10 часов сверхурочной работы с оплатой \$2 за час. На изготовление одной единицы продукции первого вида требуется 1 час рабочего времени и 3 единицы сырья, а на изготовление одной единицы продукции второго вида — 2 часа рабочего времени и 1 единица сырья. Доход от одной единицы этих продукций составляет соответственно \$10 и \$12. Обозначим через x_1 и x_2 ежедневные объемы производства продукции первого и второго видов, а через x_3 — количество используемых сверхурочных часов. Ниже приведена сформулированная задача ЛП и соответствующая симплекс-таблица с оптимальным решением.

$$\text{Максимизировать } z = 10x_1 + 12x_2 - 2x_3$$

при ограничениях

$x_1 + 2x_2 - x_3 \leq 320$, (ограничение на фонд рабочего времени)

$3x_1 + x_2 \leq 350$, (ограничение на сырье)

$x_3 \leq 10$, (ограничение на сверхурочные работы)

$x_1, x_2, x_3 \geq 0$.

Базис	x_1	x_2	x_3	s_1	s_2	s_3	Решение
z	0	0	0	26/5	8/5	16/5	2256
x_2	0	1	0	3/5	–1/5	3/5	128
x_1	1	0	0	–1/5	2/5	–1/5	74
x_3	0	0	1	0	0	1	10

a) Найдите оптимальное решение этой задачи.

b) Определите двойственные цены ресурсов и их интервалы допустимости.

c) Найдите двойственные цены для фонда рабочего времени и сверхурочных работ. Могут ли эти цены быть одинаковыми? Обоснуйте.

d) Компания может увеличить объем сверхурочных работ за дополнительную плату \$2 за час. Сколько часов такой сверхурочной работы может ввести компания?

e) Компания ежедневно может получать дополнительный объем сырья в 100 единиц по цене \$1.50. Стоит ли компании использовать этот резерв сырья? А если стоимость дополнительного сырья будет \$2 за единицу.

f) Предположим, что компания вынуждена сократить складские площади для сырья и поэтому ежедневно не может использовать более 200 единиц сырья. Найдите для этого случая новое оптимальное решение.

g) Предположим, что компания не может ежедневно использовать более 8 часов сверхурочной работы. Найдите новое оптимальное решение.

7. *Достаточное правило допустимости.* Это упрощенное правило можно использовать для проверки того, что *одновременные* изменения D_1, D_2, ..., D_m элементов вектора **b** (правых частей неравенств ограничений) сохранят допустимость текущего решения. Предположим, что правая часть b_i i-го ограничения была изменена на $b_i + D_i$, причем *независимо* от изменения правых частей других ограничений, и соответствующий интервал допустимости $p_i \leq D_i \leq q_i$ рассчитан так, как показано в примере 4.7–2. Очевидно, что $p_i \leq 0$ ($q_i \geq 0$), поскольку величина p_i (q_i) соответствует максимальному уменьшению (возрастанию) значения b_i. Положим r_i равным или отношению D_i/p_i, или D_i/q_i, в зависимости от того, будет ли величина D_i отрицательной или положительной. По определению $0 \leq r_i \leq 1$. Достаточное правило допустимости гласит, что для данных изменений D_1, D_2, ..., D_m *достаточным* (не необходимым) условием того, что текущее решение останется допустимым, будет выполнение неравенства $r_1 + r_2 + ... + r_m \leq 1$. Если это условие не выполняется, то текущее решение может быть как допустимым, так и недопустимым. Сформулированное правило неприменимо, если D_i выходят из своих интервалов допустимости.

   В действительности достаточное правило допустимости является очень слабым критерием допустимости решения и на практике применяется редко. Даже в том случае, когда допустимость решения может быть подтверждена с помощью этого правила, все равно для получения нового оптимального решения будет использовано условие допустимости прямого симплекс-метода (как в упр. 2).

   Примените данное правило к задачам b) и c) из упр. 2. В задаче b) достаточное правило допустимости не может подтвердить допустимость решения, а в задаче c) оно не применимо. Следующее упражнение должно подтвердить наши утверждения относительно этого правила.

8. Дана следующая задача ЛП.

$$\text{Максимизировать } z = x_1 + x_2$$

при ограничениях

$$2x_1 + x_2 \leq 6,$$
$$x_1 + 2x_2 \leq 6,$$
$$x_1, x_2 \geq 0.$$

a) Покажите, что оптимальное базисное решение содержит обе переменные x_1 и x_2 и что интервалы допустимости для правых частей ограничений, полученные в предположении их независимости, имеют вид $-3 \leq D_1 \leq 6$ и $-3 \leq D_2 \leq 6$.

b) Предположим, что правые части ограничений одновременно увеличиваются на величину $\Delta > 0$. Сначала докажите, что базисное решение остается допустимым для всех $\Delta > 0$. Далее покажите, что достаточное правило допустимости дает правильный ответ только тогда, когда $0 < \Delta < 3$, не дает ответа при $3 \leq \Delta \leq 6$, и не применимо — когда $\Delta > 6$.

9. Покажите, что достаточное правило допустимости является следствием неравенства $\mathbf{B}^{-1}\mathbf{b} \geq 0$.

Добавление новых ограничений. Добавление нового ограничения в существующую модель ЛП может привести к одной из следующих ситуаций.

1. Новое ограничение является *избыточным*. Это означает, что новое ограничение выполняется при текущем оптимальном решении.
2. Новое ограничение не выполняется при текущем оптимальном решении. В этом случае необходимо применить двойственный симплекс-метод, чтобы получить (или хотя бы попытаться получить) новое оптимальное решение.

Отметим, что добавление неизбыточного нового ограничения может только ухудшить текущее оптимальное значение целевой функции.

Пример 4.7–3

Предположим, что фабрика игрушек TOYCO изменила конструкцию выпускаемых моделей, и теперь для их производства необходима четвертая сборочная операция. Ежедневный фонд рабочего времени этой операции составляет 500 минут. Время выполнения этой операции при сборке одной игрушки различных видов составляет соответственно 3, 1 и 1 минуту. В результате получаем новое ограничение: $3x_1 + x_2 + x_3 \leq 500$. Это ограничение является избыточным, поскольку оно удовлетворяется при текущем оптимальном решении $x_1 = 0$, $x_2 = 100$ и $x_3 = 230$. Таким образом, текущее оптимальное решение остается неизменным.

Пример 4.7–4

Предположим, что в модели фабрики игрушек TOYCO время выполнения новой четвертой операции составляет соответственно 3, 3 и 1 минуту при сборке одной игрушки различных видов. В этом случае четвертое ограничение $3x_1 + 3x_2 + x_3 \leq 500$ не будет избыточным, и текущее оптимальное решение ему не удовлетворяет. Мы должны ввести новое ограничение в симплекс-таблицу, где представлено текущее оптимальное решение.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Решение
z	4	0	0	1	2	0	0	1350
x_2	–1/4	1	0	1/2	–1/4	0	0	100
x_3	3/2	0	1	0	1/2	0	0	230
x_6	2	0	0	–2	1	1	0	20
x_7	3	3	1	0	0	0	1	500

Поскольку переменные x_2 и x_3 являются базисными, из x_7-строки следует исключить соответствующие им коэффициенты (т.е. надо сделать их нулевыми). Для этого надо выполнить следующую операцию.

Новая x_7-строка = Старая x_7-строка – [3 × (x_2-строка) + 1 × (x_3-строка)]

В результате получим новую симплекс-таблицу.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	x_7	Решение
z	4	0	0	1	2	0	0	1350
x_2	–1/4	1	0	1/2	–1/4	0	0	100
x_3	3/2	0	1	0	1/2	0	0	230
x_6	2	0	0	–2	1	1	0	20
x_7	9/4	0	0	–3/2	1/4	0	1	–30

С помощью двойственного симплекс-метода находим новое оптимальное решение $x_1 = 0$, $x_2 = 90$, $x_3 = 230$ и $z = \$1330$ (проверьте!).

Упражнения 4.7,с

1. Пусть в модели фабрики игрушек TOYCO время выполнения четвертой операции составляет соответственно 4, 1 и 2 минуты при сборке одной игрушки различных видов. Найдите оптимальное решение задачи в предположении, что фонд рабочего времени четвертой операции составляет (a) 570 минут, (b) 548 минут.

2. *Вторичные ограничения*. Вместо решения задачи ЛП, с учетом всех ограничений, можно сначала определить так называемые *вторичные ограничения* и на первом этапе решения задачи исключить их из рассмотрения. Вторичными ограничениями являются те, которые, как мы подозреваем, лишь в малой степени влияют (или совсем не влияют) на оптимальное решение. После определения такие ограничения исключаются из множества ограничений задачи, и далее решается задача только с оставшимися ограничениями. Затем по очереди проверяются вторичные ограничения. Если полученное ранее оптимальное решение удовлетворяет вторичному ограничению, такое ограничение отбрасывается совсем. В противном случае оно вводится в множество ограничений задачи и продолжается поиск нового оптимального решения. Этот процесс выполняется до тех пор, пока не исчерпаются все вторичные ограничения.

   Примените описанную процедуру к следующей задаче ЛП.

   $$\text{Максимизировать } z = 5x_1 + 6x_2 + 3x_3$$

   при ограничениях

   $$5x_1 + 5x_2 + 3x_3 \le 50,$$
   $$x_1 + x_2 - x_3 \le 20,$$
   $$7x_1 + 6x_2 - 9x_3 \le 90,$$

$$5x_1 + 5x_2 + 5x_3 \leq 35,$$
$$12x_1 + 6x_2 \leq 90,$$
$$x_2 - 9x_3 \leq 20,$$
$$x_1, x_2, x_3 \geq 0.$$

4.7.2. Изменения, влияющие на оптимальность решения

Текущее оптимальное решение перестает быть оптимальным, если разности $z_j - c_j$ не удовлетворяют условию оптимальности. Используя вектор двойственных цен $\mathbf{Y} = \mathbf{C}_Б\mathbf{B}^{-1}$, определенный в разделе 4.6, запишем

$$z_j - c_j = \mathbf{Y}\mathbf{P}_j - c_j.$$

Отсюда следует, что на оптимальность решения влияют только коэффициенты c_j целевой функции (и, следовательно, вектор $\mathbf{C}_Б$) и/или стоимости ресурсов, представленные векторами $\mathbf{P}_j$. Рассмотрим последовательно каждый фактор, влияющий на оптимальность решения.

Изменение коэффициентов целевой функции. Для определения влияния изменений коэффициентов целевой функции следует пересчитать разности $z_j - c_j$ только для *небазисных* переменных, поскольку при любых изменениях коэффициентов c_j, соответствующих *базисным* переменным, разности $z_j - c_j$ всегда остаются равными нулю (доказательство этого утверждения см. в разделе 7.4).

Вычислительная процедура заключается в следующем.

1. Вычисляется вектор двойственных цен $\mathbf{Y} = \mathbf{C}_Б\mathbf{B}^{-1}$ для нового вектора коэффициентов $\mathbf{C}_Б$.
2. Вычисляются разности $z_j - c_j = \mathbf{Y}\mathbf{P}_j - c_j$ для текущей *небазисной* переменной x_j. При этом возможны два варианта.

   a) Если условие оптимальности выполняется, текущее решение остается оптимальным, но значение целевой функции может измениться.

   b) Если условие оптимальности не выполняется, следует применить (прямой) симплекс-метод для получения нового оптимального решения.

Пример 4.7–5

Предположим, что фабрика игрушек TOYCO проводит новую ценовую политику относительно своих изделий. В соответствии с этим доход от одной модели поезда, грузовика и легкового автомобиля составляет соответственно \$4, \$3 и \$4. Получаем новую целевую функцию для этой модели

$$\text{Максимизировать } z = 4x_1 + 3x_2 + 4x_3.$$

Поскольку текущее базисное решение $\mathbf{X}_Б$ состоит из переменных x_2, x_3 и x_6, имеем $\mathbf{C}_Б = (3, 4, 0)$. Вычислим вектор двойственных цен.

$$\mathbf{Y} = \mathbf{C}_Б\mathbf{B}^{-1} = (3, 4, 0)\begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} = \left(\frac{3}{2}, \frac{5}{4}, 0\right).$$

Разности $z_j - c_j$ для небазисных переменных x_1, x_4 и x_5 вычисляются по формуле $z_j - c_j = \mathbf{Y}\mathbf{P}_j - c_j$:

$$z_1 - c_1 = y_1 + 3y_2 + y_3 - 4 = \frac{3}{2} + 3\left(\frac{5}{4}\right) + 0 - 4 = \frac{5}{4},$$

$$z_4 - c_4 = y_1 - 0 = \frac{3}{2},$$

$$z_5 - c_5 = y_2 - 0 = \frac{5}{4}.$$

Отметим, что здесь использовалось новое значение коэффициента целевой функции $c_1 = 4$.

Вычисления показывают, что текущее решение $x_1 = 0$, $x_2 = 100$ и $x_3 = 230$ остается оптимальным. Новое значение целевой функции равно $4 \times 0 + 3 \times 100 + 4 \times 230 = \$1220$.

Предположим, что в рассматриваемой задаче целевая функция имеет следующий вид.

Максимизировать $z = 6x_1 + 3x_2 + 4x_3$.

Эта функция совпадает с предыдущей целевой функцией, за исключением того, что коэффициент при переменной x_1 теперь равен 6. Поэтому необходимо пересчитать только разность $z_1 - c_1$. В результате получаем следующее.

$$z_1 - c_1 = y_1 + 3y_2 + y_3 - 6 = \frac{3}{2} + 3\left(\frac{5}{4}\right) + 0 - 6 = -\frac{3}{4}.$$

Отсюда следует, что переменную x_1 необходимо включить в базисное решение. Имеем следующую симплекс-таблицу.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	–3/4	0	0	3/2	5/4	0	1220
x_2	–1/4	1	0	1/2	–1/4	0	100
x_3	3/2	0	1	0	1/2	0	230
x_6	2	0	0	–2	1	1	20

Новые значения разностей $z_j - c_j$ для небазисных переменных x_1, x_4 и x_5 в симплекс-таблице выделены. Все остальные элементы таблицы остались такими же, как и в исходной таблице с оптимальным решением. Для нахождения нового оптимального решения следует ввести в базис переменную x_1 и исключить из него переменную x_6. В результате получим решение $x_1 = 10$, $x_2 = 102.5$, $x_3 = 215$ и $z = \$1227.50$ (проверьте!).

Кроме того, для исследования влияния коэффициентов целевой функции на оптимальность решения можно также вычислить (по отдельности) интервалы изменения каждого коэффициента, сохраняющие оптимальность текущего решения. Для этого следует заменить текущий коэффициент c_j выражением $c_j + d_j$, где d_j — величина (положительная или отрицательная) изменения коэффициента c_j.

Ограничения на величины d_j можно определить путем вычисления новых разностей $z_j - c_j$ и наложения на них соответствующего условия оптимальности, которое зависит от того, рассматривается ли задача минимизации или максимизации (см. упр. 4.7,d(3,7)).

Упражнения 4.7,d

1. Проверьте оптимальность решения задачи о фабрике TOYCO для следующих целевых функций. Если решение неоптимально, найдите новое оптимальное решение. (Симплекс-таблица с оптимальным решением для данной задачи представлена в начале раздела 4.7.)

   a) $z = 2x_1 + x_2 + 4x_3$
   b) $z = 3x_1 + 6x_2 + x_3$
   c) $z = 8x_1 + 3x_2 + 9x_3$

2. Проверьте оптимальность решения задачи о компании Reddy Mikks (пример 4.4–1) для следующих целевых функций. Если решение неоптимально, найдите новое оптимальное решение. (Симплекс-таблица с оптимальным решением для данной задачи представлена в примере 3.3–1.)

   a) $z = 3x_1 + 2x_2$
   b) $z = 8x_1 + 10x_2$
   c) $z = 2x_1 + 10x_2$

3. Пусть в задаче о фабрике TOYCO коэффициенты целевой функции претерпели следующие изменения (каждое изменение рассматривается как отдельная задача).

   a) Доход от одной модели поезда составляет $3 + d_1$ долларов.
   b) Доход от одной модели грузовика — $2 + d_2$ долларов.
   c) Доход от одной модели легкового автомобиля — $5 + d_3$ долларов.

   Изменения d_1, d_2 и d_3 могут быть как положительные, так и отрицательные. Применяя подходящее условие оптимальности к разностям $z_j - c_j$, определите интервалы для величин d_1, d_2 и d_3, сохраняющие текущее оптимальное решение.

4. В задаче о фабрике TOYCO, используя решение из предыдущего упражнения, укажите, будет ли текущее решение оптимальным для следующих (независимых) ситуаций. Если решение изменится, найдите новое.

   a) Доход от одной модели поезда возрос от $3 до $5. До $8.
   b) Доход от одной модели поезда уменьшился от $3 до $2.
   c) Доход от одной модели грузовика увеличился от $2 до $6.
   d) Доход от одной модели легкового автомобиля уменьшился от $5 до $2.

5. Пусть в задаче о компании Reddy Mikks коэффициенты целевой функции претерпели следующие изменения (каждое изменение рассматривается как отдельная задача).

   a) Доход от одной тонны краски для наружных работ составляет $5 + d_1$ тысяч долларов.
   b) Доход от одной тонны краски для внутренних работ — $4 + d_2$ тысяч долларов.

   Изменения d_1 и d_2 могут быть как положительные, так и отрицательные. Применяя подходящее условие оптимальности к разностям $z_j - c_j$, определите интервалы для величин d_1 и d_2, сохраняющие текущее оптимальное решение.

6. В задаче о компании Reddy Mikks, используя решение из предыдущего упражнения, покажите, будет ли текущее решение оптимальным для следующих (независимых) ситуаций. Если решение изменится, найдите новое.

   a) Доход от одной тонны краски для наружных работ возрос от \$5000 до \$7000. Уменьшился от \$5000 до \$4000.
   b) Доход от одной тонны краски для внутренних работ возрос от \$4000 до \$6000. Уменьшился от \$4000 до \$3000.

7. Пусть в упр. 3 изменения d_1, d_2 и d_3 производятся *одновременно*.

   a) Найдите условия, сохраняющие текущее решение оптимальным.
   b) Используя условия, полученные в предыдущем пункте, найдите новое решение (если текущее изменилось) для следующих ситуаций.

      i) $z = 2x_1 + x_2 + 4x_3$
      ii) $z = 3x_1 + 6x_2 + x_3$
      iii) $z = 8x_1 + 3x_2 + 9x_3$

8. Пусть в упр. 5 изменения d_1 и d_2 производятся *одновременно*.

   a) Найдите условия, сохраняющие текущее решение оптимальным.
   b) Используя условия, полученные в предыдущем пункте, найдите новое решение (если текущее изменилось) для следующих ситуаций.

      i) $z = 3x_1 + 2x_2$
      ii) $z = 3x_1 + 9x_2$
      iii) $z = 5x_1 + 5x_2$

9. В задаче из упр. 4.7,b(4) найдите оптимальное решение для следующих целевых функций.

   a) $z = 40x_1 + 22x_2 + 45x_3$
   b) $z = 70x_1 + 22x_2 + 45x_3$
   c) $z = 24x_1 + 10x_2 + 45x_3$
   d) $z = 24x_1 + 20x_2 + 45x_3$
   e) $z = 24x_1 + 22x_2 + 50x_3$
   f) $z = 24x_1 + 22x_2 + 40x_3$

10. Вернитесь к задаче из упр. 4.7,b(5).

a) Найдите интервал значений удельного дохода от первой модели выпускаемых устройств, сохраняющих оптимальность текущего решения.
b) Найдите интервал значений удельного дохода от второй модели выпускаемых устройств, сохраняющих оптимальность текущего решения.
c) Найдите новое оптимальное решение в случае возрастания удельного дохода от первой модели до 6 долларов.
d) Найдите новое оптимальное решение при изменении удельного дохода от второй модели до 1 доллара.
e) Определите условия, сохраняющие текущее оптимальное решение при одновременном изменении удельных доходов от обеих моделей устройств.
f) Найдите новое оптимальное решение, если целевая функция примет вид $z = 5x_1 + 2x_2$.

11. Вернитесь к задаче из упр. 4.7,b(6).

a) Каков наименьший удельный доход от первого продукта, сохраняющий текущее оптимальное решение?
b) Найдите новое оптимальное решение при возрастании удельного дохода от первого продукта до $25.

12. *Достаточное правило оптимальности.* Правило, подобное *достаточному правилу допустимости* из упр. 4.7,b(7), можно сформулировать и для проверки оптимальности текущего решения при одновременном изменении всех коэффициентов c_j целевой функции. Для этого представим коэффициенты c_j в виде $c_j + d_j$, $j = 1, 2, ..., n$. Предположим, что для всех изменений d_j независимо получены индивидуальные интервалы $u_j \leq d_j \leq v_j$ значений, сохраняющих оптимальность текущего решения (как в упр. 3). Очевидно, что $u_j \leq 0$ ($v_j \geq 0$), поскольку эта величина соответствует максимально возможному уменьшению (увеличению) коэффициента c_j, сохраняющего текущее оптимальное решение. Для d_j, которые находятся в интервале $u_j \leq d_j \leq v_j$, определим отношение $r_j = d_j/v_j$ или $r_j = d_j/u_j$, в зависимости от того, будет величина d_j положительной или отрицательной. По определению $0 \leq r_j \leq 1$. Достаточное правило оптимальности гласит, что достаточным (но не необходимым) условием сохранения оптимальности текущего решения является выполнение неравенства $r_1 + r_2 + ... + r_n \leq 1$. Если это неравенство не выполняется, то текущее решение может быть как оптимальным, так и неоптимальным. Это правило не применимо, если величины d_j выходят за свои интервалы оптимальности.

Примените достаточное правило оптимальности к задачам из упр. 1 для определения тех измененных целевых функций, которые сохраняют текущее оптимальное решение. Сравнение результатов применения достаточного правила оптимальности и условия оптимальности, используемого при решении задач из упр. 1, показывает, что достаточное правило оптимальности слишком слабое для того, чтобы использовать его в качестве инструмента принятия решений.

13. Покажите, что достаточное правило оптимальности (упр. 12) является следствием неравенств $z_j - c_j \geq 0$ в задаче максимизации и неравенств $z_j - c_j \leq 0$ — в задаче минимизации.

Добавление в модель ЛП нового вида производственной деятельности. Введение в модель линейного программирования нового вида производственной деятельности эквивалентно добавлению новой переменной в задачу ЛП. Добавление нового вида производственной деятельности интуитивно обосновано только в том случае, если эта деятельность экономически рентабельна, т.е. улучшает оптимальное значение целевой функции. Это условие можно проверить путем вычисления для новой переменной разности $z_j - c_j = \mathbf{Y}\mathbf{P}_j - c_j$, где $\mathbf{Y}$ — вектор оптимальных значений двойственной задачи, $\mathbf{P}_j$ и c_j — соответственно ресурсы, используемые для обеспечения нового вида деятельности, и доход от единицы "выхода" этой деятельности. Если вычисленное значение разности $z_j - c_j$ удовлетворяет условию оптимальности, то новая деятельность нежелательна, поскольку не улучшает оптимального решения. Если же вычисленное значение разности $z_j - c_j$ не удовлетворяет условию оптимальности, то новый вид деятельности является рентабельным и соответствующая ему переменная должна быть включена в базисное решение.

Пример 4.7–6

Оптимальное решение задачи ЛП о фабрике игрушек TOYCO показывает, что производство моделей поездов нерентабельно. Поэтому фабрика планирует заменить производство этих моделей выпуском новых игрушек, а именно моделью пожарной машины, причем ее сборка будет осуществляться с использованием тех же производственных мощностей. Фабрика подсчитала доход от новой игрушки в $4 за одну модель. Ее время сборки на каждой из трех технологических операций составляет соответственно 1, 1 и 2 минуты.

Обозначим через x_7 объем производства новой продукции. Поскольку в этой ситуации текущий базисный вектор $\mathbf{C}_Б$ не изменился, можно для дальнейших расчетов использовать текущий вектор значений переменных двойственной задачи $\mathbf{Y} = (y_1, y_2, y_3) = (1, 2, 0)$. Вычисляем разность $z_7 - c_7$.

$$z_7 - c_7 = 1y_1 + 1y_2 + 2y_3 - 4 = 1 \times 1 + 1 \times 2 + 2 \times 0 - 4 = -1$$

Полученный результат показывает, что экономически целесообразно включить переменную x_7 в оптимальное базисное решение. Для нахождения нового оптимального решения сначала вычисляем

$$\mathbf{B}^{-1}\mathbf{P}_7 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{2} \\ 1 \end{bmatrix}.$$

Отсюда следует, что текущая симплекс-таблица должна быть приведена к следующему виду.

Базис	x_1	x_2	x_3	x_7	x_4	x_5	x_6	Решение
z	4	0	0	–1	1	2	0	1350
x_2	–1/4	1	0	1/4	1/2	–1/4	0	100
x_3	3/2	0	1	1/2	0	1/2	0	230
x_6	2	0	0	1	–2	1	1	20

Теперь новое оптимальное решение можно найти путем введения в базис переменной x_7 и исключения из него переменной x_6.

Введение в модель ЛП нового вида деятельности, как видно из приведенного выше, можно рассматривать как обобщение ситуации, когда происходит изменение в векторе ресурсов $\mathbf{P}_j$, используемых для *существующей* деятельности. Поэтому изменение параметров существующего вида деятельности отдельно мы не рассматриваем.

Упражнения 4.7,е

1. В исходной модели фабрики игрушек TOYCO производство моделей поездов не входит в оптимальный производственный план. Ситуация на рынке игрушек не позволяет увеличить отпускную цену этих моделей. Поэтому фабрика решила усовершенствовать сборочные операции данной модели. Это привело к уменьшению времени выполнения каждой из трех сборочных операций на p%. Найдите значение p, при котором выпуск моделей поездов становится рентабельным. (Симплекс-таблица с оптимальным решением данной задачи приведена в начале раздела 4.7.)
2. Предположим, что в исходной модели фабрики игрушек TOYCO время выполнения трех сборочных операций при производстве моделей поездов уменьшено соответственно до 0.5, 1 и 0.5 минут. Доход от модели этого вида остался неизменным на уровне $3 за одну игрушку. Найдите новое оптимальное решение.
3. Пусть в модели фабрики игрушек TOYCO производство модели нового вида (модель пожарной машины) требует соответственно 1, 2 и 3 минуты для выполнения каждой сборочной операции. Найдите оптимальное решение, если доход от одной модели нового вида составляет (a) $5, (b) $10.
4. Вернитесь к модели компании Reddy Mikks (модель представлена в примере 4.4–1, симплекс-таблица с ее оптимальным решением — в примере 3.3–1). Предположим, что компания рассматривает возможность производства дешевой краски для наружных работ, причем для производства тонны такой краски требуется 0.75 тонны сырья M1 и столько же сырья M2. Ситуация на рынке показывает, что производство краски для внутренних работ не должно превышать ежедневного производства *обоих* видов красок для наружных работ более чем на одну тонну. Доход от одной тонны новой краски составляет $3500. Найдите новое оптимальное решение.

4.8. Заключение

В этой главе введено понятие двойственности, описана двойственная задача ЛП и дана ее экономическая интерпретация. Здесь также углубленно рассмотрен анализ чувствительности модели ЛП при изменении ее параметров. Приведены эффективные методы получения нового оптимального решения (если оно существует) при изменении условий исходной задачи ЛП.

Литература

1. Bradley S., Hax A., Magnanti T. *Applied Mathematical Programming*, Addison-Wesley, Reading, Mass, 1977.

2. Bazaraa M., Jarvis J., Sherali M. *Linear Programming and Network Flows*, 2nd ed., Wiley, New York, 1990.
3. Nering E., Tucker A. *Linear Programming and Related Problems*, Academic Press, Boston, 1992.

Литература, добавленная при переводе

Ашманов С.А. *Линейное программирование.* — М.: Наука, 1981.

Гольштейн Е.Г. *Теория двойственности в математическом программировании и ее приложения.* — М.: Наука, 1971.

Гольштейн Е.Г., Юдин Д.Б. *Линейное программирование: Теория, методы и приложения.* — М.: Наука, 1969.

Комплексные задачи

■ **4-1.**[1] Компания MANCO производит три вида продукции: P1, P2 и P3. В производственном процессе используются материалы M1 и M2, обрабатываемые на станках C1 и C2. В следующей таблице приведены данные, характеризующие производственный процесс изготовляемой продукции.

Ресурсы	Единицы измерения	К-во ресурсов на единицу изделия			Ежедневный фонд ресурсов
		P1	P2	P3	
Время работы станка C1	Минуты	1	2	1	430
Время работы станка C2	Минуты	3	0	2	460
Материал M1	Фунты	1	4	0	420
Материал M2	Фунты	1	1	1	300

Ежедневный объем производства изделия P2 должен быть не менее 70 единиц, а изделия P3 — не более 240 единиц. Доход на единицу изделия P1, P2 и P3 составляет соответственно 300, 200 и 500 долларов.

Руководство компании разрабатывает стратегию для улучшения своего финансового положения. Имеются следующие предложения.

1. Можно увеличить на 20% доход от изделия P3, но при этом уменьшится объем его производства до 210 единиц.
2. Материал M2 является критическим фактором, ограничивающим текущее производство. Можно приобрести дополнительные объемы этого материала у сторонних поставщиков, но его цена за фунт будет на $3 выше, чем у поставщиков, которые обслуживают компанию сегодня.
3. Фонд рабочего времени станков можно увеличить на 40 минут в рабочий день, однако такое увеличение приведет к дополнительной стоимости эксплуатации каждого станка — $35 в день.

[1] Задача основана на материалах статьи D. Sheran. "Post-Optimal Analysis in Linear Programming — The Right Example", *IIE Transactions*, Vol. 16, No. 1, March 1984, pp. 99–102.

4. Отдел маркетинга обосновал необходимость увеличения минимального объема производства продукта P2 с 70 единиц до 100 единиц.
5. Время обработки единицы изделия P1 на станке C2 можно уменьшить до 2 минут с дополнительной стоимостью $4 в рабочий день.

Рассмотрите целесообразность внедрения этих предложений, учитывая, что некоторые из них можно внедрить одновременно.

■ **4-2.** Компания Reddy Mikks планирует в будущем расширить свое производство. Изучение ситуации на рынке красок показало, что компания может увеличить объем продаж на 25%. План развития производства можно разработать на основе следующих предложений. (Обратитесь к примеру 3.3–1 за детальной информацией о модели ЛП для этой компании и ее решении.)

Предложение 1. Поскольку рост продаж на 25% приведет к увеличению дохода примерно на $5250, стоимость дополнительных объемов сырья M1 и M2 составляет $750 и $500 за тонну; следовательно, для обеспечения роста объема производства потребуется $5250/(($750 + $500)/2) = 8.4 тонны сырья M1 и столько же сырья M2.

Предложение 2. Потребление сырья M1 и M2 должно возрасти на 6 и 1.5 тонны соответственно, так как эти величины соответствуют 25% текущего уровня потребления сырья (равного 24 тоннам для сырья M1 и 6 тоннам для сырья M2). Поскольку в текущем оптимальном решении оба этих ресурса дефицитны, увеличение их потребления на 25% должно привести к такому же увеличению производства краски, т.е. конечного продукта.

Какие заключения вы можете сделать относительно этих предложений? Предложите несколько подходов к решению данной проблемы.

■ **4-3.** *Анализ чувствительности одновременно на допустимость и оптимальность решения задачи ЛП.* Предположим, что в модель компании Reddy Mikks одновременно внесены следующие изменения.

1. Доход от тонны краски для наружных работ равен $1000, а краски для внутренних работ — $4000.
2. Ежедневное потребление сырья M1 и M2 ограничено 28 и 8 тоннами соответственно.

   a) Покажите, что внесенные изменения приведут к потере текущим оптимальным решением как свойства оптимальности, так и допустимости.
   b) Используя *обобщенный симплексный алгоритм* из упр. 4.5,a(7), найдите новое оптимальное допустимое решение.

Транспортные модели

5.1. Определение транспортной модели

Транспортные модели (задачи) — специальный класс задач линейного программирования. Эти модели часто описывают перемещение (перевозку) какого-либо товара из **пункта отправления** (исходный пункт, например место производства) в **пункт назначения** (склад, магазин, грузохранилище). Назначение транспортной задачи — определение объемов перевозок из пунктов отправления в пункты назначения с минимальной суммарной стоимостью перевозок. При этом должны учитываться ограничения, накладываемые на объемы грузов, имеющихся в пунктах отправления (предложения), и ограничения, учитывающие потребность грузов в пунктах назначения (спрос). В транспортной модели предполагается, что стоимость перевозки по какому-либо маршруту прямо пропорциональна объему груза, перевозимого по этому маршруту. В общем случае транспортную модель можно применять для описания ситуаций, связанных с управлением запасами, управлением движением капиталов, составлением расписаний, назначением персонала и др.

На рис. 5.1 показано представление транспортной задачи в виде сети с m пунктами отправления и n пунктами назначения, которые показаны в виде **узлов** сети. **Дуги**, соединяющие узлы сети, соответствуют маршрутам, связывающим пункты отправления и назначения. С дугой (i, j), соединяющей пункт отправления i с пунктом назначения j, соотносятся два вида данных: (1) стоимость c_{ij} перевозки единицы груза из пункта i в пункт j и (2) количество перевозимого груза x_{ij}. Объем грузов в пункте отправления i равен a_i, а объем грузов в пункте назначения j — b_j. Задача состоит в определении неизвестных величин x_{ij}, минимизирующих суммарные транспортные расходы и удовлетворяющих ограничениям, накладываемым на объемы грузов в пунктах отправления (предложения) и пунктах назначения (спрос).

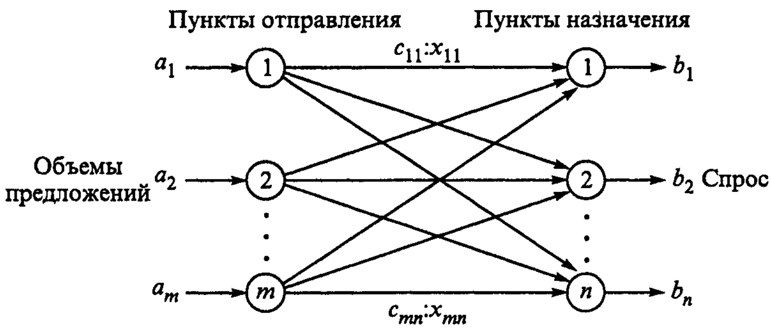


Рис. 5.1

Пример 5.1–1

Автомобильная компания MG Auto имеет три завода в Лос-Анджелесе, Детройте и Новом Орлеане и два распределительных центра в Денвере и Майами. Объемы производства заводов компании в следующем квартале составят соответственно 1000, 1500 и 1200 автомобилей. Ежеквартальная потребность распределительных центров составляет 2300 и 1400 автомобилей. Расстояния (в милях) между заводами и распределительными центрами приведены в табл. 5.1.

ТАБЛИЦА 5.1

	Денвер	Майами
Лос-Анджелес	1000	2690
Детройт	1250	1350
Новый Орлеан	1275	850

Транспортная компания оценивает свои услуги в 8 центов за перевозку одного автомобиля на расстояние в одну милю. В результате получаем следующую стоимость перевозок (с округлением до доллара) по каждому маршруту.

ТАБЛИЦА 5.2

	Денвер (1)	Майами (2)
Лос-Анджелес (1)	\$80	\$215
Детройт (2)	\$100	\$108
Новый Орлеан (3)	\$102	\$68

Основываясь на данных из табл. 5.2, формулируем следующую задачу линейного программирования.

$$\text{Минимизировать } z = 80x_{11} + 215x_{12} + 100x_{21} + 108x_{22} + 102x_{31} + 68x_{32}$$

при ограничениях

$$\begin{aligned}
&x_{11} + x_{12} = 1000 \text{ (Лос-Анджелес)},\\
&x_{21} + x_{22} = 1500 \text{ (Детройт)},\\
&x_{31} + x_{32} = 1200 \text{ (Новый Орлеан)},\\
&x_{11} + x_{21} + x_{31} = 2300 \text{ (Денвер)},\\
&x_{12} + x_{22} + x_{32} = 1400 \text{ (Майами)},\\
&x_{ij} \geq 0,\ i = 1, 2, 3,\ j = 1, 2.
\end{aligned}$$

Эти ограничения выражены в виде равенств, поскольку общий объем произведенных автомобилей (= 1000 + 1500 + 1200 = 3700) равен суммарному спросу распределительных центров (= 2300 + 1400 = 3700).

Данную задачу ЛП можно решить симплекс-методом. Но специфическая структура ограничений позволяет решить эту задачу более простым способом, с помощью так называемой **транспортной таблицы** (табл. 5.3).

ТАБЛИЦА 5.3

	Денвер	Майами	Объем производства
Лос-Анджелес	80 x_{11}	215 x_{12}	**1000**
Детройт	100 x_{21}	108 x_{22}	**1500**
Новый Орлеан	102 x_{31}	68 x_{32}	**1200**
Спрос	**2300**	**1400**	

Оптимальное решение (полученное с помощью программы TORA) показано на рис. 5.2. Оно предполагает перевозку 1000 автомобилей из Лос-Анджелеса в Денвер, 1300 автомобилей — из Детройта в Денвер, 200 автомобилей — из Детройта в Майами и 1200 — из Нового Орлеана в Майами. Минимальная стоимость перевозок составляет $313 200.

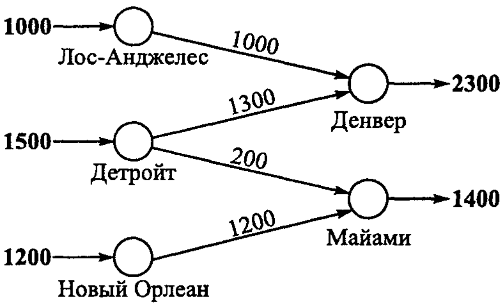


Рис. 5.2

Когда суммарный объем предложений (грузов, имеющихся в пунктах отправления) не равен общему объему спроса на товары (грузы), запрашиваемые пунктами назначения, транспортная модель называется **несбалансированной**. Далее мы последовательно будем применять прием, позволяющий любую несбалансированную транспортную задачу сделать сбалансированной. Для этого будем вводить **фиктивные** пункты назначения или отправления. Выполнение баланса транспортной задачи необходимо для того, чтобы иметь возможность применить алгоритм решения, построенный на использовании транспортных таблиц.

Пример 5.1–2

В рамках модели компании MG Auto предположим, что завод в Детройте уменьшил выпуск продукции до 1300 автомобилей (вместо 1500, как было ранее). В этом случае общее количество произведенных автомобилей (= 3500) меньше общего количества заказанных (= 3700) автомобилей. Таким образом, очевидно, что часть заказов распределительных центров Денвера и Майами не будет выполнена.

Поскольку в данной ситуации спрос превышает предложение, для восстановления баланса введем фиктивный завод (пункт отправления), производящий 200 (= 3700 – 3500) автомобилей. Назначим нулевую стоимость транспортных перевозок от фиктивного завода до пунктов назначения, поскольку такого завода не существует. В принципе, стоимость транспортных перевозок от фиктивного пункта назначения может иметь любое положительное значение. Например, чтобы гарантировать выполнение всех заказов распределительного центра Майами, можно назначить очень высокую стоимость перевозок (штраф) от фиктивного завода до Майами.

В табл. 5.4 представлена сбалансированная модель и ее оптимальное решение. Решение показывает, что фиктивный завод поставит в Майами 200 автомобилей. Это означает, что для данного распределительного центра из заказа на 1400 автомобилей не будет поставлено 200 автомобилей.

ТАБЛИЦА 5.4

	Денвер	Майами	Объем производства
Лос-Анджелес	80 **1000**	215	**1000**
Детройт	100 **1300**	108	**1500**
Новый Орлеан	102	68 **1200**	**1200**
Фиктивный завод	0	0 **200**	**200**
Спрос	**2300**	**1400**	

Если предположить, что заказ распределительного центра Денвера составляет всего 1900 автомобилей, то получим ситуацию, когда предложение превышает спрос. В этой ситуации необходимо ввести фиктивный пункт назначения, "поглощающий" избыточное предложение. Здесь также можно назначить нулевую стоимость перевозок в фиктивный пункт назначения, если не требуется выполнения каких-то особых условий. Например, если необходимо вывести всю продукцию какого-либо завода, тогда назначается очень высокая стоимость перевозок от этого завода до фиктивного пункта назначения.

В табл. 5.5 показана новая модель и ее оптимальное решение (полученное с помощью программы TORA). Решение показывает, что 400 автомобилей завода Детройта не востребованы.

Таблица 5.5

	Денвер	Майами	Фиктивный центр	
Лос-Анджелес	80 / **1000**	215	0	**1000**
Детройт	100 / **900**	108 / **200**	0 / **400**	**1500**
Новый Орлеан	102	68 / **1200**	0	**1200**
Спрос	**2300**	**1400**	**400**	

Упражнения 5.1,а

1. Истинны или ложны следующие утверждения?

   a) Для сбалансированности транспортной модели может понадобиться ввести как фиктивные пункты отправления, так и фиктивные пункты назначения.

   b) Объем перевозок в фиктивный пункт назначения равен объему превышения предложения над спросом.

   c) Объем перевозок из фиктивного пункта отправления равен разности между спросом и предложением.

2. В каждом из следующих случаев определите, следует ли ввести фиктивный пункт отправления или фиктивный пункт назначения, чтобы сбалансировать модель.

   a) Предложение: $a_1 = 10$, $a_2 = 5$, $a_3 = 4$, $a_4 = 6$. Спрос: $b_1 = 10$, $b_2 = 5$, $b_3 = 7$, $b_4 = 9$.

   b) Предложение: $a_1 = 30$, $a_2 = 44$. Спрос: $b_1 = 25$, $b_2 = 30$, $b_3 = 10$.

3. На основе табл. 5.4 из примера 5.1–2 (здесь введен фиктивный завод) интерпретируйте решение, где фиктивный завод "поставит" 150 автомобилей распределительному центру в Денвере и 50 автомобилей распределительному центру в Майами.

4. Как в табл. 5.5 из примера 5.1–2 учесть требование, что завод в Детройте должен отправить заказчикам *все* свои автомобили?

5. Пусть в примере 5.1–2 (табл. 5.4) введены штрафы в размере \$200 и \$300 за каждый недопоставленный автомобиль в распределительные центры Денвера и Майами соответственно. Кроме того, поставки из Лос-Анджелеса в Майами не планируются изначально. Постройте транспортную модель и найдите схему оптимальных перевозок с помощью программы TORA.

6. Три электрогенерирующие станции мощностью 25, 40 и 30 миллионов кВт/ч поставляют электроэнергию в три города. Максимальная потребность в электроэнергии этих городов оценивается в 30, 35 и 25 миллионов кВт/ч. Цены за миллион кВт/ч в данных городах показаны в табл. 5.6.

Таблица 5.6

		Город	
	1	2	3
Станция 1	\$600	\$700	\$400
Станция 2	\$320	\$300	\$350
Станция 3	\$500	\$480	\$450

В августе на 20% возрастает потребность в электроэнергии в каждом из трех городов. Недостаток электроэнергии города могут восполнить из другой электросети по цене \$1000 за 1 миллион кВт/ч. К сожалению, третий город не может подключиться к альтернативной электросети. Электрогенерирующие станции планируют разработать наиболее экономичный план распределения электроэнергии и восполнения ее недостатка в августе.

a) Сформулируйте эту задачу в виде транспортной модели.

b) Решите транспортную задачу с помощью программы TORA и определите оптимальный план распределения электроэнергии электрогенерирующими станциями.

c) Определите стоимость дополнительной электроэнергии для каждого из трех городов.

7. Выполните предыдущее упражнение в предположении, что 10% электроэнергии теряется при передаче по электросетям.

8. Управление национальными парками получило четыре заявки от подрядчиков на лесозаготовки в трех сосновых лесных массивах Арканзаса. Эти массивы имеют площадь 10 000, 20 000 и 30 000 акров. Каждый подрядчик может получить для разработки не более половины всех отводимых для лесозаготовки площадей. Предлагаемые подрядчиками цены за разрешение на лесозаготовки показаны в табл. 5.7.

Таблица 5.7

		Лесной массив	
	1	2	3
Подрядчик 1	\$520	\$210	\$570
Подрядчик 2	—	\$510	\$495
Подрядчик 3	\$650	—	\$240
Подрядчик 4	\$180	\$430	\$710

a) В описанной ситуации необходимо *максимизировать* общую прибыль, получаемую управлением национальными парками. Покажите, как эту проблему можно представить в виде транспортной задачи.

b) С помощью программы TORA определите площади, выделяемые каждому подрядчику для лесозаготовок.

9. Три нефтеперегонных завода с ежедневной производительностью 6, 5 и 8 миллионов галлонов бензина снабжают три бензинохранилища, ежедневная потребность которых составляет 4, 8 и 7 миллионов галлонов бензина соответственно. Бензин

транспортируется в бензинохранилища по бензопроводу. Стоимость транспортировки составляет 10 центов за 1000 галлонов на 1 милю длины трубопровода. В табл. 5.8 приведены расстояния (в милях) между заводами и хранилищами. Отметим, что первый нефтеперегонный завод не связан трубопроводом с третьим бензинохранилищем.

ТАБЛИЦА 5.8

		Бензохранилище		
		1	2	3
Завод	1	120	180	—
	2	300	100	80
	3	200	250	120

a) Сформулируйте транспортную задачу.

b) С помощью программы TORA найдите оптимальную схему транспортировки бфензина.

10. Пусть в предыдущем упражнении ежедневная производительность третьего нефтеперерабатывающего завода составляет 6 миллионов галлонов бензина, а потребности первого бензинохранилища должны выполняться в обязательном порядке. Кроме того, на недопоставки бензина во второе и третье хранилища накладываются штрафы в размере 5 центов за каждый недопоставленный галлон бензина.

a) Сформулируйте соответствующую транспортную задачу.

b) Решите сформулированную задачу с помощью программы TORA и найдите оптимальную схему поставок бензина.

11. Пусть в упр. 9 ежедневные потребности третьего бензинохранилища составляют 4 миллиона галлонов. Избыток своей продукции первый и второй нефтеперегонные заводы могут отправить на другие бензохранилища с помощью автотранспорта. В этом случае транспортные расходы на транспортировку 100 галлонов бензина составляют \$1.50 и \$2.20 для первого и второго заводов соответственно. Третий нефтеперерабатывающий завод может использовать свои излишки бензина для нужд собственного химического производства.

a) Сформулируйте соответствующую транспортную задачу.

b) Решите сформулированную задачу с помощью программы TORA и найдите оптимальную схему поставок бензина.

12. Три плодовых хозяйства поставляют апельсины в ящиках четырем оптовым покупателям. Ежедневная потребность этих покупателей составляет 150, 150, 400 и 100 ящиков соответственно. Предположим, что все три плодовых хозяйства используют только постоянную рабочую силу и могут ежедневно поставлять 150, 200 и 250 ящиков апельсинов соответственно. Первые два хозяйства могут увеличить поставки апельсинов путем привлечения дополнительных рабочих, третье хозяйство такой возможности не имеет. Транспортные расходы на один ящик апельсинов приведены в табл. 5.9.

ТАБЛИЦА 5.9

		Покупатели 1	2	3	4
	1	\$1	\$2	\$3	\$2
Хозяйства	2	\$2	\$4	\$1	\$2
	3	\$1	\$3	\$5	\$3

a) Сформулируйте соответствующую транспортную задачу.

b) Решите сформулированную задачу с помощью программы TORA.

c) Сколько дополнительных ящиков апельсинов могут поставить первое и второе хозяйства, используя временных рабочих?

13. Три распределительных центра поставляют автомобили пяти дилерам. Автомобили от распределительных центров к дилерам перевозятся на трейлерах, и стоимость перевозок пропорциональна расстоянию между пунктами отправления и назначения и не зависит от степени загрузки трейлера. В табл. 5.10 приведены расстояния между распределительными центрами и дилерами, а также соответствующие величины спроса и предложения, выраженные в *количествах* автомобилей. При полной загрузке трейлер вмещает 18 автомобилей. Транспортные расходы составляют \$25 на одну милю пути, пройденного трейлером.

ТАБЛИЦА 5.10

		Дилеры 1	2	3	4	5	Предложения
	1	100	150	200	140	35	**400**
Центры	2	50	70	60	65	80	**200**
	3	40	90	100	150	130	**150**
Спрос		**100**	**200**	**150**	**160**	**140**	

a) Сформулируйте соответствующую транспортную задачу.

b) Решите сформулированную задачу с помощью программы TORA.

14. Компания MG Auto из примера 5.1–1 производит четыре модели автомобилей. Завод в Детройте выпускает модели M1, M2 и M4. Модели M1 и M2 производятся в Новом Орлеане, M3 и M4 — в Лос-Анджелесе. Производительность каждого завода и потребности распределительных центров в автомобилях разных моделей даны в табл. 5.11.

ТАБЛИЦА 5.11

	Модель M1	M2	M3	M4	Всего
Завод					
Лос-Анджелес	—	—	700	300	1000
Детройт	500	600	—	400	1500
Новый Орлеан	800	400	—	—	1200
Распределительный центр					
Денвер	700	500	500	600	2300
Майами	600	500	200	100	1400

Расстояния между заводами и распределительными центрами приведены в примере 5.1–1, транспортные расходы составляют 8 центов за перевозку одного автомобиля (любой модели) на одну милю пути. Кроме всего прочего, есть возможность удовлетворить спрос на одну из моделей автомобилей за счет другой в соответствии с табл. 5.12.

Таблица 5.12

Распределительный центр	Заменяемая часть спроса (%)	Взаимозаменяемые модели
Денвер	10	M1, M2
	20	M3, M4
Майами	10	M1, M2
	5	M2, M4

a) Сформулируйте соответствующую транспортную задачу.

b) Решите сформулированную задачу с помощью программы TORA.

(*Совет.* Введите четыре новых пункта назначения, соответствующих комбинациям (M1, M2), (M3, M4), (M1, M2) и (M2, M4). Величины потребностей новых пунктов назначения определяются на основании данных о процентном соотношении заменяемых моделей автомобилей.)

5.2. Нетрадиционные транспортные модели

Использование транспортной модели не ограничивается задачей о транспортировке чего-либо между географическими пунктами отправления и назначения. В этом разделе описано применение транспортной модели к задаче управления запасами и задаче распределения оборудования.

Пример 5.2–1. (Управление запасами)

Фабрика Boralis производит рюкзаки для путешественников. В течение года спрос на эту продукцию, в основном, есть только в марте–июне (т.е. четыре месяца в году). Фабрика оценивает спрос в эти месяцы соответственно в 100, 200, 180 и 300 единиц изделия. Так как производство не ритмично, на фабрике рабочие работают не полный рабочий день. В течение рассматриваемых четырех месяцев фабрика может выпустить 50, 180, 280 и 270 единиц изделия. Производство и спрос в различные месяцы не совпадают, спрос в текущем месяце можно удовлетворить следующими способами.

1. Производством изделий в течение текущего месяца.
2. Избытком произведенных в прошлом месяце изделий.
3. Избытком произведенных в следующем месяце изделий в счет невыполненных заказов.

В первом случае стоимость одного рюкзака составляет \$40.00. Во втором случае возникают дополнительные расходы в расчете \$0.50 на один рюкзак за хранение в течение месяца. В третьем случае за просроченные заказы начисляются штрафы в размере \$2.00 на один рюкзак за каждый просроченный месяц. Фабрика планирует разработать оптимальный план производства на эти четыре месяца.

Описанную ситуацию можно смоделировать в виде транспортной задачи, используя следующие соответствия между элементами задачи управления запасами и транспортной модели.

Транспортная модель	Модель управления запасами
1. Пункт отправления i	1. Период производства i
2. Пункт назначения j	2. Период потребления j
3. Предложение в пункте отправления i	3. Объем производства за период i
4. Спрос в пункте назначения j	4. Объем реализации продукции за период j
5. Стоимость перевозки из пункта i в пункт j	5. Стоимость единицы продукции (производство + хранение + штрафы) за период от i до j

Результирующая транспортная модель представлена в табл. 5.13.

Таблица 5.13

	1	2	3	4	Объем производства
1	\$40.00	\$40.50	\$41.00	\$41.50	**50**
2	\$42.00	\$40.00	\$40.50	\$41.00	**180**
3	\$44.00	\$42.00	\$40.00	\$40.50	**280**
4	\$46.00	\$44.00	\$42.00	\$40.00	**270**
Спрос	**100**	**200**	**180**	**300**	

Стоимость "транспортировки" единицы изделия из периода i в период j вычисляется следующим образом.

$$c_{ij} = \begin{cases} \text{стоимость производства в } i\text{-й период}, & i = j, \\ \text{стоимость производства в } i\text{-й период} + \\ \qquad \text{стоимость хранения от } i \text{ до } j, & i < j, \\ \text{стоимость производства в } i\text{-й период} + \text{штраф от } i \text{ до } j, & i > j. \end{cases}$$

Например,

$$c_{11} = \$40.00,$$
$$c_{24} = \$40.00 + (\$0.50 + \$0.50) = \$41.00,$$
$$c_{41} = \$40.00 + (\$2.00 + \$2.00 + \$2.00) = \$46.00.$$

Оптимальное решение графически представлено на рис. 5.3. Здесь пунктирными линиями показаны задолженности производства, линией из точек представлено производство для будущего периода (для удовлетворения будущего спроса) и сплошными линиями — производство для удовлетворения спроса текущего периода.

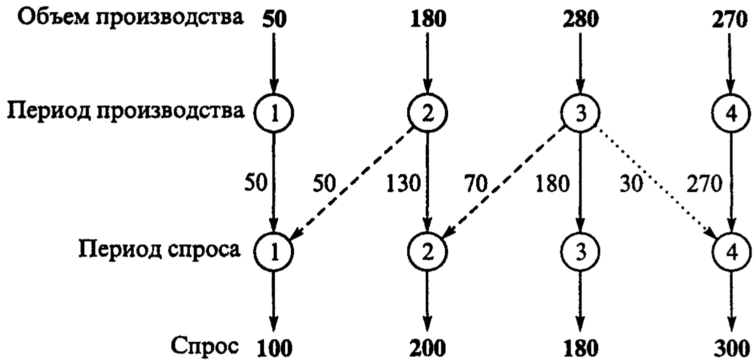


Рис. 5.3

Пример 5.2–2. (Распределение оборудования)

Лесопильный завод обрабатывает различную древесину (от мягкой сосны до твердого дуба) по утвержденному недельному производственному плану. Согласно этому плану в зависимости от типа древесины в разные рабочие дни 7-дневной рабочей недели требуется различное количество полотен для пил.

День	1 (Пн.)	2 (Вт.)	3 (Ср.)	4 (Чт.)	5 (Пт.)	6 (Сб.)	7 (Вс.)
Потребность (к-во полотен)	24	12	14	20	18	14	22

Завод может удовлетворить свои потребности в полотнах одним из следующих способов.

1. Купить новые полотна по $12 за единицу.
2. Применить ночную заточку полотен стоимостью $6 за одно заточенное полотно либо сдать пилы на 2-дневную заточку (эта услуга стоит $3 за полотно).

Данную ситуацию можно описать как транспортную модель с восьмью пунктами отправления и семью пунктами назначения. Пунктам назначения соответствуют 7 дней рабочей недели. Исходные пункты (т.е. "пункты отправления") определяются следующим образом. Первый исходный пункт соответствует покупке новых полотен; в экстремальном случае возможна такая покупка, которая удовлетворит потребность в полотнах все 7 дней рабочей недели. Исходные пункты со второго по восьмой соответствуют семи дням рабочей недели. Объем предложения каждого из этих исходных пунктов равен количеству полотен, использованных к концу соответствующего рабочего дня. Например, "предложение" второго исходного пункта (соответствует понедельнику) равно количеству полотен, необходимых для выполнения производственного плана этого рабочего дня. "Транспортные" расходы в этой модели составляют соответственно $12, $6 и $3, в зависимости от того, является ли полотно новым, получено после ночной заточки или после 2-дневного обслуживания полотен. Отметим, что для ночной заточки использованные полотна передаются в *конце* i-го рабочего дня и могут использоваться с *начала* ($i + 1$)-го или ($i + 2$)-го рабочего дня. При 2-дневной заточке использованные полотна отдаются на заточку в *конце* i-го рабочего дня и могут использоваться с *начала* ($i + 3$)-го рабочего дня или в последующие дни. В приведенной ниже таблице столбец "Остаток" соответствует фиктивному пункту назначения, в этом столбце приведено количество полотен, оставшихся не заточенными в конце каждого рабочего дня. Табл. 5.14 представляет полную транспортную модель для описанной ситуации.[1]

ТАБЛИЦА 5.14

	1 (Пн.)	2 (Вт.)	3 (Ср.)	4 (Чт.)	5 (Пт.)	6 (Сб.)	7 (Вс.)	8 (Остаток)	
1 (Новые)	$12 **24**	$12 **2**	$12	$12	$12	$12	$12	$0 **98**	**124**
2 (Пн.)	M	$6 **10**	$6 **8**	$3 **6**	$3	$3	$3	$0	**24**
3 (Вт.)	M	M	$6 **6**	$6	$3 **6**	$3	$3	$0	**12**
4 (Ср.)	M	M	M	$6 **14**	$6	$3	$3	$0	**14**
5 (Чт.)	M	M	M	M	$6 **12**	$6	$3 **8**	$0	**20**
6 (Пт.)	M	M	M	M	M	$6 **14**	$6	$0 **4**	**18**
7 (Сб.)	M	M	M	M	M	M	$6 **14**	$0	**14**
8 (Вс.)	M	M	M	M	M	M	M	$0 **22**	**22**
	24	**12**	**14**	**20**	**18**	**14**	**22**	**124**	

Оптимальное решение (полученное с помощью программы TORA) показано в следующей таблице.

Рабочий день	Новые полотна	Ночная заточка	2-дневная заточка	Остаток
Понедельник	24 (Пн.)	10 (Вт.) + 8 (Ср.)	6 (Вт.)	0
Вторник	2 (Вт.)	6 (Ср.)	6 (Пт.)	0
Среда	0	14 (Чт.)	0	0
Четверг	0	12 (Пт.)	8 (Вс.)	0
Пятница	0	14 (Сб.)	0	4
Суббота	0	14 (Вс.)	0	0
Воскресенье	0	0	0	22

Объясним полученный результат. В понедельник завод покупает 24 новых полотна для пил. В конце того же дня имеется 24 использованных полотна, из которых 18 отправляются на ночную заточку, 6 — на 2-дневную. Из 18 заточенных ночью

[1] В табл 5.14 буква M означает достаточно большое положительное число. — *Прим ред.*

полотен 10 используются во вторник, а 8 — в среду. Шесть полотен после 2-дневного обслуживания используются в четверг. Остальные позиции приведенного решения интерпретируются аналогично. В столбце "Остаток" показано количество использованных полотен, которые не передаются на заточку в конце рабочего дня.

Упражнения 5.2,а

1. Пусть в примере 5.2–1 стоимость хранения продукции зависит от месяца, в котором она произведена, и составляет для первых трех месяцев соответственно 40, 30 и 70 центов за хранение единицы продукции. Величина штрафа за просроченные заказы и величины производственных расходов остаются такими же, как и в примере 5.2–1. С помощью программы TORA найдите оптимальное решение и интерпретируйте полученный результат.

2. Пусть в примере 5.2–2 также имеется 3-дневный сервис по заточке полотен пил, использованных в понедельник и вторник; при этом стоимость заточки одного полотна составляет $1. Сформулируйте задачу заново и интерпретируйте ее оптимальное решение, полученное с помощью программы TORA.

3. Пусть в примере 5.2–2 заточенные полотна, не использованные сразу после заточки, отправляются на хранение, причем стоимость хранения одного полотна в течение дня составляет 50 центов. Переформулируйте задачу и интерпретируйте ее оптимальное решение, полученное с помощью программы TORA.

4. Компания планирует оптимизировать распределение станочного парка, состоящего из станков четырех типов, для выполнения станочных работ пяти видов. Пусть имеется 25, 30, 20 и 30 станков каждого типа. Приведем количество работ каждого вида: 20, 20, 30, 10 и 25 соответственно. Отметим, что станки четвертого типа не используются для выполнения работ четвертого вида. В табл. 5.15 представлена стоимость (в долларах) выполнения каждого вида работ на станках определенного типа. Сформулируйте транспортную задачу, решите ее с помощью программы TORA и интерпретируйте полученный результат.

Таблица 5.15

		Вид работ				
		1	2	3	4	5
Тип станка	1	10	2	3	15	9
	2	5	10	15	2	4
	3	15	5	14	7	15
	4	20	15	13	—	8

5. Спрос на некий скоропортящийся продукт в следующие 4 месяца составляет 400, 300, 420 и 380 тонн соответственно. Предложение этого товара в те же месяцы составляет 500, 600, 200 и 300 тонн. Отпускная цена на этот товар колеблется от месяца к месяцу и равна соответственно $100, $140, $120 и $150 за тонну. Поскольку товар скоропортящийся, он должен быть реализован в течение трех месяцев (включая текущий). Стоимость хранения в течение месяца тонны товара равна $3.

Природа товара такова, что невозможна задержка с выполнением заказа. Опишите данную ситуацию как транспортную модель и найдите ее оптимальное решение с помощью программы TORA.

6. Спрос на специализированные малые двигатели в следующие пять кварталов составляет 200, 150, 300, 250 и 400 единиц. Мощность производства двигателей в тот же период времени оценивается в 180, 230, 430, 300 и 300 единиц. Невыполнение заказов не допускается, при необходимости можно организовать сверхурочные работы для выпуска дополнительной продукции. Стоимость единицы продукции в каждый из следующих пяти кварталов составляет \$100, \$96, \$116, \$102 и \$106 соответственно. Стоимость дополнительно произведенной продукции увеличивается на 50% по сравнению со "стандартной" стоимостью в соответствующий период. Если двигатели, произведенные в одном квартале, реализуются в последующих, за хранение одного двигателя в течение квартала необходимо заплатить \$4. Сформулируйте транспортную задачу. С помощью программы TORA определите оптимальный план производства двигателей.

7. Периодически проводится профилактика самолетных двигателей с заменой важной детали А. В следующие 6 месяцев будут выполнены регламентные работы (с разбивкой по месяцам) на 200, 180, 300, 198, 230 и 290 двигателях. Все регламентные работы, запланированные на месяц, проводятся в течение первых двух дней месяца, когда производится замена отработанной детали А на новую или отремонтированную. Снятая деталь может быть отремонтирована в местной ремонтной мастерской, где она будет готова к началу следующего месяца, или отправлена в центральные мастерские, откуда она вернется через 3 месяца (считая месяц, в котором выполнены профилактические работы). Стоимость ремонта одной детали А в местной мастерской составляет \$120, а в центральной — только \$35. Если отремонтированная деталь будет использована в последующие месяцы, стоимость ее хранения составит \$1.50 в месяц. Новые детали можно купить по цене \$200 в первом месяце с возрастанием цены на 5% каждые 2 месяца. Представьте описанную ситуацию в виде транспортной модели и с помощью программы TORA найдите ее оптимальное решение.

5.3. Решение транспортной задачи

В данном разделе будет детально описан алгоритм решения транспортной задачи. Этот алгоритм повторяет основные шаги симплекс-метода (глава 3). Однако для представления данных, вместо обычных симплекс-таблиц, используются транспортные таблицы со специальной структурой.

Алгоритм решения транспортной задачи будет проиллюстрирован на следующем примере.

Пример 5.3–1

Транспортная компания занимается перевозкой зерна специальными зерновозами от трех элеваторов к четырем мельницам. В табл. 5.16 показаны возможности отгрузки зерна (предложения) элеваторами (в зерновозах) и потребности (спрос) мельниц (также в зерновозах), а также стоимость перевозки зерна одним зерновозом от элеваторов к мельницам. Стоимость перевозок c_{ij} приведена в сотнях долларов.

ТАБЛИЦА 5.16

		Мельницы 1	2	3	4	Предложение
Элеваторы	1	10 x_{11}	2 x_{12}	20 x_{13}	11 x_{14}	**15**
	2	12 x_{21}	7 x_{22}	9 x_{23}	20 x_{24}	**25**
	3	4 x_{31}	14 x_{32}	16 x_{33}	18 x_{34}	**10**
Спрос		**5**	**15**	**15**	**15**	

В данной задаче требуется определить структуру перевозок между элеваторами и мельницами с минимальной стоимостью. Для этого необходимо найти объемы перевозок x_{ij} между i-м элеватором и j-й мельницей.

Последовательность шагов алгоритма решения транспортной задачи в точности повторяет аналогичную последовательность этапов симплексного алгоритма.

Шаг 1. Определяем *начальное* базисное допустимое решение, затем переходим к выполнению второго шага.

Шаг 2. На основании условия оптимальности симплекс-метода среди всех небазисных переменных определяем *вводимую в базис*. Если все небазисные переменные удовлетворяют условию оптимальности, вычисления заканчиваются; в противном случае переходим к третьему шагу.

Шаг 3. С помощью условия допустимости симплекс-метода среди текущих базисных переменных определяем исключаемую. Затем находим новое базисное решение. Возвращаемся ко второму шагу.

Рассмотрим каждый описанный шаг в отдельности.

5.3.1. Определение начального решения

Общая транспортная модель с m пунктами отправления и n пунктами назначения имеет $m + n$ ограничений в виде равенств, по одному на каждый пункт отправления и назначения. Поскольку транспортная модель всегда сбалансирована (сумма предложений = сумме спроса), одно из этих равенств должно быть избыточным. Таким образом, транспортная модель имеет $m + n - 1$ независимых ограничений, отсюда вытекает, что начальное базисное решение состоит из $m + n - 1$ базисных переменных. Например, начальное решение в примере 5.3–1 содержит $3 + 4 - 1 = 6$ базисных переменных.

Специальная структура транспортной модели для построения начального решения позволяет применить следующие методы (вместо использования искусственных переменных, как это делается в симплекс-методе).

1. Метод северо-западного угла.
2. Метод наименьшей стоимости.
3. Метод Фогеля.

Различие этих методов заключается в "качестве" начального решения, т.е. насколько начальное решение ближе к оптимальному. В общем случае метод Фогеля дает наилучшее решение, а метод северо-западного угла — наихудшее. Однако метод северо-западного угла требует меньшего объема вычислений.

Метод северо-западного угла. Выполнение начинается с верхней левой ячейки (северо-западного угла) транспортной таблицы, т.е. с переменной x_{11}.

Шаг 1. Переменной x_{11} присваивается максимальное значение, допускаемое ограничениями на спрос и предложение.

Шаг 2. Вычеркивается строка (или столбец) с полностью реализованным предложением (с удовлетворенным спросом). Это означает, что в вычеркнутой строке (столбце) мы не будем присваивать значения остальным переменным (кроме переменной, определенной на первом шаге). Если одновременно удовлетворяются спрос и предложение, вычеркивается только строка или только столбец.

Шаг 3. Если не вычеркнута *только одна* строка или *только один* столбец, процесс останавливается. В противном случае переходим к ячейке справа, если вычеркнут столбец, или к нижележащей ячейке, если вычеркнута строка. Затем возвращаемся к первому шагу.

Пример 5.3–2

Если применить описанную процедуру к задаче из примера 5.3-1, получим начальное базисное решение, показанное в табл. 5.17. В этой таблице стрелками показана последовательность определения базисных переменных.

Получено следующее начальное базисное решение:

$$x_{11} = 5,\ x_{12} = 10,$$
$$x_{22} = 5,\ x_{23} = 15,\ x_{24} = 5,$$
$$x_{34} = 10.$$

Таблица 5.17

	1	2	3	4	Предложение
1	10 5	2 10	20	11	**15**
2	12	7 5	9 15	20 5	**25**
3	4	14	16	18 10	**10**
Спрос	**5**	**15**	**15**	**15**	

Соответствующая суммарная стоимость перевозок равна

$z = 5 \times 10 + 10 \times 2 + 5 \times 7 + 15 \times 9 + 5 \times 20 + 10 \times 18 = \$520.$

Метод наименьшей стоимости.[1] Данный метод находит лучшее начальное решение, чем метод северо-западного угла, поскольку выбирает переменные, которым соответствуют наименьшие стоимости. Сначала по всей транспортной таблице ведется поиск ячейки с наименьшей стоимостью. Затем переменной в этой ячейке присваивается наибольшее значение, допускаемое ограничениями на спрос и предложение. (Если таких переменных несколько, выбор произволен.) Далее вычеркивается соответствующий столбец или строка и соответствующим образом корректируются значения спроса и предложений. Если одновременно выполняются ограничения и по спросу, и по предложению, вычеркивается или строка, или столбец (точно так же, как в методе северо-западного угла). Затем просматриваются невычеркнутые ячейки, и выбирается новая ячейка с минимальной стоимостью. Описанный процесс продолжается до тех пор, пока не останется лишь одна невычеркнутая строка или столбец.

Пример 5.3–3

Применим метод наименьшей стоимости к задаче из примера 5.3–1.

1. Ячейка (1, 2) имеет наименьшую в таблице стоимость (= \$2). Наибольшее значение, которое можно присвоить переменной x_{12}, равно 15. В этом случае удовлетворяются ограничения, соответствующие первой строке и второму столбцу. Вычеркиваем второй столбец, предложение первой строки и спрос второго столбца принимают нулевые значения.
2. Следующей ячейкой с наименьшей стоимостью в незачеркнутой части таблицы будет (3, 1). Присвоим переменной x_{31} значение 5 и вычеркнем первый столбец. Ограничение по предложению, соответствующее третьей строке, станет равным $10 - 5 = 5$.
3. Продолжая процедуру, последовательно присваиваем переменной x_{23} значение 15, переменной x_{14} — значение 0; далее находим $x_{34} = 5$ и $x_{24} = 10$ (проверьте!).

Процесс поиска начального решения представлен в табл. 5.18. Стрелками показана последовательность присвоения переменным значений. Итак, получили следующее начальное базисное решение (состоящее из 6 переменных):

$$x_{12} = 15, x_{14} = 0,$$
$$x_{23} = 15, x_{24} = 10,$$
$$x_{31} = 5, x_{34} = 5.$$

Соответствующее значение целевой функции равно

$$z = 15 \times 2 + 0 \times 11 + 15 \times 9 + 10 \times 20 + 5 \times 4 + 5 \times 18 = \$475.$$

Отсюда следует, что полученное методом наименьшей стоимости начальное решение лучше, чем начальное решение, представленное методом северо-западного угла (сравните данное значение целевой функции с аналогичным значением из примера 5.3–2).

[1] Этот метод в русской математической литературе часто называют методом минимального элемента. — *Прим. ред.*

Таблица 5.18

	1	2	3	4	Предложение
1	10	2 15	20	11 0	**15**
2	12	7	9 15	20 10	**25**
3	4 5	14	16	18 5	**10**
Спрос	**5**	**15**	**15**	**15**	

Метод Фогеля. Данный метод является вариацией метода наименьшей стоимости и в общем случае находит лучшее начальное решение. Алгоритм этого метода состоит из следующих шагов.

Шаг 1. Для каждой строки (столбца), которой соответствует строго положительное предложение (спрос), вычисляется штраф путем вычитания *наименьшей* стоимости из *следующей* по величине в данной строке (столбце).

Шаг 2. Выделяется строка или столбец с наибольшим штрафом. Если таковых несколько, выбор произволен. Из выделенной строки или столбца выбирается переменная, которой соответствует минимальная стоимость, и ей присваивается наибольшее значение, позволяемое ограничениями. Затем в соответствии с присвоенным значением переменной корректируются величины оставшегося неудовлетворенным спроса и нереализованного предложения. Строка *или* столбец, соответствующие выполненному ограничению, вычеркиваются из таблицы. Если одновременно выполняются ограничения и по спросу, и по предложению, вычеркивается только строка или только столбец, причем оставшейся строке (столбцу) приписывается нулевое предложение (спрос).

Шаг 3. **a)** Если не вычеркнута только одна строка или только один столбец с нулевым спросом или предложением, вычисления заканчиваются.

b) Если не вычеркнута только одна строка (столбец) с положительным предложением (спросом), в этой строке (столбце) методом наименьшей стоимости находятся базисные переменные, и вычисления заканчиваются.

c) Если всем невычеркнутым строкам и столбцам соответствуют нулевые объемы предложения и спроса, методом наименьшей стоимости находятся *нулевые* базисные переменные, и вычисления заканчиваются.

d) Во всех остальных случаях необходимо перейти к шагу 1.

Пример 5.3–4

Применим метод Фогеля к задаче из примера 5.3–1. В табл. 5.19 показан первый набор вычисленных штрафов.

ТАБЛИЦА 5.19

	1	2	3	4		Штрафы для строк
1	10	2	20	11	15	$10-2=8$
2	12	7	9	20	25	$9-7=2$
3	4 **5**	14	16	18	10	$14-4=10$
	5	**15**	**15**	**15**		
Штрафы для столбцов	$10-4=6$	$7-2=5$	$16-9=7$	$18-11=7$		

Поскольку третья строка имеет наибольший штраф (= 10) и в этой строке наименьшая стоимость содержится в ячейке (3, 1), присваиваем переменной x_{31} значение 5. В этом случае полностью выполняется ограничение первого столбца, его вычеркиваем. Новый набор пересчитанных штрафов показан в табл. 5.20.

ТАБЛИЦА 5.20

	1	2	3	4		Штрафы для строк
1	10	2 **15**	20	11	15	$11-2=9$
2	12	7	9	20	25	$9-7=2$
3	4 **5**	14	16	18	5	$16-14=2$
	0	**15**	**15**	**15**		
Штрафы для столбцов	—	5	7	7		

Теперь первая строка имеет наибольший штраф (= 9). Поэтому мы присваиваем значение 15 переменной x_{12}, которой соответствует минимальная стоимость в первой строке. В этом случае одновременно выполняются ограничения и для первой строки, и для второго столбца. Вычеркнем второй столбец, положив объем предложений, соответствующий первой строке, равным нулю.

Продолжая этот процесс, находим, что на следующем шаге вторая строка будет иметь наибольший штраф ($20-9=11$). Поэтому переменной x_{23} присваиваем значение 15. В результате будет вычеркнут третий столбец, во второй строке останется нереализованным предложение объемом в 10 единиц. Остается невычеркнутым только четвертый столбец с положительным неудовлетворенным спросом объемом в 15 единиц. Применяя метод наименьшей стоимости к этому столбцу, последовательно получаем $x_{14} = 0$, $x_{34} = 5$ и $x_{24} = 10$ (проверьте!). Соответствующее значение целевой функции равно

$$z = 15 \times 2 + 0 \times 11 + 15 \times 9 + 10 \times 20 + 5 \times 4 + 5 \times 18 = \$475.$$

В данном случае значение целевой функции такое же, как и при методе наименьшей стоимости. Но обычно метод Фогеля дает наилучшее начальное решение для транспортной задачи.

Упражнение 5.3,а

1. Найдите начальные решения методами Фогеля, северо-западного угла и наименьшей стоимости для следующих транспортных задач.

а)

0	2	1	**6**
2	1	5	**7**
2	4	3	**7**
5	**5**	**10**	

b)

1	2	6	**7**
0	4	2	**12**
3	1	5	**11**
10	**10**	**10**	

c)

5	1	8	**12**
2	4	0	**14**
3	6	7	**4**
9	**10**	**11**	

5.3.2. Итерационный алгоритм решения транспортной задачи

После определения начального решения (с помощью одного из трех методов, описанных в предыдущем разделе) применяется алгоритм, позволяющий найти оптимальное решение транспортной задачи.

Шаг 1. На основе симплексного *условия оптимальности* среди текущего множества небазисных переменных определяется ***вводимая в базис переменная***, которая может улучшить значение целевой функции. Если условие оптимальности выполняется для всех небазисных переменных, вычисления заканчиваются, в противном случае необходимо перейти ко второму шагу.

Шаг 2. С помощью симплексного *условия допустимости* определяется *исключаемая из базиса переменная*. Происходит изменение базиса и возврат к первому шагу.

При изменении базиса в данном случае не используются вычисления, выполняемые при реализации симплекс-метода, — специальная структура транспортной таблицы позволяет значительно упростить вычисления.

Пример 5.3–5

Решим транспортную задачу из примера 5.3–1, используя начальное решение (табл. 5.21), полученное методом северо-западного угла в примере 5.3–2.

Таблица 5.21

	1	2	3	4	Предложение
1	10 / **5**	2 / **10**	20	11	**15**
2	12	7 / **5**	9 / **15**	20 / **5**	**25**
3	4	14	16	18 / **10**	**10**
Спрос	**5**	**15**	**15**	**15**	

Определение вводимой переменной среди текущих небазисных (т.е. среди тех переменных, которые не входят в начальное базисное решение) основано на вычислении коэффициентов z-строки, соответствующих небазисным переменным, с использованием **метода потенциалов** (который, как будет показано в разделе 5.4, основан на соотношениях двойственности задачи ЛП).

В методе потенциалов каждой строке i и каждому столбцу j транспортной таблицы ставятся в соответствие числа (потенциалы) u_i и v_j. Для каждой *базисной* переменной x_{ij} потенциалы u_i и v_j (как показано в разделе 5.4) удовлетворяют уравнению

$$u_i + v_j = c_{ij}.$$

В рассматриваемой задаче имеем 7 неизвестных переменных (потенциалов) и 6 уравнений, соответствующих шести базисным переменным. Чтобы найти значения потенциалов из этой системы уравнений, нужно присвоить одному из них произвольное значение (обычно полагают $u_1 = 0$) и затем последовательно вычислять значения остальных потенциалов.

Базисные переменные	Уравнения относительно потенциалов	Решение
x_{11}	$u_1 + v_1 = 10$	$u_1 = 0 \rightarrow v_1 = 10$
x_{12}	$u_1 + v_2 = 2$	$u_1 = 0 \rightarrow v_2 = 2$
x_{22}	$u_2 + v_2 = 7$	$v_2 = 2 \rightarrow u_2 = 5$
x_{23}	$u_2 + v_3 = 9$	$u_2 = 5 \rightarrow v_3 = 4$
x_{24}	$u_2 + v_4 = 20$	$u_2 = 5 \rightarrow v_4 = 15$
x_{34}	$u_3 + v_4 = 18$	$v_4 = 15 \rightarrow u_3 = 3$

Итак, имеем

$$u_1 = 0,\ u_2 = 5,\ u_3 = 3,$$
$$v_1 = 10,\ v_2 = 2,\ v_3 = 4,\ v_4 = 15.$$

Далее, используя вычисленные значения потенциалов, для каждой *небазисной* переменной вычисляются величины $u_i + v_j - c_{ij}$.

Результаты вычисления этих величин приведены в следующей таблице.

Небазисные переменные	Значения $u_i + v_j - c_{ij}$
x_{13}	$u_1 + v_3 - c_{13} = 0 + 4 - 20 = -16$
x_{14}	$u_1 + v_4 - c_{14} = 0 + 15 - 11 = 4$
x_{21}	$u_2 + v_1 - c_{21} = 5 + 10 - 12 = 3$
x_{31}	$u_3 + v_1 - c_{31} = 3 + 10 - 4 = 9$
x_{32}	$u_3 + v_2 - c_{32} = 3 + 2 - 14 = -9$
x_{33}	$u_3 + v_3 - c_{33} = 3 + 4 - 16 = -9$

Вычисленные значения совместно с нулевыми значениями для базисных переменных (поскольку $u_i + v_j - c_{ij} = 0$ для любой базисной переменной x_{ij}) фактически являются коэффициентами z-строки симплекс-таблицы.

Базис	x_{11}	x_{12}	x_{13}	x_{14}	x_{21}	x_{22}	x_{23}	x_{24}	x_{31}	x_{32}	x_{33}	x_{34}
z	0	0	–16	4	3	0	0	0	9	–9	–9	0

Поскольку в транспортной задаче ведется поиск *минимума* стоимости перевозок, вводимой в базис будет переменная, имеющая *наибольший положительный* коэффициент в z-строке. В данном случае вводимой переменной будет x_{31}.

Описанные вычисления обычно выполняются непосредственно в транспортной таблице, как показано в табл. 5.22. В этом случае нет необходимости в явном виде выписывать уравнения для потенциалов. Вычисления в транспортной таблице начинаются с присвоения потенциалу u_1 нулевого значения: $u_1 = 0$. Затем вычисляются v-потенциалы для всех столбцов, имеющих базисные переменные в первой строке. Далее на основании уравнения для потенциалов, соответствующего переменной x_{22}, вычисляется величина потенциала u_2. Зная значение потенциала u_2, вычисляем потенциалы v_3 и v_4, что позволяет найти потенциал u_3. Поскольку все потенциалы определены, далее вычисляются величины $u_i + v_j - c_{ij}$ для каждой небазисной переменной x_{ij}. Эти величины показаны в табл. 5.22 в левом нижнем углу ячеек транспортной таблицы.

ТАБЛИЦА 5.22

	$v_1 = 10$	$v_2 = 2$	$v_3 = 4$	$v_4 = 15$	Предложение
$u_1 = 0$	10 **5**	2 **10**	20 $\boxed{-16}$	11 $\boxed{4}$	**15**
$u_2 = 5$	12 $\boxed{3}$	7 **5**	9 **15**	20 **5**	**25**
$u_3 = 3$	4 $\boxed{9}$	14 $\boxed{-9}$	16 $\boxed{-9}$	18 **10**	**10**
Спрос	**5**	**15**	**15**	**15**	

Определив вводимую в базис переменную x_{31}, далее следует определить исключаемую из базиса переменную. Напомним, если какая-либо переменная вводится в базис, одна из текущих базисных переменных должна стать небазисной (и равной нулю), чтобы количество базисных переменных оставалось постоянным (в данном примере количество базисных переменных равняется $3 + 4 - 1 = 6$).

Исключаемая из базиса переменная определяется следующим образом. Выбрав в качестве вводимой переменную x_{31}, мы хотим, чтобы перевозки по маршруту, соответствующему этой переменной, уменьшили общую стоимость перевозок. Какой объем груза можно перевести по этому маршруту? Обозначим через θ количество груза, перевозимого по маршруту (3, 1) (т.е. $x_{31} = \theta$). Максимально возможное значение θ определяем из следующих условий.

1. Должны выполняться ограничения на спрос и предложение.
2. Ни по какому маршруту не должны выполняться перевозки с отрицательным объемом грузов.

Эти условия позволяют найти значение θ и определить исключаемую переменную. Сначала построим *замкнутый цикл*, который начинается и заканчивается в ячейке, соответствующей вводимой переменной (в данном примере — это ячейка (3, 1)). Цикл состоит из последовательности горизонтальных и вертикальных отрезков (но не диагональных), соединяющих ячейки, соответствующие текущим базисным переменным, и ячейку, соответствующую вводимой переменной. В табл. 5.23 показан цикл для вводимой переменной x_{31}. Для любой вводимой переменной можно построить только один замкнутый цикл.

ТАБЛИЦА 5.23

	$v_1 = 10$	$v_2 = 2$	$v_3 = 4$	$v_4 = 15$	Предложение
$u_1 = 0$	10; $5 - \theta$	2; $10 + \theta$	20; $\boxed{-16}$	11; $\boxed{4}$	**15**
$u_2 = 5$	12; $\boxed{3}$	7; $5 - \theta$	9; **15**	20; $5 + \theta$	**25**
$u_3 = 3$	4; θ; $\boxed{9}$	14; $\boxed{-9}$	16; $\boxed{-9}$	18; $10 - \theta$	**10**
Спрос	**5**	**15**	**15**	**15**	

Теперь найдем значение θ. Для того чтобы удовлетворить ограничениям по спросу и предложению, надо поочередно отнимать и прибавлять θ к значениям базисных переменных, расположенных в *угловых* ячейках цикла, как показано в табл. 5.23 (не имеет значения направление обхода цикла: по часовой стрелке или против). Новые значения базисных переменных останутся неотрицательными, если будут выполняться следующие неравенства.

$$x_{11} = 5 - \theta \geq 0,$$
$$x_{22} = 5 - \theta \geq 0,$$
$$x_{34} = 10 - \theta \geq 0.$$

Отсюда следует, что наибольшее значение, которое может принять θ, равно 5, при этом переменные x_{11} и x_{22} обращаются в нуль. Поскольку только одна переменная исключается из базиса, в качестве исключаемой можно выбрать как x_{11}, так и x_{22}. Остановим свой выбор на x_{11}.

Определив значение для вводимой переменной ($x_{31} = 5$) и выбрав исключаемую переменную, далее следует откорректировать значения базисных переменных, соответствующих угловым ячейкам замкнутого цикла, как показано в табл. 5.24. Поскольку перевозка единицы груза по маршруту (3, 1) уменьшает общую стоимость перевозок на \$9 ($= u_3 + v_1 - c_{31}$), суммарная стоимость перевозок будет на $\$9 \times 5 = \$45$ меньше, чем в предыдущем решении. Таким образом, новая суммарная стоимость перевозок будет равна \$520 – \$45 = \$475.

Имея новое базисное решение, следует повторить вычисления потенциалов, как показано в табл. 5.24. Новой вводимой в базис переменной будет x_{14}. Замкнутый цикл, соответствующий этой переменной, позволяет найти ее значение ($x_{14} = 10$) и исключаемую переменную x_{24}.

Таблица 5.24

	$v_1 = 1$	$v_2 = 2$	$v_3 = 4$	$v_4 = 15$	Предложение
$u_1 = 0$	10 $\boxed{-9}$	2 **15 − θ**	20 $\boxed{-16}$	11 **θ** $\boxed{4}$	**15**
$u_2 = 5$	12 $\boxed{-6}$	7 **0 + θ**	9 **15**	20 **10 − θ**	**25**
$u_3 = 3$	4 **5**	14 $\boxed{-9}$	16 $\boxed{-9}$	18 **5**	**10**
Спрос	**5**	**15**	**15**	**15**	

Новое решение, показанное в табл. 5.25, на $\$4 \times 10 = \$40$ уменьшает значение целевой функции. Таким образом, новая суммарная стоимость перевозок составляет \$475 – \$40 = \$435. Теперь новые значения величин $u_i + v_j - c_{ij}$ для всех небазисных переменных x_{ij} отрицательные. Поэтому решение, представленное в табл. 5.25, оптимально.

Таблица 5.25

	$v_1 = -3$	$v_2 = 2$	$v_3 = 4$	$v_4 = 11$	Предложение
$u_1 = 0$	10 $\boxed{-13}$	2 **5**	20 $\boxed{-16}$	11 **10**	**15**
$u_2 = 5$	12 $\boxed{-10}$	7 **10**	9 **15**	20 $\boxed{-4}$	**25**
$u_3 = 7$	4 **5**	14 $\boxed{-5}$	16 $\boxed{-5}$	18 **5**	**10**
Спрос	**5**	**15**	**15**	**15**	

Полученное решение, в терминах исходной задачи перевозки зерна от элеваторов до мельниц, имеет следующий смысл.

От элеватора	До мельницы	Количество зерновозов
1	2	5
1	4	10
2	2	10
2	3	15
3	1	5
3	4	5

Минимальная стоимость всех перевозок равна \$435.

Упражнения 5.3,b

1. Используя начальное решение, полученное методом северо-западного угла, найдите оптимальное решение в транспортных моделях, представленных в табл. 5.26.

ТАБЛИЦА 5.26

a)

\$0	\$2	\$1	**6**
\$2	\$1	\$5	**9**
\$2	\$4	\$3	**5**
5	**5**	**10**	

b)

\$0	\$4	\$2	**8**
\$2	\$3	\$4	**5**
\$1	\$2	\$0	**6**
7	6	6	

c)

—	\$3	\$5	**4**
\$7	\$4	\$9	**7**
\$1	\$8	\$6	**19**
5	**6**	**19**	

2. В транспортной модели (табл. 5.27) суммарный спрос превышает общий объем предложений. Пусть штрафы за необеспеченный единичный спрос составляют \$5 для первого пункта назначения, \$3 — для второго и \$2 — для третьего. Найдите оптимальное решение.

ТАБЛИЦА 5.27

\$5	\$1	\$7	**10**
\$6	\$4	\$6	**80**
\$3	\$2	\$5	**15**
75	**20**	**50**	

3. Пусть в транспортной модели из предыдущего упражнения штрафы не назначаются, но спрос третьего пункта назначения должен быть выполнен полностью. Найдите оптимальное решение.

4. В несбалансированной транспортной модели (табл. 5.28) стоимость хранения единицы груза, не отправленного из первого, второго и третьего пунктов отправления, составляет соответственно \$5, \$4 и \$3. Найдите оптимальное решение, если из второго пункта отправления груз должен быть вывезен полностью.

Таблица 5.28

\$1	\$2	\$1	**20**
\$3	\$4	\$5	**40**
\$2	\$3	\$3	**30**
30	**20**	**20**	

5. Пусть в транспортной модели размерностью 3×3 через x_{ij} обозначено количество грузов, перевозимых из пункта отправления i в пункт назначения j, а через c_{ij} — стоимость перевозки. Объемы грузов в первом, втором и третьем пунктах отправления равны соответственно 15, 30 и 85 единиц. Спрос в первом, втором и третьем пунктах назначения составляет 20, 30 и 80 единиц соответственно. Предположим, что начальное решение, полученное методом северо-западного угла, оптимально со следующими значениями потенциалов: $u_1 = -2$, $u_2 = 3$, $u_3 = 5$, $v_1 = 2$, $v_2 = 5$ и $v_3 = 10$.

   a) Найдите оптимальную стоимость транспортных расходов.

   b) Определите наименьшие значения c_{ij} для всех небазисных переменных, сохраняющих оптимальность решения, полученного методом северо-западного угла.

6. В табл. 5.29 показано *вырожденное* базисное решение некой транспортной задачи. Пусть этому решению соответствуют следующие значения потенциалов: $u_1 = 1$, $u_2 = -1$, $v_1 = 2$, $v_2 = 2$ и $v_3 = 5$. Стоимости перевозок для всех *нулевых* (базисных и небазисных) переменных x_{ij} представим в виде

$$c_{ij} = i + j\theta, \quad -\infty < \theta < \infty.$$

Таблица 5.29

10			**10**
	20	**20**	**40**
10	**20**	**20**	

   a) Найдите значение целевой функции, если данное решение оптимально.

   b) Укажите значения параметра θ, гарантирующие оптимальность данного решения. (*Совет*. Определите местоположение нулевых базисных переменных.)

7. Дана следующая задача.

$$\text{Минимизировать } z = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij}$$

при выполнении ограничений

$$\sum_{j=1}^{n} x_{ij} \geq a_i, \quad i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^{m} x_{ij} \geq b_j, \quad j = 1, 2, \ldots, n,$$

$$x_{ij} \geq 0, \text{ для всех } i \text{ и } j.$$

Логично предположить, что оптимальное решение обращает первое или второе множество неравенств в равенства (в зависимости от того, будет ли выполняться неравенство $\sum a_i \ge \sum b_j$ или $\sum a_i \le \sum b_j$). Контрпример, опровергающий это предположение, представлен в табл. 5.30 (см. [2]).

ТАБЛИЦА 5.30

\$1	\$1	\$2	5
\$6	\$5	\$1	6
2	7	1	

Покажите, что данное предположение приводит к "оптимальному" решению $x_{11} = 2$, $x_{12} = 3$, $x_{22} = 4$, $x_{23} = 2$, дающему $z = \$27$, которое хуже решения $x_{11} = 2$, $x_{12} = 7$, $x_{23} = 6$ с $z = \$15$.

5.3.3. Интерпретация метода потенциалов как симплекс-метода

Связь метода потенциалов с симплекс-методом основывается на соотношениях двойственности задач ЛП (раздел 4.4.2). Исходя из специальной структуры транспортной задачи (обратитесь к примеру 5.5–1, где показано, как транспортную задачу представить в виде стандартной задачи ЛП), двойственная ей задача будет записана в следующем виде.

$$\text{Максимизировать } z = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j$$

при ограничениях

$$u_i + v_j \le c_{ij} \quad \text{для всех } i \text{ и } j,$$
$$u_i \text{ и } v_j \text{ — свободные переменные,}$$

где

a_i — предложение (объем грузов) пункта отправления i,

b_j — спрос (заявка на грузы) пункта назначения j,

c_{ij} — стоимость перевозки единицы груза из пункта отправления i в пункт назначения j,

u_i — двойственная переменная, соответствующая ограничению на предложение пункта отправления i,

v_j — двойственная переменная, соответствующая ограничению на спрос пункта назначения j.

Из соотношений двойственности (раздел 4.3) следует, что коэффициент при переменной x_{ij} в выражении целевой функции должен быть равен разности между левой и правой частями соответствующего ограничения двойственной задачи, т.е. величине $u_i + v_j - c_{ij}$. Но как мы уже знаем, эта величина должна быть равной нулю для каждой *базисной переменной*. Другими словами, для этих переменных должно выполняться равенство

$u_i + v_j = c_{ij}$. Имея $m + n - 1$ таких равенств и решая их как систему линейных уравнений (после присвоения какой-либо переменной произвольного значения, например $u_1 = 0$), находим значения потенциалов u_i и v_j. Вычислив значения потенциалов, далее определяем вводимую в базис переменную среди всех *небазисных* как переменную, имеющую наибольшее положительное значение величины $u_i + v_j - c_{ij}$.

Присвоение одной из двойственных переменных произвольного значения (например, $u_1 = 0$) противоречит представлениям раздела 4.6, поскольку это присвоение показывает, что решение двойственной задачи (определяется вычисленными значениями двойственных переменных (потенциалов)) не единственно. В действительности, противоречия здесь нет, и решение упр. 5.3,с(2) объясняет, — почему.

Упражнения 5.3,с

1. Запишите двойственную задачу для транспортной модели из примера 5.3–5 (см. табл. 5.21). Вычислите оптимальное значение целевой функции *двойственной* задачи с помощью значений двойственных переменных, приведенных в табл. 5.25, и покажите, что найденное значение совпадает с оптимальным значением целевой функции транспортной задачи.

2. В транспортной модели одной из двойственных переменных присваивается произвольное значение. Это означает, что одному и тому же базисному решению прямой задачи соответствует не единственный набор значений двойственных переменных. Это противоречит положениям теории линейного программирования, где значения двойственных переменных вычисляются как произведение вектора коэффициентов целевой функции, стоящих при базисных переменных, на обратную матрицу соответствующего базиса (см. раздел 4.6). Покажите, что в транспортной модели обратная матрица всегда определяется однозначно, в то время как вектор коэффициентов целевой функции, стоящих при *базисных* переменных, можно определить не единственным образом. В частности, покажите, что если коэффициенты c_{ij} заменить на $c_{ij} + k$ для всех i и j, где k — произвольная константа, то оптимальные значения переменных x_{ij} не изменятся. Покажите, что присвоение одной двойственной переменной произвольного значения эквивалентно прибавлению к коэффициентам c_{ij} некой константы k.

5.4. Задача о назначениях

"Лучший работник для выполнения данной работы" — вот подходящее краткое описание задачи о назначениях. В этой задаче необходимо назначить работников на определенные работы; каждый работник может выполнять любую работу, хотя и с различной степенью мастерства. Если на некоторую работу назначается работник именно той квалификации, которая необходима для ее выполнения, тогда стоимость выполнения работы будет ниже, чем при назначении на данную работу работника неподходящей квалификации. Цель задачи — найти оптимальное (минимальной стоимости) распределение работников по всем заявленным работам.

Общая задача назначения n работников на n работ представлена в табл. 5.31.

ТАБЛИЦА 5.31

		Работы				
		1	2	...	n	
Работники	1	c_{11}	c_{12}	...	c_{1n}	1
	2	c_{21}	c_{22}	...	c_{2n}	1
	.	.	.		.	.
	.	.	.		.	.
	.	.	.		.	.
	n	c_{n1}	c_{n2}	...	c_{nn}	1
		1	1	...	1	

Коэффициент c_{ij} равен стоимости назначения работника i на работу j ($i, j = 1, 2, ..., n$). То, что количество работников равно количеству работ, не является ограничением общности, поскольку всегда можно ввести в модель фиктивных работников или фиктивные работы.

Задача о назначениях является частным случаем транспортной задачи, в которой работники соответствуют пунктам отправления, а работы — пунктам назначения. В данном случае все величины спроса и предложения равны 1. Стоимость "транспортировки" рабочего i на работу j равна c_{ij}. Задачу о назначениях можно эффективно решить точно так же, как и транспортную задачу. Вместе с тем тот факт, что все величины спроса и предложения равны 1, привел к разработке упрощенного алгоритма решения, названного **венгерским методом**. Хотя этот метод не имеет никакого отношения к транспортной задаче, он, как и метод потенциалов, все равно основан на симплекс-методе.

5.4.1. Венгерский метод

Для представления этого метода используем два примера. Интерпретация венгерского метода как симплекс-метода будет рассмотрена в следующем разделе.

Пример 5.4–1

Трое детей Джоя Клини — Джон, Карен и Терри, желают подзаработать немного денег на школьную экскурсию в местный зоопарк. М-р Клини выбрал три вида работ, которые можно выполнить детям за определенную плату: подстрижка газона, уборка гаража и мойка семейного автомобиля. Чтобы избежать ненужных споров между детьми, он опросил каждого (конечно, по секрету), сколько за каждый вид работ они хотят получить. Результаты опроса представлены в табл. 5.32.

ТАБЛИЦА 5.32

	Подстрижка газона	Уборка гаража	Мойка машины
Джон	\$15	\$10	\$9
Карен	\$9	\$15	\$10
Терри	\$10	\$12	\$8

Основываясь на этой информации, как распределить работы между детьми с минимальными (денежными) потерями для м-ра Клини?

Решим эту задачу о назначениях венгерским методом.

Шаг 1. В исходной матрице стоимостей определим в каждой строке минимальную стоимость и отнимем ее от других элементов строки.

Шаг 2. В матрице, полученной на первом шаге, найдем в каждом столбце минимальную стоимость и отнимем ее от других элементов столбца.

Шаг 3. Оптимальным назначениям будут соответствовать нулевые элементы, полученные на предыдущем шаге.

Обозначим через p_i и q_j минимальные стоимости соответственно в строке i и столбце j, определенные на первом и втором шагах описанного выше алгоритма. Минимальные стоимости по строкам находятся по исходной матрице стоимостей, как показано в табл. 5.33.

ТАБЛИЦА 5.33

	Подстрижка газона	Уборка гаража	Мойка машины	Минимумы по строкам
Джон	\$15	\$10	\$9	$p_1 = 9$
Карен	\$9	\$15	\$10	$p_2 = 9$
Терри	\$10	\$12	\$8	$p_3 = 8$

Теперь вычтем минимальные стоимости из элементов соответствующих строк, и в результате получим следующую матрицу.

ТАБЛИЦА 5.34

	Подстрижка газона	Уборка гаража	Мойка машины
Джон	6	1	0
Карен	0	6	1
Терри	2	4	0
Минимумы по столбцам	$q_1 = 0$	$q_2 = 1$	$q_3 = 0$

На втором шаге алгоритма находим минимальные значения по столбцам и вычитаем их из элементов соответствующих столбцов. В результате получим матрицу, представленную в виде табл. 5.35.

ТАБЛИЦА 5.35

	Подстрижка газона	Уборка гаража	Мойка машины
Джон	6	$\underline{0}$	0
Карен	$\underline{0}$	5	1
Терри	2	3	$\underline{0}$

В последней матрице подчеркнутые нулевые элементы определяют оптимальное решение: Джон будет убирать в гараже, Карен подстригать газон, а Терри достанется мойка машины. Эти работы обойдутся м-ру Клини в 9 + 10 + 8 = \$27. Отметим, что эта сумма всегда равна $(p_1 + p_2 + p_3) + (q_1 + q_2 + q_3) = (9 + 9 + 8) + (0 + 1 + 0) = \$27$. (Доказательство этого приведено в следующем разделе.)

Алгоритм венгерского метода в предыдущем примере завершен успешно, поскольку нулевые элементы в конечной матрице соответствуют *допустимому* решению (в том смысле, что каждому ребенку назначена в точности одна работа). В некоторых случаях нулевые элементы, полученные на первом и втором шагах венгерского метода, не позволяют непосредственно получить допустимое решение. Тогда необходимы дополнительные действия для нахождения оптимального (допустимого) решения. Следующий пример демонстрирует это.

Пример 5.4–2

Предположим, что в примере 5.4–1 представлено четыре ребенка и четыре вида работ. Табл. 5.36 соответствует матрице стоимостей для этой задачи.

ТАБЛИЦА 5.36

		Виды работ			
		1	2	3	4
Дети	1	\$1	\$4	\$6	\$3
	2	\$9	\$7	\$10	\$9
	3	\$4	\$5	\$11	\$7
	4	\$8	\$7	\$8	\$5

Применение первого и второго шагов алгоритма к исходной матрице (при этом $p_1 = 1, p_2 = 7,\ p_3 = 4,\ p_4 = 5,\ q_1 = 0,\ q_2 = 0,\ q_3 = 3$ и $q_4 = 0$) приводит к следующей матрице (проверьте!).

ТАБЛИЦА 5.37

		Виды работ			
		1	2	3	4
Дети	1	0	3	2	2
	2	2	0	0	2
	3	0	1	4	3
	4	3	2	0	0

В последней матрице расположение нулевых элементов не позволяет назначить каждому ребенку одну работу. Например, если мы назначим первому ребенку работу 1, из дальнейшего рассмотрения исключается первый столбец и тогда в строке третьего ребенка не окажется нулевых элементов. Подобные проблемы можно разрешить, добавив к описанной в примере 5.4–1 процедуре следующий шаг.

Шаг 1. Если после выполнения первого и второго шагов описанного алгоритма не получено допустимое решение, выполните следующие действия.

i) В последней матрице проведите *минимальное число* горизонтальных и вертикальных прямых по строкам и столбцам с тем, чтобы вычеркнуть в матрице все нулевые элементы.

ii) Найдите *наименьший невычеркнутый* элемент и вычтите его из остальных невычеркнутых элементов и прибавьте к элементам, стоящим на пересечении проведенных на предыдущем шаге прямых.

iii) Если новое распределение нулевых элементов не позволяет построить допустимое решение, повторите шаг 2,а. В противном случае перейдите к третьему шагу алгоритма.

В задаче данного примера выполнение шага 2,а требует проведения трех прямых и приводит к табл. 5.38.

Таблица 5.38

		Виды работ			
		1	2	3	4
Дети	1	0	3	2	2
	2	2	0	0	2
	3	0	<u>1</u>	4	3
	4	3	2	0	0

Наименьший невычеркнутый элемент (он подчеркнут) равен 1. Этот элемент вычитаем из остальных невычеркнутых элементов и прибавляем к элементам, стоящим на пересечении прямых. В результате получим матрицу, представленную в виде табл. 5.39.

Таблица 5.39

		Виды работ			
		1	2	3	4
Дети	1	**<u>0</u>**	2	1	1
	2	3	0	**<u>0</u>**	2
	3	0	**<u>0</u>**	3	2
	4	4	2	0	**<u>0</u>**

Оптимальное решение, показанное в таблице подчеркнутыми нулями, предлагает первому ребенку работу 1, второму — работу 3, третьему — работу 2 и четвертому — работу 4. Соответствующее значение целевой функции равно 1 + 10 + 5 + 5 = \$21. Такое же значение можно получить путем суммирования значений p_i и q_j и значения элемента, наименьшего среди всех невычеркнутых: (1 + 7 + 4 + 5) + (0 + 0 + 3 + 0) + (1) = \$21.

Упражнения 5.4,а

1. Решите задачи о назначениях, представленные в табл. 5.40.

Таблица 5.40

a)

\$3	\$8	\$2	\$10	\$3
\$8	\$7	\$2	\$9	\$7
\$6	\$4	\$2	\$7	\$5
\$8	\$4	\$2	\$3	\$5
\$9	\$10	\$6	\$9	\$10

b)

\$3	\$9	\$2	\$3	\$7
\$6	\$1	\$5	\$6	\$6
\$9	\$4	\$7	\$10	\$3
\$2	\$5	\$4	\$2	\$1
\$9	\$6	\$2	\$4	\$5

2. Дана следующая задача распредёления четырех рабочих по четырем видам работ. Различная квалификация рабочих обуславливает различную стоимость выполнения работ. Стоимость работ приведена в табл. 5.41. Отметим, что первый рабочий не может выполнять работу 3, а третий — работу 4. Найдите оптимальное решение.

ТАБЛИЦА 5.41

		Виды работ			
		1	2	3	4
Рабочие	1	\$50	\$50	—	\$20
	2	\$70	\$40	\$20	\$30
	3	\$90	\$30	\$50	—
	4	\$70	\$20	\$60	\$70

3. Пусть в задаче из предыдущего упражнения можно ввести нового (пятого) рабочего, способного выполнить любой вид работ со стоимостью соответственно \$60, \$45, \$30 и \$80. Будет ли экономически выгодным заменить одного из "работающих" рабочих новым?

4. Пусть в задаче из упр. 2 необходимо ввести новый вид работы, который может выполнить любой из четырех рабочих, со стоимостью соответственно \$20, \$10, \$20 и \$80. Будет ли новая работа более выгодной по сравнению с имеющимися?

5. Бизнесмен должен сделать четыре поездки туда и обратно из своего головного офиса в Далласе в филиал в Атланте (данные о поездках приведены в табл. 5.42).

ТАБЛИЦА 5.42

Дата отправления из Далласа	Дата возвращения в Даллас
Понедельник, 3 июня	Пятница, 7 июня
Понедельник, 10 июня	Среда, 12 июня
Понедельник, 17 июня	Пятница, 21 июня
Вторник, 25 июня	Пятница, 28 июня

Билет туда и обратно, т.е. Даллас-Атланта-Даллас, стоит \$400. Скидка 25% предоставляется тогда, когда даты отправления и прибытия совпадают с выходными днями (суббота и воскресенье). Если в билете между датами прибытия в Атланту и отправления из нее не менее 21 дня, то скидка увеличивается до 30%. Билет в одну сторону (в любом направлении) стоит \$250. Какую минимальную сумму может потратить на билеты бизнесмен?

6. На рис. 5.4 схематически показан план обрабатывающего цеха с четырьмя старыми станками, обозначенными квадратиками с номерами от 1 до 4. В цех будут поставлены четыре новых станка, местоположение которых обозначено кружочками с буквами a, b, c и d. Необходимо так разместить новые станки, чтобы минимизировать суммарные перемещения обрабатываемых деталей между новыми и старыми станками. В табл. 5.43 показана интенсивность перемещения деталей между новыми и старыми станками. Детали перемещаются по сторонам прямоугольника, противоположными вершинами которого будут старый и новый станки. Например, расстояние (в метрах) между станком 1 и станком, расположенным в позиции b, равно 30 + 20 = 50 м.

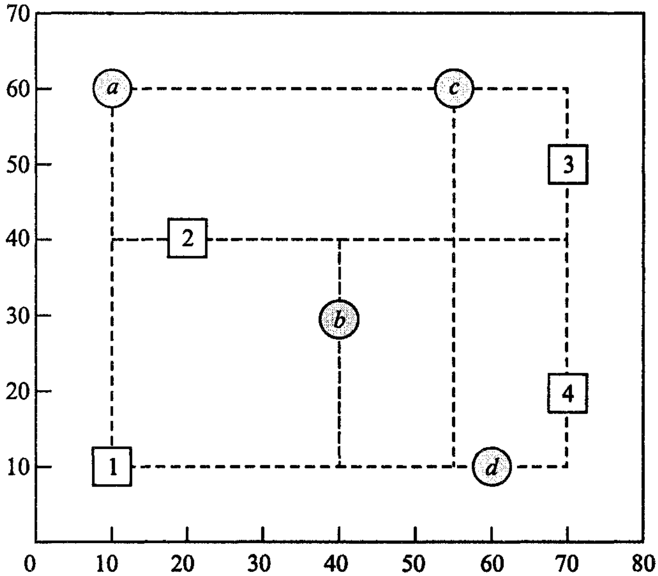


Рис. 5.4

Таблица 5.43

		Новые станки			
		1	2	3	4
Старые станки	1	10	2	4	3
	2	7	1	9	5
	3	0	8	6	2
	4	11	4	0	7

5.4.2. Интерпретация венгерского метода как симплекс-метода

Задачу о назначении n работников на n видов работ можно представить в виде задачи линейного программирования следующим образом. Обозначим через c_{ij} стоимость назначения работника i на работу j и определим

$$x_{ij} = \begin{cases} 1, & \text{если работник } i \text{ назначен на работу } j, \\ 0, & \text{в противном случае.} \end{cases}$$

Получаем следующую задачу ЛП.

$$\text{Минимизировать } z = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

при выполнении условий

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots, n,$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1, 2, ..., n,$$

$$x_{ij} = 0 \text{ или } 1.$$

Оптимальное решение данной задачи ЛП останется неизменным, если ко всем элементам какой-либо строки или столбца матрицы стоимостей (c_{ij}) прибавить константу или вычесть ее из этих элементов. Для доказательства этого обозначим через p_i и q_j константы, вычитаемые из элементов строки i и столбца j соответственно. Таким образом, стоимость c_{ij} изменится и будет равна

$$c'_{ij} = c_{ij} - p_i - q_j.$$

Теперь покажем, что при коэффициентах целевой функции c'_{ij} получим те же оптимальные значения переменных x_{ij}, что и при коэффициентах c_{ij}.

$$\sum_i \sum_j (c_{ij} - p_i - q_j)x_{ij} = \sum_i \sum_j c_{ij}x_{ij} - \sum_i p_i \left(\sum_j x_{ij} \right) - \sum_j q_j \left(\sum_i x_{ij} \right) =$$
$$= \sum_i \sum_j c_{ij}x_{ij} - \sum_i p_i (1) - \sum_j q_j (1) = \sum_i \sum_j c_{ij}x_{ij} - \text{константа}.$$

Поскольку новая целевая функция отличается от исходной только на константу, оптимальные значения переменных x_{ij} должны быть одинаковы в обоих случаях. Таким образом показано, что шаги 1 и 2 венгерского метода, где p_i вычитаются из элементов строки i, а q_j — из элементов столбца j матрицы стоимостей, приводят к эквивалентной задаче о назначениях. Поэтому, если нулевые элементы в матрице стоимостей, созданные на шаге 1 и 2 венгерского метода, позволяют найти допустимое решение, то оно должно быть оптимальным, поскольку стоимости в измененной матрице стоимостей не могут быть меньше нуля.

Если созданные нулевые элементы не могут привести к допустимому решению (как в примере 5.4–2), необходимо выполнить описанный выше шаг 2,а. Эта процедура опять основывается на симплекс-методе, и ее корректность можно обосновать, исходя из теории двойственности (глава 4) и теоремы о дополняющей нежесткости (глава 7). Пока мы не будем углубляться в это.

То, что сумма $\sum_i p_i + \sum_j q_j$ равна оптимальному значению целевой функции, вытекает из того, что данная сумма представляет собой целевую функцию двойственной задачи. Это видно из представленного в разделе 5.3.3 выражения для целевой функции задачи, двойственной транспортной задаче. (Подробности см. в [1].)

5.5. Транспортная модель с промежуточными пунктами

Транспортная модель с промежуточными пунктами соответствует реальной ситуации, когда между исходными и конечными пунктами перевозок имеются промежуточные пункты для временного хранения грузов (*транзитные* пункты). Эта модель более общая, чем обычная транспортная, где перевозки осуществляются непосредственно между пунктами отправления и назначения.

В этом разделе показано, как транспортную модель с промежуточными пунктами можно преобразовать (и решить) в обычную транспортную с помощью введения **буфера**.

Пример 5.5–1

Два автомобильных завода $P1$ и $P2$ связаны с тремя дилерами $D1$, $D2$ и $D3$, имеющими два транзитных (перевалочных) центра $T1$ и $T2$, как показано на рис. 5.5. Заводы $P1$ и $P2$ производят 1000 и 1200 автомобилей. Заказы дилеров составляют соответственно 800, 900 и 300 автомобилей. Стоимость перевозок одного автомобиля (в сотнях долларов) показана на рис. 5.5.

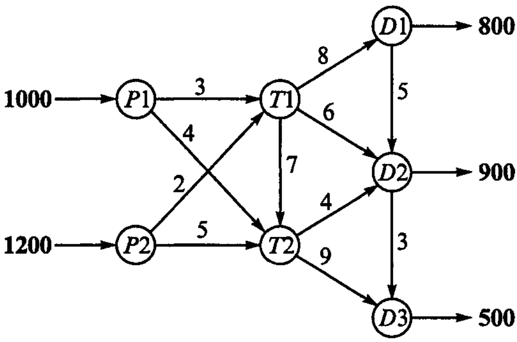


Рис. 5.5

В данной модели перевозки транзитом могут осуществляться через любые пункты (в соответствии с направлением дуг на схеме), даже через некоторые пункты назначения. Поэтому пункты, которым соответствуют как входящие, так и выходящие дуги на схеме рис. 5.5, назовем **транзитными** (пункты $T1$, $T2$, $D1$ и $D2$). Оставшиеся будут либо **истинными пунктами отправления** (пункты $P1$ и $P2$), либо **истинными пунктами назначения** (в данной схеме такой пункт только один — $D3$). Эту модель можно преобразовать в обычную транспортную модель с шестью пунктами отправления ($P1$, $P2$, $T1$, $T2$, $D1$ и $D2$) и пятью пунктами назначения ($T1$, $T2$, $D1$, $D2$ и $D3$). Объемы спроса и предложения, соответствующие этим пунктам отправления и назначения, вычисляются следующим образом.

Объем предложения *истинного пункта отправления* = объем исходного предложения.

Объем предложения *транзитного пункта* = объем исходного предложения + объем буфера.

Объем спроса *истинного пункта назначения* = объем исходного спроса.

Объем спроса *транзитного пункта* = объем исходного спроса + объем буфера.

Объем буфера должен быть таким, чтобы вместить объем всего предложения (или спроса). Обозначим через B объем буфера. Тогда

$$\begin{aligned} B &= \text{Общий объем предложения (или спроса)} = \\ &= 1000 + 1200 = (\text{или} = 800 + 900 + 500) = \\ &= 2200 \text{ автомобилей.} \end{aligned}$$

Построенная транспортная модель, эквивалентная исходной задаче, представлена в табл. 5.44. Решение этой транспортной модели (полученное с помощью программы TORA) показано на рис. 5.6. Отметим "транзитный" эффект решения: дилер *D*2 получает 1400 автомобилей, из них 900 оставляет себе (для удовлетворения своего спроса), а 500 отправляет дилеру *D*3.

Таблица 5.44

	*T*1	*T*2	*D*1	*D*2	*D*3	
*P*1	3	4	*M*	*M*	*M*	**1000**
*P*2	2	5	*M*	*M*	*M*	**1200**
*T*1	0	7	8	6	*M*	***B***
*T*2	*M*	0	*M*	4	9	***B***
*D*1	*M*	*M*	0	5	*M*	***B***
*D*2	*M*	*M*	*M*	0	3	***B***
	B	***B***	**800 + *B***	**900 + *B***	**500**	

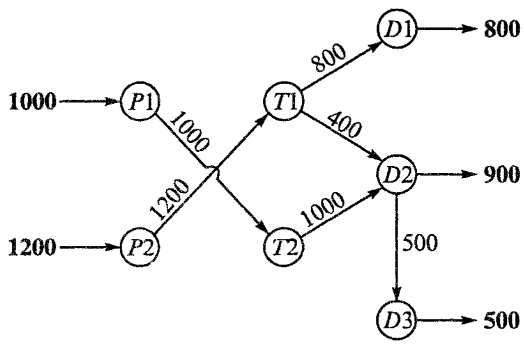


Рис. 5.6

Упражнения 5.5,а

1. В транспортной сети, показанной на рис. 5.7, осуществляются перевозки из пунктов 1 и 2 в пункты 5 и 6 через транзитные пункты 3 и 4. Стоимость перевозок показана на этом же рисунке.

   a) Постройте транспортную модель с промежуточными пунктами и соответствующую ей обычную транспортную модель.

   b) Решите транспортную задачу с помощью программы TORA и покажите на схеме маршруты доставки грузов из пунктов отправления в пункты назначения.

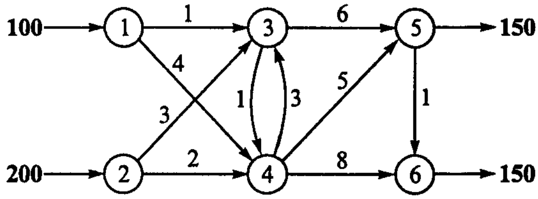


Рис. 5.7

2. Пусть в предыдущей задаче пункты 1 и 2 связаны транспортной магистралью со стоимостью перевозки в $1, а стоимость перевозки из пункта 1 в пункт 3 возросла до $5. Найдите оптимальную схему перевозок.
3. На рис. 5.8 показана транспортная сеть перевозок автомобилей между тремя заводами (пункты 1, 2 и 3) и тремя дилерами (пункты 6, 7 и 8) через два распределительных центра (пункты 4 и 5). Стоимость перевозок (в сотнях долларов) представлена на рисунке возле соответствующих дуг.

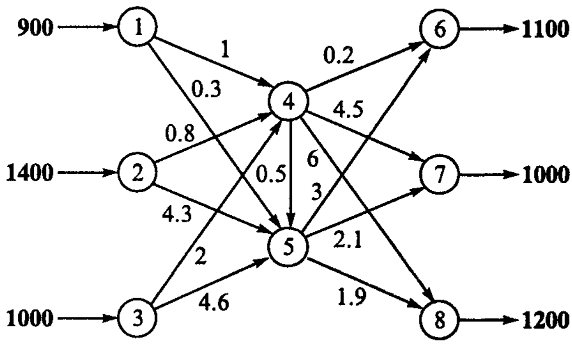


Рис. 5.8

a) Сформулируйте транспортную задачу и найдите ее оптимальное решение с помощью программы TORA.
b) Предположим, распределительный центр (пункт 4) может продать 240 автомобилей самостоятельно. Найдите новое оптимальное решение.

4. Две фабрики снабжают определенной продукцией три магазина. Объемы производства фабрик равны 200 и 300 единиц продукции, а потребности магазинов составляют 100, 200 и 50 единиц продукции. Исследуется возможность перевозок продукции через промежуточные пункты. Основываясь на величинах стоимости перевозок, приведенных в табл. 5.45, с помощью программы TORA найдите оптимальный план перевозок.
5. На рис. 5.9 показана сеть нефтепроводов. Узлы этой сети соответствуют насосным и принимающим станциям. Расстояние (в милях) между станциями приведены на схеме сети. Стоимость транспортировки одного галлона нефти между двумя станциями пропорциональна длине нефтепровода, соединяющего эти станции. Сформулируйте транспортную задачу с промежуточными пунктами и найдите (с помощью программы TORA) ее оптимальное решение для перекачки нефти из пунктов 1 и 3 в пункты 2 и 4.

ТАБЛИЦА 5.45

		Фабрики		Магазины		
		1	2	1	2	3
Фабрики	1	\$0	\$6	\$7	\$8	\$9
	2	\$6	\$0	\$5	\$4	\$3
Магазины	1	\$7	\$2	\$0	\$5	\$1
	2	\$1	\$5	\$1	\$0	\$4
	3	\$8	\$9	\$7	\$6	\$0

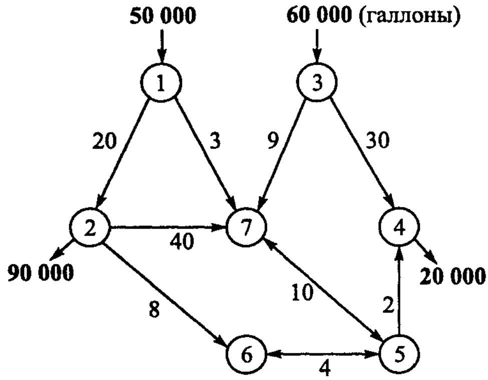


Рис. 5.9

6. *Задача нахождения кратчайшего пути.* Для нахождения кратчайшего пути между пунктами 1 и 7 транспортной сети, показанной на рис. 5.10, сформулируйте транспортную задачу с промежуточными пунктами. Расстояния между промежуточными пунктами показаны на рисунке. (*Совет.* Пусть в пункте 1 есть предложение в одну единицу, а в пункте 7 — спрос такой же величины.)

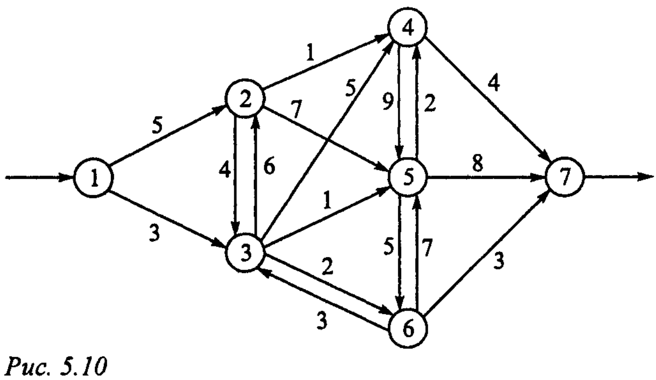


Рис. 5.10

7. В модели примера 5.5–1 обозначим через x_{ij} объем перевозок между пунктами i и j. Задачу этого примера можно сформулировать как задачу линейного программирования, у которой каждому пункту транспортной сети соответствует ограничение в виде равенства. Сформулируйте задачу ЛП и покажите, что в ограничениях коэффициенты a_{ij} при переменных x_{ij} определяются следующим образом.

$$a_{ij} = \begin{cases} 1, & \text{для ограничения } i, \\ -1, & \text{для ограничения } j, \\ 0, & \text{в противном случае.} \end{cases}$$

8. Агентство по трудоустройству имеет заявки на рабочих на следующие пять месяцев.

Месяц	1	2	3	4	5
К-во рабочих	100	120	80	170	50

Стоимость рабочей силы зависит от длительности периода трудоустройства (см. следующую таблицу).

Длительность периода трудоустройства (месяцы)	1	2	3	4	5
Стоимость трудоустройства одного рабочего ($)	100	130	180	220	250

Сформулируйте задачу линейного программирования. Затем путем алгебраических преобразований ограничений покажите, что эту задачу можно свести к транспортной. Найдите ее оптимальное решение. (*Совет.* Для преобразования общей задачи ЛП в транспортную, используйте представление коэффициентов ограничений из предыдущего упражнения.)

5.6. Заключение

Транспортная задача представляет собой частный случай общей задачи линейного программирования, специфическая структура которой позволяет разработать эффективные вычислительные методы, основанные на теории двойственности. Частными случаями транспортной задачи являются задача о назначениях и транспортная задача с промежуточными пунктами. Транспортная модель используется для описания проблем, не связанных с транспортировкой, например в задачах управления запасами и задачах производственного планирования.

Транспортные модели составляют один подкласс обобщенных сетевых моделей, рассмотренных в главе 6. Некоторые последние обзоры показывают, что более 70% практических моделей линейного программирования можно рассматривать как сетевые либо как модели, связанные с сетевыми.

Литература

1. Bazaraa M., Jarvis J., Sherali M. *Linear Programming and Network Flows*, 2nd ed., Wiley, New York, 1990.
2. Chanrnes A., Glover F., Klingman D. *A Note on a Distribution Problem*, Operation Reseach, Vol. 18, pp. 1213-1216, 1970.
3. Dantzig G. *Linear Programming and Extensions*, Princeton University Press, Princeton, N.J., 1963. (Русский перевод: Данциг Дж. *Линейное программирование, его применение и обобщение.* — М.: Прогресс, 1966.)
4. Murty K. *Network Programming*, Prentice Hall, Upper Saddle River, N.J., 1992.

Литература, добавленная при переводе

Ашманов С.А. *Линейное программирование.* — М.: Наука, 1981.

Гольштейн Е.Г., Юдин Д.Б. *Задачи линейного программирования транспортного типа.* — М.: Наука, 1969.

Гольштейн Е.Г., Юдин Д.Б. *Линейное программирование: Теория, методы и приложения.* — М.: Наука, 1969.

Комплексные задачи

■ **5-1.**[1] Компания ABC Cola имеет завод по производству безалкогольных напитков в северной части островного государства Таванда. Напитки выпускаются в трех различных упаковках: возвращаемых стеклянных бутылках, алюминиевых консервных банках и невозвращаемых пластмассовых бутылках. Возвращаемые стеклянные бутылки складируют и затем возвращают на завод для повторного использования.

Поскольку спрос на безалкогольные напитки постоянно растет, компания ABC Cola планирует построить еще один завод в центральной или южной части острова. Прогнозируемый спрос на безалкогольные напитки (выраженный в количестве упаковок) на следующие пять лет показан в табл. 5.46.

Таблица 5.46

Вид упаковки	Года				
	1	2	3	4	5
Возвращаемая стеклянная	2400	2450	2600	2800	3100
Банки	1750	2000	2300	2650	3050
Невозвращаемая пластмассовая	490	550	600	650	720

Плановые возможности существующего завода по выпуску продукции показаны в табл. 5.47.

Таблица 5.47

Вид упаковки	Года				
	1	2	3	4	5
Возвращаемая стеклянная	1800	1400	1900	2050	2150
Банки	1250	1350	1400	1500	1800
Невозвращаемая пластмассовая	350	380	400	400	450

Компания имеет 6 складов для стеклотары: N1 и N2 расположены в северной части острова, C1 и C2 — в центральной части, S1 и S2 — в южной. Доля стеклотары, обрабатываемой каждым складом, показана в табл. 5.48.

Из общего количества стеклотары примерно 60% приходится на северные склады, 15% — на центральные и 25% — на южные.

---

[1] Задача построена на материалах статьи Cheng T., Chiu C. "A Case Study of Production Expansion Planning in a Soft-Drink Manufacturing Company", *Omega*, Vol. 16, No. 6, pp. 521–532, 1988.

Таблица 5.48

Склад	Доля (в процентах)
N1	85
N2	15
C1	60
C2	40
S1	80
S2	20

Компания планирует построить новый завод по производству безалкогольных напитков либо в центре, либо на юге острова. Транспортные расходы, рассчитанные на одну упаковку напитка в стеклянных бутылках, приведены в табл. 5.49. Транспортные расходы на напитки в банках и пластмассовых бутылках составляют соответственно 60% и 70% от аналогичных расходов на транспортировку напитков в стеклянных бутылках.

Таблица 5.49

	Транспортные расходы на одну упаковку ($)		
Склад	Существующий завод	Завод в центральном районе	Завод в южном районе
N1	0.80	1.30	1.90
N2	1.20	1.90	2.90
C1	1.50	1.05	1.20
C2	1.60	0.80	1.60
S1	1.90	1.50	0.90
S2	2.10	1.70	0.80

Определите, где целесообразнее построить новый завод: в центральном районе или на юге острова?

■ **5-2.**[1] В ходе строительства международного аэропорта в г. Брисбен для укрепления болотистого грунта необходимо произвести песчаную высыпку общим объемом 1 355 000 м^3 на девяти площадках аэропорта. Песок берется из расположенных вблизи аэропорта пяти небольших карьеров. Некоторые строительные подразделения, обслуживающие стройплощадки, выделены для прокладки дорог внутри аэропорта и по его периметру. Для отсыпки этих дорог песок берется со стройплощадок. Расстояние (в сотнях метров) от карьеров до стройплощадок показано в табл. 5.50, где также приведены потребности в песке стройплощадок и возможности карьеров (в 100 м^3).

a) Управление строительством оценивает объем перемещения песка [объем (м^3) × расстояние (100 м)] в 2 495 000 соответствующих единиц при стоимости $0.65 за единицу. Достаточно ли приведенного объема песка для выполнения запланированных работ?

b) Управление строительством пришло к выводу, что подвоз песка к некоторым стройплощадкам не возможен до завершения строительства дороги вокруг аэро-

---

[1] Задача построена на материалах статьи Perry C., Ilief M. "Earth Moving on Construction Projects", *Interfaces*, Vol. 13, No. 1, pp. 79–84, 1983.

порта (длиной 900 м). В табл. 5.51 значком "x" обозначены стройплощадки, к которым не возможен подвоз песка (из определенных карьеров) до завершения строительства окружной дороги. Как организовать подвоз песка с учетом этого ограничения?

ТАБЛИЦА 5.50

	1	2	3	4	5	6	7	8	9	Возможности карьеров
1	22	26	12	10	18	18	11	8.5	20	**960**
2	20	28	14	12	20	20	13	10	22	**201**
3	16	20	26	20	1.5	28	6	22	18	**71**
4	20	22	26	22	6	∞	2	21	18	**24**
5	22	26	10	4	16	∞	24	14	21	**99**
Потребности стройплощадок	**62**	**217**	**444**	**315**	**50**	**7**	**20**	**90**	**150**	

ТАБЛИЦА 5.51

	1	2	3	4	5	6	7	8	9
1	x	x			x				
2	x	x			x				
3			x			x			
4			x					x	
5	x	x			x		x		

■ **5-3.** Десять лет назад некий предприниматель организовал доставку оптовых фармацевтических товаров с центрального склада (ЦС) на автофургонах. С тех пор значительно увеличился спрос на фармацевтическую продукцию и появились два новых товарных склада (С1 и С2). Центральный склад, традиционно имеющий больший ассортимент товаров, время от времени выручал новые склады в обеспечении кратковременных заказов. Такая взаимовыручка со временем привела к тому, что треть своих заказов новые склады выполняли за счет товаров, полученных с центрального склада. В табл. 5.52 приведено количество заказов, выполняемых тремя товарными складами для шести потребителей, обозначенных П1–П6.

Предприниматель постоянно совершенствовал способы доставки заказов, придерживаясь принципа децентрализации, когда каждому товарному складу была определена своя зона доставки товара, покрывающая всех потребителей, получающих заказы с этого склада. Но управляющие складами комплектовали заказы согласно своим "сферам влияния". Например, менеджеры центрального склада гордились тем, что поставляют товары не только обычным потребителям, но и другим товарным складам. Поэтому зачастую разные склады поставляют заказы одному и тому же потребителю.

Таблица 5.52

Маршрут		
Откуда	Куда	Количество заказов
ЦС	С1	2000
ЦС	С2	1500
ЦС	П1	4800
ЦС	П2	3000
ЦС	П3	1200
С1	П1	1000
С1	П3	1100
С1	П4	1500
С1	П5	1800
С2	П2	1900
С2	П5	600
С2	П6	2200

В табл. 5.53 приведены расстояния (в милях) между складами и потребителями. В автофургон обычно помещается 100 товарных заказов.

Таблица 5.53

	ЦС	С1	С2	П1	П2	П3	П4	П5	П6
ЦС	0	5	45	50	30	30	60	75	80
С1	5	0	80	38	70	30	8	10	60
С2	45	80	0	85	35	60	55	7	90
П1	50	38	85	0	20	40	25	30	70
П2	30	70	35	20	0	40	90	15	10
П3	30	30	60	40	40	0	10	6	90
П4	60	8	55	25	90	10	0	80	40
П5	75	10	7	30	15	6	80	0	15
П6	80	60	90	70	10	90	40	15	0

Оцените организацию перевозок предпринимателя и предложите свой план перевозок.

■ **5-4.** Авиакомпания KeeWee выполняет полеты между городами A и B согласно расписанию, приведенному в табл. 5.54. Экипаж может вернуться в свой город не ранее чем через 90 минут после прилета либо на следующий день. Составьте расписание назначения экипажей на рейсы, минимизирующее суммарное время простоев всех экипажей.

Таблица 5.54

Рейс	Отлет из города A	Прилет в город B	Рейс	Отлет из города B	Прилет в город A
A1	6:00	8:30	B1	7:30	9:30
A2	8:15	10:45	B2	9:15	11:15
A3	13:30	16:00	B3	16:30	18:30
A4	15:00	17:30	B4	20:00	22:00

Сетевые модели

6.1. Обзор применения сетевых моделей

В рамках теории исследования операций рассматривается большое количество практических задач, которые можно сформулировать и решить как сетевые модели. Приведем несколько конкретных примеров.

1. Проектирование газопровода, соединяющего буровые скважины морского базирования с находящейся на берегу приемной станцией. Целевая функция соответствующей модели должна минимизировать стоимость строительства газопровода.
2. Нахождение кратчайшего маршрута между двумя городами по существующей сети дорог.
3. Определение максимальной пропускной способности трубопровода для транспортировки угольной пульпы от угольных шахт к электростанциям.
4. Определение схемы транспортировки нефти от пунктов нефтедобычи к нефтеперерабатывающим заводам с минимальной стоимостью транспортировки.
5. Составление временн*о*го графика строительных работ (определение дат начала и завершения отдельных этапов работ).

Решение приведенных задач (как и многих других подобных задач) требует применения различных сетевых оптимизационных алгоритмов. В этой главе будут рассмотрены следующие пять алгоритмов.

1. Алгоритм нахождения минимального остовного дерева (пример 1).
2. Алгоритм нахождения кратчайшего пути (пример 2).
3. Алгоритм определения максимального потока (пример 3).
4. Алгоритм минимизации стоимости потока в сети с ограниченной пропускной способностью (пример 4).
5. Алгоритм нахождения критического пути (пример 5).

Задачи, вытекающие из перечисленных примеров, можно сформулировать и решать как задачи линейного программирования. Однако специфическая структура этих задач позволяет разработать специальные сетевые алгоритмы, более эффективные, чем стандартный симплекс-метод.

6.2. Основные определения

Сеть состоит из множества **узлов**, связанных **дугами** (или **ребрами**).[1] Таким образом, сеть описывается парой множеств (N, A), где N — множество узлов, а A — множество ребер. Например, сеть, показанная на рис. 6.1, описывается следующим образом.

$$N = \{1, 2, 3, 4, 5\},$$
$$A = \{(1, 3), (1, 2), (2, 3), (2, 4), (2, 5), (3, 4), (3, 5), (4, 5)\}.$$

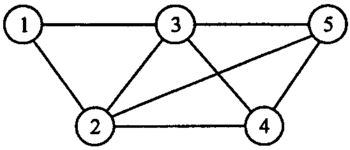


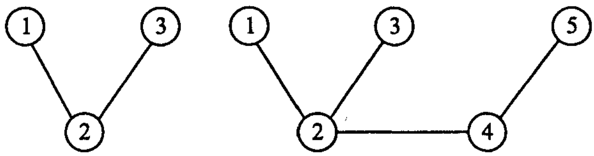
Рис. 6.1

С каждым типом сети связан определенный тип потоков (например, транспортный поток нефти в нефтепроводах или автомобильные потоки в сети городских дорог). В общем случае потоки в сети ограничены пропускной способностью ее ребер, которая может быть как конечной, так и бесконечной.

Ребро называется **направленным**, или **ориентированным** (и в этом случае ребро будем называть **дугой**), если в одном направлении возможен только положительный поток, а в противоположном — только нулевой. В **ориентированной сети** все ребра ориентированны.

Путем называется последовательность различных ребер, соединяющих два узла, независимо от направления потока в каждом ребре. Путь формирует **цикл**, если начальный и конечный узлы совпадают. Например, на рис. 6.1 ребра (2, 3), (3, 4) и (4, 2) составляют цикл. **Ориентированный цикл** — это цикл, в котором все дуги ориентированы в определенном направлении.

Связная сеть — такая сеть, у которой любые два узла связаны по крайней мере одним путем. На рис. 6.1 показан именно такой тип сети. **Деревом** называется связная сеть, содержащая подмножество узлов исходной сети и не имеющая циклов. **Остовное дерево** — это дерево, содержащее *все* узлы сети. На рис. 6.2 показаны дерево и остовное дерево для сети из рис. 6.1.



Дерево　　　　Остовное дерево

Рис. 6.2

[1] Чтобы согласовать используемую здесь терминологию с терминологией теории графов, составляющей основу рассматриваемых сетей, будем называть дугами только *ориентированные* ребра. Термин "ребро" (как более общее понятие) часто будем употреблять для обозначения как ребер, так и дуг. Также отметим очевидное соответствие между узлом сети и вершиной графа. — *Прим. ред*

Упражнения 6.2,а

1. Для каждой сети, показанной на рис. 6.3, определите a) путь, b) цикл, c) ориентированный цикл, d) дерево, e) остовное дерево.

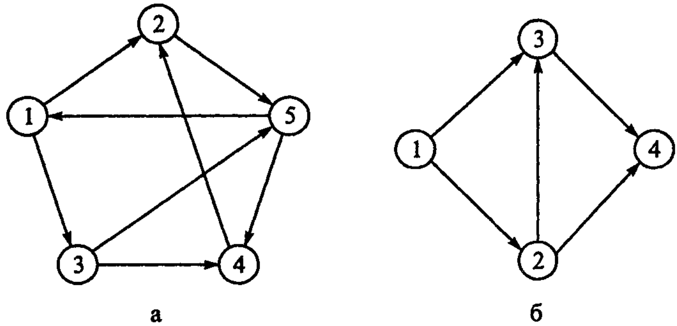


Рис. 6.3

2. Определите множества N и A для сетей, представленных на рис. 6.3.
3. Нарисуйте сеть, заданную множествами

$$N = \{1, 2, 3, 4, 5, 6\},$$
$$A = \{(1, 2), (1, 5), (2, 3), (2, 4), (3, 5), (3, 4), (4, 2), (4, 6), (5, 2), (5, 6)\}.$$

4. Дано восемь равных квадратиков, расположенных в три линии: на первой находится два квадратика, на второй — четыре и на третьей — два. Квадратики на каждой линии располагаются симметрично вертикальной оси. Необходимо приписать каждому квадратику число от 1 до 8 так, чтобы *смежные* (по вертикали, горизонтали или диагонали) квадратики не имели последовательных номеров. Сформулируйте эту задачу как сетевую, и на основе сетевого представления задачи разработайте алгоритм нахождения симметричного решения.
5. Троих заключенных в сопровождении трех стражников необходимо переправить на лодке из Сан-Франциско в тюрьму на острове Алкатраз для исполнения приговора. Лодка не может перевести более двух человек одновременно. Заключенные определенно сильнее стражников, если их количество в какой-то момент превысит количество стражников. Разработайте сетевую модель, позволяющую найти такую последовательность перевозки заключенных, чтобы не возникло каких-либо инцидентов между заключенными и стражниками.

6.3. Алгоритм построения минимального остовного дерева

Алгоритм построения минимального остовного дерева предполагает соединение всех узлов сети с помощью путей наименьшей длины. Типичной задачей, для решения которой необходим такой алгоритм, является создание (проектирование) сети дорог с твердым покрытием, соединяющих населенные пункты в сельской местности, где дороги, соединяющие два каких-либо пункта, могут проходить через другие населенные пункты. Наиболее экономичный проект дорожной системы должен минимизировать общую длину дорог с твердым покрытием, при этом желаемый результат можно получить путем применения алгоритма построения минимального остовного дерева.

Опишем процедуру выполнения этого алгоритма. Обозначим через $N = \{1, 2, ..., n\}$ множество узлов сети и введем новые обозначения:

C_k — множество узлов сети, соединенных алгоритмом после выполнения k-й итерации этого алгоритма,

$\bar{C}_k$ — множество узлов сети, не соединенных с узлами множества C_k после выполнения k-й итерации этого алгоритма.

Шаг 0. Полагаем $C_0 = \varnothing$ и $\bar{C}_0 = N$.

Шаг 1. Выбираем *любой* узел i из множества $\bar{C}_0$ и определяем $C_1 = \{i\}$, тогда $\bar{C}_1 = N - \{i\}$. Полагаем $k = 2$.

Основной шаг k. В множестве $\bar{C}_{k-1}$ выбираем узел j^*, который соединен самой короткой дугой с каким-либо узлом из множества C_{k-1}. Узел j^* присоединяется к множеству C_{k-1} и удаляется из множества $\bar{C}_{k-1}$. Таким образом, $C_k = C_{k-1} + \{j^*\}$, $\bar{C}_k = \bar{C}_{k-1} - \{j^*\}$.

Если множество $\bar{C}_k$ пусто, то выполнение алгоритма заканчивается. В противном случае полагаем $k = k + 1$ и повторяем последний шаг.

Пример 6.3–1

Телевизионная компания планирует подключение к своей кабельной сети пяти новых районов. На рис. 6.4 показана структура планируемой сети и расстояния (в милях) между районами и телецентром. Необходимо спланировать наиболее экономичную кабельную сеть.

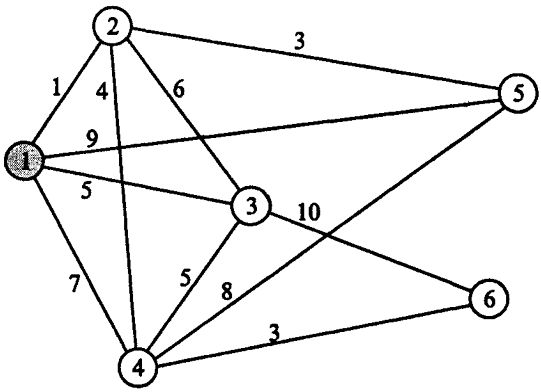


Рис. 6.4

Начнем выполнение алгоритма построения минимального остовного дерева с выбора узла 1 (или любого другого узла). Тогда

$$C_1 = \{1\} \text{ и } \bar{C}_1 = \{2, 3, 4, 5, 6\}.$$

Последовательные итерации выполнения алгоритма представлены на рис. 6.5. Здесь тонкими линиями показаны ребра, соединяющие узлы, принадлежащие множествам C_k и $\overline{C}_k$, среди которых ищется ребро с минимальной стоимостью (длиной). Это найденное ребро показано пунктирной линией. Толстыми сплошными линиями обозначены ребра, соединяющие узлы множества C_k (и которые ранее обозначались пунктирными линиями).

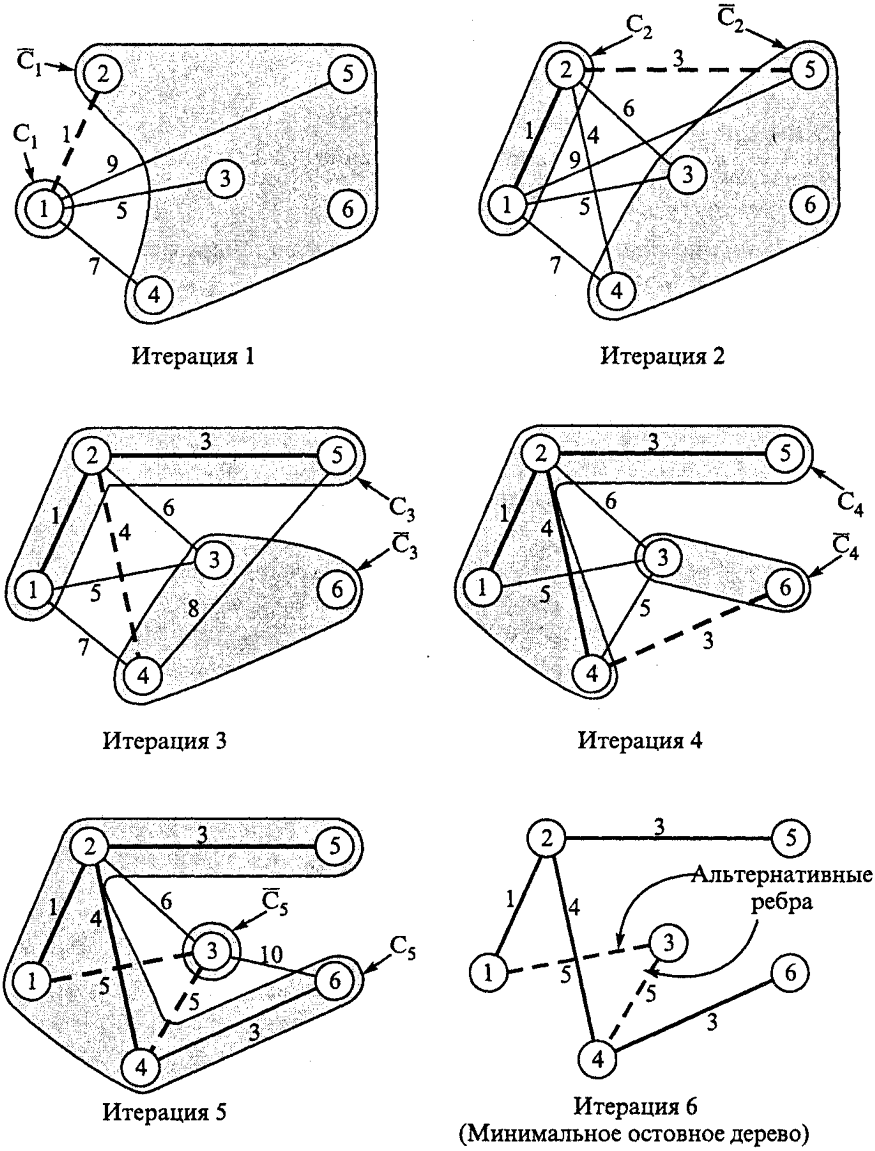


Рис. 6.5

Например, на первой итерации ребро (1, 2) имеет наименьшую стоимость (т.е. наименьшее расстояние между пунктами сети) среди всех других ребер, соединяющих узел 1 с узлами множества $\bar{C}_1$ (отметим, что узел 6 не имеет ребра, непосредственно соединяющего его с узлом 1). Поэтому $j^* = 2$ и $C_2 = \{1, 2\}$, $\bar{C}_2 = \{3, 4, 5, 6\}$.

Решение в виде минимального остовного дерева получено на 6-й итерации (рис. 6.5). Минимальная длина кабеля для построения такой сети равна 1 + 3 + 4 + 3 + 5 = 16 милям.

Упражнения 6.3,а

1. Решите задачу из примера 6.3–1, начиная с узла 5 (вместо узла 1), и убедитесь, что будет получено то же самое решение.
2. Найдите минимальное остовное дерево для сети из примера 6.3–1 при выполнении каждого из следующих условий в отдельности.
   a) Узлы 5 и 6 связаны 2-мильным кабелем.
   b) Узлы 2 и 5 не связаны.
   c) Узлы 2 и 6 связаны 4-мильным кабелем.
   d) Узлы 1 и 2 связаны кабелем длиной 8 миль.
   e) Узлы 3 и 5 связаны кабелем длиной 2 мили.
   f) Узел 2 не связан непосредственно с узлами 3 и 5.
3. В модульных перевозках груженые трейлерные платформы перевозятся по железной дороге между специальными перевалочными железнодорожными терминалами, где платформы снова присоединяются к трейлерам и далее следуют к потребителям автомобильным ходом. На рис. 6.6 показаны основные железнодорожные терминалы Соединенных Штатов и существующие железнодорожные пути между ними. Выделите сегменты железных дорог так, чтобы были связаны все железнодорожные терминалы и была минимизирована суммарная стоимость перевозок трейлерных платформ (стоимость перевозок пропорциональна длине железнодорожных путей).

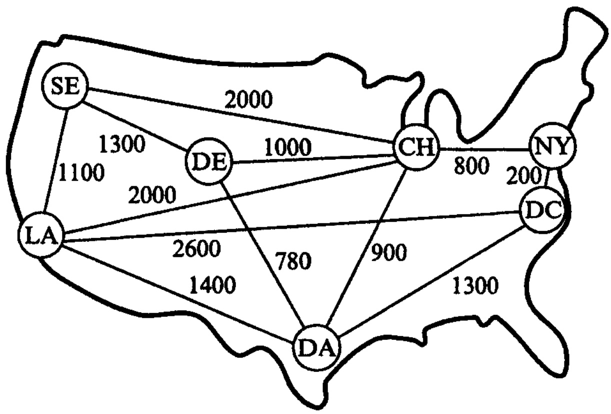


Рис. 6.6

4. На рис. 6.7 показаны расстояния между платформами, добывающими газ в открытом море, и приемным пунктом, расположенным на берегу. Поскольку платформа 1 ближе остальных к берегу, она оснащена необходимым оборудованием для перекачки газа от остальных платформ к приемному пункту. Спроектируйте сеть трубопроводов минимальной длины, соединяющую приемный пункт со всеми добывающими платформами.

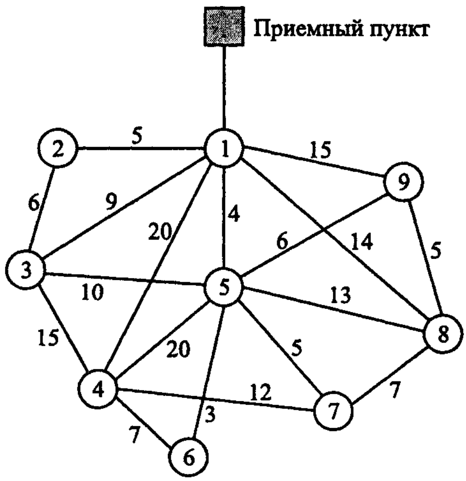


Рис. 6.7

5. Предположим, что в предыдущей задаче (рис. 6.7) все добывающие платформы разбиты на две группы в зависимости от давления газа в скважинах: к группе с высоким давлением газа относятся платформы 2, 3, 4 и 6, с низким давлением — 5, 7, 8 и 9. Из-за разницы в давлении газопроводы от платформ разных групп нельзя соединять между собой. В то же время газопроводы от этих групп могут подсоединяться к приемному пункту через платформу 1. Определите минимальную сеть трубопроводов для данной ситуации.

6. Компания производит 15 типов изделий на 10 станках. Компания планирует сгруппировать станки так, чтобы минимизировать "несходство" операций, выполняемых на каждой группе станков. Мерой "несходства" между станками i и j служит величина d_{ij}, вычисляемая по формуле

$$d_{ij} = 1 - \frac{n_{ij}}{n_{ij} + m_{ij}},$$

где n_{ij} — количество изделий, обрабатываемых как на станке i, так и на станке j, m_{ij} — количество изделий, обрабатываемых только на станке i или только на станке j.

В следующей таблице показано, изделия каких типов на каких станках обрабатываются.

Станок	Типы изделий
1	1, 6
2	2, 3, 7, 8, 9, 12, 13, 15
3	3, 5, 10, 14

Станок	Типы изделий
4	2, 7, 8, 11, 12, 13
5	3, 5, 10, 11, 14
6	1, 4, 5, 9, 10
7	2, 5, 7, 9, 10
8	3, 4, 15
9	4, 10
10	3, 8, 10, 14, 15

a) Сформулируйте данную задачу как сетевую.

b) Покажите, что разбиение множества станков на группы можно получить как решение задачи нахождения минимального остовного дерева.

c) Решите данную задачу для множества разбиения станков на две и три группы.

6.4. Задача нахождения кратчайшего пути

Данная задача состоит в определении в транспортной сети кратчайшего пути между заданными исходным пунктом и пунктом назначения. Такую модель можно использовать для описания разнообразных ситуаций, как показано в следующем разделе.

6.4.1. Практические примеры задачи нахождения кратчайшего пути

Пример 6.4–1. (Замена оборудования)

Компания по прокату автомобилей разрабатывает план по обновлению парка своих машин на следующие пять лет (2000–2004 гг.). Каждый автомобиль должен проработать не менее одного и не более трех лет. В следующей таблице приведена стоимость замены автомобиля в зависимости от года покупки и срока эксплуатации.

	Стоимость замены ($) в зависимости от срока эксплуатации		
Год покупки	1	2	3
2000	4000	5400	9800
2001	4300	6200	8700
2002	4800	7100	—
2003	4900	—	—

Задачу можно сформулировать как сетевую с пятью узлами с номерами от 1 до 5, соответствующими годам 2000–2004. Из узла 1 (2000 год) дуги идут только к узлам 2, 3 и 4, поскольку автомобиль может эксплуатироваться не менее одного и не более трех лет. Дуги из других узлов интерпретируются аналогично. Стоимости дуг равны стоимостям замены автомобилей. Решение задачи эквивалентно нахождению кратчайшего пути между узлами 1 и 5.

На рис. 6.8 показана построенная сеть. С помощью программы TORA находим кратчайший путь 1→3→5 (показан на рис. 6.8 жирными линиями) с суммарной стоимостью $12 500. Это решение означает, что автомобили, приобретенные в 2000 году (узел 1), будут эксплуатироваться 2 года, до 2002 года (узел 3), затем они будут заменены на новые, которые будут эксплуатироваться до конца 2004 года. Общая стоимость замены составит $5400 + $7100 = $12 500.

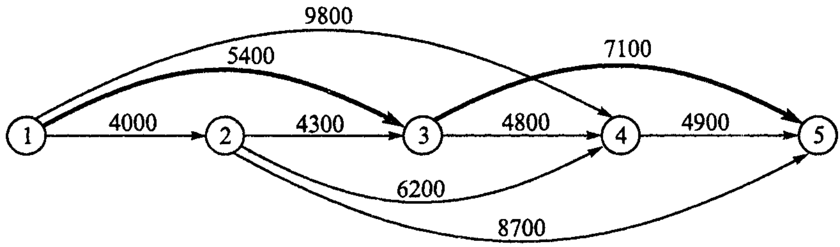


Рис. 6.8

Пример 6.4–2. (Самый надежный маршрут)

М-р Разумник ежедневно ездит на работу на автомобиле. Закончив в свое время полный курс по теории исследования операций, он легко определил самый короткий путь от дома до работы. К сожалению, данный маршрут усиленно патрулируется нарядами полиции, и автомобиль Разумника часто останавливают за превышение скорости (как ему кажется, не обоснованно). Таким образом, самый короткий путь оказался не самым быстрым. Поэтому м-р Разумник планирует разработать новый маршрут, на котором он имел бы самую высокую вероятность *не быть* остановленным полицией.

Схема сети дорог, по которой м-р Разумник может добраться от дома до работы, показана на рис. 6.9. На этой же схеме приведены вероятности *не быть остановленным* для каждого сегмента сети дорог. Вероятность не быть остановленным на всем пути следования автомобиля Разумника равна произведению вероятностей не быть остановленным на каждом сегменте выбранного пути. Например, вероятность не быть остановленным на маршруте 1→3→5→7 равна $0.9 \times 0.3 \times 0.25 = 0.0675$. Таким образом, м-ру Разумнику необходимо решить задачу выбора маршрута, который *максимизировал* бы вероятность не быть остановленным.

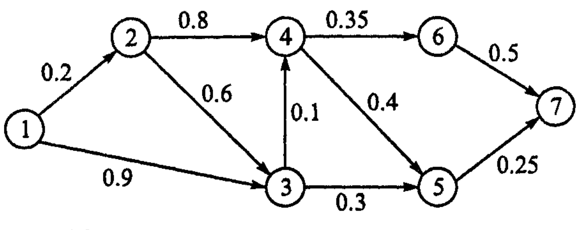


Рис. 6.9

Эту задачу можно сформулировать как задачу нахождения кратчайшего пути, если вместо вероятностей использовать логарифмы вероятностей. Тогда произведение вероятностей преобразуется в сумму логарифмов вероятностей: если $p_{1k} = p_1 \times p_2 \times \ldots \times p_k$ — вероятность не быть остановленным на маршруте $1 \to 2 \to \ldots \to k$, тогда

$$\log p_{1k} = \log p_1 + \log p_2 + \ldots + \log p_k.$$

С точки зрения математики задача максимизации вероятности p_{1k} эквивалентна задаче максимизации величины $\log p_{1k}$. Поскольку $\log p_{1k} \le 0$, задача максимизации величины $\log p_{1k}$ эквивалентна задаче минимизации $-\log p_{1k}$. Заменив на рис. 6.9 вероятности p_k на величины $-\log p_k$, получаем сеть (рис. 6.10), к которой можно применить алгоритм нахождения кратчайшего пути.

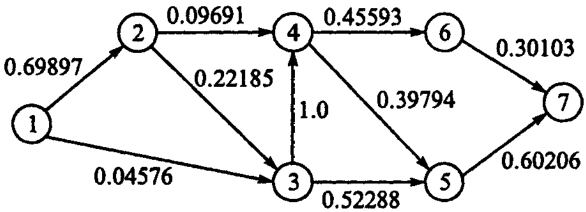


Рис. 6.10

С помощью программы TORA находим кратчайший путь для полученной сети: $1 \to 3 \to 5 \to 7$ с соответствующей “длиной” пути 1.1707 ($= -\log p_{17}$). Таким образом, максимальная вероятность не быть остановленным равна $p_{17} = 0.0675$.

Пример 6.4–3. (Головоломка о трех бидонах)

8-литровый бидон заполнен некой жидкостью и имеется еще два пустых бидона емкостью 5 и 3 литра. Необходимо разделить 8 литров жидкости на две равные части, используя только три имеющихся бидона. Какое минимальное количество переливаний из бидона в бидон надо сделать, чтобы достигнуть желаемого результата?

Вы, вероятно, уже нашли решение этой головоломки. Вместе с тем ее решение можно получить, как решение задачи о нахождении кратчайшего пути.

В этой сетевой модели каждый узел будет соответствовать объемам жидкости в 8-, 5- и 3-литровом бидонах. Начальным узлом будет (8, 0, 0), а конечным — (4, 4, 0). Новый узел получается из текущего при однократном переливании жидкости из одного бидона в другой.

На рис. 6.11 показаны различные маршруты, ведущие от начального узла (8, 0, 0) к конечному (4, 4, 0). Таким образом, наша головоломка сведена к задаче нахождения кратчайшего пути между узлами (8, 0, 0) и (4, 4, 0).

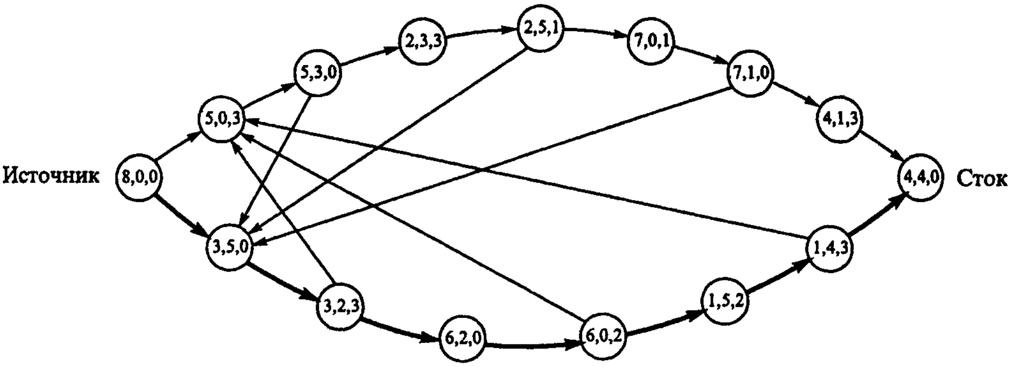


Рис. 6.11

Оптимальное решение, показанное в нижней части рис. 6.11, требует семи переливаний из бидона в бидон.

Упражнения 6.4,а

1. Создайте заново модель замены оборудования из примера 6.4–1 в предположении, что автомобиль до замены должен эксплуатироваться не менее 2-х и не более 4-х лет. Стоимость замены автомобилей в 2000–2005 годах приведена в следующей таблице.

Год покупки	Стоимость замены ($) в зависимости от срока эксплуатации: 1	2	3
2000	3800	4100	6800
2001	4000	4800	7000
2002	4200	5100	7200
2003	4800	5700	—
2004	5300	—	—

2. На рис. 6.12 показана коммуникационная сеть между двумя приемно-передающими станциями 1 и 7. Возле каждой дуги этой сети указаны вероятности передачи сообщений без потерь по этим дугам. Необходимо найти маршрут от станции 1 к станции 7 с максимальной вероятностью успешной передачи сообщений. Сформулируйте эту задачу как нахождение кратчайшего пути и решите ее с помощью программы TORA.

3. Тостер старой конструкции имеет две подпружиненные дверцы на петлях, размещенные с обеих сторон тостера. Ломтик хлеба помещается в тостер с одной стороны и дверца закрывается, затем другой ломтик хлеба загружается в тостер с противоположной стороны. После того как одна сторона ломтика хлеба подрумянится, он переворачивается. Задача заключается в определении последовательности операций (помещение ломтика хлеба, поджаривание одной стороны, переворачивание ломтика в тостере и извлечение готового хлеба из тостера), необходимых для поджаривания трех ломтиков хлеба за минимально возможное время. Сформулируйте эту задачу как нахождение кратчайшего пути, используя следующие данные о времени выполнения различных операций.

Операция	Время (секунды)
Помещение ломтика хлеба в тостер	3
Поджаривание одной стороны ломтика	30
Переворачивание ломтика в тостере	1
Извлечение ломтика из тостера	3

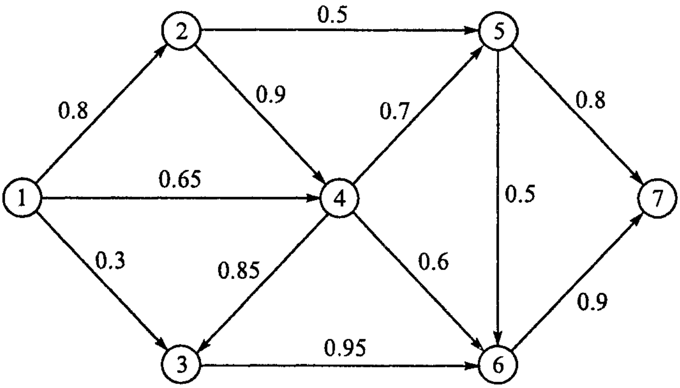


Рис. 6.12

4. *Планирование производства.* Компания продает некую продукцию, спрос на которую в следующие 4 месяца составит 100, 140, 210 и 180 единиц соответственно. Компания может удовлетворить любой помесячный спрос на продукцию и даже спрос на два и более месяца вперед. В последнем случае необходимо платить \$1.20 за хранение единицы избыточно произведенной продукции в течение месяца. Покупная цена единицы продукции в следующие 4 месяца будет равна \$15, \$12, \$10 и \$14 соответственно. Стоимость переналадки оборудования для выполнения заказа составляет \$200. Компания хочет разработать такой производственный план, чтобы минимизировать расходы на выполнение заказов, покупку продукции и ее хранение. Сформулируйте задачу как нахождение кратчайшего пути и найдите ее оптимальное решение с помощью программы TORA.

5. *Задача о рюкзаке.* Путешественник, собираясь в путь, пытается поместить в свой рюкзак, объемом 5 кубических футов, наиболее необходимые в путешествии вещи. Имеется три вещи, объемом соответственно 2, 3 и 4 кубических фута, необходимость в которых оценивается (по 100-балльной шкале) в 30, 50 и 70 баллов. Сформулируйте эту задачу как сетевую, где необходимо определить самый длинный путь, и найдите ее оптимальное решение. (*Совет.* Узел в этой сети можно определить как пару [i, v], где i — номер выбираемой вещи, а v — свободный объем рюкзака, оставшийся после выбора i-й вещи.)

6.4.2. Алгоритм нахождения кратчайшего пути

В этом разделе представлены два алгоритма для решения задачи нахождения кратчайшего пути как в сетях, имеющих циклы, так и в сетях, не имеющих циклов.

1. Алгоритм Дейкстры
2. Алгоритм Флойда

Алгоритм Дейкстры разработан для нахождения кратчайшего пути между заданным исходным узлом и любым другим узлом сети. Алгоритм Флойда более общий, поскольку он позволяет одновременно найти минимальные пути между *любыми* двумя узлами сети.

Алгоритм Дейкстры. В процессе выполнения этого алгоритма при переходе от узла i к следующему узлу j используется специальная процедура пометки ребер. Обозначим через u_i кратчайшее расстояние от исходного узла 1 до узла i, через d_{ij} — длину ребра (i, j). Тогда для узла j определим метку $[u_j, i]$ следующим образом.

$$[u_j, i] = [u_i + d_{ij}, i], \quad d_{ij} \geq 0$$

Метки узлов в алгоритме Дейкстры могут быть двух типов: *временные* и *постоянные*. Временная метка впоследствии может быть заменена на другую временную, если будет найден более короткий путь к данному узлу. Когда же станет очевидным, что не существует более короткого пути от исходного узла к данному, статус временной метки изменяется на постоянный.

Вычислительная схема алгоритма состоит из следующих шагов.

Шаг 0. Исходному узлу (узел 1) присваивается метка [0, —]. Полагаем $i = 1$.

Шаг i.

a) Вычисляются временные метки $[u_i + d_{ij}, i]$ для всех узлов j, которые можно достичь непосредственно из узла i и *которые не имеют постоянных меток*. Если узел j уже имеет метку $[u_j, k]$, полученную от другого узла k, и если $u_i + d_{ij} < u_j$, тогда метка $[u_j, k]$ заменяется на $[u_i + d_{ij}, i]$.

b) Если *все* узлы имеют постоянные метки, процесс вычислений заканчивается. В противном случае выбирается метка $[u_r, s]$ с наименьшим значением расстояния u_r среди всех *временных* меток (если таких меток несколько, то выбор произволен). Полагаем $i = r$ и повторяем шаг i.

Пример 6.4–4

На рис. 6.13 показана транспортная сеть, состоящая из пяти городов (расстояния между городами (в милях) приведены возле соответствующих дуг сети). Необходимо найти кратчайшие расстояния от города 1 (узел 1) до всех остальных четырех городов.

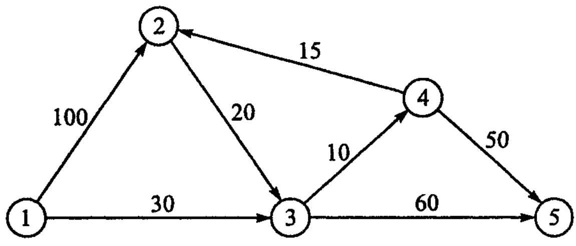


Рис. 6.13

Шаг 0. Назначаем узлу 1 *постоянную* метку [0, —].

Шаг 1. Из узла 1 можно достичь узлов 2 и 3. Вычисляем метки для этих узлов, в результате получаем следующую таблицу меток.

Узел	Метка		Статус метки
1	**[0, —]**		**Постоянная**
2	[0 + 100, 1] = [100, 1]		Временная
3	[0 + 30, 1] = [30, 1]	←	Временная

Среди узлов 2 и 3 узел 3 имеет наименьшее значение расстояния ($u_3 = 30$). Поэтому статус метки этого узла изменяется на "постоянная".

Шаг 2. Из узла 3 (последнего узла с постоянной меткой) можно попасть в узлы 4 и 5. Получаем следующий список узлов.

Узел	Метка		Статус метки
1	**[0, —]**		**Постоянная**
2	[100, 1]		Временная
3	**[30, 1]**		**Постоянная**
4	[30 + 10, 3] = [40, 3]	←	Временная
5	[30 + 60, 3] = [90, 3]		Временная

Временный статус метки [40, 3] узла 4 заменяется на постоянный ($u_4 = 40$).

Шаг 3. Из узла 4 можно достичь узлов 2 и 5. После вычисления меток получим следующий их список.

Узел	Метка		Статус метки
1	**[0, —]**		**Постоянная**
2	[40 + 15, 4] = [55, 4]	←	Временная
3	**[30, 1]**		**Постоянная**
4	**[40, 3]**		**Постоянная**
5	[90, 3] или [40 + 50, 4] = [90, 4]		Временная

Временная метка [100, 1], полученная узлом 2 на втором шаге, изменена на [55, 4]. Это указывает на то, что найден более короткий путь к этому узлу (проходящий через узел 4). На третьем шаге узел 5 получает две метки с одинаковым значением расстояния $u_5 = 90$.

Шаг 4. Из узла 2 можно перейти только в узел 3, но он уже имеет постоянную метку, которую нельзя изменить. Поэтому на данном шаге получаем такой же список меток, как и на предыдущем шаге, но с единственным изменением: метка узла 2 получает статус постоянной. С временной меткой остается только узел 5, но так как из этого узла нельзя попасть ни в какой другой, процесс вычислений заканчивается.

Алгоритм позволяет проводить вычисления непосредственно на схеме сети, как показано на рис. 6.14.

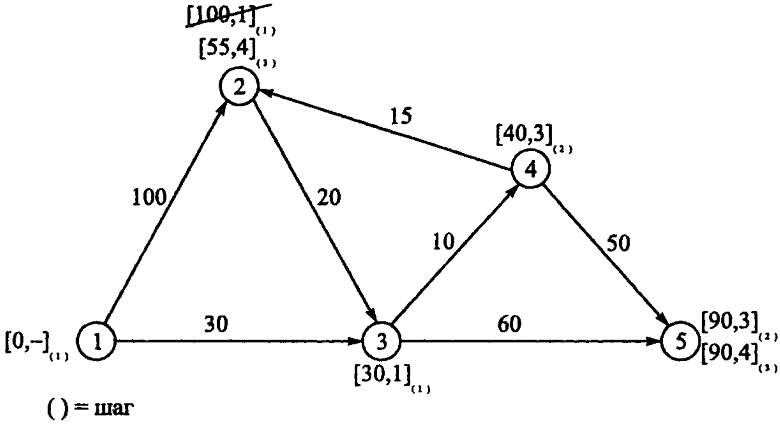


Рис. 6.14

Кратчайший маршрут между узлом 1 и любым другим узлом определяется начиная с узла назначения путем прохождения их в обратном направлении с помощью информации, представленной в постоянных метках. Например, для определения кратчайшего маршрута между узлами 1 и 2 получаем такую обратную последовательность узлов

$$(\mathbf{2}) \to [55, \mathbf{4}] \to (\mathbf{4}) \to [40, \mathbf{3}] \to (\mathbf{3}) \to [30, \mathbf{1}] \to (\mathbf{1}).$$

Таким образом, получаем путь $1 \to 3 \to 4 \to 2$ общей длиной 55 миль.

Упражнения 6.4,b

1. На рис. 6.15 показана транспортная сеть, соединяющая восемь городов, и расстояния между ними. Найдите кратчайшие маршруты между следующими городами.

   a) Между городами 1 и 8.
   b) Между городами 1 и 6.
   c) Между городами 4 и 8.
   d) Между городами 2 и 6.

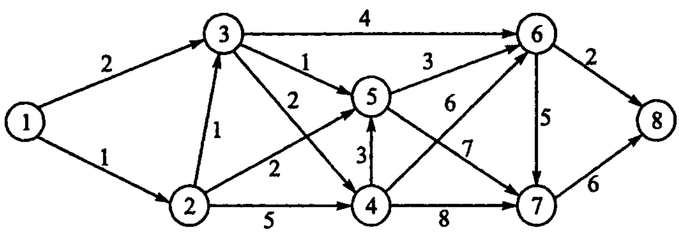


Рис. 6.15

2. Найдите кратчайшие пути между узлом 1 и всеми остальными узлами сети, представленной на рис. 6.16.

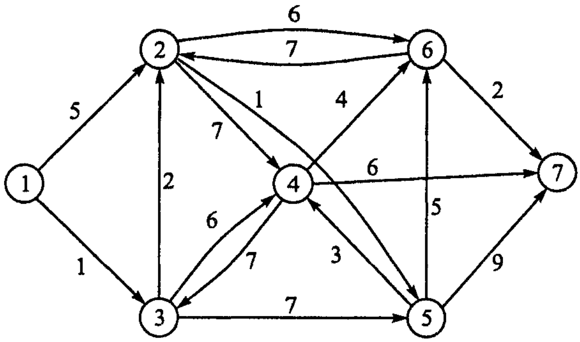


Рис. 6.16

3. Найдите оптимальное решение задачи из упр. 6.4,а(1).
4. Найдите оптимальное решение задачи из упр. 6.4,а(2).
5. Найдите оптимальное решение задачи из упр. 6.4,а(4).

Алгоритм Флойда. Этот алгоритм более общий по сравнению с алгоритмом Дейкстры, так как он находит кратчайшие пути между *любыми* двумя узлами сети. В этом алгоритме сеть представлена в виде квадратной матрицы с n строками и n столбцами. Элемент (i, j) равен расстоянию d_{ij} от узла i к узлу j, которое имеет конечное значение, если существует дуга (i, j), и равен бесконечности в противном случае.

Покажем сначала основную идею метода Флойда. Пусть есть три узла i, j и k и заданы расстояния между ними (рис. 6.17). Если выполняется неравенство $d_{ij} + d_{jk} < d_{ik}$, то целесообразно заменить путь $i \to k$ путем $i \to j \to k$. Такая замена (далее ее будем условно называть **треугольным оператором**) выполняется систематически в процессе выполнения алгоритма Флойда.

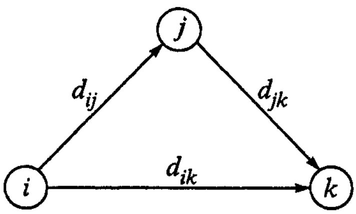


Рис. 6.17

Алгоритм Флойда требует выполнения следующих действий.

Шаг 0. Определяем начальную матрицу расстояний D_0 и матрицу последовательности узлов S_0. Диагональные элементы обеих матриц помечаются знаком "—", показывающим, что эти элементы в вычислениях не участвуют. Полагаем $k = 1$.

Основной шаг k. Задаем строку k и столбец k как *ведущую строку* и *ведущий столбец*. Рассматриваем возможность применения *треугольного оператора* ко всем элементам d_{ij} матрицы D_{k-1}. Если выполняется неравенство

$$d_{ik} + d_{kj} < d_{ij}, \quad (i \neq k, j \neq k \text{ и } i \neq j),$$

тогда выполняем следующие действия:

a) создаем матрицу D_k путем замены в матрице D_{k-1} элемента d_{ij} на сумму $d_{ik} + d_{kj}$,

b) создаем матрицу S_k путем замены в матрице S_{k-1} элемента s_{ij} на k. Полагаем $k = k + 1$ и повторяем шаг k.

$D_0 =$

	1	2	...	j	...	n
1	—	d_{12}	...	d_{1j}	...	d_{1n}
2	d_{21}	—	...	d_{2j}	...	d_{2n}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	d_{i1}	d_{i2}	...	d_{ij}	...	d_{in}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	d_{n1}	d_{n2}	...	d_{nj}	...	—

$S_0 =$

	1	2	...	j	...	n
1	—	2	...	j	...	n
2	1	—	...	j	...	n
⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	1	2	...	j	...	n
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	1	2	...	j	...	—

Поясним действия, выполняемые на k-м шаге алгоритма, представив матрицу D_{k-1} так, как она показана на рис. 6.18. На этом рисунке строка k и столбец k являются ведущими. Строка i — любая строка с номером от 1 до $k - 1$, а строка p — произвольная строка с номером от $k + 1$ до n. Аналогично столбец j представляет любой столбец с номером от 1 до $k - 1$, а столбец q — произвольный столбец с номером от $k + 1$ до n. *Треугольный оператор* выполняется следующим образом. Если сумма элементов ведущих строки и столбца (показанных в квадратиках) меньше элементов, находящихся на пересечении столбца и строки (показаны в кружках), соответствующих рассматриваемым ведущим элементам, то расстояние (элемент в кружке) заменяется на сумму расстояний, представленных ведущими элементами.

После реализации n шагов алгоритма определение по матрицам D_n и S_n кратчайшего пути между узлами i и j выполняется по следующим правилам.

1. Расстояние между узлами i и j равно элементу d_{ij} в матрице D_n.
2. Промежуточные узлы пути от узла i к узлу j определяем по матрице S_n. Пусть $s_{ij} = k$, тогда имеем путь $i \to k \to j$. Если далее $s_{ik} = k$ и $s_{kj} = j$, тогда считаем,

что весь путь определен, так как найдены все промежуточные узлы. В противном случае повторяем описанную процедуру для путей от узла i к узлу k и от узла k к узлу j.

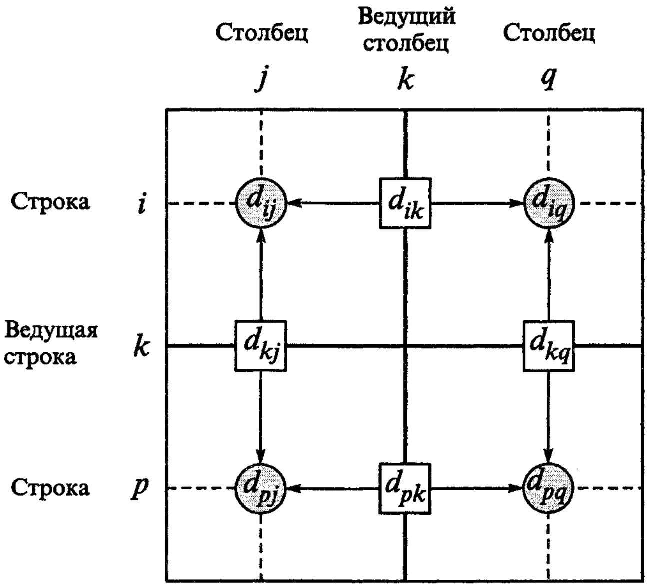


Рис. 6.18

Пример 6.4–5

Найдем для сети, показанной на рис. 6.19, кратчайшие пути между любыми двумя узлами. Расстояния между узлами этой сети проставлены на рисунке возле соответствующих ребер. Ребро (3, 5) ориентированно, поэтому не допускается движение от узла 5 к узлу 3. Все остальные ребра допускают движение в обе стороны.

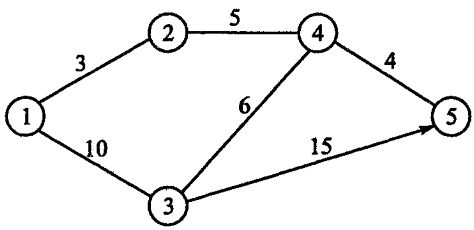


Рис. 6.19

Шаг 0. Начальные матрицы D_0 и S_0 строятся непосредственно по заданной схеме сети. Матрица D_0 симметрична, за исключением пары элементов d_{35} и d_{53}, где $d_{53} = \infty$ (поскольку невозможен переход от узла 5 к узлу 3).

	D_0				
	1	2	3	4	5
1	—	3	10	∞	∞
2	3	—	∞	5	∞
3	10	∞	—	6	15
4	∞	5	6	—	4
5	∞	∞	∞	4	—

	S_0				
	1	2	3	4	5
1	—	2	3	4	5
2	1	—	3	4	5
3	1	2	—	4	5
4	1	2	3	—	5
5	1	2	3	4	—

Шаг 1. В матрице D_0 выделены ведущие строка и столбец ($k = 1$). Двойной рамкой представлены элементы d_{23} и d_{32}, единственные среди элементов матрицы D_0, значения которых можно улучшить с помощью *треугольного оператора*. Таким образом, чтобы на основе матриц D_0 и S_0 получить матрицы D_1 и S_1, выполняем следующие действия.

1. Заменяем d_{23} на $d_{21} + d_{13} = 3 + 10 = 13$ и устанавливаем $s_{23} = 1$.

2. Заменяем d_{32} на $d_{31} + d_{12} = 10 + 3 = 13$ и устанавливаем $s_{32} = 1$.

Матрицы D_1 и S_1 имеют следующий вид.

	D_1				
	1	2	3	4	5
1	—	3	10	∞	∞
2	3	—	**13**	5	∞
3	10	**13**	—	6	15
4	∞	5	6	—	4
5	∞	∞	∞	4	—

	S_1				
	1	2	3	4	5
1	—	2	3	4	5
2	1	—	**1**	4	5
3	1	**1**	—	4	5
4	1	2	3	—	5
5	1	2	3	4	—

Шаг 2. Полагаем $k = 2$; в матрице D_1 выделены ведущие строка и столбец. Треугольный оператор применяется к элементам матриц D_1 и S_1, выделенным двойной рамкой. В результате получаем матрицы D_2 и S_2.

	D_2				
	1	2	3	4	5
1	—	3	10	**8**	∞
2	3	—	13	5	∞
3	10	13	—	6	15
4	**8**	5	6	—	4
5	∞	∞	∞	4	—

	S_2				
	1	2	3	4	5
1	—	2	3	**2**	5
2	1	—	1	4	5
3	1	1	—	4	5
4	**2**	2	3	—	5
5	1	2	3	4	—

Шаг 3. Полагаем $k = 3$; в матрице D_2 выделены ведущие строка и столбец. Треугольный оператор применяется к элементам матриц D_2 и S_2, выделенным двойной рамкой. В результате получаем матрицы D_3 и S_3.

D_3

	1	2	3	4	5
1	—	3	10	8	**25**
2	3	—	13	5	**28**
3	10	13	—	6	15
4	8	5	6	—	4
5	∞	∞	∞	4	—

S_3

	1	2	3	4	5
1	—	2	3	2	**3**
2	1	—	1	4	**3**
3	1	1	—	4	5
4	2	2	3	—	5
5	1	2	3	4	—

Шаг 4. Полагаем $k = 4$, ведущие строка и столбец в матрице D_3 выделены. Получаем новые матрицы D_4 и S_4.

D_4

	1	2	3	4	5
1	—	3	10	8	**12**
2	3	—	**11**	5	**9**
3	10	**11**	—	6	**10**
4	8	5	6	—	4
5	**12**	**9**	**10**	4	—

S_4

	1	2	3	4	5
1	—	2	3	2	**4**
2	1	—	**4**	4	**4**
3	1	**4**	—	4	**4**
4	2	2	3	—	5
5	**4**	**4**	**4**	4	—

Шаг 5. Полагаем $k = 5$, ведущие строка и столбец в матрице D_4 выделены. Никаких действий на этом шаге не выполняем; вычисления закончены.

Конечные матрицы D_4 и S_4 содержат всю информацию, необходимую для определения кратчайших путей между любыми двумя узлами сети. Например, кратчайшее расстояние между узлами 1 и 5 равно $d_{15} = 12$.

Для нахождения соответствующих маршрутов напомним, что сегмент маршрута (i, j) состоит из ребра (i, j) только в том случае, когда $s_{ij} = j$. В противном случае узлы i и j связаны, по крайней мере, через один промежуточный узел. Например, поскольку $s_{15} = 4$ и $s_{45} = 5$, сначала кратчайший маршрут между узлами 1 и 5 будет иметь вид $1 \to 4 \to 5$. Но так как $s_{14} \neq 4$, узлы 1 и 4 в определяемом пути не связаны одним ребром (но в исходной сети они могут быть связаны непосредственно). Далее следует определить промежуточный узел (узлы) между первым и четвертым узлами. Имеем $s_{14} = 2$ и $s_{24} = 4$, поэтому маршрут $1 \to 4$ заменяем $1 \to 2 \to 4$. Поскольку $s_{12} = 2$ и $s_{24} = 4$, других промежуточных узлов нет. Комбинируя определенные сегменты маршрута, окончательно получаем следующий кратчайший путь от узла 1 до узла 5: $1 \to 2 \to 4 \to 5$. Длина этого пути равна 12 милям.

Упражнения 6.4,с

1. В задаче из примера 6.4–5 определите кратчайшие пути между следующими парами узлов.
   a) От узла 5 к узлу 1.
   b) От узла 3 к узлу 5.
   c) От узла 5 к узлу 3.
   d) От узла 5 к узлу 2.

2. Примените алгоритм Флойда к сети, показанной на рис. 6.20. Заметьте, что ребра (7, 6) и (6, 4) ориентированы. Определите кратчайшие пути между следующими парами узлов.

   a) От узла 1 к узлу 7.
   b) От узла 7 к узлу 1.
   c) От узла 6 к узлу 7.

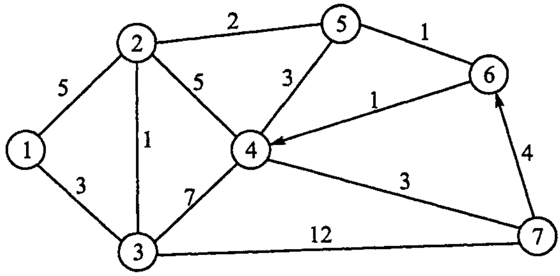


Рис. 6.20

3. Телефонная компания обслуживает шесть удаленных друг от друга районов, которые связаны сетью, показанной на рис. 6.21. Расстояния на схеме сети указаны в милях. Компании необходимо определить наиболее эффективные маршруты пересылки сообщений между любыми двумя районами.

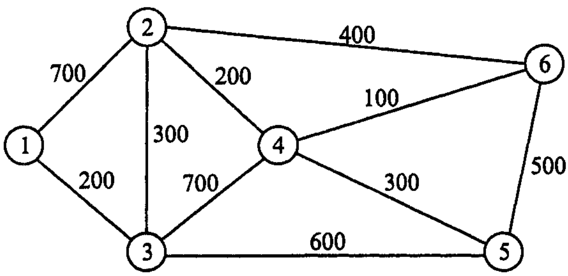


Рис. 6.21

6.5. Задача о максимальном потоке

Рассмотрим сеть трубопроводов для транспортировки сырой нефти от буровых скважин до нефтеперегонных заводов. Для перекачки нефти предусмотрены магистральные насосные станции. Каждый сегмент трубопровода имеет свою пропускную способность. Сегменты трубопровода могут быть как однонаправленные (осуществляют перекачку нефти только в одном направлении), так и двунаправленные. В однонаправленных сегментах положительная пропускная способность предполагается в одном направлении и нулевая — в другом. На рис. 6.22 показана типовая сеть нефтепроводов. Как определить максимальную пропускную способность (т.е. максимальный поток) между нефтяными скважинами и нефтеперегонными заводами?

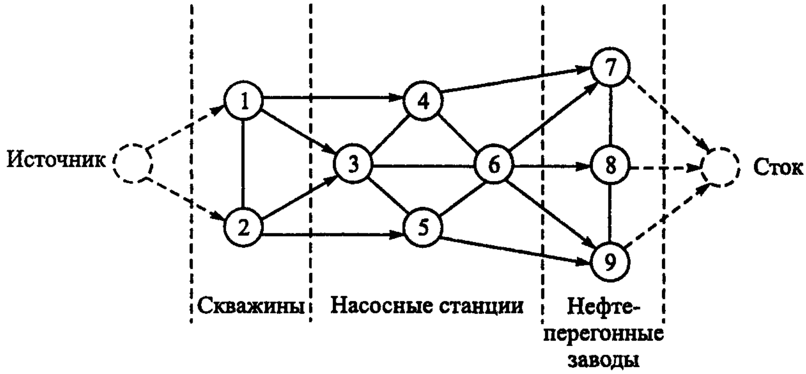


Рис. 6.22

При решении данной задачи исходную сеть необходимо свести к сети с одним источником и одним стоком. Этого можно достигнуть путем введения дополнительных дуг с бесконечной пропускной способностью от источника к скважинам и от нефтеперегонных заводов к стоку (на рис. 6.22 эти дуги показаны пунктирными линиями).

Для ребра (i, j), где $i < j$, используем запись $(\overline{C}_{ij}, \overline{C}_{ji})$ для представления пропускных способностей в направлениях $i \to j$ и $j \to i$ соответственно. Во избежание недоразумений на схеме сети $\overline{C}_{ij}$ будем располагать на ребре (i, j) ближе к узлу i, а $\overline{C}_{ji}$ — ближе к узлу j, как показано на рис. 6.23.

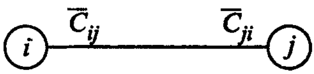


Рис. 6.23

6.5.1. Перебор разрезов

Разрез определяет множество ребер, при удалении которых из сети полностью прекращается поток от источника к стоку. **Пропускная способность разреза** равна сумме пропускных способностей "разрезанных" ребер. Среди *всех* разрезов сети разрез с *минимальной пропускной способностью* определяет максимальный поток в сети.

Пример 6.5–1

Рассмотрим сеть, показанную на рис. 6.24. На этом рисунке при обозначении пропускных способностей двунаправленных ребер придерживались соглашения, принятого ранее (рис. 6.23). Например, для ребра (3, 4) пропускная способность в направлении 3 → 4 равна 10, а в направлении 4 → 3 — 5.

Разрезы, представленные на рис. 6.24, имеют следующие пропускные способности.

Разрез	"Разрезанные" ребра	Пропускная способность
1	(1, 2), (1, 3), (1, 4)	10 + 30 + 20 = 60
2	(1, 3), (1, 4), (2, 3), (2, 5)	30 + 10 + 40 + 30 = 110
3	(2, 5), (3, 5), (4, 5)	30 + 20 + 20 = 70

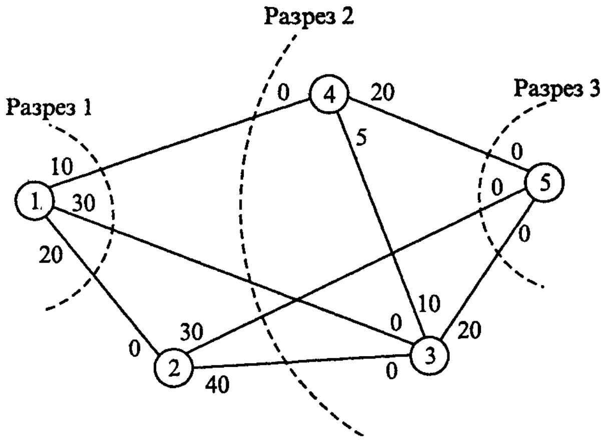


Рис. 6.24

Вывод, который можно сделать из этих трех разрезов, заключается в том, что максимальный поток не может превышать 60 единиц. Но мы не можем сказать, какой максимальный поток на самом деле, так как не перебрали *все возможные* разрезы сети. К сожалению, перебор всех разрезов является непростой задачей. Поэтому для определения максимального потока в сети не используются алгоритмы, основанные на полном переборе разрезов.

Упражнение 6.5,а

1. Для сети, показанной на рис. 6.24, проведите еще два разреза и найдите их пропускные способности.

6.5.2. Алгоритм нахождения максимального потока

Идея данного алгоритма состоит в нахождении сквозных путей с положительными потоками от источника к стоку.

Рассмотрим ребро (i, j) с (начальной) пропускной способностью $(\bar{C}_{ij}, \bar{C}_{ji})$. В процессе выполнения алгоритма части этих пропускных способностей "забираются" потоками, проходящими через данное ребро, в результате каждое ребро будет иметь **остаточную пропускную способность**. Будем использовать запись (c_{ij}, c_{ji}) для представления остаточных пропускных способностей. Сеть, где все ребра имеют остаточную пропускную способность, назовем **остаточной**.

Для произвольного узла j, получающего поток от узла i, определим метку $[a_j, i]$, где a_j — величина потока, протекающего от узла j к узлу i. Алгоритм нахождения максимального потока предполагает выполнение следующих действий:

Шаг 1. Для всех ребер (i, j) положим остаточную пропускную способность равной первоначальной пропускной способности, т.е. приравняем (c_{ij}, c_{ji}) = $(\bar{C}_{ij}, \bar{C}_{ji})$. Назначим $a_1 = \infty$ и пометим узел 1 меткой $[\infty, —]$. Полагаем $i = 1$ и переходим ко второму шагу.

Шаг 2. Определяем множество S_i как множество узлов j, в которые можно перейти из узла i по ребру с *положительной* остаточной пропускной способностью (т.е. $c_{ij} > 0$ для всех $j \in S_i$). Если $S_i \neq \varnothing$, выполняем третий шаг, в противном случае переходим к шагу 4.

Шаг 3. В множестве S_i находим узел k, такой, что

$$c_{ik} = \max_{j \in S_i} \left\{ c_{ij} \right\}.$$

Положим $a_k = c_{ik}$ и пометим узел k меткой $[a_k, i]$. Если последней меткой помечен узел стока (т.е. если $k = n$), сквозной путь найден, и мы переходим к пятому шагу. В противном случае полагаем $i = k$ и возвращаемся ко второму шагу.

Шаг 4. **(Откат назад).** Если $i = 1$, сквозной путь невозможен, и мы переходим к шагу 6. Если $i \neq 1$, находим помеченный узел r, *непосредственно* предшествующий узлу i, и удаляем узел i из множества узлов, смежных с узлом r. Полагаем $i = r$ и возвращаемся ко второму шагу.

Шаг 5. **(Определение остаточной сети).** Обозначим через $N_p = \{1, k_1, k_2, ..., n\}$ множество узлов, через которые проходит p-й найденный сквозной путь от узла источника (узел 1) до узла стока (узел n). Тогда максимальный поток, проходящий по этому пути, вычисляется как

$$f_p = \min\{a_1, a_{k1}, a_{k2}, ..., a_n\}.$$

Остаточные пропускные способности ребер, составляющих сквозной путь, *уменьшаются* на величину f_p в направлении движения потока и *увеличиваются* на эту же величину в противоположном направлении. Таким образом, для ребра (i, j), входящего в сквозной путь, текущие остаточные стоимости (c_{ij}, c_{ji}) изменятся следующим образом:

a) $(c_{ij} - f_p, c_{ji} + f_p)$, если поток идет от узла i к узлу j,

b) $(c_{ij} + f_p, c_{ji} - f_p)$, если поток идет от узла j к узлу i.

Далее восстанавливаем все узлы, удаленные на шаге 4. Полагаем $i = 1$ и возвращаемся ко второму шагу для поиска нового сквозного пути.

Шаг 6. **(Решение).**

a) При m найденных сквозных путях максимальный поток вычисляется по формуле

$$F = f_1 + f_2 + ... + f_m.$$

b) Имея значения начальных $(\bar{C}_{ij}, \bar{C}_{ji})$ и конечных (c_{ij}, c_{ji}) пропускных способностей ребра (i, j), можно вычислить оптимальный поток через это ребро следующим образом. Положим $(\alpha, \beta) = (\bar{C}_{ij} - c_{ij}, \bar{C}_{ji} - c_{ji})$. Если $\alpha > 0$, поток, проходящий через ребро (i, j), равен α. Если же $\beta > 0$, тогда поток равен β. (Случай, когда одновременно $\alpha > 0$ и $\beta > 0$, невозможен.)

Процесс отката назад на четвертом шаге выполняется тогда, когда алгоритм должен “убить” промежуточный узел до момента реализации сквозного пути. Коррекцию пропускных способностей, выполняемых на шаге 5, можно пояснить на примере простой сети, показанной на рис. 6.25. На рис. 6.25,*а* найден первый сквозной путь $N_1 = \{1, 2, 3, 4\}$ с максимальным потоком $f_1 = 5$. После этого остаточные пропускные способности ребер (1, 2), (2, 3) и (3, 4) изменятся соответственно с (5, 0) на (0, 5). На рис. 6.25,*б* показан второй сквозной путь $N_2 = \{1, 3, 2, 4\}$ с максимальным потоком $f_2 = 5$. После коррекции пропускных способностей получаем сеть, показанную на рис. 6.25,*в*, где уже невозможно построить сквозной путь. Почему так получилось? При вычислении остаточных пропускных способностей на шаге 5 при переходе от сети (*б*) к сети (*в*) невозможна организация потока в направлении 2 → 3. Получается, что алгоритм как бы "помнит", что поток в направлении 2 → 3 уже был в предыдущих сквозных путях, и поэтому снова (на пятом шаге) изменяет пропускную способность с 0 до 5 в направлении от узла 3 к узлу 2.

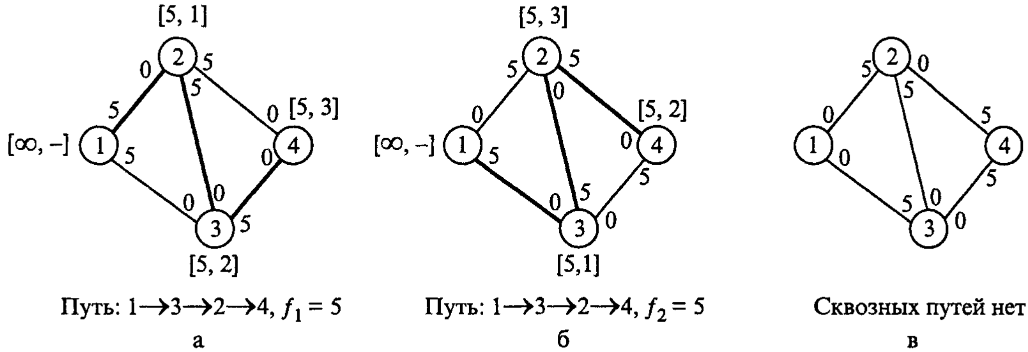


Рис. 6.25

Пример 6.5–2

Найдем максимальный поток в сети из примера 6.5–1 (рис. 6.24). На рис. 6.26 предлагается графическая иллюстрация выполнения алгоритма. Считаем полезным сравнить описание выполняемых алгоритмом вычислительных итераций с их графическим представлением.

Итерация 1

Положим остаточные пропускные способности (c_{ij}, c_{ji}) всех ребер равными первоначальным пропускным способностям $(\overline{C}_{ij}, \overline{C}_{ji})$.

Шаг 1. Назначаем $a_1 = \infty$ и помечаем узел 1 меткой [∞, —]. Полагаем $i = 1$.

Шаг 2. $S_1 = [2, 3, 4]$ $(\neq \varnothing)$.

Шаг 3. $k = 3$, поскольку $c_{13} = \max\{c_{12}, c_{13}, c_{14}\} = \max\{20, 30, 10\} = 30$. Назначаем $a_3 = c_{13} = 30$ и помечаем узел 3 меткой [30, 1]. Полагаем $i = 3$ и возвращаемся к шагу 2.

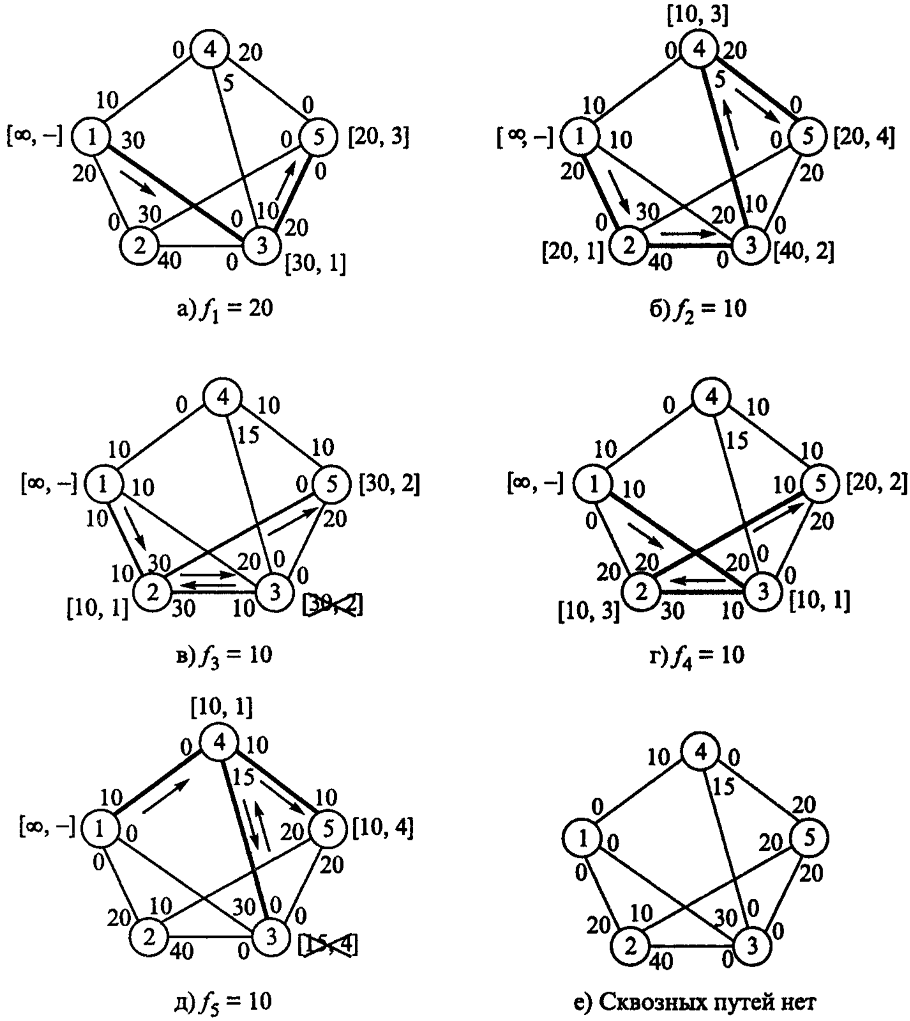
Шаг 2. $S_2 = [4, 5]$.

Шаг 3. $k = 5$ и $a_5 = c_{35} = \max\{10, 20\} = 20$. Помечаем узел 5 меткой [20, 3]. Получен сквозной путь. Переходим к шагу 5.

Шаг 5. Сквозной путь определяем по меткам, начиная с узла 5 и заканчивая узлом 1: (5) → [20, 3] → (3) → [30, 1] → (1). Таким образом, $N_1 = \{1, 3, 5\}$

и $f_1 = \min\{a_1, a_3, a_5\} = \{\infty, 30, 20\} = 20$. Вычисляем остаточные пропускные способности вдоль пути N_1:

$$(c_{13}, c_{31}) = (30 - 20, 0 + 20) = (10, 20),$$
$$(c_{35}, c_{53}) = (20 - 20, 0 + 20) = (0, 20).$$



а) $f_1 = 20$ б) $f_2 = 10$

в) $f_3 = 10$ г) $f_4 = 10$

д) $f_5 = 10$ е) Сквозных путей нет

Рис. 6.26

Итерация 2

Шаг 1. Назначаем $a_1 = \infty$ и помечаем узел 1 меткой [∞, —]. Полагаем $i = 1$.

Шаг 2. $S_1 = [2, 3, 4]$.

Шаг 3. $k = 2$, назначаем $a_2 = c_{12} = \max\{20, 10, 10\} = 20$ и помечаем узел 2 меткой [20, 1]. Полагаем $i = 2$ и возвращаемся к шагу 2.

Шаг 4. $S_2 = [3, 5]$.

Шаг 5. $k = 3$ и $a_3 = c_{23} = 40$. Помечаем узел 3 меткой [40, 2]. Полагаем $i = 3$ и возвращаемся к шагу 2.

Шаг 2. $S_3 = [4]$ (отметим, что $c_{35} = 0$, поэтому узел 5 не включается в S_3).

Шаг 3. $k = 4$, назначаем $a_4 = c_{34} = 10$ и помечаем узел 4 меткой [10, 3]. Полагаем $i = 4$ и возвращаемся к шагу 2.

Шаг 2. $S_4 = [5]$ (поскольку узлы 1 и 3 уже помечены, они не включаются в S_4).

Шаг 3. $k = 5$ и $a_5 = c_{45} = 20$. Помечаем узел 5 меткой [20, 4]. Получен сквозной путь. Переходим к шагу 5.

Шаг 5. $N_2 = \{1, 2, 3, 4, 5\}$ и $f_2 = \min\{\infty, 20, 40, 10, 20\} = 10$. Вычисляем остаточные пропускные способности вдоль пути N_2:

$$(c_{12}, c_{21}) = (20 - 10, 0 + 10) = (10, 10),$$
$$(c_{23}, c_{32}) = (40 - 10, 0 + 10) = (30, 10),$$
$$(c_{34}, c_{43}) = (10 - 10, 5 + 10) = (0, 15),$$
$$(c_{45}, c_{54}) = (20 - 10, 0 + 10) = (10, 10).$$

Итерация 3

Шаг 1. Назначаем $a_1 = \infty$ и помечаем узел 1 меткой [∞, —]. Полагаем $i = 1$.

Шаг 2. $S_1 = [2, 3, 4]$.

Шаг 3. $k = 2$, назначаем $a_2 = c_{12} = \max\{10, 10, 10\} = 10$ и помечаем узел 2 меткой [10, 1]. Полагаем $i = 2$ и возвращаемся к шагу 2.

Шаг 2. $S_2 = [3, 5]$.

Шаг 3. $k = 3$ и $a_3 = c_{23} = 30$. Помечаем узел 3 меткой [30, 2]. Полагаем $i = 3$ и возвращаемся к шагу 2.

Шаг 2. $S_3 = \varnothing$ (поскольку $c_{34} = c_{35} = 0$). Переходим к шагу 4.

Шаг 4. Метка [30, **2**] узла 3 показывает номер предшествующего узла $r = 2$. На *этой итерации* узел 3 в дальнейшем во внимание не принимается, его метку вычеркиваем. Полагаем $i = r = 2$ и возвращаемся к шагу 2.

Шаг 2. $S_4 = [5]$ (поскольку узел 3 удален из возможного сквозного пути).

Шаг 3. $k = 5$ и $a_5 = c_{25} = 30$. Помечаем узел 5 меткой [30, 2]. Получен сквозной путь. Переходим к шагу 5.

Шаг 5. $N_3 = \{1, 2, 5\}$ и $f_3 = \min\{\infty, 10, 30\} = 10$. Вычисляем остаточные пропускные способности вдоль пути N_3:

$$(c_{12}, c_{21}) = (10 - 10, 10 + 10) = (0, 20),$$
$$(c_{25}, c_{52}) = (30 - 10, 0 + 10) = (20, 10).$$

Итерация 4

На этой итерации получен путь $N_4 = \{1, 3, 2, 5\}$ с $f_4 = 10$ (проверьте!).

Итерация 5

На этой итерации получен путь $N_5 = \{1, 4, 5\}$ с $f_5 = 10$ (проверьте!).

Итерация 6

Новые сквозные пути невозможны, поскольку все ребра, исходящие из узла 1, имеют нулевые остаточные пропускные способности. Переходим к шагу 6 для определения решения.

Шаг 6. Максимальный объем потока в сети равен $F = f_1 + f_2 + \ldots + f_5 = 20 + 10 + 10 + 10 + 10 = 60$ единиц. Значения потоков по различным ребрам вычисляются путем вычитания последних значений остаточных пропускных способностей (т.е. $(c_{ij}, c_{ji})_6$) из первоначальных значений пропускных способностей $(\overline{C}_{ij}, \overline{C}_{ji})$. Результаты вычислений приведены в следующей таблице.

Ребро	$(\overline{C}_{ij}, \overline{C}_{ji}) - (c_{ij}, c_{ji})_6$	Величина потока	Направление
(1, 2)	(20, 0) – (0, 20) = (20, –20)	20	$1 \to 2$
(1, 3)	(30, 0) – (0, 30) = (30, –30)	30	$1 \to 3$
(1, 4)	(10, 0) – (0, 10) = (10, –10)	10	$1 \to 4$
(2, 3)	(40, 0) – (40, 0) = (0, 0)	0	—
(2, 5)	(30, 0) – (10, 20) = (20, –20)	20	$2 \to 5$
(3, 4)	(10, 5) – (0, 15) = (10, –10)	10	$3 \to 4$
(3, 5)	(20, 0) – (0, 20) = (20, –20)	20	$3 \to 5$
(4, 5)	(20, 0) – (0, 20) = (20, –20)	20	$4 \to 5$

Упражнения 6.5,b

1. В задаче из примера 6.5–2

   a) для всех ребер определите величины неиспользованных пропускных способностей;

   b) найдите величину потока, проходящего через узлы 2, 3 и 4;

   c) можно ли увеличить максимальный поток в сети путем повышения пропускных способностей в направлениях $3 \to 5$ и $4 \to 5$?

2. С помощью программы TORA проверьте все вычисления, выполненные в примере 6.5–2 и показанные на рис. 6.26.

3. Найдите максимальный поток и значения потоков, проходящих через каждое ребро сети, показанной на рис. 6.27.

4. Три нефтеперегонных завода транспортируют свою продукцию двум распределительным терминалам по сети трубопроводов, которая включает и насосные станции (рис. 6.28). Направления потоков в сети показаны стрелками, пропускные способности отдельных сегментов сети указаны в миллионах баррелей в день.

a) Определите ежедневную производительность каждого нефтеперегонного завода, соответствующую максимальной пропускной способности сети трубопроводов.

b) Определите ежедневную потребность каждого распределительного терминала, соответствующую максимальной пропускной способности сети трубопроводов.

c) Определите ежедневную пропускную способность каждой насосной станции, соответствующую максимальной пропускной способности сети трубопроводов.

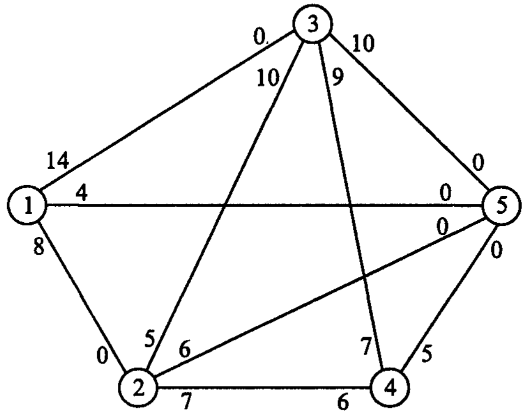


Рис. 6.27

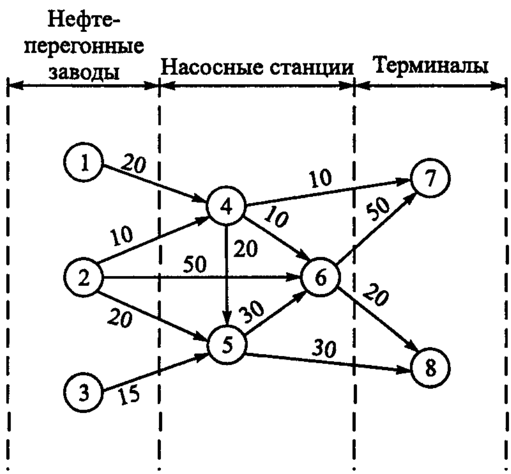


Рис. 6.28

5. Пусть в сети из предыдущего упражнения (рис. 6.28) пропускная способность насосной станции 6 ограничена 60 миллионами баррелей в день. Найдите максимальную пропускную способность сети с учетом этого ограничения.

6. Зерно из трех зернохранилищ доставляется на грузовиках четырем птицеводческим фермам, при этом некоторые зернохранилища не могут непосредственно поставлять зерно определенным фермам. Пропускная способность маршрутов от

зернохранилищ до птицеводческих ферм ограничена количеством используемых грузовиков и числом выполняемых ежедневно рейсов. В следующей таблице показаны ежедневные предложения зернохранилищ и ежедневный спрос птицеводческих ферм (в тысячах фунтов), в ячейках таблицы указаны пропускные способности соответствующих маршрутов.

		Фермы 1	2	3	4	
Зернохранилища	1	30	5	0	40	**20**
	2	0	0	5	90	**20**
	3	100	40	30	40	**200**
		200	**10**	**60**	**20**	

a) Найдите схему транспортировки зерна, обеспечивающую максимальный спрос птицеводческих ферм.

b) Существует ли при данных условиях схема транспортировки, обеспечивающая весь спрос птицеводческих ферм?

7. Пусть в задаче из предыдущего упражнения возможна транспортировка зерна между зернохранилищами 1 и 2, а также 2 и 3. Кроме того, возможна транспортировка между фермами 1 и 2, 2 и 3, 3 и 4. Максимальная пропускная способность в обоих направлениях указанных маршрутов составляет 50 тысяч фунтов. Как данные предположения скажутся на обеспечении пока неудовлетворенного спроса птицеводческих ферм?

8. Родители имеют пять детей подросткового возраста, которых ежедневно привлекают к пяти видам домашней работы. Опыт трудового воспитания детей показал, что принудительное (силовое) назначение на работу чревато конфликтами. Поэтому дети сами составили список своих предпочтений, который приведен в следующей таблице.

Ребенок	Предпочтительные работы
Ральф	3, 4 или 5
Мэй	1
Бен	1 или 2
Ким	1, 2 или 5
Кен	2

Теперь родителям осталось найти такое распределение работ, которое в наибольшей степени учитывало бы предпочтения детей. Найдите максимальное количество таких распределений работ.

9. Четыре фабрики имеют заказ на производство четырех типов игрушек. В следующей таблице показано, какие игрушки должна производить каждая фабрика.

Фабрика	Тип игрушки
1	1, 2, 3
2	2, 3
3	1, 4
4	3, 4

Ежедневные производственные возможности фабрик составляют 250, 180, 300 и 100 штук игрушек соответственно. Ежедневный спрос на игрушки четырех типов составляет 200, 150, 350 и 100 штук. Разработайте производственный план, максимально удовлетворяющий спрос на игрушки.

10. Студенческий совет университета рассматривает представления на шесть студентов, являющихся членами четырех студенческих обществ. Студенческий совет имеет три секции: естественных, гуманитарных и инженерных наук. Совет может принять не более двух студентов в каждую секцию. В следующей таблице показано членство студентов в каждом из четырех студенческих обществ.

Общество	Студент
1	1, 2, 3
2	1, 3, 5
3	3, 4, 5
4	1, 2, 4, 6

Приведем также распределение студентов по возможному принятию их в различные секции совета.

Секция	Студент
естественных наук	1, 2, 4
гуманитарных наук	3, 4
инженерных наук	4, 5, 6

Каждый студент может быть принят только в одну секцию. Могут ли быть представлены в совете все четыре студенческих общества?

11. *Максимальный и минимальный потоки в сети с нижними положительными границами пропускных способностей.* В этом разделе при выполнении алгоритма нахождения максимального потока в сети неявно предполагалось, что нижние границы пропускных способностей всех ребер равны нулю. В некоторых задачах нижние границы пропускных способностей могут быть строго положительными. В таком случае возникает интерес к нахождению как максимального, так и минимального потоков в сети (см. комплексную задачу 6-3 в конце главы). Наличие положительных нижних границ пропускных способностей вызывает определенные затруднения, так как в этом случае сеть вообще может не иметь допустимого потока. Максимальный и минимальный потоки в сети можно найти, выполняя следующие действия.

Шаг 1. Найдите начальное допустимое решение для сети с положительными нижними границами пропускных способностей.

Шаг 2. Используя допустимое решение, найденное на первом шаге, определите максимальный и минимальный потоки в сети.

а) Покажите, что поток через дугу (i, j), величина которого ограничена неравенствами $l_{ij} \le x_{ij} \le u_{ij}$, представим в виде стока со спросом l_{ij} в узле i, источником с предложением l_{ij} в узле j и потоком, ограниченным неравенствами $0 \le x'_{ij} \le u_{ij} - l_{ij}$.

b) Покажите, что нахождение допустимого решения в исходной сети эквивалентно нахождению максимального потока в сети, где (1) поток x_{ij} через каждую дугу (i, j) заменен потоком x'_{ij} с ограничениями $0 \le x'_{ij} \le u_{ij} - l_{ij}$ (как показано в предыдущем пункте); (2) все источники "слиты" в один с выходящей из него дугой, имеющей пропускную способность l_{ij}; (3) все стоки "слиты" в один с входящей в него дугой, имеющей пропускную способность l_{ij}; (4) конечный узел t соединен с узлом источника s исходной сети дугой с бесконечной пропускной способностью. Допустимое решение существует, если максимальный поток в новой сети равен сумме нижних границ пропускных способностей исходной сети. Примените описанную процедуру к следующей сети и найдите допустимое решение (поток).

Дуга (i, j)	(l_{ij}, u_{ij})
(1, 2)	(5, 20)
(1, 3)	(0, 15)
(2, 3)	(4, 10)
(2, 4)	(3, 15)
(3, 4)	(0, 20)

c) Для нахождения *минимального* потока в исходной сети используйте допустимое решение, найденное в предыдущем пункте, совместно с алгоритмом нахождения максимального потока. (*Совет.* Сначала на основе имеющегося допустимого решения найдите остаточную сеть. Далее определите максимальный поток *от конечного узла сети к начальному.* Теперь, комбинируя допустимый и максимальный потоки, найдите минимальный поток в исходной сети.)

d) Используя алгоритм нахождения *максимального* потока с допустимым решением, найденным в п. b), определите максимальный поток в исходной сети. (*Совет.* Так же, как и в п. c), начните с определения остаточной сети. Далее примените алгоритм нахождения сквозных путей к этой сети.)

6.6. Нахождение потока наименьшей стоимости

Задача нахождения потока наименьшей стоимости в сети с ограниченной пропускной способностью обобщает задачу определения максимального потока (см. предыдущий раздел) по следующим направлениям.

1. Все ребра допускают только одностороннее направление потока, т.е. являются (ориентированными) дугами.
2. Каждой дуге поставлена в соответствие (неотрицательная) стоимость прохождения единицы потока по данной дуге.
3. Дуги могут иметь положительную нижнюю границу пропускной способности.
4. Любой узел сети может выступать как в качестве источника, так и стока.

В рассматриваемой задаче необходимо найти потоки по дугам, минимизирующие общую стоимость прохождения потока по сети; при этом должны удовлетворяться ограничения на пропускные способности дуг и на величины предложений и спроса отдельных (или всех) узлов. Сначала опишем сетевую модель с заданными стоимостями прохождения потока по дугам и сформулируем соответствующую задачу линейного программирования. Эта задача далее будет использована как основа для построения специального симплексного алгоритма, предназначенного для решения данной сетевой задачи.

6.6.1. Сетевая модель

Рассмотрим сеть $G = (N, A)$ с ограниченной пропускной способностью, где N — множество узлов, A — множество дуг. Обозначим

x_{ij} — величина потока, протекающего от узла i к узлу j,
u_{ij} (l_{ij}) — верхняя (нижняя) граница пропускной способности дуги (i, j),
c_{ij} — стоимость прохождения единицы потока по дуге (i, j),
f_i — величина "чистого" результирующего потока, протекающего через узел i.

На рис. 6.29 показано, как на схемах сетей будем отображать определенные параметры дуг. Метка [f_i] указывает положительное (отрицательное) значение предложения (спроса), соответствующего узлу i.

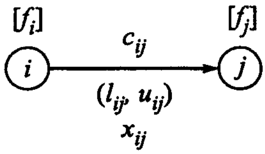


Рис. 6.29

Пример 6.6–1

Компания GrainCo снабжает зерном из трех зернохранилищ три птицеводческие фермы. Предложение зернохранилищ составляет 100, 200 и 50 тысяч бушель зерна, а спрос ферм — 150, 80 и 120 тысяч бушель соответственно. Компания может транспортировать зерно по железной дороге, за исключением трех маршрутов, где используется автомобильный транспорт.

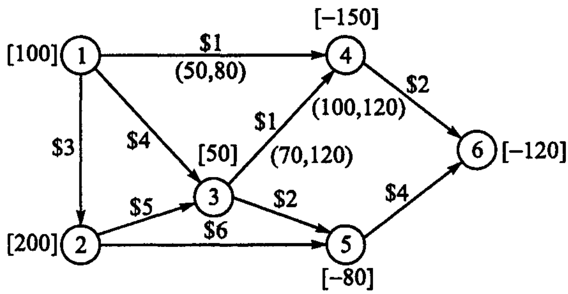


Рис. 6.30

На рис. 6.30 показаны возможные маршруты между зернохранилищами и птицеводческими фермами. Зернохранилища представлены узлами 1, 2 и 3; их предложения указаны метками [100], [200] и [50] соответственно. Фермы обозначены узлами 4, 5 и 6 с величинами спроса [−150], [−80] и [−120]. Маршруты транспортировки зерна показаны на рис. 6.30 дугами, соединяющими узлы сети. Дуги (1, 4), (3, 4) и (4, 6) соответствуют автомобильным маршрутам. Эти маршруты имеют верхние и нижние границы пропускных способностей. Например, по маршруту (1, 4) можно провести от 50 до 80 тысяч бушель зерна. Другие маршруты соответствуют железнодорожному транспорту, пропускная способность которого практически не ограничена. Стоимость транспортировки одного бушеля зерна показана возле каждой дуги.

Упражнения 6.6,а

1. Компания разрабатывает 4-этапный план производства некоего продукта согласно следующим данным.

Этап	Объем спроса	Стоимость единицы продукции ($)	Стоимость хранения единицы продукции ($)
1	100	24	1
2	110	26	2
3	95	21	1
4	125	24	2

Сформулируйте данную задачу как сетевую модель в предположении, что заказы нельзя выполнять "задним числом".

2. Пусть в предыдущем упражнении возможна недопоставка продукции со штрафом $1.50 за единицу продукции при задержке в течение одного этапа. Сформулируйте задачу как сетевую модель.

3. Пусть в упр. 1 производственные возможности каждого из четырех этапов составляют соответственно 110, 95, 125 и 100 единиц продукции, что не позволяет выполнить все заказы (без недопоставок продукции) в срок. Штраф за недопоставку единицы продукции равен $1.50 при задержке в течение одного этапа. Сформулируйте задачу как сетевую модель.

4. Химическая компания владеет двумя заводами, которые производят определенные химические компоненты для двух клиентов; потребности этих клиентов составляют 660 и 800 тонн продукции ежемесячно. Первый завод может производить от 400 до 800 тонн продукции в месяц, а второй — от 450 до 900 тонн ежемесячно. На первом заводе расходы на производство одной тонны продукции составляют $25, на втором — $28. Сырье для заводов поставляют два поставщика, которые могут поставить не менее 500 тонн сырья по цене $200 за тонну первому заводу и не менее 750 тонн по цене $210 второму заводу. Химическая компания предполагает самостоятельно транспортировать и сырье, и готовую продукцию. Стоимость транспортировки одной тонны сырья от первого поставщика к заводам составляет $10 для первого завода и $12 — для второго. Аналогичные стоимости перевозок от второго поставщика равны $9 и $13, соответственно для первого и второго заводов. Стоимость перевозки одной тонны продукции от первого завода к клиенту 1 и 2 составляет $3 и $4, от второго завода — $5 и $2 соответственно. Допустим,

из одной тонны сырья можно получить одну тонну готовой продукции. Сформулируйте задачу в виде сетевой модели.

5. В двух открытых школах необходимо изменить расовый баланс среди учащихся путем дополнительного их набора из национальных меньшинств. Учащиеся из национальных меньшинств должны составлять от 30% до 40% от общего количества. Учащиеся из национальных меньшинств живут в трех общинах, а "обычные" — в двух. В следующей таблице приведены расстояния от общин до школ (в милях).

		Расстояние от школы до				
		общины национальных меньшинств			"обычной" общины	
Школа	Количество учащихся (не более)	1	2	3	1	2
1	1500	20	12	10	4	5
2	2000	15	18	8	6	5
Количество учащихся		500	450	300	100 0	1000

Сформулируйте задачу как сетевую модель и определите расовый состав учащихся в каждой школе.

6.6.2. Сетевая модель как задача линейного программирования

Определение модели сети с ограниченной пропускной способностью как задачи линейного программирования необходимо для разработки симплексного алгоритма решения задач данного типа (алгоритм описан в следующем разделе). Используя определения, данные в разделе 6.6.1, мы можем записать задачу линейного программирования для сети с ограниченной пропускной способностью следующим образом.

$$\text{Минимизировать } z = \sum_{(i,j)\in A}\sum c_{ij}x_{ij}$$

при ограничениях

$$\sum_{\substack{k \\ (j,k)\in A}} x_{jk} - \sum_{\substack{i \\ (i,j)\in A}} x_{ij} = f_j, \quad j \in N,$$

$$l_{ij} \le x_{ij} \le u_{ij}.$$

Результирующий "чистый" поток f_j, протекающий через узел j, вычисляется по формуле

f_j = (Величина потока, выходящего из узла j) – (Величина потока, входящего в узел j).

Узел j выступает в качестве источника, если $f_j > 0$, и как сток при $f_j < 0$.

Нижнюю границу пропускной способности l_{ij} можно удалить из ограничений с помощью подстановки $x_{ij} = x'_{ij} + l_{ij}$. Для нового потока x'_{ij} верхней границей пропускной способности будет величина $u_{ij} - l_{ij}$. В этом случае результирующий поток через узел i будет равен $[f_i] - l_{ij}$, а через узел j — $[f_j] + l_{ij}$. На рис. 6.31 показаны преобразования характеристик дуги (i, j) после исключения ее нижней границы пропускной способности.

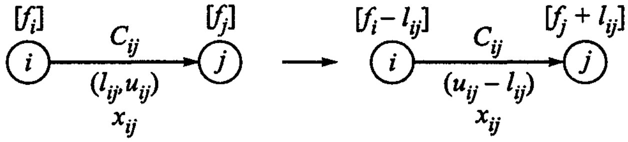


Рис. 6.31

Пример 6.6–2

Запишем задачи линейного программирования для сети из рис. 6.30 до и после исключения нижних границ пропускных способностей.

Основные ограничения формулируемой задачи линейного программирования связаны с определением входных и выходных потоков, протекающих через каждый узел, что порождает следующую задачу ЛП.

	x_{12}	x_{13}	x_{14}	x_{23}	x_{25}	x_{34}	x_{35}	x_{46}	x_{56}	
Минимизировать	3	4	1	5	6	1	2	2	4	
Узел 1	1	1	1							= 100
Узел 2	–1			1	1					= 200
Узел 3		–1		–1		1	1			= 50
Узел 4			–1			–1		1		= –150
Узел 5					–1		–1		1	= –80
Узел 6								–1	–1	= –120
Нижние границы	0	0	50	0	0	70	0	100	0	
Верхние границы	∞	∞	80	∞	∞	120	∞	120	∞	

Сделаем замечание о структуре коэффициентов, формирующих ограничения. В столбце, соответствующем переменной x_{ij}, всегда в строке i стоит +1, а в строке j — –1. Остальные коэффициенты равны нулю. Такая структура коэффициентов типична для сетевых моделей.

Для переменных, представляющих потоки через дуги, имеющие ненулевые нижние границы пропускных способностей, выполняем замену

$$x_{14} = x'_{14} + 50,$$
$$x_{34} = x'_{34} + 70,$$
$$x_{46} = x'_{46} + 100.$$

В результате получаем следующую задачу линейного программирования.

	x_{12}	x_{13}	x'_{14}	x_{23}	x_{25}	x'_{34}	x_{35}	x'_{46}	x_{56}	
Минимизировать	3	4	1	5	6	1	2	2	4	
Узел 1	1	1	1							= 50
Узел 2	–1			1	1					= 200
Узел 3		–1		–1		1	1			= –20
Узел 4			–1			–1		1		= –130
Узел 5					–1		–1		1	= –80
Узел 6								–1	–1	= –20
Верхние границы	∞	∞	30	∞	∞	50	∞	20	∞	

Соответствующая сеть после исключения нижних границ пропускных способностей дуг показана на рис. 6.32. Отметим, что данную сеть можно получить непосредственно из сети, представленной на рис. 6.30, с помощью преобразований, показанных на рис. 6.31, причем без необходимости записи в виде задачи линейного программирования.

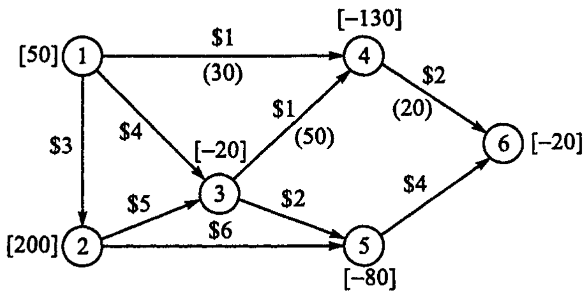


Рис. 6.32

В следующем примере представлена сетевая модель, которая исходно не удовлетворяет условию "узлового потока" (т.е. условию, что результирующий поток, проходящий через узел, равен разности выходного и входного потоков), но которую можно преобразовать в модель, удовлетворяющую этому условию, путем изменения ограничений соответствующей задачи линейного программирования.

Пример 6.6–3. (Распределение рабочих)

Агентство по найму рабочей силы имеет заказ на рабочих на 4 месяца вперед (с января по апрель) согласно следующему графику.

Месяц	Январь	Февраль	Март	Апрель
К-во рабочих	100	120	80	170

Поскольку спрос на рабочих различен в разные месяцы, возможно, экономически целесообразно нанять больше рабочих, чем требуется, в текущем месяце. Стоимость найма рабочих и удержания их в "ждущем режиме" зависит от времени трудоустройства, как показано в следующей таблице.

Время трудоустройства (месяцы)	1	2	3	4
Стоимость на одного рабочего ($)	100	130	180	220

Обозначим через x_{ij} количество рабочих, нанятых на начало i-го месяца и освобожденных на начало j-го. Например, x_{12} — это количество рабочих, нанятых в январе только на один месяц.

Чтобы сформулировать задачу линейного программирования, нужно добавить еще пятый месяц (май). Тогда, например, переменная x_{45} будет обозначать количество рабочих, нанятых в апреле на один месяц. Естественно, на май нет заказа на рабочих.

Ограничения строятся так, чтобы спрос на рабочих в месяц k можно было бы удовлетворить за счет всех рабочих x_{ij}, где $i \le k < j$. Обозначив через S_i ($S_i \ge 0$) количество рабочих, "лишних" в месяце i, получим следующую задачу линейного программирования.

	x_{12}	x_{13}	x_{14}	x_{15}	x_{23}	x_{24}	x_{25}	x_{34}	x_{35}	x_{45}	S_1	S_2	S_3	S_4	
	100	130	180	220	100	130	180	100	130	100					Min
Янв.	1	1	1	1							–1				100
Фев.		1	1	1	1	1	1					–1			120
Март			1	1		1	1	1	1				–1		80
Апр.				1			1		1	1				–1	170

В данной задаче ЛП нет той специальной структуры коэффициентов ограничений, что была в модели из примера 6.6–2. Тем не менее, эта задача имеет свою специфику, которая позволит преобразовать ее в сетевую модель. Выполним следующие действия.

1. Из n-го уравнения (ограничения) задачи создадим новое $(n + 1)$-е уравнение, умножив n-е уравнение на –1.
2. Оставляем первое уравнение без изменений.
3. Для $i = 2, 3, \ldots, n$ последовательно заменяем i-е уравнение разностью i-го и $(i - 1)$-го уравнений.

Применение описанной процедуры к задаче данного примера приводит к следующей задаче линейного программирования, которая уже имеет структуру сетевой модели.

Таким образом, задачу распределения рабочих можно представить в виде эквивалентной задачи нахождения потока минимальной стоимости в сети с ограниченной пропускной способностью (рис. 6.33). Поскольку здесь не накладываются ограничения на верхние границы пропускных способностей, эту задачу можно также решить как транспортную задачу (см. раздел 5.5).

	x_{12}	x_{13}	x_{14}	x_{15}	x_{23}	x_{24}	x_{25}	x_{34}	x_{35}	x_{45}	S_1	S_2	S_3	S_4	
	100	130	180	220	100	130	180	100	130	100					Min
Янв.	1	1	1	1							–				100
											1				
Фев.	–1				1	1	1				1	–			20
												1			
Март		–1			–1			1	1			1	–		–40
													1		
Апр.			–1			–1		–1		1			1	–	90
														1	
Май				–1			–1		–1	–1				–	–
														1	170

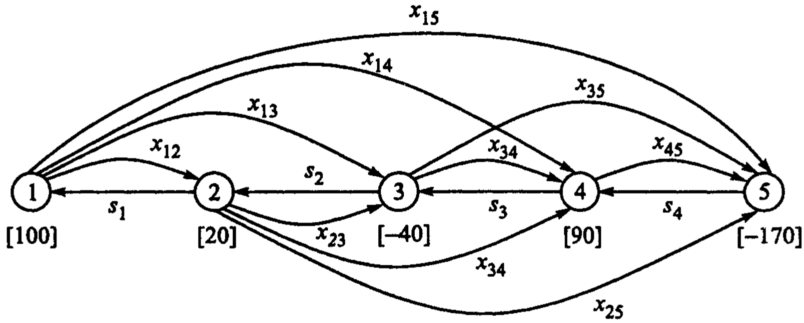


Рис. 6.33

Упражнения 6.6,b

1. Сформулируйте задачу линейного программирования, минимизирующую стоимость потока в сети, показанной на рис. 6.34, до и после исключения нижних границ пропускных способностей дуг.

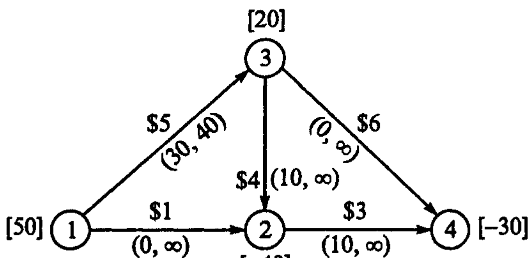


Рис. 6.34

2. С помощью непосредственной проверки найдите допустимое решение задачи определения потока минимальной стоимости в модели распределения рабочих из примера 6.6–3 (рис. 6.33). Интерпретируйте найденное решение в терминах первоначальной постановки задачи (как количество нанимаемых рабочих), проверьте выполнение ограничений модели и найдите соответствующую общую стоимость.

3. Переформулируйте задачу распределения рабочих (пример 6.6–3) в предположении, что рабочие нанимаются не менее, чем на два месяца. Запишите соответствующую задачу линейного программирования и преобразуйте ее в задачу определения потока минимальной стоимости в сети с ограниченной пропускной способностью.

4. Сформулируйте задачи ЛП и соответствующие задачи определения потока минимальной стоимости для модели распределения рабочих из примера 6.6–3, используя следующие данные о потребностях рабочих в течение последующих пяти месяцев. Стоимость найма одного рабочего в эти месяцы составит соответственно \$50, \$70, \$85, \$100 и \$130.

a)

Месяц	1	2	3	4	5
К-во рабочих	200	220	300	50	240

b)

Месяц	1	2	3	4	5
К-во рабочих	200	250	250	150	200

5. *Преобразование сети с ограниченной пропускной способностью в сеть с неограниченной пропускной способностью.* Покажите, что дугу $(i \rightarrow j)$ с ограниченным потоком $x_{ij} \le u_{ij}$ можно заменить двумя дугами, не имеющими ограничений на величину потока, — $(i \rightarrow k)$ и $(k \rightarrow j)$ с результирующим (выходным) потоком $[-u_{ij}]$ в узле k и дополнительным (входным) потоком $[+u_{ij}]$ в узле j. В результате сеть с *ограниченной* пропускной способностью преобразуется в сеть с *неограниченной* пропускной способностью, т.е. в транспортную модель (раздел 5.1). Примените описанное преобразование к сети на рис. 6.35.

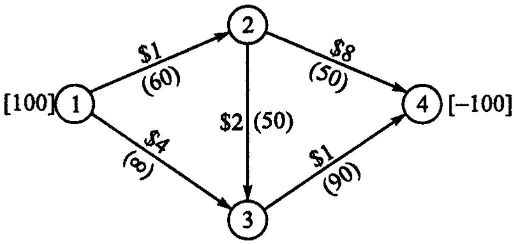


Рис. 6.35

6.6.3. Симплексный алгоритм для сетей с ограниченной пропускной способностью

Предлагаемый алгоритм повторяет в точности те же шаги, что и обычный симплекс-метод. Вместе с тем здесь учитывается специальная структура сетевой модели.

Используя определения из раздела 6.6.2, где f_i — результирующий поток через узел i, строим симплексный алгоритм на основе условия $\sum_{i=1}^{n} f_i = 0$. Это условие означает, что в сети суммарный объем предложений равен суммарному объему спроса. Мы всегда можем удовлетворить данное условие путем введения фиктивного источника или стока, которые можно связать с остальными узлами сети дугами с нулевой стоимостью и бесконечной пропускной способностью. Однако сбалансированность сети не гарантирует существования допустимого решения, поскольку этому может воспрепятствовать ограниченность пропускных способностей дуг.

Опишем шаги алгоритма нахождения потока минимальной стоимости для сетей с ограниченной пропускной способностью. Отметим, что он базируется на стандартном симплекс-методе и теории двойственности (главы 3 и 4). Здесь также полезно знакомство с симплекс-методом для задач ЛП с ограниченными переменными (раздел 7.5.2).

Шаг 0. Определяем для данной сети начальное базисное допустимое решение (множество дуг). Переходим к шагу 1.

Шаг 1. С помощью условия оптимальности симплекс-метода определяем вводимую в базис переменную (дугу). Если на основе условия оптимальности определяем, что последнее решение оптимально, вычисления заканчиваются. В противном случае переходим к шагу 2.

Шаг 2. На основе условия допустимости симплекс-метода определяем исключаемую из базиса переменную (дугу). Изменив базис, возвращаемся к шагу 1.

Сети с n узлами и нулевым результирующим потоком (т.е. при выполнении равенства $f_1 + f_2 + \ldots + f_n = 0$) соответствуют $n - 1$ *независимым* ограничениям в виде равенств. Поэтому базисное решение должно содержать $n - 1$ переменных. Можно показать, что базисное решение соответствует *остовному дереву* данной сети (см. раздел 6.3).

Вводимая переменная (дуга) на шаге 1 определяется путем вычисления разностей $z_{ij} - c_{ij}$ для всех текущих небазисных дуг (i, j). Если для всех разностей $z_{ij} - c_{ij} \le 0$, тогда текущее базисное решение оптимально. Иначе в качестве вводимой в базис переменной выбираем дугу, которой соответствует наибольшее положительное значение разности $z_{ij} - c_{ij}$.

Вычисление разностей $z_{ij} - c_{ij}$ основано на соотношениях двойственности, точно так же как в транспортной модели (см. раздел 5.3.3). Обозначим через w_i переменную задачи, двойственной к задаче ЛП (определенной в разделе 6.6.2), которая (переменная) соответствует ограничению узла i. Тогда данная двойственная задача формулируется следующим образом.

$$\text{Максимизировать } z = \sum_{i=1}^{n} f_i w_i$$

при выполнении условий

$$w_i - w_j \le c_{ij}, \quad (i, j) \in A,$$

$$\text{переменные } w_i \text{ произвольного знака } (i = 1, 2, \ldots, n).$$

Из теории линейного программирования следует, что $w_i - w_j = c_{ij}$ для любой базисной дуги (i, j). Поскольку исходная задача ЛП (раздел 6.6.2) по определению имеет одно избыточное ограничение, мы можем присвоить произвольное значение одной из переменных двойственной задачи (сравните с алгоритмом решения транспортной задачи из раздела 5.3). Для определенности положим $w_1 = 0$. Затем следует решить (базисную) систему уравнений $w_i - w_j = c_{ij}$ для нахождения остальных переменных двойственной задачи. Далее вычисляем разности $z_{ij} - c_{ij}$ для небазисных переменных согласно формуле

$$z_{ij} - c_{ij} = w_i - w_j - c_{ij}.$$

Теперь осталось показать, как определяется исключаемая из базиса переменная. Для этого рассмотрим следующий числовой пример.

Пример 6.6–4

Сеть трубопроводов связывает две станции опреснения воды с двумя городами. Ежедневное предложение опреснительных станций составляет 40 и 50 миллионов галлонов воды, города ежедневно потребляют 30 и 60 миллионов галлонов воды. Каждая станция трубопроводами соединена с каждым городом непосредственно, однако они могут также перекачивать воду в города через специальную насосную станцию. Кроме того, станция 1 может транспортировать воду на станцию 2, а город 1 — в город 2. Данная сеть сбалансирована, так как в ней суммарный спрос равен суммарному предложению. Описанная сеть показана на рис. 6.36.

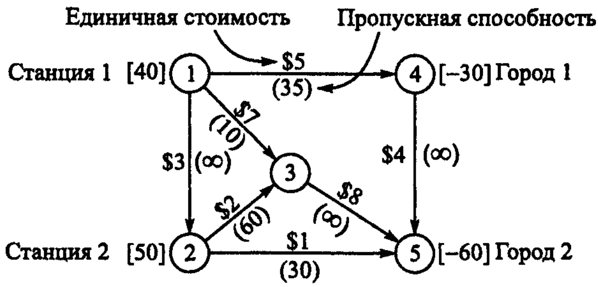
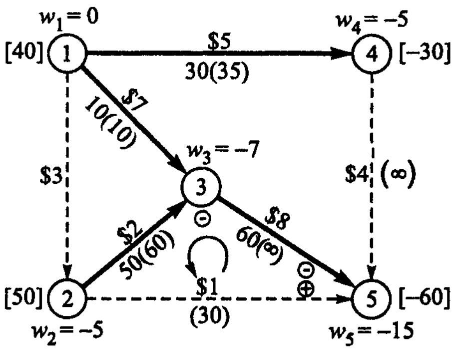


Рис. 6.36

Итерация 0

Шаг 0. *Определение начального допустимого базисного решения.* Нетрудно построить остовное дерево (на рис. 6.37 показано сплошными дугами) для рассматриваемой сети. Отсюда получаем начальное *допустимое* базисное решение. Обычно для нахождения такого решения используется метод введения искусственных переменных (подробности см. в [2]).



$z_{12} - c_{12} = 0 - (-5) - 3 = 2$

$z_{25} - c_{25} = -5 - (-15) - 1 = 9$

$z_{45} - c_{45} = -5 - (-15) - 4 = 6$

Рис. 6.37

На рис. 6.37 показано, что базисному решению соответствуют дуги (1, 3), (1, 4), (2, 3) и (3, 5) с потоками 30, 10, 50 и 60 единиц соответственно. Оставшиеся дуги (показаны пунктиром) представляют небазисные переменные. Запись $x(c)$ показывает, что через соответствующую дугу с пропускной способностью c проходит поток x.

Итерация 1

Шаг 1. *Определение вводимой в базис переменной.* При решении системы уравнений

$$w_1 = 0,$$
$$w_i - w_j = c_{ij} \text{ для базисных дуг } (i, j)$$

получим значения переменных двойственной задачи.

Дуга (1, 3): $w_1 - w_3 = 7 \Rightarrow w_3 = -7$.

Дуга (1, 4): $w_1 - w_4 = 5 \Rightarrow w_4 = -5$.

Дуга (2, 3): $w_2 - w_3 = 2 \Rightarrow w_2 = -5$.

Дуга (3, 5): $w_3 - w_5 = 8 \Rightarrow w_5 = -15$.

Теперь вычисляем разности $z_{ij} - c_{ij}$ для небазисных переменных.

Дуга (1, 2): $w_1 - w_2 - c_{12} = 0 - (-5) - 3 = 2$.

Дуга (2, 5): $w_2 - w_5 - c_{25} = (-5) - (-15) - 1 = 9$.

Дуга (4, 5): $w_4 - w_5 - c_{45} = (-5) - (-15) - 4 = 6$.

Таким образом, дуга (2, 5) будет введена в базисное решение.

Шаг 2. *Определение исключаемой из базиса переменной.* На рис. 6.37 видно, что дуга (2, 5) совместно с базисными дугами (2, 3) и (3, 5) образует цикл. *По определению остовное дерево не может содержать циклов.* Поскольку поток через дугу (2, 5) должен возрасти, необходимо выровнять потоки через дуги, составляющие цикл таким образом, чтобы новое решение осталось допустимым. Для этого поток через дугу (2, 5) пометим знаком "+", потоки через другие дуги цикла — знаком "+" или "–", в зависимости от того, будут ли совпадать направления потоков в этих дугах с направлением потока в дуге (2, 5) при обходе цикла против часовой стрелки.[1] Пометки дуг цикла показаны на рис. 6.37.

При определении максимального потока, протекающего через дугу (2, 5), необходимо придерживаться следующих правил.

1. Новый поток в текущей базисной дуге не может быть отрицательным.
2. Поток через вводимую в базис дугу не может превышать ее пропускную способность.

Применение правила 1 показывает, что потоки через дуги (2, 3) и (3, 5) нельзя уменьшить более, чем на $\min\{50, 60\} = 50$ единиц. Из правила 2 следует, что поток через дугу (2, 5) не может превышать 30 единиц

[1] Направление обхода дуг цикла всегда совпадает с направлением потока, протекающего через дугу, вводимую в базис. В данном случае это направление совпадает с направлением против часовой стрелки, но это, конечно, совсем не обязательно. — *Прим. ред.*

(пропускная способность этой дуги равна 30). Поэтому поток через дуги цикла изменится не более, чем на $\min\{30, 50\} = 30$ единиц. Таким образом, поток через дугу (2, 5) равен 30 единиц, через дугу (2, 3) $50 - 30 = 20$ единиц, а через дугу (3, 5) — $60 - 30 = 30$ единиц.

Поскольку никакая из текущих базисных переменных не приняла нулевого значения, дуга (2, 5) должна остаться небазисной, но с ненулевым значением в 30 единиц. Чтобы выполнить требование равенства нулю небазисных переменных, сделаем подстановку

$$x_{25} = 30 - x_{52},\ 0 \le x_{52} \le 30.$$

Эта подстановка изменит уравнения для потоков, протекающих через узлы 2 и 5.

Текущее уравнение для потоков узла 2: $50 + x_{12} = x_{23} + x_{25}$.

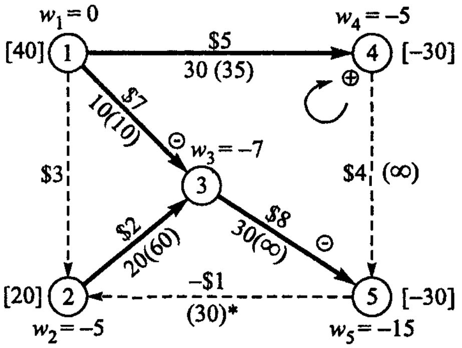
Текущее уравнение для потоков узла 5: $x_{25} + x_{35} + x_{45} = 60$.

После подстановки $x_{25} = 30 - x_{52}$ получим

Новое уравнение для потоков узла 2: $20 + x_{12} + x_{52} = x_{23}$.

Новое уравнение для потоков узла 5: $x_{35} + x_{45} = x_{52} + 30$.

Результаты этих изменений показаны на рис. 6.38. Направление потока через дугу (2, 5) изменилось на обратное (от узла 5 к узлу 2), причем, как и ожидалось, $x_{52} = 0$. Описанная подстановка также требует изменения стоимости прохождения потока по дуге (2, 5) до –\$1. Те дуги, направления потоков которых изменены на противоположные, помечены в сети звездочкой.



$z_{12} - c_{12} = 0 - (-5) - 3 = 2$

$z_{52} - c_{52} = -15 - (-5) - (-1) = -9$

$z_{45} - c_{45} = -5 - (-15) - 4 = \mathbf{6}$

Рис. 6.38

Итерация 2

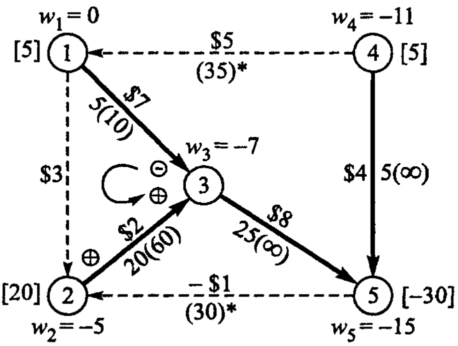
На рис. 6.38 представлены новые значения разностей $z_{ij} - c_{ij}$ (проверьте эти значения!). Очевидно, что в базис следует ввести дугу (4, 5). Введение в базис этой дуги также приводит к образованию цикла.

Величину потока через дугу (4, 5) можно увеличить до наименьшей из следующих величин.

1. *Максимальный поток* чёрез дугу (4, 5), определяемый пропускной способностью, равен ∞.
2. *Максимальное увеличение* потока через дугу (1, 4) равно 35 – 30 = 5 единиц.
3. *Максимальное уменьшение* потока через дугу (1, 3) равно 10 единиц.
4. *Максимальное уменьшение* потока через дугу (3, 5) равно 30 единиц.

Таким образом, поток через дугу (4, 5) можно увеличить до 5 единиц; эта дуга входит в базис, а дуга (1, 4) с потоком в 35 единиц исключается из базиса.

Выполнив подстановку $x_{14} = 35 - x_{41}$, получим сеть, показанную на рис. 6.39, где дуги (1, 3), (2, 3), (3, 5) и (4, 5) формируют остовное дерево сети (базисное решение). Для дуги (1, 4) с обратным направлением потока изменена стоимость прохождения потока до –\$5. Проверьте самостоятельно, что в уравнения для потоков в узлах 1 и 4 будет добавлено 5 единиц *входного* потока.

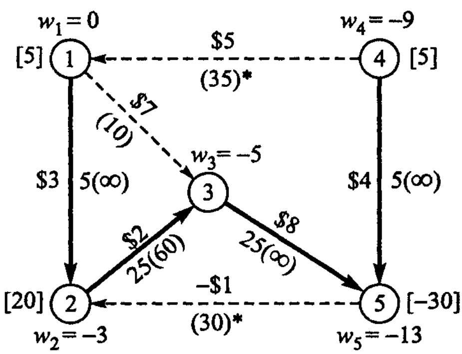


$z_{12} - c_{12} = 0 - (-5) - 3 = \mathbf{2}$
$z_{41} - c_{41} = -11 - 0 - (-5) = -6$
$z_{52} - c_{52} = -15 - (-5) - (-1) = -9$

Рис. 6.39

Итерация 3

Вычисленные новые значения разностей $z_{ij} - c_{ij}$ для небазисных дуг (1, 2), (4, 1) и (5, 2) показаны на рис. 6.39. Из этих значений вытекает, что в базис следует ввести дугу (1, 2) с потоком в 5 единиц, тогда как дуга (1, 3) исключается из базиса с нулевым значением потока. Новое решение представлено на рис. 6.40.



$z_{13} - c_{13} = 0 - (-5) - 7 = -2$
$z_{41} - c_{41} = -9 - 0 - (-5) = -4$
$z_{52} - c_{52} = -13 - (-3) - (-1) = -9$
Оптимальное решение:
$x_{12} = 5,\ x_{13} = 0$
$x_{14} = 35 - 0 = 35$
$x_{23} = 25$
$x_{25} = 30 - 0 = 30$
$x_{35} = 25,\ x_{45} = 5$
Суммарная стоимость = \$495

Рис. 6.40

Итерация 4

Из новых значений разностей $z_{ij} - c_{ij}$ (рис. 6.40) видно, что последнее решение оптимально. Значения исходных переменных получаем путем обратной подстановки, как показано на рис. 6.40.

Упражнения 6.6,с

1. Решите задачу из упр. 6.6,а(1) с помощью симплексного алгоритма для сетей с ограниченной пропускной способностью. Покажите также, что эту задачу можно решить как транспортную.
2. Решите задачу из упр. 6.6,а(2) посредством симплексного алгоритма для сетей с ограниченной пропускной способностью. Покажите также, что эту задачу можно решить как транспортную.
3. Решите задачу из упр. 6.6,а(3) с помощью симплексного алгоритма для сетей с ограниченной пропускной способностью.
4. Решите задачу из упр. 6.6,а(4) с помощью симплексного алгоритма для сетей с ограниченной пропускной способностью.
5. Решите задачу из упр. 6.6,а(5) с помощью симплексного алгоритма для сетей с ограниченной пропускной способностью.
6. Решите задачу из примера 6.6-3 с помощью симплексного алгоритма для сетей с ограниченной пропускной способностью.
7. Электрическая компания Вайоминга транспортирует по трубопроводам угольную пульпу от трех шахт (1, 2 и 3) к трем электростанциям (4, 5 и 6). Каждый трубопровод может транспортировать не более 10 тонн пульпы в час. Стоимость транспортировки одной тонны пульпы, а также предложение шахт и спрос электростанций представлены в следующей таблице.

	4	5	6	
1	\$5	\$8	\$4	**8**
2	\$6	\$9	\$12	**10**
3	\$3	\$1	\$5	**18**
	16	**6**	**14**	

Найдите оптимальную схему транспортировки угля к электростанциям.

8. На рис. 6.41 показана сеть из семи городов и расстояния между ними. С помощью симплексного алгоритма для сетей с ограниченной пропускной способностью найдите кратчайший маршрут между городами 1 и 7. (*Совет*. Положите, что узлы 1 и 7 имеют результирующие потоки [+1] и [–1] соответственно, а все остальные — нулевые результирующие потоки.)
9. Покажите, что симплексный алгоритм нахождения потока минимальной стоимости для сетей с ограниченной пропускной способностью можно применить для нахождения максимального потока в сети (см. раздел 6.5), и с его помощью решите задачу из примера 6.5–2. Для определенности положите стоимость прохождения потока от узла 4 к узлу 3 нулевой, остальные данные остаются без изменений.

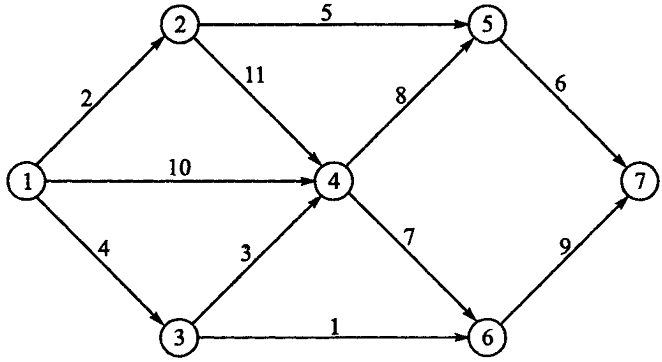


Рис. 6.41

6.7. Методы сетевого планирования

На основе сетевых моделей разработано множество методов планирования, составления временны́х расписаний и управления проектами. Наиболее известные — метод критического пути (critical path method, сокращенно CPM), а также система планирования и руководства программами разработок (program evaluation and review technique, сокращенно PERT). В этих методах проекты рассматриваются как совокупность некоторых взаимосвязанных процессов (видов деятельности, этапов или фаз выполнения проекта), каждый из которых требует определенных временны́х и других ресурсов. В методах CPM и PERT проводится анализ проектов для составления временны́х графиков распределения фаз проектов. На рис. 6.42 в обобщенной форме показаны основные этапы выполнения этих методов. На первом этапе определяются отдельные процессы, составляющие проект, их отношения предшествования (т.е. какой процесс должен предшествовать другому) и их длительность. Далее проект представляется в виде сети, показывающей отношения предшествования среди процессов, составляющих проект. На третьем этапе на основе построенной сети выполняются вычисления, в результате которых составляется временной график реализации проекта.

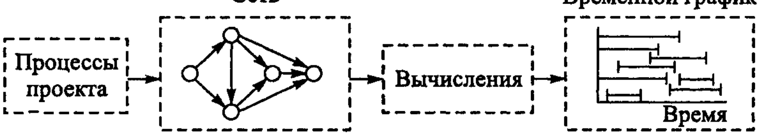


Рис. 6.42

Методы CPM и PERT, которые разрабатывались независимо друг от друга, отличаются тем, что в методе критического пути длительность каждого этапа проекта является детерминированной, тогда как в системе планирования PERT — стохастической. В этой книге будет рассмотрен только метод критического пути.

6.7.1. Построение сети проекта

Каждый процесс проекта обозначается в сети дугой, ориентированной по направлению выполнения проекта. Узлы сети (также называемые *событиями*) устанавливают отношения предшествования среди процессов проекта.

Построение сети проекта основано на следующих правилах.

Правило 1. *Каждый процесс в проекте представим одной и только одной дугой.*

Правило 2. *Каждый процесс идентифицируется двумя концевыми узлами.*

На рис. 6.43 показано, как с помощью фиктивного процесса можно представить два параллельных (конкурирующих) процесса A и B. По определению фиктивный процесс (который на схеме сети обычно обозначается пунктирной дугой) не поглощает временных или других ресурсов. Вставив фиктивный процесс одним из четырех способов, показанных на рис. 6.43, мы получаем возможность идентифицировать процессы A и B, по крайней мере, одним уникальным концевым узлом (как требует правило 2).

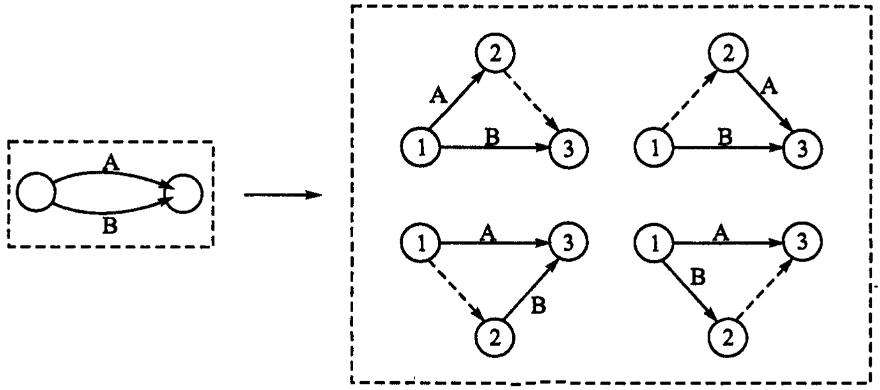


Рис. 6.43

Правило 3. *Для поддержания правильных отношений предшествования при включении в сеть любого процесса необходимо ответить на следующие вопросы.*

a) *Какой процесс непосредственно предшествует текущему?*

b) *Какой процесс должен выполняться после завершения текущего процесса?*

c) *Какой процесс конкурирует (выполняется параллельно) с текущим?*

Ответы на эти вопросы, возможно, потребуют включения в сеть фиктивных процессов, чтобы правильно отобразить последовательность выполнения процессов. Предположим, например, что четыре процесса должны удовлетворять следующим условиям.

1. Процесс C должен начаться сразу после завершения процессов A и B.
2. Процесс E должен начаться непосредственно после завершения процесса B.

На рис. 6.44,*а* показано неправильное представление наших процессов, так как из него следует, что процесс E должен начаться после завершения как процесса B, так и A. На рис. 6.44,*б* показано, как с помощью фиктивного процесса D разрешить эту коллизию.

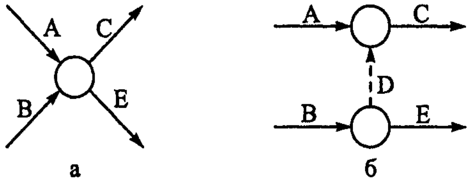


Рис. 6.44

Пример 6.7–1

Издатель имеет контракт с автором на издание его книги. Ниже представлена последовательность (упрощенная) процессов, приводящая к реализации проекта издания книги. Необходимо разработать сеть для этого проекта.

	Процесс	Предшествующий процесс	Длительность (недели)
A:	Прочтение рукописи редактором	—	3
B:	Пробная верстка отдельных страниц книги	—	2
C:	Разработка обложки книги	—	4
D:	Подготовка иллюстраций	—	3
E:	Просмотр автором редакторских правок и сверстанных страниц	A, B	2
F:	Верстка книги (создание макета книги)	E	4
G:	Проверка автором макета книги	F	2
H:	Проверка автором иллюстраций	D	1
I:	Подготовка печатных форм	G, H	2
J:	Печать и брошюровка книги	C, I	4

На рис. 6.45 показана сеть, представляющая взаимосвязь процессов данного проекта. Фиктивный процесс (2, 3) введен для того, чтобы "развести" конкурирующие процессы A и B. Номера узлов сети возрастают в направлении выполнения проектов.

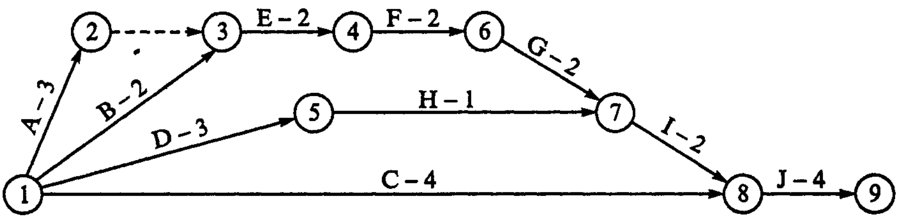


Рис. 6.45

Упражнения 6.7,а

1. Постройте сеть проекта, содержащего процессы, помеченные латинскими буквами от A до L, с учетом следующих отношений предшествования.

   a) Первые процессы A, B и C проекта могут выполняться параллельно.
   b) Процессы A и B предшествуют процессу D.
   c) Процесс B предшествует процессам E, F и H.
   d) Процессы F и C предшествуют процессу G.
   e) Процессы E и H — процессам I и J.
   f) Процессы C, D, F и J — процессу K.
   g) Процесс K предшествует процессу L.
   h) Проект заканчивается после завершения процессов I, G и L.

2. Постройте сеть проекта, содержащего процессы, помеченные латинскими буквами от A до P, с учетом следующих отношений предшествования.

   a) Первые процессы A, B и C проекта могут выполняться параллельно.
   b) Процессы D, E и F следуют за процессом A.
   c) Процессы I и G — за процессами B и D.
   d) Процесс H следует за процессами C и G.
   e) Процессы K и L — за процессом I.
   f) Процесс J следует за процессами E и H.
   g) Процессы M и N следуют за процессом F, но не могут начаться до завершения процессов E и H.
   h) Процесс O следует за процессами M и I.
   i) Процесс P — за процессами J, L и O.
   j) Проект заканчивается после завершения процессов K, N и P.

3. Фундамент здания должен состоять из четырех железобетонных блоков. При создании каждого блока выполняются следующие виды работ: (1) земляные (рытье котлована), (2) закладка арматуры, (3) заливка бетона. Земляные работы под очередной железобетонный блок не могут начаться до тех пор, пока не будут завершены все работы на предшествующем блоке. Такое же ограничение существует для бетонных работ. Постройте сеть процессов закладки фундамента.

4. Допустим, в предыдущем упражнении одновременно с земляными работами может быть выполнено 10% работ по созданию водопроводно-канализационной сети здания, но в любом случае эти работы должны начаться до заливки фундаментных блоков бетоном. Кроме того, по 5% этих работ можно выполнить при достижении 95% готовности каждого фундаментного блока. Постройте сеть строительных работ.

5. Выявление общественного мнения предполагает выполнение следующих работ: разработка и печать опросных листов, наем и транспортировка интервьюеров, выбор участников опроса, пересылка заполненных опросных листов и анализ полученных данных. Постройте сеть этих работ, предварительно сформулировав необходимые предположения о последовательности их выполнения.

6. В следующей таблице приведены работы (процессы), выполняемые при строительстве нового каркасного дома. Разработайте сеть этих работ.

	Процесс	Предшествующий процесс	Длительность (дни)
A:	Очистка строительного участка	—	1
B:	Завоз оборудования	—	2
C:	Земляные работы	A	1
D:	Заливка фундамента	C	2
E:	Наружные водопроводно-канализационные работы	B, C	6
F:	Возведение каркаса дома	D	10
G:	Прокладка электропроводки	F	3
H:	Создание перекрытий	G	1
I:	Создание каркаса крыши	F	1
J:	Внутренние водопроводно-канализационные работы	E, H	5
K:	Покрытие крыши	I	2
L:	Наружные изоляционные работы	F, J	1
M:	Вставка окон и наружных дверей	F	2
N:	Обкладка дома кирпичом	L, M	4
O:	Штукатурка стен и потолков	G, J	2
P:	Облицовка стен и потолков	O	2
Q:	Изоляция крыши	I, P	1
R:	Окончание внутренних отделочных работ	P	7
S:	Окончание наружных отделочных работ	I, N	7
T:	Ландшафтные работы	S	3

7. Компания готовит бюджет производства нового изделия. В следующей таблице представлены этапы подготовки бюджета и их длительность. Постройте сеть этапов подготовки бюджета.

	Процесс	Предшествующий процесс	Длительность (дни)
A:	Прогнозирование объема продаж	—	10
B:	Изучение рынка конкурирующих товаров	—	7
C:	Доводка изделия	A	5
D:	Подготовка производственного плана	C	3
E:	Оценка стоимости производства	D	2
F:	Определение отпускной цены	B, E	1
G:	Подготовка бюджета	E, F	14

8. Расширение участка дороги требует переноса воздушной электролинии (длиной 1700 футов). В следующей таблице приведены этапы выполнения работ по замене электролинии. Постройте соответствующую сеть.

	Процесс	Предшествующий процесс	Длительность (дни)
A:	Определение объема работ	—	1
B:	Извещение пользователей о временном отключении электросети	A	0.5
C:	Подвозка материалов и оборудования	A	1
D:	Предварительные работы	A	1
E:	Заготовка опор и материалов	C, D	3
F:	Развозка опор	E	3.5
G:	Определение нового местоположения опор	D	0.5
H:	Разметка местоположения опор	G	0.5
I:	Земляные работы для установки новых опор	H	3
J:	Установка новых опор	F, I	4
K:	Ограждение старой линии	F, I	1
L:	Прокладка новых проводов	J, K	2
M:	Обустройство новой линии	L	2
N:	Натяжка проводов	L	2
O:	Подрезка деревьев	D	2
P:	Отключение старой электролинии	B, M, N, O	0.2
Q:	Подключение новой электролинии	P	0.5
R:	Уборка территории	Q	1
S:	Возврат материалов и оборудования	I	2

9. В следующей таблице показаны этапы покупки нового автомобиля. Постройте соответствующую сеть.

	Процесс	Предшествующий процесс	Длительность (дни)
A:	Принятие окончательного решения о покупке автомобиля	—	3
B:	Поиск потенциального покупателя имеющегося автомобиля	A	14
C:	Составление списка желаемых моделей машин	A	1
D:	Исследование желаемых моделей	C	3
E:	Консультации у автомехаников	C	1

	Процесс	Предшествующий процесс	Длительность (дни)
F:	Сбор рекламных материалов продавцов автомобилей	C	2
G:	Обобщение полученной информации	D, E, F	1
H:	Выбор трех наиболее подходящих моделей	G	1
I:	Знакомство "в натуре" с выбранными моделями	H	3
J:	Сбор финансовой информации	H	2
K:	Выбор одного автомобиля	I, J	2
L:	Выбор продавца автомобиля	K	2
M:	Выбор автомобиля желаемого цвета	L	4
N:	Повторная дорожная проверка выбранной модели	L	1
O:	Покупка нового автомобиля	B, M, N	3

6.7.2. Метод критического пути

Конечным результатом применения метода критического пути будет построение временного графика выполнения проекта (см. рис. 6.42). Для этого проводятся специальные вычисления, в результате чего получаем следующую информацию.

1. Общая длительность выполнения проекта.
2. Разделение множества процессов, составляющих проект, на *критические* и *некритические*.

Процесс является **критическим**, если он не имеет "зазора" для времени своего начала и завершения. Таким образом, чтобы весь проект завершился без задержек, необходимо, чтобы все критические процессы начинались и заканчивались в строго определенное время. Для **некритического** процесса возможен некоторый "дрейф" времени его начала, но в определенных границах, когда время его начала не влияет на длительность выполнения всего проекта.

Для проведения необходимых вычислений определим **событие** как точку на временной оси, где завершается один процесс и начинается другой. В терминах сети, событие — это сетевой узел. Нам понадобятся также следующие определения и обозначения.

$\square_j$ — самое раннее возможное время наступления события j,
Δ_j — самое позднее возможное время наступления события j,
D_{ij} — длительность процесса (i, j).

Вычисление критического пути включает два этапа (прохода). При **проходе вперед** вычисляются *самые ранние времена* наступления событий, а при **проходе назад** — *самые поздние времена* наступления тех же событий.

Проход вперед. Здесь вычисления начинаются в узле 1 и заканчиваются в последнем узле n.

Начальный шаг. Полагаем $\square_1 \equiv 0$; это указывает на то, что проект начинается в нулевой момент времени.

Основной шаг j. Для узла j определяем узлы p, q, ..., v, *непосредственно* связанные с узлом j процессами (p,j), (q,j), ..., (v,j), для которых уже вычислены самые ранние времена наступления соответствующих событий. Самое раннее время наступления события j вычисляется по формуле

$$\square_j = \max\{\square_p + D_{pj}, \square_q + D_{qj}, ..., \square_v + D_{vj}\}.$$

Проход вперед завершается, когда будет вычислена величина $\leq_n$ для узла n. По определению величина $\leq_j$ равна самому длинному пути (длительности) от начала проекта до узла (события) j.

Проход назад. В этом проходе вычисления начинаются в последнем узле n и заканчиваются в узле 1.

Начальный шаг. Полагаем $\Delta_n \equiv \square_n$; это указывает, что самое раннее и самое позднее времена для завершения проекта совпадают.

Основной шаг j. Для узла j определяем узлы p, q, ..., v, *непосредственно* связанные с узлом j процессами (j, p), (j, q), ..., (j, v), для которых уже вычислены самые поздние времена наступления соответствующих событий. Самое позднее время наступления события j вычисляется по формуле

$$\Delta_j = \min\{\Delta_p - D_{jp}, \Delta_q - D_{jq}, ..., \Delta_v - D_{jv}\}.$$

Проход назад завершается при вычислении величины Δ_1 для узла

Процесс (i, j) будет *критическим*, если выполняются три условия.

1. $\Delta_i = \square_i$
2. $\Delta_j = \square_j$
3. $\Delta_j - \Delta_i = \square_j - \square_i = D_{ij}$

Если эти условия не выполняются, то процесс *некритический.*

Критические процессы должны образовывать непрерывный путь через всю сеть от начального события до конечного.

Пример 6.7–2

Найдем критический путь для сети проекта, показанной на рис. 6.46. Длительность всех процессов дана в днях.

Проход вперед

Узел 1. Полагаем $\square_1 = 0$.

Узел 2. $\square_2 = \square_1 + D_{12} = 0 + 5 = 5$.

Узел 3. $\square_3 = \max\{\square_1 + D_{13}, \square_2 + D_{23}\} = \max\{0 + 6, 5 + 3\} = 8$.

Узел 4. $\square_4 = \square_2 + D_{24} = 5 + 8 = 13$.

Узел 5. $\square_5 = \max\{\square_3 + D_{35}, \square_4 + D_{45}\} = \max\{8 + 2, 13 + 0\} = 13$.

Узел 6. $\square_6 = \max\{\square_3 + D_{36}, \square_4 + D_{46}, \square_5 + D_{56}\} =$
$= \max\{8 + 11, 13 + 1, 13 + 12\} = 25$.

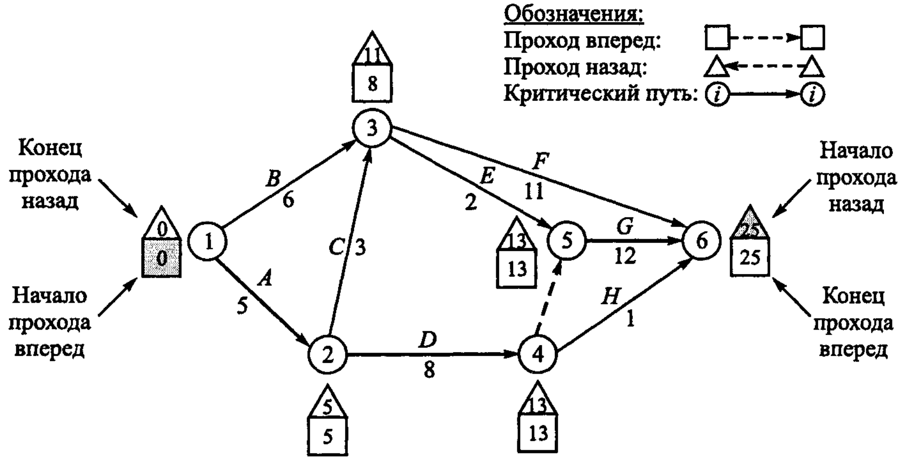


Рис. 6.46

Таким образом, расчеты показывают, что проект можно выполнить за 25 дней.

Проход назад

Узел 6. Полагаем $\Delta_6 = \square_6 = 25$.

Узел 5. $\Delta_5 = \Delta_6 - D_{56} = 25 - 12 = 13$.

Узел 4. $\Delta_4 = \min\{\Delta_6 - D_{46}, \Delta_5 - D_{45}\} = \min\{25 - 1, 13 - 0\} = 13$.

Узел 3. $\Delta_3 = \min\{\Delta_6 - D_{36}, \Delta_5 - D_{35}\} = \min\{25 - 11, 13 - 2\} = 11$.

Узел 2. $\Delta_2 = \min\{\Delta_4 - D_{24}, \Delta_3 - D_{23}\} = \min\{13 - 8, 11 - 3\} = 5$.

Узел 1. $\Delta_1 = \min\{\Delta_3 - D_{13}, \Delta_2 - D_{12}\} = \min\{11 - 6, 5 - 5\} = 0$.

Вычисления без ошибок всегда приводят к результату $\Delta_1 = 0$.

Результаты вычислений, выполняемых при проходах вперед и назад, показаны на рис. 6.46. Правила определения критических процессов показывают, что критический путь составляют процессы $1 \to 2 \to 4 \to 5 \to 6$, т.е. этот путь проходит от начального узла 1 до конечного узла 6. Сумма длительностей критических процессов (1, 2), (2, 4), (4, 5) и (5, 6) равна длительности всего проекта (т.е. 25 дней). Отметим, что процесс (4, 6) удовлетворяет первым двум условиям критического пути ($\Delta_4 = \square_4 = 13$ и $\Delta_6 = \square_6 = 25$), но не удовлетворяет третьему условию ($\square_6 - \square_4 \neq D_{46}$). Поэтому данный процесс не является критическим.

Упражнения 6.7,b

1. Найдите критический путь для сети проекта, показанной на рис. 6.47.
2. Найдите критические пути для сетей проектов, показанных на рис. 6.48.

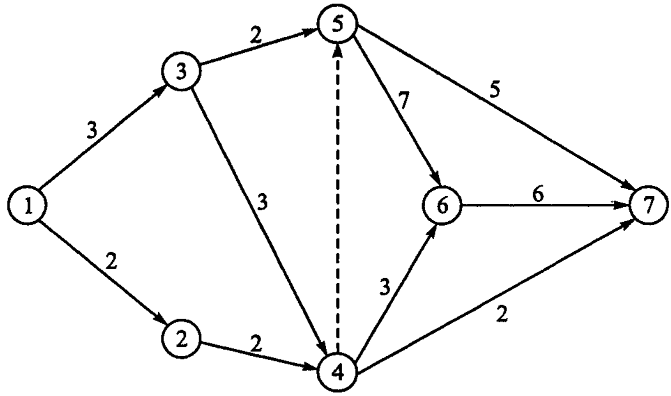
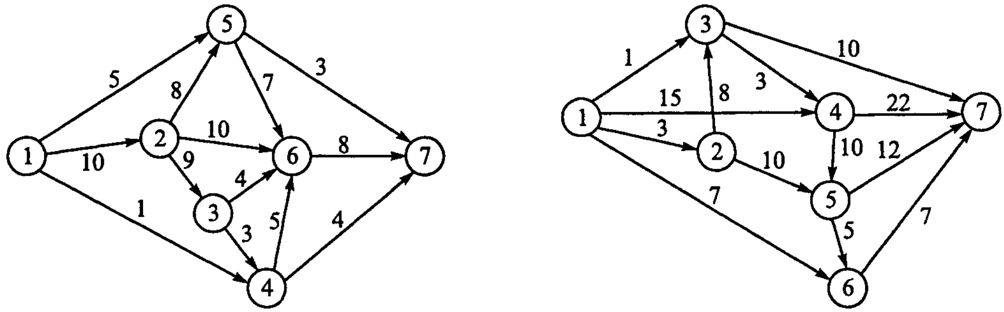


Рис. 6.47



Проект а) Проект б)

Рис. 6.48

3. Найдите критический путь в задаче из упр. 6.7,а(6).
4. Найдите критический путь в задаче из упр. 6.7,а(8).
5. Найдите критический путь в задаче из упр. 6.7,а(9).

6.7.3. Построение временного графика

В этом разделе показано, как на основе данных, полученных расчетным путем в предыдущем разделе, строится временной график последовательного выполнения проекта. Мы уже знаем, что $\square_i$ для процесса (i, j) указывает на *самое раннее время начала* этого процесса, а Δ_j — на *самое позднее время завершения* процесса. Таким образом, пара величин $(\square_i, \Delta_j)$ ограничивает максимальный интервал времени, в течение которого может выполняться процесс (i, j).

Построение предварительного графика. Метод построения предварительного временного графика выполнения проекта покажем на следующем примере.

Пример 6.7–3

Построим временн*о*й график проекта из примера 6.7–2 (рис. 6.46).

Предварительный временн*о*й график проекта можно начертить, используя максимальные интервалы выполнения каждого процесса. В результате получим график, представленный на рис. 6.49. Сделаем два замечания.

1. Критические процессы (показаны на графике сплошными линиями) располагаются последовательно друг за другом без временн*ы*х зазоров и перекрытий. Таким образом, их суммарная длительность равна длительности выполнения всего проекта (в данном случае, 25 дней).
2. Некритические процессы (показаны на графике пунктирными линиями) представлены максимальными интервалами выполнения, которые превышают реальную длительность выполнения этих процессов. Поэтому необходимо каким-то образом определиться с началом выполнения этих процессов.

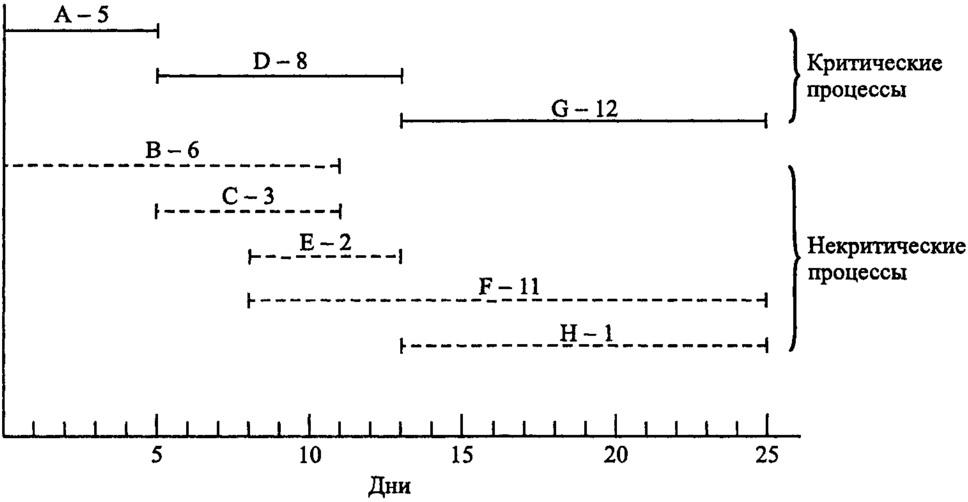


Рис. 6.49

Как выбрать время начала выполнения некритического процесса? Обычно предпочитают начинать некритические процессы (по возможности) в самый ранний срок. В этом случае остается запас времени (остаток максимального интервала выполнения), который можно использовать для решения неожиданно возникших во время выполнения процесса проблем. Вместе с тем при необходимости можно перенести начало выполнения какого-либо процесса. Допустим, если в нашем примере во время выполнения процессов E и F (рис. 6.49) используется одно и то же оборудование, причем в каждый момент времени его можно задействовать только для одного процесса, тогда можно исключить временн*о*е наложение этих процессов, начав процесс F после завершения E.

Если на некритические процессы не накладываются какие-либо дополнительные ограничения и все они начинаются в самый ранний момент времени, то временн*о*й график проекта строится автоматически. Однако в этом случае могут нарушаться некоторые отношения предшествования. В частности, в нашем примере (см. рис. 6.46) процесс C должен быть завершен до начала процесса E. Но максимальные интервалы времени выполнения этих процессов перекрываются, поэтому и реальные интервалы времени их выполнения также могут перекрываться. Поэтому необходимо предусмотреть какие-нибудь "красные флажки", которые автоматически указывали бы, когда тот или иной процесс может начинаться без нарушения отношений предшествования с другими процессами. Далее мы покажем, как для этого использовать *запасы времени* отдельных процессов.

Определение запасов времени. Запас времени некритического процесса — это часть максимального интервала времени выполнения этого процесса (который, напомним, больше реальной длительности процесса). Различают **общий запас времени** и **свободный запас времени** процесса.

На рис. 6.50 показана разность между этими запасами времени процесса (i, j) — общим (TF_{ij}) и свободным (FF_{ij}). Общий запас времени процесса (i, j) определяется как превышение над длительностью выполнения этого процесса интервала времени от *самого раннего* момента осуществления события i до *самого позднего* времени осуществления события j, т.е.

$$TF_{ij} = \Delta_j - \square_i - D_{ij}.$$

Свободный запас времени процесса (i, j) определяется как превышение над длительностью выполнения этого процесса интервала времени — от *самого раннего* момента осуществления события i до *самого раннего* времени осуществления события j, т.е.

$$FF_{ij} = \square_j - \square_i - D_{ij}.$$

По определению $FF_{ij} \le TF_{ij}$.

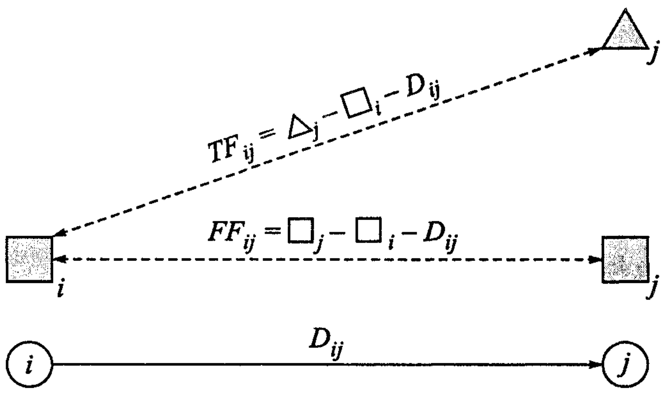


Рис. 6.50

Правило "красного флажка". *Для некритического процесса* (i, j)

а) *если* $FF_{ij} = TF_{ij}$, *тогда данный процесс может выполняться в любое время внутри максимального интервала* $(\square_i, \Delta_j)$ *без нарушения отношений следования;*

б) *если $FF_{ij} < TF_{ij}$, тогда без нарушения отношений следования данный процесс может начаться со сдвигом, не превышающим FF_{ij}, относительно самого раннего момента начала процесса $\square_i$. Сдвиг начала процесса на величину времени, превышающую FF_{ij} (но не более TF_{ij}), должен сопровождаться равным сдвигом относительно $\leq_j$ всех процессов, начинающихся с события j.*

Это правило означает, что некритический процесс (i, j) помечается "красным флажком" только тогда, когда $FF_{ij} < TF_{ij}$. Этот флажок принимается во внимание при сдвиге начала процесса относительно самого раннего времени $\leq_i$ на такую величину, при которой следует рассчитывать сдвиг процессов, следующих из узла j.

Пример 6.7–4

Вычислим запасы времени для некритических процессов в сети проекта из примера 6.7–2 и на основе этих расчетов построим окончательный временной график проекта.

Общие и свободные запасы времени некритических процессов представлены в следующей таблице. Такие расчеты можно проводить непосредственно на сети проекта, как показано на рис. 6.50.

Некритический процесс	Длительность процесса	Общий запас времени (TF)	Свободный запас времени (FF)
B (1, 3)	6	11 – 0 – 6 = 5	8 – 0 – 6 = 2
C (2, 3)	3	11 – 5 – 3 = 3	8 – 5 – 3 = 0
E (3, 5)	2	13 – 8 – 2 = 3	13 – 8 – 2 = 3
F (3, 6)	11	25 – 8 – 11 = 6	25 – 8 – 11 = 6
H (4, 6)	1	25 – 13 – 1 = 11	25 – 13 – 1 = 11

Правило "красного флажка" следует применять только к процессам B и C, поскольку для них $FF < TF$. Оставшиеся процессы (E, F и H) имеют $FF = TF$, поэтому они могут выполняться в любое время внутри своих максимальных интервалов времени выполнения.

Рассмотрим процесс B, помеченный "красным флажком". Поскольку для этого процесса $TF = 5$ дней, он может начаться в любой день из интервала 0–5 дней от начала выполнения всего проекта (рис. 6.49). Но если $FF = 2$ дня, то, поскольку процесс B начнется в 0-й, 1-й или 2-й день от начала выполнения проекта, это не окажет никакого эффекта на последующие процессы E и F. Однако если процесс B начнется в $(2 + \Delta)$-й день $(2 + \Delta < 5)$, начало выполнения процессов E и F необходимо сдвинуть от самого раннего срока их начала (8-й день от начала выполнения проекта) на величину, не меньшую Δ; только при таком условии не нарушатся отношения следования между процессами B, E и F.

Для помеченного "красным флажком" процесса C имеем $FF = 0$. Это означает, что *любой* сдвиг начала выполнения этого процесса должен сопровождаться таким же (не меньшим) сдвигом начала выполнения процессов E и F.

Упражнения 6.7,с

1. Дан процесс (i, j) длительностью D_{ij}, самый ранний срок его начала — $\square_i$ и самый поздний срок его завершения — Δ_j. Определите самый ранний срок завершения этого процесса и самый поздний срок его начала.
2. Чему равны общий и свободный запасы времени критических процессов.
3. Для следующих процессов определите максимальный сдвиг начала их выполнения (относительно самого раннего срока начала выполнения), который не нарушает никаких отношений следования с другими процессами.

   a) $TF = 10, FF = 10, D = 4$.
   b) $TF = 10, FF = 5, D = 4$.
   c) $TF = 10, FF = 0, D = 4$.

4. В задаче из примера 6.7–4 на основе значений запасов времени ответьте на следующие вопросы.

   a) Пусть процесс B начался в 1-й день от начала выполнения всего проекта, а процесс C — в 5-й день (рис. 6.49). Найдите самый ранний срок начала процессов E и F.
   b) Пусть процесс B начался на 3-й день от начала выполнения проекта, а процесс C — на 7-й. Найдите самый ранний срок начала процессов E и F.
   c) Может ли процесс B начаться после 6-го дня от начала выполнения проекта?

5. Пусть в проекте из примера 6.7–2 (рис. 6.46) длительность процессов B и F возросла до 20 и 25 дней соответственно.

   a) Определите критический путь.
   b) Найдите общие и свободные запасы времени для некритических процессов и пометьте при необходимости их "красными флажками".
   c) Пусть процесс A начался на 5-й день от начала выполнения всего проекта. Определите по возможности самые ранние сроки начала процессов C, D, E и H.
   d) Предположим, для выполнения процессов F, G и H необходимо одно и то же оборудование. Определите минимально необходимое количество этого оборудования.а

6. Вычислите запасы времени и расставьте "красные флажки" процессам проектов, показанных на рис. 6.48. Затем постройте временные графики при соблюдении следующих условий.

   Проект а

   i) Процесс (1, 5) не может начаться раньше 14-го момента времени.
   ii) Процессы (5, 6) и (5, 7) используют одинаковое оборудование, которое в любой момент времени может использоваться только в одном процессе.
   iii) Все другие процессы начинаются так рано, как только возможно.

   Проект б

   iv) Процесс (1, 3) должен начаться так рано, как только возможно, с учетом того, что процессы (1, 2), (1, 3) и (1, 6) используют одинаковое оборудование, которое в любой момент времени может использоваться только в одном процессе.

v) Все другие процессы начинаются так рано, как только возможно.

6.8. Заключение

В этой главе рассмотрены многочисленные приложения сетевых моделей. Хотя большинство сетевых задач можно решать как задачи линейного программирования, для их эффективного решения разработаны специальные методы, учитывающие специальную структуру сетевых моделей.

Симплексный метод нахождения потока наименьшей стоимости в сети с ограниченной пропускной способностью — наиболее специализированный среди рассмотренных в этой главе. Но именно его "специализация" позволяет наиболее эффективно решать указанный класс задач.

Литература

1. Ahuja R., Magnati T., Orlin J. *Network Flows: Theory, Algorithms and Applications*, Prentice Hall, Upper Saddle River, N.J., 1993.
2. Bazaraa M., Jarvis J., Sherali M. *Linear Programming and Network Flows*, 2nd ed., Wiley, New York, 1990.
3. Evans J.R., Mineka E. *Optimization Algorithms for Networks and Graphs*, 2nd ed., Marcel Dekker, New York, 1992.
4. Murty K. *Network Programming*, Prentice Hall, Upper Saddle River, N.J., 1992.

Литература, добавленная при переводе

Ахо А.В., Хопкрофт Дж. Э., Ульман Дж. Д. *Структуры данных и алгоритмы.* —М.: Издательский дом "Вильямс", 2000.[1]

Форд Л.Р., Фалкерсон Д.Р. *Потоки в сетях.* — М.: Мир, 1966.

Комплексные задачи

■ **6-1.** Любитель свежего воздуха, житель Сан-Франциско (СФ), планирует во время своего 15-дневного отпуска посетить четыре национальных парка: Йосемитский (ЙО), Йеллоустонский (ЙЕ), Гренд-Тетон (ГТ) и Маунт Рушмор (МР). Во время путешествия, которое начнется и закончится в Сан-Франциско, он планирует посетить парки в таком порядке: СФ→ЙО→ЙЕ→ГТ→МР→СФ. На осмотр каждого парка отводится 2 дня. От одного парка до другого можно добраться либо самолетом, либо автомобилем. Если пользоваться самолетом, то перелет между любыми парками (а также между парками и Сан-Франциско) занимает примерно полдня. Если путешествовать на автомобиле, то маршрут СФ – ЙО занимает полдня, ЙО – ЙЕ — 3 дня, ЙЕ – ГТ — один день пути, ГТ – МР — два дня и возвращение из МР в СФ требует 3 дня. В общем случае проезд на автомобиле дешевле перелета на самолете, но, естествен-

[1] В данной книге представлены (с вычислительной точки зрения) все рассмотренные в данной главе алгоритмы. — *Прим. ред.*

но, путешествие на автомобиле занимает больше времени. Разработайте наиболее дешевый маршрут посещения национальных парков (т.е. определите вид транспорта на каждом этапе путешествия) с учетом того, что длительность путешествия не может превышать 15 дней. В следующей таблице приведена стоимость проезда на автомобиле и перелета на самолете.

	Стоимость полета в					Стоимость проезда в				
Из	СФ	ЙО	ЙЕ	ГТ	МР	СФ	ЙО	ЙЕ	ГТ	МР
СФ	—	150	350	380	450	—	130	175	200	230
ЙО	150	—	400	290	340	130	—	200	145	180
ЙЕ	350	400	—	150	320	175	200	—	70	150
ГТ	380	290	150	—	300	200	145	70	—	100
МР	450	340	320	300	—	230	180	150	100	—

■ **6-2.**[1] Некто желает подарить большое количество ценных книг публичной библиотеке. Все книги имеют различные размеры по высоте корешка: 12, 10, 8 и 6 дюймов. Заведующий библиотекой подсчитал, что для размещения книг высотой 12 дюймов необходимы полки общей длиной 12 футов, для книг высотой 10 дюймов — полки длиной 18 футов, для книг высотой 8 дюймов — полки длиной 9 футов и для книг высотой 6 дюймов — полки длиной 10 футов. Цена книжных полок состоит из фиксированной цены и цены, рассчитываемой в зависимости от суммарной длины полок, как показано в следующей таблице.

Высота полки (дюймы)	Фиксированная цена ($)	Цена ($) за фут длины полки
12	25	5.50
10	25	4.50
8	22	3.50
6	22	2.50

Сколько и каких полок необходимо для библиотеки, если учесть, что книги меньшего размера могут храниться на полках для книг большего размера?

■ **6-3.** Пароходной компании необходимо доставить пять партий груза из портов A, B и C в порты D и E. Сроки доставки грузов приведены в следующей таблице.

Партия груза	Маршрут доставки	Срок доставки (дни)
1	из A в D	10
2	из A в E	15
3	из B в D	4
4	из B в E	5
5	из C в E	18

В следующей таблице приведено время перехода (дни) между портами (обратный переход, как правило, требует меньшего времени).

[1] Основано на материалах статьи Ravindran A. "On Compact Storage in Libraries", *Opsearch*, Vol. 8, No. 3, pp. 245–252.

	A	B	C	D	E
A				3	4
B				3	2
C				3	5
D	2	2	2		
E	3	1	4		

Компания планирует минимизировать количество судов, необходимых для перевозки данных партий груза.

■ **6-4.**[1] Несколько человек решили основать брокерскую фирму для игры на бирже с ценными бумаги. Брокеры работали по свободной финансовой системе, что позволяло им проводить многочисленные сделки между самими брокерами, включая покупку и продажу ценных бумаг, предоставление денежных ссуд и займов под проценты. Для всей этой группы брокеров, в целом, основным источником доходов были комиссионные, получаемые от продажи ценных бумаг сторонним клиентам.

Со временем эти спекулятивные сделки вышли из-под контроля, что привело всех брокеров к банкротству. К тому же, финансовое положение брокерской фирмы было таково, что все деньги брокеров были вложены во внешних клиентов и сделки между самими брокерами, причем таким образом, что практически каждый брокер стал должником другого.

Брокеры, чьи активы позволяли погасить долги, были объявлены платежеспособными. Остальные через судебные инстанции должны были погасить свои долги в интересах сторонних клиентов. Поскольку активы и авуары несостоятельных брокеров меньше общего объема долгов, долги погашаются пропорционально их объемам.

Из-за финансовых затруднений неплатежеспособной группы брокеров судебные инстанции постановили, что заключенные ранее сделки должны выполняться только для удовлетворения определенных судом требований, поскольку брокеры не имеют собственных источников капитала. В частности, судебные инстанции требуют свести количество погашаемых сделок между брокерами к минимуму. Это означает, что если брокер А должен брокеру В сумму X, а брокер В — брокеру А сумму Y, то эти взаимные долги сведутся к одному с суммой долга $|X - Y|$. Эта сумма считается долгом А перед В, если $X > Y$, и долгом В перед А, когда $X < Y$. Если $X = Y$, долг погашен. Эта идея погашения взаимных долгов распространяется на все долги между брокерами.

Каковы ваши предложения по выходу из данной финансовой ситуации? В частности, ответьте на следующие вопросы.

1. Как рассчитать доли возвращаемых долгов?
2. К какому минимальному количеству можно свести взаимные долги между брокерами?

---

[1] Задача основана на материалах статьи Taha H. "Operations Research Analysis of a Stock Market Problem", *Computers and Operations Research*, Vol. 18, No. 7, pp.597–602, 1991.

Теория линейного программирования

7.1. Введение

Эта глава представляет строгий математический фундамент теории линейного программирования (ЛП). Здесь будет обоснован симплекс-метод и теория двойственности. Кроме того, будут рассмотрены такие эффективные вычислительные алгоритмы, как модифицированный симплекс-метод, метод декомпозиции, метод решения задач, содержащих ограничения на значения переменных, и методы параметрического программирования. В заключение будет представлен алгоритм Кармаркара, полностью отличный от симплекс-метода. Этот алгоритм наиболее эффективен при решении экстремально больших задач ЛП.

В этой главе используется аппарат матричной алгебры. Читатель, не знакомый с матричной алгеброй, может обратиться к Приложению А.

7.2. Векторы и базисы

В этом разделе показано, как получить базисное решение на основе матричного представления *стандартной* задачи ЛП.

7.2.1. Матричное представление стандартной задачи ЛП

Стандартная задача ЛП, определенная в разделе 3.2.2, предполагает неотрицательность всех переменных и представление всех ограничений в виде равенств с неотрицательной правой частью. С помощью матричной формы записи стандартную задачу ЛП можно представить следующим образом.

Максимизировать или минимизировать $z = \mathbf{CX}$

при ограничениях

$$(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b},$$
$$\mathbf{X} \geq \mathbf{0},$$

где $\mathbf{I}$ — единичная матрица размером $m \times m$, $\mathbf{X} = (x_1, x_2, ..., x_n)^{\mathrm{T}}$, $\mathbf{C} = (c_1, c_2, ..., c_n)$ и

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1,n-m} \\ a_{21} & a_{22} & \dots & a_{2,n-m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{m,n-m} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Представление ограничений с применением единичной матрицы **I** возможно, если надлежащим образом расположить переменные и использовать, при необходимости, искусственные переменные.

Пример 7.2–1

Представим в матричной форме следующую задачу ЛП.

$$\text{Максимизировать } z = 2x_1 + 3x_2$$

при ограничениях

$$\begin{aligned}
&x_1 + x_2 \ge 5,\\
&x_1 + 2x_2 = 7,\\
&5x_1 - 2x_2 \le 9,\\
&x_1, x_2 \ge 0.
\end{aligned}$$

Соответствующая стандартная задача ЛП в матричной форме будет записана так:

$$\text{Максимизировать } z = (2, 3, 0, -M, -M, 0)\begin{bmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5\\ x_6 \end{bmatrix}$$

при ограничениях

$$\begin{bmatrix} 1 & 1 & -1 & 1 & 0 & 0\\ 1 & 2 & 0 & 0 & 1 & 0\\ 5 & -2 & 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1\\ x_2\\ x_3\\ x_4\\ x_5\\ x_6 \end{bmatrix} = \begin{bmatrix} 5\\ 7\\ 9 \end{bmatrix},$$

$$x_j \ge 0,\ j = 1, 2, \ldots, 6.$$

Эта задача имеет $n = 6$ переменных и $m = 3$ ограничений в виде равенств. Переменные x_3 и x_6 — это дополнительные переменные, а переменные x_4 и x_5 — искусственные.

Соответствующие векторы и матрицы определяются следующим образом.

$$\mathbf{X} = (x_1, x_2, \ldots, x_6)^T,$$
$$\mathbf{C} = (2, 3, 0, -M, -M, 0),$$
$$\mathbf{b} = (5, 7, 9)^T,$$
$$\mathbf{A} = \begin{bmatrix} 1 & 1 & -1\\ 1 & 2 & 0\\ 5 & -2 & 0 \end{bmatrix},\quad \mathbf{I} = \begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1 \end{bmatrix}.$$

Упражнения 7.2,а

1. Пусть в задаче из примера 7.2–1 первое и второе ограничения являются неравенствами типа "$\leq$", а третье неравенство имеет тип "$\geq$". Запишите стандартную задачу в матричной форме и определите соответствующие матрицы **X**, **C**, **A**, **I** и **b**.
2. Запишите следующие задачи ЛП как стандартные задачи в матричной форме и определите соответствующие матрицы **A**, **C**, **b** и **X**.

   a) Максимизировать $z = 3x_1 + 2x_2 + 5x_3$
   при ограничениях

$$7x_1 + 3x_2 - x_3 \leq 15,$$
$$2x_1 - 2x_2 + 3x_3 \leq 20,$$
$$x_1 + x_2 + x_3 \leq 5,$$
$$x_1, x_1, x_3 \geq 0.$$

   b) Минимизировать $z = 2x_1 + 5x_2$
   при ограничениях

$$3x_1 + 2x_2 \leq 5,$$
$$4x_1 - x_2 \geq 2,$$
$$x_1, x_1 \geq 0.$$

   c) Максимизировать $z = 6x_1 + 2x_2 + 3x_3$
   при ограничениях

$$5x_1 + 2x_2 + 4x_3 = 20,$$
$$3x_1 - x_2 + 2x_3 \leq 15,$$
$$x_1, x_1, x_3 \geq 0.$$

7.2.2. Векторное представление базисов

В стандартной задаче ЛП система линейных уравнений (**A**, **I**)**X** = **b** с m уравнениями и n неизвестными может быть представлена в следующей векторной форме

$$\sum_{j=1}^{n} \mathbf{P}_j x_j = \mathbf{b},$$

где вектор $\mathbf{P}_j$ — j-й столбец матрицы (**A**, **I**). Подмножество из m векторов $\mathbf{P}_j$ формирует **базис B** тогда и только тогда, когда эти векторы **линейно независимы**. Для этого необходимо (и достаточно), чтобы определитель матрицы **B**, состоящей из данных m векторов, был отличен от нуля, т.е. $\det(\mathbf{B}) \neq 0$.[1] В этом случае матрица **B** называется **невырожденной**.

[1] Здесь и далее автор обозначает через **B** как базис, т.е. совокупность базисных векторов, так и матрицу, составленную из этих векторов. Это не приводит к недоразумениям, так как из контекста всегда видно, о чем идет речь. — *Прим. ред.*

Пример 7.2–2

Следующая система, представленная в векторной форме, состоит из двух линейных уравнений с тремя неизвестными ($m = 2$ и $n = 3$). Найдем все ее базисы.

$$\begin{bmatrix}1\\2\end{bmatrix}x_1+\begin{bmatrix}3\\-2\end{bmatrix}x_2+\begin{bmatrix}-1\\-2\end{bmatrix}x_3=\begin{bmatrix}4\\2\end{bmatrix}$$

Здесь векторы $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$, $\mathbf{b}$ являются двухмерными; в общем виде они представимы как $(a_1, a_2)^T$. На рис. 7.1 на плоскости (a_1, a_2) показаны векторы данной системы. Например, вектор $\mathbf{b} = (4, 2)^T$, где $a_1 = 4$, $a_2 = 2$.

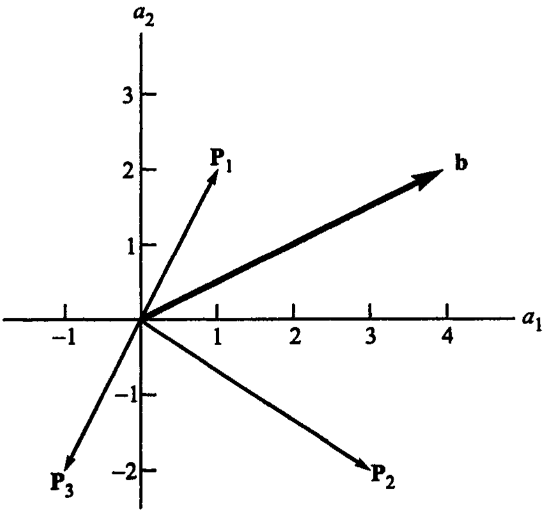


Рис. 7.1

Так как у нас только два уравнения ($m = 2$), базис должен состоять из двух векторов, выбранных среди $\mathbf{P}_1$, $\mathbf{P}_2$ и $\mathbf{P}_3$. Из рис. 7.1 видно, что пары векторов ($\mathbf{P}_1$, $\mathbf{P}_2$) и ($\mathbf{P}_2$ $\mathbf{P}_3$) могут быть базисами, поскольку векторы в этих парах независимы. Но пара векторов ($\mathbf{P}_1$, $\mathbf{P}_3$) не может составить базис, так как эти векторы линейно зависимы.

Алгебраический подход к определению базиса требует, чтобы определитель матрицы, составленной из векторов, претендующих на роль базисных, был отличен от нуля. Вычисления определителей показывают, что пары векторов ($\mathbf{P}_1$, $\mathbf{P}_2$) и ($\mathbf{P}_2$ $\mathbf{P}_3$) будут базисами, а векторы ($\mathbf{P}_1$, $\mathbf{P}_3$) не могут составить базис.

$$\det(\mathbf{P}_1,\mathbf{P}_2)=\det\begin{bmatrix}1&3\\2&-2\end{bmatrix}=(1\times(-2))-(2\times3)=-8\neq0.$$

$$\det(\mathbf{P}_2,\mathbf{P}_3)=\det\begin{bmatrix}3&-1\\-2&-2\end{bmatrix}=(3\times(-2))-(-2\times(-1))=-8\neq0.$$

$$\det(\mathbf{P}_1,\mathbf{P}_3)=\det\begin{bmatrix}1&-1\\2&-2\end{bmatrix}=(1\times(-2))-(2\times(-1))=0.$$

Упражнения 7.2,b

1. Покажите графически и определите алгебраически, какие из следующих матриц составлены из базисных векторов.

$$\mathbf{B}_1 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix},\ \mathbf{B}_2 = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},\ \mathbf{B}_3 = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix},\ \mathbf{B}_4 = \begin{bmatrix} 1 & 5 \\ 2 & 10 \end{bmatrix}$$

2. Дана следующая система уравнений.

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} x_1 + \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} x_2 + \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} x_3 + \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} x_4 = \begin{bmatrix} 3 \\ 4 \\ 2 \end{bmatrix}$$

Определите, какой из следующих наборов векторов образует базис.

a) $(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)$
b) $(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_4)$
c) $(\mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$

7.2.3. Базисные решения

Для системы из m уравнений $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$ с n неизвестными ($m < n$) обозначим через $\mathbf{X}_B$ вектор из m элементов, являющихся подмножеством n элементов вектора $\mathbf{X}$. Определим матрицу $\mathbf{B}$ размером $m \times m$, состоящую из столбцов матрицы $(\mathbf{A}, \mathbf{I})$, соответствующих элементам вектора $\mathbf{X}_B$. Если присвоить оставшимся $n - m$ элементам вектора $\mathbf{X}$ нулевые значения, то система $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$ будет сведена к $\mathbf{B}\mathbf{X}_B = \mathbf{b}$.

Если матрица $\mathbf{B}$ состоит из базисных векторов, тогда имеем *единственное* решение последней системы:

$$\mathbf{X}_B = \mathbf{B}^{-1}\mathbf{b},$$

где $\mathbf{B}^{-1}$ — матрица, обратная к матрице $\mathbf{B}$. В этом случае $\mathbf{X}_B$ является **базисным решением** системы $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$. Если выполняется неравенство $\mathbf{B}^{-1}\mathbf{b} \geq \mathbf{0}$, тогда $\mathbf{X}_B$ будет допустимым решением.

В заключение отметим, что количество базисных решений у системы из m уравнениями с n неизвестными не превышает величины

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

Пример 7.2–3

В следующей таблице показаны все базисные решения системы уравнений из примера 7.2–2 и указано, какие из них являются допустимыми. Матрица, обратная к матрице **B**, вычислена методами, приведенными в Приложении А (раздел А.2.7).

B	$BX_B = b$	Решение	Статус
(P_1, P_2)	$\begin{bmatrix}1 & 3\\2 & -2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}4\\2\end{bmatrix}$	$\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}\frac{1}{4} & \frac{3}{8}\\\frac{1}{4} & -\frac{1}{8}\end{bmatrix}\begin{bmatrix}4\\2\end{bmatrix}=\begin{bmatrix}\frac{7}{4}\\\frac{3}{4}\end{bmatrix}$	Допустимое
(P_1, P_3)	Не является базисом	—	—
(P_2, P_3)	$\begin{bmatrix}3 & -1\\-2 & -2\end{bmatrix}\begin{bmatrix}x_2\\x_3\end{bmatrix}=\begin{bmatrix}4\\2\end{bmatrix}$	$\begin{bmatrix}x_2\\x_3\end{bmatrix}=\begin{bmatrix}\frac{1}{4} & -\frac{1}{8}\\-\frac{1}{4} & -\frac{3}{8}\end{bmatrix}\begin{bmatrix}4\\2\end{bmatrix}=\begin{bmatrix}\frac{3}{4}\\-\frac{7}{4}\end{bmatrix}$	Недопустимое

Упражнения 7.2,c

1. Представленные ниже системы уравнений a) и b) имеют единственные базисные решения, система c) — бесконечно много решений, а d) не имеет решения. Докажите это, представив решения графически на плоскости. На основании данных примеров выведите общие закономерности, определяющие у системы уравнений единственное решение, бесконечно много решений или отсутствие решений.

   a) $x_1 + 3x_2 = 2,$ $3x_1 + x_2 = 3.$ b) $2x_1 + 3x_2 = 1,$ $2x_1 - x_2 = 2.$

   c) $2x_1 + 6x_2 = 4,$ $x_1 + 3x_2 = 2.$ d) $2x_1 - 4x_2 = 2,$ $-x_1 + 2x_2 = 1.$

2. Определите графически, какая из следующих систем уравнений имеет единственное решение, бесконечно много решений или не имеет решений.

   a) $\begin{bmatrix}5 & 4\\1 & -3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix}.$ b) $\begin{bmatrix}2 & -2\\1 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1\\3\end{bmatrix}.$

   c) $\begin{bmatrix}2 & 4\\1 & 3\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}-2\\-1\end{bmatrix}.$ d) $\begin{bmatrix}2 & 4\\1 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}6\\3\end{bmatrix}.$

   e) $\begin{bmatrix}-2 & 4\\1 & -2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}.$ f) $\begin{bmatrix}1 & -2\\0 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix}.$

3. Истинны или ложны следующие утверждения?

   a) Если матрица **B** не вырождена, то система **BX** = **b** имеет единственное решение.

   b) Система **BX** = **b** не имеет решения, если матрица **B** вырождена и вектор **b** независим относительно векторов, составляющих матрицу **B**.

   c) Система **BX** = **b** имеет бесконечно много решений, если матрица **B** вырождена и вектор **b** зависим относительно векторов, составляющих матрицу **B**.

7.3. Обоснование симплекс-метода[1]

В разделе 2.3 мы показали, что (конечное) оптимальное решение двухмерной задачи ЛП соответствует *крайней (угловой) точке* пространства допустимых решений. В главе 3 мы перешли от графического способа решения задач ЛП к алгебраическому симплекс-ному алгоритму путем установления соответствия между крайними точками пространства решений и базисными решениями систем уравнений, определяющих пространство допустимых решений. В этом разделе мы, наконец-то, докажем данное соответствие. Поскольку ключевую роль при доказательстве играют выпуклые множества, сначала определим эти множества математически.

7.3.1. Выпуклые множества

Множество C n-мерного пространства называется выпуклым, если отрезок прямой, соединяющий две *различные* точки этого множества, полностью принадлежит данному множеству. Математически это определение можно записать следующим образом. Если $\mathbf{X}'$ и $\mathbf{X}''$ — две различные точки множества C, тогда **выпуклая комбинация** этих точек, определяемая как

$$\mathbf{X} = \lambda\mathbf{X}' + (1-\lambda)\mathbf{X}'', \quad 0 \le \lambda \le 1,$$

должна принадлежать множеству C. На рис. 7.2 показано, что множества *а*) и *б*) выпуклые, а множество *в*) невыпуклое.

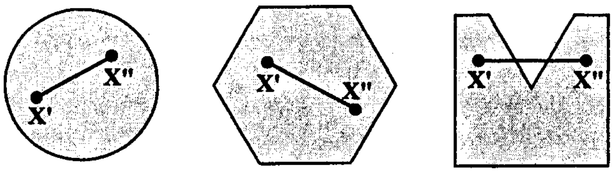


Рис. 7.2

Крайние точки выпуклого множества — это такие точки этого множества, которые нельзя представить в виде выпуклой комбинации каких-либо двух *различных* точек данного множества. На рис. 7.2 множество *а*) имеет бесконечное множество крайних точек (а именно, точки окружности, ограничивающей данный круг), множество *б*), которое типично для пространств решений задач ЛП, имеет конечное число крайних (угловых) точек (таких точек в данном случае 6).

Упражнения 7.3,а

1. Покажите, что множество $Q = \{x_1, x_2 \mid x_1 + x_2 \le 1, x_1 \ge 0, x_2 \ge 0\}$ выпуклое. Существенно ли в данном случае условие неотрицательности переменных?
2. Покажите, что множество $Q = \{x_1, x_2 \mid x_1 \ge 1 \text{ или } x_2 \ge 2\}$ не является выпуклым.

---

[1] Этот раздел можно пропустить без потери целостности изложения материала.

3. Найдите графически крайние точки выпуклого множества $Q = \{x_1, x_2 \mid x_1 + x_2 \le 2, x_1 \ge 0, x_2 \ge 0\}$. Затем покажите, что все пространство допустимых решений, совпадающее с множеством Q, можно определить как выпуклую комбинацию этих крайних точек. Таким образом будет доказано уже известное нам утверждение, что любое ограниченное пространство решений в общем случае определяется только крайними точками.

4. Представьте внутреннюю точку (3, 1) пространства решений, показанного на рис. 7.3, как выпуклую комбинацию крайних точек A, B, C и D, где каждая крайняя точка должна иметь строго положительный весовой коэффициент.

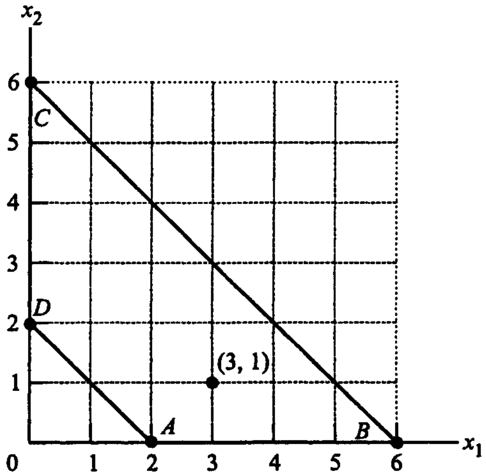


Рис. 7.3

7.3.2. Сходимость симплексного алгоритма к оптимальному решению

Доказательство того, что симплексный алгоритм приводит к оптимальному решению задачи ЛП, основано на трех взаимосвязанных теоремах. В теореме 7.3-1 доказывается, что пространство допустимых решений задачи ЛП является выпуклым множеством с конечным числом крайних точек. Теорема 7.3-2 доказывает, что оптимальное решение задачи ЛП, получаемое симплекс-методом, всегда ассоциируется с допустимой крайней точкой пространства решений. Наконец, в теореме 7.3-3 устанавливается связь между (геометрическими) крайними точками и (алгебраическими) базисными решениями задачи ЛП путем доказательства того, что все крайние точки пространства решений идентифицируются базисными решениями системы уравнений, соответствующей ограничениям задачи ЛП, записанной в стандартной форме. Отсюда следует, что оптимальное решение (стандартной) задачи ЛП является базисным (допустимым) решением системы уравнений, выражающей ограничения задачи ЛП.

Обозначим через Q множество всех допустимых решений задачи ЛП, т.е.

$$Q = \{\mathbf{X} \mid (\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}, \mathbf{X} \ge \mathbf{0}\}.$$

Предполагается, что задача ЛП имеет m ограничений в виде равенств и n неизвестных.

Теорема 7.3–1. *Множество Q всех допустимых решений задачи ЛП является выпуклым.*

Доказательство. Обозначим через $\mathbf{X}^*$ выпуклую комбинацию двух различных точек $\mathbf{X}'$ и $\mathbf{X}''$, принадлежащих множеству Q, т.е.

$$\mathbf{X}^* = \lambda\,\mathbf{X}' + (1-\lambda)\,\mathbf{X}'', \quad 0 \le \lambda \le 1.$$

Для доказательства выпуклости множества Q достаточно показать, что $\mathbf{X}^*$ принадлежит этому множеству. Заметим, что по определению $\mathbf{X}^* \ge \mathbf{0}$. Осталось показать, что $(\mathbf{A}, \mathbf{I})\mathbf{X}^* = \mathbf{b}$:

$$\begin{aligned}(\mathbf{A}, \mathbf{I})\mathbf{X}^* &= (\mathbf{A}, \mathbf{I})[\lambda\,\mathbf{X}' + (1-\lambda)\,\mathbf{X}''] = \\ &= \lambda(\mathbf{A}, \mathbf{I})\,\mathbf{X}' + (1-\lambda)(\mathbf{A}, \mathbf{I})\,\mathbf{X}'' = \\ &= \lambda\mathbf{b} + (1-\lambda)\mathbf{b} = \mathbf{b}.\end{aligned}$$

Теорема доказана.

Теорема 7.3–2. *Оптимальное решение задачи линейного программирования*

Максимизировать $z = \mathbf{CX}$ при ограничениях $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$, $\mathbf{X} \ge \mathbf{0}$

всегда достигается в крайней точке пространства Q, если оно (решение) конечное.

Доказательство. Мы можем считать, что множество Q ограничено, так как при необходимости можно ввести дополнительные ограничения $x_j \le M$ (где M — достаточно большое число) для всех неограниченных переменных x_j. В этом случае любая допустимая точка, принадлежащая множеству Q, может быть представлена в виде выпуклой комбинации крайних точек $\mathbf{X}(k)$ ($k = 1, 2, \ldots, K$) пространства решений. Определим $\mathbf{X}^*$ как крайнюю точку, в которой целевая функция принимает наибольшее значение среди всех значений, принимаемых целевой функцией на множестве крайних точек. Таким образом,

$$z^* = \mathbf{CX}^* = \max_i\{\mathbf{CX}^{(i)}\}.$$

Теперь покажем, что на точке $\mathbf{X}'$ из пространства решений Q, не являющейся крайней точкой, целевая функция не может принять большего значения, чем z^*. Поскольку точка $\mathbf{X}'$ не является крайней, она может быть представлена в виде выпуклой комбинации крайних точек множества Q, т.е.

$$\mathbf{X}' = \sum_{k=1}^{K}\lambda_k\mathbf{X}^{(k)}, \quad \lambda_k \ge 0, \quad \sum_{k=1}^{K}\lambda_k = 1.$$

Тогда

$$z' = \mathbf{CX}' = \mathbf{C}\left(\sum_{k=1}^{K}\lambda_k\mathbf{X}^{(k)}\right) = \sum_{k=1}^{K}\lambda_k\left(\mathbf{CX}^{(k)}\right) \le \mathbf{CX}^* = z^*,$$

что и завершает доказательство теоремы.

В теореме 7.3-2 доказано, что оптимальное решение задачи ЛП может быть достигнуто только в крайних точках пространства решений. Существенным в доказательстве теоремы является замена бесконечных точек пространства решений конечным числом крайних точек, которые могут быть “кандидатами” на оптимальное решение.

Теореме 7.3–3. *Точка $\mathbf{X}$ будет крайней точкой пространства решений Q тогда и только тогда, когда $\mathbf{X}$ будет базисным решением системы уравнений $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$ и при этом $\mathbf{X} \ge \mathbf{0}$.*

Доказательство ведется методом от противного. (Доказательство этой теоремы приведено в [1].)

Упражнения 7.3,b

1. Дана следующая система уравнений в векторной форме

$$\begin{bmatrix}2\\1\end{bmatrix}x_1+\begin{bmatrix}1\\2\end{bmatrix}x_2+\begin{bmatrix}1\\1\end{bmatrix}x_3+\begin{bmatrix}2\\-1\end{bmatrix}x_4+\begin{bmatrix}4\\2\end{bmatrix}x_5=\begin{bmatrix}2\\2\end{bmatrix}.$$

   Здесь столбцы соответствуют векторам $\mathbf{P}_1, ..., \mathbf{P}_5$ и вектору $\mathbf{b}$.

   a) Отобразите эти векторы на плоскости.

   b) Сколько *различных* крайних точек соответствует базисам $(\mathbf{P}_1, \mathbf{P}_3)$, $(\mathbf{P}_2, \mathbf{P}_3)$, $(\mathbf{P}_3, \mathbf{P}_4)$ и $(\mathbf{P}_3, \mathbf{P}_5)$? Поясните свой ответ.

   c) Предположим, что все переменные должны быть неотрицательными. Может ли в этом случае пара векторов $(\mathbf{P}_1, \mathbf{P}_4)$ составить *допустимый* базис?

2. Даны следующие системы линейных уравнений, в которых все переменные неотрицательные.

   a) $2x_1 + 3x_2 + x_3 = 6,$
      $x_1 + 2x_2 + x_4 = 4.$

   b) $3x_1 + 6x_2 + 5x_3 + x_4 = 12,$
      $2x_1 + 2x_2 + x_3 + 2x_5 = 8.$

   Для каждой системы путем оценки ее базисных допустимых решений найдите все допустимые крайние точки. Каково соотношение между количеством крайних точек и числом базисных решений?

3. Пусть дана задача ЛП, в которой не накладываются ограничения на знак переменных. Докажите, что после выполнения замены $x_k = x_k^+ - x_k^-$ для всех переменных x_k, где x_k^+ и x_k^- неотрицательные, на любой итерации симплекс-метода невозможна ситуация, когда переменные x_k^+ и x_k^- одновременно принимают положительные значения.

4. Докажите, что конечное оптимальное решение задачи ЛП

   Максимизировать $z = \mathbf{CX}$ при ограничениях $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$, $\mathbf{X} \geq \mathbf{0}$

   всегда достигается в крайней точке пространства решений.

5. В условиях предыдущего упражнения докажите, что если оптимальное решение достигается в нескольких крайних точках пространства решений, то значение целевой функции остается постоянным в любой выпуклой комбинации этих крайних точек.

7.4. Матричное представление симплекс-таблиц

В предыдущем разделе мы доказали, что конечное оптимальное решение задачи ЛП достигается в крайних точках пространства решений. Мы также доказали, что все крайние точки можно определить алгебраически как базисные решения системы линейных уравнений $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$, $\mathbf{X} \geq \mathbf{0}$. Таким образом, чтобы найти оптимальное решение задачи ЛП, достаточно рассматривать только базисные решения указанной системы уравнений.

При выполнении симплекс-метода мы начинали с допустимого базисного решения **B**, затем переходили к следующему допустимому базисному решению, которое улучшает (по крайней мере, не ухудшает) значение целевой функции, и так до тех пор, пока не будет достигнуто оптимальное решение. Таким образом, допустимое базисное решение **B** — это тот принципиальный элемент в симплекс-методе, вокруг которого выполняются все вычисления в симплекс-таблице. С этой точки зрения очевидна необходимость представления симплекс-таблицы в матричной форме.

В стандартной задаче ЛП

$$\text{Максимизировать } z = \mathbf{CX} \text{ при ограничениях } (\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b},\ \mathbf{X} \geq \mathbf{0},$$

разобьем вектор **X** на два — $\mathbf{X}_\mathrm{I}$ и $\mathbf{X}_\mathrm{II}$, таких, что вектор $\mathbf{X}_\mathrm{II}$ соответствует начальному базису **B**, т.е. является *начальным* допустимым базисным решением. Вектор **C** также разделим на два вектора $\mathbf{C}_\mathrm{I}$ и $\mathbf{C}_\mathrm{II}$ в соответствии с векторами $\mathbf{X}_\mathrm{I}$ и $\mathbf{X}_\mathrm{II}$. Тогда стандартную задачу ЛП можно записать следующим образом.

$$\begin{bmatrix} 1 & -\mathbf{C}_\mathrm{I} & -\mathbf{C}_\mathrm{II} \\ \mathbf{0} & \mathbf{A} & \mathbf{I} \end{bmatrix} \begin{bmatrix} z \\ \mathbf{X}_\mathrm{I} \\ \mathbf{X}_\mathrm{II} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix}$$

Для любой симплексной итерации будем обозначать через $\mathbf{X}_B$ базисный вектор переменных, а через $\mathbf{C}_B$ — вектор коэффициентов целевой функции, соответствующих этому базису. Поскольку все небазисные переменные равны нулю, стандартная задача ЛП будет сведена к задаче с целевой функцией $z = \mathbf{C}_B\mathbf{X}_B$ и ограничениями $\mathbf{BX}_B = \mathbf{b}$, где текущее решение удовлетворяет следующему уравнению.

$$\begin{bmatrix} z \\ \mathbf{X}_B \end{bmatrix} = \begin{bmatrix} 1 & -\mathbf{C}_B \\ \mathbf{0} & \mathbf{B} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{C}_B\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_B\mathbf{B}^{-1}\mathbf{b} \\ \mathbf{B}^{-1}\mathbf{b} \end{bmatrix}.$$

Здесь при обращении матрицы использовались формулы из раздела А.2.7.

Симплекс-таблица получается из исходной задачи ЛП путем вычислений по следующей формуле

$$\begin{bmatrix} 1 & \mathbf{C}_B\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\mathbf{C}_\mathrm{I} & -\mathbf{C}_\mathrm{II} \\ \mathbf{0} & \mathbf{A} & \mathbf{I} \end{bmatrix} \begin{bmatrix} z \\ \mathbf{X}_\mathrm{I} \\ \mathbf{X}_\mathrm{II} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{C}_B\mathbf{B}^{-1} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix}.$$

Выполнив вычисления по этой формуле, получаем симплекс-таблицу

Базис	$\mathbf{X}_\mathrm{I}$	$\mathbf{X}_\mathrm{II}$	Решение
z	$\mathbf{C}_B\mathbf{B}^{-1}\mathbf{A} - \mathbf{C}_\mathrm{I}$	$\mathbf{C}_B\mathbf{B}^{-1} - \mathbf{C}_\mathrm{II}$	$\mathbf{C}_B\mathbf{B}^{-1}\mathbf{b}$
$\mathbf{X}_B$	$\mathbf{B}^{-1}\mathbf{A}$	$\mathbf{B}^{-1}$	$\mathbf{B}^{-1}\mathbf{b}$

В этой таблице предварительных вычислений требует только обратная матрица $\mathbf{B}^{-1}$, так как другие элементы таблицы $\mathbf{C}_B$, $\mathbf{C}_\mathrm{I}$, $\mathbf{C}_\mathrm{II}$, **A** и **b** получаются из исходных данных задачи.

Представленная симплекс-таблица имеет большое значение, так как является основой всех вычислительных алгоритмов линейного программирования. В частности, на ней основаны модифицированный симплекс-метод, метод решения задач с ограниченными переменными и метод декомпозиции, которые будут описаны в следующем разделе.

В симплекс-методе решение переходит от одного базиса **B** к следующему $\mathbf{B}_{\text{след}}$ путем замены в **B** *базисного* вектора (*исключаемого*) на *небазисный* (*вводимый*). Определение вводимых и исключаемых векторов основано на следующих условиях оптимальности и допустимости.

Условие оптимальности симплекс-метода. Из матричного представления симплекс-таблицы следует, что коэффициент при переменной x_j в z-строке таблицы равен

$$z_j - c_j = \mathbf{C}_B\mathbf{B}^{-1}\mathbf{P}_j - c_j,$$

где $\mathbf{P}_j$ — j-й столбец матрицы (**A**, **I**), c_j — j-й элемент вектора **C**. Отметим, что разность $z_j - c_j$ всегда равна нулю для *базисных* переменных x_j (см. упр. 7.4,а(8)). Если обозначить через NB множество индексов небазисных переменных, тогда можно записать следующее уравнение для целевой функции.

$$z + \sum_{j\in NB}(z_j - c_j)x_j = \mathbf{C}_B\ \mathbf{B}^{-1}\mathbf{b}.$$

Из этого уравнения следует, что увеличение значения небазисной переменной x_j приводит к возрастанию (убыванию) значения целевой функции z выше текущего значения $\mathbf{C}_B\mathbf{B}^{-1}\mathbf{b}$ только в том случае, если разность $z_j - c_j$ строго отрицательна (положительна). В противном случае переменная x_j не может улучшить текущее решение и должна остаться небазисной с нулевым значением. (В упр. 7.4,а(9) дана более развернутая интерпретация разностей $z_j - c_j$.)

Условие допустимости симплекс-метода. Определение исключаемого из базиса вектора основано на проверке ограничения, представленного в виде равенства, соответствующего i-й базисной переменной. Это равенство имеет следующий вид.

$$(\mathbf{X}_B)_i + \sum_{j\in NB}(\mathbf{B}^{-1}\mathbf{P}_j)_i x_j = (\mathbf{B}^{-1}\mathbf{b})_i$$

Запись $(\mathbf{V})_i$ означает i-й элемент вектор-столбца **V**.

Обозначим через $\mathbf{P}_k$ вводимый вектор, определенный из условия оптимальности, а через x_k — вводимую в базис переменную, принимающую положительное значение. Поскольку все остальные небазисные переменные сохраняют нулевые значения, равенство ограничения, соответствующее базисной переменной $(\mathbf{X}_B)_i$, можно записать следующим образом.

$$(\mathbf{X}_B)_i = (\mathbf{B}^{-1}\mathbf{b})_i - (\mathbf{B}^{-1}\mathbf{P}_k)_i\, x_k$$

Это уравнение показывает, что при $(\mathbf{B}^{-1}\mathbf{P}_k)_i > 0$ возрастание переменной x_k не приведет к отрицательному значению базисной переменной $(\mathbf{X}_B)_i$ только в том случае, если будет выполняться неравенство

$$(\mathbf{B}^{-1}\mathbf{b})_i - (\mathbf{B}^{-1}\mathbf{P}_k)_i\, x_k \ge 0 \text{ для всех } i.$$

Таким образом, максимальное значение вводимой переменной x_k можно вычислить по следующей формуле.

$$x_k = \min_i \left\{ \frac{(\mathbf{B}^{-1}\mathbf{b})_i}{(\mathbf{B}^{-1}\mathbf{P}_k)_i} \middle| (\mathbf{B}^{-1}\mathbf{P}_k)_i > 0 \right\}$$

Базисная переменная, на которой достигается этот минимум, становится исключаемой.

Пример 7.4–1

Рассмотрим следующую задачу ЛП.

Максимизировать $z = x_1 + 4x_2 + 7x_3 + 5x_4$

при ограничениях

$$2x_1 + x_2 + 2x_3 + 4x_4 = 10,$$
$$3x_1 - x_2 - 2x_3 + 6x_4 = 5,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

Пусть $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_2)$ является допустимым базисом.

a) Покажем, что решение $\mathbf{B}$ не является оптимальным.

b) Найдем вводимый в базис и исключаемый из него векторы и $\mathbf{B}_{\text{след}}$.

На основании $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_2)$ имеем $\mathbf{X}_B = (x_1, x_2)^T$ и $\mathbf{C}_B = (1, 4)$. Вычисляем обратную матрицу $\mathbf{B}^{-1}$:

$$\mathbf{B}^{-1} = \begin{bmatrix} 2 & 1 \\ 3 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{3}{5} & -\frac{2}{5} \end{bmatrix}.$$

Находим текущее базисное решение

$$\mathbf{X}_B = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{B}^{-1}\mathbf{b} = \begin{bmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{3}{5} & -\frac{2}{5} \end{bmatrix} \begin{bmatrix} 10 \\ 5 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix},$$

откуда получаем $z = \mathbf{C}_B\mathbf{X}_B = 1 \times 3 + 4 \times 4 = 19$.

Для проверки оптимальности базиса $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_2)$ вычислим разности $z_j - c_j$ для текущих небазисных переменных x_3 и x_4:

$$(z_3 - c_3, z_4 - c_4) = \mathbf{C}_B\, \mathbf{B}^{-1}[\mathbf{P}_3, \mathbf{P}_4] - (c_3, c_4) =$$

$$= (1, 4)\begin{bmatrix} \frac{1}{5} & \frac{1}{5} \\ \frac{3}{5} & -\frac{2}{5} \end{bmatrix} \begin{bmatrix} 2 & 4 \\ -2 & 6 \end{bmatrix} - (7, 5) = (1, -3).$$

Поскольку в данной задаче следует максимизировать целевую функцию, разность $z_4 - c_4 = -3$ показывает, что базис $\mathbf{X}_B = (x_1, x_2)^T$ неоптимален и значение целевой функции можно увеличить, если ввести в базис переменную x_4.

Для того чтобы ввести в базис переменную x_4, из него следует исключить или переменную x_1, или x_2 (или, что то же самое, исключить или вектор $\mathbf{P}_1$, или $\mathbf{P}_2$ из базиса $\mathbf{B}$). Для определения исключаемой переменной вычисляем

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{B}^{-1}\mathbf{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \text{ и } \mathbf{B}^{-1}\mathbf{P}_4 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Поскольку вектор $\mathbf{B}^{-1}\mathbf{P}_4$ имеет только один строго положительный элемент (= 2), значение вводимой переменной x_4 равно

$$x_4 = \min_{i=1,2}\left\{\frac{3}{2}, -\right\} = \frac{3}{2}.$$

Таким образом, в базисе $\mathbf{B}$ вектор $\mathbf{P}_1$ будет заменен на $\mathbf{P}_4$, что приводит к новому базису

$$\mathbf{B}_{\text{след}} = (\mathbf{P}_4, \mathbf{P}_2) = \begin{bmatrix} 4 & 1 \\ 6 & -1 \end{bmatrix}.$$

Соответствующее новое значение целевой функции равно $z = 19 - (z_4 - c_4)x_4 = 19 - (-3) \times (3/2) = 23.5$.

Упражнения 7.4,а

1. Дана следующая задача ЛП

$$\text{Максимизировать } z = 5x_1 + 12x_2 + 4x_3$$

при ограничениях

$$x_1 + 2x_2 + x_3 + x_4 = 10,$$
$$2x_1 - 2x_2 - x_3 = 2,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

Определите, какие из следующих наборов векторов образуют базис: $(\mathbf{P}_1, \mathbf{P}_2)$, $(\mathbf{P}_2, \mathbf{P}_3)$, $(\mathbf{P}_3, \mathbf{P}_4)$.

2. Продолжите вычисления в задаче примера 7.4–1 и найдите оптимальное решение.
3. Можно ли в задаче из примера 7.4–1 выполнение симплекс-метода начать с базиса $\mathbf{B} = (\mathbf{P}_2, \mathbf{P}_3)$? Обоснуйте свой ответ.
4. Дана следующая задача ЛП

$$\text{Максимизировать } z = 4x_1 + 14x_2 - 2x_3$$

при ограничениях

$$2x_1 + 7x_2 + x_3 = 21,$$
$$7x_1 + 2x_2 + x_4 = 21,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

С помощью матричных вычислений проверьте допустимость и оптимальность следующих наборов базисных переменных.

a) (x_2, x_4); b) (x_2, x_3); c) (x_2, x_1); d) (x_1, x_4).

5. Дана следующая задача ЛП

$$\text{Минимизировать } z = 2x_1 + x_2$$

при ограничениях

$$3x_1 + x_2 - x_3 = 3,$$
$$4x_1 + 3x_2 - x_4 = 6,$$
$$x_1 + 2x_2 + x_5 = 3,$$
$$x_1, x_2, x_3, x_4, x_5 \geq 0.$$

Вычислите симплекс-таблицу, соответствующую $\mathbf{X}_B = (x_1, x_2, x_5)^T$, и определите, будет ли это решение допустимым и оптимальным.

6. Дана следующая симплекс-таблица с оптимальным решением задачи ЛП.

Базис	x_1	x_2	x_3	x_4	x_5	Решение
z	0	0	0	3	2	?
x_3	0	0	1	1	−1	2
x_2	0	1	0	1	0	6
x_1	1	0	0	−1	1	2

Здесь переменные x_3, x_4 и x_5 являются дополнительными (остаточными). С помощью матричных вычислений реконструируйте исходную задачу ЛП, затем вычислите оптимальное значение целевой функции двумя различными способами.

7. Дана следующая задача ЛП

$$\text{Максимизировать } z = c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4$$

при ограничениях

$$\mathbf{P}_1x_1 + \mathbf{P}_2x_2 + \mathbf{P}_3x_3 + \mathbf{P}_4x_4 = \mathbf{b},$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

Векторы $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{P}_3$ и $\mathbf{P}_4$ показаны на рис. 7.4. Предположим, что текущим базисом является $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_2)$.

a) Если вектор $\mathbf{P}_3$ ввести в базис, какой вектор необходимо из него исключить, чтобы полученное базисное решение было допустимым?

b) Может ли вектор $\mathbf{P}_4$ быть частью допустимого базиса?

8. Докажите, что для всех *базисных* переменных соответствующие разности $z_j - c_j$ равны нулю.

9. Докажите, что если в задаче максимизации (минимизации) для всех небазисных переменных x_j выполняется неравенство $z_j - c_j > 0$ (< 0), то задача имеет

единственное оптимальное решение. Если же для некоторых небазисных переменных x_j выполняется равенство $z_j - c_j = 0$, задача имеет не единственное оптимальное решение.

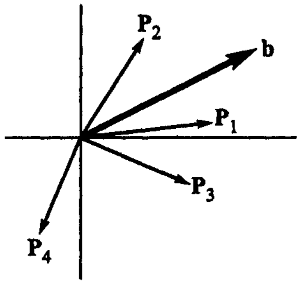


Рис. 7.4

10. Пусть начальное базисное решение задачи ЛП сформировано только из дополнительных остаточных переменных. С помощью матричного представления симплекс-таблицы покажите, что применение процедуры, описанной в разделе 3.3, где уравнение для целевой функции записано как

$$z - \sum_{j=1}^{n} c_j x_j = 0,$$

автоматически приведет к вычислению разностей $z_j - c_j$ для всех переменных в начальной симплекс-таблице.

11. Пусть начальное базисное решение задачи ЛП сформировано только из искусственных переменных. Используя матричное представление симплекс-таблицы, покажите, что применение процедуры исключения искусственных переменных из целевой функции (для чего использовались ограничения-равенства), описанной в разделе 3.4.1, приведет к вычислению разностей $z_j - c_j$ для всех переменных в начальной симплекс-таблице.

12. Пусть дана задача ЛП, в которой не накладываются ограничения на знак переменной x_k. Докажите, что после замены $x_k = x_k^+ - x_k^-$, где x_k^+ и x_k^- неотрицательны, невозможна ситуация, когда переменные x_k^+ и x_k^- заменяют одна другую в альтернативных оптимальных решениях.

13. Дана задача ЛП, записанная в стандартной форме и имеющая m ограничений (в виде равенств) и n неизвестных. Определите число *смежных* крайних точек, которые можно достичь из невырожденной крайней точки пространства решений. Почему здесь существенно условие невырожденности крайней точки?

14. Применяя условие допустимости симплекс-метода, предположим, что x_r — базисная переменная, имеющая нулевое значение, а x_j — вводимая. Объясните, почему для того, чтобы переменную x_r исключить из базиса, необходимо выполнение неравенства $(\mathbf{B}^{-1}\mathbf{P}_j)_r > 0$. Какие возникнут проблемы при выполнении неравенства $(\mathbf{B}^{-1}\mathbf{P}_j)_r \le 0$? (*Подсказка.* При ответе учитывайте, что переменная x_r должна оставаться неотрицательной.)

15. Что может указать на появление (впервые) вырожденного решения при реализации условия допустимости симплекс-метода? Какое условие приведет к повторению вырожденности решения на следующей итерации? Какое условие необходимо для того, чтобы вырожденность исчезла на следующей итерации? Обоснуйте ответы математически.

16. Каковы соотношения между крайними точками пространства решений и базисными решениями при вырожденности и невырожденности решений? Какое максимальное число итераций симплекс-метода может быть выполнено в одной и той же крайней точке?

17. Дана следующая задача ЛП: максимизировать $z = \mathbf{CX}$ при ограничениях $\mathbf{AX} \le \mathbf{b}$, $\mathbf{X} \ge \mathbf{0}$, где $\mathbf{b} \ge \mathbf{0}$. Предположим, что вводимый в базис вектор $\mathbf{P}_j$ таков, что, по крайней мере, один элемент вектора $\mathbf{B}^{-1}\mathbf{P}_j$ положителен.

    a) Пусть вектор $\mathbf{P}_j$ заменен на $\beta\mathbf{P}_j$, где β — положительное число, при этом переменная x_j остается переменной, вводимой в базис. Найдите соотношения между значениями переменной x_j, соответствующими векторам $\mathbf{P}_j$ и $\beta\mathbf{P}_j$.
    b) Ответьте на вопрос предыдущего пункта, если дополнительно вектор $\mathbf{b}$ заменен на вектор $\alpha\mathbf{b}$, где α — положительное число.

18. Дана следующая задача ЛП: максимизировать $z = \mathbf{CX}$ при ограничениях $\mathbf{AX} \le \mathbf{b}$, $\mathbf{X} \ge \mathbf{0}$, где $\mathbf{b} \ge \mathbf{0}$. Предположим, что после получения оптимального решения возникла идея сделать небазисную переменную x_j базисной (т.е. приносящей доход, если вспомнить экономическую интерпретацию задач ЛП) путем уменьшения ресурсов, расходуемых на единицу x_j, до величины $1/\beta$ от исходного значения, где β — число, большее единицы. Поскольку сокращено потребление ресурсов, ожидается, что доход на единицу x_j также уменьшится до величины $1/\beta$ от исходного значения. Может ли это изменение привести к рентабельности переменной x_j? Каковы ваши рекомендации относительно того, чтобы сделать переменную x_j приносящей доход?[1]

19. Дана следующая задача ЛП: максимизировать $z = \mathbf{CX}$ при ограничениях $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$, $\mathbf{X} \ge \mathbf{0}$. Обозначим через $\mathbf{X}_B$ текущий базисный вектор, через $\mathbf{C}_B$ — вектор коэффициентов целевой функции, соответствующих базисным переменным. Допустим, что вектор $\mathbf{C}_B$ изменен на $\mathbf{D}_B$. Докажите, что в этом случае разности $z_j - c_j$, соответствующие базисному вектору $\mathbf{X}_B$, останутся равными нулю. Дайте объяснение этому.

7.5. Эффективные вычислительные алгоритмы

В этом разделе представлены разновидности симплекс-метода, среди которых **модифицированный симплекс-метод, метод решения задач с ограниченными переменными** и **метод декомпозиции**. Эти методы изначально разрабатывались для повышения вычислительной эффективности симплекс-метода.

---

[1] Этот вопрос можно переформулировать без "экономического" подтекста: при каких условиях переменная x_j может быть представлена в оптимальном решении? — *Прим. ред.*

7.5.1. Модифицированный симплекс-метод

Модифицированный симплексный алгоритм предусматривает выполнение *точно таких же шагов*, как и обычный табличный симплекс-метод, описанный в главе 3. Отличие между ними заключается в том, что в обычном симплекс-методе при переходе от одного базиса к другому используется процедура преобразования строк симплекс-таблицы с помощью метода Гаусса-Жордана, тогда как в модифицированном симплекс-методе эти преобразования осуществляются путем вычисления обратной матрицы $\mathbf{B}^{-1}$, и основные действия связаны именно с нахождением этой матрицы.

Типовая схема модифицированного симплекс-метода представлена в примере 7.4–1. Однако там не выяснен вопрос о вычислении на каждой итерации обратной матрицы $\mathbf{B}^{-1}$. В модифицированном симплекс-методе вместо непосредственного обращения матрицы $\mathbf{B}$ используется метод вычисления обратной матрицы, основанный на мультипликативном представлении обратной матрицы (см. раздел А.7.4).

Мультипликативное представление обратной матрицы. В симплекс-методе последовательные базисы $\mathbf{B}$ и $\mathbf{B}_{\text{след}}$ различаются только одним вектор-столбцом, получаемым после замены исключаемого из базиса вектора вводимым. Это идеально для мультипликативного представления обратной матрицы.

Имея матрицу $\mathbf{B}^{-1}$, мы можем вычислить матрицу $\mathbf{B}^{-1}_{\text{след}}$ с помощью формулы

$$\mathbf{B}^{-1}_{\text{след}} = \mathbf{E}\mathbf{B}^{-1}.$$

Чтобы вычислить матрицу $\mathbf{E}$, обозначим через $\mathbf{P}_j$ и $\mathbf{P}_r$ вводимый в базис и исключаемый из него векторы, определенные на текущей итерации выполнения симплекс-метода. Тогда матрицу $\mathbf{E}$ можно определить как m-мерную единичную матрицу, у которой r-й столбец заменен следующим столбцом.

$$\xi = \frac{1}{(\mathbf{B}^{-1}\mathbf{P}_j)_r}\begin{bmatrix} -(\mathbf{B}^{-1}\mathbf{P}_j)_1 \\ -(\mathbf{B}^{-1}\mathbf{P}_j)_2 \\ \vdots \\ +1 \\ \vdots \\ -(\mathbf{B}^{-1}\mathbf{P}_j)_m \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow r\text{-й элемент} \\ \\ \\ \end{matrix}$$

Здесь предполагается, что $(\mathbf{B}^{-1}\mathbf{P}_j)_r \neq 0$, в противном случае обратной матрицы $\mathbf{B}^{-1}_{\text{след}}$ не существует.

Докажем справедливость формулы $\mathbf{B}^{-1}_{\text{след}} = \mathbf{E}\mathbf{B}^{-1}$. Обозначим через $\mathbf{F}$ m-мерную единичную матрицу, у которой r-й столбец заменен столбцом $\mathbf{B}^{-1}\mathbf{P}_j$, т.е.

$$\mathbf{F} = (\mathbf{e}_1, \ldots, \mathbf{e}_{r-1}, \mathbf{B}^{-1}\mathbf{P}_j, \mathbf{e}_{r+1}, \ldots, \mathbf{e}_m).$$

Поскольку матрица $\mathbf{B}_{\text{след}}$ отличается от матрицы $\mathbf{B}$ только r-м столбцом, который в матрице $\mathbf{B}_{\text{след}}$ совпадает с вектором $\mathbf{P}_j$, то легко проверить равенство

$$\mathbf{B}_{\text{след}} = \mathbf{B}\mathbf{F}.$$

Отсюда получаем

$$\mathbf{B}_{\text{след}}^{-1} = (\mathbf{BF})^{-1} = \mathbf{F}^{-1}\mathbf{B}^{-1}.$$

Теперь осталось положить $\mathbf{E} = \mathbf{F}^{-1}$.

Пример 7.5–1

Пусть нам известны следующие матрицы.

$$\mathbf{B} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 4 & 0 & 1 \end{bmatrix}, \qquad \mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix}.$$

Предположим, что новая матрица $\mathbf{B}_{\text{след}}$ получается из матрицы $\mathbf{B}$ путем замены третьего вектор-столбца $\mathbf{P}_3 = (0, 0, 1)^T$ на вектор-столбец $\mathbf{V}_3 = (2, 1, 5)^T$. Вычислим $\mathbf{B}_{\text{след}}^{-1}$.

Поскольку в матрице $\mathbf{B}$ заменен *третий* столбец, $r = 3$. Из

$$\mathbf{B}^{-1}\mathbf{V}_3 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} \\ \frac{1}{2} \\ \mathbf{2} \end{bmatrix} \leftarrow r = 3$$

получаем

$$\xi = \frac{1}{2}\begin{bmatrix} -\frac{3}{4} \\ -\frac{1}{2} \\ +1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{8} \\ -\frac{1}{4} \\ \frac{1}{2} \end{bmatrix}.$$

Таким образом,

$$\mathbf{B}_{\text{след}}^{-1} = \begin{bmatrix} 1 & 0 & -\frac{3}{8} \\ 0 & 1 & -\frac{1}{4} \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{4} & -\frac{5}{8} & -\frac{3}{8} \\ \frac{1}{2} & \frac{1}{4} & -\frac{1}{4} \\ -1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Упражнения 7.5,а

1. Пусть в примере 7.5–1, кроме изменения третьего столбца матрицы **B**, также заменен второй вектор-столбец $\mathbf{P}_2$ на вектор-столбец $\mathbf{V}_2 = (1, 1, 1)T$. Найдите соответствующую обратную матрицу.

2. Пусть в примере 7.5–1 третий вектор-столбец матрицы **B** заменен на вектор-столбец $\mathbf{V}_3 = \mathbf{P}_1 + 2\mathbf{P}_2$. В этом случае матрица $\mathbf{B}_{\text{след}}$ будет вырожденной. Покажите, как с помощью мультипликативного представления обратной матрицы можно обнаружить ее вырожденность.

3. С помощью мультипликативного представления обратной матрицы определите, какая из следующих систем уравнений имеет единственное решение, бесконечно много решений и не имеет решения.

   a) $x_1 + 2x_2 = 3,$
      $x_1 + 4x_2 = 2.$

   b) $x_1 + 2x_2 = 5,$
      $-x_1 - 2x_2 = -5.$

   c) $x_1 + x_3 = 5,$
      $4x_1 + x_2 + 3x_3 = 8,$
      $x_1 + 3x_2 - 2x_3 = 3.$

Вычислительная процедура модифицированного симплекс-метода. Мультипликативная форма обратной матрицы используется в модифицированном симплекс-методе следующим образом. Поскольку начальный базис всегда определяет единичную матрицу **I** (которая совпадает с обратной к себе), матрицу $\mathbf{B}_i^{-1}$, соответствующую i-й итерации, можно представить в следующем виде.

$$\mathbf{B}_i^{-1} = \mathbf{E}_i\mathbf{E}_{i-1}\ldots\mathbf{E}_1\mathbf{I}$$

Таким образом, отпадает необходимость вычисления каких-либо обратных матриц.

В остальном последовательность шагов выполнения модифицированного симплекс-метода полностью совпадает с аналогичной последовательностью стандартного табличного симплекс-метода. В этом легко убедиться, так как модифицированный симплекс-метод использует ту же знакомую нам симплекс-таблицу (в этой таблице $z_j - c_j = \mathbf{C}_B\mathbf{B}^{-1}\mathbf{P}_j - c_j$):

Базис	x_1	x_2	...	x_j	...	x_n	Решение
z	$z_1 - c_1$	$z_2 - c_2$	...	$z_j - c_j$	...	$z_n - c_n$	$\mathbf{C}_B\mathbf{B}^{-1}\mathbf{b}$
$\mathbf{X}_B$				$\mathbf{B}^{-1}\mathbf{P}_j$			$\mathbf{B}^{-1}\mathbf{b}$

Пример 7.5–2

С помощью модифицированного симплекс-метода решим заново задачу о компании Reddy Mikks из раздела 2.1. Решение этой задачи обычным симплекс-методом дано в разделе 3.3. Сравнение двух решений данной задачи показывает, что оба метода выполняют одну и ту же последовательность действий.

Представим в матричном виде рассматриваемую задачу, уже приведенную к стандартной форме:

$$\text{Максимизировать } z = (5, 4, 0, 0, 0, 0)(x_1, x_2, x_3, x_4, x_5, x_6)^T$$

при ограничениях

$$\begin{bmatrix} 6 & 4 & 1 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 24 \\ 6 \\ 1 \\ 2 \end{bmatrix}.$$

$$x_1, x_2, x_3, x_4, x_5, x_6 \geq 0.$$

Мы используем запись $\mathbf{C} = (c_1, c_2, ..., c_6)$ для представления вектора коэффициентов целевой функции и $(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_6)$ — для представления столбцов матрицы $(\mathbf{A}, \mathbf{I})$ коэффициентов левых частей ограничений. Вектор правых частей ограничений обозначим $\mathbf{b}$.

В следующих вычислениях мы приводим формулы, по которым выполняются вычисления, и конечный результат без подробных арифметических выкладок, которые читатель может выполнить самостоятельно.

Итерация 0

$$\mathbf{X}_{B_0} = (x_3, x_4, x_5, x_6)^T,\ \mathbf{C}_{B_0} = (0, 0, 0, 0).$$

$$\mathbf{B}_0 = (\mathbf{P}_3, \mathbf{P}_4, \mathbf{P}_5, \mathbf{P}_6) = \mathbf{I},\ \ \mathbf{B}_0^{-1} = \mathbf{I}.$$

Таким образом,

$$\mathbf{X}_{B_0} = \mathbf{B}_0^{-1}\mathbf{b} = (24, 6, 1, 2)^T,\ z = \mathbf{C}_{B_0}\mathbf{X}_{B_0} = 0.$$

Вычисления условия оптимальности

$$\mathbf{C}_{B_0}\mathbf{B}_0^{-1} = (0, 0, 0, 0).$$

$$\{z_j - c_j\}_{j=1,2} = \mathbf{C}_{B_0}\mathbf{B}_0^{-1}(\mathbf{P}_1, \mathbf{P}_2) - (c_1, c_2) = (-5, -4).$$

Отсюда следует, что вводимым в базис вектором будет $\mathbf{P}_1$.

Вычисления условия допустимости

$$\mathbf{X}_{B_0} = (x_3, x_4, x_5, x_6)^T = (24, 6, 1, 2)^T.$$

$$\mathbf{B}_0^{-1}\mathbf{P}_1 = (6, 1, -1, 0)^T.$$

Следовательно,

$$x_1 = \min\left\{\frac{24}{6}, \frac{6}{1}, —, —\right\} = \min\{4, 6, —, —\} = 4,$$

и вектор $\mathbf{P}_3$ определяется как исключаемый из базиса.

Результаты выполненных вычислений представим в виде знакомой симплекс-таблицы. Такое представление поможет сравнить модифицированный и обычный симплекс-методы.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	–5	–4	0	0	0	0	0
x_3	6						24
x_4	1						6
x_5	–1						1
x_6	0						2

Итерация 1

$$\mathbf{X}_{B_1} = (x_1, x_4, x_5, x_6)^T,\ \mathbf{C}_{B_1} = (5, 0, 0, 0).$$
$$\mathbf{B}_1 = (\mathbf{P}_1, \mathbf{P}_4, \mathbf{P}_5, \mathbf{P}_6).$$

Для вычисления $\mathbf{B}_1^{-1}$ заметим, что $\mathbf{B}_0^{-1}\mathbf{P}_1 = (6, 1, -1, 0)^T$. Поскольку в матрице $\mathbf{B}_0$ заменен *первый* столбец, значит, $r = 1$ (информация, необходимая для формирования столбца ξ). Вычисляем

$$\mathbf{B}_1^{-1} = \mathbf{E}\mathbf{B}_0^{-1} = \mathbf{E}\mathbf{I} = \mathbf{E} = \begin{bmatrix} \frac{\mathbf{1}}{\mathbf{6}} & 0 & 0 & 0 \\ -\frac{\mathbf{1}}{\mathbf{6}} & 1 & 0 & 0 \\ \frac{\mathbf{1}}{\mathbf{6}} & 0 & 1 & 0 \\ \mathbf{0} & 0 & 0 & 1 \end{bmatrix}.$$

Отсюда получаем

$$\mathbf{X}_{B_1} = \mathbf{B}_1^{-1}\mathbf{b} = (4, 2, 5, 2)^T,\ z = \mathbf{C}_{B_1}\mathbf{X}_{B_1} = 20.$$

Вычисления условия оптимальности

$$\mathbf{C}_{B_1}\mathbf{B}_1^{-1} = \left(\frac{5}{6}, 0, 0, 0\right).$$

$$\{z_j - c_j\}_{j=2,3} = \mathbf{C}_{B_1}\mathbf{B}_1^{-1}(\mathbf{P}_2, \mathbf{P}_3) - (c_2, c_3) = \left(-\frac{2}{3}, \frac{5}{6}\right).$$

Отсюда следует, что вводимым в базис будет вектор $\mathbf{P}_2$.

Вычисления условия допустимости

$$\mathbf{X}_{B_1} = (x_1, x_4, x_5, x_6)^T = (4, 2, 5, 2)^T.$$

$$\mathbf{B}_1^{-1}\mathbf{P}_2 = \left(\frac{2}{3}, \frac{4}{3}, \frac{5}{3}, 1\right)^T.$$

Следовательно,

$$x_2 = \min\left\{6, \frac{3}{2}, 3, 2\right\} = \frac{3}{2},$$

и вектор $\mathbf{P}_4$ определяется как исключаемый из базиса. (Постройте симплекс-таблицу, отображающую результаты вычислений этой итерации.)

Итерация 2

$$\mathbf{X}_{B_2} = (x_1, x_2, x_5, x_6)^T,\ \mathbf{C}_{B_2} = (5, 4, 0, 0).$$

$$\mathbf{B}_2 = (\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_5, \mathbf{P}_6).$$

Для вычисления $\mathbf{B}_2^{-1}$ имеем $\mathbf{B}_1^{-1}\mathbf{P}_2 = \left(\frac{2}{3}, \frac{4}{3}, \frac{5}{3}, 1\right)^T$. Поскольку в матрице $\mathbf{B}_1$ заменен *второй* столбец, значит, $r = 2$. Вычисляем

$$\mathbf{B}_2^{-1} = \mathbf{E}\mathbf{B}_1^{-1} = \begin{bmatrix} 1 & -\frac{2}{4} & 0 & 0 \\ 0 & \frac{3}{4} & 0 & 0 \\ 0 & -\frac{5}{4} & 1 & 0 \\ 0 & -\frac{3}{4} & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 \\ -\frac{1}{6} & 1 & 0 & 0 \\ \frac{1}{6} & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{8} & \frac{3}{4} & 0 & 0 \\ \frac{3}{8} & -\frac{5}{4} & 1 & 0 \\ \frac{1}{8} & -\frac{3}{4} & 0 & 1 \end{bmatrix}.$$

Отсюда получаем

$$\mathbf{X}_{B_2} = \mathbf{B}_2^{-1}\mathbf{b} = \left(3, \frac{3}{2}, \frac{5}{2}, \frac{1}{2}\right)^T,\ z = \mathbf{C}_{B_2}\mathbf{X}_{B_2} = 21.$$

Вычисления условия оптимальности

$$\mathbf{C}_{B_2}\mathbf{B}_2^{-1} = \left(\frac{3}{4}, \frac{1}{2}, 0, 0\right).$$

$$\{z_j - c_j\}_{j=3,4} = \mathbf{C}_{B_2}\mathbf{B}_2^{-1}(\mathbf{P}_3, \mathbf{P}_4) - (c_3, c_4) = \left(\frac{3}{4}, \frac{1}{2}\right).$$

Отсюда следует, что решение $\mathbf{X}_{B_2}$ оптимально. Вычисления заканчиваются.

Оптимальное решение:

$$x_1 = 3,\ x_2 = 1.5,\ z = 21.$$

Вычисления в соответствии с модифицированным симплекс-методом имеют два преимущества (по сравнению с обычным симплекс-методом).

1. Число арифметических операций может быть значительно меньше, чем в табличном симплекс-методе, в зависимости от размерности задачи (т.е. от количества ограничений и переменных) и разреженности (количества ненулевых элементов) матриц (векторов) $\mathbf{C}$, $\mathbf{A}$ и $\mathbf{b}$.

2. Поскольку в этом методе вычисления производятся, в основном, для нахождения обратной матрицы $\mathbf{B}^{-1}$ и основаны на *исходных данных*, машинную ошибку округления можно контролировать путем отслеживания точности вычисления элементов матрицы $\mathbf{B}^{-1}$. В противоположность этому, в обычном симплекс-методе машинная ошибка округления может накапливаться и распространяться по итерациям.

Упражнения 7.5,b

1. В примере 7.5–2 представьте результаты вычислений, выполненных на первой и второй итерациях, в виде симплекс-таблиц.
2. Решите модифицированным симплекс-методом следующие задачи ЛП.

   a) Максимизировать $z = 6x_1 + 2x_2 + 3x_3$
   при ограничениях
   $$2x_1 - x_2 + 2x_3 \le 2,$$
   $$x_1 + 4x_3 \le 4,$$
   $$x_1, x_2, x_3 \ge 0.$$

   b) Максимизировать $z = 2x_1 + x_2 + 2x_3$
   при ограничениях
   $$4x_1 + 3x_2 + 8x_3 \le 12,$$
   $$4x_1 + x_2 + 12x_3 \le 8,$$
   $$4x_1 - x_2 + 3x_3 \le 8,$$
   $$x_1, x_2, x_3 \ge 0.$$

   c) Минимизировать $z = 2x_1 + x_2$
   при ограничениях
   $$3x_1 + x_2 = 3,$$
   $$4x_1 + 3x_2 \ge 6,$$
   $$x_1 + 2x_2 \le 3,$$
   $$x_1, x_2 \ge 0.$$

   d) Минимизировать $z = 5x_1 - 4x_2 + 6x_3 + 8x_4$
   при ограничениях
   $$x_1 + 7x_2 + 3x_3 + 7x_4 \le 46,$$
   $$3x_1 - x_2 + x_3 + 2x_4 \le 20,$$
   $$2x_1 + 3x_2 - x_3 + x_4 \ge 18,$$
   $$x_1, x_2, x_3, x_4 \ge 0.$$

3. Решите следующую задачу ЛП с помощью модифицированного симплекс-метода, используя в качестве начального базиса вектор $\mathbf{X}_{B_0} = (x_2, x_4, x_5)$.

$$\text{Минимизировать } z = 7x_2 + 11x_3 - 10x_4 + 26x_6$$

при ограничениях
$$x_2 - x_3 + x_5 + x_6 = 6,$$
$$x_2 - x_3 + x_4 + 3x_6 = 8,$$

$$x_1 + x_2 - 3x_3 + x_4 + x_5 = 12,$$

$$x_1, x_2, x_3, x_4, x_5, x_6 \geq 0.$$

(*Совет.* Сначала примените мультипликативную форму представления обратной матрицы для нахождения $\mathbf{B}_0^{-1}$, затем выполните процедуру модифицированного симплекс-метода.)

4. Используя модифицированный симплекс-метод в схеме вычислений двухэтапного метода с искусственными переменными (раздел 3.4.2), решите следующие задачи.

   a) Задача из упр. 2(c).
   b) Задача из упр. 2(d).
   c) Задача из упр. 3 (отказавшись от предложенного начального решения $\mathbf{X}_{B_0}$).

5. *Модифицированный двойственный симплекс-метод.* Последовательность шагов выполнения модифицированного двойственного симплекс-метода (использующего матричные вычисления) можно описать следующим образом.

Шаг 0. Пусть $\mathbf{B}_0 = \mathbf{I}$ — начальный базис, причем хотя бы один из элементов вектора $\mathbf{X}_{B_0}$ отрицателен (т.е. решение недопустимо).

Шаг 1. Вычисляем текущие значения базисных переменных: $\mathbf{X}_B = \mathbf{B}^{-1}\mathbf{b}$. Выбираем в качестве исключаемой из базиса переменную, имеющую наибольшее отрицательное значение (обозначим исключаемую переменную x_r). Если все элементы в $\mathbf{X}_B$ неотрицательны, то вычисления заканчиваются, так как достигнуто допустимое решение.

Шаг 2.

a) Для всех небазисных переменных x_j вычисляем разности $z_j - c_j = \mathbf{C}_B\mathbf{B}^{-1}\mathbf{P}_j - c_j$.

b) Для всех небазисных переменных x_j вычисляем коэффициенты ограничений $(\mathbf{B}^{-1}\mathbf{P}_j)_r$, ассоциированные со строкой, соответствующей исключаемой переменной x_r.

c) Определяем вводимую в базис переменную как переменную, на которой достигается следующий минимум.

$$\theta = \min_j \left\{ \left| \frac{z_j - c_j}{(\mathbf{B}^{-1}\mathbf{P}_j)_r} \right|,\ (\mathbf{B}^{-1}\mathbf{P}_j)_r < 0 \right\}$$

Если все $(\mathbf{B}^{-1}\mathbf{P}_j)_r \geq 0$, допустимого решения не существует.

Шаг 3. Формируем новый базис путем замены в базисе вектора $\mathbf{P}_j$ на $\mathbf{P}_r$. Вычисляем обратную матрицу $\mathbf{B}_{\text{след}}^{-1} = \mathbf{E}\mathbf{B}^{-1}$, полагаем $\mathbf{B}^{-1} = \mathbf{B}_{\text{след}}^{-1}$ и переходим к шагу 1.

Примените описанный метод для решения следующей задачи.

Минимизировать $z = 2x_1 + x_2$

при ограничениях

$$3x_1 + x_2 \geq 3,$$
$$4x_1 + 3x_2 \geq 6,$$
$$x_1 + 2x_2 \leq 3,$$
$$x_1, x_2 \geq 0.$$

7.5.2. Алгоритм решения задач с ограниченными переменными

В некоторых моделях линейного программирования на значения переменных могут накладываться явные положительные верхние и нижние ограничения. Например, в производственных моделях нижние и верхние границы переменных могут соответствовать минимальным и максимальным значениям спроса на определенную продукцию. Явные ограничения, накладываемые на переменные, особенно заметны при решении задач целочисленного программирования методом ветвей и границ (см. раздел 9.3.1).

Специальный алгоритм решения задач с ограниченными переменными, который мы рассмотрим в этом разделе, особенно эффективен, поскольку он учитывает тот факт, что ограничения заданы в *явном виде*. Сначала рассмотрим более простой случай, когда заданы только нижние границы на значения переменных. При явных ограничениях $\mathbf{X} \geq \mathbf{L}$ замена (подстановка) типа

$$\mathbf{X} = \mathbf{L} + \mathbf{X}', \quad \mathbf{X}' \geq 0$$

позволяет решать задачу ЛП относительно переменных $\mathbf{X}'$, для которых нижняя граница значений равна нулю. Значения исходных переменных $\mathbf{X}$ вычисляются путем обратной подстановки; при этом, очевидно, гарантирована их неотрицательность.

Теперь рассмотрим ситуацию с верхними границами $\mathbf{X} \leq \mathbf{U}$. В данном случае прямая замена $\mathbf{X} = \mathbf{U} - \mathbf{X}''$, $\mathbf{X}'' \geq \mathbf{0}$, не возможна, поскольку обратная подстановка не гарантирует неотрицательности $\mathbf{X}$. Здесь необходимо применение других процедур.

Запишем задачу ЛП с верхними границами на значения переменных как

$$\text{Максимизировать } z = \{\mathbf{CX} \mid (\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}, \mathbf{0} \leq \mathbf{X} \leq \mathbf{U}\}.$$

При выполнении алгоритма решения задач с ограниченными переменными явно используются только ограничения $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$, $\mathbf{X} \geq \mathbf{0}$; ограничения вида $\mathbf{X} \leq \mathbf{U}$ учитываются лишь в измененном симплексном условии допустимости.

Пусть $\mathbf{X}_B = \mathbf{B}^{-1}\mathbf{b}$ — текущее базисное допустимое решение задачи ЛП с ограничениями $(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b}$, $\mathbf{X} \geq \mathbf{0}$. Обозначим через $\mathbf{P}_j$ вводимый в базис вектор, определенный на основе обычного симплексного условия оптимальности. Если предположить, что *все небазисные переменные равны нулю*, уравнение ограничения относительно базисной переменной x_i будет записано следующим образом.

$$(\mathbf{X}_B)_i = (\mathbf{B}^{-1}\mathbf{b})_i - (\mathbf{B}^{-1}\mathbf{P}_j)_i\, x_j$$

Поскольку вводимая переменная x_j примет положительное значение, величина базисной переменной $(\mathbf{X}_B)_i$ *возрастет или уменьшится*, в зависимости от того, будет ли значение $(\mathbf{B}^{-1}\mathbf{P}_j)_i$ отрицательным или положительным. Таким образом, при определении значения вводимой переменной x_j необходимо удовлетворить следующие три условия.

1. Базисная переменная $(\mathbf{X}_B)_i$ должна остаться неотрицательной, т.е. $(\mathbf{X}_B)_i \geq 0$.
2. Базисная переменная $(\mathbf{X}_B)_i$ не должна превышать своей верхней границы, т.е. должно выполняться неравенство $(\mathbf{X}_B)_i \leq (\mathbf{U}_B)_i$, где $\mathbf{U}_B$ — вектор, содержащий упорядоченные элементы вектора $\mathbf{U}$, соответствующие вектору $\mathbf{X}_B$.
3. Вводимая переменная x_j не может принять значения, большего своей верхней границы, т.е. должно выполняться неравенство $x_j \leq u_j$, где u_j — j-й элемент вектора $\mathbf{U}$.

Первое условие $(\mathbf{X}_B)_i \geq 0$ требует выполнения неравенства

$$(\mathbf{B}^{-1}\mathbf{b})_i - (\mathbf{B}^{-1}\mathbf{P}_j)_i\, x_j \geq 0,$$

которое заведомо будет выполняться, если

$$x_j \leq \theta_1 \equiv \min_i \left\{ \frac{(\mathbf{B}^{-1}\mathbf{b})_i}{(\mathbf{B}^{-1}\mathbf{P}_j)_i} \middle| \, (\mathbf{B}^{-1}\mathbf{P}_j)_i > 0 \right\}.$$

Это условие полностью соответствует условию допустимости обычного симплекс-метода.

Далее, условие $(\mathbf{X}_B)_i \leq (\mathbf{U}_B)_i$ эквивалентно неравенству

$$(\mathbf{B}^{-1}\mathbf{b})_i - (\mathbf{B}^{-1}\mathbf{P}_j)_i\, x_j \leq (\mathbf{U}_B)_i,$$

которое будет выполняться, если

$$x_j \leq \theta_2 \equiv \min_i \left\{ \frac{(\mathbf{B}^{-1}\mathbf{b})_i - (\mathbf{U}_B)_i}{(\mathbf{B}^{-1}\mathbf{P}_j)_i} \middle| \, (\mathbf{B}^{-1}\mathbf{P}_j)_i < 0 \right\}.$$

Наконец, третье условие будет выполнено, если выполняется неравенство $x_j \leq u_j$.

Все три неравенства относительно x_j можно обобщить в виде следующего условия.

$$x_j = \min\{\theta_1, \theta_2, u_j\}.$$

Базис, формируемый для следующей итерации, зависит от того, какое значение (θ_1, θ_2 или u_j) примет переменная x_j. Предполагая, что $(\mathbf{X}_B)_r$ —исключаемая переменная, имеем следующие правила формирования базиса.

1. Если $x_j = \theta_1$, то переменная $(\mathbf{X}_B)_r$ исключается из базиса и принимает нулевое значение. Новый базис формируется точно так же, как в обычном симплекс-методе с вводимой переменной x_j и исключаемой $(\mathbf{X}_B)_r$.
2. Если $x_j = \theta_2$, переменная $(\mathbf{X}_B)_r$ исключается из базиса и принимает значение *своей верхней границы*. Новый базис формируется точно так же, как в обычном симплекс-методе, но с учетом того, что переменная $(\mathbf{X}_B)_r$ равна *верхней границе*. Поскольку значения θ_1 и θ_2 получены при условии, что *все небазисные переменные равны нулю*, необходимо выполнить преобразование над новой небазисной переменной $(\mathbf{X}_B)_r$ такое, чтобы она также приняла нулевое значение. Это достигается с помощью подстановки $(\mathbf{X}_B)_r = (\mathbf{U}_B)_r - (\mathbf{X}_B)_r'$, где $(\mathbf{X}_B)_r' \geq 0$. Не имеет особого значения, когда выполняется эта подстановка: до или после вычисления нового базиса.

3. Если $x_j = u_j$, то базисный вектор $\mathbf{X}_B$ остается неизменным, поскольку в данном случае никакая из текущих базисных переменных не принимает ни нулевого значения, ни значения своей верхней границы. Таким образом, переменная x_j остается небазисной, но со *значением своего верхнего предела*. После подстановки $x_j = u_j - x_j'$ выполняется следующая итерация симплекс-метода.

В случае равенства значений θ_1, θ_2 и u_j выбор правила, в соответствии с которым переходим к следующей итерации симплекс-метода, произволен. Вместе с тем предпочтительнее использование третьего правила (случай $x_j = u_j$), так как в этом случае требуется выполнить меньший объем вычислений.

Подстановка $x_j = u_j - x_j'$ изменяет исходные значения c_j, $\mathbf{P}_j$ и $\mathbf{b}$ на $c_j' = -c_j$, $\mathbf{P}_j' = -\mathbf{P}_j$ и $\mathbf{b}' = \mathbf{b} - u_j\mathbf{P}_j$. Это означает, что при выполнении модифицированного симплекс-метода на каждой итерации все вычисления (таких величин, как $\mathbf{B}^{-1}$, $\mathbf{X}_B$ и $z_j - c_j$) должны основываться на измененных значениях $\mathbf{C}$, $\mathbf{A}$ и $\mathbf{b}$ (более подробно этот вопрос рассмотрен в упр. 7.5,с(5)).

Пример 7.5–3

Решим следующую задачу ЛП вышеописанным методом решения задач с ограниченными переменными.

$$\text{Максимизировать } z = 3x_1 + 5y + 2x_3$$

при ограничениях

$$x_1 + y + 2x_3 \le 14,$$
$$2x_1 + 4y + 3x_3 \le 43,$$
$$0 \le x_1 \le 4,\ 7 \le y \le 10,\ 0 \le x_3 \le 3.$$

Так как переменная y ограничена положительной константой не только сверху, но и снизу, делаем замену $y = x_2 + 7$, где $0 \le x_2 \le 10 - 7 = 3$.

Во избежание излишних вычислительных сложностей здесь мы применим не модифицированный симплекс-метод, а обычный симплекс-метод в виде компактных симплекс-таблиц.

Итерация 0

Базис	x_1	x_2	x_3	x_4	x_5	Решение
z	–3	–5	–2	0	0	35
x_4	1	1	2	1	0	7
x_5	2	4	3	0	1	15

Имеем $\mathbf{B} = \mathbf{B}^{-1} = \mathbf{I}$ и $\mathbf{X}_B = (x_4, x_5)^T = \mathbf{B}^{-1}\mathbf{b} = (7, 15)^T$. Принимая переменную x_2 в качестве вводимой в базис переменной ($z_2 - c_2 = -5$), получаем $\mathbf{B}^{-1}\mathbf{P}_2 = (1, 4)^T$. Отсюда следует

$$\theta_1 = \min\left\{\frac{7}{1}, \frac{15}{4}\right\} = 3.75, \text{ что соответствует переменной } x_5,$$

$\theta_2 = \infty$ (поскольку $\mathbf{B}^{-1}\mathbf{P}_2 > \mathbf{0}$).

Так как по условию $x_2 \le 3$, получаем значение, принимаемое переменной x_2.

$$x_2 = \min\{3.75, \infty, 3\} = 3 = u_2.$$

Поскольку $x_2 = u_2$, базис остается неизменным; переменная x_2 в базис не вводится (остается небазисной), но принимает значение своей верхней границы. После подстановки $x_2 = 3 - x_2'$ получаем новую симплекс-таблицу.

Базис	x_1	x_2'	x_3	x_4	x_5	Решение
z	–3	5	–2	0	0	50
x_4	1	–1	2	1	0	4
x_5	2	–4	3	0	1	3

Выполненная подстановка изменяет исходный вектор правых частей ограничений с $\mathbf{b} = (7, 15)^T$ на $\mathbf{b}' = (4, 3)^T$. Эти изменения должны учитываться в последующих вычислениях.

Итерация 1

Определяем вводимой в базис переменную x_1. Базисный вектор $\mathbf{X}_B$ и обратная матрица $\mathbf{B}^{-1}$ (= $\mathbf{I}$) такие же, как на предыдущей итерации. Вычисляем

$$\mathbf{B}^{-1}\mathbf{P}_1 = (1, 2)^T.$$

Отсюда следует

$$\theta_1 = \min\left\{\frac{4}{1}, \frac{3}{2}\right\} = 1.5, \text{ что соответствует переменной } x_5,$$

$$\theta_2 = \infty \text{ (поскольку } \mathbf{B}^{-1}\mathbf{P}_1 > \mathbf{0}).$$

Так как по условию задачи $x_1 \le 4$, получаем значение, принимаемое переменной x_1.

$$x_1 = \min\{\mathbf{1.5}, \infty, 4\} = 1.5 = \theta_1.$$

В данной ситуации переменная x_1 становится базисной со значением 1.5, переменная x_5 выводится из базиса и принимает нулевое значение. Получаем следующую симплекс-таблицу.

Базис	x_1	x_2'	x_3	x_4	x_5	Решение
z	0	–1	5/2	0	3/2	109/2
x_4	0	1	1/2	1	–1/2	5/2
x_1	1	–2	3/2	0	1/2	3/2

Итерация 2

Имеем новую обратную матрицу.

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$$

Значения базисных переменных вычисляются так: $\mathbf{X}_B = (x_4, x_1)^T = \mathbf{B}^{-1}\mathbf{b}' = (5/2, 3/2)^T$, где $\mathbf{b}' = (4, 3)^T$ — значения правых частей ограничений, вычисленные в конце нулевой итерации.

Определяем x_2' как вводимую в базис переменную. Учитывая, что $\mathbf{P}_2' = -\mathbf{P}_2$, получаем

$$\mathbf{B}^{-1}\mathbf{P}_2' = (1, -2)^T.$$

Далее вычисляем

$$\theta_1 = \min\left\{\frac{\frac{5}{2}}{1}, -\right\} = 2.5,$$ что соответствует базисной переменной x_4,

$$\theta_2 = \min\left\{-, \frac{\frac{3}{2}-4}{-2}\right\} = 1.25,$$ что соответствует базисной переменной x_1.

Так как по условию $x_2' \leq 3$, получаем значение, принимаемое переменной x_2'.

$$x_2' = \min\{2.5, \mathbf{1.25}, 3\} = 1.25 = \theta_2.$$

Таким образом, переменная x_2' вводится в базис (со значением 1.25), из которого исключается переменная x_1 с ненулевым значением, равным ее верхней границе. Получаем следующую симплекс-таблицу.

Базис	x_1	x_2'	x_3	x_4	x_5	Решение
z	–1/2	0	7/4	0	5/4	215/4
x_4	1/2	0	5/4	1	–1/4	13/4
x_2'	–1/2	1	–3/4	0	–1/4	–3/4

Поскольку переменная x_1 становится небазисной с ненулевым значением (= 4), применяем подстановку $x_1 = 4 - x_1'$, что приводит к следующей симплекс-таблице.

Базис	x_1'	x_2'	x_3	x_4	x_5	Решение
z	1/2	0	7/4	0	5/4	223/4
x_4	–1/2	0	5/4	1	–1/4	5/4
x_2'	1/2	1	–3/4	0	–1/4	5/4

Данная таблица представляет допустимое оптимальное решение. Оптимальные значения переменных x_1, x_2 и x_3 получаем обратной подстановкой: $x_1 = u_1 - x_1' = 4 - 0 = 4$, $x_2 = u_2 - x_2' = 3 - 5/4 = 7/4$ и $x_3 = 0$. Теперь вычисляем значение переменной y: $y = l_2 + x_2 = 7 + 7/4 = 35/4$. Оптимальное значение целевой функции равно $z = 223/4$.

Упражнения 7.5,с

1. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = 2x_1 + x_2$$

при ограничениях

$$x_1 + x_2 \le 3,$$
$$0 \le x_1 \le 2,\ 0 \le x_2 \le 2.$$

   a) Решите задачу графическим способом и определите последовательность крайних точек пространства решений, ведущих к оптимальному решению.
   b) Решите задачу с использованием метода решения задач с ограниченными переменными и покажите, что данный метод порождает ту же последовательность крайних точек пространства решений, приводящую к оптимальному решению, что и графический способ.
   c) Как алгоритм решения задач с ограниченными переменными распознает крайние точки пространства решений?

2. Решите следующую задачу ЛП методом решения задач с ограниченными переменными.

$$\text{Максимизировать } z = 6x_1 + 2x_2 + 8x_3 + 4x_4 + 2x_5 + 10x_6$$

при ограничениях

$$8x_1 + x_2 + 8x_3 + 2x_4 + 2x_5 + 4x_6 \le 13,$$
$$0 \le x_j \le 1,\ j = 1, 2, ..., 6.$$

3. Решите следующие задачи ЛП методом решения задач с ограниченными переменными.

   a) Минимизировать $z = 6x_1 - 2x_2 - 3x_3$
   при ограничениях

$$2x_1 + 4x_2 + 2x_3 \le 8,$$
$$x_1 - 2x_2 + 3x_3 \le 7,$$
$$0 \le x_1 \le 2,\ 0 \le x_2 \le 2,\ 0 \le x_3 \le 1.$$

   b) Максимизировать $z = 3x_1 + 5x_2 + 2x_3$
   при ограничениях

$$x_1 + 2x_2 + 2x_3 \le 10,$$
$$2x_1 + 4x_2 + 3x_3 \le 15,$$
$$0 \le x_1 \le 4,\ 0 \le x_2 \le 3,\ 0 \le x_3 \le 3.$$

4. В следующих задачах некоторые переменные имеют положительные нижние границы. Решите эти задачи методом решения задач с ограниченными переменными.

   a) Максимизировать $z = 3x_1 + 2x_2 - 2x_3$
   при ограничениях

$$2x_1 + x_2 + x_3 \le 8,$$

$$x_1 + 2x_2 - x_3 \ge 3,$$
$$1 \le x_1 \le 3,\ 0 \le x_2 \le 3,\ 2 \le x_3.$$

b) Максимизировать $z = x_1 + 2x_2$
при ограничениях

$$-x_1 + 2x_2 \ge 0,$$
$$3x_1 + 2x_2 \le 10,$$
$$-x_1 + x_2 \le 1,$$
$$1 \le x_1 \le 3,\ 0 \le x_2 \le 1.$$

c) Максимизировать $z = 4x_1 + 2x_2 + 6x_3$
при ограничениях

$$4x_1 - x_2 \le 9,$$
$$-x_1 + x_2 + 2x_3 \le 8,$$
$$-3x_1 + x_2 + 4x_3 \le 12,$$
$$1 \le x_1 \le 3,\ 0 \le x_2 \le 5,\ 0 \le x_3 \le 2.$$

5. Рассмотрим матричное представление задачи с ограниченными переменными. Разобьем вектор $\mathbf{X}$ на две части $(\mathbf{X}_z, \mathbf{X}_u)$, где элементами вектора $\mathbf{X}_u$ являются переменные, к которым в процессе выполнения алгоритма будет применена подстановка, эквивалентная их верхнему пределу. Тогда задачу ЛП с ограниченными переменными можно записать следующим образом.

$$\begin{bmatrix} 1 & -\mathbf{C}_z & -\mathbf{C}_u \\ \mathbf{0} & \mathbf{D}_z & \mathbf{D}_u \end{bmatrix} \begin{bmatrix} z \\ \mathbf{X}_z \\ \mathbf{X}_u \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{b} \end{bmatrix}.$$

Пусть $\mathbf{B}$ (и $\mathbf{X}_B$) — базисное решение текущей симплексной итерации, полученное после подстановки $\mathbf{X}_u = \mathbf{U}_u - \mathbf{X}'_u$, где $\mathbf{U}_u$ — подмножество элементов вектора значений верхних границ $\mathbf{U}$, соответствующего переменным $\mathbf{X}_u$. Покажите, что в данном случае симплекс-таблица имеет следующий вид.

Базис	$\mathbf{X}_z^T$	$\mathbf{X}_u'^T$	Решение
z	$\mathbf{C}_B\mathbf{B}^{-1}\mathbf{D}_z - \mathbf{C}_z$	$-\mathbf{C}_B\mathbf{B}^{-1}\mathbf{D}_u + \mathbf{C}_u$	$\mathbf{C}_B\mathbf{B}^{-1}\mathbf{b}' + \mathbf{C}_u\mathbf{U}_u$
$\mathbf{X}_B$	$\mathbf{B}^{-1}\mathbf{D}_z$	$-\mathbf{B}^{-1}\mathbf{D}_u$	$\mathbf{B}^{-1}\mathbf{b}'$

Здесь $\mathbf{b}' = \mathbf{b} - \mathbf{D}_u\mathbf{U}_u$.

6. В задаче из примера 7.5–3 выполните следующее.

a) На этапе итерации 1, используя матричные представления, проверьте, что $\mathbf{X}_B = (x_4, x_1)^T = (5/2, 3/2)^T$.

b) На этапе итерации 2 покажите, как на основе исходных данных задачи можно вычислить $\mathbf{B}^{-1}$. Затем, используя матричные представления, проверьте вычисленные в примере значения переменных x_4 и x'_2.

7. Решите задачу из упр. 3(а) с помощью модифицированного (использующего матричные представления) симплекс-метода.

8. *Двойственный симплекс-метод решения задач с ограниченными переменными.* Двойственный симплекс-метод (раздел 4.5) также можно модифицировать для учета явных ограничений, накладываемых на переменные. Для этого после подстановок $x_j = u_j - x'_j$ (u_j — верхняя граница переменной x_j; если u_j бесконечна, заменяем ее достаточно большим положительным числом M) исходная задача записывается в двойственной форме. Далее выполняются следующие действия.

Шаг 1. Если значение какой-либо базисной переменной $(\mathbf{X}_B)_i$ превышает ее верхнюю границу, выполняем замену $(\mathbf{X}_B)_i = (\mathbf{U}_B)_i - (\mathbf{X}_B)'_i$. Затем переходим к шагу 2.

Шаг 2. Если базисное решение допустимо, вычисления заканчиваются. В противном случае среди базисных переменных определяем исключаемую из базиса переменную x_r как имеющую наибольшее отрицательное значение. Переходим к шагу 3.

Шаг 3. В соответствии с условием оптимальности двойственного симплекс-метода определяем вводимую в базис переменную. Переходим к шагу 4.

Шаг 4. Выполняем пересчет базиса. Переходим к шагу 1.

Примените описанный алгоритм к следующим задачам.

a) Минимизировать $z = -3x_1 - 2x_2 + 2x_3$
при ограничениях

$$2x_1 + x_2 + x_3 \le 8,$$
$$-x_1 + 2x_2 + x_3 \ge 13,$$
$$0 \le x_1 \le 2,\ 0 \le x_2 \le 3,\ 0 \le x_3 \le 1.$$

b) Максимизировать $z = x_1 + 5x_2 - 2x_3$
при ограничениях

$$4x_1 + 2x_2 + 2x_3 \le 18,$$
$$x_1 + 3x_2 + 4x_3 \ge 10,$$
$$0 \le x_1 \le 2,\ 0 \le x_2 \le 3,\ 0 \le x_3\ (x_3 \text{ сверху не ограничена}).$$

7.5.3. Метод декомпозиции

Рассмотрим разработку производственного плана предприятия, состоящего из нескольких подразделений. Хотя каждое подразделение имеет собственные производственные возможности и соответствующие ограничения, производственные планы отдельных подразделений обобщаются на уровне управления предприятием. Таким образом, результирующая модель рассматриваемого предприятия будет содержать ограничения двух типов: *общие*, соответствующие уровню всего предприятия, и *независимые*, отображающие производственные ограничения отдельных подразделений. На рис. 7.5 показана структура ограничений для n подразделений. Отсутствие общих ограничений означает, что все подразделения независимы.

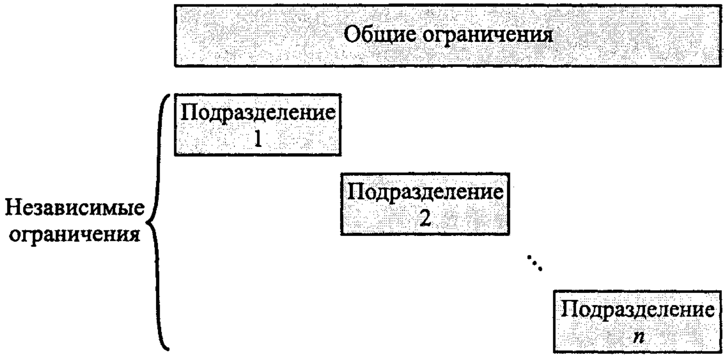


Рис. 7.5

Метод декомпозиции предлагает эффективный вычислительный алгоритм решения задач со структурой ограничений, подобной показанной на рис. 7.5, путем разбиения исходной задачи на n подзадач, которые решаются независимо друг от друга. Вместе с тем отметим, что применение этого метода особенно оправдано, когда вычисления выполняются на вычислительном устройстве с ограниченной скоростью выполнения операций и ограниченным объемом памяти. Сегодня, когда вычислительная техника имеет огромную (по сравнению с недавним прошлым) производительность, необходимость в методе декомпозиции не так очевидна. Несмотря на это мы рассмотрим данный метод, поскольку он представляет интерес в теоретическом плане.

Математическую модель, к которой применим метод декомпозиции, запишем в следующем виде.

$$\text{Максимизировать } z = \mathbf{C}_1\mathbf{X}_1 + \mathbf{C}_2\mathbf{X}_2 + \ldots + \mathbf{C}_n\mathbf{X}_n$$

при ограничениях

$$\begin{array}{lllll}
\mathbf{A}_1\mathbf{X}_1 + & \mathbf{A}_2\mathbf{X}_2 + & \ldots + & \mathbf{A}_n\mathbf{X}_n & = \mathbf{b}_0 \\
\mathbf{D}_1\mathbf{X}_1 & & & & = \mathbf{b}_1 \\
 & \mathbf{D}_2\mathbf{X}_2 & & & = \mathbf{b}_2 \\
 & & \ddots & & \ \vdots \\
 & & & \mathbf{D}_n\mathbf{X}_n & = \mathbf{b}_n
\end{array}$$

$$\mathbf{X}_j \geq \mathbf{0}, \quad j = 1, 2, \ldots, n.$$

При необходимости ограничения в виде неравенств с помощью дополнительных переменных преобразуются в равенства.

Метод декомпозиции основан на определении *крайних точек* множеств $\{\mathbf{X}_j \mid \mathbf{D}_j\mathbf{X}_j = \mathbf{b}_j, \mathbf{X}_j \geq \mathbf{0}\}$, $j = 1, 2, \ldots, n$. Для этого необходимо, чтобы каждое множество $\{\mathbf{X}_j \mid \mathbf{D}_j\mathbf{X}_j = \mathbf{b}_j, \mathbf{X}_j \geq \mathbf{0}\}$ было ограниченным. Это требование выполнимо всегда, поскольку, при необходимости, для любого множества j можно добавить искусственное ограничение $\mathbf{IX}_j \leq M$, где M — достаточно большое положительное число.

Обозначим через $\mathbf{Y}_{jk}$, $k = 1, 2, \ldots, K_j$, крайние точки множества $\{\mathbf{X}_j \mid \mathbf{D}_j\mathbf{X}_j = \mathbf{b}_j, \mathbf{X}_j \geq \mathbf{0}\}$. Тогда можно записать

$$\mathbf{X}_j = \sum_{k=1}^{K_j} \beta_{jk}\mathbf{Y}_{jk}, \quad j = 1, 2, \ldots, n,$$

где $\beta_{jk} \geq 0$ для всех k и j, причем $\sum_{k=1}^{K_j} \beta_{jk} = 1$.

Теперь переформулируем исходную задачу в терминах крайних точек и получим так называемую **главную задачу**, определяемую следующим образом.

$$\text{Максимизировать } z = \sum_{k=1}^{K_1} \mathbf{C}_1 \mathbf{Y}_{1k} \beta_{1k} + \sum_{k=1}^{K_2} \mathbf{C}_2 \mathbf{Y}_{2k} \beta_{2k} + \ldots + \sum_{k=1}^{K_n} \mathbf{C}_n \mathbf{Y}_{nk} \beta_{nk}$$

при ограничениях

$$\begin{aligned}
&\sum_{k=1}^{K_1} \mathbf{A}_1 \mathbf{Y}_{1k} \beta_{1k} + \sum_{k=1}^{K_2} \mathbf{A}_2 \mathbf{Y}_{2k} \beta_{2k} + \ldots + \sum_{k=1}^{K_n} \mathbf{A}_n \mathbf{Y}_{nk} \beta_{nk} = \mathbf{b}_0, \\
&\sum_{k=1}^{K_1} \beta_{1k} = 1, \\
&\sum_{k=1}^{K_2} \beta_{2k} = 1, \\
&\qquad \ddots \qquad \vdots \\
&\sum_{k=1}^{K_n} \beta_{nk} = 1, \\
&\beta_{jk} \geq 0 \text{ для всех } k \text{ и } j.
\end{aligned}$$

В главной задаче новыми переменными являются β_{jk}. После того как будет найдено оптимальное решение β^*_{jk} этой задачи, оптимальное решение исходной задачи вычисляется путем обратных подстановок по следующей формуле.

$$\mathbf{X}_j = \sum_{k=1}^{K_j} \beta^*_{jk} \mathbf{Y}_{jk}, \quad j = 1, 2, \ldots, n$$

Может показаться, что решение главной задачи требует предварительного нахождения *всех* крайних точек $\mathbf{Y}_{jk}$, что является очень трудной задачей. К счастью, это не так.

При решении главной задачи с использованием модифицированного симплекс-метода (раздел 7.5.1), на каждой итерации нам необходимо определить вводимую и исключаемую переменные. Начнем с определения вводимой переменной. Зная текущий базис главной задачи (и, следовательно, матрицы $\mathbf{C}_B$ и $\mathbf{B}^{-1}$), для любой небазисной переменной β_{jk} имеем

$$z_{jk} - c_{jk} = \mathbf{C}_B \mathbf{B}^{-1} \mathbf{P}_{jk} - c_{jk},$$

где

$$c_{jk} = \mathbf{C}_j \mathbf{Y}_{jk} \quad \text{и} \quad \mathbf{P}_{jk} = \begin{bmatrix} \mathbf{A}_j \mathbf{Y}_{jk} \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}.$$

Напомним: чтобы определить, какая из небазисных переменных β_{jk} должна войти в базис, мы должны найти

$$z_{j^*k^*} - c_{j^*k^*} = \min_{\text{по всем } j \text{ и } k} \{z_{jk} - c_{jk}\}.$$

Если $z_{j^*k^*} - c_{j^*k^*} > 0$, тогда в соответствии с условием оптимальности задачи максимизации переменная $\beta_{j^*k^*}$ должна войти в базис. При выполнении обратного неравенства считаем, что оптимальное решение достигнуто.

Теперь покажем, как можно вычислить разность $z_{j^*k^*} - c_{j^*k^*}$. "Секрет" вычислений заключается в равенстве

$$\min_{\text{по всем } j \text{ и } k} \{z_{jk} - c_{jk}\} = \min_{j} \left\{ \min_{k} \left\{ z_{jk} - c_{jk} \right\} \right\}.$$

Это равенство вытекает из того, что каждое выпуклое множество, определяемое ограничениями $\mathbf{D}_j\mathbf{X}_j = \mathbf{b}_j$, $\mathbf{X}_j \geq \mathbf{0}$, имеет собственное независимое множество крайних точек. В соответствии с этим равенством мы можем найти разность $z_{j^*k^*} - c_{j^*k^*}$ за два шага.

1. Для каждого выпуклого множества $\{\mathbf{X}_j \mid \mathbf{D}_j\mathbf{X}_j = \mathbf{b}_j,\ \mathbf{X}_j \geq \mathbf{0}\}$ определяем крайнюю точку $\mathbf{Y}_{jk^*}$, на которой достигается минимум разностей $z_{jk} - c_{jk}$, т.е. $z_{jk^*} - c_{jk^*} = \min_k\{z_{jk} - c_{jk}\}$.
2. Далее определяем $z_{j^*k^*} - c_{j^*k^*} = \min_j\{z_{jk^*} - c_{jk^*}\}$.

Из теории линейного программирования мы знаем, что конечное оптимальное решение ассоциируется с крайней точкой пространства решений. Поскольку каждое из множеств $\{\mathbf{X}_j \mid \mathbf{D}_j\mathbf{X}_j = \mathbf{b}_j,\ \mathbf{X}_j \geq \mathbf{0}\}$ ограничено по определению, действия, выполняемые на шаге 1, математически эквивалентны решению n задач линейного программирования вида

$$\text{Минимизировать } w_j = \{z_j - c_j \mid \mathbf{D}_j\mathbf{X}_j = \mathbf{b}_j,\ \mathbf{X}_j \geq \mathbf{0}\}.$$

Фактически здесь целевая функция w_j является линейной функцией от $\mathbf{X}_j$ (см. упр. 7.5,d(8)).

Таким образом, определение вводимой переменной $\beta_{j^*k^*}$ в главной задаче сведено к решению n задач линейного программирования (меньшего размера) для нахождения "вводимых" крайних точек $\mathbf{Y}_{j^*k^*}$. Такой подход не требует определения всех крайних точек всех n выпуклых множеств. После того как требуемая крайняя точка определена, нетрудно определить все элементы вектора $\mathbf{P}_{j^*k^*}$. На основе этой информации мы можем определить исключаемую переменную. Далее с помощью модифицированного симплекс-метода вычисляем следующую обратную матрицу $\mathbf{B}^{-1}$.

Пример 7.5–4

Решим с помощью метода декомпозиции следующую задачу ЛП.

$$\text{Максимизировать } z = 3x_1 + 5x_2 + x_4 + x_5$$

при ограничениях

$$\begin{aligned}
&x_1 + x_2 + x_4 + x_5 + x_8 = 40,\\
&5x_1 + x_2 + x_3 = 12,\\
&x_4 + x_5 - x_6 = 5,
\end{aligned}$$

$$x_4 + 5x_5 + x_7 = 50,$$
$$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 \geq 0.$$

Здесь x_3, x_6, x_7 и x_8 — дополнительные переменные.

Данную задачу можно представить в следующем виде.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
3	5	0	1	1	0	0	0	
1	1	0	1	1	0	0	1	40
5	1	1						12
			1	1	–1	0		5
			1	5	0	1		50

Данное представление показывает, что исходную задачу можно разбить на две подзадачи.

Подзадача 1 ($j = 1$)

$\mathbf{X}_1 = (x_1, x_2, x_3)^T$, $\mathbf{C}_1 = (3, 5, 0)$, $\mathbf{A}_1 = (1, 1, 0)$.

$\mathbf{D}_1 = (5, 1, 1)$, $b_1 = 12$.

Подзадача 2 ($j = 2$)

$\mathbf{X}_2 = (x_4, x_5, x_6, x_7)^T$, $\mathbf{C}_2 = (1, 1, 0, 0)$, $\mathbf{A}_2 = (1, 1, 0, 0)$.

$$\mathbf{D}_2 = \begin{bmatrix} 1 & 1 & -1 & 0 \\ 1 & 5 & 0 & 1 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 5 \\ 50 \end{bmatrix}.$$

Дополнительная переменная x_8 не входит ни в одну подзадачу, ее значение находится как часть решения главной задачи.

Если ввести искусственные переменные x_9 и x_{10}, начальную симплекс-таблицу символически можно записать следующим образом.

β_{11}	β_{12}	...	β_{1K_1}	β_{21}	β_{22}	...	β_{2K_2}	x_8	x_9	x_{10}	
$\mathbf{C}_1\mathbf{Y}_{11}$	$\mathbf{C}_1\mathbf{Y}_{12}$	...	$\mathbf{C}_1\mathbf{Y}_{1K_1}$	$\mathbf{C}_2\mathbf{Y}_{21}$	$\mathbf{C}_2\mathbf{Y}_{22}$	...	$\mathbf{C}_2\mathbf{Y}_{2K_2}$	0	$-M$	$-M$	
$\mathbf{A}_1\mathbf{Y}_{11}$	$\mathbf{A}_1\mathbf{Y}_{12}$	...	$\mathbf{A}_1\mathbf{Y}_{1K_1}$	$\mathbf{A}_2\mathbf{Y}_{21}$	$\mathbf{A}_2\mathbf{Y}_{22}$	...	$\mathbf{A}_2\mathbf{Y}_{2K_2}$	1	0	0	40
1	1	...	1	0	0	...	0	0	1	0	1
0	0	...	0	1	1	...	1	0	0	1	1
Подзадача 1				Подзадача 2				Начальное базисное решение			

Итерация 0

$$\mathbf{X}_B = (x_8, x_9, x_{10})^T = (40, 1, 1)^T,$$
$$\mathbf{C}_B = (0, -M, -M), \quad \mathbf{B} = \mathbf{B}^{-1} = \mathbf{I}.$$

Итерация 1

Подзадача 1 ($j = 1$). Имеем

$$z_1 - c_1 = \mathbf{C}_B\mathbf{B}^{-1}\begin{bmatrix}\mathbf{A}_1\mathbf{X}_1\\ 1\\ 0\end{bmatrix} - \mathbf{C}_1\mathbf{X}_1 =$$

$$= (0, -M, -M)\begin{bmatrix}(1,1,0)\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix}\\ 1\\ 0\end{bmatrix} - (3,5,0)\begin{bmatrix}x_1\\ x_2\\ x_3\end{bmatrix} =$$

$$= -3x_1 - 5x_2 - M.$$

Далее, ограничение $\mathbf{D}_1\mathbf{X}_1 = \mathbf{b}_1$ сводится к равенству $5x_1 + x_2 + x_3 = 12$. Таким образом, соответствующая задача ЛП имеет следующий вид.

$$\text{Минимизировать } w_1 = -3x_1 - 5x_2 - M$$

при ограничениях

$$5x_1 + x_2 + x_3 = 12,$$
$$x_1, x_2, x_3 \geq 0.$$

Решив эту задачу (обычным симплекс-методом), получаем

$$\mathbf{Y}_{11} = (0,\ 12,\ 0)^T,\ z_1^* - c_1^* = w_1^* = -60 - M.$$

Подзадача 2 (j = 2). Соответствующая задача ЛП имеет следующий вид.

$$\text{Минимизировать } z_2 - c_2 = \mathbf{C}_B\mathbf{B}^{-1}\begin{bmatrix}\mathbf{A}_2\mathbf{X}_2\\ 0\\ 1\end{bmatrix} - \mathbf{C}_2\mathbf{X}_2 =$$

$$= (0, -M, -M)\begin{bmatrix}(1,1,0,0)\begin{bmatrix}x_4\\ x_5\\ x_6\\ x_7\end{bmatrix}\\ 0\\ 1\end{bmatrix} - (1,1,0,0)\begin{bmatrix}x_4\\ x_5\\ x_6\\ x_7\end{bmatrix} =$$

$$= -x_4 - x_5 - M.$$

при ограничениях

$$\begin{bmatrix}1 & 1 & -1 & 0\\ 1 & 5 & 0 & 1\end{bmatrix}\begin{bmatrix}x_4\\ x_5\\ x_6\\ x_7\end{bmatrix} = \begin{bmatrix}5\\ 50\end{bmatrix},$$
$$x_4, x_5, x_6, x_7 \geq 0.$$

Находим оптимальное решение этой задачи.

$$\mathbf{Y}_{21} = (50, 0, 45, 0)^T, \; z_2^* - c_2^* = -50 - M.$$

Поскольку главная задача — задача максимизации и $z_1^* - c_1^* < z_2^* - c_2^*$, а также $z_1^* - c_1^* < 0$, отсюда следует, что переменная β_{11}, соответствующая крайней точке $\mathbf{Y}_{11}$, должна быть введена в базисное решение.

Для определения исключаемой переменной запишем

$$\mathbf{P}_{11} = \begin{bmatrix} \mathbf{A}_1\mathbf{Y}_{11} \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} (1,1,0)\begin{bmatrix} 0 \\ 12 \\ 0 \end{bmatrix} \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 12 \\ 1 \\ 0 \end{bmatrix}.$$

Отсюда $\mathbf{B}^{-1}\mathbf{P}_{11} = (12, 1, 0)^T$. Имея $\mathbf{X}_B = (x_8, x_9, x_{10})^T = (40, 1, 1)^T$, делаем вывод, что из базисного решения следует исключить переменную x_9 (искусственную переменную).

Теперь, используя мультипликативную форму представления обратной матрицы, можно найти $\mathbf{B}^{-1}_{\text{след}}$ путем удаления из базиса вектора, соответствующего переменной x_9, и введения в базис вектора $\mathbf{P}_{11}$. Получим (проверьте!)

$$\mathbf{B}^{-1} = \begin{bmatrix} 1 & -12 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

и новое базисное решение

$$\mathbf{X}_B = (x_8, \beta_{11}, x_{10})^T = \mathbf{B}^{-1}(40, 1, 1)^T = (28, 1, 1)^T,$$
$$\mathbf{C}_B = (0, \mathbf{C}_1\mathbf{Y}_{11}, -M) = (0, 60, -M).$$

Итерация 2

Подзадача 1 ($j = 1$). Вы должны проверить, что соответствующая задача ЛП останется такой же, как и на первой итерации (это простое совпадение, а не общее правило). Ее оптимальное решение дает $z_1^* - c_1^* = w_1 = 0$. Отсюда следует, что никакая из оставшихся крайних точек пространства решений подзадачи 1 не может улучшить решение главной задачи. (Оптимальным решением этой подзадачи будет крайняя точка $\mathbf{Y}_{11}$, которой соответствует переменная β_{11}, уже входящая в состав базиса. Именно поэтому $z_1^* - c_1^* = 0$.)

Подзадача 2 ($j = 2$). Снова вы должны проверить, что соответствующая задача ЛП такая же (опять совпадение), как и на первой итерации. Ее оптимальное решение

$$\mathbf{Y}_{22} = (50, 0, 45, 0)^T, \; z_2^* - c_2^* = -50 - M.$$

Отметим, что точка $\mathbf{Y}_{22}$ фактически совпадает с крайней точкой $\mathbf{Y}_{21}$, но мы используем нижний индекс 2, чтобы показать, что эта точка соответствует второй итерации.

Из решений обеих подзадач следует, что $z_2^* - c_2^* < 0$. Это указывает на то, что переменная β_{22}, соответствующая крайней точке $\mathbf{Y}_{22}$, должна войти в базисное решение.

Для определения исключаемой переменной запишем

$$\mathbf{P}_{22} = \begin{bmatrix} \mathbf{A}_2\mathbf{Y}_{22} \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (1,1,0,0)\begin{bmatrix} 50 \\ 0 \\ 45 \\ 0 \end{bmatrix} \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 50 \\ 0 \\ 1 \end{bmatrix}.$$

Отсюда $\mathbf{B}^{-1}\mathbf{P}_{22} = (50, 0, 1)^T$. Поскольку $\mathbf{X}_B = (x_8, \beta_{11}, x_{10})^T = (28, 1, 1)^T$, из базисного решения следует исключить переменную x_8.

Находим новую обратную матрицу $\mathbf{B}^{-1}$ (проверьте!)

$$\mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{50} & -\frac{12}{50} & 0 \\ 0 & 1 & 0 \\ -\frac{1}{50} & \frac{12}{50} & 1 \end{bmatrix}$$

Новое базисное решение равно

$$\mathbf{X}_B = (\beta_{22}, \beta_{11}, x_{10})^T = \mathbf{B}^{-1}(40, 1, 1)^T = (14/25, 1, 11/25)^T,$$
$$\mathbf{C}_B = (\mathbf{C}_2\mathbf{Y}_{22}, \mathbf{C}_1\mathbf{Y}_{11}, -M) = (50, 60, -M).$$

Итерация 3

Подзадача 1 ($j = 1$). Вы должны проверить, что на этой итерации целевая функция для данной подзадачи имеет следующий вид.

$$\text{Минимизировать } w_1 = \left(\frac{M}{50} - 2\right)x_1 + \left(\frac{M}{50} - 4\right)x_2 - \frac{12M}{50} + 48.$$

Соответствующее оптимальное решение дает $z_1^* - c_1^* = 0$; это указывает, что подзадача 1 не имеет улучшающей крайней точки.

Подзадача 2 ($j = 2$). Здесь целевая функция имеет следующий вид (проверьте!).

$$\text{Минимизировать } w_2 = \frac{M}{50}(x_3 + x_4) - M.$$

Находим оптимальное решение:

$$\mathbf{Y}_{23} = (5, 0, 0, 45)T,\ z_2^* - c_2^* = -\frac{9M}{10}.$$

Небазисная переменная x_8. Исходя из вида главной задачи, разность $z_j - c_j$ для переменной x_8 необходимо вычислять отдельно.

$$z_8 - c_8 = \mathbf{C}_B\mathbf{B}^{-1}\mathbf{P}_8 - c_8 =$$
$$= \left(1 + \frac{M}{50}, 48 - \frac{12M}{50}, -M\right)(1, 0, 0)^T - 0 = 1 + \frac{M}{50}.$$

Отсюда вытекает, что переменная x_8 не может улучшить решение.

Из приведенных вычислений следует, что вводимой в базис будет переменная β_{23}, соответствующая крайней точке $\mathbf{Y}_{23}$. Для определения исключаемой переменной вычисляем следующее.

$$\mathbf{P}_{23} = \begin{bmatrix} \mathbf{A}_2\mathbf{Y}_{23} \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} (1,1,0,0)\begin{bmatrix} 5 \\ 0 \\ 0 \\ 45 \end{bmatrix} \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \\ 1 \end{bmatrix}.$$

Отсюда $\mathbf{B}^{-1}\mathbf{P}_{23} = (1/10, 0, 9/10)^T$. Поскольку $\mathbf{X}_B = (\beta_{22}, \beta_{11}, x_{10})^T = (14/25, 1, 11/25)^T$, из базисного решения следует исключить искусственную переменную x_{10}.

Находим новую обратную матрицу $\mathbf{B}^{-1}$ (проверьте!).

$$\mathbf{B}^{-1} = \begin{bmatrix} \frac{1}{45} & -\frac{12}{45} & -\frac{5}{45} \\ 0 & 1 & 0 \\ -\frac{1}{45} & \frac{12}{45} & \frac{50}{45} \end{bmatrix}$$

Новое базисное решение равно

$$\mathbf{X}_B = (\beta_{22}, \beta_{11}, \beta_{23})^T = \mathbf{B}^{-1}(40, 1, 1)^T = (23/45, 1, 22/45)^T,$$
$$\mathbf{C}_B = (\mathbf{C}_2\mathbf{Y}_{22}, \mathbf{C}_1\mathbf{Y}_{11}, \mathbf{C}_2\mathbf{Y}_{23}) = (50, 60, 5).$$

Итерация 4

Подзадача 1 ($j = 1$). $w_1 = -2x_1 - 4x_2 + 48$. Получаем $z_1^* - c_1^* = w_1 = 0$.

Подзадача 2 ($j = 2$). $w_2 = 0x_4 + 0x_5 + 48 = 48$.

Небазисная переменная x_8. $z_8 - c_8 = 1$. Отсюда следует, что решение, полученное на третьей итерации, оптимально.

С помощью обратной подстановки вычисляем оптимальное решение исходной задачи.

$$\mathbf{X}_1^* = (x_1, x_2, x_3)^T = \beta_{11}\mathbf{Y}_{11} = 1(0, 12, 0)^T = (0, 12, 0)^T,$$

$$\begin{aligned}\mathbf{X}_2^* &= (x_4, x_5, x_6, x_7)^T = \beta_{22}\mathbf{Y}_{22} + \beta_{23}\mathbf{Y}_{23} = \\ &= \frac{23}{45}(50, 0, 45, 0)^T + \frac{22}{45}(5, 0, 0, 45)^T = \\ &= (28, 0, 23, 22)^T.\end{aligned}$$

Все остальные переменные равны нулю. Оптимальное значение целевой функции получим путем прямой подстановки в ее выражение значений соответствующих переменных.

Упражнения 7.5,d

1. В каждой из следующих систем ограничений графически найдите допустимые крайние точки и определите пространство допустимых решений как функцию этих крайних точек. Если пространство решений не ограничено, добавьте необходимое искусственное ограничение.

a)

$$\begin{aligned}
&x_1 + 2x_2 \le 6,\\
&2x_1 + x_2 \le 8,\\
&-x_1 + x_2 \le 1,\\
&x_2 \le 2,\\
&x_1, x_2 \ge 0.
\end{aligned}$$

b)

$$\begin{aligned}
&2x_1 + x_2 \le 2,\\
&3x_1 + 4x_2 \ge 12,\\
&x_1, x_2 \ge 0.
\end{aligned}$$

c)

$$\begin{aligned}
&x_1 - x_2 \le 10,\\
&2x_1 \le 40,\\
&x_1, x_2 \ge 0.
\end{aligned}$$

2. В примере 7.5–4 крайние точки подпространств $\{\mathbf{X}_1 \mid \mathbf{D}_1\mathbf{X}_1 = \mathbf{b}_1,\ \mathbf{X}_1 \ge \mathbf{0}\}$ и $\{\mathbf{X}_2 \mid \mathbf{D}_2\mathbf{X}_2 = \mathbf{b}_2,\ \mathbf{X}_2 \ge \mathbf{0}\}$ можно найти графически. Используйте эту информацию, чтобы сформулировать в явном виде главную задачу. Затем покажите, что применение к главной задаче модифицированного симплекс-метода порождает такую же последовательность вводимых в базис переменных β_{jk}, как и последовательное решение отдельных подзадач 1 и 2.

3. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = x_1 + 3x_2 + 5x_3 + 2x_4$$

при ограничениях

$$\begin{aligned}
&5x_1 + x_2 \le 9,\\
&x_1 + 4x_2 \le 8,\\
&5x_1 + 3x_2 + 4x_3 \ge 10,\\
&x_3 - 5x_4 \le 4,\\
&x_3 + x_4 \le 10,\\
&x_1, x_2, x_3, x_4 \ge 0.
\end{aligned}$$

Постройте в явном виде главную задачу (для этого используйте крайние точки подпространств, найденные графически) и затем решите ее модифицированным симплекс-методом.

4. Решите задачу из предыдущего упражнения методом декомпозиции и сравните процесс решения при использовании разных алгоритмов.
5. Примените метод декомпозиции к следующей задаче.

$$\text{Максимизировать } z = 6x_1 + 7x_2 + 3x_3 + 5x_4 + x_5 + x_6$$

при ограничениях

$$\begin{aligned}
&x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \le 50,\\
&x_1 + x_2 \le 10,\\
&x_2 \le 8,\\
&5x_3 + x_4 \le 12,\\
&x_5 + x_6 \ge 5,\\
&x_5 + 5x_6 \le 50,\\
&x_1, x_2, x_3, x_4, x_5, x_6 \ge 0.
\end{aligned}$$

6. Продумайте необходимые изменения в применении метода декомпозиции к задаче минимизации, затем решите следующую задачу.

$$\text{Минимизировать } z = 5x_1 + 3x_2 + 8x_3 - 5x_4$$

при ограничениях

$$\begin{aligned}
&x_1 + x_2 + x_3 + x_4 \ge 25,\\
&5x_1 + x_2 \le 20,\\
&5x_1 - x_2 \ge 5,\\
&x_3 + x_4 = 20,\\
&x_1, x_2, x_3, x_4 \ge 0.
\end{aligned}$$

7. Решите следующую задачу методом декомпозиции.

$$\text{Минимизировать } z = 10y_1 + 2y_2 + 4y_3 + 8y_4 + y_5$$

при ограничениях

$$\begin{aligned}
&y_1 + 4y_2 - y_3 \ge 8,\\
&2y_1 + y_2 + y_3 \ge 2,\\
&3y_1 + y_4 + y_5 \ge 4,\\
&y_1 + 2y_4 - y_5 \ge 10,\\
&y_1, y_2, y_3, y_4, y_5 \ge 0.
\end{aligned}$$

(*Совет.* Рассмотрите сначала задачу, двойственную к данной.)

8. Пусть при применении метода декомпозиции количество общих ограничений в исходной задаче равно r. Покажите, что целевую функцию для подзадачи j можно записать следующим образом.

$$\text{Минимизировать } w_j = z_j - c_j = (\mathbf{C}_B\mathbf{R}\mathbf{A}_j - \mathbf{C}_j)\mathbf{X}_j + \mathbf{C}_B\mathbf{V}_{r+j}$$

Здесь матрица $\mathbf{R}$ состоит из первых r столбцов матрицы $\mathbf{B}^{-1}$, а $\mathbf{V}_{r+j}$ — ее $(r + j)$-й столбец.

7.6. Двойственность

Двойственные задачи и соотношения двойственности на элементарном уровне мы изучали в главе 4. В этом разделе мы рассмотрим теорию двойственности на более строгом (доказательном) уровне, в частности, представим доказательства соотношений двойственности, которые являются основой анализа чувствительности (см. главу 4). Эти теоретические построения послужат фундаментом для описанных далее методов параметрического программирования.

7.6.1. Матричное представление двойственной задачи

Пусть прямая задача ЛП с m ограничениями и n переменными уже записана в стандартной форме.

$$\text{Максимизировать } z = (\mathbf{C}_{\text{I}}, \mathbf{C}_{\text{II}})\mathbf{X}$$

при ограничениях

$$(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{b},$$
$$\mathbf{X} \geq \mathbf{0}.$$

Вектор $\mathbf{X}$ можно представить как $(\mathbf{X}_{\text{I}}, \mathbf{X}_{\text{II}})$, где $\mathbf{X}_{\text{II}}$ — начальное базисное решение, состоящее из m элементов.

Обозначим через $\mathbf{Y} = (y_1, y_2, \ldots, y_m)$ вектор переменных двойственной задачи. В соответствии с правилами, представленными в табл. 4.2, получаем следующую двойственную задачу.

$$\text{Минимизировать } w = \mathbf{Y}\mathbf{b}$$

при ограничениях

$$\mathbf{Y}\mathbf{A} \geq \mathbf{C}_{\text{I}},$$
$$\mathbf{Y} \geq \mathbf{C}_{\text{II}},$$

$\mathbf{Y}$ — вектор свободных переменных (т.е. не имеющих ограничений в знаке).

Отметим, ограничения $\mathbf{Y} \geq \mathbf{C}_{\text{II}}$ могут наложить дополнительные условия на знаки переменных, что является более сильным ограничением, чем последнее условие неограниченности в знаке.

Упражнения 7.6,а

1. Докажите, что задача, двойственная к двойственной задаче, совпадает с исходной (прямой) задачей.
2. Пусть прямая задача имеет вид: минимизировать $z = \{\mathbf{CX} \mid \mathbf{AX} \geq \mathbf{b}, \mathbf{X} \geq \mathbf{0}\}$. Сформулируйте соответствующую двойственную задачу.

7.6.2. Оптимальное решение двойственной задачи

В этом разделе устанавливаются соотношения между прямой и двойственной задачами и показывается, как определить оптимальное решение двойственной задачи на основе оптимального решения прямой задачи. Обозначим через **B** *оптимальный* базис прямой задачи, через $\mathbf{C}_B$ — вектор коэффициентов целевой функции, соответствующих оптимальным базисным переменным $\mathbf{X}_B$.

Теорема 7.6–1. Первая теорема двойственности. *Для любой пары допустимых решений* ($\mathbf{X}$, $\mathbf{Y}$) *прямой и двойственной задач значение целевой функции в задаче* минимизации *ограничивает сверху значение целевой функции в задаче* максимизации. *Для пары оптимальных решений* ($\mathbf{X}^*$, $\mathbf{Y}^*$) *значения целевых функций совпадают.*

Доказательство. Пара допустимых решений ($\mathbf{X}$, $\mathbf{Y}$) удовлетворяет всем ограничениям обеих задач. Умножая слева обе части ограничений задачи максимизации на вектор (свободных) переменных $\mathbf{Y}$, получим

$$\mathbf{Y}(\mathbf{A}, \mathbf{I})\mathbf{X} = \mathbf{Y}\mathbf{b} = w.$$

Аналогично, умножая справа ограничения задачи минимизации на вектор (неотрицательных) переменных $\mathbf{X}$, будем иметь

$$\mathbf{Y}(\mathbf{A}, \mathbf{I})\mathbf{X} \geq (\mathbf{C}_{\mathrm{I}}, \mathbf{C}_{\mathrm{II}})\mathbf{X} = z.$$

(Неотрицательность вектора $\mathbf{X}$ здесь существенна, так как определяет "направленность" неравенства.) Комбинируя последние два выражения, получаем неравенство $z \leq w$, доказанное для пары любых *допустимых* решений ($\mathbf{X}$, $\mathbf{Y}$).

Отметим, что теорема не указывает, какая задача прямая, а какая — двойственная. Это очень важно при нахождении оптимальных решений обеих задач.

Из доказанной части теоремы ($z \leq w$ для пары любых допустимых решений) следует, что максимум функции z и минимум функции w будут достигнуты только тогда, когда обе эти функции примут одинаковое значение. Исходя из этого можно построить оценку близости полученных допустимых решений к оптимальным путем сравнения разности $z - w$ и величины $(z + w)/2$. Чем меньше отношение $2(z - w)/(z + w)$, тем ближе данные решения к оптимальным. Но, конечно, из этого эмпирического правила не следует, что оптимальные значения целевых функций равны $(z + w)/2$.

Если в одной из задач целевая функция примет бесконечное значение, что можно сказать о решении другой? Ответ очевиден: другая задача не должна иметь допустимых решений. Если это не так (т.е. обе задачи имеют допустимые решения), неравенство $z \leq w$ обязательно должно выполняться, что невозможно, например, при $z = +\infty$ или $w = -\infty$.

Рассмотрим обратную ситуацию. Если одна из задач не имеет допустимых решений, то обязательно ли решение другой задачи будет неограниченным? Не обязательно! В следующем примере и прямая, и двойственная задачи не имеют допустимых решений.

Прямая задача. Максимизировать $z = \{x_1 + x_2 \mid x_1 - x_2 \leq -1, -x_1 + x_2 \leq -1, x_1, x_2 \geq 0\}$.

Двойственная задача. Минимизировать $w = \{-y_1 - y_2 \mid y_1 - y_2 \leq -1, -y_1 + y_2 \leq -1, y_1, y_2 \geq 0\}$.

Теорема 7.6–2. *Оптимальное решение двойственной задачи равно*

$$\mathbf{Y} = \mathbf{C}_B\mathbf{B}^{-1},$$

где $\mathbf{B}$ — оптимальный *базис прямой задачи, которому соответствует вектор* $\mathbf{C}_B$ *коэффициентов целевой функции.*

Доказательство. Для доказательства этой теоремы надо показать, что допустимое решение двойственной задачи можно вычислить по формуле $\mathbf{Y} = \mathbf{C}_B\mathbf{B}^{-1}$ и что в соответствии с теоремой 7.6-1 $z = w$ для оптимальных решений.

Решение прямой задачи будет оптимальным, если выполняются неравенства $z_j - c_j \geq 0$ для всех j (см. раздел 7.4), т.е.

$$\mathbf{C}_B\mathbf{B}^{-1}\mathbf{A} - \mathbf{C}_\mathrm{I} \geq 0 \text{ и } \mathbf{C}_B\mathbf{B}^{-1} - \mathbf{C}_\mathrm{II} \geq 0.$$

Но решение $\mathbf{Y}$ допустимо, если оно удовлетворяет ограничениям $\mathbf{YA} \geq \mathbf{C}_\mathrm{I}$ и $\mathbf{Y} \geq \mathbf{C}_\mathrm{II}$. Сравнение последних неравенств доказывает равенство $\mathbf{Y} = \mathbf{C}_B\mathbf{B}^{-1}$.

Равенство оптимальных значений z и w доказывается непосредственным сравнением выражений

$$w = \mathbf{Yb} = \mathbf{C}_B\mathbf{B}^{-1}\mathbf{b},$$
$$z = \mathbf{C}_B\mathbf{X}_B = \mathbf{C}_B\mathbf{B}^{-1}\mathbf{b}.$$

Переменные двойственной задачи $\mathbf{Y} = \mathbf{C}_B\mathbf{B}^{-1}$ иногда называют *симплексными мультипликаторами*. Они также известны как *двойственные цены* — название, идущее от экономической интерпретации этих переменных (см. раздел 4.4.1).

Пример 7.6–1

В следующей задаче ЛП *оптимальный* базис равен $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_4)$. Сформулируем двойственную задачу и найдем ее оптимальное решение, используя оптимальный базис прямой задачи.

Максимизировать $z = 3x_1 + 5x_2$

при ограничениях

$$x_1 + 2x_2 + x_3 = 5,$$
$$-x_1 + 3x_2 + x_4 = 2,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

Двойственная задача имеет следующий вид.

Минимизировать $w = 5y_1 + 2y_2$

при ограничениях

$$y_1 - y_2 \geq 3,$$
$$2y_1 + 3y_2 \geq 5,$$
$$y_1, y_2 \geq 0.$$

Имеем $\mathbf{C}_B = (3, 0)$. Оптимальный базис и его обратная матрица равны следующему.

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \text{ и } \mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

Переменные прямой и двойственной задач имеют следующие значения.

$$(x_1, x_4)^T = \mathbf{B}^{-1}\mathbf{b} = (5, 7)^T,$$
$$(y_1, y_2) = \mathbf{C}_B\mathbf{B}^{-1} = (3, 0).$$

Оба решения допустимы и $z = w = 15$. Таким образом, решения оптимальны.

Обозначим через $\mathbf{P}_j$ j-й столбец матрицы $(\mathbf{A}, \mathbf{I})$. Из теоремы 7.6–2 следует, что величины

$$z_j - c_j = \mathbf{C}_B\mathbf{B}^{-1}\mathbf{P}_j - c_j = \mathbf{YP}_j - c_j$$

равны разностям между левыми и правыми частями ограничений двойственной задачи. В начале решения прямой задачи максимизации, по крайней мере, для одного j выполняется неравенство $z_j - c_j < 0$, в этом случае соответствующее решение двойственной задачи $\mathbf{Y} = \mathbf{C}_B\mathbf{B}^{-1}$ допустимо. Таким образом, пока решение прямой задачи сохраняет оптимальность, решение двойственной задачи автоматически сохраняет допустимость. На этом соотношении построен *двойственный симплекс-метод* (раздел 4.5), в котором вычисления начинаются с недопустимого решения, "лучшего", чем оптимальное. Выполнение метода продолжается до тех пор, пока не будет достигнуто допустимое решение. В противоположность этому, в "обычном" симплекс-методе (глава 3) вычисления начинаются с допустимого, но "худшего", чем оптимальное, решения и продолжаются до тех пор, пока не будет достигнуто оптимальное решение.

Упражнения 7.6,b

1. Проверьте правильность формулировки двойственной задачи в примере 7.6–1. Проверьте графически, что прямая и двойственная задачи не имеют недопустимых решений.
2. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = 50x_1 + 30x_2 + 10x_3$$

при ограничениях

$$\begin{aligned} &2x_1 + x_2 = 1, \\ &2x_2 = -5, \\ &4x_1 + x_3 = 6, \\ &x_1, x_2, x_3 \geq 0. \end{aligned}$$

   a) Сформулируйте и запишите двойственную задачу.
   b) Покажите с помощью непосредственной проверки, что прямая задача не имеет допустимого решения.
   c) Покажите, что двойственная задача не имеет ограниченного решения.
   d) На основе примеров задач из предыдущих упражнений найдите соотношение между свойствами недопустимости и неограниченности прямой и двойственной задач ЛП.

3. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = 5x_1 + 12x_2 + 4x_3$$

при ограничениях

$$\begin{aligned} &2x_1 - x_2 + 3x_3 = 2, \\ &x_1 + 2x_2 + x_3 + x_4 = 5, \\ &x_1, x_2, x_3, x_4 \geq 0. \end{aligned}$$

   a) Сформулируйте и запишите двойственную задачу.
   b) В каждом из следующих случаев сначала проверьте, что приведенный базис $\mathbf{B}$ является допустимым для прямой задачи. Затем, используя формулу $\mathbf{Y} = \mathbf{C}_B\mathbf{B}^{-1}$,

вычислите значения переменных двойственной задачи. Кроме того, определите, является ли данное решение прямой задачи оптимальным.

i) $\mathbf{B} = (\mathbf{P}_4, \mathbf{P}_3)$. iii) $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_2)$.
ii) $\mathbf{B} = (\mathbf{P}_2, \mathbf{P}_3)$. iv) $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_4)$.

4. Дана следующая задача линейного программирования.

$$\text{Максимизировать } z = 2x_1 + 4x_2 + 4x_3 - 3x_4$$

при ограничениях

$$x_1 + x_2 + x_3 = 4,$$
$$x_1 + 4x_2 + x_4 = 8,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

a) Сформулируйте и запишите двойственную задачу.
b) Путем вычисления разностей $z_j - c_j$ для всех небазисных $\mathbf{P}_j$ проверьте, что базис $\mathbf{B} = (\mathbf{P}_2, \mathbf{P}_3)$ соответствует оптимальному решению.
c) Найдите оптимальное решение двойственной задачи.

5. Модель ЛП содержит две переменные x_1 и x_2 и три ограничения типа "$\leq$". Соответствующие дополнительные (остаточные) переменные обозначены как x_3, x_4 и x_5. Предположим, что $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3)$ — оптимальный базис, а соответствующая обратная матрица имеет следующий вид.

$$\mathbf{B}^{-1} = \begin{bmatrix} 0 & -1 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{bmatrix}.$$

Ниже представлены оптимальные решения прямой и двойственной задач.

$$\mathbf{X}_B = (x_1, x_2, x_3)^T = (2, 6, 2)^T,$$
$$\mathbf{Y} = (y_1, y_2, y_3) = (0, 3, 2).$$

Найдите оптимальное значение целевой функции.

6. Докажите следующее соотношение между оптимальными решениями прямой и двойственной задач:

$$\sum_{i=1}^{m} c_i \left(\mathbf{B}^{-1}\mathbf{P}_k\right)_i = \sum_{i=1}^{m} y_i a_{ik},$$

где $\mathbf{C}_B = (c_1, c_2, ..., c_m)$ и $\mathbf{P}_k = (a_{1k}, a_{2k}, ..., a_{mk})^T$ для $k = 1, 2, ..., n$, $(\mathbf{B}^{-1}\mathbf{P}_k)_i$ — i-й элемент вектора $\mathbf{B}^{-1}\mathbf{P}_k$.

7. Запишите задачу, двойственную следующей.

$$\text{Максимизировать } z = \{\mathbf{CX} \mid \mathbf{AX} = \mathbf{b}, \mathbf{X} \text{ — свободные переменные}\}$$

8. Покажите, что задача, двойственная к задаче

$$\text{Максимизировать } z = \{\mathbf{CX} \mid \mathbf{AX} \le \mathbf{b},\ \mathbf{0} < \mathbf{L} \le \mathbf{X} \le \mathbf{U} < \infty\},$$

всегда имеет допустимое решение.

7.7. Параметрическое линейное программирование

Параметрическое линейное программирование — это расширение техники анализа чувствительности, которую мы рассмотрели в разделе 4.7. Здесь исследуются изменения в оптимальном решении задачи ЛП, являющиеся результатом *предопределенных* непрерывных изменений коэффициентов целевой функции и значений правых частей ограничений.

Может сложиться впечатление, что в параметрическом линейном программировании применяются только линейные функции, с помощью которых задаются изменения коэффициентов модели ЛП. Однако это не так, здесь могут использоваться любые функции, поддающиеся количественному определению, как линейные, так и нелинейные. Недостаток нелинейных функций заключается лишь в том, что в этом случае вычисления становятся более трудоемкие. Поэтому мы сконцентрируем свое внимание в основном на линейных функциях. (В упр. 7.7,а(5) рассмотрен случай нелинейной функции.)

Пусть задача ЛП определена следующим образом.

$$\text{Максимизировать } z = \left\{\mathbf{CX} \mid \sum_{j=1}^{n} \mathbf{P}_j x_j = \mathbf{b},\ \ \mathbf{X} \ge \mathbf{0}\right\}$$

В параметрическом программировании задаются изменения коэффициентов целевой функции и правых частей ограничений. Для этого используются линейные функции типа

$$\mathbf{C}(t) = \mathbf{C} + t\mathbf{C}',$$
$$\mathbf{b}(t) = \mathbf{b} + t\mathbf{b}',$$

где t — параметр изменения, $\mathbf{C}'$ и $\mathbf{b}'$ — заданные векторы. Для определенности полагаем, что $t \ge 0$.

Основная идея параметрического анализа заключается в следующем. Вначале находится оптимальное решение задачи ЛП при $t = 0$. Затем на основании условий оптимальности и допустимости симплекс-метода определяется интервал $0 \le t \le t_1$ значений параметра t, для которых решение, полученное при $t = 0$, остается оптимальным и допустимым. Значение t_1 называется *критическим*. Затем определяются следующие критические значения параметра t и соответствующие им оптимальные допустимые решения. Процесс заканчивается, когда будет найдено такое значение t_r, что при любых значениях $t > t_r$ последнее решение остается неизменным либо решения не существует.

7.7.1. Параметрическое изменение коэффициентов целевой функции

Пусть $\mathbf{X}_{B_i}$, $\mathbf{B}_i$ и $\mathbf{C}_{B_i}(t)$ — элементы оптимального решения, соответствующего критическому значению t_i (вычисления начались при $t_0 = 0$ с оптимальным базисом $\mathbf{B}_0$). Далее определяем следующее критическое значение t_{i+1} и соответствующий ему оптимальный базис, если он существует. Поскольку изменения в векторе коэффициентов целевой функции влияют только на оптимальность решения, текущее решение $\mathbf{X}_{B_i} = \mathbf{B}_i^{-1}\mathbf{b}$ останется оптимальным при $t \geq t_i$ до тех пор, пока будет выполняться условие оптимальности

$$z_j(t) - c_j(t) = \mathbf{C}_{B_i}(t)\mathbf{B}_i^{-1}\mathbf{P}_j - c_j(t) \geq 0 \quad \text{для всех } j.$$

Очередное критическое значение t_{i+1} определяется как наибольшее t, $t > t_i$, при котором выполняются все условия оптимальности.

Отметим, что приведенное неравенство не требует, чтобы вектор $\mathbf{C}(t)$ был линейной функцией от t; допустима функциональная зависимость любого вида, как линейная, так и нелинейная. Трудность использования нелинейных функций заключается лишь в том, что численное решение системы неравенств может быть очень трудоемким.

Пример 7.7–1

Рассмотрим задачу ЛП.

$$\text{Максимизировать } z = (3 - 6t)\,x_1 + (2 - 2t)\,x_2 + (5 + 5t)\,x_3$$

при ограничениях

$$\begin{aligned}
&x_1 + 2x_2 + x_3 \leq 40,\\
&3x_1 + 2x_3 \leq 60,\\
&x_1 + 4x_2 \leq 30,\\
&x_1, x_2, x_3 \geq 0.
\end{aligned}$$

Здесь $\mathbf{C}(t) = (3 - 6t, 2 - 2t, 5 + 5t)$, $t \geq 0$. Введем дополнительные переменные x_4, x_5 и x_6.

Оптимальное решение при $t_0 = 0$.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	4	0	0	1	2	0	160
x_2	–1/4	1	0	1/2	–1/4	0	5
x_3	3/2	0	1	0	1/2	0	30
x_6	2	0	0	–2	1	1	10

$$\mathbf{X}_{B_0} = (x_2, x_3, x_6)^T = (5, 30, 10)^T,$$
$$\mathbf{C}_{B_0}(t) = (2 - 2t, 5 + 5t, 0),$$

$$\mathbf{B}_0^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix}.$$

Условия оптимальности для текущих небазисных векторов $\mathbf{P}_1$, $\mathbf{P}_4$ и $\mathbf{P}_5$ имеют вид

$$\left\{\mathbf{C}_{B_0}(t)\mathbf{B}_0^{-1}\mathbf{P}_j - c_j(t)\right\}_{j=1,4,5} = (4 + 14t, 1 - t, 2 + 3t) \geq 0.$$

Таким образом, решение $\mathbf{X}_{B_0}$ остается оптимальным до тех пор, пока выполняются неравенства

$$4 + 14t \geq 0,$$
$$1 - t \geq 0,$$
$$2 + 3t \geq 0.$$

Первое и третье неравенства выполняются при всех положительных t (напомним, что $t \geq 0$), второе — справедливо при $t \leq 1$. Отсюда имеем, что $t_1 = 1$, т.е. решение $\mathbf{X}_{B_0}$ остается оптимальным (и допустимым) для всех t из интервала $0 \leq t \leq 1$.

При $t = 1$ разность $z_4(t) - c_4(t) = 1 - t$ равна нулю и становится отрицательной при $t > 1$. Поэтому при $t > 1$ вектор $\mathbf{P}_4$ должен войти в базис, тогда вектор $\mathbf{P}_2$ будет исключен из базиса (см. симплекс-таблицу с оптимальным решением при $t = 0$). Итак, при $t = 1$ (вследствие включения в базис вектора $\mathbf{P}_4$) получаем новое решение $\mathbf{X}_{B_1}$.

Оптимальное решение при $t_1 = 1$. Используя мультипликативную форму представления обратной матрицы (см. раздел 7.5.1, где показано, как определяется матрица $\mathbf{E}$, участвующая в дальнейших вычислениях), вычисляем новый базис.

$$\mathbf{B}_1^{-1} = \mathbf{E}\mathbf{B}_0^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Отсюда получаем

$$\mathbf{X}_{B_1} = (x_4, x_3, x_6)^T = \mathbf{B}_1^{-1}\mathbf{b} = (10, 30, 30)^T,$$
$$\mathbf{C}_{B_1}(t) = (0, 5 + 5t, 0).$$

Для небазисных векторов $\mathbf{P}_1$, $\mathbf{P}_2$ и $\mathbf{P}_5$ условия оптимальности можно записать следующим образом.

$$\left\{\mathbf{C}_{B_1}(t)\mathbf{B}_1^{-1}\mathbf{P}_j - c_j(t)\right\}_{j=1,2,5} = \left(\frac{9+27t}{2}, -2+2t, \frac{5+5t}{2}\right) \geq 0$$

В соответствии с этими условиями базисное решение $\mathbf{X}_{B_1}$ остается оптимальным для всех $t \geq 1$. Это означает, что $t_2 = \infty$ и вычислительный процесс параметрического анализа закончен. Отметим, что условие оптимальности $-2 + 2t \geq 0$ автоматически "запомнило", что решение $\mathbf{X}_{B_1}$ оптимально в интервале, который начинается со значения $t_1 = 1$. Такое "запоминание" всегда применяется при параметрическом анализе.

Оптимальные решения для всей области изменения параметра t представлены в следующей таблице. Значения целевой функции получены путем подстановки значений переменных в выражение целевой функции.

t	x_1	x_2	x_3	z
$0 \leq t \leq 1$	0	5	30	$160 + 140t$
$t \geq 1$	0	0	30	$150 + 150t$

Упражнения 7.7,а

1. Пусть в примере 7.7–1 параметр t может принимать как положительные, так и отрицательные значения. Определите интервал изменения значений параметра t, при которых решение $\mathbf{X}_{B_0}$ остается оптимальным.

2. Проведите параметрический анализ задачи ЛП из примера 7.7–1, если целевая функция задана следующими выражениями.

   a) Максимизировать $z = (3 + 3t)\,x_1 + 2x_2 + (5 - 6t)\,x_3$.
   b) Максимизировать $z = (3 - 2t)\,x_1 + (2 + t)\,x_2 + (5 + 2t)\,x_3$.
   c) Максимизировать $z = (3 + t)\,x_1 + (2 + 2t)\,x_2 + (5 - t)\,x_3$.

3. Проведите параметрический анализ оптимального решения следующей задачи ЛП, здесь $t \geq 0$.

   $$\text{Минимизировать } z = (4 - t)\,x_1 + (1 - 3t)\,x_2 + (2 - 2t)\,x_3$$

   при ограничениях

   $$3x_1 + x_2 + 2x_3 = 3,$$
   $$4x_1 + 3x_3 + 2x_3 \geq 6,$$
   $$x_1 + 2x_2 + 5x_3 \leq 4,$$
   $$x_1, x_2, x_3 \geq 0.$$

4. При выполнении параметрического анализа в этом разделе предполагалось, что оптимальное решение задачи ЛП при $t = 0$ получено обычным симплекс-методом. Однако в некоторых ситуациях предпочтительнее получение оптимального решения двойственным симплекс-методом (раздел 4.5). Разработайте схему проведения параметрического анализа для такого случая и выполните анализ задачи ЛП из примера 4.5–1, предполагая, что целевая функция задана следующим выражением.

   $$\text{Минимизировать } z = (3 + t)\,x_1 + (2 + 4t)\,x_2$$

5. Пусть в примере 7.7–1 целевая функция нелинейная по t ($t \geq 0$) и задается следующим выражением.

   $$\text{Максимизировать } z = (3 + 2t^2)\,x_1 + (2 - 2t^2)\,x_2 + (5 - t)\,x_3.$$

   Найдите первое критическое значение t_1.

7.7.2. Параметрическое изменение правых частей ограничений

Параметрическое изменение вектора правых частей ограничений $\mathbf{b}(t)$ влияет только на свойство допустимости решения. В этом случае критические значения параметра t определяются на основе следующего условия.

$$\mathbf{X}_{B_1} = \mathbf{B}_1^{-1}\mathbf{b}(t) \geq \mathbf{0}$$

Пример 7.7–2

Рассмотрим задачу ЛП.

$$\text{Максимизировать } z = 3x_1 + 2x_2 + 5x_3$$

при ограничениях

$$\begin{aligned} &x_1 + 2x_2 + x_3 \le 40 - t, \\ &3x_1 + 2x_3 \le 60 + 2t, \\ &x_1 + 4x_2 \le 30 - 7t, \\ &x_1, x_2, x_3 \ge 0. \end{aligned}$$

Предполагается, что $t \ge 0$.

Оптимальное решение этой задачи при $t = t_0 = 0$ приведено в примере 7.7–1. Имеем

$$\mathbf{X}_{B_0} = (x_2, x_3, x_6)^T = (5, 30, 10)^T,$$

$$\mathbf{B}_0^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix}.$$

Для нахождения первого критического значения t_1 используем условие допустимости $\mathbf{X}_{B_0} = \mathbf{B}_0^{-1}\mathbf{b}(t) \ge \mathbf{0}$, которое в данном случае примет вид

$$\begin{bmatrix} x_2 \\ x_3 \\ x_6 \end{bmatrix} = \begin{bmatrix} 5-t \\ 30+t \\ 10-3t \end{bmatrix} \ge \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Эти неравенства выполняются при $t \le 10/3$. Таким образом, $t_1 = 10/3$, и базис $\mathbf{B}_0$ остается допустимым при изменении параметра t в интервале $0 \le t \le 10/3$. Отметим, что здесь значения базисных переменных x_2, x_3 и x_6 изменяются при изменении параметра t.

Значение базисной переменной x_6 ($= 10 - 3t$) равно нулю при $t = t_1 = 10/3$ и становится отрицательным для $t > 10/3$. Следовательно, при $t = 10/3$ необходимо определить альтернативный (новый) базис $\mathbf{B}_1$. Для этого применим модифицированный симплекс-метод (см. упр. 7.5,b(5)). Исключаемой из базиса будет переменная x_6.

Альтернативный базис при $t = 10/3$.

Имея переменную x_6 в качестве исключаемой из базиса, определяем вводимую в базис. Так как

$$\mathbf{X}_{B_0} = (x_2, x_3, x_6)^T \text{ и } \mathbf{C}_{B_0}(t) = (2, 5, 0),$$

то

$$\{z_j - c_j\}_{j=1,4,5} = \{\mathbf{C}_{B_0}\mathbf{B}_0^{-1}\mathbf{P}_j - c_j\}_{j=1,4,5} = (4, 1, 2).$$

Далее вычисляем

$$\{(\mathbf{B}_0^{-1}\mathbf{P}_j)_{x_6}\}_{j=1,4,5} = (\text{строка в } \mathbf{B}_0^{-1}\text{, ассоциированная с } x_6)\ (\mathbf{P}_1, \mathbf{P}_4, \mathbf{P}_5) =$$

$$= (\text{3-я строка в } \mathbf{B}_0^{-1})\ (\mathbf{P}_1, \mathbf{P}_4, \mathbf{P}_5) =$$

$$= (-2, 1, 1)\ (\mathbf{P}_1, \mathbf{P}_4, \mathbf{P}_5) =$$

$$= (2, -2, 1).$$

Отсюда следует, что

$$\theta = \min\left\{-, \left|\frac{1}{-2}\right|, -\right\} = \frac{1}{2}$$

соответствует переменной x_4. Следовательно, в базис вводится вектор $\mathbf{P}_4$.

Вычисляем новый базис и обратную матрицу $\mathbf{B}_1^{-1}$.

$$\mathbf{X}_{B_1} = (x_2, x_3, x_4)^T,$$

$$\mathbf{B}_1^{-1} = \mathbf{E}\mathbf{B}_0^{-1} = \begin{bmatrix} 1 & 0 & \frac{1}{4} \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 \\ -2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}$$

Следующее критическое значение t_2 определяем из условия $\mathbf{X}_{B_1} = \mathbf{B}_1^{-1}\mathbf{b}(t) \geq \mathbf{0}$. Отсюда получаем следующее.

$$\begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \frac{30-7t}{4} \\ 30+t \\ \frac{-10+3t}{2} \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Это условие показывает, что $\mathbf{B}_1$ остается допустимым базисом для $10/3 \leq t \leq 30/7$.

При $t = t_2 = 30/7$ альтернативный базис можно получить с помощью модифицированного двойственного симплекс-метода, при этом исключаемой из базиса будет переменная x_2.

Альтернативный базис при $t_2 = 30/7$.

Определив исключаемую переменную x_2, находим вводимую переменную следующим образом. Так как

$$\mathbf{X}_{B_1} = (x_2, x_3, x_4)^T \text{ и } \mathbf{C}_{B_1}(t) = (2, 5, 0),$$

то

$$\{z_j - c_j\}_{j=1,5,6} = \{\mathbf{C}_{B_1}\mathbf{B}_1^{-1}\mathbf{P}_j - c_j\}_{j=1,5,6} = \left(5, \frac{5}{2}, \frac{1}{2}\right).$$

Далее вычисляем

$$\{(\mathbf{B}_1^{-1}\mathbf{P}_j)_{x_2}\}_{j=1,5,6} = (\text{строка в } \mathbf{B}_1^{-1}, \text{ ассоциированная с } x_2)\,(\mathbf{P}_1, \mathbf{P}_5, \mathbf{P}_6) =$$

$$= (\text{1-я строка в } \mathbf{B}_1^{-1})\,(\mathbf{P}_1, \mathbf{P}_5, \mathbf{P}_6) =$$

$$= \left(0, 0, \frac{1}{2}\right)(\mathbf{P}_1, \mathbf{P}_5, \mathbf{P}_6) =$$

$$= \left(\frac{1}{4}, 0, \frac{1}{4}\right).$$

Поскольку все элементы неотрицательны, задача не имеет допустимых решений при $t > 30/7$. Таким образом, параметрический анализ заканчивается в точке $t = t_2 = 30/7$.

Оптимальные решения при различных значениях параметра t приведены в следующей таблице.

t	x_1	x_2	x_3	z
$0 \le t \le 10/3$	0	$5 - t$	$30 + t$	$160 + 3t$
$10/3 \le t \le 30/7$	0	$(30 - 7t)/4$	$30 + t$	$165 + (3/2)t$
$t > 30/7$		Допустимых решений нет		

Упражнения 7.7,b

1. Для задачи из примера 7.7–2 найдите первое критическое значение t_1 и определите векторы, составляющие матрицу $\mathbf{B}_1$, для следующих случаев.

   a) $\mathbf{b}(t) = (40 + 2t, 60 - 3t, 30 + 6t)^T$.

   b) $\mathbf{b}(t) = (40 - t, 60 + 2t, 30 - 5t)^T$.

2. Проведите параметрический анализ оптимального решения следующей задачи ЛП, предполагая, что $t \ge 0$.

$$\text{Минимизировать } z = 4x_1 + x_2 + 2x_3$$

при ограничениях

$$3x_1 + x_2 + 2x_3 = 3 + 3t,$$
$$4x_1 + 3x_3 + 2x_3 \ge 6 + 2t,$$
$$x_1 + 2x_2 + 5x_3 \le 4 - t,$$
$$x_1, x_2, x_3 \ge 0.$$

3. При выполнении параметрического анализа в этом разделе предполагалось, что оптимальное решение задачи ЛП при $t = 0$ получено обычным симплекс-методом. Однако в некоторых ситуациях предпочтительнее получение оптимального решения двойственным симплекс-методом (раздел 4.5). Разработайте схему проведения

параметрического анализа для такого случая и выполните анализ задачи ЛП из примера 4.5–1, предполагая, что вектор правых частей ограничений задачи задан следующим выражением (считаем, что $t \geq 0$).

$$\mathbf{b}(t) = (3 + 2t, 6 - t, 3 - 4t)^T$$

4. Решите задачу из упр. 2, предполагая, что вектор правых частей ограничений задан следующим выражением.

$$\mathbf{b}(t) = (3 + 3t^2, 6 + 2t^2, 4 - t^2)^T$$

Затем решите эту же задачу при условии, что параметр t может принимать как положительные, так и отрицательные значения.

7.8. Метод Кармаркара

В симплекс-методе оптимальное решение находится путем продвижения вдоль граней пространства решений от одной угловой (крайней) точки к другой. Хотя на практике симплекс-метод применяется для решения очень больших задач, теоретически количество итераций, необходимых для достижения оптимального решения, может расти по экспоненциальному закону. Можно построить задачу ЛП с n переменными, при решении которой необходимо перебрать все $2n$ крайние точки, пока не будет достигнуто оптимальное решение.[1]

В 1984 году Н. Кармаркар (N. Karmarkar) разработал алгоритм с полиномиальным временем выполнения, который "пробегает" по внутренним точкам пространства решений. Этот алгоритм особенно эффективен при решении очень больших задач ЛП.

Мы сначала рассмотрим основную идею метода Кармаркара, а затем остановимся на вычислительной реализации алгоритма.

7.8.1. Основная идея метода Кармаркара

Рассмотрим очень простой пример.

$$\text{Максимизировать } z = x_1$$

при выполнении ограничения

$$0 \leq x_1 \leq 2.$$

Вводим дополнительную (остаточную) переменную x_2. Задача в стандартной форме будет записана следующим образом.

$$\text{Максимизировать } z = x_1$$

при ограничениях

$$x_1 + x_2 = 2,$$
$$x_1, x_2 \geq 0.$$

---

[1] Попытайтесь решить с помощью программы TORA следующую задачу ЛП: максимизировать $z = x_1 + x_2$ при ограничениях $x_1 \leq 1$, $2x_1 + x_2 \leq 3$ и $x_1, x_2 \geq 0$. Убедитесь, что в процессе решения будут проверены все $2^2 = 4$ крайние точки

Эта задача представлена на рис. 7.6. Её пространство решений совпадает с отрезком прямой AB. Направление возрастания целевой функции z совпадает с положительным направлением оси x_1.

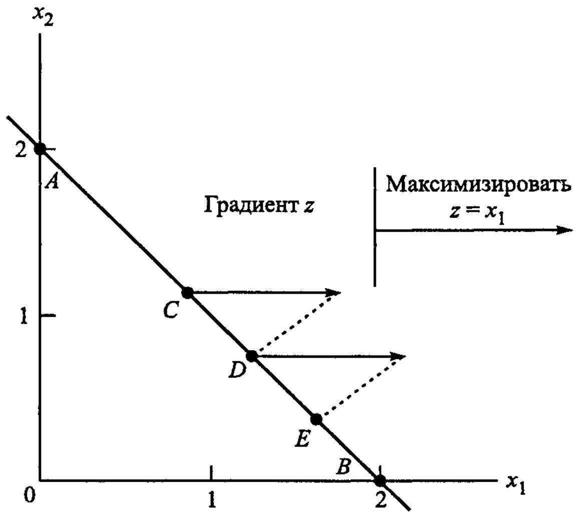


Рис. 7.6.

Начнем поиск оптимального решения с произвольной *внутренней* (не крайней) точки C пространства допустимых решений (отрезок AB). **Градиент** целевой функции $z = x_1$ в точке C показывает направление скорейшего возрастания функции z. Если мы зафиксируем какую-нибудь точку вдоль градиента и опустим из нее перпендикуляр на пространство допустимых решений, то получим новую точку D с лучшим значением целевой функции. Такое улучшение значения целевой функции получено вследствие движения по направлению проекции CD градиента. Если мы повторим описанную процедуру для точки D, получим новую точку E, которая будет еще ближе к точке оптимума B. Таким образом, если мы будем двигаться (но осторожно) в направлении **проекции градиента**, то, вероятно, рано или поздно "споткнемся" о точку оптимума B. Если же необходимо найти минимум целевой функции (вместо ее максимума), следует перемещаться в направлении, противоположном проекции градиента, т.е. от точки B к точке A, где $x_1 = 0$.

Конечно, данную процедуру нельзя считать алгоритмом (в обычном смысле), но идея очень интересная! Чтобы воплотить эту идею в алгоритм, необходимо выполнить модификации приведенной процедуры, гарантирующие, что (1) последовательность шагов, сделанных вдоль проекции градиента, не "перескочит" через точку оптимума, (2) в общем n-мерном случае направления, указанные проекциями градиентов, не приведут алгоритм к неоптимальной точке (т.е. не "зациклят" алгоритм на неоптимальной точке). Если реализовать эти условия, то получим вполне работоспособный алгоритм.

7.8.2. Алгоритм Кармаркара

Существует несколько вариантов алгоритма Кармаркара. Мы рассмотрим исходный вариант, предложенный его автором. Кармаркар предполагал, что задача ЛП приведена к следующему виду.

$$\text{Минимизировать } z = \mathbf{CX}$$

при ограничениях

$$\mathbf{AX} = \mathbf{0},$$
$$\mathbf{1X} = 1,$$
$$\mathbf{X} \geq \mathbf{0}.$$

Здесь все ограничения представлены в виде однородных уравнений, за исключением ограничения $\mathbf{1X} = \sum_{j=1}^{n} x_j = 1$, которое определяет n-мерный правильный симплекс.[1] Обоснованность алгоритма Кармаркара "покоится" на выполнении двух условий.

1. Вектор $\mathbf{X} = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)$ удовлетворяет ограничениям $\mathbf{AX} = \mathbf{0}$.
2. $\min z = 0$.

Кармаркар предложил алгебраические преобразования, приводящие общую задачу ЛП к виду, представленному выше. Эти преобразования показаны в следующем примере. Там также показано, как в результате преобразований добиться того, чтобы вектор $\mathbf{X} = (1/n, 1/n, ..., 1/n)$ являлся допустимым решением системы $\mathbf{AX} = \mathbf{0}$ (условие 1). Преобразования, необходимые для выполнения условия 2 ($\min z = 0$), мы не приводим из-за их громоздкости и трудоемкости.

Пример 7.8–1

Рассмотрим следующую задачу.

$$\text{Максимизировать } z = y_1 + y_2$$

при ограничениях

$$y_1 + 2y_2 \leq 2,$$
$$y_1, y_2 \geq 0.$$

С помощью дополнительной переменной $y_3 \geq 0$ преобразуем ограничение $y_1 + 2y_2 \leq 2$ в равенство:

$$y_1 + 2y_2 + y_3 = 2.$$

Теперь введем неравенство

$$y_1 + y_2 + y_3 \leq U,$$

где U — достаточно большое положительное число, но такое, которое не удаляло бы ни одной допустимой точки исходного пространства решений. В данном примере, исходя из равенства $y_1 + 2y_2 + y_3 = 2$, достаточно взять U, равное 5. После введения еще одной дополнительной переменной y_4 получаем

---

[1] Симплексом (n-мерным) называется выпуклая оболочка n точек $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$, не лежащих на одной $(n-2)$-мерной плоскости. Точки $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ являются вершинами симплекса, при этом любую точку симплекса можно представить как выпуклую комбинацию (см раздел 7.3.1) его вершин. Двухмерный симплекс — это отрезок, трехмерный — треугольник, четырехмерный — тетраэдр. (Здесь размерность симплекса определяется по размерности пространства, а не по размерности $(n-1)$-мерной гиперплоскости, на которой он расположен) В данном случае вершинами симплекса являются точки $\mathbf{X}_i = (0, ..., 1, ..., 0)$, $i = 1, ..., n$. Такой симплекс является частным случаем *правильного* симплекса — *Прим ред.*

$$y_1 + y_2 + y_3 + y_4 = 5.$$

Теперь можно сделать уравнение $y_1 + 2y_2 + y_3 = 2$ однородным путем умножения его правой части на $(y_1 + y_2 + y_3 + y_4)/5$, поскольку последнее соотношение равно 1. Таким образом, после приведения подобных членов получаем

$$3y_1 + 8y_2 + 3y_3 - 2y_4 = 0.$$

Чтобы преобразовать равенство $y_1 + y_2 + y_3 + y_4 = 5$ в уравнение, определяющее симплекс, введем новые переменные $x_i = y_i/5$, $i = 1, 2, 3, 4$. Получаем следующую задачу ЛП.

Максимизировать $z = 5x_1 + 5x_2$

при ограничениях

$$3x_1 + 8x_2 + 3x_3 - 2x_4 = 0,$$
$$x_1 + x_2 + x_3 + x_4 = 1,$$
$$x_i \geq 0,\ i = 1, 2, 3, 4.$$

Теперь обеспечим выполнение условия, что точка $\mathbf{X} = (1/n, 1/n, ..., 1/n)$, являющаяся центром (барицентром) симплекса, будет удовлетворять однородному уравнению. Для этого от левой части каждого однородного уравнения отнимем искусственную переменную с коэффициентом, равным алгебраической сумме всех коэффициентов левой части уравнения (в данном случае имеем $3 + 8 + 3 - 2 = 12$). Эта искусственная переменная также прибавляется к уравнению симплекса и в виде штрафа появляется в выражении целевой функции. В нашем примере искусственная переменная x_5 войдет в задачу ЛП следующим образом.

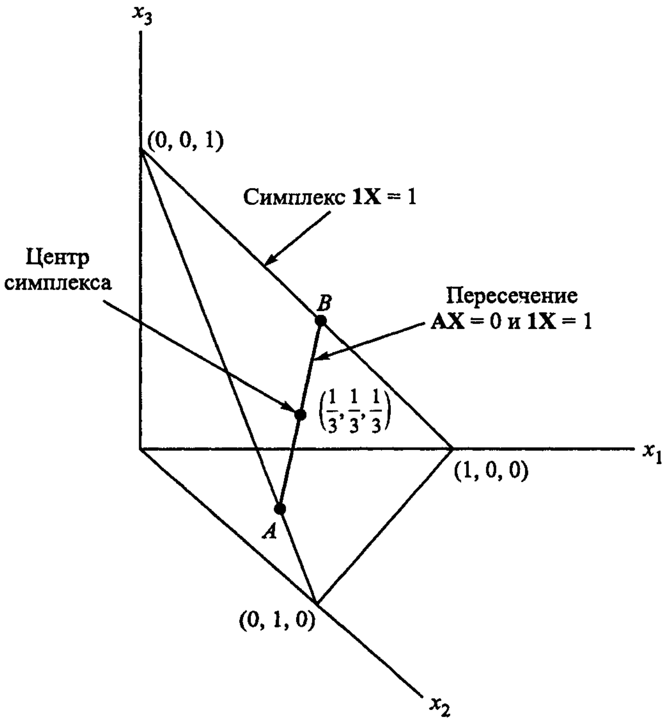
Максимизировать $z = 5x_1 + 5x_2 - Mx_5$

при ограничениях

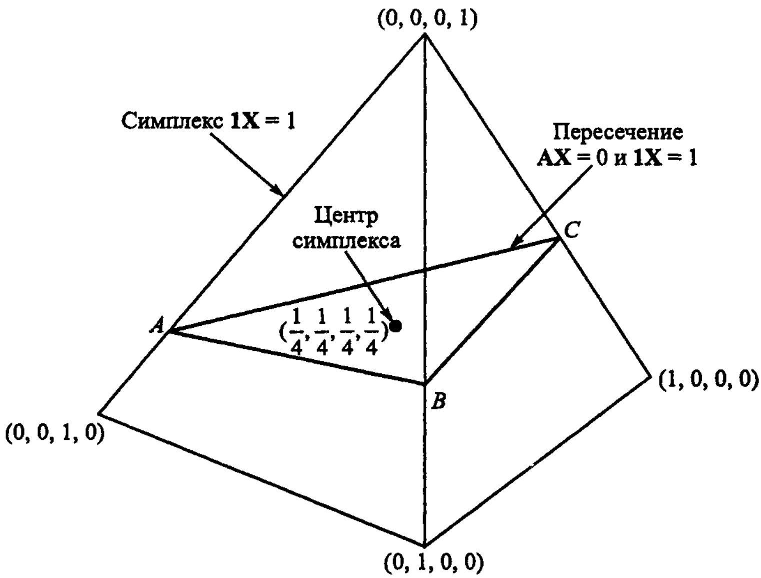
$$3x_1 + 8x_2 + 3x_3 - 2x_4 - 12x_5 = 0,$$
$$x_1 + x_2 + x_3 + x_4 + x_5 = 1,$$
$$x_i \geq 0,\ i = 1, 2, ..., 5.$$

Для этой системы уравнений новый центр симплекса (1/5, 1/5, ..., 1/5) является допустимым решением однородного уравнения. Значение константы M в выражении целевой функции должно быть достаточно большим, чтобы привести переменную x_5 к нулевому значению (сравните с M-методом из раздела 3.4.1).

Теперь опишем основные шаги алгоритма Кармаркара. На рис. 7.7,*а* показано типичное трехмерное пространство решений, определяемое однородной системой ограничений $\mathbf{AX} = \mathbf{0}$, состоящей из одного уравнения. В данном случае пространство решений совпадает с отрезком прямой AB, лежащим внутри симплекса $\mathbf{1X} = 1$, причем точка (1/3, 1/3, 1/3) является допустимой внутренней точкой пространства решений. Аналогично, на рис. 7.7,*б* показано четырехмерное пространство решений (треугольник ABC), определяемое однородной системой ограничений, также состоящей из одного уравнения. В этом случае центром симплекса является точка (1/4, 1/4, 1/4, 1/4).



а) Трехмерное пространство решений



б) Четырехмерное пространство решений

Рис. 7.7

Принципиальная идея Кармаркара заключается в том, что вычисления начинаются с внутренней точки, соответствующей центру симплекса, далее в направлении *проекции градиента* определяется новая точка решения, которая должна быть строго внутренней, т.е. все ее координаты должны быть положительными. Это является необходимым условием сходимости алгоритма.

Чтобы новая точка решения была внутренней, она не должна выходить за пределы симплекса. (Это значит, что на рис. 7.7,*а* новая точка не может быть точкой A или B, а на рис. 7.7,*б* — не может совпадать с точками отрезков AB, BC и AC.) Для того чтобы определить такую точку, построим сферу, вписанную в симплекс. В n-мерном пространстве в правильный симплекс можно вписать сферу с максимальным радиусом r, равным $1/\sqrt{n(n-1)}$.[1] Тогда пересечение сферы радиуса αr $(0 < \alpha < 1)$ и пространства решений, определяемого системой $\mathbf{AX} = \mathbf{0}$, будет содержать только внутренние точки пространства решений с положительными координатами. В таком случае для определения новой точки решения можно перемещаться вдоль проекции градиента до тех пор, пока будем находиться внутри ограниченного пространства, являющегося пересечением шара радиуса αr и пространства, определяемого системой $\mathbf{AX} = \mathbf{0}$.

Новая точка решения не обязана быть центром симплекса. Поэтому, чтобы сделать алгоритм итерационным, необходимо найти способ перенесения новой точки решения снова в центр симплекса. Для этого Кармаркар предложил оригинальную идею **проективных преобразований.** Пусть

$$y_i = \frac{\dfrac{x_i}{x_{ki}}}{\sum_{j=1}^{n} \dfrac{x_j}{x_{kj}}}, \quad i = 1, 2, \ldots, n,$$

где x_{ki} — i-й элемент текущего решения, т.е. i-я координата текущей точки решения $\mathbf{X}_k$. Такое преобразование правомерно, поскольку все $x_{ki} > 0$. Также отметим, что по определению $\sum_{i=1}^{n} y_i = 1$ (т.е. $\mathbf{1Y} = 1$). Это преобразование можно записать в матричном виде

$$\mathbf{Y} = \frac{\mathbf{D}_k^{-1}\mathbf{X}}{\mathbf{1D}_k^{-1}\mathbf{X}},$$

где $\mathbf{D}_k$ — диагональная матрица, у которой i-й диагональный элемент равен x_{ki}. Это преобразование осуществляет взаимно-однозначное отображение X-пространства в Y-пространство, поскольку нетрудно проверить, что из последней формулы следует

$$\mathbf{X} = \frac{\mathbf{D}_k\mathbf{Y}}{\mathbf{1D}_k\mathbf{Y}}.$$

По определению $\min \mathbf{CX} = 0$. Отсюда следует, что значение $\mathbf{1D}_k\mathbf{Y}$ должно быть положительным. В таком случае исходная задача ЛП преобразуется в следующую.

---

[1] Здесь надо уточнить, что центр этой сферы находится в *барицентре* симплекса. Далее в этом разделе, не оговаривая этого явно, везде предполагается, что центры всех сфер совпадают с барицентром симплекса. — *Прим. ред.*

$$\text{Минимизировать } z = \mathbf{CD}_k\mathbf{Y}$$

при ограничениях

$$\mathbf{AD}_k\mathbf{Y} = \mathbf{0},$$
$$\mathbf{1Y} = 1,$$
$$\mathbf{Y} \geq \mathbf{0}.$$

Преобразованная задача имеет тот же тип, что и исходная. Мы можем опять начать решение этой задачи с точки $\mathbf{Y} = (1/n, 1/n, ..., 1/n)$, являющейся центром симплекса, и повторить действия, выполненные на предыдущем шаге. После каждой итерации на основании полученного $\mathbf{Y}$-решения нетрудно вычислить значения исходных $\mathbf{X}$-переменных.

Теперь покажем, как определить новую точку решения для преобразованной задачи. На любой k-й итерации задача имеет следующий вид.

$$\text{Минимизировать } z = \mathbf{CD}_k\mathbf{Y}$$

при ограничениях

$$\mathbf{AD}_k\mathbf{Y} = \mathbf{0},$$
$$\mathbf{Y} \text{ лежит в шаре радиуса } \alpha r.$$

Поскольку шар радиуса αr является частью пространства допустимых решений, определяемого ограничениями $\mathbf{1X} = 1$, $\mathbf{X} \geq \mathbf{0}$, эти ограничения можно исключить из рассмотрения. Можно показать, что решение последней задачи вычисляется по следующей формуле

$$\mathbf{Y}_{\text{новое}} = \mathbf{Y}_{\text{текущее}} + \alpha r \frac{\mathbf{c}_p}{\|\mathbf{c}_p\|},$$

где $\mathbf{Y}_{\text{текущее}} = (1/n, 1/n, ..., 1/n)$, $\mathbf{c}_p$ — проекция градиента, которую можно вычислить так:

$$\mathbf{c}_p = [\mathbf{I} - \mathbf{P}^T(\mathbf{PP}^T)^{-1}\mathbf{P}](\mathbf{CD}_k)^T,$$

где $\mathbf{P} = \begin{pmatrix} \mathbf{AD}_k \\ \mathbf{1} \end{pmatrix}$.

Выбор параметра α — ключевой момент при построении алгоритма. Обычно α выбирается таким большим, насколько это возможно, чтобы ускорить сходимость к оптимальному решению. Но при выборе слишком большого значения α можно "упереться" в запрещенную границу симплекса. В настоящее время не известно оптимальное значение α. Кармаркар предлагал использовать $\alpha = (n-1)/3n$.

Итак, алгоритм Кармаркара требует выполнения следующих шагов.

Шаг 0. Алгоритм начинается с точки $\mathbf{X}_0 = (1/n, 1/n, ..., 1/n)$, вычисляются параметры $r = 1/\sqrt{n(n-1)}$ и $\alpha = (n-1)/3n$.

Основной шаг k. Определяем

$$\mathbf{D}_k = \text{diag}\{x_{k1}, ..., x_{kn}\},$$

$$\mathbf{P} = \begin{pmatrix} \mathbf{AD}_k \\ \mathbf{1} \end{pmatrix}$$

и вычисляем

$$\mathbf{Y}_{\text{новое}} = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)^T + \alpha r \frac{\mathbf{c}_p}{\|\mathbf{c}_p\|},$$

$$\mathbf{X}_{k+1} = \frac{\mathbf{D}_k \mathbf{Y}_{\text{новое}}}{\mathbf{1D}_k \mathbf{Y}_{\text{новое}}},$$

где $\mathbf{c}_p = [\mathbf{I} - \mathbf{P}^T(\mathbf{PP}^T)^{-1}\mathbf{P}](\mathbf{CD}_k)^T$.

Пример 7.8–2

Рассмотрим следующую задачу ЛП, которая уже приведена к форме, необходимой для выполнения алгоритма Кармаркара.

$$\text{Минимизировать } z = x_3$$

при ограничениях

$$x_1 - 2x_2 + x_3 = 0,$$
$$x_1 + x_2 + x_3 = 1,$$
$$x_1, x_2, x_3 \geq 0.$$

Эта задача удовлетворяет всем условиям алгоритма Кармаркара, а именно точка $\mathbf{X} = (1/3, 1/3, 1/3)$ удовлетворяет уравнению $x_1 - 2x_2 + x_3 = 0$ и оптимальное значение целевой функции равно нулю (оптимальным решением является вектор $(2/3, 1/3, 0)$).

Итерация 0

$$\mathbf{X}_0 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \quad r = \frac{1}{\sqrt{6}}, \quad \alpha = \frac{2}{9}, \quad z = \frac{1}{3} = 0.33333.$$

Итерация 1

$$\mathbf{D}_0 = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}, \quad \mathbf{CD}_0 = \left(\frac{1}{3}, -\frac{2}{3}, 0\right),$$

$$\mathbf{AD}_0 = (1, -2, 1)\begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} = \left(\frac{1}{3}, -\frac{2}{3}, \frac{1}{3}\right),$$

$$\mathbf{P} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ 1 & 1 & 1 \end{bmatrix},$$

$$\left(\mathbf{PP}^T\right)^{-1}=\begin{bmatrix}\frac{2}{3} & 0\\ 0 & 3\end{bmatrix}^{-1}=\begin{bmatrix}\frac{3}{2} & 0\\ 0 & \frac{1}{3}\end{bmatrix},$$

$$\mathbf{c}_p=\left[\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}-\begin{pmatrix}\frac{1}{3}&1\\-\frac{2}{3}&1\\\frac{1}{3}&1\end{pmatrix}\begin{pmatrix}\frac{3}{2}&0\\0.&\frac{1}{3}\end{pmatrix}\begin{pmatrix}\frac{1}{3}&-\frac{2}{3}&\frac{1}{3}\\1&1&1\end{pmatrix}\right]\begin{bmatrix}\frac{1}{3}\\-\frac{2}{3}\\0\end{bmatrix}=$$

$$=\left(\frac{1}{6},0,-\frac{1}{6}\right)^T,$$

$$\left\|\mathbf{c}_p\right\|=\sqrt{\left(\frac{1}{6}\right)^2+0+\left(\frac{1}{6}\right)^2}=0.2357,$$

$$\frac{\alpha r}{\left\|\mathbf{c}_p\right\|}=\frac{\frac{2}{9}\cdot\frac{1}{\sqrt{6}}}{0.2357}=0.384901,$$

$$\mathbf{Y}_{\text{новое}}=\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)^T+0.384901\left(\frac{1}{6},0,-\frac{1}{6}\right)^T=$$
$$=(0.397483,\ 0.333333,\ 0.269183)^T.$$

Теперь найдем $\mathbf{X}_1$ — решение исходной задачи, соответствующее $\mathbf{Y}_{\text{новое}}$. Поскольку $\mathbf{D}_0$ — диагональная матрица с диагональными элементами, равными 1/3, из равенства $\mathbf{1Y}=1$ и формулы $\mathbf{X}_1=\mathbf{D}_0\mathbf{Y}_{\text{новое}}/\mathbf{1D}_0\mathbf{Y}_{\text{новое}}$ вытекает, что $\mathbf{X}_1=\mathbf{Y}_{\text{новое}}=(0.397483, 0.333333, 0.269183)T$. Соответствующее значение целевой функции равно $z=0.269183$, что лучше предыдущего значения $z=0.333333$ при решении $\mathbf{X}_0$.

Если мы возьмем значение α, большее текущего 2/9, оптимальное решение (2/3, 1/3, 0) можно достигнуть значительно быстрее. В частности, если положить $\alpha=1$, то уже $\mathbf{Y}_{\text{новое}}$ будет соответствовать оптимальному решению. Но мы *не можем* положить $\alpha=1$, так как решение $x_3=0$ не соответствует предположениям, на которых базируются проективные преобразования. Это еще раз показывает те проблемы, которые возможны при реализации данного метода.

Итерация 2

$\mathbf{D}_1=\text{diag}\{0.397483, 0.333333, 0.269183\}$,
$\mathbf{CD}_1=(0.397483, -0.666666, 0)$,

$$\mathbf{AD}_1=(1,-2,1)\begin{bmatrix}0.397485&0&0\\0&0.333333&0\\0&0&0.269182\end{bmatrix}=$$

$$= (0.397483, -0.666666, 0.269183),$$

$$\mathbf{P} = \begin{bmatrix} 0.397845 & -0.666666 & 0.269182 \\ 1 & 1 & 1 \end{bmatrix},$$

$$\mathbf{c}_p = (0.132402, 0.018152, -0.150555)^T,$$

$$\left\| \mathbf{c}_p \right\| = 0.201312,$$

$$\frac{\alpha r}{\left\| \mathbf{c}_p \right\|} = \frac{\frac{2}{9} \cdot \frac{1}{\sqrt{6}}}{0.201312} = 0.450653,$$

$$\mathbf{Y}_{\text{новое}} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)^T + 0.450653\,(0.132402, 0.018152, -0.150555)^T =$$

$$= (0.393001, 0.341514, 0.265486)^T.$$

Теперь для вычисления $\mathbf{X}_2$ имеем

$$\mathbf{D}_1\mathbf{Y}_{\text{новое}} = \begin{pmatrix} 0.156212 \\ 0.113838 \\ 0.071464 \end{pmatrix}, \quad \mathbf{1D}_1\mathbf{Y}_{\text{новое}} = 0.341514,$$

$$\mathbf{X}_2 = \frac{\mathbf{D}_1\mathbf{Y}_{\text{новое}}}{\mathbf{1D}_1\mathbf{Y}_{\text{новое}}} = \begin{pmatrix} 0.457411 \\ 0.333333 \\ 0.209256 \end{pmatrix},$$

$$z = 0.20934.$$

Выполнив еще один шаг алгоритма, получим решение, которое будет еще ближе к оптимальному. Кармаркар предлагает дополнительный шаг, который позволяет перейти от полученного решения к оптимальной крайней точке. К сожалению, мы не имеем возможности рассмотреть его здесь.

Упражнения 7.8,а

1. Выполните еще одну итерацию в примере 7.8–2 и убедитесь, что полученное решение приблизится к оптимальному.
2. Выполните три итерации алгоритма Кармакара для решения следующей задачи.

$$\text{Максимизировать } z = -4x_1 + x_3 + x_4$$

при ограничениях

$$-2x_1 + 2x_2 + x_3 - x_4 = 0,$$
$$x_1 + x_2 + x_3 + x_4 = 1,$$
$$x_1, x_2, x_3, x_4 \geq 0.$$

3. Выполните три итерации алгоритма Кармаркара для решения следующей задачи ЛП.

Максимизировать $z = 2x_1 + x_2$

при ограничениях

$$x_1 + x_2 \leq 4,$$
$$x_1, x_2 \geq 0.$$

7.9. Заключение

В этой главе представлены методы решения задач линейного программирования: модифицированный симплекс-метод, метод решения задач с ограниченными переменными и метод декомпозиции. Особенность этих методов заключается в их вычислительной эффективности и точности полученных результатов. Первые два метода, модифицированный симплекс-метод и метод решения задач с ограниченными переменными, предлагают наиболее многообещающие решения задач ЛП. Метод декомпозиции из-за сложности вычислений не нашел такого широкого применения, как первые два. Тем более, что сегодня возможности современной вычислительной техники предлагают иные способы решения тех задач ЛП, которые ранее решались именно методом декомпозиции.

Параметрическое программирование — полезный инструмент анализа чувствительности задач ЛП. Здесь исследуется зависимость статического оптимального решения при динамическом изменении структуры модели ЛП.

Метод Кармаркара не относится к "популярным" методам решения задач линейного программирования. Практически отсутствует литература, где бы рассматривались вопросы использования этого метода. По-видимому, для более широкого распространения этого метода потребуются специальные вычислительная техника и программное обеспечение, а также более глубокое обоснование эвристических условий выполнения этого алгоритма.

Литература

1. Bazaraa M., Jarvis J., Sherali M. *Linear Programming and Network Flows*, 2nd ed., Wiley, New York, 1990.
2. Hooker J. *Karmarkar's Linear Programming Algorithm*, Interfaces, Vol. 16, No. 4, pp. 75–90, 1986.
3. Nering E., Tucker A. *Linear Programming and Related Problems*, Academic Press, Boston, 1992.

Литература, добавленная при переводе

Ашманов С.А. *Линейное программирование.* — М.: Наука, 1981.

Гольштейн Е.Г. *Теория двойственности в математическом программировании и ее приложения.* — М.: Наука, 1971.

Гольштейн Е.Г., Юдин Д.Б. *Линейное программирование: Теория, методы и приложения.* — М.: Наука, 1969.

Крона Г. *Исследование сложных систем по частям — диакоптика.* — М.: Наука, 1972.

Комплексные задачи

■ **7-1.** Имеем координаты трех точек:

$$A = (6, 4, 6, -2),\ B = (4, 12, -4, 8) \text{ и } C = (-4, 0, 8, 4).$$

Разработайте процедуру, которая позволяла бы определить, является ли данная точка выпуклой комбинацией заданных точек. С помощью этой процедуры проверьте, будут ли следующие точки выпуклыми комбинациями точек A, B и C.

a) (3, 5, 4, 2),
b) (5, 8, 4, 9).

■ **7-2.** Рассмотрим следующую задачу ЛП.

$$\text{Максимизировать } z = 3x_1 + 2x_2$$

при ограничениях

$$\begin{aligned} x_1 + 2x_2 &\le 6, \\ 2x_1 + x_2 &\le 8, \\ -x_1 + x_2 &\le 1, \\ x_1, x_2 &\ge 0. \end{aligned}$$

Найдите оптимальное решение этой задачи (можно применить программу TORA). Затем, используя информацию, представленную в симплекс-таблице с оптимальным решением, определите *второе* (по отношению к “абсолютному” оптимуму) наилучшее решение (т.е. крайнюю точку пространства решений). Проверьте ответ, решив эту задачу графически. (*Подсказка*. Искомая крайняя точка будет *смежной* с точкой, соответствующей оптимальному решению.)

■ **7-3.** *Интервальное программирование.* Рассмотрим следующую задачу ЛП.

$$\text{Максимизировать } z = \{\mathbf{CX} \mid \mathbf{L} \le \mathbf{AX} \le \mathbf{U},\ \mathbf{X} \ge \mathbf{0}\},$$

где $\mathbf{L}$ и $\mathbf{U}$ — вектор-столбцы констант. Определим вектор $\mathbf{Y} \ge \mathbf{0}$, такой, что $\mathbf{AX} + \mathbf{Y} = \mathbf{U}$. Покажите, что исходная задача эквивалентна следующей.

$$\text{Максимизировать } z = \{\mathbf{CX} \mid \mathbf{AX} + \mathbf{Y} = \mathbf{U},\ 0 \le \mathbf{Y} \le \mathbf{U} - \mathbf{L},\ \mathbf{X} \ge \mathbf{0}\}.$$

Используя показанный прием, решите следующую задачу ЛП.

$$\text{Минимизировать } z = 5x_1 - 4x_2 + 6x_3$$

при ограничениях

$$\begin{aligned} 20 \le x_1 + 7x_2 + 3x_3 &\le 46, \\ 10 \le 3x_1 - x_2 + x_3 &\le 20, \\ 18 \le 2x_1 + 3x_2 - x_3 &\le 35, \\ x_1, x_2, x_3 &\ge 0. \end{aligned}$$

■ **7-4.** Рассмотрим следующую целочисленную задачу ЛП, где переменные принимают только значения 0 и 1.

$$\text{Минимизировать } z = \{\mathbf{CX} \mid \mathbf{AX} \leq \mathbf{b}, \mathbf{X} \in (0, 1)\}.$$

Предположим, что известна верхняя граница целевой функции $z_{\min}$. Введем ограничение

$$\min_{\mu \geq 0} \max_{X=(0,1)} \{\mu(\mathbf{b} - \mathbf{AX}) + (z_{\min} - \mathbf{CX})\} \geq 0,$$

где $\mu \geq 0$. Это ограничение не нарушает ни одного из ограничений исходной задачи. Покажите, что определение μ из этого ограничения фактически сводится к решению исходной задачи ЛП. (*Совет*. В данном случае условие целочисленности $\mathbf{X} = [0, 1]$ эквивалентно "непрерывному" ограничению $\mathbf{0} \leq \mathbf{X} \leq \mathbf{1}$. Сформулируйте двойственную задачу.)

■ **7-5.** В задаче 7-2 оптимальным решением является $x_1 = 10/3$, $x_2 = 4/3$ и $z = 38/3$. Как будет изменяться оптимальное значение целевой функции в зависимости от параметра θ, если положить $x_1 = 10/3 + \theta$ (на параметр θ не накладываются ограничения в знаке)?

■ **7-6.** Предположим, что оптимальное решение задачи ЛП представлено следующим образом.

$$\text{Максимизировать } z = c_0 - \sum_{j \in NB} (z_j - c_j) x_j$$

при ограничениях

$$x_i = x_i^* - \sum_{j \in NB} \alpha_{ij} x_j, \quad i = 1, 2, ..., m,$$

$$\text{все } x_i \text{ и } x_j \geq 0,$$

где NB — множество индексов небазисных переменных. Предположим, что на текущую базисную переменную $x_i = x_i^*$ накладывается ограничение $x_i \geq d_i$, где d_i — наименьшее целое, большее x_i^*. После введения этого ограничения оцените верхнюю границу оптимального значения целевой функции. Повторите процедуру оценки верхней границы при наложении ограничения $x_i \leq e_i$, где e_i — наибольшее целое, меньшее x_i^*.

■ **7-7.** Рассмотрим следующую задачу линейного программирования.

$$\text{Минимизировать } z = (10t - 4) x_1 + (4t - 8) x_2$$

при ограничениях

$$2x_1 + 2x_2 + x_3 = 8,$$
$$4x_1 + 2x_2 + x_4 = 6 - 2t,$$
$$x_1, x_2, x_3, x_4 \geq 0,$$

где $-\infty \leq t \leq \infty$. С помощью параметрического анализа получены следующие результаты:

при $-\infty \leq t \leq -5$ оптимальный базис: $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_4)$,

при $-5 \leq t \leq -1$ оптимальный базис: $\mathbf{B} = (\mathbf{P}_1, \mathbf{P}_2)$,

при $-1 \leq t \leq 2$ оптимальный базис: $\mathbf{B} = (\mathbf{P}_2, \mathbf{P}_3)$.

Найдите все критические значения параметра t при $t \geq 2$.

Целевое программирование

8.1. Несколько целевых функций

Модели линейного программирования, рассмотренные в предыдущих главах, предполагали оптимизацию только *одной целевой функции*. Но возможны ситуации, когда в модели присутствует несколько (возможно, конфликтующих между собой) целевых функций. Например, честолюбивый политик может пообещать уменьшение национального долга и, одновременно, снижение ставок налогов. В таких ситуациях иногда невозможно найти единственное решение, оптимизирующее все конфликтующие целевые функции. Поэтому нужно искать *компромиссное* решение, учитывающее "важность" каждой целевой функции.

В этой главе представлены методы целевого программирования (многокритериальной оптимизации[1]) для решения задач линейного программирования с несколькими целевыми функциями. Основное назначение этих методов — преобразование исходной задачи с несколькими целевыми функциями в задачу ЛП с одной целевой функцией. После решения преобразованной задачи получаем так называемое **эффективное решение**, поскольку может не существовать оптимального решения, доставляющего оптимум *всем* частным целевым функциям исходной задачи.

8.2. Формулировка задачи целевого программирования

Рассмотрим пример, приводящий к задаче целевого программирования.

Пример 8.2–1

Файрвилл — небольшой городок, в котором проживает около 20 тысяч жителей. Предположим, городской совет разрабатывает ставки местного налогообложения. Ежегодная база налогообложения недвижимости составляет 550 миллионов долларов. Ежегодная база налогообложения розничных и оптовых продаж составляет 35 и 55 миллионов долларов соответственно. Ежегодное потребление городом бензина оценивается в 7.5 миллионов галлонов. Городской совет планирует разработать систему налоговых ставок, основанную на перечисленных базах налогообложения и учитывающую следующие ограничения и требования.

[1] Задачи целевого программирования являются только подклассом задач, решаемых методами многокритериальной оптимизации. Подходы и методы, применяемые в многокритериальной оптимизации, отнюдь не исчерпываются теми методами, которые описаны в этой главе (см. список литературы в конце главы). — *Прим. ред.*

1. Налоговые поступления должны составить не менее 16 миллионов долларов от всех баз налогообложения.
2. Налог с розничных продаж не может превышать 10% от суммы всех собираемых налогов.
3. Налог с оптовых продаж не может превышать 20% от суммы всех налогов.
4. Налог на бензин не может превышать 2 центов за галлон.

Обозначим через $x_н$, $x_р$ и $x_о$ ставки налогов (выраженные десятичными дробями) на недвижимость, розничную и оптовую торговлю соответственно, а через $x_б$ — налог на бензин, выраженный в центах на галлон. Пожелания городского совета можно записать следующим образом.

$550x_н + 35x_р + 55x_о + 0.075x_б \geq 16$ (Суммарные налоговые поступления)

$35x_р \leq 0.1\,(550x_н + 35x_р + 55x_о + 0.075x_б)$ (Налог на розничную торговлю)

$55x_о \leq 0.2\,(550x_н + 35x_р + 55x_о + 0.075x_б)$ (Налог на оптовую торговлю)

$x_б \leq 2$ (Налог на бензин)

$x_н, x_р, x_о, x_б \geq 0.$

После упрощения получаем три ограничения.

$$\begin{aligned}
&550x_н + 35x_р + 55x_о + 0.075x_б \geq 16,\\
&55x_н - 31.5x_р + 5.5x_о + 0.0075x_б \geq 0,\\
&110x_н + 7x_р - 44x_о + 0.015x_б \geq 0.\\
&x_б \leq 2,\\
&x_н, x_р, x_о, x_б \geq 0.
\end{aligned}$$

Каждое из этих неравенств представляет одну из целей городского совета, которую желательно добиться. Но эти цели могут конфликтовать друг с другом, и в лучшем случае мы можем попытаться достичь какого-нибудь компромиссного решения.

Способ, которым в целевом программировании достигается компромиссное решение, заключается в следующем. Сначала каждое неравенство преобразуется в более гибкую частную задачу, в рамках которой можно удовлетворить данное ограничение. В нашем случае эти частные задачи записываются так:

$$\begin{aligned}
&550x_н + 35x_р + 55x_о + 0.075x_б + s_1^+ - s_1^- = 16,\\
&55x_н - 31.5x_р + 5.5x_о + 0.0075x_б + s_2^+ - s_2^- = 0,\\
&110x_н + 7x_р - 44x_о + 0.015x_б + s_3^+ - s_3^- = 0.\\
&x_б + s_4^+ - s_4^- = 2,\\
&x_н, x_р, x_о, x_б \geq 0,\\
&s_i^+,\ s_i^- \geq 0,\ i = 1, 2, 3, 4.
\end{aligned}$$

Неотрицательные переменные s_i^+ и s_i^- называются **отклоняющими**, поскольку они показывают отклонение значений левых частей ограничений от соответствующих величин правых частей этих же ограничений.

Отклоняющие переменные s_i^+ и s_i^- зависимы по определению, поэтому они обе одновременно не могут быть базисными. Это означает, что, на любом этапе решения задачи одним из симплексных методов, только *одна* из пары отклоняющих переменных может принимать положительное значение. Если исходное i-е ограничение является неравенством типа "≤" и $s_i^+ > 0$, то это ограничение выполняется. Если же $s_i^- > 0$, то данное ограничение не выполняется. Таким образом, определенные значения отклоняющих переменных s_i^+ и s_i^- либо удовлетворяют i-е ограничение, либо нет. Это та гибкость, которая позволяет целевому программированию достичь компромиссного решения. Естественно, хорошее компромиссное решение минимизирует число невыполняемых ограничений.

В нашем примере первые три ограничения являются неравенствами типа "≥", а четвертое — неравенством типа "≤". Вследствие этого положительные значения отклоняющих переменных s_1^+, s_2^+, s_3^+ и s_4^- будут указывать на то, что соответствующие ограничения не выполняются. Поэтому ведется поиск такого компромиссного решения, которое будет удовлетворять по возможности большему числу следующих частных целей (целевых функций).

$$\text{Минимизировать } G_1 = s_1^+$$
$$\text{Минимизировать } G_2 = s_2^+$$
$$\text{Минимизировать } G_3 = s_3^+$$
$$\text{Минимизировать } G_4 = s_4^-$$

Как оптимизировать модель, имеющую несколько конфликтующих целевых функций? Для этого разработано множество разнообразных методов. В данной главе рассмотрим (1) метод весовых коэффициентов и (2) метод приоритетов. Оба метода основаны на приведении множества частных целевых функций к одной целевой функции.

Упражнения 8.2,а

1. Сформулируйте заново задачу о налогах из примера 8.2–1, предполагая, что городской совет определил дополнительное ограничение, которое заключается в том, что поступления от налога на бензин должны составлять не менее 1% от общих налоговых поступлений.

2. Руководство супермаркета планирует провести несколько специальных мероприятий для привлечения потенциальных покупателей. Два основных (наиболее популярных) мероприятия, — эстрадный концерт и выставка искусств и ремесел, — посещают практически все слои населения, которые менеджеры супермаркета условно разбивают по трем возрастным группам: тинэйджеры (подростковая группа), группа среднего возраста (сюда вошла и молодежь, вышедшая из подросткового возраста) и старшая возрастная группа. Стоимость одного концерта и одной выставки составляет \$1500 и \$3000 соответственно. Общий годовой бюджет этих мероприятий не должен превышать \$15 000. Менеджеры супермаркета оценивают посещаемость своих мероприятий следующим образом.

Мероприятие	Количество посетителей		
	Тинэйджеры	Средняя группа	Старшая группа
Концерт	200	100	0
Выставка искусств	0	400	250

Руководство супермаркета желает, чтобы их мероприятия посетило не менее 1000 подростков, 1200 людей из средней возрастной группы и не менее 800 потенциальных покупателей старшего возраста. Сформулируйте модель целевого программирования.

3. Университет Озарк проводит прием студентов. Всех поступающих можно разбить на три категории: жители данного штата, приезжие из других штатов и поступающие из других стран. Соотношения между мужчинами и женщинами среди поступающих данного штата и других штатов равны соответственно 1 : 1 и 3 : 2. Для "международных" студентов это соотношение составляет 8 : 1. При зачислении студентов на первый курс одним из основных факторов, влияющих на зачисление, является средний балл ACT (American College Test — Американский тест для поступающих в колледж). Статистика, накопленная университетом, свидетельствует, что средний балл ACT равен 27, 26 и 23 соответственно для поступающих данного штата, других штатов и других стран. Приемная комиссия университета при приеме первокурсников желает добиться следующего.

   a) На первый курс желательно принять не менее 1200 студентов.
   b) Средний балл ACT всех первокурсников должен быть не ниже 25.
   c) Студенты-неамериканцы должны составлять не менее 10% от всего количества первокурсников.
   d) Отношение количества женщин и мужчин желательно не менее 1 : 1.
   e) Среди всех принятых на первый курс жители других штатов должны составлять не менее 20%.

   Сформулируйте данную проблему как модель целевого программирования.

4. Птицефабрика ежедневно потребляет 3 тонны специальных кормов. Кормовая смесь состоит из известняка, зерна и соевой муки и должна удовлетворять требованиям рационального питания:

   кальций — не менее 0.8% и не более 1.2%,

   белок — не менее 22%,

   клетчатка — не более 5%.

   В следующей таблице приведен состав ингредиентов, составляющих кормовую смесь.

Ингредиент	Кальций	Белок	Клетчатка
	(в фунтах на фунт ингредиента)		
Известняк	0.380	0.00	0.00
Зерно	0.001	0.09	0.02
Соевая мука	0.002	0.50	0.08

Сформулируйте модель целевого программирования. Как вы думаете, стоит ли в данной ситуации применять модель целевого программирования?

5. Фабрика игрушек производит детские тележки, для которых на стадии конечной сборки необходимо четыре колеса и два сиденья. На фабрике работа организована в три смены, причем в течение одной рабочей смены производится несколько партий колес и сидений. В следующей таблице показаны объемы партий изделий в зависимости от рабочей смены.

	Количество изделий в партии	
Смена	Колеса	Сиденья
1	500	300
2	600	280
3	640	360

В идеале количество произведенных колес должно точно в два раза превышать количество произведенных сидений (напомним, что на каждую тележку идет четыре колеса и два сиденья). Но так как число произведенных изделий колеблется от смены к смене, невозможно обеспечить точный баланс между количествами произведенных колес и сидений. Фабрика планирует определить, какое количество партий изделий необходимо изготовлять каждую смену, чтобы свести к минимуму дисбаланс между произведенными колесами и сидениями. В первую смену можно произвести 4 или 5 партий изделий, во вторую — от 10 до 20 партий, а в третью — от 3 до 5. Сформулируйте задачу целевого программирования.

6. Завод продает четыре типа изделий, для производства которых используются токарный и сверлильный станки. Каждый из этих станков может работать 10 часов в рабочий день. В следующей таблице показано, сколько минут рабочего времени необходимо для изготовления изделия каждого типа.

Изделие	Токарный станок	Сверлильный станок
1	5	3
2	6	2
3	4	6
4	7	4

Завод пытается сбалансировать время использования станков таким образом, чтобы разность между полными временами работы станков не превышала 30 минут. Спрос на изделия каждого типа составляет не менее 10 единиц. Кроме того, количество изделий первого типа не может превышать количество изделий второго типа. Сформулируйте задачу целевого программирования.

7. Производство двух изделий требует двух последовательных операций. В следующей таблице показано время (в минутах) выполнения каждой операции при изготовлении изделий.

Операция	Изделие 1	Изделие 2
1	5	3
2	6	2

Ежедневная квота на производство первого и второго изделий составляет соответственно 80 и 60 единиц. На выполнение каждой операции отводится по 8 часов в рабочий день. Сверхурочные работы не желательны, хотя при необходимости, чтобы выполнить производственный план, их можно применять. Сформулируйте задачу целевого программирования.

8. Городская больница планирует организовать для больных кратковременный (до 4 дней) стационар на свободных местах. Предполагается, что в течение 4-дневного периода будет 30, 25 и 20 больных, которым потребуется 1-, 2- и 3-дневный стационар. Также рассчитано, что на тот же период будет свободно 20, 20, 24 и 30 мест соответственно. Используйте целевое программирование для нахождения максимального и минимального количества больных, которых можно принять на кратковременное лечение.

9. Семейство Траппов собирается переезжать в новый город, где оба супруга нашли новую работу. Сейчас они пытаются определить идеальное местоположение своего нового дома, составив список своих пожеланий:

   a) новый дом должен быть как можно ближе к месту работы м-ра Траппа (в пределах четверти мили),
   b) дом должен быть как можно дальше от шумного аэропорта (не ближе 10 миль),
   c) желательно, чтобы в пределах одной мили от дома был супермаркет.

   Для решения своей задачи супруги нанесли на карту города координатную сетку и пометили место работы, аэропорт и супермаркет. Получили следующие их координаты: место работы — (1, 1), аэропорт — (20, 15), супермаркет — (4, 7) (все расстояния приведены в милях). Сформулируйте задачу целевого программирования. (*Примечание.* Ограничения не обязательно должны быть линейными.)

10. *Регрессионный анализ.* Предположим, что в лабораторном эксперименте каждому значению y_i независимого параметра y соответствует n измерений x_{ij}, $i = 1, 2, ..., m$, $j = 1, 2, ..., n$, зависимой величины x. Требуется построить линию регрессии, проходящую через экспериментальные точки. Обозначим через b_j, $j = 0, 1, ..., n$, коэффициенты уравнения регрессии. Коэффициенты b_j находятся из условия минимума суммы абсолютных значений разностей наблюдаемых и расчетных (т.е. рассчитанных на основании уравнения регрессии) значений зависимой величины. Сформулируйте эту задачу в виде модели целевого программирования.

11. *Задача Чебышева.* Здесь, в отличие от предыдущей задачи, коэффициенты регрессии находятся из условия минимизации максимума абсолютных значений разностей наблюдаемых и расчетных значений зависимой величины. Сформулируйте задачу целевого программирования.

8.3. Алгоритмы целевого программирования

В этом разделе представлены два метода решения задач целевого программирования. Оба метода основаны на сведении множества частных целей к одной целевой функции. В **методе весовых коэффициентов** единственная целевая функция формируется как взвешенная сумма исходных частных целевых функций. В **методе приоритетов** на частные цели устанавливаются приоритеты в порядке их важности. Исходная задача решается путем последовательного решения ряда задач ЛП с одной целевой функцией таким образом, что решение задачи с низкоприоритетной целью не может "испортить" оптимального значения целевой функции с более высоким приоритетом.

Эти методы различны по своей природе и в общем случае дают оптимальные решения, не совпадающие между собой. Вместе с тем нельзя сказать, что один из этих методов лучше другого; в сущности, они предназначены для решения задач с разными предпочтениями в процессе принятия решений.

8.3.1. Метод весовых коэффициентов

Предположим, что модель целевого программирования имеет n целей, каждая из которых имеет следующий вид.

$$\text{Минимизировать } G_i, \ i = 1, 2, ..., n.$$

В методе весовых коэффициентов обобщенная целевая функция определяется следующим образом.

$$\text{Минимизировать } z = w_1G_1 + w_2G_2 + ... + w_nG_n.$$

Здесь w_i (i = 1, 2, ..., n) — положительные весовые коэффициенты, которые отображают предпочтения, отдаваемые каждой цели. Например, случай $w_i = 1$ для всех i говорит о равнозначности всех целей. Задание значений весовым коэффициентам очень субъективно. В настоящее время разработаны различные методы (см. [1]), которые уменьшают субъективный фактор при определении весовых коэффициентов.

Пример 8.3–1

Новое рекламное агентство, в составе которого 10 рекламных агентов, получило контракт на рекламу нового продукта. Агентство может провести рекламную акцию на радио и телевидении. В следующей таблице приведены данные о количестве людей, охватываемых тем или иным видом рекламы, стоимость этой рекламы и количество необходимых рекламных агентов. Все эти данные отнесены к одной минуте рекламного времени.

	Радио	Телевидение
Рекламная аудитория (млн чел.)	4	8
Стоимость (в тыс. долларов)	8	24
Количество рекламных агентов	1	2

Реклама на радио и телевидении должна охватить не менее 45 миллионов человек (так называемая рекламная аудитория), но контракт запрещает использовать более 6 минут рекламы на радио. Рекламное агентство может выделить на этот проект бюджет, не превышающий \$100 000. Сколько минут рекламного времени агентство должно купить на радио и сколько на телевидении?

Обозначим через x_1 и x_2 количество минут рекламного времени, закупленного соответственно на радио и телевидении. Для данной задачи целевого программирования можно задать следующие частные целевые функции.

Минимизировать $G_1 = s_1^+$ (для выполнения условия по рекламной аудитории),

Минимизировать $G_2 = s_2^-$ (для выполнения условия по бюджету)

при выполнении ограничений

$4x_1 + 8x_2 + s_1^+ - s_1^- = 45$ (условие по рекламной аудитории),

$8x_1 + 24x_2 + s_2^+ - s_2^- = 100$ (условия по бюджету),

$x_1 + 2x_2 \le 10$ (ограничение по рекламным агентам),

$x_1 \le 6$ (ограничение на рекламу по радио),

$x_1, x_2, s_1^+, s_1^-, s_2^+, s_2^- \ge 0.$

Менеджеры рекламного агентства считают, что выполнение условия по объему рекламной аудитории в два раза важнее, чем выполнение условия по бюджету. Поэтому обобщенная целевая функция будет записана следующим образом.

$$\text{Минимизировать } z = 2G_1 + G_2 = 2s_1^+ + s_2^-.$$

Оптимальное решение этой задачи, полученное с помощью программы TORA, следующее: $z = 10$, $x_1 = 5$ минут, $x_2 = 2.5$ минуты, $s_1^+ = 5$ миллионов человек. Остальные переменные равны нулю.

Тот факт, что оптимальное значение целевой функции не равно нулю, указывает, что, по крайней мере, одна из исходных целевых функций не достигла своего оптимального значения. Действительно, так как $s_1^+ = 5$, значит, объем рекламной аудитории меньше запланированного на 5 миллионов. При этом условие по бюджету выполнено, поскольку $s_2^- = 0$.

Еще раз повторим, что методы целевого программирования позволяют получить только *эффективное решение* задачи, которое не всегда будет оптимальным. Например, решение $x_1 = 6$ и $x_2 = 2$ дает такой же объем рекламной аудитории ($4 \times 6 + 8 \times 2 = 40$ миллионов человек), но при меньшей стоимости рекламной кампании ($8 \times 6 + 24 \times 2 =$ \$96 000). Существенно, что методы целевого программирования в общем случае не находят оптимум каждой целевой функции исходной модели. Этот "дефект" методов целевого программирования поднимает общий вопрос о "жизнеспособности" целевого программирования в качестве технологии оптимизации (дальнейшее обсуждение этого вопроса см. в примере 8.3–3).

Упражнения 8.3,а

1. Решите задачу из упр. 8.2,а(1) в предположении, что в обобщенную целевую функцию все частные целевые функции входят с одинаковыми весовыми коэффициентами. Будет ли в этом случае достигнут оптимум всех частных целевых функций?
2. Пусть в задаче упр. 8.2,а(2) привлечение посетителей средней возрастной группы в два раза важнее, чем привлечение посетителей других возрастных групп. Найдите соответствующее решение и проверьте, будет ли достигнут оптимум всех частных целевых функций.

3. Пусть в задаче о приеме студентов на первый курс университета (см. упр. 8.2,а(3)) ограничение, касающееся общего количества принятых первокурсников, должно быть выполнено обязательно, остальные ограничения не являются жесткими. Кроме того, считаем, что ограничение на средний балл АСТ в два раза важнее, чем остальные ограничения.

   a) Решите данную задачу и проверьте, все ли частные целевые функции достигнут своего оптимума.
   b) Пусть ограничение на количество принятых первокурсников не является жестким, но все равно имеет в два раза большую важность, чем ограничение на средний балл АСТ. Как в этом случае изменится решение?

4. Можно ли в условиях задачи из упр. 8.2,а(4) удовлетворить всем требованиям рационального питания?
5. Найдите решение задачи из упр. 8.2,а(5) и определите, можно ли сбалансировать ежедневное производство колес и сидений.
6. В задаче из упр. 8.2,а(6) выполнение ограничения, касающегося удовлетворения спроса на изделия, считается в два раза важнее, чем сбалансированность рабочего времени станков. Пусть также допустимы сверхурочные работы. Решите эту задачу и проверьте, все ли частные целевые функции достигнут своего оптимума.
7. В задаче из упр. 8.2,а(7) производственные квоты должны быть реализованы в любом случае; при необходимости можно использовать сверхурочные работы. Найдите решение этой задачи и определите объем сверхурочных работ, если они используются.
8. Пусть в задаче из упр. 8.2,а(8) все частные целевые функции в выражении обобщенной целевой функции имеют равные весовые коэффициенты. Могут ли все частные целевые функции достигнуть своего оптимума?
9. Отдел кадров компании составил следующую таблицу о своих пяти работниках, чтобы установить зависимость между заработком работника и его возрастом, образованием (выражается в количестве лет, проведенных в колледже) и опытом (количество лет работы в компании).

Возраст (года)	Образование (года)	Опыт (года)	Заработок ($)
30	4	5	40 000
39	5	10	48 000
44	2	14	38 000
48	0	18	36 000
37	3	9	41 000

   Используя формулировку задачи целевого программирования из упр. 8.2,а(10), постройте уравнение линейной регрессии $y = b_0 + b_1x_1 + b_2x_2 + b_3x_3$ на основе данных этой таблицы.

10. Решите предыдущую задачу, используя условие Чебышева из упр. 8.2,а(11).

8.3.2. Метод приоритетов

В методе приоритетов n частных целевых функций ранжируются в порядке их важности, так как их оценивает специалист по принятию решений, т.е.

Минимизировать $G_1 = \rho_1$ (наивысший приоритет),

. . .

Минимизировать $G_n = \rho_n$ (наинизший приоритет).

Переменные ρ_i — это компоненты отклоняющих переменных, т.е. s_i^+ или s_i^-, которые определяют i-ю целевую функцию. Например, в задаче из примера 8.3–1 $\rho_1 = s_1^+$ и $\rho_2 = s_2^-$.

В методе приоритетов поочередно решаются задачи с одной целевой функцией, начиная с задачи с целевой функцией, имеющей наивысший приоритет, и заканчивая задачей с целевой функцией, имеющей наинизший приоритет. *В процессе решения последовательных задач решение задачи с целевой функцией, имеющей более низкий приоритет, не может ухудшить полученные ранее решения задач с целевой функцией, имеющих более высокий приоритет.* Это означает, что если $z(G_i)$ — оптимальное значение целевой функции G_i, то для всех $i \geq 1$ оптимизация любой целевой функции G_j $(j > i)$ с меньшим приоритетом не может ухудшить значение $z(G_i)$.

В литературе по целевому программированию описан "специальный" симплекс-метод, который гарантирует неухудшаемость решений задач с целевыми функциями более высокого приоритета. Этот метод использует **правило исключения столбцов**, которое применяется для удаления из оптимальной симплекс-таблицы задачи с целевой функцией G_k *небазисной* переменной x_j с $z_j - c_j \neq 0$ до начала решения задачи с целевой функцией G_{k+1}. Это правило распознает, что небазисная переменная x_j, если она получит ненулевое значение, может ухудшить (но никогда не улучшит) оптимальное значение задачи с целевой функцией, имеющей более высокий приоритет.

К сожалению, этот метод изменения симплекс-таблиц требует более сложных вычислений, чем необходимо на самом деле. Мы покажем, что такого же результата можно достигнуть более простым способом.

Шаг 0. Определяем частные целевые функции задачи и ранжируем их в порядке приоритетов: $G_1 = \rho_1 \succ G_2 = \rho_2 \succ \ldots \succ G_n = \rho_n$. Положим $i = 1$.

Шаг i. Решаем i-ю задачу ЛП с целевой функцией G_i. Обозначим через ρ_i^* полученное оптимальное значение отклоняющей переменной ρ_i. Если $i = n$, вычисления заканчиваются, поскольку решена последняя n-я задача. В противном случае вводим в задачу новое ограничение $\rho_i = \rho_i^*$, тогда значение ρ_i не сможет измениться при решении последующих задач. Полагаем $i = i + 1$ и повторяем шаг i.

Последовательное введение дополнительных ограничений вида $\rho_i = \rho_i^*$ теоретически не так "элегантно", как *правило исключения столбцов*. Однако это приводит точно к такому же результату. Кроме того, обоснование введения дополнительных ограничений совершенно очевидно и понятно.

Определенным аргументом в пользу правила исключения столбцов может служить то, что при использовании этого правила происходит удаление переменной, т.е. уменьшается размерность задачи. В то же время описанная выше процедура увеличивает размерность задачи при добавлении новых ограничений. Но если повнимательнее присмотреться к этим ограничениям (вида $\rho_i = \rho_i^*$), то легко изменить вычисления таким образом, чтобы это ограничение учитывалось непосредственно путем подстановки значения ρ_i^* вместо переменной ρ_i, что также уменьшает количество переменных в задаче. (Можно также использовать метод решения задач ЛП с ограниченными переменными из раздела 7.5.2 для выполнения замены $\rho_i = \rho_i^*$, предполагая при этом ограничение вида $\rho_i \le \rho_i^*$.) С этой точки зрения правило исключения столбцов, если отвлечься от его теоретической привлекательности, не имеет особых вычислительных преимуществ по сравнению с описанной выше процедурой. Но ради объективности мы продемонстрируем, как работает правило исключения столбцов в примере 8.3–3. В следующем примере покажем использование описанной выше упрощенной процедуры решения задач с приоритетами.

Пример 8.3–2

Решим методом приоритетов задачу из примера 8.3–1. Предположим, что наибольший приоритет имеет частная целевая функция, соответствующая условию, накладываемому на объем рекламной аудитории.

Шаг 0. $G_1 \succ G_2$, где

G_1: Минимизировать s_1^+ (условие по рекламной аудитории),

G_2: Минимизировать s_2^- (условие по бюджету).

Шаг 1. Решаем первую задачу ЛП.

$$\text{Минимизировать } G_1 = s_1^+$$

при выполнении ограничений

$4x_1 + 8x_2 + s_1^+ - s_1^- = 45$ (условие по рекламной аудитории),

$8x_1 + 24x_2 + s_2^+ - s_2^- = 100$ (условие по бюджету),

$x_1 + 2x_2 \le 10$ (ограничение по рекламным агентам),

$x_1 \le 6$ (ограничение на рекламу по радио),

$x_1, x_2, s_1^+, s_1^-, s_2^+, s_2^- \ge 0.$

Оптимальное решение этой задачи (найденное с помощью программы TORA) составляет $x_1 = 5$ минут, $x_2 = 2.5$ минуты, $s_1^+ = 5$ миллионов человек, остальные переменные равны нулю. Решение показывает, что условие по объему рекламной аудитории не выполняется с дефицитом в 5 млн человек.

В этой задаче мы имеем $\rho_1 = s_1^+$. Поэтому в следующей задаче добавим ограничение $s_1^+ = 5$.

Шаг 2. Теперь необходимо решить вторую задачу ЛП.

$$\text{Минимизировать } G_2 = s_2^-$$

при выполнении тех же ограничений, что и в предыдущей задаче, плюс дополнительное ограничение $s_1^+ = 5$. Мы можем решить эту задачу, используя, как и на предыдущем шаге, программу TORA, и добавив с помощью опции MODIFY (Изменить) новое ограничение.

Но в данном случае в решении второй задачи нет необходимости, поскольку уже в решении первой имеем $s_2^- = 0$. Следовательно, решение первой задачи автоматически является оптимальным решением второй (можете проверить это утверждение с помощью программы TORA). Решение $s_2^- = 0$ показывает, что ограничение, касающееся бюджета рекламной компании, выполняется.

Дополнительное ограничение $s_1^+ = 5$ можно также учесть путем подстановки значения 5 вместо переменной s_1^+ в первое ограничение. В результате правая часть этого неравенства изменится со значения 45 на 40. Получим следующую задачу ЛП.

$$\text{Минимизировать } G_2 = s_2^-$$

при ограничениях

$$\begin{aligned}
&4x_1 + 8x_2 - s_1^- = 40,\\
&8x_1 + 24x_2 + s_2^+ - s_2^- = 100,\\
&x_1 + 2x_2 \le 10,\\
&x_1 \le 6,\\
&x_1, x_2, s_1^-, s_2^+, s_2^- \ge 0.
\end{aligned}$$

В новой формулировке этой задачи на одну переменную меньше, чем в первой задаче.

Теперь мы используем ту же задачу, чтобы показать, что наилучшее решение получается тогда, когда в методе приоритетов используется *оптимизация* "настоящих" целевых функций, а не тех целевых функций, которые строятся только для того, чтобы выполнялись определенные ограничения. Следующий пример также демонстрирует правило исключения столбцов при решении задач целевого программирования.

Пример 8.3–3

Цели, поставленные в задаче из примера 8.3–1, можно переформулировать следующим образом.

Цель 1. Максимизировать объем рекламной аудитории (P_1).

Цель 2. Минимизировать стоимость рекламной кампании (P_2).

Математически эти цели можно выразить с помощью следующих целевых функций.

$$\text{Максимизировать } P_1 = 4x_1 + 8x_2,$$
$$\text{Минимизировать } P_2 = 8x_1 + 24x_2.$$

Отдельные ограничения на желаемый объем рекламной аудитории и стоимость рекламной кампании, которые использовались в примерах 8.3–1 и 8.3–2, в данном случае излишни, поскольку для этих величин мы получим границы после решения соответствующих задач.

Получили новую задачу.

$$\text{Максимизировать } P_1 = 4x_1 + 8x_2,$$
$$\text{Минимизировать } P_2 = 8x_1 + 24x_2.$$

при ограничениях

$$x_1 + 2x_2 \le 10,$$
$$x_1 \le 6,$$
$$x_1, x_2 \ge 0.$$

Сначала решим эту задачу с помощью процедуры, описанной в примере 8.3–2.

Шаг 1. Решаем первую задачу ЛП.

$$\text{Максимизировать } P_1 = 4x_1 + 8x_2$$

при ограничениях

$$x_1 + 2x_2 \le 10,$$
$$x_1 \le 6,$$
$$x_1, x_2 \ge 0.$$

Оптимальное решение этой задачи (полученное с помощью программы TORA) составляет $x_1 = 0$, $x_2 = 5$ и $P_1 = 40$. Отсюда видно, что объем рекламной аудитории не может превысить 40 миллионов человек.

Шаг 2. Добавим ограничение $4x_1 + 8x_2 \ge 40$, которое гарантирует, что решение, полученное на предыдущем шаге, не будет ухудшено, и решаем следующую задачу ЛП.

$$\text{Минимизировать } P_2 = 8x_1 + 24x_2$$

при ограничениях

$$x_1 + 2x_2 \le 10,$$
$$x_1 \le 6,$$
$$4x_1 + 8x_2 \ge 40,$$
$$x_1, x_2 \ge 0.$$

Программа TORA дает следующее оптимальное решение этой задачи: $P_2 = \$96\,000$, $x_1 = 6$ минут и $x_2 = 2$ минуты. Мы получили тот же объем рекламной аудитории ($P_1 = 40$ млн чел.), но за меньшую стоимость. Это результат того, что здесь мы искали оптимальные значения соответствующих величин, а не просто удовлетворяли ограничениям, как в примере 8.3–2.

Теперь решим ту же задачу, используя правило исключения столбцов. Это правило применяется последовательно к строкам симплекс-таблицы, соответствующим частным целевым функциям.

Первая задача ЛП (Максимизация объема рекламной аудитории). При решении этой задачи симплекс-таблица содержит строки, соответствующие как целевой функции P_1, так и целевой функции P_2. Строка целевой функции P_2 пока играет пассивную роль, но будет изменена перед решением второй задачи ЛП.

Первая задача решается за две симплексные итерации, как показано в следующей таблице.

Базис	x_1	x_2	s_1	s_2	Решение
P_1	–4	–8	0	0	0
P_2	–8	–24	0	0	0
s_1	1	2	1	0	10
s_2	1	0	0	1	6
P_1	0	0	4	0	40
P_2	4	0	12	0	120
x_2	1/2	1	1/2	0	5
s_2	1	0	0	1	6

Нижняя часть этой таблицы показывает оптимальное решение $x_1 = 0$, $x_2 = 5$ и $P_1 = 40$ первой задачи.

Правило исключения столбцов применяется перед решением второй задачи для удаления из симплекс-таблицы с оптимальным решением первой задачи *небазисной* переменной x_j, для которой $z_j - c_j \neq 0$. Такие переменные, приняв положительные значения в задаче с более низким приоритетом, ухудшают решение задач с более высоким приоритетом.

Вторая задача ЛП (Минимизация стоимости рекламной кампании). Правило исключения столбцов удаляет переменную s_1, для которой $z_1 - c_1 = 4$. Из P_2-строки приведенной выше симплекс-таблицы видно, что если не удалить переменную s_1, то на первой итерации решения второй задачи она должна войти в базис, при этом из базиса будет исключена переменная x_2. После этого будет получено оптимальное решение второй задачи (в этом решении $x_1 = x_2 = 0$), которое ухудшает оптимальное решение первой, поскольку теперь $P_1 = 0$ вместо $P_1 = 40$, как было ранее.

В данном случае вторая задача ЛП является задачей минимизации. После удаления переменной s_1 в базис вводится небазисная переменная x_1 со значением разности $z_j - c_j$, равным 4 (> 0), что может улучшить значение целевой функции P_2. В следующей симплекс-таблице показаны две итерации решения второй задачи ЛП. P_1-строку можно удалить из этой таблицы, так как она не участвует в процессе нахождения оптимального решения задачи с целевой функцией P_2.

Итерация	Базис	x_1	x_2	s_1	s_2	Решение
1	P_1					40
	P_2	4	0		0	120
	x_2	1/2	1		0	5
	s_2	1	0		1	6
2	P_1					40
	P_2	0	0		–4	96
	x_2	0	1		–1/2	2
	x_1	1	0		1	6

Полученное здесь оптимальное решение ($x_1 = 6$, $x_2 = 2$) со значениями целевых функций $P_1 = 40$ и $P_2 = 96$ такое же, как и полученное ранее.

Упражнения 8.3,b

1. Пусть в задаче из примера 8.3–2 бюджет рекламной компании возрос до $110 000. Желаемый объем рекламной аудитории остался неизменным — 45 млн чел. Найдите решение данной задачи методом приоритетов.
2. Решите задачу из упр. 8.2,a(1), используя следующую иерархию приоритетов частных целевых функций: $G_1 \succ G_2 \succ G_3 \succ G_4$.
3. Вернитесь к задаче из упр. 8.2,a(2). Пусть частные целевые функции, описывающие количество привлеченных посетителей подросткового, среднего и старшего возрастов, обозначаются соответственно G_1, G_2 и G_3. Решите эту задачу при следующих иерархиях приоритетов частных целевых функций.

   a) $G_1 \succ G_2 \succ G_3$
   b) $G_3 \succ G_2 \succ G_1$
4. С помощью метода приоритетов решите задачу из упр. 8.2,a(3) в предположении, что порядок приоритетов частных целевых функций совпадает с порядком их описания в задаче.

8.4. Заключение

Задачи целевого программирования — это задачи принятия решений, имеющих несколько, возможно противоречивых, целевых функций. Основное назначение методов целевого программирования заключается в сведении задачи с несколькими целевыми функциями к одной или нескольким задачам ЛП с одной целевой функцией. Качество конечного решения задачи целевого программирования определяется способом ранжирования частных целевых функций по степени их важности, а также степенью гибкости множества ограничений задачи. Поэтому методы целевого программирования дают *эффективное решение*, которое только пытается учесть все частные целевые функции, но, как правило, не совпадает с их оптимальными решениями.

Литература

1. Cohon T.L. *Multiobjective Programming and Planning*. Academic Press, New York, 1978.
2. Ignizio J.P., Cavalier T.M. *Linear Programming*. Prentice-Hall, Upper Saddle River, N.J., 1994.
3. Steuer R.E. *Multiple Criteria Optimization: Theory, Computations, and Application*, Wiley, New York, 1986.

Литература, добавленная при переводе

Поляк Б.Т. *Введение в оптимизацию*. — М.: Наука, 1983.

Поспелов Г.С., Ириков В.А. *Программно-целевое планирование и управление*. — М.: Сов. радио, 1976.

Соболь И.М., Статников Р.Б. *Выбор оптимальных параметров в задачах со многими критериями*. — М.: Наука, 1981.

Сухарев А.Г., Тимохов А.В., Федоров В.В. *Курс методов оптимизации*. — М.: Наука, 1986.

Комплексные задачи

■ **8-1.**[1] Лесозаготовительная компания использует три участка леса площадью 100 000, 180 000 и 200 000 акров для заготовки древесины и последующего облесения вырубок. Древесная продукция компании разбита на три основные категории: пиломатериал, клееная фанера и древесная измельченная масса. Для каждого участка лесного массива возможны различные альтернативные варианты его эксплуатации, которые различаются стоимостью разработок, арендной платой, периодом чередования вырубок и лесопосадок, объемом произведенной продукции. Эти альтернативы показаны в следующей таблице.

Участок	Альтернатива	\$ в год/акр		Период чередования (годы)	$м^3$ в год/акр		
		Стоимость	Аренда		Пиломатериалы	Фанера	Древесная масса
1	A1	1000	160	20	12	0	0
	A2	800	117	25	10	0	0
	A3	1500	140	40	5	6	0
	A4	1200	195	15	4	7	0
	A5	1300	182	40	3	0	7
	A6	1200	180	40	2	0	6
	A7	1500	135	50	3	0	5
2	A1	1000	102	20	9	0	0
	A2	800	55	25	8	0	0
	A3	1500	95	40	2	5	0
	A4	1200	120	15	3	4	0
	A5	1300	100	40	2	0	5
	A6	1200	90	40	2	0	4
3	A1	1000	60	20	7	0	0
	A2	800	48	25	6	4	0
	A3	1500	60	40	2	0	4
	A4	1200	65	15	2	0	3
	A5	1300	35	40	1	0	5

Для гарантии будущего производства необходимо, чтобы на каждый акр леса, выведенный из использования, приходилось столько же выведенных из использования акров леса, каков период чередования. Арендная плата — это собираемая плата на порубку леса.

Компания преследует следующие цели.

1. Ежегодное производство пиломатериалов, фанеры и древесной массы должно быть не менее 200 000, 150 000 и 350 000 кубических метров соответственно.
2. Ежегодный бюджет на восстановление леса составляет 2.5 млн долларов.
3. Ежегодная арендная плата должна составлять \$100 за акр.

Какой вариант эксплуатации для каждого участка леса следует выбрать?

---

[1] Задача основана на материалах статьи Rustagi K.P. *Forest Management Planning for Timber Prodaction: A Goal Programming Approach*, Bulletin No. 89, Yale University, New Haven, 1976.

■ **8-2.** Благотворительная организация помогает детскому приюту. Для этого она привлекает волонтеров ежедневно с 8 до 14 часов. Волонтеры могут начать работу в любое время с 8 до 11 часов. Длительность работы волонтера не более 6 и не менее 2 часов. С 12 до 13 часов обеденное время. Благотворительная организация считает, что в приюте ежедневно необходимо 15 волонтеров для работы с 8 до 9, 16 волонтеров — с 9 до 10, 18 волонтеров — с 10 до 11, 20 волонтеров — с 11 до 12 и 16 волонтеров — с 12 до 14 часов (но это работа на один час, так как с 12 до 13 обеденное время). Здесь частными целевыми функциями могут быть волонтеры (т.е. их количество), работающие в течение каждого часа с 8 до 14. Цель всей задачи — определить количество волонтеров, приступающих к работе в начале каждого часа, и длительность их работы. Сформулируйте и решите эту задачу как задачу целевого программирования.

Целочисленное линейное программирование

9.1. Введение

Целочисленное линейное программирование (ЦЛП) ориентировано на решение задач линейного программирования, в которых все или некоторые переменные должны принимать целочисленные (или дискретные) значения. Несмотря на интенсивные исследования, проводимые на протяжении последних десятилетий, известные вычислительные методы решения задач ЦЛП далеки от совершенства. На сегодня не существует надежных вычислительных алгоритмов решения таких задач.

В настоящей главе рассматриваются сначала примеры задач целочисленного программирования, а затем методы их решения.

9.2. Примеры задач целочисленного программирования

Рассмотрим сначала простые практические задачи, которые сводятся к задачам ЦЛП, затем обратимся к более сложным. Для удобства задачи, в которых все переменные должны быть целочисленными, называются **полностью целочисленными**, а задачи, в которых лишь некоторые переменные должны принимать целочисленные значения, — **частично-целочисленными**.

Пример 9.2–1. (Распределение капиталовложений)

Оценивается пять проектов с точки зрения их возможного финансирования на предстоящий трехлетний период. Следующая таблица содержит ожидаемую прибыль от реализации каждого проекта и распределение необходимых капиталовложений по годам.

Проект	Расходы (млн долл./год)			Прибыль (млн долл.)
	1-й год	2-й год	3-й год	
1	5	1	8	20
2	4	7	10	40
3	3	9	2	20
4	7	4	1	15
5	8	6	10	30
Доступный капитал (млн долл.)	25	25	25	

Предполагается, что каждый утвержденный проект будет реализован за трехлетний период. Необходимо определить совокупность проектов, которой соответствует максимум суммарной прибыли.

Задача сводится к решению типа "да–нет" относительно каждого проекта. Определим двоичные переменные x_j:

$$x_j = \begin{cases} 1, & \text{если проект } j \text{ утвержден,} \\ 0, & \text{если проект } j \text{ не утвержден.} \end{cases}$$

Задача ЦЛП будет записана следующим образом.

$$\text{Максимизировать } z = 20x_1 + 40x_2 + 20x_3 + 15x_4 + 30x_5$$

при ограничениях

$$\begin{aligned} &5x_1 + 4x_2 + 3x_3 + 7x_4 + 8x_5 \le 25, \\ &x_1 + 7x_2 + 9x_3 + 4x_4 + 6x_5 \le 25, \\ &8x_1 + 10x_2 + 2x_3 + x_4 + 10x_5 \le 25, \\ &x_1, x_2, x_3, x_4, x_5 = 0 \text{ или } 1. \end{aligned}$$

Оптимальным целочисленным решением (полученным с помощью программы TORA) является $x_1 = x_2 = x_3 = x_4 = 1$, $x_5 = 0$ с $z = 95$ (млн долл.). Это решение означает, что необходимо выбрать для финансирования все проекты, кроме пятого.

Интересно сравнить решение данной задачи ЦЛП с решением "обычной" задачи ЛП с теми же ограничениями, но без условия целочисленности переменных. Задача линейного программирования получается в результате замены условия $x_j = 0$ или 1 на ограничение $0 \le x_j \le 1$ для всех j. Эта задача имеет решение $x_1 = 0.5789$, $x_2 = x_3 = x_4 = 1$, $x_5 = 0.7368$ и $z = 108.68$ (млн долл.). Такое решение с точки зрения целочисленной задачи лишено смысла, так как две переменные принимают дробное значение. Можно было бы попытаться округлить полученный результат, что привело бы к $x_1 = x_5 = 1$. Полученное при этом решение является недопустимым, так как нарушаются ограничения задачи. Более существенным в этой ситуации является то, что *округление применять нельзя*, так как x_j представляет решение типа "да–нет", для которого дробные значения лишены смысла.

Упражнения 9.2,а[1]

1. В модели распределения капиталовложений из примера 9.2–1 предположим, что проект 5 должен быть обязательно выбран, если выбирается либо проект 1, либо 3. Измените математическую модель, включив в нее новое ограничение, и решите полученную задачу с помощью программы TORA.
2. Рассмотрите задачу о загрузке самолета грузами пяти типов. Вес w_i, объем v_i, а также стоимость r_i единицы груза каждого типа приведены в следующей таблице.

[1] Задачи 3–6 в адаптированном виде заимствованы из книги Malba Tahan. *El Hombre Que Calculaba*, Editorial Limusa, Mexico City, 1994, pp. 39–182.

Груз i	Вес единицы груза, w_i (тонны)	Объем единицы груза, v_i (куб. ярд)	Стоимость единицы груза, r_i (в $100)
1	5	1	4
2	8	8	7
3	3	6	6
4	2	5	5
5	7	4	4

Максимальная грузоподъемность и объем самолета равны 112 тонн и 109 куб. ярдов соответственно. Сформулируйте в виде модели ЦЛП задачу определения набора грузов, обеспечивающего максимальную стоимость груза, и найдите решение с помощью программы TORA.

3. Пусть есть 7 полных бутылок вина, 7 бутылок, заполненных наполовину, и 7 пустых бутылок. Необходимо распределить 21 бутылку между тремя персонами так, чтобы каждый получил 7 бутылок. В то же время каждый должен получить одинаковое количество вина. Сформулируйте эту задачу в виде задачи ЦЛП с ограничениями в виде равенств и найдите решение, используя программу TORA. (*Совет.* Используйте фиктивную целевую функцию, все коэффициенты которой равны нулю.)

4. Эксцентричный шейх оставил завещание относительно распределения стада верблюдов между тремя детьми: Тарик получает не менее половины стада, Шариф – не менее одной трети, а Майса — по меньшей мере одну девятую часть. Остаток завещался благотворительной организации. В завещании не упоминался размер стада, говорилось лишь, что количество верблюдов — число нечетное и благотворительная организация получает в точности одного верблюда. Сколько верблюдов оставил шейх и сколько получил каждый из его детей?

5. Супружеская пара фермеров посылает трех своих сыновей на базар продать 90 яблок, для того чтобы обучить их числам и обращению с деньгами. Самый старший Джим получил для продажи 50 яблок, Билл (средний) — 30 и самый младший Джон — лишь 10. Родители поставили пять условий. 1) Цена яблок должна быть равна либо $1 за 7 яблок, либо $3 за 1 яблоко. 2) Каждый ребенок может использовать один или оба варианта цен. 3) Все дети должны вернуться с одинаковой суммой денег. 4) Каждый ребенок приносит домой сумму, которая является четным числом (без центов). 5) Сумма денег, полученная каждым из детей, должна быть максимальной при сформулированных условиях. Считается, что дети могут продать все яблоки, которые они имеют. Как дети могут выполнить требования своих родителей?

6. Когда-то был капитан торгового судна, который хотел наградить трех членов команды за их героические усилия по спасению груза корабля во время неожиданного шторма. Капитан взял некоторую сумму денег у казначея и отдал приказ старшему помощнику распределить их поровну между тремя матросами после того, как корабль достигнет берега. Однажды ночью один из матросов решил взять свою (справедливую) третью часть заранее. После деления денег на три равные части осталась одна монета, которую матрос решил оставить себе (в дополнение к третьей части денег). На следующую ночь второй матрос решил осуществить такой же план и, повторив деление на три части оставшейся суммы, присвоил себе

еще и монету, которая осталась после деления. На третью ночь третий матрос взял третью часть того, что осталось, и одну дополнительную монету, которая осталась после деления. Когда корабль достиг берега, старший помощник капитана разделил остаток денег поровну между тремя матросами и снова осталась одна монета. Старший помощник отложил эту монету в сторону и вручил матросам предназначенные им равные части. Сколько денег было в самом начале? Сформулируйте задачу в виде задачи ЦЛП и найдите оптимальное решение, используя программу TORA. (*Совет.* Задача имеет бесконечно много целочисленных решений. Предположите для удобства, что следует определить минимальную сумму денег, которая удовлетворяет условиям задачи. Увеличивая затем полученное решение на 1, рассмотрите его как нижнюю границу и получите новое минимальное решение. Продолжая таким образом, можно наметить шаблон общего решения.)

7. Имеются следующие слова, состоящие из трех букв: AFT, FAR, TVA, ADV, JOE, FIN, OSF и KEN. Предположим, что буквам алфавита приписаны числа (метки), начиная с $A = 1$ и заканчивая $Z = 26$. Каждое слово помечается числом, равным сумме числовых меток составляющих его трех букв. Например, слово AFT имеет метку $1 + 6 + 20 = 27$. Необходимо выбрать пять из заданных восьми слов таким образом, чтобы получить максимальную суммарную метку. Вместе с тем выбранные пять слов должны удовлетворять следующему условию: (сумма меток первых букв) < (сумма меток вторых букв) < (сумма меток третьих букв).

   Сформулируйте задачу в виде задачи ЦЛП и найдите ее оптимальное решение, используя программу TORA.

8. Фирма, специализирующаяся на грамзаписи песен, заключила договор с восходящей звездой эстрады на запись восьми песен. Продолжительность песен равна 8, 3, 5, 5, 9, 6, 7 и 12 минут соответственно. Фирма планирует использовать для записи двусторонние кассеты. Каждая сторона имеет длительность звучания 30 минут. Фирма намерена распределить песни на две стороны кассеты сбалансированным образом. Это значит, что продолжительность звучания песен на каждой стороне кассеты должна быть примерно одинаковой (насколько это возможно). Сформулируйте задачу в виде задачи ЦЛП и найдите оптимальное решение.

9. Рассмотрите задачу из предыдущего упражнения в предположении, что характер мелодий песен диктует условия, согласно которым третья и четвертая песни не могут быть записаны на одной стороне кассеты. Сформулируйте задачу в виде задачи ЦЛП. Возможно ли использование 25-минутной (для каждой стороны) кассеты для записи 8 песен? Если нет, используйте модель ЦЛП, чтобы определить минимальную емкость кассеты для записи.

Пример 9.2–2. (Задача с постоянными затратами)

Три телефонные компании предложили мне подписаться на их услуги по дальней связи в пределах Соединенных Штатов. Услуги компании MaBell стоят 16 долларов в месяц плюс 0.25 доллара за каждую минуту связи. Компания PaBell оценивает свои услуги в 25 долларов в месяц, но имеет поминутную оплату в 0.21 доллара. Что касается компании BabyBell, месячная плата равна 18 долларов, а стоимость минуты связи — 0.22 доллара. Мои телефонные звонки на дальние расстояния в среднем составляют обычно 200 минут в месяц. Предполагается, что я не вношу помесячной платы, если не использую телефон для звонков на дальние расстояния, и что я могу распределять звонки между тремя компаниями, как мне хочется. Как я должен использовать три компании, чтобы минимизировать свой месячный телефонный счет?

Эта задача может быть легко решена без использования методов ЦЛП. Тем не менее поучительно сформулировать ее как целочисленную задачу.

Пусть

x_1 — количество минут разговора (на дальние расстояния) в месяц через компанию MaBell,

x_2 — количество минут разговора в месяц через компанию PaBell,

x_3 — количество минут разговора в месяц через компанию BabyBell,

$y_1 = 1$, если $x_1 > 0$, и $y_1 = 0$ при $x_1 = 0$,
$y_2 = 1$, если $x_2 > 0$, и $y_2 = 0$ при $x_2 = 0$,
$y_3 = 1$, если $x_3 > 0$, и $y_3 = 0$ при $x_3 = 0$.

Для обеспечения равенства $y_j = 1$ при положительном значении переменной x_j, используем ограничение $x_j \le My_j$, $j = 1, 2, 3$, где M — достаточно большое число, которое не должно ограничивать величину x_j. Так как звонки на дальние расстояния занимают около 200 минут в месяц, т.е. $x_j \le 200$ для всех j, поэтому достаточно положить $M = 200$.

Теперь можно сформулировать следующую задачу.

$$\text{Минимизировать } z = 0.25x_1 + 0.21x_2 + 0.22x_3 + 16y_1 + 25y_2 + 18y_3$$

при ограничениях

$$\begin{aligned}
&x_1 + x_2 + x_3 = 200,\\
&x_1 \le 200y_1,\\
&x_2 \le 200y_2,\\
&x_3 \le 200y_3,\\
&x_1, x_2, x_3 \ge 0,\\
&y_1, y_2, y_3 = 0 \text{ или } 1.
\end{aligned}$$

Формулировка задачи показывает, что помесячная j-я оплата за пользование телефоном является частью целевой функции z лишь при условии $y_j = 1$, которое по определению может выполняться лишь тогда, когда $x_j > 0$. Если в оптимальном решении будет $x_j = 0$, то минимума функции z (с учетом положительности коэффициента при y_j) можно достигнуть только при равенстве нулю y_j, что и требуется.

Оптимальным решением данной задачи является $x_3 = 200$, $y_3 = 1$, все остальные переменные равны нулю. Это значит, что копания BabyBell должна быть выбрана для услуг по дальней связи. Следует подчеркнуть, что информация, которую несет значение переменной $y_3 = 1$, является избыточной, так как такой же результат следует из $x_3 > 0$ ($= 200$). В самом деле, основной причиной использования переменных y_1, y_2 и y_3 является лишь учет месячной платы за телефон. В сущности, только эти двоичные переменные преобразуют исходную нелинейную задачу в частично-целочисленную, поддающуюся аналитическому решению.

Изложенная концепция "твердого гонорара" является типичной для задачи, известной в литературе как **задача с постоянными затратами**.

Упражнения 9.2,b

1. Компания планирует производство некоторой продукции на трех станках, которой должно быть изготовлено не менее 2000 единиц. Минимальная производительность любого станка равна 500 единиц. Следующая таблица содержит необходимую информацию для рассматриваемой задачи.

Станок	Расходы на переналадку	Затраты на производство единицы продукции	Производительность (в единицах продукции)
1	300	2	600
2	100	10	800
3	200	5	1200

Сформулируйте задачу в виде задачи ЦЛП и найдите ее оптимальное решение, используя программу TORA.

2. Нефтедобывающая компания изучает вопрос о бурении четырех нефтяных скважин на двух нефтеносных участках. Издержки подготовки к бурению на каждом участке и затраты на бурение скважины j на участке i (i=1, 2; j=1, 2, 3, 4) приведены в следующей таблице.

	Затраты на бурение скважины (млн долл.)				Издержки подготовки к бурению (млн долл.)
Участок	1	2	3	4	
1	2	1	8	5	5
2	4	6	3	1	6

Сформулируйте задачу в виде задачи ЦЛП и найдите ее оптимальное решение, используя программу TORA.

3. Рассматриваются три промышленных участка для размещения обрабатывающих заводов, которые снабжают своей продукцией трех потребителей. Объемы производства продукции, объемы потребления и себестоимость перевозки продукции от заводов к потребителям содержатся в следующей таблице.

	1	2	3	Объем производства
1	\$10	\$15	\$12	**1800**
2	\$17	\$14	\$20	**1400**
3	\$15	\$10	\$11	**1300**
Спрос	**1200**	**1700**	**1600**	

В дополнение к транспортным, имеются еще фиксированные затраты в объеме 10 000, 15 000 и 12 000 долларов, связанные с 1, 2 и 3 заводом соответственно. Сформулируйте задачу в виде задачи ЦЛП и найдите ее оптимальное решение, используя программу TORA.

Пример 9.2–3. (Задача коммивояжера)

Дневной график работы предприятия, производящего краски, включает изготовление партий белой (Б), желтой (Ж), красной (К) и черной (Ч) красок. Так как предприятие использует одно и то же оборудование для изготовления всех четырех типов краски, необходима его надлежащая чистка между изготовлением различных партий краски. Приведенная ниже таблица содержит время чистки оборудования в минутах, если за изготовлением краски, указанной в строке, следует изготовление краски, указанной в столбце. Например, если за белой краской следует желтая, то время чистки равно 10 минут. Так как за определенным цветом краски не может следовать такой же цвет, то соответствующее время чистки полагается равным бесконечности. Необходимо определить оптимальную последовательность дневного производства четырех типов краски, которая минимизирует суммарное время чистки оборудования.

	Время чистки в мин., если следующий цвет			
Текущий цвет	Белый	Желтый	Черный	Красный
Белый	∞	10	17	15
Желтый	20	∞	19	18
Черный	50	44	∞	25
Красный	45	40	20	∞

На рис. 9.1 графически представлены условия данной задачи. Каждый цвет краски представлен узлом сети, числа возле дуг, соединяющих узлы, соответствуют времени чистки. Задача таким образом сводится к определению *кратчайшего контура*, который начинается в одном узле, проходит через оставшиеся три узла в точности по одному разу, перед тем как возвратиться в начальный узел. Задачи такого типа известны под общим названием **задача коммивояжера**, в "классической" формулировке которой коммивояжер пытается определить *кратчайший маршрут* для посещения *n* городов, посещая каждый из них только один раз.

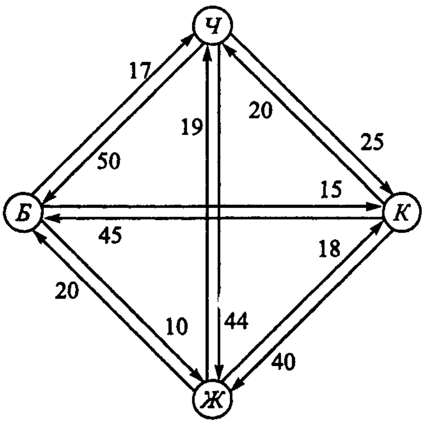


Рис. 9.1

Эту задачу можно решить путем полного перебора шести [(4–1)! = 3! = 6] возможных контуров сети. Приведенная ниже таблица показывает, что последовательность узлов Б→Ж→К→Ч→Б определяет оптимальный контур.

Производственный цикл (контур)	Суммарное время чистки
Б→Ж→Ч→К→Б	10 + 19 + 25 + 45 = 99
Б→Ж→К→Ч→Б	**10 + 18 + 20 + 50 = 98**
Б→Ч→Ж→К→Б	17 + 44 + 18 + 45 = 124
Б→Ч→К→Ж→Б	17 + 25 + 40 + 20 = 102
Б→К→Ч→Ж→Б	15 + 20 + 44 + 20 = 99
Б→К→Ж→Ч→Б	15 + 40 + 19 + 50 = 124

Метод полного перебора контуров практически применим лишь для задач малой размерности. Например, сеть с 11 узлами имеет 10! = 3 628 800 контуров. Следовательно, требуется более эффективный подход к решению таких задач.

Определим переменные x_{ij} равными 1, если узел j достигается с узла i, и 0 в противном случае. Необходимым условием для контура (маршрута) является то, что узел i связывается лишь с одним узлом и узел j достигается лишь из одного узла. Считая M достаточно большим положительным числом, мы можем записать задачу предприятия, производящего краски, следующим образом.

Минимизировать $z = Mx_{ББ} + 10x_{БЖ} + 17x_{БЧ} + 15x_{БК} + 20x_{ЖБ} + Mx_{ЖЖ} + 19x_{ЖЧ} + 18x_{ЖК} + 50x_{ЧБ} + 44x_{ЧЖ} + Mx_{ЧЧ} + 25x_{ЧК} + 45x_{КБ} + 40x_{КЖ} + 20x_{КЧ} + Mx_{КК}$

при ограничениях

$$
\begin{aligned}
&x_{ББ} + x_{БЖ} + x_{БЧ} + x_{БК} = 1,\\
&x_{ЖБ} + x_{ЖЖ} + x_{ЖЧ} + x_{ЖК} = 1,\\
&x_{ЧБ} + x_{ЧЖ} + x_{ЧЧ} + x_{ЧК} = 1,\\
&x_{КБ} + x_{КЖ} + x_{КЧ} + x_{КК} = 1,\\
&x_{ББ} + x_{ЖБ} + x_{ЧБ} + x_{КБ} = 1,\\
&x_{БЖ} + x_{ЖЖ} + x_{ЧЖ} + x_{КЖ} = 1,\\
&x_{БЧ} + x_{ЖЧ} + x_{ЧЧ} + x_{КЧ} = 1,\\
&x_{БК} + x_{ЖК} + x_{ЧК} + x_{КК} = 1,\\
&x_{ij} = 0 \text{ или } 1 \text{ для всех } i \text{ и } j,\\
&\text{решение должно быть контуром (маршрутом).}
\end{aligned}
$$

Как видим, формулировка задачи совпадает с таковой для задачи об оптимальных назначениях, за исключением требования, что решение должно образовывать маршрут. В общем случае нет гарантии, что оптимальное решение задачи коммивояжера может быть получено путем решения соответствующей задачи о назначениях, т.е., что оптимальное решение задачи о назначениях образует маршрут. Более вероятно, что оно будет содержать неполные маршруты, соединяющие вместе некоторые подмножества узлов сети. Поэтому были разработаны точные алгоритмы решения задачи коммивояжера, основу которых составляет алгоритм решения задачи о назначениях. Эти алгоритмы отличаются друг от друга сложностью и вычислительной эффективностью (см., например, [4]).

Упражнения 9.2,c

1. Менеджер проектов имеет 10 сотрудников, которые работают над шестью проектами, причем каждый работает одновременно над несколькими проектами, как показано в следующей таблице.

	Проект 1	2	3	4	5	6
1		X		X	X	
2	X		X		X	
3		X	X	X		X
4			X	X	X	
5	X	X	X			
6	X	X	X	X		X
7	X	X			X	X
8	X		X	X		
9					X	X
10	X	X		X	X	X

Менеджер должен встретиться с каждым из 10 сотрудников один раз в неделю для обсуждения их проблем. Беседа с каждым из них длится примерно 20 минут, т.е. на разговоры со всеми сотрудниками уходит 3 часа 20 минут. Предлагается проводить встречи менеджера с группами сотрудников, работающих над одним и тем же проектом, чтобы уменьшить суммарное время, затрачиваемое на совещания. Менеджер планирует составить график обсуждения проектов таким образом, чтобы уменьшить движение в офисе, т.е. сократить число сотрудников, входящих и выходящих из комнаты для совещаний. В какой последовательности необходимо рассматривать проекты?

2. Продавец книг, проживающий в городе А, один раз в месяц должен посетить своих четырех клиентов, которые проживают в городах Б, В, Г и Д соответственно. Приведенная ниже таблица содержит расстояния в милях между различными городами.

	А	Б	В	Г	Д
А	0	120	220	150	210
Б	120	0	80	110	130
В	220	110	0	160	185
Г	150	110	160	0	190
Д	210	130	185	190	0

Необходимо составить маршрут движения продавца книг, минимизирующий суммарное пройденное расстояние. Сформулируйте задачу в виде задачи ЦЛП.

3. Пусть в предыдущем упражнении города Б, В, Г и Д обозначены через 1, 2, 3 и 4, а город А разделен на два, которые обозначены через 0 и 5. Пусть маршрут продавца книг начинается в городе 0, проходит (в некотором порядке) через города 1, 2, 3 и 4 и заканчивается в городе 5. Определим переменную v_i для города i ($i = 0, 1, ..., 5$), ко-

торая удовлетворяет условиям $0 \le v_i \le 5$. Пусть $x_{ij} = 1$, если в маршруте город j следует за городом i, и 0 в противном случае. Покажите, что ограничения

$$v_i - v_j + 5x_{ij} \le 4,\ \ i = 0, 1, ..., 4,\ \ j = 1, 2, ..., 5,$$

гарантируют исключение всех неполных маршрутов в решении задачи.

Пример 9.2–4. (Задача о покрытии)

Для обеспечения безопасности студентов отдел безопасности американского университета устанавливает телефоны экстренного вызова на территории студенческого городка. Отделу желательно установить минимальное количество телефонов таким образом, чтобы на каждой из основных улиц этого городка был расположен по крайней мере один телефон. На рис. 9.2 представлены основные улицы (от A до K) студенческого городка.

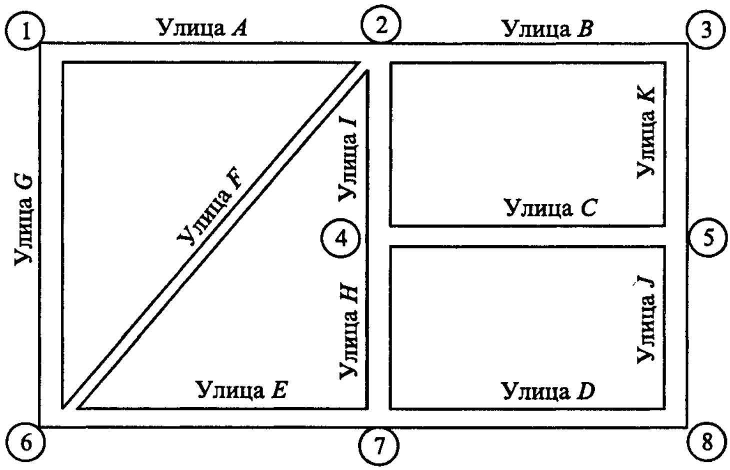


Рис. 9.2

Логично расположить телефоны на пересечениях улиц так, чтобы каждый телефон мог обслуживать по меньшей мере две улицы. Из рис. 9.2 видно, что данное расположение улиц требует не более восьми телефонов.

Определим, что переменная x_j равна 1, если телефон расположен на перекрестке j ($j = 1, 2, ..., 8$), и 0 в противном случае.

Условия задачи требуют установки по меньшей мере одного телефона на каждой из 11 улиц (от A до K). Поэтому задачу можно сформулировать следующим образом.

$$\text{Минимизировать } z = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8$$

при ограничениях

$$x_1 + x_2 \ge 1 \text{ (улица } A\text{)},$$
$$x_2 + x_3 \ge 1 \text{ (улица } B\text{)},$$
$$x_4 + x_5 \ge 1 \text{ (улица } C\text{)},$$

$$x_7 + x_8 \geq 1 \text{ (улица } D\text{)},$$
$$x_6 + x_7 \geq 1 \text{ (улица } E\text{)},$$
$$x_2 + x_6 \geq 1 \text{ (улица } F\text{)},$$
$$x_1 + x_6 \geq 1 \text{ (улица } G\text{)},$$
$$x_4 + x_7 \geq 1 \text{ (улица } H\text{)},$$
$$x_2 + x_4 \geq 1 \text{ (улица } I\text{)},$$
$$x_5 + x_8 \geq 1 \text{ (улица } J\text{)},$$
$$x_3 + x_5 \geq 1 \text{ (улица } K\text{)},$$
$$x_j = 0 \text{ или } 1, j = 1, 2, ..., 8.$$

В соответствии с оптимальным решением задачи (полученным с помощью программы TORA) необходимо установить телефоны на перекрестках 1, 2, 5 и 7.

Рассмотренная выше модель является типичным представителем общего класса задач, именуемых **задачами о покрытии**. В этой модели все переменные являются двоичными. Все коэффициенты левой части каждого ограничения равны 0 или 1, а правая часть ограничений имеет вид "≥ 1". Целевая функция всегда имеет вид $c_1x_1 + c_2x_2 + ... + c_nx_n$, где $c_j > 0$ для всех $j = 1, 2, ..., n$, и подлежит минимизации. В рассмотренном примере $c_j = 1$ для всех j. Однако если величина c_j будет равна стоимости установки телефона на j-м перекрестке, то эти коэффициенты могут принимать значения, отличные от 1.

Упражнения 9.2,d

1. Компания ABC занимается доставкой грузов пяти потребителям. Можно выбрать следующие маршруты перевозки грузов.

   Маршрут 1. 1, 2, 3, 4.

   Маршрут 2. 4, 3, 5.

   Маршрут 3. 1, 2, 5.

   Маршрут 4. 2, 3, 5.

   Маршрут 5. 1, 4, 2.

   Маршрут 6. 1, 3, 5.

   Эти маршруты определяются грузоподъемностью автомобиля, доставляющего грузы. Например, на маршруте 1 автомобиль имеет грузоподъемность, достаточную для доставки грузов лишь потребителям 1, 2, 3 и 4. Следующая таблица содержит расстояния (в милях) между терминалом компании ABC и потребителями.

	ABC	1	2	3	4	5
ABC	0	10	12	16	9	8
1	10	0	32	8	17	10
2	12	32	0	14	21	20
3	16	8	14	0	15	18
4	9	17	21	15	0	11
5	8	10	20	18	11	0

Необходимо выполнить дневные поставки пяти потребителям, минимизируя при этом пройденный суммарный путь. Оптимальное решение может быть таким, что один и тот же потребитель обслуживается более чем одним маршрутом. Но при реализации такого решения используется только один из этих маршрутов. Сформулируйте задачу в виде задачи ЦЛП и найдите оптимальное решение, используя программу TORA.

2. Американский университет формирует комитет по рассмотрению жалоб студентов. В соответствии с указаниями, полученными из администрации, в эту комиссию необходимо включить по крайней мере одну женщину, одного мужчину, одного студента, одного администратора и одного преподавателя. Выдвинуты десять кандидатур (обозначенных для удобства буквами от *a* до *j*). Принадлежность этих кандидатур к различным категориям отражена в следующей таблице.

Категория	Кандидатуры
Женщины	*a*, *b*, *c*, *d*, *e*
Мужчины	*f*, *g*, *h*, *i*, *j*
Студенты	*a*, *b*, *c*, *j*
Администраторы	*e*, *f*
Преподаватели	*d*, *g*, *h*, *i*

Университет заинтересован создать наименьшую по составу комиссию, гарантирующую представительство каждой из указанных категорий. Сформулируйте задачу в виде задачи ЦЛП и найдите ее оптимальное решение, используя программу TORA.

3. Округ Вашингтон определил шесть городов, которые нуждаются в службе скорой помощи. Станции скорой помощи могут быть расположены в некоторых или во всех шести городах. Однако в силу территориальной близости некоторых городов одна станция может обслуживать более одного населенного пункта. Единственным условием является время поездки; оно не должно превышать 15 минут. Приведенная ниже таблица содержит время поездки в минутах между шестью городами.

	1	2	3	4	5	6
1	0	23	14	18	10	32
2	23	0	24	13	22	11
3	14	24	0	60	19	20
4	18	13	60	0	55	17
5	10	22	19	55	0	12
6	32	11	20	17	12	0

Сформулируйте задачу в виде задачи ЦЛП, оптимальное решение которой определит наименьшее количество станций и их расположение. Найдите оптимальное решение, используя программу TORA.

4. Сокровища короля Тута находятся в музее в Новом Орлеане. План музея, состоящего из нескольких комнат, соединенных открытыми дверями, показан на рис. 9.3. Сторож, находящийся у двери, может наблюдать за двумя смежными комнатами. Адми-

нистрация музея заинтересована, чтобы в каждой комнате присутствовал сторож, используя при этом минимальное их число. Сформулируйте задачу в виде задачи ЦЛП и найдите ее оптимальное решение, используя программу TORA.

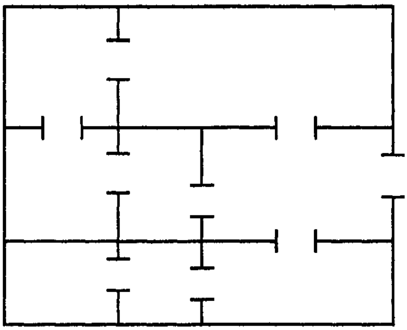


Рис. 9.3

Пример 9.2–5. (Ограничения типа "или–или")

Машиностроительная компания использует один станок для выполнения трех заказов. Время выполнения, а также срок сдачи каждого заказа даны в приведенной ниже таблице. Сроки сдачи заказов вычисляются от начальной даты, т.е. предполагаемого начала выполнения первого заказа.

Заказ	Время выполнения заказа (дни)	Срок сдачи заказа (дни)	Штраф за задержку заказа ($/день)
1	5	25	19
2	20	22	12
3	15	35	34

Требуется определить последовательность выполнения заказов, которая минимизирует штраф за задержку сдачи заказов.

Определим переменную x_j как дату завершения заказа j, измеряемую в днях от начальной даты. Задача имеет два типа ограничений: 1) ограничения, которые гарантируют, что никакие два заказа не выполняются одновременно; 2) ограничения по срокам сдачи заказов. Сначала рассмотрим первый тип ограничений.

Два заказа i и j, время выполнения которых p_i и p_j, не будут выполняться одновременно, если

$$\text{или } x_i \ge x_j + p_j, \text{ или } x_j \ge x_i + p_i,$$

в зависимости от того, будет ли заказ i предшествовать выполнению заказа j или наоборот.

Так как все математические модели имеют дело лишь с *совместными* ограничениями, мы преобразуем ограничения типа **или–или** путем введения дополнительной двоичной переменной

$$y_{ij} = \begin{cases} 1, & \text{если заказ } i \text{ предшествует заказу } j, \\ 0, & \text{если заказ } j \text{ предшествует заказу } i. \end{cases}$$

При достаточно большом M ограничение типа *или–или* преобразуется в следующие два *совместных* ограничения.

$$My_{ij} + \left(x_i - x_j\right) \geq p_j \text{ и } M\left(1 - y_{ij}\right) + \left(x_j - x_i\right) \geq p_i.$$

Указанное преобразование гарантирует, что лишь одно из двух ограничений может быть активным в любой момент времени. Если $y_{ij} = 0$, первое ограничение является активным, а второе — избыточным (так как его левая часть будет содержать величину M, которая намного больше p_i). Если $y_{ij} = 1$, первое ограничение является избыточным, а второе — активным.

Рассмотрим теперь ограничения по срокам сдачи заказов. При заданной дате d_j сдачи заказа j введем в рассмотрение неограниченную по знаку переменную s_j. Тогда соответствующее ограничение примет вид

$$x_j + p_j + s_j = d_j.$$

Если $s_j \geq 0$, то заказ сдается в срок, если же $s_j \leq 0$, получаем убытки, связанные с задержкой сдачи заказа. Используя стандартную замену

$$s_j = s_j^+ - s_j^-, \quad s_j^+, s_j^- \geq 0,$$

приводим ограничение к виду

$$x_j + s_j^+ - s_j^- = d_j - p_j.$$

Штраф за задержку сдачи заказа пропорционален s_j^-.

Математическая модель рассматриваемой задачи принимает следующий вид.

$$\text{Минимизировать } z = 19s_1^- + 12s_2^- + 34s_3^-$$

при ограничениях

$$\begin{aligned}
&x_1 - x_2 + My_{12} \geq 20,\\
&-x_1 + x_2 - My_{12} \geq 5 - M,\\
&x_1 - x_3 + My_{13} \geq 15,\\
&-x_1 + x_3 - My_{13} \geq 5 - M,\\
&x_2 - x_3 + My_{23} \geq 15,\\
&-x_2 + x_3 - My_{23} \geq 20 - M,\\
&x_1 + s_1^+ - s_1^- = 25 - 5,\\
&x_2 + s_2^+ - s_2^- = 22 - 20,\\
&x_3 + s_3^+ - s_3^- = 35 - 15,\\
&x_1, x_2, x_3, s_1^+, s_1^-, s_2^+, s_2^-, s_3^+, s_3^- \geq 0,\\
&y_{12}, y_{13}, y_{23} = 0 \text{ или } 1.
\end{aligned}$$

Целочисленные переменные y_{12}, y_{13} и y_{23} введены для преобразования ограничений типа *или–или* в совместные ограничения. Конечная задача является *частично-целочисленной* задачей линейного программирования.

Для решения задачи выберем $M = 100$ — число, которое больше суммы времени изготовления всех трех заказов.

Оптимальным решением (полученным с помощью программы TORA) является $x_1 = 20$, $x_2 = 0$ и $x_3 = 25$. Следовательно, оптимальной последовательностью выполнения заказов будет $2 \to 1 \to 3$. В соответствии с оптимальным решением заказ 2 выполняется за время $0 + 5 = 5$, заказ 1 — за время $20 + 5 = 25$ и заказ 3 — за $25 + 15 = 40$ дней. В результате выполнение заказа 3 задержано на $40 - 35 = 5$ дней, что приводит к штрафу в размере $5 \times 34 = 170$ долларов.

Упражнения 9.2,e

1. Игральная доска состоит из девяти равных квадратов, расположенных 3×3. Требуется заполнить каждый квадрат числом из интервала от 1 до 9 таким образом, чтобы сумма чисел каждой строки, каждого столбца и каждой диагонали была равна 15. Определите числа в каждом квадрате для следующих случаев.

   a) Числа в каждой строке и каждом столбце различны.

   b) Числа во всех квадратах различны.

   Запишите сформулированную задачу в виде задачи ЦЛП с ограничениями и решите ее с помощью программы TORA.

2. Станок используется для производства двух взаимозаменяемых видов продукции. Производительность станка позволяет за день изготовить не более 20 единиц продукции первого вида и 10 единиц второго вида. Существует альтернативная настройка станка, позволяющая ежедневно изготовлять не более 12 единиц продукции вида 1 и 22 единицы вида 2. Анализ рынка показывает, что максимальный суммарный спрос на два вида продукции составляет 35 единиц ежедневно. Предположим, что прибыль от производства единицы продукции первого и второго вида составляет 10 и 12 долларов соответственно. Какая из двух возможных настроек станка должна быть выбрана? Сформулируйте задачу в виде задачи ЦЛП и решите ее с помощью программы TORA. (*Примечание.* Сформулированная двухмерная задача может быть решена путем перебора возможных решений после графического построения пространства допустимых решений. Этого нельзя сделать в n-мерной задаче.)

3. Некая компания производит три вида продукции. Ежедневные временн*ы*е затраты и потребности сырья, необходимые на изготовление одной единицы продукции, приведены в следующей таблице.

Продукция	Необходимое время (час./ед.)	Необходимое сырье (фунты/ед.)
1	3	4
2	4	3
3	5	6

Доходы от производства единицы каждого вида продукции равны 25, 30 и 22 доллара соответственно. Компания имеет возможность организовать выпуск этой продукции в двух цехах своего производства. Цехи отличаются главным образом ресурсом рабочего времени и сырья, как показывает следующая таблица.

Цех	Ресурс рабочего времени (часов в рабочий день)	Ресурс сырья (фунты в день)
1	100	100
2	90	120

Сформулируйте задачу в виде задачи частично-целочисленного линейного программирования и используйте программу TORA для оптимального размещения производства по цехам.

4. Рассмотрите задачу планирования производственной линии, связанной с изготовлением двух различных изделий на одном станке. Последовательность выполнения необходимых для этого восьми операций изображена на рис. 9.4. Пусть p_j — время выполнения j-й операции ($j = 1, 2, ..., n$). Сроки сдачи изделий типа 1 и 2, которые определяются исходя из некоторого исходного момента, равны d_1 и d_2 соответственно. Предполагается, что любая выполняемая на станке операция должна быть завершена до начала другой операции. Сформулируйте задачу в виде задачи частично-целочисленного линейного программирования.

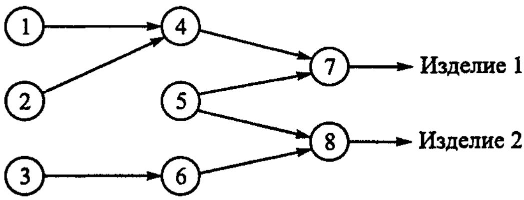


Рис. 9.4

5. Компания владеет фабрикой, которая производит изделия трех типов. Необходимые трудовые затраты и потребности сырья для производства одной единицы каждого из трех типов изделий приведены в следующей таблице.

Тип изделия	Необходимое время (час./ед.)	Необходимое сырье (фунты/ед.)
1	3	4
2	4	3
3	5	6
Наличный дневной объем	100	100

Доходы от производства единицы каждого из трех типов изделий равны 25, 30 и 45 долларов соответственно. Если будет производиться изделие типа 3, то его ежедневный объем производства должен быть не менее 5 единиц. Сформулируйте задачу в виде задачи частично-целочисленного линейного программирования и используйте программу TORA для нахождения оптимального решения.

6. Опишите невыпуклые заштрихованные области допустимых решений, которые изображены на рис. 9.5, в виде набора одновременно выполняющихся ограничений. Используйте программу TORA для нахождения оптимального решения, которое максимизирует целевую функцию $z = 2x_1 + 3x_2$ при ограничениях, определяющих область, изображенную на рис. 9.5,*а*.

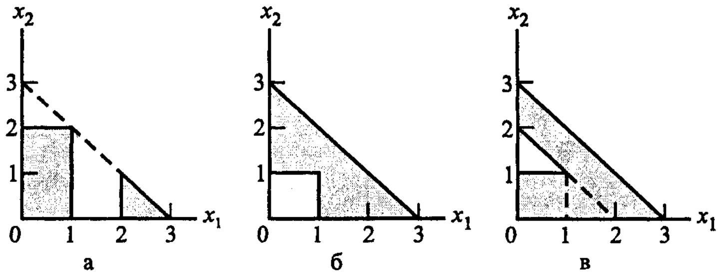


Рис. 9.5

7. Пусть требуется, чтобы *любые* k ограничений из следующих m ограничений были активными.

$$g_i(x_1, x_2, ..., x_n) \le b_i, \quad i = 1, 2, ..., m.$$

Покажите, как это сделать.

8. Правая часть следующего ограничения может принимать одно из значений b_1, b_2, ..., b_m, т.е.

$$g(x_1, x_2, ..., x_n) \le b_1, b_2, \text{ ... или } b_m.$$

Покажите, как можно представить это ограничение.

9.3. Методы решения задач целочисленного программирования

Методы решения задач целочисленного линейного программирования основаны на использовании вычислительных возможностей методов линейного программирования. Обычно алгоритмы целочисленного программирования включают три шага.

Шаг 1. "Ослабление" пространства допустимых решений задачи целочисленного линейного программирования путем замены любой двоичной переменной y непрерывным ограничением $0 \le y \le 1$ и отбрасывания требования целочисленности для всех остальных переменных. В результате получается обычная задача линейного программирования.

Шаг 2. Решение задачи линейного программирования и определение ее оптимального решения.

Шаг 3. Имея полученное (непрерывное) оптимальное решение, добавляем специальные ограничения, которые итерационным путем изменяют пространство допустимых решений задачи линейного программирования таким образом, чтобы в конечном счете получилось оптимальное решение, удовлетворяющее требованиям целочисленности.

Разработаны два общих метода генерирования специальных ограничений, о которых идет речь при реализации шага 3.

1. Метод ветвей и границ.
2. Метод отсекающих плоскостей.

Хотя ни один из упомянутых методов не дает надежных результатов при решении задачи целочисленного линейного программирования, опыт вычислений свидетельствует, что метод ветвей и границ более успешно решает задачу, чем метод отсекающих плоскостей.

9.3.1. Метод ветвей и границ

Основы этого метода объясним на численном примере.

Пример 9.3–1

Рассмотрим следующую задачу целочисленного линейного программирования.

Максимизировать $z = 5x_1 + 4x_2$

при ограничениях

$$x_1 + x_2 \le 5,$$
$$10x_1 + 6x_2 \le 45,$$
$$x_1, x_2 \ge 0 \text{ и целые.}$$

На рис. 9.6 пространство допустимых решений задачи целочисленного линейного программирования представлено точками. Соответствующая начальная задача линейного программирования (обозначим ее ЛП0) получается путем отбрасывания условий целочисленности. Ее оптимальным решением будет $x_1 = 3.75$, $x_2 = 1.25$ и $z = 23.75$.

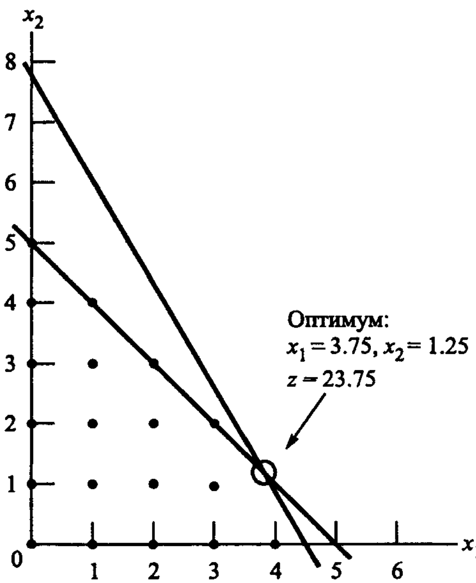


Рис. 9.6

Так как оптимальное решение задачи ЛП0 не удовлетворяет условия целочисленности, метод ветвей и границ изменяет пространство решений задачи линейного программирования так, что в конечном счете получается оптимальное решение задачи целочисленного линейного программирования. Для этого сначала выбирается одна из целочисленных переменных, значение которой в оптимальном решении задачи ЛП0 не является целочисленным. Например, выбирая x_1 (= 3.75), замечаем, что область $3 < x_1 < 4$ пространства допустимых решений задачи ЛП0 не содержит целочисленных значений переменной x_1 и, следовательно, может быть исключена из рассмотрения, как бесперспективная. Это эквивалентно замене исходной задачи ЛП0 двумя новыми задачами линейного программирования ЛП1 и ЛП2, которые определяются следующим образом:

пространство допустимых решений ЛП1 = пространство допустимых решений ЛП0 + ($x_1 \leq 3$),
пространство допустимых решений ЛП2 = пространство допустимых решений ЛП0 + ($x_1 \geq 4$).

На рис. 9.7 изображены пространства допустимых решений задач ЛП1 и ЛП2. Оба пространства содержат все допустимые решения исходной задачи ЦЛП. Это означает, что задачи ЛП1 и ЛП2 "не потеряют" решения начальной задачи ЛП0.

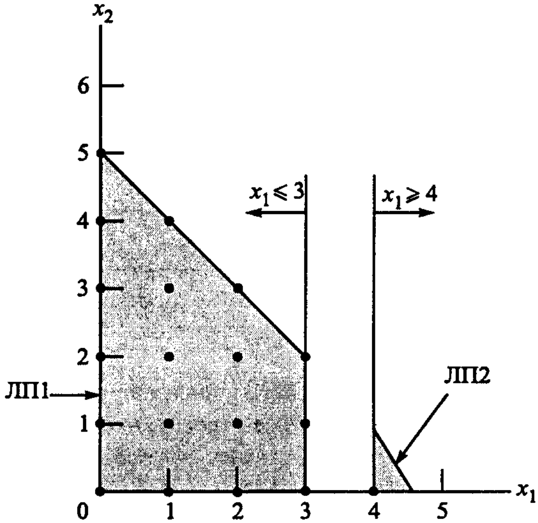


Рис. 9.7

Если мы продолжим *разумно* исключать из рассмотрения области, не содержащие целочисленных решений (такие, как $3 < x_1 < 4$), путем введения надлежащих ограничений, то в конечном счете получим задачу линейного программирования, оптимальное решение которой удовлетворяет требованиям целочисленности. Другими словами, будем решать задачу ЦЛП путем решения последовательности непрерывных задач линейного программирования.

Новые ограничения $x_1 \leq 3$ и $x_1 \geq 4$ взаимоисключаемы, так что задачи ЛП1 и ЛП2 необходимо рассматривать как независимые задачи линейного программирования, что и показано на рис. 9.8. Дихотомизация задач ЛП — основа концепции **ветвления** в методе ветвей и границ. В этом случае x_1 называется **переменной ветвления.**

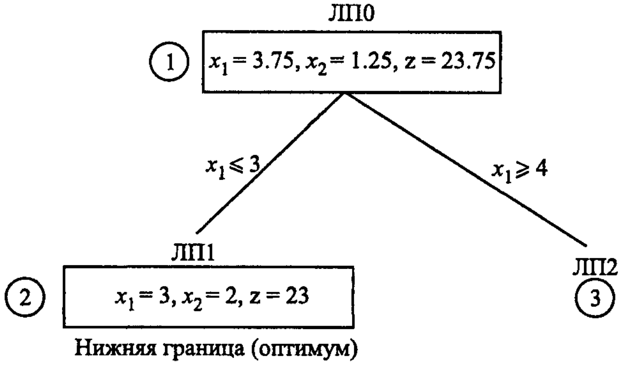


Рис. 9.8

Оптимальное решение задачи ЦЛП находится в пространстве допустимых решений либо задачи ЛП1, либо ЛП2. Следовательно, обе подзадачи должны быть решены. Выбираем сначала задачу ЛП1 (выбор произволен), имеющую дополнительное ограничение $x_1 \leq 3$.

$$\text{Максимизировать } z = 5x_1 + 4x_2$$

при ограничениях

$$\begin{aligned} &x_1 + x_2 \leq 5, \\ &10x_1 + 6x_2 \leq 45, \\ &x_1 \leq 3, \\ &x_1, x_2 \geq 0. \end{aligned}$$

Оптимальным решением задачи ЛП1 (которое можно получить с помощью метода решения задач с ограниченными переменными, изложенного в разделе 7.5.2) является $x_1 = 3$, $x_2 = 2$ и $z = 23$.

Оптимальное решение задачи ЛП1 удовлетворяет требованию целочисленности переменных x_1 и x_2. В этом случае говорят, что задача ЛП1 **прозондирована**. Это означает, что задача ЛП1 не должна больше зондироваться, так как она не может содержать *лучшего* решения задачи ЦЛП.

Мы не можем в этой ситуации оценить качество целочисленного решения, полученного из рассмотрения задачи ЛП1, ибо решение задачи ЛП2 может привести к лучшему целочисленному решению (с бо́льшим значением целевой функции z). Пока мы можем лишь сказать, что значение $z = 23$ является **нижней границей** оптимального (максимального) значения целевой функции исходной задачи ЦЛП. Это значит, что любая нерассмотренная подзадача, которая не может привести к целочисленному решению с бо́льшим значением целевой функции, должна быть исключена из рассмотрения, как бесперспективная. Если же нерассмотренная подзадача может привести к лучшему целочисленному решению, то нижняя граница должна быть надлежащим образом изменена.

При значении нижней границы $z = 23$ исследуем задачу ЛП2 (единственную оставшуюся нерассмотренную подзадачу). Так как в задаче ЛП0 оптимальное значение целевой функции равно 23.75 и *все ее коэффициенты являются целыми числами*, то невозможно получить целочисленное решение задачи ЛП2 (пространство решений которой более узко, чем в задаче ЛП0), которое будет лучше имеющегося. В результате мы отбрасываем подзадачу ЛП2 и считаем ее *прозондированной.*

Реализация метода ветвей и границ завершена, так как обе подзадачи ЛП1 и ЛП2 прозондированы (рассмотрение первой привело к нахождению целочисленного решения, а второй — к заключению, что ее возможное целочисленное решение не может быть лучше имеющегося). Следовательно, мы заключаем, что оптимальным решением задачи ЦЛП является решение, соответствующее нижней границе, а именно $x_1 = 3$, $x_2 = 2$ и $z = 23$.

Остались без ответа два вопроса, связанные с реализацией описанной вычислительной процедуры.

1. Можно ли было в задаче ЛП0 выбрать переменную x_2 в качестве *переменной ветвления* вместо x_1?
2. Можно ли было при выборе подзадачи для зондирования решить сначала задачу ЛП2 вместо ЛП1?

Ответы на оба вопроса положительные. Однако последующие вычисления могут значительно отличаться. Ситуация, когда первой решается задача ЛП2, иллюстрируется схемой вычислений (рис. 9.9), подтверждающей высказанную мысль. Оптимальным решением задачи ЛП2 является $x_1 = 4$, $x_2 = 0.83$ и $z = 23.33$ (проверьте с помощью программы TORA).

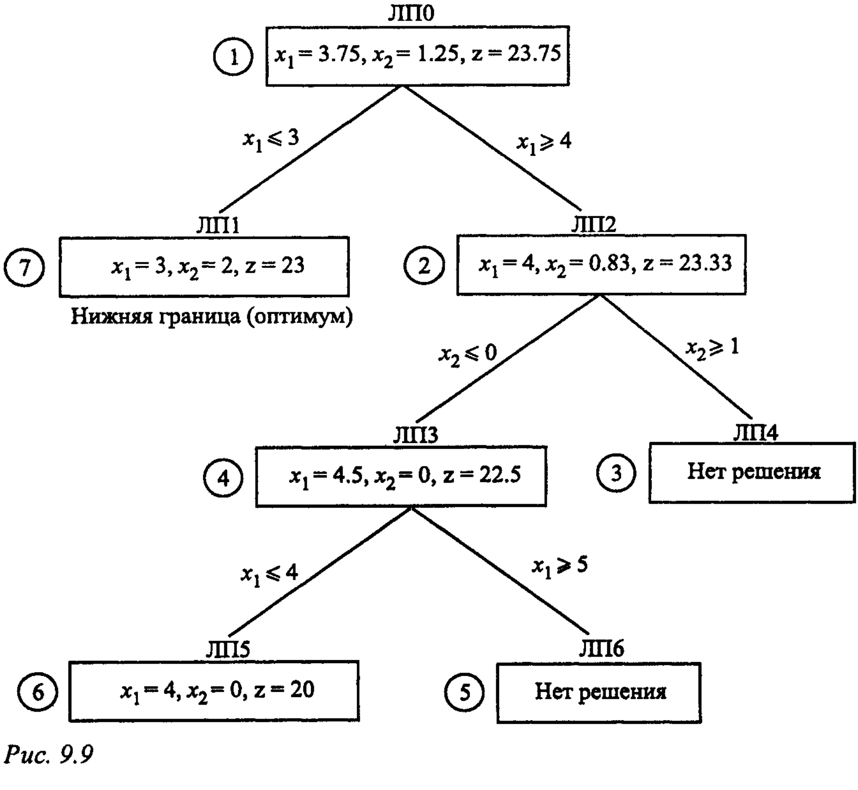


Рис. 9.9

Поскольку значение переменной x_2 (= 0.83) не является целым числом, задача ЛП2 исследуется дальше. Рассматриваем подзадачи ЛП3 и ЛП4, используя ветви $x_2 \le 0$ и $x_2 \ge 0$ соответственно. Это означает, что

пространство решений ЛП3 = пространство решений ЛП2 + ($x_2 \le 0$) =
= пространство решений ЛП0 + ($x_1 \ge 4$) +($x_2 \le 0$),
пространство решений ЛП4 = пространство решений ЛП2 + ($x_2 \ge 1$) =
= пространство решений ЛП0 + ($x_1 \ge 4$) +($x_2 \ge 1$).

У нас есть три нерассмотренные подзадачи, которые должны быть решены, — ЛП1, ЛП3 и ЛП4. Предположим, что мы произвольно выбрали первой задачу ЛП4. Эта задача не имеет решения и, следовательно, является прозондированной. В качестве следующей выберем подзадачу ЛП3. Ее оптимальным решением является $x_1 = 4.5$, $x_2 = 0$ и $z = 22.5$. Нецелочисленное значение переменной x_1 (= 4.5) порождает две ветви решений при $x_1 \le 4$ и $x_1 \ge 5$ и соответствующие им подзадачи ЛП5 и ЛП6. При этом

пространство решений ЛП5 = пространство решений ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$) + ($x_1 \le 4$),
пространство решений ЛП6 = пространство решений ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$) + ($x_1 \ge 5$).

Теперь не рассмотрены лишь подзадачи ЛП1, ЛП5 и ЛП6. Подзадача ЛП6 прозондирована, так как не имеет допустимых решений. Далее, подзадача ЛП5 имеет целочисленное решение ($x_1 = 4$, $x_2 = 0$, $z = 20$) и, следовательно, порождает нижнюю границу ($z = 20$) оптимального значения целевой функции задачи ЦЛП. Теперь остается лишь подзадача ЛП1, решение которой также является целочисленным ($x_1 = 3$, $x_2 = 2$, $z = 23$). Следовательно, нижнюю границу значений целевой функции полагаем равной 23. Так как все подзадачи прозондированы, оптимальным решением задачи ЦЛП является решение, соответствующее последней нижней границе, а именно $x_1 = 3$, $x_2 = 2$ и $z = 23$.

Последовательность решения подзадач, показанная на рис. 9.9 (ЛП0, ЛП2, ЛП4, ЛП3, ЛП6, ЛП5, ЛП1), является наихудшей; тем не менее, она встречается на практике. Этот пример указывает на основную слабость метода ветвей и границ: как выбирать следующую подзадачу для исследования и как выбирать для нее переменную ветвления?

В процессе решения, представленного на рис. 9.8, мы случайно наткнулись на хорошую нижнюю границу значений целевой функции на самой первой подзадаче ЛП1, что позволило прозондировать ЛП2 без детальных исследований и закончить вычисления. По существу, мы завершили вычисления, решив лишь одну подзадачу. В процессе решения, представленном на рис. 9.9, мы были вынуждены исследовать семь подзадач, и лишь тогда завершились вычисления метода ветвей и границ. Хотя имеются эвристические соображения, позволяющие "угадать", какая из ветвей может привести к улучшенному решению задачи ЦЛП (см., например, [4]), не существует строгой теории, которая всегда обеспечивала бы надежные результаты.

Теперь сформулируем алгоритм метода ветвей и границ в общем случае. Предположим, что рассматривается задача максимизации. Зададим нижнюю границу оптимального значения целевой функции z задачи ЦЛП равной $-\infty$. Положим $i = 0$.

Шаг 1. (Зондирование и определение границы). Выбираем i-ю подзадачу линейного программирования ЛПi для исследования. Решаем ЛПi и зондируем ее, при этом возможна одна из трех ситуаций.

a) Оптимальное значение целевой функции задачи ЛПi не может улучшить текущей нижней границы.
b) ЛПi приводит к лучшему допустимому целочисленному решению, чем текущая нижняя граница.
c) ЛПi не имеет допустимых решений.

Возможны два случая.

a) Если задача ЛПi прозондирована, нижняя граница изменяется только при условии, что найдено лучшее решение задачи ЦЛП. Если все подзадачи прозондированы, необходимо закончить вычисления: оптимальным решением задачи ЦЛП является то, которое соответствует текущей нижней границе, если такая существует. Иначе положить $i = i + 1$ и повторить шаг 1.
b) Если задача ЛПi не прозондирована, переходим к шагу 2 для выполнения ветвления.

Шаг 2. (Ветвление). Выбираем одну из целочисленных переменных x_j, оптимальное значение x_j^* которой в оптимальном решении задачи ЛПi не является целым числом. Исключаем из пространства допустимых решений область $\left[x_j^*\right] < x_j < \left[x_j^*\right] + 1$ (где $[v]$ — наибольшее целое, не превосходящее v) путем формирования двух подзадач ЛП, которые соответствуют ограничениям $x_j \le \left[x_i^*\right]$ и $x_j \ge \left[x_j^*\right] + 1$.

Положим $i = i + 1$ и переходим к шагу 1.

Описанный алгоритм применим для решения задач максимизации. Для решения задач минимизации в алгоритме необходимо заменить нижнюю границу верхней (начальное значение которой равно $z = +\infty$).

Алгоритм метода ветвей и границ непосредственно распространяется на задачи частично-целочисленного ЛП (в которых лишь некоторые из переменных должны принимать целочисленные значения). Если некоторая переменная является непрерывной, мы просто никогда не выбираем ее в качестве переменной ветвления. Допустимая подзадача определяет новую границу для значения целевой функции, если значения дискретных переменных являются целочисленными и значение целевой функции улучшено по сравнению с текущей границей.

Упражнения 9.3,а

1. Решите задачу ЦЛП из примера 9.3–1 методом ветвей и границ, начиная с переменной x_2 как переменной ветвления. Решите подзадачи с помощью программы TORA, используя опцию MODIFY (Изменить) для задания верхней и нижней границ. Начните с решения подзадачи, включающей ограничение $x_2 \le \left[x_2^*\right]$.

2. Постройте дерево подзадач, получаемое при использовании метода ветвей и границ для каждой из приведенных ниже задач. Для удобства в нулевом узле в качестве переменной ветвления всегда выбирайте x_1.

   a) Максимизировать $z = 3x_1 + 2x_2$
   при ограничениях

$$2x_1 + 2x_2 \leq 9,$$
$$3x_1 + 3x_2 \geq 18,$$
$$x_1, x_2 \geq 0 \text{ и целые.}$$

   b) Максимизировать $z = 2x_1 + 3x_2$
   при ограничениях

$$5x_1 + 7x_2 \leq 35,$$
$$4x_1 + 9x_2 \leq 36,$$
$$x_1, x_2 \geq 0 \text{ и целые.}$$

   c) Максимизировать $z = x_1 + x_2$
   при ограничениях

$$2x_1 + 5x_2 \leq 16,$$
$$6x_1 + 5x_2 \leq 27,$$
$$x_1, x_2 \geq 0 \text{ и целые.}$$

   d) Минимизировать $z = 5x_1 + 4x_2$
   при ограничениях

$$3x_1 + 2x_2 \geq 5,$$
$$2x_1 + 3x_2 \geq 7,$$
$$x_1, x_2 \geq 0 \text{ и целые.}$$

   e) Максимизировать $z = 5x_1 + 7x_2$
   при ограничениях

$$2x_1 + x_2 \leq 13,$$
$$6x_1 + 9x_2 \leq 41,$$
$$x_1, x_2 \geq 0 \text{ и целые.}$$

3. Решите задачи из предыдущего упражнения, предполагая, что на переменную x_1 не накладывается условие целочисленности.

4. Покажите графически, что следующая задача ЦЛП не имеет допустимых решений, а затем проверьте этот результат с помощью метода ветвей и границ.

$$\text{Максимизировать } z = 2x_1 + x_2$$

при ограничениях

$$10x_1 + 10x_2 \leq 9,$$
$$10x_1 + 5x_2 \geq 1,$$
$$x_1, x_2 \geq 0 \text{ и целые.}$$

5. Решите следующие задачи с использованием метода ветвей и границ.

a) Максимизировать $z = 18x_1 + 14x_2 + 8x_3 + 4x_4$
при ограничениях

$$15x_1 + 12x_2 + 7x_3 + 4x_4 + x_5 \le 37,$$
$$x_1, x_2, x_3, x_4, x_5 = 0 \text{ или } 1.$$

b) Максимизировать $z = x_1 + 2x_2 + 5x_3$
при ограничениях

$$|-x_1 + 10x_2 - 3x_3| \ge 15,$$
$$2x_1 + x_2 + x_3 \le 10,$$
$$x_1, x_2, x_3 \ge 0 \text{ и целые.}$$

9.3.2. Аддитивный алгоритм для задач с двоичными переменными

Любая целочисленная переменная x, значения которой не превышают конечной верхней границы u (т.е. $0 \le x \le u$), может быть выражена через двоичные переменные с помощью представления

$$x = 2^0 y_0 + 2^1 y_1 + 2^2 y_2 + \ldots + 2^k y_k,$$

где k — наименьшее целое число, удовлетворяющее условию $2^{k+1} - 1 \ge u$, а $y_0, y_1, \ldots, y_k$ — двоичные переменные. Это представление, а также простой алгоритм (описанный ниже) решения задач ЦЛП с двоичными переменными породили надежду, что общая задача ЦЛП может быть решена более эффективно как задача с двоичными переменными. К сожалению, это направление развития методов ЦЛП не оправдало возлагаемые на него надежды.

Впервые специальный алгоритм решения задач с двоичными переменными, названный **аддитивным**, был предложен в 1965 году, через семь лет после создания метода ветвей и границ. Первоначально алгоритм не был связан с методом ветвей и границ, так как не требовал решения задач линейного программирования; его основные операции сводились лишь к сложению и вычитанию. Однако вскоре обнаружилась связь между этими алгоритмами. Оказалось, что аддитивный алгоритм является специальным случаем метода ветвей и границ.

Замысел эвристического зондирования в аддитивном алгоритме требует представления задачи с двоичными переменными в удобной форме, удовлетворяющей следующим требованиям.

1. В выражении целевой функции все коэффициенты должны быть *неотрицательными*, и целевая функция должна подлежать минимизации.
2. Все ограничения должны быть типа "≤", возможно с отрицательными правыми частями. Эти ограничения превращаются затем в равенства путем введения непрерывных дополнительных переменных в левые части ограничений.

Любая задача с двоичными переменными может удовлетворить эти условия, что демонстрирует следующий пример.

Пример 9.3–2

Приведем следующую задачу с двоичными переменными к виду, удовлетворяющему требованиям аддитивного алгоритма.

$$\text{Максимизировать } z = 3x_1 - 5x_2$$

при ограничениях

$$x_1 + x_2 = 5,$$
$$4x_1 + 6x_2 \geq 4,$$
$$x_1, x_2 = 0 \text{ или } 1.$$

Сначала преобразуем задачу в задачу минимизации с ограничениями типа "$\leq$".

a) Умножим z на -1, чтобы получить целевую функцию $w = -3x_1 + 5x_2$, которую следует минимизировать.
b) Преобразуем ограничение–равенство в два ограничения–неравенства типа "$\leq$", для этого введем неравенства $x_1 + x_2 \leq 5$ и $-x_1 - x_2 \leq -5$.
c) Умножаем второе ограничение на -1 и получаем $-4x_1 - 6x_2 \leq -4$.

Используя в ограничениях дополнительные переменные s_1, s_2 и s_3, запишем задачу следующим образом.

$$\text{Минимизировать } w = -3x_1 + 5x_2$$

при ограничениях

$$x_1 + x_2 + s_1 = 5,$$
$$-x_1 - x_2 + s_2 = -5,$$
$$-4x_1 - 6x_2 + s_3 = -4,$$
$$x_1, x_2 = 0 \text{ или } 1,$$
$$s_1, s_2, s_3 \geq 0.$$

Чтобы обеспечить неотрицательность коэффициентов целевой функции, выполним подстановку $x_j = 1 - x'_j$ для всех переменных x_j с отрицательным коэффициентом в выражении целевой функции. В данном случае необходима подстановка $x_1 = 1 - x'_1$, после чего также надо уточнить значения правых частей ограничений. Далее аддитивный алгоритм будет иметь дело с переменными x'_1 и x_2.

Как и в методе ветвей и границ, ветвление в аддитивном алгоритме основывается на переменной ветвления x_j. Главное отличие в том, что здесь две ветви соответствуют строгим равенствам $x_j = 0$ и $x_j = 1$, так как x_j является двоичной переменной. Определение границы трактуется таким же образом, как и в методе ветвей и границ, — улучшенное целочисленное решение определяет верхнюю границу минимального значения целевой функции.

Зондирование подзадач может привести к одному из трех результатов.

1. Подзадача не может привести к допустимому решению.
2. Подзадача не может улучшить верхнюю границу целевой функции.
3. Подзадача имеет допустимое целочисленное решение.

Таким образом, возможности после решения подзадач такие же, как и в методе ветвей и границ. Основное отличие состоит в том, что здесь мы не решаем задач линейного программирования. Вместо этого используются эвристические рассуждения.

Приводимый ниже пример демонстрирует аддитивный алгоритм и его тесты зондирования подзадач.

Пример 9.3–3. (Аддитивный алгоритм)

Необходимо решить следующую задачу с двоичными переменными.

$$\text{Максимизировать } z = 3y_1 + 2y_2 - 5y_3 - 2y_4 + 3y_5$$

при ограничениях

$$\begin{aligned}
y_1 + y_2 + y_3 + 2y_4 + y_5 &\le 4,\\
7y_1 + 3y_3 - 4y_4 + 3y_5 &\le 8,\\
11y_1 - 6y_2 + 3y_4 - 3y_5 &\ge 3,\\
y_1, y_2, y_3, y_4, y_5 &= 0 \text{ или } 1.
\end{aligned}$$

Эту задачу нетрудно представить в форме, удовлетворяющей аддитивному алгоритму. Для этого нужно выполнить следующие действия (см. пример 9.3–2).

a) Умножаем целевую функцию на -1.

b) Умножаем третье ограничение на -1.

c) Вводим дополнительные переменные s_1, s_2 и s_3 для преобразования ограничений в равенства.

d) Чтобы все коэффициенты целевой функции были положительны, применяем подстановки $y_1 = 1 - x_1$, $y_2 = 1 - x_2$, $y_5 = 1 - x_5$, $y_3 = x_3$ и $y_4 = x_4$.

Указанные преобразования приводят к следующей целевой функции (проверьте!).

$$\text{Минимизировать } z' = 3x_1 + 2x_2 + 5x_3 + 2x_4 + 3y_5 - 8.$$

Для удобства будем игнорировать константу -8 и заменим $z' + 8$ на z, так что преобразованная задача принимает следующий вид.

$$\text{Минимизировать } z = 3x_1 + 2x_2 + 5x_3 + 2x_4 + 3y_5$$

при ограничениях

$$\begin{aligned}
-x_1 - x_2 + x_3 + 2x_4 - x_5 + s_1 &= 1,\\
-7x_1 + 3x_3 - 4x_4 - 3x_5 + s_2 &= -2,\\
11x_1 - 6x_2 - 3x_4 - 3x_5 + s_2 &= -1,\\
x_1, x_2, x_3, x_4, x_5 &= 0 \text{ или } 1.
\end{aligned}$$

Поскольку в преобразованной задаче ведется поиск минимума целевой функции с положительными коэффициентами, логично, что в начальном решении все двоичные переменные должны равняться нулю. В этом случае дополнительные переменные будут базисными, и их значения определяются правыми частями ограничений. Решение представлено в следующей таблице.

Базис	x_1	x_2	x_3	x_4	x_5	s_1	s_2	s_3	Решение
s_1	–1	–1	1	2	–1	1	0	0	1
s_2	–7	0	3	–4	–3	0	1	0	–2
s_3	11	–6	0	–3	–3	0	0	1	–1
Коэффициенты целевой функции	3	2	5	2	3				

Как следует из дальнейших действий, применение тестов зондирования к каждой подзадаче требует лишь изменения крайнего правого столбца таблицы.

Так как все двоичные переменные равны нулю, дополнительные переменные принимают следующие значения:

$$(s_1, s_2, s_3) = (1, -2, -1),$$

при этом $z = 0$. Если бы все дополнительные переменные были неотрицательными, мы сделали бы вывод, что рассматриваемое решение, в котором все двоичные переменные равны нулю, является оптимальным. Поскольку некоторые из дополнительных переменных являются недопустимыми (так как отрицательные), необходимо увеличить значение одной или нескольких двоичных переменных до 1, чтобы получить допустимое решение (или прийти к выводу, что задача допустимого решения не имеет).

Увеличение значений двоичных переменных до 1 в аддитивном алгоритме происходит *по отдельности*. Выбранная переменная называется переменной ветвления, ее выбор основан на использовании специальных тестов.

Переменная ветвления должна уменьшить (по абсолютной величине) отрицательные значения дополнительных переменных. Как следует из приведенной таблицы, переменная x_3 не может быть выбрана в качестве переменной ветвления, так как ее коэффициенты во втором и третьем ограничениях неотрицательны. Следовательно, положив $x_3 = 1$, мы лишь увеличим (по абсолютной величине) отрицательные значения дополнительных переменных s_2 и s_3. Каждая же из оставшихся двоичных переменных имеет по крайней мере один отрицательный коэффициент во втором и третьем ограничениях. Следовательно, комбинация этих переменных может привести к положительным значениям дополнительных переменных. Таким образом, лишь переменные x_1, x_2, x_4 и x_5 можно рассматривать в качестве возможных кандидатов на переменную ветвления.

Выбор переменной ветвления из возможных претендентов основан на использовании *меры недопустимости дополнительной переменной*. Эта мера, основанная на предположении, что значение двоичной переменной x_j будет увеличено до 1, определяется соотношением

$$I_j = \sum_{\text{по всем } i} \min\{0, s_i - a_{ij}\},$$

где s_i — текущее значение дополнительной переменной, a_{ij} — коэффициент при переменной x_i в i-м ограничении.

В действительности I_j есть не что иное, как сумма значений отрицательных дополнительных переменных, являющаяся результатом увеличения значения переменной x_j до 1. Сложная на вид формула может быть упрощена следующим образом.

$$I_j = \sum_{\text{по всем } i} (\text{отрицательное значение } s_i \text{ при заданном } x_j = 1).$$

Например, если мы положим $x_1 = 1$, то получим $s_1 = 1 - (-1) = 2$, $s_2 = -2 - (-7) = 5$ и $s_3 = -1 - 11 = -12$. Следовательно, $I_1 = -12$. Аналогично, $I_2 = -2$, $I_4 = -1$ и $I_5 = 0$ (напомним, что переменная x_3 была исключена из претендентов на переменную ветвления, как бесперспективная). Так как I_5 имеет наименьшую меру недопустимости, переменная x_5 выбирается в качестве переменной ветвления. На рис. 9.10 изображены две ветви, соответствующие $x_5 = 1$ и $x_5 = 0$, и образованные при этом узлы 1 и 2. Вершина 1 дает допустимое значение дополнительных переменных $(s_1, s_2, s_3) = (2, 1, 2)$ и $z = 3$. Следовательно, узел 1 прозондирован, и значение $\overline{z} = 3$ определяет текущую *верхнюю границу* оптимального значения целевой функции.

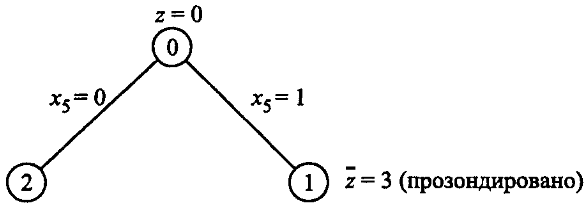


Рис. 9.10

Прозондировав узел 1, переходим к узлу 2, для которого $x_5 = 0$. Здесь имеем $(s_1, s_2, s_3) = (1, -2, -1)$ и $z = 0$, т.е. решение недопустимо. Переменные x_1, x_2, x_3 и x_4 являются возможными кандидатами на переменную ветвления. (Заметим, что хотя решения в узлах 0 и 2 идентичны, узел 2 отличается тем, что переменная x_5 не является больше претендентом на ветвление.) Как и в вершине 0, здесь переменная x_3 бесперспективна, так как не уменьшает (по абсолютной величине) отрицательных значений дополнительных переменных s_2 и s_3. Кроме того, значение $x_3 = 1$ приводит к значению целевой функции, равному 5, что хуже текущего значения верхней границы $\overline{z} = 3$. Переменная x_1 также бесперспективна, так как соответствующий коэффициент в целевой функции равен 3, поэтому значение $x_1 = 1$ не приводит к улучшению имеющегося значения целевой функции. Для оставшихся переменных x_2 и x_4 вычисляем меру недопустимости: $I_2 = -2$, $I_4 = -1$. Следовательно, в узле 2 переменной ветвления будет x_4.

На рис. 9.11 показаны ветви $x_4 = 1$ и $x_1 = 0$, которые ведут к узлам 3 и 4 соответственно. В узле 3 (определенном ограничениями $x_5 = 0$ и $x_4 = 1$) имеем решение $(s_1, s_2, s_3) = (-1, 2, 2)$, $z = 2$, которое не является допустимым. Кандидатами на ветвление являются переменные x_1, x_2 и x_3. Однако, увеличивая значение любой из них до 1, мы ухудшим значение целевой функции z по сравнению с текущей верхней границей ($\overline{z} = 3$). Следовательно, все переменные исключены из кандидатов на ветвление, и узел 3 прозондирован.

Далее в узле 4, определенном ограничениями $x_5 = x_4 = 0$, имеем $(s_1, s_2, s_3) = (1, -2, -1)$, $z = 0$. Переменные x_1 и x_3 исключаются из претендентов на ветвление в силу теста на верхнюю границу. Оставшуюся переменную x_2 нельзя исключить из рассмотрения на основании тестов. Следовательно, x_2 является переменной ветвления.

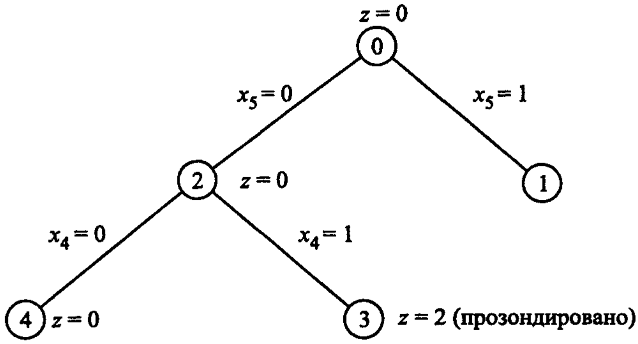


Рис. 9.11

На рис. 9.12 показаны узлы 5 и 6, исходящие из узла 4. В узле 5 имеем $(s_1, s_2, s_3) = (2, -2, 5)$, $z = 2$, и переменные x_1 и x_3 являются кандидатами на ветвление. Переменная x_1 исключается тестом на верхнюю границу, а x_3 — как тестом на верхнюю границу, так и тестом на допустимость дополнительных переменных. Это значит, что узел 5 прозондирован. Узел 6 также прозондирован, ибо ни x_1, ни x_3 не могут привести к улучшению допустимого решения.

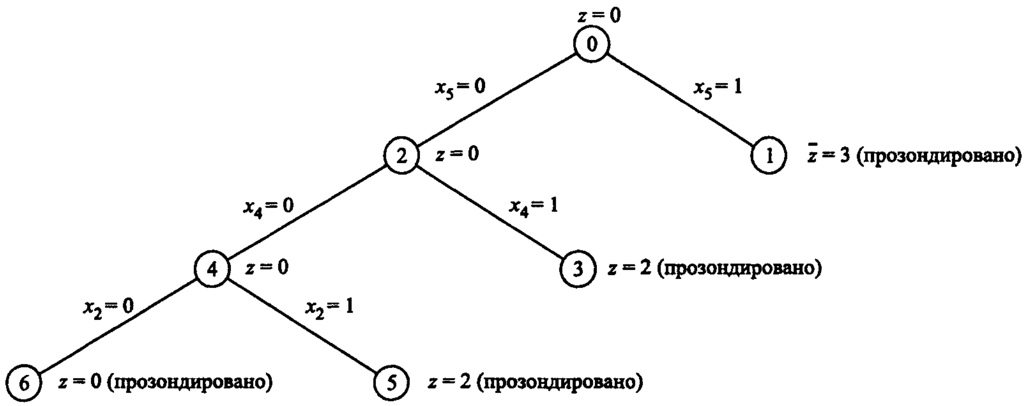


Рис. 9.12

Теперь все узлы (рис. 9.12) прозондированы, поэтому метод ветвей и границ заканчивает работу. Оптимальное решение найдено в узле 1, т.е. $x_5 = 1$, $z = 3$, все остальные переменные равны нулю. Отсюда получаем решение исходной задачи: $y_1 = y_2 = 1, y_3 = y_4 = y_5 = 0, w = 5$.

Из рис. 9.12 видно, чем меньше ветвей, ведущих к прозондированному узлу, тем эффективнее алгоритм. Например, вершина 1 определена посредством фиксирования одной ветви ($x_5 = 1$), и ее зондирование автоматически отвечает за $2^{5-1} = 16$ двоичных решений (все решения, для которых $x_5 = 1$). Узел 3 определен путем фиксирования двух двоичных переменных, и его зондирование отвечает лишь за $2^{5-2} = 8$ двоичных решений.

Тесты, предложенные в примере 9.3–3, являются явно эвристическими, и их эффективность в исключении переменных, которые не могут инициировать процесс ветвления, зависит от того, насколько "умными" мы их конструируем. В действительности полный аддитивный алгоритм включает более сильные тесты, чем те, которые показаны в примере 9.3–3. Однако все они основаны на эвристических рассуждениях.

Упражнения 9.3,b

1. Приведенные ниже задачи решите с помощью аддитивного алгоритма.

   a) Максимизировать $z = 3x_1 + x_2 + 3x_3$
   при ограничениях

$$
\begin{aligned}
&-x_1 + 2x_2 + x_3 \le 4,\\
&4x_2 - 3x_3 \le 2,\\
&x_1 - 3x_2 + 2x_3 \le 3,\\
&x_1, x_2, x_3 = 0 \text{ или } 1.
\end{aligned}
$$

   b) Минимизировать $z = 2x_1 + 4x_2 + 6x_3$
   при ограничениях

$$
\begin{aligned}
&8x_1 - 4x_2 - x_3 \ge 5,\\
&6x_1 - 3x_2 - 2x_3 \ge 2,\\
&-2x_1 + 9x_2 + 7x_3 \ge 4,\\
&x_1, x_2, x_3 = 0 \text{ или } 1.
\end{aligned}
$$

   c) Минимизировать $z = -5x_1 + 7x_2 + 10x_3$
   при ограничениях

$$
\begin{aligned}
&-x_1 - 3x_2 + 5x_3 \ge 0,\\
&2x_1 + 6x_2 - 3x_3 \ge 4,\\
&-x_2 + 2x_3 \ge 2,\\
&x_1, x_2, x_3 = 0 \text{ или } 1.
\end{aligned}
$$

2. Покажите с помощью аддитивного алгоритма, что следующая задача не имеет допустимых решений.

   Максимизировать $z = 2x_1 + x_2$

   при ограничениях

$$
\begin{aligned}
&10x_1 + 10x_2 \le 9,\\
&10x_1 + 5x_2 \ge 1,\\
&x_1, x_2 = 0 \text{ или } 1.
\end{aligned}
$$

3. Решите задачу из примера 9.2–1 с помощью аддитивного алгоритма.

4. Решите следующую задачу при условии, что должно выполняться только одно из имеющихся ограничений.

   Максимизировать $z = x_1 + 2x_2 - 3x_3$

при ограничениях

$$20x_1 + 15x_2 - x_3 \le 10,$$
$$12x_1 - 3x_2 + 4x_3 \le 20,$$
$$x_1, x_2, x_3 = 0 \text{ или } 1.$$

9.3.3. Метод отсекающих плоскостей

Данный метод, как и метод ветвей и границ, начинает работу с оптимального решения "обычной" (непрерывной) задачи линейного программирования. Однако вместо ветвления и построения границ этот метод видоизменяет пространство допустимых решений, последовательно прибавляя специальным образом построенные ограничения (именуемые **отсечениями**). Рассмотрим сначала идею этого метода на графическом примере, а затем покажем, как отсечения строятся алгебраически.

Пример 9.3–4

Продемонстрируем применение метода отсекающих плоскостей для решения следующей задачи ЦЛП.

$$\text{Максимизировать } z = 7x_1 + 10x_2$$

при ограничениях

$$-x_1 + 3x_2 \le 6,$$
$$7x_1 + x_3 \le 35,$$
$$x_1, x_2 \ge 0 \text{ и целые.}$$

Данный метод путем добавления отсечений (отсекающих плоскостей) преобразует пространство допустимых решений соответствующей задачи линейного программирования в выпуклый многогранник, вершина которого, соответствующая оптимуму, является целочисленной и представляет решение исходной задачи. На рис. 9.13 показан пример двух таких отсечений. Мы начинаем с оптимального решения непрерывной задачи линейного программирования $(x_1, x_2) = (41, 31)$ и $z = 661$. Затем прибавляем отсечение I, которое вместе с ограничениями исходной задачи линейного программирования приводит к оптимальному решению $(x_1, x_2) = (4\frac{4}{7}, 3)$ с $z = 62$. После этого прибавляется отсечение II, которое вместе с отсечением I и исходными ограничениями приводит к оптимальному решению $(x_1, x_2) = (4, 3)$ с $z = 58$. Последнее решение является полностью целочисленным, как и требуется.

Прибавленные отсечения не отбрасывают ни одной исходной допустимой целочисленной точки, но должны проходить по меньшей мере через одну целочисленную точку (допустимую или недопустимую). Этим основным требованиям должно удовлетворять любое отсечение.

В общем случае может потребоваться любое (конечное) число отсечений для достижения полностью целочисленной экстремальной точки. В действительности количество необходимых для этого отсечений не зависит от размерности задачи в том смысле, что для решения задачи с небольшим количеством переменных и ограничений может потребоваться больше отсечений, чем для задачи большой размерности.

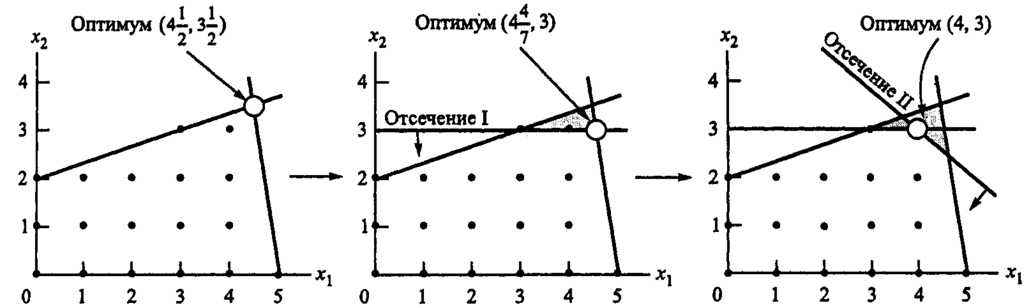


Рис. 9.13

Упражнения 9.3,c

1. В примере 9.3–4 покажите графически, может ли каждое из следующих ограничений служить в качестве правильного отсечения.

   a) $x_1 + 2x_2 \le 10$.
   b) $2x_1 + x_2 \le 10$.
   c) $3x_2 \le 10$.
   d) $3x_1 + x_2 \le 15$.

2. В примере 9.3–4 покажите графически, как следующие два правильных отсечения могут привести к оптимальному целочисленному решению.

   Отсечение I. $x_1 + 2x_2 \le 10$.

   Отсечение II. $3x_1 + x_2 \le 15$.

Алгебраический способ определения отсечений. Метод отсекающих плоскостей начинает работу с решения непрерывной задачи линейного программирования. В симплекс-таблице, соответствующей оптимальному решению задачи линейного программирования, следует выбрать одну из строк (называемую **производящей**), для которой базисная переменная нецелочисленная. Искомое отсечение строится на основании дробных составляющих коэффициентов производящей строки. По этой причине его называют **дробным отсечением**.

Сейчас мы выведем уравнение дробного отсечения. Построение отсечения объясняется на численном примере, а не на сложных математических конструкциях.

Пример 9.3–5

При дополнительных переменных x_3 и x_4 для ограничений 1 и 2 задачи из примера 9.3–4 оптимальная симплекс-таблица соответствующей задачи линейного программирования имеет следующий вид.

Базис	x_1	x_2	x_3	x_4	Решение
z	0	0	63/22	31/22	66.5
x_2	0	1	7/22	1/22	3.5
x_1	1	0	−1/22	3/22	4.5

Оптимальным непрерывным решением является $x_1 = 41$, $x_2 = 31$, $x_3 = 0$, $x_4 = 0$ и $z = 661$. Целочисленное отсечение получается в предположении, что все переменные задачи являются целочисленными. Заметим также, так как коэффициенты исходной целевой функции являются целочисленными, то и значение z, соответствующее целочисленному решению, должно быть целочисленным.

Информация, содержащаяся в симплекс-таблице, соответствующей оптимальному решению, может быть записана в виде следующих уравнений.

$$
\begin{aligned}
&z\text{-уравнение:} && z + \tfrac{63}{22}x_3 + \tfrac{31}{22}x_4 = 66\tfrac{1}{2},\\
&x_2\text{-уравнение:} && x_2 + \tfrac{7}{22}x_3 + \tfrac{1}{22}x_4 = 3\tfrac{1}{2},\\
&x_1\text{-уравнение:} && x_1 - \tfrac{1}{22}x_3 + \tfrac{3}{22}x_4 = 4\tfrac{1}{2}.
\end{aligned}
$$

Так как в этом примере z, x_1 и x_2 должны быть целочисленными и все они на данный момент имеют дробные значения в оптимальной симплекс-таблице, любое из трех уравнений можно использовать в качестве производящей строки для построения отсечения. Выберем (произвольно) для этой цели z-уравнение.

$$z + \tfrac{63}{22}x_3 + \tfrac{31}{22}x_4 = 66\tfrac{1}{2} \quad \text{(производящая } z\text{-строка).}$$

Для построения дробного отсечения каждый из *нецелочисленных* коэффициентов раскладывается на целую и дробную части при условии, что *дробная часть является строго положительной.* Например,

$$
\begin{aligned}
\tfrac{5}{2} &= \left(2 + \tfrac{1}{2}\right),\\
-\tfrac{7}{3} &= \left(-3 + \tfrac{2}{3}\right).
\end{aligned}
$$

Разложение коэффициентов производящей z-строки приводит к следующему уравнению.

$$z + \left(2 + \tfrac{19}{22}\right)x_3 + \left(1 + \tfrac{9}{22}\right)x_4 = \left(66 + \tfrac{1}{2}\right).$$

При переносе всех целочисленных слагаемых в левую часть уравнения, а всех дробных слагаемых в правую часть получаем следующее.

$$z + 2x_3 + x_4 - 66 = \tfrac{1}{2} - \tfrac{19}{22}x_3 - \tfrac{9}{22}x_4.$$

Поскольку все переменные в рассматриваемой задаче принимают целочисленные значения, левая часть последнего уравнения должна быть целочисленной, откуда следует, что и правая часть также должна принимать целые значения. Перепишем ее следующим образом.

$$\tfrac{1}{2} - \tfrac{19}{22}x_3 - \tfrac{9}{22}x_4 = \tfrac{1}{2} - \left(\tfrac{19}{22}x_3 + \tfrac{9}{22}x_4\right).$$

Поскольку $x_3, x_4 \geq 0$, выражение в скобках является неотрицательным. Поэтому величина $\tfrac{1}{2} - \tfrac{19}{22}x_3 - \tfrac{9}{22}x_4$, будучи целочисленной, не может превышать 1. Следовательно, необходимое условие целочисленности можно записать следующим образом.

$$\tfrac{1}{2} - \tfrac{19}{22}x_3 - \tfrac{9}{22}x_4 \leq 0.$$

Это и есть *отсечение*, порожденное z-строкой. Нетрудно убедиться, что ранее найденное оптимальное непрерывное решение $x_1 = 41$, $x_2 = 31$, $x_3 = 0$, $x_4 = 0$ не удовлетворяет отсечению. Действительно, так как $x_3 = x_4 = 0$, отсечение не удовлетворяется (оно приводит к неравенству $1 \le 0$). Следовательно, если мы присоединим отсечение к конечной симплекс-таблице, то оптимальное решение новой симплекс-таблицы будет "двигаться" в направлении выполнения ограничения целочисленности.

Можно таким же образом построить отсечение, исходя из производящей x_1-строки или x_2-строки. Рассмотрим сначала x_1-строку. Имеем

$$x_1 - \tfrac{1}{22}x_3 + \tfrac{3}{22}x_4 = 4\tfrac{1}{2} \quad \text{(производящая } x_1\text{-строка).}$$

Операция разложения приводит к уравнению

$$x_1 + \left(-1 + \tfrac{21}{22}\right)x_3 + \left(0 + \tfrac{3}{22}\right)x_4 = \left(4 + \tfrac{1}{2}\right).$$

Соответствующим отсечением является

$$-\tfrac{21}{22}x_3 - \tfrac{3}{22}x_4 + \tfrac{1}{2} \le 0.$$

Аналогично,

$$x_2 + \tfrac{7}{22}x_3 + \tfrac{1}{22}x_4 = 3\tfrac{1}{2} \quad \text{(производящая } x_2\text{-строка)}$$

записывается в виде

$$x_2 + \left(0 + \tfrac{7}{22}\right)x_3 + \left(0 + \tfrac{1}{22}\right)x_4 = \left(3 + \tfrac{1}{2}\right).$$

Следовательно, в данном случае отсечение имеет вид

$$-\tfrac{7}{22}x_3 - \tfrac{1}{22}x_4 + \tfrac{1}{2} \le 0.$$

Любое из трех отсечений может быть использовано на первой итерации метода отсекающих плоскостей. Поэтому нет необходимости строить все три отсечения перед выбором одного из них.

Выбирая (произвольно) отсечение, порожденное x_2-строкой, записываем его следующим образом.

$$-\tfrac{7}{22}x_3 - \tfrac{1}{22}x_4 + s_1 = -\tfrac{1}{2},\ s_1 \ge 0 \quad \text{(отсечение 1).}$$

Это ограничение добавляется в качестве дополнительного в оптимальную симплекс-таблицу задачи линейного программирования.

Базис	x_1	x_2	x_3	x_4	s_1	Решение
z	0	0	63/22	31/22	0	66.5
x_2	0	1	7/22	1/22	0	3.5
x_1	1	0	–1/22	3/22	0	4.5
s_1	0	0	–7/22	–1/22	1	–0.5

Таблица представляет оптимальное, но недопустимое решение. Для восстановления допустимости решения применим двойственный симплекс-метод (см. раздел 4.5), что приведет к следующей симплекс-таблице.

Базис	x_1	x_2	x_3	x_4	s_1	Решение
z	0	0	0	1	9	62
x_2	0	1	0	0	1	3
x_1	1	0	0	1/7	–1/7	$4\frac{4}{7}$
x_3	0	0	1	1/7	–22/7	$1\frac{4}{7}$

Из-за дробных значений переменных x_1 и x_3 последнее решение все еще нецелочисленное. Выберем x_1-строку в качестве производящей, т.е.

$$x_1 + \left(0 + \tfrac{1}{7}\right)x_4 + \left(-1 + \tfrac{6}{7}\right)s_1 = \left(4 + \tfrac{4}{7}\right).$$

Соответствующее отсечение имеет вид

$$-\tfrac{1}{7}x_4 - \tfrac{6}{7}s_1 + s_2 = -\tfrac{4}{7},\ \ s_2 \geq 0\ \ (\text{отсечение } 2).$$

Присоединяя отсечение 2 к последней симплекс-таблице, получаем следующее.

Базис	x_1	x_2	x_3	x_4	s_1	s_2	Решение
z	0	0	0	1	9	0	62
x_2	0	1	0	0	1	0	3
x_1	1	0	0	1/7	–1/7	0	$4\frac{4}{7}$
x_3	0	0	1	1/7	–22/7	0	$1\frac{4}{7}$
s_2	0	0	0	–1/7	–6/7	1	–4/7

Применение двойственного симплекс-метода приводит к следующей симплекс-таблице.

Базис	x_1	x_2	x_3	x_4	s_1	s_2	Решение
z	0	0	0	0	3	7	58
x_2	0	1	0	0	1	0	3
x_1	1	0	0	0	–1	1	4
x_3	0	0	1	0	–4	1	1
x_4	0	0	0	1	6	–7	4

Оптимальное решение ($x_1 = 4$, $x_2 = 3$, $z = 58$), определяемое последней симплекс-таблицей, является целочисленным. То, что все элементы данной симплекс-таблицы являются целочисленными, не случайность. Это типичное явление при использовании дробных отсечений.

Важно подчеркнуть, что применение дробного отсечения предполагает целочисленность всех переменных, включая дополнительные. Это значит, что данный метод применим только к решению полностью целочисленных задач.

Важность этого продемонстрируем на следующем примере. Рассмотрим ограничение

$$x_1 + \frac{1}{3}x_2 \le \frac{13}{2},$$

$$x_1, x_2 \ge 0 \text{ и целые.}$$

С точки зрения решения соответствующей задачи ЦЛП это ограничение преобразуется в уравнение путем введения неотрицательной дополнительной переменной s_1, т.е.

$$x_1 + \frac{1}{3}x_2 + s_1 = \frac{13}{2}.$$

Применение дробного отсечения предполагает, что ограничение имеет допустимое целочисленное решение по всем переменным, т.е. x_1, x_2 и s_1. Рассмотрев это уравнение, можем сказать, что оно может иметь допустимое целочисленное решение переменных x_1 и x_2 *лишь в том случае, если переменная* s_1 *принимает нецелочисленные значения.* Следовательно, применение дробного отсечения приведет к недопустимому целочисленному решению, так как *все* переменные x_1, x_2 и s_1 не могут одновременно быть целочисленными. Тем не менее ограничение имеет допустимые целочисленные решения для рассматриваемых переменных x_1 и x_2.

Есть две возможности исправить эту ситуацию.

1. Можно умножить все ограничения на соответствующую константу для устранения дробей. Например, приведенное выше ограничение умножается на 6, что приводит к неравенству $6x_1 + 2x_2 \le 39$. Любое целочисленное решение переменных x_1 и x_2 автоматически дает целочисленное значение дополнительной переменной. Однако этот тип преобразования применим лишь для простых ограничений, так как значения необходимых целочисленных коэффициентов в некоторых случаях могут быть чрезвычайно большими.
2. Можно использовать специальные отсечения, именуемые **частично-целочисленными**. Они ориентированы на решение задач, в которых лишь часть переменных должна принимать целочисленные значения, а остальные (включая дополнительные) остаются непрерывными. Детальное изложение таких отсечений в этой главе не рассматривается (см. [4]).

Упражнения 9.3,d

1. Запишите отсечения I и II в примере 9.3–5 через уравнения для переменных x_1 и x_2 и покажите, что они совпадают с отсечениями, которые графически показаны на рис. 9.13.
2. Покажите, что дробное отсечение не приводит к допустимому решению в приведенной ниже задаче, если не устранены все дроби в ограничении.

$$\text{Максимизировать } z = x_1 + 2x_2$$

при ограничениях

$$x_1 + \frac{1}{2}x_2 \le \frac{13}{4},$$

$$x_1, x_2 \ge 0 \text{ и целые.}$$

3. Решите следующие задачи методом дробного отсечения и сравните оптимальное целочисленное решение с решением, полученным путем округления соответствующего оптимального непрерывного решения.

   a) Максимизировать $z = 4x_1 + 6x_2 + 2x_3$
   при ограничениях

$$4x_1 - 4x_2 \le 5,$$
$$-x_1 + 6x_2 \le 5,$$
$$-x_1 + x_2 + x_3 \le 5,$$
$$x_1, x_2, x_3 \ge 0 \text{ и целые.}$$

   b) Максимизировать $z = 3x_1 + x_2 + 3x_3$
   при ограничениях

$$-x_1 + 2x_2 + x_3 \le 4,$$
$$4x_2 - 3x_3 \le 2,$$
$$x_1 - 3x_2 + 2x_3 \le 3,$$
$$x_1, x_2, x_3 = 0 \text{ или } 1.$$

9.4. Заключение

Наиболее важным фактором, влияющим на процесс вычислений в целочисленном программировании, является количество целочисленных переменных. Так как имеющиеся алгоритмы не решают задачу ЦЛП последовательно, т.е. не позволяют получать промежуточные целочисленные решения, отличные от оптимального, поэтому, с вычислительной точки зрения, в задаче ЦЛП необходимо ограничиться (по возможности) меньшим числом переменных, принимающих целочисленные значения. Следующие советы могут оказаться полезными при решении практических задач.

1. Аппроксимировать целочисленные переменные непрерывными, где это возможно.
2. Сузить, насколько возможно, интервалы допустимых значений целочисленных переменных.
3. Избегать в модели ЦЛП использования нелинейных элементов.

Уровень развития методов решения целочисленных задач не удовлетворяет требованиям, которые диктуются их практической важностью. Маловероятно, что в целочисленном программировании будет получен новый теоретический прорыв. Представляется более вероятным, что новые технологические достижения в области компьютерной техники могут предложить новые пути повышения эффективности алгоритмов решения задач ЦЛП.

Литература

1. Nemhauser G., Wolsey L. *Integer and Combinatorial Optimization,* Wiley, New York, 1988.
2. Parker G., Rardin R. *Discrete Optimization,* Academic Press, Orlando, FL, 1988.
3. Salkin H., Mathur K. *Foundations of Integer Programming*, North–Holland, New York, 1989.
4. Taha H. *Integer Programming: Theory, Applications, and Computations*, Academic Press, Orlando, FL, 1975.

Литература, добавленная при переводе

Белоусов Е.Г. *Введение в выпуклый анализ и целочисленное программирование.* — М.: Изд-во МГУ, 1977.

Корбут А.А., Финкельштейн Ю.Ю. *Дискретное программирование.* — М.: Наука, 1968.

Ху Т. *Целочисленное программирование и потоки в сетях.* — М.: Мир, 1974.

Комплексные задачи

■ **9-1.** Развивающаяся компания владеет 90 акрами земли в растущей столичной зоне, где планирует построить офисные здания и торговый центр. Созданная собственность сдается в аренду на 7 лет и затем продается. Цена каждого здания оценивается в 10 раз выше суммы чистого дохода, полученного за последний год от сдачи здания в аренду. Компания оценивает, что проект будет включать торговый центр площадью в 4.5 миллиона квадратных футов. Бизнес-план предусматривает строительство трех высотных офисных зданий и четырех зданий с садом.

Перед компанией стоит задача составления расписания работ. Если строительство завершается очень рано, построенные здания могут остаться невостребованными; если строительство завершается очень поздно, могут быть потеряны потенциальные арендаторы. Потребности в офисных площадях на следующие 7 лет, вычисленные на основе соответствующего изучения рынка, показаны в следующей таблице.

Год	Потребность (тысячи кв. футов)	
	В высотных зданиях	В офисах с садом
1	200	100
2	220	110
3	242	121
4	266	133
5	293	146
6	322	161
7	354	177

Следующая таблица содержит предложения по площадям семи зданий.

Здания с садом	Площадь (кв. футы)	Высотные здания	Площадь (кв. футы)
1	60 000	1	350 000
2	60 000	2	450 000
3	75 000	3	350 000
4	75 000	—	—

Суммарный доход от ренты оценивается в 25 долларов за квадратный фут. Текущие затраты равны \$5.75 и \$9.75 за квадратный фут для офисов с садом и высотных зданий соответственно. Затраты на строительство равны \$70 и \$105 за квадратный фут соответственно. Считается, что и стоимость строительства, и рентный доход возрастают примерно в соответствии с процентом инфляции, равным 4%.

Каким образом компания должна спланировать строительство семи зданий?

■ 9-2[1]. В Национальной университетской ассоциации женской спортивной гимнастики проводятся соревнования, которые включают четыре вида спортивных состязаний: опорный прыжок, упражнения на разновысоких брусьях и бревнах, а также вольные упражнения. Каждая команда может выставить шесть участников на каждый вид. Гимнастки оцениваются по десятибалльной шкале. Статистика по университетской команде дает следующие показатели по отдельным гимнасткам.

	Показатели гимнасток					
	1	2	3	4	5	6
Опорный прыжок	6	9	8	8	4	10
Брусья	7	9	7	8	9	5
Бревно	9	8	10	9	9	8
Вольные упражнения	6	6	5	9	10	9

Общее число очков команды определяется путем суммирования пяти лучших индивидуальных показателей для каждого вида состязаний. Участник соревнований может выступать в одном виде или во всех четырех сразу. Эти две возможности взаимно исключают друг друга. Спортсмену разрешается участвовать максимум в трех видах, и по меньшей мере четыре участника команды должны соревноваться по четырем видам. Сформулируйте задачу ЦЛП, которую можно использовать для формирования гимнастической команды, и найдите ее оптимальное решение, используя программу TORA.

■ 9-3[2]. В 1990 году в Соединенных Штатах на телевидении функционировало примерно 180 000 центров рекламы, что позволяло предоставить работу 2 миллионам человек. Исследования показывают, что в 2000 году более 700 000 компаний будут использовать труд примерно 8 миллионов человек для рекламы своей продукции на телевидении. Вопросами первостепенной важности являются необходимое количество центров рекламы на телевидении и их размещение.

Компания ABC занимается решением поставленных вопросов. Рекламный центр может быть расположен в одном из нескольких районов, выбранных компанией, и может обслуживать (частично или полностью) одну или несколько географических зон. Обычно понятие географической зоны ассоциируется с одним или несколькими телефонными кодами. Центры рекламы на телевидении, которыми занимается компания ABC, расположены в восьми зонах с кодами 501, 918, 316, 417, 314, 816, 502 и 606. Следующая таблица содержит информацию о кандидатах на размещение центра; зонах, которые они обслуживают; и о стоимости организации центра.

Размещение центра	Коды обслуживаемых зон	Стоимость (долл.)
Даллас	501, 918, 316, 417	500 000
Атланта	314, 816, 502, 606	800 000
Луисвилл	918, 316, 417, 314, 816	400 000

[1] Задача основана на материалах статьи Ellis P., Corn R. "Using Bivalent Integer Programming to Select Teams of Intercollegiate Women's Gymnastic Competition", *Interfaces*, Vol.14, No. 3, 1984, pp. 41–46.

[2] Задача базируется на работе Spenser T., Brigandi A., Dargon D., Sheehan M. "AT&T's Telemarketing Site Selection System Offers Customer Support", *Interfaces*, Vol. 20, No. 1, 1990, pp. 83–96.

Размещение центра	Коды обслуживаемых зон	Стоимость (долл.)
Денвер	501, 502, 606	900 000
Литл-Рок	417, 314, 816, 502	300 000
Мемфис	606, 501, 316, 417	450 000
Сан-Луис	816, 502, 606, 314	550 000

Клиенты всех зон имеют доступ к каждому из рекламных центров 24 часа в сутки. Стоимость связи между центрами и соответствующими зонами (в долларах за час) приведена в следующей таблице.

	Из зоны с кодом							
В центр	501	918	316	417	314	816	502	606
Даллас	14	35	29	32	25	13	14	20
Атланта	18	18	22	18	26	23	12	15
Луисвилл	22	25	12	19	30	17	26	25
Денвер	24	30	19	14	12	16	18	30
Литл-Рок	19	20	23	16	23	11	28	12
Мемфис	23	21	17	21	20	23	20	10
Сан-Луис	17	18	12	10	19	22	16	22

Компания АВС собирается выбрать три или четыре рекламных центра. Где они должны быть расположены?

Детерминированные модели динамического программирования

10.1. Введение

Динамическое программирование (ДП) определяет оптимальное решение n-мерной задачи путем ее декомпозиции на *n этапов*, каждый из которых представляет подзадачу относительно одной переменной. Вычислительное преимущество такого подхода состоит в том, что мы занимаемся решением одномерных оптимизационных подзадач вместо большой n-мерной задачи. Фундаментальным принципом ДП, составляющим основу декомпозиции задачи на этапы, является **оптимальность**. Так как природа каждого этапа решения зависит от конкретной оптимизационной задачи, ДП не предлагает вычислительных алгоритмов непосредственно для каждого этапа. Вычислительные аспекты решения оптимизационных подзадач на каждом этапе проектируются и реализуются по отдельности (что, конечно, не исключает применения единого алгоритма для всех этапов).

10.2. Рекуррентная природа вычислений в ДП

Вычисления в ДП выполняются рекуррентно в том смысле, что оптимальное решение одной подзадачи используется в качестве исходных данных для следующей. Решив последнюю подзадачу, мы получим оптимальное решение исходной задачи. Способ выполнения рекуррентных вычислений зависит от того, как выполняется декомпозиция исходной задачи. В частности, подзадачи обычно связаны между собой некоторыми общими ограничениями. Если осуществляется переход от одной подзадачи к другой, то должны учитываться эти ограничения.

Пример 10.2–1. (Задача о кратчайшем пути)

Предположим, необходимо выбрать кратчайший путь между двумя городами. Сеть дорог, показанная на рис. 10.1, представляет возможные маршруты между исходным городом, находящимся в узле 1, и конечным пунктом, который находится в узле 7. Маршруты проходят через промежуточные города, обозначенные на сети узлами с номерами 2–6.

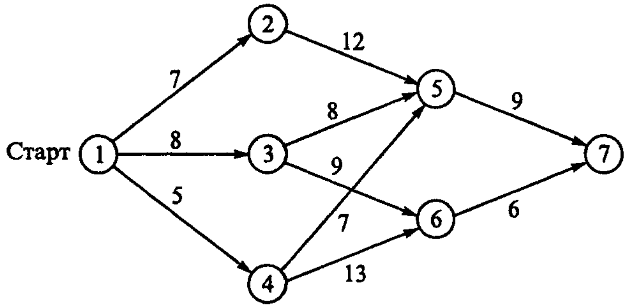


Рис. 10.1

Мы можем решить эту задачу посредством полного перебора всех маршрутов между узлами 1 и 7 (имеется пять таких маршрутов). Однако в большой сети полный перебор является неэффективным с вычислительной точки зрения.

Чтобы решить эту задачу с помощью методов динамического программирования, сначала разделим ее на *этапы*. Вертикальные пунктирные линии на рис. 10.2 очерчивают три этапа задачи. Далее выполняются вычисления для каждого этапа в отдельности.

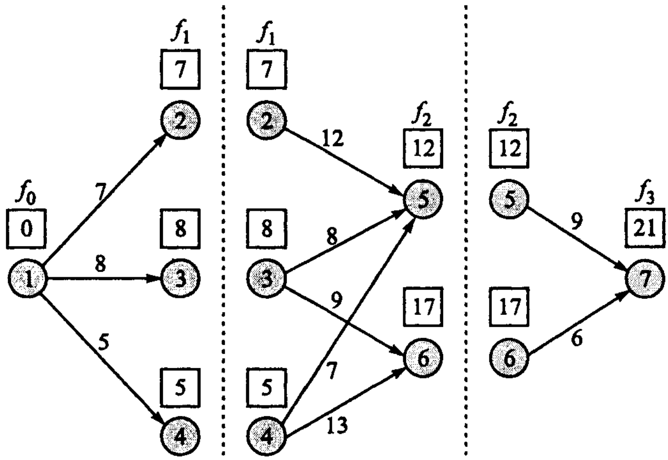


Рис. 10.2

Общая задача состоит в вычислении кратчайших (постепенно накапливаемых) расстояний ко всем вершинам этапа с последующим использованием этих расстояний в качестве исходных данных для следующего этапа. Рассматривая узлы, относящиеся к первому этапу, замечаем, что каждый из узлов 2, 3 и 4 связан с начальным узлом 1 единственной дугой (рис. 10.2). Следовательно, для первого этапа имеем следующее.

Этап 1. Итоговые результаты.

Кратчайший путь к узлу 2 равен 7 миль (из узла 1),
Кратчайший путь к узлу 3 равен 8 миль (из узла 1),
Кратчайший путь к узлу 4 равен 5 миль (из узла 1).

Далее переходим ко второму этапу для вычисления кратчайших (накопленных) расстояний к узлам 5 и 6. Рассматривая узел 5 первым, из рис. 10.2 замечаем, что есть три возможных маршрута, по которым можно достичь узла 5, а именно (2, 5), (3, 5) и (4, 5). Эта информация вместе с кратчайшими расстояниями к узлам 2, 3, и 4 определяет кратчайшее (накопленное) расстояние к узлу 5 следующим образом.

$$\begin{pmatrix}\text{Кратчайший}\\ \text{путь к узлу 5}\end{pmatrix} = \min_{i=2,3,4}\left\{\begin{pmatrix}\text{Кратчайший}\\ \text{путь к узлу } i\end{pmatrix} + \begin{pmatrix}\text{Расстояние от}\\ \text{узла } i \text{ к узлу 5}\end{pmatrix}\right\} =$$

$$= \min\begin{Bmatrix}7+12=19\\ 8+\ \ 8=16\\ 5+\ \ 7=12\end{Bmatrix} = 12 \quad (\text{из узла } 4).$$

Аналогично для узла 6 имеем следующее.

$$\begin{pmatrix}\text{Кратчайший}\\ \text{путь к узлу 6}\end{pmatrix} = \min_{i=3,4}\left\{\begin{pmatrix}\text{Кратчайший}\\ \text{путь к узлу } i\end{pmatrix} + \begin{pmatrix}\text{Расстояние от}\\ \text{узла } i \text{ к узлу 6}\end{pmatrix}\right\} =$$

$$= \min\begin{Bmatrix}8+\ \ 9=17\\ 5+\ 13=18\end{Bmatrix} = 17 \quad (\text{из узла } 3).$$

Этап 2. Итоговые результаты.

Кратчайший путь к узлу 5 равен 12 миль (из узла 4),
Кратчайший путь к узлу 6 равен 17 миль (из узла 3).

Последним шагом является третий этап. Конечный узел 7 можно достигнуть как из узла 5, так и 6. Используя итоговые результаты этапа 2 и расстояния от узлов 5 и 6 к узлу 7, получаем следующее.

$$\begin{pmatrix}\text{Кратчайший}\\ \text{путь к узлу 7}\end{pmatrix} = \min\begin{Bmatrix}12+9=21\\ 17+6=23\end{Bmatrix} = 21 \quad (\text{из узла } 5).$$

Этап 3. Итоговые результаты.

Кратчайший путь к узлу 7 равен 21 миле (из узла 5).

Приведенные вычисления показывают, что кратчайшее расстояние между узлами 1 и 7 равно 21 миле. Города, через которые проходит кратчайший маршрут, определяются следующим образом. Из итоговых результатов третьего этапа следует, что узел 7 связывается с узлом 5. Далее из итоговых результатов второго этапа следует, что узел 4 связывается с узлом 5. Наконец, из итоговых результатов первого этапа следует, что узел 4 связывается с узлом 1. Следовательно, оптимальным маршрутом является последовательность 1–4–5–7.

Теперь покажем, как рекуррентные вычисления динамического программирования можно выразить математически. Пусть $f_i(x_i)$ — кратчайшее расстояние до вершины x_i на этапе i, $d(x_{i-1},x_i)$ — расстояние от узла x_{i-1} до узла x_i. Тогда f_i вычисляется из f_{i-1} с помощью следующего рекуррентного уравнения.

$$f_i(x_i) = \min_{\substack{\text{все допустимые} \\ (x_{i-1}, x_i)\text{-маршруты}}} \{d(x_{i-1}, x_i) + f_{i-1}(x_{i-1})\}, \quad i = 1, 2, 3.$$

При $i = 1$ полагаем $f_0(x_0) \equiv 0$. Это уравнение показывает, что кратчайшие расстояния $f_i(x_i)$ на этапе i должны быть выражены как функции следующего узла x_i. В терминологии динамического программирования x_i именуется *состоянием* системы на этапе i. В действительности *состояние* системы на этапе i — это информация, связывающая этапы между собой, при этом оптимальные решения для оставшихся этапов могут приниматься без повторной проверки того, как были получены решения на предыдущих этапах. Такое определение *состояния* системы позволяет рассматривать каждый этап отдельно и гарантирует, что решение является допустимым на каждом этапе.

Определение *состояния* системы приводит к следующему унифицированному положению.

Принцип оптимальности. На каждом этапе оптимальная стратегия определяется независимо от стратегий, примененных на предыдущих этапах.

Применение принципа оптимальности демонстрируется вычислениями из примера 10.2–1. Например, на этапе 3 мы используем кратчайшие пути к узлам 5 и 6 и не интересуемся, как эти узлы были достигнуты из узла 1.

Упражнения 10.2,а

1. Решите задачу из примера 10.2–1 в предположении, что используются следующие длины маршрутов:

$$d(1, 2) = 5,\ d(1, 3) = 9,\ d(1, 4) = 8,$$
$$d(2, 5) = 10,\ d(2, 6) = 17,$$
$$d(3, 5) = 4,\ d(3, 6) = 10,$$
$$d(4, 5) = 9,\ d(4, 6) = 9,$$
$$d(5, 7) = 8,$$
$$d(6, 7) = 9.$$

2. Я — заядлый турист. Прошлым летом мой друг и я отправились в пятидневный поход по прекрасным Белым Горам в штате Нью-Гемпшир. Мы решили ограничить наше путешествие территорией, на которой находится три хорошо известные вершины: Вашингтон, Джефферсон и Адамс. Гора Вашингтон имеет шестимильную тропу от подножия до вершины. Аналогичные тропы гор Джефферсона и Адамса имеют длину 4 и 5 миль соответственно. Тропы, соединяющие подножия этих трех гор, имеют следующую длину: 3 мили между вершинами Вашингтона и Джефферсона, 2 мили между вершинами Джефферсона и Адамса и 5 миль между вершинами Адамса и Вашингтона. В первый день мы стартовали от подножия вершины Вашингтона и вернулись в эту же точку к концу пятого дня. Нашей целью было пройти как можно более длинный путь. Мы также решили подниматься каждый день только на одну вершину и располагаться лагерем у подножия той горы, на которую мы решили восходить на следующий день. Кроме того, мы решили, что не будем подниматься на одну и ту же вершину в течение двух дней подряд. Каким было расписание нашего похода?

10.3. Рекуррентные алгоритмы прямой и обратной прогонки

В примере 10.2–1 вычисления проводились последовательно: от первого этапа до третьего. Такая последовательность вычислений известна как **алгоритм прямой прогонки**. Этот же пример может быть решен с помощью **алгоритма обратной прогонки,** в соответствии с которым вычисления проводятся от третьего этапа до первого.

Алгоритмы прямой и обратной прогонки приводят к одному и тому же решению. Несмотря на то что алгоритм прямой прогонки представляется более логичным, в специальной литературе, посвященной динамическому программированию, неизменно используется алгоритм обратной прогонки. Причина этого в том, что в общем случае алгоритм обратной прогонки может быть более эффективным с вычислительной точки зрения. Продемонстрируем использование алгоритма обратной прогонки на примере 10.2–1. Мы также представим вычисления динамического программирования в компактной табличной форме.

Пример 10.3–1

Рекуррентное уравнение для алгоритма обратной прогонки в примере 10.2–1 имеет вид

$$f_i(x_i) = \min_{\substack{\text{все допустимые} \\ (x_i, x_{i+1})\text{-маршруты}}} \left\{ d(x_i, x_{i+1}) + f_{i+1}(x_{i+1}) \right\}, \quad i = 1, 2, 3,$$

где $f_4(x_4) \equiv 0$ для $x_4 = 7$. Соответствующей последовательностью вычислений будет $f_3 \to f_2 \to f_1$.

Этап 3. Так как узел 7 ($x_4 = 7$) связан с узлами 5 и 6 ($x_3 = 5$ и 6) в точности одним маршрутом, альтернативы для выбора отсутствуют, а результаты третьего этапа можно подытожить следующим образом.

	$d(x_3, x_4)$	Оптимальное решение	
x_3	$x_4 = 7$	$f_3(x_3)$	x_4^*
5	9	9	7
6	6	6	7

Этап 2. Так как маршрута (2, 6) не существует, соответствующая альтернатива не рассматривается. Используя значения $f_3(x_3)$, полученные на третьем этапе, мы можем сравнить допустимые альтернативные решения, как показано в следующей таблице.

	$d(x_2, x_3) + f_3(x_3)$		Оптимальное решение	
x_2	$x_3 = 5$	$x_3 = 6$	$f_2(x_2)$	x_3^*
2	$12 + 9 = 21$	—	21	5
3	$8 + 9 = 17$	$9 + 6 = 15$	15	6
4	$7 + 9 = 16$	$13 + 6 = 19$	16	5

Оптимальное решение второго этапа означает следующее. Если вы находитесь в узле (городе) 2 или 4, кратчайший путь к узлу 7 проходит через узел 5, а если находитесь в узле 3, кратчайший путь к узлу 7 проходит через узел 6.

Этап 1. Из узла 1 мы имеем три альтернативных маршрута: (1, 2), (1, 3) и (1, 4). Используя значения $f_2(x_2)$, полученные на втором этапе, вычисляем данные следующей таблицы.

x_1	$d(x_1, x_2) + f_2(x_2)$			Оптимальное решение	
	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$f_1(x_1)$	x_2^*
1	7 + 21 = 28	8 + 15 = 23	5 + 16 = 21	21	4

Оптимальное решение на первом этапе показывает, что кратчайший путь проходит через город 4. Далее из оптимального решения на втором этапе следует, что из города 4 необходимо двигаться в город 5. Наконец, из оптимального решения на третьем этапе следует, что город 5 связан с городом 7. Следовательно, полным маршрутом, имеющим кратчайшую длину, является 1–4–5–7, и его длина равна 21 миле.

Упражнения 10.3,а

1. Для задачи из упр. 10.2,а(1) получите рекуррентное соотношение обратной прогонки и используйте его для получения оптимального решения.
2. Для задачи из упр. 10.2,а(2) получите рекуррентное соотношение обратной прогонки и используйте его для получения оптимального решения.
3. Определите кратчайший маршрут между городами 1 и 7 на сети дорог, представленной на рис. 10.3. Определите этапы и состояния системы с помощью алгоритма обратной прогонки, а затем решите задачу.

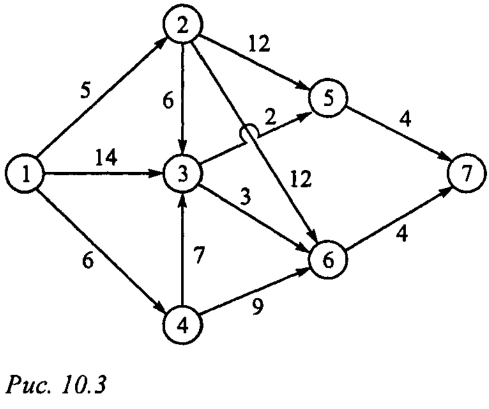


Рис. 10.3

10.4. Некоторые приложения динамического программирования

В данном разделе рассмотрено четыре примера, каждый из которых выбран для демонстрации методов динамического программирования. При рассмотрении каждого примера особое внимание обратите на три основных элемента моделей динамического программирования.

1. Определение *этапов.*
2. Определение на каждом этапе *вариантов решения* (альтернатив).
3. Определение *состояний* на каждом этапе.

Из перечисленных выше элементов понятие *состояния*, как правило, представляется весьма сложным для восприятия. Рассмотренные в этом разделе приложения последовательно показывают, что определение состояния меняется в зависимости от моделируемой ситуации. При рассмотрении каждого приложения полезно ответить на следующие вопросы.

1. Какие соотношения связывают этапы вместе?
2. Какая информация необходима для того, чтобы получить допустимые решения на текущем этапе без повторной проверки решений, принятых на предыдущих этапах?

Мой опыт преподавания показывает, что понятие *состояния* удается глубже уяснить, если поставить под сомнение определение состояния, которое предложено в настоящей книге. Рекомендуем воспользоваться каким-либо другим определением, которое покажется вам "более логичным", и применить его в рекуррентных вычислениях. В конечном счете вы сможете убедиться, что приведенные здесь определения обеспечивают корректное решение задач. В то же время предложенный подход будет способствовать вашему пониманию самой концепции состояния.

10.4.1. Задача о загрузке

Задача о загрузке — это задача о рациональной загрузке судна (самолета, автомашины и т.п.), которое имеет ограничения по объему или грузоподъемности. Каждый помещенный на судно груз приносит определенную прибыль. Задача состоит в определении загрузки судна такими грузами, которые приносят наибольшую суммарную прибыль.

Перед тем как представить соотношения динамического программирования, заметим, что рассматриваемая здесь задача известна также как **задача о снаряжении**, где пилот реактивного самолета должен определить наиболее ценные (необходимые) предметы, которые следует взять на борт самолета, или как **задача о рюкзаке**, в которой солдат (или турист) должен определить наиболее ценные предметы, подлежащие загрузке в ранец (рюкзак). Кажется, что три упомянутых названия для одной и той же задачи были выбраны для того, чтобы гарантировать равное представительство военно-морского флота, воздушных сил и армии!

Рекуррентное уравнение процедуры обратной прогонки выводится для общей задачи загрузки судна грузоподъемностью W предметов (грузов) n наименований. Пусть m_i — количество предметов i-го наименования, подлежащих загрузке, r_i — прибыль, которую приносит один загруженный предмет i-го наименования, w_i — вес одного предмета i-го наименования. Общая задача имеет вид следующей целочисленной задачи линейного программирования.

$$\text{Максимизировать } z = r_1m_1 + r_2m_2 + \dots + r_nm_n$$

при условии, что

$$w_1m_1 + w_2m_2 + \dots + w_nm_n \le W,$$
$$m_1, m_2, \dots, m_n \ge 0 \text{ и целые.}$$

Три элемента модели динамического программирования определяются следующим образом.

1. *Этап* i ставится в соответствие предмету i-го наименования, $i = 1, 2, ..., n$.
2. *Варианты решения* на этапе i описываются количеством m_i предметов i-го наименования, подлежащих загрузке. Соответствующая прибыль равна $r_i m_i$. Значение m_i заключено в пределах от 0 до $[W/w_i]$, где $[W/w_i]$ — целая часть числа W/w_i.
3. *Состояние* x_i на этапе i выражает суммарный вес предметов, решения о погрузке которых приняты на этапах $i, i + 1, ..., n$. Это определение отражает тот факт, что ограничение по весу является единственным, которое связывает n этапов вместе.

Пусть $f_i(x_i)$ — максимальная суммарная прибыль от этапов $i, i + 1, ..., n$ при заданном состоянии x_i. Проще всего рекуррентное уравнение определяется с помощью следующей двухшаговой процедуры.

Шаг 1. Выразим $f_i(x_i)$ как функцию $f_{i+1}(x_{i+1})$ в виде

$$f_i(x_i) = \max_{\substack{m_i = 0,1,\ldots,\left[\frac{W}{w_i}\right] \\ x_i = 0,1,\ldots,W}} \left\{ r_i m_i + f_{i+1}(x_{i+1}) \right\}, \quad i = 1, 2, ..., n,$$

где $f_{n+1}(x_{n+1}) \equiv 0$.

Шаг 2. Выразим x_{i+1} как функцию x_i для гарантии того, что левая часть последнего уравнения является функцией лишь x_i. По определению $x_i — x_{i+1}$ представляет собой вес, загруженный на этапе i, т.е. $x_i - x_{i+1} = w_i m_i$ или $x_{i+1} = x_i - w_i m_i$. Следовательно, рекуррентное уравнение приобретает следующий вид.

$$f_i(x_i) = \max_{\substack{m_i = 0,1,\ldots,\left[\frac{W}{w_i}\right] \\ x_i = 0,1,\ldots,W}} \left\{ r_i m_i + f_{i+1}(x_i - w_i m_i) \right\}, \quad i = 1, 2, ..., n.$$

Пример 10.4–1

В 4-тонный самолет загружаются предметы трех наименований. Приведенная ниже таблица содержит данные о весе одного предмета w_i (в тоннах) и прибыли r_i (в тысячах долларов), получаемой от одного загруженного предмета. Как необходимо загрузить самолет, чтобы получить максимальную прибыль?

Предмет i	w_i	r_i
1	2	31
2	3	47
3	1	14

Так как вес одного предмета w_i для всех наименований и максимальный вес W принимают целочисленные значения, состояние x_i может принимать лишь целочисленные значения.

Этап 3. Точный вес, который может быть загружен на этапе 3 (предмет наименования 3), заранее неизвестен, но он должен принимать одно из значений 0, 1, ..., 4 (так как $W = 4$ тонны). Состояния $x_3 = 0$ и $x_3 = 4$ представляют собой крайние случаи, когда предметы третьего наименования совсем не загружаются или загружают самолет полностью. Остальные значения x_3 (= 1, 2 или 3) предполагают частичную загрузку самолета предметами третьего наименования. Действительно, при этих значениях x_3 все оставшиеся емкости самолета могут быть заполнены предметами третьего наименования.

Так как вес w_3 одного предмета третьего типа равен 1 тонне, максимальное количество единиц этого типа, которое может быть загружено, равно $[4/1] = 4$. Это означает, что возможными значениями x_3 будут 0, 1, 2, 3 и 4. Решение m_3 является допустимым лишь при условии, что $w_3 m_3 \le x_3$. Следовательно, все недопустимые альтернативы (те, для которых $w_3 m_3 > x_3$) исключены. Следующее уравнение является основой для сравнения альтернатив на этапе 3.

$$f_3(x_3) = \max_{m_3}\{14m_3\}, \quad \max\{m_3\} = \left[\frac{4}{1}\right] = 4.$$

В следующей таблице сравниваются допустимые решения для каждого значения x_3.

	$14m_3$					Оптимальное решение	
x_3	$m_3 = 0$	$m_3 = 1$	$m_3 = 2$	$m_3 = 3$	$m_3 = 4$	$f_3(x_3)$	m_3^*
0	0	—	—	—	—	0	0
1	0	14	—	—	—	14	1
2	0	14	28	—	—	28	2
3	0	14	28	42	—	42	3
4	0	14	28	42	56	56	4

Этап 2.

$$f_2(x_2) = \max_{m_2}\{47m_2 + f_3(x_2 - 3m_2)\}, \quad \max\{m_2\} = \left[\frac{4}{3}\right] = 1.$$

	$47m_2 + f_3(x_2 - 3m_2)$		Оптимальное решение	
x_2	$m_2 = 0$	$m_2 = 1$	$f_2(x_2)$	m_2^*
0	$0 + 0 = 0$	—	0	0
1	$0 + 14 = 14$	—	14	0
2	$0 + 28 = 28$	—	28	0
3	$0 + 42 = 42$	$47 + 0 = 47$	47	1
4	$0 + 56 = 56$	$47 + 14 = 61$	61	1

Этап 1.

$$f_1(x_1) = \max_{m_1}\{31m_1 + f_2(x_1 - 2m_1)\}, \quad \max\{m_1\} = \left[\frac{4}{2}\right] = 2.$$

x_1	$31m_1 + f_2(x_1 - 2m_1)$			Оптимальное решение	
	$m_1 = 0$	$m_1 = 1$	$m_1 = 2$	$f_1(x_1)$	m_1^*
0	$0 + 0 = 0$	—	—	0	0
1	$0 + 14 = 14$	—	—	14	0
2	$0 + 28 = 28$	$31 + 0 = 31$	—	31	1
3	$0 + 47 = 47$	$31 + 14 = 45$	—	47	0
4	$0 + 61 = 61$	$31 + 28 = 59$	$62 + 0 = 62$	62	2

Оптимальное решение определяется теперь следующим образом. Из условия $W = 4$ следует, что первый этап решения задачи при $x_1 = 4$ дает оптимальное решение $m_1^* = 2$, которое означает, что два предмета первого наименования будут загружены в самолет. Эта загрузка оставляет $x_2 = x_1 - 2m_1^* = 4 - 2 \times 2 = 0$. Решение на втором этапе при $x_2 = 0$ приводит к оптимальному решению $m_2^* = 0$, которое, в свою очередь, дает $x_3 = x_2 - 3m_2^* = 0 - 3 \times 0 = 0$. Далее этап 3 при $x_3 = 0$ приводит к $m_3^* = 0$. Следовательно, оптимальным решением задачи является $m_1^* = 2$, $m_2^* = 0$ и $m_3^* = 0$. Соответствующая прибыль равна 62 000 долларов.

В таблице для первого этапа нам, по существу, необходимо получить оптимальное решение лишь для $x_1 = 4$, так как это последний этап, подлежащий рассмотрению. Однако в таблицу включены также вычисления для $x_1 = 0$, 1, 2 и 3, которые позволяют провести анализ чувствительности решения. Например, что произойдет, если максимальная грузоподъемность самолета будет 3 тонны вместо 4? Новое оптимальное решение может быть определено, начиная с $x_1 = 3$ на первом этапе и продолжая так, как мы поступали при $x_1 = 4$.

Задача о загрузке является типичным представителем задачи *распределения ресурсов*, в которой ограниченный ресурс распределяется между конечным числом видов (экономической) деятельности. При этом целью является максимизация соответствующей функции прибыли. В таких моделях определение состояния на каждом этапе будет аналогично приведенному для задачи о загрузке: состоянием на этапе i является суммарное количество ресурса, распределяемого на этапах $i, i + 1, \ldots, n$.

Упражнения 10.4,а

1. В задаче примера 10.4–1 определите оптимальное решение, предполагая, что максимальная грузоподъемность самолета составляет 3 тонны.
2. Решите задачу о загрузке из примера 10.4–1 для каждого из следующих случаев.

   a) $w_1 = 4, r_1 = 70, w_2 = 1, r_2 = 20, w_3 = 2, r_3 = 40, W = 6.$
   b) $w_1 = 1, r_1 = 30, w_2 = 2, r_2 = 60, w_3 = 3, r_3 = 80, W = 4.$

3. Турист собирается в путешествие по дикой местности и должен упаковать в рюкзак предметы трех наименований: пищу, средства первой помощи и одежду. Объем рюкзака составляет 3 кубических фута. Каждая единица пищи занимает 1 ку-

бический фут, упаковка средств первой помощи — четверть кубического фута, а отдельный предмет одежды — примерно половину кубического фута. Турист определил свои предпочтения весовыми коэффициентами 3, 4 и 5 — для пищи, средств первой помощи и одежды соответственно. Это означает, что одежда является самым ценным предметом среди остальных. Опыт подсказывает туристу, что он должен взять не менее одного предмета каждого наименования и не более двух комплектов средств первой помощи. Сколько единиц каждого наименования возьмет турист в поход?

4. Студент должен выбрать 10 факультативных курсов на четырех различных факультетах, причем на каждом факультете должен быть выбран по меньшей мере один курс. Эти курсы распределяются между факультетами таким образом, чтобы максимизировать объем "знаний". Студент оценивает знания по шкале в сто баллов и приходит к выводам, представленным в следующей таблице.

Факультет	Номер курса						
	1	2	3	4	5	6	≥7
I	25	50	60	80	100	100	100
II	20	70	90	100	100	100	100
III	40	60	80	100	100	100	100
IV	10	20	30	40	50	60	70

Какие курсы следует выбрать студенту?

5. У меня во дворе имеется небольшой огород 10×20 футов. Этой весной я собираюсь посадить овощи трех видов: помидоры, зеленые бобы и кукурузу. Огород разбит на ряды, длина которых равна 20 футам. Кукуруза и помидоры занимают ряды шириной 2 фута, а зеленые бобы — 3 фута. Помидоры мне нравятся больше, а бобы меньше. По 10-балльной шкале предпочтений я бы присвоил помидорам 10 баллов, кукурузе — 7 баллов и зеленым бобам — 3 балла. Независимо от моих предпочтений, жена настаивает, чтобы я посадил не менее одного ряда зеленых бобов и не более двух рядов помидоров. Сколько рядов каждого вида овощей следует мне посадить?

6. "Жилище для Человечества" — прекрасная благотворительная организация, которая строит дома для бедствующих семей силами добровольцев. Такая семья может выбрать себе дом из трех типоразмеров: 1000, 1100 и 1200 квадратных футов. Дом каждого типоразмера требует выполнения определенного объема работ силами добровольцев. Филиал организации в городе Файтвилл получил пять заявок на предстоящие шесть месяцев. Комитет по надзору дает оценку каждой заявке в численном виде, принимая во внимание различные факторы. Более высокая оценка означает более острую потребность в жилье. В течение предстоящих шести месяцев филиал организации в этом городе может привлечь к работе максимум 23 добровольца. Следующая таблица содержит оценку каждой заявки и необходимое число добровольцев для ее выполнения. Какие заявки следует утвердить комитету?

Заявка	Размер дома (футов^2)	Оценка	Необходимое число добровольцев
1	1200	78	7
2	1000	64	4
3	1100	68	6
4	1000	62	5
5	1200	85	8

7. Шериф округа Вашингтон принимает участие в переизбрании на следующий срок. Денежные средства на предвыборную кампанию составляют примерно 10 000 долларов. Хотя комитет по переизбранию хотел бы провести кампанию во всех пяти избирательных участках округа, ограниченность денежных средств предписывает действовать по-другому. Приведенная ниже таблица содержит данные о числе избирателей и денежных средствах, необходимых для проведения успешной кампании по каждому избирательному участку. Каждый участок может либо использовать все предназначенные деньги, либо вовсе их не использовать. Как следует распределить денежные средства?

Участок	Число избирателей	Необходимые средства ($)
1	3100	3500
2	2600	2500
3	3500	4000
4	2800	3000
5	2400	2000

8. Конструируется электронный прибор, состоящий из трех основных компонентов. Все компоненты соединены последовательно, поэтому выход из строя одного из них влечет за собой отказ всего прибора. Надежность (вероятность безаварийной работы) прибора можно повысить путем дублирования каждого компонента. Конструкция прибора допускает использование одного или двух резервных (параллельных) блоков, т.е. каждый компонент прибора может содержать до трех блоков, соединенных параллельно. Следующая таблица содержит данные о надежности r и стоимости компонентов прибора.

Число параллельных блоков	Компонент 1		Компонент 2		Компонент 3	
	r_1	c_1 ($)	r_2	c_2 ($)	r_3	c_3 ($)
1	0.6	1000	0.7	3000	0.5	2000
2	0.8	2000	0.8	5000	0.7	4000
3	0.9	3000	0.9	6000	0.9	5000

Общая сумма, выделенная на конструирование прибора, равна 10 000 долларов. Как следует сконструировать прибор? (*Совет.* Наша задача состоит в максимизации надежности $r_1r_2r_3$ прибора. Это значит, что целевая функция является мультипликативной, а не аддитивной.)

9. Решите следующую задачу с помощью метода динамического программирования.

$$\text{Максимизировать } z = y_1 y_2 \ldots y_n$$

при условиях

$$y_1 + y_2 + \ldots + y_n = c,$$
$$y_i \geq 0,\ i = 1, 2, \ldots, n.$$

(*Подсказка.* Это упражнение аналогично предыдущему упражнению, но с той лишь разницей, что переменные y_i являются непрерывными.)

10. Решите следующую задачу с использованием метода динамического программирования.

$$\text{Минимизировать } z = y_1^2 + y_2^2 + \ldots + y_n^2$$

при условиях

$$y_1 y_2 \ldots y_n = c,$$
$$y_i > 0,\ i = 1, 2, \ldots, n.$$

11. Решите следующую задачу посредством метода динамического программирования.

$$\text{Максимизировать } z = (y_1 + 2)^2 + y_2 y_3 + (y_4 - 5)^2$$

при условиях

$$y_1 + y_2 + y_3 + y_4 \leq 5,$$
$$y_i \geq 0 \text{ и целые},\ i = 1, 2, 3, 4.$$

12. Решите следующую задачу с помощью метода динамического программирования.

$$\text{Минимизировать } z = \max\{f(y_1), f(y_2), \ldots, f(y_n)\}$$

при условиях

$$y_1 + y_2 + \ldots + y_n = c,$$
$$y_i \geq 0,\ i = 1, 2, \ldots, n.$$

Найдите решение задачи при условии, что $n = 3$, $c = 10$, $f(y_1) = y_1 + 5$, $f(y_2) = 5y_2 + 3$ и $f(y_3) = y_3 - 2$.

10.4.2. Задача планирования рабочей силы

При выполнении некоторых проектов число рабочих, необходимых для выполнения какого-либо проекта, регулируется путем их найма и увольнения. Поскольку как наем, так и увольнение рабочих связано с дополнительными затратами, необходимо определить, каким образом должна регулироваться численность рабочих в период реализации проекта.

Предположим, что проект будет выполняться в течение n недель и минимальная потребность в рабочей силе на протяжении i-й недели составит b_i рабочих. При идеальных условиях хотелось бы на протяжении i-й недели иметь в точности b_i рабочих. Однако в зависимости от стоимостных показателей может быть более выгодным отклонение численности рабочей силы как в одну, так и в другую сторону от минимальных потребностей. Если x_i — количество работающих на протяжении i-й недели, то возможны затраты двух видов: 1) $C_1(x_i - b_i)$ — затраты, связанные с необходимостью содержать избыток $x_i - b_i$ рабочей силы и 2) $C_2(x_i - x_{i-1})$ — затраты, связанные с необходимостью дополнительного найма $x_i - x_{i-1}$ рабочих.

Элементы модели динамического программирования определяются следующим образом.

1. *Этап* i представляется порядковым номером недели i, $i = 1, 2, \ldots, n$.
2. *Вариантами решения* на i-м этапе являются значения x_i — количество работающих на протяжении i-й недели.
3. *Состоянием* на i-м этапе является x_{i-1} — количество работающих на протяжении $(i - 1)$-й недели (этапа).

Рекуррентное уравнение динамического программирования представляется в виде

$$f_i\left(x_{i-1}\right) = \min_{x_i \geq b_i}\left\{C_1\left(x_i - b_i\right) + C_2\left(x_i - x_{i-1}\right) + f_{i+1}\left(x_i\right)\right\}, \quad i = 1, 2, \ldots, n,$$

где $f_{n+1}(x_n) \equiv 0$. Вычисления начинаются с этапа n при $x_n = b_n$ и заканчиваются на этапе 1.

Пример 10.4–2

Строительный подрядчик оценивает минимальные потребности в рабочей силе на каждую из последующих пяти недель следующим образом: 5, 7, 8, 4 и 6 рабочих соответственно. Содержание избытка рабочей силы обходится подрядчику в 300 долларов за одного рабочего в неделю, а наем рабочей силы на протяжении одной недели обходится в 400 долларов плюс 200 долларов за одного рабочего в неделю.

Выражая C_1 и C_2 в сотнях долларов, имеем следующее.

$$b_1 = 5,\ b_2 = 7,\ b_3 = 8,\ b_4 = 4,\ b_5 = 6,$$
$$C_1(x_i - b_i) = 3(x_i - b_i),\ \ x_i > b_i,\ \ i = 1, 2, \ldots, 5,$$
$$C_2(x_i - x_{i-1}) = 4 + 2(x_i - x_{i-1}),\ \ x_i > x_{i-1},\ \ i = 1, 2, \ldots, 5.$$

Этап 5. ($b_5 = 6$).

	$C_1(x_5 - 6) + C_2(x_5 - x_4)$	Оптимальное решение	
x_4	$x_5 = 6$	$f_5(x_4)$	x_5^*
4	$3(0) + 4 + 2(2) = 8$	8	6
5	$3(0) + 4 + 2(1) = 6$	6	6
6	$3(0) + 0 = 0$	0	6

Этап 4. ($b_4 = 4$).

	$C_1(x_4 - 4) + C_2(x_4 - x_3) + f_5(x_4)$			Оптимальное решение	
x_3	$x_4 = 4$	$x_4 = 5$	$x_4 = 6$	$f_4(x_3)$	x_4^*
8	$3(0) + 0 + 8 = 8$	$3(1) + 0 + 6 = 9$	$3(2) + 0 + 0 = 6$	6	6

Этап 3. ($b_3 = 8$).

x_2	$C_1(x_3 - 8) + C_2(x_3 - x_2) + f_4(\tilde{o}_3)$	Оптимальное решение	
	$x_3 = 8$	$f_3(x_2)$	x_3^*
7	$3(0) + 4 + 2(1) + 6 = 12$	12	8
8	$3(0) + 0 + 6 = 6$	6	8

Этап 2. ($b_2 = 7$).

x_1	$C_1(x_2 - 7) + C_2(x_3 - x_2) + f_3(x_2)$		Оптимальное решение	
	$x_2 = 7$	$x_2 = 8$	$f_2(x_1)$	x_2^*
5	$3(0) + 4 + 2(2) + 12 = 20$	$3(1) + 4 + 2(3) + 6 = 19$	19	8
6	$3(0) + 4 + 2(1) + 12 = 18$	$3(1) + 4 + 2(2) + 6 = 17$	17	8
7	$3(0) + 0 + 12 = 12$	$3(1) + 4 + 2(1) + 6 = 15$	12	7
8	$3(0) + 0 + 12 = 12$	$3(1) + 0 + 6 = 9$	9	8

Этап 1. ($b_1 = 5$).

x_0	$C_1(x_1 - 5) + C_2(x_1 - x_0) + f_2(x_1)$				Оптимальное решение	
	$x_1 = 5$	$x_1 = 6$	$x_1 = 7$	$x_1 = 8$	$f_1(x_0)$	x_1^*
0	$3(0) + 4 + 2(5) + 19 = 33$	$3(1) + 4 + 2(6) + 17 = 36$	$3(2) + 4 + 2(7) + 12 = 36$	$3(2) + 4 + 2(8) + 9 = 35$	33	5

Оптимальное решение определяется последовательно следующим образом.

$$x_0 = 0 \to x_1^* = 5 \to x_2^* = 8 \to x_3^* = 8 \to x_4^* = 6 \to x_5^* = 6.$$

Полученному решению соответствует следующий план.

Номер недели (i)	Минимум рабочей силы (b_i)	Количество фактически работающих (x_i)	Решение
1	5	5	Нанять 5 рабочих
2	7	8	Нанять 3 рабочих
3	8	8	Ничего не менять
4	4	6	Уволить 2 рабочих
5	6	6	Ничего не менять

Упражнения 10.4,b

1. Решите задачу из примера 10.4–2 при следующих минимальных потребностях в рабочей силе.

   a) $b_1 = 6, b_2 = 5, b_3 = 3, b_4 = 6, b_5 = 8$.

   b) $b_1 = 8, b_2 = 4, b_3 = 7, b_4 = 8, b_5 = 2$.

2. Пусть в примере 10.4–2 каждому уволенному рабочему выплачивается выходное пособие в размере 100 долларов. Найдите оптимальное решение задачи.

3. Туристическое агентство организовывает недельные поездки в Египет. В соответствии с договором на ближайшие четыре недели агентство должно обеспечить туристические группы арендными автомобилями в количестве семь, четыре, семь и восемь штук соответственно. Агентство заключает договор с местным дилером по прокату автомобилей. Дилер назначает арендную плату за один автомобиль 220 долларов в неделю плюс 500 долларов за любую арендную сделку. Агентство, однако, может не возвращать арендованные автомобили в конце недели, и в этом случае оно должно будет только арендную плату в 220 долларов. Каково оптимальное решение проблемы, связанной с арендой автомобилей?

4. Компания на следующие четыре года заключила контракт на поставку авиационных двигателей, по 4 двигателя в год. Доступные производственные мощности и стоимость производства меняются от года к году. Компания может изготовить пять двигателей за 1-й год, шесть — за 2-й, три — за 3-й и пять — за 4-й. Стоимость производства одного двигателя на протяжении следующих четырех лет равна соответственно 300 000, 330 000, 350 000 и 420 000 долларов. В течение года компания может произвести больше двигателей, чем необходимо, но в этом случае двигатели должны надлежащим образом храниться до их отгрузки потребителю. Стоимость хранения одного двигателя также меняется от года к году и оценивается в 20 000 долларов для первого года, 30 000 долларов — для второго, 40 000 долларов — для третьего и 50 000 долларов — для четвертого. В начале первого года компания имеет один двигатель, готовый к отгрузке. Разработайте оптимальный план производства двигателей.

5. Фирма выпускает пять типов электронных игр ($E1$, $E2$, ..., $E5$) и пять типов механических игрушек ($M1$, $M2$, ..., $M5$). На рынке порядок предпочтения электронных игр таков: $E1 \to E2 \to \ldots \to E5$. Это означает, что клиент будет покупать игру с более высоким предпочтением, если она имеется в продаже. Известен также порядок предпочтения механических игрушек: $M1 \to M2 \to \ldots \to M5$. Недельный спрос на пять типов электронных игр равен 100, 180, 90, 250 и 190 единиц соответственно. Аналогичные показатели для механических игрушек равны 300, 190, 240, 280 и 260 единиц соответственно. Производство одной игры $E1$, $E2$, ..., $E5$ обходится в 10, 12, 8, 9 и 6 долларов соответственно. Изготовление же одной игрушки $M1$, $M2$, ..., $M5$ обходится фирме в 4, 5, 3, 2 и 3 доллара соответственно. Организация производства каждой электронной игры или игрушки обходится в 500 долларов. Определите оптимальный план производства игрушек.

10.4.3. Задача замены оборудования

Чем дольше механизм эксплуатируется, тем выше затраты на его обслуживание и ниже его производительность. Когда срок эксплуатации механизма достигает определенного уровня, может оказаться более выгодной его замена. Задача замены оборудования, таким образом, сводится к определению оптимального срока эксплуатации механизма.

Предположим, что мы занимаемся заменой механизмов на протяжении n лет. В *начале* каждого года принимается решение либо об эксплуатации механизма еще один год, либо о замене его новым. Обозначим через $r(t)$ и $c(t)$ прибыль от эксплуатации t-летнего механизма на протяжении года и затраты на его обслуживание за этот же период. Далее пусть $s(t)$ — стоимость продажи механизма, который эксплуатировался t лет. Стоимость приобретения нового механизма остается неизменной на протяжении всех лет и равна I.

Элементы модели динамического программирования таковы.

1. *Этап* i представляется порядковым номером года i, $i = 1, 2, ..., n$.
2. *Вариантами решения* на i-м этапе (т.е. для i-го года) являются альтернативы: *продолжить эксплуатацию* или *заменить* механизм в *начале* i-го года.
3. *Состоянием* на i-м этапе является срок эксплуатации t (возраст) механизма к началу i-го года.

Пусть $f_i(t)$ — максимальная прибыль, получаемая за годы от i до n при условии, что в начале i-го года имеется механизм t-летнего возраста.

Рекуррентное уравнение имеет следующий вид.

$$f_i(t) = \max \left\{ \begin{array}{ll} r(t) - c(t) + f_{i+1}(t+1), & \text{если эксплуатировать механизм,} \\ r(0) + s(t) - I - c(0) + f_{i+1}(1), & \text{если заменить механизм,} \end{array} \right\},$$

где $f_{n+1}(.) \equiv 0$.

Пример 10.4–3

Компания планирует определить оптимальную политику замены имеющегося в настоящее время трехлетнего механизма на протяжении следующих 4 лет ($n = 4$), т.е. вплоть до *начала* пятого года. Приведенная таблица содержит относящиеся к задаче данные. Компания требует замены механизма, который находится в эксплуатации 6 лет. Стоимость нового механизма равна 100 000 долларов.

Возраст t (года)	Прибыль $r(t)$ (\$)	Стоимость обслуживания $c(t)$ (\$)	Остаточная стоимость $s(t)$ (\$)
0	20 000	200	—
1	19 000	600	80 000
2	18 500	1200	60 000
3	17 200	1500	50 000
4	15 500	1700	30 000
5	14 000	1800	10 000
6	12 200	2200	5 000

Определение допустимых значений возраста механизма на каждом этапе является нетривиальной задачей. На рис. 10.4 представлена рассматриваемая задача замены оборудования в виде сети. В начале первого года имеется механизм трехлетнего возраста. Мы можем либо заменить его (*З*), либо эксплуатировать (*С*) на протяжении следующего года. Если механизм заменили, то в начале второго года его возраст будет равен одному году, в противном случае его возраст будет 4 года. Такой же подход используется в начале каждого года, начиная со второго по четвертый.

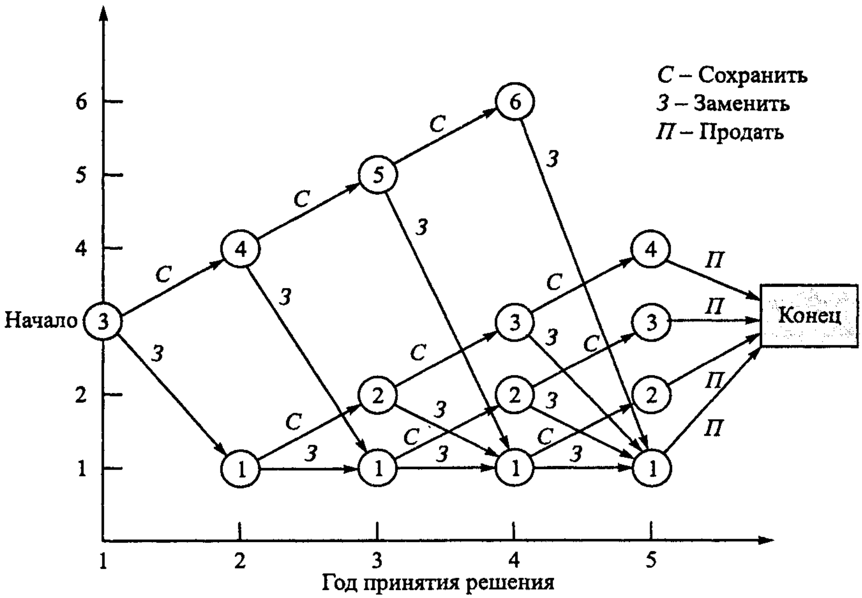


Рис. 10.4

Если однолетний механизм заменяется в начале второго или третьего года, то заменивший его механизм к началу следующего года также будет однолетним. К тому же, в начале 4-го года 6-летний механизм обязательно должен быть заменен, если он еще эксплуатируется; в конце 4-го года все механизмы продаются (*П*) в обязательном порядке. На схеме сети также видно, что в начале второго года возможны только механизмы со сроком эксплуатации 1 или 4 года. В начале третьего года механизм может иметь возраст 1, 2 или 5 лет, а в начале четвертого — 1, 2, 3 или 6 лет.

Решение данной задачи эквивалентно нахождению маршрута максимальной длины (т.е. приносящего максимальную прибыль) от начала первого года к концу четвертого в сети, показанной на рис. 10.4. При решении этой задачи используем табличную форму записи. (Числовые данные в таблице кратны тысячам долларов.)

Этап 4.

	С	*З*	Оптимум	
t	$r(t) + s(t+1) - c(t)$	$r(0) + s(t) + s(1) - c(0) - I$	$f_4(t)$	Решение
1	19.0 + 60 – 0.6 = 78.4	20 +80 + 80 – 0.2 – 100 = 79.8	79.8	*З*
2	18.5 + 50 – 1.2 = 67.3	20 + 60 + 80 – 0.2 – 100 = 59.8	67.3	*С*
3	17.2 + 30 – 1.5 = 45.7	20 + 50 + 80 – 0.2 – 100 = 49.8	49.8	*З*
6	Необходима замена	20 + 5 + 80 – 0.2 – 100 = 4.8	4.8	*З*

Этап 3.

	С	*З*	Оптимум	
t	$r(t) - c(t) + f_4(t+1)$	$r(0) + s(t) - c(0) - I + f_4(1)$	$f_3(t)$	Решение
1	19.0 – 0.6 + 67.3 = 85.7	20 + 80 – 0.2 – 100 + 79.8 = 79.6	85.7	*С*
2	18.5 – 1.2 + 49.8 = 67.1	20 + 60 – 0.2 – 100 + 79.8 = 59.6	67.1	*С*
5	14.0 – 1.8 + 4.8 = 17.0	20 + 10 – 0.2 – 100 + 79.8 = 9.6	17.0	*С*

Этап 2.

	С	*З*	Оптимум	
t	$r(t) - c(t) + f_3(t+1)$	$r(0) + s(t) - c(0) - I + f_3(1)$	$f_2(t)$	Решение
1	$19.0 - 0.6 + 67.1 = 85.5$	$20 + 80 - 0.2 - 100 + 85.7 = 85.5$	85.5	*С* или *З*
4	$15.5 - 1.7 + 17.0 = 30.8$	$20 + 30 - 0.2 - 100 + 85.7 = 35.5$	35.5	*З*

Этап 1.

	С	*З*	Оптимум	
t	$r(t) - c(t) + f_2(t+1)$	$r(0) + s(t) - c(0) - I + f_2(1)$	$f_1(t)$	Решение
1	$17.2 - 1.5 + 35.5 = 51.2$	$20 + 50 - 0.2 - 100 + 85.5 = 55.3$	55.3	*З*

На рис. 10.5 показана последовательность получения оптимального решения. В начале первого года оптимальным решением при $t = 3$ является замена механизма. Следовательно, новый механизм к началу второго года будет находиться в эксплуатации 1 год. При $t = 1$ в начале второго года оптимальным решением будет либо использование, либо замена механизма. Если он заменяется, то новый к началу третьего года будет находиться в эксплуатации 1 год, иначе механизм будет иметь возраст 2 года. Описанный процесс продолжается до тех пор, пока не будет определено оптимальное решение для четвертого года.

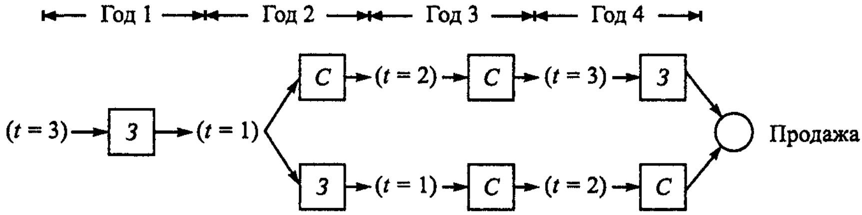


Рис. 10.5

Следовательно, начиная с первого года эксплуатации механизма, альтернативными оптимальными стратегиями относительно замены механизма будут (*З, С, С, З*) и (*З, З, С, С*). Общая прибыль составит 55 300 долларов.

Упражнения 10.4,с

1. Постройте сеть и найдите оптимальное решение в задаче из примера 10.4–3 в каждом из следующих случаев.

   a) В начале первого года имеется механизм, находящийся в эксплуатации 2 года.

   b) В начале первого года имеется механизм, находящийся в эксплуатации 1 год.

   c) В начале первого года куплен новый механизм.

2. Мой тринадцатилетний сын занимается собственным бизнесом — косит газоны десяти клиентам. Каждому клиенту он косит траву три раза в год, получая за один скошенный газон 50 долларов. Он купил косилку за 200 долларов. На протяжении первого года затраты на содержание и использование косилки равны 120 долла-

ров, и через год они увеличиваются на 20%. Одногодичная косилка может быть продана за 150 долларов, и с каждым годом ее стоимость уменьшается на 10%. Мой сын планирует продолжить свой бизнес, пока ему не исполнится 16 лет, и считает, что более выгодно менять косилку через каждые два года. Он объясняет это тем, что цена новой косилки увеличивается за год лишь на 10%. Справедливо ли его решение?

3. Группа ферм владеет трактором двухлетней давности и планирует разработать стратегию его замены на следующие пять лет. Трактор должен эксплуатироваться не менее двух и не более пяти лет. В настоящее время новый трактор стоит 40 000 долларов, и эта цена за год увеличивается на 10%. Текущая годичная стоимость эксплуатации трактора составляет 1300 долларов и, как ожидается, будет увеличиваться на 10% в год.

   a) Сформулируйте задачу в виде задачи о кратчайшем пути.
   b) Постройте соответствующее рекуррентное уравнение.
   c) Определите оптимальную стратегию замены трактора на следующие пять лет.

4. Рассмотрим задачу замены оборудования на протяжении n лет. Цена новой единицы оборудования равна c долларов, а стоимость продажи после t лет эксплуатации равна $s(t) = 2(n - t)$ при $n > t$ и нулю — в противном случае. Годичная прибыль от эксплуатации является функцией возраста оборудования t и равна $r(t) = n^2 - t^2$ при $n > t$ и нулю — в противном случае.

   a) Сформулируйте задачу как модель динамического программирования.
   b) Определите оптимальную стратегию замены оборудования двухгодичной давности при c = 10 000 долларов, считая, что $n = 5$.

5. Решите задачу из предыдущего упражнения, предполагая, что возраст оборудования составляет 1 год и $n = 4$, $c = 6000$ долларов, $r(t) = n/(n + 1)$.

10.4.4. Задача инвестирования

Предположим, что в начале каждого из следующих n лет необходимо сделать инвестиции $P_1, P_2, ..., P_n$ соответственно. Вы имеете возможность вложить капитал в два банка: первый банк выплачивает годовой сложный процент r_1, а второй — r_2. Для поощрения депозитов оба банка выплачивают новым инвесторам премии в виде процента от вложенной суммы. Премиальные меняются от года к году, и для i-го года равны q_{i1} и q_{i2} в первом и втором банках соответственно. Они выплачиваются в конце года, на протяжении которого сделан вклад, и могут быть инвестированы в один из двух банков на следующий год. Это значит, что лишь указанные проценты и новые деньги могут быть инвестированы в один из двух банков. Размещенный в банке вклад должен находиться там до конца рассматриваемого периода. Необходимо разработать стратегию инвестиций на следующие n лет.

Элементы модели динамического программирования следующие.

1. *Этап* i представляется порядковым номером года i, $i = 1, 2, ..., n$.
2. *Вариантами решения* на i-м этапе (для i-го года) являются суммы I_i и $\overline{I}_i$ инвестиций в первый и второй банк соответственно.

3. *Состоянием* x_i на i-м этапе является сумма денег на начало i-го года, которые могут быть инвестированы.

Заметим, что по определению $\overline{I}_i = x_i - I_i$. Следовательно,

$$x_i = P_i + q_{i-1,1}I_{i-1} + q_{i-1,2}(x_{i-1} - I_{i-1}) =$$
$$= P_i + (q_{i-1,1} - q_{i-1,2})I_{i-1} + q_{i-1,2}x_{i-1},$$

где $i = 2, 3, ..., n$, $x_1 = P_1$. Сумма денег x_i, которые могут быть инвестированы, включает лишь новые деньги и премиальные проценты за инвестиции, сделанные на протяжении $(i-1)$-го года.

Пусть $f_i(x_i)$ — оптимальная сумма инвестиций для интервала от i-го до n-го года при условии, что в начале i-го года имеется денежная сумма x_i. Далее обозначим через s_i накопленную сумму к концу n-го года при условии, что I_i и $(x_i - I_i)$ — объемы инвестиций на протяжении i-го года в первый и второй банк соответственно. Обозначая $\alpha_i = (1 + r_i)$, $i = 1, 2$, мы можем сформулировать задачу в следующем виде.

$$\text{Максимизировать } z = s_1 + s_2 + ... + s_n,$$

где

$$s_i = I_i\alpha_1^{n+1-i} + (x_i - I_i)\alpha_2^{n+1-i} = \left(\alpha_1^{n+1-i} - \alpha_2^{n+1-i}\right)I_i + \alpha_2^{n+1-i}x_i, \quad i = 1, 2, ..., n-1,$$
$$s_n = (\alpha_1 + q_{n1} - \alpha_2 - q_{n2})I_n + (\alpha_2 + q_{n2})x_n.$$

Так как премиальные за n-й год являются частью накопленной денежной суммы от инвестиций, в выражения для s_n добавлены q_{n1} и q_{n2}.

Итак, в данном случае рекуррентное уравнение для обратной прогонки в алгоритме динамического программирования имеет вид

$$f_i(x_i) = \max_{0 \le I_i \le x_i}\left\{s_i + f_{i+1}(x_{i+1})\right\}, \quad i = 1, 2, ..., n-1,$$

где x_{i+1} выражается через x_i в соответствии с приведенной выше формулой, а $f_{n+1}(x_{n+1}) \equiv 0$.

Пример 10.4–4

Предположим, вы хотите инвестировать 4000 долларов сейчас и 2000 долларов в начале каждого года, от второго до четвертого, считая от текущего года. Первый банк выплачивает годовой сложный процент 8% и премиальные на протяжении следующих четырех лет в размере 1.8%, 1.7%, 2.1% и 2.5% соответственно. Годовой сложный процент, предлагаемый вторым банком, на 0.2% ниже, чем предлагает первый банк, но его премиальные на 0.5% выше. Задача состоит в максимизации накопленного капитала к концу четвертого года.

Используя введенные выше обозначения, имеем следующее.

$$P_1 = \$4\,000,\ P_2 = P_3 = P_4 = \$2\,000,$$
$$\alpha_1 = (1 + 0.08) = 1.08,$$
$$\alpha_2 = (1 + 0.078) = 1.078,$$
$$q_{11} = 0.018,\ q_{21} = 0.017,\ q_{31} = 0.021,\ q_{41} = 0.025,$$
$$q_{12} = 0.023,\ q_{22} = 0.022,\ q_{32} = 0.026,\ q_{42} = 0.030.$$

Этап 4.

$$f_4(x_4) = \max_{0 \le I_4 \le x_4} \{s_4\},$$

где $s_4 = (\alpha_1 + q_{41} - \alpha_2 - q_{42})I_4 + (\alpha_2 + q_{42})x_4 = -0.003I_4 + 1.108x_4$.

Функция s_4 является линейной по I_4 в области $0 \le I_4 \le x_4$, и, следовательно, ее максимум достигается при $I_4 = 0$ из-за отрицательного коэффициента при I_4. Следовательно, оптимальное решение для этапа 4 может быть представлено в следующем виде.

Состояние	Оптимальное решение	
	$f_4(x_4)$	I_4^*
x_4	$1.108x_4$	0

Этап 3.

$$f_3(x_3) = \max_{0 \le I_3 \le x_3} \{s_3 + f_4(x_4)\},$$

где

$$s_3 = (1.08^2 - 1.078^2)I_3 + 1.078^2 x_3 = 0.00432I_3 + 1.1621x_3,$$
$$x_4 = 2000 - 0.005I_3 + 0.026x_3.$$

Следовательно,

$$f_3(x_3) = \max_{0 \le I_3 \le x_3} \{0.00432I_3 + 1.1621x_3 + 1.108(2000 - 0.005I_3 + 0.026x_3)\} =$$
$$= \max_{0 \le I_3 \le x_3} \{2216 - 0.00122I_3 + 1.1909x_3\}.$$

Состояние	Оптимальное решение	
	$f_3(x_3)$	I_3^*
x_3	$2216 + 1.1909x_3$	0

Этап 2.

$$f_2(x_2) = \max_{0 \le I_2 \le x_2} \{s_2 + f_3(x_3)\},$$

где

$$s_2 = (1.08^3 - 1.078^3)I_2 + 1.078^3 x_2 = 0.006985I_2 + 1.25273x_2,$$
$$x_3 = 2000 - 0.005I_2 + 0.022x_2.$$

Следовательно,

$$f_2(x_2) = \max_{0 \le I_2 \le x_2} \{0.006985I_2 + 1.2527x_2 + 2216 + 1.1909(2000 - 0.005I_2 + 0.022x_2)\} =$$
$$= \max_{0 \le I_2 \le x_2} \{4597.8 + 0.0010305I_2 + 1.27893x_2\}.$$

Состояние	Оптимальное решение	
	$f_2(x_2)$	I_2^*
x_2	$4597.8 + 1.27996x_2$	x_2

Этап 1.

$$f_1(x_1) = \max_{0 \le I_1 \le x_1} \{s_1 + f_2(x_2)\},$$

где

$$s_1 = (1.08^4 - 1.078^4)I_1 + 1.078^4 x_1 = 0.01005I_1 + 1.3504x_1,$$
$$x_2 = 2000 - 0.005I_1 + 0.023x_1.$$

Следовательно,

$$f_1(x_1) = \max_{0 \le I_1 \le x_1} \{0.01005I_1 + 1.3504x_1 + 4597.8 + 1.27996(2000 - 0.005I_1 + 0.023x_1)\} =$$
$$= \max_{0 \le I_1 \le x_1} \{7157.7 + 0.00365I_1 + 1.37984x_1\}.$$

Состояние	Оптимальное решение	
	$f_1(x_1)$	I_1^*
$x_1 = \$4000$	$7157.7 + 1.38349x_1$	\$4000

При вычислениях в обратном направлении получаем следующее.

$$x_2 = 2000 - 0.005 \times 4000 + 0.023 \times 4000 = \$2072,$$
$$x_3 = 2000 - 0.005 \times 2072 + 0.022 \times 2072 = \$2035.22,$$
$$x_4 = 2000 - 0.005 \times 0 + 0.026 \times 2035.22 = 2052.92.$$

Следовательно, оптимальное решение будет записано следующим образом.

Оптимальное решение	Решение, принимаемое инвестором
$I_1^* = x_1$	Инвестировать x_1 = \$4000 в первый банк
$I_2^* = x_2$	Инвестировать x_2 = \$2072 в первый банк
$I_3^* = 0$	Инвестировать x_3 = \$2035.22 во второй банк
$I_4^* = 0$	Инвестировать x_4 = \$2052.92 во второй банк

Упражнения 10.4,d

1. Решите задачу из примера 10.4–4 в предположении, что $r_1 = 0.085$, $r_2 = 0.08$. Кроме того, пусть $P_1 = \$5000$, $P_2 = \$4000$, $P_3 = \$3000$ и $P_4 = \$2000$.
2. Некий инвестор с начальным капиталом в 10 000 долларов должен решить в конце каждого года, сколько денег истратить и сколько инвестировать. Каждый инвестированный доллар возвращает $\alpha = 1.09$ доллара в конце года. Истраченные y долларов на протяжении каждого года приносят удовлетворение, определяемое количественно как эквивалент получения $g(y) = \sqrt{y}$ долларов. Решите задачу с помощью методов динамического программирования для периода в $n = 5$ лет.

3. Фермер имеет k овец. В конце каждого года он принимает решение, сколько овец продать и сколько оставить. Прибыль от продажи одной овцы в i-й год равна p_i. Количество овец в конце i-го года удваивается к концу $(i + 1)$-го года. Фермер планирует в конце n-го года полностью продать овец.

   a) Получите общее рекуррентное уравнение для решения задачи.
   b) Решите задачу при следующих данных: $n = 3$ года, $k = 2$ овцы, $p_1 = \$100$, $p_2 = \$130$, $p_3 = \$120$.

10.4.5. Модели управления запасами

Важной областью применения методов динамического программирования являются задачи управления запасами. В главах 11 и 16 рассмотрены некоторые задачи этого класса, при этом в главе 11 рассматриваются детерминированные модели, а в главе 16 — стохастические.

10.5. Проблема размерности

Во всех рассмотренных выше задачах динамического программирования *состояние* системы на любом этапе описывалось единственной переменной. Например, в задаче о загрузке (раздел 10.4.1) вес предмета является единственным ограничением, которое учитывается при его погрузке. Вместе с этим объем судна также может быть ограничительной величиной. В этом случае говорят, что *состояние* системы является двухмерным, так как формируется двумя переменными: весом и объемом.

Увеличение числа переменных состояния системы влечет за собой увеличение объема вычислений на каждом этапе. Особенно это заметно в моделях динамического программирования при вычислениях с использованием таблиц, так как количество строк каждой таблицы должно соответствовать всем возможным комбинациям значений переменных состояния. Эти вычислительные трудности настолько значительны в динамическом программировании, что в литературе на них ссылаются как на **проклятие размерности**.

Следующий пример приводится для иллюстрации *проклятия размерности*. Он также демонстрирует возможность решения задачи линейного программирования методами динамического программирования.

Пример 10.5–1

Предприятие обрабатывающей промышленности выпускает два вида продукции. Производственный процесс составляет 430 минут в день. Для производства единицы продукции первого вида требуется 2 минуты, а второго — 1 минута. На дневной объем производства продукции первого вида ограничений нет (кроме возможностей производственного процесса), максимальный ежедневный спрос на второй вид продукции равен 230 единиц. Реализация единицы продукции первого вида приносит прибыль в 2 доллара, а второго — 5 долларов. Необходимо найти оптимальное решение задачи максимизации прибыли методами динамического программирования.

Данная задача является следующей задачей линейного программирования.

$$\text{Максимизировать } z = 2x_1 + 5x_2$$

при ограничениях

$$
\begin{aligned}
&2x_1 + x_2 \le 430,\\
&x_2 \le 230,\\
&x_1, x_2 \ge 0.
\end{aligned}
$$

Элементы модели динамического программирования таковы.

1. *Этап* i соответствует продукции i, $i = 1, 2$.
2. *Альтернативой* x_i на i-м этапе является объем производства продукции i, $i = 1, 2$.
3. *Состояние* (v_1, w_1) представляет количество ресурсов, необходимое для производства продукции вида 1 и 2 (производственное время и ограничение на спрос) и используемое на этапах 1 и 2.
4. *Состояние* (v_2, w_2) представляет количество ресурсов, необходимое для производства продукции вида 1 и 2 (производственное время и ограничение на спрос) и используемое на этапе 2.

Этап 2.

Пусть $f_2(v_2, w_2)$ представляет максимальную прибыль для этапа 2 (прибыль от выпуска продукции вида 2) при заданном состоянии (v_2, w_2). Тогда

$$f_2(v_2, w_2) = \max_{\substack{0 \le x_2 \le v_2 \\ 0 \le x_2 \le w_2}} \{5x_2\}.$$

Следовательно, $\max\{5x_2\}$ имеет место при $x_2 = \min\{v_2, w_2\}$. Имеем следующее решение для второго этапа.

Состояние	Оптимальное решение	
	$f_2(v_2, w_2)$	x_2
(v_2, w_2)	$5\min\{v_2, w_2\}$	$\min\{v_2, w_2\}$

Этап 1.

$$f_1(v_1, w_1) = \max_{0 \le 2x_1 \le v_1} \{2x_1 + f_2(v_1 - 2x_1, w_1)\} = \max_{0 \le 2x_1 \le v_1} \{2x_1 + 5\min(v_1 - 2x_1, w_1)\}.$$

Оптимизация на первом этапе требует решения минимаксной задачи, что в общем случае является достаточно сложным делом. Для рассматриваемой задачи имеем $v_1 = 430$ и $w_1 = 230$, что дает $0 \le 2x_1 \le 430$. Так как $\min(430 - 2x_1, 230)$ представляет собой нижнюю огибающую двух пересекающихся прямых (проверьте!), отсюда следует, что

$$\min(430 - 2x_1, 230) = \begin{cases} 230, & 0 \le x_1 \le 100, \\ 430 - 2x_1, & 100 \le x_1 \le 215 \end{cases}$$

и

$$f_1(430, 230) = \max_{x_1} \begin{cases} 2x_1 + 1150, & 0 \le x_1 \le 100, \\ -8x_1 + 2150, & 100 \le x_1 \le 215. \end{cases}$$

Графически можно легко проверить, что функция $f_1(430, 230)$ достигает максимального значения при $x_1 = 100$. Итак, получаем следующее.

Состояние	Оптимальное решение	
	$f_1(v_1, w_1)$	x_1
(430, 230)	1350	100

Для определения оптимального значения x_2 заметим, что

$$v_2 = v_1 - 2x_1 = 430 - 200 = 230,$$
$$w_2 = w_1 - 0 = 230.$$

Следовательно,

$$x_2 = \min\{v_2, w_2\} = 230.$$

Итак, оптимальное решение имеет вид:

$x_1 = 100$ единиц, $x_2 = 230$ единиц, $z = 1350$ долларов.

Упражнения 10.5,а

1. Решите следующие задачи линейного программирования методами динамического программирования.

   a) Максимизировать $z = 4x_1 + 14x_2$
   при ограничениях
   $$2x_1 + 7x_2 \leq 21,$$
   $$7x_1 + 2x_2 \leq 21,$$
   $$x_1, x_2 \geq 0.$$

   b) Максимизировать $z = 8x_1 + 7x_2$
   при ограничениях
   $$2x_1 + x_2 \leq 8,$$
   $$5x_1 + 2x_2 \leq 15,$$
   $$x_1, x_2 \geq 0 \text{ и целые.}$$

   c) Максимизировать $z = 7x_1^2 + 6x_1 + 5x_2^2$
   при ограничениях
   $$x_1 + 2x_2 \leq 10,$$
   $$x_1 - 3x_2 \leq 9,$$
   $$x_1, x_2 \geq 0.$$

2. Пусть в задаче из примера 10.4–1 о загрузке предметов n наименований ограничения самолета по весу и объему представлены величинами W и V соответственно. Величины w_i, v_i и r_i представляют соответственно вес, объем и прибыль, отнесенные к одному предмету наименования i. Необходимо записать рекуррентное уравнение обратной прогонки для алгоритма динамического программирования решения сформулированной задачи.

10.6. Заключение

В этой главе рассмотрены некоторые примеры детерминированных моделей динамического программирования. Принцип оптимальности является основой поэтапного решения задачи динамического программирования. Хотя этот принцип не содержит информации о способах решения подзадач на каждом этапе, его применение существенно облегчает решение многих сложных задач, которые нельзя решить другими методами.

Литература

Bertsekas D. *Dynamic Programming: Deterministic and Stochastic Models*, Prentice Hall, Upper Saddle River, N.J., 1987.

Denardo E. *Dynamic Programming Theory and Application*, Prentice Hall, Upper Saddle River, N.J., 1982.

Dreyfus S., Law A. *The Art and Theory of Dynamic Programming*, Academic Press, New York, 1977.

Sniedovich M. *Dynamic Programming*, Marcel Dekker, New York, 1991.

Литература, добавленная при переводе

Беллман Р. *Динамическое программирование.* — М.: ИЛ, 1960.

Беллман Р., Дрейфус С. *Прикладные задачи динамического программирования.* — М.: Наука, 1965.

Вентцель Е.С. *Исследование операций.* — М.: Сов. радио, 1972.

Романовский И.В. *Алгоритмы решения экстремальных задач.* — М.: Наука, 1977.

Комплексная задача

■ **10–1.** Компания проверяет состояние оборудования в конце каждого года и на основании этого принимает следующее решение: либо использовать его еще один год, либо заменить. Однако оборудование, которое находилось в эксплуатации три года, подлежит замене в обязательном порядке. Компания планирует разработать стратегию замены имеющегося оборудования на следующие 10 лет. Соответствующая информация содержится в приведенной ниже таблице. Считается, что в начале первого года все оборудование новое.

Год покупки	Стоимость покупки ($)	Стоимость содержания ($) для данного возраста (лет)			Стоимость продажи ($) для данного возраста (лет)		
		0	1	2	1	2	3
1	10 000	200	500	600	9 000	7 000	5 000
2	12 000	250	600	680	11 000	9 500	8 000
3	13 000	280	550	600	12 000	11 000	10 000
4	13 500	320	650	700	12 000	11 500	11 000
5	13 800	350	590	630	12 000	11 800	11 200
6	14 200	390	620	700	12 500	12 000	11 200
7	14 800	410	600	620	13 500	12 900	11 900
8	15 200	430	670	700	14 000	13 200	12 000
9	15 500	450	700	730	15 500	14 500	13 800
10	16 000	500	710	720	15 800	15 000	14 500

Детерминированные модели управления запасами

11.1. Введение

Как в бизнесе, так и в производстве обычно принято поддерживать разумный запас материальных ресурсов или комплектующих для обеспечения непрерывности производственного процесса. Традиционно запас рассматривается как неизбежные издержки, когда слишком низкий уровень запаса приводит к дорогостоящим остановкам производства, а слишком высокий — к "омертвлению" капитала. Задача управления запасами определяет уровень запаса, который уравновешивает два упомянутых крайних случая.

Важным фактором, определяющим формулировку и решение задачи управления запасами, является то, что объем спроса на хранимый запас (в единицу времени) может быть или *детерминированным* (достоверно известным), или *вероятностным* (описанным вероятностным распределением). В этой главе рассматриваются детерминированные модели управления запасами. Вероятностные модели (обычно более сложные) обсуждаются в главе 16.

11.2. Обобщенная модель управления запасами

Природа задачи управления запасами определяется неоднократным размещением и получением заказов заданных объемов продукции (в дальнейшем — хранимых запасов) в определенные моменты времени. С этой точки зрения **стратегия управления запасами** должна отвечать на следующие два вопроса.

1. Какое количество хранимого запаса следует заказать?
2. Когда заказывать?

Ответ на первый вопрос определяет *экономичный размер заказа* путем минимизации следующей функции затрат.

$$\begin{pmatrix}\text{Суммарные}\\ \text{затраты системы}\\ \text{управления}\\ \text{запасами}\end{pmatrix}=\begin{pmatrix}\text{Затраты на}\\ \text{приобретение}\end{pmatrix}+\begin{pmatrix}\text{Затраты на}\\ \text{оформление}\\ \text{заказа}\end{pmatrix}+\begin{pmatrix}\text{Затраты на}\\ \text{хранение}\\ \text{заказа}\end{pmatrix}+\begin{pmatrix}\text{Потери от}\\ \text{дефицита}\\ \text{запаса}\end{pmatrix}.$$

Все эти стоимости должны быть выражены как функции искомого объема заказа и интервала времени между заказами.

1. *Затраты на приобретение* определяются стоимостью единицы приобретаемой продукции (хранимого запаса). Эта стоимость может быть постоянной или со скидкой, которая зависит от объема заказа.
2. *Затраты на оформление заказа* представляют собой постоянные расходы, связанные с его размещением (для изготовления продукции) на других производствах. Эти затраты не зависят от объема заказа.
3. *Затраты на хранение запаса* представляют собой затраты на содержание запаса на складе. Этот вид затрат включает как процент на инвестированный капитал, так и стоимость хранения, содержания и ухода.
4. *Потери от дефицита запаса* — это расходы, обусловленные отсутствием запаса необходимой продукции. Они включают как потенциальные потери прибыли, так и более субъективную стоимость, связанную с потерей доверия клиентов.

Ответ на второй вопрос (когда заказывать?) зависит от типа системы управления запасами, с которой мы имеем дело. Если система предусматривает **периодический контроль** состояния запаса (например, каждую неделю или месяц), момент поступления нового заказа совпадает с началом периода. Если же в системе предусмотрен **непрерывный контроль** состояния запаса, новые заказы размещаются тогда, когда уровень запаса опускается до заранее определенного значения, называемого **точкой возобновления заказа**.

Модели управления запасами, рассматриваемые в этой главе, охватывают два типа детерминированных моделей: статические и динамические. В статических моделях рассматриваются ситуации, когда объем спроса на хранимую продукцию (запас) является постоянным во времени. В динамических моделях объем спроса является функцией времени.

11.3. Статические модели управления запасами

В этом разделе рассмотрены три разновидности модели управления запасами, позволяющие определить экономичные размеры заказа со статическим объемом спроса.

11.3.1. Классическая задача экономичного размера заказа

Простейшие модели управления запасами характеризуются постоянным во времени спросом, мгновенным пополнением запаса и отсутствием дефицита. Введем обозначения:

y — объем заказа (количество единиц продукции),
D — интенсивность спроса (измеряется в единицах продукции на единицу времени),
t_0 — продолжительность цикла заказа (измеряется во временных единицах).

Уровень запаса изменяется в соответствии с функцией, показанной на рис. 11.1, где использованы приведенные выше обозначения. Заказ объема y единиц размещается и пополняется мгновенно, когда уровень запаса равен нулю. Затем запас равномерно расходуется с постоянной интенсивностью спроса D. Продолжительность цикла заказа для этого примера равна

$$t_0 = \frac{y}{D} \text{ единиц времени.}$$

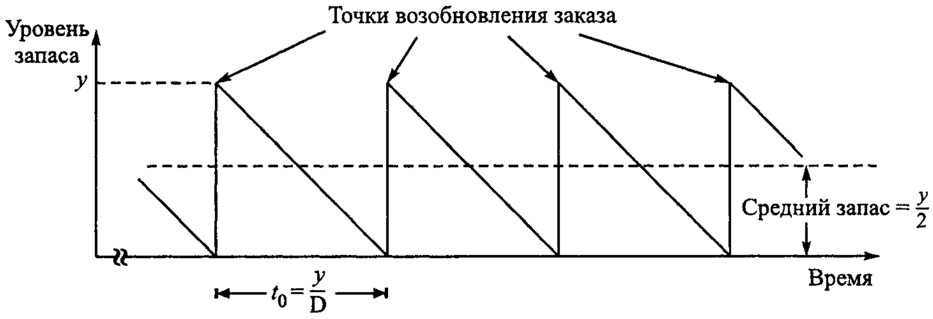


Рис. 11.1

Средний уровень запаса определяется соотношением

$$\text{средний уровень запаса} = \frac{y}{2} \text{ единиц.}$$

Для построения функции затрат требуется два стоимостных параметра.

K — затраты на оформление, связанные с размещением заказа,

h — затраты на хранение (затраты на единицу складируемой продукции в единицу времени).

Суммарные затраты *в единицу времени* (обозначается TCU[1]) можно представить как функцию от y в следующем виде.

$$\begin{aligned} TCU(y) &= \text{Затраты на оформление заказа в единицу времени} + \\ &+ \text{Затраты на хранение запаса в единицу времени} = \\ &= \frac{\text{Затраты на оформление} + \text{Затраты на хранение за цикл } t_0}{t_0} = \\ &= \frac{K + h\left(\frac{y}{2}\right)t_0}{t_0} = \frac{K}{\frac{y}{D}} + h\left(\frac{y}{2}\right). \end{aligned}$$

Оптимальное значение объема заказа y определяется путем минимизации по y функции $TCU(y)$. Предполагая, что y является непрерывной переменной, получаем необходимое условие минимума (в виде уравнения), из которого можно найти оптимальное значение y

$$\frac{dTCU(y)}{dy} = -\frac{KD}{y^2} + \frac{h}{2} = 0.$$

[1] TCU — сокращение от Total Cost per Unit time, т.е. суммарные затраты в единицу времени. — *Прим ред.*

Это условие является также и достаточным, так как функция $TCU(y)$ выпуклая. Решение данного уравнения определяет экономичный объем заказа y^*.

$$y^* = \sqrt{\frac{2KD}{h}}.$$

Оптимальная стратегия управления запасами для рассмотренной модели формулируется следующим образом:

Заказывать $y^* = \sqrt{\frac{2KD}{h}}$ единиц продукции через каждые $t_0^* = \frac{y^*}{D}$ единиц времени.

В действительности пополнение запаса не может произойти мгновенно в момент размещения заказа, как предполагалось ранее. Для большинства реальных ситуаций существует положительный **срок выполнения** заказа L (временное запаздывание) от момента его размещения до реальной поставки, как показано на рис. 11.2. В этом случае **точка возобновления заказа** имеет место, когда уровень запаса опускается до LD единиц.

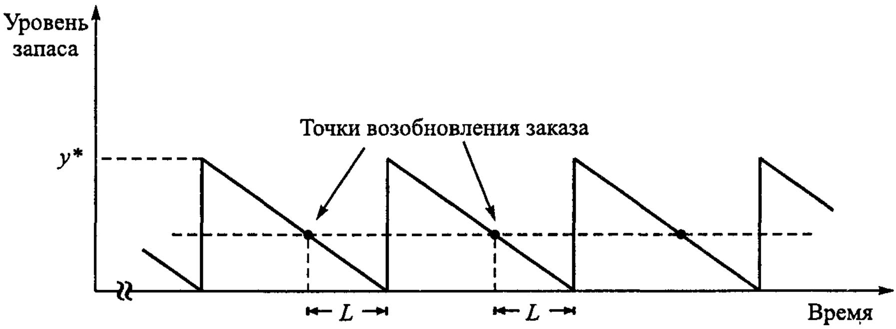


Рис. 11.2

На рис. 11.2 представлено изменение уровня запаса во времени в предположении, что срок выполнения заказа L меньше продолжительности цикла заказа t_0^*, что в общем случае выполняется не всегда. В противном случае определяется *эффективный* срок L_e выполнения заказа в виде

$$L_e = L - n\,t_0^*,$$

где n — наибольшее целое, не превышающее L/t_0^*. Такое решение оправдывается тем, что после n циклов (длиной t_0^* каждый) ситуация управления запасами становится такой же, как если бы интервал между размещением одного заказа и получением другого был равен L_e. Следовательно, точка возобновления заказа имеет место при уровне запаса L_eD единиц продукции, и стратегия управления запасами может быть переформулирована следующим образом.

Заказывать y* единиц продукции, как только уровень запаса опускается до LeD единиц.

Пример 11.3–1

Неоновые лампы в университетском городке заменяются с интенсивностью 100 штук в день. Подразделение материального обеспечения городка заказывает эти лампы с определенной периодичностью. Стоимость размещения заказа на покупку ламп составляет 100 долларов. Стоимость хранения лампы на складе оценивается в 0.02 доллара в день. Срок выполнения заказа от момента его размещения до реальной поставки равен 12 дней. Требуется определить оптимальную стратегию заказа неоновых ламп.

На основании приведенных данных имеем следующее.

$D = 100$ единиц в день,
$K = 100$ долларов за заказ,
$h = 0.02$ доллара за хранение одной лампы в день,
$L = 12$ дней.

Следовательно,

$$y^* = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2\times 100\times 100}{0.02}} = 1000 \text{ ламп.}$$

Соответствующая длина цикла составляет

$$t_0^* = \frac{y^*}{D} = \frac{1000}{100} = 10 \text{ дней.}$$

Так как срок выполнения заказа $L = 12$ дней превышает продолжительность цикла t_0^* (= 10 дней), необходимо вычислить L_e. Число целых циклов, заключенных в L, равно

$$n = (\text{наибольшее целое} \le L/t_0^*) = (\text{наибольшее целое} \le 12/10) = 1.$$

Следовательно,

$$L_e = L - n t_0^* = 12 - 1 \times 10 = 2 \text{ дня.}$$

Поэтому точка возобновления заказа имеет место при уровне запаса

$$L_e D = 2 \times 100 = 200 \text{ неоновых ламп.}$$

Оптимальная стратегия заказа неоновых ламп может быть сформулирована следующим образом.

Заказать 1000 ламп, как только уровень их запаса уменьшается до 200 единиц.

Дневные расходы, связанные с содержанием запаса в соответствии с оптимальной стратегией, равны

$$TCU(y) = \frac{K}{\frac{y}{D}} + h\left(\frac{y}{2}\right) = \frac{100}{\frac{1000}{100}} + 0.02\left(\frac{1000}{2}\right) = 20 \text{ долларов в день.}$$

Упражнения 11.3,а

1. В каждом из следующих случаев дефицит не допускается, а время выполнения заказа от момента его размещения до реальной поставки равно 30 дней. Требуется определить оптимальную стратегию управления запасами и соответствующие дневные затраты.

   a) $K = \$100$, $h = \$0.05$, $D = 30$ единиц в день.
   b) $K = \$50$, $h = \$0.05$, $D = 30$ единиц в день.
   c) $K = \$100$, $h = \$0.01$, $D = 40$ единиц в день.
   d) $K = \$100$, $h = \$0.04$, $D = 20$ единиц в день.

2. Ресторан заказывает мясной фарш в начале каждой недели для удовлетворения недельного спроса в 300 фунтов. Фиксированная стоимость размещения заказа равна 20 долларов. Стоимость замораживания и хранения одного фунта фарша обходится ресторану примерно в 0.03 доллара в день.

   a) Определите недельные затраты ресторана, связанные с существующей стратегией создания запаса.
   b) Определите оптимальную стратегию управления запасами в предположении, что время выполнения заказа от момента его размещения до реальной поставки равно нулю.
   c) Вычислите разность между текущими недельными затратами ресторана и теми, которые определяются оптимальной стратегией управления запасами.

3. Компания хранит на складе продукцию, которая потребляется с интенсивностью 50 единиц в день. За размещение заказа компания каждый раз платит 20 долларов. Стоимость хранения единицы продукции на складе обходится в \$0.35 в неделю.

   a) Определите оптимальную стратегию управления запасами, если предположить, что время выполнения заказа от момента его размещения до реальной поставки равно 1 неделе.
   b) Определите оптимальное количество заказов в течение года (считая, что год имеет 365 дней).

4. Отдел снабжения компании предложил две стратегии управления запасами.

   Стратегия 1. Объем заказа 150 единиц при точке возобновления заказа в 50 единиц и времени выполнения заказа 10 дней.

   Стратегия 2. Объем заказа 200 единиц при точке возобновления заказа в 75 единиц и времени выполнения заказа 15 дней.

   Затраты на оформление заказа равны 20 долларов, а стоимость хранения единицы продукции на складе обходится в \$0.02 в день.

   a) Какую из двух стратегий следует утвердить?
   b) Если бы вы отвечали за разработку стратегии управления запасами, какова была бы ваша рекомендация?

5. Магазин прессует и складывает в поддоны пустые картонные упаковочные коробки для их последующей переработки. За день штабелируется пять поддонов. Стоимость хранения одного поддона на заднем дворе магазина составляет 0.10

доллара в день. Компания, которая перевозит поддоны в перерабатывающий центр, устанавливает оплату в 100 долларов за аренду своего погрузочного оборудования плюс 3 доллара за перевозку каждого поддона. Изобразите графически изменение количества поддонов с течением времени и разработайте оптимальную стратегию доставки поддонов в перерабатывающий центр.

6. Отель использует внешнюю прачечную для стирки полотенец. За день в отеле накапливается 600 грязных полотенец. Прачечная забирает эти полотенца и заменяет их чистыми через постоянные промежутки времени. Стоимость однократной доставки полотенец в прачечную и обратно равна 81 доллар. Стирка одного полотенца обходится в \$0.60. Стоимость хранения в отеле грязного и чистого полотенец равна \$0.02 и \$0.01 соответственно. Как часто следует отелю пользоваться службой доставки полотенец? (*Подсказка.* В этой задаче имеется два типа складируемых предметов. Если количество грязных полотенец возрастает, то количество чистых уменьшается с равной интенсивностью.)

7. Дана задача управления запасами, в которой склад пополняется равномерно (вместо мгновенного пополнения) с интенсивностью a. Продукция потребляется с интенсивностью D. Так как потребление происходит наряду с периодом пополнения, необходимо, чтобы было $a > D$. Стоимость размещения заказа равна K, а стоимость хранения единицы продукции в единицу времени — h. Покажите, что если y — объем заказа и отсутствует дефицит, то

   a) максимальный объем запаса равен $y(1 - D/a)$,

   b) общие затраты в единицу времени при заданном y равны

$$TCU(y) = \frac{KD}{y} + \frac{h}{2}\left(1 - \frac{D}{a}\right)y,$$

   c) экономичный объем заказа равен

$$y^* = \sqrt{\frac{2KD}{h\left(1 - \frac{D}{a}\right)}}, \quad D < a,$$

   d) формулу экономичного объема заказа при мгновенном пополнении запаса можно получить из формулы в п. c).

8. Фирма может производить изделие или покупать его у подрядчика. Если фирма сама выпускает изделие, то каждый запуск его в производство обходится в 20 долларов. Мощность производства составляет 100 единиц в день. Если изделие закупается, затраты на размещение каждого заказа равны 15 долларов. Затраты на содержание изделия на складе, независимо от того, закупается оно или производится на фирме, равны \$0.02 в день. Потребление изделия фирмой оценивается в 260 000 единиц в год. Если предположить, что фирма работает без дефицита, определите, что выгоднее — закупать или производить изделия?

9. Предположим, что в упр. 7 допускается дефицит и удельные потери от него составляют p долл. в единицу времени. Если w — величина дефицита и y — объем заказа, покажите, что имеют место следующие соотношения.

$$TCU(y,w)=\frac{KD}{y}+\frac{h\left\{y\left(1-\frac{D}{a}\right)-w\right\}^2+pw^2}{2\left(1-\frac{D}{a}\right)y},$$

$$y^*=\sqrt{\frac{2KD(p+h)}{ph\left(1-\frac{D}{a}\right)}},$$

$$w^*=\sqrt{\frac{2KDh\left(1-\frac{D}{a}\right)}{p(p+h)}}.$$

11.3.2. Задача экономичного размера заказа с разрывами цен

Представленная в этом разделе модель управления запасами отличается от рассмотренной в разделе 11.3.1 только тем, что продукция может быть приобретена со скидкой, если объем заказа y превышает некоторый фиксированный уровень q; таким образом, стоимость единицы продукции c определяется как

$$c=\begin{cases}c_1, & \text{если } y\le q,\\ c_2, & \text{если } y>q,\end{cases}$$

где $c_1 > c_2$. Следовательно,

$$\text{затраты на приобретение продукции в единицу времени} = \begin{cases}\dfrac{c_1y}{t_0}=\dfrac{c_1y}{\left(\dfrac{y}{D}\right)}=Dc_1, & y\le q,\\[2ex] \dfrac{c_2y}{t_0}=\dfrac{c_1y}{\left(\dfrac{y}{D}\right)}=Dc_2, & y>q.\end{cases}$$

Используя обозначения из раздела 11.3.1, запишем общие затраты в единицу времени следующим образом.

$$TCU(y)=\begin{cases}TCU_1(y)=Dc_1+\dfrac{KD}{y}+\dfrac{h}{2}y, & y\le q,\\[2ex] TCU_2(y)=Dc_2+\dfrac{KD}{y}+\dfrac{h}{2}y, & y>q.\end{cases}$$

Графики функций TCU_1 и TCU_2 представлены на рис. 11.3. Так как значения этих функций отличаются только на постоянную величину, то точки их минимума совпадают и находятся в точке

$$y_m = \sqrt{\frac{2KD}{h}}.$$

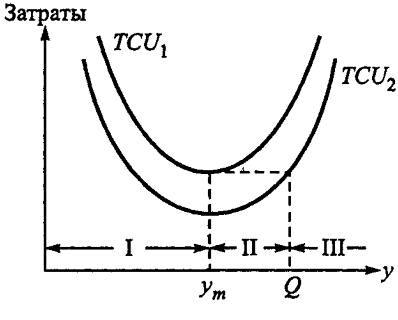


Рис. 11.3

График функции затрат $TCU(y)$, если идти от минимальных значений аргументов, совпадает с графиком функции $TCU_1(y)$ до точки $y = q$, в которой меняется цена продукции, а затем совпадает с графиком функции $TCU_2(y)$. Рис. 11.3 показывает, что определение оптимального объема заказа y^* зависит от того, где находится точка разрыва цены q по отношению к указанным на рисунке зонам I, II и III, которые определены как интервалы $[0, y_m)$, $[y_m, Q)$ и $[Q, \infty)$ соответственно. Величина Q ($> y_m$) определяется из уравнения

$$TCU_2(Q) = TCU_1(y_m).$$

Рис. 11.4 показывает, как определяется оптимальное значение y^*.

$$y^* = \begin{cases} y_m, & \text{если } q \text{ находится в зоне I или III,} \\ q, & \text{если } q \text{ находится в зоне II.} \end{cases}$$

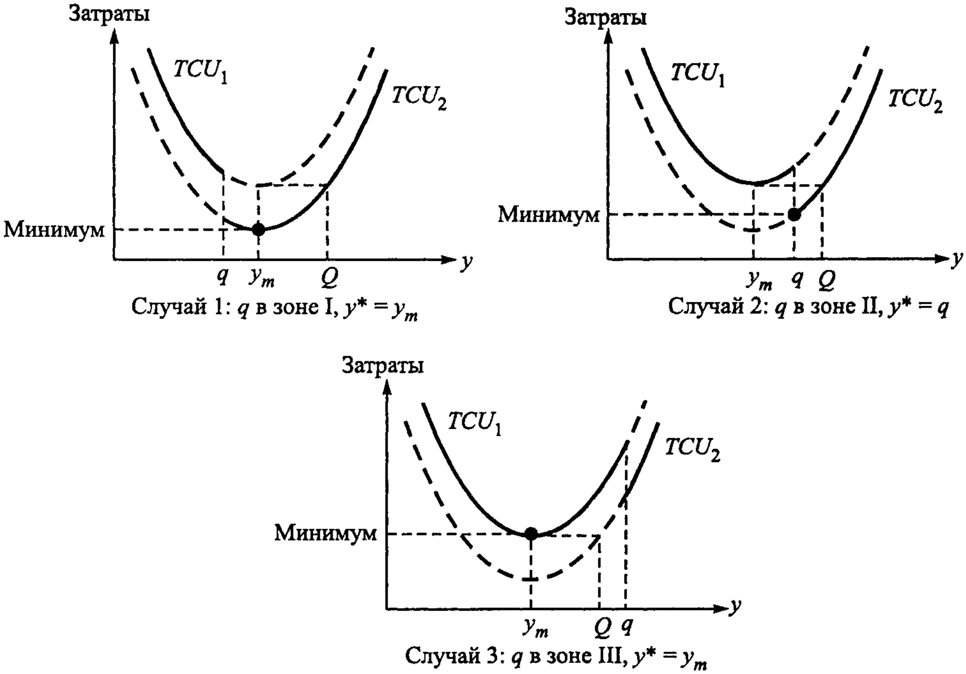
Алгоритм определения y^* можно сформулировать в следующем виде.

Шаг 1. Вычисляем $y_m = \sqrt{\frac{2KD}{h}}$. Если q попадает в зону I, полагаем $y^* = y_m$. В противном случае переходим к шагу 2.

Шаг 2. Находим Q из уравнения $TCU_2(Q) = TCU_1(y_m)$ и определяем зоны II и III. Если q находится в зоне II, полагаем $y^* = q$. Иначе q находится в зоне III, тогда $y^* = y_m$.

Пример 11.3–2

Автомобильная мастерская специализируется на быстрой замене масла в автомобилях. Мастерская покупает автомобильное масло в большом количестве по 3 доллара за галлон. Цена может быть снижена до 2.50 долларов за галлон при условии, что мастерская покупает более 1000 галлонов. За день в мастерской обслуживается около 150 автомобилей, и на каждый из них для замены требуется 1.25 галлона масла. Мастерская хранит на складе большие объемы масла, что обходится в 0.02 доллара в день за один галлон. Стоимость размещения заказа на большой объем масла равна 20 долларов. Срок выполнения заказа равен 2 дня. Требуется определить оптимальную стратегию управления запасами.



Случай 1: q в зоне I, $y^* = y_m$

Случай 2: q в зоне II, $y^* = q$

Случай 3: q в зоне III, $y^* = y_m$

Рис. 11.4

Дневное потребление масла равно

$$D = 125 \text{ автомобилей} \times 1.25 \text{ галлона} = 187.5 \text{ галлона в день.}$$

Также имеем

$h = \$0.02$ за галлон в день,

$K = \$20$ за заказ,

$L = 2$ дня,

$c_1 = \$3$ за галлон,

$c_2 = \$2.50$ за галлон,

$q = 1000$ галлонов.

Шаг 1. Вычисляем

$$y_m = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2\times 20\times 187.5}{0.02}} = 612.37 \text{ галлонов.}$$

Так как $q = 1000$ больше $y_m = 612.37$, переходим к шагу 2.

Шаг 2. Вычисляем Q из уравнения

$$TCU_2(Q) = TCU_1(y_m)$$

или

$$c_2 D + \frac{KD}{Q} + \frac{hQ}{2} = c_1 D + \frac{KD}{y_m} + \frac{hy_m}{2}.$$

Подставляя данные в это уравнение, получаем

$$2.5\times187.5+\frac{20\times187.5}{Q}+\frac{0.02Q}{2}=3\times187.5+\frac{20\times187.5}{612.37}+\frac{0.02\times612.37}{2}.$$

После упрощений имеем

$$0.01Q^2-106Q+3750=0.$$

Решением этого уравнения будет $Q = 10564.5$ ($> y_m$). Следовательно,

$$\text{Зона II} = [612.37,\ 10564.5),$$
$$\text{Зона III} = [10564.5,\ \infty).$$

Так как q ($= 1000$) находится в зоне II, оптимальный объем заказа равен $y^* = q = 1000$ галлонов.

При заданном сроке выполнения заказа в 2 дня точкой возобновления заказа является $2D = 2\times187.5 = 375$ галлонов. Следовательно, оптимальная стратегия управления запасами формулируется следующим образом.

Заказать 1000 галлонов масла, когда уровень запаса понижается до 375 галлонов.

Упражнения 11.3,b

1. Вернитесь к задаче из упр. 11.3,a(6). Стоимость стирки одного грязного полотенца равна 0.60 доллара, но она может быть снижена до 0.50 доллара, если отель поставляет в прачечную по меньшей мере 2500 единиц полотенец. Следует ли отелю воспользоваться скидкой?

2. Продукция используется с интенсивностью 30 единиц в день. Стоимость хранения единицы продукции равна 0.05 доллара в день, стоимость размещения заказа составляет 100 долларов. Предположим, что дефицит продукции не допускается, стоимость закупки равна 10 долларов за единицу продукции, если объем закупки не превышает 500 единиц, и 8 долларов в противном случае. Определите оптимальную стратегию управления запасами при условии, что срок выполнения заказа равен 21 день.

3. Комплектующие продаются по 25 долларов за единицу, но предлагается 10% скидка при покупке партии от 150 единиц и выше. Компания в день использует 20 единиц комплектующих. Стоимость размещения заказа равна 50 долларов, стоимость хранения единицы товара составляет 0.30 доллара в день. Следует ли компании воспользоваться скидкой?

4. В предыдущем упражнении определите пределы изменения скидки на цену комплектующих в процентах (предлагаемую за партию от 150 единиц и выше), при которых компания не получит никакой финансовой выгоды.

5. В модели управления запасами, рассмотренной в этом разделе, предположите, что стоимость хранения единицы товара в единицу времени равна h_1, если объем хранимого товара меньше q единиц, и h_2 в противном случае, $h_1 > h_2$. Покажите, как в этом случае можно определить экономичный размер партии хранимого товара.

11.3.3. Многопродуктовая статическая модель с ограниченной вместимостью склада

Эта модель рассматривает задачу управления запасами n различных товаров, которые хранятся на одном складе ограниченной вместимости. Характер изменения запаса каждого товара в отдельности определяется функцией, показанной на рис. 11.1; предполагаем, что дефицит отсутствует. Отличие от ранее рассмотренных моделей состоит в том, что товары конкурируют между собой за ограниченное складское пространство.

Определим для товара i, $i = 1, 2, ..., n$, следующие параметры.

D_i — интенсивность спроса,
K_i — стоимость размещения заказа,
h_i — стоимость хранения единицы товара в единицу времени,
y_i — объем заказа,
a_i — необходимое пространство для хранения единицы товара,
A — максимальное складское пространство для хранения товаров n видов.

При отсутствии дефицита математическая модель сформулированной задачи имеет следующий вид.

$$\text{Минимизировать } TCU(y_1, y_2, ..., y_n) = \sum_{i=1}^{n}\left(\frac{K_i D_i}{y_i} + \frac{h_i y_i}{2}\right)$$

при ограничениях

$$\sum_{i=1}^{n} a_i y_i \le A,$$

$$y_i > 0,\ i = 1, 2, ..., n.$$

Алгоритм решения этой задачи можно описать следующим образом.

Шаг 1. Вычисляются оптимальные объемы заказов *без учета ограничения* по вместимости склада:

$$y_i^* = \sqrt{\frac{2K_i D_i}{h_i}},\quad i = 1, 2, ..., n.$$

Шаг 2. Осуществляется проверка, удовлетворяют ли найденные значения y_i^* ограничению по вместимости склада. Если "Да", вычисления заканчиваются, при этом значения y_i^*, $i = 1, 2, ..., n$ являются оптимальными. В противном случае следует перейти к шагу 3.

Шаг 3. Ограничение по вместимости склада должно удовлетворяться в форме равенства. Используется метод множителей Лагранжа для определения оптимальных объемов заказа для задачи с ограничением.

На шаге 3 строится функция Лагранжа

$$L(\lambda, y_1, y_2, ..., y_n) = TCU(y_1, y_2, ..., y_n) - \lambda\left(\sum_{i=1}^{n} a_i y_i - A\right) =$$

$$= \sum_{i=1}^{n}\left(\frac{K_i D_i}{y_i} + \frac{h_i y_i}{2}\right) - \lambda\left(\sum_{i=1}^{n} a_i y_i - A\right),$$

где λ (< 0) — множитель Лагранжа[1].

Так как функция Лагранжа является выпуклой, оптимальные значения y_i и λ находятся из следующих уравнений, которые представляют собой необходимые условия экстремума функции Лагранжа.

$$\frac{\partial L}{\partial y_i} = -\frac{K_i D_i}{y_i^2} + \frac{h_i}{2} - \lambda a_i = 0,$$

$$\frac{\partial L}{\partial \lambda} = -\sum_{i=1}^{n} a_i y_i + A = 0.$$

Второе уравнение показывает, что ограничение по вместимости склада в оптимальной точке должно удовлетворяться в форме равенства.

Из первого уравнения следует, что

$$y_i^* = \sqrt{\frac{2K_i D_i}{h_i - 2\lambda^* a_i}}.$$

Полученная формула показывает, что y_i^* зависит от оптимального значения λ^* множителя Лагранжа. Кроме того, при $\lambda^* = 0$ значение y_i^* является решением задачи без ограничения.

Значение λ^* может быть найдено следующим образом. Так как по определению в поставленной выше задаче минимизации $\lambda^* < 0$, мы последовательно уменьшаем λ на достаточно малую величину и используем ее в данной формуле для вычисления соответствующего значения y_i^*. Искомое значение λ^* приводит к значениям y_i^*, $i = 1, 2, ..., n$, которые удовлетворяют ограничению по вместимости склада в форме равенства.

Пример 11.3–3

Рассмотрим задачу управления запасами, исходные данные для которой приведены в следующей таблице.

Товар i	K_i (\$)	D_i (единиц в день)	h_i (\$)	a_i (кв. футы)
1	10	2	0.3	1
2	5	4	0.1	1
3	15	4	0.2	1
Общая площадь склада = 25 футов^2				

[1] В разделе 20.2.1 детально рассмотрен метод множителей Лагранжа. Применение метода в данном случае является корректным, так как здесь функция $TCU(y_1, y_2, ..., y_n)$ выпуклая, задача имеет единственное линейное ограничение и, следовательно, выпуклое пространство решений. Метод может оказаться некорректным при других ограничениях или при наличии более одного ограничения (см. раздел 20.2.1).

Уменьшаем λ с шагом 0.1. Результаты вычислений приведены в следующей таблице.

λ	y_1	y_2	y_3	$\sum_{i=1}^{3} a_i y_i - A$
0	11.5	20.0	24.5	+31.0
−0.1	8.9	11.5	17.3	+12.7
−0.2	7.6	8.9	14.1	+5.6
−0.3	6.7	7.6	12.2	+1.5
−0.4	6.0	6.7	11.0	−1.3

Последний столбец показывает, что ограничение по вместимости склада в форме равенства выполняется в интервале $-0.3 > \lambda > -0.4$. Дальше можно использовать надлежащие методы численного анализа для нахождения соответствующего значения λ из указанного интервала. Используя такую процедуру, получаем $\lambda^* \approx -0.345$, что дает

$$y_1^* \approx 6.35 \text{ единицы}, \quad y_2^* \approx 7.11 \text{ единицы}, \quad y_3^* \approx 11.6 \text{ единицы}.$$

Как показывает предыдущий пример, вычисления всегда начинаются со значения $\lambda = 0$, что может привести к неэффективной вычислительной процедуре. Это положение может быть улучшено путем надлежащего выбора начального значения λ. Для этого можно использовать вычислительную процедуру, предложенную Доном Делом (Don Deal) из университета города Хьюстон, о которой идет речь в упр. 11.3,с(4).

Упражнения 11.3,с

1. Решите задачу из примера 11.3–3 в предположении, что сумма средних запасов всех предметов должна быть меньше 25 единиц.
2. Приведенные ниже данные относятся к задаче управления запасами для четырех видов продукции. Компания желает определить экономичный объем заказа для каждого из четырех видов продукции таким образом, чтобы суммарное количество заказов в год (365 дней) было не более 150.

Продукция i	K_i ($)	D_i (единиц в день)	h_i ($)
1	100	10	0.1
2	50	20	0.2
3	90	5	0.2
4	20	10	0.1

3. Решите предыдущее упражнение в предположении, что единственным ограничением является денежная сумма в 10 000 долларов, которая может быть инвестирована на приобретение запасов продукции. Стоимость закупки единицы продукции вида 1, 2, 3 и 4 равна соответственно 10, 5, 10 и 10 долларов.

4. На основе уравнения в частных производных задачи управления запасами этой главы покажите, что в качестве начального значения λ в процедуре поиска оптимального значения этого параметра можно взять величину

$$\lambda^* \approx \frac{\bar{h}}{2\bar{a}} - \frac{n^2 \bar{a} \overline{KD}}{A^2},$$

где

$$\bar{h} = \frac{\sum_{i=1}^{n} h_i}{n}, \quad \bar{a} = \frac{\sum_{i=1}^{n} a_i}{n}, \quad \overline{KD} = \frac{\sum_{i=1}^{n} K_i D_i}{n}.$$

Точное значение λ находится выше или ниже λ^*. Примените это начальное значение в задаче из примера 11.3–3.

11.4. Динамические задачи экономичного размера заказа

Рассматриваемые здесь модели отличаются от представленных в разделе 11.3. Во-первых, уровень запаса контролируется периодически на протяжении конечного числа одинаковых периодов. Во-вторых, объем спроса на протяжении периода хотя и является детерминированным, но в то же время он динамический, поскольку может периодически меняться.

Ситуация, в которой имеет место переменный детерминированный спрос, называется **планированием потребностей ресурсов**. Подход к решению такой задачи рассмотрим на примере. Предположим, что на протяжении следующего года квартальный спрос на модели *M*1 и *M*2 некоторой продукции равен 100 и 150 единиц соответственно. Поставки квартальных партий реализуются в конце каждого квартала. Срок выполнения заказа на модели *M*1 и *M*2 равен 2 месяца и 1 месяц соответственно. Для изготовления каждой единицы модели *M*1 и *M*2 используется 2 единицы комплектующих деталей *S*. Срок изготовления комплектующих равен одному месяцу.

На рис. 11.5 схематически представлено календарное планирование производства моделей *M*1 и *M*2. Построение плана начинается с отображения в виде сплошных стрелок квартального спроса на две модели, который имеет место в конце 3-го, 6-го, 9-го и 12-го месяцев. Затем при известных квартальных сроках пунктирные стрелки указывают начало производства каждой партии продукции *M*1 и *M*2 в 1-й и 2-й месяцы.

Чтобы вовремя начать производство партий двух рассматриваемых моделей, поставка комплектующих *S* должна совпадать с началом производства *M*1 и *M*2, т.е. с пунктирными стрелками в планах их производства. Эта информация представлена сплошными стрелками на *S*-схеме, где учитывается, что спрос на комплектующие *S* равен 2 единицам на каждую единицу продукции *M*1 и *M*2. Если учесть, что срок изготовления комплектующих равен одному месяцу, пунктирные стрелки на *S*-схеме определяют план производства комплектующих. Исходя из указанных двух планов, можно определить соответствующий суммарный спрос на *S*, как это показано в нижней части рис. 11.5. Результирующий *переменный* (но известный) спрос на комплектующие *S* представляет собой типичную ситуацию, когда применяются динамические модели экономичного размера заказа. При указанном переменном спросе на комплектующие S задача, по существу, сводится к определению объемов производства в начале каждого месяца для уменьшения затрат, связанных с производством и хранением продукции.

В этом разделе представлены две модели. В первой не учитывается стоимость размещения заказа, а вторая модель учитывает такие затраты. Эта "маленькая" деталь порождает соответствующие отличия в сложности моделей.

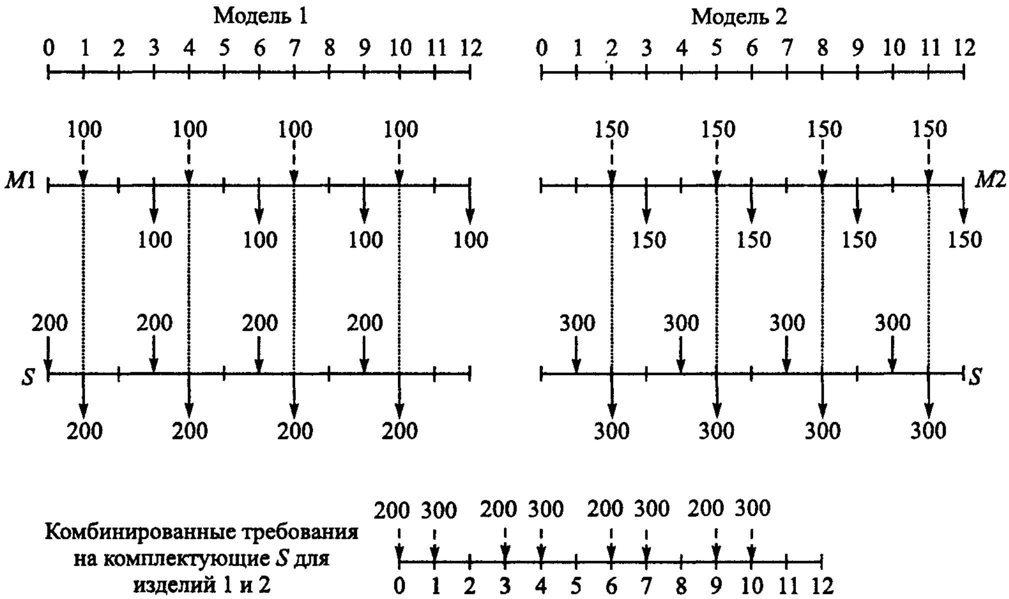


Рис. 11.5

Упражнение 11.4,а

1. Определите суммарные потребности в комплектующих S в соответствии с рис. 11.5 в каждом из следующих случаев.
   a) Поставка продукции $M1$ осуществляется каждый квартал.
   b) Поставка продукции $M1$ осуществляется раз в три квартала.

11.4.1. Модель при отсутствии затрат на оформление заказа

В этой модели рассматривается задача календарного планирования производства, рассчитанная на n равных периодов. Возможные объемы производства в каждый из периодов ограничены, однако они могут включать несколько уровней (например, два возможных объема производства могут определяться обычным режимом работы и сверхурочными работами соответственно). На протяжении текущего периода могут производиться изделия для последующих периодов, но в этом случае должны учитываться затраты на их хранение.

Основные предположения модели состоят в следующем.

1. Отсутствие затрат на оформление заказа в любой период планирования.
2. Отсутствие (недопустимость) дефицита.
3. Стоимость производства единицы продукции в любой период либо является постоянной, либо имеет возрастающие предельные затраты (т.е. соответствующая функция затрат является выпуклой).
4. Стоимость хранения единицы продукции в каждый период является постоянной величиной.

Предположение об отсутствии дефицита означает, что спрос на продукцию на протяжении текущего периода не может быть удовлетворен за счет ее производства в последующие периоды. Это предположение по крайней мере требует, чтобы суммарные возможности производства за периоды 1, 2, ..., i были равны суммарному спросу на продукцию за это же время.

На рис. 11.6 показано, когда производственные затраты на единицу продукции возрастают с увеличением уровня производства. Например, при двух возможных объемах производства, которые определяются обычным режимом работы и сверхурочными работами, стоимость производства единицы продукции, производимой в сверхурочное время, выше, чем при обычном режиме работы.

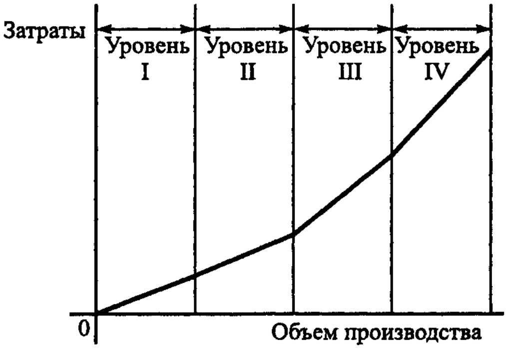


Рис. 11.6

Рассматриваемую задачу n-этапного планирования можно сформулировать в виде транспортной задачи (см. главу 5) с kn пунктами производства и n потребителями, где k — количество возможных уровней производства на протяжении периода (например, если на протяжении каждого периода используется регулярный и сверхурочный режимы работы, то $k = 2$). Производственные возможности каждого из kn пунктов производства определяют объемы поставок. Объемы потребления определяются объемом спроса для каждого периода. Себестоимость "перевозки" от пункта производства до пункта назначения определяется суммой затрат используемого производственного процесса и стоимости хранения единицы продукции. Оптимальное решение такой транспортной задачи определит объемы производства продукции для каждого производственного уровня, которые минимизируют суммарные затраты на производство и хранение.

Эту задачу можно решить без использования метода решения транспортных задач, представленного в главе 5. Обоснованность нового метода решения, показанного далее, следует из упомянутых предположений об отсутствии дефицита и выпуклости функции затрат на производство.

Пример 11.4–1

Компания производит специальные вытяжки, которые используются в домашних каминах в период с декабря по март. В начале отопительного сезона спрос на эту продукцию низкий, в середине сезона он достигает своего пика и уменьшается к концу сезона. Учитывая популярность продукции, компания может использовать сверхурочные работы для удовлетворения спроса на свою продукцию. Следующая таблица содержит данные о производственных мощностях компании и объемах спроса на протяжении четырех месяцев.

Месяц	Возможности производства: Обычный режим работы (единицы)	Возможности производства: Сверхурочные (единицы)	Спрос (единицы)
1	90	50	100
2	100	60	190
3	120	80	210
4	110	70	160

Стоимость производства единицы продукции равна 6 долларов в условиях обычного режима работы и 9 долларов при сверхурочных работах. Стоимость хранения единицы продукции на протяжении месяца равна 0.10 доллара.

Чтобы гарантировать допустимое решение при отсутствии дефицита, требуется, чтобы суммарное предложение продукции (возможности производства) к началу каждого месяца по меньшей мере равнялось суммарному спросу. Об этом свидетельствует следующая таблица.

Месяц	Суммарное предложение	Суммарный спрос
1	90 + 50 = 140	100
2	140 + 100 + 60 = 300	100 + 190 = 290
3	300 + 120 + 80 = 500	290 + 210 = 500
4	500 + 110 + 70 = 680	500 + 160 = 660

В табл. 11.1 содержатся данные, относящиеся к рассматриваемой задаче, и ее решение. Здесь R_i и O_i соответствуют уровням производства в обычном и сверхурочном режиме работы на протяжении периода i, $i = 1, 2, 3, 4$. Так как суммарное предложение в четвертом периоде превышает суммарный спрос, то введен искусственный пункт потребления (избыток), чтобы сбалансировать модель (это показано в табл. 11.1). Все "транспортные" маршруты из предыдущего в текущий период заблокированы, так как дефицит отсутствует.

Себестоимости "перевозок" продукции вычисляются в виде суммы затрат на производство и хранение. Например, соответствующая себестоимость от R_1 до первого периода равна лишь стоимости изготовления в 6 долларов, себестоимость от O_1 до четвертого периода — стоимости изготовления плюс стоимость хранения от первого периода до четвертого, т.е. $9 + (0.1 + 0.1 + 0.1) = 9.30$ доллара. Наконец, себестоимость перевозки до искусственного пункта потребления (*избыток*) равна нулю.

Таблица 11.1

	1	2	3	4	Избыток	
R_1	[6] 90	[6.1]	[6.2]	[6.3]	[0]	90
O_1	[9] 10	[9.1] 30	[9.2] 10	[9.3]	[0]	50→40→10
R_2		[6] 100	[6.1]	[6.2]	[0]	100
O_2		[9] 60	[9.1]	[9.2]	[0]	60
R_3			[6] 120	[6.1]	[0]	120
O_3			[9] 80	[9.1]	[0]	80
R_4				[6] 110	[0]	110
O_4				[9] 50	[0] 20	70→20
	100 ↓ 10	190 ↓ 90 ↓ 30	210 ↓ 90 ↓ 10	160 ↓ 50	20	

Оптимальное решение получается в один проход, начиная с первого столбца в направлении к столбцу "Избыток". Для каждого перспективного столбца спрос удовлетворяется с использованием самого дешевого маршрута[1].

Начиная с первого столбца маршрут $(R_1, 1)$ имеет самую дешевую себестоимость перевозки, и мы назначаем перевозку максимально возможного объема, а именно $\min(90, 100) = 90$ единиц, что оставляет 10 единиц неудовлетворенного спроса в первом столбце. Далее мы переходим к следующему по себестоимости маршруту $(O_1, 1)$ первого столбца и определяем перевозку $\min(50, 10) = 10$ единиц, что теперь полностью удовлетворяет спрос для первого периода.

После удовлетворения спроса для первого периода мы переходим ко второму столбцу. Определение перевозок в этом столбце происходит следующим образом: 100 единиц по маршруту $(R_2, 2)$, 60 единиц по маршруту $(O_2, 2)$ и 30 единиц по маршруту $(O_1, 2)$. Этим маршрутам соответствуют себестоимости "перевозок" в 6,

[1] Доказательство оптимальности этой процедуры приведено в работе Johnson S.M. "Sequential Production Planning over Time at Minimum Cost", *Management Science*, Vol.3, 1957, pp. 435–437.

9 и 9.10 долларов. При этом маршрут $(R_1, 2)$, транспортные расходы на единицу продукции для которого равны 6.10 долларов, не рассматривается, так как весь запас R_1 был израсходован для первого периода.

Продолжая аналогичным образом, мы удовлетворяем спрос для третьего, а затем и четвертого столбцов. Оптимальное решение, выделенное жирным шрифтом в табл. 11.1, интерпретируется следующим образом.

Период 1 (обычный режим работы). Изготовить 90 единиц продукции для первого периода.

Период 1 (сверхурочный режим работы). Изготовить 40 единиц продукции: 10 для периода 1, 30 для периода 2 и 10 для периода 3.

Период 2 (обычный режим работы). Изготовить 100 единиц продукции для второго периода.

Период 2 (сверхурочный режим работы). Изготовить 60 единиц продукции для периода 2.

Период 3 (обычный режим работы). Изготовить 120 единиц продукции для третьего периода.

Период 3 (сверхурочный режим работы). Изготовить 80 единиц продукции для периода 3.

Период 4 (обычный режим работы). Изготовить 110 единиц продукции для четвертого периода.

Период 4 (сверхурочный режим работы). Изготовить 50 единиц продукции для периода 4; осталась неиспользованной производственная мощность на 20 единиц продукции.

Соответствующие суммарные затраты при этом равны

$$90 \times 6 + 10 \times 9 + 30 \times 9.10 + 100 \times 6 + 60 \times 9 + 10 \times 9.20 +$$
$$120 \times 6 + 80 \times 9 + 110 \times 6 + 50 \times 9 = 4685 \text{ долларов.}$$

Упражнения 11.4,b

1. Решите задачу из примера 11.4–1 в предположении, что стоимости производства и хранения единицы продукции имеют значения, приведенные в следующей таблице.

Период i	Стоимость единицы продукции при обычном режиме работы ($)	Стоимость единицы продукции при сверхурочном режиме ($)	Стоимость хранения единицы продукции до периода $i+1$
1	5.00	7.50	0.10
2	3.00	4.50	0.15
3	4.00	6.00	0.12
4	1.00	1.50	0.20

2. Изделие производится для удовлетворения заданного спроса на четырех временных этапах в соответствии со следующими данными.

Диапазон объема производства (единицы)	Удельные производственные затраты на этапах ($)			
	1	2	3	4
1–3	1	2	2	3
4–11	1	4	5	4
12–15	2	4	7	5
16–25	5	6	10	7
Затраты на хранение одного изделия до следующего этапа ($)	0.30	0.35	0.20	0.25
Суммарный спрос (единицы)	11	4	17	29

a) Найдите оптимальное решение, определяющее количество изделий, которые необходимо изготовить на каждом из четырех этапов.

b) Предположим, что на этапе 4 требуется 10 дополнительных изделий. На каких этапах следует их изготовить?

3. В течение последующих пяти этапов спрос на некоторое изделие можно удовлетворить при обычном режиме работы, сверхурочных работах и субподряде. Субподряды можно использовать лишь при нехватке мощностей сверхурочных работ. Данные о производственных мощностях и объемах спроса приведены в следующей таблице.

	Производственные мощности (единицы)			
Этап	Обычный режим работы	Сверхурочные работы	Субподряд	Спрос
1	100	50	30	153
2	40	60	80	200
3	90	80	70	150
4	60	50	20	200
5	70	50	100	203

Предполагается, что затраты на производство единицы продукции на всех этапах одинаковы и составляют 4, 6 и 7 долларов при обычном режиме работы, сверхурочных работах и субподряде соответственно. Затраты на хранение единицы продукции на каждом этапе равны 0.50 доллара. Требуется найти оптимальное решение.

11.4.2. Модель с затратами на оформление заказа

В рассматриваемой модели предполагается, что дефицит не допускается и затраты на оформление заказа учитываются всякий раз, когда начинается производство новой партии продукции. Здесь будут рассмотрены два метода решения этой задачи: точный метод динамического программирования и эвристический.

Данная задача управления запасами схематически представлена на рис. 11.7. На этом рисунке использованы обозначения следующих величин, определенных для каждого этапа i, $i = 1, 2, ..., n$.

z_i — количество заказанной продукции (объем заказа),
D_i — потребность в продукции (спрос),
x_i — объем запаса на начало этапа i.

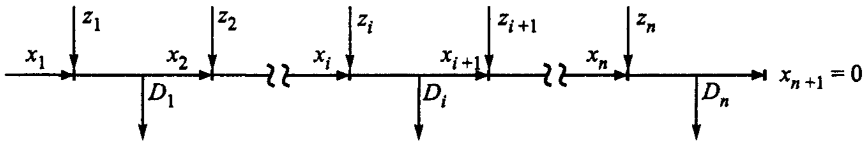


Рис. 11.7

Стоимостные элементы в рассматриваемой задаче определяются так:

K_i — затраты на оформление заказа,
h_i — затраты на хранение единицы продукции, переходящей из этапа i в этап $i + 1$.
Соответствующая функция производственных затрат для этапа i задается формулой

$$C_i(z_i) = \begin{cases} 0, & z_i = 0, \\ K_i + c_i(z_i), & z_i > 0, \end{cases}$$

где $c_i(z_i)$ — функция предельных производственных затрат при заданном значении z_i.

Алгоритм динамического программирования с общей функцией стоимости. Так как дефицит не допускается, задача управления запасами сводится к нахождению значений z_i, минимизирующих суммарные затраты, связанные с размещением заказов, закупкой и хранением продукции на протяжении n этапов. Затраты на хранение на i-м этапе для простоты предполагаются пропорциональными величине

$$x_{i+1} = x_i + z_i - D_i,$$

которая представляет собой объем запаса, переходящего из этапа i в этап $i + 1$.

Для построения модели динамического программирования можно воспользоваться рекуррентными соотношениями процедуры прямой или обратной прогонки (см. главу 10). Мы воспользуемся рекуррентными соотношениями процедуры прямой прогонки, так как они более эффективны при анализе важного частного случая рассматриваемой задачи, связанного с невозрастающими предельными затратами.

Для рекуррентного уравнения процедуры прямой прогонки *состояние* на *этапе* (периоде) i определяется как объем запаса x_{i+1} на конец этапа, где, как следует из рис. 11.7,

$$0 \le x_{i+1} \le D_{i+1} + ... + D_n.$$

Это неравенство означает, что в предельном случае запас x_{i+1} может удовлетворить спрос на всех последующих этапах.

Пусть $f_i(x_{i+1})$ — минимальные общие затраты на этапах $1, 2, ..., i$ при заданной величине запаса x_{i+1} на конец этапа i. Тогда рекуррентное уравнение алгоритма прямой прогонки будет записано следующим образом.

$$f_1(x_2)=\min_{0\le z_1\le D_1+x_2}\left\{C_1(z_1)+h_1x_2\right\},$$

$$f_i(x_{i+1})=\min_{0\le z_i\le D_i+x_{i+1}}\left\{C_i(z_i)+h_ix_{i+1}+f_{i-1}(x_{i+1}+D_i-z_i)\right\},\quad i=2,3,\ldots,n.$$

Пример 11.4–2

Требуется найти оптимальную стратегию в трехэтапной системе управления запасами, которая формулируется ниже. Начальный запас равен $x_1 = 1$ единице продукции. Предполагается, что предельные затраты на приобретение продукции составляют 10 долларов за каждую единицу для первых трех единиц и 20 долларов — за каждую дополнительную единицу.

Период i	Спрос, D_i (единицы)	Затраты на оформление заказа, K_i ($)	Затраты на хранение, h_i ($)
1	3	3	1
2	2	7	3
3	4	6	2

Функция производственных затрат для периода i равна $C_i(z_i) = K_i + c_i(z_i)$ для $z_i > 0$, где

$$c_i(z_i)=\begin{cases}10z_i, & 0\le z_i\le 3,\\ 30+20(z_i-3), & z_i\ge 4.\end{cases}$$

Этап 1. $D_1 = 3,\ 0 \le x_2 \le 2 + 4 = 6.$

		$C_1(z_1)+h_1x_2$							Оптимальное решение	
		$z_1 = 2$	3	4	5	6	7	8		
x_2	h_1x_2	$C_1(z_1) = 23$	33	53	73	93	113	133	$f_1(x_2)$	z_1^*
0	0	23							23	2
1	1		34						34	3
2	2			55					55	4
3	3				76				76	5
4	4					97			97	6
5	5						118		118	7
6	6							139	139	8

Так как $x_1 = 1$, минимальное значение z_1 равно $D_1 - x_1 = 3 - 1 = 2$.

Этап 2. $D_2 = 2, 0 \le x_3 \le 4$.

		$C_2(z_2) + h_2x_3 + f_1(x_3 + D_2 - z_2)$							Оптимальное решение	
		$z_2 = 0$	1	2	3	4	5	6		
x_3	h_2x_3	$C_2(z_2) = 0$	17	27	37	57	77	97	$f_2(x_3)$	z_2^*
0	0	0 + 55 = 55	17 + 34 = 51	27 + 23 = 50					50	2
1	3	3 + 76 = 79	20 + 55 = 75	30 + 34 = 64	40 + 23 = 63				63	3
2	6	6 + 97 = 103	23 + 76 = 99	33 + 55 = 88	43 + 34 = 77	63 + 23 = 86			77	3
3	9	9 + 118 = 127	26 + 97 = 123	36 + 76 = 112	46 + 55 = 101	66 + 34 = 100	86 + 23 = 109		100	4
4	12	12 + 139 = 151	29 + 118 = 147	39 + 97 = 136	49 + 76 = 125	69 + 55 = 124	89 + 34 = 123	109 + 23 = 132	123	5

Этап 3. $D_3 = 4, x_4 = 0$.

		$C_3(z_3) + h_3x_4 + f_2(x_4 + D_3 - z_3)$					Оптимальное решение	
		$z_3 = 0$	1	2	3	4		
x_4	h_3x_4	$C_3(z_3) = 0$	16	26	36	56	$f_3(x_4)$	z_3^*
0	0	0 + 123 = 123	16 + 100 = 116	26 + 77 = 103	36 + 63 = 99	56 + 50 = 106	99	3

Оптимальное решение определяется следующими значениями искомых переменных: $z_1^* = 2$, $z_2^* = 3$ и $z_3^* = 3$. При этом общие затраты составляют 99 долларов.

Упражнения 11.4,с

1. Вернитесь к задаче из примера 11.4–2.

   a) Имеет ли смысл рассматривать значение $x_4 > 0$?

   b) Для каждого из следующих случаев определите допустимые интервалы значений для z_1, z_2, z_3, x_1, x_2 и x_3. (Полезно представить каждый случай графически, как сделано на рис. 11.7.)

      i) $x_1 = 4$, остальные исходные данные не меняются,

      ii) $x_1 = 0$, $D_1 = 5$, $D_2 = 4$ и $D_3 = 5$.

2. Найдите оптимальное решение следующей четырехэтапной задачи управления запасами.

Период i	Спрос, Di (единицы)	Затраты на оформление заказа, K_i ($)	Затраты на хранение, h_i ($)
1	5	5	1
2	2	7	1
3	3	9	1
4	3	7	1

Затраты на приобретение первых шести единиц продукции составляют 1 доллар за каждую единицу и 2 доллара за каждую дополнительную единицу.

3. Пусть стоимость хранения запаса определяется *средним* его объемом на протяжении периода. Получите соответствующее рекуррентное уравнение для алгоритма прямой прогонки.

4. Получите рекуррентное уравнение для алгоритма обратной прогонки и с его помощью решите задачу из примера 11.4–2.

5. Получите рекуррентное уравнение для алгоритма обратной прогонки в предположении, что стоимость хранения запаса определяется *средним* его объемом на протяжении периода.

Алгоритм динамического программирования для задачи с постоянными или невозрастающими предельными затратами. Рассмотренную выше модель динамического программирования можно использовать при любых функциях затрат. Важным частным случаем этой модели является такая модель, когда на этапе i как стоимость закупки единицы продукции, так и затраты на ее хранение есть *невозрастающими* (вогнутыми) функциями объема закупаемой и хранимой продукции соответственно. Такая же ситуация возникает, когда функция стоимости, отнесенная к единице продукции, является постоянной или когда имеет место оптовая скидка.

При указанных выше условиях можно доказать следующее.[1]

1. При заданном начальном нулевом уровне запаса ($x_1 = 0$) для любого этапа i оптимальной стратегией является удовлетворение спроса за счет либо новой закупленной продукции, либо запаса, но не с обоих источников, т.е. $z_i x_i = 0$. (При положительном начальном уровне запаса ($x_1 > 0$) этот объем может быть списан из спроса последующих этапов, пока он не исчерпается.)

2. Оптимальный объем заказа z_i на любом этапе i должен либо равняться нулю, либо в точности соответствовать спросу одного или более последующих этапов.

Использование указанных двух свойств с рекуррентным уравнением для алгоритма прямой прогонки динамического программирования позволяет упростить схему вычислений.

---

[1] Подробное доказательство изложено в работе Wagner H. and Whitin T. "Dynamic Version of the Economic Lot Size Model", *Management Science*, Vol.5, 1958, pp. 89–96. Доказательство получено при ограничивающем предположении, что затраты на единицу продукции постоянны и идентичны на всех этапах. Этот результат в дальнейшем был обобщен А. Вейноттом (A. Veinott) из Стэнфордского университета для вогнутых функций затрат, имеющих место на каждом этапе.

Пример 11.4–3

Рассмотрим четырехэтапную модель управления запасами при следующих исходных данных.

Период i	Спрос, D_i (единицы)	Затраты на оформление заказа, K_i ($)
1	76	98
2	26	114
3	90	185
4	67	70

Начальный уровень запаса равен $x_1 = 15$ единиц. Затраты на закупку единицы продукции и ее хранение в течение одного периода для всех этапов одинаковы и составляют 2 доллара и 1 доллар соответственно. (Затраты на закупку и хранение единицы продукции приняты одинаковыми для всех этапов исключительно в целях упрощения.)

Решение определяется обычным алгоритмом прямой прогонки, за исключением того, что величины x_{i+1} и z_i допускают "общие" платежи, как это следует из свойств функции затрат. Так как начальный запас $x_1 = 15$, спрос на первом этапе уменьшается на эту величину и составляет 76 – 15 = 61 единицу.

Этап 1. $D_1 = 61$.

		$C_1(z_1) + h_1x_2$				Оптимальное решение	
		$z_1 = 61$	87	177	244		
x_2	h_1x_2	$C_1(z_1) = 220$	272	452	586	$f_1(x_2)$	z_1^*
0	0	220	—	—	—	220	61
26	26	—	298	—	—	298	87
116	116	—	—	568	—	568	177
183	183	—	—	—	769	769	244
Заказ на этапе 1 для этапов		1	1, 2	1, 2, 3	1, 2, 3, 4		

Этап 2. $D_2 = 26$.

		$C_2(z_2) + h_2x_3 + f_1(x_3 + D_2 - z_2)$				Оптимальное решение	
		$z_2 = 0$	26	116	183		
x_3	h_2x_3	$C_2(z_2) = 0$	166	346	480	$f_2(x_3)$	z_2^*
0	0	0 + 298 = 298	166 + 220 = 386	—	—	298	0
90	90	90 + 568 = 658	—	436 + 220 = 656	—	656	116
157	157	157 + 769 = 926	—	—	637 + 220 = 857	857	183
Заказ на этапе 2 для этапов		—	2	2, 3	2, 3, 4		

Этап 3. $D_3 = 90$.

		$C_3(z_3) + h_3x_4 + f_2(x_4 + D_3 - z_3)$			Оптимальное решение	
		$z_3 = 0$	90	157		
x_4	h_3x_4	$C_3(z_3) = 0$	365	499	$f_3(x_4)$	z_3^*
0	0	0 + 656 = 656	365 + 298 = 663	—	656	0
67	67	67 + 857 = 924	—	566 + 298 = 864	864	157
Заказ на этапе 3 для этапов		—	3	3, 4		

Этап 4. $D_4 = 67$

		$C_4(z_4) + h_4x_5 + f_3(x_5 + D_4 - z_4)$		Оптимальное решение	
		$z_4 = 0$	67		
x_5	h_4x_5	$C_4(z_4) = 0$	204	$f_4(x_5)$	z_4^*
0	0	0 + 864 = 864	204 + 656 = 860	860	67
Заказ на этапе 4 для этапов		—	4		

Таким образом, оптимальная стратегия задается следующими значениями: $z_1^* = 61$, $z_2^* = 116$, $z_3^* = 0$ и $z_4^* = 67$ при суммарных затратах 860 долларов.

Частным случаем описанной модели с невозрастающей (вогнутой) функцией затрат является модель, где производственные затраты на этапе i определяются линейной функцией

$$C_i(z_i) = K_i + c_i z_i,\ z_i > 0,\ i = 1, 2, ..., n,$$

при $c_1 \ge c_2 \ge ... \ge c_n$. При этих условиях алгоритм прямой прогонки можно модифицировать таким образом, чтобы обеспечивать дальнейшее сокращение объема вычислений.

Модифицированный алгоритм связывает этап i с соответствующим периодом планирования, когда заказы на всех предыдущих этапах делаются для последующих, включая текущий этап i. Математически это выражается следующим образом:

$$f_i = \min\begin{cases} C_1 + h_1\left(D_2 + ... + D_i\right) + h_{i-1}D_i & \left(\text{заказ на этапе } 1\right), \\ C_2 + h_2\left(D_3 + ... + D_i\right) + h_{i-1}D_i + f_1 & \left(\text{заказ на этапе } 2\right), \\ \vdots & \\ C_{i-1} + h_{i-1}D_i + f_{i-2} & \left(\text{заказ на этапе } i-1\right), \\ C_i + f_{i-1} & \left(\text{заказ на этапе } i\right), \end{cases}$$

где f_i — минимальные суммарные затраты на этапах от 1-го до i-го включительно, $i = 1, 2, ..., n$, C_k — суммарные затраты на размещение заказа, плюс затраты на этапе k на закупку продукции в объеме $z_k = D_k + ... + D_i$ для этапов от k-го до i-го, $k \le i$.

В модифицированном алгоритме состояние системы не определяется, так как оно в точности соответствует числу предыдущих этапов i.

При использовании модифицированного алгоритма объем вычислений можно дополнительно сократить на основе следующей теоремы.

Теорема о горизонте планирования. *Если на этапе i^* минимум затрат достигается таким образом, что спрос на этапе i^* удовлетворяется за счет размещения заказа на предыдущем этапе $i^{**} < i^*$, то для всех последующих этапов $i > i^*$ достаточно определить оптимальную стратегию управления запасами, основанную на размещении заказов только на этапах i^{**}, $i^{**} + 1$, ..., i. В частности, если оптимальная стратегия требует размещения заказа на этапе i^* для удовлетворения спроса этого же этапа (т.е. $i^* = i^{**}$), то для любого последующего этапа $i > i^*$ всегда оптимально размещать заказ на этапе i^*, не учитывая спрос в будущем. В этом случае i^* — начало горизонта планирования.*

Пример 11.4–4

Рассмотрим задачу определения оптимальной стратегии в шестиэтапной модели управления запасами со следующими данными.

Период i	Спрос, D_i (единицы)	Затраты на оформление заказа, K_i (\$)	Затраты на хранение, h_i (\$)
1	10	20	1
2	15	17	1
3	7	10	1
4	20	18	3
5	13	5	1
6	25	50	1

Закупочная цена единицы продукции равна 2 доллара на всех этапах.

Результаты вычислений для этого примера приведены в табл. 11.2. Они выполняются построчно, начиная со строки 1. Каждый столбец представляет альтернативное решение, определяющее этап k, на котором удовлетворяется спрос для этапов k, $k + 1$, ..., и i. Каждая строка представляет предельный этап, до которого удовлетворяется спрос. Таким образом, при каждом i оптимальное значение f_i получается путем анализа всех допустимых альтернативных решений k ($\leq i$) и последующего выбора альтернативы, обеспечивающей минимальные суммарные затраты. Так, например, при $i = 3$ существует три альтернативы: заказывать на этапе 1 для этапов 1, 2 и 3; на этапе 2 — для этапов 2 и 3; на этапе 3 — для этапа 3. Элементы табл. 11.2, расположенные над главной диагональю, недопустимы, так как дефицит в данном случае не разрешается.

Подтверждением положений теоремы о горизонте планирования служит появление f_3 в строке 3 ($i = 3$) на этапе 2 ($k = 2$). Это означает, что оптимальным решением для этапа 3 (и этапа 2) является размещение заказа на этапе 2. Это эквивалентно тому, что $i^{**} = 2$ и $i^* = 3$. Согласно сформулированной теореме, все вычисления для последующих строк ($i > 3$) можно выполнять начиная со второго этапа. Таким образом, этап 2 соответствует началу *подгоризонта.* При переходе к строке 4 видим, что величина f_4 появляется на этапе 4, а значит, оптимальное решение

Таблица 11.2

		$k = 1$[a]	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$
$i = 1$	(1)[б]	20					
	(2)	$10 \times 2 = 20$					
	(3)	0					
	(4)	$\underline{0}$					
		$f_1 \to 40^*$					
$i = 2$	(1)	20	17				
	(2)	$(10 + 15) \times 2 = 50$	$15 \times 2 = 30$				
	(3)	$15 \times 1 = 15$	0				
	(4)	$\underline{0}$	$f_1 = \underline{40}$				
		$f_1 \to 85^*$	87				
$i = 3$	(1)	20	17	10			
	(2)	$(10 + 15 + 7) \times 2 = 64$	$(15 + 7) \times 2 = 44$	$7 \times 2 = 14$			
	(3)	$22 \times 1 + 7 \times 1 = 29$	$7 \times 1 = 7$	0			
	(4)	$\underline{0}$	$f_1 = \underline{40}$	$f_2 = \underline{85}$			
		113	$f_3 \to 108^*$	109			
$i = 4$	(1)		17	10	18		
	(2)		$(15 + 7 + 20) \times 2 = 84$	$(7 + 20) \times 2 = 54$	$20 \times 2 = 40$		
	(3)		$27 \times 1 + 20 \times 1 = 47$	$20 \times 1 = 20$	0		
	(4)		$f_1 = \underline{40}$	$f_2 = \underline{85}$	$f_3 = \underline{108}$		
			188	169	$f_4 \to 166^*$		
$i = 5$	(1)				18	5	
	(2)				$(20 + 13) \times 2 = 66$	$13 \times 2 = 26$	
	(3)				$13 \times 3 = 39$	0	
	(4)				$f_3 = \underline{108}$	$f_4 = \underline{166}$	
					231	$f_5 \to 197^*$	
$i = 6$	(1)					5	50
	(2)					$(13 + 25) \times 2 = 76$	$25 \times 2 = 50$
	(3)					$25 \times 1 = 25$	0
	(4)					$f_4 = \underline{166}$	$f_5 = \underline{197}$
						$f_6 \to 272^*$	297

[a] Размещение заказов на этапе k для этапов по i-й включительно.

[б] (1) Затраты на оформление заказа, (2) затраты на закупку, (3) затраты на хранение, (4) оптимальные суммарные затраты предыдущих этапов.

предусматривает размещение заказа для этапа 4 на самом этапе 4. Следовательно, значения $i^{**} = i^{*} = 4$ и $i = 4$ определяют начало горизонта планирования. Поэтому в последующих строках 5 и 6 элементы для этапов 1–3 вычислять не нужно. Продолжая аналогичным образом, мы видим из табл. 11.2, что новый горизонт планирования начинается на этапе 5. В результате, вычисления в строке 6 необходимо выполнить только для этапов 5 и 6. Таким образом, этапы 1, 4 и 5 определяют начало трех горизонтов планирования задачи.

Преимущество теоремы о горизонте планирования теперь очевидно, так как нет необходимости проводить вычисления в незаполненных ячейках под главной диагональю таблицы. Это свидетельствует о том, что объем вычислений сократился, и, следовательно, вычисления стали более эффективными.

Оптимальная стратегия управления запасами получается при рассмотрении последней строки табл. 11.2. Величина f_6 показывает, что оптимальным является решение о размещении на этапе 5 заказа объемом $z_5 = 38$ единиц для удовлетворения спроса на этапах 5 и 6. Далее из строки 4 (= 5 – 1) в соответствии с f_4 следует, что требуется размещение заказа $z_4 = 20$ только для этапа 4. В строке 3 (= 4 – 1), согласно значению f_3, на втором этапе требуется разместить заказ размером $z_2 = 22$ единицы для удовлетворения спроса на этапе 2 и 3. Наконец, на первом этапе размещается заказ $z_1 = 10$. При этом суммарные затраты составляют 272 доллара.

Упражнения 11.4,d

1. Решите задачу из примера 11.4–3 в предположении, что начальный запас равен 80 единиц.

2. Решите следующую 10-этапную детерминированную задачу управления запасами в предположении, что исходный запас равен 50 единиц.

Этап i	Спрос, D_i (единицы)	Стоимость единицы продукции ($)	Затраты на хранение единицы продукции ($)	Затраты на оформление заказа ($)
1	150	6	1	100
2	100	6	1	100
3	20	4	2	100
4	40	4	1	200
5	70	6	2	200
6	90	8	3	200
7	130	4	1	300
8	180	4	4	300
9	140	2	2	300
10	50	6	1	300

3. Определите оптимальную стратегию управления запасами в следующей 5-этапной задаче. Стоимость единицы продукции равна 20 долларов для первых 30 единиц и 10 долларов для любого дополнительного количества (оптовая скидка). Стоимость хранения единицы продукции на протяжении периода равна 1 доллару.

Этап i	Спрос, D_i (единицы)	Затраты на оформление заказа, K_i ($)
1	50	80
2	70	70
3	100	60
4	30	80
5	60	60

4. В задаче из примера 11.4–4 определите оптимальную стратегию управления запасами (непосредственно из табл. 11.2), предполагая, что задача решается только для первых пяти этапов.
5. Решите упр. 2, предполагая, что затраты на закупку единицы продукции равны 6 долларов и являются постоянными для всех этапов. Определите горизонты и подгоризонты планирования.
6. Решите упр. 3, предполагая, что затраты на закупку единицы продукции равны 25 долларов и являются постоянными для всех этапов.

Эвристический подход Сильвера–Мила. Данный подход применим к решению только тех задач управления запасами, в которых затраты на закупку единицы продукции постоянны и одинаковы для всех этапов. Поэтому эвристический подход стремится сбалансировать лишь стоимости размещения заказа и затраты на хранение.

Эвристический метод определяет последующие этапы, потребности которых можно удовлетворить за счет размещения заказа на протяжении текущего периода. Задача планирования заключается в минимизации затрат, которые связаны с размещением заказа и хранением продукции и отнесены к одному периоду.

Предположим, на этапе i размещается заказ для периодов i, $i + 1$, ..., и t $(i \le t)$. Пусть $TC(i, t)$ — соответствующая стоимость размещения заказов и хранения продукции для этих же этапов. С использованием обозначений, принятых для моделей динамического программирования, математически это можно выразить следующим образом.

$$TC(i,t) = \begin{cases} K_i, & t = i, \\ K_i + h_i D_{i+1} + (h_i + h_{i+1}) D_{i+2} + \ldots + (h_i + h_{i+1} + \ldots + h_{t-1}) D_t, & t > i. \end{cases}$$

Обозначим далее через $TCU(i, t)$ соответствующие затраты за период, т.е.

$$TCU(i,t) = \frac{TC(i,t)}{t - i + 1}.$$

Таким образом, для заданного текущего этапа i эвристический метод определяет t^*, которое минимизирует функцию $TCU(i, t)$.

Функция $TC(i, t)$ определяется с помощью рекуррентных соотношений.

$$TC(i, i) = K_i,$$

$$TC(i, t) = TC(i, t-1) + (h_i + h_{i+1} + \ldots + h_{t-1}) D_t, \quad t = i + 1, i + 2, \ldots, n.$$

Алгоритм эвристического метода состоит из следующих шагов.

Шаг 0. Полагаем $i = 1$.

Шаг 1. Определяем локальный минимум t^* функции $TCU(i, t)$, который должен удовлетворять неравенствам

$$TCU(i, t^* - 1) \geq TCU(i, t^*),$$
$$TCU(i, t^* + 1) \geq TCU(i, t^*).$$

Тогда в соответствии с эвристическим подходом на этапе i размещается заказ объемом $(D_i + D_{i+1} + \ldots + D_{t^*})$ для этапов $i, i + 1, \ldots, t^*$.

Шаг 2. Полагаем $i = t^* + 1$. Если $i > n$, вычисления заканчиваются; рассмотрен весь плановый период. Иначе следует перейти к шагу 1.

Пример 11.4–5

Применим эвристический метод для решения задачи из примера 11.4–4, в котором стоимость закупки единицы продукции равна 2 доллара для всех этапов, а стоимость хранения единицы продукции — 1 доллар для всех этапов, кроме четвертого, где она равна 3 доллара.

Итерация 1 ($i = 1$, $K_1 = 20$ долларов).

Функция $TC(1, t)$ определяется рекуррентно по t. Например, при заданном значении $TC(1, 1) = \$20$, $TC(1, 2) = TC(1, 1) + h_1D_2 = 20 + 1 \times 15 = \$35$.

Этап t	D_t	$TC(1, t)$	$TCU(1, t)$
1	10	20	20/1 = 20.00
2	15	20 + 1 × 15 = 35	35/2 = 17.50
3	7	35 + (1 + 1) × 7 = 49	49/3 = 16.33
4	20	49 + (1 + 1 + 1) × 20 = 109	109/4 = 27.25

Локальный минимум достигается при $t^* = 3$, что означает необходимость размещения на первом этапе заказа объемом $10 + 15 + 7 = 32$ единицы для этапов 1, 2, 3. Полагаем $i = t^* + 1 = 3 + 1 = 4$.

Итерация 2 ($i= 4$, $K_4 = 18$ долларов).

Этап t	D_t	$TC(4, t)$	$TCU(4, t)$
4	10	18	18/1 = 18.00
5	13	18 + 3 × 13 = 57	52/2 = 28.50

Значение $t^* = 1$ означает, что на четвертом этапе необходимо разместить заказ объемом 20 единиц для этапа 4. Полагаем $i = 4 + 1 = 5$.

Итерация 3 ($i = 5$, $K_5 = 5$ долларов).

Этап t	D_t	$TC(5, t)$	$TCU(5, t)$
5	13	5	5/1 = 5
6	25	5 + 1 × 25 = 30	30/2 = 15

Так как $t^* = 1$, на пятом этапе заказывается 13 единиц для этапа 5. Полагаем далее $i = 5 + 1 = 6$. Так как $i = 6$, то это последний этап планирования. Мы должны заказать на шестом этапе 25 единиц для этого же этапа.

В следующей таблице сравниваются решения, полученные эвристическим методом и точным методом динамического программирования. Мы исключили из части таблицы, содержащей результаты динамического программирования, стоимость закупки единицы продукции, так как этот параметр в вычислениях с помощью эвристического метода не учитывается.

Этап	Эвристический метод		Метод динамического программирования	
	Закуплено единиц	Стоимость ($)	Закуплено единиц	Стоимость ($)
1	32	49	10	20
2	0	0	22	24
3	0	0	0	0
4	20	18	20	18
5	13	5	38	30
6	25	50	0	0
Всего	90	122	90	92

Стоимость производственного плана, предложенного эвристическим методом, примерно на 32% превышает стоимость аналогичного плана, полученного методами динамического программирования (122 доллара против 90). "Неадекватность" результата эвристического метода может быть обусловлена данными, которые использовались в задаче. В частности, причиной этого может быть чрезвычайная неравномерность стоимостей размещения заказов для этапов 5 и 6. Тем не менее, этот пример показывает, что эвристический метод не обладает способностью "смотреть вперед" в поисках лучшего производственного плана. Например, размещение на пятом этапе заказов для этапов 5 и 6 (вместо размещения их в отдельности) может сэкономить 25 долларов, что уменьшит суммарные затраты производственного плана, предложенного эвристическим методом, до 97 долларов.

Упражнения 11.4,e

1. Спрос на удилища достигает минимума в декабре, а максимума — в апреле. Компания, которая изготовляет удилища, оценивает декабрьский спрос на них в 50 единиц. Затем спрос увеличивается на 10 удилищ в месяц, достигая максимального значения 90 единиц в апреле. После этого объем спроса уменьшается на 5 удилищ в месяц. Стоимость размещения заказа на изготовление партии удилищ равна 250 долларов на протяжении всех месяцев, за исключением февраля, марта и апреля, когда она составляет 300 долларов. Стоимость производства одного удилища является примерно постоянной на протяжении всего года и составляет 15 долларов, а стоимость хранения одного удилища равна 1 доллару в месяц. Требуется составить план производства удилищ.

2. На протяжении последующих 12 месяцев небольшое издательство переиздает роман в целях удовлетворения спроса на него: 100, 120, 50, 70, 90, 105, 115, 95, 80, 85, 100 и 110 экземпляров. Стоимость размещения заказа на переиздание равна 200 долларов, а стоимость хранения книги на протяжении одного месяца — 1.20 доллара. Определите план переиздания книги издательством.

11.5. Заключение

В задаче управления запасами рассматривается, *когда* заказывать и *какое количество* продукции, которая будет храниться на складе. Решения таких задач находятся путем минимизации соответствующей функции затрат, которая учитывает стоимость размещения заказа, хранения, а также потери, связанные с дефицитом. В этой главе рассмотрены различные детерминированные модели указанной задачи. Степень сложности той или иной модели определяется главным образом характером спроса. В частности, математическая модель задачи при статическом спросе является более простой. Задачи с динамическим спросом требуют применения методов динамического программирования и, следовательно, более сложные. В главе 16 будут рассмотрены задачи управления запасами, в которых учитывается случайный характер спроса.

Литература

1. Silver E. and Peterson R. *Decision Systems for Inventory Management and Production Control,* 2nd ed., Wiley, New York, 1985.
2. Tersine R. *Principles of Inventory and Materials Management*, 3rd ed., North Holland, New York, 1988.
3. Waters C. *Inventory Control and Management*, Wiley, New York, 1992.

Литература, добавленная при переводе

Кофман А. *Методы и модели исследования операций.* — М.: Мир, 1966.

Комплексные задачи

■ **11–1.** Распределительный центр универмага специализируется на ежедневной покупке и хранении предметов торговли, вышедших из моды. Постоянный спрос на такие предметы поступает от многочисленных торговых точек универмага. В прошлом решения относительно того, *когда* и *сколько* товара заказывать, перекладывались на отдел поставки, главная задача которого состояла в том, чтобы приобрести продукцию в достаточно больших объемах, дабы гарантировать низкие закупочные цены. Эта стратегия применялась без надлежащего рассмотрения фактора хранения продукции. Действительно, решения относительно того, сколько товара закупать, основывались на годовой стоимости спроса на товар на уровне распределительного центра. Например, если единица продукции приобретается по цене 25 долларов и в год используется 10 000 единиц, то годовая стоимость спроса на этот товар составляет 250 000 долларов. Отдел поставки руководствовался основным принципом: чем выше годовая стоимость спроса на товар, тем больше должен быть уровень его запаса в распределительном центре. Этот принцип затем выражался в объеме запаса продукции, который должен храниться в распределительном центре в период между пополнениями. Например, отдел поставки мог закупать заранее определенное количество продукции каждые три месяца.

Чтобы улучшить стратегию управления запасами, руководство универмага решило прибегнуть к услугам консультанта по исследованию операций. Изучив ситуацию, он пришел к выводу, что интенсивность потребления большинства видов продукции в распределительном центре с практической точки зрения является постоянной и что проводится политика отсутствия дефицита. Дальнейшее изучение показало, что стоимость хранения всех рассматриваемых видов продукции составляет один и тот же постоянный процент от закупочной цены. Кроме того, стоимость размещения заказа для всех рассматриваемых видов продукции является одинаковой. С помощью этой информации консультант смог построить для каждого вида продукции соответствующую кривую, которая устанавливает связь годовой стоимости спроса на товар со средним временем между пополнениями товара. Эта кривая была затем использована для того, чтобы выяснить, по какой продукции в настоящее время имеется излишний запас, а по какой — недостаточный. Как консультант сделал это?

■ **11–2**. Компания производит конечный продукт с использованием единственного комплектующего блока, который она закупает у внешнего поставщика. Спрос на конечный продукт является постоянным и равным примерно 20 изделий в неделю. Для изготовления каждой единицы конечного продукта требуется два комплектующих блока. Имеется также следующая информация для рассматриваемой задачи управления запасами.

	Комплектующие	Продукция
Стоимость размещения заказа ($)	80	100
Стоимость хранения единицы в неделю ($)	2	5
Срок изготовления (недели)	2	3

Неудовлетворенный спрос на готовую продукцию является задолженностью компании и приносит ей потери в 8 долларов в неделю за единицу спроса. Разработайте стратегию как размещения заказов на комплектующие, так и производства конечной продукции.

■ **11–3.** Компания производит сезонную продукцию, спрос на которую ощутимо меняется от месяца к месяцу. Данные об объемах спроса (в количестве единиц продукции) за последние пять лет приведены в следующей таблице.

	Год				
Месяц	1	2	3	4	5
Январь	10	11	10	12	11
Февраль	50	52	60	50	55
Март	8	10	9	15	10
Апрель	99	100	105	110	120
Май	120	100	110	115	110
Июнь	100	105	103	90	100
Июль	130	129	125	130	130
Август	70	80	75	75	78
Сентябрь	50	52	55	54	51
Октябрь	120	130	140	160	180
Ноябрь	210	230	250	280	300
Декабрь	40	46	42	41	43

Принимая во внимание колебания спроса на продукцию, менеджер по управлению запасами избрал стратегию, в соответствии с которой заказы на продукцию размещаются поквартально: 1 января, 1 апреля, 1 июля и 1 октября. При этом объем заказа покрывает объем спроса соответствующего квартала. Срок между размещением заказа и его получением равен 3 месяца. Оценки объема спроса на текущий год принимаются равными соответствующим показателям пятого года плюс дополнительно 10% в качестве безопасности.

Новый сотрудник верит в то, что можно достичь более эффективной стратегии управления запасами, используя экономичный объем заказа, основанный на среднемесячном объеме спроса на продукцию за год. Колебания спроса могут быть сглажены путем размещения заказов, которые по объему примерно равны экономичному объему партии и покрывают спрос нескольких последовательных месяцев. В отличие от менеджера, новый сотрудник считает, что оценки объема спроса на следующий год должны основываться на усредненных показателях для четвертого и пятого годов.

Во всех вычислениях, связанных с хранением продукции, компания считает, что затраты на хранение единицы продукции на протяжении месяца равны 0.50 доллара. Стоимость размещения нового заказа равна 55 долларов.

Предложите стратегию управления запасами для компании.

ЧАСТЬ II

ВЕРОЯТНОСТНЫЕ МОДЕЛИ

Основы теории вероятностей

12.1. Введение

Все методы решения задач исследования операций, изложенные в предыдущих главах, предполагают, что необходимые для их реализации данные известны точно. Однако это предположение выполняется не во всех случаях. Например, потребность в электроэнергии в летние месяцы может меняться от года к году в зависимости от погодных условий. В таких случаях представление потребности в виде постоянной детерминированной величины неприемлемо. Вместо этого можно использовать наблюденные или известные из исторических источников данные для описания потребности с помощью вероятностного распределения.

12.2. Законы теории вероятностей

Понятие вероятности ассоциируется с проведением **эксперимента**, результаты которого, именуемые **исходами**, изменяются случайным образом. Множество всех возможных исходов эксперимента называется **пространством событий**, а любое подмножество этого пространства — **событием**. Например, в эксперименте с бросанием игральной шестигранной кости исход соответствует грани кости, т.е. может принимать значение от 1 до 6. Следовательно, пространство событий представляет собой множество $\{1, 2, 3, 4, 5, 6\}$. Примером события в этом эксперименте может, например, быть выпадение четного числа (2, 4 или 6).

Эксперимент может быть связан также с непрерывным пространством событий. Например, время между отказами некоторого электронного устройства может принимать любое неотрицательное значение.

Если в эксперименте, состоящем из n опытов, событие E имеет место m раз, то вероятность $P\{E\}$ появления события E математически определяется соотношением

$$P\{E\} = \lim_{n\to\infty}\frac{m}{n}.$$

Приведенное определение означает, что если эксперимент повторяется *бесконечное число раз* ($n \to \infty$), искомая вероятность представляется граничным значением дроби m/n. Это определение можно проверить посредством эксперимента с бросанием монеты, исходами которого являются выпадение герба (Γ) и решетки (P). Чем большее число раз бросается монета (проводится эксперимент), тем ближе оценки $P\{P\}$ и $P\{\Gamma\}$ к теоретическому значению 0.5.

По определению

$$0 \le P\{E\} \le 1,$$

где вероятность $P\{E\}$ равна 0, если событие E невозможно, и 1, если оно достоверно. Например, в эксперименте с шестигранной игральной костью выпадение семерки является невозможным событием, тогда как любое целое число от 1 до 6 — событие достоверное.

Упражнения 12.2,а

1. В ходе анализа, проведенного в средних школах штата Арканзас в целях изучения зависимости между успеваемостью по математике и поступлением в технические колледжи, получены следующие данные: из 1000 опрошенных выпускников 400 изучали математику. Из тех, кто изучал математику, лишь 150 были приняты в технические колледжи.

   a) Определите вероятность того, что студент, изучавший математику, 1) поступит в технический колледж, 2) не поступит в технический колледж.

   b) Определите вероятность того, что студент, не изучавший математику, не поступит в технический колледж.

2. Рассмотрим случайную совокупность из n человек. Определите наименьшее n такое, что более вероятным будет событие, состоящее в совпадении дней рождения нескольких человек (т.е. более вероятным по сравнению с событием, что у всех индивидуумов в данной совокупности даты рождения различны). (*Совет.* Предположите, что нет високосных годов и все дни года с равной вероятностью могут быть днем рождения каждого человека.)

12.2.1. Закон сложения вероятностей

Для данных двух событий E и F запись $E + F$ обозначает их **объединение**, а EF — **пересечение**. События E и F называются **несовместными (взаимно исключающими)**, если они не пересекаются, т.е. наступление одного события исключает возможность реализации другого. При принятых определениях закон сложения вероятностей определяется соотношением

$$P\{E+F\} = \begin{cases} P\{E\}+P\{F\}, & \text{если } E \text{ и } F \text{ несовместные,} \\ P\{E\}+P\{F\}-P\{EF\}, & \text{в противном случае.} \end{cases}$$

Пример 12.2–1

Рассмотрим эксперимент с игральной костью. В данном случае пространство событий представляет собой множество $\{1, 2, 3, 4, 5, 6\}$. Для симметричной кости имеем

$$P\{1\} = P\{2\} = P\{3\} = P\{4\} = P\{5\} = P\{6\} = \frac{1}{6}.$$

Определим события

$$E = \{1, 2, 3 \text{ или } 4\},$$
$$F = \{3, 4 \text{ или } 5\}.$$

Исходы 3 и 4 являются общими для событий E и F. Следовательно, $EF = \{3 \text{ или } 4\}$. Имеем

$$P\{E\} = P\{1\} + P\{2\} + P\{3\} + P\{4\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{2}{3},$$

$$P\{F\} = P\{3\} + P\{4\} + P\{5\} = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

$$P\{EF\} = P\{3\} + P\{4\} = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Отсюда следует, что

$$P\{E + F\} = P\{E\} + P\{F\} - P\{EF\} = \frac{2}{3} + \frac{1}{2} - \frac{1}{3} = \frac{5}{6}.$$

Чисто интуитивно этот результат понятен, так как событие $(E + F) = \{1, 2, 3, 4, 5\}$, очевидно, имеет вероятность 5/6.

Упражнения 12.2,b

1. В эксперименте с игральной костью вычислите вероятность события $E + F$, считая, что событие E — выпадение четного числа, меньшего 6, а F — выпадение нечетного числа, большего 1. Являются ли события E и F несовместными?
2. Анна, Джим, Джон и Лиза участвуют в теннисном турнире. Вероятность того, что Анна победит Джима, в два раза выше вероятности обратного результата, а мастерство Джима оценивается на том же уровне, что и мастерство Джона. В прошлом Лиза выигрывала у Джона примерно один раз из трех.

   a) Какова вероятность того, что Джим выиграет турнир?

   b) Какова вероятность того, что турнир выиграет женщина?

12.2.2. Условные вероятности

Для данных двух событий E и F условная вероятность события E при условии, что наступило событие F, обозначается как $P\{E \mid F\}$ и определяется по формуле

$$P\{E \mid F\} = \frac{P\{EF\}}{P\{F\}}, \quad P\{F\} > 0.$$

Если событие E содержится в событии F (т.е. множество исходов E является подмножеством множества исходов F), тогда $P\{EF\} = P\{E\}$.

Два события E и F являются *независимыми* тогда и только тогда, когда выполняется равенство $P\{E \mid F\} = P\{E\}$. В этом случае формула условной вероятности сводится к следующему

$$P\{EF\} = P\{E\}P\{F\}.$$

Пример 12.2–2

Предположим, вы участвуете в игре, в которой другой человек бросает игральную кость. Вы не можете видеть игральную кость, но вам сообщается некоторая информация об исходах бросания кости. Вам необходимо предсказать возможный исход каждого бросания кости. Определим вероятность того, что исходом будет число 6 при условии, что вам сообщили о том, что исходом бросания кости является четное число.

Пусть $E = \{6\}$; определим $F = \{2, 4 \text{ или } 6\}$. Следовательно,

$$P\{E \mid F\} = \frac{P\{EF\}}{P\{F\}} = \frac{P\{E\}}{P\{F\}} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{1}{2}\right)} = \frac{1}{3}.$$

Заметим, что $P\{EF\} = P\{E\}$, так как множество исходов E является подмножеством множества исходов F.

Упражнения 12.2,с

1. Вернитесь к примеру 12.2–2. Предположим, вам сообщили, что исход бросания кости меньше 6. Определите вероятность выпадения четного числа.
2. Бросаются независимо две игральные кости и записываются выпавшие числа.
   a) Какова вероятность того, что оба числа являются четными?
   b) Какова вероятность того, что сумма двух чисел равна 10?
   c) Какова вероятность того, что два числа отличаются по меньшей мере на 3?
3. Можно бросать симметричную монету до 7 раз и выиграть 100 долларов, если появится по крайней мере три герба до появления решетки. Каковы шансы выиграть 100 долларов?
4. *Теорема Байеса.* Покажите, что для двух заданных событий A и B имеет место соотношение
   $$P\{A \mid B\} = \frac{P\{B \mid A\} P\{A\}}{P\{B\}},$$
   где $P\{B\} > 0$.
5. Завод A поставляет в магазин 75% продаваемых аккумуляторов, а завод B — 25%. Процент бракованных аккумуляторов равен 1% и 2% соответственно для заводов A и B. Клиент купил в магазине аккумулятор.
   a) Какова вероятность того, что аккумулятор бракованный?
   b) Если купленный аккумулятор является бракованным, какова вероятность того, что он изготовлен на заводе A? (*Совет.* Используйте теорему Байеса из предыдущего упражнения.)
6. Статистика свидетельствует, что 70% мужчин болеют какой-нибудь формой рака предстательной железы. Американский тест PSA в 90% случаев дает положительный результат для пораженных болезнью мужчин и в 10% — для здоровых. Какова вероятность того, что мужчина, имеющий положительный результат теста, имеет рак предстательной железы?

12.3. Случайные величины и распределения вероятностей

Исходы эксперимента (испытания) обычно либо выражаются в числовом виде, либо им можно поставить в соответствие некоторые действительные числа. Например, исходы бросания игральной кости выражаются в виде целых чисел от 1 до 6. А проверка на брак некоторого изделия дает два исхода: некачественное и качественное. В этом случае можно использовать число 0 для представления исхода "некачественный" и 1 – для исхода "качественный". Численное представление исходов эксперимента — это то, что именуется **случайной величиной.**[1]

Случайная величина x может быть **дискретной** или **непрерывной**. Например, случайная величина, связанная с бросанием игральной кости, является дискретной со значениями от 1 до 6, тогда как время между поступлениями заявок в систему обслуживания выражается непрерывной случайной величиной с положительными значениями.

Как непрерывная, так и дискретная случайная величина имеет **плотность распределения вероятностей**, которая часто именуется просто плотностью вероятности и обозначается как $f(x)$ (для непрерывной случайной величины) или $p(x)$ (для дискретной случайной величины). Плотность вероятности ставит в соответствие случайной величине вероятностную меру. Плотности вероятностей должны удовлетворять условиям, перечисленным в следующей таблице.

Характеристики плотности	Случайная величина x	
	Дискретная	Непрерывная
Область определения	$x = a, a + 1, \ldots, b$	$a \le õ \le b$
Условия неотрицательности и нормировки	$p(x) \ge 0$, $\sum_{x=a}^{b} p(x) = 1$	$f(x) \ge 0$, $\int_a^b f(x)dx = 1$

Условие неотрицательности для непрерывных и дискретных распределений означает, что плотность вероятности не может принимать отрицательные значения (в противном случае вероятность некоторых событий могла бы быть отрицательной). Условие нормировки показывает, что сумма вероятностей по всему пространству событий должна быть равна единице.

Самой важной вероятностной характеристикой случайной величины является **функция распределения**, определяемая следующим образом:

$$P\{x \le X\} = \begin{cases} P(X) = \sum_{x=a}^{X} p(x) & \text{для дискретной случайной величины } x, \\ F(X) = \int_a^X f(x)dx & \text{для непрерывной случайной величины } x. \end{cases}$$

[1] Случайную величину можно считать функцией, отображающей пространство элементарных исходов на пространство действительных чисел. — *Прим. ред*

Пример 12.3–1

Рассмотрим случай с бросанием игральной кости. Пусть $x \in \{1, 2, 3, 4, 5, 6\}$ — случайная величина, представляющая количество выпавших очков. Тогда плотность вероятности и функция распределения вероятности случайной величины x определяются следующим образом.

$$p(x) = \frac{1}{6}, \quad x = 1, 2, \ldots, 6,$$

$$P(X) = \frac{X}{6}, \quad X = 1, 2, \ldots, 6.$$

На рис. 12.1 приведены графики этих двух функций. Плотность вероятности $p(x)$ является равномерной дискретной функцией, так как любые значения случайной величиной принимаются с одинаковыми вероятностями.

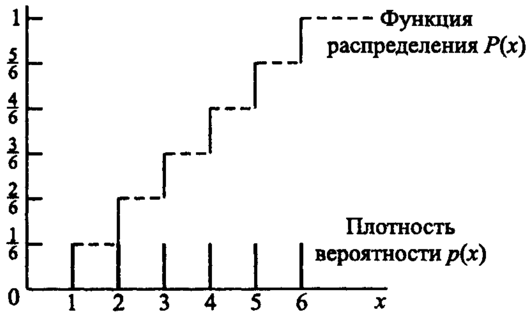


Рис. 12.1

Непрерывный аналог равномерной плотности вероятности можно получить на основе следующего эксперимента. Стрелка длиной l закреплена подвижно на оси в центре круга, радиус которого также равен l. На окружности выбирается точка отсчета, стрелка вращается в направлении часовой стрелки, по окружности измеряется расстояние x, пройденное стрелкой от точки отсчета. Такая случайная величина x является непрерывной, принимающей значения из интервала $0 \le x \le \pi l$. Нет никаких оснований считать, что стрелка будет иметь тенденцию останавливаться в некоторой области окружности чаще, чем в других областях. Поэтому все значения x из интервала $0 \le x \le \pi l$ могут приниматься с равной вероятностью и распределение x должно быть равномерным.

В данном случае плотность вероятности $f(x)$ случайной величины x определяется следующим образом.

$$f(x) = \frac{1}{\pi l}, \quad 0 \le x \le \pi l.$$

Функция распределения $F(X)$ случайной величины x вычисляется по формуле

$$F(X) = P\{x \le X\} = \int_0^X \frac{1}{\pi l} dx = \frac{X}{\pi l}, \quad 0 \le X \le \pi l.$$

На рис. 12.2 представлены графики этих двух функций.

Рис. 12.2

Упражнения 12.3,а

1. Некоторая величина принимает случайным образом целочисленное значение x из интервала $[1, 5]$. Плотность вероятности $p(x)$ этой величины прямо пропорциональна значению x с коэффициентом пропорциональности K.

   a) Определите плотность вероятности и функцию распределения данной случайной величины, нарисуйте графики полученных функций.

   b) Определите вероятность того, что случайная величина примет значение, равное четному числу.

2. Дана следующая функция:

$$f(x) = \frac{k}{x^2}, \quad 10 \le x \le 20.$$

   a) Найдите значение константы k, при котором функция $f(x)$ будет плотностью вероятности.

   b) Найдите функцию распределения случайной величины x и определите вероятность того, что случайная величина x примет значение: (i) больше 12, (ii) между 13 и 14.

3. Дневная потребность в бензине без свинца является равномерно распределенной случайной величиной, изменяющейся в интервале от 750 до 1250 галлонов. Бензоцистерна емкостью 11000 галлонов наполняется ежедневно в полночь. Какова вероятность того, что цистерна будет пустой как раз перед заполнением ее бензином?

12.4. Математические ожидания и моменты случайной величины

Пусть x — случайная величина, $h(x)$ — некоторая функция от x. **Математическим ожиданием** значений функции $h(x)$, которое обозначается как $M\{h(x)\}$, называется средняя величина, взвешенная по отношению к плотности вероятности случайной величины x. При заданной плотности вероятности ($p(x)$ или $f(x)$ для дискретной и непрерывной случайных величин соответственно) величина $M\{h(x)\}$ вычисляется следующим образом:

$$M\{h(x)\}=\begin{cases}\sum\limits_{x=a}^{b} h(x)\,p(x), & \text{если } x \text{ — дискретная случайная величина,}\\ \int\limits_{a}^{b} h(x)\,f(x)\,dx, & \text{если } x \text{ — непрерывная случайная величина.}\end{cases}$$

Пример 12.4–1

В течение первой недели каждого месяца я, как и большинство людей, оплачиваю все свои счета и отвечаю на некоторые письма. С этой целью я обычно покупаю 20 почтовых марок. Число используемых марок является случайной величиной, принимающей значения от 10 до 24 с равными вероятностями. Найдем, чему равно среднее число оставшихся марок.

Пусть x — количество используемых марок, тогда плотность вероятности x такова:

$$p(x)=\frac{1}{15},\quad x=10,11,\ldots,24.$$

Количество оставшихся марок определяется соотношением

$$h(x)=\begin{cases}20-x, & x=10,11,\ldots,19,\\ 0, & \text{в противном случае.}\end{cases}$$

Следовательно,

$$M\{h(x)\}=\left(\frac{1}{15}\right)\left[(20-10)+(20-11)+(20-12)+\ldots+(20-19)\right]+\left(\frac{5}{15}\right)(0)=3\frac{2}{3}.$$

Произведение $\left(\frac{5}{15}\right)(0)$ необходимо для завершения вычисления математического ожидания. Вероятность того, что вообще не останется марок, равна

$$P\{x\geq 20\}=P(20)+P(21)+P(22)+P(23)+P(24)=5\left(\frac{1}{15}\right)=\frac{5}{15}.$$

Упражнения 12.4,а

1. В задаче из примера 12.4–1 вычислите среднее количество оставшихся марок при условии, что ежемесячно покупается число марок, соответствующее максимально возможной потребности в них.
2. Результаты решения задачи из примера 12.4–1 и предыдущего упражнения приводят к *положительным* значениям средних величин как при избытке, так и недостаче марок. Являются ли эти результаты противоречивыми? Дайте объяснение.

12.4.1. Математическое ожидание и дисперсия случайной величины

Для общей характеристики свойств одномерной случайной величины x обычно используется две числовые характеристики: **математическое ожидание** (среднее) $M\{x\}$ и **дисперсия** $D\{x\}$. Математическое ожидание является характеристикой положения распределения случайной величины x на числовой оси относительно начала координат, а дисперсия — мерой ее разброса относительно математического ожидания $M\{x\}$. Большее значение дисперсии свидетельствует о более высокой степени неопределенности в описании случайной величины.

Формулы для математического ожидания и дисперсии случайной величины x могут быть получены из общей формулы для математического ожидания путем подстановки $h(x) = x$ для $M\{x\}$ и $h(x) = (x - M\{x\})^2$ для $D\{x\}$. Следовательно,

$$M\{x\} = \begin{cases} \sum\limits_{x=a}^{b} xp(x), & \text{если } x \text{ — дискретная случайная величина,} \\ \int\limits_{a}^{b} xf(x)dx, & \text{если } x \text{ — непрерывная случайная величина,} \end{cases}$$

$$D\{x\} = \begin{cases} \sum\limits_{x=a}^{b} (x - M\{x\})^2 p(x), & \text{если } x \text{ — дискретная случайная величина,} \\ \int\limits_{a}^{b} (x - M\{x\})^2 f(x)dx, & \text{если } x \text{ — непрерывная случайная величина.} \end{cases}$$

Обоснованность вывода указанных формул легче просматривается для дискретного распределения. В этом случае $M\{x\}$ представляет собой взвешенную сумму дискретных значений x, $D\{x\}$ — взвешенную сумму квадратов отклонения случайной величины x от $M\{x\}$. Случай с непрерывно распределенной случайной величиной можно интерпретировать аналогично, заменив интегрирование суммированием.

Пример 12.4–2

Вычислим математическое ожидание и дисперсию для каждого из двух экспериментов, рассмотренных в примере 12.3–1.

Случай 1 (бросание игральной кости)

Здесь плотность вероятности равна $p(x) = \frac{1}{6}$, $x = 1, 2, ..., 6$.

Следовательно,

$$M\{x\} = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = 3.5,$$

$$D\{x\} = \frac{1}{6}\left[(1-3,5)^2 + (2-3,5)^2 + (3-3,5)^2 + (4-3,5)^2 + (5-3,5)^2 + (6-3,5)^2\right] = 2.917.$$

Случай 2 (вращение стрелки)

Пусть длина стрелки равна единице. Тогда

$$f(x) = \frac{1}{3.14}, \quad 0 \le x \le 3.14.$$

Математическое ожидание и дисперсия вычисляются следующим образом.

$$M\{x\} = \int_0^{3,14} x\frac{1}{3.14}dx = 1.57,$$

$$D\{x\} = \int_0^{3,14} (x-1.57)^2 \frac{1}{3.14}dx = 0.822.$$

Упражнения 12.4,b

1. Вычислите математическое ожидание и дисперсию случайной величины, определенной в упр. 12.3,a(1).
2. Вычислите математическое ожидание и дисперсию случайной величины, определенной в упр. 12.3,a(2).
3. Покажите, что математическое ожидание и дисперсия случайной величины x, равномерно распределенной на интервале $a \le x \le b$, равны

$$M\{x\} = \frac{b+a}{2},$$

$$D\{x\} = \frac{(b-a)^2}{12}.$$

4. Докажите, что для случайной величины x, определенной на интервале $a \le x \le b$ с заданной плотностью вероятности $f(x)$, имеет место соотношение

$$D\{x\} = M\{x^2\} - (M\{x\})^2.$$

5. Пусть для случайной величины x, определенной на интервале $a \le x \le b$, задана плотность вероятности $f(x)$ и $y = cx + d$, где c и d — константы. Докажите, что имеют место соотношения

$$M\{y\} = cM\{x\} + d,$$

$$D\{y\} = c^2 D\{x\}.$$

12.4.2. Совместные распределения вероятностей

Рассмотрим две непрерывно распределенные случайные величины x_1 и x_2, определенные соответственно на интервалах $a_1 \le x_1 \le b_1$ и $a_2 \le x_2 \le b_2$. Обозначим через $f(x_1, x_2)$ **плотность совместного распределения вероятностей** величин x_1 и x_2, а через $f_1(x_1)$ и $f_2(x_2)$ — **маргинальные (частные) плотности распределения вероятностей** величин x_1 и x_2 соответственно. Тогда

$$f(x_1, x_2) \ge 0, \ a_1 \le x_1 \le b_1, \ a_2 \le x_2 \le b_2,$$

$$\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 f(x_1, x_2) = 1,$$

$$f_1(x_1) = \int_{a_2}^{b_2} f(x_1, x_2) dx_2,$$

$$f_2(x_2) = \int_{a_1}^{b_1} f(x_1, x_2) dx_1,$$

$$f(x_1, x_2) = f_1(x_1) f_2(x_2), \text{ если } x_1 \text{ и } x_2 \text{ независимы.}$$

Такие же формулы используются для дискретно распределенных случайных величин, которые получаются в результате замены интегрирования суммированием.

Далее в этом разделе рассматриваются функции от нескольких случайных переменных. В частности, рассмотрим два случая

1. $y = x_1 x_2$,
2. $y = c_1 x_1 + c_2 x_2$,

где x_1 и x_2 — случайные величины, плотность совместного распределения вероятностей которых задана функцией $f(x_1, x_2)$.

Если x_1 и x_2 независимые случайные величины, то

$$M\{x_1 x_2\} = M\{x_1\} M\{x_2\}.$$

Для суммы случайных величин x_1 и x_2, независимо от того, зависимы ли эти величины, можно доказать, что

$$M\{c_1 x_1 + c_2 x_2\} = c_1 M\{x_1\} + c_2 M\{x_2\}.$$

Кроме того,

$$D\{c_1 x_1 + c_2 x_2\} = c_1^2 D\{x_1\} + c_2^2 D\{x_2\} + 2c_1 c_2 \operatorname{cov}\{x_1, x_2\},$$

где *ковариация* $\operatorname{cov}\{x_1, x_2\}$ случайных величин x_1 и x_2 вычисляется по формуле

$$\begin{aligned}\operatorname{cov}\{x_1, x_2\} = {} & M\{(x_1 - M\{x_1\})(x_2 - M\{x_2\})\} = \\ & M\{x_1 x_2 - x_1 M\{x_2\} - x_2 M\{x_1\} + M\{x_1\} M\{x_2\}\} = \\ & M\{x_1 x_2\} - M\{x_1\} M\{x_2\}.\end{aligned}$$

Если x_1 и x_2 — *независимые* случайные величины, то $M\{x_1 x_2\} = M\{x_1\} M\{x_2\}$ и $\operatorname{cov}\{x_1, x_2\} = 0$. Обратное утверждение неверно в том смысле, что две *зависимые* случайные величины могут иметь ковариацию, равную нулю.

Пример 12.4–3

Партия изделий содержит четыре дефектных (*Д*) и шесть качественных (*К*) изделий. Случайным образом выбирается и проверяется одно изделие. Затем, не возвращая его, выбирается и тестируется другое. Пусть случайные величины x_1 и x_2 представляют исходы тестирования первого и второго изделий соответственно.

a) Определим плотность совместного распределения вероятностей случайных величин x_1 и x_2.

b) Найдем маргинальную плотность вероятности случайной величины x_2.

c) Предположим, мы получаем 5 долларов за каждое качественное изделие из выбранных и платим 6 долларов, если изделие бракованное. Найдем математическое ожидание и дисперсию выигрыша после двух испытаний.

Обозначим через $p(x_1, x_2)$ плотность совместного распределения вероятностей случайных величин x_1 и x_2, а через $p_1(x_1)$ и $p_2(x_2)$ — их маргинальные плотности вероятностей. Определим сначала $p_1(x_1)$ как

$$p_1(К) = \frac{6}{10}, \quad p_1(Д) = \frac{4}{10}.$$

Мы знаем, что исход x_2 второго испытания зависит от x_1. По этой причине для определения $p_2(x_2)$ сначала определим плотность $p(x_1, x_2)$ совместного распределения вероятностей случайных величин x_1 и x_2, после чего можно будет определить маргинальную плотность вероятности $p_2(x_2)$. Имеем

$$P\{x_2 = К \mid x_1 = К\} = \frac{5}{9},$$

$$P\{x_2 = К \mid x_1 = Д\} = \frac{6}{9},$$

$$P\{x_2 = Д \mid x_1 = К\} = \frac{4}{9},$$

$$P\{x_2 = Д \mid x_1 = Д\} = \frac{3}{9}.$$

Для определения $p(x_1, x_2)$ воспользуемся формулой $P\{AB\} = P\{A \mid B\}P\{B\}$ (см. раздел 12.2.2). Получаем следующее.

$$P\{x_2 = К, x_1 = К\} = \frac{5}{9} \times \frac{6}{10} = \frac{5}{15},$$

$$P\{x_2 = К, x_1 = Д\} = \frac{6}{9} \times \frac{4}{10} = \frac{4}{15},$$

$$P\{x_2 = Д, x_1 = К\} = \frac{4}{9} \times \frac{6}{10} = \frac{4}{15},$$

$$P\{x_2 = Д, x_1 = Д\} = \frac{3}{9} \times \frac{4}{10} = \frac{2}{15}.$$

Представим теперь плотность совместного распределения следующим образом.

		$x_2 = К$	$x_2 = Д$	$p_1(x_1)$
$p(x_1, x_2) =$	$x_1 = К$	5/15	4/15	9/15
	$x_1 = Д$	4/15	2/15	6/15
	$p_2(x_2)$	9/15	6/15	

Маргинальные плотности распределения вероятностей $p_1(x_1)$ и $p_2(x_2)$ могут быть определены посредством суммирования элементов, соответственно, столбцов или строк в таблице, представляющей значения плотности совместного распределения. Интересно отметить, что, вопреки интуиции, здесь $p_1(x_1) = p_2(x_2)$.

Математическое ожидание выигрыша можно определить из совместного распределения, если принять, что изделие *К* дает 5 долларов, а изделие *Д* — 6 долларов. Следовательно,

$$\text{ожидаемый выигрыш} = (\$5+\$5)\left(\frac{5}{15}\right)+(\$5-\$6)\left(\frac{4}{15}\right)+$$
$$+(-\$6+\$5)\left(\frac{4}{15}\right)+(-\$6-\$6)\left(\frac{2}{15}\right)=\$1.20.$$

Тот же результат можем получить, принимая во внимание, что математическое ожидание выигрыша после двух испытаний равно сумме математических ожиданий после каждого испытания в отдельности.

$$\text{ожидаемый выигрыш} = (\text{ожидаемый выигрыш после 1-го испытания})+$$
$$+(\text{ожидаемый выигрыш после 2-го испытания})=$$
$$=\left[\$5\left(\frac{9}{15}\right)-\$6\left(\frac{6}{15}\right)\right]+\left[\$5\left(\frac{9}{15}\right)-\$6\left(\frac{6}{15}\right)\right]=$$
$$=\$0.60+\$0.60=\$1.20.$$

Для вычисления дисперсии общего выигрыша заметим, что

$$D\{\text{выигрыша}\} = D\{\text{1-го выигрыша}\} + D\{\text{2-го выигрыша}\} +$$
$$+ 2\operatorname{cov}\{\text{1-го выигрыша, 2-го выигрыша}\}.$$

Так как $p_1(x_1) = p_2(x_2)$, то $D\{\text{1-го выигрыша}\} = D\{\text{2-го выигрыша}\}$. Для вычисления дисперсии воспользуемся следующей формулой (см. упр. 12.4,b(4)).

$$D\{x\} = M\{x^2\} - (M\{x\})^2.$$

Следовательно,

$$D\{\text{1-го выигрыша}\}=\left[(5)^2\left(\frac{9}{15}\right)+(-6)^2\left(\frac{6}{15}\right)\right]-0.6^2=29.04.$$

Далее для вычисления ковариации применим формулу

$$\operatorname{cov}\{x_1,x_2\}=M\{x_1x_2\}-M\{x_1\}M\{x_2\}.$$

При вычислении значения $M\{x_1x_2\}$ нужно знать плотность совместного распределения вероятностей случайных величин x_1 и x_2. Имеем

$$\text{ковариация}=\left[(5\times5)\left(\tfrac{5}{15}\right)+(5\times(-6))\left(\tfrac{4}{15}\right)+(-6\times5)\left(\tfrac{4}{15}\right)+(-6\times(-6))\left(\tfrac{2}{15}\right)\right]-$$
$$-0.6\times0.6=-3.23.$$

Итак,

$$\text{дисперсия}=29.04+29.04+2(-3.23)=51.62.$$

Упражнение 12.4,c

1. Плотность совместного распределения вероятностей $p(x_1, x_2)$ случайных величин x_1 и x_2 имеет следующий вид.

$p(x_1, x_2) =$

	$x_2 = 3$	$x_2 = 5$	$x_2 = 7$
$x_1 = 1$	0.2	0	0.2
$x_1 = 2$	0	0.2	0
$x_1 = 3$	0.2	0	0.2

   a) Найдите маргинальные плотности вероятностей $p_1(x_1)$ и $p_2(x_2)$.
   b) Являются ли случайные величины x_1 и x_2 независимыми?
   c) Определите $M\{x_1 + x_2\}$.
   d) Найдите $\text{cov}\{x_1, x_2\}$.
   e) Вычислите $D\{5x_1 - 6x_2\}$.

12.5. Некоторые распределения вероятностей

В разделах 12.3 и 12.4 мы рассмотрели равномерные распределения (дискретные и непрерывные). В этом разделе рассматриваются еще четыре распределения случайных величин, которые часто используются в теории исследования операций, — дискретные (биномиальное и Пуассона) и непрерывные (экспоненциальное и нормальное).

12.5.1. Биномиальное распределение

Предположим, предприниматель изготавливает некоторые изделия партиями по n единиц в каждой. Предыдущий опыт свидетельствует, что вероятность появления бракованного изделия в каждой партии равна p. Необходимо определить плотность распределения числа бракованных изделий в партии.

Имеется $C_n^x = \dfrac{x!(n-x)!}{n!}$ различных возможностей получить x бракованных изделий в партии из n изделий; вероятность реализации каждой такой комбинации равна $p^x(1-p)^{n-x}$. Отсюда следует, что вероятность того, что в партии из n изделий имеется k бракованных, равна (что следует из закона сложения вероятностей)

$$P\{x = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 1, 2, \ldots, n.$$

Это формула плотности вероятности биномиального распределения с параметрами n и p. Математическое ожидание и дисперсия для этого распределения равны

$$M\{x\} = np, \; D\{x\} = np(1-p).$$

Пример 12.5–1

Некая работа сопряжена с десятью поездками на автомобиле между двумя городами. Выполнив все 10 поездок, работник может отдыхать остаток дня, что является достаточным стимулом для превышения скорости. Опыт показывает, что вероятность получения штрафа за превышение скорости в каждой поездке туда и обратно равна 40%.

a) Какова вероятность того, что рабочий день закончится без получения штрафа?

b) Если штраф равен 80 долларам, то каково среднее значение дневного штрафа?

Вероятность быть оштрафованным в одной поездке равна $p = 0.4$. Следовательно, вероятность того, что рабочий день закончится без штрафа, равна

$$P\{x = 0\} = C_{10}^{0}(0.4)^{0}(0.6)^{10} = 0.006047.$$

Это значит, что шанс закончить рабочий день без штрафа меньше одного процента. Средний дневной штраф равен $80M\{x\} = 80(np) = 80 \times 10 \times 0.4 = 320$ (долларов).

Упражнения 12.5,а

1. Симметричная игральная кость бросается 10 раз. Какова вероятность того, что ни разу не выпадет четное число очков?
2. Пусть независимо бросаются пять симметричных монет. Какова вероятность того, что в точности одна из монет выпадет одной стороной, а остальные четыре — другой?
3. Гадалка утверждает, что по почерку она может предсказать, достигнет ли человек благосостояния на протяжении всей своей жизни. Для проверки этого попросили 10 миллионеров и 10 профессоров предоставить образцы их почерка. Затем эти образцы были представлены гадалке попарно — по одной подписи миллионера и профессора в каждой паре. Считается, что утверждение гадалки является правильным, если она сделала, по меньшей мере, восемь правильных предсказаний. Какова вероятность того, что утверждение гадалки будет "удачным"?
4. Докажите приведенные выше формулы для математического ожидания и дисперсии биномиального распределения.

12.5.2. Распределение Пуассона

Люди приходят в банк или магазин "совершенно случайным" образом. Это означает, что нет никакой возможности предсказать, когда и кто придет. Плотность распределения случайной величины, которая равна количеству таких посещений на протяжении определенного периода времени, описывается с помощью распределения Пуассона.

Пусть x — количество событий (например, посещений банка или магазина), которые происходят на протяжении единицы времени (например, минуты или часа). Плотность вероятности распределения Пуассона задается формулой

$$P\{x = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 1, 2, \ldots .$$

Математическое ожидание и дисперсия распределения Пуассона равны соответственно $M\{x\} = \lambda$ и $D\{x\} = \lambda$. Из интуитивных соображений формула $M\{x\} = \lambda$ должно означать среднее количество событий, происходящих за единицу времени. По существу, это так и есть: параметр λ определяет скорость, с которой происходят события (количество событий за единицу времени).

Распределение Пуассона широко используется в теории массового обслуживания (см. главу 17).

Пример 12.5–2

Заказы на ремонт небольших электродвигателей поступают в мастерскую случайным образом, примерно 10 заказов в день.

a) Каково среднее количество электродвигателей, которые поступают ежедневно в мастерскую?

b) Какова вероятность того, что на протяжении одного часа не поступит ни одного заказа, если мастерская открыта 8 часов в день?

Среднее количество работ, которые поступают ежедневно в мастерскую, равно $\lambda = 10$. Для вычисления вероятности того, что на протяжении одного часа не поступит ни одного заказа, необходимо подсчитать скорость поступления заказов в час, т.е. в среднем $\lambda_{час} = 10/8 = 1.25$ заказа в час. Следовательно,

$$P\{\text{нет заказов за 1 час}\} = \frac{(\lambda_{час})^0 e^{-\lambda_{час}}}{0!} = \frac{1.25^0 e^{-1.25}}{1} = 0.2865.$$

Упражнения 12.5,b

1. Клиенты прибывают в учреждение обслуживания в соответствии с распределением Пуассона со скоростью четыре клиента в минуту. Какова вероятность того, что по крайней мере один клиент прибудет в любой заданный 30-секундный интервал времени?

2. Распределение Пуассона с параметром λ аппроксимирует биномиальное распределение с параметрами n и p при условии, что n — достаточно большое положительное число, p — очень малое число, а λ примерно равно np (с математической точки зрения это означает, что $n \to \infty$, $p \to 0$ и $np \to \lambda$). Продемонстрируйте это на ситуации, когда известно, что изготовленная партия изделий содержит 1% брака. Вычислите вероятность того, что выборка объемом 10 изделий содержит не более одного бракованного изделия, использовав для этого сначала (точное) биномиальное распределение, а затем (приближенное) распределение Пуассона. Покажите, что такое приближение будет неприемлемым, если величина p будет увеличена, скажем, до 0.5.

3. Докажите приведенные выше формулы для математического ожидания и дисперсии распределения Пуассона.

12.5.3. Отрицательное экспоненциальное распределение

Если *число заявок*, поступивших в учреждение за определенный период времени, удовлетворяет распределению Пуассона, то распределение *интервалов времени* между последовательными поступлениями заявок должно следовать *отрицательному экспоненциальному* (или просто *экспоненциальному*) распределению. В частности, если λ есть скорость появления событий в распределении Пуассона, то распределение времени x между последовательными поступлениями определяется плотностью вероятности

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

На рис. 12.3 показан график функции $f(x)$.

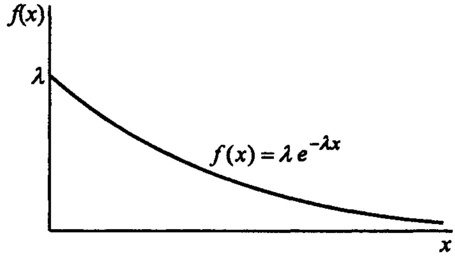


Рис. 12.3

Математическое ожидание и дисперсия экспоненциального распределения равны

$$M\{x\} = \frac{1}{\lambda}, \; D\{x\} = \frac{1}{\lambda^2}.$$

Математическое ожидание $M\{x\}$ согласуется с определением λ. Если λ — *скорость*, с которой происходят события, то $1/\lambda$ — средний интервал между последовательными событиями.

Пример 12.5–3

Автомобили прибывают на заправочную станцию случайно, в среднем каждые 2 минуты. Вычислите вероятность того, что интервал между последовательными прибытиями автомобилей не превысит 1 минуты.

Искомая вероятность имеет вид $P\{x \le A\}$, здесь $A = 1$ минута. Вычисление требуемой вероятности эквивалентно вычислению значения функции распределения случайной величины x.

$$P\{x \le A\} = \int_0^A \lambda e^{-\lambda x} dx = -e^{-\lambda x} \Big|_0^A = 1 - e^{-\lambda A}.$$

Вычисляем скорость прибытия автомобилей.

$$\lambda = \frac{1}{2} \text{ прибытия в минуту.}$$

Следовательно,

$$P\{x \le 1\} = 1 - e^{-\frac{1}{2}(1)} = 0.39.$$

Упражнения 12.5,с

1. Магазин посещают жители как городской местности, так и сельской. Городские покупатели прибывают со скоростью 5 посетителей в минуту, а сельские — 7 посетителей в минуту. Прибытия покупателей являются случайными событиями. Определите вероятность того, что время между последовательными прибытиями посетителей будет меньше 5 секунд.

2. Докажите приведенные выше формулы для математического ожидания и дисперсии экспоненциального распределения.

12.5.4. Нормальное распределение

Нормальное распределение описывает многие случайные явления, которые происходят в каждодневной жизни, включая анализ счетов, распределение веса и роста людей и многое другое. Плотность вероятности нормального распределения задается формулой

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad -\infty<x<\infty,$$

где $M\{x\} = \mu$, $D\{x\} = \sigma^2$. Нормальное распределение с математическим ожиданием μ и стандартным отклонением σ обозначается как $N(\mu, \sigma)$.

На рис. 12.4 показан график плотности $f(x)$ нормального распределения. Как видим, плотность распределения является симметричной функцией относительно математического ожидания μ.

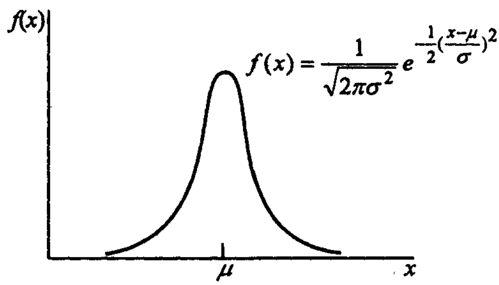


Рис. 12.4

Важность нормального распределения объясняется также тем, что распределение среднего достаточно большой выборки, имеющей любое распределение, можно асимптотически аппроксимировать нормальным распределением. Это следует из представленной ниже теоремы.

Центральная предельная теорема. *Пусть* $x_1, x_2, \ldots, x_n$ *— независимые, одинаково распределенные случайные величины с математическим ожиданием* μ *и стандартным отклонением* σ *каждая. Определим сумму*

$$s_n = x_1 + x_2 + \ldots + x_n.$$

При возрастании n ($n \to \infty$) распределение случайной величины s_n становится асимптотически нормальным с математическим ожиданием $n\mu$ и дисперсией $n\sigma^2$ независимо от начального распределения величин $x_1, x_2, ..., x_n$.

Центральная предельная теорема свидетельствует, в частности, о том, что среднее выборки объемом n, имеющей любое распределение, асимптотически является нормальным с математическим ожиданием μ и дисперсией σ^2/n.

Функцию нормального распределения трудно представить в виде формулы, пригодной для практических расчетов. В связи с этим составлены специальные таблицы функции нормального распределения (см. табл. 1 в Приложении Д). Эти таблицы созданы для **стандартного нормального распределения** с нулевым математическим ожиданием и дисперсией, равной единице. Любую нормально распределенную случайную величину x с математическим ожиданием μ и дисперсией σ^2 можно привести к стандартному виду путем замены

$$z = \frac{x-\mu}{\sigma}.$$

Отметим, что около 99.98% площади под кривой плотности нормального распределения находится в интервале $\mu - 3\sigma \le x \le \mu + 3\sigma$. Этот факт известен под названием "правило трех сигм".

Пример 12.5–4

Внутренний диаметр цилиндра должен иметь размер (спецификацию) 1±0.3 дюйма. Процесс механической обработки таких деталей подчиняется нормальному распределению с математическим ожиданием 1 и стандартным отклонением 0.1 дюйма. Требуется определить процент продукции, удовлетворяющей требованиям спецификации.

Пусть случайная величина x равна диаметру цилиндра. Вероятность того, что цилиндр будет удовлетворять требованиям спецификации, равна

$$P\{1 - 0.03 \le x \le 1 + 0.03\} = P\{0.97 \le x \le 1.03\}.$$

При $\mu = 1$ и $\sigma = 0.1$ эту вероятность можно выразить через стандартное нормальное распределение следующим способом.

$$P\{0.97 \le x \le 1.03\} = P\left\{\frac{0.97-1}{0.1} \le z \le \frac{1.03-1}{0.1}\right\} = P\{-0.3 \le z \le 0.3\} =$$
$$= P\{z \le 0.3\} - P\{z \le -0.3\} = P\{z \le 0.3\} - [1 - P\{z \le 0.3\}] =$$
$$= 2P\{z \le 0.3\} - 1 = 2 \times 0.6179 - 1 = 0.2358.$$

На рис. 12.5 заштрихованная область соответствует искомой вероятности. Заметим, что равенство $P\{z \le -0.3\} = 1 - P\{z \le 0.3\}$ имеет место в силу симметрии функции плотности вероятностей. Величина 0.6179 ($= P\{z \le 0.3\}$) взята из таблицы для нормального распределения (табл. 1 Приложения Д).

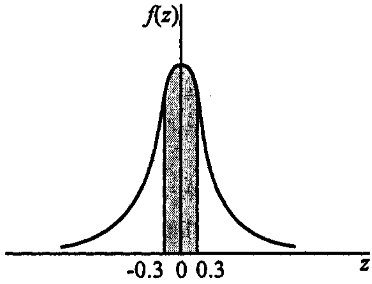


Рис. 12.5

Упражнения 12.5,d

1. Инженерный колледж американского университета набирает студентов из числа выпускников средней школы, которые по стандартному тесту ACT для поступающих в колледжи имеют не менее 26 баллов. Результаты тестирования выпускников являются нормально распределенной случайной величиной с математическим ожиданием 22 балла и стандартным отклонением 2 балла.
   a) Определите процент выпускников средней школы, которые являются потенциальными студентами инженерного колледжа.
   b) Определите процент выпускников школы, которые не будут приняты в инженерный колледж, если университет не примет ни одного из них с количеством баллов, меньшим 17.
2. Вес людей, которые хотят совершить прогулку на вертолете в парке аттракционов, является случайной величиной с математическим ожиданием 180 фунтов и стандартным отклонением 15 фунтов. Вместимость вертолета составляет пять человек, максимальная грузоподъемность — 1000 фунтов. Какова вероятность того, что вертолет не взлетит с пятью пассажирами на борту? (*Совет.* Используйте центральную предельную теорему.)
3. Внутренний диаметр цилиндра является нормально распределенной случайной величиной с математическим ожиданием 1 дюйм и стандартным отклонением 0.01 дюйма. При сборке твердый стержень вставляется внутрь каждого цилиндра. Диаметр стержня является нормально распределенной случайной величиной с математическим ожиданием 0,99 и стандартным отклонением 0.01 дюйма. Требуется определить процент пар цилиндр–стержень, которые не подойдут для сборки. (*Подсказка.* Разность двух нормально распределенных случайных величин также является нормально распределенной величиной.)

12.6. Эмпирические распределения

В предыдущих разделах мы рассмотрели свойства плотностей вероятностей, функции распределения случайных величин и привели примеры четырех типов распределений. Как определяются эти распределения на практике?

Определение (или, точнее, оценка) любой плотности вероятности случайной величины содержится в необработанной информации, которая собирается в процессе изучения исследуемой ситуации. Например, для оценки плотности вероятности времени между приходом покупателей в бакалейную лавку, сначала фиксируется время их прихода. Искомое время между приходом покупателей находится путем вычитания времен последовательных их приходов.

В этом разделе мы рассмотрим, как собранные данные (именуемые *статистическим рядом* или *выборкой*) можно преобразовать в плотность вероятности случайных величин с помощью следующих шагов.

Шаг 1. Отображаем данные в виде подходящей частотной гистограммы и подбираем эмпирическую функцию плотности вероятности.

Шаг 2. Используем критерий согласия, чтобы проверить, совпадает ли полученная эмпирическая функция плотности вероятности с одной из известных плотностей вероятностей.

Рассмотрим детали этой процедуры.

Гистограмма частот. Данная гистограмма строится на основе статистического ряда путем деления области изменения исходных данных (от минимального до максимального значения) на непересекающиеся интервалы. При заданных границах (I_{i-1}, I_i) интервала i соответствующая частота определяется как число наблюденных значений x, которые удовлетворяют неравенству $I_{i-1} < x \le I_i$.

Пример 12.6–1

Данные, приведенные в следующей таблице, представляют время обслуживания (в минутах) 60 посетителей в некотором центре услуг.

0.7	0.4	3.4	4.8	2.0	1.0	5.5	6.2	1.2	4.4
1.5	2.4	3.4	6.4	3.7	4.8	2.5	5.5	0.3	8.7
2.7	0.4	2.2	2.4	0.5	1.7	9.3	8.0	4.7	5.9
0.7	1.6	5.2	0.6	0.9	3.9	3.3	0.2	0.2	4.9
9.6	1.9	9.1	1.3	10.6	3.0	0.3	2.9	2.9	4.8
8.7	2.4	7.2	1.5	7.9	11.7	6.3	3.8	6.9	5.3

Минимальное и максимальное значения приведенных данных соответственно равны 0.2 и 11.7. Поэтому выбираем двенадцать интервалов длиной в 1 минуту (полный интервал изменений равен [0, 12]). Надлежащий выбор размера интервала является решающим фактором в определении формы эмпирического распределения. Хотя не существует жестких правил выбора оптимального размера интервала, общим правилом, которого следует придерживаться, является выбор от 10 до 20 интервалов. На практике было бы неплохо попробовать различные размеры интервала для построения подходящей гистограммы (в этом отношении будут полезными интерактивные возможности программы TORA).

Приведенная ниже таблица суммирует информацию для рассматриваемого статистического ряда, необходимую для построения гистограммы. Столбец относительной частоты f_i вычисляется путем деления соответствующих значений столбца наблюденной частоты o_i на общий объем наблюдений ($n = 60$). Например $f_1 = 11/60 = 0.1833$. Значения F_i в столбце накопленных частот вычисляются посредством последовательного суммирования величин f_i. Так, $F_1 = f_1 = 0.1833$ и $F_2 = F_1 + f_2 = 0.1833 + 0.1333 = 0.3166$.

Интервал	Подсчет наблюдений	Наблюденная частота, o_i	Относительная частота, f_i	Накопленная относительная частота, F_i
(0, 1)	₩ ₩ I	11	0.1833	0.1833
(1, 2)	₩ III	8	0.1333	0.3166
(2, 3)	₩ IIII	9	0.1500	0.4666
(3, 4)	₩ II	7	0.1167	0.5833
(4, 5)	₩ I	6	0.1000	0.6833
(5, 6)	₩	5	0.0833	0.7666
(6, 7)	IIII	4	0.0667	0.8333
(7, 8)	II	2	0.0333	0.8666
(8, 9)	III	3	0.0500	0.9166
(9, 10)	III	3	0.0500	0.9666
(10, 11)	I	1	0.0167	0.9833
(11, 12)	I	1	0.0167	1.0000
Всего		60	1.0000	

Величины f_i и F_i являются дискретными эквивалентами плотности вероятности и функции распределения времени обслуживания t. Так как гистограмма частот дает дискретную версию непрерывного времени обслуживания, можно преобразовать дискретную функцию распределения в непрерывную кусочно-линейную функцию, соединяя полученные точки отрезками прямых. На рис. 12.6 представлена эмпирическая плотность вероятности и функция распределения для рассматриваемого примера. Здесь функция распределения определена в средних точках соответствующего интервала значений.

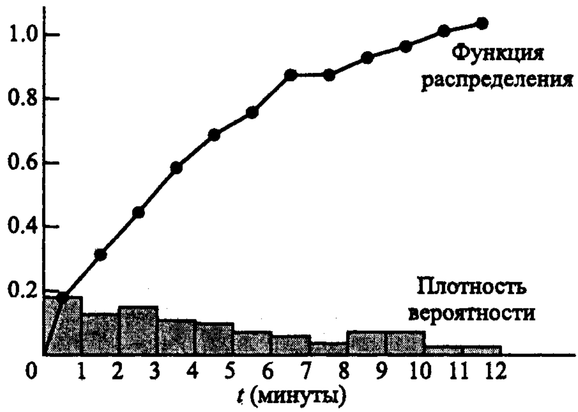


Рис. 12.6

Теперь можно оценить математическое ожидание $\bar{t}$ и дисперсию s_t^2 эмпирического распределения. Пусть N — число интервалов в гистограмме; обозначим через $\bar{t}_i$ среднюю точку интервала i. Тогда

$$\bar{t} = \sum_{i=1}^{N} f_i \bar{t}_i,$$

$$s_t^2 = \sum_{i=1}^{N} f_i \left(\bar{t}_i - \bar{t}\right)^2.$$

Применяя эти формулы для рассматриваемого примера, получаем следующее.

$\bar{t} = 0.1833 \times 0.5 + 0.133 \times 1.5 + ... + 0.0167 \times 11.5 = 3.934$ минуты,

$s_t^2 = 0.1833 \times (0.5 - 3.934)^2 + 0.133 \times (1.5 - 3.934)^2 + ... + 0.0167 \times (11.5 - 3.934)^2 = 8.646$ минут^2.

Критерий согласия. С помощью этого критерия можно проверить, является ли выборка, на основе которой определено эмпирическое распределение, представителем конкретного вероятностного распределения. Начальную оценку можно сделать путем сравнения значений эмпирической функции распределения и предполагаемой теоретической функции распределения. Если значения этих функций чрезмерно не отличаются друг от друга, то, вероятно, рассматриваемая выборка получена из предложенного теоретического распределения. Это начальное "предчувствие" может быть в дальнейшем подтверждено посредством применения критерия согласия.

Пример 12.6–2

Проверим данные из примера 12.6–1 на принадлежность предполагаемому экспоненциальному распределению.

Первой задачей является уточнение параметров плотности вероятности и функции распределения, которые определяют теоретическое распределение. Из примера 12.6–1 следует, что $\bar{t} = 3.934$ минуты, поэтому $\lambda = 1/3.934 = 0.2542$ для предполагаемого экспоненциального распределения (см. раздел 12.5.3). Соответствующая плотность вероятности и функция распределения имеют следующий вид.

$$f(t) = 0.2542e^{-0.2542t}, \quad t > 0,$$

$$F(T) = \int_0^T f(t)dt = 1 - e^{-0.2542T}, \quad T > 0.$$

Используем теперь функцию распределения $F(T)$ для вычисления ее значений в точках $T = 0.5, 1.5, ..., 11.5$ и сравнения их с эмпирическими значениями F_i, $i = 1, 2, ..., 12$, которые вычислены в примере 12.6–1. Например,

$$F(0.5) = 1 - e^{-(0.2542 \times 0.5)} \approx 0.12.$$

На рис. 12.7 представлены результаты сравнения. Просмотрев два графика, можем сделать вывод, что экспоненциальное распределение действительно приемлемо для аппроксимации распределения имеющихся данных.

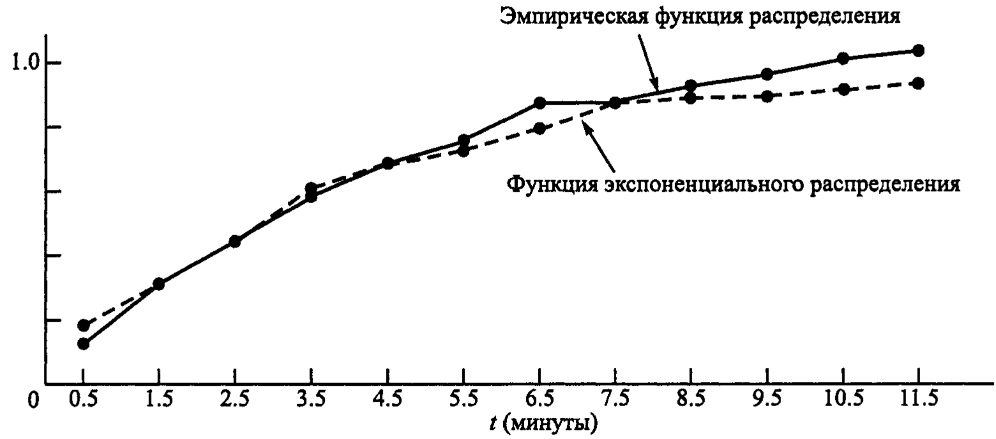


Рис. 12.7

Следующий шаг состоит в применении критерия согласия. Имеется два таких критерия: 1) критерий Колмогорова–Смирнова и 2) критерий χ^2 (критерий хи-квадрат). Здесь мы ограничимся обсуждением критерия χ^2.

Критерий χ^2 основан на измерении отклонений между эмпирическими и теоретическими частотами, соответствующими различным интервалам построенной гистограммы. В частности, теоретическая частота n_i, соответствующая наблюденной частоте o_i интервала i, вычисляется по формуле

$$n_i = n\int_{I_{i-1}}^{I_i} f(t)dt = n\left(F(I_i) - F(I_{i-1})\right) = 60\left(e^{-0.2542 I_{i-1}} - e^{-0.2542 I_i}\right).$$

При заданных o_i и n_i для каждого интервала i гистограммы мера отклонения между эмпирическими и теоретическими частотами определяется следующей формулой.

$$\chi^2 = \sum_{i=1}^{N} \frac{(o_i - n_i)^2}{n_i}.$$

Когда количество интервалов $N \to \infty$, величина χ^2 асимптотически стремится к плотности вероятности χ^2–распределения с $N-k-1$ степенями свободы, где k — число параметров, оцененных на основе исходной информации и использованных для определения теоретического распределения.

Нулевая гипотеза, утверждающая, что наблюденная выборка получена из теоретического распределения $f(t)$, принимается, если $\chi^2 < \chi^2_{N-k-1,1-\alpha}$, где $\chi^2_{N-k-1,1-\alpha}$ — значение χ^2 при $N-k-1$ степенях свободы, α — уровень значимости критерия.

Вычисления в соответствии с критерием показаны в следующей таблице.

Интервал	Наблюденная частота, o_i	Теоретическая частота, n_i	$\frac{(o_i - n_i)^2}{n_i}$
(0, 1)	11	13.47	0.435
(1, 2)	8	10.44	0.570
(2, 3)	9	8.10	0.100
(3, 4)	7	6.28	0.083
(4,5)	6	4.87	
(5,6)	5	3.88	
(6,7)	4	2.93	
(7,8)	2	2.27	
(8,9)	3 } 25	1.76 } 21.71	0.499
(9,10)	3	1.37	
(10,11)	1	1.06	
(11,12)	1	0.82	
(12,∞)	1	2.75	
Всего	$n = 60$	$n = 60$	Величина χ^2 равна 1.705

Существует практическое правило: ожидаемое значение теоретической частоты для любого интервала должно быть не менее 5. Это правило всегда можно выполнить путем объединения последовательных интервалов. В приведенной таблице правило требует формирования единого интервала (4, ∞). Количество интервалов становится равным $N = 5$. Поскольку на основе исходных данных оценивается только один параметр (а именно λ), степень свободы величины χ^2 должна равняться 5 – 1 – 1 = 3. Если выберем уровень значимости $\alpha = 0.05$, таблица значений величины χ^2 (табл. 3 из Приложения Д) дает критическое значение $\chi^2_{3;0.05} = 7.815$. Так как значение величины χ^2 (= 1.705) меньше критического, мы принимаем гипотезу, что выборка получена из экспоненциального распределения.

Упражнения 12.6,а

1. Следующие данные представляют время (в минутах) между прибытием клиентов в некий центр обслуживания.

4.3	3.4	0.9	0.7	5.8	3.4	2.7	7.8
4.4	0.8	4.4	1.9	3.4	3.1	5.1	1.4
0.1	4.1	4.9	4.8	15.9	6.7	2.1	2.3
2.5	3.3	3.8	6.1	2.8	5.9	2.1	2.8
3.4	3.1	0.4	2.7	0.9	2.9	4.5	3.8
6.1	3.4	1.1	4.2	2.9	4.6	7.2	5.1
2.6	0.9	4.9	2.4	4.1	5.1	11.5	2.6

0.1	10.3	4.3	5.1	4.3	1.1	4.1	6.7
2.2	2.9	5.2	8.2	1.1	3.3	2.1	7.3
3.5	3.1	7.9	0.9	5.1	6.2	5.8	1.4
0.5	4.5	6.4	1.2	2.1	10.7	3.2	2.3
3.3	3.3	7.1	6.9	3.1	1.6	2.1	1.9

a) Постройте три гистограммы с длиной интервалов 0.5, 1 и 1.5 минуты соответственно (для удобства воспользуйтесь программой TORA).

b) Сравните графически эмпирическую функцию распределения с аналогичной функцией экспоненциального распределения.

c) Проверьте гипотезу о том, что заданная выборка взята из экспоненциального распределения. Используйте 95%-ный доверительный уровень (т.е. 5%-ный уровень значимости).

d) Какая из трех гистограмм является "наилучшей" для проверки нулевой гипотезы о том, что выборочные значения подчиняются экспоненциальному закону?

2. Следующие данные представляют время (в секундах), необходимое для передачи сообщения.

25.8	67.3	35.2	36.4	58.7
47.9	94.8	61.3	59.3	93.4
17.8	34.7	56.4	22.1	48.1
48.2	35.8	65.3	30.1	72.5
5.8	70.9	88.9	76.4	17.3
77.4	66.1	23.9	23.8	36.8
5.6	36.4	93.5	36.4	76.7
89.3	39.2	78.7	51.9	63.6
89.5	58.6	12.8	28.6	82.7
38.7	71.3	21.1	35.9	29.2

При 95%-ном доверительном уровне проверьте гипотезу о том, что данная выборка имеет равномерное распределение, при этом используйте следующую дополнительную информацию о теоретическом равномерном распределении.

a) Распределение сосредоточено на интервале от 0 до 100.

b) Интервал, на котором сосредоточено распределение, определяется из данных выборки.

c) Верхний предел интервала, на котором сосредоточено распределение, равен 100, а нижний должен быть определен из данных выборки.

3. Автоматический прибор используется для определения интенсивности движения на оживленном перекрестке. Прибор фиксирует время прибытия автомобиля на перекресток по непрерывной временной шкале, начиная с нуля. Приведенная ниже таблица содержит (накопленное) время (в минутах) прибытия на перекресток первых 60 автомобилей. Постройте подходящую гистограмму для проверки гипотезы о том, что данные выборки имеют экспоненциальное распределение. Используйте 95%-ный доверительный уровень.

Прибытие	Время прибытия	Прибытие	Время прибытия	Прибытие	Время прибытия
1	5.2	21	97.2	41	180.1
2	6.7	22	97.9	42	188.8
3	9.1	23	111.5	43	201.2
4	12.5	24	116.7	44	218.4
5	18.9	25	117.3	45	219.9
6	22.6	26	118.2	46	227.8
7	27.4	27	124.1	47	233.5
8	29.9	28	127.4	38	239.8
9	35.4	29	127.6	49	243.6
10	35.7	30	127.8	50	250.5
11	44.4	31	132.7	51	255.8
12	47.1	32	142.3	52	256.5
13	47.5	33	145.2	53	256.9
14	49.7	34	154.3	54	270.3
15	67.1	35	155.6	55	275.1
16	67.6	36	166.2	56	277.1
17	69.3	37	169.2	57	278.1
18	78.6	38	169.5	58	283.6
19	86.6	39	172.4	59	299.8
20	91.3	40	175.3	60	300.0

12.7. Заключение

Эта глава дает краткий обзор основ теории вероятностей. Здесь также приведены основные характеристики наиболее распространенных дискретных и непрерывных вероятностных распределений. Заканчивается глава рассмотрением эмпирических распределений.

Литература

Feller W. *An Introduction to Probability Theory and Its Applications,* 2nd ed., Vols. 1 and 2, Wiley, New York, 1967. (Имеется русский перевод первого издания: Феллер В. *Введение в теорию вероятностей и ее приложения*: В 2-х томах. — М.: Мир, 1967.)

Papoulis A. *Probability and Statistics,* Prentice Hall, Upper Saddle River, N.J., 1990.

Parzen E. *Modern Probability Theory and Its Applications,* Wiley, New York, 1960.

Ross S. *Introduction to Probability Models,* 5th ed., Academic Press, New York, 1993.

Литература, добавленная при переводе

Бендат Дж., Пирсол А. *Прикладной анализ случайных данных.* — М.: Мир, 1989.

Пугачев В.С. *Теория вероятностей и математическая статистика.* — М.: Наука, 1979.

Чистяков В.П. *Курс теории вероятностей.* — М.: Высш. школа, 1982.

Методы прогнозирования

13.1. Введение

Принимая решения, мы определяем планы на будущее. Следовательно, используемые при этом данные должны соответствовать будущим событиям. Например, в теории управления запасами мы обосновываем наши решения посредством спроса на определенные виды продукции в течение определенного планового периода. Аналогично в финансовом планировании необходимо предсказать структуру денежного потока в будущем на основе структуры текущих денежных потоков.

В этой главе рассматриваются три методики прогнозирования изменений интересующих нас переменных как функций времени.

1. Прогнозирование с использованием скользящего среднего.
2. Прогнозирование путем экспоненциального сглаживания.
3. Регрессионное прогнозирование.

В этой главе используются такие основные обозначения:

y_t — действительное (или наблюденное) значение случайной величины y в момент времени t,

y_t^* — расчетное значение (оценка) случайной величины y в момент времени t,

ε_t — случайный компонент (или шум) в момент времени t.

13.2. Прогнозирование с использованием скользящего среднего

При использовании этой методики основное предположение состоит в том, что временной ряд является устойчивым в том смысле, что его члены есть реализациями следующего случайного процесса:

$$y_t = b + \varepsilon_t,$$

где b — неизвестный постоянный параметр, который оценивается на основе представленной информации. Предполагается, что случайная ошибка ε_t имеет нулевое математическое ожидание и постоянную дисперсию. Кроме того, предполагается, что данные для различных периодов времени не коррелированы.

Метод с использованием скользящего среднего предполагает, что последние n наблюдений являются равнозначно важными для оценки параметра b. Другими словами, если в текущий момент времени t последними n наблюдениями есть $y_{t-n+1}, y_{t-n+2},\ldots, y_t$, тогда оцениваемое значение для момента $t + 1$ вычисляется по формуле

$$y^*_{t+1} = \frac{y_{t-n+1} + y_{t-n+2} + \ldots + y_t}{n}.$$

Не существует четкого правила для выбора числа n — базы метода, использующего скользящее среднее. Если есть весомые основания полагать, что наблюдения в течение достаточно длительного времени удовлетворяют модели $y_t = b + \varepsilon_t$, то рекомендуется выбирать большие значения n. Если же наблюдаемые значения удовлетворяют приведенной модели в течение коротких периодов времени, может быть приемлемым и малое значение n. На практике величина n обычно принимается в пределах от 2 до 10.

Пример 13.2–1

В табл. 13.1 представлены объемы спроса на некое изделие за прошедшие 24 месяца. Необходимо с помощью методики скользящего среднего дать прогноз объема спроса на следующий месяц (здесь $t = 25$).

Таблица 13.1

Месяц t	Спрос y_t	Месяц t	Спрос y_t
1	46	13	54
2	56	14	42
3	54	15	64
4	43	16	60
5	57	17	70
6	56	18	66
7	67	19	57
8	62	20	55
9	50	21	52
10	56	22	62
11	47	23	70
12	56	24	72

Чтобы проверить применимость метода скользящего среднего, проанализируем приведенные данные. На рис. 13.1 нанесены значения временного ряда y_t. График показывает, что наблюдается тенденция к возрастанию значений y_t с течением времени. Это, вообще-то, означает, что скользящее среднее не будет хорошим предсказателем для будущего спроса. В частности использование большой базы n для скользящего среднего неприемлемо в этом случае, так как это приведет к подавлению наблюдаемой тенденции в изменении данных. Следовательно, если мы используем небольшое значение для базы n, то будем находиться в лучшем положении с точки зрения отображения упомянутой тенденции в изменении данных.

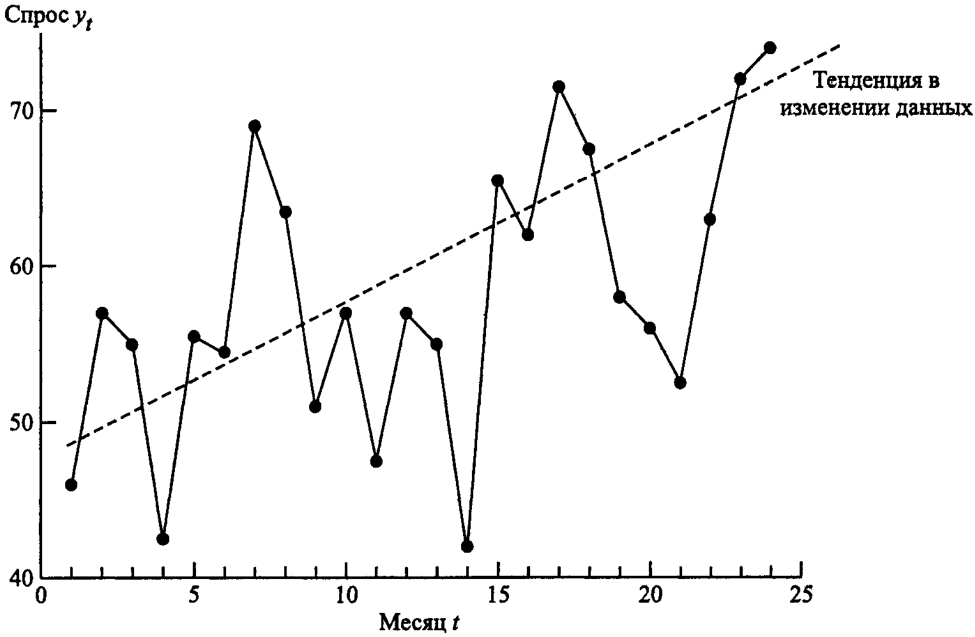


Рис. 13.1

Если мы используем значение $n = 3$ в качестве базы скользящего среднего, то оценка спроса на следующий месяц ($t = 25$) будет равна средней величине спроса за 22, 23 и 24 месяцы:

$$y_{25}^{*} = \frac{62+70+72}{3} = 68 \text{ единиц.}$$

Оценка величины спроса в 68 единиц для 25 месяца будет использоваться также при прогнозе спроса для $t = 26$:

$$y_{26}^{*} = \frac{70+72+68}{3} = 70 \text{ единиц.}$$

Когда значение *реального* спроса в 25 месяце будет известно, его следует использовать для вычисления новой оценки объема спроса для 26 месяца в виде средней величины спроса 23, 24 и 25 месяцев.

Упражнения 13.2,а

1. В примере 13.2–1 оцените объем спроса для $t = 25$, используя n=12 в качестве базы скользящего среднего. Какой эффект имеет большее значение n с точки зрения подавления тенденции изменения данных?

2. Число кондиционеров, проданных за последние 24 месяца, приведено в табл. 13.2. Проанализируйте эти данные с точки зрения применимости метода скользящего среднего.

Таблица 13.2

Месяц	Продажа	Месяц	Продажа
1	25	13	40
2	15	14	35
3	30	15	50
4	38	16	60
5	58	17	66
6	62	18	90
7	85	19	105
8	88	20	85
9	60	21	60
10	40	22	55
11	40	23	50
12	38	24	45

3. В табл. 13.3 содержатся данные за десятилетний период о количестве людей, посетивших туристическую зону на автомобиле и воздушном транспорте. Проанализируйте эти данные с точки зрения применимости метода скользящего среднего.

Таблица 13.3

Год	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Автомобиль	1042	1182	1224	1338	1455	1613	1644	1699	1790	1885
Самолет	500	522	540	612	715	790	840	900	935	980

4. В табл. 13.4 представлены данные об объемах продажи универмага (в миллионах долларов). Проанализируйте эти данные с точки зрения применимости метода скользящего среднего.

Таблица 13.4

Год	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
Продажа	21.0	23.2	23.2	24.0	24.9	25.6	26.6	27.4	28.5	29.6

5. Университет предлагает курсы лекций (вне своей территории) в пяти населенных пунктах штата. В табл. 13.5 приведены данные о числе слушателей курсов на протяжении шести лет. Данные относительно каждого года разбиты по семестрам: осень (1), весна (2) и лето (3). Необходимо использовать эти данные для оценки числа слушателей в следующем году. Проанализируйте приведенные данные с точки зрения применимости метода скользящего среднего.

Таблица 13.5

Семестр		Населенный пункт, где проводятся курсы лекций				
		1	2	3	4	5
1989						
	1	288	136	48	165	59
	2	247	150	49	168	46
	3	117	69	14	61	15
1990						
	1	227	108	41	108	28
	2	239	106	46	128	43
	3	101	50	15	54	16
1991						
	1	240	126	31	104	46
	2	261	134	19	83	38
	3	138	48	9	56	32
1992						
	1	269	149	17	90	51
	2	301	113	25	54	28
	3	119	50	14	17	6
1993						
	1	226	102	22	16	30
	2	241	110	16	0	24
	3	125	46	7	0	12
1994						
	1	231	88	2	0	1
	2	259	66	3	0	27
	3	102	23	0	0	0

13.3. Экспоненциальное сглаживание

Прогнозирование путем экспоненциального сглаживания (метод экспоненциального сглаживания) предполагает, что вероятностный процесс определяется моделью $y_t = b + \varepsilon_t$; это предположение использовалось и при рассмотрении метода скользящего среднего. Метод экспоненциального сглаживания разработан для того, чтобы устранить недостаток метода скользящего среднего, который состоит в том, что все данные, используемые при вычислении среднего, имеют одинаковый вес. В частности, метод экспоненциального сглаживания приписывает больший весовой коэффициент самому последнему наблюдению.

Определим величину α ($0 < \alpha < 1$) как **константу сглаживания**, и пусть известны значения временного ряда для прошедших t моментов времени $y_1, y_2, ..., y_t$. Тогда оценка y^*_{t+1} для момента времени $t + 1$ вычисляется по формуле

$$y^*_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots .$$

Коэффициенты при y_t, y_{t-1}, y_{t-2}, ... постепенно уменьшаются, тем самым эта процедура приписывает больший вес последним (по времени) данным.

Формулу для вычисления y^*_{t+1} можно привести к следующему (более простому) виду:

$$y^*_{t+1} = \alpha y_t + (1-\alpha)\left\{\alpha y_{t-1} + \alpha(1-\alpha) y_{t-2} + \alpha(1-\alpha)^2 y_{t-3} + \ldots\right\} = \alpha y_t + (1-\alpha) y^*_t.$$

Таким образом, значение y^*_{t+1} можно вычислить рекуррентно на основании значения y^*_t. Вычисления в соответствии с этим рекуррентным уравнением начинаются с того, что пропускается оценка y^*_1 для $t = 1$ и в качестве оценки для $t = 2$ принимается наблюденная величина для $t = 1$, т.е. $y^*_2 = y_1$. В действительности же для начала можно использовать любую разумную процедуру. Например, часто в качестве оценки y^*_0 берется усредненное значение y_t по "приемлемому" числу периодов в начале временного ряда.

Выбор константы сглаживания α является решающим моментом при вычислении значения прогнозируемой величины. Большее значение α приписывает больший вес последним наблюдениям. На практике значение α берут в пределах от 0.01 до 0.30.

Пример 13.3–1

Применим метод экспоненциального сглаживания к данным из примера 13.2–1 при $\alpha = 0.1$.

В табл. 13.6 содержатся результаты вычислений. При вычислениях пропускается y^*_1 и принимается, что $y^*_2 = y_1 = 46$ единиц.

ТАБЛИЦА 13.6

i	y_i	y^*_i	i	y_i	y^*_i
1	46	—	13	54	0.1×56 + 0.9×51.63 = 52.07
2	56	46	14	42	0.1×54 + 0.9×52.07 = 52.26
3	54	0.1×56 + 0.9×46 = 47	15	64	0.1×42 + 0.9×52.26 = 51.23
4	43	0.1×54 + 0.9×47 = 47.7	16	60	0.1×64 + 0.9×51.23 = 52.50
5	57	0.1×43 + 0.9×47.7 = 47.23	17	70	0.1×60 + 0.9×52.50 = 53.26
6	56	0.1×57 + 0.9×47.23 = 48.21	18	66	0.1×70 + 0.9×53.26 = 54.93
7	67	0.1×56 + 0.9×48.21 = 48.98	19	57	0.1×66 + 0.9×54.93 = 56.04
8	62	0.1×67 + 0.9×48.98 = 50.79	20	55	0.1×57 + 0.9×56.04 = 56.14
9	50	0.1×62 + 0.9×50.79 = 51.91	21	52	0.1×55 + 0.9×56.14 = 56.02
10	56	0.1×50 + 0.9×51.91 = 51.72	22	62	0.1×52 + 0.9×56.02 = 55.62
11	47	0.1×56 + 0.9×51.72 = 52.15	23	70	0.1×62 + 0.9×55.62 = 56.26
12	56	0.1×47 + 0.9×52.15 = 51.63	24	72	0.1×70 + 0.9×56.26 = 57.63

Из приведенных данных следует, что оценка для $t = 25$ равна

$$y^*_{25} = \alpha y_{24} + (1-\alpha) y^*_{24} = 0.1(72) + 0.9(57.63) = 59.07 \text{ единиц.}$$

Эта оценка значительно отличается от полученной с помощью метода скользящего среднего (68 единиц). Большее значение для α даст оценку, более близкую к оценке метода скользящего среднего.

Упражнения 13.3,а

1. Примените метод экспоненциального сглаживания для данных из упр. 13.2,а(2) при $\alpha = 0.2$.
2. Примените метод экспоненциального сглаживания для данных из упр. 13.2,а(3) при $\alpha = 0.2$.
3. Примените метод экспоненциального сглаживания для данных из упр. 13.2,а(4) при $\alpha = 0.2$.
4. Примените метод экспоненциального сглаживания для данных из упр. 13.2,а(5) при $\alpha = 0.2$.

13.4. Регрессионный анализ

Регрессионный анализ определяет связь между *зависимой* переменной (например, спросом на продукцию) и *независимой* переменной (например, временем). Часто применяемая формула регрессии, описывающая зависимость между переменной y и независимой переменной x, имеет вид

$$y = b_0 + b_1x + b_2x^2 + \ldots + b_nx^n + \varepsilon,$$

где $b_0, b_1, \ldots, b_n$ — неизвестные параметры. Случайная ошибка ε имеет нулевое математическое ожидание и постоянную дисперсию (т.е. дисперсия случайной величины ε одинакова для всех наблюдаемых значений y).

Самая простая регрессионная модель предполагает, что зависимая переменная линейна относительно независимой переменной, т.е.

$$y^* = .\ bx.$$

Константы a и b определяются из временного ряда с использованием **метода наименьших квадратов**, в соответствии с которым находятся значения этих констант, доставляющих минимум сумме квадратов разностей между наблюденными и вычисленными величинами. Пусть (y_i, x_i) представляет i-ю точку исходных данных временного ряда, $i = 1, 2, \ldots, n$. Определим сумму квадратов отклонений между наблюденными и вычисленными величинами.

$$S = \sum_{i=1}^{n} \left(y_i - a - bx_i\right)^2 .$$

Значения коэффициентов a и b определяются из соответствующих условий минимума функции S, которые представимы в виде следующих уравнений.

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n} \left(y_i - a - bx_i\right) = 0,$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{n} \left(y_i - a - bx_i\right)x_i = 0.$$

После алгебраических преобразований получаем следующее решение данных уравнений.

$$b = \frac{\sum_{i=1}^{n} y_i x_i - n\overline{y}\,\overline{x}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2},$$

$$a = \overline{y} - b\overline{x},$$

где $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$.

Приведенные соотношения показывают, что сначала необходимо вычислить b, а затем величину коэффициента a.

Вычисленные значения a и b имеют силу при любом вероятностном распределении случайных величин y_i. Однако если y_i является нормально распределенной случайной величиной с постоянным стандартным отклонением, можно установить доверительный интервал для среднего значения оценки при $x = x^0$ (т.е. для $y^0 = a + bx^0$) в виде интервала

$$\left(a + bx^0\right) \pm t_{\alpha/2,\,n-2}\sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{n-2}}\sqrt{\frac{1}{n} + \frac{\left(x^0 - \overline{x}\right)^2}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}}.$$

Выражение $(y_i - y_i^*)$ представляет собой отклонение i-го наблюденного значения зависимой переменной от его соответствующей оценки.

Мы заинтересованы в установлении для прогнозируемых значений зависимой переменной y соответствующих им **интервалов предсказания** (скорее чем доверительного интервала для среднего значения оценки). Как и следовало ожидать, интервал предсказания для значения прогнозируемой величины является более широким, чем доверительный интервал для среднего значения оценки. Действительно, формула для интервала предсказания такая же, как и для доверительного интервала, но с той лишь разницей, что член $1/n$ под вторым квадратным корнем заменен на $(n + 1)/n$.

Чтобы проверить, насколько линейная модель $y^* = a + bx$ соответствует исходным данным, необходимо вычислить *коэффициент корреляции* r согласно формуле:

$$r = \frac{\sum_{i=1}^{n} y_i x_i - n\overline{y}\,\overline{x}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)\left(\sum_{i=1}^{n} y_i^2 - n\overline{y}^2\right)}},$$

где $-1 \le r \le 1$.

Если $r = \pm 1$, тогда линейная модель идеально подходит для описания зависимости между y и x. В общем случае, чем ближе $|r|$ к 1, тем лучше подходит линейная модель. Если же $r = 0$, величины y и x могут быть независимыми. В действительности равенство $r = 0$ является лишь необходимым, но не достаточным условием независимости, так как возможен случай, когда для двух *зависимых* величин коэффициент корреляции будет равен нулю.

Іример 13.4–1

Применим модель линейной регрессии к данным из примера 13.2–1, которые для удобства приведены в табл. 13.7.

ТАБЛИЦА 13.7

Месяц, x_i	Спрос, y_i	Месяц, x_i	Спрос, y_i
1	46	13	54
2	56	14	42
3	54	15	64
4	43	16	60
5	57	17	70
6	56	18	66
7	67	19	57
8	62	20	55
9	50	21	52
10	56	22	62
11	47	23	70
12	56	24	72

Из данных этой таблицы получаем следующее.

$$\sum_{i=1}^{24} y_i x_i = 17842,\quad \sum_{i=1}^{24} x_i = 300,\quad \sum_{i=1}^{24} x_i^2 = 4900,\quad \sum_{i=1}^{24} y_i = 1374,\quad \sum_{i=1}^{24} y_i^2 = 80254.$$

Следовательно,

$$\overline{x} = 12.5,\quad \overline{y} = 57.25,$$

$$b = \frac{17842 - 24\times 57.25\times 12.5}{4900 - 24\times 12.5^2} = 0.58,$$

$$a = 57.25 - 0.58 \times 12.5 = 50.$$

Таким образом, оценка спроса представляется формулой

$$y^* = 50 + 0.58x.$$

Например, при $x = 25$ получаем $y^* = 50 + 0.58 \times 25 = 64.5$ единицы.
Вычисляем коэффициент корреляции:

$$r = \frac{17842 - 24\times 57.25\times 12.5}{\sqrt{\left(4900 - 24\times 12.5^2\right)\left(80.254 - 24\times 57.25^2\right)}} = 0.493.$$

Относительно малое значение коэффициента корреляции r указывает на то, что линейная модель $y^* = 50 + 0.58x$ является не совсем подходящей для исходных данных. Считается, как правило, что линейная модель подходит для исходных данных, если $0.75 \le |r| \le 1$.

Предположим, необходимо вычислить 95%-ный доверительный интервал для полученной линейной оценки. Для этого надо сначала вычислить сумму квадратов отклонений от аппроксимирующей прямой. В табл. 13.8 приведены результаты этих вычислений.

ТАБЛИЦА 13.8

õ	ó	ó*	$(ó - ó^*)^2$
1	46	50.58	20.98
2	56	51.16	23.43
3	54	51.74	5.11
4	43	54.32	86.86
5	57	52.90	16.81
6	56	53.48	6.35
7	67	54.06	152.77
8	62	54.64	54.17
9	50	55.22	27.25
10	56	55.80	0.04
11	47	56.38	87.98
12	56	56.96	0.92
13	54	57.54	12.53
14	42	58.12	259.85
15	64	58.70	28.09
16	60	59.28	0.52
17	70	59.86	102.82
18	66	60.44	30.91
19	57	61.02	16.16
20	55	61.60	43.56
21	52	62.18	103.63
22	62	62.76	0.58
23	70	63.34	44.53
24	72	63.92	65.29

$$\sum_{i=1}^{24}\left(y_i - y_i^*\right)^2 = 1088.70$$

Из табл. 2 Приложения Д имеем $t_{0.025;\,22} = 2.074$. Следовательно, искомый доверительный интервал имеет вид

$$\left(50+0.58x^0\right)\pm 2.074\sqrt{\frac{1088.7}{24-2}}\sqrt{\frac{1}{24}+\frac{\left(x^0-12.5\right)^2}{4900-24\times 12.5^2}}.$$

Это выражение можно упростить, в результате получим следующее.

$$\left(50+0.58x^0\right)\pm 14.59\sqrt{0.042+\frac{\left(x^0-12.5\right)^2}{1150}}.$$

Чтобы продемонстрировать применение этой формулы, вычислим интервал предсказания для оценки спроса на следующий месяц ($x^0 = 25$). В этом случае коэффициент 0.042 должен быть заменен на 1.042[1], и соответствующий интервал предсказания определяется как (64.5 ± 15.82) или (46.68, 80.32). Следовательно, можно сказать, что с вероятностью 95% спрос для $x = 25$ будет находиться между 46.68 и 80.32 единицами.

Упражнения 13.4,а

1. Примените метод линейной регрессии к данным из упр. 13.2,а(2).
2. Примените метод линейной регрессии к данным из упр. 13.2,а(3).
3. Примените метод линейной регрессии к данным из упр. 13.2,а(4).
4. Примените метод линейной регрессии к данным из упр. 13.2,а(5).
5. Докажите, что при линейной регрессии сумма разностей между расчетными и предсказанными величинами по всем исходным данным равна нулю, т.е. выполняется равенство

$$\sum_{i=1}^{n}\left(y_i - y_i^*\right) = 0.$$

13.5. Заключение

В этой главе рассмотрены три метода прогнозирования: прогнозирование с использованием скользящего среднего, прогнозирование путем экспоненциального сглаживания и регрессионное прогнозирование. Каждый метод применим при определенных характеристиках временного ряда, представляющего исходные данные. Существуют другие методики прогнозирования, с которыми можно ознакомиться в дополнительной литературе.

Литература

Brown R.G. *Smoothing, Forecasting, and Prediction of Discrete Time Series*, Prentice Hall, Upper Saddle River, N.J., 1972.

Montgomery D. and Peck E. *Introduction to Linear Regression Analysis*, Wiley, New York, 1991.

Willis R.E. *A Guide to Forecasting for Planners and Managers*, Prentice Hall, Upper Saddle River, N.J., 1987.

Литература, добавленная при переводе

Айвазян С.А., Енюков И.С., Мешалкин Л.Д. *Прикладная статистика: Исследование зависимостей.* — М.: Финансы и статистика, 1985.

Алгоритмы и программы восстановления зависимостей/Под ред. В.Н. Вапника. — М.: Наука, 1984.

Бендат Дж., Пирсол А. *Прикладной анализ случайных данных.* — М.: Мир, 1989.

Пугачев В.С. *Теория вероятностей и математическая статистика.* — М.: Наука, 1979.

[1] Напомним, что определение интервала предсказания основано на формуле, определяющей доверительный интервал, где под корнем слагаемое $1/n$ заменено на $(n + 1)/n$. — *Прим ред.*

Теория игр и принятия решений

14.1. Условия принятия решений

В теории принятия решений используются "разумные" процедуры выбора наилучшей из нескольких возможных альтернатив. Доброкачественность выбранного решения зависит от качества данных, используемых при описании ситуации, в которой принимается решение. С этой точки зрения процесс принятия решений может принадлежать к одному из трех возможных условий.

1. Принятие решений в условиях **определенности**, когда данные известны точно.
2. Принятие решений в условиях **риска**, когда данные можно описать с помощью вероятностных распределений.
3. Принятие решений в условиях **неопределенности**, когда данным нельзя приписать относительные веса (весовые коэффициенты), которые представляли бы степень их значимости в процессе принятия решений.

По существу, в условиях определенности данные надежно определены, в условиях неопределенности они не определены.[1] Принятие решений в условиях риска, следовательно, представляет "промежуточный" случай.

В этой главе представлены некоторые модели принятия решений, относящиеся к указанным трем условиям.

14.2. Принятие решений в условиях определенности

Модели линейного программирования (главы 2–8) являются примером принятия решений в условиях определенности. Эти модели применимы лишь в тех случаях, когда альтернативные решения можно связать между собой точными линейными функциями. В этом разделе рассматривается иной подход к принятию решений в ситуациях, когда, например, для идей, чувств, эмоций определяются некоторые количественные показатели, обеспечивающие числовую шкалу предпочтений для возможных альтернативных решений. Этот подход известен как метод анализа иерархий.

---

[1] Это не значит, что в условии неопределенности полностью отсутствует информация о задаче. Речь идет о том, что имеющиеся данные трудно или невозможно классифицировать по степени значимости их для принятия решения и что для этих данных, рассматриваемых как реализации случайных величин или процессов, неизвестна или не может быть определена их функция распределения или другие статистические характеристики. — *Прим. ред.*

14.2.1. Метод анализа иерархий

Перед тем как изложить детали данного метода, рассмотрим пример, демонстрирующий способ, с помощью которого оцениваются различные альтернативные решения.

Пример 14.2–1

Мартин Ганс — выпускник-отличник средней школы, который получил полную стипендию от трех университетов: A, B и C. В целях выбора университета Мартин сформулировал два основных критерия: местонахождение университета и его академическая репутация. Будучи отличным учеником, он оценивает академическую репутацию университета в пять раз выше, чем его местонахождение. Это приводит к тому, что репутации университета приписывается вес примерно 83%, а его местонахождению — 17%. Далее Мартин использует системный анализ (сущность его излагается ниже) для оценки трех университетов с точки зрения их местонахождения и репутации. Проведенный анализ дает следующие оценки.

	Университет		
	A	B	C
Местонахождение	12.9%	27.7%	59.4%
Репутация	54.5%	27.3%	18.2%

Структура задачи принятия решений приведена на рис. 14.1. Задача имеет единственный иерархический уровень с двумя критериями (местонахождение и репутация) и три альтернативных решения (университеты A, B и C).

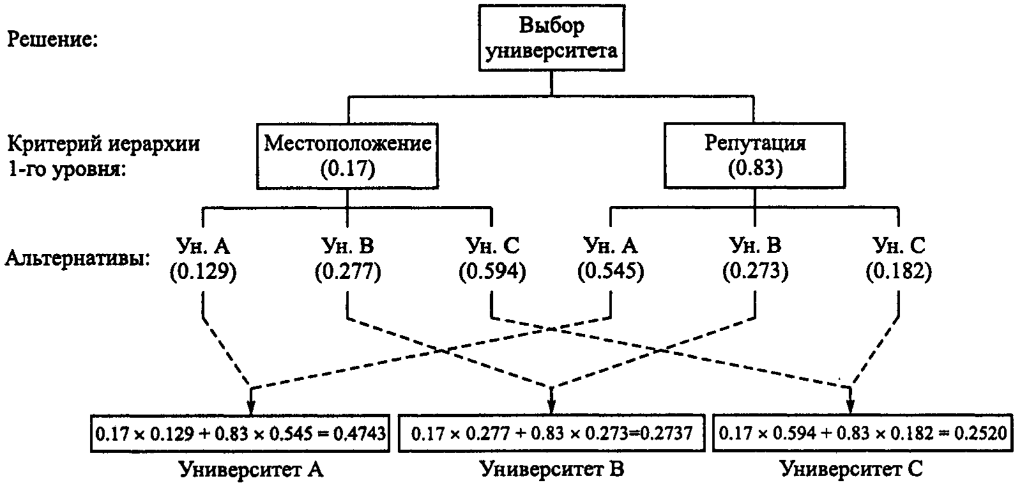


Рис. 14.1

Оценка трех университетов основана на вычислении *комбинированного* весового коэффициента для каждого из них.

Университет A: $0.17 \times 0.129 + 0.83 \times 0.545 = \mathbf{0.4743}$.
Университет B: $0.17 \times 0.277 + 0.83 \times 0.273 = 0.2737$.
Университет C: $0.17 \times 0.594 + 0.83 \times 0.182 = 0.2520$.

На основе этих вычислений университет А получает наивысший комбинированный вес и, следовательно, является наиболее оптимальным выбором Мартина.

Общая структура метода анализа иерархий может включать несколько иерархических уровней со своими критериями. Предположим в примере 14.2–1, что сестра-близнец Мартина Джейн также получила полную стипендию от трех университетов. Однако их родители ставят условие, что дети должны учиться в одном университете, тогда они смогут пользоваться одним автомобилем. На рис. 14.2 приведена структура задачи выбора решения, которая теперь включает два иерархических уровня со своими критериями. Величины p и q (предположительно равные) на первом иерархическом уровне представляют собой весовые коэффициенты, которые приписываются точке зрения Мартина и Джейн относительно процесса выбора, соответственно. Второй иерархический уровень использует веса (p_1, p_2) и (q_1, q_2) для отображения индивидуальных точек зрения Мартина и Джейн относительно критериев местонахождения и академической репутации каждого университета. Остальная часть структуры принятия решения может быть интерпретирована аналогично предыдущему примеру. Заметим, что $p + q = 1$, $p_1 + p_2 = 1$, $q_1 + q_2 = 1$, $p_{11} + p_{12} + p_{13} = 1$, $p_{21} + p_{22} + p_{23} = 1$, $q_{11} + q_{12} + q_{13} = 1$, $q_{21} + q_{22} + q_{23} = 1$. Определение комбинированного веса для университета А, представленное на рис. 14.2, демонстрирует, каким образом вычисляются эти показатели.

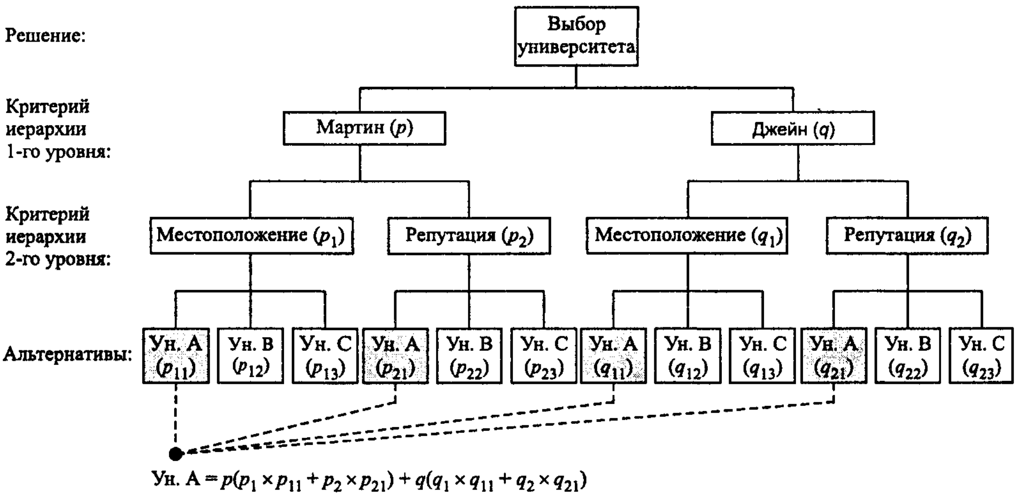


Рис. 14.2

Упражнение 14.2, а

1. Пусть для задачи выбора университета Мартином и Джейн установлены следующие значения весовых коэффициентов.

$p = 0.5, q = 0.5,$

$p_1 = 0.17, p_2 = 0.83,$

$p_{11} = 0.129, p_{12} = 0.277, p_{13} = 0.594,$

$p_{21} = 0.545, p_{22} = 0.273, p_{23} = 0.182,$

$q_1 = 0.3, q_2 = 0.7,$

$q_{11} = 0.2, q_{12} = 0.3, q_{13} = 0.5,$

$q_{21} = 0.5, q_{22} = 0.2, q_{23} = 0.3.$

Основываясь на этой информации, оцените с помощью комбинированных весов каждый из трех университетов.

Определение весовых коэффициентов. Сложность метода анализа иерархий заключается в определении относительных весовых коэффициентов (таких, какие использованы в примере 14.2–1) для оценки альтернативных решений. Если имеется n критериев на заданном уровне иерархии, соответствующая процедура создает матрицу **A** размерности $n \times n$, именуемую *матрицей парных сравнений*, которая отражает суждение лица, принимающего решение, относительно важности разных критериев. Парное сравнение выполняется таким образом, что критерий в строке i ($i = 1, 2, ..., n$) оценивается относительно каждого из критериев, представленных n столбцами. Обозначим через a_{ij} элемент матрицы **A**, находящийся на пересечении i-й строки и j-го столбца. В соответствии с методом анализа иерархий для описания упомянутых оценок используются целые числа от 1 до 9. При этом $a_{ij} = 1$ означает, что i-й и j-й критерии *одинаково важны*, $a_{ij} = 5$ отражает мнение, что i-й критерий *значительно важнее*, чем j-й, а $a_{ij} = 9$ указывает, что i-й критерий *чрезвычайно важнее* j-го. Другие промежуточные значения между 1 и 9 интерпретируются аналогично. Согласованность таких обозначений обеспечивается следующим условием: если $a_{ij} = k$, то автоматически $a_{ji} = 1/k$. Кроме того, все диагональные элементы a_{ii} матрицы **A** должны быть равны 1, так как они выражают оценку критерия относительно самих себя.

Пример 14.2–2

Покажем, как определяется матрица сравнения **A** для задачи выбора Мартина из примера 14.2–1. Начнем с главного иерархического уровня, который имеет дело с критериями академической репутации университета и его местонахождения. С точки зрения Мартина академическая репутация университета *значительно важнее* его местонахождения. Следовательно, он приписывает элементу (1, 2) матрицы **A** значение 5, т.е. $a_{12} = 5$. Это автоматически предполагает, что $a_{21} = 1/5$. Обозначив через R и L критерии репутации университета и его местонахождения, можно записать матрицу сравнения следующим образом.

$$\mathbf{A} = \begin{array}{c} \\ R \\ L \end{array} \begin{array}{c} \begin{array}{cc} R & L \end{array} \\ \begin{bmatrix} 1 & 5 \\ \frac{1}{5} & 1 \end{bmatrix} \end{array}.$$

Относительные веса критериев R и L могут быть определены путем деления элементов каждого столбца на сумму элементов этого же столбца. Следовательно, для нормализации матрицы **A** делим элементы первого столбца на величину

$1 + 1/5 = 1.2$, элементы второго — на величину $5 + 1 = 6$. Искомые относительные веса w_R и w_L критериев вычисляются теперь в виде средних значений элементов соответствующих строк нормализованной матрицы **A**. Следовательно,

$$\mathbf{N} = \begin{array}{c} \\ R \\ L \end{array} \begin{array}{c} \begin{array}{cc} R & L \end{array} \\ \begin{bmatrix} 0.83 & 0.83 \\ 0.17 & 0.17 \end{bmatrix} \end{array} \qquad \begin{array}{l} \text{Средние значения элементов строк} \\ w_R = (0.83+0.83)/2 = 0.83, \\ w_L = (0.17+0.17)/2 = 0.17. \end{array}$$

В результате вычислений получили $w_R = 0.83$ и $w_L = 0.17$, т.е. те веса, которые показаны на рис. 14.1. Столбцы матрицы **N** одинаковы, что имеет место лишь в случае, когда лицо, принимающее решение, проявляет идеальную *согласованность* в определении элементов матрицы **A**. Этот тезис детальнее обсуждается ниже.

Относительные веса альтернативных решений, соответствующих университетам A, B и C, вычисляются в пределах каждого критерия R и L с использованием следующих двух матриц сравнения.

$$\mathbf{A}_R = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 2 & 3 \\ \frac{1}{2} & 1 & \frac{3}{2} \\ \frac{1}{3} & \frac{2}{3} & 1 \end{bmatrix} \end{array},$$

$$\text{Сумма элементов столбцов} = [1.83,\ \ 3.67,\ \ 5.5],$$

$$\mathbf{A}_L = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{5} \\ 2 & 1 & \frac{1}{2} \\ 5 & 2 & 1 \end{bmatrix} \end{array},$$

$$\text{Сумма элементов столбцов} = [8,\ \ 3.5,\ \ 1.7].$$

Элементы матриц $\mathbf{A}_R$ и $\mathbf{A}_L$ определены на основе суждений Мартина, касающихся относительной важности трех университетов.

При делении элементов каждого столбца матриц $\mathbf{A}_R$ и $\mathbf{A}_L$ на сумму элементов этих же столбцов получаем следующие нормализованные матрицы.

$$\mathbf{N}_R = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0.545 & 0.545 & 0.545 \\ 0.273 & 0.273 & 0.273 \\ 0.182 & 0.182 & 0.182 \end{bmatrix} \end{array}, \qquad \begin{array}{l} \text{Средние значения элементов строк} \\ w_{RA} = (0.545+0.545+0.545)/3 = 0.545, \\ w_{RB} = (0.273+0.273+0.273)/3 = 0.273, \\ w_{RC} = (0.182+0.182+0.182)/3 = 0.182, \end{array}$$

$$\mathbf{N}_L = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0.125 & 0.143 & 0.118 \\ 0.250 & 0.286 & 0.294 \\ 0.625 & 0.571 & 0.588 \end{bmatrix} \end{array}, \qquad \begin{array}{l} \text{Средние значения элементов строк} \\ w_{LA} = (0.125+0.143+0.118)/3 = 0.129, \\ w_{LB} = (0.250+0.286+0.294)/3 = 0.277, \\ w_{LC} = (0.625+0.571+0.588)/3 = 0.594. \end{array}$$

Величины $(w_{RA}, w_{RB}, w_{RC}) = (0.545, 0.273, 0.182)$ дают соответствующие веса для университетов A, B и C с точки зрения академической репутации. Аналогично величины $(w_{LA}, w_{LB}, w_{LC}) = (0.129, 0.277, 0.594)$ являются относительными весами, касающимися местонахождения университетов.

Согласованность матрицы сравнений. В примере 14.2–2 мы отмечали, что все столбцы нормализованных матриц $\mathbf{N}$ и $\mathbf{N}_R$ идентичны, а столбцы матрицы $\mathbf{N}_L$ таковыми не являются. Одинаковые столбцы указывают на то, что результирующие относительные веса сохраняют одно и то же значение независимо от того, как выполняется сравнение. В этом случае говорят, что исходные матрицы сравнения $\mathbf{A}$ и $\mathbf{A}_R$ являются *согласованными*. Следовательно, матрица $\mathbf{A}_L$ не является таковой.

Согласованность означает, что решение будет согласовано с определениями парных сравнений критериев или альтернатив. С математической точки зрения согласованность матрицы $\mathbf{A}$ означает, что $a_{ij}a_{jk} = a_{ik}$ для всех i, j и k. Например, в матрице $\mathbf{A}_R$ из примера 14.2–2 $a_{13} = 3$ и $a_{12}a_{23} = 2 \times 3/2 = 3$. Свойство согласованности требует линейной зависимости столбцов (и строк) матрицы $\mathbf{A}$. В частности, столбцы любой матрицы сравнений размерностью 2×2 являются зависимыми, и, следовательно, такая матрица всегда является согласованной. Не все матрицы сравнений являются согласованными. Действительно, принимая во внимание, что такие матрицы строятся на основе человеческих суждений, можно ожидать некоторую степень несогласованности, и к ней следует относиться терпимо при условии, что она не выходит за определенные "допустимые" рамки.

Чтобы выяснить, является ли уровень согласованности "допустимым", необходимо определить соответствующую количественную меру для матрицы сравнений $\mathbf{A}$. В примере 14.2–2 мы видели, что идеально согласованная матрица $\mathbf{A}$ порождает нормализованную матрицу $\mathbf{N}$, в которой все столбцы одинаковы.

$$\mathbf{N} = \begin{bmatrix} w_1 & w_1 & \cdots & w_1 \\ w_2 & w_2 & \cdots & w_2 \\ \vdots & \vdots & \vdots & \vdots \\ w_n & w_n & \cdots & w_n \end{bmatrix}.$$

Отсюда следует, что матрица сравнений $\mathbf{A}$ может быть получена из матрицы $\mathbf{N}$ путем деления элементов i-го столбца на w_i (это процесс, обратный к нахождению матрицы $\mathbf{N}$ из $\mathbf{A}$). Итак, получаем следующее.

$$\mathbf{A} = \begin{bmatrix} 1 & \dfrac{w_1}{w_2} & \cdots & \dfrac{w_1}{w_n} \\ \dfrac{w_2}{w_1} & 1 & \cdots & \dfrac{w_2}{w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \dfrac{w_n}{w_1} & \dfrac{w_n}{w_2} & \cdots & 1 \end{bmatrix}.$$

Используя приведенное определение матрицы **A**, имеем

$$\begin{bmatrix} 1 & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & 1 & \cdots & \frac{w_2}{w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} nw_1 \\ nw_2 \\ \vdots \\ nw_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}.$$

В компактной форме условие согласованности матрицы **A** формулируется следующим образом. Матрица **A** будет согласованной тогда и только тогда, когда

$$\mathbf{A}\mathbf{w} = n\mathbf{w},$$

где **w** — вектор-столбец относительных весов w_i, $i = 1, 2, ..., n$.

Когда матрица **A** не является согласованной, относительный вес w_i аппроксимируется средним значением n элементов i-й строки нормализованной матрицы **N** (см. пример 14.2–2). Обозначив через $\overline{\mathbf{w}}$ вычисленную оценку (среднее значение), можно показать, что

$$\mathbf{A}\overline{\mathbf{w}} = n_{\max}\overline{\mathbf{w}},$$

где $n_{\max} \geq n$. В этом случае, чем ближе $n_{\max}$ к n, тем более согласованной является матрица сравнения **A**. В результате в соответствии с методом анализа иерархий вычисляется *коэффициент согласованности* в виде

$$CR = \frac{CI}{RI},$$

где

$CI = \dfrac{n_{\max} - n}{n-1}$ — коэффициент согласованности матрицы **A**,

$RI = \dfrac{1.98(n-2)}{n}$ — стохастический коэффициент согласованности матрицы **A**.

Стохастический коэффициент согласованности RI определяется эмпирическим путем как среднее значение коэффициента CI для большой выборки генерированных случайным образом матриц сравнения **A**.

Коэффициент согласованности CR используется для проверки согласованности матрицы сравнения **A** следующим образом. Если $CR \leq 0.1$, уровень несогласованности является приемлемым. В противном случае уровень несогласованности матрицы сравнения **A** является высоким и лицу, принимающему решение, рекомендуется проверить элементы парного сравнения a_{ij} матрицы **A** в целях получения более согласованной матрицы.

Значение $n_{\max}$ вычисляется на основе матричного уравнения $\mathbf{A}\overline{\mathbf{w}} = n_{\max}\overline{\mathbf{w}}$, при этом нетрудно заметить, что i-е уравнение этой системы имеет вид:

$$\sum_{j=1}^{n} a_{ij}\overline{w}_j = n_{\max}\overline{w}_i, \quad i = 1, 2, ..., n.$$

Поскольку $\sum_{i=1}^{n} \overline{w}_i = 1$, легко проверить, что

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}\overline{w}_j\right) = n_{\max}\sum_{i=1}^{n}\overline{w}_i = n_{\max}.$$

Это значит, что величину $n_{\max}$ можно определить путем вычисления вектор-столбца $\mathbf{A}\overline{\mathbf{w}}$ с последующим суммированием его элементов.

Пример 14.2–3

В примере 14.2–2 матрица $\mathbf{A}_L$ является несогласованной, так как столбцы матрицы $\mathbf{N}_L$ неодинаковы. Требуется исследовать согласованность матрицы $\mathbf{A}_L$.

Вычислим значение $n_{\max}$. Из данных примера 14.2–2 имеем

$$\overline{w}_1 = 0.129,\ \overline{w}_2 = 0.277,\ \overline{w}_3 = 0.594.$$

Следовательно,

$$\mathbf{A}_L\overline{\mathbf{w}} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{5} \\ 2 & 1 & \frac{1}{2} \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0.129 \\ 0.277 \\ 0.594 \end{bmatrix} = \begin{bmatrix} 0.3863 \\ 0.8320 \\ 1.7930 \end{bmatrix}.$$

Отсюда получаем

$$n_{\max} = 0.3863 + 0.8320 + 1.7930 = 3.0113.$$

Следовательно, для $n = 3$ имеем

$$CI = \frac{n_{\max} - n}{n-1} = \frac{3.0113 - 3}{3-1} = 0.00565,$$

$$RI = \frac{1.98(n-2)}{n} = \frac{1.98\times 1}{3} = 0.66,$$

$$CR = \frac{CI}{RI} = \frac{0.00565}{0.66} = 0.00856.$$

Так как $CR < 0.1$, уровень несогласованности матрицы $\mathbf{A}_L$ является приемлемым.

Упражнения 14.2, b

1. Отдел кадров фирмы сузил поиск будущего сотрудника до трех кандидатур: Стив (S), Джейн (J) и Майса (M). Конечный отбор основан на трех критериях: собеседование (C), опыт работы (O) и рекомендации (P). Отдел кадров использует матрицу $\mathbf{A}$ (приведенную ниже) для сравнения трех критериев. После проведенного собеседования с тремя претендентами, сбора данных, относящихся к опыту их работы и рекомендациям, построены матрицы $\mathbf{A}_C$, $\mathbf{A}_O$ и $\mathbf{A}_P$. Какого из трех кандидатов следует принять на работу? Оцените согласованность данных.

$$\mathbf{A} = \begin{array}{c} \\ C \\ O \\ P \end{array}\!\!\begin{array}{c} \begin{array}{ccc} C & O & P \end{array} \\ \begin{bmatrix} 1 & 2 & \frac{1}{4} \\ \frac{1}{2} & 1 & \frac{1}{5} \\ 4 & 5 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_C = \begin{array}{c} \\ S \\ J \\ M \end{array}\!\!\begin{array}{c} \begin{array}{ccc} S & J & M \end{array} \\ \begin{bmatrix} 1 & 3 & 4 \\ \frac{1}{3} & 1 & \frac{1}{5} \\ \frac{1}{4} & 5 & 1 \end{bmatrix} \end{array},$$

$$\mathbf{A}_O = \begin{array}{c} \\ S \\ J \\ M \end{array}\!\!\begin{array}{c} \begin{array}{ccc} S & J & M \end{array} \\ \begin{bmatrix} 1 & \frac{1}{3} & 2 \\ 3 & 1 & \frac{1}{2} \\ \frac{1}{2} & 2 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_P = \begin{array}{c} \\ S \\ J \\ M \end{array}\!\!\begin{array}{c} \begin{array}{ccc} S & J & M \end{array} \\ \begin{bmatrix} 1 & \frac{1}{2} & 1 \\ 2 & 1 & \frac{1}{2} \\ 1 & 2 & 1 \end{bmatrix} \end{array}.$$

2. Кевин и Джун Парки (*К* и *Д*) покупают новый дом. Рассматриваются три варианта *A*, *B* и *C*. Парки согласовали два критерия для выбора дома: площадь зеленой лужайки (*Л*) и близость к месту работы (*Б*), а также разработали матрицы сравнений, приведенные ниже. Необходимо оценить три дома в порядке их приоритета и вычислить коэффициент согласованности каждой матрицы.

$$\mathbf{A} = \begin{array}{c} \\ К \\ Д \end{array}\!\!\begin{array}{c} \begin{array}{cc} К & Д \end{array} \\ \begin{bmatrix} 1 & 2 \\ \frac{1}{2} & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_К = \begin{array}{c} \\ Л \\ Б \end{array}\!\!\begin{array}{c} \begin{array}{cc} Л & Б \end{array} \\ \begin{bmatrix} 1 & \frac{1}{3} \\ 3 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_Д = \begin{array}{c} \\ Л \\ Б \end{array}\!\!\begin{array}{c} \begin{array}{cc} Л & Б \end{array} \\ \begin{bmatrix} 1 & 4 \\ \frac{1}{4} & 1 \end{bmatrix} \end{array},$$

$$\mathbf{A}_{КЛ} = \begin{array}{c} \\ A \\ B \\ C \end{array}\!\!\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 2 & 3 \\ \frac{1}{2} & 1 & 2 \\ \frac{1}{3} & \frac{1}{2} & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_{КБ} = \begin{array}{c} \\ A \\ B \\ C \end{array}\!\!\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 2 & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{3} \\ 2 & 3 & 1 \end{bmatrix} \end{array},$$

$$\mathbf{A}_{ДЛ} = \begin{array}{c} \\ A \\ B \\ C \end{array}\!\!\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & 4 & 2 \\ \frac{1}{4} & 1 & 3 \\ \frac{1}{2} & \frac{1}{3} & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_{ДБ} = \begin{array}{c} \\ A \\ B \\ C \end{array}\!\!\begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 1 & \frac{1}{2} & 4 \\ 2 & 1 & 3 \\ \frac{1}{4} & \frac{1}{3} & 1 \end{bmatrix} \end{array}.$$

Автор книги по исследованию операций определил три критерия для выбора издательства, которое будет печатать его книгу: процент авторского гонорара (*R*), уровень маркетинга (*M*) и размер аванса (А). Издательства *H* и *P* проявили интерес к изданию книги. Используя приведенные ниже матрицы сравнения, необходимо дать оценку двум издательствам и оценить согласованность решения.

$$\mathbf{A} = \begin{array}{c} \\ R \\ M \\ A \end{array} \begin{array}{c} \begin{array}{ccc} R & M & A \end{array} \\ \begin{bmatrix} 1 & 1 & \frac{1}{4} \\ 1 & 1 & \frac{1}{5} \\ 4 & 5 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_R = \begin{array}{c} \\ H \\ P \end{array} \begin{array}{c} \begin{array}{cc} H & P \end{array} \\ \begin{bmatrix} 1 & 2 \\ \frac{1}{2} & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_M = \begin{array}{c} \\ H \\ P \end{array} \begin{array}{c} \begin{array}{cc} H & P \end{array} \\ \begin{bmatrix} 1 & \frac{1}{2} \\ 2 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_A = \begin{array}{c} \\ H \\ P \end{array} \begin{array}{c} \begin{array}{cc} H & P \end{array} \\ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{array}.$$

4. Профессор политологии планирует предсказать исход выборов в местный школьный совет. Кандидаты I, B и S баллотируются на одно место. Профессор делит всех избирателей на три категории: левые (L), центристы (C) и правые (R). Оценка кандидатов основывается на трех факторах: педагогический опыт (O), отношение к детям ($Д$) и характер (X). Ниже приведены матрицы сравнения для первого иерархического уровня, связанного с градацией избирателей (левые, центристы и правые).

$$\mathbf{A} = \begin{array}{c} \\ L \\ C \\ R \end{array} \begin{array}{c} \begin{array}{ccc} L & C & R \end{array} \\ \begin{bmatrix} 1 & 2 & \frac{1}{2} \\ \frac{1}{2} & 1 & \frac{1}{5} \\ 2 & 5 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_L = \begin{array}{c} \\ O \\ Д \\ X \end{array} \begin{array}{c} \begin{array}{ccc} O & Д & X \end{array} \\ \begin{bmatrix} 1 & 3 & \frac{1}{2} \\ \frac{1}{3} & 1 & \frac{1}{3} \\ 2 & 3 & 1 \end{bmatrix} \end{array},$$

$$\mathbf{A}_C = \begin{array}{c} \\ O \\ Д \\ X \end{array} \begin{array}{c} \begin{array}{ccc} O & Д & X \end{array} \\ \begin{bmatrix} 1 & 2 & 2 \\ \frac{1}{2} & 1 & 1 \\ \frac{1}{2} & 1 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_R = \begin{array}{c} \\ O \\ Д \\ X \end{array} \begin{array}{c} \begin{array}{ccc} O & Д & X \end{array} \\ \begin{bmatrix} 1 & 1 & 9 \\ 1 & 1 & 8 \\ \frac{1}{9} & \frac{1}{8} & 1 \end{bmatrix} \end{array}.$$

Профессор сгенерировал еще девять матриц сравнения для трех кандидатов на втором иерархическом уровне, связанном с педагогическим опытом, отношением к детям и характером. Затем был использован метод анализа иерархий для сведения этих матриц к следующим относительным весам.

	Левые			Центристы			Правые		
Кандидат	O	$Д$	X	O	$Д$	X	O	$Д$	X
I	0.1	0.2	0.3	0.3	0.5	0.2	0.7	0.1	0.3
B	0.5	0.4	0.2	0.4	0.2	0.4	0.1	0.4	0.2
S	0.4	0.4	0.5	0.3	0.3	0.4	0.2	0.5	0.5

Используя эту информацию, необходимо определить, кто из кандидатов выиграет выборы, и оценить согласованность решения.

5. Школьный округ крайне заинтересован в сокращении своих расходов, что вызвано очередным уменьшением бюджетного финансирования начальных школ. Есть две возможности решить эту проблему: ликвидировать программу физического воспитания ($Ф$) или программу музыкального образования (M). Управляющий округа сформировал комитет с равным представительством от местного школьного

совета (*С*) и ассоциации родителей и учителей (*Р*) для изучения ситуации и выработки предложения. Комитет принял решение изучить ситуацию с точки зрения ограничения бюджета (*Б*) и потребностей учеников (*П*). Проведенный анализ дал следующие матрицы сравнения.

$$\mathbf{A}_C = \begin{array}{c} \\ Б \\ П \end{array} \begin{array}{c} Б \quad П \\ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_P = \begin{array}{c} \\ Б \\ П \end{array} \begin{array}{c} Б \quad П \\ \begin{bmatrix} 1 & 2 \\ \frac{1}{2} & 1 \end{bmatrix} \end{array},$$

$$\mathbf{A}_{СБ} = \begin{array}{c} \\ Ф \\ М \end{array} \begin{array}{c} Ф \quad М \\ \begin{bmatrix} 1 & \frac{1}{2} \\ 2 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_{СП} = \begin{array}{c} \\ Ф \\ М \end{array} \begin{array}{c} Ф \quad М \\ \begin{bmatrix} 1 & \frac{1}{3} \\ 3 & 1 \end{bmatrix} \end{array},$$

$$\mathbf{A}_{РБ} = \begin{array}{c} \\ Ф \\ М \end{array} \begin{array}{c} Ф \quad М \\ \begin{bmatrix} 1 & \frac{1}{3} \\ 3 & 1 \end{bmatrix} \end{array}, \quad \mathbf{A}_{РП} = \begin{array}{c} \\ Ф \\ М \end{array} \begin{array}{c} Ф \quad М \\ \begin{bmatrix} 1 & 2 \\ \frac{1}{2} & 1 \end{bmatrix} \end{array}.$$

Требуется проанализировать ситуацию, связанную с принятием решения, и выработать соответствующее предложение.

6. Решив купить автомобиль, человек сузил свой выбор до трех моделей: *М*1, *М*2 и *М*3. Факторами, влияющими на его решение, являются: стоимость автомобиля (С), стоимость обслуживания (О), стоимость поездки по городу (Г) и сельской местности (М). Следующая таблица содержит необходимые данные, соответствующие трехгодичному сроку эксплуатации автомобиля.

Модель автомобиля	С ($)	О ($)	Г ($)	М ($)
*М*1	6 000	1800	4500	1500
*М*2	8 000	1200	2250	750
*М*3	10 000	600	1125	600

Используйте указанные стоимости для построения матриц сравнений. Оцените согласованность матриц и определите модель автомобиля, которую следует выбрать.

14.3. Принятие решений в условиях риска

Если решение принимается в условиях риска, то стоимости альтернативных решений обычно описываются вероятностными распределениями. По этой причине принимаемое решение основывается на использовании *критерия ожидаемого значения*, в соответствии с которым альтернативные решения сравниваются с точки зрения максимизации ожидаемой прибыли или минимизации ожидаемых затрат. Такой подход имеет свои недостатки, которые не позволяют использовать его в некоторых ситуациях. Для них разработаны модификации упомянутого критерия. В этой главе рассматриваются часто используемые подходы к принятию решений в условиях риска.

14.3.1. Критерий ожидаемого значения

Критерий ожидаемого значения сводится либо к максимизации ожидаемой (средней) прибыли, либо к минимизации ожидаемых затрат. В данном случае предполагается, что прибыль (затраты), связанная с каждым альтернативным решением, является случайной величиной.

Дерево решений. В приведенном ниже примере рассматривается простая ситуация, связанная с принятием решения при наличии конечного числа альтернатив и точных значений матрицы доходов.

Пример 14.3–1

Предположим, что вы хотите вложить на фондовой бирже 10 000 долларов в акции одной из двух компаний: A или B. Акции компании A являются рискованными, но могут принести 50% прибыли от суммы инвестиции на протяжении следующего года. Если условия фондовой биржи будут неблагоприятны, сумма инвестиции может обесцениться на 20%. Компания B обеспечивает безопасность инвестиций с 15% прибыли в условиях повышения котировок на бирже и только 5% — в условиях понижения котировок. Все аналитические публикации, с которыми можно познакомиться (а они всегда есть в изобилии в конце года), с вероятностью 60% прогнозируют повышение котировок и с вероятностью 40% — понижение котировок. В какую компанию следует вложить деньги?

Информация, связанная с принятием решения, суммирована в следующей таблице.

Альтернативное решение	Прибыль от инвестиции за один год	
	При повышении котировок ($)	При понижении котировок ($)
Акции компании A	5000	–2000
Акции компании B	1500	500
Вероятность события	0.6	0.4

Эта задача может быть также представлена в виде **дерева решений**, показанного на рис. 14.3. На этом рисунке используется два типа вершин: квадратик представляет *"решающую" вершину*, а кружок — *"случайную"*. Таким образом, из вершины 1 ("решающая") выходят две ветви, представляющие альтернативы, связанные с покупкой акций компании A или B. Далее две ветви, выходящие из "случайных" вершин 2 и 3, соответствуют случаям повышения и понижения котировок на бирже с вероятностями их появления и соответствующими платежами.

Исходя из схемы рис. 14.3, получаем ожидаемую прибыль за год для каждой из двух альтернатив.

Для акций компании A: $5000 × 0.6 + (–2000) × 0.4 = **$2 200**.

Для акций компании B: $1500 × 0.6 + $500 × 0.4 = $1 100.

Вашим решением, основанным на этих вычислениях, является покупка акций компании A.

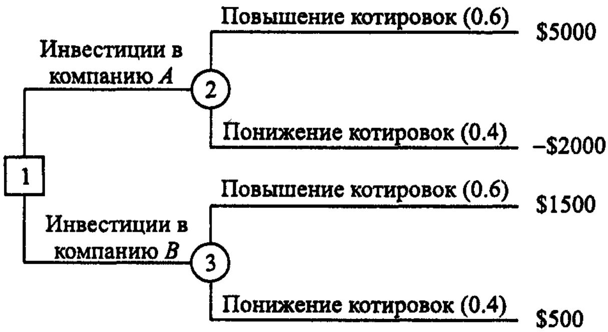


Рис. 14.3

В теории принятия решений повышение и понижение котировок на бирже именуются **состояниями природы**, возможные реализации которых являются случайными событиями (в данном случае с вероятностями 0.6 и 0.4). В общем случае задача принятия решений может включать n состояний природы и m альтернатив. Если p_j — вероятность j-го состояния природы, а a_{ij} — платеж, связанный с принятием решения i при состоянии природы j ($i = 1, 2, ..., m, j = 1, 2, ..., n$), тогда ожидаемый платеж для решения i вычисляется в виде

$$MV_i = a_{i1}p_1 + a_{i2}p_2 + ... + a_{in}p_n,\ i = 1, 2, ..., n,$$

где по определению $p_1 + p_2 + ... + p_n = 1$.

Наилучшим решением будет то, которое соответствует $MV_i^* = \max_i\{MV_i\}$ или $MV_i^* = \min_i\{MV_i\}$, в зависимости от того, является ли платеж в задаче доходом (прибылью) или убытком (затратами).

Упражнения 14.3,а

1. Вас пригласили на телевизионную игру *Колесо фортуны*. Колесо управляется электронным образом с помощью двух кнопок, которые сообщают колесу сильное (*В*) или слабое (*Н*) вращение. Само колесо разделено на равные области — белую (*Б*) и красную (*К*). Вам сообщили, что в белой области колесо останавливается с вероятностью 0.3, а в красной — 0.7. Плата, которую вы получаете за игру, равна (в долларах) следующему.

	Б	*К*
В	800	200
Н	–2500	1000

Изобразите соответствующее дерево решений.

2. Фермер Мак-Кой может выращивать либо кукурузу, либо соевые бобы. Вероятность того, что цены на будущий урожай этих культур повысятся, останутся на том же уровне или понизятся, равна соответственно 0.25, 0.30 и 0.45. Если цены

возрастут, урожай кукурузы даст 30 000 долларов чистого дохода, а урожай соевых бобов — 10 000 долларов. Если цены останутся неизменными, Мак-Кой лишь покроет расходы. Но если цены станут ниже, урожай кукурузы и соевых бобов приведет к потерям в 35 000 и 5 000 долларов соответственно.

a) Представьте данную задачу в виде дерева решений.
b) Какую культуру следует выращивать Мак-Кою?

3. Допустим, у вас имеется возможность вложить деньги в три инвестиционных фонда открытого типа: простой, специальный (обеспечивающий максимальную долгосрочную прибыль от акций мелких компаний) и глобальный. Прибыль от инвестиции может измениться в зависимости от условий рынка. Существует 10%-ная вероятность, что ситуация на рынке ценных бумаг ухудшится, 50%-ная — что рынок останется умеренным и 40%-ная — рынок будет возрастать. Следующая таблица содержит значения процентов прибыли от суммы инвестиции при трех возможностях развития рынка.

Альтернатива (фонды)	Процент прибыли от инвестиции (%)		
	Ухудшающийся рынок	Умеренный рынок	Растущий рынок
Простой	+5	+7	+8
Специальный	−10	+5	+30
Глобальный	+2	+7	+20

a) Представьте задачу в виде дерева решений.
b) Какой фонд открытого типа вам следует выбрать?

4. Предположим, у вас имеется возможность вложить деньги либо в 7.5%-ные облигации, которые продаются по номинальной цене, либо в специальный фонд, который выплачивает лишь 1% дивидендов. Если существует вероятность инфляции, процентная ставка возрастет до 8%, и в этом случае номинальная стоимость облигаций увеличится на 10%, а цена акций фонда — на 20%. Если прогнозируется спад, то процентная ставка понизится до 6%. При этих условиях ожидается, что номинальная стоимость облигаций поднимется на 5%, а цена акций фонда увеличится на 20%. Если состояние экономики останется неизменным, цена акций фонда увеличится на 8%, а номинальная стоимость облигаций не изменится. Экономисты оценивают в 20% шансы наступления инфляции и в 15% — наступление спада. Ваше решение относительно инвестиций принимается с учетом экономических условий следующего года.

a) Представьте задачу в виде дерева решений.
b) Будете ли вы покупать акции фонда или облигации?

5. Фирма планирует производство новой продукции быстрого питания в национальном масштабе. Исследовательский отдел убежден в большом успехе новой продукции и хочет внедрить ее немедленно, без рекламной кампании на рынках сбыта фирмы. Отдел маркетинга положение вещей оценивает иначе и предлагает провести интенсивную рекламную кампанию. Такая кампания обойдется в 100 000 долларов, а в случае успеха принесет 950 000 долларов годового дохода. В случае неуспеха рек-

ламной кампании (вероятность этого составляет 30%) годовой доход оценивается лишь в 200 000 долларов. Если рекламная кампания не проводится вовсе, годовой доход оценивается в 400 000 долларов при условии, что покупателям понравится новая продукция (вероятность этого равна 0.8), и в 200 000 долларов с вероятностью 0.2, если покупатели останутся равнодушными к новой продукции.

a) Постройте соответствующее дерево решений.
b) Как должна поступить фирма в связи с производством новой продукции?

6. Симметричная монета подбрасывается три раза. Вы получаете один доллар за каждое выпадение герба (*Г*) и дополнительно 0.25 доллара за каждые два последовательных выпадения герба (заметим, что выпадение *ГГГ* состоит из двух последовательностей *ГГ*). Однако Вам приходится платить 1.1 доллара за каждое выпадение решетки (*Р*). Вашим решением является участие или неучастие в игре.

   a) Постройте соответствующее дерево решений для описанной игры.
   b) Будете ли вы играть в эту игру?

7. Предположим, у вас имеется возможность сыграть в игру следующего содержания. Симметричная игральная кость бросается два раза, при этом возможны четыре исхода: 1) выпадает два четных числа, 2) выпадает два нечетных числа, 3) выпадает сначала четное, затем нечетное число, 4) выпадает сначала нечетное, затем четное число. Вы можете делать одинаковые ставки на два исхода. Например, вы можете поставить на два четных числа (исход 1) и два нечетных (исход 2). Выигрыш на каждый доллар, поставленный на первый исход, равен 2 доллара, на второй и третий исходы — 1.95 доллара, на четвертый — 1.50 доллара.

   a) Постройте дерево решений для описанной игры.
   b) На какие исходы следует делать ставки?
   c) Можно ли иметь стабильный выигрыш в этой игре?

8. Фирма производит партии продукции с 0.8%, 1%, 1.2% и 1.4% бракованных изделий с вероятностями 0.4, 0.3, 0.25 и 0.05 соответственно. Три потребителя А, В и С заключили контракт на получение партий изделий с процентом некачественных изделий не выше 0.8%, 1.2% и 1.4% соответственно. Фирма штрафуется в сумме 1000 долларов за каждый пункт процента[1] в случае, когда процент некачественных изделий выше указанного. Наоборот, поставка партий изделий с меньшим процентом бракованных изделий, чем оговорено в контракте, приносит фирме прибыль в 500 долларов за каждый пункт процента. Предполагается, что партии изделий перед отправкой не проверяются.

   a) Постройте соответствующее дерево решений.
   b) Какой из потребителей должен иметь наивысший приоритет при получении своего заказа?

9. Фирма планирует открыть новое предприятие в Арканзасе. В настоящее время имеется возможность построить либо крупное предприятие, либо небольшое, которое через два года можно будет расширить при условии высокого спроса на вы-

[1] Пункт процента — это одна десятая процента. — *Прим. ред.*

пускаемую им продукцию. Рассматривается задача принятия решений на десятилетний период. Фирма оценивает, что на протяжении этих 10 лет вероятность высокого и низкого спроса на производимую продукцию будет равна 0.75 и 0.25 соответственно. Стоимость немедленного строительства крупного предприятия равна 5 миллионов долларов, а небольшого — один миллион долларов. Расширение малого предприятия через два года обойдется фирме в 4.2 миллиона долларов. Прибыль, получаемая от функционирования производственных мощностей на протяжении 10 лет, приводится в следующей таблице.

	Ожидаемый доход за год (тысячи долл.)	
Альтернатива	Высокий спрос	Низкий спрос
Крупное предприятие сейчас	1000	300
Небольшое предприятие сейчас	250	200
Расширенное предприятие через 2 года	900	200

a) Постройте соответствующее дерево решений, принимая во внимание, что через два года фирма может либо расширить небольшое предприятие, либо не расширять его.

b) Сформулируйте стратегию строительства для фирмы на планируемый 10-летний период. (Для простоты не принимайте во внимание возможную инфляцию.)

10. Решите предыдущее упражнение в предположении, что ежегодная учетная ставка равна 10% и что решение принимается с учетом инфляции. (*Совет.* Для решения задачи необходимы таблицы сложных процентных ставок.)

11. Решите упр. 9 в предположении, что спрос может быть высоким, средним и низким с вероятностями 0.7, 0.2 и 0.1 соответственно. Расширение небольшого предприятия будет проведено лишь в том случае, если на протяжении первых двух лет спрос будет высоким. Следующая таблица содержит данные о прибылях за год.

	Ожидаемый доход за год (тысячи долл.)		
Альтернатива	Высокий спрос	Средний спрос	Низкий спрос
Крупное предприятие сейчас	1000	500	300
Небольшое предприятие сейчас	400	280	150
Расширенное предприятие через 2 года	900	600	200

Более сложные ситуации выбора решений. Для демонстрации других возможностей применения критерия ожидаемого значения рассмотрим ситуации принятия решений, в которых плата является математической функцией альтернативных решений. В этом случае представление задачи в виде дерева решений хотя и является возможным, но может быть не столь полезным, как в предыдущих примерах.

Пример 14.3–2

Электроэнергетическая компания использует парк из 20 грузовых автомобилей для обслуживания электрической сети. Компания планирует периодический профилактический ремонт автомобилей. Вероятность p_t поломки автомобиля по истечении t месяцев после профилактического ремонта оценивается следующим образом.

t	1	2	3	4	5	6	7	8	9	≥ 10
p_t	0.05	0.07	0.10	0.13	0.18	0.23	0.33	0.43	0.50	0.55

Случайная поломка одного грузового автомобиля обходится компании в 200 долларов, а планируемый профилактический ремонт в 50 долларов. Необходимо определить оптимальный период (в месяцах) между планируемыми профилактическими ремонтами.

Обозначим через N искомое число месяцев между профилактическими ремонтами. На протяжении N-месячного цикла могут иметь место два вида расходов: 1) затраты, связанные с устранением поломки автомобиля на протяжении первых $N-1$ месяцев и 2) затраты на профилактический ремонт в конце цикла. Затраты второго вида (профилактический ремонт) составляют $\$50 \times 20$ автомобилей, т.е. 1000 долларов на цикл. Затраты, связанные с устранением поломок автомобилей, должны основываться на среднем количестве автомобилей, вышедших из строя на протяжении первых $N-1$ месяцев цикла. Здесь мы имеем два состояния по истечении месяца t: поломка автомобиля с вероятностью p_t и ее отсутствие с вероятностью $1-p_t$. Следовательно, ожидаемое число поломок по истечении месяца t равно количеству автомашин в парке, умноженному на p_t, т.е. $20p_t$. Используя этот результат, подсчитаем ожидаемое общее число сломавшихся автомобилей на протяжении первых $N-1$ месяцев цикла в виде суммы соответствующих величин для каждого месяца в отдельности, т.е. $20p_1 + 20p_2 + \ldots + 20p_{N-1} = 20(p_1 + p_2 + \ldots + p_{N-1})$. Обозначив через $EC(N)$ общую ожидаемую стоимость для цикла между профилактическими ремонтами, имеем следующее.

$$EC(N) = \$1000 + \$200 \times 20(p_1 + p_2 + \ldots + p_{N-1}).$$

Задача выбора решения компанией сводится таким образом к определению длины цикла N, которая минимизирует общие ожидаемые затраты за один месяц $ECPM(N)$, т.е. величину

$$ECPM(N) = \frac{EC(N)}{N} = \frac{1000 + 4000\sum_{t=1}^{N-1} p_t}{N}.$$

Минимизацию функции $ECPM(N)$ нельзя выполнить в явной форме. Вместо этого используется следующая табличная форма нахождения решения.

	N	p_N	$\sum_{t=1}^{N-1} p_t$	$ECPM(N)$
	1	0.05	0.00	\$1000.00
	2	0.07	0.05	600.00
	3	0.10	0.12	493.33
Оптимальное $N \rightarrow$	4	0.13	0.22	470.00
	5	0.18	0.35	480.00
	6	0.23	0.53	520.00

Вычисления показывают, что $ECPM(N)$ достигает своего минимума при $N = 4$. Следовательно, профилактический ремонт автомобилей нужно выполнять каждые четыре месяца.

Задачу выбора решения в примере 14.3–2 можно также представить в виде дерева решений. Вам предлагается сделать это в упр. 14.3,b(1).

Упражнения 14.3,b

1. В задаче из примера 14.3–2 стоимость профилактического ремонта одного автомобиля равна 75 долларов, а стоимость устранения поломки — 200 долларов. Вероятность поломки автомобиля в первый месяц равна 0.03 и увеличивается на 0.01 для каждого последующего месяца, по десятый включительно. Начиная с одиннадцатого месяца и далее, вероятность поломки сохраняется постоянной на уровне 0.13.

   a) Постройте соответствующее дерево решений.

   b) Определите оптимальную длину цикла для профилактического ремонта.

2. Ежедневный спрос на булочки в продовольственном магазине задается следующим распределением вероятностей.

n	100	150	200	250	300
p_n	0.20	0.25	0.30	0.15	0.10

   Магазин покупает булочку по 55 центов, а продает по 1.20 доллара. Если булочка не продана в тот же день, то к концу дня она может быть реализована за 25 центов. Величина запаса булочек может принимать одно из возможных значений спроса, которые перечислены выше.

   a) Постройте соответствующее дерево решений.

   b) Сколько булочек необходимо заказывать ежедневно?

3. Пусть в предыдущем упражнении временной интервал, для которого необходимо решить задачу принятия решений, составляет два дня. Альтернативы для второго дня зависят от объема реализации булочек в первый день. Если реализован в точности весь запас первого дня, магазин закажет такое же количество булочек и на второй день. Если потребность в булочках в первый день превышает имеющийся запас, то для второго дня магазин может заказать любой из объемов спроса на булочки, который превышает запас первого дня. И наконец, если в первый день реализовано меньше булочек, чем было закуплено, то для второго дня магазин может заказать любой из объемов спроса на булочки, который меньше запаса первого дня. Постройте соответствующее дерево решений и определите оптимальную стратегию заказа.

4. Автомат производит α тысяч единиц некоего продукта ежедневно. Если α увеличивается, доля брака p, будучи случайной величиной, возрастает в соответствии со следующей функцией плотности распределения:

$$f(p)=\begin{cases}\alpha p^{\alpha-1}, & 0\le p\le 1,\\ 0 & \text{в противном случае.}\end{cases}$$

Каждое бракованное изделие приносит убыток в 50 долларов, а качественное изделие — прибыль в 5 долларов.

a) Объясните, почему трудно построить дерево решений для этой задачи.

b) Определите значение α, при котором ожидаемая прибыль принимает максимальное значение.

5. Наружный диаметр d цилиндра, производимого автоматом, имеет верхнее и нижнее допустимые значения $d + t_U$ и $d - t_L$ соответственно. Производственный процесс настроен так, что величина диаметра является нормально распределенной случайной величиной с математическим ожиданием μ и стандартным отклонением σ. Каждый цилиндр со значением диаметра, превышающим верхнее допустимое значение, доводится до нужных размеров за c_1 долларов. Цилиндр, диаметр которого меньше установленной нижней нормы, реализуется с убытком c_2 долларов. Определите оптимальное значение настройки для автомата.

6. В производственном процессе периодически производится заточка режущих инструментов. Если инструмент не затачивается достаточно часто, то возрастает процент брака, и наоборот. Пусть S_U и S_L — соответственно верхний и нижний допустимые пределы измеримого размера изделия, обрабатываемого с помощью режущего инструмента в данном производственном процессе. Далее пусть $\mu(t)$ — среднее значение результата производственного процесса (например, размера изделия, получаемого в результате обработки режущим инструментом) после того, как прошло t единиц времени после последней заточки инструмента; здесь $\mu(0)$ соответствует идеальной начальной настройке инструмента. Стоимость заточки инструмента равна c_1, а стоимость бракованного изделия — c_2. Продукция производится партиями объемом Q со скоростью r единиц продукции в единицу времени, и результат процесса описывается нормальным распределением со средним значением $\mu(t)$ и не зависящим от времени стандартным отклонением σ.

a) Найдите выражение для ожидаемой стоимости заточки инструмента и переделки дефектных изделий как функцию интервала времени T между последовательными заточками.

b) Покажите, что оптимальное значение T не зависит от Q, и объясните это.

c) Найдите численное значение T при следующих данных: $c_1 = 10$ долларов, $c_2 = 48.85$ долларов, $r = 10$ единиц продукции за час, $\mu(t) = \mu(0) + t$ и $\sigma = 1$.

(*Совет.* Аппроксимируйте число заточек инструмента при изготовлении партии объемом Q с помощью числа Q/rT. Кроме того, используйте численное интегрирование для нахождения оптимального значения T.)

7. *Критерий предельного уровня.* Фирма для технических целей использует в одном из своих производственных процессов химические препараты (химикалии). Срок годности этих препаратов составляет один месяц, после чего оставшаяся их часть уничтожается. Объем используемых фирмой химических препаратов (в галлонах) является случайной величиной, изменяющейся в соответствии со следующим распределением.

$$f(x)=\begin{cases}\frac{200}{x^2}, & 100\le x\le 200,\\ 0, & \text{в противном случае.}\end{cases}$$

Реально химикалии используются на протяжении месяца в соответствии с равномерным распределением. Фирма планирует определить количество химических препаратов, удовлетворяющих двум конфликтующим критериям (или предельным уровням).

1) Среднее число оставшихся химикалий не превышает 20 галлонов в месяц.
2) Среднее количество недостающих химикалий не превышает 40 галлонов в месяц.

14.3.2. Другие критерии ожидаемого значения

В этом разделе рассматриваются три модификации критерия ожидаемого значения. Первая состоит в определении *апостериорных вероятностей* на основе эксперимента над исследуемой системой, вторая — в *полезности* реальной стоимости денег, а третья модифицирует критерий ожидаемого значения таким образом, что он может быть использован для принятия решений при краткосрочном планировании.

Апостериорные вероятности Байеса. Распределения вероятностей, которые используются при формулировке критерия ожидаемого значения, получаются, как правило, из накопленной ранее информации (см. раздел 12.6). В некоторых случаях оказывается возможным модифицировать эти вероятности с помощью текущей и/или полученной ранее информации, которая обычно основывается на исследовании выборочных (или экспериментальных) данных. Получаемые при этом вероятности называют **апостериорными** (или **Байесовскими**), в отличие от **априорных**, полученных из исходной информации. Следующий пример показывает, как рассмотренный в разделе 14.3.1 критерий ожидаемого значения можно модифицировать так, чтобы воспользоваться новой информацией, содержащейся в апостериорных вероятностях.

Пример 14.3–3

В примере 14.3–1 априорные вероятности 0.6 и 0.4 повышения и понижения котировок акций на бирже были определены из наличных публикаций финансового характера. Предположим, вместо того чтобы полностью полагаться на эти публикации, вы решили провести личное исследование путем консультаций с другом, который хорошо разбирается в вопросах, касающихся фондовой биржи. Друг высказывает общее мнение "за" или "против" инвестиций. Это мнение в дальнейшем определяется количественно следующим образом. При повышении котировок его мнение с 90%-ной вероятностью будет "за", при снижении котировок вероятность его мнения "за" уменьшится до 50%. Каким образом можно извлечь пользу из этой дополнительной информации?

Мнение друга фактически представляет условные вероятности "за–против" при заданных состояниях природы в виде повышения и понижения котировок. Введем следующие обозначения:

v_1 — мнение "за",
v_2 — мнение "против",

m_1 — повышение котировок,
m_2 — понижение котировок.

Мнение друга можно записать в виде вероятностных соотношений следующим образом.

$$P\{v_1 \mid m_1\} = 0.9,\ P\{v_1 \mid m_2\} = 0.1,$$
$$P\{v_2 \mid m_1\} = 0.5,\ P\{v_2 \mid m_2\} = 0.5.$$

С помощью этой дополнительной информации задачу выбора решения можно сформулировать следующим образом.

1. Если мнение друга "за", акции какой компании следует покупать — *A* или *B*?
2. Если мнение друга "против", то, опять-таки, — акции какой компании следует покупать — *A* или *B*?

Рассматриваемую задачу можно представить в виде дерева решений, показанного на рис. 14.4. Узлу 1 здесь соответствует случайное событие, мнение друга, с соответствующими вероятностями "за" и "против". Узлы 2 и 3 представляют выбор между компаниями *A* и *B* при известном мнении друга "за" или "против" соответственно. Узлы 4–7 соответствуют случайным событиям, связанным с повышением и понижением котировок.

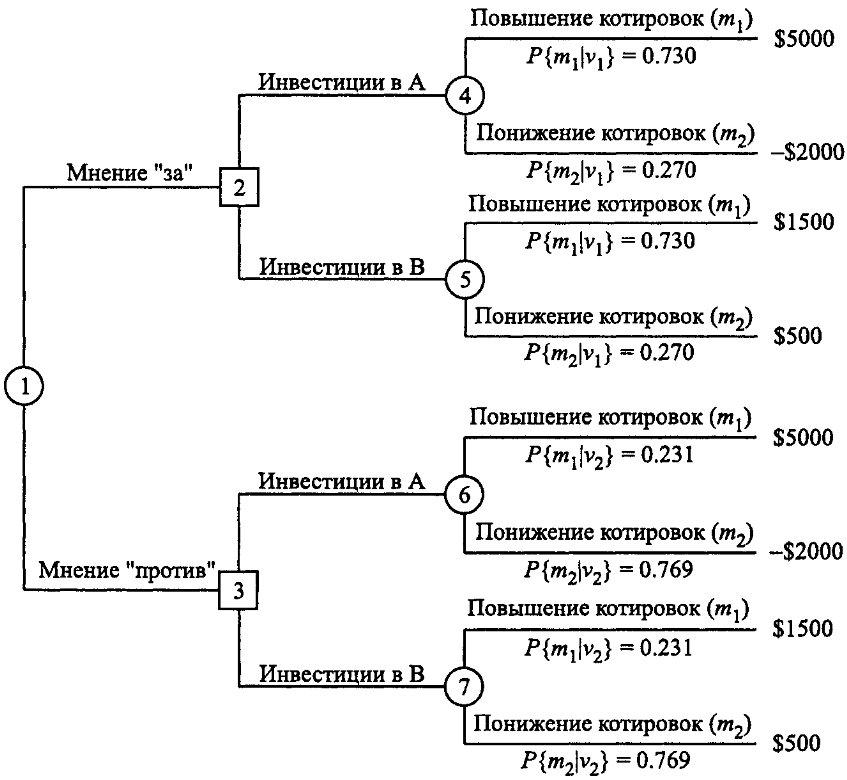


Рис. 14.4

Для оценки различных альтернатив, показанных на рис. 14.4, необходимо вычислить апостериорные вероятности $P\{m_i | v_j\}$, указанные на соответствующих ветвях, выходящих из узлов 4–7. Эти апостериорные вероятности вычисляются с учетом дополнительной информации, содержащейся в рекомендациях друга, с помощью следующих действий.

Шаг 1. Условные вероятности $P\{v_j | m_i\}$ для данной задачи запишем следующим образом.

$P\{v_j | m_i\} =$

	v_1	v_2
m_1	0.9	0.1
m_2	0.5	0.5

Шаг 2. Вычисляем вероятности совместного появления событий.

$$P\{m_i, v_j\} = P\{v_j | m_i\}P\{m_i\} \text{ для всех } i \text{ и } j.$$

При заданных *априорных* вероятностях $P\{m_1\} = 0.6$ и $P\{m_2\} = 0.4$ вероятности совместного появления событий определяются умножением первой и второй строк таблицы, полученной на шаге 1, на 0.6 и 0.4 соответственно. В результате имеем следующее.

$P\{m_i, v_j\} =$

	v_1	v_2
m_1	0.54	0.06
m_2	0.20	0.20

Сумма всех элементов этой таблицы равна 1.

Шаг 3. Вычисляем абсолютные вероятности.

$$P\{v_j\} = \sum_{\text{все } i} P\{m_i, v_j\}, \text{ для всех } j.$$

Эти вероятности получаются путем суммирования элементов соответствующих столбцов таблицы, полученной на шаге 2. В итоге имеем следующее.

$P\{v_1\}$	$P\{v_2\}$
0.74	0.26

Шаг 4. Определяем искомые апостериорные вероятности по формуле

$$P\{m_i | v_j\} = \frac{P\{m_i, v_j\}}{P\{v_j\}}.$$

Эти вероятности вычисляются в результате деления каждого столбца таблицы, полученной на шаге 2, на элемент соответствующего столбца таблицы, вычисленной на шаге 3, что приводит к следующим результатам (округленным до трех десятичных знаков).

$P\{v_j | m_i\} =$

	v_1	v_2
m_1	0.730	0.231
m_2	0.270	0.769

Это те вероятности, которые показаны на рис. 14.4. Они отличаются от исходных априорных вероятностей $P\{m_1\} = 0.6$ и $P\{m_2\} = 0.4$.

Теперь можно оценить альтернативные решения, основанные на ожидаемых платежах для узлов 4–7.

Мнение "за"

Доход от акций компании A в узле 4 = 5000 × 0.730 + (–2000) × 0.270 = **$3110**.
Доход от акций компании B в узле 5 = 1500 × 0.730 + 500 × 0.270 = $1230.
Решение. Инвестировать в акции компании A.

Мнение "против"

Доход от акций компании A в узле 6 = 5000 × 0.231 + (–2000) × 0.769 = –$383.
Доход от акций компании B в узле 7 = 1500 × 0.231 + 500 × 0.769 = **$731**.

Решение. Инвестировать в акции компании B.

Заметим, что предыдущие решения эквивалентны утверждению, что ожидаемые платы в узлах 2 и 3 равны 3110 и 731 доллар соответственно (рис. 14.4). Следовательно, при известных вероятностях $P\{v_1\} = 0.74$ и $P\{v_2\} = 0.26$, вычисленных на шаге 3, можно вычислить ожидаемую плату для всего дерева решений (упр. 14.3,c(3)).

Упражнения 14.3,c

1. Несмотря на сезон дождей, Джим Боб планирует завтра идти на рыбалку, но только если не будет дождя. Из данных о погоде прошлых лет следует, что имеется 70%-ная вероятность, что в сезон дождей будет идти дождь. В шесть часов вечера синоптики предсказали с 85%-ной вероятностью, что завтра будет дождь. Следует ли Джиму Бобу планировать рыбалку на завтра?

2. Фирма Электра получает 75% электронных деталей от поставщика A и 25% — поставщика B. Доля брака в продукции поставщиков A и B составляет 1% и 2% соответственно. При проверке пяти деталей из полученной партии обнаружена лишь одна дефектная. Определите вероятность того, что партия получена от поставщика A. Проведите аналогичные вычисления относительно поставщика B. (*Подсказка*. Вероятность появления бракованной детали в партии подчиняется биномиальному закону распределения.)

3. Предположим, что в задаче из примера 14.3–3 есть дополнительный выбор, связанный с инвестированием 10 000 долларов в надежный депозит, который приносит 8% прибыли. Совет вашего друга, по-прежнему, относится к инвестированию через биржу.

   a) Постройте соответствующее дерево решений.

   b) Какое оптимальное решение в этом случае? (*Совет*. Используйте вероятности $P\{v_1\}$ и $P\{v_2\}$, полученные на шаге 3 в примере 14.3–3, для вычисления ожидаемой суммы инвестирования через биржу.)

4. Допустим, вы являетесь автором романа, который обещает быть популярным. У вас имеется возможность либо самостоятельно напечатать роман, либо через издательство. Издательство предлагает вам 20 000 долларов за подписание контрак-

та. Если роман будет пользоваться спросом, будет продано 200 000 экземпляров, в противном случае — лишь 10 000 экземпляров. Издательство выплачивает авторский гонорар в сумме один доллар за экземпляр. Исследование рынка, проведенное издательством, свидетельствует о том, что существует 70%-ная вероятность, что роман будет популярным. Если же вы сами напечатаете роман, то понесете потери в сумме 90 000 долларов, связанные с печатанием и маркетингом, но в этом случае каждый проданный экземпляр принесет вам прибыль в два доллара.

a) Принимая во внимание имеющуюся информацию, примете ли вы предложение издательства или будете печатать роман самостоятельно?

b) Предположим, что вы заключили договор с литературным агентом на исследование, связанное с потенциальным успехом романа. Исходя из предыдущего опыта, компания извещает вас, что если роман будет пользоваться спросом, то исследование предскажет неверный результат в 20% случаев. Если же роман будет не популярен, то исследование предскажет верный результат в 85% случаев. Как эта информация повлияет на ваше решение?

5. Вернитесь к проблеме выбора решения фермером Мак-Коем из упр. 14.3,a(2). Фермер имеет дополнительный выбор, связанный с использованием земли как пастбища, что гарантированно принесет ему прибыль в 7500 долларов. Фермер получил также дополнительную информацию от брокера, касающуюся степени стабильности будущих цен на продукцию. Оценки брокера "благоприятный — неблагоприятный" затем выражаются количественно в виде следующих условных вероятностей.

$P\{a_j \mid s_i\} =$

	a_1	a_2
s_1	0.15	0.85
s_2	0.50	0.50
s_3	0.85	0.15

В данном случае a_1 и a_2 — оценки брокера "благоприятный" и "неблагоприятный", а s_1, s_2 и s_3 представляют изменение в будущих ценах: соответственно "понижение", "такие же", "повышение".

a) Постройте соответствующее дерево решений.

b) Найдите оптимальное решение задачи.

6. Пусть в упр. 14.3,a(5) дирекция компании решила провести пробную продажу своей продукции в выбранных населенных пунктах. Результатом пробной продажи являются оценки "хорошо" (a_1) или "плохо" (a_2). Тест дает следующие условные вероятности с проведением рекламной кампании и без нее.

$P\{a_j \mid v_i\}$ с рекламной кампанией

	a_1	a_2
v_1	0.95	0.05
v_2	0.3	0.7

$P\{a_j \mid v_i\}$ без рекламной кампании

	a_1	a_2
w_1	0.8	0.2
w_2	0.4	0.6

Здесь символы v_1 и v_2 обозначают, соответственно, "успех" и "неуспех", а w_1 и w_2 — "восприимчивый" и "невосприимчивый" покупатель.

a) Постройте соответствующее дерево решений.

b) Определите оптимальный план действий фирмы.

7. Статистические данные о работе компании показывают, что с вероятностью 5% произведенная партия продукции будет неприемлемой (плохой). Плохая партия содержит 15% дефектных изделий, а хорошая — лишь 4%. Пусть значение переменной $a = a_1$ ($= a_2$) обозначает, что партия изделий является хорошей (плохой). Тогда соответствующие априорные вероятности равны соответственно $P\{a = a_1\} = 0.95$ и $P\{a = a_2\} = 0.05$.

   Вместо того чтобы отправить партии продукции с характеристиками, основанными на априорных вероятностях, из каждой партии проверяются два изделия. Возможны следующие результаты проверки.

   Оба изделия являются доброкачественными (s_1).

   Одно изделие является доброкачественным (s_2).

   Оба изделия являются бракованными (s_3).

   a) Определите апостериорные вероятности $P\{a_i \mid s_j\}$, $i = 1, 2; j = 1, 2, 3$.

   b) Предположим, что фирма отправляет партии продукции двум потребителям A и B. Контракты с ними определяют, что процент бракованных изделий в поставках не должен превышать 5% и 8% соответственно. Предусматривается штраф в 100 долларов за превышение на один процент максимально допустимого лимита бракованных изделий. Поставка партий лучшего качества, чем указано в контракте, приносит производителю прибыль в 80 долларов за каждый процент уменьшения доли бракованных изделий. Постройте соответствующее дерево решений и определите приоритетную стратегию отправки партий продукции.

Функции полезности. В предыдущих примерах критерий ожидаемого значения применялся лишь в тех ситуациях, где платежи выражались в виде *реальных* денег. Имеются многочисленные случаи, когда при анализе следует использовать скорее *полезность*, чем реальную величину платежей. Для демонстрации этого предположим, что имеется шанс 50 на 50, что инвестиция в 20 000 долларов или принесет прибыль в 40 000 долларов, или будет полностью потеряна. Соответствующая ожидаемая прибыль равна $40000 \times 0.5 - 20000 \times 0.5 = 10000$ долларов. Хотя здесь ожидается прибыль в виде чистого дохода, разные люди могут по-разному интерпретировать полученный результат. Инвестор, который идет на риск, может сделать инвестицию, чтобы с вероятностью 50% получить прибыль в 40 000 долларов. Наоборот, осторожный инвестор может не выразить желания рисковать потерей 20 000 долларов. С этой точки зрения очевидно, что разные индивидуумы проявляют разное отношение к риску, т.е. они проявляют разную *полезность* по отношению к риску.

Определение полезности является субъективным. Оно зависит от нашего отношения к риску. В этом разделе мы представляем систематизированную процедуру числовой оценки отношения к риску лица, принимающего решение. Конечным результатом является функция полезности, которая занимает место реальных денег.

В примере, приведенном выше, наилучший платеж равен \$40 000, а наихудший — -\$20 000. Следовательно, мы устанавливаем произвольную, но логическую шкалу полезности U, изменяющуюся от 0 до 100, где 0 соответствует полезности –\$20 000, а 100 — \$40000, т.е. $U(-20000) = 0$ и $U(40000) = 100$. Далее определяем полезность в точках между –\$20000 и \$40000 для определения общего вида функции полезности.

Если отношение лица, принимающего решение, беспристрастно к риску, то результирующая функция полезности является прямой линией, соединяющей точки (0, –\$20000) и (100, \$40000). В этом случае как реальные деньги, так и их полезность дают совпадающие решения. В более реальных ситуациях функция полезности может принимать другой вид, отражающий отношение к риску лица, принимающего решение. Рис. 14.5 иллюстрирует вид функции полезности для трех индивидуумов X, Y и Z. Индивидуум X **не расположен к риску** (осторожен), так как проявляет бóльшую чувствительность к потере, чем к прибыли. Индивидуум Z — противоположность в этом отношении индивиду X; он **настроен на риск**. Это следует из того, что для индивидуума X при изменении в 10 000 долларов вправо и влево от точки, соответствующей 0 долларов, увеличение прибыли изменяет полезность на величину ab, которая меньше изменения полезности bc, обусловленной потерями такой же величины, т.е. $ab < bc$. В то же время такие же изменения в $\pm$10000 долларов, относящиеся к индивидууму Z, обнаруживают противоположное поведение; здесь $de > ef$. Далее, индивидуум Y является **нейтральным к риску**, так как упомянутые изменения порождáют одинаковые изменения полезности. В общем случае индивидуум может быть как не расположен к риску, так и настроен на риск, в зависимости от суммы риска. В этом случае соответствующая кривая полезности будет иметь вид удлиненной буквы S.

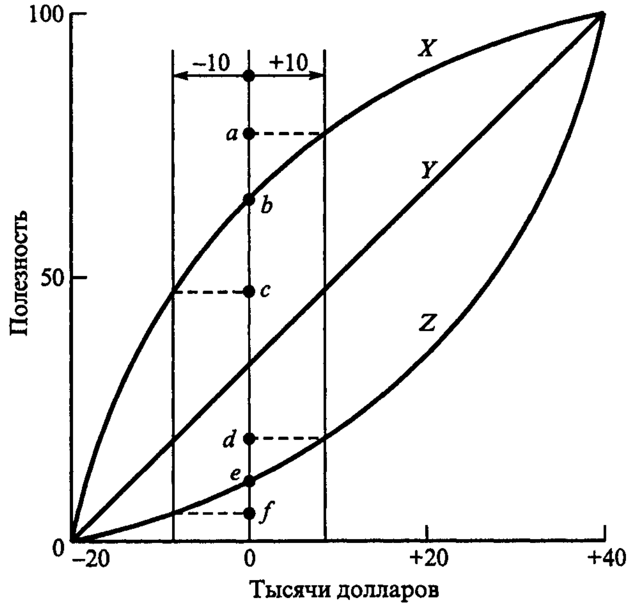


Рис. 14.5

Кривые полезности, аналогичные изображенным на рис. 14.5, определены с помощью количественного показателя, характеризующего отношение к риску лица, принимающего решение, для различных значений уровня реальных денег в пределах установленного интервала. Так в рассмотренном примере установленным интервалом является (–\$20000, \$40000), соответствующая полезность изменяется в интервале (0, 100). Необходимо определить полезность, соответствующую таким промежуточным значениям, например, как –\$10 000, \$0, \$10 000, \$20 000 или \$30 000. Соответствующая процедура построения функции полезности начинается с того, что организовывается **лотерея** для определения суммы реальных денег x, для которой ожидаемое значение полезности будет вычислено по следующей формуле.

$$U(x) = pU(-20000) + (1-p)U(40000) = 0p + 100(1-p) = 100 - 100p, \quad 0 \le p \le 1.$$

Для определения значения $U(x)$ просят лицо, принимающее решение, сообщить свое предпочтение между *гарантированной* наличной суммой x и возможностью сыграть в лотерею, в которой с вероятностью p реализуется проигрыш в сумме \$20000 и с вероятностью $1-p$ имеет место выигрыш в \$40000. При этом под предпочтением понимается выбор значения "нейтральной" вероятности p, при котором с точки зрения лица, принимающего решение, возможности сыграть в лотерею и получить гарантированную сумму x являются одинаково привлекательными. Например, если x = \$20000, лицо, принимающее решение, может заявить, что гарантированные 20000 долларов наличными и лотерея одинаково привлекательны при $p = 0.8$. В этом случае вычисляется полезность для x = \$20000 по следующей формуле.

$$U(20000) = 100 - 100 \times 0.8 = 20.$$

Эта процедура продолжается до тех пор, пока не будет получено достаточное количество точек $(x, U(x))$ для определения формы функции полезности. Затем можно определить искомую функцию полезности путем регрессионного анализа или просто линейной интерполяции между полученными точками.

Хотя здесь применяется количественная процедура для определения функции полезности, сам подход далек от того, чтобы быть научно обоснованным. То, что процедура полностью определяется мнением лица, принимающего решение, порождает сомнения относительно надежности описанного процесса. Процедура, в частности, неявно предполагает, что лицо, принимающее решение, является рационально мыслящим — требование, которое не всегда может быть согласовано с вариациями в поведении и настроении, что является типичным для человеческой личности. В этом отношении лицо, принимающее решение, должно придерживаться концепции полезности в широком смысле, в соответствии с которой денежные величины не должны быть единственным решающим фактором в теории принятия решений.

Упражнения 14.3,d

1. Допустим, вы — студент университета штата Арканзас и имеете сильное желание присутствовать на следующем баскетбольном матче. Проблема в том, что входной билет стоит 10 долларов, а у вас есть лишь 5 долларов. Вы можете рискнуть 5 долларами в игре в покер с шансами 50 на 50 удвоить свою сумму или совсем ее проиграть.

   a) Будете ли вы, исходя из реальной стоимости денег, искушать судьбу, играя в покер?

   b) Учитывая ваше сильное желание присутствовать на матче, переведите наличные деньги в функцию полезности.

c) Основываясь на функции полезности, которую вы построили, примете ли вы участие в игре в покер?

2. Семья переехала в местность, где возможны землетрясения, и собирается построить дом. Решается вопрос, стоит ли строить дом в соответствии с высокими стандартами, рассчитанными на сейсмическую зону. Строительство дома в соответствии с такими стандартами обойдется в 850 000 долларов, а без их учета — в 350 000 долларов. В случае землетрясения (его вероятность равна 0.001) восстановление дома, построенного без соответствующих стандартов, обойдется в 900 000 долларов. Примените в этой ситуации рассмотренную выше процедуру использования лотереи, предполагая, что шкала полезности изменяется от 0 до 100.

3. Инвестиция в 10 000 долларов в предприятие с высоким уровнем риска имеет шанс 50 на 50 увеличить эту сумму до 14 000 долларов на протяжении следующего года либо уменьшить ее до 8 000 долларов. Это значит, что чистый доход составит либо 4000 долларов, либо –2000 долларов.

a) Принимая позицию нейтрального к риску инвестора и шкалу полезности от 0 до 100, определите полезность 0 долларов чистого дохода и соответствующую "нейтральную" вероятность.

b) Пусть два инвестора A и B определили следующие "нейтральные" вероятности.

	Вероятность	
Чистая прибыль ($)	Инвестор A	Инвестор B
–2000	1.00	1.00
–1000	0.30	0.90
0	0.20	0.80
1000	0.15	0.70
2000	0.10	0.50
3000	0.05	0.40
4000	0.00	0.00

Нарисуйте графики функций полезности для инвесторов A и B и охарактеризуйте их отношение к риску.

c) Пусть инвестор A имеет возможность сделать инвестицию в одно из двух рискованных предприятий: I или II. Инвестиция в предприятие I может принести прибыль в сумме 3000 долларов с вероятностью 0.4 или убыток в 1000 долларов с вероятностью 0.6. Инвестиция в предприятие II может принести прибыль в 2000 долларов с вероятностью 0.6 или вовсе не принести прибыли с вероятностью 0.4. Используя функцию полезности инвестора A, построенную в предыдущем пункте, и критерий ожидаемой полезности, определите предприятие, которое следует выбрать инвестору A. Каково ожидаемое денежное значение, соответствующее выбранному предприятию (используйте линейную интерполяцию функции полезности)?

d) Повторите упражнение предыдущего пункта для инвестора B.

14.4. Принятие решений в условиях неопределенности

Принятие решений в условиях неопределенности, как и в условиях риска, требует определения альтернативных действий, которым соответствуют платежи, зависящие от (случайных) *состояний природы*. Матрицу платежей в задаче принятия решений с m возможными действиями и n состояниями природы можно представить следующим образом.

	s_1	s_2	...	s_n
a_1	$v(a_1, s_1)$	$v(a_1, s_2)$	...	$v(a_1, s_n)$
a_2	$v(a_2, s_1)$	$v(a_2, s_2)$	...	$v(a_2, s_n)$
.	.	.	.	.
.	.	.	.	.
a_m	$v(a_m, s_1)$	$v(a_m, s_2)$	...	$v(a_m, s_n)$

Элемент a_i представляет i-е возможное решение, а элемент s_j — j-е состояние природы. Плата (или доход), связанная с решением a_i и состоянием s_j, равна $v(a_i, s_j)$.

Отличие между принятием решений в условиях риска и неопределенности состоит в том, что в условиях неопределенности вероятностное распределение, соответствующее состояниям s_j, $j = 1, 2, ..., n$, либо неизвестно, либо не может быть определено. Этот недостаток информации обусловил развитие следующих критериев для анализа ситуации, связанной с принятием решений.

1. Критерий Лапласа.
2. Минимаксный критерий.
3. Критерий Сэвиджа.
4. Критерий Гурвица.

Эти критерии отличаются по степени консерватизма, который проявляет индивидуум, принимающий решение, перед лицом неопределенности.

Критерий Лапласа опирается на **принцип недостаточного основания**[1], который гласит, что поскольку распределение вероятностей состояний $P(s_j)$ неизвестно, нет причин считать их различными. Следовательно, используется *оптимистическое* предположение, что вероятности всех состояний природы равны между собой, т.е. $P\{s_1\} = P\{s_2\} = ... = P\{s_n\} = 1/n$. Если при этом $v(a_i, s_j)$ представляет получаемую прибыль, то наилучшим решением является то, которое обеспечивает

$$\max_{a_i}\left\{\frac{1}{n}\sum_{j=1}^{n} v\left(a_i, s_j\right)\right\}.$$

Если величина $v(a_i, s_j)$ представляет расходы лица, принимающего решение, то оператор "max" заменяется на "min".

Максиминный (минимаксный) критерий основан на консервативном осторожном поведении лица, принимающего решение, и сводится к выбору наилучшей альтернативы из наихудших. Если величина $v(a_i, s_j)$ представляет получаемую прибыль, то в соответствии с *максиминным* критерием в качестве оптимального выбирается решение, обеспечивающее

[1] Этот принцип впервые сформулирован Я. Бернулли. — *Прим. перев.*

$$\max_{a_i}\left\{\min_{s_j} v(a_i, s_j)\right\}.$$

Если величина $v(a_i, s_j)$ представляет потери, используется *минимаксный* критерий, который определяется следующим соотношением.

$$\min_{a_i}\left\{\max_{s_j} v(a_i, s_j)\right\}.$$

Критерий Сэвиджа стремится смягчить консерватизм *минимаксного* (*максиминного*) критерия путем замены матрицы платежей (выигрышей или проигрышей) $v(a_i, s_j)$ матрицей *потерь* $r(a_i, s_j)$, которая определяется следующим образом.

$$r(a_i, s_j) = \begin{cases} \max\limits_{a_k}\left\{v(a_k, s_j)\right\} - v(a_i, s_j), & \text{если } v - \text{доход}, \\ v(a_i, s_j) - \min\limits_{a_k}\left\{v(a_k, s_j)\right\}, & \text{если } v - \text{потери}. \end{cases}$$

Чтобы показать, как критерий Сэвиджа "смягчает" минимаксный (максиминный) критерий, рассмотрим следующую матрицу платежей $v(a_i, s_j)$:

	s_1	s_2	Максимум строк
a_1	\$11 000	\$90	\$11 000
a_2	\$10 000	\$10 000	**\$10 000** ← Минимакс

Применение минимаксного критерия приводит к тому, что решение a_2 с фиксированными потерями в 10000 долларов является предпочтительным. Однако можно выбрать a_1, так как в этом случае имеется возможность потерять лишь 90 долларов, если реализуется состояние s_2.

Посмотрим, какой результат получится, если в минимаксном критерии вместо матрицы платежей $v(a_i, s_j)$ используем матрицу потерь $r(a_i, s_j)$.

	s_1	s_2	Максимум строк
a_1	\$1000	\$0	**\$1000** ← Минимакс
a_2	\$0	\$9910	\$9910

Как видим, минимаксный критерий, применяемый к матрице потерь, приводит к выбору решения a_1 в качестве предпочтительного.

Рассмотрим теперь **критерий Гурвица**. Этот критерий охватывает ряд различных подходов к принятию решений — от наиболее оптимистичного до наиболее пессимистичного (консервативного). Пусть $0 \le \alpha \le 1$ и величины $v(a_i, s_j)$ представляют доходы. Тогда решению, выбранному по критерию Гурвица, соответствует

$$\max_{a_i}\left\{\alpha \max_{s_j} v(a_i, s_j) + (1-\alpha)\min_{s_j} v(a_i, s_j)\right\}.$$

Параметр α — **показатель оптимизма**. Если $\alpha = 0$, критерий Гурвица становится консервативным, так как его применение эквивалентно применению обычного минимаксного критерия. Если $\alpha = 1$, критерий Гурвица становится слишком оптимистичным, ибо рассчитывает на *наилучшие из наилучших* условий. Мы можем конкретизировать степень оптимизма (или пессимизма) надлежащим выбором величины α из интервала [0, 1]. При отсутствии ярко выраженной склонности к оптимизму или пессимизму выбор $\alpha = 0.5$ представляется наиболее разумным.

Если величины $v(a_i, s_j)$ представляют потери, то критерий принимает следующий вид:

$$\min_{a_i}\left\{\alpha \min_{s_j} v\left(a_i, s_j\right) + \left(1-\alpha\right)\max_{s_j} v\left(a_i, s_j\right)\right\}.$$

Пример 14.4–1

Национальная школа выживания подбирает место для строительства летнего лагеря в центре Аляски в целях тренировки людей на выживание в условиях дикой природы. Школа считает, что число участников сбора может быть 200, 250, 300 или 350 человек. Стоимость летнего лагеря будет минимальной, поскольку он строится для удовлетворения только точно определенных небольших потребностей. Отклонения в сторону уменьшения или увеличения относительно идеальных уровней потребностей влекут за собой дополнительные затраты, обусловленные строительством избыточных (неиспользуемых) мощностей или потерей возможности получить прибыль в случае, когда некоторые потребности не удовлетворяются. Пусть переменные a_1–a_4 представляют возможные размеры лагеря (на 200, 250, 300 или 350 человек), а переменные s_1–s_4 — соответствующее число участников сбора. Следующая таблица содержит матрицу стоимостей (в тысячах долларов), относящуюся к описанной ситуации.

	s_1	s_2	s_3	s_4
a_1	5	10	18	25
a_2	8	7	12	23
a_3	21	18	12	21
a_4	30	22	19	15

Описанная ситуация анализируется с точки зрения четырех рассмотренных выше критериев.

Критерий Лапласа. При заданных вероятностях $P\{s_j\} = 1/4$, $j = 1, 2, 3, 4$, ожидаемые значения затрат для различных возможных решений вычисляются следующим образом.

$M\{a_1\} = (1/4)(5 + 10 + 18 + 25) = \$14\,500$,
$M\{a_2\} = (1/4)(8 + 7 + 12 + 23) =$ **\$12 500** $\leftarrow$ оптимум,
$M\{a_3\} = (1/4)(21 + 18 + 12 + 21) = \$18\,000$,
$M\{a_4\} = (1/4)(30 + 22 + 19 + 15) = \$21\,500$.

Минимаксный критерий. Этот критерий использует исходную матрицу стоимостей.

	s_1	s_2	s_3	s_4	Максимум строк
a_1	5	10	18	25	25
a_2	8	7	12	23	23
a_3	21	18	12	21	**21** $\leftarrow$ минимакс
a_4	30	22	19	15	30

Критерий Сэвиджа. Матрица потерь определяется посредством вычитания чисел 5, 7, 12 и 15 из элементов столбцов от первого до четвертого соответственно. Следовательно,

	s_1	s_2	s_3	s_4	Максимум строк
a_1	0	3	6	10	10
a_2	3	0	0	8	**8** ← минимакс
a_3	16	11	0	6	16
a_4	25	15	7	0	25

Критерий Гурвица. Результаты вычислений содержатся в следующей таблице.

Альтернатива	Минимум строк	Максимум строк	α(Минимум строки) + $(1-\alpha)$(Максимум строки)
a_1	5	25	$25-20\alpha$
a_2	7	23	$23-16\alpha$
a_3	12	21	$21-9\alpha$
a_4	15	30	$30-15\alpha$

Используя подходящее значение для α, можно определить оптимальную альтернативу. Например, при $\alpha = 0.5$ оптимальными являются либо альтернатива a_1, либо a_2, тогда как при $\alpha = 0.25$ оптимальным является решение a_3.

Упражнения 14.4,а

1. Хенк — прилежный студент, который обычно получает хорошие отметки благодаря, в частности, тому, что имеет возможность повторить материал в ночь перед экзаменом. Перед завтрашним экзаменом Хенк столкнулся с небольшой проблемой. Его сокурсники организовали на всю ночь вечеринку, в которой он не хочет участвовать. Хенк имеет три альтернативы:

   a_1 — участвовать в вечеринке всю ночь,

   a_2 — половину ночи участвовать в вечеринке, а половину — учиться,

   a_3 — учиться всю ночь.

   Профессор, принимающий завтрашний экзамен, непредсказуем в том смысле, что экзамен может быть легким (s_1), средним (s_2) или трудным (s_3). В зависимости от сложности экзамена и времени, затраченного Хенком на повторение, можно ожидать следующие экзаменационные баллы.

	s_1	s_2	s_3
a_1	85	60	40
a_2	92	85	81
a_3	100	88	82

a) Порекомендуйте Хенку, какой выбор он должен сделать (основываясь на каждом из четырех критериев принятия решений в условиях неопределенности).

b) Предположим, что Хенк более заинтересован в оценке (в буквенном выражении), которую он получит на экзамене. Буквенным оценкам от A до D, означающим сдачу экзамена, соответствует 90, 80, 70 и 60 баллов. Иначе при числе баллов ниже 60 студент получает оценку F, которая свидетельствует о том, что экзамен не сдан. Изменит ли такое отношение к оценкам выбор Хенка?

2. В приближении посевного сезона фермер Мак-Кой имеет четыре альтернативы:

a_1 — выращивать кукурузу,

a_2 — выращивать пшеницу,

a_3 — выращивать соевые бобы,

a_4 — использовать землю под пастбища.

Платежи, связанные с указанными возможностями, зависят от количества осадков, которые условно можно разделить на четыре категории:

s_1 — сильные осадки,

s_2 — умеренные осадки,

s_3 — незначительные осадки,

s_4 — засушливый сезон.

Платежная матрица (в тысячах долларов) оценивается следующим образом.

	s_1	s_2	s_3	s_4
a_1	–20	60	30	–5
a_2	40	50	35	0
a_3	–50	100	45	–10
a_4	12	15	15	10

Что должен посеять Мак-Кой?

3. Один из N станков должен быть выбран для изготовления Q единиц определенной продукции. Минимальная и максимальная потребность в продукции равна Q^* и Q^{**} соответственно. Производственные затраты TC_i на изготовление Q единиц продукции на станке i включают фиксированные затраты K_i и удельные затраты c_i на производство единицы продукции и выражаются формулой $Tc_i = K_i + c_iQ$.

a) Решите задачу с помощью каждого из четырех критериев принятия решений в условиях неопределенности.

b) Решите задачу при следующих данных, предполагая, что $1000 \le Q \le 4000$.

Станок i	K_i ($)	c_i ($)
1	100	5
2	40	12
3	150	3
4	90	8

14.5. Теория игр

В теории игр рассматриваются ситуации, связанные с принятием решений, в которых два разумных противника имеют конфликтующие цели. К числу типичных примеров относится рекламирование конкурирующих товаров и планирование военных стратегий противоборствующих армий. Эти ситуации принятия решений отличаются от рассмотренных ранее, где природа не рассматривается в роли недоброжелателя.

В игровом конфликте участвуют два противника, именуемые **игроками**, каждый из которых имеет некоторое множество (конечное или бесконечное) возможных выборов, которые называются **стратегиями**. С каждой парой стратегий связан **платеж**, который один из игроков выплачивает другому. Такие игры известны как **игры двух лиц с нулевой суммой**, так как выигрыш одного игрока равен проигрышу другого. В такой игре достаточно задать результаты в виде платежей для одного из игроков. При обозначении игроков через A и B с числом стратегий m и n соответственно, игру обычно представляют в виде матрицы платежей игроку A:

$$
\begin{array}{c|cccc|}
 & B_1 & B_2 & \ldots & B_n \\
\hline
A_1 & a_{11} & a_{12} & \ldots & a_{1n} \\
A_2 & a_{21} & a_{22} & \ldots & a_{2n} \\
. & . & . & . & . \\
. & . & . & . & . \\
. & . & . & . & . \\
A_m & a_{m1} & a_{m2} & \ldots & a_{mn} \\
\hline
\end{array}
$$

Такое представление матричной игры означает, что если игрок A использует стратегию i, а игрок B — стратегию j, то платеж игроку A составляет a_{ij} и, следовательно, игроку B — $-a_{ij}$.

14.5.1. Оптимальное решение игры двух лиц с нулевой суммой

Поскольку игры берут свое начало в конфликте интересов, оптимальным решением игры является одна или несколько таких стратегий для каждого из игроков, при этом любое отклонение от данных стратегий не улучшает плату тому или другому игроку. Эти решения могут быть в виде единственной **чистой** стратегии или нескольких стратегий, которые являются **смешанными** в соответствии с заданными вероятностями. Рассматриваемые ниже примеры демонстрируют перечисленные случаи.

Пример 14.5–1

Две компании A и B продают два вида лекарств против гриппа. Компания A рекламирует продукцию на радио (A_1), телевидении (A_2) и в газетах (A_3). Компания B, в дополнение к использованию радио (B_1), телевидения (B_2) и газет (B_3), рассылает также по почте брошюры (B_4). В зависимости от умения и интенсивности проведения рекламной кампании, каждая из компаний может привлечь на свою сторону часть клиентов конкурирующей компании. Приведенная ниже матрица характеризует процент клиентов, привлеченных или потерянных компанией A.

	B_1	B_2	B_3	B_4	Минимумы строк
A_1	8	–2	9	–3	–3
A_2	6	5	6	8	**5** ← Максимин
A_3	–2	4	–9	5	–9
Максимумы столбцов	8	**5**	9	8	
		Минимакс			

Решение игры основано на обеспечении *наилучшего результата из наихудших* для каждого игрока. Если компания A выбирает стратегию A_1, то, независимо от того, что предпринимает компания B, наихудшим результатом является потеря компанией A 3% рынка в пользу компании B. Это определяется минимумом элементов первой строки матрицы платежей. Аналогично при выборе стратегии A_2 наихудшим исходом для компании A является увеличение рынка на 5% за счет компании B. Наконец, наихудшим исходом при выборе стратегии A_3 является потеря компанией A 9% рынка в пользу компании B. Эти результаты содержатся в столбце "Минимумы строк". Чтобы достичь наилучшего результата из наихудших, компания A выбирает стратегию A_2, так как она соответствует наибольшему элементу столбца "Минимумы строк".

Рассмотрим теперь стратегии компании B. Так как элементы матрицы являются платежами компании A, критерий *наилучшего результата из наихудших* для компании B соответствует выбору минимаксного значения. В результате приходим к выводу, что выбором компании B является стратегия B_2.

Оптимальным решением игры является выбор стратегий A_2 и B_2, т.е. обеим компаниям следует проводить рекламу на телевидении. При этом выигрыш будет в пользу компании A, так как ее рынок увеличится на 5%. В этом случае говорят, что **цена игры** равна 5% и что компании A и B используют стратегии, соответствующие **седловой точке**.

Решение, соответствующее седловой точке, гарантирует, что ни одной компании нет смысла пытаться выбрать другую стратегию. Действительно, если компания B переходит к другой стратегии (B_1, B_3 или B_4), то компания A может сохранить свой выбор стратегии A_2, что приведет к большей потере рынка компанией B (6% или 8%). По тем же причинам компании A нет резона использовать другую стратегию, ибо если она применит, например, стратегию A_3, то компания B может использовать свою стратегию B_3 и увеличить свой рынок на 9%. Аналогичные выводы имеют место, если компания A будет использовать стратегию A_1.

Оптимальное решение игры, соответствующее седловой точке, не обязательно должно характеризоваться чистыми стратегиями. Вместо этого оптимальное решение может требовать смешивания случайным образом двух или более стратегий, как это сделано в следующем примере.

Пример 14.5–2

Два игрока A и B играют в игру, основанную на подбрасывании монеты. Игроки одновременно и независимо друг от друга выбирают герб (*Г*) или решку (*Р*). Если результаты двух подбрасываний монеты совпадают (т.е. *ГГ* или *РР*), то игрок A получает один доллар от игрока B. Иначе игрок A платит один доллар игроку B.

Следующая матрица платежей игроку A показывает величины минимальных элементов строк и максимальных элементов столбцов, соответствующих стратегиям обоих игроков.

	$B_Г$	B_P	Минимумы строк
$A_Г$	1	–1	–1
A_P	–1	1	–1
Максимумы столбцов	1	1	

Максиминная и минимаксная величины (цены) для этой игры равны –1 доллар и 1 доллар соответственно. Так как эти величины не равны между собой, игра не имеет решения в чистых стратегиях. В частности, если игрок A использует стратегию $A_Г$, игрок B выберет стратегию B_P, чтобы получить от игрока A один доллар. Если это случится, игрок A может перейти к стратегии A_P, чтобы изменить исход игры и получить один доллар от игрока B. Постоянное искушение каждого игрока перейти к другой стратегии указывает на то, что решение в виде чистой стратегии неприемлемо. Вместо этого оба игрока должны использовать надлежащую случайную комбинацию своих стратегий. В рассматриваемом примере оптимальное значение цены игры находится где-то между максиминной и минимаксной ценами для этой игры:

максиминная (нижняя) цена ≤ цена игры ≤ минимаксная (верхняя) цена.

Следовательно, в данном случае цена игры должна лежать в интервале [–1, 1], измеряемом в долларах.

Упражнения 14.5,а

1. Определите решение, определяемое седловой точкой, соответствующие чистые стратегии и цену игры для следующих игр, в которых платежи заданы для игрока A.

a)

	B_1	B_2	B_3	B_4
A_1	8	6	2	8
A_2	8	9	4	5
A_3	7	5	3	5

b)

	B_1	B_2	B_3	B_4
A_1	4	–4	–5	6
A_2	–3	–4	–9	–2
A_3	6	7	–8	–9
A_4	7	3	–9	5

2. В следующих играх заданы платежи игроку A. Укажите область значений для параметров p и q, при которых пара (2, 2) будет седловой точкой в каждой игре.

a)

	B_1	B_2	B_3
A_1	1	q	6
A_2	p	5	10
A_3	6	2	3

b)

	B_1	B_2	B_3
A_1	2	4	5
A_2	10	7	q
A_3	4	p	6

3. Укажите область, которой принадлежит цена игры в каждом из следующих случаев, предполагая, что платежи заданы для игрока A.

a)

	B_1	B_2	B_3	B_4
A_1	1	9	6	0
A_2	2	3	8	4
A_3	–5	–2	10	–3
A_4	7	4	–2	–5

b)

	B_1	B_2	B_3	B_4
A_1	–1	9	6	8
A_2	–2	10	4	6
A_3	5	3	0	7
A_4	7	–2	8	4

c)

	B_1	B_2	B_3
A_1	3	6	1
A_2	5	2	3
A_3	4	2	–5

d)

	B_1	B_2	B_3	B_4
A_1	3	7	1	3
A_2	4	8	0	–6
A_3	6	–9	–2	4

4. Две фирмы производят два конкурирующих товара. Каждый товар в настоящее время контролирует 50% рынка. Улучшив качество товаров, обе фирмы собираются развернуть рекламные кампании. Если они не будут этого делать, то существующее состояние рынка не изменится. Однако если какая-либо фирма будет более активно рекламировать свои товары, то другая фирма потеряет соответствующий процент своих потребителей. Исследование рынка показывает, что 50% потенциальных потребителей получают информацию посредством телевидения, 30% — через газеты и 20% — посредством радио.

   a) Сформулируйте задачу в виде игры двух лиц с нулевой суммой и выберите подходящие средства рекламы для каждой фирмы.

b) Укажите интервал значений, которому принадлежит цена игры. Может ли каждая фирма действовать с единственной чистой стратегией?

5. Пусть a_{ij} — (i, j)-й элемент платежной матрицы с m стратегиями игрока A и n стратегиями игрока B. Элементы платежной матрицы представляют собой платежи игроку A. Докажите, что

$$\max_i \min_j a_{ij} \le \min_j \max_i a_{ij}.$$

14.5.2. Решение матричных игр в смешанных стратегиях

Решение матричных игр в смешанных стратегиях может быть найдено либо графически, либо методами линейного программирования. Графический метод применим для решения игр, в которых хоть один игрок имеет две чистые стратегии. Этот метод интересен в том плане, что графически объясняет понятие седловой точки. Методами линейного программирования может быть решена любая игра двух лиц с нулевой суммой.

Графическое решение игр. Рассмотрим игру $2 \times n$, в которой игрок A имеет две стратегии.

	y_1	y_2	...	y_n
	B_1	B_2	...	B_n
x_1: A_1	a_{11}	a_{12}	...	a_{1n}
$1 - x_1$: A_2	a_{21}	a_{22}	...	a_{2n}

Игра предполагает, что игрок A смешивает стратегии A_1 и A_2 с соответствующими вероятностями x_1 и $1 - x_1$, $0 \le x_1 \le 1$. Игрок B смешивает стратегии B_1, B_2, ..., B_n с вероятностями $y_1, y_2, ..., y_n$, где $y_j \ge 0$, $j = 1, 2, ..., n$, и $y_1 + y_2 + ... + y_n = 1$. В этом случае ожидаемый выигрыш игрока A, соответствующий j-й чистой стратегии игрока B, вычисляется в виде

$$(a_{1j} - a_{2j})\, x_1 + a_{2j}, \quad j = 1, 2, ..., n.$$

Следовательно, игрок A ищет величину x_1, которая максимизирует минимум ожидаемых выигрышей

$$\max_{x_1} \min_j \left\{ \left(a_{1j} - a_{2j}\right) x_1 + a_{2j} \right\}.$$

Пример 14.5–3

Рассмотрим следующую игру 2×4, в которой платежи выплачиваются игроку A.

	B_1	B_2	B_3	B_4
A_1	2	2	3	−1
A_2	4	3	2	6

Игра не имеет решения в чистых стратегиях, и, следовательно, стратегии должны быть смешанными. Ожидаемые выигрыши игрока A, соответствующие чистым стратегиям игрока B, приведены в следующей таблице.

Чистые стратегии игрока B	Ожидаемые выигрыши игрока A
1	$-2x_1 + 4$
2	$-x_1 + 3$
3	$x_1 + 2$
4	$-7x_1 + 6$

На рис. 14.6 изображены четыре прямые линии, соответствующие чистым стратегиям игрока B. Чтобы определить *наилучший результат из наихудших*, построена *нижняя* огибающая четырех указанных прямых (изображенная на рисунке толстыми линейными сегментами), которая представляет минимальный (наихудший) выигрыш для игрока A независимо от того, что делает игрок B. Максимум (наилучшее) нижней огибающей соответствует максиминному решению в точке $x_1^* = 0.5$. Эта точка определяется пересечением прямых 3 и 4. Следовательно, оптимальным решением для игрока A является смешивание стратегий A_1 и A_2 с вероятностями 0.5 и 0.5 соответственно. Соответствующая цена игры v определяется подстановкой $x_1 = 0.5$ в уравнение либо прямой 3, либо 4, что приводит к следующему.

$$v = \begin{cases} \frac{1}{2} + 2 = \frac{5}{2}, & \text{из уравнения прямой 3,} \\ -7\left(\frac{1}{2}\right) + 6 = \frac{5}{2}, & \text{из уравнения прямой 4.} \end{cases}$$

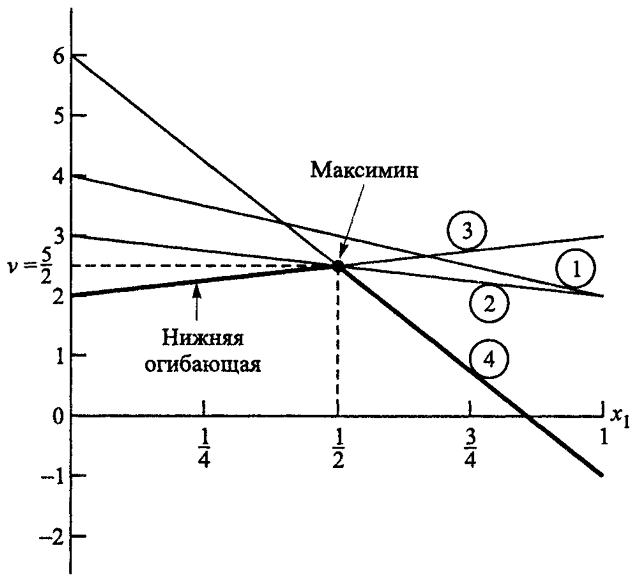


Рис. 14.6

Оптимальная смешанная стратегия игрока B определяется двумя стратегиями, которые определяют нижнюю огибающую графика. Это значит, что игрок B может смешивать стратегии B_3 и B_4, в этом случае $y_1 = y_2 = 0$ и $y_4 = 1 - y_3$. Следовательно, ожидаемые платежи игрока B, соответствующие чистым стратегиям игрока A, имеют следующий вид.

Чистые стратегии игрока A	Ожидаемые платежи игрока B
1	$4\acute{o}_3 - 1$
2	$-4\acute{o}_3 + 6$

Наилучшее решение из наихудших для игрока B представляет собой точку минимума *верхней* огибающей заданных двух прямых (построение прямых и определение верхней огибающей будет для Вас поучительным). Эта процедура эквивалентна решению уравнения

$$4y_3 - 1 = -4y_3 + 6.$$

Его решением будет $y_3 = 7/8$, что определяет цену игры $v = 4 \times (7/8) - 1 = 5/2$.

Таким образом, решением игры для игрока A является смешивание стратегий A_1 и A_2 с равными вероятностями 0.5 и 0.5, а для игрока B — смешивание стратегий B_3 и B_4 с вероятностями 7/8 и 1/8. (В действительности игра имеет альтернативное решение для игрока B, так как максиминная точка на рис. 14.6 определяется более чем двумя прямыми. Любая выпуклая линейная комбинация этих альтернативных решений также является решением задачи.)

Для игры, в которой игрок A имеет m стратегий, а игрок B — только две, решение находится аналогично. Главное отличие состоит в том, что здесь строятся графики функций, представляющих ожидаемые платежи второго игрока, соответствующие чистым стратегиям игрока A. В результате ведется поиск минимаксной точки *верхней огибающей* построенных прямых.

Упражнения 14.5,b

1. Решите графически игру с подбрасыванием монет из примера 14.5-2.
2. Робин часто путешествует между двумя городами. При этом есть возможность выбирать один из двух маршрутов: маршрут A представляет собой скоростное шоссе в четыре полосы, маршрут B — длинную обдуваемую ветром дорогу. Патрулирование дорог осуществляется ограниченным числом полицейских. Если все полицейские расположены на одном маршруте, Робин с ее страстным желанием ездить очень быстро, несомненно, получит штраф в 100 долларов за превышение скорости. Если полицейские патрулируют на двух маршрутах в соотношении 50 на 50, то имеется 50%-ная вероятность, что Робин получит штраф в 100 долларов на маршруте A и 30%-ная вероятность, что она получит такой же штраф на маршруте B. Кроме того, маршрут B длиннее, поэтому бензина расходуется на 15 долларов больше, чем на маршруте A. Определите стратегию как для Робин, так и для полиции.
3. Решите графически следующие игры, в которых платежи выплачиваются игроку A.

   а)

	B_1	B_2	B_3
A_1	1	−3	7
A_2	2	4	−6

b)

	B_1	B_2
A_1	5	8
A_2	6	5
A_3	5	7

4. Дана следующая игра двух лиц с нулевой суммой.

	B_1	B_2	B_3
A_1	5.0	50.0	50.0
A_2	1.0	1.0	0.1
A_3	10.0	1.0	10.0

a) Проверьте, что смешанные стратегии с вероятностями (1/6, 0, 5/6) для игрока A и с вероятностями (49/54, 5/54, 0) для игрока B являются оптимальными, и определите цену игры.

b) Покажите, что цена игры равна

$$\sum_{i=1}^{3}\sum_{j=1}^{3} a_{ij}x_i y_j.$$

Решение матричных игр методами линейного программирования. Теория игр находится в тесной связи с линейным программированием, так как любую конечную игру двух лиц с нулевой суммой можно представить в виде задачи линейного программирования и наоборот. Дж. Данциг [2] отмечает, что когда в 1947 году создатель теории игр Дж. фон Нейман впервые ознакомился с симплекс-методом, он сразу установил эту взаимосвязь и обратил особое внимание на концепцию *двойственности* в линейном программировании. Этот раздел иллюстрирует решение матричных игр методами линейного программирования.

Оптимальные значения вероятностей x_i, $i = 1, 2, \ldots, m$, игрока A могут быть определены путем решения следующей максиминной задачи.

$$\max_{x_i}\left\{\min\left(\sum_{i=1}^{m} a_{i1}x_i, \sum_{i=1}^{m} a_{i2}x_i, \ldots, \sum_{i=1}^{m} a_{in}x_i\right)\right\},$$

$$x_1 + x_2 + \ldots + x_m = 1,$$
$$x_i \geq 0, \quad i = 1, 2, \ldots, m.$$

Чтобы сформулировать эту задачу в виде задачи линейного программирования, положим

$$v = \min\left(\sum_{i=1}^{m} a_{i1}x_i, \sum_{i=1}^{m} a_{i2}x_i, \ldots, \sum_{i=1}^{m} a_{in}x_i\right).$$

Отсюда вытекает, что

$$\sum_{i=1}^{m} a_{ij}x_i \geq v, \quad j = 1, 2, \ldots, n.$$

Следовательно, задача игрока A может быть записана в виде

Максимизировать $z = v$

при ограничениях

$$\sum_{i=1}^{m} a_{ij}x_i \geq v, \ \ j = 1, 2, ..., n,$$

$$x_1 + x_2 + ... + x_m = 1,$$
$$x_i \geq 0, \ \ i = 1, 2, ..., m,$$
$$v \text{ не ограничено в знаке.}$$

Отметим последнее условие, что цена игры v может быть как положительной, так и отрицательной.

Оптимальные стратегии $y_1, y_2, ..., y_n$ игрока B определяются путем решения задачи

$$\min_{y_j} \left\{ \max \left(\sum_{j=1}^{n} a_{1j} y_j, \sum_{j=1}^{n} a_{2j} y_j, ..., \sum_{j=1}^{n} a_{mj} y_j \right) \right\},$$

$$y_1 + y_2 + ... + y_n = 1,$$
$$y_j \geq 0, \ \ j = 1, 2, ..., n.$$

Используя процедуру, аналогичную приведенной выше для игрока A, приходим к выводу, что задача для игрока B сводится к следующему.

Минимизировать $w = v$

при ограничениях

$$\sum_{j=1}^{n} a_{ij} y_j \leq v, \ \ i = 1, 2, ..., m,$$

$$y_1 + y_2 + ... + y_n = 1,$$
$$y_j \geq 0, \ \ j = 1, 2, ..., n,$$
$$v \text{ не ограничена в знаке.}$$

Две полученные задачи оптимизируют одну и ту же (не ограниченную в знаке) переменную v, которая является ценой игры. Причиной этого является то, что задача игрока B является двойственной к задаче игрока A (вам предлагается доказать это утверждение в упр. 14.5,c(6), используя определение двойственности из главы 4). Это означает, что оптимальное решение одной из задач автоматически определяет оптимальное решение другой.

Пример 14.5–4

Решим следующую матричную игру методами линейного программирования.

	B_1	B_2	B_3	Минимумы строк
A_1	3	–1	–3	–3
A_2	–2	4	–1	–2
A_3	–5	–6	2	–6
Максимумы столбцов	3	4	2	

Значение цены игры v находится между –2 и 2.

Задача линейного программирования для игрока A

Максимизировать $z = v$

при ограничениях

$$
\begin{aligned}
&3x_1 - 2x_2 - 5x_3 - v \geq 0,\\
&-x_1 + 4x_2 - 6x_3 - v \geq 0,\\
&-3x_1 - x_2 + 2x_3 - v \geq 0,\\
&x_1 + x_2 + x_3 = 1,\\
&x_1, x_2, x_3 \geq 0,\\
&v \text{ не ограничено в знаке.}
\end{aligned}
$$

Оптимальным решением, полученным с помощью программы TORA, является $x_1 = 0.3945$, $x_2 = 0.3119$, $x_3 = 0.2936$ и $v = -0.9083$.

Соответствующими двойственными переменными являются $y_1 = -0.3211$, $y_2 = -0.0826$, $y_3 = -0.5963$. Причина того, что переменные y_1, y_2, y_3 не являются положительными, как это должно быть, заключается в том, что задача линейного программирования для игрока A является задачей максимизации с ограничениями вида "$\geq$". При этих условиях, как известно, соответствующие двойственные переменные должны быть отрицательными (см. главу 4). Чтобы убедиться в том, что причина именно в этом, преобразуем все ограничения вида "$\geq$" в задаче линейного программирования для игрока A в ограничения вида "$\leq$" путем умножения каждого неравенства на -1. Соответствующие двойственные переменные будут неотрицательными, как и требуется (см. упр. 14.5,с(1)). Действительно, построение двойственной задачи непосредственно из задачи линейного программирования для игрока A показывает (см. упр. 14.5,с(6)), что в двойственной задаче, являющейся соответствующей задачей линейного программирования для игрока B, должны быть $y_j \leq 0$, но в то же время требуется выполнение условия $-y_1 - y_2 - \ldots - y_n = 1$, что равносильно требованиям $y_j \geq 0$ и $y_1 + y_2 + \ldots + y_n = 1$. К счастью, проблем, связанных со знаками, можно избежать, преобразуя ограничения–неравенства вида "$\geq$" в задаче линейного программирования для игрока A в ограничения–неравенства вида "$\leq$".

Задача линейного программирования для игрока B

$$\text{Минимизировать } z = v$$

при ограничениях

$$
\begin{aligned}
&3y_1 - y_2 - 3y_3 - v \leq 0,\\
&-2y_1 + 4y_2 - y_3 - v \leq 0,\\
&-5y_1 - 6y_2 + 2y_3 - v \leq 0,\\
&y_1 + y_2 + y_3 = 1,\\
&y_1, y_2, y_3 \geq 0,\\
&v \text{ не ограничено в знаке.}
\end{aligned}
$$

Оптимальным решением, полученным с помощью программы TORA, является $y_1 = 0.3211$, $y_2 = 0.0826$, $y_3 = 0.5963$ и $v = -0.9083$. Соответствующими двойственными переменными являются $x_1 = -0.3945$, $x_2 = -0.3119$, $x_3 = -0.2936$. Двойственные переменные отрицательны, ибо, подобно разобранной выше задаче для игрока A, в этом случае задача минимизации линейного программирования имеет ограничения–неравенства вида "$\leq$". Приведение ограничений к виду "$\geq$" исправит ситуацию со знаками.

Упражнения 14.5,с

1. Покажите в задаче из примера 14.5–4, что если ограничения–неравенства задачи линейного программирования для игрока A приведены к виду "$\leq$", то аналогичные ограничения задачи для игрока B преобразуются в вид "$\geq$" и двойственные переменные, полученные из одной или другой задачи, будут неотрицательными.

2. На загородном пикнике две команды, по два человека в каждой, играют в прятки. Есть четыре места, где можно спрятаться (А, Б, В и Г), и два члена прячущейся команды могут спрятаться каждый отдельно в любых двух из четырех мест. Затем другая команда имеет возможность проверить любые два места. Команда, которая ищет, получает премию, если будут обнаружены оба участника прячущейся команды, если же не обнаружен ни один участник, то она выплачивает премию. Иначе игра заканчивается вничью.

   a) Сформулируйте задачу в виде игры двух лиц с нулевой суммой.
   b) Определите оптимальные стратегии и цену игры.
   c) Имеет ли задача альтернативные решения?
   d) Является ли эта игра справедливой, т.е. имеет ли она цену, равную нулю?

3. Университетские команды UA и DU определяют свои стратегии игры в национальном чемпионате по баскетболу для колледжей. Оценивая возможности своих "запасных скамеек", каждый тренер разработал по четыре варианта замены игроков на протяжении игры. Способность каждой команды выполнять двух-, трехочковые и штрафные броски является основным фактором, определяющим результат игры. Приведенная ниже таблица содержит очки чистого выигрыша команды UA на протяжении одного владения мячом в зависимости от стратегий, планируемых каждой командой.

	DU_1	DU_2	DU_3	DU_4
UA_1	3	–2	1	2
UA_2	2	3	–3	0
UA_3	–1	2	–2	2
UA_4	–1	–2	4	1

   a) Решите игру методами линейного программирования и определите выигрышные стратегии.
   b) Исходя из имеющейся информации, какая из двух команд может выиграть чемпионат?
   c) Пусть за всю игру имеется 60 возможностей владения мячом (30 владений для каждой команды). Предскажите ожидаемое количество очков, с которым будет выиграна игра чемпионата.

4. Армия полковника Блотто сражается с вражеской армией за контроль над двумя стратегически важными позициями. Полковник имеет в своем распоряжении два полка, а его противник — три. Каждый из противников может посылать на любую позицию только целое число полков или совсем не посылать. Позиция будет захвачена армией, которая атакует бóльшим количеством полков. Иначе результат сражения является ничейным.

a) Сформулируйте задачу в виде игры двух лиц с нулевой суммой и решите игру методами линейного программирования.

b) Какая армия выиграет сражение?

5. В игре двух лиц, именуемой двухпальцевой игрой Морра, каждый игрок показывает один или два пальца и одновременно отгадывает число пальцев, которые покажет его противник. Игрок, который угадал, выигрывает сумму, равную суммарному числу показанных противниками пальцев. Иначе игра заканчивается вничью. Сформулируйте задачу в виде игры двух лиц с нулевой суммой и решите игру методами линейного программирования.

6. Покажите, что задача, двойственная к задаче линейного программирования для игрока A, является задачей линейного программирования для игрока B и что следующие два утверждения не противоречат друг другу.

a) Задача линейного программирования для игрока A записана в форме, приведенной в разделе 14.5.2.

b) Задача линейного программирования для игрока A записана в форме, упомянутой в п. а), в которой все ограничения вида "$\geq$" приведены к виду "$\leq$".

14.6. Заключение

В этой главе рассмотрены критерии выбора оптимальных решений как в ситуациях с точными (детерминированными), так и неполными (вероятностными или неопределенными) данными. Дальнейшие приложения критериев выбора оптимальных решений в ситуациях с неполными данными будут рассмотрены в последующих главах. В частности, критерий ожидаемой величины используется в главах, посвященных управлению запасами, теории массового обслуживания, имитационному моделированию и марковским процессам принятия решений. Это вовсе не значит, что другие критерии неприменимы. Критерий ожидаемой величины используется скорее традиционно в силу своей простоты.

Литература

Chen S. and Hwang C. *Fuzzy Multiple Decision Making*, Springer–Verlag, Berlin, 1992.

Dantzig G.B. *Linear Programming and Extension*, Princeton University Press, Princeton, N.J., 1963. (Имеется перевод: Данциг Дж. *Линейное программирование, его применения и обобщения*. — М.: Прогресс, 1966.)

Meyerson R. *Game Theory: Analysis of Conflict*, Harvard University Press, Camridge, Mass., 1991.

Saaty T.L. *Fundamentals of Decision Making*, RWS Publications, Pittsburg, 1994.

Литература, добавленная при переводе

Вилкас Э.Й., Майминас Е.З. *Решения: теория, информация, моделирование*. — М.: Радио и связь, 1981.

Карлин С. *Математические методы в теории игр, программировании и экономике*. — М.: Мир, 1964.

Ларичев О.И. *Наука и искусство принятия решений*. — М.: Наука, 1979.

Мулен Э. *Теория игр с примерами из математической экономики.* — М.: Мир, 1985.
Фишберн П. *Теория полезности для принятия решений.* — М.: Наука, 1978.
Фон Нейман Дж., Моргенштерн О. *Теория игр и экономическое поведение.* — М.: Наука, 1970.

Комплексные задачи

■ **14–1**[1]. Руководитель цеха рассматривает три возможных решения относительно существующего фрезерного станка.

a) Модифицировать имеющийся станок, установив на нем автоматическую подачу (АП).
b) Купить новый станок с программным управлением (ПУ).
c) Заменить станок обрабатывающим центром (ОЦ).

Три альтернативы оцениваются на основе двух критериев: денежный и функциональный. Следующая таблица содержит необходимые данные.

Критерий	Единицы	АП	ПУ	ОЦ
Денежный				
Начальная стоимость ($)		12 000	25 000	120 000
Стоимость обслуживания ($)		2 000	4 000	15 000
Стоимость обучения персонала ($)		3 000	8 000	20 000
Функциональный				
Производительность	Изделий/день	8	14	40
Время наладки	Минут	30	20	3
Металлические отходы	Фунты/день	440	165	44

Руководитель считает, что денежный критерий в полтора раза важнее функционального. Кроме того, производительность в два раза важнее времени наладки и в три раза важнее, чем количество получаемых металлических отходов. Показатель, связанный с временем наладки, считается в четыре раза важнее показателя, связанного с количеством металлических отходов. Что же касается денежного критерия, то руководитель считает, что стоимость обслуживания и стоимость обучения персонала одинаково важны, а начальная стоимость в два раза важнее каждого из этих двух показателей.

Проанализируйте описанную ситуацию и дайте соответствующие рекомендации.

■ **14–2**[2]. Компания использует каталог товаров для продажи, включающий более 200 000 наименований, хранящихся на многих региональных складах. В прошлом компания считала важным иметь точный перечень запасов на каждом складе. Как следствие этого, каждый год проводился переучет — интенсивная и неприятная работа, которая неохотно выполнялась всеми складами. Компания для проверки качества складских операций в регионе сопровождала каждый переучет ревизией, которая охватывала около 100 наименований на каждом складе. Результаты проверки обнаружили, что в среднем лишь

[1] Этот пример взят из работы Weber S. "A Modified Analytic Hierarchy Process for Automated Manufacturing Decisions", *Interfaces*, Vol.23, No. 4, 1993, pp. 75–84.

[2] Этот пример взят из работы Millet I. "A Novena to Saint Anthony, or How to Find Inventory by Not Looking", *Interfaces*, Vol. 24, No. 2, 1994, pp.69–75.

64% наименований на каждом складе соответствовали действительной инвентарной описи, что является неприемлемым. Дабы исправить ситуацию, компания распорядилась чаще проводить переучет дорогих и быстро реализуемых товаров. Системному аналитику была поставлена задача разработать процедуры для реализации этих планов.

Вместо того чтобы напрямую заняться выполнением задания компании, системный аналитик решил установить причину возникшей проблемы. Он перешел в своем исследовании от формулировки "Как мы можем увеличить частоту переучетов?" к "Как можно повысить точность переучетов?". Изучение проблемы под таким углом зрения свелось к следующему анализу. Предполагая, что доля точно сосчитанных наименований на складе равна p, аналитик затем предположил, что есть основания считать, что имеется 95%-ная вероятность того, что если изделие было правильно учтено в первый раз, то будет правильно переучтено и при последующем переучете. Для части $1-p$ товаров, которая не была точно учтена в первом раунде проверки, доля правильного учета во втором раунде равна 80%. Используя эту информацию, аналитик с помощью дерева решений построил график безубыточности, который сравнил точность учета в первом и втором раундах проверки. Конечный результат сводился к тому, что склады, на которых уровень точности выше порога безубыточности, не требовали переучета. Удивительным результатом предложенного решения было рьяное усилие со стороны каждого склада сделать правильный учет за первый раз, что привело к повышению точности учета на всех складах.

Как аналитик убедил руководство в жизнеспособности предложенного порога безубыточности для повторного переучета?

■ **14–3**[1]. В аэрофлоте рабочие часы устанавливаются в соответствии с договорами, заключенными с профсоюзными организациями. В частности, максимальная продолжительность работы может быть ограничена 16 часами для полетов на Боинге–747 (B–747) и 14 часами — на Боинге–707 (B–707). Если эти пределы превышаются в силу неожиданных задержек, экипаж должен быть заменен новым. Аэрофлот содержит резервные экипажи для таких случаев. Средняя годовая стоимость содержания члена резервного экипажа оценивается в 30 000 долларов. Задержка полета на одну ночь, обусловленная отсутствием резервного экипажа, может стоить 50 000 долларов. Член экипажа находится по вызову непрерывно 12 часов в сутки 4 дня в неделю и может *не* находиться по вызову три оставшихся дня недели. Самолет B–747 может обслуживаться двумя экипажами для самолета B–707.

Следующая таблица содержит вероятности вызова резервных экипажей, вычисленные на основании трехлетнего опыта.

		Вероятность вызова	
Категория рейса	Рейс (время вылета)	B–747	B–707
1	14.0	0.014	0.072
2	13.0	0.000	0.019
3	12.5	0.000	0.006
4	12.0	0.016	0.006
5	11.5	0.003	0.003
6	11.0	0.002	0.003

[1] Этот пример взят из работы Gaballa A. "Planning Callout Reserves for Aircraft Delays", *Interfaces*, Vol. 9, No. 2, Part 2, 1979, pp.78–86.

Приведенные данные свидетельствуют, например, что для 14-часового рейса вероятность вызова равна 0.014 для B–747 и 0.072 — для B–707.

Типичная пиковая часть расписания дня имеет следующий вид.

Время дня	Самолет	Категория рейса
8:00	707	3
9:00	707	6
	707	2
10:00	707	3
11:00	707	2
	707	4
15:00	747	6
16:00	747	4
19:00	747	1

Существующая политика относительно резервных экипажей состоит в использовании двух экипажей (по семь членов каждый) от 5:00 до 11:00, четырех — от 11:00 до 17:00 и двух — от 17:00 до 23:00.

Оцените эффективность существующей политики относительно резервных экипажей. В частности, является ли число резервных экипажей очень большим, очень малым или таким, как необходимо?

Вероятностное динамическое программирование

15.1. Введение[1]

Вероятностное динамическое программирование (ДП) отличается от детерминированного динамического программирования, описанного в главе 10, тем, что состояния и прибыли на каждом этапе являются случайными. Модели вероятностного ДП возникают, в частности, при рассмотрении стохастических моделей управления запасами и в теории марковских процессов принятия решений. Этим двум темам посвящены главы 16 и 19, поэтому в настоящей главе они не рассматриваются. В этой главе описываются некоторые примеры достаточно общего содержания, призванные показать стохастическую природу ДП.

15.2. Азартная игра

Одна из разновидностей игры в русскую рулетку состоит во вращении колеса, на котором по его периметру нанесены n последовательных чисел от 1 до n. Вероятность того, что колесо в результате одного вращения остановится на цифре i, равна p_i. Игрок платит x долларов за возможность осуществить m вращений колеса. Сам же игрок получает сумму, равную удвоенному числу, которое выпало при *последнем* вращении колеса. Поскольку игра повторяется достаточно много раз (каждая до m вращений колеса), требуется разработать оптимальную стратегию для игрока.

Мы сформулируем задачу в виде модели ДП, используя следующие определения.

1. *Этап* i соответствует i-му вращению колеса, $i = 1, 2, ..., m$.
2. *Альтернативы* на каждом этапе состоят в следующем — либо покрутить колесо еще раз, либо прекратить игру.
3. *Состояние* системы j на каждом этапе i представляется одним из чисел от 1 до n, которое выпало в результате *последнего* вращения колеса.

Пусть $f_i(j)$ — максимум ожидаемой прибыли при условии, что игра находится на этапе (вращении) i и исходом последнего вращения есть число j. Имеем следующее.

[1] Эта глава является продолжением главы 10, посвященной детерминированным моделям динамического программирования.

$$\begin{pmatrix}\text{Ожидаемая прибыль на этапе } i \text{ при условии,}\\ \text{что исходом последнего вращения есть } j\end{pmatrix} = \begin{cases} 2j, & \text{если игра заканчивается,}\\ \sum_{k=1}^{n} p_k f_{i+1}(k), & \text{если игра продолжается.}\end{cases}$$

Рекуррентное уравнение для $f_i(j)$ можно записать следующим образом.

$$f_{m+1}(j) = 2j,$$

$$f_i(j) = \max\begin{cases}\text{конец игры:} & 2j,\\ \text{продолжение игры:} & \sum_{k=1}^{n} p_k f_{i+1}(k), \quad i = 2, 3, \ldots, m,\end{cases}$$

$$f_1(0) = \sum_{k=1}^{n} p_k f_2(k).$$

Обоснование рекуррентного уравнения сводится к следующему. При первом вращении колеса ($i = 1$) состоянием системы является $j = 0$, ибо игра только началась. Следовательно, $f_1(0) = p_1 f_2(1) + p_2 f_2(2) + \ldots + p_n f_2(n)$. После выполнения последнего вращения колеса ($i = m$) имеется лишь один выбор — закончить игру независимо от исхода j m-го вращения. Следовательно, $f_{m+1}(j) = 2j$.

Рекуррентные вычисления начинаются с f_{m+1}, заканчиваются при $f_1(0)$ и сводятся таким образом к $m + 1$ вычислительному этапу. Так как $f_1(0)$ представляет собой ожидаемую прибыль от всех m вращений колеса, а игра обходится игроку в x долларов, имеем следующее.

$$\text{Ожидаемая прибыль} = f_1(0) - x.$$

Пример 15.2–1

Предположим, что по периметру колеса русской рулетки расставлены числа от 1 до 5. Вероятности p_i остановки колеса на числе i соответственно равны следующему: $p_1 = 0.3$, $p_2 = 0.25$, $p_3 = 0.2$, $p_4 = 0.15$, $p_5 = 0.1$. Игрок платит 5 долларов за возможность сделать не более четырех вращений колеса. Определим оптимальную стратегию игрока для каждого из четырех вращений и найдем соответствующий ожидаемый выигрыш.

Этап 5.

$$f_5(j) = 2j.$$

Исход 4-го вращения	Оптимальное решение	
j	$f_5(j)$	Решение
1	2	Закончить
2	4	Закончить
3	6	Закончить
4	8	Закончить
5	10	Закончить

Этап 4.

$$f_4(j) = \max\{2j, p_1f_5(1) + p_2f_5(2) + p_3f_5(3) + p_4f_5(4) + p_5f_5(5)\} =$$
$$= \max\{2j, 0.3 \times 2 + 0.25 \times 4 + 0.2 \times 6 + 0.15 \times 8 + 0.1 \times 10\} =$$
$$= \max\{2j, 5\}.$$

Исход 3-го вращения	Ожидаемая прибыль		Оптимальное решение	
j	Закончить	Вращать	$f_4(j)$	Решение
1	2	5	5	Вращать
2	4	5	5	Вращать
3	6	5	6	Закончить
4	8	5	8	Закончить
5	10	5	10	Закончить

Этап 3.

$$f_3(j) = \max\{2j, p_1f_4(1) + p_2f_4(2) + p_3f_4(3) + p_4f_4(4) + p_5f_4(5)\} =$$
$$= \max\{2j, 0.3 \times 5 + 0.25 \times 5 + 0.2 \times 6 + 0.15 \times 8 + 0.1 \times 10\} =$$
$$= \max\{2j, 6.15\}.$$

Исход 2-го вращения	Ожидаемая прибыль		Оптимальное решение	
j	Закончить	Вращать	$f_3(j)$	Решение
1	2	6.15	6.15	Вращать
2	4	6.15	6.15	Вращать
3	6	6.15	6.15	Вращать
4	8	6.15	8.00	Закончить
5	10	6.15	10.00	Закончить

Этап 2.

$$f_2(j) = \max\{2j, p_1f_3(1) + p_2f_3(2) + p_3f_3(3) + p_4f_3(4) + p_5f_3(5)\} =$$
$$= \max\{2j, 0.3 \times 6.15 + 0.25 \times 6.15 + 0.2 \times 6.15 + 0.15 \times 8 + 0.1 \times 10\} =$$
$$= \max\{2j, 6.8125\}.$$

Исход 1-го вращения	Ожидаемая прибыль		Оптимальное решение	
j	Закончить	Вращать	$f_2(j)$	Решение
1	2	6.81	6.81	Вращать
2	4	6.81	6.81	Вращать
3	6	6.81	6.81	Вращать
4	8	6.81	8.00	Закончить
5	10	6.81	10.00	Закончить

Этап 1.

$$f_1(0) = p_1f_2(1) + p_2f_2(2) + p_3f_2(3) + p_4f_2(4) + p_5f_2(5) =$$
$$= 0.3 \times 6.8125 + 0.25 \times 6.8125 + 0.2 \times 6.8125 + 0.15 \times 8 + 0.1 \times 10 = 7.31.$$

В начале игры единственным выбором является вращение колеса.

Из предыдущих таблиц следует, что оптимальным решением будет следующая последовательность действий.

Номер вращения	Оптимальная стратегия
1	Начало игры; вращать
2	Продолжить игру, если исходом первого вращения есть 1, 2 или 3; иначе закончить игру
3	Продолжить игру, если исходом второго вращения есть 1, 2 или 3; иначе закончить игру
4	Продолжить игру, если исходом третьего вращения есть 1 или 2; иначе закончить игру

Ожидаемая прибыль от игры составляет 7.31 – 5 = 2.31 доллара.

Упражнения 15.2,а

1. Пусть в задаче из примера 15.2–1 по периметру колеса расставлены числа от 1 до 8 и вероятности остановки колеса на каждом из этих чисел одинаковы. Предполагая, что каждая игра состоит не более чем из пяти вращений колеса, определите оптимальную стратегию игры.

2. Я хочу продать свой подержанный автомобиль тому, кто предложит наивысшую цену. Изучая автомобильный рынок, я пришел к выводу, что с равными вероятностями мне за автомобиль могут предложить очень низкую цена (около 1050 долларов), просто низкую цену (около 1900 долларов), среднюю цену (около 2500 долларов) либо высокую цену (примерно 3000 долларов). Я решил помещать объявление о продаже автомобиля на протяжении не более трех дней подряд. В конце каждого дня мне следует решить, принять ли наилучшее предложение, поступившее в течение этого дня. Какой должна быть моя оптимальная стратегия относительно принятия предложенной цены за автомобиль?

15.3. Задача инвестирования

Некто планирует инвестировать C тысяч долларов через фондовую биржу в течение последующих n лет. Инвестиционный план состоит в покупке акций в начале года и продаже их в конце этого же года. Накопленные деньги затем могут быть снова инвестированы (все или их часть) в начале следующего года. Степень риска инвестиции представлена тем, что прибыль имеет вероятностный характер. Изучение рынка свидетельствует о том, что прибыль от инвестиции зависит от m условий рынка (благоприятных или неблагоприятных). При этом условие i приводит к прибыли r_i с вероятностью p_i, $i = 1, 2, ..., m$. Как следует инвестировать C тысяч долларов для наибольшего накопления к концу n лет?

Обозначим

x_i — сумма денежных средств, доступных для инвестирования в начале i-го года ($x_1 = C$),
y_i — сумма реальной инвестиции в начале i-го года ($y_i \le x_i$).

Элементы модели ДП можно описать следующим образом.

1. *Этап i* представляет i-й год инвестирования.
2. *Альтернативами* на этапе i являются величины y_i.
3. *Состояние* системы на этапе i описывается величиной x_i.

Пусть $f_i(x_i)$ — максимальная ожидаемая сумма поступления денежных средств за года от i до n при условии, что в начале i-го года имеется сумма x_i. Для k-го условия рынка имеем следующее.

$$x_{i+1} = (1 + r_k)\, y_i + (x_i - y_i) = r_k y_i + x_i, \quad k = 1, 2, \ldots, m.$$

Так как вероятность k-го условия рынка равна p_k, рекуррентное уравнение динамического программирования имеет следующий вид.

$$f_i\left(x_i\right) = \max_{0 \le y_i \le x_i} \left\{ \sum_{k=1}^{m} p_k f_{i+1}\left(x_i + r_k y_i\right) \right\}, \quad i = 1, 2, \ldots, n,$$

где $f_{n+1}(x_{n+1}) = x_{n+1}$, так как после n-го года инвестиции нет. Отсюда следует, что

$$f_n\left(x_n\right) = \max_{0 \le y_n \le x_n} \left\{ \sum_{k=1}^{m} p_k \left(x_n + r_k y_n\right) \right\} = x_n \sum_{k=1}^{m} p_k \left(1 + r_k\right) = x_n \left(1 + p_1 r_1 + p_2 r_2 + \ldots + p_m r_m\right),$$

поскольку функция в фигурных скобках является линейной по y_n и, следовательно, достигает своего максимума при $y_n = x_n$.

Пример 15.3–1

Пусть в предыдущей модели инвестирования объем инвестиции составляет C = 10 000 долларов на 4-летний период. Существует 40%-ная вероятность того, что вы удвоите деньги, 20%-ная — останетесь при своих деньгах и 40% — потеряете весь объем инвестиции. Необходимо разработать оптимальную стратегию инвестирования.

Используя принятые в модели обозначения, имеем следующее.

C = \$10 000, $n = 4$, $m = 3$,
$p_1 = 0.4$, $p_2 = 0.2$, $p_3 = 0.4$,
$r_1 = 2$, $r_2 = 0$, $r_3 = -1$.

Этап 4.

$$f_4(x_4) = x_4(1 + 0.4 \times 2 + 0.2 \times 0 + 0.4 \times (-1)) = 1.4x_4.$$

Отсюда получаем

Состояние	Оптимальное решение	
	$f_4(x_4)$	y_4^*
x_4	$1.4\, x_4$	x_4

Этап 3.

$$f_3(x_3) = \max_{0 \le y_3 \le x_3} \{p_1 f_4(x_3 + r_1 y_3) + p_2 f_4(x_3 + r_2 y_3) + p_3 f_4(x_3 + r_3 y_3)\} =$$

$$= \max_{0 \le y_3 \le x_3} \{0.4 \times 1.4(x_3 + 2y_3) + 0.2 \times 1.4(x_3 + 0y_3) + 0.4 \times 1.4[x_3 + (-1)y_3]\} =$$

$$= \max_{0 \le y_3 \le x_3} \{1.4x_3 + 0.56y_3\} = 1.96x_3.$$

Поэтому имеем

Состояние	Оптимальное решение	
	$f_3(x_3)$	y_3^*
x_3	$1.96\, x_3$	x_3

Этап 2.

$$f_2(x_2) = \max_{0 \le y_2 \le x_2} \{p_1 f_3(x_2 + r_1 y_2) + p_2 f_3(x_2 + r_2 y_2) + p_3 f_3(x_2 + r_3 y_2)\} =$$

$$= \max_{0 \le y_2 \le x_2} \{0.4 \times 1.96(x_2 + 2y_2) + 0.2 \times 1.96(x_2 + 0y_2) + 0.4 \times 1.96[x_2 + (-1)y_2]\} =$$

$$= \max_{0 \le y_2 \le x_2} \{1.96x_2 + 0.784y_2\} = 2.744x_2.$$

Отсюда следует

Состояние	Оптимальное решение	
	$f_2(x_2)$	y_2^*
x_2	$2.744\, x_2$	x_2

Этап 1.

$$f_1(x_1) = \max_{0 \le y_1 \le x_1} \{p_1 f_2(x_1 + r_1 y_1) + p_2 f_2(x_1 + r_2 y_1) + p_3 f_2(x_1 + r_3 y_1)\} =$$

$$= \max_{0 \le y_1 \le x_1} \{0.4 \times 2.744(x_1 + 2y_1) + 0.2 \times 2.744(x_1 + 0y_1) + 0.4 \times 2.744[x_1 + (-1)y_1]\} =$$

$$= \max_{0 \le y_1 \le x_1} \{2.744x_1 + 1.0976y_2\} = 3.8416x_1.$$

Имеем

Состояние	Оптимальное решение	
	$f_1(x_1)$	y_1^*
x_1	$3.8416\, x_1$	x_1

Оптимальную инвестиционную политику можно сформулировать следующим образом. Так как $y_i^* = x_i$ для $i = 1, 2, 3, 4$, то оптимальным решением является инвестирование всех наличных денежных средств в начале каждого года. Накопленные денежные средства к концу четырех лет составят $3.8416x_1 = 3.8416 \times 10000 =$ 38 416 долларов.

Упражнения 15.3,а

1. Определите оптимальную инвестиционную политику в примере 15.3–1 в предположении, что вероятности p_k и прибыли r_k для следующих 4 лет принимают такие значения.

Год	r_1	r_2	r_3	p_1	p_2	p_3
1	2	1	0.5	0.1	0.4	0.5
2	1	0	–1	0.4	0.4	0.2
3	4	–1	–1	0.2	0.4	0.4
4	0.8	0.4	0.2	0.6	0.2	0.2

2. Камера объемом 10 кубических метров предназначена для хранения изделий трех наименований. Одно изделие наименований 1, 2, 3 занимает соответственно 2, 1 и 3 кубических метра. Вероятности спроса на эти изделия приведены в следующей таблице.

	Вероятность спроса		
Количество единиц	Наименование 1	Наименование 2	Наименование 3
1	0.5	0.3	0.3
2	0.5	0.4	0.2
3	0.0	0.2	0.5
4	0.0	0.1	0.0

Стоимость хранения единицы изделия наименований 1, 2, 3 равна 8, 10 и 15 долларов соответственно. Сколько единиц изделий каждого наименования следует хранить в камере?

3. Фирма с высокотехнологичным производством начала выпуск самых современных суперкомпьютеров в расчете на трехлетний период. Годовой спрос D на новый суперкомпьютер описывается распределением

$$p(D = 1) = 0.5,\ p(D = 2) = 0.3,\ p(D = 3) = 0.2.$$

Производственная мощность завода составляет три суперкомпьютера в год стоимостью 5 миллионов долларов каждый. Количество произведенных за год суперкомпьютеров может не совпадать в точности с объемом спроса. На нереализованный к концу года суперкомпьютер требуются затраты в 1 миллион долларов, связанные с его хранением и содержанием в исправности. Фирма терпит убытки в 2 миллиона долларов, если поставка суперкомпьютера откладывается на один год. Фирма не будет принимать новых заказов позже четвертого года, но будет продолжать выпуск суперкомпьютеров на протяжении пятого года, чтобы выполнить все заказы, оказавшиеся невыполненными к концу четвертого года. Определите оптимальные годичные объемы производства суперкомпьютеров.

4. Компания владеет тремя спортивными центрами в деловой части города. На Пасху популярны велосипедные прогулки на открытом воздухе. В компании имеется восемь велосипедов, которые она может распределить между тремя

центрами для их проката в целях максимизации доходов. Спрос на велосипеды и часовая стоимость их аренды зависят от месторасположения центра и характеризуются следующими данными.

Количество велосипедов	Вероятность спроса		
	Центр 1	Центр 2	Центр 3
0	0.10	0.02	0
1	0.20	0.03	0.15
2	0.30	0.10	0.25
3	0.20	0.25	0.30
4	0.10	0.30	0.15
5	0.10	0.15	0.10
6	0	0.05	0.025
7	0	0.05	0.025
8	0	0.05	0
Арендная плата ($/час)	6	7	5

Как компании распределить восемь велосипедов между тремя спортивными центрами?

15.4. Максимизация вероятности достижения цели

В разделе 15.3 рассматривалась задача, связанная с максимизацией ожидаемой прибыли. Иным полезным критерием для рассмотренной задачи является максимизация вероятности достижения определенного уровня дохода. Продемонстрируем этот подход на примере модели инвестирования, которая описана в разделе 15.3.

Используя обозначения из раздела 15.3, оставим без изменения определение *этапа* i, *альтернативы* y_i и *состояния* x_i. Эти модели отличаются только определением критерия; здесь нашей целью является максимизация вероятности достижения некоторой накопленной денежной суммы S по истечении n лет. С этой точки зрения определим функцию $f_i(x_i)$ — вероятность накопления суммы S, если в начале i-го года имеются денежные средства в сумме x_i и для последующих лет $i, i+1, \ldots, n$ используется оптимальное инвестирование.

Рекуррентное уравнение динамического программирования имеет вид

$$f_n(x_n) = \max_{0 \le y_n \le x_n} \left\{ \sum_{k=1}^{m} p_k P\{x_n + r_k y_n \ge S\} \right\},$$

$$f_i(x_i) = \max_{0 \le y_i \le x_i} \left\{ \sum_{k=1}^{m} p_k f_{i+1}(x_i + r_k y_i) \right\}, \quad i = 1, 2, \ldots, n-1.$$

Рекуррентная формула основана на формуле условной вероятности

$$P\{A\} = \sum_{j=1}^{m} P\{A \mid B_j\} P\{B_j\}.$$

В нашем случае $f_{i+1}(x_i + r_k y_i)$ играет роль вероятности $P\{A \mid B_j\}$.

Пример 15.4–1

Некий индивидуум планирует инвестировать 2 000 долларов. Имеющиеся варианты позволяют удвоить эту сумму с вероятностью 0.3 или потерять ее с вероятностью 0.7. Акции продаются в конце года, а в начале следующего года все деньги или их часть снова инвестируются. Этот процесс повторяется на протяжении трех лет. Целью является максимизация вероятности достижения суммы в 4 000 долларов в конце третьего года.

В соответствии с обозначениями данной модели, имеем $r_1 = 1$ с вероятностью 0.3 и $r_2 = -1$ с вероятностью 0.7.

Этап 3.

На этом этапе состояние x_3 может изменяться от 0 до 8 000 долларов. Минимальное значение возможно, когда вся инвестиция потеряна, а максимальное — когда инвестиция удваивается в конце каждого из двух первых лет. Следовательно, рекуррентное уравнение для этапа 3 записывается в следующем виде:

$$f_3(x_3) = \max_{y_3=0,1,\ldots,x_3} \left\{0.3P\left\{x_3 + y_3 \geq 4\right\} + 0.7P\left\{x_3 - y_3 \geq 4\right\}\right\},$$

где $x_3 = 0, 1, \ldots, 8$.

Приведенная ниже таблица содержит детали вычислений для данного этапа.

Все заштрихованные ячейки таблицы являются неподходящими, так как не удовлетворяют условию $y_3 \leq x_3$. Кроме того, при выполнении вычислений можно заметить, что

$$P\{x_3 + y_3 \geq 4\} = 0, \text{ если } x_3 + y_3 < 4,$$
$$P\{x_3 - y_3 \geq 4\} = 0, \text{ если } x_3 - y_3 < 4.$$

В противном случае эти вероятности равны 1.

Хотя приведенная таблица и свидетельствует о том, что существуют альтернативные оптимумы для $x_3 = 1, 3, 4, 5, 6, 7$ и 8, оптимальный (последний) столбец содержит лишь наименьшие оптимальные значения y_3. Это объясняется тем, что инвестор не собирается инвестировать больше того, что необходимо для достижения поставленной цели.

Этап 2.

$$f_2(x_2) = \max_{y_2=0,1,\ldots,x_2} \left\{0.3f_3(x_2 + y_2) + 0.7f_3(x_2 - y_2)\right\}.$$

	$0.3f_3(x_2 + y_2) + 0.7f_3(x_2 - y_2)$					Оптимум	
x_2	$y_2 = 0$	$y_2 = 1$	$y_2 = 2$	$y_2 = 3$	$y_2 = 4$	f_2	y_2
0	$0.3 \times 0 + 0.7 \times 0 = 0$					0	0
1	$0.3 \times 0 + 0.7 \times 0 = 0$	$0.3 \times 0.3 + 0.7 \times 0 = 0.09$				0.09	1
2	$0.3 \times 0.3 + 0.7 \times 0.3 = 0.3$	$0.3 \times 0.3 + 0.7 \times 0 = 0.09$	$0.3 \times 1 + 0.7 \times 0 = 0.3$			0.30	0
3	$0.3 \times 0.3 + 0.7 \times 0.3 = 0.3$	$0.3 \times 1 + 0.7 \times 0.3 = 0.51$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$		0.51	1
4	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 0.3 = 0.51$	$0.3 \times 1 + 0.7 \times 0.3 = 0.51$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	1	0

x_3	$0.3P\{x_3 + y_3 \geq 4\} + 0.7P\{x_3 - y_3 \geq 4\}$									Оптимум	
	$y_3 = 0$	$y_3 = 1$	$y_3 = 2$	$y_3 = 3$	$y_3 = 4$	$y_3 = 5$	$y_3 = 6$	$y_3 = 7$	$y_3 = 8$	f_3	y_3
0	$0.3 \times 0 + 0.7 \times 0 = 0$									0	0
1	$0.3 \times 0 + 0.7 \times 0 = 0$	$0.3 \times 0 + 0.7 \times 0 = 0$								0	0
2	$0.3 \times 0 + 0.7 \times 0 = 0$	$0.3 \times 0 + 0.7 \times 0 = 0$	$0.3 \times 1 + 0.7 \times 0 = 0.3$							0.3	2
3	$0.3 \times 0 + 0.7 \times 0 = 0$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$						0.3	1
4	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$					1	0
5	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$				1	0
6	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$			1	0
7	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$		1	0
8	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 1 = 1$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	$0.3 \times 1 + 0.7 \times 0 = 0.3$	1	0

Этап 1.

$$f_1(x_1) = \max_{y_1=0,1,2}\{0.3f_2(x_1+y_1)+0.7f_2(x_1-y_1)\}.$$

	$0.3f_2(x_1+y_1)+0.7f_2(x_1-y_1)$			Оптимум	
x_1	$y_1 = 0$	$y_1 = 1$	$y_1 = 2$	f_1	y_1
2	$0.3\times0.3+0.7\times0.3 = 0.3$	$0.3\times0.51+0.7\times0.09 = 0.216$	$0.3\times1+0.7\times0 = 0.3$	0.3	0

Оптимальная стратегия определяется следующим образом. При заданной начальной сумме $x_1 = 2000$ долларов вычисления для первого этапа дают $y_1 = 0$. Это означает, что в первый год не следует делать инвестиций. Данное решение оставляет инвестора с 2000 долларов к началу второго года. Из таблицы, соответствующей второму этапу, при $x_2 = 2$ получаем $y_2 = 0$; это снова означает, что на протяжении второго года также не следует делать инвестиций. Далее использование значения $x_3 = 2$ на третьем этапе приводит к $y_3 = 2$, а это означает, что на третий год следует инвестировать всю имеющуюся в распоряжении сумму. Соответствующая максимальная вероятность достижения цели $S = 4$ равна $f_1(2) = 0.3$.

Упражнения 15.4,а

1. В примере 15.4–1 этап 1 решения задачи показывает, что существует два альтернативных оптимума: $y_1 = 0$ и $y_1 = 2$. Покажите, что применение стратегии $y_1 = 2$ (т.е. инвестировать все деньги в начале первого года) не изменяет результата инвестиционной политики на протяжении трех лет, а именно, соответствующая максимальная вероятность достижения цели сохраняется равной 0.3.

2. Решите задачу из примера 15.4–1, если целью инвестора является максимизация вероятности достижения, по меньшей мере, суммы в 6 000 долларов к концу третьего года. Инвестор имеет в своем распоряжении 1000 долларов, и вероятность удвоения суммы на протяжении каждого года равна 0.6.

3. Вы и ваш друг хотите сыграть в казино в следующую игру. Вы делаете определенную ставку, и каждый из вас независимо подбрасывает симметричную монету. За каждый доллар суммы ставки казино заплатит три доллара (что дает чистую прибыль в 2 доллара), если в результате подбрасывания выпадут две решки. Иначе вы теряете сумму ставки. Если вы с другом имеете в сумме один доллар, определите стратегию игры, считая, что целью является максимизация вероятности окончания трех игр с суммой в 4 доллара.

Литература

Bertsekas D. *Dynamic Programming: Deterministic and Stochastic Models*, Prentice Hall, Upper Saddle River, N.J., 1987.

Cooper L. and Cooper M. *Introduction to Dynamic Programming*, Pergamon Press, New York, 1981.

Smith D. *Dynamic Programming: A Practical Introduction*, Ellis Horwood, London, 1991.

Литература, добавленная при переводе

Беллман Р., Дрейфус С. *Прикладные задачи динамического программирования.* — М.: Наука, 1965.

Романовский И.В. *Алгоритмы решения экстремальных задач.* — М.: Наука, 1977.

Комплексные задачи

■ **15–1.** Компания использует грузовые автомобили для доставки заказов покупателям и планирует заменить свои автомобили на протяжении последующих пяти лет. Годовые затраты, связанные с использованием нового грузовика, являются нормально распределенной случайной величиной с математическим ожиданием 300 долларов и среднеквадратическим отклонением 50 долларов. Математическое ожидание и среднеквадратическое отклонение годовых эксплуатационных затрат через год возрастают на 10%. Стоимость нового грузового автомобиля в настоящее время равна 20 000 долларов и через год возрастет, как ожидается, на 12%. Грузовые автомобили используются чрезвычайно интенсивно, поэтому существует вероятность того, что каждый из них может сломаться в любое время, после чего ремонту не подлежит. Имеется возможность сдать старый автомобиль при покупке нового. При этом стоимость старого автомобиля зависит от того, находится ли он в рабочем состоянии. В начале шестого года автомобиль подлежит продаже по цене, которая также зависит от его состояния (аварийное или рабочее). Приведенная ниже таблица содержит данные, описывающие ситуацию в зависимости от возраста автомобиля.

Возраст автомобиля (года)	0	1	2	3	4	5	6
Вероятность поломки	0.01	0.05	0.10	0.16	0.25	0.40	0.60

Если автомобиль использовался 1 год и находится в рабочем состоянии, то его стоимость равна 70% от начальной и за год уменьшается на 15%. Если же он находится в аварийном состоянии, то соответствующие показатели уменьшаются в два раза. Стоимость автомобиля в виде испорченного имущества в начале шестого года составляет 200 долларов, если он находится в рабочем состоянии, и 50 долларов, если он аварийный. Разработайте оптимальную политику замены автомобилей.

Вероятностные модели управления запасами

16.1. Введение[1]

В главе 11 изложены основы теории управления запасами в условиях определенности. В этой главе рассматриваются вероятностные модели управления запасами, в которых значение спроса является случайной величиной с известным распределением вероятностей. Рассмотренные модели подразделяются на модели с *непрерывным* и *периодическим* контролем уровня запаса. При этом класс моделей с периодическим контролем включает как одноэтапные, так и многоэтапные модели.

16.2. Модель с непрерывным контролем уровня запаса

В этом разделе рассмотрены две модели управления запасами: 1) обобщение детерминированной модели экономичного размера заказа (см. раздел 11.3.1) на вероятностный случай, в которой используется буферный запас, отвечающий за случайный спрос и 2) более точная вероятностная модель экономичного размера заказа, которая учитывает вероятностный характер спроса непосредственно в постановке задачи.

16.2.1. "Рандомизированная" модель экономичного размера заказа

Некоторые специалисты пытались адаптировать детерминированную модель экономичного размера заказа (см. раздел 11.3.1) для учета вероятностной природы спроса, используя при этом приближенный метод, который предполагает существование постоянного буферного запаса на протяжении всего планового периода. Размер резерва устанавливается таким образом, чтобы вероятность истощения запаса в течение *периода выполнения заказа* (интервала между моментом размещения заказа и его поставкой) не превышала наперед заданной величины.

Введем следующие обозначения.

L — срок выполнения заказа, т.е. время от момента размещения заказа до его поставки,

x_L — случайная величина, представляющая величину спроса на протяжении срока выполнения заказа,

μ_L — средняя величина спроса на протяжении срока выполнения заказа,

[1] Данная глава продолжает тему главы 11, посвященной детерминированным моделям управления запасами.

σ_L — среднеквадратическое отклонение величины спроса на протяжении срока выполнения заказа,

B — размер резервного запаса,

α — максимально возможное значение вероятности истощения запаса на протяжении срока выполнения заказа.

Основным предположением при построении модели является то, что величина спроса x_L на протяжении срока выполнения заказа L является нормально распределенной случайной величиной со средним μ_L и стандартным отклонением σ_L, т.е. имеет распределение $N(\mu_L, \sigma_L)$.

На рис. 16.1 показана зависимость между размером резервного запаса B и параметрами детерминированной модели экономичного размера заказа, которая включает срок выполнения заказа L, среднюю величину спроса μ_L на протяжении срока выполнения заказа и экономичный размер заказа y^*. Заметим, что L должно быть равно *эффективному* времени выполнения заказа, как это определено в разделе 11.3.1.

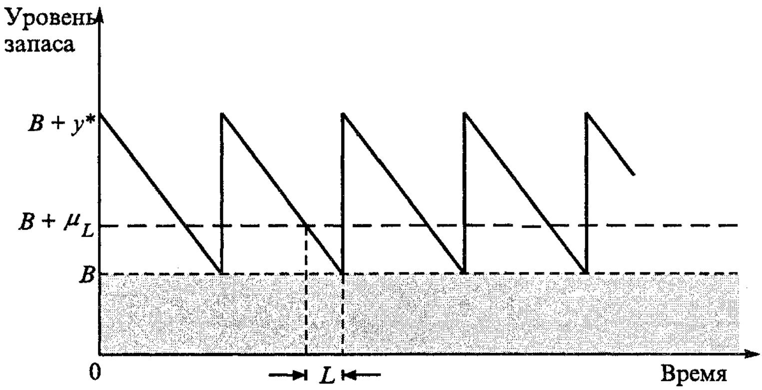


Рис. 16.1

Вероятностное условие, которое определяет размер резервного запаса B, имеет вид

$$P\{x_L \geq B + \mu_L\} \leq \alpha.$$

По определению (см. раздел 12.5.4) случайная величина

$$z = \frac{x_L - \mu_L}{\sigma_L}$$

является нормированной нормально распределенной случайной величиной, т.е. имеет распределение $N(0, 1)$. Следовательно,

$$P\left\{z \geq \frac{B}{\sigma_L}\right\} \leq \alpha.$$

На рис. 16.2 показана величина K_α, которая определяется из таблицы стандартного нормального распределения (см. Приложение Д), так что

$$P\{z \geq K_\alpha\} = \alpha.$$

Следовательно, размер резервного запаса должен удовлетворять неравенству

$$B \geq \sigma_L K_\alpha.$$

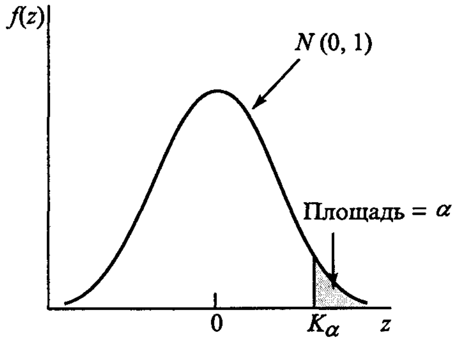


Рис. 16.2

Величина спроса на протяжении срока выполнения заказа L обычно описывается плотностью распределения вероятностей, отнесенной к *единице времени* (например, к дню или неделе), из которой можно определить распределение спроса на протяжении периода L. В частности, если спрос за единицу времени является нормально распределенной случайной величиной со средним D и стандартным отклонением σ, то общий спрос на протяжении срока выполнения заказа L будет иметь распределение $N(\mu_L, \sigma_L)$, где $\mu_L = DL$ и $\sigma_L = \sqrt{\sigma^2 L}$. Формула для σ_L получена на основании того, что значение L является целым числом (или же округлено до целого числа).

Пример 16.2–1

В примере 11.3–1, где речь шла об управлении запасом неоновых ламп в университетском городке, был определен экономичный размер заказа в 1000 ламп. Требуется определить размер резервного запаса таким образом, чтобы вероятность истощения запаса не превышала $\alpha = 0.05$ при условии, что дневной спрос является нормально распределенной случайной величиной с математическим ожиданием $D = 100$ ламп и среднеквадратическим отклонением $\sigma = 10$ ламп, т.е. имеет распределение $N(100, 10)$.

Как следует из примера 11.3–1, эффективное время выполнения заказа L равно 2 дня. Следовательно,

$$\mu_L = DL = 100 \times 2 = 200 \text{ единиц},$$

$$\sigma_L = \sqrt{\sigma^2 L} = \sqrt{10^2 \times 2} = 14.14 \text{ единиц}.$$

Из таблицы стандартного нормального распределения (Приложение Д) определяем $K_{0\,05} = 1.64$. Следовательно, размер резервного запаса вычисляется следующим образом.

$$B \geq 14.14 \times 1.64 \approx 23 \text{ неоновые лампы}.$$

При экономичном размере заказа $y^* = 1000$ единиц оптимальная политика управления запасами с объемом резерва B состоит в заказе 1000 ламп, как только объем запаса уменьшается до 223 единиц ($= B + \mu_L = 23 + 2 \times 100$).

Упражнения 16.2,а

1. В примере 16.2–1 определите оптимальное управление запасами для каждого из следующих случаев.

   a) Время выполнения заказа равно 15 дней.
   b) Время выполнения заказа равно 23 дня.
   c) Время выполнения заказа равно 8 дней.
   d) Время выполнения заказа равно 10 дней.

2. Музыкальный магазин продает популярный компакт-диск. Распределение дневного спроса на диск можно аппроксимировать нормальным распределением с математическим ожиданием 200 дисков и стандартным отклонением 20 дисков. Стоимость хранения диска в магазине составляет 0.04 доллара за один день. Размещение нового заказа обходится магазину в 100 долларов. Поставщик обычно устанавливает семидневный срок для выполнения заказа. Предположим, что магазин хочет ограничить вероятность истощения запаса дисков на протяжении срока выполнения заказа величиной, не превышающей 0.02. Определите оптимальное управление запасами для магазина.

3. Дневной спрос на кинопленку в подарочном магазине курортной зоны является нормально распределенной случайной величиной с математическим ожиданием 30 пленок и стандартным отклонением 5 пленок. Стоимость хранения катушки с пленкой в магазине составляет 0.02 доллара. Размещение нового заказа на кинопленку каждый раз обходится магазину в 30 долларов. Стратегия магазина по управлению запасами состоит в размещении заказа на 150 кинопленок, как только уровень запаса опускается до 80 единиц, тем самым поддерживается одновременно постоянный резерв в 20 кинопленок.

   a) Для указанной стратегии магазина по управлению запасами определите вероятность истощения запаса на протяжении срока выполнения заказа.
   b) Разработайте рекомендации для магазина относительно стратегии по управлению запасами, предполагая, что вероятность истощения запаса кинопленок на протяжении срока выполнения заказа не превышает 0.10.

16.2.2. Стохастический вариант модели экономичного размера заказа

Нет оснований полагать, что "рандомизированная" модель экономичного размера заказа, рассмотренная в разделе 16.2.1, определит оптимальную политику управления запасами. Подтверждением этого является то, что существенная информация, имеющая отношение к вероятностной природе спроса, при этом подходе первоначально не учитывается, а используется лишь независимо на последнем этапе вычислений. Чтобы исправить такую "нездоровую" ситуацию, в этом разделе рассматривается более точная модель, в которой вероятностная природа спроса учитывается непосредственно в постановке задачи.

В отличие от случая, рассмотренного в разделе 16.2.1, в новой модели допускается неудовлетворенный спрос, как это показано на рис. 16.3. В рассматриваемой модели заказ размером y размещается тогда, когда объем запаса достигает уровня R. Как и в детерминированном случае, уровень R, при котором снова размещается заказ, является функцией периода времени между размещением заказа и его выполнением. Оптимальные значения y и R определяются путем минимизации ожидаемых затрат системы управления запасами, отнесенных к единице времени, которые включают как расходы на размещение заказа и его хранение, так и потери, связанные с неудовлетворенным спросом.

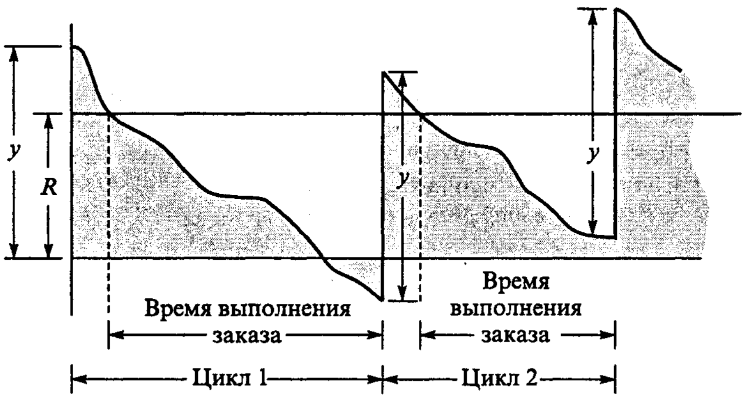


Рис. 16.3

В рассматриваемой модели приняты три допущения.

1. Неудовлетворенный в течение срока выполнения заказа спрос накапливается.
2. Разрешается не более одного невыполненного заказа.
3. Распределение спроса в течение срока выполнения заказа является стационарным (неизменным) во времени.

Для определения функции, отражающей суммарные затраты, отнесенные к единице времени, введем следующие обозначения.

$f(x)$ — плотность распределения спроса x в течение срока выполнения заказа,
D — ожидаемое значение спроса в единицу времени,
h — удельные затраты на хранение (на единицу продукции за единицу времени),
p — удельные потери от неудовлетворенного спроса (на единицу продукции за единицу времени),
K — стоимость размещения заказа.

Основываясь на этих определениях, вычислим компоненты функции затрат.

1. *Стоимость размещения заказов*. Приближенное число заказов в единицу времени равно D/y, так что стоимость размещения заказов в единицу времени равна KD/y.

2. *Ожидаемые затраты на хранение.* Средний уровень запаса равен

$$I = \frac{(y + M\{R - x\}) + M\{R - x\}}{2} = \frac{y}{2} + R - M\{x\}.$$

Следовательно, ожидаемые затраты на хранение за единицу времени равны hI.

Приведенная формула получена в результате усреднения ожидаемых запасов в начале и конце временного цикла, т.е. величин $y + M\{R - x\}$ и $M\{R - x\}$ соответственно. При этом игнорируется случай, когда величина $R - M\{x\}$ может быть отрицательной, что является одним из упрощающих допущений рассматриваемой модели.

3. *Ожидаемые потери, связанные с неудовлетворенным спросом.* Дефицит возникает при $x > R$. Следовательно, ожидаемый дефицит за единицу времени равен

$$S = \int_R^\infty (x - R) f(x)\,dx.$$

Так как в модели предполагается, что p пропорционально лишь объему дефицита, ожидаемые потери, связанные с неудовлетворенным спросом, за один цикл равны pS. Поскольку единица времени содержит D/y циклов, то ожидаемые потери, обусловленные дефицитом, составляют pDS/y за единицу времени.

Результирующая функция общих потерь за единицу времени TCU имеет следующий вид.

$$TCU(y, R) = \frac{DK}{y} + h\left(\frac{y}{2} + R - M\{x\}\right) + \frac{pD}{y}\int_R^\infty (x - R) f(x)\,dx.$$

Оптимальные значения y^* и R^* определяются из представленных ниже уравнений.

$$\frac{\partial TCU}{\partial y} = -\left(\frac{DK}{y^2}\right) + \frac{h}{2} - \frac{pD}{y^2}S = 0,$$

$$\frac{\partial TCU}{\partial R} = h - \left(\frac{pD}{y}\right)\int_R^\infty f(x)\,dx = 0.$$

Следовательно, имеем

$$y^* = \sqrt{\frac{2D(K + pS)}{h}}, \tag{1}$$

$$\int_{R^*}^\infty f(x)\,dx = \frac{hy^*}{pD}. \tag{2}$$

Так как из уравнений (1) и (2) y^* и R^* нельзя определить в явном виде, для их нахождения используется численный алгоритм, предложенный Хедли и Уайтин (Hadley, Whitin) [1]. Доказано, что алгоритм сходится за конечное число итераций при условии, что допустимое решение существует.

При $R = 0$ последние два уравнения соответственно дают следующее.

$$\bar{y} = \sqrt{\frac{2D(K + pM\{x\})}{h}},$$

$$\tilde{y} = \frac{pD}{h}.$$

Если $\tilde{y} \geq \bar{y}$, тогда существуют единственные оптимальные значения для y и R. Вычислительная процедура определяет, что наименьшим значением y^* является $\sqrt{2KD/h}$, которое достигается при $S = 0$.

Алгоритм состоит из следующих шагов.

Шаг 0. Принимаем начальное решение $y_1 = y^* = \sqrt{2KD/h}$ и считаем $R_0 = 0$. Полагаем $i = 1$ и переходим к шагу i.

Шаг i. Используем значение y_i для определения R_i из уравнения (2). Если $R_i \approx R_{i-1}$, вычисления заканчиваются; оптимальным решением считаем $y^* = y_i$ и $R^* = R_i$. Иначе используем значение R_i в уравнении (1) для вычисления y_i. Полагаем $i = i + 1$ и повторяем шаг i.

Пример 16.2–2

Электротехническая компания использует в производственном процессе канифоль в количестве 1000 галлонов в месяц. Размещение заказа на новую поставку канифоли обходится фирме в 100 долларов. Стоимость хранения одного галлона канифоли на протяжении одного месяца равна 2 доллара, а удельные потери от ее дефицита — 10 долларов за один галлон. Статистические данные свидетельствуют о том, что спрос в период поставки является случайной величиной, равномерно распределенной от 0 до 100 галлонов. Определите оптимальную политику управления запасами для компании.

Используя принятые в модели обозначения, имеем следующее.

$D = 1000$ галлонов в месяц,
$K = 100$ долл. за размещение заказа,
$h = 2$ долл. за один галлон в месяц,
$p = 10$ долл. за один галлон,
$f(x) = 1/100,\ 0 \leq x \leq 100$,
$M\{x\} = 50$ галлонов.

Сначала необходимо проверить, существует ли допустимое решение задачи. Используя уравнения для $\bar{y}$ и $\tilde{y}$, получаем следующее.

$$\bar{y} = \sqrt{\frac{2\times1000(100+10\times50)}{2}} = 774.6 \text{ галлонов},$$

$$\tilde{y} = \frac{10\times1000}{2} = 5000 \text{ галлонов}.$$

Так как $\tilde{y} \geq \bar{y}$, значит, существует единственное решение для y^* и R^*.

Выражение для S записывается в следующем виде:

$$S = \int_R^{100} (x-R)\frac{1}{100}dx = \frac{R^2}{200} - R + 50.$$

Используя в уравнениях (1) и (2) выражение для S, получаем следующее.

$$y_i = \sqrt{\frac{2\times1000(100+10S)}{2}} = \sqrt{100000+10000S} \text{ галлонов,} \quad (3)$$

$$\int_{R_i}^{100} \frac{1}{100} dx = \frac{2y_i}{10\times1000}.$$

Из последнего уравнения имеем

$$R_i = 100 - \frac{y_i}{50}. \quad (4)$$

Теперь используем уравнения (3) и (4) для нахождения решения.

Итерация 1.

$$y_1 = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2\times1000\times100}{2}} = 316.23 \text{ галлонов,}$$

$$R_1 = 100 - \frac{316.23}{50} = 93.68 \text{ галлонов.}$$

Итерация 2.

$$S = \frac{R_1^2}{200} - R_1 + 50 = 0.19971 \text{ галлонов.}$$

$$y_2 = \sqrt{100000+10000\times0.19971} = 319.37 \text{ галлонов.}$$

Следовательно,

$$R_2 = 100 - \frac{319.37}{50} = 93.612 \text{ галлонов.}$$

Итерация 3.

$$S = \frac{R_2^2}{200} - R_2 + 50 = 0.20399 \text{ галлонов.}$$

$$y_3 = \sqrt{100000+10000\times0.20399} = 319.44 \text{ галлонов.}$$

Следовательно,

$$R_3 = 100 - \frac{319.44}{50} = 93.611 \text{ галлонов.}$$

Так как значения R_2 и R_3 примерно одинаковы, приближенное оптимальное решение определяется значениями $R^* \approx 93.61$ галлонов, $y^* \approx 319.4$ галлонов.

Следовательно, оптимальное управление запасами состоит в размещении заказа примерно на 320 галлонов, как только запас уменьшается до 94 галлонов.

Упражнения 16.2,b

1. Для данных, приведенных в примере 16.2–2, определите следующее.
   a) Приближенное число заказов в месяц.
   b) Ожидаемое значение месячной стоимости размещения заказов.
   c) Ожидаемое значение месячных затрат на хранение.
   d) Ожидаемое значение месячных затрат, связанных с дефицитом.
   e) Вероятность истощения запаса в течение периода выполнения заказа.
2. Решите задачу из примера 16.2–2 в предположении, что спрос в период выполнения заказа является случайной величиной, равномерно распределенной на интервале [0, 50] (галлонов).
3. В задаче из примера 16.2–2 предположите, что спрос в период поставки является случайной величиной, равномерно распределенной на интервале [40, 60] (галлонов). Сравните решение, полученное при этих условиях, с решением, полученным в примере 16.2–2, и интерпретируйте результаты. (*Подсказка.* В обеих задачах величина $M\{x\}$ одинакова, но дисперсия в этой задаче меньше.)
4. Найдите оптимальное решение задачи из примера 16.2–2, если спрос в период поставки является нормально распределенной случайной величиной со средним 100 галлонов и стандартным отклонением 2 галлона. Предположите также, что $D = 10000$ галлонов в месяц, $h = 2$ доллара за галлон в месяц, $p = 4$ доллара за галлон и $K = 20$ долларов.

16.3. Одноэтапные модели

Одноэтапные модели управления запасами отражают ситуацию, когда для удовлетворения спроса в течение определенного периода продукция заказывается только один раз. Например, модный сезонный товар устаревает к концу сезона, и, следовательно, заказы на него могут не возобновляться. В данном разделе рассматривается два типа таких моделей: с учетом и без учета затрат на оформление заказов.

При изложении данного материала используются следующие обозначения.

c — стоимость закупки (или производства) единицы продукции,

K — стоимость размещения заказа,

h — удельные затраты на хранение единицы продукции в течение рассматриваемого периода,

p — удельные потери от неудовлетворенного спроса (на единицу продукции за рассматриваемый период),

D — величина случайного спроса за рассматриваемый период,

$f(D)$ — плотность вероятности спроса за рассматриваемый период,

y — объем заказа,

x — наличный запас продукта перед размещением заказа.

Модель определяет оптимальный объем заказа y, который минимизирует суммарные ожидаемые затраты, связанные с закупкой (или производством), хранением и неудовлетворенным спросом. При известном оптимальном значении y (обозначается y^*) оптимальное управление запасами состоит в размещении заказа объемом $y^* - x$, если $x < y$; в противном случае заказ не размещается.

16.3.1. Модель при отсутствии затрат на оформление заказа

В этой модели принято следующее.

1. Спрос удовлетворяется мгновенно в начале периода непосредственно после получения заказа.
2. Затраты на размещение заказа отсутствуют.

Рис. 16.4 иллюстрирует состояние запаса после удовлетворения спроса D. Если $D < y$, запас $y - D$ хранится на протяжении периода. Если же $D > y$, возникает дефицит объема $D - y$.

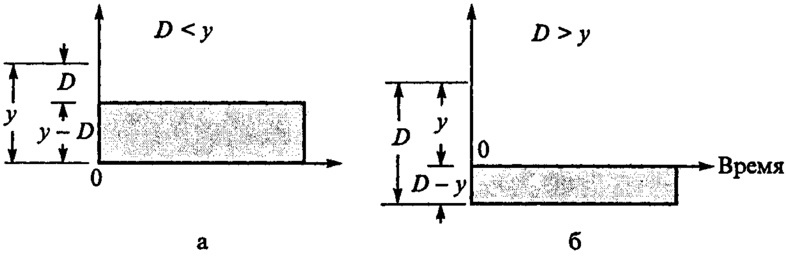


Рис. 16.4

Ожидаемые затраты $M\{C(y)\}$ на период выражаются следующей формулой.

$$M\{C(y)\} = c(y-x) + h\int_0^y (y-D)f(D)dD + p\int_y^\infty (D-y)f(D)dD.$$

Можно показать, что функция $M\{C(y)\}$ является выпуклой по y и, таким образом, имеет единственный минимум. Следовательно, вычисляя первую производную функции $M\{C(y)\}$ по y и приравнивая ее к нулю, получим

$$c + h\int_0^y f(D)dD - p\int_y^\infty f(D)dD = 0$$

или

$$c + hP\{D \le y\} + p(1 - P\{D \le y\}) = 0.$$

Отсюда имеем

$$P\{D \le y^*\} = \frac{p-c}{p+h}.$$

Правая часть последней формулы известна как **критическое отношение**. Значение y^* определено только при условии, что критическое отношение неотрицательно, т.е. $p \ge c$. Случай, когда $p < c$, является бессмысленным, так как это предполагает, что стоимость закупки единицы продукции выше потери от неудовлетворенного спроса.

Ранее предполагалось, что спрос D является непрерывной случайной величиной. Если же D является дискретной величиной, то плотность распределения вероятностей $f(D)$ определена лишь в дискретных точках и функция затрат определяется в соответствии с формулой

$$M\{C(y)\} = c(y-x) + h\sum_{D=0}^{y}(y-D)f(D) + p\sum_{D=y+1}^{\infty}(D-y)f(D).$$

Необходимыми условиями оптимальности являются неравенства

$$M\{C(y-1)\} \geq M\{C(y)\} \text{ и } M\{C(y+1)\} \geq M\{C(y)\}.$$

Эти условия в данном случае являются достаточными, так как функция $M\{C(y)\}$ выпукла. Применение этих условий после некоторых алгебраических преобразований приводит к следующим неравенствам для определения y^*.

$$P\left\{D \leq y^* - 1\right\} \leq \frac{p-c}{p+h} \leq P\left\{D \leq y^*\right\}.$$

Пример 16.3–1

Владелец газетного киоска должен определить количество экземпляров газеты *USA Now*, которые должны быть в продаже в начале каждого дня. Он покупает экземпляр газеты за 30 центов, а продает за 75 центов. Продажа газеты обычно происходит с 7.00 до 8.00 часов утра. Оставшиеся к концу дня экземпляры газеты повторно выставляются для продажи по цене 5 центов за экземпляр. Сколько экземпляров газеты должен закупить владелец каждое утро, если дневной спрос описывается одним из следующих вероятностных распределений.

1. Нормальным распределением с математическим ожиданием 300 экземпляров и стандартным отклонением 20 экземпляров.
2. Дискретной плотностью распределения $f(D)$, заданной в виде следующей таблицы.

D	200	220	300	320	340
$f(D)$	0.1	0.2	0.4	0.2	0.1

Стоимости хранения и потери, обусловленные дефицитом, в этой ситуации не определены в явном виде. Однако данные задачи свидетельствуют о том, что каждый непроданный экземпляр газеты обходится владельцу в 30 – 5 = 25 центов и что потери, связанные с истощением запаса газет, равны 75 центов за экземпляр. Следовательно, в терминах, принятых в модели управления запасами, мы можем предполагать, что $c = 30$ центов за экземпляр, $h = 25$ центов за экземпляр, $p = 75$ центов за экземпляр в день.

Сначала определяем критическое отношение

$$\frac{p-c}{p+h} = \frac{75-30}{75+25} = 0.45.$$

Случай 1.

Спрос D распределен по нормальному закону $N(300, 20)$. Определим нормированную нормально распределенную случайную величину (с законом распределения $N(0, 1)$)

$$z = \frac{D - 300}{20}.$$

Тогда

$$\frac{y^* - 300}{20} = -0.125.$$

Из таблицы стандартного нормального распределения находим (см. Приложение Д)

$$P\{z \le -0.125\} \approx 0.45.$$

Следовательно, оптимальный объем заказа равен $y^* = 297.5$ (или примерно 298) экземпляров.

Случай 2.

Спрос D описывается дискретной плотностью распределения $f(D)$. Сначала найдем функцию распределения $P\{D \le y\}$.

y	200	220	300	320	340
$P\{D \le y\}$	0.1	0.3	0.7	0.9	1.0

Для вычисленного критического отношения 0.45 имеем неравенства

$$P\{D \le 220\} \le 0.45 \le P\{D \le 300\}.$$

Отсюда следует, что $y^* = 300$ экземпляров.

Упражнения 16.3,а

1. Покажите, что для одноэтапной модели с дискретной величиной спроса оптимальный объем заказа определяется из соотношения

$$P\left\{D \le y^* - 1\right\} \le \frac{p - c}{p + h} \le P\left\{D \le y^*\right\}.$$

2. Спрос на продукцию в течение единственного периода удовлетворяется мгновенно в начале периода. Соответствующая плотность распределения вероятностей является экспоненциальной со средним 10 единиц. В силу сложности оценки стоимостных параметров объем заказа определяется таким образом, что вероятность либо излишка, либо дефицита не превышает 0.1. Можно ли удовлетворить оба условия одновременно?

3. В одноэтапной модели управления запасами стоимость закупки единицы продукции равна 10 долларов, а стоимость ее хранения — 1 доллар. Найдите допустимую область значений удельных потерь от неудовлетворенного спроса в оптимальном случае, если объем заказа равен 4 единицы. Также предполагается, что спрос удовлетворяется мгновенно в начале периода и что плотность распределения величины спроса представляется следующей таблицей.

D	0	1	2	3	4	5	6	7	8
$f(D)$	0.05	0.1	0.1	0.2	0.25	0.15	0.05	0.05	0.05

4. Профессор Портер Стоун на своих лекциях по экономике учит студентов "играть" в фондовую биржу. Эта игра продолжается 10 дней; с самого начала предполагается, что стоимость избранных акций будет возрастать ежедневно на 1%. В любой день существует также вероятность, что биржа пойдет на убыль (снизятся котировки акций), о чем свидетельствуют данные следующей таблицы.

День	1	2	3	4	5	6	7	8	9	10
Процент снижения	5.0	5.0	0.8	0.1	1.4	0.6	1.2	0.2	0.3	0.1
Вероятность снижения	0.06	0.02	0.05	0.07	0.10	0.13	0.15	0.20	0.12	0.10

Задача — максимизировать накопленную стоимость акций. Если бы вы были студентом профессора Стоуна, когда бы вы начали игру?

5. Книжный магазин университета предлагает ксерокопии конспектов лекций университетских профессоров. Профессор Ятаха читает лекции первокурсникам; на первый курс принимается от 200 до 250 студентов, причем ожидаемое количество первокурсников подчиняется равномерному распределению. Изготовление каждой копии обходится в 10 долларов, магазин продает их студентам по 25 долларов за копию. Студенты покупают конспекты в начале семестра. Каждая непроданная копия конспекта профессора Ятахи выставляется для продажи по частям. Между тем, если в магазине заканчиваются копии конспектов, дополнительные копии не печатаются и студенты сами заботятся о том, чтобы достать конспекты из других источников. Сколько копий конспектов лекций следует напечатать магазину, если он заинтересован в максимизации своих доходов?

6. Магазин быстрого обслуживания предлагает посетителям кофе и орехи с шести часов утра каждый день. Магазин покупает орехи по 7 центов за порцию, а продает по 25 центов за порцию до 8 часов утра. После этого времени орехи продаются по 5 центов за порцию. Число посетителей, которые ежедневно покупают орехи, является случайной величиной, равномерно распределенной на интервале [30, 50]. Каждый посетитель обычно заказывает три порции орехов с кофе. Сколько примерно порций орехов следует закупать магазину каждое утро в целях максимизации своих доходов?

7. Магазин одежды создает запас теплых пальто для приближающейся зимы. Каждое пальто закупают по 50 долларов, а продают со 100%-ной наценкой. В конце сезона проводится распродажа и предлагается пальто по цене 55 долларов. Спрос на пальто в течение зимнего сезона является случайной величиной, равномерно распределенной на интервале от 20 до 30. Так как зимний сезон является коротким, затраты на хранение в расчет не принимаются. Управляющий магазина считает также, что не будет потерь, вызванных дефицитом товара. Определите оптимальный объем заказа, который максимизирует доходы магазина.

8. Пусть в рамках одноэтапной модели спрос в течение периода меняется равномерно (а не удовлетворяется мгновенно в начале периода). Разработайте соответствующую стоимостную модель и определите оптимальный объем заказа.
9. Решите задачу из примера 16.3–1 в предположении, что спрос непрерывен и равномерен в течение периода; плотность вероятности спроса является постоянной на интервале от 0 до 100. (*Совет.* Воспользуйтесь результатами упр. 8.)
10. Дана задача максимизации ожидаемой прибыли в рамках одноэтапной модели. Спрос удовлетворяется мгновенно в *конце* периода. Пусть c, r и v представляют соответственно стоимость покупки, цену продажи и цену распродажи единицы продукции. Предполагая, что спрос D является непрерывным и описывается плотностью распределения вероятностей $f(D)$, постройте соответствующее выражение для суммарной ожидаемой прибыли и определите оптимальный объем заказа.

16.3.2. Модель при наличии затрат на оформление заказа

Рассматриваемая модель отличается от представленной в предыдущем разделе тем, что учитывается стоимость K размещения заказа. Используя обозначения, введенные выше, получаем следующее выражение для суммарной ожидаемой стоимости.

$$M\{\overline{C}(y)\} = K + M\{C(y)\} = K + c(y-x) + h\int_0^y (y-D)f(D)dD + p\int_y^\infty (D-y)f(D)dD.$$

Как показано в разделе 16.3.1, оптимальное значение y^* должно удовлетворять соотношению

$$P\{y \le y^*\} = \frac{p-c}{p+h}.$$

Так как K является константой, минимум величины $M\{\overline{C}(y)\}$ также должен достигаться при y^*, как показано на рис. 16.5.

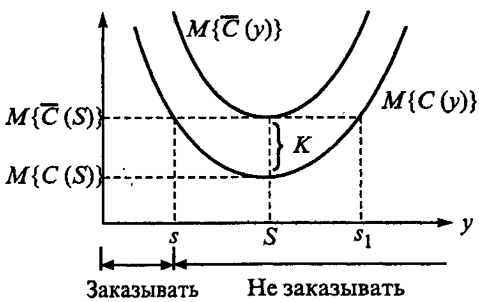


Рис. 16.5

На рис. 16.5 $S = y^*$ и величина s ($< S$) определяются из уравнения

$$M\{C(s)\} = M\{\overline{C}(S)\} = K + M\{C(S)\}, \quad s < S.$$

(Отметим, что это уравнение имеет и другое решение $s_1 > S$, которое не рассматривается.)

Задача формулируется следующим образом. Какое количество продукции необходимо заказывать, если наличный запас перед размещением заказа составляет x единиц? Ответ на этот вопрос рассматривается по отдельности при выполнении следующих условий.

1. $x < s$.
2. $s \le x \le S$.
3. $x > S$.

Случай 1 ($x < s$). Так как в наличии имеется x единиц продукции, соответствующие издержки содержания запаса составляют $M\{C(x)\}$. Если заказывается любое дополнительное количество продукции y ($y > x$), то соответствующие затраты при заданной величине y равны величине $M\{\overline{C}(y)\}$, которая учитывает стоимость K размещения заказа. Из рис. 16.5 следует, что

$$\min_{y>x} M\{\overline{C}(y)\} = M\{\overline{C}(S)\} < M\{C(x)\}.$$

Следовательно, оптимальной стратегией управления запасами в этом случае будет заказ в $S - x$ единиц.

Случай 2 ($s \le x \le S$). Из рис. 16.5 видно, что

$$M\{C(x)\} \le \min_{y>x} M\{\overline{C}(y)\} = M\{\overline{C}(S)\}.$$

Следовательно, в данном случае дополнительных затрат не возникает, если новый заказ *не размещается*. Поэтому $y^* = x$.

Случай 3 ($x > S$). Из рис. 16.5 видно, что при $y > x$

$$M\{C(x)\} < M\{\overline{C}(y)\}.$$

Это неравенство показывает, что в данном случае экономнее будет не размещать заказ, т.е. $y^* = x$.

Описанная стратегия управления запасами, часто именуемая **(s-S)-стратегией**, определяется следующим правилом.

Если $x < s$, делать заказ объемом $S - x$,
если $x \ge s$, заказывать не следует.

(Оптимальность (s-S)-стратегии следует из того, что соответствующая функция затрат является выпуклой. Если это свойство не выполняется, данная стратегия перестает быть оптимальной.)

Пример 16.3–2

Дневной спрос на продукцию в течение одного периода удовлетворяется мгновенно в начале периода. Спрос является случайной величиной, равномерно распределенной от 0 до 10 единиц. Стоимость хранения единицы продукции на протяжении периода равна 0.50 доллара, а штраф за дефицит единицы продукции — 4.50 доллара. Стоимость единицы продукции равна 0.50 доллара, стоимость размещения заказа — 25 долларов. Необходимо определить оптимальную стратегию заказа продукции.

Определим сначала y^*. Имеем

$$\frac{p-c}{p+h}=\frac{4.5-0.5}{4.5+0.5}=0.8.$$

Так как

$$P\{D\le y^*\}=\int_0^{y^*}\frac{1}{10}dD=\frac{y^*}{10},$$

то $S=y^*=8$.

Ожидаемое значение функции затрат определяется следующим образом.

$$M\{C(y)\}=0.5(y-x)+0.5\int_0^y\frac{1}{10}(y-D)dD+4.5\int_y^{10}\frac{1}{10}(D-y)dD=$$

$$=0.5(y-x)+0.05\left[yD-\frac{D^2}{2}\right]_0^y+0.45\left[\frac{D^2}{2}-Dy\right]_y^{10}=0.25y^2-4y+22.5-0.5x.$$

Величина s определяется из уравнения

$$M\{C(s)\}=K+M\{C(S)\}.$$

Отсюда получаем

$$0.25s^2-4s+22.5-0.5x=25+0.25S^2-4S+22.5-0.5x.$$

При $S=8$ это уравнение сводится к виду

$$s^2-16s-36=0.$$

Решением данного уравнения является $s=-2$ или $s=18$. Значение $s=18$ (превышающее S) следует отбросить. Так как оставшееся значение является отрицательным ($=-2$), то s не имеет допустимого значения. Следовательно, оптимальной стратегией является отказ от размещения заказа (рис. 16.6). Такая ситуация обычно возникает тогда, когда функция затрат является "плоской" или когда затраты на размещение заказа превышают другие затраты модели.

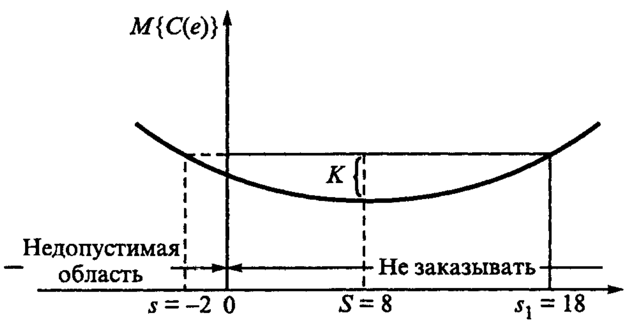


Рис. 16.6

Упражнения 16.3,b

1. Определите оптимальную стратегию управления запасами для ситуации, описанной в примере 16.3–2, если предположить, что стоимость размещения заказа составляет 5 долларов.
2. Пусть в одноэтапной модели, рассмотренной в разделе 16.3.1, необходимо максимизировать прибыль, при этом следует учитывать стоимость размещения заказа K. Постройте выражение для ожидаемого значения прибыли и определите оптимальный объем заказа, используя при этом информацию из раздела 16.3.1 и предполагая, что стоимость продажи единицы продукции равна r. Решите задачу при следующих данных: $r = \$3$, $c = \$2$, $p = \$4$, $h = \$1$ и $K = \$10$. Плотность вероятности спроса является постоянной на интервале [0, 10].
3. Решите упр. 16.3,а(6) в предположении, что доставка орехов связана с постоянными затратами в 10 долларов.

16.4. Многоэтапные модели

В этом разделе рассматривается многоэтапная модель в предположении, что не учитывается стоимость размещения заказа. Кроме того, в модели предусматривается возможность задолженности и нулевое время поставки. Предполагается также, что спрос D в каждый период описывается стационарной (независящей от времени) плотностью вероятности $f(D)$.

В многоэтапной модели учитывается приведенная стоимость денег. Если α (< 1) — коэффициент дисконтирования (процент скидки) для одного этапа, то сумма A спустя n этапов будет эквивалентна сумме $\alpha^n A$ в настоящий момент.

Предположим, что горизонт планирования охватывает n этапов и неудовлетворенный спрос может оставаться таковым лишь на протяжении одного этапа. Пусть $F_i(x_i)$ — максимальная суммарная ожидаемая прибыль для этапов от i до n, определенная при условии, что x_i — уровень имеющегося запаса перед размещением заказа на i-м этапе.

Используя обозначения из раздела 16.3 и предполагая, что r — удельный доход от реализации единицы продукции, сформулируем задачу управления запасами в виде следующей задачи динамического программирования (см. главу 15).

$$F_i(x_i) = \max_{y_i \geq x_i}\{-c(y_i - x_i) + \int_0^{y_i}[rD - h(y_i - D)]f(D)dD +$$
$$+\int_{y_i}^{\infty}[ry_i + \alpha r(D - y_i) - p(D - y_i)]f(D)dD +$$
$$+\alpha\int_0^{\infty}F_{i+1}(y_i - D)f(D)dD\},\quad i = 1, 2, \ldots, n,$$

где $F_{n+1}(y_n - D) \equiv 0$. Величина x_i может принимать отрицательные значения, так как неудовлетворенный спрос может накапливаться. Величина $\alpha r(D - y_i)$ включена во второй интеграл, поскольку $D - y_i$ представляет собой неудовлетворенный спрос на i-м этапе, который должен быть удовлетворен на этапе $i + 1$.

Задачу можно решить рекуррентно методами динамического программирования. Если число этапов является бесконечным (бесконечный горизонт планирования), приведенное выше рекуррентное уравнение сводится к следующему.

$$F(x) = \max_{y \ge x} \{-c(y-x) + \int_0^y [rD - h(y-D)] f(D) dD +$$
$$+ \int_y^\infty [ry + \alpha r(D-y) - p(D-y)] f(D) dD +$$
$$+ \alpha \int_0^\infty F(y-D) f(D) dD\},$$

где x и y представляют собой уровни запаса на каждом этапе до и после получения заказа соответственно.

Оптимальное значение y можно определить из приведенного ниже необходимого условия, которое в данном случае есть также достаточным, так как функция ожидаемой прибыли $F(x)$ является вогнутой.

$$\frac{\partial(.)}{\partial y} = -c - h\int_0^y f(D) dD + \int_y^\infty [(1-\alpha)r + p] f(D) dD + \alpha \int_0^\infty \frac{\partial F(y-D)}{\partial y} f(D) dD = 0.$$

Величина

$$\frac{\partial F(y-D)}{\partial y}$$

определяется следующим образом. Если на начало следующего этапа уровень запаса еще составляет β (> 0) единиц, то прибыль на этом этапе возрастает на величину $c\beta$, так как объем последующего заказа уменьшается именно на эту величину. Это означает, что

$$\frac{\partial F(y-D)}{\partial y} = c.$$

Следовательно, необходимое условие принимает вид

$$-c - h\int_0^y f(D) dD + [(1-\alpha)r + p]\left(1 - \int_0^y f(D) dD\right) + \alpha c \int_0^\infty f(D) dD = 0.$$

Поэтому оптимальный уровень заказа y^* определяется из уравнения

$$\int_0^{y^*} f(D) dD = \frac{p + (1-\alpha)(r-c)}{p + h + (1-\alpha)r}.$$

Оптимальная стратегия каждого этапа при заданном исходном запасе x выражается следующим правилом.

Если $x < y^*$, делать заказ объемом $y^* - x$,
если $x \ge y^*$, заказа не делать.

Упражнения 16.4,а

1. Рассмотрим двухэтапную вероятностную модель управления запасами, в которой спрос накапливается, а заказы поступают с нулевым запаздыванием. Спрос в каждый период описывается постоянной плотностью вероятности на интервале от 0 до 10. Стоимостные параметры модели таковы:

   цена продажи единицы продукции = \$2,
   цена покупки единицы продукции = \$1,
   стоимость хранения единицы продукции = \$0.10,
   штраф за дефицит единицы продукции = \$3,
   коэффициент дисконтирования = 0.8.

   Определите оптимальную стратегию для двух этапов, предполагая, что начальный запас для первого периода равен нулю.

2. Плотность распределения величины спроса для каждого этапа в модели управления запасами при бесконечном горизонте планирования имеет вид $f(D) = 0.08D$, $0 \le D \le 5$. Стоимостные параметры, отнесенные к единице продукции, следующие:

   цена продажи = \$10,
   цена покупки = \$8,
   штраф за дефицит = \$1,
   коэффициент дисконтирования = 0.9.

   Определите оптимальную стратегию управления запасами в предположении, что заказы выполняются с нулевым запаздыванием и неудовлетворенный спрос накапливается.

3. Дана модель управления запасами при бесконечном горизонте планирования, в которой заказы выполняются с нулевым запаздыванием, а неудовлетворенный спрос накапливается. Определите оптимальную стратегию управления запасами, основанную на минимизации ожидаемых затрат, если

   затраты на хранение z единиц продукции = hz^2,
   штраф за дефицит z единиц продукции = pz^2.

   Покажите, что в частном случае, когда $h = p$, оптимальное решение не зависит от конкретного вида вероятностного распределения спроса.

16.5. Заключение

В моделях управления запасами, рассмотренных в этой главе, спрос является случайным. Предложен широкий спектр методов решения построенных моделей — от вероятностной (рандомизированной) версии детерминированной модели экономичного размера заказа до более сложных, связанных с применением методов динамического программирования. Учет стохастической природы спроса приводит к сложным моделям, которые, вероятно, не являются полезными на практике. Тем не менее известны случаи успешного применения вероятностных моделей управления запасами.

Литература

Hadley G. and Whitin T., *Analysis of Inventory Systems*, Prentice Hall, Upper Saddle River, N.J., 1963. (Имеется перевод: Хедли Дж., Уайтин Т. *Анализ систем управления запасами*. — М.: Наука, 1969.)

Silver E. and Peterson R. *Decision Systems for Inventory Management and Production Planning*, 2nd ed., Wiley, New York, 1985.

Tersine R. *Principles of Inventory and Materials Management*, North Holland, New York, 1982.

Литература, добавленная при переводе

Кофман А. *Методы и модели исследования операций*. — М.: Мир, 1966.

Комплексные задачи

■ **16–1**[1]. Телефонная компания управляет *телефонными центрами*, которые предоставляют услуги клиентам по месту жительства в своих регионах. Имеется более 60 моделей телефонных аппаратов, которые предлагаются клиентам. В настоящее время каждый телефонный центр содержит запас телефонных аппаратов, рассчитанный на срок от 15 до 75 дней. Управляющий считает такие уровни запаса чрезмерными, так как они ежедневно пополняются из центрального товарного склада. В то же время управляющий должен гарантировать, что в телефонных центрах поддерживается уровень запаса, достаточный для обеспечения обслуживания клиентов на уровне 95%. Изучающий проблему коллектив специалистов начал свою работу со сбора соответствующих данных. Задача этого коллектива — определить оптимальный уровень запаса для каждой модели телефонного аппарата. Следующая таблица содержит количество установленных за день настольных телефонных аппаратов зеленого цвета с вращающимся наборным диском.

Установленные аппараты	0	1	2	3	4
Частота	189	89	20	4	1

Аналогичные таблицы были построены для всех моделей телефонных аппаратов.

Стоимостные параметры, необходимые для определения оптимального уровня запаса каждой модели телефонного аппарата, трудно поддаются оценке, и, следовательно, применение традиционных моделей управления запасами невозможно. Поэтому исследовательский коллектив решил использовать более основательный подход к определению соответствующего уровня запаса для различных моделей телефонных аппаратов. В результате исследования был сделан вывод, что как регрессионный анализ, так и анализ временных рядов не смогли обнаружить заметных тенденций спроса.

Предложите метод определения соответствующих уровней запаса для различных моделей телефонных аппаратов. Сформулируйте все предположения, необходимые для получения решения.

[1] Этот пример взят из работы Cohen R. and Dunford F. "Forecasting for Inventory Control: An Example of When "Simple" Means "Better", *Interfaces*, Vol. 16, No. 6, pp. 95–99, 1986.

■ 16–2[1]. Менеджер по снабжению небольших магазинов розничной торговли размещает заказы на продукцию с тем, чтобы получить выгоду за счет специальных цен или объединения заказов, полученных от одного поставщика. В результате как объем заказа, так и продолжительность цикла (период между последовательными заказами) являются случайными. Более того, так как стратегия менеджера, в основном, определяется не соображениями управления запасами, объем заказа и продолжительность цикла могут рассматриваться независимо в том смысле, что из меньшей длительности циклов не следует (с необходимостью) меньший объем заказов и наоборот.

Приведенная ниже таблица содержит типичные данные для трех наименований продукции, которые были заказаны одновременно. Эти данные показывают, что как объем заказа, так и продолжительность цикла являются случайными величинами. Более того, даже беглый анализ данных таблицы обнаруживает отсутствие корреляции между объемом заказа и продолжительностью цикла.

	Объем заказа (единицы)		
Длина цикла (месяцы)	Продукция 1	Продукция 2	Продукция 3
2.3	10	8	1
2.6	4	6	0
0.4	1	4	2
2.0	8	6	2
1.2	7	0	2
1.4	0	10	1
1.7	1	2	0
1.3	0	5	2
1.1	9	4	3
1.8	4	6	2
1.6	2	0	0
0.5	5	3	1
2.1	10	7	2
2.3	4	12	4
2.4	8	9	3
2.1	10	8	5
2.2	9	13	2
1.8	12	8	4
0.7	6	4	2
2.1	5	4	0

Проверка критерия согласия для всех приведенных данных (см. главу 12) показывает, что распределение коэффициентов спроса (объем заказа, деленный на длину цикла) для трех наименований продукции является распределением Вейбулла $f(r)$, плотность вероятности которого задается формулой

$$f(r)=\frac{2r}{\alpha}e^{-r^2/\alpha},\quad r\ge 0,$$

[1] Этот пример взят из работы Holt A. "Multi-Item Inventory Control for Fluctuating Reorder Intervals", *Interfaces*, Vol. 16, No. 3, pp. 60–67, 1986.

где r — коэффициент спроса на продукцию. Дальнейший анализ показывает, что распределение *обратной величины* к длине цикла $s(x)$ является экспоненциальным

$$s(x) = \beta e^{-\beta(x-a)},\ x \geq a,$$

где a — минимальное значение, которое может принимать x.

Определение оптимального объема заказа основано на максимизации ожидаемой прибыли в месяц, которая определяется формулой

$$\text{Ожидаемая прибыль} = \int\left[\frac{1}{t}\int u(q,r,t)f(r)dr\right]g(t)dt = \int\left[x\int u\left(q,r,\frac{1}{x}\right)f(r)dr\right]s(x)dx,$$

где t и $g(t)$ представляют соответственно длину цикла и ее плотность вероятности. Функция прибыли $u(q, r, t)$ определяется чистым доходом p от единицы продукции, стоимостью хранения h единиц продукции на протяжении месяца и фиксированными затратами K, связанными с размещением заказа.

a) Определите плотность вероятности спроса для каждого наименования продукции на основе данных таблицы.
b) Используйте приведенные в таблице значения длины цикла для определения $s(x)$.
c) Получите в математической форме выражение для $u(q, r, t)$.
d) Определите оптимальный объем заказа для трех наименований продукции при следующих данных: $p_1 = \$100$, $p_2 = \$150$, $p_3 = \$125$, $h_1 = \$2$, $h_2 = \$1.20$, $h_3 = \$1.65$ и $K = \$30$.

Системы массового обслуживания

17.1. Введение

Ожидание того или иного вида обслуживания является частью нашей повседневной жизни. Мы ожидаем, чтобы пообедать в ресторане, мы стоим в очереди к кассам в продовольственных магазинах и выстраиваемся в очередь в почтовых отделениях. Однако феномен ожидания характерен не только для людей: работы, поставленные в очередь для выполнения на станке; группа пассажирских самолетов, ожидающих разрешения на посадку в аэропорту; автомобили, движение которых приостановлено сигналом светофора на пути их следования и т.п. К сожалению, феномен ожидания нельзя исключить без чрезмерных расходов. И лишь на одно мы можем надеяться — на возможность сокращения времени нежелательного ожидания в очереди до некоторых терпимых пределов.

Изучение очередей в системах массового обслуживания позволяет определить критерии функционирования обслуживающей системы, среди которых наиболее значимыми являются среднее время ожидания в очереди и средняя длина очереди. Эта информация используется затем для выбора надлежащего уровня обслуживания, что продемонстрировано в следующем примере.

Пример 17.1–1

Посетители ресторана быстрого питания жалуются на медленное обслуживание. В настоящее время в ресторане работают три кассира. Управляющий поручил консалтинговой фирме провести расследование жалобы. В результате была обнаружена следующая зависимость между числом кассиров и временем ожидания обслуживания.

Число кассиров	1	2	3	4	5	6	7
Среднее время ожидания (минуты)	16.2	10.3	6.9	4.8	2.9	1.9	1.3

Приведенные данные свидетельствуют о том, что при работающих в настоящее время трех кассирах среднее время ожидания обслуживания примерно равно 7 минут. Управляющий хочет уменьшить его примерно до трех минут.

Как следует из этих же данных, среднее время ожидания становится меньше 3 минут, если число кассиров больше или равно пяти. Действительно, при пяти работающих кассирах среднее время ожидания равно 2.9 минуты.

Результаты исследования системы обслуживания также можно использовать для оптимизации модели со стоимостными характеристиками, в которой минимизируется сумма затрат, связанных с предоставлением услуг, и потерь, обусловленных задержками в их предоставлении. На рис. 17.1 изображена типичная стоимостная модель системы обслуживания (в долларах за единицу времени), где затраты на обслуживание возрастают с ростом его уровня. В то же время потери, обусловленные задержками в предоставлении услуг, уменьшаются с возрастанием уровня обслуживания. Главной проблемой, связанной с применением стоимостных моделей, является трудность оценки потерь в единицу времени, обусловленных задержками в предоставлении услуг. В частности, это особенно ощутимо, когда услуги предоставляются индивидууму, чье поведение может не совпадать с интересами функционирования системы обслуживания (см. раздел 17.9).

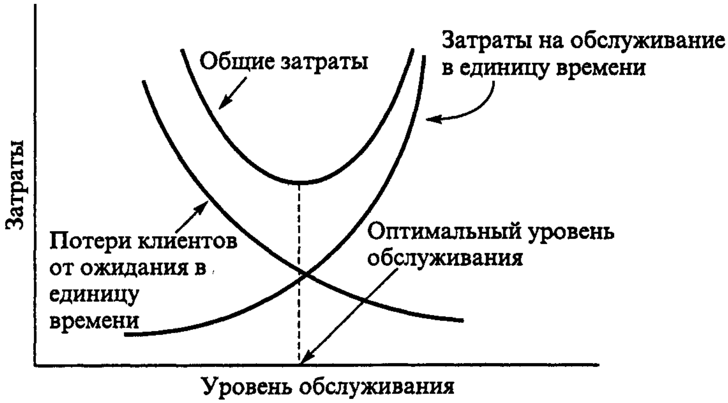


Рис. 17.1

Упражнения 17.1,а

1. Предположим, что дальнейший анализ работы ресторана быстрого питания привел к следующим дополнительным данным.

Число кассиров	1	2	3	4	5	6	7
Простой (%)	0	8	12	18	29	36	42

   a) Какова эффективность обслуживания (выраженная в процентах времени, когда служащие заняты работой), если число кассиров равно пяти?

   b) Управляющий желает поддерживать одновременно среднее время ожидания на уровне примерно трех минут и эффективность использования персонала на уровне 90%. Может ли он достичь этого? Приведите надлежащее объяснение.

2. Металлообрабатывающая мастерская покупает перфоратор многоцелевого назначения. Подходящими являются модели *A* и *B*, эксплуатационные затраты при их использовании равны 18 и 25 долларов в час соответственно. Перфоратор модели *A* работает медленнее, чем перфоратор модели *B*. Анализ систем обслуживания с такими устройствами показывает, что в случае использования модели *A* средняя длина очереди работ, ожидающих обслуживания, равна четырем, что на 30% вы-

ше аналогичного показателя при использовании модели *B*. Задержка в выполнении работы приводит к потере прибыли, которая оценивается в 10 долларов за час ожидания в очереди. Какую модель перфоратора следует приобрести мастерской?

17.2. Основные компоненты моделей массового обслуживания

Основными элементами модели массового обслуживания являются **клиент** (заявка или требование на обслуживание либо просто "объект обслуживания") и **сервис** (обслуживающее устройство[1], средства обслуживания и т.п.). Клиенты поступают в систему обслуживания из **источника**. Поступив в **сервис**, они могут сразу же попасть на обслуживание или ожидать в **очереди**, если сервис занят. После завершения процедуры обслуживания сервис автоматически "выбирает" из очереди (если она имеется) одного из клиентов с тем, чтобы приступить к его обслуживанию. Если же очередь отсутствует, то сервис становится незанятым до прибытия нового клиента.

Поступление клиентов в систему обслуживания характеризуется **интервалом между их последовательными поступлениями**, а обслуживание — **временем обслуживания** клиента. В общем случае эти параметры могут быть и случайными, как, например, в работе почтового отделения, и детерминированными, как, например, прибытие на собеседование кандидатов на вакантную должность.

В анализе систем обслуживания определенную роль играет **длина очереди**, которая может быть конечной, как в буферной зоне между двумя последовательными обслуживающими устройствами, и бесконечной, как у обслуживающих операторов почтовых отделений.

Важным фактором при анализе систем обслуживания является **дисциплина очереди**, или **принцип построения очереди**, определяющие порядок, в соответствии с которым выбираются клиенты из очереди для обслуживания. Наиболее распространенный принцип построения очереди основан на правиле "первым пришел — первым обслуживаешься" (это правило часто обозначается аббревиатурой FIFO — от английского First-In-First-Out, т.е. "первым вошел — первым вышел"). Среди других правил, определяющих принципы построения очередей, укажем правило "последним пришел — первым обслуживаешься" (обычно обозначается как LIFO — от английского Last-In-First-Out, т.е. "последним вошел — первым вышел") и дисциплину очереди, определяемую случайным правилом отбора клиентов. Кроме того, клиенты могут выбираться из очереди в соответствии с заданным **приоритетом**. Например, в производственном цехе срочные работы выполняются раньше обычных.

При анализе систем с очередями важным фактором является также поведение индивидуума, нуждающегося в обслуживании. Такие индивидуумы, выступающие в роли клиентов, при наличии параллельного обслуживания могут **перейти** из одной очереди в другую в надежде сократить продолжительность своего вынужденного ожидания. Они могут также **отказаться** от ожидания в очереди, так как люди обычно не переносят длительного бездействия, или **покинуть** очередь, простояв в ней какое-то время и придя к выводу, что и так уж слишком много времени потеряно.

---

[1] Очень часто, особенно в системах с одним сервисом, понятия "обслуживающее устройство" и "обслуживающая система" являются синонимами. — *Прим. перев.*

Структура обслуживающей системы может включать один сервис или несколько таких средств обслуживания, работающих параллельно (например, работа нескольких клерков почтового отделения). Кроме того, сервисы могут быть расположены последовательно (например, обслуживание представляет собой комплекс работ, которые выполняются последовательно на различных станках).

Источник, генерирующий "клиентов", подлежащих обслуживанию, может иметь конечную или бесконечную мощность. **Источник конечной мощности** ограничивает число клиентов, поступающих на обслуживание (например, в цехе, располагающем N станками, суммарное количество потенциальных заявок на их ремонт не превышает N). Наоборот, **источник бесконечной мощности** всегда имеет клиентов в "изобилии" (например, звонки, поступающие на телефонную станцию).

Можно построить множество моделей систем массового обслуживания, варьируя перечисленные выше операционные характеристики систем. В этой главе рассматривается ряд таких моделей.

Упражнения 17.2,а

1. В каждой из следующих ситуаций определите клиента и сервис (средство обслуживания).

   a) Самолеты, прибывающие в аэропорт.
   b) Стоянка такси.
   c) Обрабатывающие инструменты в цехе механической обработки.
   d) Письма, обрабатываемые в почтовом отделении.
   e) Регистрация на учебу в университете.
   f) Судебные дела судьи.
   g) Процесс контроля в супермаркете.
   h) Обслуживание площадки, отведенной под автостоянку.

2. В каждой ситуации, описанной в предыдущем упражнении, определите следующее: (a) мощность источника "клиентов" (конечная или бесконечная), (b) характер поступающих "клиентов" (индивидуальные или групповые), (c) тип интервалов времени между последовательными поступлениями заявок (случайный или детерминированный), (d) тип времени обслуживания, (e) вместимость очереди (конечная или бесконечная) и (f) дисциплину очереди.

3. Проанализируйте описанные ниже ситуации и опишите их с помощью моделей массового обслуживания. В каждом случае определите клиентов, сервисы, дисциплину очереди, время обслуживания, максимальную длину очереди, а также источник клиентов.

   a) В мастерскую поступают заказы на выполнение работ. При их приемке диспетчер указывает, является ли заказ срочным или обычным. Для выполнения некоторых заказов требуется использование одного из нескольких одинаковых станков, которыми располагает мастерская. Остальные заказы выполняются на двухэтапной производственной линии; их в мастерской имеется две. В каждой из двух групп оборудования мастерской один станок предназначается для выполнения срочных работ.

b) Заявки, поступающие на некоторое обслуживающее устройство, выполняются в порядке их поступления. Выполненные заказы до момента их отправки клиентам складируются в специальном помещении зоны готовой продукции, которое имеет ограниченную вместимость.

c) Режущими инструментами различные станки обеспечиваются из центрального инструментального склада. При выходе станка из строя вызывается механик из централизованной службы ремонта для проведения ремонтных работ. Станки, выполняющие срочные заказы, всегда имеют преимущество как при получении нового инструмента со склада, так и при выполнении ремонтных работ.

4. Истинны или ложны следующие утверждения?

a) Если у ожидающего обслуживания клиента не хватает терпения, то он может отказаться от дальнейшего пребывания в системе обслуживания.

b) Если продолжительное время ожидания неприемлемо, поступивший клиент может отказаться от обслуживания.

c) Переход клиента из одной очереди в другую объясняется тем, что он надеется за счет этого сократить время ожидания.

5. В каждой ситуации, описанной в упр. 1, обсудите возможности клиентов, осуществляющих переход из очереди в очередь и отказывающихся от обслуживания по причине длительности ожидания, а также от дальнейшего пребывания в обслуживающей системе из-за медленного обслуживания.

17.3. Экспоненциальное распределение в системах массового обслуживания

В большинстве систем массового обслуживания поступление клиентов происходит случайным образом. Это означает, что наступление события (например, поступление клиента или завершение обслуживания) не зависит от времени, прошедшего с момента наступления предыдущего события.

Время между последовательными поступлениями клиентов и время их обслуживания, будучи случайными, при моделировании систем массового обслуживания количественно описываются **экспоненциальным распределением**, плотность вероятности которого имеет вид

$$f(t) = \lambda e^{-\lambda t}, t > 0,$$

где $M\{t\} = 1/\lambda$ (подробнее см. раздел 12.5.3). То, что экспоненциальное распределение является **совершенно случайным**, иллюстрируется следующим примером. Если сейчас 8:20 утра и некое событие имело место в 8:02 утра, то в соответствии с экспоненциальным законом распределения вероятность того, что следующее аналогичное событие произойдет в 8:29, является функцией лишь интервала времени от 8:20 до 8:29 и не зависит от интервала времени, прошедшего с момента наступления последнего события (от 8:02 до 8:20). Данное свойство экспоненциального распределения обычно называют **отсутствием последействия** или **отсутствием памяти**.

17.3.1. Свойство отсутствия последействия

Пусть время t наступления какого-либо события распределено по экспоненциальному закону с функцией плотности $f(t)$. Если S — время, прошедшее с момента наступления предыдущего события, то свойство отсутствия последействия выражается соотношением

$$P\{t > T + S \mid t > S\} = P\{t > T\}.$$

Для доказательства этого равенства заметим, что

$$P\{t > Y\} = 1 - P\{t < Y\} = e^{-\lambda Y}.$$

Следовательно,

$$P\{t > T + S \mid t > S\} = \frac{P\{t > T + S, t > S\}}{P\{t > S\}} = \frac{P\{t > T + S\}}{P\{t > S\}} = \frac{e^{-\lambda(T+S)}}{e^{-\lambda S}} = e^{-\lambda T} = P\{t > T\}.$$

Пример 17.3–1

При обслуживании сложного агрегата всегда имеется запасной блок для немедленной замены в случае поломки. Время выхода из строя агрегата (или его запасного блока) является случайной величиной, распределенной по экспоненциальному закону, и в среднем происходит каждые 40 минут. Оператор, обслуживающий агрегат, утверждает, что агрегат "имеет привычку" выходить из строя каждый вечер около 8:30. Проанализируем утверждение оператора.

Средняя интенсивность отказов агрегата равна $\lambda = 60/40 = 1.5$ отказа в час. Следовательно, плотность экспоненциального распределения времени отказа имеет вид

$$f(t) = 1.5e^{-1.5t}, \quad t > 0.$$

Что касается заявления оператора, то и без вычислений видно, что оно не может соответствовать действительности, так как не согласуется с тем, что время между отказами агрегата распределено по экспоненциальному закону и, следовательно, является случайным. Для подтверждения или опровержения заявления оператора нельзя использовать вероятность того, что отказ будет происходить в 8:30 вечера, так как вероятность такого события зависит от времени дня (относительно 8:30 вечера), когда эта вероятность вычисляется. Например, если вычисления выполняются в 8:20 вечера, то вероятность того, что утверждение оператора окажется справедливым этим вечером, равна

$$P\left\{t < \frac{10}{60}\right\} = 1 - e^{-1.5(10/60)} = 0.22,$$

т.е. является очень малой. Если вычисления выполняются в 7:00 вечера, то вероятность того, что отказ будет иметь место в 8:30, возрастает примерно до 0.9 (проверьте!). Эти два крайних значения вероятности показывают, что достоверность утверждения оператора нельзя проанализировать на основе полученных вероятностей; в данной ситуации мы должны полагаться только на характеристики экспоненциального распределения (точнее, на его свойство отсутствия последействия).

Упражнения 17.3,а

1. a) Объясните связь между интенсивностью поступления заявок на обслуживание λ и средним временем между последовательными их поступлениями. В каких единицах измеряется каждая из этих величин?

   b) В каждом из следующих случаев определите среднюю интенсивность (в час) поступлений заявок на обслуживание λ и среднее время между их последовательными поступлениями.

      i) Каждые 10 минут происходит одно поступление.
      ii) Каждые 6 минут происходит два поступления.
      iii) Число поступлений за 30 минут равно 10.
      iv) Средний интервал между последовательными поступлениями равен 0.5 часа.

   c) В каждом из следующих случаев определите среднюю (в час) интенсивность обслуживания μ и среднее время обслуживания.

      i) Каждые 12 минут выполняется одно обслуживание.
      ii) Каждые 15 минут две обслуженные заявки покидают систему.
      iii) Число обслуженных клиентов за 30 минут равно 5.
      iv) Среднее время обслуживания равно 0.3 часа.

2. В примере 17.3–1 определите следующие показатели.

   a) Среднее число отказов за одну неделю, если система обслуживания функционирует 7 дней в неделю по 24 часа в день.
   b) Вероятность по крайней мере одного отказа за два часа.
   c) Вероятность того, что следующий отказ не произойдет на протяжении трех часов.
   d) Если после последнего отказа на протяжении трех часов других отказов не было, то какова вероятность того, что время между последовательными отказами системы равно по крайней мере 4 часа?

3. Время между последовательными поступлениями клиентов в Управление департамента государственных сборов распределено по экспоненциальному закону со средним значением 0.05 часа. Управление начинает работу в 8:00 утра.

   a) Определите плотность вероятности экспоненциального распределения, описывающего время между последовательными поступлениями клиентов.
   b) Определите вероятность того, что до 8:15 утра в Управлении клиентов не будет.
   c) Сейчас 8:35 утра. Последний клиент прибыл в Управление в 8:26. Какова вероятность того, что следующий клиент прибудет до 8:38 утра? Какова вероятность того, что следующего клиента не будет до 8:40 утра?
   d) Чему равно среднее число посетителей, которые прибудут в Управление от 8:10 до 8:45 утра?

4. Пусть время между отказами механизма распределено по экспоненциальному закону со средним 6 часов. Если механизм работал безотказно на протяжении последних трех часов, то какова вероятность того, что не будет отказа на протяжении следующего часа? Найдите также вероятность того, что на протяжении следующего получаса произойдет отказ.

17.3.2. Определение экспоненциального распределения

Чтобы уяснить свойства экспоненциального распределения, сформулируем основные аксиомы, на которых базируется это распределение.

Теорема 17.3–1. *Экспоненциальное распределение основано на трех аксиомах.*

Аксиома 1. *Если $N(t)$ — число событий, происшедших на протяжении интервала времени $(0, t)$, то вероятностный процесс, описывающий $N(t)$, имеет независимые стационарные приращения; вероятность наступления события в интервале $(T, T + S)$ зависит лишь от его длины S.*

Аксиома 2. *Вероятность того, что событие наступит на достаточно малом временном интервале $h > 0$, положительна, но меньше 1.*

Аксиома 3. *На достаточно малом временном интервале $h > 0$ может осуществиться не более одного события, т.е. $P\{N(h) > 1\} = 0$.*

Доказательство. Обозначим через $p_n(t)$ вероятность наступления n событий на временном интервале длиной t. В соответствии с аксиомой 1 вероятность того, что на интервале длиной $t + h$, где $h > 0$ и достаточно мало, не наступит ни одного события, равна

$$p_0(t+h) = p_0(t)\,p_0(h).$$

Опираясь на две другие аксиомы, можно показать, что решением приведенного выше уравнения есть следующее

$$p_0(t) = e^{-\lambda t},\ t > 0,$$

где λ — положительная константа. (Детали доказательства можно найти в книге [4].)

Пусть $f(t)$ — плотность вероятности распределения длины временного интервала между последовательными наступлениями случайного события ($t > 0$). Очевидно, что

$$P\left\{\begin{matrix}\text{интервал времени между моментами}\\ \text{наступления двух последовательных}\\ \text{случайных событий не меньше } T\end{matrix}\right\} = P\left\{\begin{matrix}\text{в течение } T\\ \text{событие не}\\ \text{наступает}\end{matrix}\right\}.$$

В математической форме это выглядит следующим образом:

$$\int_T^{\infty} f(t)dt = p_0(T),\ \ T > 0.$$

Подставив выражение для $p_0(T)$ и выполнив небольшие преобразования, получаем следующее.

$$\int_0^{T} f(t)dt = 1 - e^{-\lambda T},\ \ T > 0.$$

Дифференцируя правую и левую части этого соотношения по T, получаем выражение для плотности экспоненциального распределения

$$f(t) = \lambda e^{-\lambda t},\ t > 0.$$

Математическое ожидание экспоненциального распределения равно $1/\lambda$ временных единиц, где λ — интенсивность (количество событий в единицу времени), с которой появляются события.

Упражнения 17.3,b

1. Истинны или ложны следующие утверждения?
   a) В условиях экспоненциального распределения на протяжении бесконечно малого интервала времени могут произойти два события.
   b) Если интервалы времени между последовательными поступлениями заявок в систему распределены по экспоненциальному закону, то вероятность того, что заявка не поступит на протяжении интервала h, положительна, независимо от того, насколько малым является интервал h.
2. Время между последовательными поступлениями клиентов в игротеку распределено по экспоненциальному закону с математическим ожиданием 10 минут.
   a) Какова интенсивность прихода клиентов в час?
   b) Какова вероятность того, что на протяжении следующих 15 минут в игротеку не придет ни один клиент?
   c) Какова вероятность того, что на протяжении следующих 20 минут в игротеку придет по крайней мере один клиент?
3. Помогите управляющему нового ресторана быстрого питания описать количественно процесс поступления посетителей, оценивая долю интервалов времени между их приходами, которые будут (а) меньше двух минут, (б) больше трех минут и (в) от двух до трех минут. Интенсивность поступления посетителей в рестораны подобного типа равна 35 клиентов в час. Время между приходами последовательных посетителей распределено по экспоненциальному закону.
4. Официанты O_1 и O_2 ресторана быстрого питания в ожидании посетителей заняты следующей игрой: если в течение одной минуты в ресторан не прибудет ни один посетитель, O_1 платит 2 цента O_2, в противном случае 2 цента от O_2 получает O_1. Требуется вычислить средний выигрыш официанта O_1 за восьмичасовой период. Время между последовательными прибытиями посетителей распределено по экспоненциальному закону со средним значением 1.5 минуты.
5. Пусть в предыдущей ситуации (упр. 4) правила игры таковы, что официант O_1 платит 2 цента O_2, если следующий посетитель прибывает через 1.5 минуты после предыдущего, а официант O_2 платит такую же сумму O_1, если очередной промежуток между последовательными прибытиями посетителей не превышает одной минуты. Если последовательные прибытия посетителей происходят в пределах от 1 до 1.5 минут, игра заканчивается вничью. Требуется вычислить средний выигрыш официанта O_1 за восьмичасовой период.
6. Предположим, что в ситуации из упр. 4 официант O_2 платит 2 цента O_1, если следующий посетитель прибывает после предыдущего в пределах 1 минуты, и 3 цента, если это происходит в пределах от 1 до 1.5 минут. Официант O_2 получает 5 центов от O_1, если интервал между прибытиями следующего и предыдущего посетителей находится в пределах от 1.5 до 2 минут, и 6 центов, если это время больше 2 минут. Требуется вычислить средний выигрыш официанта O_2 за восьмичасовой период.

7. Посетитель ресторана быстрого питания, который приходит в пределах четырехминутного интервала после предыдущего посетителя, обслуживается без очереди. Если же время между последовательными приходами посетителей составляет от 4 до 5 минут, время ожидания будет около 1 минуты. Если же время между последовательными приходами посетителей больше 5 минут, время ожидания составляет около 2 минут. Время между последовательными приходами посетителей является случайной величиной, распределенной по экспоненциальному закону со средним значением 6 минут.

   a) Определите вероятность того, что прибывающий посетитель не будет ожидать в очереди.
   b) Определите среднее время ожидания для прибывающего посетителя.

8. Известно, что время между отказами в работе холодильника является случайной величиной, распределенной по экспоненциальному закону со средним значением 9000 часов (примерно один год эксплуатации), и производящая компания выдает на холодильник годичную гарантию. Какова вероятность того, что холодильник потребует ремонта во время гарантийного срока?

9. В университетском городке функционируют две автобусные линии: красная и зеленая. Красная обслуживает северную часть городка, зеленая — южную. Автобусные линии связаны пересадочной станцией. Время между прибытиями автобусов зеленой линии на пересадочную станцию является случайной величиной, распределенной по экспоненциальному закону со средним значением 10 минут. Аналогичный показатель для автобусов красной линии равен 7 минут.

   a) Определите распределение времени ожидания студента, который прибывает по красной линии для пересадки на зеленую.
   b) Определите распределение времени ожидания студента, который прибывает по зеленой линии для пересадки на красную.

10. Докажите, что математическое ожидание и стандартное отклонение для экспоненциального распределения совпадают.

17.4. Модели рождения и гибели (связь между экспоненциальным и пуассоновским распределениями)

В данном разделе рассматриваются две модели обслуживающих систем: в первой представлены только поступления клиентов (модель чистого рождения), во второй — только выход клиентов из системы (модель чистой гибели). Примером модели чистого рождения является процесс оформления свидетельств о рождении детей. В качестве модели чистой гибели может служить случайное изъятие хранящихся на складе запасов.

Данные модели строятся на основе аксиом теоремы 17.3–1. Побочным продуктом этих построений является демонстрация тесной связи между экспоненциальным распределением и распределением Пуассона в том смысле, что одно из них автоматически определяет другое.

17.4.1. Модель чистого рождения

При заданной интенсивности λ поступлений клиентов в систему обслуживания и достаточно малом интервале времени $h > 0$ из теоремы 17.3–1 следует, что

$$p_0(h) = e^{-\lambda h} = 1 - \lambda h + \frac{(\lambda h)^2}{2!} - \ldots = 1 - \lambda h + O(h^2).$$

В соответствии с аксиомой 3 на достаточно малом временном интервале $h > 0$ может наступить не более одного события (поступления клиента). Следовательно, при $h \to 0$

$$p_1(h) = 1 - p_0(h) \approx \lambda h.$$

Этот результат показывает, что вероятность поступления клиента на протяжении интервала h прямо пропорциональна h с коэффициентом пропорциональности, равным интенсивности поступлений λ.

Чтобы на основе аксиом теоремы 17.3–1 получить распределение Пуассона, обозначим через $p_n(t)$ вероятность поступления n клиентов на протяжении времени t. Следовательно, при достаточно малом $h > 0$ имеем следующее.

$$p_n(t + h) \approx p_n(t)(1 - \lambda h) + p_{n-1}(t)\lambda h, \quad n > 0,$$
$$p_0(t + h) \approx p_0(t)(1 - \lambda h), \quad n = 0.$$

Из первого уравнения следует, что поступление n клиентов на протяжении времени $t + h$ возможно в двух случаях: если имеется n поступлений на протяжении времени t и нет поступлений за время h или имеется $n - 1$ поступлений за время t и одно поступление за время h. Любые другие комбинации невозможны вследствие аксиомы 3 теоремы 17.3–1. В соответствии с условием независимости стационарных приращений аксиомы 1 к правой части уравнения применим закон умножения вероятностей. Во втором уравнении отсутствие поступлений клиентов на протяжении интервала $t + h$ может иметь место лишь в случае, когда нет поступлений клиентов за время h.

Перегруппировывая члены и переходя к пределу при $h \to 0$, получаем следующее.

$$p_n'(t) = \lim_{h \to 0} \frac{p_n(t+h) - p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n > 0,$$

$$p_0'(t) = \lim_{h \to 0} \frac{p_0(t+h) - p_0(t)}{h} = -\lambda p_0(t),$$

где $p_n'(t)$ — производная по t функции $p_n(t)$.

Решение приведенных выше разностно-дифференциальных уравнений имеет следующий вид:

$$p_n(t) = \frac{(\lambda t)^n e^{-\lambda t}}{n!}, \quad n = 0, 1, 2, \ldots .$$

В данном случае мы получили дискретную плотность вероятности **распределения Пуассона** с математическим ожиданием $M\{n \mid t\} = \lambda t$ поступлений за время t. Дисперсия распределения Пуассона также равна λt.

Полученный результат означает, что всякий раз, когда временные интервалы между моментами последовательных поступлений заявок распределены по экспоненциальному закону с математическим ожиданием $1/\lambda$, число поступлений заявок в интервале, равном t единиц времени, характеризуется распределением Пуассона с математическим ожиданием λt. Верным является и обратное утверждение.

Пример 17.4–1

В небольшом штате дети рождаются с интенсивностью одно рождение каждые 12 минут. Время между рождениями распределено по экспоненциальному закону. Требуется определить следующее:

a) среднее число рождений за год;
b) вероятность того, что на протяжении одного дня не будет ни одного рождения;
c) вероятность выдачи 50 свидетельств о рождении к концу третьего часа, если известно, что на протяжении последних двух часов было выдано 40 таких свидетельств.

Вычислим интенсивность рождений за день: $\lambda = \dfrac{24 \times 60}{12} = 120$ рождений за день.

Интенсивность рождений в штате за год равна $\lambda t = 120 \times 365 = 43800$ рождений.

Вероятность того, что на протяжении одного дня не будет ни одного рождения, вычисляется с использованием пуассоновского распределения

$$p_0(1) = \frac{(120 \times 1)^0 e^{-120\times 1}}{0!} \approx 0.$$

Для вычисления вероятности выдачи 50 свидетельств о рождении к концу третьего часа при условии, что на протяжении последних двух часов было выдано 40 таких свидетельств, заметим, что поскольку распределение числа рождений является пуассоновским, искомая вероятность сводится к вероятности появления 10 (= 50 – 40) рождений за один (= 3 – 2) час. Так как $\lambda = 60/12 = 5$ рождений за час, то

$$p_{10}(1) = \frac{(5 \times 1)^{10} e^{-5\times 1}}{10!} = 0{,}01813.$$

Формулы для вычисления функциональных параметров систем обслуживания включают громоздкие вычисления, поэтому для их выполнения желательно использовать программное обеспечение TORA. На рис. 17.2 показаны выходные данные, полученные с помощью программы TORA, для модели чистого рождения с интенсивностью $\lambda t = 5 \times 1 = 5$ рождений за день.

Упражнения 17.4,а

1. Пусть в примере 17.4–1 служащий, который вводит информацию из свидетельств о рождении в компьютер, обычно ожидает пока не накопится, по крайней мере, пять сертификатов. Определите вероятность того, что служащий будет вводить новый пакет данных каждый час.

```
Problem title: Example 17.4-1
Scenario 1 -- Pure Birth Model
------------------------------------------------------------
Poisson with Lambda*t =  5.00000
------------------------------------------------------------
Values of p(n) for n=0 to 17, else p(n) < .00001

 0  0.00674    1  0.03369    2  0.08422    3  0.14037    4  0.17547
 5  0.17547    6  0.14622    7  0.10444    8  0.06528    9  0.03627
10  0.01813   11  0.00824   12  0.00343   13  0.00132   14  0.00047
15  0.00016   16  0.00005   17  0.00001

Cumulative values of p(n) for n=0 to 17

 0  0.00674    1  0.04043    2  0.12465    3  0.26503    4  0.44049
 5  0.61596    6  0.76218    7  0.86663    8  0.93191    9  0.96817
10  0.98630   11 0.99455   12  0.99798   13 0.99930   14  0.99977
15  0.99993   16 0.99998   17  0.99999
```

Рис. 17.2

2. Коллекционер произведений искусства в среднем раз в месяц ездит на художественные аукционы. Каждая поездка гарантирует в точности одну покупку. Время между поездками имеет экспоненциальное распределение. Определите следующие параметры.

   a) Вероятность того, что коллекционер на протяжении трехмесячного периода не купит ни одного произведения искусства.
   b) Вероятность того, что коллекционер приобретет не более восьми произведений искусства на протяжении года.
   c) Вероятность того, что интервал между двумя последовательными поездками коллекционера превысит один месяц.

3. В течение очень малого временного интервала h в обслуживающую систему банка может поступить не более одной заявки на обслуживание. Вероятность того, что поступление заявки действительно будет иметь место, прямо пропорциональна h с коэффициентом пропорциональности 2. Требуется определить следующие параметры.

   a) Среднее число заявок на протяжении пяти временных единиц.
   b) Вероятность того, что в течение 0.5 единицы времени не поступит ни одной заявки.
   c) Вероятность того, что по крайней мере одна заявка поступит в течение 0.5 единицы времени.
   d) Вероятность того, что промежуток времени между двумя последовательными поступлениями заявок на обслуживание равен по меньшей мере трем временным единицам. (*Совет.* Используйте аксиому 3 теоремы 17.3–1.)

4. Время между прибытиями посетителей в ресторан распределено по экспоненциальному закону со средним значением 5 минут. Ресторан открывается в 11:00 утра. Требуется вычислить следующее.

   a) Вероятность того, что в 11:12 в ресторане окажется 10 посетителей при условии, что в 11:05 в ресторане было 8 посетителей.
   b) Вероятность того, что новый посетитель прибудет в ресторан в интервале между 11:28 и 11:33, если известно, что предыдущий посетитель прибыл в ресторан в 11:25.

5. Заказанные публичной библиотекой книги поступают в соответствии с пуассоновским распределением со средним значением 25 книг в день. На каждой полке в хранилищах можно разместить 100 книг. Требуется вычислить следующее.

   a) Среднее количество полок, ежемесячно заполняемых новыми книгами.
   b) Вероятность того, что ежемесячно для размещения поступающих книг потребуется более 10 книжных шкафов, если известно, что каждый книжный шкаф состоит из пяти полок.

6. В университетском городке функционируют две автобусные линии: красная и зеленая. Красная обслуживает северную часть городка, зеленая — южную. Автобусные линии связаны пересадочной станцией. Время между прибытиями автобусов зеленой линии на пересадочную станцию является случайной величиной, распределенной по пуассоновскому закону с математическим ожиданием 10 минут. Аналогичный показатель для автобусов красной линии равен 7 минут.

   a) Какова вероятность того, что оба автобуса остановятся на пересадочной станции на протяжении пятиминутного интервала?
   b) У студента, чье общежитие находится рядом с пересадочной станцией, через 10 минут лекция. Любой автобус доставит студента к учебному зданию. Поездка занимает 5 минут, затем студент должен около 3 минут идти пешком к учебному зданию. Какова вероятность того, что студент вовремя придет на лекцию?

7. Докажите, что среднее и дисперсия распределения Пуассона на интервале t равны λt, где λ — интенсивность поступления заявок.

8. Получите распределение Пуассона из разностно-дифференциальных уравнений модели чистого рождения. *Совет.* Решением общего дифференциального уравнения

$$y + a(t)y = b(t)$$

является функция

$$y = e^{-\int a(t)dt}\left\{\int b(t)e^{\int a(t)dt}dt + \text{константа}\right\}.$$

17.4.2. Модель чистой гибели

В данной модели предполагается, что система начинает функционировать, когда в момент времени 0 в ней имеется N клиентов и не допускается ни одного нового поступления клиента. Клиенты после завершения их обслуживания выбывают из системы с интенсивностью μ клиентов в единицу времени. Пусть $p_n(t)$ — вероятность того, что после t

временных единиц в системе остается n клиентов. Для получения разностно-дифференциальных уравнений относительно $p_n(t)$ обычно следуют логике рассуждений, использованных в модели чистого рождения (раздел 17.4–1). Поэтому имеем

$$p'_N(t) = -\mu p_N(t),$$

$$p'_n(t) = -\mu p_n(t) + \mu p_{n+1}(t), \quad 0 < n < N,$$

$$p'_0(t) = \mu p_1(t).$$

Эти уравнения имеют решение

$$p_n(t) = \frac{(\mu t)^{N-n} e^{-\mu t}}{(N-n)!}, \quad n = 1, 2, \ldots, N,$$

$$p_0(t) = 1 - \sum_{n=1}^{N} p_n(t),$$

которое называется **усеченным распределением Пуассона**.

Пример 17.4–2

Секция цветов бакалейно-гастрономического магазина складирует 18 дюжин роз в начале каждой недели. В среднем продается 3 дюжины роз в день (за один раз продается дюжина роз), но действительный спрос подчиняется распределению Пуассона. Как только уровень запаса снижается до 5 дюжин, делается новый заказ на поставку 18 дюжин роз в начале следующей недели. Запасы по своей природе таковы, что все неиспользованные до конца недели розы приходят в негодность и ликвидируются. Требуется вычислить следующие параметры системы.

a) Вероятность размещения заказа к концу каждого дня недели.

b) Среднее количество роз, которые будут ликвидированы к концу недели.

Так как розы покупаются с интенсивностью $\mu = 3$ дюжины в день, то вероятность того, что заказ будет размещен в конце дня t, равна

$$p_{n \le 5}(t) = p_0(t) + p_1(t) + \ldots + p_5(t) = p_0(t) + \sum_{n=1}^{5} \frac{(3t)^{18-n} e^{-3t}}{(18-n)!}, \quad t = 1, 2, \ldots, 7.$$

Используя многочисленные сценарии программного обеспечения TORA, соответствующие $t = 1, 2, \ldots, 7$, получаем следующее.

t (дни)	1	2	3	4	5	6	7
μt	3	6	9	12	15	18	21
$p_{n \le 5}(t)$	0.0000	0.0088	0.1242	0.4240	0.7324	0.9083	0.9755

Среднее количество роз, которые будут ликвидированы к концу недели ($t = 7$), вычисляется (с использованием программы TORA) следующим образом.

$$M\{n \mid t = 7\} = \sum_{n=0}^{18} n p_n(7) = 0.664 \text{ дюжины.}$$

Упражнения 17.4,b

1. На основе примера 17.4–2 выполняется следующее.

   a) Используйте программу TORA для проверки значений $p_{n \le 5}(t)$ при $t = 1, 2, ..., 7$.

   b) Используйте программу TORA для вычисления $p_n(7)$, $n = 1, 2, ..., 18$, и затем проверьте, что эти вероятности дают значение $M\{n \mid t = 7\} = 0.664$ дюжины.

2. В примере 17.4–2 определите следующее.

   a) Вероятность того, что за три дня запас роз будет исчерпан.

   b) Среднее количество роз, оставшихся к концу второго дня.

   c) Вероятность того, что по крайней мере одна дюжина роз будет продана в течение четвертого дня, если последняя покупка роз была в конце третьего дня.

   d) Вероятность того, что интервал времени до следующей покупки роз не превышает полдня, если последняя покупка была днем раньше.

   e) Вероятность того, что на протяжении первого дня не будет продано ни одной дюжины роз.

3. Джазовый оркестр средней школы дает концерт в зале на 400 мест. Местные фирмы покупают билеты блоками по 10 билетов и дарят их молодежным организациям. Этим фирмам билеты продаются лишь 4 часа в день перед концертом. Процесс заказа билетов по телефону является пуассоновским со средним значением, равным 10 звонков в час. Билеты, оставшиеся после закрытия кассы, продаются со скидкой (как "срочные") за час перед началом концерта. Требуется определить следующее.

   a) Вероятность того, что можно будет купить "срочные" билеты.

   b) Среднее значение количества "срочных" билетов.

4. Каждое утро в холодильник небольшой мастерской помещается два ящика (по 24 банки) безалкогольных напитков для десяти работников. Они могут утолять свою жажду в любой момент на протяжении восьмичасового рабочего дня (с 8:00 до 16:00). Процесс потребления напитков является случайным (распределение Пуассона), но известно, что в среднем каждый работник употребляет примерно 2 банки в день. Какова вероятность того, что запас напитков исчерпается к полудню? К моменту закрытия мастерской?

5. Студент-первокурсник ежемесячно получает от родителей банковский депозит на 100 долларов для покрытия текущих расходов. Получение студентом денег чеками по 20 долларов каждый на протяжении месяца происходит случайным образом в соответствии с экспоненциальным законом со средним значением 1 раз в неделю. Определите вероятность того, что к концу месяца (т.е. к концу четвертой недели) у студента не будет денег на текущие расходы.

6. На складе находится 80 единиц продукции, которая изымается в соответствии с распределением Пуассона с интенсивностью 5 единиц в день. Требуется определить следующее.

   a) Вероятность того, что за два дня из склада будет изъято 10 единиц продукции.

   b) Вероятность того, что к концу четвертого дня на складе не останется ни одной единицы продукции.

   c) Среднее количество изъятых единиц продукции на протяжении четырех дней.

7. Ремонтный цех предприятия только что складировал 10 комплектов запасных частей для ремонта автомобилей данного предприятия. Пополнение запаса в таком же объеме происходит каждые 7 дней. Время между поломками автомобилей является случайной величиной, распределенной по экспоненциальному закону со средним значением, равным одному дню. Определите вероятность того, что автомобиль 2 дня будет находиться в неисправном состоянии по причине отсутствия запасных частей.

8. Объем спроса на изделие является случайной величиной, распределенной по закону Пуассона со средним значением 3 единицы в день. Максимальная вместимость склада равна 25 единицам. Склад полностью заполняется каждый понедельник сразу же после получения нового заказа. Объем заказа зависит от количества изделий, оставшихся к концу недели в субботу (воскресенье — выходной день). Требуется определить следующие параметры.

   a) Средний недельный объем заказа.
   b) Вероятность отсутствия запаса утром в пятницу.
   c) Вероятность того, что недельный объем заказа превысит 10 единиц.

9. Докажите, что в модели чистой гибели распределение времени между удалениями (подчиняющимися усеченному распределению Пуассона) клиентов из системы является экспоненциальным с математическим ожиданием $1/\mu$ единиц времени.

10. Получите усеченное распределение Пуассона из разностно-дифференциальных уравнений модели чистой гибели с помощью метода индукции. (*Подсказка.* См. указание к упр. 17.4,a(8).)

17.5. Обобщенная модель системы массового обслуживания

В данном разделе рассматриваются общие системы массового обслуживания, в которых имеется как входной поток клиентов, так и выходной поток обслуженных клиентов. Время между последовательными поступлениями клиентов и время обслуживания являются экспоненциально распределенными случайными величинами. Эта модель является основой при рассмотрении специализированных моделей Пуассона, которым посвящен раздел 17.6.

При рассмотрении общих систем массового обслуживания предполагается, что система функционирует в течение достаточно большого интервала времени, по истечении которого в ее работе наступает **стационарный** режим. Этот режим функционирования обслуживающей системы противопоставляется **переходному** (или **неустановившемуся**) режиму, который превалирует в самый начальный период функционирования системы. В этой главе не рассматриваются переходные режимы работы систем массового обслуживания, поскольку, во-первых, это связано с серьезными математическими трудностями, а, во-вторых, на практике данные системы обычно предназначаются для работы в течение весьма длительного времени.

В рассматриваемой в этом разделе общей модели системы массового обслуживания предполагается, что и интенсивность поступления клиентов, и интенсивность выходного потока *зависят от состояния* системы, что означает их зависимость от числа клиентов в системе обслуживания. Например, сборщик платы за проезд по автомагистрали в часы интенсивного движения стремится ускорить сбор пошлины. Или в мастерской с фиксированным количеством станков интенсивность их поломки убывает по мере возрастания числа аварийных станков, ибо лишь работающие станки могут выходить из строя.

Введем следующие обозначения.

n — число клиентов в системе обслуживания (в очереди и на обслуживании),

λ_n — интенсивность поступления в систему клиентов при условии, что в системе уже находится n клиентов,

μ_n — интенсивность выходного потока обслуженных клиентов при условии, что в системе находится n клиентов,

p_n — вероятность того, что в системе находится n клиентов.

В общей модели системы массового обслуживания устанавливается функциональная зависимость вероятностей p_n от λ_n и μ_n. Эти вероятности используются затем при определении функциональных характеристик обслуживающей системы, таких как средняя длина очереди, среднее время ожидания и средний коэффициент использования сервисов.

Вероятности p_n определяются из **диаграммы интенсивностей переходов**, представленной на рис. 17.3. Обслуживающая система находится в состоянии n, если в ней имеется n клиентов. Из аксиом пуассоновского процесса, приведенных в теореме 17.3–1, следует, что вероятность появления более одного нового клиента на протяжении малого промежутка времени h стремится к нулю при $h \to 0$. Это означает, что при $n > 0$ состояние n может быть изменено в двух возможных направлениях: $n - 1$, когда с интенсивностью μ_n обслуженный клиент выбывает из системы, и $n + 1$, когда имеет место поступление клиента с интенсивностью λ_n. Состояние 0 может измениться лишь к состоянию 1, когда имеет место поступление клиента с интенсивностью λ_0. Заметим, что μ_0 не определено, так как не может происходить выбывания клиентов из пустой системы обслуживания.

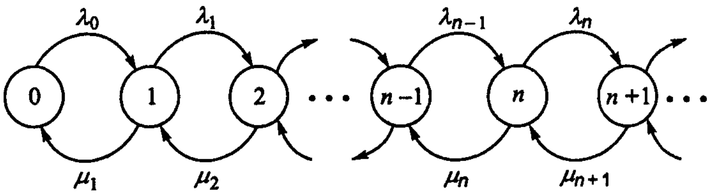


Рис. 17.3

При выполнении условий стационарности *ожидаемые* интенсивности входного и выходного потоков в состоянии n ($n > 0$) должны быть равны. Так как состояние n может изменяться лишь к состояниям $n - 1$ и $n + 1$, отсюда следует

$$\begin{pmatrix}\text{Ожидаемая интенсивность}\\ \text{входного потока в состоянии } n\end{pmatrix} = \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1}.$$

Аналогично,

$$\begin{pmatrix}\text{Ожидаемая интенсивность}\\ \text{выходного потока в состоянии } n\end{pmatrix} = (\lambda_n + \mu_n) p_n.$$

Отсюда, приравнивая эти две интенсивности, получаем следующее **уравнение баланса**.

$$\lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1} = (\lambda_n + \mu_n)p_n, \quad n = 1, 2, \ldots .$$

Как видно из рис. 17.3, уравнение баланса, соответствующее $n = 0$, имеет вид

$$\lambda_0 p_0 = \mu_1 p_1.$$

Уравнения баланса решаются рекуррентно, последовательно выражая вероятности p_i через p_0 следующим образом: для $n = 0$ имеем

$$p_1 = \left(\frac{\lambda_0}{\mu_1}\right)p_0.$$

Далее, для $n = 1$ получаем

$$\lambda_0 p_0 + \mu_2 p_2 = (\lambda_1 + \mu_1)p_1.$$

Подставляя сюда $p_1 = (\lambda_0/\mu_1)p_0$ и упрощая полученное выражение, имеем (проверьте!)

$$p_2 = \left(\frac{\lambda_1\lambda_0}{\mu_2\mu_1}\right)p_0.$$

Методом индукции можно показать, что

$$p_n = \left(\frac{\lambda_{n-1}\lambda_{n-2}\ldots\lambda_0}{\mu_n\mu_{n-1}\ldots\mu_1}\right)p_0, \quad n = 1, 2, \ldots .$$

Значение p_0 определяется из уравнения $\sum_{n=0}^{\infty} p_n = 1$.

Пример 17.5–1

Бакалейный магазин работает с тремя кассами. Вывеска возле касс извещает покупателей, что в любой момент будет открыта дополнительная касса, как только число покупателей в любой очереди превысит 3. Это означает, что если число покупателей меньше четырех, то работать будет лишь одна касса. Если число покупателей от четырех до шести, то будет работать две кассы. Если имеется больше шести покупателей, будут открыты все три кассы. Покупатели подходят к кассам в соответствии с распределением Пуассона с математическим ожиданием 10 человек в час. Время расчета одного покупателя в кассе распределено по экспоненциальному закону со средним 12 минут. Определим в установившемся режиме вероятность p_n нахождения n покупателей возле касс.

Из формулировки задачи имеем следующее.

$$\lambda_n = \lambda = 10 \text{ покупателей в час, } n = 0, 1, \ldots,$$

$$\mu_n = \begin{cases} \frac{60}{12} = 5 & \text{покупателей в час, } n = 0,1,2,3, \\ 2\times5 = 10 & \text{покупателей в час, } n = 4,5,6, \\ 3\times5 = 15 & \text{покупателей в час, } n = 7,8,\ldots . \end{cases}$$

Следовательно,

$$p_1 = \left(\frac{10}{5}\right) p_0 = 2p_0,$$

$$p_2 = \left(\frac{10}{5}\right)^2 p_0 = 4p_0,$$

$$p_3 = \left(\frac{10}{5}\right)^3 p_0 = 8p_0,$$

$$p_4 = \left(\frac{10}{5}\right)^3 \left(\frac{10}{10}\right) p_0 = 8p_0,$$

$$p_5 = \left(\frac{10}{5}\right)^3 \left(\frac{10}{10}\right)^2 p_0 = 8p_0,$$

$$p_6 = \left(\frac{10}{5}\right)^3 \left(\frac{10}{10}\right)^3 p_0 = 8p_0,$$

$$p_n = \left(\frac{10}{5}\right)^3 \left(\frac{10}{10}\right)^3 \left(\frac{10}{15}\right)^{n-6} p_0 = 8\left(\frac{2}{3}\right)^{n-6} p_0, \quad n = 7, 8, \ldots .$$

Значение p_0 определяется из уравнения

$$p_0 + p_0\left[2+4+8+8+8+8+8\left(\frac{2}{3}\right)+8\left(\frac{2}{3}\right)^2+8\left(\frac{2}{3}\right)^3+\ldots\right]=1,$$

или, что равносильно,

$$p_0\left\{31+8\left[1+\left(\frac{2}{3}\right)+\left(\frac{2}{3}\right)^2+\ldots\right]\right\}=1.$$

Используя формулу суммы бесконечной геометрической прогрессии

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}, \quad |x| < 1,$$

получаем следующее.

$$p_0\left[31+8\left(\frac{1}{1-\frac{2}{3}}\right)\right]=1.$$

Следовательно, $p_0 = 1/35$.

Зная p_0, можно определить любую вероятность, имеющую отношение к задаче. Например, вероятность того, что будет работать лишь одна касса, вычисляется как вероятность нахождения в системе не больше трех клиентов, т.е.

$$p_0 + p_1 + p_2 + p_3 = (1 + 2 + 4 + 8)(1/35) \approx 0.273.$$

Здесь предполагается, что в бакалейном магазине будет открыта, как минимум, одна касса, даже в том случае, когда вовсе нет покупателей.

Вероятности p_n можно использовать для определения численных значений функциональных характеристик рассматриваемой системы. Например, среднее количество неработающих касс равно

$$3p_0 + 2(p_1 + p_2 + p_3) + 1(p_4 + p_5 + p_6) + 0(p_7 + p_8 + ...) = 1 \text{ касса.}$$

Упражнения 17.5,а

1. В примере 17.5–1 определите следующие характеристики системы.
   a) Распределение вероятностей количества работающих касс.
   b) Среднее количество работающих касс.
   c) Среднее количество неработающих касс.

2. Пусть в задаче о бакалейном магазине из примера 17.5–1 время между подходом покупателей к кассам распределено по экспоненциальному закону со средним 5 минут, а время расчета одного покупателя в кассе распределено по такому же закону со средним 10 минут. Предположим далее, что будет установлена четвертая касса и все кассы будут открываться при увеличении числа покупателей на два. Требуется определить следующие величины.
   a) Вероятности p_n установившегося режима для всех n.
   b) Вероятность того, что четвертая касса будет открыта.
   c) Среднее количество неработающих касс.

3. Пусть в задаче о бакалейном магазине из примера 17.5–1 три кассы всегда открыты и обслуживание организовано так, что покупатель будет подходить к первой свободной кассе. Определите следующие величины.
   a) Вероятность того, что покупатели будут у всех трех касс.
   b) Вероятность того, что подошедший к кассам покупатель не будет ожидать обслуживания.

4. Банк имеет один пункт, где клиенты обслуживаются банковским автоматом не выходя из автомобиля. Автомобили прибывают в соответствии с распределением Пуассона с интенсивностью 12 автомобилей в час. Время, необходимое для обслуживания клиента банкоматом, распределено по экспоненциальному закону со средним, равным 6 минут. Максимальная вместимость полосы обслуживания банкоматом составляет 10 автомобилей. При заполненной полосе прибывающие клиенты должны искать обслуживание в другом банке. Определите следующие величины.
   a) Вероятность того, что прибывающий клиент не сможет воспользоваться услугами банковского автомата из-за того, что полоса обслуживания будет заполнена.
   b) Вероятность того, что прибывающий клиент не сможет воспользоваться услугами банковского автомата без ожидания.
   c) Среднее количество автомобилей в полосе обслуживания.

5. Вы слышали когда-либо, чтобы кто-то повторял противоречивое заявление: "Это место настолько переполнено, что больше туда никто не ходит"? Это утверждение можно интерпретировать в том смысле, что возможность отказа от обслуживания возрастает с ростом числа клиентов, которые нуждаются в обслуживании. Возможной основой для моделирования этой ситуации является утверждение о том, что интенсивность входного в систему потока клиентов уменьшается, если число клиентов в системе возрастает. Для конкретности рассмотрим упрощенную модель бильярдного клуба, куда посетители обычно приходят парами для игры в бильярд. Нормальная интенсивность прихода клиентов равна шести парам в час. Однако если число пар в бильярдном клубе превышает восемь, интенсивность поступления клиентов уменьшается до 5 пар в час. Предполагается, что входной поток подчиняется распределению Пуассона. Время игры каждой пары является случайной величиной, распределенной по экспоненциальному закону с математическим ожиданием 30 минут. Бильярдный клуб имеет в своем распоряжении 5 бильярдных столов и одновременно может расположить не более 12 пар. Определите следующие величины.

   a) Вероятность того, что клиенты начнут отказываться от сервиса.
   b) Вероятность того, что все бильярдные столы заняты.
   c) Среднее количество используемых бильярдных столов.
   d) Среднее число пар, ожидающих освобождения бильярдного стола.

6. Парикмахерская в любой момент времени может обслужить только одного клиента. Имеется также три места для ожидающих клиентов. Это значит, что в парикмахерской одновременно не могут находиться более четырех человек. Клиенты приходят в соответствии с распределением Пуассона со средним значением 4 человека в час. Время обслуживания является случайной величиной, распределенной по экспоненциальному закону с математическим ожиданием 15 минут. Определите следующие величины.

   a) Вероятности установившегося режима.
   b) Ожидаемое число клиентов в парикмахерской.
   c) Вероятность того, что клиент уйдет в поисках другой парикмахерской, поскольку все места заняты.

7. Рассмотрите систему обслуживания с одним сервисом, в которой интенсивности входного и выходного потоков имеют следующий вид.

$$\lambda_n = 10 - n,\ \ n = 0, 1, 2, 3,$$

$$\mu_n = \frac{n}{2} + 5,\ \ n = 1, 2, 3, 4.$$

   Эта ситуация эквивалентна снижению интенсивности входного потока и увеличению интенсивности выходного потока при увеличении числа n клиентов в системе.

   a) Постройте диаграмму переходов и уравнение баланса для описанной ситуации.
   b) Определите вероятности установившегося режима.

8. Рассмотрите простую систему обслуживания, когда в ней может находиться лишь один клиент. Прибывающие клиенты, которые застают систему занятой, покидают ее и больше к ее услугам не обращаются. Пусть поступления клиентов происходят в соответствии с распределением Пуассона со средним λ человек в единицу времени, а время обслуживания является экспоненциально распределенной случайной величиной со средним $1/\mu$ временных единиц.

   a) Постройте диаграмму переходов и уравнения баланса.

   b) Определите вероятности установившегося режима.

   c) Определите среднее число клиентов в системе.

9. Процедура получения общего решения для системы обслуживания общего вида с использованием метода индукции выполняется следующим образом. Рассмотрим соотношения

$$p_k = \prod_{i=0}^{k-1}\left(\frac{\lambda_i}{\mu_{i+1}}\right)p_0, \quad k = n, n-1, n-2.$$

   Для получения искомого выражения для p_n следует подставить выражения для p_{n-1} и p_{n-2} в общее разностное уравнение, включающее p_n, p_{n-1} и p_{n-2}. Проверьте корректность этой процедуры.

17.6. Специализированные системы обслуживания с пуассоновским распределением

На рис. 17.4 схематически представлена специализированная система обслуживания пуассоновского типа, в которой параллельно функционируют c идентичных сервисов (средств обслуживания). Ожидающий клиент выбирается из очереди для обслуживания на первом свободном сервисе. Интенсивность поступления клиентов в систему равна λ клиентов в единицу времени. Все параллельные сервисы являются идентичными; это означает, что интенсивность обслуживания каждого сервиса равна μ клиентов в единицу времени. Число клиентов, находящихся в *системе обслуживания*, включает тех, кто уже *обслуживается*, и тех, кто находится в *очереди*.

Обозначения, наиболее подходящие для характеристик системы обслуживания (рис. 17.4), имеют следующую структуру:

$$(a / b / c) : (d / e / f),$$

где

a — тип распределения моментов времени поступления клиентов в систему,

b — тип распределения времени между появлением элементов выходного потока (времени обслуживания),

c — количество параллельно работающих сервисов (= 1, 2, ..., ∞),

d — дисциплина очереди,

e — максимальная емкость (конечная или бесконечная) системы (количество клиентов в очереди плюс число клиентов, принятых на обслуживание сервисами),

f — емкость (конечная или бесконечная) источника, генерирующего клиентов.

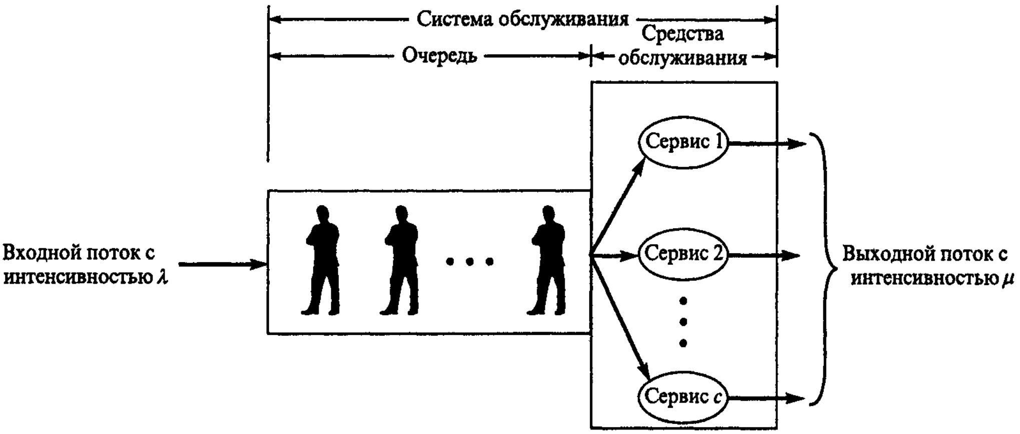


Рис. 17.4

Стандартными обозначениями для типов распределений входного и выходного потоков (символы *a* и *b*) являются следующие.

M — марковское (или пуассоновское) распределение моментов поступления клиентов в систему или их выхода из нее (или эквивалентное экспоненциальное распределение интервалов времени между моментами последовательных поступлений или продолжительностей обслуживания клиентов),

D — детерминированный (фиксированный) интервал времени между моментами последовательных поступлений в систему клиентов или детерминированная (фиксированная) продолжительность обслуживания клиентов,

E_k — распределение Эрланга, или гамма-распределение интервалов времени (или, что то же самое, распределение суммы независимых случайных величин, имеющих экспоненциальное распределение),

GI — произвольный (общий) тип распределения моментов поступления клиентов на обслуживание,

G — произвольный (общий) тип распределения продолжительности обслуживания клиентов.

Для дисциплины очереди (символ *d*) используются следующие обозначения.

FCFS — первым пришел — первым обслуживаешься,

LCFS — последним пришел — первым обслуживаешься,

SIRO — случайный отбор клиентов,

GD — произвольный (общий) тип дисциплины.

Для иллюстрации рассмотрим структуру системы обслуживания, которая соответствует модели (M / D / 10) : (GD / N / ∞). В соответствии с принятыми обозначениями здесь речь идет о системе (и, соответственно, модели) массового обслуживания с пуассоновским входным потоком (или экспоненциальным распределением интервалов времени между моментами последовательных поступлений клиентов), фиксированным временем обслуживания и десятью параллельно функционирующими сервисами. При этом дисциплина очереди не регламентирована, и максимальное количество допускаемых в систему клиентов равно N. Наконец, источник, порождающий клиентов, имеет неограниченную емкость.

В качестве исторической справки заметим, что первые три элемента $(a / b / c)$ рассмотренного обозначения были введены Кендаллом (D. G. Kendall) в 1953 году, и в литературе по теории массового обслуживания они фигурируют как **обозначения Кендалла**. Позднее в 1966 году Ли (A. M. Lee) добавил к ним символы d и e. Автором этой книги в 1968 году был введен последний символ принятых обозначений — f.

Перед детальным рассмотрением системы обслуживания пуассоновского типа покажем, как с помощью полученных в разделе 17.5 вероятностей p_n, соответствующих стационарному режиму, можно получить функциональные характеристики системы.

17.6.1. Функциональные характеристики стационарных систем обслуживания

Наиболее употребляемыми функциональными характеристиками систем массового обслуживания являются следующие.

L_s — среднее число находящихся в *системе* клиентов,
L_q — среднее число клиентов в *очереди*,
W_s — средняя продолжительность пребывания клиента в *системе*,
W_q — средняя продолжительность пребывания клиента в *очереди*,
$\bar{c}$ — среднее количество занятых средств обслуживания (сервисов).

Напомним, что *система* включает как *очередь*, так и *средства обслуживания*.

Покажем, как перечисленные функциональные характеристики получаются (прямо или косвенно) из вероятностей p_n — вероятностей того, что в системе находится n клиентов. В частности, имеем следующее.

$$L_s = \sum_{n=1}^{\infty} np_n,$$

$$L_q = \sum_{n=c+1}^{\infty} (n-c)p_n.$$

Зависимость между L_s и W_s (а также между L_q и W_q) известна в литературе по теории массового обслуживания как **формула Литтла** и имеет вид

$$L_s = \lambda_{эфф} W_s,$$
$$L_q = \lambda_{эфф} W_q.$$

Эти соотношения справедливы при достаточно общих условиях. Параметр $\lambda_{эфф}$ представляет собой *эффективную* интенсивность поступления клиентов в систему обслуживания. Он равен (исходной) интенсивности поступления клиентов λ, когда все прибывающие клиенты имеют возможность попасть в обслуживающую систему. Если же некоторые клиенты не имеют такой возможности по той причине, что она заполнена (например, заполненная автостоянка), то $\lambda_{эфф} < \lambda$. Позже мы покажем, как вычисляется $\lambda_{эфф}$.

Существует также прямая зависимость между величинами W_s и W_q. По определению

$$\begin{pmatrix}\text{Средняя продолжительность}\\ \text{пребывания в системе}\end{pmatrix}=\begin{pmatrix}\text{Среднее время}\\ \text{пребывания в очереди}\end{pmatrix}+\begin{pmatrix}\text{Среднее время}\\ \text{обслуживания}\end{pmatrix}.$$

Математически это записывается в следующем виде

$$W_s = W_q + \frac{1}{\mu}.$$

Теперь можно получить формулу, связывающую L_s и L_q, умножая обе части последнего соотношения на $\lambda_{эфф}$ и используя формулу Литтла. В результате получаем

$$L_s = L_q + \frac{\lambda_{эфф}}{\mu}.$$

По определению разность между средним числом находящихся в системе клиентов L_s и средним числом клиентов в очереди L_q равна среднему количеству занятых узлов обслуживания $\bar{c}$. Следовательно, имеем

$$\bar{c} = L_s - L_q = \frac{\lambda_{эфф}}{\mu}.$$

Поэтому процент использования узлов обслуживания вычисляется как $(c/\bar{c}) \times 100$.

Пример 17.6–1

Автостоянка для посетителей колледжа имеет всего пять мест. Автомобили прибывают на стоянку в соответствии с распределением Пуассона с интенсивностью шесть автомобилей в час. Время пребывания автомобилей на стоянке является экспоненциально распределенной случайной величиной со средним 30 минут. Посетители, которые не могут найти свободного места на стоянке непосредственно по прибытии, могут временно ожидать освобождения места на территории стоянки. Таких мест для ожидания на стоянке имеется три. Если и стоянка, и все места для ожидания заполнены, то прибывшие автомобили вынуждены искать другую автостоянку. Требуется определить следующее:

a) вероятность p_n того, что в системе находится n автомобилей,
b) эффективную интенсивность поступления автомобилей на стоянку,
c) среднее количество автомобилей на стоянке,
d) среднее время нахождения автомобиля в очереди на территории стоянки,
e) среднее количество занятых мест на автостоянке.

Прежде всего, заметим, что место для стоянки в рассматриваемой ситуации выступает в роли сервиса, так что система имеет всего $c = 5$ средств обслуживания. Максимальная вместимость системы равна $5 + 3 = 8$ автомобилей.

Вероятность p_n может быть определена как частный случай из общей модели, рассмотренной в разделе 17.5. В частности, имеем

$$\lambda_n = 6 \text{ автомобилей в час}, \ n = 0, 1, 2, ..., 8,$$

$$\mu_n = \begin{cases} n\left(\dfrac{60}{30}\right) = 2n \text{ автомобилей в час}, \ n = 1,2,...,5, \\ 5\left(\dfrac{60}{30}\right) = 10 \text{ автомобилей в час}, \ n = 6,7,8. \end{cases}$$

Следовательно, из соотношений, полученных в разделе 17.5, получаем

$$p_n = \begin{cases} \dfrac{\left(\dfrac{6}{2}\right)^n}{n!} p_0, & n = 1,2,....,5, \\ \dfrac{\left(\dfrac{6}{2}\right)^n}{5!5^{n-5}} p_0, & n = 6,7,8. \end{cases}$$

Значение p_0 вычисляется путем подстановки значений для p_n, $n = 1, 2, ..., 8$, в уравнение $p_0 + p_1 + ... + p_8 = 1$. В результате получаем

$$p_0 + p_0\left(\frac{3}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} + \frac{3^4}{4!} + \frac{3^5}{5!} + \frac{3^6}{5!5^1} + \frac{3^7}{5!5^2} + \frac{3^8}{5!5^3}\right) = 1.$$

Решением этого уравнения является $p_0 = 0.04812$ (проверьте!). Найденное значение p_0 позволяет вычислить все вероятности от p_1 до p_8.

n	1	2	3	4	5	6	7	8
p_n	0.14436	0.21654	0.21654	0.16240	0.09744	0.05847	0.03508	0.02105

Эффективную интенсивность поступления автомобилей на стоянку $\lambda_{эфф}$ можно вычислить с использованием принципиальной схемы (рис. 17.5), в соответствии с которой клиенты из источника поступают с интенсивностью λ. Прибывающий автомобиль может поступить на стоянку с интенсивностью $\lambda_{эфф}$ или уехать в поисках другой стоянки с интенсивностью $\lambda_{потери}$, т.е. $\lambda = \lambda_{эфф} + \lambda_{потери}$. Автомобиль не может въехать на стоянку, если там уже имеется 8 автомобилей. Это значит, что часть автомобилей, которые *не смогут* попасть на стоянку, пропорциональна p_8. Следовательно,

$$\lambda_{потери} = \lambda p_8 = 6 \times 0.02105 = 0.1263 \text{ автомобилей в час},$$
$$\lambda_{эфф} = \lambda - \lambda_{потери} = 6 - 0.1263 = 5.8737 \text{ автомобилей в час}.$$

Среднее количество автомобилей на стоянке (тех, которые занимают места стоянки, и тех, которые ожидают места) определяется значением L_s — средним числом клиентов в системе. Значение L_s определяется через p_n следующим образом.

$$L_s = 0p_0 + 1p_1 + ... + 8p_8 = 3.1286 \text{ автомобилей}.$$

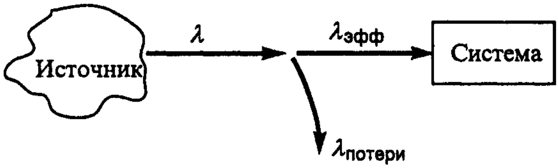


Рис. 17.5

Автомобиль, ожидающий место для стоянки, фактически находится в очереди. Следовательно, время его ожидания освобождения места на стоянке равно величине W_q. Для вычисления W_q используем определение

$$W_q = W_s - \frac{1}{\mu}.$$

Так как

$$W_s = \frac{L_s}{\lambda_{эфф}} = \frac{3.1286}{5.8737} = 0.53265 \text{ часа},$$

то

$$W_q = 0.53265 - \frac{1}{2} = 0.03265 \text{ часа}.$$

Среднее количество занятых мест на автостоянке равно среднему значению "занятых сервисов" и поэтому вычисляется следующим образом.

$$\overline{c} = L_s - L_q = \frac{\lambda_{эфф}}{\mu} = \frac{5.8737}{2} = 2.9368 \text{ мест}.$$

Упражнения 17.6,a

1. В задаче из примера 17.6–1 выполните следующее.
   a) Вычислите L_q, используя для этого формулу $\sum_{n=c+1}^{\infty}(n-c)p_n$.
   b) На основании найденного значения L_q вычислите W_s.
   c) Вычислите среднее количество автомобилей, которые не смогут въехать на территорию стоянки на протяжении восьмичасового периода.
   d) Покажите, что среднее количество свободных мест на стоянке равно $\sum_{n=0}^{c-1}(c-n)p_n$.
2. Решите задачу из примера 17.6–1, если количество мест для стоянки автомобилей равно 6, количество временных мест — 4, $\lambda = 10$ автомобилей в час и среднее время нахождения автомобиля на месте стоянки равно 45 минут.

17.6.2. Модели с одним сервисом

В этом разделе представлены две модели обслуживающей системы с одним средством обслуживания (т.е. $c = 1$). Предполагается, что клиенты поступают с постоянной интенсивностью λ. Интенсивность обслуживания также постоянна и равна μ клиентов в единицу времени. Первая модель не устанавливает ограничений на вместимость системы, во второй модели предполагается, что вместимость системы является ограниченной. В этих двух моделях источник, "порождающий" клиентов, имеет неограниченную емкость.

Рассматриваемые здесь модели обслуживающих систем (как, по сути, и все остальные модели раздела 17.6) являются частными случаями систем обслуживания общего вида, рассмотренных в разделе 17.5.

Используем обозначения Кендалла для описания характеристик системы обслуживания в каждом случае. Так как вывод выражения для p_n в разделе 17.5 и всех функциональных характеристик обслуживающей системы в разделе 17.6.1 выполнен независимо от конкретной дисциплины очереди, в обозначениях будем использовать символ GD (дисциплина очереди не регламентирована).

Модель (M/M/1) : (GD/∞/∞). Используя обозначения общей модели, имеем

$$\lambda_n = \lambda \text{ и } \mu_n = \mu \text{ для всех } n = 0, 1, 2, \dots .$$

Поскольку отсутствуют ограничения на емкость очереди и, следовательно, все прибывающие клиенты могут попасть в систему обслуживания, то $\lambda_{\text{эфф}} = \lambda$ и $\lambda_{\text{потери}} = 0$.

Обозначим $\rho = \lambda/\mu$. Тогда выражение для вероятности p_n в общей модели принимает следующий вид:

$$p_n = \rho^n p_0,\ n = 0, 1, 2, \dots .$$

Для определения величины p_0 используется тождество

$$p_0(1 + \rho + \rho^2 + \dots) = 1.$$

Предполагаем, что $\rho < 1$, тогда геометрический ряд имеет конечную сумму $1/(1-\rho)$, поэтому $p_0 = 1 - \rho$.

Следовательно, общая формула для p_n имеет вид

$$p_n = (1-\rho)\rho^n,\ \ n = 1, 2, \dots, (\rho < 1).$$

Эти значения вероятностей p_n (включая вероятность p_0) соответствуют геометрическому распределению.

При выводе формулы для p_n предполагалось, что $\rho < 1$. Это означает, что для достижения системой стационарного режима функционирования необходимо, чтобы интенсивность поступления клиентов λ была строго меньше интенсивности обслуживания μ. Если $\lambda \geq \mu$, геометрический ряд является расходящимся, и, следовательно, вероятности p_n стационарного состояния не существуют. В этом случае система обслуживания будет функционировать в нестационарном режиме, когда длина очереди со временем неограниченно возрастает.

Среднее число находящихся в системе клиентов L_s, как функциональная характеристика обслуживающей системы, вычисляется по следующей формуле:

$$L_s = \sum_{n=0}^{\infty} np_n = \sum_{n=0}^{\infty} n(1-\rho)\rho^n = (1-\rho)\rho\frac{d}{d\rho}\sum_{n=0}^{\infty}\rho^n = (1-\rho)\rho\frac{d}{d\rho}\left(\frac{1}{1-\rho}\right) = \frac{\rho}{1-\rho}.$$

Так как в рассматриваемой модели $\lambda_{\text{эфф}} = \lambda$, то остальные функциональные характеристики обслуживающей системы вычисляются с использованием соотношений из раздела 17.6.1, что приводит к следующим результатам.

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\rho)},$$

$$W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)},$$

$$L_q = \lambda W_q = \frac{\rho^2}{1-\rho},$$

$$\bar{c} = L_s - L_q = \rho.$$

Пример 17.6–2

Автоматическая мойка для автомобилей имеет только один моечный бокс. Автомобили прибывают в соответствии с распределением Пуассона со средним 4 машины в час и могут ожидать обслуживания на стоянке рядом с автомойкой. Время мойки автомобиля является экспоненциально распределенной случайной величиной с математическим ожиданием 10 минут. Автомобили, которые не помещаются на стоянке, могут ожидать на прилегающей к автомойке улице. Это значит, что практически нет ограничений на емкость системы обслуживания. Хозяин автомойки хочет определить количество мест на стоянке для автомобилей.

Для рассматриваемой задачи имеем $\lambda = 4$ автомобиля в час и $\mu = 60/10 = 6$ автомобилей в час. Так как $\rho = \lambda/\mu < 1$, то система может функционировать в стационарном режиме.

Результаты использования программы TORA для решения рассматриваемой задачи, представленные на рис. 17.6, получены путем введения данных в следующем порядке: $\lambda = 4$, $\mu = 6$, $c = 1$, емкость системы равна ∞ и емкость источника также равна ∞.

Результаты решения показывают, что среднее количество автомашин, ожидающих в очереди, равно $L_q = 1.33$ автомашины. Мы не можем рассматривать L_q в качестве единственного аргумента при определении искомого количества мест на стоянке, ибо при расчете должна учитываться максимально возможная длина очереди. Например, можно рассчитать количество мест на стоянке, при котором по меньшей мере 90% прибывших автомобилей найдет место на стоянке.

Пусть неизвестная переменная s представляет искомое количество мест на стоянке. Тогда система имеет емкость $s + 1$ (очередь плюс место на мойке). Прибывающий автомобиль в 90% случаев найдет место на стоянке, если в системе находится самое большее s автомобилей. Это условие эквивалентно следующему вероятностному утверждению:

$$p_0 + p_1 + \ldots + p_s \geq 0.9.$$

Из листинга, показанного на рис. 17.6, следует, что *суммы* вероятностей p_n равны 0.86831 и 0.91221 при $n = 4$ и $n = 5$ соответственно. Это значит, что условие выполняется при $s \geq 5$ мест на стоянке.

Количество мест на стоянке s может быть также определено с использованием формулы, определяющей p_n. Получаем

$$(1-\rho)(1 + \rho + \rho^2 + \ldots + \rho^s) \geq 0.9.$$

Сумма усеченного геометрического ряда равна $(1 - \rho^{s+1})/(1 - \rho)$. Следовательно, последнее выражение приводится к виду

$$(1 - \rho^{s+1}) \geq 0.9.$$

```
Problem title: Example 17.6-2
Scenario 1 -- (M/M/1):(GD/*/*)
--------------------------------------------------------------------
Lambda =      4.00000          Lambda eff =      4.00000
Mu =      6.00000              Rho =      0.66667

Ls =      2.00000              Lq =      1.33333
Ws =      0.50000              Wq =      0.33333
--------------------------------------------------------------------
Values of p(n) for n=0 to 25, else p(n) < .00001

 0 0.33333  1 0.22222  2 0.14815  3 0.09877  4 0.06584
 5 0.04390  6 0.02926  7 0.01951  8 0.01301  9 0.00867
10 0.00578 11 0.00385 12 0.00257 13 0.00171 14 0.00114
15 0.00076 16 0.00051 17 0.00034 18 0.00023 19 0.00015
20 0.00010 21 0.00007 22 0.00004 23 0.00003 24 0.00002
25 0.00001

Cumulative values of p(n) for n=0 to 25

 0 0.33333  1 0.55556  2 0.70370  3 0.80247  4 0.86831
 5 0.91221  6 0.94147  7 0.96098  8 0.97399  9 0.98266
10 0.98844 11 0.99229 12 0.99486 13 0.99657 14 0.99772
15 0.99848 16 0.99898 17 0.99932 18 0.99955 19 0.99970
20 0.99980 21 0.99987 22 0.99991 23 0.99994 24 0.99996
25 0.99997
```

Рис. 17.6

Упрощая это неравенство, получаем

$$\rho^{s+1} \leq 0.1.$$

Логарифмируя обе части последнего неравенства, получаем следующее.

$$s \geq \frac{\ln(0.1)}{\ln\left(\frac{4}{6}\right)} - 1 = 4.679.$$

Таким образом, необходимо $s \geq 5$ мест на стоянке.

Упражнения 17.6,b

1. В задаче из примера 17.6–2 выполните следующее.

   a) Определите процент использования автомойки.

   b) Определите вероятность того, что прибывающий автомобиль должен ожидать на стоянке, прежде чем попасть в моечный бокс.

   c) Определите вероятность того, что прибывающий автомобиль найдет свободное место на стоянке при условии, что там имеется семь мест.

   d) Сколько должно быть мест на стоянке, чтобы прибывающий автомобиль в 99% случаев нашел место на стоянке?

2. Джон является студентом университета. Он выполняет случайные работы для улучшения своего материального положения. Интервал времени между последовательными поступлениями заявок на работу является экспоненциально распределенной случайной величиной со средним значением пять дней. Время, необходимое для выполнения работы, также является экспоненциально распределенной случайной величиной со средним значением четыре дня.

   a) Какова вероятность того, что Джон будет без работы?

   b) Если за каждую работу Джон получает примерно 50 долларов, то каков его среднемесячный заработок?

   c) Если в конце семестра Джон решает передоверить невыполненные работы другому лицу по 40 долларов за каждую работу, то каково среднее значение суммы, которую должен уплатить Джон?

3. На протяжении многих лет детектив Коломбо из отделения полиции города Фейетвилл демонстрирует феноменальный успех в расследовании каждого криминального дела, за которое он берется. Для него раскрытие любого криминального дела — это всего лишь вопрос времени. Коломбо соглашается, что время раскрытия каждого отдельного случая является "совершенно случайным", но в среднем каждое расследование занимает около полторы недели. Криминальные дела в мирном городке, где работает Коломбо, явление не очень частое. Они происходят случайным образом с интенсивностью одно преступление в месяц. Проанализируйте "производительность" работы детектива Коломбо; в частности, найдите следующие показатели.

   a) Среднее число случаев, которые ожидают расследования.

   b) Процент времени, когда детектив занят расследованиями.

   c) Среднее время, необходимое для раскрытия преступления.

4. Автомобили прибывают к пропускному пункту туннеля Линкольна, где взимается плата за проезд, в соответствии с распределением Пуассона со средним 90 единиц в час. Время прохождения пропускного пункта автомобилями является случайной величиной, распределенной по экспоненциальному закону со средним 38 секунд. Водители жалуются на долгое время ожидания, и власти планируют сократить среднее время прохождения пропускного пункта до 30 секунд путем установки автоматического устройства для взимания транспортной пошлины, если только выполняются два условия: 1) среднее количество ожидающих автомобилей превы-

шает 5 единиц при существующей системе взимания пошлины и 2) процент времени простоя нового устройства, установленного на пропускном пункте, не будет превышать 10%. Может ли быть оправдана установка нового устройства?

5. Ресторан быстрого питания имеет один пункт обслуживания, где клиенты обслуживаются, не выходя из автомашины. Машины прибывают в соответствии с распределением Пуассона с интенсивностью 2 клиента за каждые 5 минут. Возле пункта обслуживания может расположиться не больше 10 автомашин, включая ту, которую обслуживают. Другие автомашины при необходимости могут ожидать обслуживания за пределами этого пространства. Время обслуживания одного клиента распределено по экспоненциальному закону со средним значением 1.5 минуты. Определите следующие показатели.

   a) Вероятность того, что пункт обслуживания свободен.
   b) Среднее число клиентов, ожидающих обслуживания.
   c) Среднее время ожидания клиента до того момента, когда он делает заказ.
   d) Вероятность того, что очередь превысит десятиместное пространство перед пунктом обслуживания.

6. Банк располагает одним пунктом обслуживания, где клиенты обслуживаются, не выходя из автомашины. Клиенты прибывают в соответствии с распределением Пуассона со средним значением 10 клиентов в час. Время обслуживания одного клиента распределено по экспоненциальному закону со средним значением 5 минут. Напротив пункта обслуживания имеется место для трех автомобилей, включая и тот, что обслуживается. Другие прибывающие автомашины выстраиваются в очередь вне этого пространства.

   a) Какова вероятность того, что прибывающий автомобиль может занять одно из трех мест возле пункта обслуживания?
   b) Какова вероятность того, что прибывающий автомобиль будет ожидать обслуживания вне зоны для трех автомобилей?
   c) Каково среднее время ожидания прибывающего клиента до того момента, когда его начнут обслуживать?
   d) Сколько мест для автомобилей должно быть возле обслуживающего пункта обслуживания, чтобы прибывающий клиент мог найти там место по крайней мере в 20% случаев?

7. В модели $(M/M/1):(GD/\infty/\infty)$ системы обслуживания покажите, что в общем случае L_s не равно $L_q + 1$. При каком условии имеет место это равенство?

8. Для модели $(M/M/1):(GD/\infty/\infty)$ системы обслуживания получите выражение для L_q, используя формулу для $\sum_{n=2}^{\infty}(n-1)p_n$.

9. Для модели $(M/M/1):(GD/\infty/\infty)$ системы обслуживания покажите, что

   a) среднее число клиентов в очереди, если она не пуста, равно $1/(1-\rho)$,
   b) среднее время ожидания в очереди равно $1/(\mu-\lambda)$.

Распределение времени ожидания в модели (M/M/1):(FCFS/∞/∞)[1]. Вывод формулы для вероятностей p_n в общей модели системы обслуживания, рассмотренной в разделе 17.5, *полностью* не зависит от дисциплины очереди. Это значит, что математические ожидания всех функциональных параметров W_s, W_q, L_s и L_q справедливы для системы с любой дисциплиной очереди.

Хотя *среднее* время ожидания в системе обслуживания не зависит от дисциплины очереди, плотность вероятности его распределения зависит от дисциплины очереди. Проиллюстрируем это утверждение путем построения плотности вероятности времени ожидания для модели ($M/M/1$) с дисциплиной очереди FCFS ("первым пришел — первым обслуживаешься").

Обозначим через τ количество времени, которое *только что прибывший* клиент проведет в *системе* от момента прибытия до завершения обслуживания. Исходя из дисциплины очереди FCFS, если в системе уже находится n клиентов, которые поступили в систему ранее только что прибывшего, то

$$\tau = t_1' + t_2 + \ldots + t_{n+1},$$

где t_1' — время, необходимое для завершения обслуживания клиента, который уже находится в средстве обслуживания системы, а t_2, t_3, ..., t_n — интервалы времени, которые потребуются для обслуживания $n - 1$ клиентов, которые находятся в очереди. Величина t_{n+1} представляет собой время обслуживания только что прибывшего клиента.

Обозначим через $w(\tau \mid n + 1)$ условную плотность вероятности τ, где условием служит наличие в обслуживающей системе n клиентов на момент прибытия нового. Поскольку время обслуживания в системе распределено по экспоненциальному закону, в силу свойства отсутствия последействия этого распределения (раздел 17.4) величина t_1' также распределена по экспоненциальному закону. Следовательно, τ представляет собой сумму $n + 1$ независимых случайных величин, каждая из которых подчиняется одному и тому же экспоненциальному распределению. Как известно из теории вероятностей, функция $w(\tau \mid n + 1)$ будет плотностью вероятности гамма-распределения с параметрами μ и $n + 1$. Отсюда следует, что

$$w(\tau) = \sum_{n=0}^{\infty} w(\tau \mid n+1) p_n = \sum_{n=0}^{\infty} \frac{\mu (\mu t)^n e^{-\mu t}}{n!} (1-\rho)\rho^n =$$

$$= (1-\rho)\mu e^{-\mu\tau} \sum_{n=0}^{\infty} \frac{(\lambda\tau)^n}{n!} = \mu(1-\rho)e^{-\mu(1-\rho)\tau}, \quad \tau > 0.$$

Таким образом показано, что случайная величина τ имеет экспоненциальное распределение с математическим ожиданием $W_s = 1/\mu(1 - \rho)$.

Пример 17.6–3

В модели работы автомойки из примера 17.6–2 вполне обоснованным является предположение о том, что обслуживание выполняется в соответствии с дисциплиной очереди FCFS. Определим надежность использования величины W_s в качестве оценки времени ожидания в системе. Для этого оценим часть клиентов, время ожидания которых превышает W_s. Используя формулу $W_s = 1/\mu(1 - \rho)$, получаем

$$P\{\tau > W_s\} = 1 - \int_0^{W_s} w(\tau) d\tau = e^{-\mu(1-\rho)W_s} = e^{-1} \approx 0.368.$$

[1] Этот пункт можно пропустить без ущерба для понимания дальнейшего материала.

Следовательно, при дисциплине очереди FCFS для 37% клиентов время ожидания превысит значение W_s. Этот процент кажется чрезмерно большим, особенно если учесть, что текущее значение W_s для рассматриваемой системы обслуживания также является большим (= 0.5 часа). Заметим, что найденная вероятность ($= e^{-1} \approx 0.368$) не зависит от интенсивностей λ и μ модели $(M/M/1)$: (FCFS/∞/∞); это в свою очередь означает, что ее величина не может быть уменьшена. Следовательно, если мы проектируем систему на основании средней величины W_s, то следует ожидать, что для 36,8% клиентов время ожидания превысит среднее время ожидания в системе.

Существует две возможности улучшения ситуации: (1) можно увеличить интенсивность обслуживания клиентов μ для *уменьшения* значения W_s до приемлемого уровня или (2) можно выбрать интенсивность обслуживания клиентов таким образом, чтобы вероятность того, что время ожидания превысит заранее определенную величину (скажем, 10 минут), была меньше приемлемо малой величины (например, 10%). Первая возможность эквивалентна нахождению такого значения μ, что $W_s < \overline{T}$, а вторая — нахождению такого значения μ, что $P\left\{\tau > \overline{T}\right\} < \alpha$, где $\overline{T}$ и α должны определяться пользователем.

Упражнения 17.6,c

1. В условиях упр. 17.6,b(3) определите вероятность того, что для раскрытия преступления детективу Коломбо потребуется больше одной недели.
2. В примере 17.6–3 вычислите следующее.
   a) Среднеквадратическое отклонение для времени τ ожидания в системе.
   b) Вероятность того, что время ожидания в системе изменится на половину среднеквадратического отклонения относительно своего среднего значения.
3. В примере 17.6–3 определите интенсивность обслуживания клиентов μ таким образом, чтобы выполнялось условие $W_s < 10$ минут.
4. В примере 17.6–3 определите такое значение интенсивности обслуживания клиентов μ, при котором будет выполняться условие $P\{\tau > 10$ минут$\} < 0.1$.
5. Вернитесь к упр. 17.6,b(5). Для привлечения большего числа посетителей администрация ресторана решила предлагать бесплатно порцию прохладительного напитка каждому клиенту, который вынужден ожидать более 5 минут. Если стоимость порции прохладительного напитка равняется 50 центам, то в какую сумму в среднем обойдется ресторану ежедневное угощение прохладительными напитками? Предполагается, что ресторан открыт для клиентов 12 часов в сутки.
6. Покажите, что для модели $(M/M/1)$: (FCFS/∞/∞) системы обслуживания плотность вероятности времени пребывания клиентов в очереди имеет следующий вид:

$$w_q(t) = \begin{cases} 1-\rho, & t = 0 \\ \mu\rho(1-\rho)e^{-(\mu-\lambda)t}, & t > 0. \end{cases}$$

   С помощью этой формулы найдите W_q.

Модель (M/M/1) : (GD/N/∞). Эта модель отличается от рассмотренной выше только тем, что система вмещает не более N клиентов (максимальная длина очереди равняется $N-1$). Примерами обслуживающей системы такого типа служат производственные ситуации, когда станок может иметь ограниченную зону складирования заготовок, а также рестораны быстрого питания с одним пунктом обслуживания клиентов на автомобилях и т.д.

Ситуация в рассматриваемой модели такова, что как только число клиентов в системе достигает N, ни один из дополнительных клиентов на обслуживание не принимается. Из этого условия следует, что

$$\lambda_n = \begin{cases} \lambda, & n = 0,1,\ldots,N-1, \\ 0, & n = N, N+1,\ldots, \end{cases}$$

$$\mu_n = \mu,\ n = 0,\ 1,\ \ldots .$$

Используя обозначение $\rho = \lambda/\mu$, имеем

$$p_n = \begin{cases} \rho^n p_0, & n \le N, \\ 0, & n > N. \end{cases}$$

Значение вероятности p_0 определяется из уравнения $\sum_{n=0}^{\infty} p_n = 1$, которое принимает следующий вид:

$$p_0(1 + \rho + \rho^2 + \ldots + \rho^N) = 1.$$

Отсюда получаем

$$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{N+1}}, & \rho \ne 1, \\ \dfrac{1}{N+1}, & \rho = 1. \end{cases}$$

Следовательно,

$$p_n = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{N+1}}, & \rho \ne 1, \\ \dfrac{1}{N+1}, & \rho = 1, \end{cases} \qquad n = 0,\ 1,\ \ldots,\ N.$$

Заметим, что в этой модели значение параметра $\rho = \lambda/\mu$ не обязательно должно быть меньше единицы, так как поступления клиентов в систему контролируются максимальной емкостью системы N. Это значит, что в данном случае в качестве интенсивности поступления клиентов скорее выступает $\lambda_{\text{эфф}}$, нежели λ. Так как клиенты будут потеряны в том случае, если в системе находится N клиентов, тогда, как показано на рис. 17.5,

$$\lambda_{\text{потери}} = \lambda p_n,$$
$$\lambda_{\text{эфф}} = \lambda - \lambda_{\text{потери}} = \lambda(1 - p_N).$$

Следует ожидать, что $\lambda_{\text{эфф}}$ будет меньше μ.

Среднее число клиентов в системе вычисляется по формуле

$$L_s = \sum_{n=0}^{N} np_n = \frac{1-\rho}{1-\rho^{N+1}} \sum_{n=0}^{N} n\rho_n = \frac{(1-\rho)\rho}{1-\rho^{N+1}} \frac{d}{d\rho}\left(\frac{1-\rho^{N+1}}{1-\rho}\right) =$$

$$= \frac{\rho\left\{1-(N+1)\rho^N + N\rho^{N+1}\right\}}{(1-\rho)\left(1-\rho^{N+1}\right)}, \quad \rho \neq 1.$$

При $\rho = 1$ $L_s = N/2$ (проверьте!). Используя значения L_s и $\lambda_{эфф}$, можно также получить выражения для W_s, W_q и L_q, как это сделано в разделе 17.6.1.

Использование калькулятора для вычислений по формулам теории массового обслуживания является в лучшем случае громоздким (в последующих моделях формулы еще более сложные!). Поэтому автор рекомендует использовать программное обеспечение TORA при проведении этих вычислений.

Пример 17.6–4

Рассмотрим ситуацию с автомойкой автомобилей из примера 17.6–2. Пусть станция имеет четыре места для стоянки автомобилей. Если все места на стоянке заняты, вновь прибывающие автомобили вынуждены искать другую автомойку. Хозяин хочет определить влияние ограниченного количества мест для стоянки автомобилей на потери клиентов.

В обозначениях, принятых в модели, максимальная вместимость системы равна $N = 4 + 1 = 5$. Исходными данными для программы TORA являются числа 4, 6, 1, 5 и ∞ соответственно, а выходные данные представлены на рис. 17.7.

```
Problem title: Example 17.6-4
Scenario 1 -- (M/M/1):(GD/5/*)
--------------------------------------------------------------------
Lambda =     4.00000        Lambda eff =     3.80752
Mu =     6.00000            Rho =     0.66667

Ls =     1.42256            Lq =     0.78797
Ws =     0.37362            Wq =     0.20695
--------------------------------------------------------------------
Values of p(n) for n=0 to 5, else p(n) < .00001

0 0.36541    1 0.24361    2 0.16241    3 0.10827    4 0.07218
5 0.04812

Cumulative values of p(n) for n=0 to 5

 0 0.36541    1 0.60902    2 0.77143    3 0.87970    4 0.95188
 5 1.00000
```

Рис. 17.7

Так как емкость системы равняется $N = 5$, доля потерянных клиентов составляет $p_5 = 0.04812$, что при круглосуточной работе моечной станции эквивалентно потере $(\lambda p_5) \times 24 = 4 \times 0.04812 \times 24 = 4.62$ автомобилей в день. Решение относительно увеличения количества мест для стоянки автомобилей должно основываться на сумме потерь автомойки.

Анализируя ситуацию с другой стороны, замечаем, что среднее время пребывания клиента в обслуживающей системе $W_s = 0.3736$ (примерно 22 минуты), т.е. меньше 30 минут, как это было в примере 17.6–3, когда всем прибывающим автомобилям разрешалось встать в очередь. Уменьшение этого показателя обслуживающей системы примерно на 25% обеспечено за счет потери около 4.8% потенциальных клиентов из-за ограниченного количества мест на стоянке автомобилей.

Упражнения 17.6,d

1. В примере 17.6–4 определите следующие величины.

   a) Вероятность того, что прибывший автомобиль сразу же попадет в моечный бокс.

   b) Среднее время ожидания клиентов до начала обслуживания.

   c) Среднее количество свободных мест на стоянке автомобилей.

   d) Вероятность того, что все места на стоянке автомобилей заняты.

   e) Уменьшение (в процентах) среднего времени обслуживания клиентов при уменьшении среднего времени пребывания клиента в системе примерно до 10 минут. (*Совет.* Используйте метод проб и ошибок в работе с программой TORA для решения этого упражнения).

2. Рассмотрите ситуацию с автомойкой из примера 17.6–4. Определите такое количество мест для стоянки автомобилей, при котором процент автомобилей, которые не могут найти место на стоянке, ограничен значением 1%.

3. Парикмахер Иосиф обслуживает клиентов в соответствии с экспоненциальным распределением со средним значением 12 минут. Иосиф очень популярен среди клиентов, поэтому они прибывают (в соответствии с распределением Пуассона) с интенсивностью, намного большей, чем он может обслужить (шесть клиентов в час). На самом деле парикмахеру хотелось бы, чтобы интенсивность поступления клиентов уменьшилась примерно до четырех клиентов в час. Поэтому он пришел к мысли ограничить количество стульев в зале ожидания, чтобы вновь прибывающие клиенты, обнаружив, что все стулья заняты, уходили в поисках иного обслуживания. Сколько стульев следует Иосифу разместить в зале ожидания, чтобы реализовать свой план?

4. Конечная сборка электрических генераторов на электропредприятии проходит в соответствии с распределением Пуассона со средним значением 10 генераторов в час. Затем генераторы с помощью ленточного конвейера транспортируются в отдел технического контроля для испытаний. На конвейере может находиться максимум 7 генераторов. Электронный датчик автоматически останавливает конвейер, как только он заполнен, прекращая таким образом работу сборочного цеха до появления свободного места на конвейере. Время проверки генераторов имеет экспоненциальное распределение со средним значением 15 минут.

   a) Какова вероятность того, что сборочный цех прекратит сборку генераторов?

b) Чему равняется среднее количество генераторов на конвейере?

c) Инженер утверждает, что перерывы в работе сборочного цеха можно уменьшить за счет увеличения производительности конвейера до такого уровня, который обеспечивает сборочному цеху возможность работать 95% времени без перерывов. Обосновано ли такое утверждение?

5. В кафетерии имеется не более 50 мест. Посетители прибывают в соответствии с пуассоновским распределением с интенсивностью 10 человек в час и обслуживаются (каждый в отдельности) с интенсивностью 12 человек в час.

   a) Какова вероятность того, что очередной посетитель не сможет пообедать в кафетерии по причине отсутствия свободных мест?

   b) Предположим, что три посетителя кафетерия (каждый из которых прибывает случайным образом) хотели бы сидеть за одним столиком. Какова вероятность того, что их желание может быть выполнено? (Здесь предполагается, что всегда есть возможность посадить упомянутых посетителей вместе, если в кафетерии имеется больше трех свободных мест.)

6. Пациенты прибывают в клинику в соответствии с распределением Пуассона с интенсивностью 20 пациентов в час. В комнате ожидания могут разместиться не больше 14 человек. Время осмотра клиентов является экспоненциально распределенной случайной величиной с математическим ожиданием 8 минут.

   a) Какова вероятность того, что очередной пациент не будет ожидать?

   b) Какова вероятность того, что очередной пациент найдет свободный стул в комнате ожидания?

   c) Каково среднее время пребывания пациента в клинике?

7. Ниже приведены значения вероятностей p_n нахождения n клиентов в обслуживающей системе модели $(M/M/1)$: $(GD/5/\infty)$.

n	0	1	2	3	4	5
p_n	0.399	0.249	0.156	0.097	0.061	0.038

   Интенсивность λ поступления клиентов равняется пяти клиентам в час. Интенсивность μ обслуживания клиентов равняется восьми клиентам в час.

   a) Вычислите вероятность того, что очередной клиент сможет попасть в систему обслуживания.

   b) Вычислите интенсивность, при которой прибывающие клиенты не смогут попасть в систему обслуживания.

   c) Вычислите среднее число клиентов в обслуживающей системе.

   d) Вычислите среднее время ожидания клиентов в очереди.

8. Покажите, что при $\rho = 1$ в модели $(M/M/1)$: $(GD/N/\infty)$ среднее число клиентов в системе L_s равно $N/2$. (*Подсказка.* $1 + 2 + ... + i = i(i + 1)/2$.)

9. Покажите, что $\lambda_{эфф}$ в модели $(M/M/1)$: $(GD/N/\infty)$ можно вычислить по формуле $\lambda_{эфф} = \mu(L_s - L_q)$.

17.6.3. Модели с параллельными сервисами

В этом разделе рассматриваются три модели систем массового обслуживания с несколькими параллельно работающими средствами обслуживания (сервисами). Первые две модели представляют собой обобщение моделей, рассмотренных в разделе 17.6.2, на случай нескольких параллельно работающих сервисов. В третьей модели рассматривается случай бесконечного количества параллельно работающих сервисов.

Модель (M/M/c) : (GD/∞/∞). Эта модель предусматривает работу c параллельных средств обслуживания. Интенсивность входного потока клиентов равна λ, а интенсивность обслуживания клиентов — μ для каждого сервиса. Поскольку отсутствуют ограничения на количество клиентов в системе, то $\lambda_{эфф} = \lambda$.

Результатом использования c параллельных сервисов является пропорциональное увеличение интенсивности обслуживания клиентов системой до $n\mu$, если $n \le c$, и до $c\mu$, если $n > c$. Следовательно, в терминах общей модели системы обслуживания (раздел 17.5) λ_n и μ_n определяются следующим образом.

$$\lambda_n = \lambda, \quad n \ge 0,$$

$$\mu_n = \begin{cases} n\mu, & n \le c, \\ c\mu, & n > c. \end{cases}$$

Следовательно,

$$p_n = \begin{cases} \dfrac{\lambda^n}{\mu(2\mu)(3\mu)\dots(n\mu)} p_0 = \dfrac{\lambda^n}{n!\mu^n} p_0, & n \le c, \\ \dfrac{\lambda^n}{\mu(2\mu)\dots(c-1)\mu(c\mu)^{n-c+1}} p_0 = \dfrac{\lambda^n}{c!c^{n-c}\mu^n} p_0, & n > c. \end{cases}$$

Значение вероятности p_0 определяется из уравнения $\sum_{n=0}^{\infty} p_n = 1$. Полагая $\rho = \lambda/\mu$ и предполагая, что $\rho/c < 1$, приходим к следующей формуле для p_0:

$$p_0 = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \sum_{n=c}^{\infty} \left(\frac{\rho}{c} \right)^{n-c} \right\}^{-1} = \left\{ \sum_{n=0}^{c-1} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \left(\frac{1}{1 - \frac{\rho}{c}} \right) \right\}^{-1}, \quad \frac{\rho}{c} < 1.$$

Выражение для L_q можно найти следующим образом.

$$L_q = \sum_{n=c}^{\infty} (n-c) p_n = \sum_{k=0}^{\infty} k p_{k+c} = \sum_{k=0}^{\infty} k \frac{\rho^{k+c}}{c^k c!} p_0 = \frac{\rho^c \rho}{c!c} \sum_{k=0}^{\infty} k \left(\frac{\rho}{c} \right)^{k-1} p_0 = \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} p_0.$$

Поскольку $\lambda_{эфф} = \lambda$, то $L_s = L_q + \rho$; значения для W_s и W_q можно найти путем деления на λ значений L_s и L_q.

Пример 17.6–5

В небольшом городке функционируют две службы такси. Каждая из них располагает двумя автомобилями, и по имеющейся информации заказы на обслуживание делятся службами практически поровну. Это подтверждается тем фактом, что заказы в диспетчерские отделения обеих служб поступают с одной и той же интенсивностью, равной 8 вызовам в час. Среднее время выполнения одной заявки составляет 12 минут. Заявки на обслуживание поступают в соответствии с распределением Пуассона, а время обслуживания клиентов распределено по экспоненциальному закону. Недавно обе службы были приобретены инвестором, который заинтересован в их объединении с единым диспетчерским пунктом для обеспечения более быстрого обслуживания клиентов. Необходимо проанализировать предложения нового хозяина.

С точки зрения теории массового обслуживания, такси представляют собой обслуживающие устройства, а вызов такси является сервисом. Каждая служба такси может быть представлена моделью $(M/M/2) : (GD/\infty/\infty)$ с $\lambda = 8$ вызовов в час и $\mu = 60/12 = 5$ поездок *на одно такси* в час. Объединение служб такси приведет к модели $(M/M/4) : (GD/\infty/\infty)$ с $\lambda = 2 \times 8 = 16$ вызовов в час и $\mu = 5$ поездок *на одно такси* в час.

Подходящей мерой для сравнения двух моделей обслуживания является среднее время ожидания клиентом такси от момента его вызова до момента прибытия автомобиля, т.е. W_q. На рис. 17.8 представлены выходные данные программы TORA для двух описанных моделей. Результаты показывают, что время ожидания клиентом приезда автомобиля равняется 0.356 часа (примерно 21 минута) для модели обслуживания с двумя таксомоторными службами и 0.149 часа (примерно 9 минут) для модели обслуживания в объединенном варианте. Значительное уменьшение (более чем на 50%) функционального показателя рассмотренной обслуживающей системы делает очевидной целесообразность объединения двух служб такси.

Problem title: Example 17-8
Comparative measures: Nbr of scenarios = 2

Nbr	c	Lambda	Mu	l'da_eff	Ls	Ws	Lq	Wq
1	2	8.000	5.000	8.000	4.444	0.556	2.844	0.356
2	4	16.000	5.000	16.000	5.586	0.349	2.386	

Рис. 17.8

Из приведенного анализа следует, что **объединение систем обслуживания** всегда обеспечивает более эффективный режим работы. Этот вывод остается справедливым даже в том случае, когда загруженность всех сервисов очень высока.

Упражнения 17.6,е

1. В примере 17.6–5 определите следующие параметры.

a) Вероятность того, что все автомобили в каждой из двух служб такси находятся на вызове.

b) Вероятность того, что все такси в объединенной компании будут на вызове.

c) Среднее количество свободных такси в каждой из двух моделей.

d) Количество машин, которое следует иметь объединенной компании для того, чтобы время ожидания клиентом приезда автомобиля по вызову составляло не более пяти минут.

2. В примере 17.6–5 предположите, что среднее время обслуживания клиента фактически равняется 14 минут, так что коэффициент загруженности автомобилей ($= \lambda/\mu c$) для режимов работы с двумя и четырьмя такси возрастает до 93% и более. Имеет ли смысл при этих условиях объединять две компании в одну?

3. Определите минимальное количество сервисов в каждой из следующих ситуаций (предполагается пуассоновское распределение поступления клиентов и обслуживания), которое гарантирует стационарный режим работы системы массового обслуживания (в этом случае длина очереди не будет неограниченно возрастать).

   a) Клиенты прибывают каждые 5 минут, а обслуживаются с интенсивностью 10 клиентов в час.

   b) Среднее время между последовательными прибытиями клиентов равняется 2 минутам, а среднее время обслуживания — 6 минутам.

   c) Интенсивность входного потока равняется 30 клиентов в час, а интенсивность обслуживания одним сервисом — 40 клиентов в час.

4. Посетители прибывают в банк в соответствии с распределением Пуассона с математическим ожиданием 45 клиентов в час. Длительность деловых операций с одним клиентом имеет экспоненциальное распределение с математическим ожиданием примерно пять минут. Банк планирует использовать одноканальный многокассовый режим работы, подобный тем, которые применяются в аэропортах и почтовых отделениях.[1] Управляющий отдает себе отчет в том, что клиенты могут обратиться в другие банки, если они чувствуют, что их ожидание в очереди является "чрезмерным". По этой причине управляющий хочет уменьшить среднее время ожидания в очереди до 30 секунд, не более. Сколько кассиров должен иметь банк?

5. Ресторан быстрого питания имеет три кассира. Посетители прибывают в ресторан в соответствии с распределением Пуассона каждые три минуты и образуют одну очередь, чтобы быть обслуженным первым освободившимся кассиром. Время до момента размещения заказа экспоненциально распределено со средним, равным примерно пяти минутам. Вместимость зала ожидания внутри ресторана ограничена. Однако ресторан имеет хорошую кухню и при необходимости посетители готовы выстраиваться в очередь и вне ресторана. Определите размер зала ожидания внутри ресторана, кроме мест возле касс, таким образом, чтобы с вероятностью не менее 0.999 следующий посетитель не ожидал обслуживания вне ресторана.

6. Небольшое почтовое отделение имеет два обслуживающих окна. Клиенты прибывают на почтовое отделение в соответствии с распределением Пуассона с интенсивностью 1 клиент в каждые три минуты. Однако лишь 80% из них нуждаются в обслуживании возле окон. Время обслуживания клиента подчиняется экспоненциальному закону со средним значением 5 минут. Все прибывающие клиенты образуют одну очередь и подходят к свободному окну в соответствии с дисциплиной "первым пришел — первым обслуживаешься".

---

[1] Здесь имеется в виду, что есть одна очередь, но клиенты обслуживаются несколькими сервисами (в данном примере сервисы — это кассы). — *Прим. ред.*

a) Какова вероятность того, что очередной клиент будет ожидать в очереди?
b) Какова вероятность того, что оба обслуживающих окна свободны?
c) Какова средняя длина очереди?
d) Можно ли предложить приемлемое обслуживание лишь с одним окном? Приведите аргументы.

7. Вычислительный центр университета состоит из четырех одинаковых больших ЭВМ коллективного пользования. Число работающих в центре пользователей в любой момент времени равно 25. Каждый пользователь готовит свою программу для ее машинной реализации через терминал, куда она сразу же передается. Время подготовки программ имеет экспоненциальное распределение со средним значением 15 минут. Поступающие программы автоматически размещаются для реализации на первую свободную ЭВМ. Время выполнения программы имеет экспоненциальное распределение со средним значением 2 минуты. Вычислите следующие показатели.

   a) Вероятность того, что программа не будет выполнена сразу же, как только она поступила на терминал.
   b) Среднее время до получения пользователем результатов машинной реализации программы.
   c) Среднее количество программ, ожидающих машинной реализации.
   d) Процент времени, когда все ЭВМ вычислительного центра свободны.
   e) Среднее количество свободных ЭВМ.

8. Аэропорт обслуживает пассажиров трех категорий: городских жителей, жителей пригородов и транзитных пассажиров. Прибытие в аэропорт пассажиров всех трех категорий во времени происходит в соответствии с распределением Пуассона со средней интенсивностью 15, 10 и 7 пассажиров в час соответственно. Время регистрации пассажиров подчиняется экспоненциальному распределению с математическим ожиданием 6 минут. Определите количество стоек для регистрации пассажиров, которыми должен располагать аэропорт в каждом из следующих случаев.

   a) Среднее время пребывания пассажира в режиме ожидания и регистрации не должно превышать 15 минут.
   b) Процент свободных регистрационных стоек не превышает 10%.
   c) Вероятность того, что все регистрационные стойки свободны, не превышает 0.17.

9. В Соединенных Штатах Америки одноканальный многосервисный режим работы обслуживающих систем обычно используется в почтовых отделениях и аэропортах при регистрации пассажиров. Однако продовольственные магазины и банки (особенно в небольших населенных пунктах) отдают предпочтение использованию одноканальных систем обслуживания с одним сервисом, несмотря на то что одноканальный многосерверный режим работы является более эффективным. Прокомментируйте это наблюдение.

10. В 1994 году ведущая телефонная компания США по дальней связи известила через хорошо разрекламированные по телевидению объявления, что абоненты, пользующиеся услугами компании с оплатой в кредит, могут получить значительную экономию, если будут заказывать переговоры по централизованному теле-

фонному номеру 1-800, вместо того чтобы делать это через региональных операторов связи. Использование абонентами номера 1-800 поощряется, так как эта услуга приносит финансовую выгоду компании. Почему это происходит?

11. Для модели $(M/M/c):(GD/\infty/\infty)$ обслуживающей системы Морс (Morse) [3] предложил аппроксимацию $L_q = \rho/(c-\rho)$ при условии, что $\rho/c \to 1$. С использованием этой информации покажите, что отношение среднего времени ожидания в модели $(M/M/c):(GD/\infty/\infty)$ к аналогичному параметру в модели $(M/M/1):(GD/\infty/\infty)$ стремится к $1/c$ при $\rho/c \to 1$. Следовательно, при $c=2$ среднее время ожидания может быть уменьшено на 50%. Из этого следует вывод, что объединение систем обслуживания целесообразно всегда.
12. В процедуре вывода формулы для вероятности p_n в модели $(M/M/c):(GD/\infty/\infty)$ укажите, какая ее часть требует условия $\rho/c<1$. Объясните значение этого условия. Что произойдет, если это условие не будет выполнено?
13. Используя формулу $L_q = \sum_{n=c+1}^{\infty}(n-c)p_n$, докажите, что $L_s = L_q + \overline{c}$, где $\overline{c}$ — среднее количество работающих сервисов. Далее покажите, что $\overline{c} = \lambda_{эфф}/\mu$.
14. Покажите, что формулу для вероятностей p_n в модели $(M/M/1):(GD/\infty/\infty)$ можно получить из аналогичной формулы для модели $(M/M/c):(GD/\infty/\infty)$ при $c=1$.
15. Покажите, что в модели $(M/M/c):(GD/\infty/\infty)$ имеет место следующая формула:

$$L_q = \frac{c\rho}{(c-\rho)^2}p_c.$$

16. Покажите, что для модели $(M/M/c):(GD/\infty/\infty)$ обслуживающей системы справедливы следующие утверждения.
    a) Вероятность того, что клиент ожидает, равняется $p_c\rho/(c-\rho)$.
    b) Если имеется очередь, то ее средняя длина равна $c/(c-\rho)$.
    c) Среднее время ожидания в очереди тех клиентов, которые вынуждены ждать, равно $1/\mu(c-\rho)$.
17. Покажите, что для модели $(M/M/c):(GD/\infty/\infty)$ обслуживающей системы плотность вероятности времени ожидания в очереди имеет следующий вид:

$$w_q(T) = \begin{cases} 1 - \dfrac{\rho^c}{(c-1)!(c-\rho)}p_0, & T=0, \\ \dfrac{\mu\rho^c e^{-\mu(c-\rho)T}}{(c-1)!}p_0, & T>0. \end{cases}$$

(*Совет.* Преобразуйте систему обслуживания с c каналами в *эквивалентную* одноканальную, для которой

$$P\{t>T\} = P\left\{\min_{1\le i\le c} t_i > T\right\} = \left\{e^{-\mu T}\right\}^c = e^{-\mu c T},$$

где t — время обслуживания в эквивалентной одноканальной системе обслуживания.)

18. Докажите, что если плотность вероятности $w_q(T)$ задается формулой из предыдущего упражнения, то

$$P\{T > y\} = P\{T > 0\}e^{-(c\mu-\lambda)y},$$

где $P\{T > 0\}$ — вероятность того, что поступающий в систему клиент будет ждать обслуживания.

19. Докажите, что в модели $(M/M/c)$: (FCFS/∞/∞) системы обслуживания плотность вероятности времени ожидания клиента в очереди имеет вид

$$w(\tau) = \mu e^{-\mu\tau} + \frac{\rho^c \mu e^{-\mu\tau}}{(c-1)!(c-\rho-1)}\left\{\frac{1}{c-\rho} - e^{-\mu(c-\rho-1)\tau}\right\}p_0, \quad \tau \geq 0.$$

(*Подсказка.* Распределение случайной величины τ представляет собой свертку распределений времени ожидания в очереди T (см. упр. 17) и времени обслуживания.)

Модель (M/M/c) : (GD/N/∞), c ≤ N. Эта модель обслуживающей системы отличается от модели $(M/M/c)$: (GD/∞/∞) тем, что емкость системы ограничена сверху значением N (тогда максимальная длина очереди равна $N - c$). Интенсивности поступления и обслуживания клиентов равны λ и μ соответственно. Эффективная интенсивность поступления заявок в систему обслуживания $\lambda_{эфф}$ меньше λ в силу ограниченности емкости системы значением N.

Параметры λ_n и μ_n общей модели обслуживающей системы в данной модели определяются следующим образом:

$$\lambda_n = \begin{cases} \lambda, & 0 \leq n \leq N, \\ 0, & n > N, \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 0 \leq n \leq c, \\ c\mu, & c \leq n \leq N. \end{cases}$$

Подставляя λ_n и μ_n в общее выражение для p_n из раздела 17.5 и используя обозначение $\rho = \lambda/\mu$, получаем

$$p_n = \begin{cases} \dfrac{\rho^n}{n!}p_0, & 0 \leq n \leq c, \\ \dfrac{\rho^n}{c!c^{n-c}}p_0, & c \leq n \leq N, \end{cases}$$

где

$$p_0 = \begin{cases} \left[\sum_{n=0}^{c-1}\dfrac{\rho^n}{n!} + \dfrac{\rho^c\left(1-\left(\dfrac{\rho}{c}\right)^{N-c+1}\right)}{c!\left(1-\dfrac{\rho}{c}\right)}\right]^{-1}, & \dfrac{\rho}{c} \neq 1, \\ \left[\sum_{n=0}^{c-1}\dfrac{\rho^n}{n!} + \dfrac{\rho^c}{c!}(N-c+1)\right]^{-1}, & \dfrac{\rho}{c} = 1. \end{cases}$$

Далее мы вычисляем L_q для случая, когда $\rho/c \neq 1$:

$$L_q = \sum_{n=c}^{N}(n-c)p_n = \sum_{j=0}^{N-c} jp_{j+c} = \frac{\rho^c \rho}{c!c}\sum_{j=0}^{N-c} j\left(\frac{\rho}{c}\right)^{j-1} p_0 = \frac{\rho^{c+1}}{cc!}\frac{d}{d\left(\frac{\rho}{c}\right)}\sum_{j=0}^{N-c}\left(\frac{\rho}{c}\right)^j =$$

$$= \frac{\rho^{c+1}}{(c-1)!(c-\rho)^2}\left\{1-\left(\frac{\rho}{c}\right)^{N-c+1} - (N-c+1)\left(1-\frac{\rho}{c}\right)\left(\frac{\rho}{c}\right)^{N-c}\right\}p_0.$$

Может также быть показано, что для случая, когда $\rho/c = 1$, выражение для L_q имеет следующий вид:

$$L_q = \frac{\rho^c (N-c)(N-c+1)}{2c!}p_0, \quad \frac{\rho}{c} = 1.$$

Для определения W_q и, следовательно, W_s и L_s, необходимо получить выражение для $\lambda_{\text{эфф}}$. Поскольку ни один клиент не может попасть в систему после того, как достигнут лимит N по ее вместимости, то

$$\lambda_{\text{потери}} = \lambda p_N,$$
$$\lambda_{\text{эфф}} = \lambda - \lambda_{\text{потери}} = (1 - p_N)\lambda.$$

Пример 17.6–6

Пусть в задаче, связанной с объединением служб такси, которая рассматривалась в примере 17.6–5, известно, что объединенная служба такси не имеет финансовых возможностей для покупки новых автомашин. Друг нового хозяина советует ему, что одной из возможностей уменьшения времени ожидания клиентами заказанного такси является информирование их диспетчерской службой о возможных задержках с прибытием заказанной автомашины, как только список ожидающих клиентов достигает шести. Эта мера, несомненно, заставит новых клиентов искать обслуживания в другом месте, что уменьшит время ожидания тех клиентов, которые уже ожидают в очереди. Требуется исследовать, насколько правдоподобным является совет друга.

Ограничение списка ожидающих в очереди до 6 клиентов равносильно тому, что емкость системы становится равной $N = 6 + 4 = 10$ клиентов. Следовательно, мы имеем дело с системой обслуживания модели $(M/M/4)$: $(GD/10/\infty)$ с $\lambda = 16$ клиентов в час и $\mu = 5$ поездок в час. На рис. 17.9 представлены выходные данные, полученные с помощью программы TORA для этой модели.

Среднее время ожидания W_q, в случае отсутствия ограничения на емкость системы, равняется 0.149 часа ($\approx$ 9 минут) (см. рис. 17.8), что почти в два раза больше значения 0.075 часа ($\approx$ 4.5 минуты) аналогичного показателя при наличии ограничения на емкость системы. Это существенное уменьшение функциональной характеристики системы достигнуто за счет потери примерно 3.6% потенциальных клиентов. Этот результат, однако, не отражает возможной потери расположения клиентов к деятельности службы такси.

```
Problem title: Example 17-9
Scenario 1 -- (M/M/4):(GD/10/*)
------------------------------------------------------------------
Lambda =    16.00000        Lambda eff =    15.42815
Mu =     5.00000            Rho =     3.20000

Ls =     4.23984            Lq =     1.15421
Ws =    0.27481             Wq =     0.07481
------------------------------------------------------------------
Values of p(n) for n=0 to 10, else p(n) < .00001

 0 0.03121   1 0.09986   2 0.15977   3 0.17043   4 0.13634
 5 0.10907   6 0.08726   7 0.06981   8 0.05584   9 0.04468
10 0.03574

Cumulative values of p(n) for n=0 to 10

 0 0.03121   1 0.13106   2 0.29084   3 0.46126   4 0.59760
 5 0.70667   6 0.79393   7 0.86374   8 0.91958   9 0.96426
10 1.00000
```

Рис. 17.9

Упражнения 17.6,f

1. В примере 17.6–6 определите следующие показатели.
   a) Среднее количество свободных такси.
   b) Вероятность того, что клиент, вызывающий такси, будет последним из тех, кто ставится в очередь.
   c) Максимальное число ожидающих в очереди клиентов при условии, что время ожидания не превышает трех минут.
2. Газозаправочная станция для автомобилей располагает двумя газовыми насосами. В очереди, ведущей к насосам, могут расположиться не более пяти автомашин, включая те, которые обслуживаются. Если уже нет места, прибывающие автомобили уезжают искать другую заправку. Распределение прибывающих автомобилей является пуассоновским с математическим ожиданием 20 автомобилей в час. Время обслуживания клиентов имеет экспоненциальное распределение с математическим ожиданием 6 минут. Определите следующие величины.
   a) Процент автомобилей, которые будут искать другую заправку.
   b) Процент времени, когда используется только один из насосов.
   c) Процент времени использования двух насосов.
   d) Вероятность того, что прибывающий автомобиль найдет свободное место в очереди.

e) Емкость очереди, которая обеспечит потерю в среднем не более 10% потенциальных клиентов.
f) Емкость очереди, при которой вероятность того, что оба насоса свободны, не превышает 0.05.

3. Небольшая ремонтная мастерская имеет трех механиков. В начале марта каждого года клиенты приносят в мастерскую свои культиваторы и газонокосилки для ремонта и технического обслуживания. Мастерская стремится принять все, что приносят клиенты. Однако когда очередной клиент видит на полу мастерской массу механизмов, ожидающих обслуживания, он уходит в другое место в поисках более быстрого обслуживания. На полу мастерской размещается не более 15 культиваторов или газонокосилок, не учитывая тех, которые уже ремонтируются. Клиенты прибывают в мастерскую в среднем каждые 10 минут, а на выполнение механиком одного ремонта уходит в среднем 30 минут. Как время между последовательными приходами клиентов, так и время выполнения работы подчиняются экспоненциальному распределению. Определите следующие величины.

a) Среднее число незанятых механиков.
b) Число потерянных потенциальных клиентов на протяжении десятичасового рабочего дня по причине ограниченной емкости мастерской.
c) Вероятность того, что следующий клиент будет обслужен в мастерской.
d) Вероятность того, что по крайней мере один механик будет свободен.
e) Среднее количество культиваторов и газонокосилок, которые ожидают обслуживания.
f) Показатель общей производительности мастерской.

4. Студенты первого курса одного из американских университетов приезжают на лекции на своих автомобилях (даже несмотря на то, что большинство из них нуждаются в проживании на территории университета и могут пользоваться удобной университетской бесплатной транспортной системой). На протяжении первых двух недель осеннего семестра на университетской территории преобладает беспорядок в транспортном движении, так как первокурсники отчаянно пытаются найти места для стоянки автомашин. С необычной самоотверженностью студенты терпеливо ожидают на пешеходных дорожках возле стоянок для автомашин, когда кто-нибудь заберет свою автомашину, чтобы можно было поставить на стоянку свои авто. Рассмотрим следующий характерный сценарий. Автостоянка имеет 30 мест, но может также расположить еще 10 автомашин на пешеходных дорожках. Эти 10 автомашин не могут постоянно оставаться на пешеходных дорожках и должны ожидать, пока хоть одно место на стоянке освободится. Первокурсники прибывают к автостоянке в соответствии с распределением Пуассона с математическим ожиданием 20 автомашин в час. Время пребывания автомашины на стоянке подчиняется экспоненциальному распределению со средним значением примерно 60 минут.

a) Каков процент первокурсников, вынужденных повернуть обратно по той причине, что они не смогли поставить автомашину на стоянку?

b) Какова вероятность того, что прибывающий автомобиль будет ожидать на пешеходной дорожке?

c) Какова вероятность того, что прибывающий автомобиль займет единственное оставшееся место на стоянке?

d) Определите среднее количество занятых мест на стоянке.

e) Определите среднее количество занятых мест на пешеходных дорожках.

f) Определите среднее число первокурсников, которые не попадут на лекции на протяжении восьмичасового периода, так как стоянка будет полностью заполнена.

5. Проверьте правильность формулы для p_0 в модели $(M/M/c)$: (GD/N/∞) для случая, когда $\rho/c \neq 1$.

6. Для модели $(M/M/c)$: (GD/N/∞) докажите равенство $\lambda_{эфф} = \mu\bar{c}$, где $\bar{c}$ — среднее количество занятых сервисов.

7. Проверьте правильность формул для p_0 и L_q в модели $(M/M/c)$:(GD/N/∞) для случая, когда $\rho/c = 1$.

8. Для модели $(M/M/c)$: (GD/N/∞) при $N = c$ из соотношений для общей модели (раздел 17.4) определите λ_n и μ_n, затем покажите, что формула для p_n имеет такой вид:

$$p_n = \frac{\rho^n}{n!} p_0, \quad n = 1, 2, .., c,$$

где

$$p_0 = \left\{ 1 + \sum_{n=1}^{c} \frac{\rho^n}{n!} \right\}^{-1}.$$

Модель самообслуживания (M/M/∞) : (GD/∞/∞). В этой модели количество сервисов является неограниченным, так как клиент выступает одновременно и в роли сервиса. Типичным примером модели самообслуживания является сдача письменной части экзамена на право вождения автомобиля. Газозаправочные станции автомобилей с самообслуживанием и банковские автоматы с 24-часовым режимом работы не вписываются в рассматриваемую здесь модель, так как обслуживающими устройствами в этих случаях являются, по существу, насосы и банковские автоматы соответственно.

В рассматриваемой модели предполагается, что интенсивность поступления клиентов λ является постоянной. Интенсивность обслуживания μ также является постоянной. Воспользовавшись общей моделью из раздела 17.5, имеем

$$\lambda_n = \lambda, \quad n = 0, 1, 2, \ldots,$$
$$\mu_n = n\mu, \quad n = 0, 1, 2, \ldots.$$

Следовательно,

$$p_n = \frac{\lambda^n}{n!\mu^n} p_0 = \frac{\rho^n}{n!} p_0, \quad n = 0, 1, 2, \ldots.$$

Из равенства $\sum_{n=0}^{\infty} p_n = 1$ вытекает, что

$$p_0 = \frac{1}{1+\rho+\frac{\rho^2}{2!}+\ldots} = \frac{1}{e^{\rho}} = e^{-\rho}.$$

В результате получаем

$$p_n = \frac{e^{-\rho}\rho^n}{n!}, \quad n = 0, 1, 2, \ldots .$$

Эти вероятности совпадают с вероятностями распределения Пуассона с математическим ожиданием $L_s = \rho$. Как и следовало ожидать, здесь (вследствие принципа самообслуживания) $L_q = W_q = 0$.

Пример 17.6–7

Инвестор вкладывает 1000 долларов в месяц в специальный тип облигаций фондовой биржи. Так как инвестор должен ждать возможности хорошей "покупки", фактическое время совершения этой покупки является случайным. Инвестор обычно держит облигации в среднем три года, но продаст их в случайный момент времени, когда представится такая возможность. Хотя инвестор известен как хитрый биржевой игрок, опыт прошлого показывает, что около 25% облигаций теряют в цене примерно 20% в год. Остальные 75% облигаций повышаются в цене примерно на 12% в год. Оценим среднюю стоимость акций инвестора на протяжении длительного периода.

Эту ситуацию можно представить в виде модели $(M/M/\infty) : (GD/\infty/\infty)$, так как инвестор не должен ждать в очереди, чтобы купить или продать облигации. Среднее время между размещениями заказа равняется 1 месяц, что дает значение $\lambda = 12$ облигаций в год. Интенсивность продажи облигаций равна $\mu = 1/3$ облигаций в год.

При указанных значениях λ и μ получаем

$$L_s = \rho = \frac{\lambda}{\mu} = 36 \text{ облигаций.}$$

Средняя годовая стоимость облигаций инвестора на протяжении длительного периода оценивается следующей величиной:

$$(0.25L_s \times 1000)(1 - 0.20) + (0.75L_s \times 1000)(1 + 0.12) = 63\,990 \text{ долларов.}$$

Упражнения 17.6,g

1. В примере 17.6–7 вычислите следующие показатели.

   a) Вероятность того, что инвестор продаст все свои облигации.
   b) Вероятность того, что инвестор будет иметь больше 10 облигаций.
   c) Вероятность того, что инвестор будет иметь от 30 до 40 облигаций.
   d) Оцените среднюю стоимость акций инвестора на протяжении длительного периода, если только 10% облигаций теряют в цене 30% в год, а остальные 90% повышаются в цене на 15% в год.

2. Новые водители перед тем, как им выдадут задание по дорожному вождению, должны сдать письменную часть (тесты) экзамена на право вождения автомобиля. Эти тесты обычно проводятся городским управлением полиции. Статистика показывает, что среднее количество письменных тестов за 8-часовой день равняется 100. Среднее время, необходимое для выполнения теста, равно примерно 30 минут. Однако фактическое прибытие каждого экзаменующегося и время, которое он тратит на сдачу экзамена, являются случайными величинами. Необходимо определить следующее.

   a) Среднее количество посадочных мест в зале для сдачи экзамена, которое должно обеспечить управление полиции.
   b) Вероятность того, что число экзаменующихся превысит среднее количество посадочных мест в зале для сдачи экзамена.
   c) Вероятность того, что в какой-нибудь день не будет проведено ни одного экзамена.

3. Покажите (используя программное обеспечение TORA), что для малого параметра $\rho = 0.1$ значения величин W_s, L_s, W_q, L_q и p_n в модели $(M/M/c)$: (GD/∞/∞) при $c \geq 10$ сервисов можно надежно оценить с помощью менее громоздких формул для модели $(M/M/\infty)$: (GD/∞/∞).

4. Повторите предыдущее упражнения для значения $\rho = 9$ и покажите, что в этом случае значение c должно быть больше 20. Какой общий вывод можно сделать относительно использования модели $(M/M/\infty)$: (GD/∞/∞) для оценки показателей модели $(M/M/c)$: (GD/∞/∞), исходя из результатов, полученных в упр. 3 и 4?

17.6.4. Модель $(M/M/R)$: $(GD/K/K)$ при $R < K$

Базовым примером для этой модели является цех, насчитывающий K станков. Всякий раз, когда станки выходят из строя, прибегают к услугам одного из механиков, бригада которых состоит из R человек. Интенсивность поломок, отнесенная *к одному станку*, равняется λ поломок в единицу времени. Механик ремонтирует сломанные станки с интенсивностью μ станков в единицу времени. Предполагается, что моменты времени поломок и время ремонта подчиняются распределению Пуассона.

Эта модель отличается от всех рассмотренных ранее тем, что мощность источника, генерирующая "клиентов", конечна. Например, в модели цеха источник может породить конечное количество заявок на ремонт. Это положение становится очевидным, если предположить, что все станки в цехе сломаны, тогда больше не поступит ни одной заявки на ремонт. По существу, лишь работающие станки могут сломаться и, следовательно, генерировать заявки на ремонт.

При заданной интенсивности λ поломок на один станок интенсивность поломок во *всем цехе* пропорциональна количеству станков в рабочем состоянии. В терминологии систем обслуживания наличие n станков *в системе* означает, что n станков сломаны. Следовательно, интенсивность поломок во всем цехе вычисляется так:

$$\lambda_n = (K - n)\lambda, \quad 0 \leq n \leq K.$$

В обозначениях общей модели системы обслуживания из раздела 17.5 имеем следующее:

$$\lambda_n = \begin{cases} (K-n)\lambda, & 0 \le n \le K, \\ 0, & n \ge K, \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 0 \le n \le R, \\ R\mu, & R \le n \le K, \\ 0, & n > K. \end{cases}$$

Теперь из общей модели можно получить (проверьте!) следующие выражения:

$$p_n = \begin{cases} \binom{K}{n}\rho^n p_0, & 0 \le n \le R, \\ \binom{K}{n}\dfrac{n!\rho^n}{R!R^{n-R}} p_0, & R \le n \le K, \end{cases}$$

$$p_0 = \left\{ \sum_{n=0}^{R} \binom{K}{n}\rho^n + \sum_{n=R+1}^{K} \binom{K}{n}\frac{n!\rho^n}{R!R^{n-R}} \right\}^{-1}.$$

В этой модели трудно получить в замкнутой форме выражение для L_s или L_q, и, следовательно, они должны вычисляться в соответствии с их определением:

$$L_s = \sum_{n=0}^{K} n p_n.$$

Значение $\lambda_{эфф}$ можно вычислить в следующей форме:

$$\lambda_{эфф} = M\{\lambda(K-n)\} = \lambda(K-L_s).$$

Используя формулы из раздела 17.6.1, можно вычислить оставшиеся функциональные показатели W_s, W_q и L_q.

Пример 17.6–8

Компания располагает цехом, насчитывающим 22 станка. Известно, что каждый станок выходит из строя в среднем один раз в два часа. На его ремонт уходит в среднем 12 минут. Как время между поломками станков, так и время, необходимое для ремонта, являются экспоненциально распределенными случайными величинами. Компания заинтересована в определении минимального числа механиков, необходимых для обеспечения "плавной" работы цеха.

Описанную ситуацию можно проанализировать путем исследования производительности станков как функции числа механиков. Такая мера производительности может быть определена в следующем виде:

$$\begin{pmatrix} \text{Производительность} \\ \text{станков} \end{pmatrix} = \frac{\text{Количество станков} - \text{сломанные станки}}{\text{Количество станков}} \times 100 =$$

$$= \frac{22 - L_s}{22} \times 100.$$

На рис. 17.10 представлены полученные с помощью программы TORA сравнительные характеристики обслуживающей системы для $R = 1, 2, 3, 4$ при $\lambda = 0.5$ поломки в час на один станок и $\mu = 5$ ремонтов в час. Соответствующие значения производительности представлены в следующей таблице.

Количество механиков, R	1	2	3	4
Производительность станков (в %)	45.44	80.15	88.79	90.45
Рост производительности (в %)	—	34.71	8.64	1.66

Problem title: Example 17.6-8
Comparative measures: Nbr of scenarios = 4

Nbr	c	Lambda	Mu	l'da_eff	Ls	Ws	Lq	Wq
1	1	0.500	5.000	4.998	12.004	2.402	11.004	2.202
2	2	0.500	5.000	8.816	4.368	0.495	2.604	0.295
3	3	0.500	5.000	9.767	2.466	0.252	0.513	0.052
4	4	0.500	5.000	9.950	2.100	0.211	0.110	0.011

Рис. 17.10

Приведенные результаты показывают, что неприемлемым является наличие лишь одного механика, так как в этом случае будет очень низкая производительность (= 45.44%). Если число механиков увеличивается до двух, то производительность станков возрастает на 34.71% и достигает 80.15%. Когда имеем трех механиков, то производительность станков возрастает примерно на 8.64% и достигает значения 88.79%, в то время как при наличии четырех механиков производительность возрастает лишь на 1.66% и достигает значения 90.45%.

Судя по этим результатам, использование двух механиков может быть оправданным. Ситуация с тремя механиками является не такой убедительной, так как производительность при этом возрастает лишь на 8.64%. Возможно, что сравнение в денежном эквиваленте содержания третьего механика и прибыли, обусловленной ростом производительности станков, на 8.64% можно использовать для решения этого вопроса (см. раздел 17.10, где рассматриваются стоимостные характеристики модели). Что касается приема на работу четвертого механика, то скудный рост производительности станков на 1.66%, который при этом достигается, не оправдывает данного шага.

Упражнения 17.6,h

1. В примере 17.6–8 выполните следующее.

   a) Проверьте значения величин $\lambda_{эфф}$, которые приведены в листинге на рис. 17.10.
   b) Подсчитайте среднее число свободных механиков при $R = 4$.
   c) Вычислите вероятность того, что все механики свободны при $R = 3$.
   d) Вычислите вероятность того, что большинство механиков свободны при $R = 3$ и $R = 4$.

2. В примере 17.6–8 определите и затем вычислите производительность работы механиков при $R = 1, 2, 3, 4$. Используйте эту информацию вместе с показателем производительности станков для решения вопроса о количестве механиков, которых компании следует принять на работу.
3. В вычислениях, которые приведены в листинге на рис. 17.10, может вызвать недоразумение то обстоятельство, что средняя интенсивность поломки станков в цехе $\lambda_{эфф}$ увеличивается с ростом R. Объясните, почему следует ожидать увеличения $\lambda_{эфф}$.
4. Оператор обслуживает пять автоматических станков. После того как каждый станок завершает выполнение пакета программ, оператор должен его перенастроить на выполнение нового пакета. Время выполнения пакета программ является экспоненциально распределенной случайной величиной со средним значением 45 минут. Время настройки также описывается экспоненциальным распределением с математическим ожиданием 8 минут.

   a) Определите среднее количество станков, которые ожидают настройки.
   b) Вычислите вероятность того, что все станки работают.
   c) Определите среднее время простоя станка.

5. Обслуживающая фирма выполняет разнообразные работы, такие как уборка в саду, хозяйственные работы, подрезка деревьев и покраска домов. Четыре сотрудника фирмы оставляют контору, каждый получив свое первое задание. После выполнения задания сотрудник может позвонить в контору и получить информацию относительно следующей заявки для выполнения. Время выполнения задания имеет экспоненциальное распределение со средним значением 45 минут. Время переезда между последовательными работами также имеет экспоненциальное распределение со средним значением 20 минут.

   a) Определите среднее число сотрудников фирмы, которые перемещаются между последовательными работами.
   b) Вычислите вероятность того, что нет ни одного сотрудника в процессе перемещения.

6. После долгого ожидания, благодаря чудесам медицины, семья Ньюборнов была вознаграждена рождением пяти близнецов: двух мальчиков и трех девочек. На протяжении пяти первых месяцев жизнь малышей состояла из двух состояний: бодрствующего (и, в основном, с криком) и спящего. По словам Ньюборнов эти состояния для малышей никогда не совпадают и проявляются совершенно случайно. На самом деле госпожа Ньюборн, статистик по профессии, считает, что время, когда каждый из малышей кричит, экспоненциально распределено со средним значением 30 минут. Время сна каждого малыша также является экспоненциально распределенной случайной величиной со средним значением 2 часа. Определите следующее.

   a) Среднее число малышей, которые бодрствуют в некоторый момент времени.
   b) Вероятность того, что все малыши спят.
   c) Вероятность того, что Ньюборны не будут счастливы по той причине, что бодрствующих малышей (которые при этом кричат) больше, чем спящих.

7. Проверьте корректность выражения для вероятностей p_n в модели $(M/M/R) : (GD/K/K)$.

8. Покажите, что интенсивность поломок в цехе можно вычислить по формуле $\lambda_{эфф} = \mu \overline{R}$, где $\overline{R}$ — среднее число занятых механиков.

9. Проверьте следующие формулы для случая наличия одного механика ($R = 1$):

$$p_n = \frac{K!\rho^n}{(K-n)!} p_0,$$

$$p_0 = \left\{1 + \sum_{n=1}^{K} \frac{K!\rho^n}{(K-n)!}\right\}^{-1},$$

$$L_s = K - \frac{(1-p_0)}{\rho}.$$

17.7. Модель $(M/G/1):(GD/\infty/\infty)$. Формула Поллачека–Хинчина

Анализ моделей массового обслуживания, в которых входные и выходные потоки не подчиняются пуассоновскому распределению, является сложным. Вообще в таких случаях в качестве альтернативного аппарата для анализа моделей обслуживания целесообразно использовать методы имитационного моделирования (см. главу 18).

В этом разделе рассматривается один из немногих случаев системы массового обслуживания, не подчиняющейся пуассоновскому распределению, для которого могут быть получены аналитические результаты. Речь идет о том случае, когда время обслуживания t имеет произвольное распределение с математическим ожиданием $M\{t\}$ и дисперсией $D\{t\}$. Для такой модели известны аналитические формулы для основных функциональных характеристик обслуживающей системы, таких как L_s, L_q, W_s и W_q. Однако в силу аналитических трудностей невозможно получить в замкнутой форме выражения для вероятностей p_n.

Пусть λ — интенсивность поступления клиентов в системе обслуживания с одним сервисом. При заданных значениях $M\{t\}$ и $D\{t\}$ для времени обслуживания и условии $\lambda M\{t\} < 1$ с использованием сложного анализа, связанного с применением аппарата цепей Маркова, можно показать, что[1]

$$L_s = \lambda M\{t\} + \frac{\lambda^2\left(M^2\{t\} + D\{t\}\right)}{2\left(1 - \lambda M\{t\}\right)}, \quad \lambda M\{t\} < 1.$$

Так как $\lambda_{эфф} = \lambda$, то остальные функциональные характеристики обслуживающей системы (L_q, W_s и W_q) можно получить из формулы для L_s, как это сделано в разделе 17.6.1.

Пример 17.7–1

Пусть в задаче об автомойке из примера 17.6–2 сделано дополнительное предположение, что установлено новое оборудование, поэтому время обслуживания для всех автомобилей является постоянным и равным 10 минутам. Как новое оборудование влияет на функционирование автомойки?

[1] Данная формула, собственно, и называется формулой Поллачека–Хинчина — *Прим. ред.*

На основании данных примера 17.6–2 имеем, что $\lambda_{эфф} = \lambda = 4$ автомобиля в час. Время обслуживания является постоянным, так что $M\{t\} = 10/60 = 1/6$ часа и $D\{t\} = 0$. Следовательно,

$$L_s = 4\left(\frac{1}{6}\right) + \frac{4^2\left[\left(\frac{1}{6}\right)^2 + 0\right]}{2\left(1-\frac{4}{6}\right)} = 1.333 \text{ автомобилей},$$

$$L_q = 1.333 - \left(\frac{4}{6}\right) = 0.667 \text{ автомобиля},$$

$$W_s = \frac{1.333}{4} = 0.333 \text{ часа},$$

$$W_q = \frac{0.667}{4} = 0.167 \text{ часа}.$$

Эти результаты можно также получить с использованием программы TORA.

Интересно отметить, что несмотря на то, что интенсивности как поступления клиентов, так и обслуживания в рассматриваемой модели такие же, как и в пуассоновском случае из примера 17.6–2 ($\lambda = 4$ автомобиля в час и $\mu = 1/M\{t\} = 6$ автомобилей в час), среднее время ожидания в рассматриваемом случае меньше, так как время обслуживания является постоянным, о чем свидетельствует следующая таблица.

	$(M/M/1) : (GD/\infty/\infty)$	$(M/D/1) : (GD/\infty/\infty)$
W_s (часы)	0.5	0.333
W_q (часы)	0.333	0.167

В этих результатах есть смысл, поскольку постоянное время обслуживания подразумевает *большую определенность* в функционировании системы.

Упражнения 17.7,а

1. В примере 17.7–1 вычислите процент времени, когда оборудование простаивает.
2. Решите задачу из примера 17.7–1 в предположении, что распределение времени обслуживания задается следующим образом.
   a) Равномерное на интервале от 8 до 20 минут.
   b) Нормальное с $\mu = 12$ минут и $\sigma = 3$ минуты.
   c) Дискретное со значениями, равными 4, 8 и 15 минут, и вероятностями 0.2, 0.6 и 0.2 соответственно.
3. Фирма устанавливает деревянные крыши в новых и старых жилых домах в штате Арканзас. Будущие клиенты обращаются за услугами фирмы случайным образом с интенсивностью 9 работ за месяц (30 дней) и заносятся в очередь для выполнения в соответствии с принципом "первым пришел — первым обслуживаешься". Дома отличаются своими размерами, но есть основания утверждать, что площади крыш равномерно распределены между 150 и 300 кв. единицами (1 кв. едини-

ца = 100 футов^2). Рабочая бригада за день может выполнить работы на 75 кв. единицах площади крыши. Определите следующие показатели.

a) Среднее количество невыполненных заказов по установке крыш.
b) Среднее время, которое клиент вынужден ждать до завершения установки крыши.
c) Если рабочая бригада за день сможет выполнить работы на 150 кв. единицах площади крыш, как это повлияет на среднее время ожидания завершения установки крыши.

4. Фирма *Оптика* изготавливает очки в соответствии с рецептами, которые получает от своих клиентов. Каждый рабочий специализируется на изготовлении определенных типов очков. Фирма испытывает некоторые затруднения с изготовлением бифокальных и трифокальных типов очков. Рабочие получают 30 заказов на восьмичасовой рабочий день. Время изготовления очков по рецепту нормально распределено с математическим ожиданием 12 минут и стандартным отклонением 3 минуты. Затем рабочий проверяет очки, затрачивая на это от 2 до 4 минут с равномерным распределением соответствующего времени, после чего может приступить к выполнению нового заказа. Определите следующие показатели.

a) Процент времени простоя рабочего.
b) Среднее количество невыполненных заказов на изготовление бифокальных и трифокальных типов очков.
c) Среднее время выполнения заказа.

5. Изделие поступает на обработку в соответствии с распределением Пуассона с интенсивностью одно изделие в единицу времени. Обработка изделия требует выполнения двух последовательных операций, за которыми следит один рабочий. Для выполнения первой операции используется полуавтоматический станок, который выполняет свои операции точно за 28 минут. На второй операции осуществляются незначительные изменения и регулировка, и время ее выполнения зависит от качества изделия после первой операции. Время выполнения второй операции равномерно распределено на интервале от 3 до 6 минут. Так как выполнение каждой операции требует полного внимания рабочего, нельзя начать обработку нового изделия на полуавтоматическом станке до тех пор, пока не будет выполнена вторая операция для обрабатываемого изделия.

a) Определите количество изделий, ожидающих обработки на полуавтоматическом станке.
b) Чему равен процент времени простоя рабочего?
c) Сколько в среднем требуется времени для обработки изделия обеими операциями?

6. Модель $(M/D/1)$: (GD/∞/∞). Покажите, что для случая, когда время обслуживания постоянно (в этом случае $D\{t\} = 0$), формула Поллачека–Хинчина принимает следующий вид:

$$L_s = \rho + \frac{\rho^2}{2(1-\rho)},$$

где $\mu = 1/M\{t\}$ и $\rho = \lambda/\mu = \lambda M\{t\}$.

7. Модель $(M/E_m/1):(GD/\infty/\infty)$. Время обслуживания подчиняется распределению Эрланга с параметрами m и μ (здесь $M\{t\} = m/\mu$ и $D\{t\} = m/\mu^2$). Покажите, что в данном случае формула Поллачека–Хинчина принимает такой вид:

$$L_s = m\rho + \frac{m(1+m)\rho^2}{2(1-m\rho)}.$$

8. Покажите, что формула Поллачека–Хинчина сводится к формуле для L_s в модели $(M/M/1):(GD/\infty/\infty)$, когда время обслуживания подчиняется экспоненциальному закону с математическим ожиданием $1/\mu$ единиц времени.

9. a) Пусть в системе обслуживания с c параллельными сервисами клиенты прибывают в соответствии с распределением Пуассона со средней интенсивностью λ. Прибывающие клиенты распределяются между сервисами (занятыми или свободными) на основе строгого чередования. Найдите распределение времени между последовательными поступлениями клиентов на сервисы.

   b) В п. а) этого упражнения предположите, что прибывающие клиенты случайным образом распределяются между c сервисами с вероятностями α_i, $\alpha_i \geq 0$, $i = 1, 2, ..., c$ и $\alpha_1 + \alpha_2 + ... + \alpha_c = 1$. Найдите распределение времени между последовательными поступлениями клиентов на сервисы.

17.8. Другие модели массового обслуживания

В предыдущих разделах основное внимание было сосредоточено на пуассоновских системах массового обслуживания. Однако в литературе по теории массового обслуживания рассматривается множество других моделей. В частности, системы массового обслуживания с приоритетами и системы обслуживания непуассоновского типа составляют существенную часть соответствующей научной литературы. С такими моделями можно познакомиться в большинстве специальных книг по теории массового обслуживания.

17.9. Модели принятия решений в теории массового обслуживания

Уровень обслуживания в системе является функцией интенсивности обслуживания μ и количества c параллельно работающих сервисов. В этом разделе рассматриваются две модели принятия решений для определения "подходящих" уровней обслуживания для систем массового обслуживания: (1) модель со стоимостными характеристиками и (2) модель предпочтительного уровня обслуживания. В обеих моделях более высокий уровень обслуживания подразумевает уменьшение времени ожидания в системе. В этих моделях для поиска равновесия между конфликтующими факторами (уровнем обслуживания и временем ожидания в системе) используются функциональные показатели обслуживающей системы, которые получены ранее для различных моделей.

17.9.1. Модель со стоимостными характеристиками

Модели со стоимостными характеристиками стремятся уравновесить два конфликтующих стоимостных показателя.

1. Затраты на обслуживание.
2. Потери, обусловленные задержками в предоставлении услуг (время ожидания клиента).

Эти два вида затрат конфликтуют между собой, так как увеличение одного из них автоматически ведет к уменьшению другого и наоборот (см. рис. 17.1).

Пусть *уровень обслуживания* представляет переменная x, равная μ или c. Тогда модель со стоимостными характеристиками можно представить в следующем виде:

$$\mathrm{COC}(x) = \mathrm{CTC}(x) + \mathrm{CTO}(x),$$

где

СОС — средняя общая стоимость *в единицу времени*,
СТС — средняя стоимость обслуживания *в единицу времени*,
СТО — средняя стоимость ожидания *в единицу времени*.

Простейшим видом функций СТС и СТО являются линейные функции:

$$\mathrm{CTC}(x) = C_1 x,$$
$$\mathrm{CTO}(x) = C_2 L_s,$$

где

C_1 — *удельная* стоимость на единицу x в единицу времени,
C_2 — "цена" ожидания в единицу времени на (ожидающего) клиента.

Следующие два примера иллюстрируют использование стоимостной модели. В первом примере предполагается, что x равняется интенсивности обслуживания μ, во втором — количеству параллельных сервисов c.

Пример 17.9–1

Издательская фирма покупает высокоскоростной копировальный аппарат для коммерческих целей. Продавцы предложили четыре модели копировальных аппаратов, характеристики которых приведены в следующей таблице.

Модель копировального аппарата	Эксплуатационные затраты ($/час)	Скорость печати (стр./минута)
1	15	30
2	20	36
3	24	50
4	27	66

Заказы поступают на фирму в соответствии с пуассоновским распределением с математическим ожиданием четыре работы на протяжении 24-часового дня. Объем работы является случайной величиной, но в среднем составляет примерно 10 000 страниц. Договоры с клиентами предусматривают штраф в сумме 80 долларов (за одну работу) за задержку выполнения заказа на один день. Какой копировальный аппарат следует купить фирме?

Пусть индекс i представляет номер модели копировального аппарата i ($i = 1, 2, 3, 4$). Общая ожидаемая стоимость обслуживания *в день*, связанная с использованием копировального аппарата i, представлена в следующем виде:

$$\mathrm{COC}_i = \mathrm{CTC}_i + \mathrm{CTO}_i = C_{1i} \times 24 + C_{2i}L_{si} = 24C_{1i} + 80L_{si}, \quad i = 1, 2, 3, 4.$$

Значения величин C_{1i} приведены в постановке задачи. Определим значения L_{si}, считая, что с практической точки зрения каждый копировальный аппарат может рассматриваться как модель $(M/M/1):(\mathrm{GD}/\infty/\infty)$ обслуживающей системы. При этом интенсивность поступления заявок $\lambda = 4$ работы в день, а интенсивности μ_i обслуживания, соответствующие модели i копировального аппарата, приведены в следующей таблице.

Модель i	Интенсивность обслуживания μ_i (работы/день)
1	4.32
2	5.18
3	7.20
4	9.50

Проиллюстрируем определение интенсивности обслуживания на примере модели 1.

$$\text{Среднее время выполнения работы} = \frac{10000}{30} \times \frac{1}{60} = 5.56 \text{ часа.}$$

Следовательно,

$$\mu_1 = \frac{24}{5.56} = 4.32 \text{ работы в день.}$$

Значения L_{si}, которые вычислены с помощью программы TORA, приведены в следующей таблице.

Модель i	λ_i (работы/день)	μ_i (работы/день)	L_{si} (работы)
1	4	4.32	12.50
2	4	5.18	3.39
3	4	7.20	1.25
4	4	9.50	0.73

Получены следующие стоимостные характеристики для четырех рассматриваемых моделей копировальных аппаратов.

Модель i	CTC_i (\$)	CTO_i (\$)	COC_i (\$)
1	360.00	1000.00	1360.00
2	480.00	271.20	751.20
3	**576.00**	**100.00**	**676.00**
4	648.00	58.40	706.40

Следовательно, для третьей модели копировального аппарата стоимость самая низкая.

Упражнения 17.9,а

1. В примере 17.9–1 выполните следующее.

   a) Проверьте правильность вычисленных в примере значений μ_2, μ_3 и μ_4.

   b) Предположим, что за задержку выполнения заказа штрафом в сумме 80 долларов за работу в день облагаются лишь те работы, которые не выполняются к концу дня. Для какой модели копировального аппарата общая стоимость в день будет самой низкой?

2. Механическая мастерская нанимает механика для обслуживания парка из 10 станков. Рассматриваются две кандидатуры. Первый кандидат может выполнять ремонтные работы с интенсивностью 5 станков в час, и ему придется платить 15 долларов в час. Второй претендент более квалифицирован, он может ремонтировать 8 станков в час, и ему придется платить 20 долларов в час. Руководство мастерской считает, что каждый час вышедшего со строя станка приносит мастерской убыток в 50 долларов. Предполагая, что станки выходят со строя в соответствии с распределением Пуассона со средним 3 станка в час и что время ремонта имеет экспоненциальное распределение, определите, которого из претендентов следует нанять мастерской.

3. Компания, владеющая бакалейными магазинами, открывает новый магазин, где в кассовом блоке будут установлены сканеры. Мистер Браун, один из владельцев компании, ограничил выбор сканеров двумя типами: сканер A может обработать 10 изделий в минуту, а сканер более высокого качества B — 15 изделий в минуту. Дневная (10 часов) стоимость эксплуатации и технического обслуживания сканеров моделей A и B равна 25 и 35 долларов соответственно. Покупатели, выбрав покупки, подходят к каждой из касс в соответствии с распределением Пуассона с интенсивностью 10 покупателей в час. Тележка каждого из покупателей содержит от 25 до 35 предметов (равномерное распределение). Мистер Браун считает, что средняя стоимость минуты ожидания клиента равна примерно 10 центам. Сканеры какого типа следует установить в кассовом блоке? (*Подсказка.* Время обслуживания одного покупателя подчиняется не экспоненциальному закону, а равномерному.)

4. Промышленная компания производит станки специального типа, производительность которых можно изменять в соответствии со спецификацией заказчиков. Компания увеличивает цену станка на 50 долларов за наращивание его производительности на одну единицу. Хозяин одной из мастерских планирует купить один из таких станков и решает вопрос об определении оптимальной (с экономической точки зрения) производительности станка. Опираясь на предыдущий опыт, хозяин мастерской считает, что заказы от клиентов поступают в соответствии с распределением Пуассона с интенсивностью три заказа в час. Объем каждого заказа в среднем составляет 500 изделий. Договорами между хозяином и клиентами предусматривается штраф в сумме 100 долларов за задержку выполнения заказа на один час.

   a) Предполагая, что время выполнения заказа распределено по экспоненциальному закону, постройте общую стоимостную модель как функцию производительности станка μ.

   b) Исходя из стоимостной модели, построенной в предыдущем пункте, найдите формулу для оптимального значения производительности станка.

c) Исходя из данных задачи, определите оптимальное значение производительности станка, которое должен указать хозяин мастерской промышленной компании при оформлении заказа.

5. Работы поступают в механическую мастерскую в соответствии с распределением Пуассона с интенсивностью 80 заказов в неделю. Автоматический станок представляет собой узкое место в работе мастерской. Считается, что увеличение мощности этого станка на одну единицу обойдется в 250 долларов в неделю. Не вовремя выполненные работы приносят убытки в 500 долларов на одну работу в неделю. Основываясь на приведенной информации, определите соответствующую оптимальную производительность автоматического станка.

6. Фирма продает рестораны двух типов. Ресторан типа A рассчитан на 80 групп посетителей, а ресторан типа B — на 100 групп (группой считаются посетители, занимающие один столик). Месячные расходы, связанные с обслуживанием клиентов в ресторанах типа A и B, равны 12 000 и 16 000 долларов соответственно. Потенциальный покупатель планирует открыть ресторан, специализирующийся на приготовлении пиццы. Он оценивает, что группы посетителей прибывают в соответствии с распределением Пуассона с интенсивностью 25 групп в час. Если все столики заняты, то дополнительно прибывающие посетители уходят, не дожидаясь обслуживания. Ресторан типа A позволяет обслужить 10 групп посетителей в час, а ресторан типа B — 13 групп в час. Так как группы посетителей различаются как по количественному составу, так и по типу заказов, то время их обслуживания распределено по экспоненциальному закону. Покупатель считает, что средние потери, обусловленные не обслуженными посетителями, составляют в расчете на одного посетителя 3 доллара в час. Задержка в обслуживании клиентов, которые уже разместились в ресторане, обходится в 0.50 доллара в расчете на одного посетителя в час.

   a) Постройте общую стоимостную модель, которая учитывает потери, обусловленные не обслуженными посетителями (в дополнение к расходам, связанным с обслуживанием клиентов в ресторане и задержкой в обслуживании клиентов).

   b) Предположим, что ресторан будет работать 10 часов в сутки, какой тип ресторана вы бы рекомендовали потенциальному покупателю?

7. Пусть в условиях предыдущего упражнения покупатель может выбрать любую желаемую вместимость ресторана на основе определенной граничной стоимости за каждую затребованную дополнительную единицу вместимости ресторана. Постройте соответствующую этому случаю общую стоимостную модель, определите все ее элементы и характеристики.

8. Фирма занимается комиссионной торговлей товарами, которые она получает от своих клиентов на основе поручительства. Функционирование фирмы можно рассматривать как задачу управления запасами, в которой товары поступают на склад и извлекаются из него случайным образом в соответствии с распределением Пуассона с интенсивностями λ и μ соответственно. Каждый раз, когда некоторое изделие отсутствует на складе, фирма теряет C_1 долларов по причине нереализованной возможности, а если изделие находится на складе, то за хранение необходимо платить C_2 долларов.

a) Получите формулу для ожидаемой общей стоимости в единицу времени.

b) Найдите оптимальное значение параметра $\rho = \lambda/\mu$. Какое условие должно быть наложено на величины C_1 и C_2 для того, чтобы полученный результат был совместимым с предположениями модели $(M/M/1) : (\mathrm{GD}/\infty/\infty)$?

Пример 17.9–2

На инструментальный склад, где работают несколько служащих, поступают заявки на замену режущего инструмента в соответствии с распределением Пуассона со средним значением 17.5 заявки в час. Каждый служащий может выполнить в среднем 10 заявок в час. Стоимость найма нового служащего на склад составляет 12 долларов в час. Потери, связанные с ожиданием станка, оцениваются примерно в 50 долларов в час. Необходимо определить число служащих для рассматриваемой системы обслуживания.

Описанная ситуация соответствует модели $(M/M/c)$, в которой требуется определить оптимальное значение c. Следовательно, $x = c$, и соответствующая стоимостная модель имеет вид

$$\mathrm{COC}(c) = C_1 c + C_2 L_s(c) = 12c + 50L_s(c).$$

Заметим, что $L_s(c)$ является функцией количества служащих в мастерской.

Используем модель $(M/M/c) : (\mathrm{GD}/\infty/\infty)$ с $\lambda = 17.5$ заявок в час и $\mu = 10$ заявок в час. В этом отношении модель достигнет устойчивого состояния лишь при условии, что $c > \lambda/\mu$, т.е. в рассматриваемом примере c должно быть по крайней мере равно 2. Приведенная ниже таблица содержит результаты вычислений для определения оптимального значения c. Значения $L_s(c)$ определены с помощью программы TORA.

c	$L_s(c)$ (заявки)	COC(c) (\$)
2	7.467	397.35
3	2.217	142.35
4	**1.842**	**140.10**
5	1.769	148.45
6	1.754	159.70

Следовательно, оптимальным числом сотрудников для мастерской является 4.

Упражнения 17.9,b

1. Решите задачу из примера 17.9–2 в предположении, что $C_1 = \$20$, $C_2 = \$45$.
2. Компания владеет насосной станцией трубопровода, агрегаты которой работают в непрерывном режиме. Время между последовательными выходами из строя каждого из агрегатов экспоненциально распределено со средним значением 20 часов. Время ремонта каждого агрегата также имеет экспоненциальное распределение со средним значением 10 часов. На станции имеются 10 агрегатов, которые обслуживаются двумя механиками по ремонту. Заработная плата каждого из них составляет 18 долларов в час. Потери, связанные с выходом из строя одного агрегата, оцениваются в 30 долларов в час. Компания изучает возможность принятия на работу еще одного механика по ремонту агрегатов.

a) Будет ли какая-либо экономия в результате принятия на работу еще одного механика по ремонту агрегатов?

b) Чему равны потери, связанные с выходом из строя одного агрегата, когда работают два механика? А если работают три механика?

3. Компания арендует для служебных целей автоматическую телефонную линию по цене 1500 долларов в месяц. Сотрудники аппарата компании используют телефонную линию в рабочее время, что в сумме составляет 200 часов в месяц. Остальное время телефонная линия используется для других целей и для компании является недоступной. В течение рабочего дня арендуемой телефонной линией пользуются 100 сотрудников компании, каждому из которых она может понадобиться в любое время, но в среднем 2 раза на протяжении восьмичасового рабочего дня с экспоненциальным распределением времени между звонками. Служащий компании во всех случаях ждет освобождения телефонной линии, если она занята, при этом испытываемое им неудобство оценивается в 1 цент за минуту ожидания. Предполагается, что во время ожидания служащим освобождения линии не возникает потребности в других звонках. Обычная стоимость минуты разговора (без использования арендованной телефонной линии) составляет в среднем 50 центов, длительность телефонных разговоров имеет экспоненциальное распределение со средним значением 6 минут. Компания считает, что используемая линия телефонной связи перегружена звонками и рассматривает вопрос об аренде (по той же цене) второй автоматической телефонной линии.

   a) Приносит ли аренда одной автоматической телефонной линии экономическую выгоду компании по сравнению с ситуацией, когда телефонные линии вообще не арендуются? Какую сумму компания выигрывает или проигрывает в месяц, арендуя одну автоматическую телефонную линию?

   b) Следует ли компании арендовать вторую автоматическую телефонную линию? Какую сумму компания выиграет или проиграет в месяц, арендуя вторую автоматическую телефонную линию, по сравнению со случаем аренды лишь одной линии?

4. Механический цех насчитывает 20 станков, которые обслуживаются тремя механиками. Неполадки в работающих станках возникают случайным образом в соответствии с распределением Пуассона. Время устранения неполадки на одном станке подчиняется экспоненциальному распределению с математическим ожиданием 6 минут. Анализ рассматриваемой системы обслуживания показывает, что во всем цехе на протяжении восьмичасового рабочего дня возникает в среднем 57.8 неполадок в работе станочного парка и что к вышедшему из строя станку механик подходит в среднем через 4.5 минуты. Пусть интенсивность работы каждого станка составляет 20 единиц продукции в час, каждая из которых приносит прибыль в 2 доллара. Далее, пусть зарплата каждого механика равняется 20 долларов в час. Сопоставьте плату за наем механиков с потерями, обусловленными потерей прибыли в случае выхода из строя станков.

5. Необходимым условием того, что величина $\mathrm{COC}(c)$ (определение которой приведено выше) достигает своего минимума при $c = c^*$, являются неравенства

$$\mathrm{COC}(c^* - 1) \geq \mathrm{COC}(c^*) \text{ и } \mathrm{COC}(c^* + 1) \geq \mathrm{COC}(c^*).$$

Покажите, что эти неравенства сводятся к следующим:

$$L_s\left(c^*\right)-L_s\left(c^*+1\right)\leq\frac{C_1}{C_2}\leq L_s\left(c^*-1\right)-L_s\left(c^*\right).$$

6. Примените результат предыдущего упражнения к примеру 17.9–2 и покажите, что это приводит к значению $c^* = 4$.

17.9.2. Модель предпочтительного уровня обслуживания

Жизнеспособность модели обслуживающей системы со стоимостными характеристиками зависит от того, насколько хорошо мы можем оценить параметры стоимости. В общем случае оценить эти параметры довольно сложно, особенно, если стоимость связана с ожиданием клиента. В моделях с предпочтительным уровнем обслуживания делается попытка обойти эту проблему, оперируя непосредственно функциональными показателями обслуживающей системы. Идея состоит в определении приемлемого интервала изменения для уровня обслуживания (параметры μ или c) путем нахождения разумных пределов для *конкурирующих* экономических показателей, которые характеризуют процесс обслуживания. Эти пределы представляют собой *уровни предпочтительного обслуживания*, которых стремится достичь лицо, принимающее управленческое решение.

Проиллюстрируем применение этой процедуры для модели системы обслуживания с несколькими сервисами, в которой необходимо определить "приемлемое" количество сервисов c^*. Для этого рассмотрим два (конкурирующих) экономических показателя процесса обслуживания.

1. Среднее время ожидания в системе W_s.
2. Процент простоя сервисов X.

Значение W_s можно вычислить с использованием программного обеспечения TORA для модели $(M/M/c)$. Процент простоя средств обслуживания можно вычислить следующим образом.

$$X=\frac{c-\bar{c}}{c}\times100=\frac{c-\left(L_s-L_q\right)}{c}\times100=\left(1-\frac{\lambda_{\text{эфф}}}{c\mu}\right)\times100.$$

(Доказательство этого соотношения см. в упр. 17.6,е(13).)

Задача сводится к определению такого количества сервисов c^*, что

$$W_s\leq\alpha \text{ и } X\leq\beta,$$

где α и β — уровни предпочтительного обслуживания, определенные лицом, которое принимает решение. Например, можно поставить условие, что $\alpha = 3$ минуты и $\beta = 10\%$.

Решение задачи можно найти, построив графики W_s и X как функции количества сервисов c, как показано на рис. 17.11. Отмечая на графиках значения α и β, мы определяем приемлемый интервал изменения для уровня обслуживания c^*. Если оба упомянутых выше условия нельзя удовлетворить одновременно, необходимо ослабить один или оба уровня предпочтительности, пока не будет получен приемлемый интервал изменения количества сервисов.

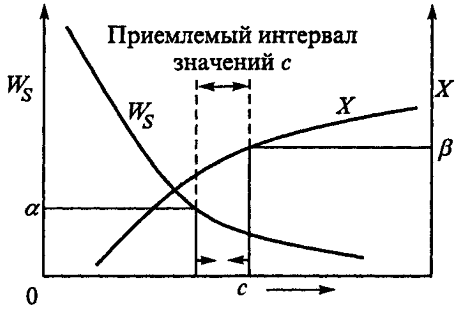


Рис. 17.11

Пример 17.9–3

Допустим, что в примере 17.9–2 необходимо определить такое количество служащих на складе, чтобы среднее время ожидания получения инструмента было меньше 5 минут. Одновременно требуется, чтобы процент времени, в течение которого персонал мастерской остается свободным, не превышал 20%.

Заметим, что и без вычислений очевидным является тот факт, что уровень предпочтительности в 5 минут для времени ожидания до получения инструмента (т.е. здесь $W_s \le 5$) является недостижимым, так как из данных задачи следует, что среднее время обслуживания равно 6 минут. Приведенная ниже таблица содержит значения W_s и X как функций от c и подтверждает это замечание. Действительно, при $c \ge 5$ имеем $W_s = 6$ минут. Это означает, что время ожидания в очереди будет отсутствовать. Таблица также показывает, что дальнейшее увеличение числа служащих может лишь увеличить процент их простоя X.

c	2	3	4	5	6	7	8
W_s (минуты)	25.6	7.6	6.3	6.0	6.0	6.0	6.0
X (%)	12.5	41.7	56.3	65.0	70.8	75.0	78.0

В этой ситуации мы фактически ничего не можем сделать, так как задачу нельзя решить путем увеличения количества служащих на складе. Необходимо либо уменьшить время обслуживания в системе, либо признать, что в реальной ситуации, которой соответствует математическая модель, интенсивность запросов на инструменты чрезвычайно высока ($\lambda = 17.5$ запросов в час). Весьма вероятно, что именно на последний факт должно быть обращено внимание. Например, можно провести исследование причины такого высокого спроса на замену инструмента. Может ли быть причиной этого неадекватная конструкция самого инструмента? Или же причиной является то, что операторы станков целенаправленно стараются подорвать производство, таким образом выражая свое недовольство?

Упражнения 17.9,с

1. Цех использует 10 одинаковых станков. Каждый станок выходит из строя в среднем один раз в 7 часов. Ремонт сломанного станка длится в среднем 4 часа. Как процесс выхода станков из строя, так и процесс ремонта подчиняются распределению Пуассона. Определите следующие показатели.

   a) Необходимое число механиков для ремонта станков, при котором среднее количество неработающих станков будет меньше 4.

   b) Такое число механиков для ремонта станков, чтобы ожидаемое время задержки, обусловленное ремонтом станка, было меньше четырех часов.

2. В стоимостной модели системы обслуживания, рассмотренной в разделе 17.9.1, в общем случае трудно оценить стоимостный параметр C_2 (стоимость ожидания). Может быть полезным вычисление стоимости ожидания C_2, предполагаемое уровнями предпочтительности обслуживания. В этом случае считаем, что значение C_1 известно. Используя модель предпочтительного уровня обслуживания для определения c^*, можно оценить предполагаемое значение C_2 с помощью следующего неравенства:

$$L_s\left(c^*\right)-L_s\left(c^*+1\right)\le\frac{C_1}{C_2}\le L_s\left(c^*-1\right)-L_s\left(c^*\right).$$

   (Вывод этой формулы см. в упр. 17.9,b(5).) Примените эту процедуру для задачи из примера 17.9.2, предполагая, что $c^* = 3$ и $C_1 = \$12$.

17.10. Заключение

Теория массового обслуживания позволяет строить модели для анализа функционирования обслуживающих систем, когда входной и выходной потоки клиентов являются случайными. Важную роль в теории массового обслуживания играют пуассоновское и экспоненциальное распределения вероятностей. Разумеется, модели систем обслуживания могут анализироваться также с использованием других вероятностных распределений, однако возникающие при этом математические задачи обычно неразрешимы в аналитическом виде.

Анализ систем обслуживания, по существу, не решает самих задач. Он лишь позволяет определить функциональные показатели обслуживающей системы, которые могут затем использоваться в рамках некоторых моделей принятия решений.

Литература

1. Hall R. *Queueing Methods for Service and Manufacturing,* Prentice Hall, Upper Saddle River, N.J., 1991.
2. Lipsky L. *Queueing Theory, A Linear Algebraic Approach*, Macmillan, New York, 1992.
3. Morse P. *Queues, Invéntories, and Maintenance*, Wiley, New York, 1958.
4. Parzen E. *Stochastic Processes*, Holden–Day, San Francisco, 1962.
5. Saaty T. *Elements of Queueing Theory with Applications*, Dover, New York, 1983.
6. Tijms H.C. *Stochastic Models — An Algorithmic Approach*, Wiley, New York, 1994.

Литература, добавленная при переводе

Боровков А.А. *Вероятностные процессы в теории массового обслуживания.* — М.: Наука, 1972.

Гихман И.И., Скороход А.В. *Введение в теорию случайных процессов.* — М.: Наука, 1965.

Кофман А. *Методы и модели исследования операций.* — М.: Мир, 1966.

Саати Т. *Элементы теории массового обслуживания и ее приложения.* — М.: Сов. радио, 1965.

Комплексные задачи

■ **17–1.**[1] Банк в своей деятельности использует традиционную линию, где клиенты обслуживаются, не выходя из автомашины, и две "автоматические" линии, которые соединены с внутренней частью банка пневматическим контейнером. Банк хотел бы расширить существующие средства обслуживания, чтобы прибывающий на автомашине клиент смог завершить свое дело в банке в среднем не более чем за 4 минуты. Это ограничение по времени было основано на психологических исследованиях, которые показывают, что нетерпение клиентов проявляется на основе движения минутной стрелки часов между двумя метками, что для большинства часов, установленных в банках, соответствует пяти минутам. Для сбора необходимой информации группа специалистов по исследованию операций изучила действия кассиров, которые работают в банке. После некоторого ознакомления с работой системы, один из членов группы отметил заметное отличие времени, которое тратит клиент в линии обслуживания, и времени, когда кассир выполняет необходимые банковские операции. Действительно, время, которое автомобиль проводит в системе, состоит из (1) осознания, что впереди стоящий автомобиль уехал, (2) движения к окну кассира, (3) передачи кассиру необходимых указаний, (4) необходимых действий кассира и (5) выхода из системы. На протяжении первой, второй и пятой составляющих этого времени кассир является вынужденно свободным. Действительно, обследование показало, что на протяжении цикла обслуживания одного клиента кассир занят им лишь 40% времени этого цикла. На основе этой информации группа сделала вывод о наличии возможности уменьшения стоимости обслуживания в существующей системе.

В чем состояло предложение специалистов по исследованию операций по улучшению обслуживания в банке с использованием линии, где клиенты обслуживаются, не выходя из автомашины? Обсудите все последствия предложения.

■ **17–2.** Государственный центр по защите детей работает ежедневно с 9 часов утра до 9 часов вечера. Звонки, сигнализирующие о необходимости вмешательства для защиты детей, поступают в центр, как и следовало ожидать, случайным образом. Приведенная ниже таблица содержит количество звонков на протяжении 7 дней, разбитых на интервалы длиной в 1 час.

---

[1] Задача базируется на работе Foote B.L. "A Queueing Case Study in Drive-In Banking", *Interfaces*, Vol.6, No. 4, pp. 31–37. 1976.

Время	Количество звонков в день						
	1	2	3	4	5	6	7
9:00	4	6	8	4	5	3	4
10:00	6	5	5	3	6	4	7
11:00	3	9	6	8	4	7	5
12:00	8	11	10	5	15	12	9
13:00	10	9	8	7	10	16	6
14:00	8	6	10	12	12	11	10
15:00	10	9	12	4	10	6	8
16:00	8	6	9	14	12	10	7
17:00	5	10	10	8	10	10	9
18:00	5	4	6	5	6	7	5
19:00	3	4	6	2	3	4	5
20:00	4	3	6	2	2	3	4
21:00	1	2	2	3	3	5	3

Таблица не содержит звонков, потерянных по той причине, что во время звонка линия оказалась занятой. Длительность каждого из полученных звонков является случайной величиной со средним, равным 7 минутам, но иногда достигает и 12 минут. Данные свидетельствуют о том, что количество звонков в центр возрастает на 15% в год. Центр планирует определить количество телефонных линий, которые необходимо установить для надлежащего обслуживания в настоящем и в будущем. В частности, особое внимание уделяется уменьшению неблагоприятных результатов, когда звонки не поступают из-за занятости телефонной линии.

■ **17–3.** Промышленная компания использует три грузовика для перевозки материалов шести своим филиалам. Клиенты требуют выделения для этих целей четвертого грузовика для уменьшения проблемы, связанной с чрезмерными задержками поставок. Грузовики не имеют постоянного гаража, откуда их можно было бы вызывать. Администрация считает более эффективной систему, когда грузовики находятся в непрерывном движении. Представитель филиала, которому необходим грузовик, должен дождаться его прибытия поблизости завода. Если грузовик свободен, он ответит на звонок. Иначе представитель филиала должен ожидать следующего грузовика. Приведенная таблица содержит частоты появления соответствующего количества звонков на протяжении часа.

Количество звонков в час	Частота
0	30
1	90
2	99
3	102
4	120
5	100
6	60
7	47
8	30
9	20

Количество звонков в час	Частота
10	12
11	10
12	4

Время обслуживания для каждого филиала (в минутах) примерно одинаковое. Приведенная ниже таблица содержит типичную гистограмму времени обслуживания для одного из филиалов.

Время обслуживания, t	Частота
$0 \le t < 10$	61
$10 \le t < 20$	34
$20 \le t < 30$	15
$30 \le t < 40$	5
$40 \le t < 50$	8
$50 \le t < 60$	4
$60 \le t < 70$	4
$70 \le t < 80$	3
$80 \le t < 90$	2
$90 \le t < 100$	2

Выработайте рекомендации для руководства компании.

■ **17–4.** Джон, молодой инженер, недавно принят на работу в машиностроительную компанию. Компания владеет цехом, насчитывающим 30 станков, для поддержания которых в рабочем состоянии содержит штат из шести механиков. Цех работает в одну смену с 8:00 до 16:00. Джон получил первое задание: изучить эффективность ремонтной службы цеха. С этой целью Джон собрал следующие данные из протокола регистрации ремонта трех станков, выбранных случайным образом.

Станок 5		Станок 18		Станок 23	
Время поломки	Время ремонта	Время поломки	Время ремонта	Время поломки	Время ремонта
8:05	8:15	8:01	8:09	8:45	8:58
10:02	10:14	9:10	9:18	9:55	10:06
10:59	11:09	11:03	11:16	10:58	11:08

Станок 5		Станок 18		Станок 23	
Время поломки	Время ремонта	Время поломки	Время ремонта	Время поломки	Время ремонта
12:22	12:35	12:58	13:06	12:21	12:32
14:12	14:22	13:49	13:58	12:59	13:07
15:09	15:21	14:30	14:43	14:32	14:43
15:33	15:42	14:57	15:09	15:09	15:17
15:48	15:59	15:32	15:42	15:50	16:00

В дополнение к этой информации, путем отбора записей протокола регистрации ремонта для пяти случайно выбранных дней, Джон собрал данные о количестве сломанных станков (включая и те, которые находились в ремонте) в начале каждого часа рабочего дня.

	Общее количество сломанных станков по состоянию на							
Дата	8:00	9:00	10:00	11:00	12:00	13:00	14:00	15:00
2/10	6	6	9	6	8	8	7	7
29/10	9	8	5	9	5	5	6	8
4/11	6	6	5	7	7	8	6	5
1/12	9	5	9	7	5	7	5	5
19/1	6	5	8	5	9	8	8	6

Джон встретился со своим руководителем Ребеккой и обсудил с ней собранную информацию. Джон заявил, что он уверен в том, что процесс выхода из строя и ремонта станков в цехе является совершенно случайным, и можно предположить, что эту ситуацию можно представить в виде системы массового обслуживания пуассоновского типа. Ребекка подтвердила, что ее многолетний опыт работы в цехе согласуется с предположением, что рассматриваемая ситуация является совершенно случайной. Основываясь на этом предположении, Ребекка проверила собранные Джоном данные и, сделав некоторые вычисления, сказала, что в данных содержатся ошибки. Как Ребекка пришла к такому выводу?

■ **17–5.** Таксомоторная компания владеет четырьмя машинами. Служба такси работает 10 часов каждый день. Заявки на обслуживание поступают в диспетчерскую службу в соответствии с распределением Пуассона в среднем 20 заявок в час. Известно, что длительность поездки такси является экспоненциально распределенной случайной величиной с математическим ожиданием 11.5 минут. В силу высокого уровня спроса на обслуживание компания ограничивает список клиентов в диспетчерской службе, ожидающих обслуживания, до 16 заявок. Когда этот предел достигается, клиентам советуют искать обслуживание в другом месте в силу того, что среднее время ожидания такси будет длительным.

Управляющий компании Стоун опасается, что он, возможно, теряет много клиентов и поэтому хочет рассмотреть вопрос об увеличении парка машин компании. Стоун считает, что средняя прибыль, получаемая компанией от обслуживания одного клиента, составляет примерно 5 долларов. Новый автомобиль по его оценке можно купить за 18 000 долларов. Новое такси используется на протяжении пяти лет и затем продается за 3 500 долларов. Стоимость содержания и обслуживания такси на протяжении одного года составляет 20 000 долларов. Может ли мистер Стоун обосновать увеличение парка автомобилей компании, и если да, то на сколько единиц? При анализе предположите 10% ежегодного прироста клиентов.

Имитационное моделирование

18.1. Что такое имитационное моделирование

Имитационное моделирование является мощным инструментом исследования поведения реальных систем. Методы имитационного моделирования позволяют собрать необходимую информацию о поведении системы путем создания ее компьютеризованной модели. Эта информация используется затем для проектирования системы. Имитационное моделирование не решает оптимизационных задач, а скорее представляет собой технику оценки значений функциональных характеристик моделируемой системы.

Имитационное моделирование применяется в различных областях науки, техники, экономики и т.д., о чем свидетельствует представленный ниже краткой список решаемых с помощью имитационного моделирования задач.

1. Задачи в различных областях естественных наук (математика, физика, химия):
   a) вычисление площадей фигур, ограниченных кривыми, или, в более общем случае, вычисление кратных интегралов;
   b) вычисление констант (например π, равной 3.14159...);
   c) обращение матриц;
   d) изучение диффузионных процессов.
2. Практические задачи:
   a) производственно-технологические задачи, возникающие в процессе создания систем массового обслуживания, систем связи, в сфере управления запасами, а также при анализе химических процессов;
   b) экономические и коммерческие задачи, включая оценки поведения потребителя, определение цен, экономическое прогнозирование деятельности фирм;
   c) социальные и социально-психометрические задачи, например проблемы динамики народонаселения, влияния экологии на здоровье, эпидемиологических исследований, а также прогнозирование группового поведения;
   d) задачи биомедицинских систем, например проблема баланса жидкости в организме человека, размножения клеток крови, деятельности мозга;
   e) задачи анализа той или иной военной стратегии и тактики.

Вычисление результатов имитации базируется на *случайной выборке* (подобным образом мы поступаем и при наблюдении реальной ситуации). Это означает, что любой результат, полученный путем имитационного моделирования, подвержен экспериментальным ошибкам и, следовательно, как в любом статистическом эксперименте, должен основываться на результатах соответствующих статистических проверок. Это важное замечание мы подчеркиваем на протяжении всей главы.

18.2. Метод Монте–Карло

Основная идея данного метода состоит в использовании выборки случайных чисел для получения искомых оценок. Этот метод в известном смысле является предшественником современного имитационного моделирования.

Для демонстрации метода Монте–Карло рассмотрим следующий пример. В этом примере особо подчеркнута статистическая природа имитационного эксперимента.

Пример 18.2–1

Используем метод Монте–Карло для оценки площади круга, уравнение окружности которого имеет вид

$$(x-1)^2+(y-2)^2=25.$$

Круг имеет радиус $r = 5$ см, и его центр находится в точке $(x, y) = (1, 2)$.

Процедура оценки площади требует заключения круга в описанный около него квадрат, сторона которого равна диаметру круга, как это показано на рис. 18.1. Вершины квадрата определяются непосредственно из геометрических свойств фигуры.

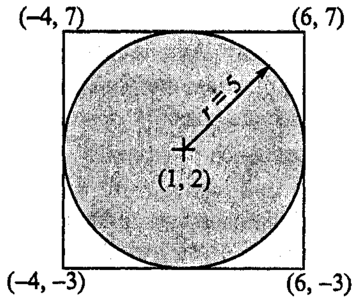


Рис. 18.1

Оценка площади круга основана на предположении, что все точки квадрата равновероятны. Предположим, что выборка состоит из наблюдений n точек квадрата и m из них попали внутрь круга или на окружность. Тогда

$$\text{площадь круга } \approx \frac{m}{n}(\text{площадь квадрата}) = \frac{m}{n}(10\times 10).$$

Здесь координаты x и y точек квадрата представлены как равномерно распределенные случайные величины с плотностями вероятностей

$$f(x)=\frac{1}{10}, \quad -4 \le x \le 6,$$

$$f(y)=\frac{1}{10}, \quad -3 \le y \le 7.$$

Обе функции равны нулю вне указанных интервалов.

Пусть R_1 и R_2 — различные случайные числа из интервала [0, 1]. Тогда координаты (x, y) точек квадрата можно выразить через эти случайные числа:

$$x = -4 + [6 - (-4)]R_1 = -4 + 10R_1,$$
$$y = -3 + [7 - (-3)]R_2 = -3 + 10R_2.$$

В табл. 18.1 приведен небольшой список случайных чисел из интервала [0, 1]. Эти числа сгенерированы с использованием специальных методов, которые описаны в разделе 18.5.

Таблица 18.1

0.0589	0.3529	0.5869	0.3455	0.7900	0.6307
0.6733	0.3646	0.1281	0.4871	0.7698	0.2346
0.4799	0.7676	0.2867	0.8111	0.2871	0.4220
0.9486	0.8931	0.8216	0.8912	0.9534	0.6991
0.6139	0.3919	0.8261	0.4291	0.1394	0.9745
0.5933	0.7876	0.3866	0.2302	0.9025	0.3428
0.9341	0.5199	0.7125	0.5954	0.1605	0.6037
0.1782	0.6358	0.2108	0.5423	0.3567	0.2569
0.3473	0.7472	0.3575	0.4208	0.3070	0.0546
0.5644	0.8954	0.2926	0.6975	0.5513	0.0305

Используя приведенные формулы, мы можем сгенерировать равномерно распределенную случайную точку (x, y) квадрата для каждой пары случайных чисел (R_1,R_2). Сгенерированная точка (x', y') попадает внутрь круга, если

$$(x'-1)^2 + (y'-2)^2 \le 25.$$

Например, если $R_1 = 0.0589$ и $R_2 = 0.6733$, то

$$x' = -4 + 10R_1 = -4 + 10 \times 0.0589 = -3.411,$$
$$y' = -3 + 10R_2 = -3 + 10 \times 0.6733 = 3.733.$$

Так как величина $(-3.411 - 1)^2 + (3.733 - 2)^2 = 22.46$ меньше 25, следовательно, точка (x', y') попадает внутрь круга.

Исследуем теперь влияние случайной выборки на точность оценки площади круга. На рис. 18.2 показаны оценки площади круга для различных объемов выборки, изменяющихся от $n = 100$ до $n = 10\,000$. При каждом n эксперимент повторялся 10 раз, при этом использовались различные последовательности случайных чисел из интервала [0, 1]. На рисунке показаны оценки площади лишь для первого и второго экспериментов (прогонов) вместе со средними и стандартными отклонениями для всех 10 прогонов.

Исходя из результатов эксперимента, можно сделать следующие выводы.

1. Оценка площади круга улучшается с увеличением числа генерируемых точек (с увеличением объема выборки), о чем свидетельствуют результаты прогонов 1 и 2 на рис. 18.2.

2. Усреднение результатов 10 прогонов для каждой выборки объемом n дает более точную оценку площади, чем любой из прогонов. На рис. 18.2 видно, что среднее 10 экспериментов ближе к точному значению площади (= 78.54 см^2), чем оценки, полученные в каждом отдельном прогоне.
3. Уменьшение величины стандартного отклонения свидетельствует о том, что "точность" среднего 10 экспериментов повышается с увеличением объема выборки n.

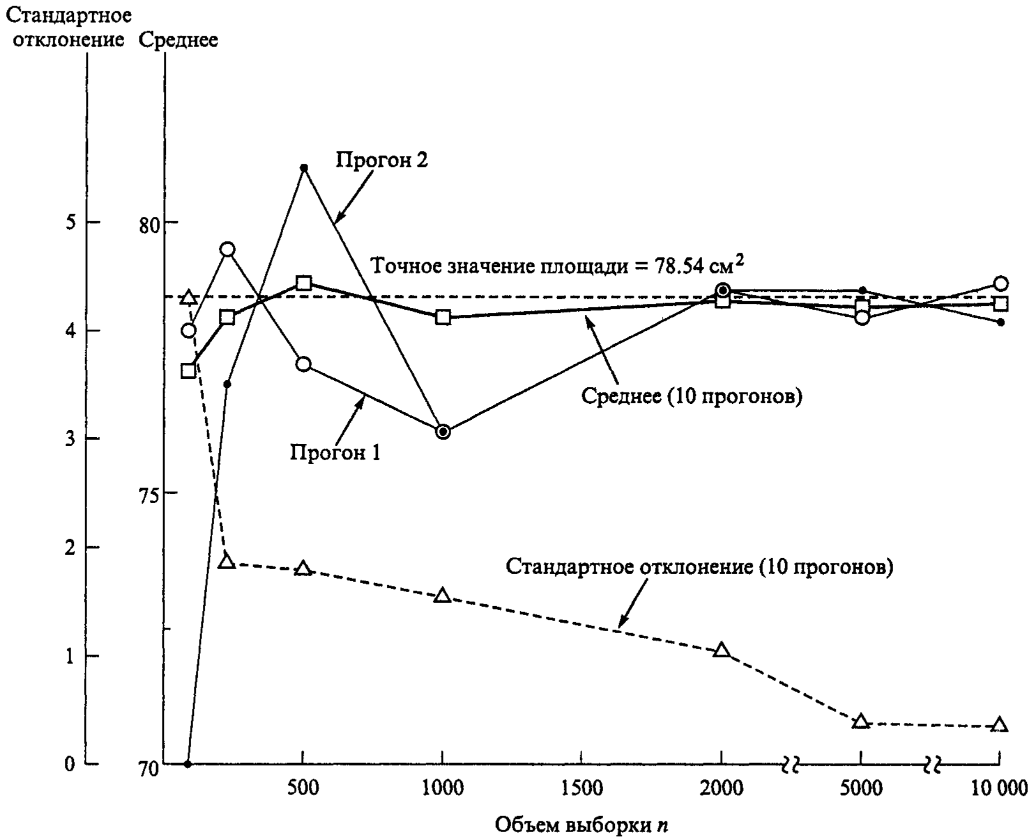


Рис. 18.2

Ввиду того, что оценки площади имеют разброс, важно, чтобы результаты эксперимента, связанного с моделированием, были выражены в виде доверительных интервалов, показывающих величину отклонения от точного значения. В рассматриваемом примере, если A представляет собой точное значение площади, а $\bar{A}$ и s^2 — среднее и дисперсию N наблюдений, то $100(1-\alpha)\%$-й доверительный интервал для A задается в виде

$$\bar{A} - \frac{s}{\sqrt{N}} t_{\alpha/2,N-1} \le A \le \bar{A} + \frac{s}{\sqrt{N}} t_{\alpha/2,N-1},$$

где $t_{\alpha/2,N-1}$ — $(100\alpha/2)$%-я точка t-распределения (распределения Стьюдента) с $N-1$ степенями свободы (см. Приложение Д). Заметим, что N обозначает число наблюдений (прогонов), и его следует отличать от n, которое обозначает объем выборки (продолжительность прогона модели). В рассматриваемом примере мы заинтересованы в установлении доверительного интервала, полученного для выборки наибольшего объема (т.е. $n = 10\,000$). При $N = 10$, $\bar{A} = 78{,}57$ см^2 и $s = 0.47$ см^2, результирующим 95%-м доверительным интервалом является $78.23 \le A \le 78.9$.

Рассмотренный пример ставит два вопроса, характерные для любого эксперимента, связанного с моделированием.

1. Каким должен быть объем выборки n для достижения необходимого значения доверительных интервалов?
2. Сколько для этого требуется прогонов N?

Ответы зависят от природы эксперимента, связанного с моделированием. Как и в любом статистическом эксперименте, большие значения n и N обеспечивают более надежные результаты. Препятствием может быть стоимость проведения эксперимента, которая возрастает пропорционально увеличению n и N.[1]

Упражнения 18.2,а

1. В условиях примера 18.2–1 выполните следующее.

   a) Вычислите площадь круга, используя из табл. 18.1 два первых столбца в качестве источника случайных чисел из интервала [0, 1]. (Для удобства выбирайте R_1 из первого столбца, а R_2 — из второго, двигаясь при этом сверху вниз.)

   b) Пусть уравнение окружности имеет вид

   $$(x-3)^2 + (y+2)^2 = 16.$$

   Найдите соответствующие плотности вероятностей $f(x)$ и $f(y)$. Покажите, как с помощью пары случайных чисел (R_1, R_2) из интервала [0, 1] можно получить выборочную точку (x, y).

2. Примените метод Монте–Карло для вычисления площади озера, показанного на рис. 18.3. Для получения значений случайных чисел из интервала [0, 1] используйте два первых столбца табл. 18.1.

3. Рассмотрим игру, в которой два игрока А и Б поочередно подбрасывают правильную (симметричную) монету. Если выпадает лицевая сторона монеты, игрок А получает 10 долларов от игрока Б, иначе игрок Б выигрывает у игрока А 10 долларов.

---

[1] Определение достаточных значений объема выборки n и количества прогонов N в общем случае является нетривиальной задачей. Многое определяется выбранной методикой построения доверительных интервалов. Например, доверительные интервалы данного примера построены в предположении, что выборочная оценка исследуемого параметра (здесь — площадь круга) имеет приближенно нормальное распределение. Для этого в данном случае достаточно, чтобы объем выборки был не менее 100. Вместе с тем существуют исследования, показывающие, что если распределение выборочной оценки одновершинное, то уже при объеме выборки, превышающем 20, можно построить 95%-й доверительный интервал по "правилу $3S$" (аналог "правила 3σ"). — *Прим ред.*

a) Как смоделировать эту игру в виде эксперимента Монте–Карло?

b) Проведите эксперимент в 5 прогонов с 10 подбрасываниями монеты в каждом прогоне. Используйте пять первых столбцов табл. 18.1 для получения значений случайных чисел из интервала [0, 1], при этом каждый столбец будет соответствовать одному прогону.

c) Вычислите 95%-й доверительный интервал для выигрышей игрока А.

d) Сравните доверительный интервал, полученный в предыдущем пункте, с ожидаемым теоретическим выигрышем игрока А.

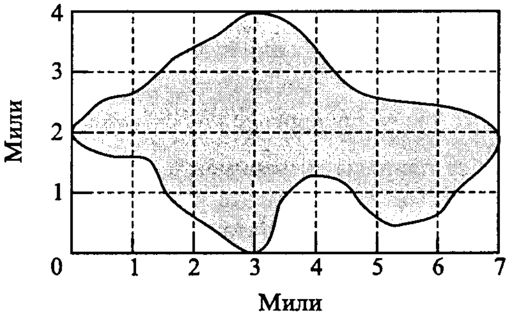


Рис. 18.3

4. Рассмотрим определенный интеграл $\int_0^1 x^2 dx$.

a) Оцените значение этого интеграла с помощью метода Монте–Карло.

b) Используйте четыре первых столбца табл. 18.1 для получения оценки значения интеграла, основанной на 4 прогонах объемом 10 каждый. Вычислите 95%-й доверительный интервал и сравните его с точным значением интеграла.

5. Смоделируйте ситуацию с пятью выигрышами или проигрышами в следующей игре в кости. Игрок бросает две симметричные игральные кости. Если выпавшая сумма равна 7 или 11, игрок выигрывает 10 долларов. Иначе он запоминает выпавшую сумму (называемую *очком*) и продолжает бросать кости до тех пор, пока выпавшая сумма не совпадет с очком, после чего игрок получает 10 долларов. Но если выпавшая сумма равна 7, игрок проигрывает 10 долларов.

6. Цикл исполнения заказа на некую продукцию с равной вероятностью составляет 1 или 2 дня. Предполагается, что *ежедневный* спрос равен 0, 1 и 2 единицы этой продукции с вероятностями 0.2, 0.5 и 0.3 соответственно. Используйте значения случайных чисел из табл. 18.1 (начиная с первого столбца) для оценки совместного распределения спроса и цикла исполнения заказа. Исходя из полученного совместного распределения, оцените плотность вероятности спроса в течение цикла исполнения заказа. (*Подсказка.* Спрос во время исполнения заказа может принимать значения 0, 1, 2, 3 и 4.)

7. Рассмотрим эксперимент Бюффона с иглой. Горизонтальная плоскость разделена параллельными прямыми, расположенными на расстоянии D см одна от другой. Игла длиной d см ($d < D$) случайным образом бросается на плоскость. Необходимо найти вероятность того, что игла коснется или пересечет одну из прямых. Введем следующие обозначения.

h — расстояние от центра иглы до ближайшей прямой,

θ — угол, составленный иглой с этой прямой.

a) Покажите, что игла коснется или пересечет прямую, если

$$h \le \frac{d}{2}\sin\theta,\ \ 0 \le h \le \frac{D}{2},\ \ 0 \le \theta \le \pi.$$

b) Составьте план эксперимента Монте–Карло, обеспечивающий нахождение оценки искомой вероятности.

c) Используйте первые четыре столбца табл. 18.1 для нахождения оценки искомой вероятности, основанной на 4 прогонах объемом 10 каждый. Вычислите 95%-й доверительный интервал для оценки.

d) Покажите, что вероятность интересующего нас события равна

$$p = \frac{2d}{\pi D}.$$

Используйте эту формулу вместе с результатом, полученным при решении задачи из п. c), для оценки значения числа π.

18.3. Типы имитационных моделей

Использование современных имитационных моделей основано, в основном, на идее метода Монте–Карло. Отличие состоит в том, что имитационная модель обычно связана с изучением реально существующей системы, поведение которой является функцией времени. Существует два типа имитационных моделей.

1. Непрерывные модели используются для систем, поведение которых изменяется *непрерывно* во времени. Типичным примером непрерывной имитационной модели является изучение динамики народонаселения мира. Непрерывные имитационные модели обычно представляются в виде разностно-дифференциальных уравнений, которые описывают взаимодействие между различными элементами системы.

2. Дискретные модели имеют дело с системами, поведение которых изменяется лишь в заданные моменты времени. Типичным примером такой модели является очередь, когда задача моделирования состоит в оценивании операционных характеристик обслуживающей системы, таких, например, как среднее время ожидания или средняя длина очереди. Такие характеристики системы массового обслуживания изменяют свои значения либо в момент появления клиента, либо при завершении обслуживания. В других случаях в системе ничего существенного (с точки зрения имитационного моделирования) не происходит.

Те моменты времени, в которые в системе происходят изменения, определяют **события** модели (например, приход или уход клиента). То, что эти события происходят в дискретные моменты, указывает, что процесс протекает в дискретном времени, откуда и появилось название **дискретное моделирование**.

Несмотря на то что как непрерывные, так и дискретные имитационные модели являются важным инструментом решения практических задач, в исследовании операций всегда используется дискретный тип имитационных моделей. Это объясняется, в частности, тем, что дискретная имитационная модель очень тесно связана с моделями массового обслуживания (см. главу 17). На самом деле есть все основания считать, что практически все дискретные имитационные модели в той или иной форме могут быть описаны, как модели массового обслуживания.

В настоящей главе основное внимание уделено обсуждению основ дискретного моделирования. Начнем с описания событий и того, как они могут быть сгенерированы в имитационной модели. Далее рассмотрим процедуры сбора статистических данных на основе имитационной модели и обсудим статистические аспекты имитационного эксперимента. Мы также подчеркнем важную роль компьютера и языков имитационного моделирования при реализации имитационных моделей.

Упражнения 18.3,а

1. Распределите по категориям приведенные ниже ситуации с точки зрения их принадлежности к дискретному или непрерывному типу (они также могут быть комбинацией обоих типов). В каждом случае укажите цель создания имитационной модели.
   a) Заказы на деталь поступают на склад случайным образом. Заказ, который не может быть выполнен сразу из наличного запаса, должен ожидать прибытия новых поставок.
   b) На численность населения земного шара влияет наличие полезных ископаемых, производство пищи, экологические условия, уровень образования, здравоохранения и капитальные вложения.
   c) Товары прибывают на приемную платформу автоматизированного склада на поддонах. Поддоны погружаются на нижний ленточный конвейер и поднимаются лифтом на верхний конвейер, который перемещает их к коридорам. Коридоры обслуживаются кранами, которые снимают поддоны с конвейера и помещают их в складские бункеры.
2. Объясните, почему вы согласны (или не согласны) со следующим утверждением: "Большинство моделей дискретного моделирования в той или иной форме можно обнаружить в системах массового обслуживания, состоящих из источников, которые поставляют клиентов; очередей, где клиенты ожидают; и сервисов, где клиенты обслуживаются".

18.4. Элементы дискретного моделирования

В этом разделе показано, как используется концепция события и как собираются статистические данные на основе имитационных моделей систем.

18.4.1. Общее определение событий

Все имитационные модели с дискретными событиями описывают прямо или косвенно ситуации с очередью, в которую клиенты прибывают, при необходимости ожидают в ней, потом обслуживаются перед тем, как оставить систему. В общем случае любая модель с дискретными событиями состоит из сети взаимосвязанных очередей.

Имитационная модель с дискретными событиями в действительности является композицией очередей. В целях сбора статистических данных (показателей функционирования системы) заметим, что изменения в системе (например, изменение длины очереди или состояния средств обслуживания) могут иметь место лишь в случае, когда клиент поступает в очередь или когда покидает систему после обслуживания. Это означает, что двумя главными событиями в любой дискретной имитационной модели являются прибытие и уход клиентов. Это единственные показатели, по которым необходимо исследовать систему. В другие моменты времени никаких изменений, влияющих на статистические данные системы, не происходит.

Пример 18.4–1

Металлообрабатывающий цех получает два вида работ: обычную и срочную. Все работы выполняются последовательно на двух обрабатывающих центрах с обширными буферными зонами. Считается, что срочные работы всегда имеют приоритет перед обычными. Охарактеризуем события в описанной ситуации.

Рассматриваемая ситуация состоит из двух сдвоенных очередей, соответствующих двум обрабатывающим центрам. На первый взгляд можно согласиться со следующим определением событий в описанной ситуации.

А11: обычная работа поступает в центр 1;
А21: срочная работа поступает в центр 1;
Д11: обычная работа уходит из центра 1;
Д21: срочная работа уходит из центра 1;
А12: обычная работа поступает в центр 2;
А22: срочная работа поступает в центр 2;
Д12: обычная работа уходит из центра 2;
Д22: срочная работа уходит из центра 2.

В действительности мы точно имеем лишь два события: прибытие (новой) работы в цех и выход (выполненной) работы из цеха. Заметим сначала, что события Д11 и А12 совпадают и являются неразличимыми. Это же замечание применимо к событиям Д21 и А22. Далее в дискретной имитационной модели мы можем использовать одно событие (прибытие или уход) для обоих типов работ, просто приписывая событию "ярлык", **атрибут** которого указывает тип работы. (В данном случае можно рассматривать в качестве атрибута *персональный идентификационный номер работы*.) Принимая эту аргументацию, приходим к выводу, что события в модели сводятся к 1) поступлению А (в цех) и 2) уходу Д (из каждого центра). Действия, связанные с уходом, будут зависеть от обрабатывающего центра, на котором это происходит.

Определив основные события имитационной модели, покажем теперь, как модель функционирует. Рис. 18.4 дает схематическое представление типичных местонахождений событий на шкале времени имитации. После выполнения всех действий, связанных с текущим событием, имитационная модель "перепрыгивает" к другому событию, которое непосредственно за ним следует. По сути, имитация имеет место в те моменты, когда происходят события.

Как имитационное моделирование определяет время наступления событий? В системе события, связанные с прибытием, определяются временем между поступлениями клиентов, а события, связанные с их уходом, — временем обслуживания.

Время наступления этих событий может быть детерминированным (например, прибытие поездов метро на станцию каждые пять минут) или случайным (например, прибытие клиентов в банк). Если время между наступлениями событий является детерминированным, то процедура определения времени их наступления проста. Если же указанное время является случайным. то используется специальная процедура для получения выборочных значений времени между событиями в системе, соответствующей заданному вероятностному распределению. Содержание такой процедуры рассматривается в следующем разделе.

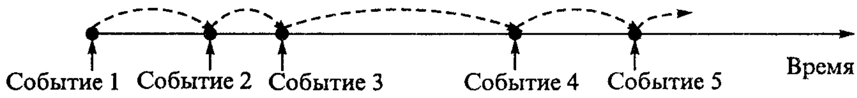


Рис. 18.4

Упражнения 18.4,а

1. Опишите дискретные события, необходимые для моделирования следующей ситуации. Два вида работ поступают из двух различных источников. Все работы выполняются на единственной машине, причем преимущество имеют работы, поступающие из первого источника.
2. Работы поступают с постоянной скоростью на карусельный конвейер. Три станции обслуживания (серверы) расположены равномерно вокруг конвейера. Если станция свободна, то работа снимается с конвейера для выполнения. Иначе работа находится на вращающемся конвейере до тех пор, пока не освободится какой-либо сервер. Выполненная работа складируется в прилегающей зоне для отправления. Определите дискретные события, необходимые для моделирования этой ситуации.
3. Автомобили прибывают по двум линиям в банк, где клиенты обслуживаются, не выходя из машины. Максимальная емкость каждой линии составляет четыре машины. Если обе линии заполнены, то прибывший автомобиль уезжает в поисках другого банка для обслуживания. Если в любой момент на одной из линий по меньшей мере на два автомобиля больше, чем на другой, то последний автомобиль из более "длинной" линии перемешается на последнюю позицию "короткой" линии. В этом режиме банк работает с 8:00 до 15:00 каждый рабочий день. Определите дискретные события для описанной ситуации.
4. Кафетерий начальной школы обеспечивает всех своих учеников завтраком, который помещается на одном подносе. Дети подходят к раздаточному окну каждые 30 секунд. Для получения подноса с завтраком необходимо 18 секунд. Нанесите на временную шкалу события прибытия–ухода для первых пяти учеников.

18.4.2. Генерирование выборочных значений

Случайность в имитационных моделях возникает тогда, когда интервал времени t между последовательными событиями является случайным. В этом разделе рассматриваются методы получения последовательных случайных значений $t = t_1, t_2, \ldots$, имеющих заданное распределение вероятности $f(x)$. Все нижеописанные методы основаны на использовании независимых одинаково распределенных случайных чисел, имеющих равномерное распределение на интервале [0, 1].

Метод обратных функций. Пусть необходимо получить значение x случайной величины y, имеющей непрерывную или дискретную плотность вероятности $f(x)$. Согласно методу обратных функций, сначала находится функция распределения $F(x) = P\{y \le x\}$, где $0 \le F(x) \le 1$ для всех значений x. Пусть R — случайное число, полученное из равномерного на интервале [0, 1] распределения, и пусть F^{-1} — функция, обратная к функции F. Метод обратных функций требует выполнения следующих действий.

Шаг 1. Генерируется случайное число R из интервала [0, 1].

Шаг 2. Вычисляется искомое случайное число $x = F^{-1}(R)$.

На рис. 18.5 проиллюстрирована описанная процедура для непрерывного и дискретного распределений. Равномерно распределенная случайная величина R_1 из интервала [0, 1] проектируется с вертикальной оси $F(x)$ на горизонтальную, определяя при этом искомое значение x_1.

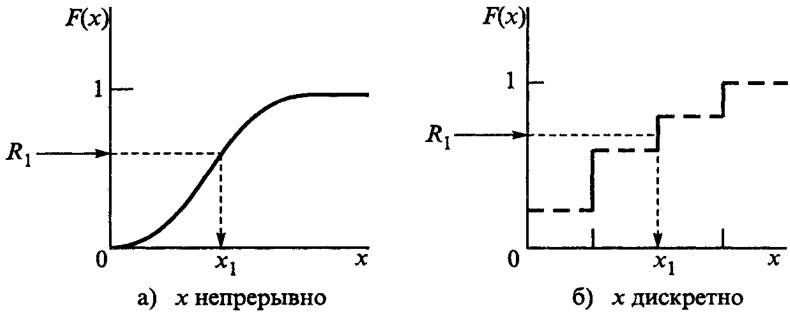


Рис. 18.5

Корректность предложенной процедуры основана на том, что случайная переменная $z = F(x)$ является равномерно распределенной на интервале $0 \le z \le 1$, что доказывается в следующей теореме.

Теорема 18.4-1. *Для заданной функции распределения $F(x)$ случайной величины x, $-\infty < x < \infty$, случайная величина $z = F(x)$, $0 \le z \le 1$, имеет плотность вероятности*

$$f(z) = 1,\ 0 \le z \le 1,$$

т.е. является случайной величиной, равномерно распределенной на интервале [0, 1].

Доказательство. Случайная величина z является равномерно распределенной на интервале [0, 1] тогда и только тогда, когда

$$P\{z \le Z\} = Z,\ \ 0 \le Z \le 1.$$

Это соотношение непосредственно следует из следующих равенств:

$$P\{z \le Z\} = P\{F(x) \le Z\} = P\{x \le F^{-1}(Z)\} = F(F^{-1}(Z)) = Z\,.$$

При этом $0 \le Z \le 1$, так как $0 \le P\{z \le Z\} \le 1$.

Пример 18.4–2. (Экспоненциальное распределение)

Предположим, что время t между прибытиями клиентов в парикмахерскую распределено по экспоненциальному закону с математическим ожиданием $M\{t\} = 1/\lambda$ единиц времени, т.е. плотность вероятности задается формулой

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0.$$

Найдем случайное значение времени t.

Функция распределения вычисляется стандартным образом:

$$F(t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}, \quad t > 0.$$

Если R — случайное число из интервала [0, 1], то, полагая $R = F(t)$, получим

$$t = -\left(\frac{1}{\lambda}\right) \ln(1 - R).$$

Так как R — случайное число из интервала [0, 1], то и $(1 - R)$ представляет собой случайное число из того же интервала, поэтому можно заменить $(1 - R)$ на R.

Пусть в имитационной модели события происходят через t единиц времени. Тогда, например при $\lambda = 4$ посетителя в час и $R = 0.9$, интервал времени между прибытиями посетителей вычисляется как

$$t_1 = -\left(\frac{1}{4}\right) \ln(1 - 0.9) = 0.577 \text{ часа } = 34.5 \text{ минуты.}$$

Значения R, используемые для получения последовательных случайных чисел, должны выбираться случайным образом из интервала [0, 1], подчиняясь равномерному закону распределения. В разделе 18.5 мы покажем, как генерируются такие случайные числа.

Упражнения 18.4,b

1. Пусть в модели из примера 18.4–2 первый клиент поступает в начальный момент времени 0. Используя первые три случайных числа первого столбца табл. 18.1, сгенерируйте время прибытия следующих трех клиентов и нанесите полученные события на временную шкалу.

2. *Равномерное распределение.* Пусть время, необходимое для обработки детали на станке, равномерно распределено на интервале $[a, b]$, $a < b$, т.е.

$$f(t) = \frac{1}{b - a}, \quad a \le t \le b.$$

Найдите выражение для вычисления случайного числа t при заданном значении случайного числа R.

3. В небольшой цех с одним станком заказы поступают случайным образом. Время между поступлениями заказов распределено по экспоненциальному закону с математическим ожиданием 2 часа. Время, необходимое для выполнения заказа, является случайной величиной, равномерно распределенной на интервале [1.1, 2], измеряемом в часах. Пусть первый заказ поступает в момент времени, равный нулю. Определите время поступления и выполнения первых пяти заказов, используя случайные числа из интервала [0, 1], помещенные в первом столбце табл. 18.1.

4. Запрос на дорогую запасную деталь для пассажирского самолета равен 0, 1, 2 или 3 единицы в день с вероятностями 0.2, 0.3, 0.4 и 0.1 соответственно. Цех технического обеспечения полетов имеет на складе 5 упомянутых деталей и немедленно восстановит запас до этого же уровня, если их останется на складе меньше двух единиц.

   a) Разработайте процедуру получения случайных значений величины запроса.
   b) Сколько дней пройдет до первого пополнения запаса деталей? Используйте последовательные значения случайных чисел R из первого столбца табл. 18.1.

5. В имитационной модели телевизионные блоки проверяются на наличие дефектов. В 80% случаев блок успешно проходит проверку и упаковывается, иначе он ремонтируется. Символически эту ситуацию можно представить одним из двух способов.

   Выполнить ремонт с вероятностью 0.2, упаковать с вероятностью 0.8.

   Упаковать с вероятностью 0.8, выполнить ремонт с вероятностью 0.2.

   Эти два способа кажутся эквивалентными. Однако при моделировании этой ситуации с помощью одной и той же заданной последовательности случайных чисел из интервала [0, 1] эти представления могут дать различные результаты (в виде “выполнить ремонт” или “упаковать”). Объясните, почему.

6. Игрок подбрасывает симметричную монету до тех пор, пока не появится ее лицевая сторона (герб). Результирующая выплата равна 2^n, где n — число подбрасываний до появления лицевой стороны монеты.

   a) Разработайте имитационную модель игры.
   b) Используйте случайные числа из первого столбца табл. 18.1 для определения накопленного значения выплаты после того, как лицевая сторона монеты появится два раза.

7. *Треугольное распределение.* В имитационном моделировании при недостатке данных часто невозможно определить вероятностное распределение, соответствующее моделируемым ситуациям. Во многих таких случаях может оказаться полезным описание распределения случайной переменной путем оценки ее наименьшего, наиболее вероятного и наибольшего значений. Этих трех величин достаточно для определения треугольного распределения, которое можно использовать в качестве “черновой” оценки истинного распределения.

   a) Разработайте формулу для получения случайных значений, соответствующих треугольному распределению, параметрами которого являются константы a, b и c, $a < b < c$, и плотность вероятности которого задается формулой

$$f(x)=\begin{cases}\dfrac{2(x-a)}{(b-a)(c-a)}, & a \le x \le b,\\ \dfrac{2(c-x)}{(c-b)(c-a)}, & b \le x \le c.\end{cases}$$

b) Получите три значения, соответствующие треугольному распределению с параметрами (1, 3, 7), используя для этого три первых случайных числа первого столбца табл. 18.1.

8. Разработайте процедуру получения случайных значений, подчиняющихся распределению, плотность вероятности которого состоит из прямоугольника, граничащего слева и справа с двумя прямоугольными треугольниками (вертикальными катетами этих треугольников являются стороны прямоугольника). Соответствующие основания левого треугольника, прямоугольника и треугольника справа равны $[a, b]$, $[b, c]$ и $[c, d]$, $a < b < c < d$. Каждый треугольник имеет высоту, равную высоте прямоугольника.

   Определите пять случайных значений, которые соответствуют описанному выше распределению с набором параметров $(a, b, c, d) = (1, 2, 4, 6)$, используя при этом пять первых случайных чисел из первого столбца табл. 18.1.

9. *Геометрическое распределение.* Покажите, как можно получить случайное значение, подчиняющееся геометрическому распределению

   $$f(x) = p(1-p)^x,\ \ x = 0, 1, 2, \ldots,$$

   где x — число неудач в схеме Бернулли до первого появления успеха, p — вероятность успеха, $0 < p < 1$. Сгенерируйте пять случайных значений при $p = 0.6$.

10. *Распределение Вейбулла.*[1] Покажите, как можно получить случайное значение, подчиняющееся распределению Вейбулла, плотность вероятности которого имеет вид

   $$f(x)=\alpha\beta^{-\alpha}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}},\ \ x>0,$$

   где $\alpha > 0$ — параметр формы, $\beta > 0$ — параметр масштаба.

Метод обратных функций дает хорошие результаты для непрерывных распределений, функция распределения которых имеет аналитическое представление. Такие распределения, как нормальное, гамма-распределение и распределение Пуассона, к упомянутому классу не принадлежат. Для получения случайных значений, имеющих эти распределения, служат методы, рассмотренные далее в этой главе.

Метод сверток. Основная идея данного метода состоит в том, чтобы выразить искомую случайную величину в виде суммы других случайных величин, для которых легко получить реализации случайных значений. Типичными среди таких распределений являются распределения Эрланга и Пуассона, которые можно получить из экспоненциального распределения.

---

[1] В русской математической литературе это распределение также имеет название "распределение Вейбулла–Гнеденко". — *Прим ред.*

Пример 18.4–3. (Распределение Эрланга)

Случайная величина, имеющая распределение Эрланга размерности m, определяется как сумма (свертка) m независимых случайных величин, каждая из которых имеет экспоненциальное распределение с параметром λ. Пусть y — случайная величина, подчиняющаяся распределению Эрланга размерности m. Тогда

$$y = y_1 + y_2 + \dots + y_m,$$

где y_i ($i = 1, 2, \dots, m$) — независимые экспоненциально распределенные случайные величины, плотность вероятности которых задается формулой

$$f(y_i) = \lambda e^{-\lambda y_i}, \quad y_i > 0, \quad i = 1, 2, \dots, m.$$

Как следует из примера 18.4–2, случайное значение, имеющее (i-е) экспоненциальное распределение, вычисляется по формуле

$$y_i = -\left(\frac{1}{\lambda}\right)\ln(R_i), \quad i = 1, 2, \dots, m.$$

Следовательно, значение случайной величины Эрланга размерности m можно вычислить как

$$y = -\left(\frac{1}{\lambda}\right)\left[\ln(R_1) + \ln(R_2) + \dots + \ln(R_m)\right] = -\left(\frac{1}{\lambda}\right)\ln(R_1 R_2 \dots R_m).$$

В качестве иллюстрации применения этой формулы предположим, что $m = 3$ и $\lambda = 4$ события в час. Первые три случайных числа из первого столбца табл. 18.1 дают такой результат: $R_1R_2R_3 = 0.0589 \times 0.6733 \times 0.4799 = 0.0190$, что в свою очередь приводит к значению $y = -(1/4)\ln(0.019) = 0.991$ часа.

Пример 18.4–4. (Распределение Пуассона)

Если время между некоторыми событиями представляет собой случайную величину, распределенную по экспоненциальному закону, то распределение числа событий, происходящих в единицу времени, будет пуассоновским и наоборот. Этот факт используется для получения случайных значений, подчиняющихся распределению Пуассона.

Пусть для рассматриваемого распределения Пуассона среднее количество событий в единицу времени равно λ. Тогда время между событиями является случайной величиной, имеющей экспоненциальное распределение со средним, равным $1/\lambda$ единиц времени. Это означает, что на протяжении t единиц времени будет иметь место n событий (число n подчиняется распределению Пуассона) тогда и только тогда, когда

время до реализации события $n \le t \le$ время до реализации события $n + 1$.

Это условие можно записать следующим образом:

$$t_1 + t_2 + \dots + t_n \le t \le t_1 + t_2 + \dots + t_{n+1}, \; n > 0,$$
$$0 \le t \le t_1, \; n = 0,$$

где t_i — случайная величина, подчиняющаяся экспоненциальному распределению со средним $1/\lambda$. Принимая во внимание результат примера 18.4–3, имеем

$$-\left(\frac{1}{\lambda}\right)\ln\left(R_1R_2...R_n\right) \le t < -\left(\frac{1}{\lambda}\right)\ln\left(R_1R_2...R_{n+1}\right),\ \ n>0,$$

$$0 \le t < -\left(\frac{1}{\lambda}\right)\ln\left(R_1\right),\ \ n=0.$$

Отсюда получаем

$$R_1R_2...R_n \ge e^{-\lambda t} > R_1R_2...R_{n+1},\ \ n>0,$$

$$1 \ge e^{-\lambda t} > R_1,\ \ n=0.$$

Проиллюстрируем описанный подход на следующем примере. Пусть необходимо получить случайное значение, соответствующее распределению Пуассона со средней частотой $\lambda = 4$ события в час, при $t = 0.5$ часа. Это нам дает $e^{-\lambda t} = 0.1353$. Используя случайные числа первого столбца табл. 18 1, замечаем, что $R_1 = 0.0589$ меньше $e^{-\lambda t} = 0.1353$. Следовательно, $n = 0$.

Пример 18.4–5. (Нормальное распределение)

Центральная предельная теорема утверждает, что сумма n одинаково распределенных случайных величин стремится к нормально распределенной величине при бесконечном увеличении n. Мы используем этот результат для получения значений, соответствующих нормальному распределению с математическим ожиданием μ и стандартным отклонением σ.

Обозначим $x = R_1 + R_2 + ... + R_n$, где $R_1, R_2, ..., R_n$ — случайные числа, равномерно распределенные в интервале [0, 1]. В соответствии с центральной предельной теоремой случайная величина x является асимптотически нормальной величиной со средним $n/2$ и дисперсией $n/12$. Следовательно, случайная величина y, подчиняющаяся нормальному распределению $N(\mu, \sigma)$ с математическим ожиданием μ и стандартным отклонением σ, может быть получена из случайной величины x по формуле

$$y = \mu + \sigma\left(\frac{x - \frac{n}{2}}{\sqrt{\frac{n}{12}}}\right).$$

Для удобства в практических задачах n обычно выбирается равным 12, что приводит предыдущую формулу к виду $y = \mu + \sigma(x - 6)$.

Для демонстрации этого метода предположим, что необходимо получить случайное значение, соответствующее нормальному распределению $N(10, 2)$ (математическое ожидание $\mu = 10$ и стандартное отклонение $\sigma = 2$). Вычисляя сумму первых 12 случайных чисел из первого и второго столбцов табл. 18.1, получаем $x = 6.1094$. Следовательно, $y = 10 + 2(6.1094 - 6) = 10.2188$.

Неудобство этой процедуры состоит в том, что необходимо генерировать 12 случайных чисел из интервала [0, 1] для получения одного выборочного значения, соответствующего рассматриваемому нормальному распределению, что делает ее малоэффективной с вычислительной точки зрения. В соответствии с более эффективной процедурой решения этой же задачи необходимо использовать преобразование

$$x = \sqrt{-2\ln(R_1)}\cos(2\pi R_2).$$

Бокс и Мюллер (Box and Muller) [1] доказали, что случайная величина x является стандартной нормально распределенной случайной величиной, т.е. имеет распределение $N(0, 1)$. Следовательно, $y = \mu + \sigma x$ дает значение, подчиняющееся нормальному распределению $N(\mu, \sigma)$. Эта процедура является более эффективной, так как требует генерирования всего двух случайных чисел из интервала [0, 1].

В действительности данный метод (метод Бокса–Мюллера) является еще более эффективным, так как Бокс и Мюллер доказали, что предыдущая формула дает другое значение, также имеющее нормальное распределение $N(0, 1)$, если $\cos(2\pi R_2)$ заменить на $\sin(2\pi R_2)$. Это значит, что два случайных числа R_1 и R_2 из интервала [0,1] можно использовать для одновременного получения *двух* значений, соответствующих нормальному распределению $N(0, 1)$.[1]

В качестве иллюстрации применим метод Бокса–Мюллера для нахождения значений, подчиняющихся нормальному распределению $N(10, 2)$. Два первых случайных числа первого столбца табл. 18.1 приводят к следующим значениям, подчиняющимся нормальному распределению $N(0, 1)$:

$$x_1 = \sqrt{-2\ln(0.0589)}\cos(2\pi\times 0.6733) \approx -1.103,$$

$$x_2 = \sqrt{-2\ln(0.0589)}\sin(2\pi\times 0.6733) \approx -2.108.$$

Следовательно, соответствующие значения, имеющие нормальное распределение $N(10, 2)$, равны

$$y_1 = 10 + 2(-1.103) = 7.794,$$
$$y_2 = 10 + 2(-2.108) = 5.7823.$$

Упражнения 18.4,с[2]

1. В задаче примера 18.4–3 вычислите случайное значение, соответствующее распределению Эрланга при $m = 4$ и $\lambda = 4$.
2. В задаче примера 18.4–4 сгенерируйте три случайных значения, соответствующих распределению Пуассона для двухчасового периода при математическом ожидании 5 событий в час.
3. В задаче примера 18.4–5 сгенерируйте два случайных значения, соответствующих нормальному распределению $N(8, 1)$, используя как метод сверток, так и метод Бокса–Мюллера.

[1] Интересно отметить, что две случайные величины, получаемые в методе Бокса–Мюллера по схожим формулам, причем зависимые от одних и тех же равномерно распределенных случайных величин, независимы между собой. — *Прим. ред.*

[2] Для всех приведенных здесь задач используйте случайные числа из табл. 18.1, начиная с первого столбца.

4. Работы поступают в металлообрабатывающий цех в соответствии с распределением Пуассона с математическим ожиданием шесть работ в день. Цех имеет пять обрабатывающих центров, на которые контролер направляет полученные работы в строгом соответствии с очередностью. Определите одно случайное значение интервала между получением работ на первом обрабатывающем центре.

5. Проведен стандартный тест ACT среди выпускников одного класса средней школы небольшого городка. Результаты теста являются нормально распределенной случайной величиной с математическим ожиданием 27 баллов и стандартным отклонением 3 балла. Используя метод Бокса–Мюллера, получите случайные значения показателей шести выпускников этого класса.

6. Профессор психологии Ятаха проводит обучающий эксперимент, в котором мыши приучаются находить путь внутри лабиринта. Основой лабиринта является квадрат. Мышь впускают в лабиринт через один из четырех его углов, и она должна найти путь через лабиринт таким образом, чтобы выйти из него через этот же угол. Конструкция лабиринта такова, что до своего выхода из лабиринта мышь должна пройти через оставшиеся три угла в точности по одному разу. Многовариантные пути лабиринта соединяют четыре его вершины строго в направлении вращения часовой стрелки. Профессор считает, что время, которое мышь затрачивает для перехода от одной вершины лабиринта до другой, является случайной величиной, равномерно распределенной на интервале от 10 до 20 секунд. Опишите процедуру получения случайного значения для времени пребывания мыши в лабиринте.

7. Пусть в ситуации, описанной в предыдущем упражнении, как только одна мышь покидает лабиринт, другая в тот же миг входит в него. Опишите процедуру получения случайного значения для количества мышей, которые пройдут лабиринт на протяжении 5 минут.

8. *Отрицательное биномиальное распределение (распределение Паскаля)*. Покажите, как можно вычислить случайное значение, имеющее отрицательное биномиальное распределение, плотность вероятности которого имеет вид

$$f(x)=\binom{y+x-1}{x}p^y(1-p)^x,\quad x=0,1,2,\dots,$$

где x — число неудач в последовательности независимых испытаний Бернулли до получения y-го успеха, p — вероятность успеха, $0<p<1$. (*Подсказка.* Случайная величина, имеющая отрицательное биномиальное распределение, является сверткой (суммой) независимых случайных величин, подчиняющихся геометрическому распределению. См. также упр. 18.4,b(9).)

Метод отбора.[1] Данный метод разработан для получения значений случайных величин со сложными функциями плотностей вероятностей, к которым нельзя применить изложенные выше методы. Общая идея данного метода сводится к замене сложной плотности вероятности $f(x)$ более удобной, с аналитической точки зрения, плотностью вероятности $h(x)$. Затем значения, соответствующие плотности $h(x)$, используются для получения значений, соответствующих исходной плотности $f(x)$.

---

[1] В русской математической литературе этот метод иногда также называют методом отказов, что больше соответствует английскому названию *acceptance-rejection method*. — *Прим. ред.*

Для плотности вероятности $f(x)$ определяем **мажорирующую функцию** $g(x)$, такую, что

$$g(x) \geq f(x), \quad -\infty < x < \infty.$$

Теперь определим плотность вероятности $h(x)$ путем нормализации функции $g(x)$:

$$h(x) = \frac{g(x)}{\int_{-\infty}^{\infty} g(y)\,dy}, \quad -\infty < x < \infty.$$

В методе отбора последовательно выполняются следующие действия.

Шаг 1. С помощью метода обратной функции или метода свертки получаем случайное значение $x = x_1$, соответствующее плотности вероятности $h(x)$.

Шаг 2. Генерируем случайное число R из интервала [0, 1].

Шаг 3. Если $R \leq f(x_1)/g(x_1)$, следует принять x_1 как искомое значение, соответствующее распределению $f(x)$. Иначе необходимо вернуться к шагу 1, отбросив значение x_1.

Обоснованность этого метода вытекает из следующего равенства:

$$P\{x \leq a \mid x = x_1 \text{ принимается}, \ -\infty < x_1 < \infty\} = \int_{-\infty}^{a} f(y)\,dy, \quad -\infty < a < \infty.$$

Это вероятностное соотношение означает, что значение $x = x_1$, удовлетворяющее условию шага 3, в действительности является значением, соответствующим исходной плотности вероятности $f(x)$, что и требуется.

Эффективность предложенного метода можно повысить путем уменьшения вероятности отклонения значения x_1 на шаге 3. Эта вероятность зависит от выбранной мажорирующей функции $g(x)$ и должна уменьшаться с выбором такой функции $g(x)$, которая более "точно" мажорирует функцию $f(x)$.

Пример 18.4–5. (Бета–распределение)

Используем метод отбора для нахождения значения, соответствующего бета–распределению, плотность вероятности которого задается формулой

$$f(x) = 6x(1-x), \quad 0 \leq x \leq 1.$$

На рис. 18.6 изображены функция $f(x)$ и мажорирующая ее функция $g(x)$.

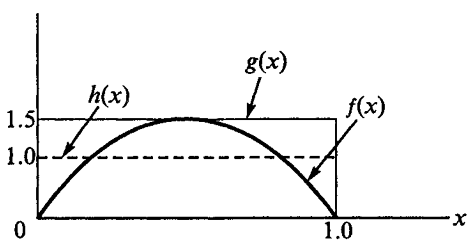


Рис. 18.6

Высота мажорирующей функции $g(x)$ равна максимальному значению функции $f(x)$, которого она достигает в точке $x = 0.5$. Это значит, что $g(x) = 1.5,\ 0 < x < 1$.

Функция плотности "замещающего" распределения $h(x)$, также представленная на рис. 18.6, вычисляется согласно соотношению

$$h(x) = \frac{g(x)}{\text{площадь под } g(x)} = \frac{1.5}{1\times 1.5} = 1,\ \ 0 < x < 1.$$

Следующие действия демонстрируют применение процедуры метода отбора с использованием последовательности случайных чисел из табл. 18.1.

Шаг 1. Использование числа $R = 0.0589$ приводит к случайному значению $x = 0.0589$, соответствующему плотности $h(x)$.

Шаг 2. Выбираем из табл. 18.1 следующее число $R = 0.6733$.

Шаг 3. Так как $f(0.0589)/g(0.0589) = 0.3326/1.5 = 0.2217$ меньше $R = 0.6733$, мы отбрасываем значение $x = 0.0589$.

Для получения второго значения повторяем действия.

Шаг 1. Использование числа $R = 0.4799$ (из первого столбца табл. 18.1) приводит к случайному значению $x = 0.4799$, соответствующему плотности $h(x)$.

Шаг 2. Выбираем из табл. 18.1 следующее число $R = 0.9486$.

Шаг 3. Так как $f(0.4799)/g(0.4799) = 0.9984$ больше $R = 0.9486$, мы принимаем значение $x = 0.4799$ как соответствующее бета-распределению.

Из этого примера видно, что итерации метода отбора должны повторяться с новыми значениями случайных чисел, равномерно распределенных на интервале [0, 1], до тех пор, пока не будут выполнены условия шага 3.

Эффективность метода отбора повышается в результате выбора такой мажорирующей функции $g(x)$, которая "облегала" бы функцию $f(x)$ как можно более плотно, порождая в то же время приемлемую с аналитической точки зрения аппроксимирующую функцию $h(x)$. Например, метод будет более эффективным, если прямоугольную мажорирующую функцию $g(x)$ на рис. 18.6 заменить ступенчатой (см. упр. 18.4,d(2)). С увеличением числа ступеней функция $g(x)$ становится все более прилегающей к функции $f(x)$ и, следовательно, возрастает вероятность принятия полученного случайного значения как искомого. Однако получение тесно прилегающей мажорирующей функции влечет за собой дополнительные вычисления, которые могут стать чрезмерными, что, в свою очередь, может свести на нет преимущества высокой вероятности принятия полученных значений.

Упражнения 18.4,d

1. В примере 18.4–5 продолжите выполнение процедуры метода отбора до получения следующего приемлемого случайного значения.
2. Рассмотрим плотность вероятности бета-распределения из примера 18.4–5. Определите двухступенчатую "пирамидальную" мажорирующую функцию $g(x)$ с двумя равными скачками величины 0.75. Получите случайное значение, соответствую-

щее бета–распределению, с использованием новой мажорирующей функции и тех же случайных чисел из интервала [0, 1] (табл. 18.1), которые использовались в примере 18.4–5. Вывод, который можно здесь сделать, сводится к тому, что использование более точной мажорирующей функции повышает вероятность принятия полученного значения как искомого. Заметьте, однако, что объем вычислений, связанных с использованием новой функции, увеличился.

3. Определите функции $g(x)$ и $h(x)$ для применения метода отбора к плотности распределения следующего вида:

$$f(x)=\frac{\sin(x)+\cos(x)}{2},\quad 0\le x\le\frac{\pi}{2}.$$

Используйте случайные числа из первого столбца табл. 18.1 для получения двух значений, соответствующих плотности распределения $f(x)$. (*Совет.* Для удобства используйте прямоугольную функцию $g(x)$ над областью определения функции $f(x)$.)

4. Время между приходом клиентов в парикмахерскую описывается следующим распределением:

$$f_1(t)=\frac{k_1}{t},\quad 12\le t\le 20.$$

Время стрижки является случайной величиной, плотность вероятности которой

$$f_2(t)=\frac{k_2}{t^2},\quad 18\le t\le 22.$$

Константы k_1 и k_2 выбираются из условия, что функции $f_1(x)$ и $f_2(x)$ являются плотностями вероятностей. Используйте метод отбора (и случайные числа из табл. 18.1) для определения времени ухода первого клиента из парикмахерской и времени прихода второго клиента. Предположите, что первый клиент приходит в момент времени $T = 0$.

18.5. Генерирование случайных чисел

Равномерно распределенные случайные числа из интервала [0, 1] играют ключевую роль в получении выборок из любого вероятностного распределения. Истинные случайные числа из интервала [0, 1] можно генерировать лишь с помощью электронных приборов. Так как имитационные модели реализуются на компьютере, использование электронных приборов для генерации случайных чисел слишком бы замедлило процедуру имитационного моделирования. Кроме того, электронные приборы активизируются случайным образом. Следовательно, невозможно по желанию воспроизвести одну и ту же последовательность случайных чисел. Этот факт чрезвычайно важен, так как для отладки, проверки и утверждения имитационной модели часто требуется дублирование одной и той же последовательности случайных чисел.

В имитационном моделировании единственным подходящим методом генерации случайных чисел из интервала [0, 1] является метод, основанный на арифметических операциях. Такие числа не являются истинно случайными, так как они могут быть определены заранее, поэтому их называют **псевдослучайными**.

Наиболее часто используется **мультипликативный метод сравнений**, который генерирует случайные числа из интервала [0, 1] с использованием арифметических операций. В соответствии с этим методом псевдослучайное число R_n при заданных значениях параметров u_0, b, c, и m можно вычислить по следующей формуле:

$$u_n = (bu_{n-1} + c)\bmod(m), \quad n = 1, 2, \dots,$$

$$R_n = \frac{u_n}{m}, \quad n = 1, 2, \dots .$$

Начальное значение параметра u_0 обычно называют **начальным числом** генератора случайных чисел.

Пример 18.5–1

Используя мультипликативный метод сравнений, сгенерируем три случайных числа при следующих начальных данных: $b = 9$, $c = 5$, $u_0 = 11$ и $m = 12$.

$$u_1 = (9 \times 11 + 5)\bmod 12 = 8, \quad R_1 = \frac{8}{12} = 0.6667,$$

$$u_2 = (9 \times 8 + 5)\bmod 12 = 5, \quad R_2 = \frac{5}{12} = 0.4167,$$

$$u_3 = (9 \times 5 + 5)\bmod 12 = 4, \quad R_3 = \frac{4}{12} = 0.3333.$$

Конкретный выбор параметров u_0, b, c, и m является решающим фактором, определяющим статистические качества генератора случайных чисел, а также длины его цикла (по окончании цикла генерируемая последовательность начинает повторять себя). Использование параметров, выбранных “наобум”, не дает хорошего генератора случайных чисел. Надежные генераторы, наряду с достаточно большой длиной цикла генерируемых случайных чисел, должны пройти соответствующие статистические проверки, чтобы гарантировать равномерное распределение на интервале [0, 1] полученной последовательности. На это условие нужно обращать внимание при использовании непроверенного программного обеспечения в качестве генератора случайных чисел.

Возможны также вариации мультипликативного метода сравнений, которые улучшают качество генератора. Подробное обсуждение этого вопроса можно найти в работе [2].

Упражнения 18.5,а

1. Покажите, что значения параметров b, c, u_0 и m, использованные в примере 18.5–1, быстро приводят к повторению случайных чисел.
2. Найдите программу генератора случайных чисел для вашего компьютера и сгенерируйте с его помощью 500 случайных чисел из интервала [0, 1]. Постройте гистограмму полученных чисел и визуально убедите себя в том, что есть веские основания считать, что они подчинены равномерному распределению из интервала

[0, 1]. Впрочем, чтобы проверить последовательность надлежащим образом, вам необходимо использовать следующие тесты: критерий согласия хи–квадрат (раздел 12.6), тест на независимость и корреляционный тест (см. работу [1]).

18.6. Механика дискретной имитации

Как указывалось выше, все дискретные имитационные модели в той или иной форме представляют ситуации, связанные с очередями, в которых есть два типа основных событий: прибытие и уход. Эти события определяют моменты, в которые могут происходить изменения в статистике системы.

В этом разделе детально обсуждаются вопросы сбора статистических данных, полученных в процессе наблюдений над имитационной моделью. Все объяснения проводятся на модели простой очереди.

Пример 18.6–1

Время между приходом клиентов в парикмахерскую является случайной величиной, изменяющейся по экспоненциальному закону с математическим ожиданием 12 минут. В парикмахерской работает один мастер, который выполняет два вида стрижки: обычную мужскую, которая занимает около 10 минут, и квалифицированную, которая занимает около 15 минут. В среднем один из четырех клиентов заказывает обычную стрижку, хотя все происходит случайно. Требуется определить следующие параметры работы парикмахерской.

1. Среднюю занятость парикмахера.
2. Среднее количество ожидающих клиентов.
3. Среднее время ожидания клиента в очереди.

Логику работы имитационной модели можно описать в терминах событий, связанных с прибытием и уходом клиентов, следующим образом.

Событие, связанное с прибытием клиента

1. Сгенерировать и сохранить время события, связанного с прибытием следующего клиента (= текущее время моделирования + промежуток времени между приходами клиентов).
2. Если средство обслуживания (парикмахер) свободно
   a) начать обслуживание поступившего клиента, изменить состояние системы на рабочее и скорректировать данные использования системы;
   b) сгенерировать и сохранить хронологически время события, связанного с уходом клиента (= текущее время моделирования + время обслуживания).
3. Если средство обслуживания занято, поставить поступившего клиента в очередь и увеличить ее длину на единицу.

Событие, связанное с уходом клиента (окончание обслуживания)

1. Если очередь пуста, объявить систему свободной. Скорректировать данные использования системы.

2. Если очередь не является пустой

   a) а) начать обслуживание первого в очереди клиента. Уменьшить длину очереди на единицу и скорректировать данные использования системы;
   b) б) сгенерировать и сохранить хронологически время события, связанного с уходом клиента (= текущее время моделирования + время обслуживания).

В рассматриваемом примере время между приходом клиентов изменяется по экспоненциальному закону с математическим ожиданием 12 минут, а время обслуживания является дискретным (10 минут с вероятностью 0.25 или 15 минут с вероятностью 0.75). Обозначим через p и q случайные значения времени между прибытием клиентов и обслуживанием соответственно. Используя результаты раздела 18.4.2, получаем

$$p = -10\ln(R) \text{ минут, } 0 \le R \le 1,$$

$$q = \begin{cases} 10 \text{ минут}, & 0 \le R \le 0.25, \\ 15 \text{ минут}, & 0.25 \le R \le 1. \end{cases}$$

В этом примере используем случайные числа R из табл. 18.1, начиная с первого столбца. Обозначим через T время моделирования (время наблюдения над моделью). Предположим также, что первый клиент приходит в момент времени $T = 0$ и средство обслуживания свободно.

Событие, связанное с прибытием клиента в момент $T = 0$ (клиент 1). Второй клиент прибудет в момент времени $T = 0 + p_1$, где $p_1 = -10\ln(0.0589) = 28.3$ минуты.

Так как при $T = 0$ система свободна, немедленно начинается обслуживание первого клиента. Соответствующее время обслуживания вычисляется с использованием равномерно распределенного случайного числа $R = 0.6733$, что дает $q_1 = 15$ минут. Это значит, что первый клиент уйдет из парикмахерской в момент времени $T = 0 + 15 = 15$. Время прихода второго клиента и время ухода первого ($T = 28.3$ и $T = 15$) порождают следующий хронологический перечень будущих событий.

1. Уход в $T = 15$.
2. Приход в $T = 28.3$.

Следовательно, наиболее близким является событие, связанное с уходом клиента в момент времени $T = 15$.

Событие, связанное с уходом клиента в момент $T = 15$ (клиент 1). Так как очередь пуста, средство обслуживания становится свободным. В то же время фиксируем, что система была занята от момента времени $T = 0$ до $T = 15$. Имеем следующий скорректированный перечень будущих событий (в данном случае имеется только одно событие).

1. Приход второго клиента в $T = 28.3$.

Таким образом, следующим в хронологическом порядке является событие, связанное с приходом клиента в момент времени $T = 28.3$.

Событие, связанное с прибытием клиента в момент $T = 28.3$ (клиент 2). Третий клиент прибудет через $p_2 = -10\ln(0.4799) = 7.34$ минуты, т.е. в момент времени $T = 28.3 + 7.34 = 35.64$. Так как средство обслуживания свободно, начинается обслуживание второго клиента и система объявляется занятой. При $R = 0.9486$ соответствующее время обслуживания равно $q_2 = 15$ минут. Следовательно, второй клиент оставит систему в момент времени $T = 28.3 + 15 = 43.3$. Список будущих событий изменяется следующим образом.

1. Приход в $T = 35.64$.
2. Уход в $T = 43.3$.

Следующим событием в хронологическом порядке является приход клиента в момент времени $T = 35.64$.

Событие, связанное с прибытием клиента в момент T = 35.64 (клиент 3). Клиент 4 прибудет в момент времени $T = 35.64 + [-10 \ln(0.6139)] = 40.52$.

Так как система в это время занята (она занята до $T = 43.3$), третий клиент помещается в очередь в момент $T = 35.64$. Откорректированный список будущих событий принимает следующий вид.

1. Приход в $T = 40.52$.
2. Уход в $T = 43.3$.

Событие, связанное с прибытием клиента в момент T = 40.52 (клиент 4). Клиент 5 прибудет в момент времени $T = 40.52 + [-10 \ln(0.5933)] = 45.74$.

Так как система в это время все еще занята (она занята до $T = 43.3$), четвертый клиент помещается вторым в очередь в момент времени $T = 40.52$. Откорректированный список будущих событий принимает следующий вид.

1. Уход в $T = 43.3$.
2. Приход в $T = 45.74$.

Событие, связанное с уходом клиента в момент T = 43.3 (клиент 2). Первый клиент в очереди (третий клиент) начинает обслуживаться. При $R = 0.9341$ соответствующее событие, связанное с его уходом, произойдет в момент времени $T = 43.3 + 15 = 58.3$.

Система в это время все еще остается занятой, длина очереди уменьшается на единицу в момент $T = 43.3$. Откорректированный список будущих событий принимает такой вид.

1. Приход в $T = 45.74$.
2. Уход в $T = 58.3$.

Событие, связанное с прибытием клиента в момент T = 45.74 (клиент 5). Клиент 6 прибудет в момент времени $T = 45.74 + [-10 \ln(0.1782)] = 62.99$.

Пятый клиент помещается последним в очередь в момент $T = 45.74$.

Описанные вычисления демонстрируют все возможности, которые заложены в логике функционирования имитационной модели. Здесь мы прекращаем вычисления и покажем, как вычисляются требуемые характеристики рассматриваемой модели.

Предположим, что имитация системы выполняется до момента времени $T = 50$ минут. На рис. 18.7 показаны изменения длины очереди и использование сервиса (занятость системы) как функции времени имитации.

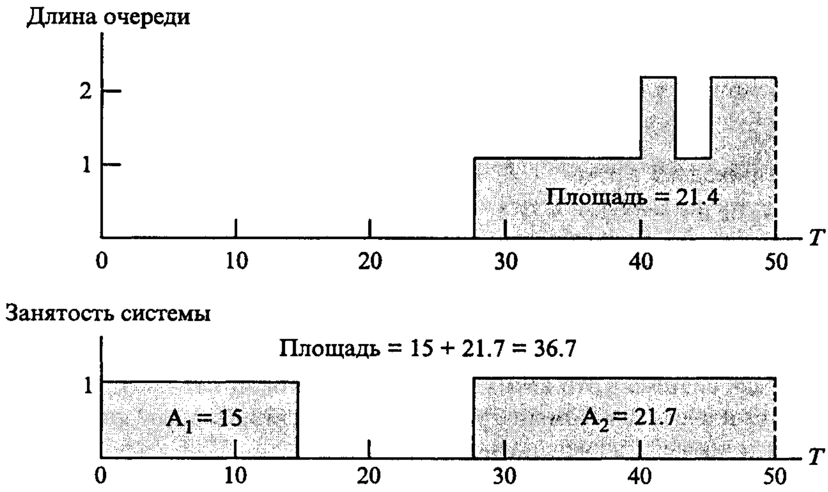


Рис. 18.7

Длина очереди и занятость системы (то бишь парикмахера) являются переменными, зависящими от времени. Поэтому их средние значения вычисляются следующим образом.

$$\left(\begin{array}{c}\text{Среднее значение переменной,}\\ \text{зависящей от времени}\end{array}\right) = \frac{\text{Площадь под кривой}}{\text{Период имитации}}.$$

Применяя эту формулу для данных, показанных на рис. 18.7, получаем

Средняя длина очереди = 21.4/50 = 0.428 клиента,
Средняя занятость системы = 36.7/50 = 0.734.

Среднее время ожидания в очереди является переменной, зависящей от количества происшедших событий, и ее значение вычисляется по формуле

$$\left(\begin{array}{c}\text{Среднее значение переменной,}\\ \text{зависящей от количества событий}\end{array}\right) = \frac{\text{Сумма наблюдений}}{\text{Количество событий}}.$$

Из рис. 18.7 видно, что площадь под кривой длины очереди в действительности равна сумме времен ожидания трех клиентов, которые формировали очередь. Именно первый ожидающий клиент (клиент 3) находился в очереди с момента $T = 35.64$ до $T = 43.3$, второй (клиент 4) был отправлен в очередь в момент времени $T = 40.52$ и оставался в ней до конца времени имитации ($T = 50$), третий ожидающий клиент (клиент 5) поступил в очередь в момент времени $T = 45.74$ и также оставался в ней в момент $T = 50$. Следовательно, сумма времен ожидания трех клиентов равна $(43.3 - 35.64) + (50 - 40.52) + (50 - 45.74) = 21.4$ минуты, что равняется площади под кривой длины очереди.

Среднее время ожидания в очереди равно 21.4/3 = 7.13 минуты. Эта величина имеет отношение лишь к тем клиентам, которые ожидали в очереди. Среднее время ожидания для всех клиентов (= 5), независимо от того, ожидали они в очереди или нет, вычисляется так: 21.4/5 = 4.28 минуты.

Утомительные вычисления примера 18.6–1 указывают на необходимость использования компьютера как существенного инструмента для реализации имитационных моделей. И действительно, было разработано несколько специализированных языков программирования (например, SIMNET II, GPSS, SIMAN), ориентированных на автоматизацию таких типов вычислений, которые встречаются в имитационном моделировании. В разделе 18.8 содержится краткая информация о языках программирования, ориентированных на имитационное моделирование.

Упражнения 18.6,а

1. В примере 18.6–1 продолжите моделирование до тех пор, пока не будет обслужен пятый клиент. Используйте случайные числа из табл. 18.1 в той же последовательности, в которой они используются в примере. Вычислите результирующую среднюю длину очереди, среднюю занятость парикмахера и среднее время ожидания в очереди.
2. Предположим, что в парикмахерской, о которой шла речь в примере 18.6–1, работают два парикмахера и клиенты обслуживаются согласно принципу "первым пришел — первым обслуживаешься". Предположим также, что время стрижки является случайной величиной, равномерно распределенной на интервале от 15 до 30 минут; время между приходом клиентов распределено по экспоненциальному закону с математическим ожиданием 10 минут. Смоделируйте вручную работу системы на протяжении 75 единиц времени. Из полученных результатов имитации определите среднее время ожидания клиента, среднее число ожидающих клиентов и среднюю занятость парикмахеров. Используйте случайные числа из табл. 18.1.
3. Классифицируйте следующие переменные как зависящие либо от количества событий, либо от времени.

   a) Время отказа электронного прибора.
   b) Объем запаса некоторого изделия.
   c) Объем заказа на некоторый товар, внесенный в инвентарную опись.
   d) Количество бракованных изделий в партии.
   e) Время, необходимое для оценки результатов теста.
   f) Количество автомобилей на стоянке агентства по прокату автомобилей.
4. Следующая таблица представляет изменение числа ожидающих в очереди клиентов в зависимости от времени имитации.

Время имитации T (часы)	Число ожидающих клиентов
$0 \le T \le 3$	0
$3 < T \le 4$	1
$4 < T \le 6$	2
$6 < T \le 7$	1
$7 < T \le 10$	0
$10 < T \le 12$	2
$12 < T \le 18$	3
$18 < T \le 20$	2
$20 < T \le 25$	1

Определите следующие величины.

a) Среднюю длину очереди.
b) Среднее время ожидания в очереди.

5. Предположим, что в парикмахерской из примера 18.6–1 работают три парикмахера; их занятость парикмахеров характеризуется следующей таблицей.

Время имитации T (часы)	Число занятых парикмахеров
$0 < T \leq 10$	0
$10 < T \leq 20$	1
$20 < T \leq 30$	2
$30 < T \leq 40$	0
$40 < T \leq 60$	1
$60 < T \leq 70$	2
$70 < T \leq 80$	3
$80 < T \leq 90$	1
$90 < T \leq 100$	0

Определите следующие величины.

a) Среднее число работающих парикмахеров.
b) Среднее время работы парикмахеров.
c) Среднее время простоя парикмахеров.

18.7. Методы сбора статистических данных

Имитационное моделирование представляет собой статистический эксперимент. Его результаты должны основываться на соответствующих статистических проверках (с использованием, например, доверительных интервалов и методов проверки гипотез). Для выполнения этой задачи получаемые наблюдения и имитационный эксперимент должны удовлетворять следующим трем требованиям.

1. Наблюдения имеют стационарные распределения, т.е. распределения не изменяются во время проведения эксперимента.
2. Наблюдения подчиняются нормальному распределению.
3. Наблюдения независимы.

Иногда на практике результаты имитационного моделирования не удовлетворяют ни одному из этих требований. Тем не менее, их выполнение гарантирует наличие корректных способов сбора наблюдений над имитационной моделью.

Рассмотрим сначала вопрос о стационарности распределений (первое требование). В примере 18.2–1 мы видели, что точность оценки площади круга методом Монте–Карло возрастает с увеличением объема выборки. Действительно, графики на рис. 18.2 показывают, что оценка площади неустойчива для выборки малого объема, так как результаты имитации не являются стационарными. Результаты стабилизируются в области, близкой к точному значению площади круга, при увеличении объема выборки. Следовательно, нестационарность процесса имитации можно преодолеть путем использования выборки достаточно большого объема.

Результаты наблюдений над моделью зависят от продолжительности периода имитации. Начальный период неустойчивого поведения модели (системы) обычно называется **переходным**. Когда результаты имитационного эксперимента стабилизируются, говорят, что система работает в **установившемся режиме**. Продолжительность переходного периода определяется в значительной степени начальными характеристиками модели и невозможно предсказать, когда наступит установившийся режим. В общем случае, чем длиннее продолжительность прогона модели, тем выше шанс достижения установившегося состояния.

Рассмотрим теперь второе требование, состоящее в том, что наблюдения над имитационной моделью должны иметь нормальное распределение. Это требование можно выполнить, если привлечь *центральную предельную теорему* (см. раздел 12.5.4), утверждающую, что распределение среднего выборки является асимптотически нормальным независимо от распределения генеральной совокупности, из которой взята выборка. Центральная предельная теорема, таким образом, есть главное средство удовлетворения требования о нормальности распределения.

Третье требование касается независимости наблюдений. Природа имитационного эксперимента не гарантирует независимости между последовательными наблюдениями над моделью. Однако использование выборочных средних для представления отдельных наблюдений позволяет смягчить проблему, связанную с отсутствием независимости. Для этого, в частности, следует увеличивать интервал времени имитации для получения выборочного среднего.

Понятия переходного и установившегося состояний имеют силу в ситуациях, именуемых **незаканчивающейся имитацией**, т.е. имитацией, применяемой к системам, которые функционируют бесконечно долго. В случае **заканчивающейся имитации** (например. работа банка, если он обычно работает восемь часов в день) переходное поведение является частью нормального функционирования системы и, следовательно, не может игнорироваться. Единственным выходом в такой ситуации является увеличение, насколько это возможно, числа наблюдений.

Обсудив "подводные камни" имитационного эксперимента и средства, с помощью которых их можно обойти, рассмотрим теперь три наиболее общих метода сбора информации в процессе имитационного моделирования:

1) метод подынтервалов,

2) метод повторения,

3) метод циклов.

18.7.1. Метод подынтервалов

На рис. 18.8 проиллюстрирована идея метода подынтервалов. Предположим, что имитация длится на протяжении T единиц времени (т.е. длина прогона модели равна T) и требуется получить n наблюдений. В соответствии с методом подынтервалов необходимо сначала "обрезать" информацию, относящуюся к переходному периоду, а затем разделить остаток результатов имитации на n равных подынтервалов (групп). Среднее значение искомой величины (например, длины очереди или времени ожидания в очереди) внутри каждого подынтервала используется затем в качестве единственного наблюдения. Отбрасывание начального переходного периода означает, что статистические данные, собранные на протяжении этого периода, не используются.

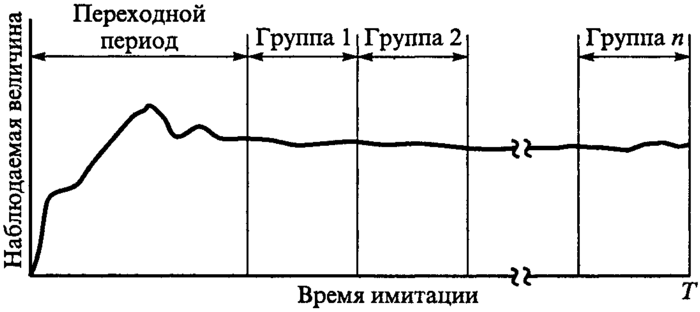


Рис. 18.8

Преимущество данного метода состоит в том, что влияние переходных (нестационарных) условий уменьшается, в частности, на те данные, которые собраны в конце времени имитации. Недостаток заключается в том, что последовательные группы с общей границей являются коррелированными, что приводит к невыполнению предположения о независимости. Влияние корреляции может быть уменьшено путем увеличения интервала времени для каждой группы.

Пример 18.7–1

На рис. 18.9 показано изменение длины очереди в системе обслуживания с одним сервисом (модель простой очереди) как функции времени. Период имитации составляет $T = 35$ часов, а длина переходного периода оценивается в 5 часов. Необходимо получить 5 наблюдений, т.е. $n = 5$. Соответствующая длина интервала времени для каждой группы равна $(35 - 5)/5 = 6$ часов.

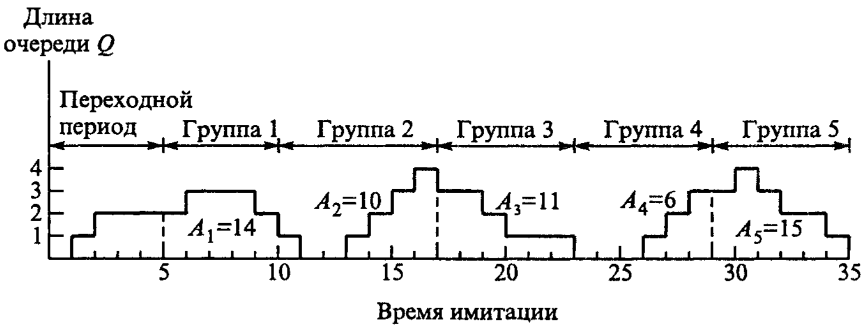


Рис. 18.9

Пусть $\bar{Q}_i$ представляет среднюю длину очереди в группе i. Так как длина очереди является переменной, зависящей от времени, то

$$\bar{Q}_i = \frac{A_i}{t}, \quad i = 1, 2, \ldots, 5,$$

где A_i — площадь под кривой длины очереди, t — длина интервала времени для группы. В рассматриваемом примере $t = 6$ часов.

Анализ данных, приведенных на рис. 18.9, приводит к следующей таблице.

Наблюдение i	1	2	3	4	5
A_i	14	10	11	6	15
$\bar{Q}_i$	2.33	1.67	1.83	1.00	2.5
Выборочное среднее = 3.32			Выборочная дисперсия = 0.35		

Выборочные среднее и дисперсию можно использовать, если это необходимо, для нахождения доверительного интервала.

Вычисление выборочной дисперсии в этом примере основано на использовании следующей хорошо известной формулы:

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

Эта формула является лишь приближением точного значения дисперсии, так как не учитывает эффекта автокорреляции между последовательными группами. Точную формулу можно найти в работе [2].

18.7.2. Метод повторения

В данном методе каждое наблюдение представляется независимым прогоном модели, в котором переходной период не учитывается, как показано на рис. 18.10. Вычисление средних величин выборки для каждой группы проводится точно так, как и в методе подынтервалов. Единственное отличие в том, что в данном случае стандартная формула для дисперсии применима, так как группы не коррелированы между собой.

Преимуществом этого метода является то, что каждый имитационный прогон модели определяется своей последовательностью случайных чисел из интервала [0, 1], что действительно обеспечивает статистическую независимость получаемых наблюдений. Недостаток состоит в том, что все наблюдения могут оказаться под сильным влиянием начальных переходных условий. Этот недостаток можно смягчить в результате увеличения длины прогона модели.

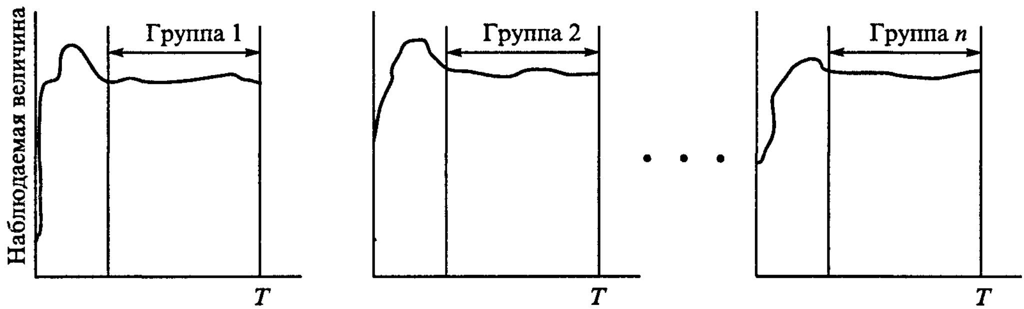


Рис. 18.10

18.7.3. Метод циклов

Этот метод можно рассматривать как расширенный вариант метода подынтервалов. Мотивацией данного метода является попытка уменьшить влияние автокорреляции, которая характерна для метода подынтервалов, путем выбора групп таким образом, чтобы обеспечить одинаковые начальные условия для каждой из них. Например, если в качестве переменной рассматривается длина очереди, то каждая группа должна начинаться в тот момент, когда длина очереди равна нулю. В отличие от метода подынтервалов, в методе циклов длины интервалов каждой группы могут оказаться различными.

Хотя метод циклов и позволяет уменьшить влияние автокорреляции, его недостатком является меньшее, по сравнению с методом подынтервалов, число получаемых наблюдений при заданной длине прогона модели. Это следует из того, что нельзя заранее сказать, когда новая группа (цикл) начинается и какова продолжительность каждого цикла. Однако можно ожидать, что в стационарных условиях начальные точки последовательных циклов будут расположены более или менее равномерно.

Вычисление среднего для цикла i в рассматриваемом методе определяется в виде отношения двух случайных величин a_i и b_i, т.е. в виде $x_i = a_i/b_i$. Определение величин a_i и b_i зависит от вычисляемой переменной. Например, если переменная является *функцией времени*, то a_i представляет площадь под кривой, а b_i — длительность соответствующего интервала времени. Если же переменная является *функцией количества событий*, то a_i — общая сумма наблюдений этой величины в пределах цикла i, а b_i — общее число событий внутри соответствующего цикла.

Так как x_i является отношением двух случайных величин, можно показать, что в данном случае несмещенная оценка выборочного среднего определяется формулой

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

где

$$y_i = \frac{n\overline{a}}{\overline{b}} - \frac{(n-1)(n\overline{a} - a_i)}{n\overline{b} - b_i}, \quad i = 1, 2, \ldots, n,$$

$$\overline{a} = \frac{\sum_{i=1}^{n} a_i}{n},$$

$$\overline{b} = \frac{\sum_{i=1}^{n} b_i}{n}.$$

В этом случае доверительный интервал для математического ожидания можно найти с помощью выборочного среднего $\overline{y}$ и стандартного отклонения величин y_i.

Пример 18.7–2

На рис. 18.11 показано число занятых обслуживающих устройств в одноканальной системе обслуживания с тремя параллельными обслуживающими устройствами. Длина периода имитации равна 35 единиц времени, а длина переходного периода — 4 единицы времени. Требуется оценить среднее значение использования сервисов методом циклов.

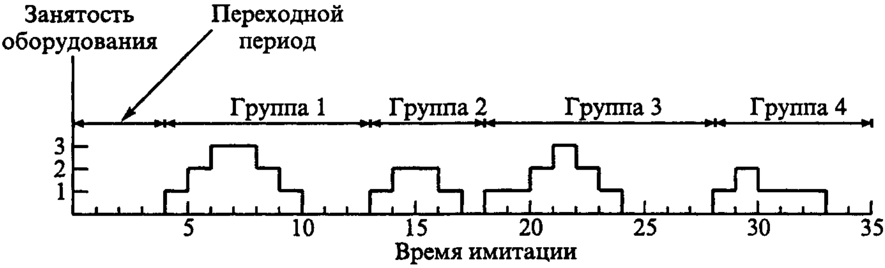


Рис. 18.11

После отбрасывания переходного периода получаем четыре цикла, общей характеристикой начала каждого из которых является незанятость всех трех обслуживающих устройств. Результаты вычислений приведены в следующей таблице.

Цикл i	a_i	b_i	y_i
1	12	9	1.32
2	6	5	1.24
3	10	6	1.65
4	6	5	1.24
	$\bar{a} = 8.5$	$\bar{b} = 6.25$	$\bar{y} = 1.363,\ s_y = 0.195$

Вычисление y_i проводится в соответствии со следующей формулой:

$$y_i = \frac{4\times 8.5}{6.25} - \frac{(4-1)(4\times 8.5 - a_i)}{4\times 6.25 - b_i} = 5.44 - \frac{3(34-a_i)}{25-b_i}.$$

Упражнения 18.7,а

1. В примере 18.7–1 используйте метод подынтервалов для вычисления среднего времени ожидания в очереди для тех клиентов, которые вынуждены ожидать.
2. Пусть в имитационной модели используется метод подынтервалов для вычисления средних величин в циклах. Переходной период равен 100 единиц времени, длина каждого цикла также равняется 100 единиц времени. Используя приведенные ниже данные, которые представляют время ожидания клиентов, как функцию времени имитации, определите 95%-й доверительный интервал для среднего времени ожидания.

Интервал времени	Времена ожидания
0 – 100	10, 20, 13, 14, 8, 15, 6, 8
100 – 200	12, 30, 10, 14, 16
200 – 300	15, 17, 20, 22
300 – 400	10, 20, 30, 15, 25, 31
400 – 500	15, 17, 20, 14, 13
500 – 600	25, 30, 15

3. Пусть в условиях примера 18.7–2 начальные точки циклов совпадают с теми моментами времени, когда все три обслуживающих устройства становятся незанятыми. На рис. 18.11 эти точки соответствуют моментам времени $t = 10$, 17, 24 и 33. Определите 95%-й доверительный интервал для занятости обслуживающих устройств, основываясь на новом определении точек начала циклов.

4. Для системы обслуживания с одним обслуживающим устройством проводится имитация ее работы на протяжении 100 часов. Результаты имитации показывают, что обслуживающее устройство было занято на протяжении таких интервалов времени: (0, 10), (15, 20), (25, 30), (35, 60), (70, 80) и (90, 95). Остальное время из интервала имитации обслуживающее устройство было свободно. Длина переходного периода равна 10 часов.

   a) Определите начальные точки, необходимые для применения метода циклов.
   b) Методом циклов определите 95%-й доверительный интервал для времени занятости обслуживающего устройства.
   c) С помощью метода подынтервалов решите ту же задачу при числе выборок $n = 5$. Определите соответствующий 95%-й доверительный интервал и сравните его с результатом, полученным методом циклов.

18.8. Языки имитационного моделирования

Реализация имитационных моделей влечет за собой два различных типа вычислений: 1) манипуляции регистрацией, которые имеют дело с хронологическим накоплением и обработкой событий модели и 2) вычисления, связанные с генерированием случайных чисел и сбором статистических данных, относящихся к модели. Вычисления первого типа основываются на различных логических методах обработки списков, а вычисления второго типа обычно очень громоздки и занимают много времени. Природа этих вычислений делает компьютер важным инструментом в реализации имитационных моделей и, в свою очередь, стимулирует создание специализированных языков программирования, что позволяет выполнять эти вычисления более удобным и эффективным способом.

Доступные языки дискретного имитационного моделирования делятся на две большие категории.

1. Языки, ориентированные на планирование событий.
2. Языки, ориентированные на обработку процессов (процедур).

При использовании языков, ориентированных на планирование событий, пользователю необходимо указать действия, связанные с каждым событием, происходящим в системе, аналогично тому, как они были представлены в примере 18.6–1. Основная роль программы в этом случае сводится к автоматизации процесса получения случайных значений, имеющих соответствующее распределение, хронологическому накоплению, обработке событий и сбору данных, относящихся к модели.

Процедурно-ориентированные языки используют блоки (или узлы), которые можно соединять для формирования сети, которая описывает движение **транзакций** или **объектов** (т.е. клиентов) в системе. Например, наиболее известными типами узлов в любом языке имитационного моделирования являются *источник*, в котором транзакции создаются, *оче-*

редь, где при необходимости они могут ожидать, и *сервисы*, где выполняется обслуживание. Каждый из этих узлов при его определении обеспечивается всей необходимой информацией, позволяющей выполнять имитацию автоматически. Например, если время между поступлениями заказов для источника определено, процедурно-ориентированный язык автоматически "знает", когда будут иметь место события, связанные с прибытием заказа. В действительности каждый узел имеет установленные инструкции, т.е. точно определяет *как* и *когда* транзакции перемещаются по имитационной сети.

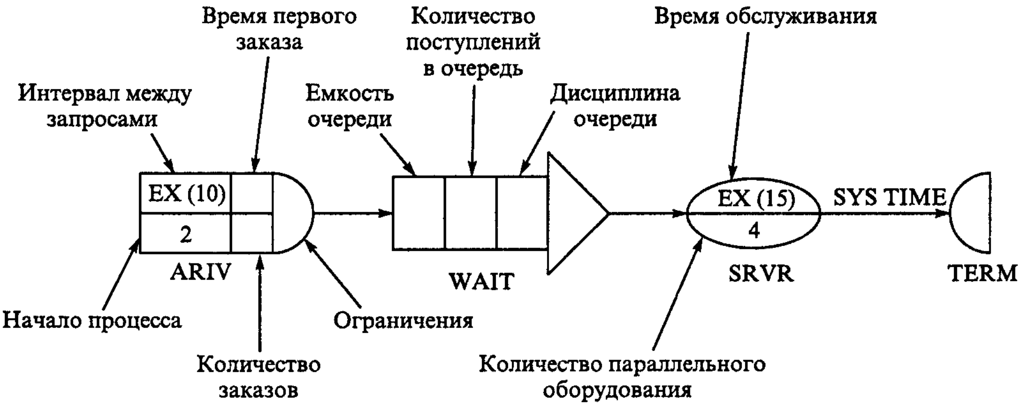
Процедурно-ориентированные языки управляются теми же действиями, что и языки, ориентированные на планирование событий. Отличие состоит в том, что эти действия автоматизированы для освобождения пользователя от утомительных вычислительных и логических деталей. В некотором отношении можно рассматривать процедурно-ориентированные языки как основанные на концепции "черного ящика", имеющего заданные "вход" и "выход". Это означает, что процедурно-ориентированные языки просты и легки в использовании благодаря гибкости процесса моделирования.

Наиболее известными языками программирования, ориентированными на планирование событий, являются SIMSCRIPT, SLAM и SIMAN. Развитие этих языков на протяжении многих лет привело к тому, что они включают и возможности процедурно-ориентированных языков. Все эти языки позволяют пользователю создавать модели (или их отдельные части) на языках высокого уровня, таких как FORTRAN и C. Это необходимо для того, чтобы дать возможность пользователю программировать сложные логические операции, которые невозможно или трудно осуществить обычными средствами этих языков. Главной причиной этого является ограничительная и, возможно, запутанная процедура, с помощью которой данные языки перемещают транзакции (или объекты) между очередью и сервисами, присутствующими в модели.

Первым процедурно-ориентированным языком был GPSS. Этот язык, первая версия которого появилась в начале 60-х годов, совершенствовался на протяжении нескольких лет, чтобы удовлетворить новым требованиям, связанным с моделированием сложных систем. Чтобы эффективно использовать этот язык, пользователю необходимо настроить примерно восемьдесят различных блоков. Несмотря на многие годы использования GPSS, язык все еще имеет некоторые трудно объяснимые особенности моделирования. Примером может служить необходимость аппроксимировать непрерывные распределения вероятностей их кусочно-линейными аналогами. Справедливости ради заметим, что некоторые последние версии этого языка обеспечивают возможности использования и непрерывных распределений (например, экспоненциального и нормального). Однако при имеющихся громадных возможностях современных компьютеров трудно понять, почему такое препятствие продолжает так долго существовать.

Другой процедурно-ориентированный язык программирования, именуемый SIMNET II, разработан для непосредственного моделирования сложных ситуаций. SIMNET II использует три типа узлов: *источник*, который генерирует заказы, *очередь*, где заказы могут ожидать, и *средство обслуживания*, где выполняется обслуживание. В этом языке имеется еще четвертый *дополнительный* тип узла, который рассматривается как средство обслуживания неограниченной емкости, что призвано усилить моделирующие возможности языка.

На рис. 18.12 показана схема применения SIMNET II к модели с одним обслуживающим устройством. Каждому узлу дается произвольное имя, после которого ставится символ *S, *Q, *F или *A, означающий, соответственно, что узел является источником, очередью, средством обслуживания или дополнительным. Транзакции поступают из источника ARIV, если необходимо, входят в узел-очередь WAIT и обслуживаются в узле средства обслуживания SRVR перед тем, как оставить систему через узел TERM. Время между поступлениями транзакций равно EX(10), т.е. имеет экспоненциальное распределение с математическим ожиданием 10 единиц времени. Аналогичная интерпретация обозначения EX(15) как времени обслуживания в узле SRVR. Перед тем как транзакция оставляет систему, подсчитывается время ее пребывания в системе на дуге (*B), ведущей в узел TERM. Имитация выполняется при значении параметра RUN-LENGTH (Длина выполнения), равного 480 единиц времени. Приложение Б содержит более подробную информацию о языке SIMNET II. Версия этого языка программирования для студентов содержится на Web-странице данной книги по адресу: www.williamspublishing.com.




```
$PROJECT;MODEL MM1;2/18/86;TAHA:
$DIMENSION;ENTITY(80),A(2):
$VARIABLES:SYS TIME;OBS.BASED;TRANSIT(2):
$BEGIN:
    ARVL    *S;EX(10);;2:
    WAIT    *Q:
    SRVR    *F;;EX(15);4:
           *B;TERM:/4/SYS TIME%:
$END:
$RUN-LENGTH=480:
$TRANSIT-PERIOD=0:
$TRACE=0-100:
$STOP:
```

Рис. 18.12

Упражнения 18.9,а[1]

1. Клиенты случайным образом прибывают на почтовое отделение, в котором работают три служащих. Время между их приходами является случайной величиной, распределенной по экспоненциальному закону с математическим ожиданием 5

[1] Решите эти упражнения, используя язык имитационного моделирования SIMNET II (Приложение Б), другие языки моделирования либо языки BASIC, FORTRAN или C.

минут. Время, затрачиваемое служащим на обслуживание клиента, имеет экспоненциальное распределение с математическим ожиданием 10 минут. Все прибывающие клиенты формируют единую очередь и ожидают первого освободившегося служащего. Используйте имитационную модель системы для исследования ее работы на протяжении 480 минут, чтобы определить

a) среднее число клиентов, ожидающих в очереди;
b) среднюю занятость служащих.

2. Телевизионные блоки прибывают для проверки на ленточный конвейер с постоянной скоростью пять единиц в час. Время проверки блока является случайной величиной, равномерно распределенной между 10 и 15 минутами. Предыдущий опыт показывает, что 20% проверенных блоков должны быть отрегулированы и отправлены на повторную проверку. Время регулировки также является случайной величиной, равномерно распределенной между 6 и 8 минутами. Используйте имитационную модель системы для исследования ее работы на протяжении 480 минут, чтобы вычислить

a) среднее время, необходимое для проверки одного блока;
b) среднее количество повторных проверок, которые должен пройти телевизионный блок перед выходом из системы.

3. Мышь находится в лабиринте и отчаянно пытается из него выбраться. Проведя в попытках выбраться от 1 до 3 минут с равномерным распределением, в 30% случаев она находит выход из лабиринта. В случае неудачи она бродит бесцельно от 2 до 3 минут с равномерным распределением и, в конце концов, останавливается в месте, откуда она начинала, но лишь для того, чтобы попробовать еще раз. Мышь может пытаться выбраться на свободу столько раз, сколько она хочет, но всему есть предел. Истратив много энергии на многочисленные попытки выбраться на свободу, мышь непременно умрет, если не успеет выбраться за период времени, имеющий нормальное распределение с математическим ожиданием 10 минут и стандартным отклонением 2 минуты. Постройте имитационную модель для оценки вероятности того, что мышь будет свободной. С этой целью предположите, что в модели будет использовано 100 мышей.

4. На финальной стадии сборки автомобиль движется по конвейеру между двумя параллельными рабочими местами, чтобы можно было выполнять работы как с левой, так и с правой стороны одновременно. Время выполнения работ с каждой стороны является равномерно распределенной случайной величиной с интервалом изменения от 15 до 20 минут и от 18 до 22 минут соответственно. Конвейер прибывает на рабочие места каждые 20 минут. Постройте имитационную модель работы сборочного конвейера на протяжении 480 минут для определения времени использования левого и правого рабочих мест.

5. Время между прибытием машин на пункт мойки с одним обслуживающим устройством является случайной величиной, распределенной по экспоненциальному закону с математическим ожиданием 10 минут. Автомобили выстраиваются в одну очередь, которая может поместить максимум пять ожидающих автомобилей. Если очередь заполнена, вновь прибывший автомобиль уезжает. Время мойки

машины является случайной величиной, равномерно распределенной на интервале от 10 до 15 минут. Постройте имитационную модель работы системы на протяжении 960 минут и найдите время пребывания автомобиля на пункте мойки.

18.9. Заключение

Имитационное моделирование представляет собой достаточно гибкий инструмент исследования, который можно эффективно использовать при анализе сложных систем. Его недостаток состоит в том, что любой результат, полученный при имитационном моделировании, подвержен экспериментальным ошибкам и, следовательно, должен интерпретироваться через призму надлежащих статистических тестов. В силу специфической природы имитационного моделирования задача получения наблюдений, которые являются одновременно независимыми и репрезентативными в стационарных условиях, достаточно трудна. Тем не менее, использование специальных методик сбора данных позволяет смягчить эти трудности.

Все языки программирования, ориентированные на решение задач имитационного моделирования, оперируют понятием события при построении имитационной модели. Языки программирования, ориентированные на планирование дискретных событий, более сложны в использовании, чем процедурно-ориентированные языки.

Литература

Box G. and Muller M. "A Note on the Generation of Random Normal Deviates", *Annals of Mathematical Statistics,* Vol.29, pp.610–611, 1958.

Law A. and Kelton W. *Simulation Modeling & Analysis,* 2nd ed., McGraw–Hill, New York, 1991.

Ross S. *A Course of Simulation,* Macmillan, New York, 1990.

Taha A. *Simulation Modeling and SIMNET,* Prentice Hall, Upper Saddle River, N.J., 1988.

Литература, добавленная при переводе

Ермаков С.М. *Метод Монте–Карло и смежные вопросы.* — М.: Наука, 1975.

Нейлор Т. *Машинные имитационные эксперименты с моделями экономических систем.* — М.: Мир, 1978.

Соболь И.М. *Численные методы Монте–Карло.* — М.: Наука, 1973.

Шеннон Р. *Имитационное моделирование систем — искусство и наука.* — М.: Мир, 1978.

Марковские процессы принятия решений

19.1. Марковская задача принятия решений[1]

В этой главе рассмотрены приложения методов динамического программирования к решению стохастических задач, где процесс принятия решений можно представить конечным числом состояний. Переходные вероятности между состояниями описывают марковскую цепь. Структура вознаграждений в подобном процессе представима в виде матрицы, элементами которой являются величины дохода (или затраты), возникающие при переходе из одного состояния в другое. Матрица переходных вероятностей и матрица доходов зависят от альтернатив решения, которыми располагает лицо, принимающее решение. Целью задачи является нахождение оптимальной стратегии, максимизирующей ожидаемый доход от процесса, имеющего конечное или бесконечное число этапов.

Мы используем простой пример, который будет служить нам на протяжении всей главы. Несмотря на простоту, парафраза этого примера находит применение во многих приложениях в области управления запасами, замены оборудования, контроля и регулирования денежных потоков и др.

Каждый год в начале сезона садовник проводит химический анализ состояния почвы в своем саду. В зависимости от результатов анализа продуктивность сада на новый сезон оценивается как 1) хорошая, 2) удовлетворительная или 3) плохая.

В результате наблюдений на протяжении многих лет, садовник заметил, что продуктивность в текущем году зависит только от состояния почвы в предыдущем году. Поэтому вероятности перехода почвы из одного состояния продуктивности в другое для каждого года можно представить как вероятности перехода в следующей цепи Маркова.

$$\begin{array}{lcc}
 & & \overbrace{\begin{array}{ccc}1 & 2 & 3\end{array}}^{\text{Состояние системы в следующем году}} \\
\text{Состояние системы в текущем году} & \left\{\begin{array}{c}1\\2\\3\end{array}\right. & \begin{bmatrix}0.2 & 0.5 & 0.3\\ 0 & 0.5 & 0.5\\ 0 & 0 & 1\end{bmatrix} = \mathbf{P}^1
\end{array}$$

[1] Обзор теории цепей Маркова приведен в разделе 19.6.

Переходные вероятности в матрице $\mathbf{P}^1$ показывают, что продуктивность почвы в текущем году не лучше, чем в предыдущем. Например, если состояние почвы в текущем году удовлетворительное (состояние 2), то в следующем году оно может остаться удовлетворительным с вероятностью 0.5 или стать плохим (состояние 3) с той же вероятностью.

В результате различных агротехнических мероприятий садовник может изменить переходные вероятности $\mathbf{P}^1$. Обычно для повышения продуктивности почвы применяются удобрения. Эти мероприятия приводят к новой матрице переходных вероятностей $\mathbf{P}^2$.

$$\mathbf{P}^2 = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.05 & 0.4 & 0.55 \end{bmatrix} \end{array}.$$

Чтобы рассмотреть задачу принятия решений в перспективе, садовник связывает с переходом из одного состояния почвы в другое функцию дохода (или структуру вознаграждения), которая определяет прибыль или убыток за одногодичный период в зависимости от состояний, между которыми осуществляется переход. Так как садовник может принять решение использовать или не использовать удобрения, его доход или убыток будет изменяться в зависимости от принятого решения. Матрицы $\mathbf{R}^1$ и $\mathbf{R}^2$ определяют функции дохода (в сотнях долларов) и соответствуют матрицам переходных вероятностей $\mathbf{P}^1$ и $\mathbf{P}^2$.

$$\mathbf{R}^1 = \left\| r_{ij}^1 \right\| = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & -1 \end{bmatrix} \end{array},$$

$$\mathbf{R}^2 = \left\| r_{ij}^2 \right\| = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 6 & 5 & -1 \\ 7 & 4 & 0 \\ 6 & 3 & -2 \end{bmatrix} \end{array}.$$

Элементы r_{ij}^2 матрицы $\mathbf{R}^2$ учитывают затраты, связанные с применением удобрения. Например, если система находится в состоянии 1 и остается в этом состоянии и в следующем году, то доход составит $r_{11}^2 = 6$, если же удобрения не используются, $r_{11}^1 = 7$.

Какая же задача принятия решений стоит перед садовником? Сначала необходимо установить, будет ли деятельность садовника продолжаться ограниченное число лет или бесконечно. В соответствии с этим рассматриваются задачи принятия решения **с конечным и бесконечным числом этапов.** В обоих случаях, имея результаты химического анализа почвы (состояние системы), садовник должен выбрать *наилучшую* стратегию поведения (удобрять или не удобрять почву). При этом оптимальность принятого решения основывается на максимизации ожидаемого дохода.

Садовника может также интересовать оценка ожидаемого дохода при заранее определенной стратегии поведения в случае того или иного состояния системы. Например, он может принять решение всегда применять удобрения, если состояние почвы плохое (состояние 3). В таком случае говорят, что процесс принятия решений описывается **стационарной стратегией.**

Каждой стационарной стратегии соответствуют свои матрицы переходных вероятностей и доходов, которые можно построить на основе матриц $\mathbf{P}^1$, $\mathbf{P}^2$, $\mathbf{R}^1$ и $\mathbf{R}^2$. Например, для стационарной стратегии, требующей применения удобрения только тогда, когда состояние почвы плохое (состояние 3), результирующие матрицы переходных вероятностей и доходов задаются следующими выражениями.

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0.05 & 0.4 & 0.55 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 6 & 3 & -2 \end{bmatrix}.$$

Эти матрицы отличаются от $\mathbf{P}^1$ и $\mathbf{R}^1$ только третьей строкой, включенной в них из матриц $\mathbf{P}^2$ и $\mathbf{R}^2$. Причина этого в том, что $\mathbf{P}^2$ и $\mathbf{R}^2$ — матрицы, соответствующие случаю, когда удобрения применяются при *любом* состоянии почвы.

Упражнения 19.1,а

1. В задаче садовника определите матрицы **P** и **R**, соответствующие стационарной стратегии, требующей использования удобрений при удовлетворительном и плохом состояниях почвы.
2. Определите все стационарные стратегии в примере с садовником.

19.2. Модель динамического программирования с конечным числом этапов

Предположим, что садовник планирует прекратить занятие садоводством через N лет. В этом случае необходимо определить стратегию поведения (удобрять или не удобрять почву) для каждого года при конечном горизонте планирования. Очевидно, что оптимальной стратегией будет такая, при которой садовник получит наибольший ожидаемый доход через N лет.

Пусть $k = 1$ или 2 обозначает две возможные (альтернативные) стратегии поведения садовника. Матрицы $\mathbf{P}^k$ и $\mathbf{R}^k$, представляющие переходные вероятности и функцию дохода для альтернативы k, определены в разделе 19.1; здесь они приводятся для удобства.

$$\mathbf{P}^1 = \left\|p_{ij}^1\right\| = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R}^1 = \left\|r_{ij}^1\right\| = \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & -1 \end{bmatrix},$$

$$\mathbf{P}^2 = \left\|p_{ij}^2\right\| = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.05 & 0.4 & 0.55 \end{bmatrix}, \quad \mathbf{R}^2 = \left\|r_{ij}^2\right\| = \begin{bmatrix} 6 & 5 & -1 \\ 7 & 4 & 0 \\ 6 & 3 & -2 \end{bmatrix}.$$

Задачу садовника можно представить как задачу динамического программирования (ДП) с конечным числом этапов следующим образом. Пусть число состояний для каждого этапа (года) равно m (= 3 в примере с садовником). Обозначим через $f_n(i)$ оптимальный *ожидаемый* доход, полученный на этапах от n до N включительно при условии, что система находится в начале этапа n в состоянии i.

Обратное рекуррентное уравнение, связывающее f_n и f_{n+1}, можно записать в виде

$$f_n(i) = \max_k \left\{ \sum_{j=1}^{m} p_{ij}^k \left[r_{ij}^k + f_{n+1}(j) \right] \right\}, \quad n = 1, 2, \ldots, N,$$

где $f_{N+1}(j) \equiv 0$ для всех j.

Приведенное уравнение основано на том, что *накапливающийся* доход $r_{ij}^k + f_{n+1}(j)$ получается в результате перехода из состояния i на этапе n в состояние j на этапе $n+1$ с вероятностью p_{ij}^k. Введя обозначение

$$v_i^k = \sum_{j=1}^{m} p_{ij}^k r_{ij}^k,$$

рекуррентное уравнение ДП можно записать следующим образом:

$$f_N(i) = \max_k \left\{ v_i^k \right\},$$

$$f_n(i) = \max_k \left\{ v_i^k + \sum_{j=1}^{m} p_{ij}^k f_{n+1}(j) \right\}, \quad n = 1, 2, \ldots, N-1.$$

Проиллюстрируем использование рекуррентного уравнения для вычисления величин v_i^k в задаче садовника для случая, когда удобрения не применяются ($k = 1$).

$$v_1^1 = 0.2 \times 7 + 0.5 \times 6 + 0.3 \times 3 = 5.3,$$

$$v_2^1 = 0 \times 0 + 0.5 \times 5 + 0.5 \times 1 = 3,$$

$$v_3^1 = 0 \times 0 + 0 \times 0 + 1 \times (-1) = -1.$$

Эти значения показывают, что если состояние почвы в начале года оказывается хорошим (состояние 1), то при одном переходе ожидаемый годовой доход составляет 5.3. Аналогично, если состояние почвы удовлетворительное, ожидаемый годовой доход составит 3, а в случае плохого будет равен –1.

Пример 19.2–1

В этом примере задача садовника решается при данных, заданных матрицами $\mathbf{P}^1$, $\mathbf{P}^2$, $\mathbf{R}^1$ и $\mathbf{R}^2$. Предполагается, что горизонт планирования включает 3 года ($N = 3$).

Так как значения v_i^k многократно используются в вычислениях, для удобства они сведены в таблицу. Напомним, что значение $k = 1$ соответствует решению "не удобрять", $k = 2$ — "удобрять".

i	v_i^1	v_i^2
1	5.3	4.7
2	3	3.1
3	–1	0.4

Этап 3

	v_i^k		Оптимальное решение	
i	$k = 1$	$k = 2$	$f_3(i)$	k^*
1	5.3	4.7	5.3	1
2	3	3.1	3.1	2
3	–1	0.4	0.4	2

Этап 2

	$v_i^k + p_{i1}^k f_3(1) + p_{i2}^k f_3(2) + p_{i3}^k f_3(3)$		Оптимальное решение	
i	$k = 1$	$k = 2$	$f_2(i)$	k^*
1	$5.3 + 0.2 \times 5.3 + 0.5 \times 3.1 + 0.3 \times 0.4 = 8.03$	$4.7 + 0.3 \times 5.3 + 0.6 \times 3.1 + 0.1 \times 0.4 = 8.19$	8.19	2
2	$3 + 0 \times 5.3 + 0.5 \times 3.1 + 0.5 \times 0.4 = 4.75$	$3.1 + 0.1 \times 5.3 + 0.6 \times 3.1 + 0.3 \times 0.4 = 5.61$	5.61	2
3	$-1 + 0 \times 5.3 + 0 \times 3.1 + 1 \times 0.4 = -0.6$	$0.4 + 0.05 \times 5.3 + 0.4 \times 3.1 + 0.55 \times 0.4 \approx 2.13$	2.13	2

Этап 1

	$v_i^k + p_{i1}^k f_2(1) + p_{i2}^k f_2(2) + p_{i3}^k f_2(3)$		Оптимальное решение	
i	$k = 1$	$k = 2$	$f_1(i)$	k^*
1	$5.3 + 0.2 \times 8.19 + 0.5 \times 5.61 + 0.3 \times 2.13 \approx 10.38$	$4.7 + 0.3 \times 8.19 + 0.6 \times 5.61 + 0.1 \times 2.13 \approx 10.74$	10.74	2
2	$3 + 0 \times 8.19 + 0.5 \times 5.61 + 0.5 \times 2.13 = 6.87$	$3.1 + 0.1 \times 8.19 + 0.6 \times 5.61 + 0.3 \times 2.13 \approx 7.92$	7.92	2
3	$-1 + 0 \times 8.19 + 0 \times 5.61 + 1 \times 2.13 = 1.13$	$0.4 + 0.05 \times 8.19 + 0.4 \times 5.61 + 0.55 \times 2.13 \approx 4.23$	4.23	2

Оптимальное решение показывает, что в 1-й и 2-й годы садовник должен применять удобрения ($k^* = 2$), независимо от состояния системы (состояния почвы по данным химического анализа). На 3-й год садовнику следует применять удобрения только тогда, когда система находится в состояниях 2 или 3 (т.е. при удовлетворительном или плохом состоянии почвы). Суммарный ожидаемый доход за три года составит $f_1(1) = 10.74$ при хорошем состоянии системы в 1-й год, $f_1(2) = 7.92$ — при удовлетворительном состоянии системы в 1-й год и $f_1(3) = 4.23$ — при плохом состоянии.

Задача при конечном горизонте планирования (решенная выше задача садовника) может быть обобщена в двух направлениях. Во-первых, переходные вероятности и функции дохода не обязательно должны быть одинаковы для каждого года. Во-вторых, можно использовать коэффициент переоценки (дисконтирования) ожидаемых доходов для последовательных этапов, вследствие чего значения $f_1(i)$ будут представлять собой *приведенные величины* ожидаемых доходов по всем этапам.

В первом случае значения доходов r_{ij}^k и переходные вероятности p_{ij}^k должны быть функциями этапа n. Здесь рекуррентное уравнение ДП принимает вид

$$f_N(i) = \max_k \left\{v_i^{k,N}\right\},$$

$$f_n(i) = \max_k \left\{v_i^{k,n} + \sum_{j=1}^{m} p_{ij}^{k,n} f_{n+1}(j)\right\}, \quad n = 1, 2, \dots, N-1,$$

где

$$v_i^{k,n} = \sum_{j=1}^{m} p_{ij}^{k,n} r_{ij}^{k,n}.$$

Второе обобщение заключается в следующем. Пусть α (< 1) — годовой коэффициент переоценки (дисконтирования), тогда D долларов будущего года равны αD долларам настоящего года. При введении коэффициента переоценки исходное рекуррентное уравнение преобразуется в следующее:

$$f_N(i) = \max_k \left\{v_i^k\right\},$$

$$f_n(i) = \max_k \left\{v_i^k + \alpha \sum_{j=1}^{m} p_{ij}^k f_{n+1}(j)\right\}, \quad n = 1, 2, \dots, N-1.$$

Упражнения 19.2,а

1. Фирма ежегодно оценивает положение со сбытом одного из видов своей основной продукции и дает ему удовлетворительную (состояние 1) или неудовлетворительную оценку (состояние 2). Необходимо принять решение о целесообразности рекламирования этой продукции в целях расширения ее сбыта. Приведенные ниже матрицы $\mathbf{P}^1$ и $\mathbf{P}^2$ определяют переходные вероятности при наличии рекламы и без нее в течение любого года. Соответствующие доходы заданы матрицами $\mathbf{R}^1$ и $\mathbf{R}^2$. Найдите оптимальные решения для последующих трех лет.

$$\mathbf{P}^1 = \begin{bmatrix} 0.9 & 0.1 \\ 0.6 & 0.4 \end{bmatrix}, \quad \mathbf{R}^1 = \begin{bmatrix} 2 & -1 \\ 1 & -3 \end{bmatrix},$$

$$\mathbf{P}^2 = \begin{bmatrix} 0.7 & 0.3 \\ 0.2 & 0.8 \end{bmatrix}, \quad \mathbf{R}^2 = \begin{bmatrix} 4 & 1 \\ 2 & -1 \end{bmatrix}.$$

2. Компания может провести рекламную акцию с помощью одного из трех средств массовой информации: радио, телевидения или газеты. Недельные затраты на рекламу с помощью этих средств оцениваются в 200, 900 и 300 долларов соответст-

венно. Компания оценивает недельный объем сбыта своей продукции по трехбалльной шкале как удовлетворительный (1), хороший (2) и отличный (3). Ниже указаны переходные вероятности, соответствующие каждому из трех средств массовой информации.

$$\text{Радио}\quad \begin{array}{c|c} & \begin{array}{ccc}1 & 2 & 3\end{array} \\ \begin{array}{c}1\\2\\3\end{array} & \begin{bmatrix}0.4 & 0.5 & 0.1\\0.1 & 0.7 & 0.2\\0.1 & 0.2 & 0.7\end{bmatrix}\end{array} \qquad \text{Телевидение}\quad \begin{array}{c|c} & \begin{array}{ccc}1 & 2 & 3\end{array} \\ \begin{array}{c}1\\2\\3\end{array} & \begin{bmatrix}0.7 & 0.2 & 0.1\\0.3 & 0.6 & 0.1\\0.1 & 0.7 & 0.2\end{bmatrix}\end{array} \qquad \text{Газета}\quad \begin{array}{c|c} & \begin{array}{ccc}1 & 2 & 3\end{array} \\ \begin{array}{c}1\\2\\3\end{array} & \begin{bmatrix}0.2 & 0.5 & 0.3\\0 & 0.7 & 0.3\\0 & 0.2 & 0.8\end{bmatrix}\end{array}$$

Соответствующие недельные доходы (в тысячах долларов) равны следующему.

$$\text{Радио}\ \begin{bmatrix}400 & 520 & 600\\300 & 400 & 700\\200 & 250 & 500\end{bmatrix} \quad \text{Телевидение}\ \begin{bmatrix}1000 & 1300 & 1600\\800 & 1000 & 1700\\600 & 700 & 1100\end{bmatrix} \quad \text{Газета}\ \begin{bmatrix}400 & 530 & 710\\350 & 450 & 800\\250 & 400 & 650\end{bmatrix}$$

Найдите оптимальную стратегию рекламы для последующих трех недель.

3. Фирма выпускает на рынок новый вид продукции. Если объем сбыта высокий, то с вероятностью 0.5 он останется таким же в следующем месяце. Если он невысокий, то вероятность того, что в следующем месяце он станет высоким, равна только 0.2. Фирма может провести рекламную кампанию. Если она примет это решение при высоком объеме сбыта, то вероятность того, что он останется высоким и в следующем месяце, возрастает до 0.8. При низком уровне реклама сбыта повышает эту вероятность только до 0.4.

   Если при высоком уровне сбыта реклама не используется, то ожидаемый доход составит 10 при условии, что объем сбыта останется высоким в следующем месяце, и 4 — в противном случае. Если первоначально наблюдается высокий уровень сбыта, то соответствующие доходы равны 7 и –2. При использовании рекламы доход равен 7, если первоначально уровень сбыта высокий, и становится равным 6, если уровень сбыта снижается. Если начальный уровень сбыта низкий, то доходы равны 3 и –5 в зависимости от того, повышается он или нет.

   Определите оптимальную стратегию фирмы для последующих трех месяцев.

4. *Задача управления запасами.* Магазин электротоваров в целях быстрого удовлетворения спроса покупателей на холодильники может размещать заказы в начале каждого месяца. Каждое размещение заказа приводит к постоянным затратам в 100 долларов. Затраты на хранение одного холодильника в течение месяца равны 5 долларов. Потери магазина при отсутствии холодильников оцениваются в 150 долларов за каждый холодильник в месяц. Месячный спрос на холодильники задается следующим распределением вероятностей.

Спрос x	0	1	2
$p(x)$	0.2	0.5	0.3

Магазин реализует следующую стратегию: максимальный уровень запаса не должен превышать двух холодильников в течение любого месяца.

a) Определите переходные вероятности при различных альтернативах решения этой задачи.
b) Определите ожидаемые месячные затраты на хранение запаса как функцию состояния системы и альтернативных решений.
c) Определите оптимальную стратегию размещения заказов на последующие 3 месяца.

5. Выполните задания предыдущего упражнения, предполагая, что плотности вероятностей спроса на следующий квартал определяются значениями из следующей таблицы.

	Месяц		
Спрос, x	1	2	3
0	0.1	0.3	0.2
1	0.4	0.5	0.4
2	0.5	0.2	0.4

6. Рыночная цена подержанного автомобиля составляет 2000 долларов. Владелец полагает, что он может получить больше этой суммы, но при этом он намерен согласиться с ценой, предложенной первыми тремя потенциальными покупателями, которые откликнулись на его объявление (это означает, что он должен принять решение не позже момента получения третьего предложения). Предположим, что с равными вероятностями будут предложены цены 2000, 2200, 2400 и 2600 долларов. Естественно, что после принятия какой-либо из предложенных цен все последующие предложения теряют смысл. Задача продавца заключается в том, чтобы установить пороговое значение цены, которым он будет пользоваться после получения первых трех предложений. В качестве такого значения можно выбрать 2000, 2200, 2400 и 2600 долларов. Найдите оптимальную стратегию владельца автомашины.

19.3. Модель с бесконечным числом этапов

Установившееся состояние марковского процесса характеризуется его независимостью от начального состояния системы. В этом случае говорят, что система достигла *установившегося состояния.* Поэтому нас интересует оценка стратегий, для которых соответствующие марковские цепи допускают существование установившегося решения. (В разделе 19.6 определены условия, при которых марковская цепь может иметь установившиеся состояния.)

Существует два метода решения задачи с бесконечным числом этапов. Первый метод основан на переборе *всех возможных* стационарных стратегий в задаче принятия решений. Этот подход, по существу, эквивалентен методу *полного перебора,* и его можно использовать только тогда, когда общее число стационарных стратегий с точки зрения практических вычислений достаточно мало.

Второй метод, называемый методом итераций по стратегиям, как правило, более эффективен, так как определяет оптимальную стратегию итерационным путем.

19.3.1. Метод полного перебора

Предположим, что в задаче принятия решений имеется S стационарных стратегий. Пусть $\mathbf{P}^s$ и $\mathbf{R}^s$ — матрицы переходных (одношаговых) вероятностей и доходов, соответствующие применяемой стратегии, $s = 1, 2, ..., S$. Метод перебора включает следующие шаги.

Шаг 1. Вычисляем v_i^s — ожидаемый доход, получаемый за один этап при стратегии s для заданного состояния i, $i = 1, 2, ..., m$.

Шаг 2. Вычисляем π_i^s — долгосрочные стационарные вероятности матрицы переходных вероятностей $\mathbf{P}^s$, соответствующие стратегии s. Эти вероятности (если они существуют) находятся из уравнений

$$\boldsymbol{\pi}^s\mathbf{P}^s = \boldsymbol{\pi}^s,$$

$$\pi_1^s + \pi_2^s + ... + \pi_m^s = 1,$$

где $\boldsymbol{\pi}^s = \left(\pi_1^s, \pi_2^s, ..., \pi_m^s\right)$.

Шаг 3. Вычисляем по следующей формуле E^s, ожидаемый доход за один шаг (этап) при выбранной стратегии s.

$$E^s = \sum_{i=1}^{m} \pi_i^s v_i^s.$$

Шаг 4. Оптимальная стратегия s^* определяется из условия, что

$$E^{s^*} = \max_s \left\{E^s\right\}.$$

Проиллюстрируем этот метод на примере задачи садовника в случае бесконечного горизонта планирования.

Пример 19.3–1

В задаче садовника имеется восемь стационарных стратегий, представленных в следующей таблице.

Стационарная стратегия, s	Действия
1	Не применять удобрения вообще
2	Применять удобрения независимо от состояния почвы
3	Применять удобрения, если почва находится в состоянии 1
4	Применять удобрения, если почва находится в состоянии 2
5	Применять удобрения, если почва находится в состоянии 3
6	Применять удобрения, если почва находится в состоянии 1 или 2
7	Применять удобрения, если почва находится в состоянии 1 или 3
8	Применять удобрения, если почва находится в состоянии 2 или 3

Матрицы $\mathbf{P}^s$ и $\mathbf{R}^s$ для стратегий от 3 до 8 получаются из аналогичных матриц для стратегий 1 и 2. Таким образом, имеем

$$\mathbf{P}^1 = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R}^1 = \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 0 & 0 & -1 \end{bmatrix},$$

$$\mathbf{P}^2 = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0.05 & 0.4 & 0.55 \end{bmatrix}, \quad \mathbf{R}^2 = \begin{bmatrix} 6 & 5 & -1 \\ 7 & 4 & 0 \\ 6 & 3 & -2 \end{bmatrix},$$

$$\mathbf{P}^3 = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R}^3 = \begin{bmatrix} 6 & 5 & -1 \\ 0 & 5 & 1 \\ 0 & 0 & -1 \end{bmatrix},$$

$$\mathbf{P}^4 = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R}^4 = \begin{bmatrix} 7 & 6 & 3 \\ 7 & 4 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

$$\mathbf{P}^5 = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0 & 0.5 & 0.5 \\ 0.05 & 0.4 & 0.55 \end{bmatrix}, \quad \mathbf{R}^5 = \begin{bmatrix} 7 & 6 & 3 \\ 0 & 5 & 1 \\ 6 & 3 & -2 \end{bmatrix},$$

$$\mathbf{P}^6 = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0.1 & 0.6 & 0.3 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{R}^6 = \begin{bmatrix} 6 & 5 & -1 \\ 7 & 4 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

$$\mathbf{P}^7 = \begin{bmatrix} 0.3 & 0.6 & 0.1 \\ 0 & 0.5 & 0.5 \\ 0.05 & 0.4 & 0.55 \end{bmatrix}, \quad \mathbf{R}^7 = \begin{bmatrix} 6 & 5 & -1 \\ 0 & 5 & 1 \\ 6 & 3 & -2 \end{bmatrix},$$

$$\mathbf{P}^8 = \begin{bmatrix} 0.2 & 0.5 & 0.3 \\ 0.1 & 0.6 & 0.3 \\ 0.05 & 0.4 & 0.55 \end{bmatrix}, \quad \mathbf{R}^8 = \begin{bmatrix} 7 & 6 & 3 \\ 7 & 4 & 0 \\ 6 & 3 & -2 \end{bmatrix}.$$

Результаты вычислений значений v_i^s приведены в следующей таблице.

	v_i^s		
s	$i=1$	$i=2$	$i=3$
1	5.3	3	−1
2	4.7	3.1	0.4
3	4.7	3	−1
4	5.3	3.1	−1
5	5.3	3	0.4
6	4.7	3.1	−1
7	4.7	3	0.4
8	5.3	3.1	0.4

Стационарные вероятности находятся из уравнений

$$\boldsymbol{\pi}^s \mathbf{P}^s = \boldsymbol{\pi}^s,$$

$$\pi_1 + \pi_2 + \ldots + \pi_m = 1.$$

Для иллюстрации применения этих уравнений рассмотрим стратегию $s = 2$. Соответствующие уравнения имеют следующий вид.

$$0.3\pi_1 + 0.1\pi_2 + 0.05\pi_3 = \pi_1,$$

$$0.6\pi_1 + 0.6\pi_2 + 0.4\pi_3 = \pi_2,$$

$$0.1\pi_1 + 0.3\pi_2 + 0.55\pi_3 = \pi_3,$$

$$\pi_1 + \pi_2 + \pi_3 = 1.$$

(Отметим, что одно из первых трех уравнений избыточно.) Решением системы будет

$$\pi_1^2 = \frac{6}{59}, \quad \pi_2^2 = \frac{31}{59}, \quad \pi_3^2 = \frac{22}{59}.$$

В данном случае ожидаемый годовой доход равен

$$E^2 = \sum_{i=1}^{3} \pi_i^2 v_i^2 = \frac{1}{59}\left(6\times 4.7 + 31\times 0.31 + 22\times 0.4\right) = 2.256.$$

Результаты вычислений π^s и E^s для всех стационарных стратегий приведены в следующей таблице. (Отметим, что хотя каждая из стратегий 1, 3, 4 и 6 имеет поглощающее состояние (состояние 3), это никоим образом не влияет на результаты вычислений. По этой причине $\pi_1 = \pi_2 = 0$ и $\pi_3 = 1$ для всех этих стратегий.)

s	π_1^s	π_2^s	π_3^s	E^s
1	0	0	1	–1.0
2	6/59	31/59	22/59	2.256
3	0	0	1	0.4
4	0	0	1	–1.0
5	5/154	69/154	80/154	1.724
6	0	0	1	–1.0
7	5/137	62/137	70/137	1.734
8	12/135	69/135	54/135	2.216

Из этой таблицы видно, что стратегия 2 дает наибольший ожидаемый годовой доход. Следовательно, оптимальная долгосрочная стратегия требует применения удобрений независимо от состояния системы.

Упражнения 19.3,а

1. Решите задачу из упр. 19.2,а(1) методом полного перебора при бесконечном числе этапов.
2. Решите задачу из упр. 19.2,а(2) при бесконечном горизонте планирования методом полного перебора.
3. Решите задачу из упр. 19.2,а(3) методом полного перебора, предполагая, что горизонт планирования бесконечен.

19.3.2. Метод итераций по стратегиям без дисконтирования

Чтобы оценить трудности, связанные с применением метода полного перебора, предположим, что у садовника вместо двух имеется четыре стратегии поведения (альтернативы): не удобрять, удобрять один раз в сезон, удобрять дважды и удобрять трижды в сезон. В этом случае общее число стратегий, имеющихся в распоряжении садовника, составляет $4^4 = 256$ стационарных стратегий. Таким образом, при увеличении числа альтернатив с 2 до 4 число стационарных стратегий возрастает по экспоненте с 8 до 256. Трудно не только перечислить в явном виде все эти стратегии, но и может оказаться также недопустимо большим объем вычислений, требуемых для оценивания всего множества стратегий.

Метод итераций по стратегиям основывается на следующем. Как показано в разделе 19.2, для любой конкретной стратегии ожидаемый суммарный доход за n-й этап определяется рекуррентным уравнением

$$f_n(i) = v_i + \sum_{j=1}^{m} p_{ij} f_{n+1}(j), \quad i = 1, 2, \dots, m.$$

Это уравнение и служит основой метода итераций по стратегиям. Однако, чтобы сделать возможным изучение асимптотического поведения процесса, вид уравнения нужно немного изменить. В отличие от величины n, которая фигурирует в уравнении и соответствует n-му этапу, обозначим через η число *оставшихся* для анализа этапов. Тогда рекуррентное уравнение записывается в виде

$$f_\eta(i) = v_i + \sum_{j=1}^{m} p_{ij} f_{\eta-1}(j), \quad i = 1, 2, \dots, m.$$

Здесь f_η — суммарный ожидаемый доход при условии, что остались не рассмотренными η этапов. При таком определении η можно изучить асимптотическое поведение процесса, полагая при этом, что $\eta \to \infty$.

Обозначим через $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_m)$ вектор установившихся вероятностей состояний с матрицей переходных вероятностей $\mathbf{P} = \|p_{ij}\|$ и пусть $E = \pi_1 v_1 + \pi_2 v_2 + \dots + \pi_m v_m$ — ожидаемый доход за этап, вычисленный по схеме раздела 19.3.1, тогда можно показать, что при достаточно большом η

$$f_\eta(i) = \eta E + f(i),$$

где $f(i)$ — постоянный член, описывающий асимптотическое поведение функции $f_\eta(i)$ при заданном состоянии i.

Так как $f_\eta(i)$ представляет суммарный оптимальный доход за η этапов при заданном состоянии i, а E — ожидаемый доход *за один этап*, то интуитивно понятно, почему величина $f_\eta(i)$ равна сумме ηE и поправочного числа $f(i)$, учитывающего определенное состояние i. При этом, конечно, предполагается, что число η достаточно велико.

Теперь рекуррентное уравнение можно записать в следующем виде

$$\eta E + f(i) = v_i + \sum_{j=1}^{m} p_{ij}\left\{(\eta-1)E + f(j)\right\},\ \ i = 1,2,\ldots,m.$$

Упростив это уравнение, получаем

$$E = v_i + \sum_{j=1}^{m} p_{ij} f(j) - f(i),\ \ i = 1,2,\ldots,m,$$

т.е. имеем m уравнений с $m+1$ неизвестными $f(1), f(2), \ldots, f(m)$ и E.

Здесь, как и в разделе 19.3.1, конечной целью является определение оптимальной стратегии, приводящей к максимальному значению E. Так как имеется m уравнений с $m+1$ неизвестными, оптимальное значение E нельзя определить за один шаг. В связи с этим используется итеративная процедура, начинающаяся с произвольной стратегии, а затем определяется новая стратегия, дающая лучшее значение E. Итеративный процесс заканчивается, если две последовательно получаемые стратегии совпадают.

Итеративный процесс состоит из двух основных шагов.

1. *Шаг оценивания параметров.* Выбираем произвольную стратегию s. Используя соответствующие ей матрицы $\mathbf{P}^s$ и $\mathbf{R}^s$ и полагая (произвольно) $f^s(m) = 0$, решаем уравнения

$$E^s = v_i^s + \sum_{j=1}^{m} p_{ij}^s f^s(j) - f^s(i),\ \ i = 1,2,\ldots,m,$$

   относительно неизвестных $E^s, f^s(1), \ldots, f^s(m-1)$. Переходим к следующему шагу.

2. *Шаг улучшения стратегии.* Для каждого состояния i определяем альтернативу k, обеспечивающую

$$\max_k \left\{ v_i^k + \sum_{j=1}^{m} p_{ij}^k f^s(j) \right\},\ \ i = 1,2,\ldots,m.$$

   (Здесь используются значения $f^s(j)$, $j = 1, 2, \ldots, m$, определенные на шаге оценивания параметров.) Результирующие оптимальные решения для состояний $1, 2, \ldots, m$ формируют новую стратегию t. Если s и t идентичны, то алгоритм заканчивается; в этом случае t — оптимальная стратегия. В противном случае полагаем $s = t$ и возвращаемся к шагу оценивания параметров.

Оптимизационная задача на шаге улучшения стратегии нуждается в пояснении. Целью этого шага является получение максимального значения E. Как показано выше,

$$E = v_i + \sum_{j=1}^{m} p_{ij} f(j) - f(i).$$

Поскольку $f(i)$ не зависит от альтернатив k, задача максимизации на шаге улучшения стратегии эквивалентна максимизации E по альтернативам k.

Пример 19.3–2

Решим задачу садовника методом итераций по стратегиям.

Начнем с произвольной стратегии, исключающей применение удобрений. Имеем соответствующие матрицы

$$\mathbf{P}=\begin{bmatrix}0.2 & 0.5 & 0.3\\ 0 & 0.5 & 0.5\\ 0 & 0 & 1\end{bmatrix},\quad \mathbf{R}=\begin{bmatrix}7 & 6 & 3\\ 0 & 5 & 1\\ 0 & 0 & -1\end{bmatrix}.$$

Уравнения шага оценивания параметров принимают вид

$$\begin{aligned}&E+f(1)-0.2f(1)-0.5f(2)-0.3f(3)=5.3,\\ &E+f(2)\qquad\qquad\;\; -0.5f(2)-0.5f(3)=3,\\ &E+f(3)\qquad\qquad\qquad\qquad\quad\;\; -f(3)=-1.\end{aligned}$$

Полагая $f(3) = 0$, получаем решение этих уравнений

$$E=-1,\ f(1)\approx 12.88,\ f(2)=8,\ f(3)=0.$$

Перейдем к шагу улучшения стратегии. Результаты соответствующих вычислений приведены в следующей таблице.

	$v_i^k+p_{i1}^k f(1)+p_{i2}^k f(2)+p_{i3}^k f(3)$		Оптимальное решение	
i	$k=1$	$k=2$	$f(i)$	k^*
1	$5.3+0.2\times 12.88+0.5\times 8+0.3\times 0=11.875$	$4.7+0.3\times 12.88+0.6\times 8+0.1\times 0=13.36$	13.36	2
2	$3+0\times 12.88+0.5\times 8+0.5\times 0=7$	$3.1+0.1\times 12.88+0.6\times 8+0.3\times 0=9.19$	9.19	2
3	$-1+0\times 12.88+0\times 8+1\times 0=-1$	$0.4+0.05\times 12.88+0.4\times 8+0.55\times 0=4.24$	4.24	2

Новая стратегия предусматривает применение удобрений независимо от состояния почвы. Так как новая стратегия отличается от предыдущей, повторяется шаг оценивания параметров. Новой стратегии соответствуют матрицы

$$\mathbf{P}=\begin{bmatrix}0.3 & 0.6 & 0.1\\ 0.1 & 0.6 & 0.3\\ 0.05 & 0.4 & 0.55\end{bmatrix},\quad \mathbf{R}=\begin{bmatrix}6 & 5 & -1\\ 7 & 4 & 0\\ 6 & 3 & -2\end{bmatrix}.$$

Эти матрицы определяют следующие уравнения:

$$\begin{aligned}&E+f(1)-0.3f(1)-0.6f(2)-0.1f(3)=4.7,\\ &E+f(2)-0.1f(1)-0.6f(2)-0.3f(3)=3.1,\\ &E+f(3)-0.05f(1)-0.4f(2)-0.55f(3)=0.4.\end{aligned}$$

Снова полагая $f(3) = 0$, получаем решение

$$E = 2.26,\ f(1) = 6.75,\ f(2) = 3.79,\ f(3) = 0.$$

Результаты вычислений на шаге улучшения стратегии приведены в следующей таблице.

	$v_i^k + p_{i1}^k f(1) + p_{i2}^k f(2) + p_{i3}^k f(3)$		Оптимальное решение	
i	$k=1$	$k=2$	$f(i)$	k^*
1	$5.3 + 0.2\times 6.75 + 0.5\times 3.79 + 0.3 \times 0 = 8.54$	$4.7 + 0.3\times 6.75 + 0.6\times 3.79 + 0.1 \times 0 = 8.99$	8.99	2
2	$3 + 0\times 6.75 + 0.5\times 3.79 + 0.5 \times 0 = 4.89$	$3.1 + 0.1\times 6.75 + 0.6\times 3.79 + 0.3 \times 0 = 6.05$	6.05	2
3	$-1 + 0\times 6.75 + 0\times 3.79 + 1 \times 0 = -1$	$0.4 + 0.05\times 6.75 + 0.4\times 3.79 + 0.55 \times 0 = 2.25$	2.25	2

Новая стратегия, требующая применения удобрений независимо от состояния системы, идентична предыдущей, поэтому последняя стратегия оптимальна и итеративный процесс заканчивается. Естественно, что этот результат совпадает с результатом, полученным методом полного перебора (раздел 19.3.1). Однако следует отметить, что метод итераций по стратегиям достаточно быстро сходится к оптимальному решению, что является его характерной особенностью.

Упражнения 19.3,b

1. Пусть в задаче упр. 19.2,a(1) горизонт планирования бесконечен. Решите эту задачу методом итераций по стратегиям.
2. Решите задачу упр. 19.2,a(2) методом итераций по стратегиям, предполагая, что горизонт планирования бесконечен.
3. Решите задачу упр. 19.2,a(3) методом итераций по стратегиям, предполагая, что горизонт планирования бесконечен.

19.3.3. Метод итераций по стратегиям с дисконтированием

Описанный метод итераций по стратегиям можно обобщить на случай дисконтирования. Если обозначить через α (< 1) коэффициент дисконтирования (переоценки), то рекуррентное уравнение при конечном числе этапов можно записать в следующем виде (см. раздел 19.2)

$$f_\eta(i) = \max_k \left\{ v_i^k + \alpha \sum_{j=1}^{m} p_{ij}^k f_{\eta-1}(j) \right\}.$$

(Отметим, что η равно числу этапов, которые необходимо рассмотреть.) Можно показать, что при $\eta \to \infty$ (модель с бесконечным числом этапов) $f_\eta(i) = f(i)$, где $f(i)$ — приведенный к текущему моменту времени дисконтированный доход при условии, что система находится в состоянии i и функционирует на бесконечном интервале времени. Таким образом, долгосрочное поведение $f_\eta(i)$ при $\eta \to \infty$ не зависит от значения η. В этом состоит отличие от случая без дисконтирования, когда $f_\eta(i) = \eta E + f(i)$. Этого следовало ожидать, так как в случае с дисконтированием влияние будущих доходов асимптотически уменьшается до нуля. Действительно, приведенный доход $f(i)$ должен стремиться к постоянной величине при $\eta \to \infty$.

С учетом вышеизложенного в данном случае при использовании метода итераций по стратегиям выполняются следующие действия.

1. *Шаг оценивания параметров.* Для произвольной стратегии s с матрицами $\mathbf{P}^s$ и $\mathbf{R}^s$ решаем систему из m уравнений

$$f^s(i) = v_i^s + \alpha \sum_{j=1}^{m} p_{ij}^s f^s(j), \quad i = 1, 2, \ldots, m,$$

относительно m неизвестных $f^s(1), f^s(2), \ldots, f^s(m)$.

2. *Шаг улучшения стратегии.* Для каждого состояния i определяем альтернативу k, обеспечивающую

$$\max_k \left\{ v_i^k + \alpha \sum_{j=1}^{m} p_{ij}^k f^s(j) \right\}, \quad i = 1, 2, \ldots, m,$$

где $f^s(j)$ имеют значения, определенные на шаге оценивания параметров. Если полученная стратегия t совпадает со стратегией s, то алгоритм закончен; в этом случае стратегия t оптимальна. В противном случае полагаем $s = t$ и повторяем шаг оценивания параметров.

Пример 19.3–3

Решим задачу из примера 19.3–2 с учетом коэффициента дисконтирования $\alpha = 0.6$.

Выберем произвольную стратегию, например $s = \{1, 1, 1\}$. Матрицы $\mathbf{P}$ и $\mathbf{R}$ ($\mathbf{P}^1$ и $\mathbf{R}^1$ в примере 19.3–1) определяют уравнения

$$\begin{aligned}
f(1) - 0.6[0.2f(1) + 0.5f(2) + 0.3f(3)] &= 5.3,\\
f(2) - 0.6[\qquad\qquad 0.5f(2) + 0.5f(3)] &= 3,\\
f(3) - 0.6[\qquad\qquad\qquad\qquad\quad + f(3)] &= -1.
\end{aligned}$$

Решение этих уравнений дает

$$f_1 \approx 6.6, f_2 \approx 3.21, f_3 = -2.5.$$

Результаты вычислений итерации по улучшению стратегии приведены в следующей таблице.

	$v_i^k + 0.6[p_{i1}^k f(1) + p_{i2}^k f(2) + p_{i3}^k f(3)]$		Оптимальное решение	
i	$k = 1$	$k = 2$	$f(i)$	k^*
1	$5.3 + 0.6[0.2 \times 6.6 + 0.5 \times 3.21 + 0.3 \times (-2.5)] = 6.61$	$4.7 + 0.6[0.3 \times 6.6 + 0.6 \times 3.21 + 0.1 \times (-2.5)] = 6.89$	6.89	2
2	$3 + 0.6[0 \times 6.6 + 0.5 \times 3.21 + 0.5 \times (-2.5)] = 3.21$	$3.1 + 0.6[0.1 \times 6.6 + 0.6 \times 3.21 + 0.3 \times (-2.5)] = 4.2$	4.2	2
3	$-1 + 0.6[0 \times 6.6 + 0 \times 3.21 + 1 \times (-2.5)] = -2.5$	$0.4 + 0.6[0.05 \times 6.6 + 0.4 \times 3.21 + 0.55 \times (-2.5)] = 0.54$	0.54	2

Шаг оценивания параметров, выполненный на основе матриц $\mathbf{P}^2$ и $\mathbf{R}^2$ (пример 19.3–1), приводит к следующим уравнениям.

$$f(1) - 0.6\left[0.3f(1) + 0.6f(2) + 0.1f(3)\right] = 4.7,$$

$$f(2) - 0.6\left[0.1f(1) + 0.6f(2) + 0.3f(3)\right] = 3.1,$$

$$f(3) - 0.6\left[0.05f(1) + 0.4f(2) + 0.55f(3)\right] = 0.4.$$

Решением этих уравнений будет

$$f_1 = 8.88, f_2 = 6.62, f_3 = 3.37.$$

Результаты, полученные на шаге улучшения стратегии, приведены в следующей таблице.

	$v_i^k + 0.6[p_{i1}^k f(1) + p_{i2}^k f(2) + p_{i3}^k f(3)]$		Оптимальное решение	
i	$k = 1$	$k = 2$	$f(i)$	k^*
1	$5.3 + 0.6[0.2 \times 8.88 + 0.5 \times 6.62 + 0.3 \times 3.37] = 8.95$	$4.7 + 0.6[0.3 \times 8.88 + 0.6 \times 6.62 + 0.1 \times 3.37] = 8.88$	8.95	1
2	$3 + 0.6[0 \times 8.88 + 0.5 \times 6.62 + 0.5 \times 3.37] = 5.99$	$3.1 + 0.6[0.1 \times 8.88 + 0.6 \times 6.62 + 0.3 \times 3.37] = 6.62$	6.62	2
3	$-1 + 0.6[0 \times 8.88 + 0 \times 6.62 + 1 \times 3.37] = 1.02$	$0.4 + 0.6[0.05 \times 8.88 + 0.4 \times 6.62 + 0.55 \times 3.37] = 3.37$	3.37	2

Так как новая стратегия $\{1, 2, 2\}$ отличается от предыдущей, повторяем шаг оценивания параметров с использованием матриц $\mathbf{P}^8$ и $\mathbf{R}^8$ (пример 19.3–1). Получаем следующие уравнения.

$$f(1) - 0.6\left[0.2f(1) + 0.5f(2) + 0.3f(3)\right] = 5.3,$$

$$f(2) - 0.6\left[0.1f(1) + 0.6f(2) + 0.3f(3)\right] = 3.1,$$

$$f(3) - 0.6\left[0.05f(1) + 0.4f(2) + 0.55f(3)\right] = 0.4.$$

Решением этих уравнений будет

$$f_1 = 8.98, f_2 = 6.63, f_3 = 3.38.$$

Результаты, полученные на шаге улучшения стратегии, приведены в следующей таблице.

	$v_i^k + 0.6[p_{i1}^k f(1) + p_{i2}^k f(2) + p_{i3}^k f(3)]$		Оптимальное решение	
i	$k=1$	$k=2$	$f(i)$	k^*
1	$5.3 + 0.6[0.2 \times 8.98 + 0.5 \times 6.63 + 0.3 \times 3.38] = 8.98$	$4.7 + 0.6[0.3 \times 8.98 + 0.6 \times 6.63 + 0.1 \times 3.38] = 8.91$	8.98	1
2	$3 + 0.6[0 \times 8.98 + 0.5 \times 6.63 + 0.5 \times 3.38] = 6.00$	$3.1 + 0.6[0.1 \times 8.98 + 0.6 \times 6.63 + 0.3 \times 3.38] = 6.63$	6.63	2
3	$-1 + 0.6[0 \times 8.98 + 0 \times 6.63 + 1 \times 3.38] = 1.03$	$0.4 + 0.6[0.05 \times 8.98 + 0.4 \times 6.63 + 0.55 \times 3.38] = 3.37$	3.37	2

Так как новая стратегия $\{1, 2, 2\}$ идентична предыдущей, то она оптимальна. Заметим, что при дисконтировании оптимальная стратегия исключает применение удобрений при хорошем состоянии системы (состояние 1).

Упражнение 19.3,с

1. Решите указанные ниже задачи, приняв коэффициент дисконтирования равным $\alpha = 0.9$
   a) Задача упр. 19.3,b(1).
   b) Задача упр. 19.3,b(2).
   c) Задача упр. 19.3,b(3).

19.4. Применение методов линейного программирования

Марковскую задачу принятия решений при бесконечном числе этапов как с дисконтированием, так и без него можно сформулировать и решить как задачу линейного программирования. Рассмотрим сначала случай без дисконтирования.

В разделе 19.3.1 было показано, что марковская задача без дисконтирования при бесконечном числе этапов сводится к нахождению оптимальной стратегии s^*, соответствующей

$$\max_{s \in S} \left\{ \sum_{j=1}^{m} \pi_i^s v_i^s \mid \boldsymbol{\pi}^s \mathbf{P}^s = \boldsymbol{\pi}^s,\ \ \pi_1^s + \pi_2^s + \ldots + \pi_m^s = 1,\ \ \pi_i^s \geq 0,\ \ i = 1, 2, \ldots, m \right\},$$

где S — множество всех возможных стратегий. Здесь π_i^s, $i = 1, 2, \ldots, m$, представляют установившиеся вероятности марковской цепи $\mathbf{P}^s$. Подобная задача была решена в разделе 19.3.1 полным перебором всех стратегий s.

Приведенная задача служит основой для формулировки марковской задачи принятия решений в виде задачи линейного программирования. Однако необходимо преобразовать переменные задачи таким образом, чтобы оптимальное решение *ав-*

томатически определяло оптимальное действие (альтернативу) k, когда система находится в состоянии i. Совокупность всех оптимальных действий определяет оптимальную стратегию s^*.

Это реализуется следующим образом. Введем обозначение: q_i^k — условная вероятность выбора альтернативы k, когда система находится в состоянии i. Тогда задачу можно представить в следующем виде.[1]

$$\text{Максимизировать } E = \sum_{i=1}^{m} \pi_i \left(\sum_{k=1}^{K} q_i^k v_i^k \right)$$

при ограничениях

$$\pi_j = \sum_{i=1}^{m} \pi_i p_{ij}, \quad j = 1, 2, \ldots, m,$$

$$\pi_1 + \pi_2 + \ldots + \pi_m = 1,$$

$$q_i^1 + q_i^2 + \ldots + q_i^K = 1, \quad i = 1, 2, \ldots, m,$$

$$\pi_i \geq 0, \quad q_i^k \geq 0 \text{ для всех } i \text{ и } k.$$

Заметим, что вероятности p_{ij} являются функциями выбранной стратегии и, следовательно, конкретных альтернатив k этой стратегии.

Ниже будет показано, что эту задачу можно преобразовать в задачу линейного программирования путем соответствующих подстановок, включающих q_i^k. Но следует отметить, что приведенная выше формулировка эквивалентна исходной формулировке задачи из раздела 19.3.1 только при условии, что $q_i^k = 1$ для *одного* k при каждом i, так как только при этом сумма $\sum_{k=1}^{K} q_i^k v_i^k$ будет сведена к $v_i^{k^*}$, где k^* — выбранная оптимальная альтернатива. Предлагаемая здесь линейная задача учитывает это условие автоматически.

Обозначим $w_{ik} = \pi_i q_i^k$ для всех i и k. По определению величина w_{ik} представляет собой *совместную* вероятность пребывания в состоянии i и принятия решения k. Из теории вероятностей известно, что

$$\pi_i = \sum_{k=1}^{K} w_{ik}.$$

Следовательно,

$$q_i^k = \frac{w_{ik}}{\sum_{k=1}^{K} w_{ik}}.$$

Поэтому очевидно, что ограничение $\sum_{i=1}^{m} \pi_i = 1$ можно записать в виде

$$\sum_{i=1}^{m} \sum_{k=1}^{K} w_{ik} = 1.$$

Ограничение $\sum_{k=1}^{K} q_i^k = 1$ также автоматически вытекает из способа определения q_i^k через w_{ik}. (Проверьте!) Таким образом, задачу можно записать в следующем виде.

---

[1] Здесь и далее K — количество возможных альтернатив. — *Прим. ред.*

$$\text{Максимизировать } E = \sum_{i=1}^{m}\sum_{k=1}^{K} v_i^k w_{ik}$$

при ограничениях

$$\sum_{k=1}^{K} w_{jk} - \sum_{i=1}^{m}\sum_{k=1}^{K} p_{ij}^k w_{ik} = 0,\ \ j = 1, 2, \ldots, m,$$

$$\sum_{i=1}^{m}\sum_{k=1}^{K} w_{ik} = 1,$$

$$w_{ik} \geq 0,\ \ i = 1, 2, \ldots, m;\ \ k = 1, 2, \ldots, K.$$

Сформулированная задача представляет собой задачу линейного программирования с переменными w_{ik}. Покажем, что ее оптимальное решение автоматически гарантирует, что $q_i^k = 1$ для одного k при любом i. Заметим, что в задаче линейного программирования имеется m независимых уравнений (одно уравнение, соответствующее $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, избыточно). Следовательно, задача должна включать m базисных переменных. Однако можно показать, что w_{ik} должно быть строго положительным по меньшей мере при одном k для каждого i. Из этих двух утверждений можно заключить, что величина

$$q_i^k = \frac{w_{ik}}{\sum_{k=1}^{K} w_{ik}}$$

может принимать только два значения (0 или 1), что и требовалось доказать. (Фактически полученный выше результат показывает также, что $\pi_i = \sum_{k=1}^{K} w_{ik} = w_{ik^*}$, где k^* – альтернатива, соответствующая $w_{ik} > 0$.)

Пример 19.4–1

Ниже приведена формулировка задачи садовника без дисконтирования в виде задачи линейного программирования

$$\text{максимизировать } E = 5.3w_{11} + 4.7w_{12} + 3w_{21} + 3.1w_{22} - w_{31} + 0.4w_{32}$$

при ограничениях

$$w_{11} + w_{12} - (0.2w_{11} + 0.3w_{12} + 0.1w_{22} + 0.05w_{32}) = 0,$$

$$w_{21} + w_{22} - (0.5w_{11} + 0.6w_{12} + 0.52w_{21} + 0.6w_{22} + 0.4w_{32}) = 0,$$

$$w_{31} + w_{32} - (0.3w_{11} + 0.1w_{12} + 0.5w_{21} + 0.3w_{22} + w_{31} + 0.55w_{32}) = 0,$$

$$w_{11} + w_{12} + w_{21} + w_{22} + w_{31} + w_{32} = 1,$$

$$w_{ik} \geq 0 \text{ при любых } i \text{ и } k.$$

Оптимальным решением является $w_{11} = w_{21} = w_{31} = 0$, $w_{12} = 6/59$, $w_{22} = 31/59$ и $w_{32} = 22/59$. Это означает, что $q_1^2 = q_2^2 = q_3^2 = 1$. Таким образом, оптимальная стратегия требует выбора альтернативы 2 ($k = 2$) при $i = 1$, 2 и 3. Оптимальное значение E равно 4.7(6/59) + 3.1(31/59) + 0.4(22/59) = 2.256. Интересно отметить, что положительные значения w_{ik} в точности равны значениям π_i, соответствующим оптимальной стратегии, определенной с помощью метода полного перебора в примере 19.3–1, что доказывает непосредственную взаимосвязь между двумя методами решения.

Рассмотрим теперь марковскую задачу принятия решений с дисконтированием. В разделе 19.3.2 эта задача описывается рекуррентным уравнением

$$f(i)=\max_k\left\{v_i^k+\alpha\sum_{j=1}^m p_{ij}^k f(j)\right\},\quad i=1,2,\dots,m.$$

Это уравнение эквивалентно следующему:

$$f(i)\ge\alpha\sum_{j=1}^m p_{ij}^k f(j)+v_i^k,\ \text{ для всех } i \text{ и } k,$$

при условии, что при любом i функция $f(i)$ достигает минимального значения. Рассмотрим теперь целевую функцию

$$\text{минимизировать } \sum_{i=1}^m b_i f(i),$$

где b_i (> 0 при всех i) — произвольные константы. Можно показать, что оптимизация этой функции (при ограничениях в виде приведенных выше неравенств) дает, что и требуется, минимальное значение $f(i)$. Таким образом, задачу можно записать следующим образом.

$$\text{Минимизировать } \sum_{i=1}^m b_i f(i)$$

при ограничениях

$$f(i)-\alpha\sum_{j=1}^m p_{ij}^k f(j)\ge v_i^k,\ \text{ при любых } i \text{ и } k,$$

$$f(i) \text{ не ограничены по знаку, } i=1,2,\dots,m.$$

Двойственной к приведенной выше задаче является задача

$$\text{максимизировать } \sum_{i=1}^m\sum_{k=1}^K v_i^k w_{ik}$$

при ограничениях

$$\sum_{k=1}^K w_{jk}-\alpha\sum_{i=1}^m\sum_{k=1}^K p_{ij}^k w_{ik}=b_j,\ j=1,2,\dots,m,$$

$$w_{ik}\ge 0 \text{ при } i=1,2,\dots,m;\ k=1,2,\dots,K.$$

Пример 19.4–2

Рассмотрим задачу садовника с коэффициентом дисконтирования $\alpha = 0.6$. Если положить $b_1 = b_2 = b_3 = 1$, то двойственную задачу линейного программирования можно записать в следующем виде

$$\text{максимизировать } 5.3w_{11} + 4.7w_{12} + 3w_{21} + 3.1w_{22} - w_{31} + 0.4w_{32}$$

при ограничениях

$$\begin{aligned}
&w_{11} + w_{12} - 0.6[0.2w_{11} + 0.3w_{12} + 0.1w_{22} + 0.05w_{32}] = 1,\\
&w_{21} + w_{22} - 0.6[0.5w_{11} + 0.6w_{12} + 0.5w_{21} + 0.6w_{22} + 0.4w_{32}] = 1,\\
&w_{31} + w_{32} - 0.6[0.3w_{11} + 0.1w_{12} + 0.5w_{21} + 0.3w_{22} + w_{31} + 0.55w_{32}] = 1,\\
&w_{ik} \geq 0 \text{ при любых } i \text{ и } k.
\end{aligned}$$

Оптимальным решением будет $w_{12} = w_{21} = w_{31} = 0$, $w_{11} = 1.5678$, $w_{22} = 3.3528$ и $w_{32} = 2.8145$. Из этого решения следует, что оптимальной стратегией является $\{1, 2, 2\}$.

Упражнения 19.4,а

1. Сформулируйте указанные ниже задачи в виде задач линейного программирования.
   a) Задача упр. 19.3,b(1).
   b) Задача упр. 19.3,b(2).
   c) Задача упр. 19.3,b(3).
2. Сформулируйте задачи упр. 19.3,c(1) в виде задач линейного программирования.

19.5. Заключение

В этой главе рассмотрены модели марковской задачи принятия решений, включая модели с конечным числом этапов. Решения этих задач можно получить непосредственно из рекуррентных уравнений динамического программирования. Здесь также показано, что в моделях с бесконечным числом этапов метод полного перебора неприменим к практическим задачам большой размерности. С вычислительной точки зрения более эффективным является метод итераций по стратегиям, основанный на использовании рекуррентных уравнений динамического программирования. Продемонстрировано, что при дисконтировании возможны другие оптимальные стратегии по сравнению со случаем без дисконтирования. Этот вывод справедлив для моделей как с конечным, так и бесконечным числом этапов.

Формулировка задачи принятия решений в виде задачи линейного программирования представляет интерес, но в вычислительном отношении алгоритм линейного программирования менее эффективен, чем метод итераций по стратегиям, особенно при больших значениях m и k.

19.6. Приложение: обзор теории цепей Маркова

Рассмотрим дискретные моменты времени $\{t_k\}$, $k = 1, 2, \ldots$, и пусть ξ_{t_k} — случайная величина, характеризующая состояние системы в момент t_k. Семейство случайных величин $\{\xi_{t_k}\}$ образует **стохастический процесс**. Состояния в момент времени t_k, в которых может находиться в этот момент система, формируют полную и взаимно исключающую группу событий. Число состояний системы может быть конечным или бесконечным. Так, например, распределение Пуассона

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad n = 0,1,2,...$$

представляет стохастический процесс с бесконечным числом состояний. Здесь случайная величина n соответствует числу наблюдаемых событий в интервале от 0 до t (начальным моментом считается 0). Таким образом, состояния системы в любой момент t задаются случайной величиной $n = 0, 1, 2, ...$.

Еще одним примером может служить игра с k подбрасываниями монеты. Каждое испытание (подбрасывание монеты) можно рассматривать как некоторый момент времени. Полученная последовательность испытаний образует случайный процесс. Состоянием системы при любом испытании является либо "герб", либо "решка".

В данном разделе приводятся сведения о таких случайных процессах, как **марковские процессы** и **марковские цепи**. Цепь Маркова является частным случаем марковских процессов. Цепи Маркова используются при изучении краткосрочного и долгосрочного поведения стохастических систем с дискретным множеством состояний.

19.6.1. Марковские процессы

Марковский процесс описывает поведение стохастической системы, в которой наступление очередного состояния зависит только от непосредственно предшествующего состояния системы. Если $t_0 < t_1 < ... < t_n$ ($n = 0, 1, 2, ...$) — моменты времени, то семейство случайных величин $\{\xi_{t_n}\}$ будет процессом Маркова тогда и только тогда, когда оно обладает **марковским свойством**

$$P\{\xi_{t_n} = x_n \mid \xi_{t_{n-1}} = x_{n-1}, ..., \xi_{t_0} = x_0\} = P\{\xi_{t_n} = x_n \mid \xi_{t_{n-1}} = x_{n-1}\}$$

для всех возможных значений случайных величин $\xi_{t_0}, \xi_{t_1}, ..., \xi_{t_n}$.

Вероятность $p_{x_{n-1}, x_n} = P\{\xi_{t_n} = x_n \mid \xi_{t_{n-1}} = x_{n-1}\}$ называется **переходной**. Она представляет собой *условную* вероятность того, что система будет находиться в состоянии x_n в момент t_n, если в момент t_{n-1} она находилась в состоянии x_{n-1}. Эту вероятность называют также **одношаговой переходной**, поскольку она описывает изменение состояния системы между последовательными моментами времени t_{n-1} и t_n. Аналогично m-шаговая переходная вероятность определяется формулой

$$p_{x_n, x_{n+m}} = P\{\xi_{t_{n+m}} = x_{n+m} \mid \xi_{t_n} = x_n\}.$$

19.6.2. Цепи Маркова

Пусть $E_1, E_2, ..., E_j$ ($j = 0, 1, 2, ...$) — полная и взаимно исключающая группа состояний некоторой системы в любой момент времени. В исходный момент t_0 система может находиться в одном из этих состояний. Пусть $a_j^{(0)}$ $(j = 0,1,2,...)$ — вероятность того, что в момент t_0 система находится в состоянии E_j. Предположим также, что рассматриваемая система является марковской.

Определим $p_{ij} = P\left\{\xi_{t_n} = j \mid \xi_{t_{n-1}} = i\right\}$ как одношаговую вероятность перехода системы из состояния i в момент времени t_{n-1} в состояние j в момент t_n и допустим, что эти вероятности постоянны во времени. Удобнее представить вероятности перехода из состояния E_i в состояние E_j в матричном виде

$$\mathbf{P} = \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & \cdots \\ p_{10} & p_{11} & p_{12} & p_{13} & \cdots \\ p_{20} & p_{21} & p_{22} & p_{23} & \cdots \\ p_{30} & p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix}.$$

Матрица **P** называется **однородной матрицей переходов** (переходных вероятностей), поскольку все переходные вероятности p_{ij} фиксированы и не зависят от времени. Вероятности p_{ij} должны удовлетворять условиям

$$\sum_j p_{ij} = 1 \text{ для всех } i,$$

$$p_{ij} \geq 0 \text{ для всех } i \text{ и } j.$$

Теперь можно определить **цепь Маркова** (или **марковскую цепь**). *Матрица переходных вероятностей* **P** *совместно с исходными вероятностями состояний полностью определяет марковскую цепь.* Обычно считается, что цепь Маркова описывает переходный режим некоторой системы на одинаковых интервалах времени. Однако иногда интервалы времени между переходами зависят от характеристик системы и, следовательно, могут быть неодинаковыми.

Абсолютные и переходные вероятности. При заданных вероятностях состояний $\left\{a_j^{(0)}\right\}$ и матрице переходных вероятностей **P** абсолютные вероятности состояний системы после определенного числа переходов определяются следующим образом. Пусть $\left\{a_j^{(n)}\right\}$ — абсолютные вероятности состояний системы после n переходов, т.е. в момент t_n. Величины $\left\{a_j^{(n)}\right\}$ можно выразить в общем виде через $\left\{a_j^{(0)}\right\}$ и **P** посредством следующего соотношения:

$$a_j^{(1)} = a_1^{(0)} p_{1j} + a_2^{(0)} p_{2j} + a_3^{(0)} p_{3j} + \ldots = \sum_i a_i^{(0)} p_{ij}.$$

Далее,

$$a_j^{(2)} = \sum_i a_i^{(1)} p_{ij} = \sum_i \left(\sum_k a_k^{(0)} p_{ki} \right) p_{ij} = \sum_k a_k^{(0)} \left(\sum_i p_{ki} p_{ij} \right) = \sum_k a_k^{(0)} p_{kj}^{(2)},$$

где $p_{kj}^{(2)} = \sum_i p_{ki} p_{ij}$ — **двухшаговая вероятность, или переходная вероятность второго порядка,** т.е. вероятность перехода из состояния k в состояние j в точности за два шага.

Подобным образом по индукции можно показать, что

$$a_j^{(n)} = \sum_i a_i^{(0)} \left(\sum_k p_{ik}^{(n-1)} p_{kj} \right) = \sum_i a_i^{(0)} p_{ij}^{(n)},$$

где $p_{ij}^{(n)}$ — n-шаговая переходная вероятность (или переходная вероятность n-го порядка), определяемая рекуррентной формулой

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(n-1)} p_{kj}.$$

В общем виде для произвольных i и j имеем

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(n-m)} p_{kj}^{(m)}, \quad 0 < m < n.$$

Эти уравнения известны как **уравнения Колмогорова–Чепмена**.

Элементы матриц переходов высших порядков $\left\| p_{ij}^{(n)} \right\|$ можно получить непосредственно путем перемножения матриц. Так, например,

$$\left\| p_{ij}^{(2)} \right\| = \left\| p_{ij} \right\| \left\| p_{ij} \right\| = \mathbf{P}^2,$$

$$\left\| p_{ij}^{(3)} \right\| = \left\| p_{ij}^{(2)} \right\| \left\| p_{ij} \right\| = \mathbf{P}^3$$

и в общем виде

$$\left\| p_{ij}^{(n)} \right\| = \mathbf{P}^{n-1}\mathbf{P} = \mathbf{P}^n.$$

Следовательно, если абсолютные вероятности состояний определены в векторной форме как

$$\mathbf{a}^{(n)} = \left\{ a_1^{(n)}, a_2^{(n)}, a_3^{(n)}, \ldots \right\},$$

то

$$\mathbf{a}^{(n)} = \mathbf{a}^{(0)}\mathbf{P}^n.$$

Пример 19.6–1

Рассмотрим цепь Маркова с двумя состояниями. Матрица переходных вероятностей имеет вид

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix}$$

и $\mathbf{a}^{(0)} = (0.7, 0.3)$. Найдем $\mathbf{a}^{(1)}$, $\mathbf{a}^{(4)}$ и $\mathbf{a}^{(8)}$.

$$\mathbf{P}^2 = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.52 & 0.48 \\ 0.36 & 0.64 \end{bmatrix},$$

$$\mathbf{P}^4 = \mathbf{P}^2\mathbf{P}^2 = \begin{bmatrix} 0.52 & 0.48 \\ 0.36 & 0.64 \end{bmatrix} \begin{bmatrix} 0.52 & 0.48 \\ 0.36 & 0.64 \end{bmatrix} \approx \begin{bmatrix} 0.443 & 0.557 \\ 0.417 & 0.583 \end{bmatrix},$$

$$\mathbf{P}^8 = \mathbf{P}^4\mathbf{P}^4 = \begin{bmatrix} 0.443 & 0.557 \\ 0.417 & 0.583 \end{bmatrix} \begin{bmatrix} 0.443 & 0.557 \\ 0.417 & 0.583 \end{bmatrix} \approx \begin{bmatrix} 0.4281 & 0.5719 \\ 0.4274 & 0.5726 \end{bmatrix}.$$

Таким образом,

$$\mathbf{a}^{(1)} = (0.7, 0.3)\begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix} = (0.32, 0.68),$$

$$\mathbf{a}^{(4)} = (0.7, 0.3)\begin{bmatrix} 0.443 & 0.557 \\ 0.417 & 0.583 \end{bmatrix} = (0.435, 0.565),$$

$$\mathbf{a}^{(8)} = (0.7, 0.3)\begin{bmatrix} 0.4281 & 0.5719 \\ 0.4274 & 0.5726 \end{bmatrix} = (0.4279, 0.5721).$$

Отметим, что строки матрицы $\mathbf{P}^8$ незначительно отличаются друг от друга. Кроме того, вектор $\mathbf{a}^{(8)}$ также несущественно отличается от каждой из строк матрицы $\mathbf{P}^8$. Это связано с тем, что абсолютные вероятности состояний после выполнения нескольких переходов практически не зависят от начальных вероятностей $\mathbf{a}^{(0)}$. В том случае, когда они действительно не будут зависеть от начальных вероятностей, их называют **установившиеся**.

КЛАССИФИКАЦИЯ СОСТОЯНИЙ МАРКОВСКИХ ЦЕПЕЙ. При рассмотрении цепей Маркова нас может интересовать поведение системы на коротком отрезке времени. В таком случае абсолютные вероятности вычисляются так, как показано в предыдущем разделе. Однако более важно изучить поведение системы на большом интервале времени, т.е. в условиях, когда число переходов стремится к бесконечности. В этом случае изложенный выше метод непригоден; требуется систематический подход, позволяющий прогнозировать долгосрочное поведение системы. Ниже вводятся определения состояний марковских цепей, которые необходимы для изучения долгосрочного поведения системы.

Неприводимая марковская цепь. Цепь Маркова называется **неприводимой**, если любое состояние E_j может быть достигнуто из любого другого состояния E_i за конечное число переходов, т.е. при $i \neq j$ $p_{ij}^{(n)} > 0$ для $1 \leq n < \infty$. В этом случае все состояния цепи называются **сообщающимися**.

Замкнутое множество состояний и поглощающие состояния. Множество C состояний цепи Маркова называется **замкнутым**, если система, однажды оказавшаяся в одном из состояний этого множества, будет находиться в множестве C в течение бесконечного интервала времени. Частным случаем замкнутого множества является единственное состояние E_j с переходной вероятностью $p_{jj} = 1$. В этом случае состояние E_j называется **поглощающим**. Все состояния неприводимой цепи должны образовывать замкнутое множество, и ни одно подмножество этого множества не может быть замкнутым. Замкнутое множество C удовлетворяет всем условиям, характеризующим марковскую цепь, и, следовательно, его можно подвергнуть независимому анализу.

Пример 19.6–2

Рассмотрим цепь Маркова со следующей матрицей переходных вероятностей

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}.$$

Эта цепь графически представлена на рис. 19.1. Как видно из рисунка, четыре состояния *не составляют* неприводимую цепь, поскольку состояний 0, 1 и 2 нельзя достигнуть из состояния 3. Состояние 3 образует замкнутое множество и, таким образом, является поглощающим. Можно также утверждать, что состояние 3 соответствует неприводимой цепи Маркова.

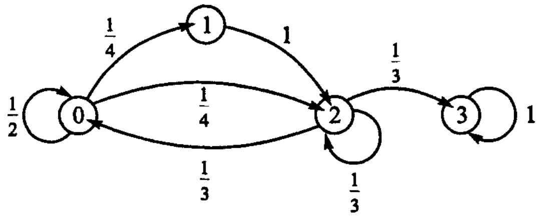


Рис. 19.1

Первое время возвращения. Важным понятием в теории марковских цепей является **первое время возвращения**. Система, первоначально находящаяся в состоянии E_j, может вернуться *в первый раз* в это же состояние через n шагов, $n > 1$. Число шагов, за которое система возвращается в состояние E_j, называется первым временем возвращения.

Обозначим через $f_{jj}^{(n)}$ вероятность того, что первое возвращение в состояние E_j состоится на n-м шагу. Тогда при заданной матрице переходных вероятностей $\mathbf{P} = \|p_{ij}\|$ $f_{jj}^{(n)}$ можно определить следующим образом:

$$p_{jj} = f_{jj}^{(1)},$$

$$p_{jj}^{(2)} = f_{jj}^{(2)} + f_{jj}^{(1)} p_{jj},$$

или

$$f_{jj}^{(2)} = p_{jj}^{(2)} - f_{jj}^{(1)} p_{jj}.$$

По индукции нетрудно показать, что

$$p_{jj}^{(n)} = f_{jj}^{(n)} + \sum_{m=1}^{n-1} f_{jj}^{(m)} p_{jj}^{(n-m)},$$

откуда получается искомое выражение для вероятности $f_{jj}^{(n)}$:

$$f_{jj}^{(n)} = p_{jj}^{(n)} - \sum_{m=1}^{n-1} f_{jj}^{(m)} p_{jj}^{(n-m)}.$$

Отсюда следует, что вероятность *по крайней мере* одного возвращения в состояние E_j задается формулой

$$f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)}.$$

Следовательно, система обязательно вернется в состояние j, если $f_{jj} = 1$. Обозначив через μ_{jj} среднее время возвращения, получаем

$$\mu_{jj} = \sum_{n=1}^{\infty} n f_{jj}^{(n)}.$$

Если $f_{jj} < 1$, то неизвестно, вернется ли система в состояние E_j, и, следовательно, $\mu_{jj} = \infty$.

Исходя из определения первого времени возвращения, состояния марковской цепи можно классифицировать следующим образом.

1. Состояние является **невозвратным**, если $f_{jj} < 1$, т.е. $\mu_{jj} = \infty$.
2. Состояние является **возвратным**, если $f_{jj} = 1$.
3. Возвратное состояние является **нулевым**, если $\mu_{jj} = \infty$, и **ненулевым**, когда $\mu_{jj} < \infty$ (т.е. конечно).
4. Состояние называется **периодическим** с периодом t, если возвращение в него возможно только через число шагов, кратное t: t, $2t$, $3t$, ... Это означает, что если n не делится на t без остатка, то $p_{jj}^{(n)} = 0$.
5. Возвратное состояние является **эргодическим**, если оно ненулевое и апериодическое.

Эргодические марковские цепи. Если все состояния цепи Маркова являются эргодическими, то цепь *неприводима*. В этом случае распределение абсолютных вероятностей состояний

$$\mathbf{a}^{(n)} = \mathbf{a}^{(0)}\mathbf{P}^n$$

всегда однозначно сходится к предельному распределению при $n \to \infty$, где предельное распределение не зависит от начальных вероятностей $\mathbf{a}^{(0)}$.

Справедлива следующая теорема.

Теорема 19.6–1. *Все состояния в неприводимой бесконечной цепи Маркова могут принадлежать к одному и только одному из следующих трех классов: невозвратных, возвратных нулевых или возвратных ненулевых состояний. Во всех случаях все состояния являются сообщающимися и имеют один и тот же период. В частном случае, когда в цепи конечное число состояний, она не может содержать только невозвратные или какие-либо нулевые состояния.*

Предельные распределения в неприводимых цепях Маркова. Из примера 19.6–1 видно, что с ростом числа переходов абсолютная вероятность состояний становится независимой от начального распределения. В данном разделе описывается метод вычисления предельного распределения вероятностей состояний в неприводимой цепи. Мы ограничимся рассмотрением только апериодических состояний, так как эго единственный класс состояний, используемый в дальнейшем. Кроме того, анализ периодических состояний является довольно сложным.

Существование предельного распределения в неприводимой апериодической цепи зависит от класса ее состояний. Таким образом, рассматривая три класса состояний, указанных в теореме 19.6–1, можно сформулировать следующую теорему.

Теорема 19.6–2. *Если в неприводимой апериодической цепи Маркова*

a) *все состояния невозвратные или нулевые, то* $p_{ij}^{(n)} \to 0$ *при* $n \to \infty$ *для всех* i *и* j *и предельного распределения не существует,*

b) *все состояния эргодические, то*

$$\lim_{n\to\infty} a_j^{(n)} = \beta_j, \quad j = 0, 1, 2, \ldots,$$

где β_j *— предельное (установившееся) распределение. Вероятности* β_j *определяются однозначно и не зависят от* $a_j^{(0)}$. *Величины* β_j *можно определить из системы уравнений*[1]

$$\beta_j = \sum_i \beta_i p_{ij}, \quad 1 = \sum_j \beta_j.$$

Среднее время возвращения в состояние j *при этом определяется формулой*

$$\mu_{jj} = \frac{1}{\beta_j}.$$

Пример 19.6–3

Рассмотрим задачу из примера 19.6–1. Для определения установившегося распределения вероятностей используем соотношения

$$\beta_1 = 0.2\beta_1 + 0.6\beta_2,$$
$$\beta_2 = 0.8\beta_1 + 0.4\beta_2,$$
$$1 = \beta_1 + \beta_2.$$

Решением будет $\beta_1 = 0.4286$ и $\beta_2 = 0.5714$. Эти результаты очень близки к значениям элементов вектора $\mathbf{a}^{(8)}$ (и строкам матрицы $\mathbf{P}^8$) из примера 19.6–1. Далее получаем

$$\mu_{11} = \frac{1}{\beta_1} = 2.3, \quad \mu_{22} = \frac{1}{\beta_2} = 1.75,$$

так что значения среднего времени возвращения в первое и второе состояния равны 2.3 и 1.75 шагов соответственно.

---

[1] Заметим, что одно из уравнений $\beta_j = \sum_i \beta_i p_{ij}$ является избыточным.

Пример 19.6–4

Рассмотрим следующую цепь Маркова с тремя состояниями:

$$
\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array}
\begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}.
$$

Такая матрица называется **дважды стохастической**, так как

$$\sum_{i=1}^{s} p_{ij} = \sum_{j=1}^{s} p_{ij} = 1,$$

где s — число состояний цепи Маркова. В таких случаях установившиеся вероятности равны $\beta_j = 1/s$ для всех j. Поэтому для данной задачи $\beta_0 = \beta_1 = \beta_2 = 1/3$.

Упражнения 19.6,а

1. Определите класс состояний приведенных ниже цепей Маркова и найдите их стационарные распределения.

a) $$\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

b) $$\begin{bmatrix} q & p & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ q & 0 & 0 & p & 0 \\ q & 0 & 0 & 0 & p \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad p + q = 1.$$

2. Найдите среднее время возвращения в каждое состояние цепи Маркова, заданной следующей матрицей переходных вероятностей.

$$\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{3}{5} & \frac{1}{5} \end{bmatrix}.$$

Литература

Derman C. *Finite State Markovian Decision Process*, Academic Press, New York, 1970.
Howard R. *Dynamic Programming and Markov Processes*, MIT Press, Cambridge, Mass., 1960. (Имеется перевод: Ховард Р. *Динамическое программирование и марковские процессы*. — М.: Сов. радио, 1964.)

Литература, добавленная при переводе

Дынкин Е.Б., Юшкевич А.А. *Теоремы и задачи о процессах Маркова*. — М.: Наука, 1967.
Кемени Дж., Снелл Дж. *Конечные цепи Маркова*. — М.: Наука, 1970.

ЧАСТЬ III

НЕЛИНЕЙНЫЕ МОДЕЛИ

Классическая теория оптимизации

20.1. Введение

В классической теории оптимизации для нахождения точек максимума и минимума (экстремальных точек) функций в условиях как отсутствия, так и наличия ограничений на переменные широко используется аппарат дифференциального исчисления. Получаемые при этом методы не всегда оказываются удобными при их численной реализации. Однако соответствующие теоретические результаты лежат в основе большинства алгоритмов решения задач нелинейного программирования (см. главу 21).

В этой главе изложены необходимые и достаточные условия существования экстремумов функций при отсутствии ограничений на переменные задачи, методы *Якоби* и *Лагранжа* для решения задач с ограничениями на переменные в форме равенств, а также условия *Куна–Таккера* для задач с ограничениями в виде неравенств.

20.2. Экстремальные задачи без ограничений

Экстремальная точка функции $f(\mathbf{X})$ определяет либо ее максимальное, либо минимальное значение. С математической точки зрения точка $\mathbf{X}_0 = (x_1, ..., x_j, ..., x_n)$ является точкой максимума функции $f(\mathbf{X})$, если неравенство

$$f(\mathbf{X}_0 + \mathbf{h}) \le f(\mathbf{X}_0)$$

выполняется для всех $\mathbf{h} = (h_1, ..., h_j, ..., h_n)$ таких, что $|h_j|$ достаточно малы при всех j. Другими словами, точка $\mathbf{X}_0$ является точкой максимума, если значения функции f в окрестности точки $\mathbf{X}_0$ не превышают $f(\mathbf{X}_0)$. Аналогично, точка $\mathbf{X}_0$ является точкой минимума функции $f(\mathbf{X})$, если для определенного выше вектора $\mathbf{h}$ имеет место неравенство

$$f(\mathbf{X}_0 + \mathbf{h}) \ge f(\mathbf{X}_0).$$

На рис. 20.1 показаны точки максимума и минимума функции одной переменной $f(x)$ на интервале $[a, b]$. Точки x_1, x_2, x_3, x_4 и x_6 составляют множество экстремальных точек функции $f(x)$. Здесь точки x_1, x_3 и x_6 являются точками максимума, а точки x_2 и x_4 — точками минимума функции $f(x)$. Поскольку

$$f(x_6) = \max\{f(x_1),\ f(x_3),\ f(x_6)\},$$

значение $f(x_6)$ называется **глобальным** или **абсолютным** максимумом, а значения $f(x_1)$ и $f(x_3)$ — **локальными** или **относительными** максимумами. Подобным образом, значение $f(x_4)$ является локальным, а $f(x_2)$ — глобальным минимумом функции $f(x)$.

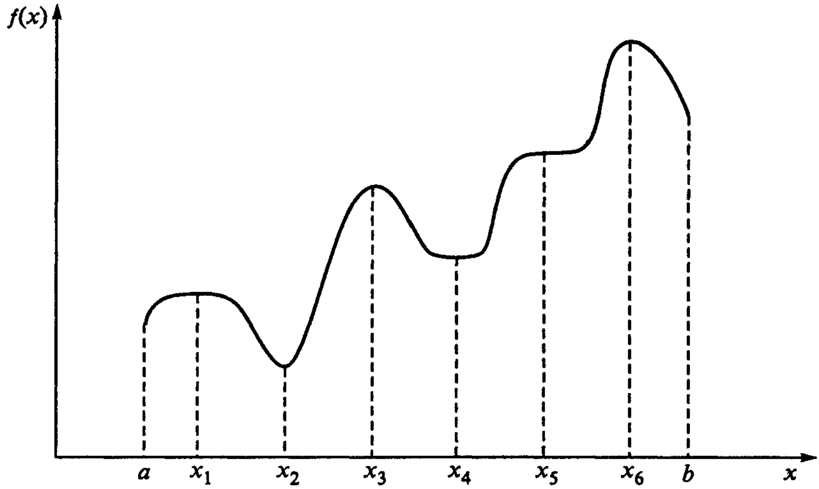


Рис. 20.1

Заметим, что хотя точка x_1 является точкой максимума функции $f(x)$ (рис. 20.1), она отличается от остальных локальных максимумов $f(x)$ тем, что по крайней мере в одной точке ее окрестности значение функции $f(x)$ совпадает с $f(x_1)$. Точка x_1 по этой причине называется **нестрогим (слабым) максимумом** функции $f(x)$, в отличие, например, от точки x_3, которая является **строгим максимумом** $f(x)$. Нестрогий максимум, следовательно, подразумевает наличие (бесконечного количества) различных точек, которым соответствует одно и то же максимальное значение функции. Аналогичные результаты имеют место в точке x_4, где функция $f(x)$ имеет нестрогий минимум. В общем случае $\mathbf{X}_0$ является точкой нестрогого максимума функции $f(x)$, если $f(\mathbf{X}_0 + \mathbf{h}) \le f(\mathbf{X}_0)$, и точкой ее строгого максимума, если $f(\mathbf{X}_0 + \mathbf{h}) < f(\mathbf{X}_0)$, где $\mathbf{h}$ — вектор, определенный выше.

На рис. 20.1 легко заметить, что первая производная функции f (тангенс угла наклона касательной к графику функции) равна 0 во всех ее экстремальных точках. Однако это условие выполняется и в точках **перегиба** и **седловых** точках функции f, таких как точка x_5. Если точка, в которой угол наклона касательной к графику функции (градиент функции) равен нулю, не является в то же время точкой экстремума (максимума или минимума), то она автоматически должна быть точкой перегиба или седловой точкой.

20.2.1. Необходимые и достаточные условия существования экстремума

В этом разделе излагаются необходимые и достаточные условия существования экстремумов функции n переменных $f(\mathbf{X})$. При этом предполагается, что первые и вторые частные производные функции $f(\mathbf{X})$ непрерывны в каждой точке $\mathbf{X}$.

Теорема 20.2–1. *Необходимым условием того, что точка* $\mathbf{X}_0$ *является экстремальной точкой функции* $f(\mathbf{X})$, *является равенство*

$$\nabla f(\mathbf{X}_0) = 0.$$

Доказательство. Из теоремы Тейлора следует, что при $0 < \theta < 1$ имеет место разложение функции $f(\mathbf{X})$

$$f\left(\mathbf{X}_0+\mathbf{h}\right)-f\left(\mathbf{X}_0\right)=\nabla f\left(\mathbf{X}_0\right)\mathbf{h}+\frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h}\,|_{\mathbf{X}_0+\theta\mathbf{h}},$$

где $\mathbf{h}$ — вектор, определенный выше.

Для достаточно малых значений $|h_j|$ остаточный член $\frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h}$ является величиной порядка h_j^2. Следовательно,

$$f\left(\mathbf{X}_0+\mathbf{h}\right)-f\left(\mathbf{X}_0\right)=\nabla f\left(\mathbf{X}_0\right)\mathbf{h}+O\left(h_j^2\right)\approx\nabla f\left(\mathbf{X}_0\right)\mathbf{h}.$$

Пусть $\mathbf{X}_0$ — точка минимума функции $f(\mathbf{X})$. Докажем от противного, что градиент $\nabla f(\mathbf{X}_0)$ функции $f(\mathbf{X})$ в точке минимума $\mathbf{X}_0$ равен нулю. Пусть это условие не выполняется; тогда для некоторого j должно выполняться условие

$$\frac{\partial f(\mathbf{X}_0)}{\partial x_j}<0 \text{ или } \frac{\partial f(\mathbf{X}_0)}{\partial x_j}>0.$$

Знак h_j всегда можно выбрать таким образом, чтобы

$$h_j\frac{\partial f(\mathbf{X}_0)}{\partial x_j}<0.$$

Полагая остальные h_j равными нулю, из разложения Тейлора получаем неравенство

$$f(\mathbf{X}_0+\mathbf{h})<f(\mathbf{X}_0).$$

Этот результат противоречит предположению, что $\mathbf{X}_0$ — точка минимума. Следовательно, величина $\nabla f(\mathbf{X}_0)$ должна равняться нулю. Доказательство для точки максимума проводится аналогично.

Так как необходимое условие выполняется также в точках перегиба и седловых точках, точки, удовлетворяющие уравнению $\nabla f(\mathbf{X}_0)=0$, называют **стационарными**. Следующая теорема устанавливает достаточные условия того, что стационарная точка $\mathbf{X}_0$ является экстремальной.

Теорема 20.2–2. *Для того чтобы стационарная точка* $\mathbf{X}_0$ *была экстремальной, достаточно, чтобы матрица Гессе* $\mathbf{H}$ *в точке* $\mathbf{X}_0$ *была*

i) *положительно определенной (тогда* $\mathbf{X}_0$ *— точка минимума);*
ii) *отрицательно определенной (тогда* $\mathbf{X}_0$ *— точка максимума).*

Доказательство. В силу теоремы Тейлора при $0<\theta<1$ имеем

$$f\left(\mathbf{X}_0+\mathbf{h}\right)-f\left(\mathbf{X}_0\right)=\nabla f\left(\mathbf{X}_0\right)\mathbf{h}+\frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h}\,|_{\mathbf{X}_0+\theta\mathbf{h}}.$$

Поскольку $\mathbf{X}_0$ — стационарная точка, в силу теоремы 20.2–1 $\nabla f(\mathbf{X}_0)=0$. Таким образом,

$$f\left(\mathbf{X}_0+\mathbf{h}\right)-f\left(\mathbf{X}_0\right)=\frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h}\,|_{\mathbf{X}_0+\theta\mathbf{h}}.$$

Если $\mathbf{X}_0$ — точка минимума, то

$$f(\mathbf{X}_0+\mathbf{h})>f(\mathbf{X}_0)$$

для всех ненулевых векторов $\mathbf{h}$. Следовательно, в точке минимума $\mathbf{X}_0$ должно выполняться неравенство

$$\frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h}\,|_{\mathbf{X}_0+\theta\mathbf{h}} > 0.$$

Непрерывность вторых частных производных функции $f(\mathbf{X})$ гарантирует, что величина $\frac{1}{2}\mathbf{h}^T\mathbf{H}\mathbf{h}$ имеет один и тот же знак как в точке $\mathbf{X}_0$, так и $\mathbf{X}_0+\theta\mathbf{h}$. Так как $\mathbf{h}^T\mathbf{H}\mathbf{h}\,|_{\mathbf{X}_0}$ представляет собой квадратичную форму (см. раздел А.3), ее значение (и, следовательно, $\mathbf{h}^T\mathbf{H}\mathbf{h}\,|_{\mathbf{X}_0+\theta\mathbf{h}}$) положительно тогда и только тогда, когда $\mathbf{H}\,|_{\mathbf{X}_0}$ — положительно определенная матрица. Это означает, что положительная определенность матрицы Гессе в стационарной точке $\mathbf{X}_0$ является достаточным условием существования в этой точке минимума. Путем аналогичных рассуждений доказывается, что стационарная точка является точкой максимума, если матрица Гессе в этой точке отрицательно определена.

Пример 20.2–1

Рассмотрим функцию

$$f(x_1, x_2, x_3) = x_1 + 2x_3 + x_2x_3 - x_1^2 - x_2^2 - x_3^2.$$

Необходимое условие экстремума $\nabla f(\mathbf{X}_0) = 0$ здесь принимает следующий вид.

$$\frac{\partial f}{\partial x_1} = 1 - 2x_1 = 0,$$

$$\frac{\partial f}{\partial x_2} = x_3 - 2x_2 = 0,$$

$$\frac{\partial f}{\partial x_3} = 2 + x_2 - 2x_3 = 0.$$

Решением этой системы уравнений является точка $\mathbf{X}_0 = (1/2, 2/3, 4/3)$.

Для проверки выполнения условия достаточности вычислим

$$\mathbf{H}\,|_{\mathbf{X}_0} = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_3} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \dfrac{\partial^2 f}{\partial x_2 \partial x_3} \\ \dfrac{\partial^2 f}{\partial x_3 \partial x_1} & \dfrac{\partial^2 f}{\partial x_3 \partial x_2} & \dfrac{\partial^2 f}{\partial x_3^2} \end{bmatrix}_{\mathbf{X}_0} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -2 \end{bmatrix}.$$

Угловые миноры матрицы $\mathbf{H}\,|_{\mathbf{X}_0}$ равны -2, 4 и -6 соответственно. В этом случае $\mathbf{H}\,|_{\mathbf{X}_0}$ является отрицательно определенной матрицей (см. раздел А.3), откуда следует, что точка $\mathbf{X}_0 = (1/2, 2/3, 4/3)$ является точкой максимума.

В общем случае, когда матрица $\mathbf{H}|_{\mathbf{X}_0}$ является неопределенной, точка $\mathbf{X}_0$ должна быть седловой. Если же матрица $\mathbf{H}|_{\mathbf{X}_0}$ оказывается полуопределенной, то соответствующая точка $\mathbf{X}_0$ может как быть, так и не быть экстремальной. При этом формулировка достаточного условия существования экстремума значительно усложняется, ибо для этого необходимо учитывать члены более высоких порядков в разложении Тейлора. (Иллюстрацией вышесказанного в случае функции одной переменной служит теорема 20.2–3.) В некоторых случаях такие сложные процедуры не являются необходимыми, поскольку диагонализация матрицы $\mathbf{H}$ позволяет сделать вывод о наличии или отсутствии экстремума. Проиллюстрируем это следующим примером.

Пример 20.2–2

Рассмотрим функцию

$$f(x_1, x_2) = 8x_1x_2 + 3x_2^2,$$

необходимые условия экстремума для которой принимают вид

$$\nabla f(x_1, x_2) = (8x_2,\ 8x_1 + 6x_2) = (0,\ 0).$$

Отсюда следует, что стационарной является точка $\mathbf{X}_0 = (0, 0)$. При этом матрица Гессе в точке $\mathbf{X}_0$ имеет вид

$$\mathbf{H} = \begin{bmatrix} 0 & 8 \\ 8 & 6 \end{bmatrix}$$

и не дает информации о наличии или отсутствии экстремума в рассматриваемой точке. Используя один из методов диагонализации матриц, получаем преобразованную матрицу Гессе

$$\mathbf{H}_t = \begin{bmatrix} -\frac{64}{6} & 0 \\ 0 & 6 \end{bmatrix},$$

проверка угловых миноров которой свидетельствует, что матрица $\mathbf{H}_t$ (и, следовательно, $\mathbf{H}$) является неопределенной.[1] Следовательно, $\mathbf{X}_0$ — седловая точка.

Применим достаточные условия, полученные в теореме 20.2–2, к функции одной переменной. Пусть y_0 — стационарная точка функции f, тогда

i) неравенство $f''(y_0) < 0$ является достаточным условием существования максимума в точке y_0;

ii) неравенство $f''(y_0) > 0$ является достаточным условием существования минимума в точке y_0.

Если же для функции одной переменной $f''(y_0) = 0$, то необходимо исследовать производные высших порядков в соответствии со следующей теоремой.

[1] Если матрица $\mathbf{H}$ приведена к диагональному виду путем преобразования подобия (а это здесь подразумевается), то нет необходимости проверять угловые миноры матрицы $\mathbf{H}_t$, так как в этом случае диагональные элементы матрицы $\mathbf{H}_t$ являются собственными значениями матрицы $\mathbf{H}$. Поэтому данная матрица будет неопределенной, если диагональные элементы матрицы $\mathbf{H}_t$ имеют разные знаки. — *Прим. ред.*

Теорема 20.2–3. *Если в стационарной точке y_0 функции $f(y)$ первые $(n-1)$ ее производных равны нулю и $f^{(n)}(y_0) \neq 0$, то в точке $y = y_0$ функция $f(y)$ имеет*

i) *точку перегиба, если n — нечетное;*
ii) *экстремальную точку, если n — четное. Эта экстремальная точка является точкой максимума при $f^{(n)}(y_0) < 0$ и точкой минимума при $f^{(n)}(y_0) > 0$.*

Пример 20.2–3

Рассмотрим функции $f(y) = y^4$ и $g(y) = y^3$. Для функции $f(y) = y^4$ имеем

$$f'(y) = 4y^3 = 0,$$

откуда получаем стационарную точку $y_0 = 0$. Далее находим

$$f'(0) = f''(0) = f^{(3)}(0) = 0.$$

Так как $f^{(4)}(0) = 24 > 0$, то $y_0 = 0$ является точкой минимума (рис. 20.2).

Для функции $g(y) = y^3$ имеем

$$g'(y) = 3y^2 = 0.$$

Следовательно, точка $y_0 = 0$ является стационарной точкой. Поскольку $g^{(n)}(0)$ не обращается в нуль при $n = 3$, точка $y_0 = 0$ является точкой перегиба.

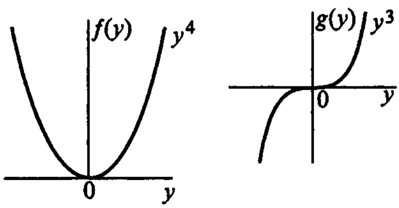


Рис. 20.2

Упражнения 20.2,а

1. Найдите экстремальные точки следующих функций.

   a) $f(x) = x^3 + x$.
   b) $f(x) = x^4 + x^2$.
   c) $f(x) = 4x^4 - x^2 + 5$.
   d) $f(x) = (3x-2)^2(2x-3)^2$.
   e) $f(x) = 6x^5 - 4x^3 + 10$.

2. Найдите экстремальные точки следующих функций.

   a) $f(\mathbf{X}) = x_1^3 + x_2^3 - 3x_1x_2$.
   b) $f(\mathbf{X}) = 2x_1^2 + x_2^2 + x_3^2 + 6(x_1 + x_2 + x_3) + 2x_1x_2x_3$.

3. Проверьте, что функция

$$f(x_1, x_2, x_3) = 2x_1x_2x_3 - 4x_1x_3 - 2x_2x_3 + x_1^2 + x_2^2 + x_3^2 - 2x_1 - 4x_2 + 4x_3$$

имеет стационарные точки (0, 3, 1), (0, 1, –1), (1, 2, 0), (2, 1, 1) и (2, 3, –1). Используйте достаточные условия для нахождения экстремумов функции.

4. Решите следующую систему уравнений путем превращения ее в задачу минимизации нелинейной целевой функции при отсутствии ограничений на переменные.

$$x_2 - x_1^2 = 0,$$

$$x_2 - x_1 = 2.$$

(*Подсказка.* $\min f^2(x_1, x_2)$ имеет место при $f(x_1, x_2) = 0$.)

5. Докажите теорему 20.2–3.

20.2.2. Метод Ньютона–Рафсона

В общем случае использование необходимого условия экстремума $\nabla f(\mathbf{X}) = 0$ для нахождения стационарных точек функции $f(\mathbf{X})$ может быть сопряжено с трудностями, возникающими при численном решении соответствующей системы уравнений. Метод Ньютона–Рафсона предлагает итерационную процедуру решения системы нелинейных уравнений. Несмотря на то что данный метод рассматривается в этом разделе именно в указанном контексте, на самом деле он относится к числу **градиентных методов** численного поиска экстремумов функций при отсутствии ограничений (см. раздел 21.1.2).

Рассмотрим систему уравнений

$$f_i(\mathbf{X}) = 0, \; i = 1, 2, ..., m.$$

Пусть $\mathbf{X}^k$ — некоторая фиксированная точка. Используя разложение Тейлора, имеем

$$f_i(\mathbf{X}) \approx f_i(\mathbf{X}^k) + \nabla f_i(\mathbf{X}^k)(\mathbf{X} - \mathbf{X}^k), \; i = 1, 2, ..., m.$$

Следовательно, исходная система уравнений приближенно представима в виде

$$f_i(\mathbf{X}^k) + \nabla f_i(\mathbf{X}^k)(\mathbf{X} - \mathbf{X}^k) = 0, \; i = 1, 2, ..., m.$$

Эти уравнения можно записать в матричной форме:

$$\mathbf{A}_k + \mathbf{B}_k(\mathbf{X} - \mathbf{X}^k) = 0.$$

Предположим, что векторы $f_i(\mathbf{X})$ независимы, тогда матрица $\mathbf{B}_k$ является невырожденной. В результате из предыдущего уравнения получим

$$\mathbf{X} = \mathbf{X}^k - \mathbf{B}_k^{-1}\mathbf{A}_k.$$

Идея метода Ньютона–Рафсона состоит в следующем. На первом шаге выбирается начальная точка $\mathbf{X}^0$. С помощью полученного уравнения по известной точке $\mathbf{X}^k$ можно вычислить координаты новой точки $\mathbf{X}^{k+1}$. Процесс вычислений завершается в точке $\mathbf{X}^m$, которая считается приближенным решением исходной системы, если $\mathbf{X}^m \approx \mathbf{X}^{m-1}$.

Геометрическая интерпретация данного метода для функции одной переменной $f(x)$ приведена на рис. 20.3. Связь между точками x^k и x^{k+1} в этом случае выражается формулой

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}$$

или

$$f'(x^k) = \frac{f(x^k)}{x^k - x^{k+1}}.$$

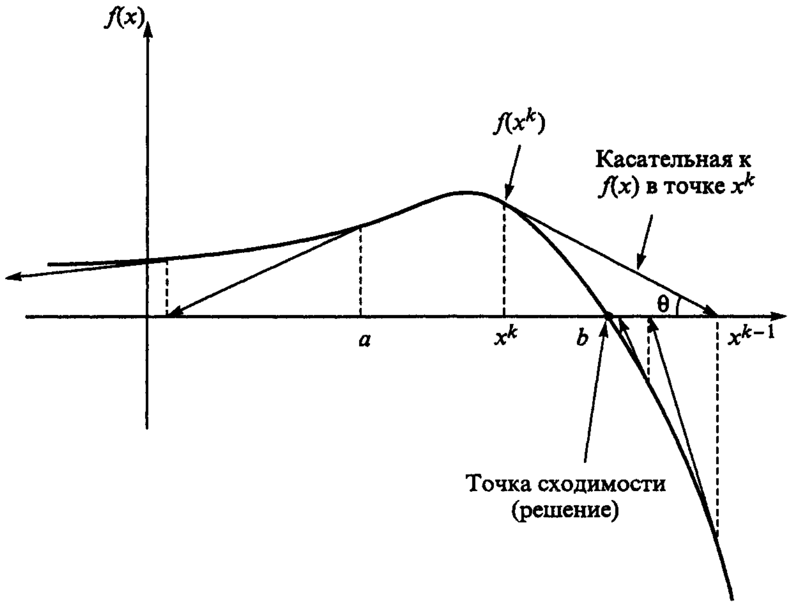


Рис. 20.3

На рис. 20.3 видно, что положение точки x^{k+1} определяется углом θ наклона касательной к графику функции $f(x)$ в точке x^k, где $\mathrm{tg}\theta = f'(x^k)$.

Одним из недостатков изложенного метода является то, что для функций общего вида не всегда гарантируется его сходимость. На рис. 20.3 видно, что при выборе в качестве x^0 точки a итерационный процесс расходится. Простых рецептов для выбора "хорошего" начального приближения не существует.

Упражнение 20.2,b

1. Примените метод Ньютона–Рафсона для решения упр. 20.2,a(1,c) и 20.2,a(2,b).

20.3. Задачи на экстремум при наличии ограничений

В настоящем разделе рассматриваются задачи поиска экстремумов непрерывных функций при наличии ограничений на переменные. В разделе 20.3.1 рассматривается случай, когда дополнительные ограничения имеют вид равенств, а в разделе 20.3.2 — неравенств. Основная часть материала раздела 20.3.1 изложена в соответствии с [2].

20.3.1. Ограничения в виде равенств

В данном разделе рассматривается два метода решения оптимизационных задач при наличии ограничений в виде равенств: **метод Якоби** и **метод Лагранжа**. Метод Лагранжа можно логически получить из метода Якоби. Эта связь позволяет дать интересную экономическую интерпретацию метода множителей Лагранжа.

Метод приведенного градиента (метод Якоби). Рассмотрим задачу

$$\text{Минимизировать } z = f(\mathbf{X})$$

при ограничениях

$$\mathbf{g}(\mathbf{X}) = \mathbf{0},$$

где

$$\mathbf{X} = (x_1, x_2, \ldots, x_n),$$
$$\mathbf{g} = (g_1, g_2, \ldots, g_m)^T.$$

Функции $f(\mathbf{X})$ и $g_i(\mathbf{X})$, $i = 1, 2, \ldots, m$, предполагаются дважды непрерывно дифференцируемыми.

Идея использования приведенного градиента заключается в том, чтобы найти замкнутое аналитическое выражение для первых частных производных функции $f(\mathbf{X})$ во всех точках, удовлетворяющих ограничениям $\mathbf{g}(\mathbf{X}) = \mathbf{0}$. Соответствующие стационарные точки определяются из условия равенства нулю указанных частных производных. Затем можно использовать достаточные условия, сформулированные в разделе 20.2, для классификации найденных стационарных точек.

Для пояснения изложенной идеи рассмотрим функцию $f(x_1, x_2)$, график которой представлен на рис. 20.4. Предположим, что эту функцию необходимо минимизировать при ограничении

$$g_1(x_1, x_2) = x_2 - b = 0,$$

где b — некоторая константа. На рис. 20.4 видно, что кривая, которая проходит через точки A, B и C, состоит из значений функции $f(x_1, x_2)$, для которых заданное ограничение выполнено. В соответствии с рассматриваемым методом определяются компоненты приведенного градиента функции $f(x_1, x_2)$ в каждой точке кривой ABC. Точка B, в которой приведенная производная обращается в нуль, является стационарной для рассматриваемой задачи с ограничением.

Теперь рассмотрим общую математическую формулировку метода. Из теоремы Тейлора следует, что для точек $\mathbf{X} + \Delta\mathbf{X}$ из окрестности точки $\mathbf{X}$ имеем

$$f(\mathbf{X}+\Delta\mathbf{X}) - f(\mathbf{X}) = \nabla f(\mathbf{X})\Delta\mathbf{X} + O(\Delta x_j^2),$$
$$\mathbf{g}(\mathbf{X}+\Delta\mathbf{X}) - \mathbf{g}(\mathbf{X}) = \nabla \mathbf{g}(\mathbf{X})\Delta\mathbf{X} + O(\Delta x_j^2).$$

При $\Delta x_j \to 0$ эти уравнения принимают вид

$$\partial f(\mathbf{X}) = \nabla f(\mathbf{X})\partial\mathbf{X},$$
$$\partial \mathbf{g}(\mathbf{X}) = \nabla \mathbf{g}(\mathbf{X})\partial\mathbf{X}.$$

Поскольку $\mathbf{g}(\mathbf{X}) = \mathbf{0}$, $\partial\mathbf{g}(\mathbf{X}) = \mathbf{0}$ в допустимой области. Отсюда следует, что

$$\partial f(\mathbf{X}) - \nabla f(\mathbf{X})\partial\mathbf{X} = 0,$$
$$\nabla \mathbf{g}(\mathbf{X})\partial\mathbf{X} = \mathbf{0}.$$

Как видим, задача сводится к решению $m + 1$ уравнений с $n + 1$ неизвестными, которыми являются $\partial f(\mathbf{X})$ и $\partial \mathbf{X}$. Неизвестную величину $\partial f(\mathbf{X})$ можно определить, как только будет найден вектор $\partial \mathbf{X}$. Это означает, что, по существу, имеется m уравнений с n неизвестными.

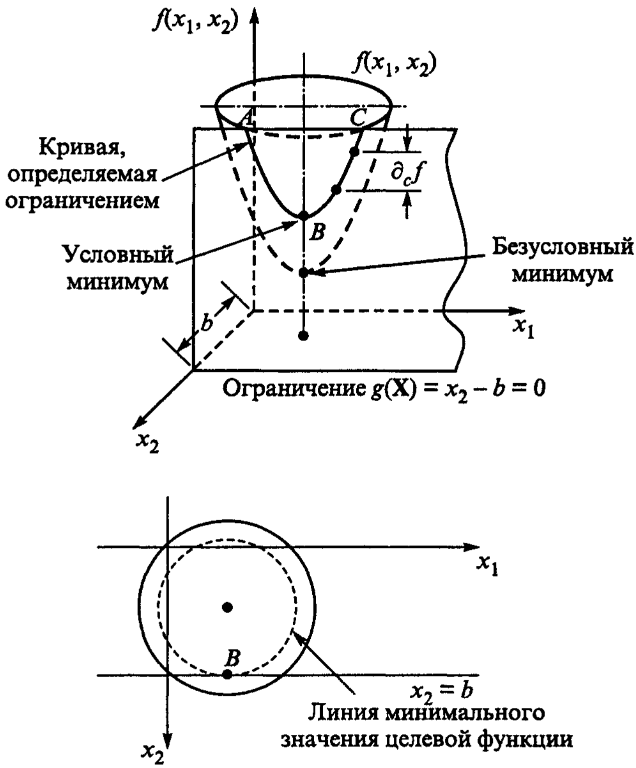


Рис. 20.4

При $m > n$ по меньшей мере $m - n$ уравнений системы являются избыточными. После устранения избыточности количество независимых уравнений в системе становится равным m ($\leq n$). Если $m = n$, решением является $\partial \mathbf{X} = 0$. При этом точка $\mathbf{X}$ не имеет допустимой окрестности, и, следовательно, пространство решений задачи состоит из единственной точки. Такая ситуация является тривиальной. Ситуацию, когда $m < n$, рассмотрим подробно.

Пусть

$$\mathbf{X} = (\mathbf{Y}, \mathbf{Z}),$$

где

$$\mathbf{Y} = (y_1, y_2, \ldots, y_m) \text{ и } \mathbf{Z} = (z_1, z_2, \ldots, z_{n-m})$$

называются *зависимыми* и *независимыми* переменными соответственно. Переписывая градиенты функций f и $\mathbf{g}$ в новых обозначениях, получим

$$\nabla f(\mathbf{Y}, \mathbf{Z}) = (\nabla_\mathbf{Y} f, \nabla_\mathbf{Z} f),$$

$$\nabla \mathbf{g}(\mathbf{Y}, \mathbf{Z}) = (\nabla_\mathbf{Y} \mathbf{g}, \nabla_\mathbf{Z} \mathbf{g}).$$

Введем в рассмотрение матрицы

$$\mathbf{J} = \nabla_{\mathbf{Y}}\mathbf{g} = \begin{bmatrix} \nabla_{\mathbf{Y}} g_1 \\ \vdots \\ \nabla_{\mathbf{Y}} g_m \end{bmatrix},$$

$$\mathbf{C} = \nabla_{\mathbf{Z}}\mathbf{g} = \begin{bmatrix} \nabla_{\mathbf{Z}} g_1 \\ \vdots \\ \nabla_{\mathbf{Z}} g_m \end{bmatrix}.$$

Матрица $\mathbf{J}_{m\times m}$ называется **матрицей Якоби**, а $\mathbf{C}_{m\times(n-m)}$ — **матрицей управления**. Матрица Якоби $\mathbf{J}$ предполагается невырожденной. Это всегда можно обеспечить, поскольку рассматриваемые m уравнений являются независимыми по определению. Поэтому компоненты вектора $\mathbf{Y}$ можно выбрать среди компонентов вектора $\mathbf{X}$ таким образом, что матрица $\mathbf{J}$ окажется невырожденной.

Исходную систему уравнений с неизвестными $\partial f(\mathbf{X})$ и $\partial \mathbf{X}$ можно переписать в следующем виде

$$\partial f(\mathbf{Y}, \mathbf{Z}) = \nabla_{\mathbf{Y}} f \partial \mathbf{Y} + \nabla_{\mathbf{Z}} f \partial \mathbf{Z},$$
$$\mathbf{J}\partial \mathbf{Y} = -\mathbf{C}\partial \mathbf{Z}.$$

Так как матрица $\mathbf{J}$ невырожденная, существует обратная матрица $\mathbf{J}^{-1}$. Следовательно,

$$\partial \mathbf{Y} = -\mathbf{J}^{-1}\mathbf{C}\partial \mathbf{Z}.$$

Подставляя это выражение $\partial \mathbf{Y}$ в уравнение для $\partial f(\mathbf{Y}, \mathbf{Z})$, можно выразить ∂f через $\partial \mathbf{Z}$ в следующем виде

$$\partial f(\mathbf{Y}, \mathbf{Z}) = (\nabla_{\mathbf{Z}} f - \nabla_{\mathbf{Y}} f \mathbf{J}^{-1}\mathbf{C})\partial \mathbf{Z}.$$

Из этого уравнения получаем формулу для производных функции f по вектору независимых переменных $\mathbf{Z}$

$$\nabla_c f = \frac{\partial_c f(\mathbf{Y}, \mathbf{Z})}{\partial_c \mathbf{Z}} = \nabla_{\mathbf{Z}} f - \nabla_{\mathbf{Y}} f \mathbf{J}^{-1}\mathbf{C},$$

где $\nabla_c f$ представляет вектор **приведенного градиента** функции f по $\mathbf{Z}$. Следовательно, вектор $\nabla_c f(\mathbf{Y}, \mathbf{Z})$ должен обращаться в нуль в стационарных точках.

Достаточные условия экстремума в стационарной точке аналогичны изложенным в разделе 20.2. В этом случае элементы матрицы Гессе будут соответствовать компонентам вектора независимых переменных $\mathbf{Z}$. Между тем элементы матрицы Гессе должны быть *приведенными* вторыми производными. Чтобы показать, как они вычисляются, положим

$$\nabla_c f = \nabla_{\mathbf{Z}} f - \mathbf{WC}.$$

Отсюда следует, что i-й строкой приведенной матрицы Гессе является вектор $\partial \nabla_c f / \partial z_i$. Заметим, что $\mathbf{W}$ — функция от $\mathbf{Y}$, а $\mathbf{Y}$, в свою очередь, — функция от $\mathbf{Z}$. Следовательно, при вычислении частной производной $\nabla_c f$ по z_i следует применять правило дифференцирования сложной функции, а именно

$$\frac{\partial w_j}{\partial z_i} = \frac{\partial w_j}{\partial y_j}\frac{\partial y_j}{\partial z_i}.$$

Пример 20.3–1

Рассмотрим следующую задачу.

$$f(\mathbf{X}) = x_1^2 + 3x_2^2 + 5x_1x_3^2,$$

$$g_1(\mathbf{X}) = x_1x_3 + 2x_2 + x_2^2 - 11 = 0,$$

$$g_2(\mathbf{X}) = x_1^2 + 2x_1x_2 + x_3^2 - 14 = 0.$$

Известна допустимая точка $\mathbf{X}^0 = (1, 2, 3)$. Требуется оценить приращение функции $f(= \partial_c f)$ в допустимой окрестности точки $\mathbf{X}^0$.

Пусть $\mathbf{Y} = (x_1, x_3)$ и $\mathbf{Z} = x_2$. Имеем

$$\nabla_{\mathbf{Y}} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_3}\right) = \left(2x_1 + 5x_3^2, 10x_1x_3\right),$$

$$\nabla_{\mathbf{Z}} f = \frac{\partial f}{\partial x_2} = 6x_2,$$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} x_3 & x_1 \\ 2x_1 + 2x_2 & 2x_3 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_2 + 2 \\ 2x_1 \end{bmatrix}.$$

Оценка приращения $\partial_c f$ в окрестности допустимой точки $\mathbf{X}^0 = (1, 2, 3)$, вызванного малым изменением $\partial x_2 = 0.01$, получается следующим образом:

$$\mathbf{J}^{-1}\mathbf{C} = \begin{bmatrix} 3 & 1 \\ 6 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{6}{12} & -\frac{1}{12} \\ -\frac{6}{12} & \frac{3}{12} \end{bmatrix} \begin{bmatrix} 6 \\ 2 \end{bmatrix} \approx \begin{bmatrix} 2.83 \\ -2.50 \end{bmatrix}.$$

Следовательно, приращение функции f с учетом ограничений равно

$$\partial_c f = \left(\nabla_{\mathbf{Z}} f - \nabla_{\mathbf{Y}} f \mathbf{J}^{-1}\mathbf{C}\right)\partial\mathbf{Z} = \left(6 \times 2 - (47, 30)\begin{bmatrix} 2.83 \\ -2.5 \end{bmatrix}\right)\partial x_2 \approx -46\partial x_2 = -0.46.$$

Если указана величина изменения ∂x_2 *независимой* переменной x_2, то допустимые значения ∂x_1 и ∂x_3 зависимых переменных x_1 и x_3 определяются по формуле

$$\partial\mathbf{Y} = -\mathbf{J}^{-1}\mathbf{C}\partial\mathbf{Z}.$$

При $\partial x_2 = 0.01$ получаем

$$\begin{bmatrix} \partial x_1 \\ \partial x_3 \end{bmatrix} = -\mathbf{J}^{-1}\mathbf{C}\partial x_2 = \begin{bmatrix} -0.0283 \\ 0.0250 \end{bmatrix}.$$

Чтобы проверить точность полученной выше оценки для $\partial_c f$, можно вычислить значения функции f в точках $\mathbf{X}^0$ и $\mathbf{X}^0 + \partial\mathbf{X}$. Получаем

$$\mathbf{X}^0 + \partial\mathbf{X} = (1 - 0.0283,\ 2 + 0.01,\ 3 + 0.025) = (0.9717,\ 2.01,\ 3.025).$$

Отсюда следует, что

$$f(\mathbf{X}^0) = 58 \text{ и } f(\mathbf{X}^0 + \partial\mathbf{X}) = 57.523$$

или

$$\partial_c f = f(\mathbf{X}^0 + \partial\mathbf{X}) - f(\mathbf{X}^0) = -0.477.$$

Полученный результат свидетельствует об уменьшении значения функции f по сравнению с вычисленным выше в соответствии с формулой для $\partial_c f$. Это различие между двумя полученными результатами (–0.477 и –0.46) является следствием линейной аппроксимации в окрестности точки $\mathbf{X}^0$. Поэтому приведенная выше формула дает хорошие результаты лишь тогда, когда отклонения от точки $\mathbf{X}^0$ малы.

Упражнение 20.3,a

1. Вернитесь к задаче из примера 20.3–1.

   a) Вычислите величину $\partial_c f$ двумя способами, как это было проделано в примере, используя $\partial x_2 = 0.001$ вместо $\partial x_2 = 0.01$. Будет ли влияние линейной аппроксимации менее существенным при уменьшении величины ∂x_2?

   b) Установите зависимость между приращениями ∂x_1, ∂x_2 и ∂x_3 в допустимой точке $\mathbf{X}^0 = (1, 2, 3)$ при условии, что точка $(x_1^0 + \partial x_1, x_2^0 + \partial x_2, x_3^0 + \partial x_3)$ также является допустимой.

   c) Если $\mathbf{Y} = (x_2, x_3)$ и $\mathbf{Z} = x_1$, то каким должно быть значение ∂x_1, чтобы величина приращения $\partial_c f$ была такая же, как в рассмотренном примере?

   d) Проверьте, что результат, полученный в п. c), приведет к приращению $\partial_c f = -0.46$.

Пример 20.3–2

Данный пример иллюстрирует процедуру использования метода приведенного градиента. Рассмотрим задачу

$$\text{минимизировать } f(\mathbf{X}) = x_1^2 + x_2^2 + x_3^2$$

при ограничениях

$$g_1(\mathbf{X}) = x_1 + x_2 + 3x_3 - 2 = 0,$$
$$g_2(\mathbf{X}) = 5x_1 + 2x_2 + x_3 - 5 = 0.$$

Определяем экстремальные точки целевой функции при наличии ограничений следующим образом. Пусть $\mathbf{Y} = (x_1, x_2)$ и $\mathbf{Z} = x_3$. Тогда

$$\nabla_{\mathbf{Y}} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right) = (2x_1, 2x_2), \ \nabla_{\mathbf{Z}} f = \frac{\partial f}{\partial x_3} = 2x_3,$$

$$\mathbf{J} = \begin{bmatrix} 1 & 1 \\ 5 & 2 \end{bmatrix}, \quad \mathbf{J}^{-1} = \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{5}{3} & -\frac{1}{3} \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Следовательно,

$$\nabla_c f = \frac{\partial_c f}{\partial_c x_3} = 2x_3 - (2x_1, 2x_2)\begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{5}{3} & -\frac{1}{3} \end{bmatrix}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \frac{10}{3}x_1 - \frac{28}{3}x_2 + 2x_3.$$

В стационарной точке выполняется равенство$\nabla_c f = 0$, которое вместе с ограничениями $g_1(\mathbf{X}) = 0$ и $g_2(\mathbf{X}) = 0$ определяет искомую стационарную точку (или точки). В данном случае система уравнений

$$\begin{bmatrix} 10 & -28 & 6 \\ 1 & 1 & 3 \\ 5 & 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix}$$

имеет решение $\mathbf{X}^0 \approx (0.81, 0.35, 0.28)$.

Далее устанавливаем тип полученной стационарной точки путем проверки выполнения достаточных условий экстремума. Так как x_3 — независимая переменная, из равенства $\nabla_c f = 0$ следует, что

$$\frac{\partial_c^2 f}{\partial_c x_3^2} = \frac{10}{3}\left(\frac{dx_1}{dx_3}\right) - \frac{28}{3}\left(\frac{dx_2}{dx_3}\right) + 2 = \left(\frac{10}{3}, -\frac{28}{3}\right)\begin{bmatrix} \frac{dx_1}{dx_3} \\ \frac{dx_2}{dx_3} \end{bmatrix} + 2.$$

В соответствии с методом Якоби получаем

$$\begin{bmatrix} \frac{dx_1}{dx_3} \\ \frac{dx_2}{dx_3} \end{bmatrix} = -\mathbf{J}^{-1}\mathbf{C} = \begin{bmatrix} \frac{5}{3} \\ -\frac{14}{3} \end{bmatrix}.$$

Теперь путем подстановки находим, что $\frac{\partial_c^2 f}{\partial_c x_3^2} = \frac{460}{9} > 0$. Следовательно, $\mathbf{X}^0$ — точка минимума.

Использование описанного метода Якоби в общем случае может быть затруднено вычислением обратной матрицы $\mathbf{J}^{-1}$, если количество ограничений задачи велико. Однако этого можно избежать, воспользовавшись правилом Крамера, которое позволяет выразить ∂f через $\partial\mathbf{Z}$. Если z_j представляет собой j-ю компоненту вектора $\mathbf{Z}$, а y_i — i-я компонента вектора $\mathbf{Y}$, то можно показать, что

$$\frac{\partial_c f}{\partial_c z_j}=\frac{\dfrac{\partial(f,g_1,\ldots,g_m)}{\partial(z_j,y_1,\ldots,y_m)}}{\dfrac{\partial(g_1,\ldots,g_m)}{\partial(y_1,\ldots,y_m)}},$$

где

$$\frac{\partial(f,g_1,\ldots,g_m)}{\partial(z_j,y_1,\ldots,y_m)}=\begin{vmatrix}\dfrac{\partial f}{\partial z_j} & \dfrac{\partial f}{\partial y_1} & \cdots & \dfrac{\partial f}{\partial y_m}\\ \dfrac{\partial g_1}{\partial z_j} & \dfrac{\partial g_1}{\partial y_1} & \cdots & \dfrac{\partial g_1}{\partial y_m}\\ \cdots & \cdots & \cdots & \cdots\\ \dfrac{\partial g_m}{\partial z_j} & \dfrac{\partial g_m}{\partial y_1} & \cdots & \dfrac{\partial g_m}{\partial y_m}\end{vmatrix}$$

и

$$\frac{\partial(g_1,\ldots,g_m)}{\partial(y_1,\ldots,y_m)}=\begin{vmatrix}\dfrac{\partial g_1}{\partial y_1} & \cdots & \dfrac{\partial g_1}{\partial y_m}\\ \cdots & \cdots & \cdots\\ \dfrac{\partial g_m}{\partial y_1} & \cdots & \dfrac{\partial g_m}{\partial y_m}\end{vmatrix}=|\mathbf{J}|.$$

Следовательно, необходимые условия экстремума принимают вид

$$\frac{\partial_c f}{\partial_c z_j}=0,\qquad j=1,2,\ldots,n-m.$$

Аналогично в матричном равенстве

$$\frac{\partial\mathbf{Y}}{\partial\mathbf{Z}}=-\mathbf{J}^{-1}\mathbf{C}$$

(i, j)-й элемент вычисляется по формуле

$$\frac{\partial y_i}{\partial z_j}=-\frac{\dfrac{\partial(g_1,\ldots,g_m)}{\partial(y_1,\ldots,y_{i-1},z_j,y_{i+1},\ldots,y_m)}}{\dfrac{\partial(g_1,\ldots,g_m)}{\partial(y_1,\ldots,y_m)}},$$

которая представляет собой скорость изменения зависимой переменной y_i, вызванного изменением независимой переменной z_j.

Для получения достаточного условия наличия экстремума i-й элемент вектора $\mathbf{W} \equiv \nabla_{\mathbf{Y}} f \mathbf{J}^{-1}$ выразим в следующем виде

$$w_i = \frac{\dfrac{\partial(g_1, \ldots, g_{i-1}, f, g_{i+1}, \ldots, g_m)}{\partial(y_1, \ldots, y_m)}}{\dfrac{\partial(g_1, \ldots, g_m)}{\partial(y_1, \ldots, y_m)}}.$$

В качестве иллюстрации использования описанного выше метода установим необходимое условие экстремума в задаче из примера 20.3–2. В этом случае имеем

$$\frac{\partial_c f}{\partial_c x_3} = \frac{\begin{vmatrix} 2x_3 & 2x_1 & 2x_2 \\ 3 & 1 & 1 \\ 1 & 5 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ 5 & 2 \end{vmatrix}} = \frac{10}{3}x_1 - \frac{28}{3}x_2 + 2x_3.$$

Анализ чувствительности с помощью метода Якоби. Метод Якоби можно использовать для анализа чувствительности оптимального значения целевой функции f к малым изменениям правых частей ограничений оптимизационной задачи. В частности, можно определить, как повлияет на оптимальное значение функции f замена ограничения $g_i(\mathbf{X}) = 0$ на $g_i(\mathbf{X}) = \partial g_i$. Исследование такого типа именуется **анализом чувствительности** и в некотором смысле аналогично соответствующей процедуре, которая была рассмотрена в линейном программировании (глава 4). Следует однако заметить, что анализ чувствительности в нелинейном программировании приводит к результатам, справедливым лишь в малой окрестности экстремальной точки. Знакомство с такими процедурами будет полезно при изучении метода множителей Лагранжа.

Выше было показано, что

$$\partial f(\mathbf{Y}, \mathbf{Z}) = \nabla_{\mathbf{Y}} f \partial \mathbf{Y} + \nabla_{\mathbf{Z}} f \partial \mathbf{Z},$$
$$\partial \mathbf{g} = \mathbf{J}\, \partial \mathbf{Y} + \mathbf{C}\, \partial \mathbf{Z}.$$

Предположим, что $\partial \mathbf{g} \neq \mathbf{0}$, тогда

$$\partial \mathbf{Y} = \mathbf{J}^{-1} \partial \mathbf{g} - \mathbf{J}^{-1} \mathbf{C}\, \partial \mathbf{Z}.$$

Подставляя последнее выражение в уравнение для $\partial f(\mathbf{Y}, \mathbf{Z})$, получаем

$$\partial f(\mathbf{Y}, \mathbf{Z}) = \nabla_{\mathbf{Y}_0} f \mathbf{J}^{-1} \partial \mathbf{g} + \nabla_c f \partial \mathbf{Z},$$

где

$$\nabla_c f = \nabla_{\mathbf{Z}} f - \nabla_{\mathbf{Y}_0} f \mathbf{J}^{-1} \mathbf{C},$$

как было определено ранее. Полученное выражение для $\partial f(\mathbf{Y}, \mathbf{Z})$ можно использовать при анализе изменений целевой функции f в малой окрестности допустимой точки $\mathbf{X}^0$, обусловленных малыми изменениями $\partial \mathbf{g}$ и $\partial \mathbf{Z}$.

В экстремальной (фактически любой стационарной) точке $\mathbf{X}_0 = (\mathbf{Y}_0, \mathbf{Z}_0)$ приведенный градиент $\nabla_c f$ должен быть равен нулю. Следовательно, в точке $\mathbf{X}_0$ имеем

$$\partial f(\mathbf{Y}_0, \mathbf{Z}_0) = \nabla_{\mathbf{Y}_0} f \mathbf{J}^{-1} \partial \mathbf{g}(\mathbf{Y}_0, \mathbf{Z}_0)$$

или

$$\frac{\partial f}{\partial \mathbf{g}} = \nabla_{\mathbf{Y}_0} f \mathbf{J}^{-1},$$

вычисленному в точке $\mathbf{X}_0$. Это значит, что влияние малых изменений $\partial \mathbf{g}$ на *оптимальное* значение функции f можно оценить через скорость изменения функции f по отношению к изменениям $\mathbf{g}$. Эти величины принято называть **коэффициентами чувствительности.**

Пример 20.3–3

Рассмотрим задачу из примера 20.3–2. Оптимальной точкой является $\mathbf{X}_0 = (x_1^0, x_2^0, x_3^0) = (0.81,\ \ 0.35,\ \ 0.28)$. Так как $\mathbf{Y}_0 = (x_1^0, x_2^0)$, то

$$\nabla_{\mathbf{Y}_0} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right) = (2x_1^0, 2x_2^0) = (1.62, 0.70).$$

Следовательно,

$$\left(\frac{\partial f}{\partial g_1}, \frac{\partial f}{\partial g_2} \right) = \nabla_{\mathbf{Y}_0} f \mathbf{J}^{-1} = (1.62, 0.7) \begin{bmatrix} -\frac{2}{3} & \frac{1}{3} \\ \frac{5}{3} & -\frac{1}{3} \end{bmatrix} = (0.0876, 0.3067).$$

Это свидетельствует о том, что изменение $\partial g_1 = 1$ приводит к увеличению значения целевой функции f *приблизительно* на 0.0867. Аналогично при $\partial g_2 = 1$ имеет место увеличение значения f *приблизительно* на 0.3067.

Использование метода Якоби в задачах линейного программирования. Рассмотрим задачу линейного программирования.

$$\text{Максимизировать } z = 2x_1 + 3x_2$$

при ограничениях

$$\begin{aligned} x_1 + x_2 + x_3 &= 5, \\ x_1 - x_2 + x_4 &= 3, \\ x_1, x_2, x_3, x_4 &\ge 0. \end{aligned}$$

Для каждого условия неотрицательности $x_j \ge 0$ введем соответствующую (неотрицательную) дополнительную переменную w_j^2. Таким образом, $x_j - w_j^2 = 0$ или $x_j = w_j^2$. При такой замене условия неотрицательности становятся неявными, и исходная задача принимает вид

$$\text{Максимизировать } \ z = 2w_1^2 + 3w_2^2$$

при ограничениях

$$\begin{aligned} w_1^2 + w_2^2 + w_3^2 &= 5, \\ w_1^2 - w_2^2 + w_4^2 &= 3. \end{aligned}$$

Для метода Якоби положим $\mathbf{Y} = (w_1, w_2)$ и $\mathbf{Z} = (w_3, w_4)$. (В терминологии линейного программирования векторы $\mathbf{Y}$ и $\mathbf{Z}$ образованы базисными и небазисными переменными соответственно.) Следовательно,

$$\mathbf{J} = \begin{bmatrix} 2w_1 & 2w_2 \\ 2w_1 & -2w_2 \end{bmatrix}, \quad \mathbf{J}^{-1} = \begin{bmatrix} \frac{1}{4w_1} & \frac{1}{4w_1} \\ \frac{1}{4w_2} & \frac{-1}{4w_2} \end{bmatrix}, \quad w_1 \text{ и } w_2 \neq 0,$$

$$\mathbf{C} = \begin{bmatrix} 2w_3 & 0 \\ 0 & 2w_4 \end{bmatrix}, \nabla_{\mathbf{Y}} f = (4w_1, 6w_2), \nabla_{\mathbf{Z}} f = (0,0),$$

так что

$$\nabla_c f = (0,0) - (4w_1, 6w_2) \begin{bmatrix} \frac{1}{4w_1} & \frac{1}{4w_1} \\ \frac{1}{4w_2} & \frac{-1}{4w_2} \end{bmatrix} \begin{bmatrix} 2w_3 & 0 \\ 0 & 2w_4 \end{bmatrix} = (-5w_3, w_4).$$

Решение системы уравнений, состоящей из уравнения $\nabla_c f = \mathbf{0}$ и ограничений задачи, позволяет определить стационарную точку $w_1 = 2$, $w_2 = 1$, $w_3 = 0$, $w_4 = 0$. При этом матрица Гессе имеет вид

$$\mathbf{H}_c = \begin{bmatrix} \frac{\partial_c^2 f}{\partial_c w_3^2} & \frac{\partial_c^2 f}{\partial_c w_3 \partial_c w_4} \\ \frac{\partial_c^2 f}{\partial_c w_3 \partial_c w_4} & \frac{\partial_c^2 f}{\partial_c w_4^2} \end{bmatrix} = \begin{bmatrix} -5 & 0 \\ 0 & 1 \end{bmatrix}.$$

Поскольку $\mathbf{H}_c$ является неопределенной матрицей, стационарная точка не будет точкой максимума.

На самом деле полученный результат не является неожиданным, поскольку (небазисные) переменные w_3 и w_4 (и, следовательно, x_3 и x_4) равняются нулю, как и предусмотрено теорией линейного программирования. Следовательно, решение, полученное с помощью метода Якоби, определяет ту или иную крайнюю точку пространства допустимых решений рассматриваемой задачи, которая определяется конкретным выбором векторов $\mathbf{Y}$ и $\mathbf{Z}$. При этом найденное решение может как быть, так и не быть оптимальным. Однако метод Якоби позволяет распознать оптимальное решение задачи путем использования достаточных условий экстремума.

Приведенные рассуждения свидетельствуют о том, что для нахождения оптимального решения задачи линейного программирования необходимо менять конкретный выбор $\mathbf{Y}$ и $\mathbf{Z}$, пока не будут выполнены достаточные условия. Пусть, например, $\mathbf{Y} = (w_2, w_4)$ и $\mathbf{Z} = (w_1, w_3)$. Тогда соответствующий приведенный градиент принимает вид

$$\nabla_c f = (4w_1, 0) - (6w_2, 0) \begin{bmatrix} \frac{1}{2w_2} & 0 \\ \frac{1}{2w_4} & \frac{1}{2w_4} \end{bmatrix} \begin{bmatrix} 2w_1 & 2w_3 \\ 2w_1 & 0 \end{bmatrix} = (-2w_1, 6w_3).$$

Соответствующая стационарная точка определяется значениями $w_1 = 0$, $w_2 = \sqrt{5}$, $w_3 = 0$, $w_4 = \sqrt{8}$. Так как матрица

$$\mathbf{H}_c = \begin{bmatrix} -2 & 0 \\ 0 & -6 \end{bmatrix}$$

является отрицательно определенной, то указанная точка — точка максимума.

Полученный результат можно проверить графически, используя рис. 20.5. Первое из найденных решений ($x_1 = 4$, $x_2 = 1$) не является оптимальным, в то время как второе ($x_1 = 0$, $x_2 = 5$) оказывается таковым. Читатель имеет возможность проверить, что две оставшиеся крайние точки пространства допустимых решений не являются точками максимума. Кроме того, с использованием достаточных условий экстремума можно показать, что в точке $x_1 = 0$, $x_2 = 0$ имеет место минимум целевой функции рассматриваемой задачи.

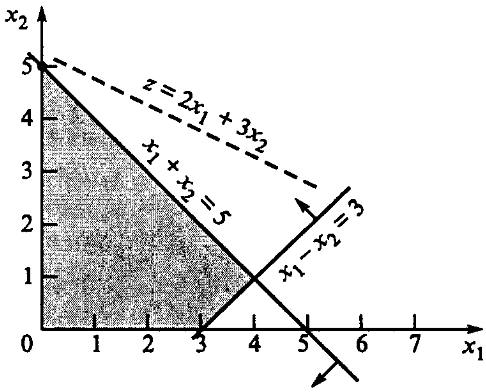


Рис. 20.5

Заметим, что введенные ранее коэффициенты чувствительности $\nabla_{\mathbf{Y}_0} f\mathbf{J}^{-1}$ в случае задачи линейного программирования являются переменными двойственной задачи. Чтобы проиллюстрировать это на рассматриваемом численном примере, обозначим через u_1 и u_2 соответствующие переменные двойственной задачи. В оптимальной точке $w_1 = 0$, $w_2 = \sqrt{5}$, $w_3 = 0$, $w_4 = \sqrt{8}$ переменные двойственной задачи вычисляются по формуле

$$(u_1, u_2) = \nabla_{\mathbf{Y}_0} \mathbf{J}^{-1} = (6w_2, 0)\begin{bmatrix} \dfrac{1}{2w_2} & 0 \\ \dfrac{1}{2w_4} & \dfrac{1}{2w_4} \end{bmatrix} = (3, 0).$$

В точке (3, 0) целевая функция двойственной задачи равна $5u_1 + 3u_2 = 15$ и совпадает с оптимальным значением целевой функции прямой задачи. Полученное решение также удовлетворяет ограничениям двойственной задачи, т.е. является допустимым и, следовательно, оптимальным. Это значит, что коэффициенты чувствительности совпадают с переменными двойственной задачи линейного программирования.

Из процедуры применения метода Якоби к решению задач линейного программирования можно сделать несколько общих выводов. Рассмотренный численный пример показывает, что в силу необходимых условий экстремума независимые переменные задачи равны нулю. Достаточные же условия указывают, что матрица Гессе является диагональной. Следовательно, все диагональные элементы этой матрицы должны быть положительными, если в рассматриваемой точке имеет место минимум, и отрицательными — в случае максимума.

Из вышесказанного следует, что необходимое условие экстремума равносильно указанию на то, что оптимальное решение задачи следует искать только среди ее базисных решений. В этом случае в качестве небазисных переменных задачи линейного программирования выступают независимые переменные. Достаточное же условие приводит к выводу, что между диагональными элементами матрицы Гессе и разностями $z_j - c_j$ задачи линейного программирования (см. раздел 7.4), получаемыми с помощью симплекс-метода, существует строгое соответствие[1].

Упражнения 20.3,b

1. Предположим, что задача из примера 20.3–2 решена следующим образом. Сначала из ограничений задачи выражаем переменные x_1 и x_2 через x_3; затем используем полученные соотношения для представления целевой функции в зависимости лишь от переменной x_3. Дифференцируя полученную целевую функцию по x_3, находим точки минимума и максимума.

   a) Будут ли производные новой целевой функции (выраженные через переменную x_3) отличаться от полученных с помощью метода Якоби?

   b) В чем основное отличие описанной процедуры решения задачи от метода Якоби?

2. Примените метод Якоби к решению задачи из примера 20.3–1, выбирая $\mathbf{Y} = (x_2, x_3)$ и $\mathbf{Z} = (x_1)$.

3. С помощью метода Якоби решите следующую задачу.

$$\text{Минимизировать } f(\mathbf{X}) = \sum_{i=1}^{n} x_i^2$$

   при ограничении

$$\prod_{i=1}^{n} x_i = C,$$

   где C — положительная константа. Пусть теперь правая часть ограничения равна $C + \delta$, где δ — малое положительное число. Найдите соответствующее изменение оптимального значения целевой функции f.

4. Решите методом Якоби следующую задачу.

$$\text{Минимизировать } f(\mathbf{X}) = 5x_1^2 + x_2^2 + 2x_1x_2$$

---

[1] Формальное доказательство справедливости этих утверждений для общей задачи линейного программирования можно найти в статье Taha H. and Curry G. “Classical Derivation of the Necessary and Sufficient Conditions for Optimal Linear Programs”, *Operations Research*, Vol.19, 1971, pp.1045–1049. В этой статье показано, что все основные положения симплекс-метода могут быть получены с помощью метода Якоби.

при ограничении

$$g(\mathbf{X}) = x_1x_2 - 10 = 0.$$

a) Найдите изменение оптимального значения целевой функции $f(\mathbf{X})$, если ограничение задачи примет вид $x_1x_2 - 9.99 = 0$.

b) Найдите изменение значения целевой функции $f(\mathbf{X})$ в окрестности допустимой точки (2, 5) при условии, что $x_1x_2 = 9.99$ и $\partial x_1 = 0.01$.

5. Решите следующую задачу.

$$\text{Максимизировать } f(\mathbf{X}) = x_1^2 + 2x_2^2 + 10x_3^2 + 5x_1x_2$$

при ограничениях

$$g_1(\mathbf{X}) = x_1 + x_2^2 + 3x_2x_3 - 5 = 0,$$

$$g_2(\mathbf{X}) = x_1^2 + 5x_1x_2 + x_3^2 - 7 = 0.$$

Используйте метод Якоби для нахождения $\partial f(\mathbf{X})$ в окрестности допустимой точки (1, 1, 1). Пусть указанная окрестность определяется условиями $\partial g_1 = -0.01$, $\partial g_2 = 0.02$ и $\partial x_1 = 0.01$.

6. Решите следующую задачу.

$$\text{Минимизировать } f(\mathbf{X}) = x_1^2 + x_2^2 + x_3^2 + x_4^2$$

при ограничениях

$$g_1(\mathbf{X}) = x_1 + 2x_2 + 3x_3 + 5x_4 - 10 = 0,$$
$$g_2(\mathbf{X}) = x_1 + 2x_2 + 5x_3 + 6x_4 - 15 = 0.$$

Покажите, что при выборе x_3 и x_4 в качестве независимых переменных метод Якоби не дает оптимального решения. Затем решите задачу, выбрав x_1 и x_3 в качестве независимых переменных, и примените достаточные условия для исследования полученной стационарной точки. Найдите коэффициенты чувствительности для этой задачи.

7. Решите следующую задачу линейного программирования.

$$\text{Максимизировать } f(\mathbf{X}) = \sum_{j=1}^{n} c_j x_j$$

при ограничениях

$$g_i(\mathbf{X}) = \sum_{j=1}^{n} a_{ij}x_j - b_i = 0, \quad i = 1, 2, \ldots, m,$$

$$x_j \geq 0, \; j = 1, 2, \ldots, n.$$

Покажите, что если в этой задаче пренебречь условиями неотрицательности переменных, то приведенные производные $\nabla_c f(\mathbf{X})$ совпадают с разностями $\{z_j - c_j\}$, определяемыми условием оптимальности задачи линейного программирования (раздел 7.5.1), а именно

$$\{z_j - c_j\} = \{\mathbf{C}_B\mathbf{B}^{-1}\mathbf{P}_j - c_j\} \text{ для всех } j.$$

Можно ли непосредственно применять метод приведенного градиента к решению задачи линейного программирования? Почему можно или почему нельзя?

Метод множителей Лагранжа. Коэффициенты чувствительности метода Якоби

$$\frac{\partial f}{\partial \mathbf{g}} = \nabla_{\mathbf{Y}_0} \mathbf{J}^{-1}$$

можно использовать для решения задач с ограничениями в виде равенств. Пусть

$$\boldsymbol{\lambda} = \nabla_{\mathbf{Y}_0} \mathbf{J}^{-1} = \frac{\partial f}{\partial \mathbf{g}}.$$

Следовательно,

$$\partial f - \boldsymbol{\lambda}\, \partial \mathbf{g} = 0.$$

Это уравнение соответствует необходимым условиям стационарности точек, так как формула для $\dfrac{\partial f}{\partial \mathbf{g}}$ получена с учетом того, что $\nabla_c f = \mathbf{0}$. Представленное уравнение можно записать в более удобной форме, если перейти к частным производным по всем переменным x_j, что приводит к системе уравнений

$$\frac{\partial}{\partial x_j}\left(f - \boldsymbol{\lambda}\mathbf{g}\right) = 0, \quad j = 1, 2, \ldots, n.$$

Полученные уравнения вместе с ограничениями задачи $\mathbf{g} = \mathbf{0}$ формируют систему, которая позволяет определить допустимые векторы $\mathbf{X}$ и $\boldsymbol{\lambda}$, удовлетворяющие *необходимым* условиям стационарности.

Описанная процедура составляет основу *метода множителей Лагранжа,* который позволяет определять стационарные точки задачи оптимизации с ограничениями в виде *равенств.* Формально схема этого метода представима в следующем виде. Пусть

$$L(\mathbf{X}, \boldsymbol{\lambda}) = f(\mathbf{X}) - \boldsymbol{\lambda}\, \partial \mathbf{g}(\mathbf{X}).$$

Функция L называется **функцией Лагранжа**, а параметры $\boldsymbol{\lambda}$ — **множителями Лагранжа**. Как следует из определения, эти множители имеют тот же смысл, что и коэффициенты чувствительности, фигурирующие в методе Якоби.

Уравнения

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = \mathbf{0} \quad \text{и} \quad \frac{\partial L}{\partial \mathbf{X}} = \mathbf{0}$$

дают необходимые условия для определения стационарных точек функции $f(\mathbf{X})$ при ограничениях $\mathbf{g}(\mathbf{X}) = \mathbf{0}$. Достаточные условия, используемые в методе множителей Лагранжа, будут сформулированы без доказательства. Определим матрицу

$$\mathbf{H}^B = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{P} \\ \hline \mathbf{P}^T & \mathbf{Q} \end{array}\right]_{(m+n)\times(m+n)},$$

где

$$\mathbf{P} = \begin{bmatrix} \nabla g_1(\mathbf{X}) \\ \vdots \\ \nabla g_m(\mathbf{X}) \end{bmatrix}_{m \times n} \quad \text{и} \quad \mathbf{Q} = \left\| \frac{\partial^2 L(\mathbf{X}, \boldsymbol{\lambda})}{\partial x_i \partial x_j} \right\|_{n \times n} \quad \text{для всех } i \text{ и } j.$$

Матрица $\mathbf{H}^B$ называется **окаймленной матрицей Гессе.**

Пусть имеется стационарная точка $(\mathbf{X}_0, \boldsymbol{\lambda}_0)$ функции Лагранжа $L(\mathbf{X}, \boldsymbol{\lambda})$ и окаймленная матрица Гессе $\mathbf{H}^B$ вычислена в точке $(\mathbf{X}_0, \boldsymbol{\lambda}_0)$. Тогда $\mathbf{X}_0$ является следующим.

1. Точкой максимума, если, начиная с углового минора порядка $2m + 1$, *последующие* $n - m$ угловых миноров окаймленной матрицы Гессе $\mathbf{H}^B$ образуют знакопеременный числовой ряд, в котором знак первого члена определяется множителем $(-1)^{m+1}$.
2. Точкой минимума, если, начиная с углового минора порядка $2m + 1$, *последующие* $n - m$ угловых миноров матрицы $\mathbf{H}^B$ имеют знаки, определяемые множителем $(-1)^m$.

Эти условия являются достаточными для определения экстремальной точки. Другими словами, экстремальной может оказаться стационарная точка, не удовлетворяющая этим условиям.

Существуют другие условия определения экстремальных точек задачи, которые являются как необходимыми, так и достаточными. Однако их практическое использование часто связано со значительными вычислительными трудностями. Определим матрицу

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{0} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{Q} - \mu \mathbf{I} \end{bmatrix},$$

вычисленную в стационарной точке $(\mathbf{X}_0, \boldsymbol{\lambda}_0)$, где μ — неизвестный параметр. Пусть $|\boldsymbol{\Delta}|$ — определитель матрицы $\boldsymbol{\Delta}$, тогда все $n - m$ действительных корней u_i полинома

$$|\boldsymbol{\Delta}| = 0$$

должны быть

1. отрицательными, если $\mathbf{X}_0$ — точка максимума,
2. положительными, если $\mathbf{X}_0$ — точка минимума.

Пример 20.3–4

Рассмотрим задачу из примера 20.3–2. Функция Лагранжа имеет вид

$$L(\mathbf{X}, \boldsymbol{\lambda}) = x_1^2 + x_2^2 + x_3^2 - \lambda_1(x_1 + x_2 + 3x_3 - 2) - \lambda_2(5x_1 + 2x_2 + x_3 - 5).$$

Необходимые условия экстремума записываются в виде следующей системы уравнений.

$$\frac{\partial L}{\partial x_1} = 2x_1 - \lambda_1 - 5\lambda_2 = 0,$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - \lambda_1 - 2\lambda_2 = 0,$$

$$\frac{\partial L}{\partial x_3} = 2x_3 - 3\lambda_1 - \lambda_2 = 0,$$

$$\frac{\partial L}{\partial \lambda_1} = -(x_1 + x_2 + 3x_3 - 2) = 0,$$

$$\frac{\partial L}{\partial \lambda_2} = -(5x_1 + 2x_2 + x_3 - 5) = 0.$$

Решением этой системы являются векторы

$$\mathbf{X}_0 = (x_1, x_2, x_3) = (0.81, 0.35, 0.28),$$

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2) = (0.0867, 0.3067).$$

В этом решении объединены результаты примеров 20.3–2 и 20.3–3. Как видим, значения множителей Лагранжа $\boldsymbol{\lambda}$ совпадают со значениями коэффициентов чувствительности, полученными в примере 20.3–3. Это указывает на то, что эти коэффициенты не зависят от выбора вектора зависимых переменных $\mathbf{Y}$ при реализации метода Якоби.

Чтобы показать, что найденная точка является точкой минимума, рассмотрим матрицу

$$\mathbf{H}^B = \begin{bmatrix} 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 5 & 2 & 1 \\ 1 & 5 & 2 & 0 & 0 \\ 1 & 2 & 0 & 2 & 0 \\ 3 & 1 & 0 & 0 & 2 \end{bmatrix}.$$

Здесь $n = 3$, $m = 2$ и $n - m = 1$. Следовательно, необходимо проверить только определитель матрицы $\mathbf{H}^B$, знак которого должен определяться множителем $(-1)^2$ в точке минимума. Так как $\det\mathbf{H}^B = 460 > 0$, то $\mathbf{X}_0$ является точкой минимума.

При решении систем уравнений, выражающих необходимые условия экстремума, иногда удобно применять метод, который заключается в последовательном выборе числовых значений $\boldsymbol{\lambda}$, после чего данная система решается относительно $\mathbf{X}$. Процедура повторяется до тех пор, пока вектор $\mathbf{X}$, соответствующий некоторому значению $\boldsymbol{\lambda}$, не будет удовлетворять всем ограничениям. Этот метод был продемонстрирован в главе 11 при решении задачи управления запасами с одним ограничением (см. пример 11.3–4). В вычислительном плане эта процедура является громоздкой. Можно использовать и другие численные методы решения систем уравнений, например метод Ньютона–Рафсона (раздел 20.2.2).

Пример 20.3–5

Рассмотрим задачу

$$\text{минимизировать} \quad z = x_1^2 + x_2^2 + x_3^2$$

при ограничении

$$4x_1 + x_2^2 + 2x_3 - 14 = 0.$$

Функция Лагранжа имеет вид

$$L(\mathbf{X}, \lambda) = x_1^2 + x_2^2 + x_3^2 - \lambda\left(4x_1 + x_2^2 + 2x_3 - 14\right).$$

Отсюда получаем необходимые условия экстремума в виде системы

$$\frac{\partial L}{\partial x_1} = 2x_1 - 4\lambda = 0,$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - 2\lambda x_2 = 0,$$

$$\frac{\partial L}{\partial x_3} = 2x_3 - 2\lambda = 0,$$

$$\frac{\partial L}{\partial \lambda} = -\left(4x_1 + x_2^2 + 2x_3 - 14\right) = 0,$$

решениями которой являются векторы

$(\mathbf{X}_0, \lambda_0)_1 = (2, 2, 1, 1)$,

$(\mathbf{X}_0, \lambda_0)_2 = (2, -2, 1, 1)$,

$(\mathbf{X}_0, \lambda_0)_3 = (2.8, 0, 1.4, 1.4)$.

Используя достаточные условия, вычислим элементы матрицы

$$\mathbf{H}^B = \begin{bmatrix} 0 & 4 & 2x_2 & 2 \\ 4 & 2 & 0 & 0 \\ 2x_2 & 0 & 2-2\lambda & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}.$$

Так как $m = 1$ и $n = 3$, то для того, чтобы стационарная точка была точкой минимума, знак последних $3 - 1 = 2$ угловых миноров должен определяться знаком множителя $(-1)^m = -1$. Таким образом, в точке $(\mathbf{X}_0, \lambda_0)_1 = (2, 2, 1, 1)$ имеем

$$\begin{vmatrix} 0 & 4 & 4 \\ 4 & 2 & 0 \\ 4 & 0 & 0 \end{vmatrix} = -32 < 0 \quad \text{и} \quad \begin{vmatrix} 0 & 4 & 4 & 2 \\ 4 & 2 & 0 & 0 \\ 4 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \end{vmatrix} = -64 < 0.$$

В точке $(\mathbf{X}_0, \lambda_0)_2 = (2, -2, 1, 1)$

$$\begin{vmatrix} 0 & 4 & -4 \\ 4 & 2 & 0 \\ -4 & 0 & 0 \end{vmatrix} = -32 < 0 \quad \text{и} \quad \begin{vmatrix} 0 & 4 & -4 & 2 \\ 4 & 2 & 0 & 0 \\ -4 & 0 & 0 & 0 \\ 2 & 0 & 0 & 2 \end{vmatrix} = -64 < 0.$$

Наконец, в точке $(\mathbf{X}_0, \lambda_0)_3 = (2.8, 0, 1.4, 1.4)$

$$\begin{vmatrix} 0 & 4 & 0 \\ 4 & 2 & 0 \\ 0 & 0 & -0.8 \end{vmatrix} = 12.8 > 0 \quad \text{и} \quad \begin{vmatrix} 0 & 4 & 0 & 2 \\ 4 & 2 & 0 & 0 \\ 0 & 0 & -0.8 & 0 \\ 2 & 0 & 0 & 2 \end{vmatrix} = 32 > 0.$$

Отсюда следует, что $(\mathbf{X}_0)_1$ и $(\mathbf{X}_0)_2$ — точки минимума. Тот факт, что в точке $(\mathbf{X}_0)_3$ не удовлетворяются достаточные условия максимума или минимума, не означает, что эта точка не является экстремальной. Это следствие того, что используемые условия являются лишь достаточными.

Для иллюстрации схемы проверки достаточных условий экстремума, которая использует корни полинома, рассмотрим матрицу

$$\Delta = \begin{bmatrix} 0 & 4 & 2x_2 & 2 \\ 4 & 2-\mu & 0 & 0 \\ 2x_2 & 0 & 2-2\lambda-\mu & 0 \\ 2 & 0 & 0 & 2-\mu \end{bmatrix}.$$

В точке $(\mathbf{X}_0, \lambda_0)_1 = (2, 2, 1, 1)$ имеем

$$|\Delta| = 9\mu^2 - 26\mu + 16 = 0,$$

откуда получаем $\mu = 2$ и $\mu = 8/9$. Так как все $\mu > 0$, то $(\mathbf{X}_0)_1 = (2, 2, 1)$ — точка минимума. В точке $(\mathbf{X}_0, \lambda_0)_2 = (2, -2, 1, 1)$

$$|\Delta| = 9\mu^2 - 26\mu + 16 = 0,$$

что совпадает с рассмотренным выше случаем. Следовательно, $(\mathbf{X}_0)_2 = (2, -2, 1)$ является точкой минимума. Наконец, в точке $(\mathbf{X}_0, \lambda_0)_3 = (2.8, 0, 1.4, 1.4)$

$$|\Delta| = 5\mu^2 - 6\mu - 8 = 0,$$

откуда получаем $\mu = 2$ и $\mu = -0.8$, что свидетельствует о том, что точка $(\mathbf{X}_0)_3 = (2.8, 0, 1.4)$ не является экстремальной.

Упражнения 20.3,с

1. Методом Якоби и методом множителей Лагранжа решите следующую задачу линейного программирования.

   Максимизировать $f(\mathbf{X}) = 5x_1 + 3x_2$

при ограничениях

$$g_1(\mathbf{X}) = x_1 + 2x_2 + x_3 - 6 = 0,$$

$$g_2(\mathbf{X}) = 3x_1 + x_2 + x_4 - 9 = 0,$$

$$x_1, x_2, x_3, x_4 \geq 0.$$

2. Найдите оптимальное решение следующей задачи.

$$\text{Минимизировать } f(\mathbf{X}) = x_1^2 + 2x_2^2 + 10x_3^2$$

при ограничениях

$$g_1(\mathbf{X}) = x_1 + x_2^2 + x_3 - 5 = 0,$$

$$g_2(\mathbf{X}) = x_1 + 5x_2 + x_3 - 7 = 0.$$

Пусть $g_1(\mathbf{X}) = 0.01$ и $g_2(\mathbf{X}) = 0.02$. Найдите величину соответствующего изменения оптимального значения функции $f(\mathbf{X})$.

3. Решите упр. 20.3,b(6) методом множителей Лагранжа и проверьте, что полученные значения множителей Лагранжа совпадают со значениями коэффициентов чувствительности, полученными ранее при решении этого упражнения.

20.3.2. Ограничения в виде неравенств

В данном разделе показано, как метод множителей Лагранжа можно обобщить на случай задачи с ограничениями в виде неравенств. Основную часть раздела составляет вывод условий Куна–Таккера, которые играют важную роль в нелинейном программировании.

Обобщенный метод множителей Лагранжа. Предположим, что дана задача, в которой требуется

$$\text{максимизировать } z = f(\mathbf{X})$$

при ограничениях

$$g_i(\mathbf{X}) \leq 0, \quad i = 1, 2, \ldots, m.$$

Если в рассматриваемой задаче имеются условия неотрицательности переменных $\mathbf{X} \geq \mathbf{0}$, то предполагается, что они включены в указанные m ограничений.

Основная идея обобщенного метода множителей Лагранжа состоит в том, что если точка *безусловного* оптимума функции $f(\mathbf{X})$ не удовлетворяет всем ограничениям задачи, то оптимальное решение задачи с ограничениями должно достигаться в граничной точке области допустимых решений. Следовательно, одно или несколько из m ограничений задачи должны выполняться как равенства. Поэтому вычислительная процедура обобщенного метода множителей Лагранжа включает следующие шаги.

Шаг 1. Решается задача без учета ограничений

$$\text{максимизировать } z = f(\mathbf{X}).$$

Если полученное оптимальное решение удовлетворяет всем ограничениям задачи, вычисления заканчиваются, так как все ограничения являются избыточными. Иначе следует положить $k = 1$ и перейти к шагу 2.

Шаг 2. Активизируются любые k ограничений задачи (т.е. преобразовываются в равенства) и оптимизируется функция $f(\mathbf{X})$ при наличии k ограничений методом множителей Лагранжа. Если оптимальное решение этой задачи является допустимым по отношению к остальным ограничениям исходной задачи, вычисления заканчиваются: полученная точка является *локальным* оптимумом[1]. Иначе нужно сделать активными другие k ограничений исходной задачи и повторить данный шаг. Если *все* подмножества, состоящие из k активных ограничений, не приводят к допустимому решению, следует перейти к шагу 3.

Шаг 3. Если $k = m$, вычисления прекращаются: задача не имеет допустимых решений. Иначе необходимо положить $k = k + 1$ и перейти к шагу 2.

В описанной вычислительной процедуре часто игнорируется то важное обстоятельство, что она *не гарантирует* получения глобального оптимума даже в тех случаях, когда задача имеет *единственное* оптимальное решение. Другим важным моментом является ошибочность интуитивного представления, что если $p < q$, то оптимальное значение целевой функции $f(\mathbf{X})$ при p ограничениях–равенствах всегда лучше, чем при q ограничениях–равенствах. В общем случае это имеет место лишь в том случае, когда набор из q ограничений является подмножеством набора из p ограничений. Рассматриваемый ниже пример служит иллюстрацией сказанного.

Пример 20.3–6

$$\text{Максимизировать } z = -(2x_1 - 5)^2 - (2x_2 - 1)^2$$

при ограничениях

$$x_1 + 2x_2 \le 2,$$
$$x_1, x_2 \ge 0.$$

Графическое представление данной задачи (рис. 20.6) призвано помочь в понимании аналитических выкладок. Как видим, целевая функция задачи является вогнутой, а область допустимых решений выпуклой. Отсюда следует, что эффективный алгоритм должен гарантировать нахождение глобального оптимума. Однако обобщенный метод множителей Лагранжа, как будет показано, позволяет найти лишь точку локального максимума.

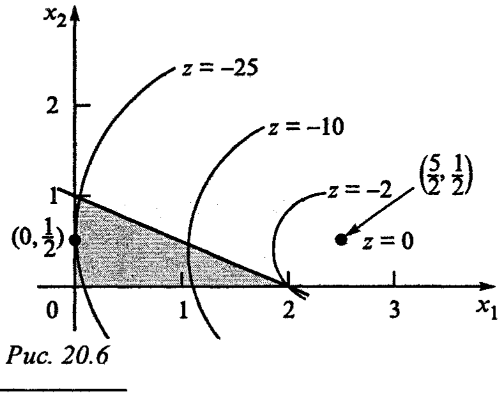


Рис. 20.6

[1] *Локальный* оптимум определяется из множества оптимумов, полученных в результате оптимизации целевой функции $f(\mathbf{X})$ при учете всех комбинаций из k ограничений–равенств, $k = 1, 2, ..., m$.

Точка безусловного экстремума находится как решение уравнений

$$\frac{\partial z}{\partial x_1} = -4(2x_1 - 5) = 0,$$

$$\frac{\partial z}{\partial x_2} = -4(2x_2 - 1) = 0.$$

Отсюда имеем $(x_1, x_2) = (5/2, 1/2)$. Так как это решение не удовлетворяет условию $x_1 + 2x_2 \le 2$, ограничения поочередно активизируются. Рассмотрим ограничение $x_1 = 0$. Функция Лагранжа в этом случае примет вид

$$L(x_1, x_2, \lambda) = -(2x_1 - 5)^2 - (2x_2 - 1)^2 - \lambda x_1.$$

Следовательно,

$$\frac{\partial L}{\partial x_1} = -4(2x_1 - 5) - \lambda = 0,$$

$$\frac{\partial L}{\partial x_2} = -4(2x_2 - 1) = 0,$$

$$\frac{\partial L}{\partial \lambda} = -x_1 = 0.$$

Решением этой системы будет точка $(x_1, x_2) = (0, 1/2)$, которая является точкой максимума, в чем можно убедиться, используя достаточное условие. Поскольку эта точка удовлетворяет остальным ограничениям исходной задачи, вычислительная процедура завершается: $(x_1, x_2) = (0, 1/2)$ — точка локального максимума задачи. (Заметим, что если поочередно активизировать оставшиеся ограничения задачи $x_2 \ge 0$ и $x_1 + 2x_2 \le 2$, то получаются недопустимые решения.) Оптимальное значение целевой функции $z = -25$.

Как видно из рис. 20.6, допустимому решению $(x_1, x_2) = (2, 0)$ рассматриваемой задачи, которое определяется точкой пересечения *двух* прямых $x_2 = 0$ и $x_1 + 2x_2 = 2$, соответствует значение целевой функции $z = -2$. Оно лучше значения функции z, которое получено с учетом одного активного ограничения.

Как следует из описанной вычислительной процедуры, при реализации обобщенного метода множителей Лагранжа на большее, чем получение приемлемого допустимого решения задачи, не следует рассчитывать. Особенно это проявляется тогда, когда целевая функция задачи не является одновершинной. Если в оптимизационной задаче имеется единственный (глобальный) условный экстремум (как в примере 20.3–6), то для его нахождения можно использовать откорректированный обобщенный метод множителей Лагранжа. Для этого необходимо провести сравнение безусловного оптимума и условных экстремумов, полученных с учетом *всех* возможных наборов активных ограничений, т.е. когда сначала отдельные ограничения делаются активными поочередно, затем рассматриваются пары активных ограничений и так до тех пор, пока все m ограничений рассматриваемой задачи не станут активными. Наилучший из *всех* таких *допустимых* экстремумов является глобальным.

Если эту процедуру применить к решению задачи из примера 20.3–6, то для определения глобального экстремума необходимо решить семь задач. Это свидетельствует о том, что возможности использования обобщенного метода множителей Лагранжа для решения практических задач ограничены.

Условия Куна–Таккера. Данный раздел посвящен рассмотрению *необходимых* условий Куна–Таккера, которые позволяют определять стационарные точки в задаче нелинейного программирования с ограничениями в виде неравенств. В основу изложения положен метод множителей Лагранжа. Эти условия являются также и достаточными, если выполняются определенные правила, которые описаны ниже.

Рассмотрим следующую задачу.

$$\text{Максимизировать } z = f(\mathbf{X})$$

при ограничениях

$$\mathbf{g}(\mathbf{X}) \le \mathbf{0}.$$

Ограничения–неравенства можно преобразовать в равенства с помощью *неотрицательных* дополнительных переменных. Пусть $S_i^2\,(\ge 0)$ — дополнительная переменная, которая прибавляется к левой части i-го ограничения $g_i(\mathbf{X}) \le 0$ и пусть

$$\mathbf{S} = (S_1, S_2, \ldots, S_m)^T \quad \text{и} \quad \mathbf{S}^2 = (S_1^2, S_2^2, \ldots, S_m^2)^T,$$

где m — общее количество ограничений–неравенств. Следовательно, функция Лагранжа записывается в виде

$$L(\mathbf{X}, \mathbf{S}, \boldsymbol{\lambda}) = f(\mathbf{X}) - \boldsymbol{\lambda}[\mathbf{g}(\mathbf{X}) + \mathbf{S}^2].$$

При ограничениях $\mathbf{g}(\mathbf{X}) \le \mathbf{0}$ необходимым условием оптимальности в задаче максимизации (минимизации) является неотрицательность (соответственно, неположительность) множителей $\boldsymbol{\lambda}$. Приведем обоснование этого. Рассмотрим задачу максимизации. Так как множители $\boldsymbol{\lambda}$ выражают скорость изменения целевой функции f по отношению к изменениям $\mathbf{g}$, т.е.

$$\boldsymbol{\lambda} = \frac{\partial f}{\partial \mathbf{g}},$$

то как только правая часть ограничения $\mathbf{g}(\mathbf{X}) \le \mathbf{0}$ увеличивается и становится больше нуля, область допустимых решений задачи расширяется и, следовательно, оптимальное значение целевой функции не может уменьшиться. Это значит, что $\boldsymbol{\lambda} \ge \mathbf{0}$. Аналогично в случае задачи минимизации при увеличении правой части ограничения оптимальное значение функции f не может увеличиться, откуда следует, что $\boldsymbol{\lambda} \le \mathbf{0}$. Если же ограничения задачи имеют вид равенств, т.е. $\mathbf{g}(\mathbf{X}) = \mathbf{0}$, то компоненты вектора $\boldsymbol{\lambda}$ по знаку не ограничены (см. упр. 20.3,d(2)).

Указанные выше ограничения на вектор $\boldsymbol{\lambda}$ должны рассматриваться как часть необходимых условий Куна–Таккера. Теперь получим остальные условия.

Вычисляя частные производные функции L по $\mathbf{X}$, $\mathbf{S}$ и $\boldsymbol{\lambda}$ и приравнивая их к нулю, получаем

$$\frac{\partial L}{\partial \mathbf{X}} = \nabla f(\mathbf{X}) - \boldsymbol{\lambda}\nabla \mathbf{g}(\mathbf{X}) = \mathbf{0},$$

$$\frac{\partial L}{\partial S_i} = -2\lambda_i S_i = 0, \quad i = 1, 2, \ldots, m,$$

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = -\left(\mathbf{g}(\mathbf{X}) + \mathbf{S}^2\right) = \mathbf{0}.$$

Из второй группы уравнений следуют такие результаты.

1. Если λ_i не равняется нулю, то $S_i^2 = 0$. Это означает, что соответствующий этому ограничению ресурс является дефицитным и, следовательно, полностью исчерпан (ограничение имеет вид равенства).
2. Если $S_i^2 > 0$, то $\lambda_i = 0$. Это значит, что i-й ресурс дефицитным не является и, следовательно, не влияет на значение целевой функции f (т.е. $\lambda_i = \frac{\partial f}{\partial g_i} = 0$).

Из второй и третьей групп уравнений следует, что

$$\lambda_i g_i(\mathbf{X}) = 0, \ i = 1, 2, \ldots, m.$$

Полученные условия фактически подтверждают предыдущий результат, ибо если $\lambda_i > 0$, то $g_i(\mathbf{X}) = 0$ или $S_1^2 = 0$. Аналогично при $g_i(\mathbf{X}) < 0$ $S_1^2 > 0$ и, следовательно, $\lambda_i = 0$.

Теперь для задачи максимизации можно сформулировать условия Куна–Таккера, необходимые для того, чтобы векторы $\mathbf{X}$ и $\boldsymbol{\lambda}$ определяли стационарную точку:

$$\boldsymbol{\lambda} \geq \mathbf{0},$$
$$\nabla f(\mathbf{X}) - \boldsymbol{\lambda}\nabla \mathbf{g}(\mathbf{X}) = \mathbf{0},$$
$$\lambda_i g_i(\mathbf{X}) = 0, \ i = 1, 2, \ldots, m,$$
$$\mathbf{g}(\mathbf{X}) \leq \mathbf{0}.$$

Читатель имеет возможность самостоятельно убедиться, что эти условия применимы также к задаче минимизации, за тем лишь исключением, что вектор $\boldsymbol{\lambda}$ должен быть неположительным, как это было установлено ранее. Как в случае задачи максимизации, так и задачи минимизации множители Лагранжа, соответствующие ограничениям в виде равенств, по знаку не ограничены.

Достаточность условий Куна–Таккера. Если целевая функция и область допустимых решений рассматриваемой задачи обладают определенными свойствами, связанными с выпуклостью и вогнутостью, то необходимые условия Куна–Таккера являются также достаточными. Упомянутые свойства перечислены в табл. 20.1.

Таблица 20.1

Тип оптимизации	Требуемые свойства: Целевая функция	Требуемые свойства: Область допустимых решений
Максимизация	Вогнутая	Выпуклое множество
Минимизация	Выпуклая	Выпуклое множество

Процедура проверки выпуклости или вогнутости некоторой функции является более простой, чем доказательство выпуклости множества допустимых решений экстремальной задачи. Так как выпуклость допустимого множества может быть установлена путем непосредственной проверки выпуклости или вогнутости функций ограничений, по этой причине мы приводим перечень требований, которые легче использовать на практике. Чтобы установить эти требования, рассмотрим задачу нелинейного программирования в общей постановке.

Максимизировать или минимизировать $z = f(\mathbf{X})$

при ограничениях

$$g_i(\mathbf{X}) \le 0, \quad i = 1, 2, \ldots, r,$$
$$g_i(\mathbf{X}) \ge 0, \quad i = r + 1, \ldots, p,$$
$$g_i(\mathbf{X}) = 0, \quad i = p + 1, \ldots, m.$$

При этом функция Лагранжа имеет вид

$$L(\mathbf{X},\mathbf{S},\boldsymbol{\lambda}) = f(\mathbf{X}) - \sum_{i=1}^{r} \lambda_i \left[g_i(\mathbf{X}) + S_i^2 \right] - \sum_{i=r+1}^{p} \lambda_i \left[g_i(\mathbf{X}) - S_i^2 \right] - \sum_{i=p+1}^{m} \lambda_i g_i(\mathbf{X}),$$

где λ_i — множитель Лагранжа, соответствующий i-му ограничению. Требования, которые устанавливают достаточность условий Куна–Таккера, приведены в табл. 20.2.

В табл. 20.2 представлена лишь часть условий, упомянутых в табл. 20.1. Это связано с тем, что область допустимых решений задачи может быть выпуклой и в то же время функции ограничений, которые ее формируют, не соответствуют требованиям, перечисленным в табл. 20.2.

ТАБЛИЦА 20.2

Тип оптимизации	Требуемые свойства			
	$f(\mathbf{X})$	$g_i(\mathbf{X})$	λ_i	
Максимизация	Вогнутая	Выпуклая	≥ 0	$(1 \le i \le r)$
		Вогнутая	≤ 0	$(r+1 \le i \le p)$
		Линейная	Нет ограничений	$(p+1 \le i \le m)$
Минимизация	Выпуклая	Выпуклая	≤ 0	$(1 \le i \le r)$
		Вогнутая	≥ 0	$(r+1 \le i \le p)$
		Линейная	Нет ограничений	$(p+1 \le i \le m)$

Законность положений табл. 20.2 основана на том, что при сформулированных условиях функция Лагранжа $L(\mathbf{X}, \mathbf{S}, \lambda)$ является вогнутой в случае задачи максимизации и выпуклой — в случае задачи минимизации. Это подтверждается тем, что если $g_i(\mathbf{X})$ — выпуклая функция, то функция $\lambda_i g_i(\mathbf{X})$ оказывается выпуклой при $\lambda_i \ge 0$ и вогнутой — при $\lambda_i \le 0$. Аналогичные рассуждения подтверждают все остальные условия. Заметим также, что линейная функция является как выпуклой, так и вогнутой. Наконец, если функция f вогнутая, то функция $-f$ является выпуклой, и наоборот.

Пример 20.3–7

Рассмотрим следующую задачу.

$$\text{Минимизировать } f(\mathbf{X}) = x_1^2 + x_2^2 + x_3^2$$

при ограничениях

$$g_1(\mathbf{X}) = 2x_1 + x_2 - 5 \le 0,$$
$$g_2(\mathbf{X}) = x_1 + x_3 - 2 \le 0,$$
$$g_3(\mathbf{X}) = 1 - x_1 \le 0,$$
$$g_4(\mathbf{X}) = 2 - x_2 \le 0,$$
$$g_5(\mathbf{X}) = -x_3 \le 0.$$

Поскольку рассматривается задача минимизации, то $\boldsymbol{\lambda} \le \mathbf{0}$. Поэтому условия Куна–Таккера записываются в следующем виде.

$$(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) \le \mathbf{0},$$

$$(2x_1, 2x_2, 2x_3) - (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5)\begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \mathbf{0},$$

$$\lambda_1 g_1 = \lambda_2 g_2 = \ldots = \lambda_5 g_5 = 0,$$
$$\mathbf{g}(\mathbf{X}) \le \mathbf{0}.$$

Эти условия можно упростить и привести к виду

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \le 0,$$
$$2x_1 - 2\lambda_1 - \lambda_2 + \lambda_3 = 0,$$
$$2x_2 - \lambda_1 + \lambda_4 = 0,$$
$$2x_3 - \lambda_2 + \lambda_5 = 0,$$
$$\lambda_1(2x_1 + x_2 - 5) = 0,$$
$$\lambda_2(x_1 + x_3 - 2) = 0,$$
$$\lambda_3(1 - x_1) = 0,$$
$$\lambda_4(2 - x_2) = 0,$$
$$\lambda_5 x_3 = 0,$$
$$2x_1 + x_2 \le 5,$$
$$x_1 + x_3 \le 2,$$
$$x_1 \ge 1, x_2 \ge 2, x_3 \ge 0.$$

Отсюда получаем решение: $x_1 = 1$, $x_2 = 2$, $x_3 = 0$, $\lambda_1 = \lambda_2 = \lambda_5 = 0$, $\lambda_3 = -2$, $\lambda_4 = -4$. Поскольку и целевая функция $f(\mathbf{X})$, и область допустимых решений $\mathbf{g}(\mathbf{X}) \le \mathbf{0}$ являются выпуклыми, функция $L(\mathbf{X}, \mathbf{S}, \boldsymbol{\lambda})$ также должна быть выпуклой, и полученная стационарная точка определяет глобальный минимум задачи. Рассмотренный пример показывает также, что решение системы, порожденной условиями Куна–Таккера, в явной форме может быть затруднительным. Следовательно, для численных расчетов описанная процедура не подходит. Тем не менее условия Куна–Таккера играют основополагающую роль при рассмотрении алгоритмов решения задач нелинейного программирования в главе 21.

Упражнения 20.3,d

1. Покажите, что условия Куна–Таккера для задачи

$$\text{максимизировать } f(\mathbf{X})$$

при ограничениях

$$\mathbf{g}(\mathbf{X}) \geq \mathbf{0}$$

совпадают с условиями, сформулированными в разделе 20.3.2, с той лишь разницей, что множители Лагранжа $\boldsymbol{\lambda}$ должны быть неположительными.

2. Дана следующая задача.

$$\text{Максимизировать } f(\mathbf{X})$$

при ограничениях

$$\mathbf{g}(\mathbf{X}) = \mathbf{0}.$$

Покажите, что условия Куна–Таккера для этой задачи имеют вид

$$\nabla f(\mathbf{X}) - \boldsymbol{\lambda}\nabla g(\mathbf{X}) = \mathbf{0},$$

$$\mathbf{g}(\mathbf{X}) = \mathbf{0},$$

$$\boldsymbol{\lambda} \text{ по знаку не ограничен.}$$

3. Запишите необходимые условия Куна–Таккера для следующих задач.

a) Максимизировать $f(\mathbf{X}) = x_1^3 - x_2^2 + x_1 x_3^2$

при ограничениях

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

$$5x_1^2 - x_2^2 - x_3 \geq 0,$$

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

b) Минимизировать $f(\mathbf{X}) = x_1^4 + x_2^2 + 5x_1 x_2 x_3$

при ограничениях

$$x_1^2 - x_2^2 + x_3^3 \leq 10,$$

$$x_1^3 + x_2^2 + 4x_3^2 \geq 20.$$

4. Дана задача

$$\text{максимизировать } f(\mathbf{X})$$

при ограничениях

$$\mathbf{g}(\mathbf{X}) = \mathbf{0}.$$

Пусть $f(\mathbf{X})$ — вогнутая функция, а $g_i(\mathbf{X})$ ($i = 1, 2, \ldots, m$) — *линейные* функции. Покажите, что в этом случае необходимые условия Куна–Таккера оказываются также и достаточными. Верно ли последнее утверждение, если $g_i(\mathbf{X})$ является выпуклой *нелинейной* функцией для всех i? Почему?

5. Дана задача

$$\text{максимизировать } f(\mathbf{X})$$

при ограничениях

$$g_1(\mathbf{X}) \geq 0,\ g_2(\mathbf{X}) = 0,\ g_3(\mathbf{X}) \leq 0.$$

Сформулируйте условия Куна–Таккера для данной задачи и установите требования к функциям, при выполнении которых выполняются указанные условия.

20.4. Заключение

В этой главе изложена классическая теория нахождения максимума и минимума в задачах нелинейного программирования как с ограничениями на переменные, так и без них. Как правило, классическая теория мало подходит для практического применения. Из этого правила существует несколько исключений, когда теоретические положения Куна–Таккера служат основой для создания эффективных вычислительных алгоритмов. В этом смысле *квадратичное программирование,* которое рассматривается в следующей главе, является примером использования необходимых условий Куна–Таккера.

Для задач нелинейного программирования с ограничениями в виде неравенств нет достаточных условий существования экстремума (подобных условиям для задач без ограничений и задач с ограничениями в виде *равенств*). Следовательно, не существует какого-либо приемлемого способа проверить, к какому экстремуму сходится тот или иной алгоритм нелинейного программирования — к локальному или глобальному, за исключением тех случаев, которые перечислены в табл. 20.1 и 20.2, когда такие условия можно установить *заранее.*

Литература

1. Bazaraa M., Shrali H. and Shetty C. *Nonlinear Programming Theory and Algorithms*, 2nd ed., Wiley, New York, 1993. (Имеется перевод первого издания: Базара М., Шетти К. *Нелинейное программирование. Теория и алгоритмы.* — М.: Мир, 1982.)
2. Beightler C., Phillips D. and Wilde D. *Foundations of Optimization*, 2nd ed., Prentice Hall, N.J., 1979.

Литература, добавленная при переводе

Банди Б. *Методы оптимизации. Вводный курс.* — М.: Радио и связь, 1988.

Зангвилл У.И. *Нелинейное программирование.* — М.: Сов. радио, 1973.

Сухарев А.Г., Тимохов А.В., Федоров В.В. *Курс методов оптимизации.* — М.: Наука, 1986.

Поляк Б.Т. *Введение в оптимизацию.* — М.: Наука, 1983.

Химмельблау Д. *Прикладное нелинейное программирование.* — М.: Мир, 1975.

Алгоритмы нелинейного программирования

21.1. Алгоритмы решения задач без ограничений

В этом разделе представлены два метода решения задач нелинейного программирования без ограничений на переменные: метод *прямого поиска* и *градиентный* метод. В первом реализуется прямой поиск оптимальной точки в заданной области, тогда как во втором методе для нахождения экстремальной точки используется градиент целевой функции.

21.1.1. Методы прямого поиска

Методы прямого поиска применяются главным образом для нахождения экстремумов функций одной переменной. Несмотря на то что такие задачи могут показаться тривиальными, в разделе 21.1.2 показано, что методы прямого поиска находят применение и при решении многомерных задач.

Идея методов прямого поиска проста. Сначала определяется **интервал неопределенности,** который содержит искомую точку оптимума. Затем длина интервала последовательно уменьшается до тех пор, пока не будет найдена точка оптимума. Процедура строится таким образом, что длину интервала, содержащего оптимальное решение, можно сделать сколь угодно малой.

В данном разделе рассматривается метод **дихотомического поиска**. Здесь на интервале $a \le x \le b$ ищется максимум функции $f(x)$, имеющей на этом интервале только один максимум. (Такая функция называется *одновершинной*.) Определим две точки x_1 и x_2, симметрично расположенные относительно точек a и b, таким образом, чтобы интервалы $a \le x \le x_2$ и $x_1 \le x \le b$ пересекались на некотором конечном отрезке Δ (рис. 21.1).

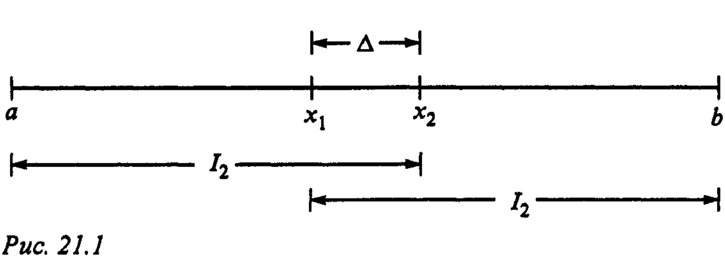


Рис. 21.1

Вычисляем значения $f(x_1)$ и $f(x_2)$. При этом возможны три случая.

1. Если $f(x_1) > f(x_2)$, то точка экстремума x^* должна лежать между a и x_2.
2. Если $f(x_1) < f(x_2)$, то $x_1 < x^* < b$.
3. Если $f(x_1) = f(x_2)$, то $x_1 < x^* < x_2$.

Эти утверждения непосредственно следуют из строгой одновершинности функции $f(x)$. В каждом из этих случаев из дальнейшего рассмотрения исключается интервал, не содержащий точку x^*.

В результате одного шага поиска сужается интервал, в котором находится точка максимума функции $f(x)$. Далее новый интервал делится на два (пересекающихся) интервала точно так же, как это было сделано с интервалом $[a, b]$. Продолжая процесс деления интервалов указанным способом, можно уменьшить длину интервала, содержащего точку локального максимума, до любой приемлемо малой величины Δ.

Пример 21.1–1

Имеем следующую задачу.

$$\text{Максимизировать} \quad f(x) = \begin{cases} 3x, & 0 \le x \le 2, \\ -\dfrac{x}{3} + \dfrac{20}{3}, & 2 \le x \le 3. \end{cases}$$

Ясно, что $\max f(x)$ достигается в точке $x = 2$. Пусть точки x_L и x_R определяют левую и правую границы *текущего* интервала. На первом шаге $x_L = 0$ и $x_R = 3$. Пусть x_1 и x_2 — точки, которые делят интервал таким образом, что соответствующие пересекающиеся отрезки описываются неравенствами $x_L \le x \le x_2$ и $x_1 \le x \le x_R$, где $x_1 - x_L = x_R - x_2$ и $\Delta = x_2 - x_1$. Это значит, что

$$x_1 = x_L + \frac{x_R - x_L - \Delta}{2},$$

$$x_2 = x_L + \frac{x_R - x_L + \Delta}{2}.$$

В табл. 21.1 содержатся результаты вычислений при $\Delta = 0.001$. Как следует из этой таблицы, на последнем шаге $x_L = 1.99737$ и $x_R = 2.00423$. Это значит, что максимум функции $f(x)$ достигается в точке x^*, удовлетворяющей неравенству $1.99737 \le x^* \le 2.00423$. Если в качестве приближенного решения задачи взять среднюю точку полученного интервала, то получим значение $x = 2.0008$, которое близко к точному оптимуму $x^* = 2$.

Таблица 21.1. Результаты вычислений методом дихотомического поиска[a]

x_L	x_R	x_1	x_2	$f(x_1)$	$f(x_2)$
0	3	1.4995^L	1.5005	4.4985	4.5015^b
1.4995	3	2.24925	2.25025^R	5.91692^b	5.91658
1.4995	2.25025	1.87437^L	1.87537	5.62312	5.62612^b
1.87437	2.25025	2.06181	2.06281^R	5.97939^b	5.97906
1.87437	2.06281	1.96809^L	1.96909	5.90427	5.90727^b
1.96809	2.06281	2.01495	2.01595^R	5.99502^b	5.99447

x_L	x_R	x_1	x_2	$f(x_1)$	$f(x_2)$
1.96809	2.01595	1.99152^L	1.99252	5.97456	5.97756^b
1.99152	2.01595	2.00323	2.00423^R	5.99892^b	5.99859
1.99152	2.00423	1.99737^L	1.99837	5.99213	5.99511^b
1.99737	2.00423				

[a] Верхние символы L и R показывают, что на следующем шаге x_L (соответственно x_R) равно x_1 (x_2).

[b] $\max\{f(x_1), f(x_2)\}$.

Упражнения 21.1,а

1. Решите задачу из примера 21.1–1, положив $\Delta = 0.01$. Сравните точность полученных результатов с данными из табл. 21.1.
2. Методом дихотомического поиска найдите максимумы следующих функций, полагая при этом, что $\Delta = 0.05$.

   a) $f(x) = \dfrac{1}{\left|(x-3)^3\right|}, \quad 2 \le x \le 4;$

   b) $f(x) = x\cos x, \ 0 \le x \le \pi;$

   c) $f(x) = x\sin \pi x, \ 1.5 \le x \le 2.5;$

   d) $f(x) = -(x-3)^2, \ 2 \le x \le 4;$

   e) $f(x) = \begin{cases} 4x, & 0 \le x \le 2, \\ 4-x, & 2 \le x \le 4. \end{cases}$

3. Получите формулу для определения максимального числа итераций, выполняемых в методе дихотомического поиска при заданной величине Δ и длине начального интервала неопределенности $I_0 = b - a$.

21.1.2. Градиентный метод

В этом разделе рассматривается метод решения задач нелинейного программирования с дважды непрерывно дифференцируемыми функциями. Основная идея метода заключается в построении последовательности точек с учетом направления градиента исследуемой функции.

Одним из градиентных методов является изложенный в разделе 20.2.2 метод Ньютона–Рафсона, который ориентирован на решение систем уравнений. Здесь рассматривается еще один метод, который называется методом *наискорейшего подъема*[1].

Согласно градиентному методу, вычисления завершаются при нахождении точки, в которой градиент функции равен нулю. Но это лишь необходимое условие для того, чтобы в данной точке находился оптимум. Чтобы удостовериться, что найдена именно точка оптимума, обычно привлекаются свойства вогнутости или выпуклости функции $f(x)$.

---

[1] В русской математической литературе описываемый метод называется *методом наискорейшего спуска* и ориентирован на решение задач безусловной минимизации. — *Прим. перев.*

Рассмотрим задачу максимизации функции $f(\mathbf{X})$. Пусть $\mathbf{X}^0$ — начальная точка, с которой начинается реализация метода, $\nabla f(\mathbf{X}^k)$ — градиент функции f в k-й точке $\mathbf{X}^k$. Идея метода сводится к определению в данной точке направления p, вдоль которого производная по направлению df/dp достигает своего максимума. Это достигается в том случае, когда последовательные точки $\mathbf{X}^k$ и $\mathbf{X}^{k+1}$ связаны соотношением

$$\mathbf{X}^{k+1} = \mathbf{X}^k + r^k\nabla f(\mathbf{X}^k),$$

где r^k — параметр, называемый **длиной шага.**

Обычно значение параметра r^k выбирается из условия, чтобы в точке $\mathbf{X}^{k+1}$ наблюдалось максимальное увеличение значения целевой функции f. Другими словами, если функция $h(r)$ определяется соотношением

$$h(r) = f(\mathbf{X}^k + r\nabla f(\mathbf{X}^k)),$$

то r^k равен значению r, на котором функция $h(r)$ достигает максимума. Поскольку $h(r)$ является функцией одной переменной, для нахождения ее максимума можно применить метод поиска экстремума, описанный в разделе 21.1.1, если, конечно, функция $h(r)$ строго одновершинная.

Описанная процедура заканчивается, когда две последовательные точки $\mathbf{X}^k$ и $\mathbf{X}^{k+1}$ близки. Это эквивалентно тому, что $r^k\nabla f(\mathbf{X}^k) \approx 0$. Поскольку $r^k \neq 0$, последнее соотношение означает, что в точке $\mathbf{X}^k$ выполняется необходимое условие экстремума $\nabla f(\mathbf{X}^k) = 0$.

Пример 21.1–2

Рассмотрим задачу максимизации функции

$$f(x_1, x_2) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2.$$

Функция $f(x_1, x_2)$ является квадратичной и нам известно, что ее абсолютный максимум достигается в точке $(x_1^*, x_2^*) = \left(\frac{1}{3}, \frac{4}{3}\right)$. Решим эту задачу методом наискорейшего подъема. На рис. 21.2 изображена последовательность получаемых точек. Градиенты в двух последовательных итерационных точках с необходимостью являются ортогональными (перпендикулярными) друг к другу. Для данной функции градиент вычисляется по формуле

$$\nabla f(\mathbf{X}) = (4 - 4x_1 - 2x_2,\ 6 - 2x_1 - 4x_2).$$

Пусть начальной точкой является $\mathbf{X}^0 = (1, 1)$.

Первая итерация

$$\nabla f(\mathbf{X}^0) = (-2, 0).$$

Следующая точка $\mathbf{X}^1$ находится на основании формулы

$$\mathbf{X} = (1, 1) + r(-2, 0) = (1 - 2r, 1).$$

Отсюда получаем выражение для функции $h(r)$:

$$h(r) = f(1 - 2r, 1) = -2(1 - 2r)^2 + 2(1 - 2r) + 4.$$

Оптимальная длина шага r, при которой функция $h(r)$ достигает максимума, равна 1/4. Таким образом, мы нашли, что $\mathbf{X}^1 = (1/2, 1)$.

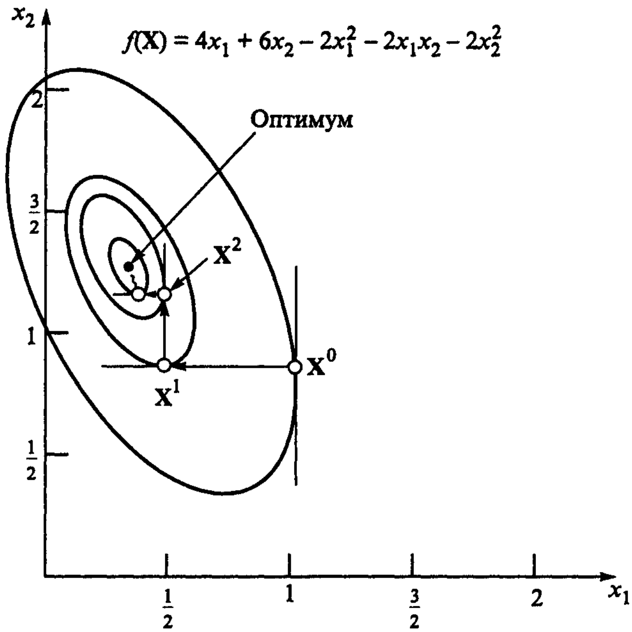


Рис. 21.2

Вторая итерация

$$\nabla f(\mathbf{X}^1) = (0,\ 1).$$

Теперь следующая точка $\mathbf{X}^2$ определяется выражением

$$\mathbf{X} = (\frac{1}{2},\ 1) + r(0,\ 1) = (\frac{1}{2},\ 1 + r).$$

Следовательно,

$$h(r) = -2(1 + r)^2 + 5(1 + r) + \frac{3}{2}.$$

Отсюда получаем $r = 1/4$ и $\mathbf{X}^2 = (1/2,\ 5/4)$.

Третья итерация

$$\nabla f(\mathbf{X}^2) = (-\frac{1}{2},\ 0).$$

Точка $\mathbf{X}^3$ определяется выражением

$$\mathbf{X} = \left(\frac{1}{2}, \frac{5}{4}\right) + r\left(-\frac{1}{2}, 0\right) = \left(\frac{1-r}{2}, \frac{5}{4}\right).$$

Следовательно,

$$h(r) = -\frac{1}{2}(1 - r)^2 + \frac{3}{4}(1 - r) + \frac{35}{8}.$$

Отсюда получаем $r = 1/4$ и $\mathbf{X}^3 = (3/8,\ 5/4)$.

Четвертая итерация

$$\nabla f(\mathbf{X}^3) = (0, \frac{1}{4}).$$

Точка $\mathbf{X}^3$ определяется выражением

$$\mathbf{X} = \left(\frac{3}{8}, \frac{5}{4}\right) + r\left(0, \frac{1}{4}\right) = \left(\frac{3}{8}, \frac{5+r}{4}\right).$$

Поэтому

$$h(r) = -\left(\frac{1}{8}\right)(5+r)^2 + \left(\frac{21}{16}\right)(5+r) + \frac{39}{32}.$$

Отсюда имеем $r = 1/4$ и $\mathbf{X}^4 = (3/8, 21/16)$.

Пятая итерация

$$\nabla f(\mathbf{X}^4) = (-\frac{1}{8}, 0).$$

Точка $\mathbf{X}^5$ определяется выражением

$$\mathbf{X} = \left(\frac{3}{8}, \frac{21}{16}\right) + r\left(-\frac{1}{8}, 0\right) = \left(\frac{3-r}{8}, \frac{21}{16}\right).$$

Следовательно,

$$h(r) = -\left(\frac{1}{32}\right)(3-r)^2 + \left(\frac{11}{64}\right)(3-r) + \frac{567}{128}.$$

Отсюда получаем $r = 1/4$ и $\mathbf{X}^5 = (11/32, 21/16)$.

Шестая итерация

$$\nabla f(\mathbf{X}^5) = (0, \frac{1}{16}).$$

Так как $\nabla f(\mathbf{X}^5) \approx \mathbf{0}$, вычисления можно закончить. Получена *приближенная* точка максимума $\mathbf{X}^5 = (0.3437, 1.3125)$. Точный максимум достигается в точке $\mathbf{X}^* = (0.3333, 1.3333)$.

Упражнения 21.1,b

1. Покажите, что метод Ньютона–Рафсона (раздел 20.2.2), применяемый к решению задачи максимизации строго вогнутой квадратичной функции, в общем случае сходится в точности за один шаг. Примените указанный метод к задаче максимизации функции

$$f(\mathbf{X}) = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2.$$

2. Выполните не более пяти итераций метода наискорейшего спуска (подъема) для каждой из следующих задач. Во всех случаях положите $\mathbf{X}^0 = \mathbf{0}$.

a) $\min f(\mathbf{X}) = (x_2 - x_1^2)^2 + (1 - x_1)^2$.

b) $\max f(\mathbf{X}) = \mathbf{cX} + \mathbf{X}^T\mathbf{AX}$, где

$$\mathbf{c} = (1, 3, 5),$$

$$\mathbf{A}=\begin{bmatrix} -5 & -3 & -\frac{1}{2} \\ -3 & -2 & 0 \\ -\frac{1}{2} & 0 & -\frac{1}{2} \end{bmatrix}.$$

c) $\max f(x) = 3 - x^2 - x^4$.

d) $\min f(\mathbf{X}) = x_1 - x_2 + x_1^2 - x_1 x_2$.

21.2. Алгоритмы решения задач с ограничениями

Общая задача нелинейного программирования с ограничениями записывается в виде:

максимизировать (или минимизировать) $z = f(\mathbf{X})$

при ограничениях

$$\mathbf{g}(\mathbf{X}) \le \mathbf{0}.$$

Условия неотрицательности переменных $\mathbf{X} \ge \mathbf{0}$ составляют часть заданных ограничений общего вида. Предполагается также, что по крайней мере одна из функций $f(\mathbf{X})$ или $\mathbf{g}(\mathbf{X})$ является нелинейной и все функции непрерывно дифференцируемы.

Универсальных алгоритмов решения задач нелинейного программирования не существует, и связано это, главным образом, с разнообразием нелинейных функций. Возможно, наиболее общим результатом, имеющим отношение к рассматриваемым задачам, являются условия Куна–Таккера. Как показано в разделе 20.3.2, эти условия являются лишь необходимыми для существования экстремума, за исключением, когда функции $f(\mathbf{X})$ и $\mathbf{g}(\mathbf{X})$ являются выпуклыми или вогнутыми.

В этом разделе рассматривается ряд алгоритмов решения задач нелинейного программирования, которые условно можно разделить на *непрямые* и *прямые*. В непрямых методах решение задачи нелинейного программирования сводится к решению одной или нескольких *линейных* задач, порожденных исходной. Прямые методы непосредственно имеют дело с исходной нелинейной задачей и позволяют построить последовательность точек, сходящуюся к точке экстремума.

В данном разделе непрямые методы представлены алгоритмами сепарабельного, квадратичного, геометрического и стохастического программирования. К числу прямых методов относится метод линейных комбинаций и метод последовательной безусловной максимизации, изложенный далее вкратце. С другими важными методами решения задач нелинейного программирования можно познакомиться, обратившись к приведенному в конце главы списку литературы.

21.2.1. Сепарабельное программирование

Функция $f(x_1, x_2, \ldots, x_n)$ называется **сепарабельной**[1] (разделимой), если она представляется в виде суммы n функций одной переменной $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, т.е.

---

[1] Дотошность математика требует сделать замечание, что понятие сепарабельности здесь не имеет ничего общего (кроме названия) с "классическим" его определением из функционального анализа. — *Прим. ред*

$$f(x_1, x_2, ..., x_n) = f_1(x_1) + f_2(x_2) + ... + f_n(x_n).$$

К примеру, линейная функция

$$h(x_1, x_2, ..., x_n) = a_1x_1 + a_2x_2 + ... + a_nx_n$$

(здесь a_i, $i = 1, 2, ..., n$ — константы) является сепарабельной. Функция же

$$h(x_1, x_2, x_3) = x_1^2 + x_1 \sin(x_2 + x_3) + x_2e^{x_3}$$

таковой не является.

Некоторые нелинейные функции сепарабельными непосредственно не являются, однако могут быть приведены к такому виду путем соответствующих подстановок. Рассмотрим, к примеру, задачу максимизации функции $z = x_1x_2$. Если ввести обозначение $y = x_1x_2$, то $\ln y = \ln x_1 + \ln x_2$ и задача принимает следующий вид.

Максимизировать $z = y$

при ограничении

$$\ln y = \ln x_1 + \ln x_2,$$

т.е. она является сепарабельной. При такой замене предполагается, что переменные x_1 и x_2 принимают *положительные* значения, иначе логарифмическая функция не определена.

В случае, когда переменные x_1 и x_2 принимают и нулевые значения (т.е. $x_1, x_2 \geq 0$), можно поступить следующим образом. Пусть δ_1 и δ_2 — положительные константы, введем новые переменные $w_1 = x_1 + \delta_1$ и $w_2 = x_2 + \delta_2$. Эти переменные принимают только положительные значения. Теперь имеем

$$x_1x_2 = w_1w_2 - \delta_2w_1 - \delta_1w_2 + \delta_1\delta_2.$$

Пусть $y = w_1w_2$, тогда исходная задача эквивалентна следующей.

Максимизировать $z = y - \delta_2w_1 - \delta_1w_2 + \delta_1\delta_2$

при ограничениях

$$\ln y = \ln w_1 + \ln w_2, \; w_1 \geq \delta_1, \; w_2 \geq \delta_2.$$

В этой задаче все функции сепарабельные.

Примерами других функций, которые в результате замены переменных приводятся к сепарабельным, могут служить $e^{x_1+x_2}$ и $x_1^{x_2}$. В этих случаях для реализации свойства сепарабельности следует применить соответствующий вариант описанной выше процедуры.

В сепарабельном программировании рассматриваются задачи нелинейного программирования, в которых как целевая функция, так и функции ограничений являются сепарабельными. В этом разделе рассматриваются методы приближенного решения задачи сепарабельного программирования, основанные на линейной аппроксимации функций и на симплекс-методе линейного программирования.

Функцию одной переменной $f(x)$ можно аппроксимировать кусочно-линейной функцией с помощью методов частично-целочисленного программирования (глава 9). Пусть требуется аппроксимировать функцию $f(x)$ на интервале $[a, b]$. Обозначим через a_k, $k = 1, 2, ..., K$, k-ю точку разбиения интервала $[a, b]$, причем $a = a_1 < a_2 < ... < a_K = b$. Тогда функция $f(x)$ аппроксимируется следующей кусочно-линейной функцией:

$$f(x) \approx \sum_{k=1}^{K} f(a_k)t_k, \quad x = \sum_{k=1}^{K} a_k t_k,$$

где t_k — неотрицательный весовой коэффициент, связанный с k-й точкой разбиения интервала. Весовые коэффициенты удовлетворяют условию

$$\sum_{k=1}^{K} t_k = 1.$$

Методы частично-целочисленного программирования гарантируют корректность такой аппроксимации. В частности, аппроксимация является обоснованной в следующих случаях.

1. Если не больше двух весовых коэффициентов t_k имеют положительные значения.
2. Если значение t_k больше нуля, то положительным может оказаться лишь один из смежных весовых коэффициентов t_{k+1} или t_{k-1}.

Чтобы показать, как выполняются перечисленные условия, рассмотрим общую задачу сепарабельного программирования.

$$\text{Максимизировать (минимизировать) } z = \sum_{i=1}^{n} f_i(x_i)$$

при ограничениях

$$\sum_{i=1}^{n} g_i^j(x_i) \le b_j, \quad j = 1, 2, \ldots, m.$$

Эту задачу можно привести к задаче частично-целочисленного программирования следующим образом. Обозначим через K_i число точек разбиения для i-й переменной x_i, а через a_i^k — k-ю точку разбиения. Пусть t_i^k — весовой коэффициент, ассоциируемый с k-й точкой разбиения для i-й переменной. Тогда эквивалентная задача частично-целочисленного программирования имеет вид

$$\text{максимизировать (или минимизировать) } z = \sum_{i=1}^{n}\sum_{k=1}^{K_i} f_i(a_i^k)t_i^k$$

при ограничениях

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} g_i^j(a_i^k)t_i^k \le b_j, \quad j = 1, 2, \ldots, m,$$

$$0 \le t_i^1 \le y_i^1,$$

$$0 \le t_i^k \le y_i^{k-1} + y_i^k, \quad k = 2, 3, \ldots, K_i - 1,$$

$$0 \le t_i^{K_i} \le y_i^{K_i - 1},$$

$$\sum_{k=1}^{K_i - 1} y_i^k = 1,$$

$$\sum_{k=1}^{K_i} t_i^k = 1,$$

$$y_i^k = 0 \text{ или } 1, \quad k = 1, 2, \ldots, K_i, \quad i = 1, 2, \ldots, n.$$

В аппроксимирующей задаче переменными являются t_i^k и y_i^k.

Данное представление свидетельствует о том, что в принципе любая задача сепарабельного программирования может быть решена методами частично-целочисленного программирования. Трудность такого подхода к решению задачи связана с быстрым ростом числа ограничений при увеличении количества точек разбиения. В частности, проблематичной является вычислительная реализация процедуры, поскольку нет эффективных компьютерных программ для решения задач частично-целочисленного программирования большой размерности.

Для решения аппроксимирующей задачи можно использовать также обычный симплекс-метод (глава 3), дополненный правилом **ограниченного ввода в базис**. В этом случае игнорируются вспомогательные ограничения, содержащие y_i^k. Согласно данному правилу, в базисе может находиться *не более двух положительных* весовых коэффициентов t_i^k. Более того, два коэффициента t_i^k могут быть положительными лишь тогда, когда они являются смежными. Таким образом, строгое условие оптимальности симплекс-метода *только тогда* используется для выбора переменной t_i^k, подлежащей введению в число базисных, когда она удовлетворяет указанным выше требованиям. В противном случае анализ разностей $\left(z_i^k - c_i^k\right)$ позволяет выбрать следующую переменную t_i^k, которая может быть сделана базисной. Процесс повторяется до тех пор, пока не будет выполнен критерий оптимальности или же установлена невозможность введения в базис новой переменной t_i^k без нарушения правила ограниченного ввода. В обоих случаях последняя симплексная таблица дает приближенное оптимальное решение исходной задачи.

Метод частично-целочисленного программирования позволяет найти глобальный экстремум аппроксимирующей задачи, тогда как симплекс-метод с учетом правила ограниченного ввода в базис может гарантировать нахождение лишь локального оптимума. Кроме того, приближенное решение, полученное любым из двух упомянутых методов, может быть недопустимым для исходной задачи, поскольку при решении аппроксимирующей задачи могут быть обнаружены дополнительные экстремальные точки, которые для исходной задачи экстремальными не являются. Это зависит от точности линейной аппроксимации исходной задачи.

Пример 21.2–1

Рассмотрим следующую задачу.

$$\text{Максимизировать } z = x_1 + x_2^4$$

при ограничениях

$$3x_1 + 2x_2^2 \le 9,$$

$$x_1, x_2 \ge 0.$$

Опишем применение симплекс-метода совместно с правилом ограниченного ввода в базис.

Точное оптимальное решение этой задачи находится непосредственной проверкой: $x_1 = 0$, $x_2 = 2.12$ и $z^* = 20.25$. Чтобы продемонстрировать использование метода аппроксимации, рассмотрим отдельные функции

$$f_1(x_1) = x_1,$$

$$f_2(x_2) = x_2^4,$$
$$g_1^1(x_1) = 3x_1,$$
$$g_1^2(x_2) = 2x_2^2.$$

Функции $f_1(x_1)$ и $g_1^1(x_1)$ остаются в исходном виде, поскольку они уже являются линейными. В этом случае x_1 рассматривается как одна из переменных. Для функций $f_2(x_2)$ и $g_1^2(x_2)$ полагаем, что количество точек разбиения равно четырем ($K_2 = 4$). Так как значение x_2 не может превышать 3, отсюда вытекают данные, приведенные в следующей таблице.

k	a_2^k	$f_2(a_2^k)$	$g_1^2(a_2^k)$
1	0	0	0
2	1	1	2
3	2	16	8
4	3	81	18

Отсюда получаем

$$f_2(x_2) \approx t_2^1 f_2(a_2^1) + t_2^2 f_2(a_2^2) + t_2^3 f_2(a_2^3) + t_2^4 f_2(a_2^4) =$$
$$= 0 \times t_2^1 + 1 \times t_2^2 + 16 \times t_2^3 + 81 \times t_2^4 = t_2^2 + 16t_2^3 + 81t_2^4.$$

Аналогично

$$g_1^2(x_2) \approx 2t_2^2 + 8t_2^3 + 18t_2^4.$$

Следовательно, аппроксимирующая задача принимает вид

$$\text{максимизировать } z = x_1 + t_2^2 + 16t_2^3 + 81t_2^4$$

при ограничениях

$$3x_1 + 2t_2^2 + 8t_2^3 + 18t_2^4 \le 9,$$
$$t_2^1 + t_2^2 + t_2^3 + t_2^4 = 1,$$
$$t_2^k \ge 0, \quad k = 1, 2, 3, 4,$$
$$x_1 \ge 0.$$

Исходная симплекс-таблица (в которой осуществлена перестановка столбцов для получения начального решения) выглядит следующим образом.

Базис	x_1	t_2^2	t_2^3	t_2^4	S_1	t_2^1	Решение
z	–1	–1	–16	–81	0	0	0
S_1	3	2	8	18	1	0	9
t_2^1	0	1	1	1	0	1	1

Здесь S_1 (≥ 0) — дополнительная (остаточная) переменная. (В этой задаче начальное решение является очевидным. В общем же случае для его получения могут понадобиться искусственные переменные, см. раздел 3.4.)

Коэффициенты z-строки указывают на то, что в число базисных необходимо ввести переменную t_2^4. Поскольку при этом переменная t_2^1 уже входит в базисные, в силу правила ограниченного ввода в базис, она должна быть выведена из числа базисных перед тем, как базисной станет переменная t_2^4. Вместе с тем, согласно условию допустимости, при введении в базисные переменной t_2^4 переменная S_1 должна выводиться из базиса. Это значит, что t_2^4 нельзя сделать базисной. Следующей подходящей для введения в базисные является переменная t_2^3. Снова при этом переменная t_2^1 должна быть выведена из числа базисных. Условие допустимости на этот раз позволяет вывести из числа базисных переменную t_2^1, что и требуется. Таким образом, получаем новую симплекс-таблицу.

Базис	x_1	t_2^2	t_2^3	t_2^4	S_1	t_2^1	Решение
z	–1	15	0	–65	0	16	16
S_1	3	–6	0	10	1	–8	1
t_2^3	0	1	1	1	0	1	1

Очевидно, что базисной следует сделать переменную t_2^4. То, что переменная t_2^3 уже является базисной, не противоречит правилу ограниченного ввода в базис. В соответствии с симплекс-методом переменная S_1 исключается из числа базисных. Получаем следующую симплекс-таблицу.

Базис	x_1	t_2^2	t_2^3	t_2^4	S_1	t_2^1	Решение
z	37/2	–24	0	0	13/2	–36	45/2
t_2^4	3/10	–6/10	0	1	1/10	–8/10	1/10
t_2^3	–3/10	16/10	1	0	–1/10	18/10	9/10

Из этой таблицы следует, что в базис могут вводиться переменные t_2^1 и t_2^2. Переменная t_2^1 не может быть базисной, поскольку не является смежной с переменными t_2^3 и t_2^4, которые уже находятся среди базисных. Далее, переменную t_2^2 также нельзя сделать базисной, поскольку t_2^4 нельзя исключить из числа базисных. Поэтому процедура решения задачи на этом завершается, и полученное решение является наилучшим допустимым решением аппроксимирующей задачи.

Воспользуемся найденными значениями $t_2^3 = 9/10$ и $t_2^4 = 1/10$, чтобы выразить полученное решение через переменные x_1 и x_2. Получаем

$$x_2 \approx 2t_2^3+3t_2^4=2\left(\frac{9}{10}\right)+3\left(\frac{1}{10}\right)=2.1$$

и $x_1 = 0$, $z = 22.5$. Полученное приближенное оптимальное значение для x_2 (= 2.1) мало отличается от истинного оптимального значения (= 2.12).

Сепарабельное выпуклое программирование. Задача выпуклого сепарабельного программирования имеет место в том случае, когда функции $g_i^j(x_i)$ являются выпуклыми, обеспечивая, таким образом, выпуклость области допустимых решений. Кроме того, считаем, что функции $f_i(x_i)$ являются выпуклыми в задаче минимизации и вогнутыми в задаче максимизации (см. табл. 20.2). При этих условиях можно использовать следующий упрощенный метод аппроксимации.

Рассмотрим задачу минимизации. Пусть функция $f_i(x_i)$ имеет такой вид, как показано на рис. 21.3. Обозначим точки разбиения для функции $f_i(x_i)$ через $x_i = a_{ki}$, $k = 0, 1, ..., K_i$. Пусть x_{ki} — величина изменения переменной x_i на интервале $(a_{k-1,i}, a_{ki})$, $k = 1, 2, ..., K_i$, а ρ_{ki} — тангенс угла наклона линейного сегмента на том же интервале.

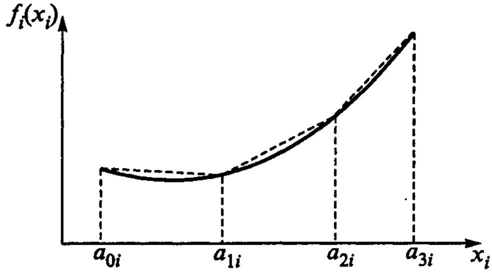


Рис. 21.3

Тогда

$$f_i(x_i) \approx \sum_{k=1}^{K_i} \rho_{ki} x_{ki} + f_i(a_{0i}),\ x_i = \sum_{k=1}^{K_i} x_{ki}\ .$$

Здесь предполагается, что $0 \le x_{ki} \le a_{ki} - a_{k-1,i}$, $k = 1, 2, ..., K_i$.

Так как функция $f_i(x_i)$ выпуклая, то $\rho_{1i} < \rho_{2i} < ... < \rho_{K_i i}$. Это означает, что в задаче минимизации при $p < q$ большее влияние на значение целевой функции оказывает переменная x_{pi}, чем x_{qi}. Следовательно, переменная x_{pi} всегда будет входить в решение до того, как в него войдет x_{qi} При этом значения каждой переменной x_{ki} должны быть ограничены сверху величиной $(a_{ki} - a_{k-1,i})$.

Выпуклые функции ограничений $g_i^j(x_i)$ аппроксимируются аналогичным образом. Пусть ρ_{ki}^j — тангенс угла наклона k-го линейного сегмента, соответствующего функции $g_i^j(x_i)$. Тогда функция $g_i^j(x_i)$ аппроксимируется формулой

$$g_i^j(x_i) \approx \sum_{k=1}^{K_i} \rho_{ki}^j x_{ki} + g_i^j(a_{0i}).$$

Следовательно, рассматриваемая задача принимает вид

$$\text{минимизировать } z = \sum_{i=1}^{n} \left(\sum_{k=1}^{K_i} \rho_{ki} x_{ki} + f_i(a_{0i}) \right),$$

при ограничениях

$$\sum_{i=1}^{n} \left(\sum_{k=1}^{K_i} \rho_{ki}^j x_{ki} + g_i^j(a_{0i}) \right) \le b_j,\ j = 1, 2, ..., m,$$

$$0 \le x_{ki} \le a_{ki} - a_{k-1,i},\ k = 1, 2, ..., K_i,\ i = 1, 2, ..., n,$$

где

$$\rho_{ki} = \frac{f_i(a_{ki}) - f_i(a_{k-1,i})}{a_{ki} - a_{k-1,i}},$$

$$\rho_{ki}^j = \frac{g_i^j(a_{ki}) - g_i^j(a_{k-1,i})}{a_{ki} - a_{k-1,i}}.$$

Задача максимизации преобразуется аналогично. В этом случае $\rho_{1i} > \rho_{2i} > ... > \rho_{K_i i}$, откуда следует, что при $p < q$ переменная x_{pi} всегда будет входить в решение раньше x_{qi} (см. упр. 21.2,а(7)).

Полученную задачу можно решить симплекс-методом, предназначенным для решения задач с ограниченными сверху переменными (раздел 7.5.2). При этом исчезает необходимость в соблюдении правила ограниченного ввода в базис, поскольку выпуклость (вогнутость) функций гарантирует надлежащий выбор переменных.

Пример 21.2–2

Рассмотрим задачу

$$\text{минимизировать } z = x_1^2 + x_2^2 + 5$$

при ограничениях

$$3x_1^4 + x_2 \le 243,$$
$$x_1 + 2x_2^2 \le 32,$$
$$x_1, x_2 \ge 0.$$

Отдельными (сепарабельными) функциями этой задачи являются

$$f_1(x_1) = x_1^2,\quad f_2(x_2) = x_2^2 + 5,$$
$$g_1^1(x_1) = 3x_1^4,\quad g_2^1(x_2) = x_2,$$
$$g_1^2(x_1) = x_1,\quad g_2^2(x_2) = 2x_2^2.$$

Эти функции выпуклые, что и требуется в задачах минимизации.

Область допустимых значений переменных x_1 и x_2, как следует из ограничений задачи, определяется неравенствами $0 \le x_1 \le 3$ и $0 \le x_2 \le 4$. Разбиваем эти интервалы изменения переменных x_1 и x_2. Пусть $K_1 = 3$ и $K_2 = 4$ при $a_{01} = a_{02} = 0$. Значения тангенсов углов наклона, соответствующих отдельным функциям рассматриваемой задачи, приведены в следующих таблицах.

Для $i = 1$ имеем

k	a_{k1}	ρ_{k1}	ρ_{k1}^1	ρ_{k1}^2	x_{k1}
0	0	—	—	—	—
1	1	1	3	1	x_{11}
2	2	3	45	1	x_{21}
3	3	5	195	1	x_{31}

Для $i = 2$ имеем

k	a_{k2}	ρ_{k2}	ρ_{k2}^1	ρ_{k2}^2	x_{k2}
0	0	—	—	—	—
1	1	1	1	2	x_{12}
2	2	3	1	6	x_{22}
3	3	5	1	10	x_{32}
4	4	7	1	14	x_{42}

Теперь задача принимает следующий вид.

$$\text{Минимизировать } z \approx x_{11} + 3x_{21} + 5x_{31} + x_{12} + 3x_{22} + 5x_{32} + 7x_{42} + 5$$

при ограничениях

$$3x_{11} + 45x_{21} + 195x_{31} + x_{12} + x_{22} + x_{32} + x_{42} \le 243,$$
$$x_{11} + x_{21} + x_{31} + 2x_{12} + 6x_{22} + 10x_{32} + 14x_{42} \le 32,$$
$$0 \le x_{k1} \le 1,\ k = 1, 2, 3,$$
$$0 \le x_{k2} \le 1,\ k = 1, 2, 3, 4.$$

Пусть x_{k1}^* и x_{k2}^* — оптимальные значения соответствующих переменных, полученные с помощью описанного выше варианта симплекс-метода. Тогда оптимальные значения переменных x_1 и x_2 вычисляются по формулам

$$x_1^* = \sum_{k=1}^{3} x_{k1}^*, \quad x_2^* = \sum_{k=1}^{4} x_{k2}^*.$$

Упражнения 21.2,а

1. Для следующей задачи постройте аппроксимирующую модель в виде задачи частично-целочисленного программирования.

$$\text{Максимизировать } z = e^{-x_1} + x_1 + (x_2 + 1)^2$$

при ограничениях

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

2. Решите аппроксимирующую задачу из предыдущего упражнения, используя правило ограниченного ввода в базис. Затем найдите оптимальное решение исходной задачи.
3. Рассмотрите следующую задачу.

$$\text{Максимизировать } z = x_1 x_2 x_3$$

при ограничениях

$$x_1^2 + x_2 + x_3 \le 4,$$
$$x_1, x_2, x_3 \ge 0.$$

Постройте аппроксимирующую задачу в виде задачи линейного программирования с учетом дальнейшего использования правила ограниченного ввода в базис.

4. Покажите, каким образом приведенная ниже задача может быть приведена к сепарабельному виду.

$$\text{Максимизировать } z = x_1x_2 + x_3 + x_1x_3$$

при ограничениях

$$x_1x_2 + x_2 + x_1x_3 \le 10,$$
$$x_1, x_2, x_3 \ge 0.$$

5. Покажите, каким образом следующую задачу можно преобразовать в задачу сепарабельного программирования.

$$\text{Минимизировать } z = e^{2x_1+x_2^2} + (x_3 - 2)^2$$

при ограничениях

$$x_1 + x_2 + x_3 \le 6,$$
$$x_1, x_2, x_3 \ge 0.$$

6. Покажите, как следующую задачу можно преобразовать в задачу сепарабельного программирования.

$$\text{Максимизировать } z = e^{x_1x_2} + x_2^2x_3 + x_4$$

при ограничениях

$$x_1 + x_2x_3 + x_3 \le 10,$$
$$x_1, x_2, x_3 \ge 0,$$
$$x_4 \text{ не ограничена в знаке.}$$

7. Покажите, что при реализации метода выпуклого сепарабельного программирования (раздел 21.2.1) оптимальное решение не содержит переменной $x_{ki} > 0$, если переменная $x_{k-1,i}$ не достигает своей верхней границы.

8. Решите представленную ниже задачу как задачу выпуклого сепарабельного программирования.

$$\text{Минимизировать } z = x_1^4 + 2x_2 + x_3^2$$

при ограничениях

$$x_1^2 + x_2 + x_3^2 \le 4,$$
$$|x_1 + x_2| \le 0,$$
$$x_1, x_3 \ge 0,$$
$$x_2 \text{ не ограничена в знаке.}$$

9. Решите как задачу выпуклого сепарабельного программирования следующую задачу.

$$\text{Минимизировать } z = (x_1 - 2)^2 + 4(x_2 - 6)^2$$

при ограничениях

$$6x_1 + 3(x_2 + 1)^2 \le 12,$$
$$x_1, x_2 \ge 0.$$

21.2.2. Квадратичное программирование

Задача квадратичного программирования имеет следующий вид.

$$\text{Максимизировать (или минимизировать) } z = \mathbf{CX} + \mathbf{X}^T\mathbf{DX}$$

при ограничениях

$$\mathbf{AX} \le \mathbf{b},\ \mathbf{X} \ge \mathbf{0},$$

где

$$\mathbf{X} = (x_1, x_2, \ldots, x_n)^T,$$
$$\mathbf{C} = (c_1, c_2, \ldots, c_n),$$
$$\mathbf{b} = (b_1, b_2, \ldots, b_m)^T,$$
$$\mathbf{A} = \begin{bmatrix} a_{11} \cdots a_{1n} \\ \vdots \qquad \vdots \\ a_{m1} \cdots a_{mn} \end{bmatrix},$$
$$\mathbf{D} = \begin{bmatrix} d_{11} \cdots d_{1n} \\ \vdots \qquad \vdots \\ d_{n1} \cdots d_{nn} \end{bmatrix}.$$

Функция $\mathbf{X}^T\mathbf{DX}$, где $\mathbf{D}$ — симметрическая матрица, является квадратичной формой (см. раздел А.3). Матрица $\mathbf{D}$ будет отрицательно определенной в задаче максимизации и положительно определенной — в задаче минимизации. Это означает, что функция z является строго выпуклой по переменным $\mathbf{X}$ в случае задачи минимизации и строго вогнутой — в задаче максимизации. Ограничения в этой задаче предполагаются линейными, что гарантирует выпуклость области допустимых решений.

Решение сформулированной задачи основано на использовании необходимых условий Куна–Таккера (раздел 20.3.2). Так как целевая функция z строго выпуклая (или вогнутая) и область допустимых решений задачи является выпуклым множеством, эти условия (как показано в разделе 20.3.2) также достаточны для установления наличия глобального оптимума.

Задачу квадратичного программирования рассмотрим для случая, когда целевая функция подлежит максимизации. Изменение формулировки задачи при минимизации целевой функции является тривиальным. Общая постановка задачи имеет следующий вид.

$$\text{Максимизировать } z = \mathbf{CX} + \mathbf{X}^T\mathbf{DX}$$

при ограничениях

$$\mathbf{G}(\mathbf{X}) = \begin{bmatrix} \mathbf{A} \\ -\mathbf{I} \end{bmatrix}\mathbf{X} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \le \mathbf{0}.$$

Обозначим через $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)^T$ и $\mathbf{U} = (\mu_1, \mu_2, \ldots, \mu_n)^T$ множители Лагранжа, соответствующие ограничениям $\mathbf{AX} - \mathbf{b} \le \mathbf{0}$ и $-\mathbf{X} \le \mathbf{0}$ соответственно. Применяя условия Куна–Таккера, получаем

$$\boldsymbol{\lambda} \ge \mathbf{0},\ \mathbf{U} \ge \mathbf{0},$$
$$\nabla z - (\boldsymbol{\lambda}^T, \mathbf{U}^T)\nabla\mathbf{G}(\mathbf{X}) = \mathbf{0},$$

$$\lambda_i\left(b_i-\sum_{j=1}^{n}a_{ij}x_j\right)=0,\ \ i=1,2,\ldots,m,$$

$$\mu_j x_j = 0,\ j = 1, 2, \ldots, n,$$

$$\mathbf{AX} \le \mathbf{b},\ -\mathbf{X} \le \mathbf{0}.$$

Отсюда имеем

$$\nabla z = \mathbf{C} + 2\mathbf{X}^T\mathbf{D},$$

$$\nabla \mathbf{G}(\mathbf{X}) = \begin{bmatrix} \mathbf{A} \\ -\mathbf{I} \end{bmatrix}.$$

Обозначим через $\mathbf{S} = \mathbf{b} - \mathbf{AX} \ge \mathbf{0}$ вектор дополнительных (остаточных) переменных. Тогда приведенные выше условия принимают следующий вид.

$$-2\mathbf{X}^T\mathbf{D} + \boldsymbol{\lambda}^T\mathbf{A} - \mathbf{U}^T = \mathbf{C},$$

$$\mathbf{AX} + \mathbf{S} = \mathbf{b},$$

$$\mu_j x_j = 0 = \lambda_i S_i \text{ для всех } i \text{ и } j,$$

$$\boldsymbol{\lambda}, \mathbf{U}, \mathbf{X}, \mathbf{S} \ge \mathbf{0}.$$

Так как $\mathbf{D}^T = \mathbf{D}$, в результате транспонирования первой системы уравнений получим

$$-2\mathbf{DX} + \mathbf{A}^T\boldsymbol{\lambda} - \mathbf{U} = \mathbf{C}^T.$$

Следовательно, необходимые условия могут быть записаны в виде

$$\begin{bmatrix} -2\mathbf{D} & \mathbf{A}^T & -\mathbf{I} & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \boldsymbol{\lambda} \\ \mathbf{U} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} \mathbf{C}^T \\ \mathbf{b} \end{bmatrix},$$

$$\mu_j x_j = 0 = \lambda_i S_i \text{ для всех } i \text{ и } j,$$

$$\boldsymbol{\lambda}, \mathbf{U}, \mathbf{X}, \mathbf{S} \ge \mathbf{0}.$$

Все уравнения, за исключением $\mu_j x_j = 0 = \lambda_i S_i$, являются линейными относительно переменных $\mathbf{X}$, $\boldsymbol{\lambda}$, $\mathbf{U}$ и $\mathbf{S}$. Следовательно, исходная задача сводится к нахождению решения системы линейных уравнений, удовлетворяющего дополнительным условиям $\mu_j x_j = 0 = \lambda_i S_i$. Поскольку функция z строго вогнутая, а область допустимых решений представляет собой выпуклое множество, *допустимое* решение, удовлетворяющее всем этим условиям, должно быть единственным и оптимальным.

Решение рассматриваемой системы находится путем реализации этапа I двухэтапного метода (раздел 3.4.2). При этом единственным ограничением является необходимость удовлетворения условий $\lambda_i S_i = 0 = \mu_j x_j$. Выполнение этого условия означает, что если переменная λ_i в базисном решении принимает *положительное значение*, то переменная S_i не может быть базисной и принимать положительное значение. Аналогично переменные μ_j и x_j не могут одновременно принимать положительные значения. Это обстоятельство подобно правилу *ограниченного ввода в базис*, которое было использовано в разделе 21.2.1. Этап I завершается обращением в нуль всех искусственных переменных только в том случае, когда исходная задача имеет непустое множество допустимых решений.

Пример 21.2–3

Рассмотрим следующую задачу.

$$\text{Максимизировать } z = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2$$

при ограничениях

$$x_1 + 2x_2 \le 2,$$
$$x_1, x_2 \ge 0.$$

Эту задачу можно записать в матричной форме

$$\text{максимизировать } z = (4,6)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + (x_1, x_2)\begin{bmatrix} -2 & -1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

при ограничениях

$$(1,2)\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \le 2,$$
$$x_1, x_2 \ge 0.$$

Условия Куна–Таккера принимают следующий вид.

$$\begin{bmatrix} 4 & 2 & 1 & -1 & 0 & 0 \\ 2 & 4 & 2 & 0 & -1 & 0 \\ 1 & 2 & 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \lambda_1 \\ \mu_1 \\ \mu_2 \\ S_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix}.$$

Начальная симплекс-таблица для этапа I строится путем введения искусственных переменных R_1 и R_2. Имеем

Базис	x_1	x_2	λ_1	μ_1	μ_2	R_1	R_2	S_1	Решение
r	6	6	3	–1	–1	0	0	0	10
R_1	4	2	1	–1	0	1	0	0	4
R_2	2	4	2	0	–1	0	1	0	6
S_1	1	2	0	0	0	0	0	1	2

Первая итерация

Поскольку $\mu_1 = 0$, вводимой в базис переменной может быть x_1, при этом из числа базисных исключается переменная R_1. Получаем следующую симплекс-таблицу.

Базис	x_1	x_2	λ_1	μ_1	μ_2	R_1	R_2	S_1	Решение
r	0	3	3/2	1/2	–1	–3/2	0	0	4
x_1	1	1/2	1/4	–1/4	0	1/4	0	0	1
R_2	0	3	3/2	1/2	–1	–1/2	1	0	4
S_1	0	3/2	–1/4	1/4	0	–1/4	0	1	1

Вторая итерация

Так как $\mu_2 = 0$, вводимой в базис переменной является x_2. В результате приходим к следующей таблице.

Базис	x_1	x_2	λ_1	μ_1	μ_2	R_1	R_2	S_1	Решение
r	0	0	2	0	–1	–1	0	–2	2
x_1	1	0	1/3	–1/3	0	1/3	0	–1/3	2/3
R_1	0	0	2	0	–1	0	1	–2	2
x_2	0	1	–1/6	1/6	0	–1/6	0	2/3	2/3

Третья итерация

Так как $S_1 = 0$, в число базисных можно ввести переменную λ_1. В результате получаем симплекс-таблицу.

Базис	x_1	x_2	λ_1	μ_1	μ_2	R_1	R_2	S_1	Решение
r	0	0	0	0	0	–1	–1	0	0
x_1	1	0	0	–1/3	1/6	1/3	–1/6	0	1/3
λ_1	0	0	1	0	–1/2	0	1/2	–1	1
x_2	0	1	0	1/6	–1/12	–1/6	1/12	1/2	5/6

Последняя таблица дает оптимальное решение рассматриваемой на этапе I задачи. Так как $r = 0$, полученное решение $x_1 = 1/3$, $x_2 = 5/6$ является допустимым. Оптимальное значение z вычисляется подстановкой полученного решения в выражение для целевой функции исходной задачи и равняется 4.16.

Упражнения 21.2,b

1. Дана задача, в которой требуется

$$\text{максимизировать } z = 6x_1 + 3x_2 - 4x_1x_2 - 2x_1^2 - 3x_2^2$$

при ограничениях

$$x_1 + x_2 \le 1,$$
$$2x_1 + 3x_2 \le 4,$$
$$x_1, x_2 \ge 0.$$

Покажите, что z — строго вогнутая функция, и решите задачу, используя алгоритм квадратичного программирования.

2. Дана следующая задача.

$$\text{Минимизировать } z = 2x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 2x_2x_3 + x_1 - 3x_2 - 5x_3$$

при ограничениях

$$x_1 + x_2 + x_3 \ge 1,$$

$$3x_1 + 2x_2 + x_3 \le 6,$$
$$x_1, x_2, x_3 \ge 0.$$

Покажите, что функция z — строго выпуклая, и найдите решение задачи методом квадратичного программирования.

21.2.3. Геометрическое программирование

В *геометрическом программировании* рассматриваются задачи нелинейного программирования специального вида. Подход геометрического программирования позволяет находить решение исходной задачи с помощью соответствующей двойственной задачи (которая будет сформулирована ниже).

Методами геометрического программирования решаются нелинейные задачи, в которых как целевая функция, так и функции ограничений имеют следующий вид:

$$z = f(\mathbf{X}) = \sum_{j=1}^{N} U_j,$$

где

$$U_j = c_j \prod_{i=1}^{n} x_i^{a_{ij}}, \quad j = 1, 2, \ldots, N.$$

Предполагается, что здесь все $c_j > 0$, а число N является конечным. Показатели степени a_{ij} по знаку не ограничены. Функция $f(\mathbf{X})$ имеет вид полинома, если не учитывать того, что показатели степени a_{ij} могут принимать отрицательные значения. По этой причине, а также в связи с тем, что все коэффициенты $c_j > 0$, функция $f(\mathbf{X})$ называется **позиномом.**

В этом разделе будут рассмотрены задачи геометрического программирования без ограничений. Исследование задач геометрического программирования с ограничениями выходит за рамки настоящей главы. Детальное рассмотрение излагаемых здесь вопросов содержится в книге [2], глава 6.

Рассмотрим задачу *минимизации* позиномиальной функции $f(\mathbf{X})$. Будем называть эту задачу *прямой*. Предполагается, что переменные x_i принимают *строго положительные* значения, т.е. область $x_i \le 0$ недопустима. Далее будет показано, что требование $x_i \ne 0$ является существенным при получении основных результатов.

Первые частные производные функции z в точке минимума должны обращаться в нуль. Следовательно,

$$\frac{\partial z}{\partial x_k} = \sum_{j=1}^{N} \frac{\partial U_j}{\partial x_k} = \sum_{j=1}^{N} c_j a_{kj} (x_k)^{a_{kj}-1} \prod_{i \ne k} (x_i)^{a_{ij}} = 0, \quad k = 1, 2, \ldots, n.$$

Поскольку по предположению все $x_k > 0$, имеем

$$\frac{\partial z}{\partial x_k} = 0 = \frac{1}{x_k} \sum_{j=1}^{N} a_{kj} U_j, \quad k = 1, 2, \ldots, n.$$

Пусть z^* — минимальное значение функции z. Очевидно, что $z^* > 0$, так как z — позином и все $x_k > 0$. Обозначим

$$y_j = \frac{U_j^*}{z^*}.$$

Из определения следует, что $y_j > 0$ и $\sum_{j=1}^{N} y_j = 1$. Величина y_j характеризует относительный вклад j-го слагаемого U_j в оптимальное значение целевой функции z^*. *Необходимые* условия экстремума теперь можно записать в следующем виде.

$$\sum_{j=1}^{N} a_{kj} y_j = 0, \quad k = 1, 2, \ldots, n,$$

$$\sum_{j=1}^{N} y_j = 1, \quad y_j > 0 \text{ для всех } j.$$

Эти уравнения, именуемые **условиями ортогональности** и **нормировки**, определяют единственное решение для y_j при условии, что все уравнения независимы и $n + 1 = N$. Задача усложняется при $N > n + 1$, так как в этом случае решение для y_j не является единственным. Ниже показано, что и в этом случае существует возможность однозначно определить y_j для минимизации функции z.

Пусть y_j^* — элементы единственного решения представленной выше системы уравнений. Тогда величины z^* и x_i^*, $i = 1, 2, \ldots, n$, определяются следующим образом. Рассмотрим величину

$$z^* = \left(z^*\right)^{\sum_{j=1}^{N} y_j^*}.$$

Поскольку $z^* = \frac{U_j^*}{y_j^*}$, отсюда следует, что

$$z^* = \left(\frac{U_1^*}{y_1^*}\right)^{y_1^*} \left(\frac{U_2^*}{y_2^*}\right)^{y_2^*} \cdots \left(\frac{U_N^*}{y_N^*}\right)^{y_N^*} = \left\{\prod_{j=1}^{N} \left(\frac{c_j}{y_j^*}\right)^{y_j^*}\right\} \left\{\prod_{j=1}^{N} \left(\prod_{i=1}^{n} (x_i^*)^{a_{ij}}\right)^{y_j^*}\right\} =$$

$$= \left\{\prod_{j=1}^{N} \left(\frac{c_j}{y_j^*}\right)^{y_j^*}\right\} \left\{\prod_{i=1}^{n} \left(x_i^*\right)^{\sum_{j=1}^{N} a_{ij} y_j^*}\right\} = \prod_{j=1}^{N} \left(\frac{c_j}{y_j^*}\right)^{y_j^*}.$$

В последнем выражении учтено условие $\sum_{j=1}^{N} a_{ij} y_j = 0$. Следовательно, значение z^* можно вычислить, как только будут определены все y_j. При известных y_j^* и z^* можно определить $U_j^* = y_j^* z^*$. Значения x_i^* находятся как решение системы уравнений

$$U_j^* = c_j \prod_{i=1}^{n} \left(x_i^*\right)^{a_{ij}}, \quad j = 1, 2, \ldots, N.$$

Рассмотренная процедура показывает, что исходная задача минимизации позинома z может быть сведена к решению системы линейных уравнений относительно y_j. При этом значения всех y_j^* определяются из необходимых условий существования минимума. Можно показать, что эти условия являются также и достаточными. Доказательство этого приводится в книге [2].

Фактически переменные y_j определяют двойственные переменные, соответствующие прямой задаче минимизации z. Чтобы установить эту зависимость, запишем целевую функцию прямой задачи в виде

$$z = \sum_{j=1}^{N} y_j \left(\frac{U_j}{y_j} \right).$$

Определим теперь функцию

$$w = \prod_{j=1}^{N} \left(\frac{U_j}{y_j} \right)^{y_j} = \prod_{j=1}^{N} \left(\frac{c_j}{y_j} \right)^{y_j}.$$

Поскольку $\sum_{j=1}^{N} y_j = 1$ и $y_j>0$, в силу неравенства Коши[1] имеем $w \le z$.

Функция w с переменными $y_1, y_2, ..., y_N$ определяет задачу, двойственную к исходной. Поскольку w является нижней границей значений z и рассматривается задача минимизации z, отсюда следует, что, максимизируя функцию w, мы получим

$$w^* = \max_{y_j} w = \min_{x_i} z = z^*.$$

Это значит, что максимальное значение w $(= w^*)$ на множестве допустимых значений y_j совпадает с минимальным значением z $(= z^*)$ на множестве допустимых значений x_i.

Пример 21.2–4

В этом примере рассматривается задача, где $N = n + 1$, так что решение, получаемое из условий ортогональности и нормировки, является единственным. (Следующий пример иллюстрирует другой случай, когда $N > n + 1$.)

Рассмотрим задачу

$$\text{минимизировать } z = 7x_1x_2^{-1} + 3x_2x_3^{-2} + 5x_1^{-3}x_2x_3 + x_1x_2x_3.$$

Эту функцию можно записать в виде

$$z = 7x_1^1x_2^{-1}x_3^0 + 3x_1^0x_2^1x_3^{-2} + 5x_1^{-3}x_2^1x_3^1 + x_1^1x_2^1x_3^1,$$

так что здесь

$$(c_1, c_2, c_3, c_4) = (7, 3, 5, 1),$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -3 & 1 \\ -1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \end{bmatrix}.$$

Условия ортогональности и нормировки приводят к системе уравнений

$$\begin{bmatrix} 1 & 0 & -3 & 1 \\ -1 & 1 & 1 & 1 \\ 0 & -2 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

---

[1] Неравенство Коши (неравенство между средним арифметическим и средним геометрическим) устанавливает, что при $z_j > 0$ выполняется $\sum_{j=1}^{N} w_j z_j \ge \prod_{j=1}^{N} (z_j)^{w_j}$, где $w_j > 0$ и $\sum_{j=1}^{N} w_j = 1$.

Эта система имеет единственное решение

$$y_1^* = \frac{12}{24},\ y_2^* = \frac{4}{24},\ y_3^* = \frac{5}{24},\ y_4^* = \frac{3}{24}.$$

Следовательно,

$$z^* = \left(\frac{7}{\left(\frac{12}{24}\right)}\right)^{\frac{12}{24}} \left(\frac{3}{\left(\frac{4}{24}\right)}\right)^{\frac{4}{24}} \left(\frac{5}{\left(\frac{5}{24}\right)}\right)^{\frac{5}{24}} \left(\frac{1}{\left(\frac{3}{24}\right)}\right)^{\frac{3}{24}} = 15.22.$$

Из уравнения $U_j^* = y_j^* z^*$ следует, что

$$7x_1x_2^{-1} = U_1 = \left(\frac{1}{2}\right)(15.22) = 7.61,$$

$$3x_2x_3^{-2} = U_2 = \left(\frac{1}{6}\right)(15.22) = 2.54,$$

$$5x_1^{-3}x_2x_3 = U_3 = \left(\frac{5}{24}\right)(15.22) = 3.17,$$

$$x_1x_2x_3 = U_4 = \left(\frac{1}{8}\right)(15.22) = 1.90.$$

Эта система имеет решение $x_1^* = 1.315$, $x_2^* = 1.21$, $x_3^* = 1.2$, которое является оптимальным решением прямой задачи.

Пример 21.2–5

Рассмотрим задачу

$$\text{минимизировать}\quad z = 5x_1x_2^{-1} + 2x_1^{-1}x_2 + 5x_1 + x_2^{-1}.$$

Условия ортогональности и нормировки приводят к системе уравнений

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ -1 & 1 & 0 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Поскольку $N > n + 1$, эти уравнения не позволяют непосредственно определить искомые значения y_j. Выражая y_1, y_2 и y_3 через y_4, имеем

$$\begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ y_4 \\ 1 - y_4 \end{bmatrix},$$

или

$$y_1 = \frac{(1-3y_4)}{2},$$

$$y_2 = \frac{(1-y_4)}{2},$$

$$y_3 = y_4.$$

Двойственную задачу можно записать в следующем виде.

$$\text{Максимизировать } w = \left[\frac{5}{0.5(1-3y_4)}\right]^{0.5(1-3y_4)} \left[\frac{2}{0.5(1-y_4)}\right]^{0.5(1-y_4)} \left(\frac{5}{y_4}\right)^{y_4} \left(\frac{1}{y_4}\right)^{y_4}.$$

Максимизация w эквивалентна максимизации функции $\ln w$, переход к которой упрощает вычисления. Получим

$$\ln w = 0.5(1 - 3y_4)\{\ln 10 - \ln(1 - 3y_4)\} + 0.5(1 - y_4)\{\ln 4 - \ln(1 - y_4)\} + y_4\{\ln 5 - \ln y_4 + \ln 1 - \ln y_4\}.$$

Значение y_4, максимизирующее функцию $\ln w$, должно быть единственным, поскольку прямая задача имеет единственный минимум. Следовательно,

$$\frac{\partial \ln w}{\partial y_4} = \left(\frac{-3}{2}\right)\ln 10 - \left\{\left(\frac{-3}{2}\right) + \left(\frac{-3}{2}\right)\ln(1-3y_4)\right\} + \left(\frac{-1}{2}\right)\ln 4 -$$

$$- \left\{\left(\frac{-1}{2}\right) + \left(\frac{-1}{2}\right)\ln(1-y_4)\right\} + \ln 5 - \{1 + \ln y_4\} + \ln 1 - \{1 + \ln y_4\} = 0.$$

После упрощений получаем

$$-\ln\left(\frac{2\times 10^{\frac{3}{2}}}{5}\right) + \ln\left[\frac{(1-3y_4)^{\frac{3}{2}}(1-y_4)^{\frac{1}{2}}}{y_4^2}\right] = 0$$

или

$$\frac{\sqrt{(1-3y_4)^3(1-y_4)}}{y_4^2} = 12.6,$$

откуда $y_4^* \approx 0.16$. Следовательно, $y_3^* = 0.16$, $y_2^* = 0.42$ и $y_1^* = 0.26$.

Значение z^* вычисляется по формуле

$$z^* = w^* = \left(\frac{5}{0.26}\right)^{0.26} \left(\frac{2}{0.42}\right)^{0.42} \left(\frac{5}{0.16}\right)^{0.16} \approx 9.661.$$

Следовательно,

$$U_3 = 0.16 \times 9.661 = 1.546 = 5x_1,$$

$$U_4 = 0.16 \times 9.661 = 1.546 = x_2^{-1}.$$

Отсюда имеем $x_1^* = 0.309$ и $x_2^* = 0.647$.

Упражнения 21.2,с

1. Решите следующую задачу методом геометрического программирования.

$$\text{Минимизировать } z = 2x_1^{-1}x_2^2 + x_1^4x_2^{-2} + 4x_1^2 \text{ при } x_1, x_2 > 0.$$

2. Решите следующую задачу методом геометрического программирования.

$$\text{Минимизировать } z = 5x_1x_2^{-1}x_3^2 + x_1^{-2}x_3^{-1} + 10x_2^3 + 2x_1^{-1}x_2x_3^{-3} \text{ при } x_1, x_2, x_3 > 0.$$

3. Решите следующую задачу методом геометрического программирования.

$$\text{Минимизировать } z = 2x_1^2x_2^{-3} + 8x_1^{-3}x_2 + 3x_1x_2 \text{ при } x_1, x_2 > 0.$$

4. Решите следующую задачу методом геометрического программирования.

$$\text{Минимизировать } z = 2x_1^3x_2^{-3} + 4x_1^{-2}x_2 + x_1x_2 + 8x_1x_2^{-1} \text{ при } x_1, x_2 > 0.$$

21.2.4. Стохастическое программирование

В стохастическом программировании рассматриваются задачи, в которых отдельные или все параметры являются случайными величинами. Такие ситуации типичны в реальных задачах, когда трудно определить точные значения параметров.

Основной подход к решению задач стохастического программирования состоит в преобразовании исходной вероятностной задачи в эквивалентную детерминированную. В данном разделе рассматривается оптимизационная задача с **вероятностными ограничениями**, которая имеет следующий вид.

$$\text{Максимизировать } z = \sum_{j=1}^{n} c_j x_j$$

при ограничениях

$$P\left\{\sum_{j=1}^{n} a_{ij}x_j \le b_i\right\} \ge 1-\alpha_i, \quad i = 1, 2, \ldots, m; \quad x_j \ge 0 \text{ для всех } j.$$

Название "вероятностные ограничения" обусловлено тем, что каждое ограничение задачи должно выполняться с вероятностью не меньше, чем $1 - \alpha_i$, $0 < \alpha_i < 1$. Предполагается, что все коэффициенты a_{ij} и b_i являются случайными величинами. Далее рассматриваются три случая. Первые два соответствуют предположениям о том, что только или a_{ij}, или b_i являются случайными величинами, а третий объединяет эти два случая. Во всех трех ситуациях предполагается, что параметры являются нормально распределенными случайными величинами с известными математическими ожиданиями и дисперсиями.

Случай 1. Предполагается, что все a_{ij} являются нормально распределенными случайными величинами с математическими ожиданиями $M\{a_{ij}\}$ и дисперсиями $D\{a_{ij}\}$. Также известны ковариации $\text{cov}\{a_{ij}, a_{i'j'}\}$ случайных величин a_{ij} и $a_{i'j'}$.

Рассмотрим i-е ограничение задачи

$$P\left\{\sum_{j=1}^{n} a_{ij}x_j \le b_i\right\} \ge 1-\alpha_i$$

и введем обозначение

$$h_i = \sum_{j=1}^{n} a_{ij} x_j .$$

Случайная величина h_i имеет нормальное распределение с математическим ожиданием $M\{h_i\} = \sum_{j=1}^{n} M\{a_{ij}\} x_j$ и дисперсией $D\{h_i\} = \mathbf{X}^T \mathbf{D}_i \mathbf{X}$, где $\mathbf{X} = (x_1, x_2, \ldots, x_n)^T$,

$$\mathbf{D}_i = i\text{-я матрица ковариаций } = \begin{bmatrix} D\{a_{i1}\} & \cdots & \operatorname{cov}\{a_{i1}, a_{in}\} \\ \vdots & & \vdots \\ \operatorname{cov}\{a_{in}, a_{i1}\} & \cdots & D\{a_{in}\} \end{bmatrix}.$$

Теперь имеем

$$P\{h_i \le b_i\} = P\left\{ \frac{h_i - M(h_i)}{\sqrt{D\{h_i\}}} \le \frac{b_i - M(h_i)}{\sqrt{D\{h_i\}}} \right\} \ge 1 - \alpha_i ,$$

где $\dfrac{h_i - M(h_i)}{\sqrt{D\{h_i\}}}$ — нормированная нормально распределенная случайная величина с нулевым математическим ожиданием и единичной дисперсией. Это означает, что

$$P\{h_i \le b_i\} = \Phi\left(\frac{b_i - M(h_i)}{\sqrt{D\{h_i\}}} \right),$$

где через Φ обозначена функция распределения стандартного нормального распределения.

Пусть K_{α_i} — значение стандартной нормально распределенной случайной величины, определяемое из уравнения

$$\Phi\left(K_{\alpha_i}\right) = 1 - \alpha_i .$$

В этом случае неравенство $P\{h_i \le b_i\} \ge 1 - \alpha_i$ выполняется тогда и только тогда, когда

$$\frac{b_i - M(h_i)}{\sqrt{D\{h_i\}}} \ge K_{\alpha_i} .$$

Это приводит к детерминированному нелинейному ограничению, которое эквивалентно исходному вероятностному

$$\sum_{j=1}^{n} M\{a_{ij}\} x_j + K_{\alpha_i} \sqrt{\mathbf{X}^T \mathbf{D}_i \mathbf{X}} \le b_i .$$

В частности, если a_{ij} — независимые нормально распределенные случайные величины, тогда $\operatorname{cov}\{a_{ij}, a_{i'j'}\} = 0$ и последнее неравенство принимает вид

$$\sum_{j=1}^{n} M\{a_{ij}\} x_j + K_{\alpha_i} \sqrt{\sum_{j=1}^{n} D\{a_{ij}\} x_j^2} \le b_i .$$

Данное ограничение можно записать в виде ограничений задачи сепарабельного программирования (раздел 21.2.1), для чего используется замена переменных

$$y_i = \sqrt{\sum_{j=1}^{n} D\{a_{ij}\}x_j^2} \quad \text{для всех } i.$$

Таким образом, исходное ограничение эквивалентно неравенству

$$\sum_{j=1}^{n} M\{a_{ij}\}x_j + K_{\alpha_i} y_i \le b_i$$

и уравнению

$$\sum_{j=1}^{n} D\{a_{ij}\}x_j^2 - y_i^2 = 0.$$

Случай 2. Здесь предполагается, что только b_i являются нормально распределенными случайными величинами с математическим ожиданием $M\{b_i\}$ и дисперсией $D\{b_i\}$. Анализ этой ситуации проводится аналогично случаю 1. Рассмотрим стохастическое ограничение

$$P\left\{b_i \ge \sum_{j=1}^{n} a_{ij}x_j\right\} \ge \alpha_i.$$

Как и в первом случае, имеем

$$P\left\{\frac{b_i - M(b_i)}{\sqrt{D\{b_i\}}} \ge \frac{\sum_{j=1}^{n} a_{ij}x_j - M(b_i)}{\sqrt{D\{b_i\}}}\right\} \ge \alpha_i.$$

Это ограничение выполнимо лишь при выполнении неравенства

$$\frac{\sum_{j=1}^{n} a_{ij}x_j - M(b_i)}{\sqrt{D\{b_i\}}} \le K_{\alpha_i}.$$

Итак, исходное вероятностное ограничение эквивалентно детерминированному линейному

$$\sum_{j=1}^{n} a_{ij}x_j \le M\{b_i\} + K_{\alpha_i}\sqrt{D\{b_i\}}.$$

Случай 3. Теперь предположим, что все параметры a_{ij} и b_i являются нормально распределенными случайными величинами. Перепишем ограничение

$$\sum_{j=1}^{n} a_{ij}x_j \le b_i$$

в виде

$$\sum_{j=1}^{n} a_{ij}x_j - b_i \le 0.$$

Так как все a_{ij} и b_i распределены по нормальному закону, в соответствии с известными результатами математической статистики величина $\sum_{j=1}^{n} a_{ij}x_j - b_i$ также имеет нормальное распределение. Отсюда следует, что данный вариант подобен случаю 1 и может быть рассмотрен аналогичным образом.

Пример 21.2–6

Рассмотрим задачу с вероятностными ограничениями

$$\text{максимизировать } z = 5x_1 + 6x_2 + 3x_3$$

при ограничениях

$$P\{a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \le 8\} \ge 0.95,$$

$$P\{5x_1 + x_2 + 6x_3 \le b_2\} \ge 0.10,$$

при чем все $x_j \ge 0$. Пусть a_{1j} — *независимые* нормально распределенные случайные величины со следующими значениями математических ожиданий и дисперсий

$$M\{a_{11}\} = 1,\ M\{a_{12}\} = 3,\ M\{a_{13}\} = 9,$$

$$D\{a_{11}\} = 25,\ D\{a_{12}\} = 16,\ D\{a_{13}\} = 4.$$

Пусть параметр b_2 является нормально распределенной случайной величиной с математическим ожиданием 7 и дисперсией 9.

По таблице функции распределения стандартного нормального закона (Приложение В) находим

$$K_{\alpha_1} = K_{0.05} \approx 1.645,\ \ K_{\alpha_2} = K_{0.10} \approx 1.285.$$

Первое ограничение задачи эквивалентно детерминированному неравенству

$$x_1 + 3x_2 + 9x_3 + 1.645\sqrt{25x_1^2 + 16x_2^2 + 4x_3^2} \le 8,$$

а второе ограничение — неравенству

$$5x_1 + x_2 + 6x_3 \le 7 + 1.285 \times 3 = 10.855.$$

Если положить

$$y^2 = 25x_1^2 + 16x_2^2 + 4x_3^2,$$

исходная задача принимает следующий вид.

$$\text{Максимизировать } z = 5x_1 + 6x_2 + 3x_3$$

при ограничениях

$$x_1 + 3x_2 + 9x_3 + 1.645y \le 8,$$

$$25x_1^2 + 16x_2^2 + 4x_3^2 - y^2 = 0,$$

$$5x_1 + x_2 + 6x_3 \le 10.855,$$

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

и может быть решена методами сепарабельного программирования.

Упражнения 21.2,d

1. Преобразуйте следующую задачу стохастического программирования в эквивалентную детерминированную модель.

$$\text{Максимизировать } z = x_1 + 2x_2 + 5x_3$$

при ограничениях

$$P\{a_1x_1 + 3x_2 + a_3x_3 \le 10\} \ge 0.9,$$

$$P\{7x_1 + 5x_2 + x_3 \le b_2\} \ge 0.1,$$

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

Пусть a_1 и a_3 являются независимыми нормально распределенными случайными величинами с математическими ожиданиями $M\{a_1\} = 2$ и $M\{a_3\} = 5$ и дисперсиями $D\{a_1\} = 9$ и $D\{a_3\} = 16$. Предполагается также, что b_2 — нормально распределенная случайная величина с математическим ожиданием 15 и дисперсией 25.

2. Дана следующая задача стохастического программирования.

$$\text{Максимизировать } z = x_1 + x_2^2 + x_3$$

при ограничениях

$$P\left\{x_1^2 + a_2x_2^3 + a_3\sqrt{x_3} \le 10\right\} \ge 0.9,$$

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

Пусть параметры a_2 и a_3 — независимые нормально распределенные случайные величины с математическими ожиданиями 5 и 2 и дисперсиями 16 и 25 соответственно. Преобразуйте данную задачу в детерминированную задачу сепарабельного программирования.

21.2.5. Метод линейных комбинаций

Этот метод ориентирован на решение задач, в которых все ограничения являются линейными:

$$\text{максимизировать } z = f(\mathbf{X})$$

при ограничениях

$$\mathbf{AX} \le \mathbf{b}, \mathbf{X} \ge \mathbf{0}.$$

Основой метода линейных комбинаций является градиентный метод наискорейшего подъема (раздел 21.1.2). Известно, что в этом методе движение по направлению градиента может вывести за пределы области допустимых решений. Кроме того, в точке условного оптимума градиент не обязательно обращается в нуль. Поэтому метод наискорейшего подъема необходимо модифицировать в целях его применения к рассматриваемой задаче с ограничениями.

Пусть $\mathbf{X}^k$ — *допустимая* точка, полученная на k-й итерации. Разложим целевую функцию $f(\mathbf{X})$ в окрестности точки $\mathbf{X}^k$ в ряд Тейлора. В результате получим

$$f(\mathbf{X}) \approx f(\mathbf{X}^k) + \nabla f(\mathbf{X}^k)(\mathbf{X} - \mathbf{X}^k) = f(\mathbf{X}^k) - \nabla f(\mathbf{X}^k)\mathbf{X}^k + \nabla f(\mathbf{X}^k)\mathbf{X}.$$

Нам необходимо определить допустимую точку $\mathbf{X} = \mathbf{X}^*$, в которой достигается максимум функции $f(\mathbf{X})$ при выполнении (линейных) ограничений задачи. Так как $f(\mathbf{X}^k) - \nabla f(\mathbf{X}^k)\mathbf{X}^k$ — постоянное слагаемое, задача определения $\mathbf{X}^*$ сводится к решению следующей задачи линейного программирования:

$$\text{максимизировать } w_k(\mathbf{X}) = \nabla f(\mathbf{X}^k)\mathbf{X}$$

при ограничениях

$$\mathbf{AX} \le \mathbf{b},\ \mathbf{X} \ge \mathbf{0}.$$

Так как целевая функция w_k определяется градиентом функции $f(\mathbf{X})$ в точке $\mathbf{X}^k$, улучшение имеющегося решения может быть обеспечено лишь тогда, когда $w_k(\mathbf{X}^*) > w_k(\mathbf{X}^k)$. Из разложения Тейлора следует, что выполнение этого условия не гарантирует, что $f(\mathbf{X}^*) > f(\mathbf{X}^k)$, так как $\mathbf{X}^*$ не обязательно находится в малой окрестности точки $\mathbf{X}^k$. Однако при выполнении условия $w_k(\mathbf{X}^*) > w_k(\mathbf{X}^k)$ на отрезке $(\mathbf{X}^k, \mathbf{X}^*)$ должна существовать точка $\mathbf{X}^{k+1}$, такая, что $f(\mathbf{X}^{k+1}) > f(\mathbf{X}^k)$. Следовательно, задача сводится к нахождению такой точки $\mathbf{X}^{k+1}$. Определим эту точку следующим образом:

$$\mathbf{X}^{k+1} = (1 - r)\mathbf{X}^k + r\mathbf{X}^* = \mathbf{X}^k + r(\mathbf{X}^* - \mathbf{X}^k),\ 0 < r \le 1.$$

Отсюда следует, что точка $\mathbf{X}^{k+1}$ является **линейной комбинацией** точек $\mathbf{X}^k$ и $\mathbf{X}^*$. Так как $\mathbf{X}^k$ и $\mathbf{X}^*$ — точки *выпуклой* области допустимых решений, поэтому точка $\mathbf{X}^{k+1}$ также является допустимой. Если сравнить эту процедуру с методом наискорейшего подъема (раздел 21.1.2), параметр r можно рассматривать как длину шага.

Точка $\mathbf{X}^{k+1}$ определяется из условия максимума функции $f(\mathbf{X})$. Так как $\mathbf{X}^{k+1}$ зависит лишь от переменной r, то $\mathbf{X}^{k+1}$ определяется путем максимизации функции

$$h(r) = f[\mathbf{X}^k + r(\mathbf{X}^* - \mathbf{X}^k)].$$

Процедура повторяется до тех пор, пока на k-й итерации не будет выполнено условие $w_k(\mathbf{X}^*) \le w_k(\mathbf{X}^k)$. В этой точке дальнейшее улучшение имеющегося решения невозможно, процесс вычислений завершается, точка $\mathbf{X}^k$ считается наилучшим решением.

Задачи линейного программирования, которые решаются на каждой итерации описанной процедуры, отличаются друг от друга лишь коэффициентами целевой функции. Поэтому для повышения эффективности вычислений можно использовать методы анализа чувствительности, рассмотренные в разделе 4.7.

Пример 21.2–7

Рассмотрим задачу квадратичного программирования из примера 21.2–3.

$$\text{Максимизировать } z = 4x_1 + 6x_2 - 2x_1^2 - 2x_1x_2 - 2x_2^2$$

при ограничениях

$$x_1 + 2x_2 \le 2,$$
$$x_1, x_2 \ge 0.$$

Пусть $\mathbf{X}^0 = (1/2,\ 1/2)$ — начальная точка, которая является допустимой для рассматриваемой задачи. Имеем

$$\nabla f(\mathbf{X}) = (4 - 4x_1 - 2x_2,\ 6 - 2x_1 - 4x_2).$$

Первая итерация

$$\nabla f(\mathbf{X}^0) = (1,\ 3).$$

Соответствующая задача линейного программирования сводится к максимизации функции $w_1 = x_1 + 3x_2$ с учетом ограничений исходной задачи. Ее оптимальным решением есть точка $\mathbf{X}^* = (0, 1)$. Значения функции w_1 в точках $\mathbf{X}^0$ и $\mathbf{X}^*$ равны 2 и 3 соответственно. Следовательно, новая точка ищется в виде

$$\mathbf{X}^1 = \left(\frac{1}{2}, \frac{1}{2}\right) + r\left[(0,1) - \left(\frac{1}{2}, \frac{1}{2}\right)\right] = \left(\frac{1-r}{2}, \frac{1+r}{2}\right).$$

Максимизация функции

$$h(r) = f\left(\frac{1-r}{2}, \frac{1+r}{2}\right)$$

дает $r = 1$. Таким образом, $\mathbf{X}^1 = (0, 1)$ и $f(\mathbf{X}^1) = 4$.

Вторая итерация

$$\nabla f(\mathbf{X}^1) = (2, 2).$$

Целевой функцией новой задачи линейного программирования является $w_2 = 2x_1 + 2x_2$. Оптимальное решение этой задачи — $\mathbf{X}^* = (2, 0)$. Поскольку значения w_2 в точках $\mathbf{X}^1$ и $\mathbf{X}^*$ равны 2 и 4, следует найти новую точку. Рассматриваем точку

$$\mathbf{X}^2 = (0, 1) + r[(2, 0) - (0, 1)] = (2r, 1 - r).$$

Максимизация функции

$$h(r) = f(2r, 1 - r)$$

дает $r = 1/6$. Таким образом, $\mathbf{X}^2 = (1/3, 5/6)$, при этом $f(\mathbf{X}^2) \approx 4.16$.

Третья итерация

$$\nabla f(\mathbf{X}^2) = (1, 2).$$

Здесь целевая функция имеет вид $w_3 = x_1 + 2x_2$. Оптимальным решением задачи линейного программирования будут альтернативные точки $\mathbf{X}^* = (0, 1)$ и $\mathbf{X}^* = (2, 0)$. Значение w_3 в обоих случаях совпадает со значением w_3 в точке $\mathbf{X}^2$. Следовательно, улучшить имеющееся решение невозможно. Получено *приближенное* оптимальное решение задачи: $\mathbf{X}^2 = (1/3, 5/6)$ с $f(\mathbf{X}^2) \approx 4.16$. В данном случае оно совпадает с точным оптимумом.

Упражнение 21.2,е

1. Решите следующую задачу методом линейных комбинаций.

$$\text{Минимизировать } f(\mathbf{X}) = x_1^3 + x_2^3 - 3x_1x_2$$

при ограничениях

$$3x_1 + x_2 \le 3,$$
$$5x_1 - 3x_2 \le 5,$$
$$x_1, x_2 \ge 0.$$

21.2.6. Алгоритм последовательной безусловной максимизации

В этом разделе приводится описание градиентного метода более общего вида. Предполагается, что целевая функция $f(\mathbf{X})$ является вогнутой, а каждая функция ограничений $g_i(\mathbf{X})$ — выпуклой. Кроме того, допустимая область задачи должна содержать внутренние точки. Это исключает как явное, так и неявное использование ограничений в виде *равенств.*

Алгоритм последовательной безусловной максимизации основан на преобразовании исходной задачи с ограничениями в эквивалентную задачу *без* ограничений. Процедура в некоторой степени подобна применению метода множителей Лагранжа. Преобразованная задача затем может быть решена методом наискорейшего подъема (раздел 21.1.2).

Введем новую целевую функцию

$$p(\mathbf{X},t)=f(\mathbf{X})+t\left(\sum_{i=1}^{m}\frac{1}{g_i(\mathbf{X})}-\sum_{j=1}^{n}\frac{1}{x_j}\right),$$

где t — неотрицательный параметр. Наличие второй суммы обусловлено учетом требований неотрицательности переменных исходной задачи, которые следует записать в виде $-x_j\le 0$ для их согласования с остальными ограничениями вида $g_i(\mathbf{X})\le 0$. Поскольку функция $g_i(\mathbf{X})$ является выпуклой, $1/g_i(\mathbf{X})$ будет вогнутой. Отсюда следует, что $p(\mathbf{X}, t)$ является вогнутой функцией от переменных $\mathbf{X}$. Следовательно, функция $p(\mathbf{X}, t)$ имеет единственный максимум. Поэтому решение исходной оптимизационной задачи с ограничениями эквивалентно максимизации функции $p(\mathbf{X}, t)$.

Выполнение алгоритма начинается с произвольного выбора начального *неотрицательного* значения t. Начальная точка $\mathbf{X}^0$ выбирается в качестве первого пробного решения. Эта точка должна быть внутренней точкой области допустимых решений, т.е. не должна находиться на ее границе. При заданном значении t оптимальное решение (точка максимума) функции $p(\mathbf{X}, t)$ находится методом наискорейшего подъема.

Точка, представляющая новое решение, всегда будет внутренней, ибо если решение находится близко к границе области, то по крайней мере одна из функций $1/g_i(\mathbf{X})$ или $-1/x_j$ примет очень большое отрицательное значение. Поскольку рассматривается задача максимизации $p(\mathbf{X}, t)$, то такие решения автоматически исключаются. Основной результат состоит в том, что последовательно получаемые решения всегда будут внутренними точками допустимой области. Следовательно, исходная задача может всегда рассматриваться как задача без ограничений.

Когда получено оптимальное решение, соответствующее заданному значению t, выбирается новое значение t и процесс вычислений (методом наискорейшего подъема) повторяется. Если t' — текущее значение t, то следующее значение t'' необходимо выбирать так, чтобы выполнялись неравенства $0<t''<t'$.

Решение задачи методом последовательной безусловной максимизации завершается, когда для двух последовательных значений t соответствующие *оптимальные* решения $\mathbf{X}$, полученные в результате максимизации функции $p(\mathbf{X}, t)$, примерно совпадают. В этом случае дальнейшие вычисления лишь незначительно улучшат полученное решение.

Практическая реализация алгоритма последовательной безусловной максимизации имеет ряд нюансов, которые здесь не рассматриваются. В частности, существенным фактором является выбор начального значения параметра t, который может повлиять на скорость сходимости алгоритма. Далее для определения начальной внутренней точки области допустимых решений могут потребоваться специальные алгоритмы. Соответствующие детали можно найти в книге [3].

21.3. Заключение

Методы решения задач нелинейного программирования можно классифицировать как *прямые* и *непрямые*. Примерами прямых методов являются градиентные алгоритмы, в которых поиск максимума (минимума) в оптимизационной задаче ведется по направлению наискорейшего увеличения (уменьшения) значения целевой функции. В непрямых методах исходная задача оптимизации заменяется вспомогательной, оптимальное решение которой принимается за решение исходной. Частными случаями таких методов являются квадратичное, сепарабельное и стохастическое программирование.

Литература

1. Bazaraa M., Sherall H. and Shetty C. *Nonlinear Programming, Theory and Algorithms*, 2nd ed.,Wiley, New York, 1993. (Имеется перевод первого издания: Базара М., Шетти К. *Нелинейное программирование. Теория и алгоритмы.* — М.: Мир, 1982.)
2. Beightler C., Phillips D. and Wilde D. *Foundations of Optimization*, 2nd ed., Prentice Hall, Upper Saddle River, N.J., 1979.
3. Fiacco A. and McCormick G. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968. (Имеется перевод: Фиакко А., Мак-Кормик Г. *Нелинейное программирование. Методы последовательной безусловной минимизации.* — М.: Мир, 1972.)

Литература, добавленная при переводе

Банди Б. *Методы оптимизации. Вводный курс.* — М.: Радио и связь, 1988.
Гилл Ф., Мюррей У., Райт М. *Практическая оптимизация.* — М.: Мир, 1985.
Зангвилл У.И. *Нелинейное программирование.* — М.: Сов. радио, 1973.
Сухарев А.Г., Тимохов А.В., Федоров В.В. *Курс методов оптимизации.* — М.: Наука, 1986.
Химмельблау Д. *Прикладное нелинейное программирование.* — М.: Мир, 1975.

Краткий обзор теории матриц

А.1. Векторы

А.1.1. Определение вектора

Пусть $p_1, p_2, ..., p_n$ — произвольные действительные числа. Обозначим через $\mathbf{P}$ упорядоченное множество этих чисел: $\mathbf{P} = (p_1, p_2, ..., p_n)$. В этом случае $\mathbf{P}$ называется n-мерным вектором (или просто вектором), а p_i — i-м элементом вектора $\mathbf{P}$. Например, $\mathbf{P} = (1, 2)$ — 2-мерный вектор.

А.1.2. Сложение и вычитание векторов

Пусть $\mathbf{P} = (p_1, p_2, ..., p_n)$ и $\mathbf{Q} = (q_1, q_2, ..., q_n)$ — два n-мерных вектора. Тогда элементы вектора $\mathbf{R} = (r_1, r_2, ..., r_n)$, равного $\mathbf{R} = \mathbf{P} \pm \mathbf{Q}$, определяются соотношениями $r_i = p_i \pm q_i$.

В общем случае для любых векторов $\mathbf{P}$, $\mathbf{Q}$ и $\mathbf{S}$, имеющих одинаковую размерность, выполняются следующие соотношения.

$\mathbf{P} \pm \mathbf{Q} = \mathbf{Q} \pm \mathbf{P}$ (Свойство коммутативности)

$(\mathbf{P} + \mathbf{Q}) + \mathbf{S} = \mathbf{P} + (\mathbf{Q} + \mathbf{S})$ (Свойство ассоциативности)

$\mathbf{P} + (-\mathbf{P}) = \mathbf{0}$ (Существование нулевого вектора)

А.1.3. Умножение вектора на скаляр

Для произвольного вектора $\mathbf{P}$ и скаляра θ (произвольного действительного числа) произведение вектора $\mathbf{P}$ на скаляр θ определяет новый вектор $\mathbf{Q}$, задаваемый соотношением

$$\mathbf{Q} = \theta\mathbf{P} = (\theta p_1, \theta p_2, ..., \theta p_n).$$

В общем случае для любых векторов $\mathbf{P}$ и $\mathbf{S}$, имеющих одинаковую размерность, и произвольных скалярных величин θ и γ выполняются следующие соотношения.

$\theta(\mathbf{P} + \mathbf{S}) = \theta\mathbf{P} + \theta\mathbf{S}$ (Свойство дистрибутивности)

$\theta(\gamma\mathbf{P}) = (\theta\gamma)\mathbf{P}$ (Свойство ассоциативности)

А.1.4. Линейная независимость векторов

Векторы $\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_n$ называются *линейно независимыми*, если равенство

$$\sum_{j=1}^{n} \theta_j \mathbf{P}_j = \mathbf{0}$$

выполняется тогда и только тогда, когда все θ_j равны нулю (θ_j — произвольные действительные числа). Если указанное равенство выполняется при некоторых $\theta_j \neq 0$, то векторы $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n$ называются *линейно зависимыми*. Например, векторы $\mathbf{P}_1 = (1, 2)$ и $\mathbf{P}_2 = (2, 4)$ линейно зависимы, поскольку существуют ненулевые числа $\theta_1 = 2$ и $\theta_2 = -1$, при которых выполняется равенство

$$\theta_1\mathbf{P}_1 + \theta_2\mathbf{P}_2 = \mathbf{0}.$$

А.2. Матрицы

А.2.1. Определение матриц

Матрицей называется прямоугольный массив элементов, структурированный строками и столбцами. В матрице $\mathbf{A}$ элемент a_{ij} расположен на пересечении i-й строки и j-го столбца массива. Говорят, что матрица имеет порядок (размерность) $m \times n$, если она состоит из m строк и n столбцов. Например, матрица

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} = \left\| a_{ij} \right\|_{4\times3}$$

имеет размерность 4×3.

А.2.2. Типы матриц

1. *Квадратная матрица* — это матрица, имеющая одинаковое количество строк и столбцов (т.е. $m = n$).
2. *Единичная матрица* — квадратная матрица, у которой все диагональные элементы равны 1, а все недиагональные — нулю. Например, единичная матрица порядка 3×3 имеет вид

$$\mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

3. *Вектор-строка* — матрица, имеющая одну строку и n столбцов.
4. *Вектор-столбец* — матрица, имеющая m строк и один столбец.
5. Матрица $\mathbf{A}^T$ называется *транспонированной* к матрице $\mathbf{A}$, если элемент a_{ij} матрицы $\mathbf{A}^T$ равен элементу a_{ji} матрицы $\mathbf{A}$. Например, если

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix},$$

то

$$\mathbf{A}^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}.$$

6. Матрица $\mathbf{B}$ называется *нулевой* ($\mathbf{B} = \mathbf{0}$), если все ее элементы равны нулю.
7. Две матрицы $\mathbf{A} = \|a_{ij}\|$ и $\mathbf{B} = \|b_{ij}\|$ равны тогда и только тогда, когда они имеют одинаковую размерность и $a_{ij} = b_{ij}$ для всех i и j.

А.2.3. Арифметические операции над матрицами

Для матриц определены только операции сложения (вычитания) и умножения. Операция деления матриц не определена, но в некоторых случаях ее может заменить операция обращения матриц (см. раздел А.2.6.).

Сложение и вычитание матриц. Сложение (вычитание) двух матриц $\mathbf{A} = \|a_{ij}\|$ и $\mathbf{B} = \|b_{ij}\|$ возможно только тогда, когда они имеют одинаковую размерность. Матрица суммы получается путем сложения элементов матриц $\mathbf{A}$ и $\mathbf{B}$, т.е.

$$\mathbf{D} = \mathbf{A} + \mathbf{B} = \|d_{ij}\|_{m\times n} = \|a_{ij} + b_{ij}\|_{m\times n}.$$

Для произвольных матриц $\mathbf{A}$, $\mathbf{B}$ и $\mathbf{C}$, имеющих одинаковую размерность, справедливы следующие соотношения.

$\mathbf{A} \pm \mathbf{B} = \mathbf{B} \pm \mathbf{A}$ (Свойство коммутативности)

$\mathbf{A} \pm (\mathbf{B} \pm \mathbf{C}) = (\mathbf{A} \pm \mathbf{B}) \pm \mathbf{C}$ (Свойство ассоциативности)

$(\mathbf{A} \pm \mathbf{B})^T = \mathbf{A}^T \pm \mathbf{B}^T$

Произведение матриц. Произведение $\mathbf{AB}$ матриц $\mathbf{A} = \|a_{ij}\|$ и $\mathbf{B} = \|b_{ij}\|$ определено тогда и только тогда, когда количество столбцов матрицы $\mathbf{A}$ равно количеству строк матрицы $\mathbf{B}$. Таким образом, если матрица $\mathbf{A}$ имеет размерность $m \times r$, матрица В должна иметь размерность $r \times n$, где m и n — произвольные положительные целые числа.

Пусть $\mathbf{D} = \mathbf{AB}$. Тогда матрица $\mathbf{D}$ имеет размерность $m \times n$, и ее элементы d_{ij} для всех i и j определяются формулой

$$d_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj}.$$

Например, если

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \text{ и } \mathbf{B} = \begin{bmatrix} 5 & 7 & 9 \\ 6 & 8 & 0 \end{bmatrix},$$

тогда

$$\mathbf{D} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 & 9 \\ 6 & 8 & 0 \end{bmatrix} = \begin{bmatrix} (1\times 5 + 3\times 6) & (1\times 7 + 3\times 8) & (1\times 9 + 3\times 0) \\ (2\times 5 + 4\times 6) & (2\times 7 + 4\times 8) & (2\times 9 + 4\times 0) \end{bmatrix} =$$

$$= \begin{bmatrix} 23 & 31 & 9 \\ 34 & 46 & 18 \end{bmatrix}.$$

Произведение матриц обладает следующими свойствами.

$\mathbf{I}_m\mathbf{A} = \mathbf{A}\mathbf{I}_n = \mathbf{A}$, где $\mathbf{I}$ — единичная матрица,

$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$,

$$\mathbf{C}(\mathbf{A}+\mathbf{B}) = \mathbf{CA}+\mathbf{CB},$$

$$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC},$$

$$\alpha(\mathbf{AB}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B}),\ \alpha \text{ — скаляр.}$$

Умножение блочных матриц. Пусть матрицы **A** и **B** имеют размерности $m \times r$ и $r \times n$ соответственно. Предположим, что эти матрицы представимы в виде совокупности подматриц (блоков):

$$\mathbf{A} = \left[\begin{array}{c|c|c} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} \end{array}\right] \text{ и } \mathbf{B} = \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \\ \hline \mathbf{B}_{31} & \mathbf{B}_{32} \end{array}\right],$$

причем для всех i и j число столбцов в блоке $\mathbf{A}_{ij}$ равно числу строк в блоке $\mathbf{B}_{ji}$. Тогда

$$\mathbf{A}\times\mathbf{B} = \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{B}_{11}+\mathbf{A}_{12}\mathbf{B}_{21}+\mathbf{A}_{13}\mathbf{B}_{31} & \mathbf{A}_{11}\mathbf{B}_{12}+\mathbf{A}_{12}\mathbf{B}_{22}+\mathbf{A}_{13}\mathbf{B}_{32} \\ \hline \mathbf{A}_{21}\mathbf{B}_{11}+\mathbf{A}_{22}\mathbf{B}_{21}+\mathbf{A}_{23}\mathbf{B}_{31} & \mathbf{A}_{21}\mathbf{B}_{12}+\mathbf{A}_{22}\mathbf{B}_{22}+\mathbf{A}_{23}\mathbf{B}_{32} \end{array}\right].$$

Например,

$$\left[\begin{array}{c|cc} 1 & 2 & 3 \\ \hline 1 & 0 & 5 \\ 2 & 5 & 6 \end{array}\right] \left[\begin{array}{c} 4 \\ \hline 1 \\ 8 \end{array}\right] = \left[\begin{array}{c} (1)(4)+(2\quad 3)\begin{bmatrix}1\\8\end{bmatrix} \\ \hline \begin{bmatrix}1\\2\end{bmatrix}(4)+\begin{bmatrix}0 & 5\\5 & 6\end{bmatrix}\begin{bmatrix}1\\8\end{bmatrix} \end{array}\right] = \left[\begin{array}{c} 4+2+24 \\ \hline \begin{bmatrix}4\\8\end{bmatrix}+\begin{bmatrix}40\\53\end{bmatrix} \end{array}\right] = \begin{bmatrix}30\\44\\61\end{bmatrix}.$$

А.2.4. Определитель квадратной матрицы

Для квадратной матрицы

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

порядка n рассмотрим произведение ее элементов

$$P_{j_1 j_2 \ldots j_n} = a_{1j_1} a_{2j_2} \ldots a_{nj_n},$$

где каждый столбец и каждая строка матрицы **A** представлены в точности одним элементом. Определим величину $\in_{j_1 j_2 \ldots j_n}$, равную $+1$, если множество индексов $j_1, j_2, \ldots, j_n$ получено из множества натуральных чисел $1, 2, \ldots, n$ четным числом парных перестановок, и равную -1 — в противном случае. Тогда скалярная величина

$$\sum_{\rho} \in_{j_1 j_2 \ldots j_n} P_{j_1 j_2 \ldots j_n},$$

где суммирование ведется по всем $n!$ перестановкам индексов $j_1, j_2, \ldots, j_n$, называется *определителем* (*детерминантом*) матрицы **A**. Определитель матрицы обычно обозначается как $\det \mathbf{A}$ или $|\mathbf{A}|$.

Например, если

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

тогда $|\mathbf{A}| = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{31}a_{23}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$.

Определители обладают следующими свойствами.

1. Если все элементы какого-нибудь столбца или строки матрицы равны нулю, то определитель этой матрицы равен нулю.
2. Определитель транспонированной матрицы равен определителю исходной матрицы, т.е. $|\mathbf{A}^T| = |\mathbf{A}|$.
3. Если матрица **B** получена из матрицы **A** путем перестановки двух каких-либо строк (или двух столбцов), тогда $|\mathbf{B}| = -|\mathbf{A}|$.
4. Если две строки (или два столбца) в матрице одинаковы, то ее определитель равен нулю.
5. Значение определителя не изменится, если какую-либо строку матрицы (столбец) умножить на скаляр и затем прибавить ее к другой строке (столбцу).
6. Если каждый элемент какой-либо строки (столбца) умножить на скаляр α, то значение определителя также будет умножено на это число α.
7. Если **A** и **B** — две квадратные матрицы порядка n, тогда

$$|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}|.$$

Определение минора. Минором M_{ij} элемента a_{ij} в определителе n-го порядка $|\mathbf{A}|$ называется определитель $(n-1)$-го порядка, получаемый после вычеркивания в матрице **A** i-й строки и j-го столбца. Например, в определителе матрицы

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

имеем миноры

$$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix},\ M_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \ldots .$$

Присоединенная матрица. Обозначим через $A_{ij} = (-1)^{i+j} M_{ij}$ *алгебраическое дополнение* элемента a_{ij} квадратной матрицы **A**. Тогда матрица, присоединенная к матрице **A**, определяется соотношением

$$\operatorname{adj}\mathbf{A} = \left\| A_{ij} \right\|^T = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}.$$

Например, если

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{bmatrix},$$

тогда $A_{11} = (-1)^2(3 \times 4 - 2 \times 3) = 6$, $A_{12} = (-1)^3(2 \times 4 - 2 \times 3) = -2$ и т.д. Отсюда получаем

$$\operatorname{adj}\mathbf{A} = \begin{bmatrix} 6 & 1 & -5 \\ -2 & -5 & 4 \\ -3 & 3 & -1 \end{bmatrix}.$$

А.2.5. Невырожденная матрица

Рангом матрицы называется порядок наибольшей квадратной подматрицы, определитель которой отличен от нуля. *Квадратная* матрица, определитель которой отличен от нуля, называется **матрицей полного ранга, или невырожденной матрицей**. Например, матрица

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{bmatrix},$$

является вырожденной, поскольку

$$|\mathbf{A}| = 1(21 - 20) - 2(14 - 12) + 3(10 - 9) = 0.$$

Вместе с тем матрица **A** имеет ранг $r = 2$, так как

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = -1 \neq 0.$$

А.2.6. Обратная матрица

Если **B** и **C** — две квадратные матрицы порядка n, причем такие, что $\mathbf{BC} = \mathbf{CB} = \mathbf{I}$, тогда матрица **B** называется обратной к матрице **C**, при этом матрица **C** также будет обратной к матрице **B**. Обратные матрицы обозначаются как $\mathbf{B}^{-1}$ и $\mathbf{C}^{-1}$.

Теорема. *Если* $\mathbf{BC} = \mathbf{I}$ *и* **B** *— невырожденная матрица, тогда* $\mathbf{C} = \mathrm{B}^{-1}$, *причем матрица* **C** *определяется единственным образом.*

Доказательство. По условию теоремы $\mathbf{BC} = \mathbf{I}$. Тогда умножая это равенство справа на $\mathbf{B}^{-1}$, получим $\mathbf{B}^{-1}\mathbf{BC} = \mathbf{B}^{-1}\mathbf{I}$, откуда следует, что $\mathbf{IC} = \mathbf{B}^{-1}$ или $\mathbf{C} = \mathbf{B}^{-1}$. Теорема доказана.

Для невырожденных матриц справедливы также следующие результаты.

1. Если **A** и **B** невырожденные квадратные матрицы одинаковой размерности, тогда $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
2. Если **A** — невырожденная матрица, тогда из равенства $\mathbf{AB} = \mathbf{AC}$ вытекает, что $\mathbf{B} = \mathbf{C}$.

Обратные матрицы находят применение при решении систем линейных уравнений. Рассмотрим систему из n линейных *независимых* уравнений

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

где x_i — неизвестные, a_{ij} и b_i — заданные константы. Эта система в матричной форме запишется

$$\mathbf{AX} = \mathbf{b}.$$

Поскольку уравнения системы линейно независимы, матрица **A** будет невырожденной и, следовательно, будет существовать обратная к ней матрица. Таким образом, имеем

$$\mathbf{A}^{-1}\mathbf{AX} = \mathbf{A}^{-1}\mathbf{b},$$

откуда получаем решение системы: $\mathbf{X} = \mathbf{A}^{-1}\mathbf{b}$.

А.2.7. Методы вычисления обратных матриц

Метод присоединенной матрицы. Для невырожденной матрицы **A** порядка n справедлива формула

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\operatorname{adj}\mathbf{A} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix}.$$

Например, для матрицы

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{bmatrix},$$

имеем $\operatorname{adj}\mathbf{A} = \begin{bmatrix} 6 & 1 & -5 \\ -2 & -5 & 4 \\ -3 & 3 & -1 \end{bmatrix}$ и $|\mathbf{A}| = -7$. Поэтому

$$\mathbf{A}^{-1} = \frac{1}{-7}\begin{bmatrix} 6 & 1 & -5 \\ -2 & -5 & 4 \\ -3 & 3 & -1 \end{bmatrix} = \begin{bmatrix} -\frac{6}{7} & -\frac{1}{7} & \frac{5}{7} \\ \frac{2}{7} & \frac{5}{7} & -\frac{4}{7} \\ \frac{3}{7} & -\frac{3}{7} & \frac{1}{7} \end{bmatrix}.$$

Метод последовательных исключений (метод Гаусса–Жордана). Рассмотрим блочную матрицу $(\mathbf{A} \mid \mathbf{I})$, где **A** — невырожденная матрица. Умножая слева эту матрицу на матрицу $\mathbf{A}^{-1}$, получим

$$(\mathbf{A}^{-1}\mathbf{A} \mid \mathbf{A}^{-1}\mathbf{I}) = (\mathbf{I} \mid \mathbf{A}^{-1}).$$

Таким образом, при последовательном преобразовании строк исходной матрицы, обеспечивающем преобразование матрицы **A** в **I**, одновременно матрица **I** преобразуется в $\mathbf{A}^{-1}$.

Рассмотрим систему линейных уравнений

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}.$$

Вектор решений **X** и матрицу, обратную к матрице данной системы, можно получить из соотношения

$$\mathbf{A}^{-1}(\mathbf{A} \mid \mathbf{I} \mid \mathbf{b}) = (\mathbf{I} \mid \mathbf{A}^{-1} \mid \mathbf{A}^{-1}\mathbf{b}).$$

Реализация метода последовательных исключений приводит к следующей последовательности действий. Исходная матрица имеет вид

$$\left[\begin{array}{ccc|ccc|c} 1 & 2 & 3 & 1 & 0 & 0 & 3 \\ 2 & 3 & 2 & 0 & 1 & 0 & 4 \\ 3 & 3 & 4 & 0 & 0 & 1 & 5 \end{array}\right].$$

Итерация 1

$$\left[\begin{array}{ccc|ccc|c} 1 & 2 & 3 & 1 & 0 & 0 & 3 \\ 0 & -1 & -4 & -2 & 1 & 0 & -2 \\ 0 & -3 & -5 & -3 & 0 & 1 & -4 \end{array}\right].$$

Итерация 2

$$\left[\begin{array}{ccc|ccc|c} 1 & 0 & -5 & -3 & 2 & 0 & -1 \\ 0 & 1 & 4 & 2 & -1 & 0 & 2 \\ 0 & 0 & 7 & 3 & -3 & 1 & 2 \end{array}\right].$$

Итерация 3

$$\left[\begin{array}{ccc|ccc|c} 1 & 0 & 0 & -\frac{6}{7} & -\frac{1}{7} & \frac{5}{7} & \frac{3}{7} \\ 0 & 1 & 0 & \frac{2}{7} & \frac{5}{7} & -\frac{4}{7} & \frac{6}{7} \\ 0 & 0 & 1 & \frac{3}{7} & -\frac{3}{7} & \frac{1}{7} & \frac{2}{7} \end{array}\right].$$

Таким образом, получили решение системы $x_1 = 3/7$, $x_2 = 6/7$ и $x_3 = 2/7$. Обратная матрица $\mathbf{A}^{-1}$ приведена справа от единичной матрицы и совпадает с обратной матрицей, полученной методом присоединенной матрицы.

Метод блочных матриц. Пусть две невырожденные матрицы **A** и **B** представимы в следующем блочном виде, причем подматрица $\mathbf{A}_{11}$ невырожденная.

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ (p\times p) & (p\times q) \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \\ (q\times p) & (q\times q) \end{array}\right] \text{ и } \mathbf{B} = \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ (p\times p) & (p\times q) \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \\ (q\times p) & (q\times q) \end{array}\right].$$

Пусть $\mathbf{B}$ будет матрицей, обратной к матрице $\mathbf{A}$. Тогда из равенства $\mathbf{AB} = \mathbf{I}_n$ следует, что

$$\mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} = \mathbf{I}_p,$$
$$\mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} = \mathbf{0}.$$

Аналогично из равенства $\mathbf{BA} = \mathbf{I}_n$ получаем

$$\mathbf{B}_{21}\mathbf{A}_{11} + \mathbf{B}_{22}\mathbf{A}_{21} = \mathbf{0},$$
$$\mathbf{B}_{21}\mathbf{A}_{12} + \mathbf{B}_{22}\mathbf{A}_{22} = \mathbf{I}_q.$$

Так как подматрица $\mathbf{A}_{11}$ невырожденная, то $|\mathbf{A}_{11}| \neq 0$ и существует обратная матрица $\mathbf{A}_{11}^{-1}$. Тогда из приведенных уравнений получаем

$$\mathbf{B}_{11} = \mathbf{A}_{11}^{-1} + (\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{D}^{-1}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1}),$$
$$\mathbf{B}_{12} = -(\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{D}^{-1},$$
$$\mathbf{B}_{21} = -\mathbf{D}^{-1}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1}),$$
$$\mathbf{B}_{22} = \mathbf{D}^{-1},$$

где $\mathbf{D} = \mathbf{A}_{22} - \mathbf{A}_{21}(\mathbf{A}_{11}^{-1}\mathbf{A}_{12})$.

Для иллюстрации применения этих формул разобьем матрицу

$$\mathbf{A} = \left[\begin{array}{c|cc} 1 & 2 & 3 \\ \hline 2 & 3 & 2 \\ 3 & 3 & 4 \end{array}\right]$$

на блоки $\mathbf{A}_{11} = (1)$, $\mathbf{A}_{12} = (2, 3)$, $\mathbf{A}_{21} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ и $\mathbf{A}_{22} = \begin{bmatrix} 3 & 2 \\ 3 & 4 \end{bmatrix}$. Здесь $\mathbf{A}_{11}^{-1} = 1$ и

$$\mathbf{D} = \begin{bmatrix} 3 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix}(1)(2,\ 3) = \begin{bmatrix} -1 & -4 \\ -3 & -5 \end{bmatrix},$$

$$\mathbf{D}^{-1} = -\frac{1}{7}\begin{bmatrix} -5 & 4 \\ 3 & -1 \end{bmatrix} = \begin{bmatrix} \frac{5}{7} & -\frac{4}{7} \\ -\frac{3}{7} & \frac{1}{7} \end{bmatrix}.$$

Далее вычисляем

$$\mathbf{B}_{11} = \begin{bmatrix} -\frac{6}{7} \end{bmatrix},\ \mathbf{B}_{12} = \begin{bmatrix} -\frac{1}{7} & \frac{5}{7} \end{bmatrix},\ \mathbf{B}_{21} = \begin{bmatrix} \frac{2}{7} \\ \frac{3}{7} \end{bmatrix} \text{ и } \mathbf{B}_{22} = \begin{bmatrix} \frac{5}{7} & -\frac{4}{7} \\ -\frac{3}{7} & \frac{1}{7} \end{bmatrix}.$$

Теперь нетрудно получить матрицу $\mathbf{B} = \mathbf{A}^{-1}$.

А.3. Квадратичные формы

Пусть $\mathbf{X} = (x_1, x_2, ..., x_n)^T$ и

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

Тогда функция

$$Q(\mathbf{X}) = \mathbf{X}^T \mathbf{A} \mathbf{X} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$

называется *квадратичной формой.* Всегда можно считать, что матрица $\mathbf{A}$ симметрическая. В самом деле, значение квадратичной формы не изменится, если каждый коэффициент из пары a_{ij} и a_{ji} $(i \neq j)$ заменить на $(a_{ij} + a_{ji})/2$. В дальнейшем свойство симметричности матрицы $\mathbf{A}$ будет предполагаться.

Для примера приведем квадратичную форму

$$Q(\mathbf{X}) = (x_1, x_2, x_3) \begin{bmatrix} 1 & 0 & 1 \\ 2 & 7 & 6 \\ 3 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

которая совпадает с формой

$$Q(\mathbf{X}) = (x_1, x_2, x_3) \begin{bmatrix} 1 & 1 & 2 \\ 1 & 7 & 3 \\ 2 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Отметим, что во второй квадратичной форме матрица симметрическая.

Различают следующие типы квадратичных форм.

1. Квадратичная форма называется *положительно определенной,* если $Q(\mathbf{X}) > 0$ для всех $\mathbf{X} \neq \mathbf{0}$.
2. Квадратичная форма называется *положительно полуопределенной,* если $Q(\mathbf{X}) \geq 0$ для всех $\mathbf{X}$ и существует вектор $\mathbf{X} \neq 0$ такой, что $Q(\mathbf{X}) = 0$.
3. Квадратичная форма $Q(\mathbf{X})$ называется *отрицательно определенной,* если квадратичная форма $-Q(\mathbf{X})$ является положительно определенной.
4. Квадратичная форма $Q(\mathbf{X})$ называется *отрицательно полуопределенной,* если квадратичная форма $-Q(\mathbf{X})$ является положительно полуопределенной.
5. Во всех остальных случаях квадратичная форма называется *неопределенной.*

Можно доказать, что необходимыми и достаточными условиями того, что квадратичная форма будет принадлежать одному из перечисленных выше типов, являются следующие утверждения.

1. Квадратичная форма $Q(\mathbf{X})$ будет положительно определенной (полуопределенной), если значения всех угловых миноров определителя $|\mathbf{A}|$ положительны (неотрицательны).[1] В этом случае матрица **A** называется положительно определенной (полуопределенной).[2]

2. Квадратичная форма $Q(\mathbf{X})$ является отрицательно определенной, если значения k-х угловых миноров определителя $|\mathbf{A}|$ отличны от нуля и имеют знак $(-1)^k$, $k = 1, 2, ..., n$. В этом случае матрица **A** называется отрицательно определенной.

3. Квадратичная форма $Q(\mathbf{X})$ является отрицательно полуопределенной, если значения k-х угловых миноров определителя $|\mathbf{A}|$ равны нулю либо имеют знак $(-1)^k$, $k = 1, 2, ..., n$.

А.4. Выпуклые и вогнутые функции

Функция $f(\mathbf{X})$ называется строго выпуклой, если для произвольных двух различных точек $\mathbf{X}_1$ и $\mathbf{X}_2$ выполняется неравенство

$$f(\lambda\mathbf{X}_1 + (1-\lambda)\mathbf{X}_2) < \lambda f(\mathbf{X}_1) + (1-\lambda)f(\mathbf{X}_2),$$

где $0 < \lambda < 1$. Обратно, функция $f(\mathbf{X})$ называется строго вогнутой, если функция $-f(\mathbf{X})$ — строго выпукла.

Специальным случаем выпуклой (вогнутой) функции является квадратичная форма[3]

$$f(\mathbf{X}) = \mathbf{CX} + \mathbf{X}^T\mathbf{AX},$$

где **C** — вектор констант, а **A** — симметрическая матрица. Можно показать, что функция $f(\mathbf{X})$ будет строго выпуклой, если матрица **A** положительно определенная, и строго вогнутой, когда **A** — отрицательно определенная матрица.

Литература

1. Hadley G. *Matrix Algebra*, Addison-Wesley, Reading, Mass., 1961.
2. Hohn F. *Elementary Matrix Algebra*, 2nd ed., Macmillan, New York, 1964.

---

[1] *k-м угловым минором* определителя матрицы $\mathbf{A}_{n\times n}$ называется определитель вида

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

[2] В этой формулировке допущена неточность относительно положительно полуопределенных форм. квадратичная форма будет положительно полуопределенной тогда и только тогда, когда все *главные миноры* определителя матрицы **A** будут неотрицательны. Главными минорами называются определители, полученные путем вычеркивания из матрицы **A** строк и столбцов с одинаковыми номерами, т.е. главные миноры симметричны относительно главной диагонали. — *Прим. ред.*

[3] Строго говоря, здесь приведена не квадратичная форма, а сумма квадратичной и линейной форм — *Прим. ред.*

Литература, добавленная при переводе

Ефимов Н.В. *Квадратичные формы и матрицы.* — М.: Наука, 1967.
Хорн Р., Джонсон Ч. *Матричный анализ.* — М.: Мир, 1989.

Задачи

■ **А–1.** Покажите, что следующие векторы являются линейно зависимыми.

a) $\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}, \begin{bmatrix} -2 \\ 4 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}.$

b) $\begin{bmatrix} 2 \\ -3 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ -6 \\ 8 \\ 10 \end{bmatrix}.$

■ **А–2.** Для данных матриц

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 9 \\ 2 & 5 & -8 \\ 3 & 7 & 2 \end{bmatrix} \text{ и } \mathbf{B} = \begin{bmatrix} 7 & -1 & 2 \\ 9 & 4 & 8 \\ 3 & 6 & 10 \end{bmatrix}$$

найдите

a) $\mathbf{A} + 7\mathbf{B}$;
b) $2\mathbf{A} - 3\mathbf{B}$;
c) $(\mathbf{A} + 7\mathbf{B})^T$.

■ **А–3.** Для матриц из задачи А–2 покажите, что $\mathbf{AB} \neq \mathbf{BA}$.

■ **А–4.** Даны блочные матрицы

$$\mathbf{A} = \left[\begin{array}{c|cc} 1 & 5 & 7 \\ 2 & -6 & 9 \\ \hline 3 & 7 & 2 \\ 4 & 9 & 1 \end{array}\right] \text{ и } \mathbf{B} = \left[\begin{array}{cccc} 2 & 3 & -4 & 5 \\ \hline 1 & 2 & 6 & 7 \\ 3 & 1 & 0 & 9 \end{array}\right].$$

Найдите произведение **AB**, используя блочную структуру матриц.

■ **А–5.** Для матриц из задачи А–2 найдите $\mathbf{A}^{-1}$ и $\mathbf{B}^{-1}$

a) методом присоединенной матрицы,
b) методом последовательных исключений,
c) методом блочных матриц.

■ **А–6.** Проверьте правильность формул вычисления обратных матриц с блочной структурой, приведенных в разделе А.2.7.

■ **А–7.** Найдите матрицу, обратную к матрице

$$\mathbf{A} = \begin{bmatrix} 1 & \mathbf{G} \\ \mathbf{H} & \mathbf{B} \end{bmatrix},$$

где **B** — невырожденная матрица.

■ **А–8.** Покажите, что следующая квадратичная форма является отрицательно определенной.

$$Q(x_1; x_2) = 6x_1 + 3x_2 - 4x_1x_2 - 2x_1^2 - 3x_2^2 - \frac{27}{4}.$$

■ **А–9.** Покажите, что следующая квадратичная форма является положительно определенной.

$$Q(x_1, x_2, x_3) = 2x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 2x_2x_3.$$

■ **А–10.** Покажите, что функция $f(x) = e^x$ строго выпукла на всей действительной оси.

■ **А–11.** Покажите, что квадратичная форма

$$f(x_1, x_2, x_3) = 5x_1^2 + 5x_2^2 + 4x_3^2 + 4x_1x_2 + 2x_2x_3.$$

является строго выпуклой.

■ **А–12.** В условиях задачи А–11 покажите, что функция $-f(x_1, x_2, x_3)$ строго вогнута.

Введение в SIMNET II[1]

Б.1. Сетевые модели

Язык программирования SIMNET II предназначен для имитации дискретных моделей, представимых в виде системы очередей. Язык основан на **сетевом подходе**, который использует три типа узлов: узел **источника** (source), из которого исходят транзакции[2], узел **очереди** (queue), где транзакции ожидают обслуживания, и узел **средств обслуживания** (facility). Существует еще четвертый **дополнительный** тип узла (auxiliary), который предназначен для расширения возможностей языка.

Узлы в SIMNET II соединены **дугами** (branches). При перемещении транзакций по дугам, они (дуги) выполняют важные функции, в частности, (1) управляют потоком транзакций в сети, (2) собирают статистические данные и (3) выполняют арифметические вычисления.

В процессе имитационного моделирования SIMNET II отслеживает последовательную обработку транзакций, помещая информацию о них в специальные **файлы**. Такие файлы имеют структуру двухмерного массива, в котором каждая строка хранит информацию об одной транзакции. Столбцы массива представляют **атрибуты**, которые являются уникальными характеристиками каждой транзакции.

В SIMNET II используются файлы трех типов.

1. Календарь событий (event calendar).
2. Очередь (queue).
3. Средства обслуживания (facility).

Календарь событий (E.FILE) — основной файл, определяющий имитацию. Он автоматически обновляется и хранит информацию о всех событиях, происшедших в модели, в хронологическом порядке. Обработка файлов очередей и средств обслуживания также поддерживается автоматически средствами языка SIMNET II.

Б.2. Операторы SIMNET II

Общий формат операторов языка SIMNET II следующий

```
идентификатор; поле 1; поле 2; ...; поле m:
```

Идентификатор *узла* состоит из имени, определяемого пользователем (не более 12 символов), за которым следует один из кодов *S, *Q, *F или *A, которые указывают, что

---

[1] В этом приложении представлены только базовые средства языка SIMNET II. Ограниченный объем книги не позволяет описать все средства этого языка. Полная документация по SIMNET II приведена в книге Taha H. *Simulation with SIMNET II*, 2nd ed., SimTec Inc., Fayetteville, AR, 1995.

[2] Транзакция — обобщенный термин, обозначающий "то, что обрабатывается" в данной модели. Например, это могут быть запросы на обслуживание или клиенты в моделях массового обслуживания и т.п. — *Прим. ред.*

имя принадлежит соответственно узлу источника, узлу очереди, узлу средств обслуживания либо дополнительному узлу. Идентификатор дуги состоит только из кода *B. Поля отделяются точкой с запятой, а оператор обязательно должен заканчиваться двоеточием.

Для примера рассмотрим следующие операторы.

```
ARIV  *S;EX(5);;;;LIM=100:
      *B;QQ;;A(1)=1%:
  QQ  *Q:
```

Здесь источник ARIV создает 100 транзакций, время между созданиями транзакций подчиняется экспоненциальному распределению с математическим ожиданием 5 временных единиц. Дуга соединяет источник ARIV с очередью QQ. При переходе транзакции по этой дуге перед поступлением в очередь QQ первый атрибут A(1) транзакции устанавливается равным 1.

Все поля в операторе *позиционные* в том смысле, что они определяются своей позицией в операторе. Если поля не используются или имеют значения по умолчанию, то они замещаются последовательными точками с запятыми.

В языке SIMNET II код имеет свободное форматирование, и он не чувствителен к регистру символов. Оператор можно сегментировать, перенося части оператора на отдельные строки, при этом части операторов на каждой строке (кроме последней) должны заканчиваться знаком амперсанда (&). Например, оператор, задающий источник ARIV, можно записать следующим образом.

```
ARIV    *S;EX(5);;;;LI&      !Строка 1 оператора ARIV
                 M=100:      !Строка 2 оператора ARIV
```

В код языка SIMNET II можно включать **комментарии**, префиксом для которых служит восклицательный знак (!).

Б.2.1. Узел источника

Узел источника используется для создания в моделируемой сети последовательных транзакций. Определения полей этого узла приведены в следующей таблице; графическое представление узла источника показано на рис. Б.1.

ИМЯ_ИСТОЧНИКА *S; F1; F2; F3; MULT = F4; LIM = F5; F6; F7; *T:			
	Идентификатор поля		Значение по умолчанию
F1		Время между появлениями транзакций (выражение)[a]	0
F2		Время создания первой транзакции	0
F3		Указывает номер атрибута, при этом атрибут принимает значение времени создания транзакции, если F3 > 0, или значение порядкового номера (т.е. последовательных целых чисел), если перед номером атрибута стоит знак "минус"	Нет
F4	/m/	Указывает на возможность одновременного создания нескольких транзакций (константа или переменная)[b,c]	1

ИМЯ_ИСТОЧНИКА *S; F1; F2; F3; MULT = F4; LIM = F5; F6; F7; *T:			
	Идентификатор поля		Значение по умолчанию
F5	/L/	Указывает верхнюю границу количества созданий транзакций (F5 > 0) или верхнюю границу времени генерации транзакций (F5 < 0) (константа или переменная)[c]	∞
F6	/s/	Определяет правило выбора выходной дуги (см. раздел Б.5.1)	Нет
F7	/r/	Определяет ресурсы, возвращаемые источником	Нет
*T		Список узлов, которые можно достичь непосредственно из источника (см. раздел Б.5.2)[d]	Нет

[a] Математическое выражение языка SIMNET II (см. раздел Б.3).

[b] Для полей F4 ... F7 номер поля n в идентификаторе поля /n/ можно заменить "описательными" словами /Multiple/ (Многократно), /Limit/ (Предел), /Select/ (Выбор) и /Resources/ (Ресурсы) или, соответственно, /M/, /L/, /S/, /R/.

[c] Переменная может быть без индекса или элементом массива (см. раздел Б.3).

[d] Здесь звездочку можно заметить словом GOTO- (Перейти).

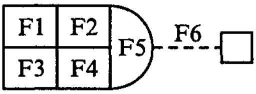


Рис. Б.1

Примеры

1. Количество автомобилей, прибывающих на пункт регистрации, можно смоделировать следующим образом.

```
CUSTMS  *S;EX(12);10;-1:
```

Здесь время создания транзакций (прибытие автомобилей на пункт регистрации) подчиняется экспоненциальному распределению с математическим ожиданием 12; первая транзакция будет сгенерирована в момент времени 10. Значение третьего поля (`;-1:`) показывает, что при генерации каждой транзакции атрибуту A(1) будут присваиваться различные *последовательные номера*. Если значение этого поля изменить на (`;1:`), тогда атрибуту A(1) будут присваиваться значения *времени создания* транзакций.

2. Со сборочного конвейера каждые 5 минут готовый телевизионный приемник поступает на упаковку. Эта ситуация моделируется следующим источником.

```
TVS  *S;5;goto-PAKAGE:
```

Здесь "goto-" — это поле *T из описания узла источника, PAKAGE — имя узла, имитирующего упаковку телевизоров; в этот узел переходят транзакции. Отметим, что поле *T является "плавающим" в том смысле, что оно всегда будет последним полем в описании источника независимо от количества полей, предшествующих ему (или пропущенных в описании, если поля принимают значения по умолчанию).

3. По контракту на лесопильный завод прибудет 100 лесовозов, каждый из которых вмещает по 50 бревен. Пилорама завода одновременно может обрабатывать только одно бревно. Поступление лесоматериалов моделируется следующим оператором.

```
TRKS  *S;45;/m/MULT=50;LIM=100;goto-MILL:
```

Поле LIM указывает, что источником TRKS будет сгенерировано только 100 транзакций (лесовозов). В свою очередь, поле MULT показывает, что при генерировании каждая транзакция будет повторена 50 раз. Применение предшествующего этому полю идентификатора /m/ очень удобно, так как это позволяет не ставить точки с запятыми для тех полей, значения которых установлены по умолчанию. Значение последнего поля "goto-MILL" указывает, что сгенерированные транзакции перейдут в узел MILL (Пилорама).

Б.2.2. Узел очереди

Поля узла очереди описаны в следующей таблице и показаны на рис. Б.2.

ИМЯ_ОЧЕРЕДИ *Q; F1(SUBF1); F2(SUBF2); F3; F4; F5; *T:			
	Идентификатор поля		Значение по умолчанию
F1		Максимальный размер очереди (константа или переменная)[a]	∞
SUBF1		Начальное количество транзакций в очереди (константа или переменная)	0
F2		Количество ожидающих транзакций, необходимое для создания *одной* покидающей очередь транзакции (константа или переменная)	1
SUBF2		Если F2 > 1, это подполе указывает правило вычисления атрибутов транзакций, покидающих очередь, — SUM, PROD, FIRST, LAST, HI(#), LO(#), где # — номер атрибута[1]	LAST
F3	/d/	Принцип организации (дисциплина) очереди: FIFO, LIFO, RAN, HI(#), LO(#), где # — номер атрибута[2]	FIFO
F4	/s/	Определяет правило выбора выходной дуги (см. раздел Б.5.1)[b]	Нет
F5	/r/	Определяет ресурсы, возвращаемые очередью	Нет
*T		Список узлов, которые можно достичь непосредственно из очереди (см. раздел Б.5.2)[c]	Нет

[1] Так как эти правила вычисления атрибутов в дальнейшем не объясняются, кратко поясним их: SUM — суммирование, PROD — произведение, FIRST — первый (т.е. определение транзакции, стоящей первой в очереди), LAST — последний (определение последней в очереди транзакции), HI(#) — наивысший (определение транзакции, имеющей наибольшее значение атрибута с номером #), LO(#) — наинизший (определение транзакции, имеющей наименьшее значение атрибута с номером #). — *Прим. ред.*

[2] Объясним аббревиатуры, применяемые здесь для обозначения принципов организации очередей. FIFO (от First-In-First-Out) — принцип организации очереди, когда первым из очереди выходит тот элемент, который первый в нее вошел. LIFO (от Last-In-First-Out) — из очереди первым выходит тот элемент, который в нее вошел последним (очереди с таким принципом организации чаще называют *стеками*). HI(#) — первым из очереди выходит элемент, имеющий наибольшее значение атрибута с номером #. LO(#) — первым из очереди выходит элемент, имеющий наименьшее значение атрибута с номером #. — *Прим. ред.*

[a] Переменная может быть без индекса или элементом массива (см. раздел Б.3).

[b] Для полей F3, F4 и F5 номер поля n в идентификаторе поля /n/ можно заменить "описательными" словами /Discipline/ (Дисциплина очереди), /Select/ (Выбор) и /Resources/ (Ресурсы) или, соответственно, /D/, /S/, /R/.

[c] Здесь звездочку можно заметить словом GOTO- (Перейти).

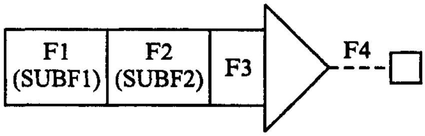


Рис. Б.2

Примеры

1. В магазине розничной торговли срочная и обычная работа, как правило, появляется случайным образом. Очередь работ, где приоритет отдается срочной работе, определяется следующим образом.

   ```
   JOBQ    *Q;;;HI(1):
   ```

   Атрибут 1 определяет тип работы, при этом значения A(1) = 0 или A(1) = 1 обозначают, соответственно, обычную и срочную работу. Работы в очереди JOBQ упорядочены по значениям атрибутов A(1) в соответствии с принципом организации очереди HI(1). В данном операторе можно опустить точки с запятыми, записав его следующим образом.

   ```
   JOBQ    *Q;/d/HI(1):
   ```

2. Некая продукция пакетируется по четыре единицы в одну упаковку. Участок упаковки не может хранить более 75 единиц продукции. Первоначально на участке упаковки находится 30 единиц продукции. Эта ситуация моделируется посредством следующей очереди.

   ```
   QUNIT    *Q;75(30);4:
   ```

   Первое поле устанавливает максимальную емкость очереди в 75 единиц, причем в начале процесса имитации в очереди уже находится 30 единиц. Второе поле показывает, что 4 единицы продукции пакетируются в одну упаковку (один выход из очереди). По умолчанию эту упаковку формируют последние 4 единицы (значение SUBF2 по умолчанию LAST). Очередь организована по принципу FIFO (значение по умолчанию третьего поля).

Б.2.3. Узел средств обслуживания

Данный узел состоит из одного или нескольких (идентичных) параллельных сервисов. Если все сервисы заняты, то новые транзакции не могут поступить на узел средств обслуживания. Когда сервисы освобождаются, узел *автоматически* загружает новую транзакцию из предшествующей очереди, если, конечно, такая транзакция существует. В противном случае средства обслуживания находятся в режиме ожидания, пока не поступит новая транзакция. Описания полей узла средств обслуживания даны в следующей таблице, а его графическое представление показано на рис. Б.3.

ИМЯ_СРЕДСТВ_ОБСЛ *F; F1; F2; F3(SUBF3); F4; F5; *T:			
	Идентификатор поля		Значение по умолчанию
F1		Правило выбора входной очереди (см. раздел Б.5.1)	Нет
F2		Время обслуживания (выражение)[a]	0
F3		Количество параллельных сервисов (константа или переменная)[b]	1
SUBF3		Количество первоначально занятых сервисов (константа или переменная)	0
F4	/s/	Определяет правило выбора выходных дуг (см. раздел Б.5.1)[c]	Нет
F5	/r/	Определяет ресурсы, получаемые/освобождаемые средством обслуживания	Нет
*T		Список узлов, которые можно достичь непосредственно из узла средств обслуживания (см. раздел Б.5.2)[d]	Нет

[a] Математическое выражение языка SIMNET II (см. раздел Б.3).

[b] Переменная может быть без индекса или элементом массива (см. раздел Б.3).

[c] Для полей F4 и F5 номер поля n в идентификаторе поля /n/ можно заменить "описательными" словами /Select/ (Выбор) и /Resources/ (Ресурсы) или, соответственно, /S/ и /R/.

[d] Здесь звездочку можно заметить словом GOTO- (Перейти).

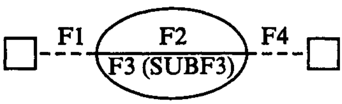


Рис. Б.3

Примеры

1. Средство обслуживания с тремя сервисами начинает работу, когда два сервиса заняты. Время обслуживания имеет экспоненциальное распределение со средним 15 минут. Элементы, обслуживание которых закончено, удаляются из системы (пункт назначения TERM). Описанная ситуация моделируется следующим оператором.

```
SRVR    *F;;EX(15);3(2);goto-TERM:
```

Здесь первое поле не использовано, поскольку значения ему присваиваются тогда, когда есть несколько входных очередей. Распределение времени обслуживания EX(15) задано во втором поле (здесь можно использовать любое математическое выражение). Третье поле показывает, что имеется 3 сервиса, два из которых заняты. Поле "goto" завершает существование транзакции, используя для этого зарезервированное (служебное) слово TERM (от TERMinate — завершить) в качестве названия пункта назначения транзакции.

2. В небольшой мастерской имеется один станок и 10 деталей, которые требуют обработки на этом станке. Кроме того, есть еще одна деталь, которая уже находится в процессе обработки на этом станке. Время обработки всех деталей имеет равномерное распределение на интервале от 20 до 30 минут. Эта ситуация описывается следующими операторами.

```
QJOB   *Q;(10):
FJOB   *F;;UN(20,30);(1);goto-TERM:
```

Поскольку средство обслуживания (станок) FJOB в начальный момент занято, на что указывает значение (1) в поле 3, процесс моделирования начинается с получения случайного значения, равномерно распределенного на интервале (20, 30) (значение UN(20,30) в поле 2), для определения времени обработки детали, которая уже находится на станке. После того как обработка детали завершена, она удаляется из системы (значение `goto-TERM` в поле *T). Затем узел FJOB *автоматически* "загружает" новую работу из очереди QJOB. Этот процесс повторяется 10 раз, пока не будут обработаны все детали.

Б.2.4. Дополнительный узел

Дополнительный узел имеет бесконечную емкость, что позволяет вместить все генерируемые в системе транзакции. Этот узел обычно используется для организации задержек транзакций в процессе прохождения их по системе. Поскольку дополнительный узел — это единственный тип узла, который может свой выход направить на свой же вход, он очень полезен при организации в имитационной модели петель (циклов). Поля дополнительного узла описаны в следующей таблице и показаны на рис. Б.4.

ИМЯ_ДОПОЛН_УЗЛА *A; F1; F2; F3; *T:			
	Идентификатор поля		Значение по умолчанию
F1		Время задержки (выражение)[a]	0
F2	/s/	Определяет правило выбора выходной дуги (см. раздел Б.5.1)[b]	Нет
F3	/r/	Определяет ресурсы, возвращаемые дополнительным узлом	Нет
*T		Список узлов, которые можно достичь непосредственно из дополнительного узла (см. раздел Б.5.2)[c]	Нет

[a] Математическое выражение языка SIMNET II (см. раздел Б.3).

[b] Для полей F2 и F3 номер поля n в идентификаторе поля /n/ можно заменить "описательными" словами /Select/ (Выбор) и /Resources/ (Ресурсы) или, соответственно, /S/ и /R/.

[c] Здесь звездочку можно заметить словом GOTO- (Перейти).

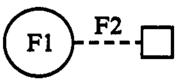


Рис. Б.4

Пример

1. В бюро по трудоустройству желающие получить работу появляются через промежутки времени, имеющие экспоненциальное распределение с математическим ожиданием 25 минут. Каждый прибывший должен заполнить специальную форму, затем подождать в очереди вызова на собеседование. Время заполнения формы приблизительно составляет 15 минут. Описанная ситуация моделируется следующим образом.

```
ARIV   *S;EX(25):
FORM   *A;15:
WAIT   *Q:
```

Заполнению формы здесь соответствует дополнительный узел FORM с бесконечной емкостью, поскольку формы для заполнения доступны любому сразу же по его приходу в бюро трудоустройство.

Б.2.5. Основные правила работы с узлами

1. Узел источника может не иметь входа от других узлов.
2. Узел очереди может не иметь выхода непосредственно на другие очереди, также запрещается замыкать выход очереди на ее вход.
3. Узлы средств обслуживания можно соединять последовательно без применения промежуточных очередей. Запрещается выход узла средств обслуживания замыкать на свой вход.
4. Только в дополнительном узле можно выход узла замыкать на свой вход.
5. Транзакция поступает в очередь, если очередь не заполнена (т.е. еще не достигнут предел емкости очереди).
6. Если узел средств обслуживания следует за узлом очереди, он автоматически принимает ожидающую транзакцию из очереди, если есть свободные сервисы. Если предшествующая очередь пуста, узел средств обслуживания переходит в режим ожидания до тех пор, пока не появится новая транзакция.
7. Перемещение транзакций в очередь и из нее должно инициализироваться *другим* узлом. Сам узел очереди не может инициализировать перемещение транзакций.
8. Если несколько узлов средств обслуживания следуют друг за другом либо промежуточные узлы очередей имеют ограниченную емкость, возможна ситуация, когда транзакция, закончившая обслуживание на одном сервисе, оказывается **заблокированной**, если следующим узлом будет заполненная очередь (с ограниченной емкостью) или узел средств обслуживания с занятыми сервисами. **Разблокировка** транзакции происходит автоматически при появлении возможности перемещения транзакции в следующий узел.

Упражнения Б.2,а

1. Определите значения атрибута A(1), соответствующие первым трем транзакциям для следующих узлов источников.

a) ARIV *S;5;;1:

b) ARIV *S;5;3;1:

c) ARIV *S;5;3;-1:

2. Сколько транзакций будет сгенерировано следующими источниками в течение первых 20 единиц времени имитации?

a) ARIV *S;5;/L/LIM=3:

b) ARIV *S;5;/m/MULT=2:

3. Первые пять транзакций, поступивших в очередь QQ, имеют следующие атрибуты.

Транзакция	A(1)	A(2)
1	4	9
2	7	–3
3	1	10
4	3	14
5	2	6

Покажите, как эти транзакции будут упорядочены в очереди QQ, описываемой следующими операторами.

a) QQ *Q:

b) QQ *Q;/d/LIFO:

c) QQ *Q;/d/HI(1):

d) QQ *Q;/d/LO(2):

4. В условиях предыдущего упражнения определите транзакции, покидающие очередь QQ, и их атрибуты, предполагая, что очередь QQ организована по принципу FIFO и задана следующими операторами.

a) QQ *Q;;2(SUM):

b) QQ *Q;;4(FIRST):

c) QQ *Q;;3:

d) QQ *Q;;2(LO(1)):

e) QQ *Q;;2(HI(2)):

5. Рассмотрите представленную часть кода имитационной модели и ответьте на следующие вопросы.

```
QQ    *Q;(3):
FF    *F;;2;(1);goto-TERM:
```

a) Сколько транзакций пройдет через средство обслуживания FF?

b) Определите моменты времени, когда каждая транзакция покинет узел FF.

c) Повторите пп. a) и b) в предположении, что представленная выше часть кода имитационной модели изменена следующим образом.

Изменение 1

```
QQ   *Q:
FF   *F;;2;(1);goto-TERM:
```

Изменение 2

```
QQ   *Q;(3):
FF   *F;;2;goto-TERM:
```

Изменение 3

```
QQ   *Q;(3):
FF   *F;;2;2(1);goto-TERM:
```

6. Дан следующий оператор, определяющий узел источника.

   ```
   SS  *S;UN(10,15):
   ```

   Предположим, что в процессе реализации имитационной модели получена следующая последовательность случайных чисел, равномерно распределенных на интервале [0, 1],

   ```
   0.1111, 0.2342, 0.6712, 0.8923, 0.4687, 0.3526, ...
   ```

   Определите моменты времени создания источником SS первых трех транзакций.

Б.3. Математические выражения в SIMNET II

В SIMNET II правила построения и выполнения математических выражений точно такие же, как в языке программирования FORTRAN. Множество арифметических операций включает следующие операторы: сложения (+), вычитания (–), умножения (*), деления (/) и возведения в степень (**). Математическое выражение может включать любые допустимые комбинации следующих элементов.

1. Определенные пользователем переменные: неиндексированные и индексированные, т.е. элементы массивов (допускаются одно- и двухмерные массивы).
2. Любые допустимые алгебраические и тригонометрические функции (табл. Б.1).
3. Переменные имитации SIMNET II, которые определяют состояние имитационного процесса при его реализации (табл. Б.2).
4. Генерируемые случайные значения, имеющие различные вероятностные распределения (табл. Б.3).

В табл. Б.2 и Б.3 представлена только часть переменных имитации и вероятностных распределений, которые можно моделировать в SIMNET II. Полные их списки даны в книге [1].

ТАБЛИЦА Б.1

Встроенные функции SIMNET II

Алгебраические	
С одним аргументом	INT, ABS, EXP, SQRT, SIGN, LOG, LOG10
С двумя аргументами	MOD

Окончание табл. Б.1

С произвольным числом аргументов	MAX, MIN
Тригонометрические (с одним аргументом)	
Обычные	SIN, COS, TAN
Обратные	ASIN, ACOS, ATAN
Гиперболические	SINH, COSH, TANH

* Аргументом функций может быть любое математическое выражение SIMNET II.

Таблица Б.2

Переменные имитации SIMNET II

Переменная	Описание
LEN (имя дополнительного узла)	Текущее количество транзакций, находящихся в дополнительном узле
LEN/HLEN/LLEN/ALEN (имя узла)	Текущая/Наибольшая/Наименьшая/Средняя длина (LENght) очереди или средства обслуживания
VAL/HVAL/LVAL/AVAL (имя переменной)	Текущее/Наибольшее/Наименьшее/Среднее значение (VALue) статистической переменной (см. раздел Б.7)
COUNT (узел или имя переменной)	Количество транзакций, прошедших через узел с начала имитации, либо общее число сделанных изменений статистической переменной
RUN.LEN	Длина текущей реализации
CUR.TIME	Текущее время имитации

Таблица Б.3

Моделируемые SIMNET II функции вероятностных распределений

Функция*	Описание
EX(p_1, RS)	Экспоненциальное распределение со средним p_1
NO(p_1, p_2, RS)	Нормальное распределение с математическим ожиданием p_1 и среднеквадратическим отклонением p_2
PO(p_1, RS)	Распределение Пуассона со средним p_1
RND(RS)	Равномерное распределение на интервале [0, 1]
TR(p_1, p_2, p_3, RS)	Треугольное распределение на интервале [p_1, p_3] с модой в p_2
UN(p_1, p_2)	Равномерное распределение на интервале [p_1, p_2]

* Аргументами p_1, p_2, p_3 и RS могут быть любые математические выражения SIMNET II. Аргумент RS[1] должен быть целым числом (отличным от нуля) из интервала [–50, +50], он определяет одну из ста последовательностей случайных чисел, генерируемых SIMNET II. Если параметр RS отрицательный, то при генерировании случайных чисел с любым распределением используются числа вида 1 – R, где R — значение равномерно распределенной на [0, 1] случайной величины. По умолчанию RS = 1.

[1] Имя RS этого параметра является сокращением от *Random Stream* — случайный поток. — *Прим. ред.*

Б.4. Пример модели, созданной в SIMNET II

Клиенты приходят в почтовое отделение в случайные моменты времени, где их обслуживают три клерка. Длина интервала времени между приходами клиентов имеет экспоненциальное распределение с математическим ожиданием 5 минут. Время обслуживания одного клиента также подчиняется экспоненциальному закону распределения со средним 10 минут. Все пришедшие клиенты становятся в одну очередь и обслуживаются свободным клерком по принципу FIFO (первый пришел — первый обслужен).

На рис. Б.5 показана сетевая модель, соответствующая описанной ситуации. Модель описывается следующими операторами SIMNET II.

```
$PROJECT;Post Office;2 April 1990;Nancy Sloan:
$DIMENSION;ENTITY(30):
$BEGIN:
    ARVL    *S;EX(5):                    !Приход клиентов
    LINE    *Q:                          !Очередь клиентов
    CLRKS   *F;;EX(10);3;goto-TERM:      !Обслуживание клиентов
$END:
$RUN-LENGTH=480:                         !Время имитации 480 минут
$STOP:
```

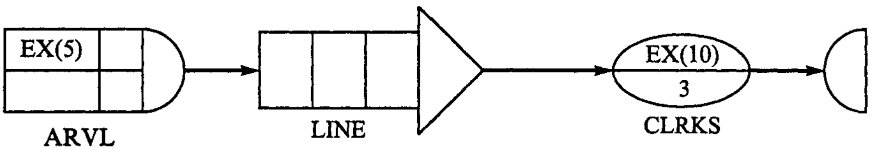


Рис. Б.5

В операторе $PROJECT указано название модели, дата и имя разработчика модели. Оператор $DIMENSION использует опцию ENTITY(30) для *оценки* памяти, требуемой для файлов модели. Данный оператор указывает, что в системе в любой момент времени реализации модели ожидается не более 30 транзакций. Если в процессе имитации потребуется больший объем памяти, на экране появится соответствующее сообщение.

Операторы, описывающие непосредственно модель, заключены в операторные скобки $BEGIN и $END. Транзакции создаются автоматически источником ARVL через определенные промежутки времени в соответствии с законом распределения EX(5). Первая транзакция будет создана в момент времени 0 (значение по умолчанию поля 2). Созданные транзакции поступают в очередь LINE, если все три клерка заняты. В противном случае транзакция "перескакивает" очередь. После того как транзакция получит обслуживание средством CLRKS (Клерки), она удаляется из системы. Средство обслуживания CLRKS "смотрит назад" на очередь LINE и принимает транзакцию, стоящую первой в очереди (очередь по умолчанию организована по принципу FIFO).

Маршрут транзакции определяется автоматически последовательностью операторов, заключенных в операторные скобки $BEGIN и $END. Как будет показано ниже, транзакции могут проходить по любому маршруту, организованному с помощью полей *T, а также дугами сети.

Операторы, стоящие между $END и $STOP, управляют выполнением имитации. В частности, в данном случае оператор $RUN-LENGTH указывает, что имитация должна выполняться в течение 480 минут.[1]

Стандартный выходной листинг результатов имитации показан на рис. Б.6. Очередь LINE имеет бесконечную емкость, о чем свидетельствуют звездочки в качестве значения свойства CAPACITY (Емкость) в области QUEUES (Очереди) листинга. Значение 1 : 1 свойства IN:OUT RATIO (Отношение вход/выход) показывает, что каждая существующая транзакция соответствует *одной* ожидающей транзакции (значение по умолчанию поля 2 в описании очереди LINE). Значение 1.31 свойства AVERAGE LENGTH (Средняя длина) соответствует среднему количеству транзакций (средней длине очереди), ожидаемых в очереди, причем среднее считается по всему времени выполнения имитации. В столбце MIN/MAX/LAST LEN свойств очереди показаны значения минимальной, максимальной и последней длины очереди LINE (здесь они равны, соответственно, 0, 12, 0). Среднее время ожидания *всех* транзакций (включая те, которые не стояли в очереди) дано в столбце AV. DELAY (ALL) (Средняя задержка (все)) и в данном случае равно 6.51 минуты. Следующий столбец AV. DELAY (+VE WAIT) (Средняя задержка (+Значение ожидания)) показывает, что среднее время ожидания для тех транзакций, которые *стояли* в очереди, равно 15.63 минуты. Наконец, последний показатель % ZERO WAIT TRANSACTION (Процент транзакций с нулевым временем ожидания) дает процент транзакций, которые "перескочили" через очередь; в данном случае этот показатель равен 58.33%.

В области FACILITIES (Средства обслуживания) представлены свойства и показатели имитации узлов средств обслуживания модели. В данном случае средство обслуживания CLRKS имеет три параллельных сервиса (свойство NBR SRVRS — Количество сервисов). Второй столбец показывает, что средство CLRKS начало работу со свободными клерками, затем был период времени, когда были заняты все три клерка, и закончило работу (имитацию) с одним занятым клерком. Третий столбец AV. GROSS UTILIZ (Среднее общее использование) показывает, что в среднем было занято 1.9955 сервиса (из 3-х), в процентном отношении это составляет $1.9955/3 \times 100 = 66.5\%$. В столбце AVERAGE BLOCKAGE (Среднее количество блокированных сервисов) записан средний *непродуктивный* простой средства обслуживания (выраженный как количество простаиваемых сервисов), вызванный блокировкой последующих очередей или средств обслуживания. В нашем примере такая ситуация невозможна, поэтому данный показатель равен нулю. В общем случае *чистое использование* средства обслуживания считается как разность между средним общим использованием и средним непродуктивным простоем. Показатель AVERAGE BLKGE TIME (Среднее время блокировки) содержит среднее время, которое средство обслуживания провело в режиме блокировки (в данном примере оно равно нулю). В последних двух столбцах (AVERAGE IDLE TIME — Среднее время простоя и AVERAGE BUSY TIME — Среднее время полной занятости) показаны средние длительности периодов простоя (из-за отсутствия входных транзакций) и полной занятости средства обслуживания. В представленном примере каждый клерк оставался свободным примерно 8.44 минуты. Среднее время занятости никогда не может быть меньше среднего времени обслуживания одним сервисом. В нашем случае среднее время обслуживания по условию равно 10 минут, а среднее время полной занятости — 17.06 минут. Отсюда можно сделать вывод, что каждый сервис обслуживает примерно $17.06/10 = 1.706$ клиента, после чего простаивает 8.44 минуты.

---

[1] Здесь указывается "внутреннее" время модели, а не реальное время имитации, которое, очевидно, зависит от производительности компьютера, на котором выполняется моделирование. — *Прим. ред.*

```
PROJECT: Post Office      RUN LENGTH =    480.00        NBR RUNS =  1
DATE: 2 April 1990        TRANSIENT PERIOD =      .00   OBS/RUN =   1
ANALYST: Nancy Sloan      TIME BASE/OBS =   480.00

          ***INDEPENDENT RUNS DATA***

*** RUN 1:
                          ---------------------
                               QUEUES
                          ---------------------
        CAPA-  IN:OUT  AVERAGE  MIN/MAX/  AV. DELAY  AV. DELAY  % ZERO WAIT
        CITY   RATIO   LENGTH   LAST LEN  (ALL)      (+VE WAIT) TRANSACTION
LINE    ****    1: 1    1.31    0/ 12/ 0    6.51       15.63        58.33
                          ---------------------
                             FACILITIES
                          ---------------------
         NBR    MIN/MAX/   AV. GROSS AVERAGE  AVERAGE     AVERAGE   AVERAGE
         SRVRS LAST UTILZ  UTILIZ    BLOCKAGE BLKGE TIME IDLE TIME BUSY TIME
CLRKS     3     0/ 3/ 1    1.9955      .0000        .00     8.44      17.06

*** TRANSACTIONS COUNT AT T =   478.8 OF RUN  1:
NODE      IN    OUT   RESIDING   SKIPPING    UNLINKED/LINKED   TERMINATED
                                 (BLOCKED)   (DESTROYED)
*S:
ARVL             96                          (   0)                      0
*Q:
LINE      40     40      0          56          0/       0               0
*F:
CLRKS     96     95      1       (   0)      (   0)                     95
```

Рис. Б.6

Раздел TRANSACTION COUNT (Подсчет транзакций) в конце листинга отчета об имитации содержит полную хронику потока транзакций в процессе имитации. Эти данные могут быть полезны при отладке, настройке или модификации модели. Например, большое количество транзакций, стоящих в очереди, может указать на узкое место в системе. В данном примере за 480 минут (точнее — 478.8 минут) времени имитации источником ARVL было создано 96 транзакций, из которых 40 вынуждены были "стоять в очереди", остальные 56 получили обслуживание сразу после создания. На вход средства обслуживания CLRKS поступило 96 транзакций, из которых получили полное обслуживание 95, одна транзакция осталась в процессе обслуживания на момент окончания имитации. Данные столбца UNLINKED/LINKED (Не связано/Связано) используются, когда при имитации модели применяются средства задания начальных данных для очередей и средств обслуживания (см. раздел Б.9.3). Значения в столбце (BLOCKED) (Заблокировано) положительны только тогда, когда при имитации блокируются какие-либо средства обслуживания. Также значения в столбце (DESTROYED) (Уничтожены) отличны от нуля только в случае, когда транзакции уничтожаются (удаляются из системы). В нашем примере таких ситуаций не было.

Б.5. Маршрутизация транзакций

Из текущего узла транзакция может перейти к одному из следующих узлов сети.

1. К следующему узлу в заданной последовательности узлов.
2. К узлу, определенному полем выбора /s/.

3. К узлу, указанному в поле *T с помощью опции goto-.
4. К определенному узлу по дугам, исходящим из узла.

Переход транзакции к следующему узлу в заданной последовательности узлов рассмотрен в предыдущих разделах. В данном разделе описаны следующие два способа перехода из перечисленных выше. Переход по дугам представлен в разделе Б.6.

Б.5.1. Выбор маршрута

Каждый из четырех типов узлов, имеющихся в SIMNET II, имеет **поле выбора** (см. раздел Б.2), в котором с помощью заданного правила определяется один из нескольких возможных узлов назначения транзакции. Узел средств обслуживания имеет как входное поле выбора, так и выходное, т.е. такой узел может получить транзакцию от одной из нескольких входных очередей и послать ее на один из принимающих узлов. В табл. Б.4 приведен неполный список правил выбора SIMNET II; полный список приведен в книге [1].

ТАБЛИЦА Б.4

Правила выбора SIMNET II

Правило	Описание
Для всех узлов и пункта TERM	
формат: `правило(узел`$_1$`, узел`$_2$`, ..., узел`$_m$`)`	
POR	Задан порядок просмотра узлов назначения, из них первый свободный узел принимает транзакцию
ROT	Просмотр заданных узлов назначения осуществляется циклически, начиная от узла, следующего за узлом, который был использован для последней транзакции
Для очередей и средств обслуживания	
формат: `правило(файл`$_1$`, файл`$_2$`, ..., файл`$_m$`)`	
HTE (LTE)	Выбирается файл (узел), который наибольшее (наименьшее) время был пустым (незанятым)
Для очередей и средств обслуживания	
формат: `правило(файл`$_{11}\pm$`файл`$_{12}\pm$` ..., ..., файл`$_{m1}\pm$`файл`$_{m2}\pm$` ...)`	
HBC (LBC)	Выбираются файлы (узлы), которые имеют наибольшее (наименьшее) суммарное значение (в соответствии с заданными формулами правила) переменной имитации LEN (см. табл. Б.2)

Примеры

1. `SS     *S;UN(10,20);/s/POR(Q1,Q2,A1):`

   Перед тем как транзакция покинет источник SS, просматриваются узлы в заданном порядке Q1→Q2→A1 и транзакция посылается на *первый* узел, способный ее принять.

2. `QQ     *Q;/s/HTE(F1,F2):`

   Транзакция, покидающая очередь QQ, пересылается на один из узлов средств обслуживания F1 или F2, а именно на тот, который наибольшее время был незагруженным (пустым).

3. `AAX *A;/s/LBC(Q1+F1,Q2-Q1,Q3):`

Транзакция, покидающая дополнительный узел AAX, поступит на один из *ведущих узлов* Q1, Q2 или Q3, в зависимости от того, какое из следующих значений будет наименьшим: LEN(Q1) + LEN(F1), LEN(Q2) – LEN(Q1) или LEN(Q3), где LEN — переменная SIMNET II, определяющая длину файла-очереди (см. табл. Б.2).

4. `FF *F;HBC(Q1,Q2,Q3);EX(10);3;LBC(F3,Q4,Q5):`

В данном случае узел средств обслуживания FF использует *входное* и *выходное* поля выбора. Входное поле выбора (поле 2) "смотрит назад" на очереди Q1, Q2 и Q3 и получает транзакцию от той из них, которая имеет наибольшую длину (правило HBC). При выходе из узла FF транзакция передается на тот из узлов F3, Q4 или Q5, который имеет наименьшую длину.

Пример Б.5–1. (Модель банка)

Банк, обслуживающий клиентов прямо в автомобиле, имеет два пункта обслуживания. Время прибытия клиентов подчиняется экспоненциальному закону с математическим ожиданием 2 минуты. Прибывающий клиент выбирает полосу обслуживания (правую или левую), ведущую к одному из двух пунктов обслуживания, причем ту, где очередь меньше. Время обслуживания клиента на каждом пункте имеет равномерное распределение на интервале [3, 4]. Каждая полоса обслуживания может вместить не более трех автомобилей, включая тот, который обслуживается.

Ниже представлены операторы, описывающие эту ситуацию; соответствующая сетевая модель графически показана на рис. Б.7. Транзакции (клиенты) создаются источником CARS (Автомобили) в моменты времени, имеющие распределение EX(2). Поле выбора этого источника содержит правило выбора LBC(QL+WL,QR+WR), показывающее, что сгенерированная транзакция поступает в очередь QL (Левая полоса обслуживания), если количество автомобилей на левой полосе, включая обслуживаемый автомобиль (средство обслуживания WL), не превышает количества автомобилей на правой полосе (очередь QR и средство обслуживания WR). В противном случае выбирается правая полоса обслуживания. Если обе очереди QL и QR заполнены (т.е. когда LEN(QL) = 3 и LEN(QR) = 3), транзакция уничтожается системой.

```
$PROJECT;bank model;3 April 1990;Taha:
$DIMENSION;ENTITY(50):
$BEGIN:
  CARS *S;EX(2);/S/LBC(QL+WL,QR+WR):    !Выбор очереди
  QL   *Q;3:                            !Левая полоса
                                        !обслуживания
  WL   *F;;UN(3,4);GOTO-TERM:
  QR   *Q;3:                            !Правая полоса
                                        !обслуживания
  WR   *F;;UN(3,4);GOTO-TERM:
$END:
$RUN-LENGTH=480:        !Время имитации 480 минут
$RUNS=1:                !Реализация только одной имитации
$STOP
```

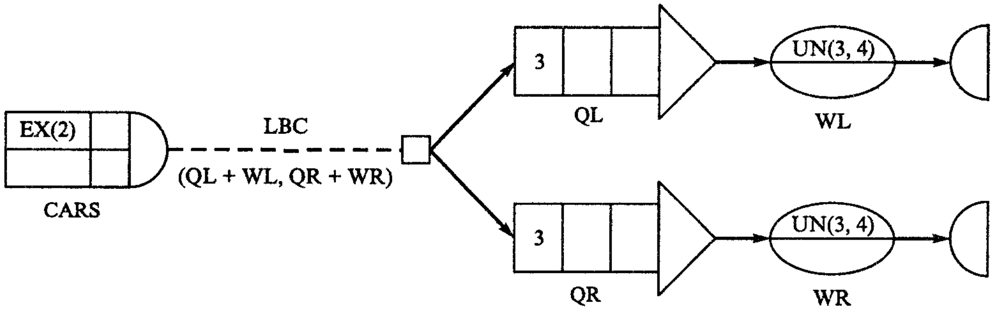


Рис. Б.7

Б.5.2. Маршрутизация с помощью поля *T

Поле *T предоставляет дополнительные возможности по управлению переходом транзакций по узлам сети. Это поле имеет следующий формат.

```
GOTO-имя_узла/тип_перехода[,имя_узла/тип_перехода,...]
```

Поле *T всегда последнее в операторе, задающем узел сети. В табл. Б.5 представлены некоторые наиболее общие типы переходов. Такие типы переходов, как исключающий и зависимый, здесь не представлены (более подробно о них см. в [1]).

Таблица Б.5

Типы переходов

Символ типа	Описание
A	Переход, выполняемый всегда (безусловный переход). Этот тип применяется по умолчанию
Число ($0 \le Ð \le 1$)	Переход с заданной вероятностью $Ð$
L	Осуществляется переход к последнему узлу в списке узлов

Пример Б.5–2. (Проверка телевизоров)

Телевизионные приемники поступают на проверку со сборочной линии со скоростью 5 единиц в час. Время проверки одного телевизора имеет равномерное распределение UN(10,15) (минуты). Наблюдения показывают, что примерно 20% проверенных приемников требуют незначительных настроек, после чего они опять поступают на проверку. Время, затрачиваемое на “доводку” телевизора, подчиняется равномерному распределению UN(6,8) (минуты).

Операторы SIMNET II, описывающие данную ситуацию, представлены ниже. Соответствующая сетевая модель показана на рис. Б.8. Маршрут транзакций TVS-QINSP-FINSP определяется последовательностью соединений узлов. Из узла FINSP (Проверка) осуществляется вероятностный переход: с вероятностью 20% транзакция переходит в очередь QADJ (для “доводки” телевизора) и с вероятностью 80% удаляется из системы (пункт TERM). После завершения работ по настройке в узле средств обслуживания FADJ транзакция (телевизор) снова переходит в очередь QINSP (значение поля *T узла FADJ по умолчанию имеет тип перехода A). Отметим, что сумма вероятностей переходов из любого узла всегда должна быть равной 1.

```
$PROJECT;TV inspection;4-3-1990;Sloan:
$DIMENSION;ENTITY(50):
$BEGIN:
   TVS   *S;12:         !Поступление телевизоров на проверку
   QINSP *Q:            !Очередь ожидания проверки
   FINSP *F;;UN(10,15);*QADJ/.20,TERM/.80:
                        !Проверка, после чего 20% телевизоров
                        !поступают в очередь QADJ,
                        !остальные 80% удаляются из системы
   QADJ  *Q:            !Очередь ожидания наладки
   FADJ  *F;;UN(6,8);GOTO-QINSP:
                        !Наладка и возврат в очередь QINSP
$END:
$RUN-LENGTH=480:
$STOP:
```

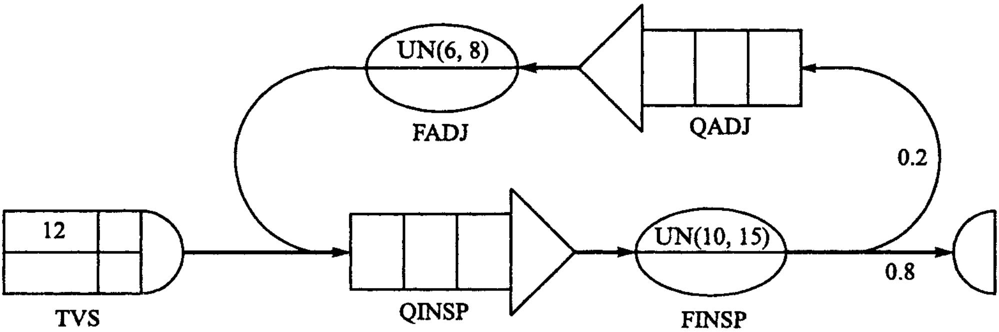


Рис. Б.8

Пример Б.5–3. (Измененная модель банка)

Изменим модель банка, построенную в примере Б.5.1. Пусть 90% клиентов, которые не могут воспользоваться обслуживанием непосредственно из автомобиля, заходят в банк для обслуживания у простого банковского кассира. Для парковки автомобиля и захода внутрь банка отводится три минуты. Время обслуживания у кассира распределено равномерно на интервале от 2 до 3 минут.

На рис. Б.9 показана новая сетевая модель. Соответствующий код этой модели приведен ниже. Здесь в случае заполненности очередей QL и QR запись GOTO-AAX/L в поле *T узла источника CARS указывает на переход к дополнительному узлу AAX (с нулевым временем задержки), где клиент решает вопрос, уехать из банка или получить обслуживание внутри банка. Спецификация узла AAX показывает, что с вероятностью 0.9 клиент будет обслужен внутри банка и с вероятностью 0.1 — покинет банк.

```
$PROJECT;bank model;2/8/96;Taha:
$DIMENSION;ENTITY(50):
$BEGIN:
      CARS  *S;EX(2);/S/LBC(QL+WL,QR+WR);goto-AAX/L:
      QL    *Q;3:
      WL    *F;;UN(3,4);GOTO-TERM:
```

```
        QR     *Q;3:
        WR     *F;;UN(3,4);GOTO-TERM:
        AAX    *A;GOTO-PARK/.9,TERM/.1:
        PARK   *A;3:
        QIN    *Q:
        FIN    *F;;UN(2,3);GOTO-TERM:
$END:
$RUN-LENGTH=480:
$STOP:
```

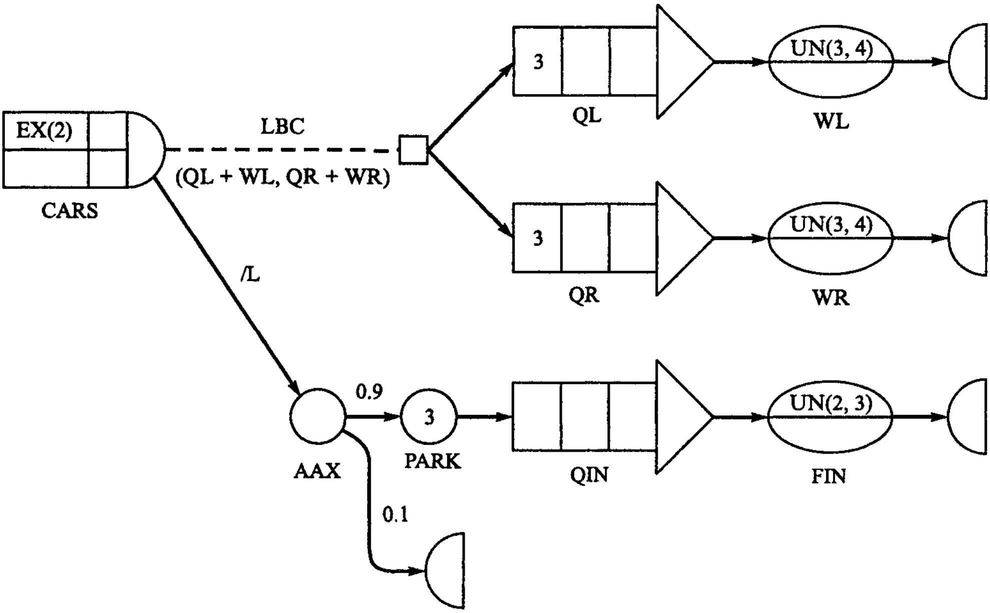


Рис. Б.9

Упражнения Б.5,а

1. Выполните имитацию модели банка из примера Б.5–1 и определите следующие показатели.
   a) Среднее количество автомобилей на каждой полосе обслуживания.
   b) Процент автомобилей на каждой полосе обслуживания, которые не обслужены вследствие больших очередей.
   c) Отношение числа автомобилей (в процентах), прошедших по правой и левой полосам обслуживания.
2. Выполните имитацию модели проверки телевизоров из примера Б.5–2 в течение 480 минут.
3. Выполните имитацию модели из примера Б.5–3 и определите процент загрузки банковского кассира (работающего внутри банка).

4. Средствами SIMNET II создайте модель, описывающую следующую ситуацию. Новые клиенты заходят в банк в среднем каждые 18 минут, причем время между их приходами распределено по экспоненциальному закону. Каждый клиент сначала обращается к банковскому офицеру для открытия нового счета. На это уходит от 15 до 20 минут (равномерное распределение). Далее клиенты с помощью одного из трех банковских кассиров делают первый вклад на свой новый счет. Здесь время обслуживания подчиняется экспоненциальному распределению со средним 5 минут. Все клиенты обслуживаются по правилу FIFO (первый пришел — первый обслужен). Выполните имитацию этой модели в течение 480 минут и определите следующие параметры модели.

   a) Среднюю длину каждой очереди в банке.
   b) Среднее время ожидания в каждой очереди.
   c) Процент клиентов в каждой очереди, которые были обслужены сразу (без ожидания).
   d) Максимальную длину каждой очереди.
   e) Процент загрузки банковского офицера и кассиров.

Б.6. Задание дуг в сетевых моделях

Дуги в сетевых моделях управляют переходом транзакций от одного узла к другому точно так же, как это делается с помощью поля *T (см. раздел Б.5.2). Вместе с тем непосредственное задание дуг используется также для выполнения следующих функций.

1. Проверка выполнения логических условий перед пересылкой транзакции в следующий узел.
2. Выполнение операторов присваивания.
3. Сбор статистических данных с помощью (определенных пользователем) переменных.
4. Возврат ресурсов в начальное состояние.

Эти функции задаются в соответствующих полях оператора описания дуги и приведены в следующей таблице.

*B; F1/SUBF1; F2?; F3%; F4%; F5:			
	Идентификатор поля		Значение по умолчанию
F1		Имя узла назначения	Нет
SUBF1		Тип дуги (см. табл. Б.6)	A
F2	/c/	Условия перехода по этой дуге (запись в поле должна заканчиваться знаком ?)[a]	Нет
F3	/a/	Поле присваивания (запись в поле должна заканчиваться знаком %)	Нет
F4	/v/	Список статистических переменных (запись в поле должна заканчиваться знаком %) (см. раздел Б.7)	Нет
F5	/r/	Определяет ресурсы, возвращаемые дугой	Нет

[a] Для полей F2 ... F5 номер поля n в идентификаторе поля /n/ можно заменить "описательными" словами /Conditions/ (Условия), /Assignments/ (Присваивания), /Variables/ (Переменные) и /Resources/ (Ресурсы) или, соответственно, /C/, /A/, /V/ и /R/.

Общий формат поля условия следующий.

```
условие_1,AND/OR,…,AND/OR,условие_m?
```

Запись в этом поле обязательно должна заканчиваться знаком вопроса (?). Условия могут быть арифметическими или логическими выражениями. Логические выражения могут содержать операторы логических переключателей, которые обсуждаются в разделе Б.8. Условия в виде арифметических выражений имеют следующий общий формат.

```
левое_выражение(=, >, <, >=, <= или <>)правое_выражение
```

Как и в других языках программирования, в SIMNET II логический оператор AND (логическое И) имеет более высокий приоритет, чем логический оператор OR (логическое ИЛИ). Это ограничение можно обойти путем группирования части условий с помощью фигурных скобок {}. Следующие примеры демонстрируют типичные условия, основанные на арифметических выражениях.

```
(A(I)+SAMPLE(J))**2<SUM,AND,LEN(QQ)=0?
{{V(I)>J,OR,K<M},AND,XX=LEN(QQ)}?
```

Поле присваивания F3 используется для выполнения операторов присвоения языка SIMNET II. Это поле может содержать любое количество операторов присвоения согласно следующему формату.

```
присваивание_1,присваивание_2,…,присваивание_m%
```

Запись в этом поле обязательно должна заканчиваться знаком процентов (%). Операторы присвоения, записанные в этом поле, выполняются последовательно.

Операторы присвоения в SIMNET II могут выполняться только при перемещении транзакций. Если условия в поле F2 не выполняются или узел, в который направлена дуга, не может принять транзакцию (например, средство обслуживания занято), тогда операторы присвоения не выполняются.

Присваивания могут выполняться как безусловно, так и по условию, а также в цикле. Синтаксис *оператора условного присваивания* следующий.

IF,*условия*,THEN,*присваивания*,ELSE,*присваивания*,ENDIF

Операторы IF–ENDIF могут вкладываться друг в друга на любую глубину, как показано в следующих примерах.

```
1. IF,{A(1)=1,OR,SUM=0},AND,K=LEN(QQ),
            THEN,A(1)+A(1)+1,SUM=(K+1)**2,
            ELSE,I=LEN(QQ),J=I+1,
      ENDIF
2. IF,I=1,
            THEN,K=K+1,
            IF,K>2,THEN,J=1,A(K**2)=1,ENDIF,
```

```
        ELSE,
        IF,K<=2,THEN,J=0,ENDIF,
   ENDIF
```

Присваивание в цикле осуществляется так же, как в языке FORTRAN, и имеет следующий формат.

```
FOR,счетчик=предел_1,TO,предел_2,STEP,шаг,DO,
      присваивание_1,
      присваивание_2,
            .
            .
            .
      присваивание_m,
NEXT
```

Счетчик — произвольная переменная без индексов. Параметры *предел_1*, *предел_2* и *шаг* могут быть произвольными математическими выражениями, причем значение параметра *шаг* может быть как положительным, так и отрицательным. Если конструкция STEP,*шаг* не используется, то по умолчанию значение параметра *шаг* принимается равным 1.

В языке SIMNET II внутри цикла FOR–NEXT допускаются два специальных оператора присваивания (позаимствованных из языка С).

1. LOOP=BREAK позволяет сразу выйти из цикла, как только значение переменной *счетчик* превысит значение *предел_2*.
2. LOOP=CONTINUE позволяет "перескочить" через оставшиеся операторы присваивания во время выполнения текущей итерации цикла.

Приведем пример оператора FOR–NEXT.

```
IF,I=J,THEN,
      FOR,K=1,TO,J+4,DO,
            IF,K=10,THEN,LOOP=BREAK,ENDIF,
            IF,K=12,THEN,A(1)=A(2),ENDIF,
            FOR,L=K+2,TO,1,STEP,-1,DO,
                  nbr_jobs=nbr_jobs+1,
            NEXT,
      NEXT,
ELSE,
      nbr_jobs=nbr_jobs-1,
      FOR,K=1,TO,3,DO,
         A(K)=A(K+1),
      NEXT,
ENDIF
```

Типы дуг, задаваемые в подполе SUBF1, приведены в табл. Б.6. Переход по дугам типа A, P и L осуществляется так же, как переход из узлов, имеющих такие же значения в поле *T (см. табл. Б.5). Переход по дугам типа S синхронизирован с полем выбора соответствующего узла. Переход по дуге типа C возможен только тогда, когда выполнено указанное условие.

ТАБЛИЦА Б.6

Типы дуг

Символ типа	Описание	Возможность условного перехода
S	Переход в соответствии с заданным правилом выбора	Возможно, но не обязательно
A	Переход только в указанный узел (значение по умолчанию)	Нет
Число (C = 1)	Переход по условию	Обязательно
Число ($0 \le Ð \le 1$)	Переход с заданной вероятностью $Ð$	Возможно, но не обязательно
L	Переход по такой дуге выполняется последним (после просмотра других возможных дуг)	Возможно, но не обязательно

Отметим, что условия перехода нельзя задавать для дуг типа A, эти условия обязательны для дуг типа C и необязательны (но возможны) в описании дуг других типов.

В SIMNET II из узла может исходить любое количество дуг различных типов. Эти дуги упорядочены последовательностью их описаний, при этом следует придерживаться двух правил.

1. Дуги типа S должны помещаться вверху списка дуг, так как они должны быть синхронизированы с полем выбора узла до того, как будут рассмотрены любые другие дуги.
2. Дуги типа L необходимо помещать в конец списка дуг.

Пример Б.6–1

Транзакции, генерируемые источником SS, выбирают ту из двух очередей Q1 и Q2, которая имеет наименьшую длину. Однако в течение первых 100 единиц времени имитации эти очереди не доступны и транзакции должны поступать в очередь Q3. Эта ситуация моделируется следующими операторами (фрагмент модели показан на рис. Б.10).

```
SS     *S;EX(10);/S/LBC(Q1,Q2):
        *B;Q1/S;CUR.TIME>100?:
        *B;Q2/S;CUR.TIME>100?:
        *B;Q3/L:
```

Условие выбора LBC(Q1,Q2) направляет транзакции в ту очередь (Q1 или Q2), которая имеет наименьшую длину. Но прежде чем выбрать такую очередь, SIMNET II проверяет дуги типа S, которые исходят из узла SS и у которых узлом назначения является узел, совпадающий с одним из узлов условия LBC(Q1,Q2). Если по условию LBC(Q1,Q2) выбран узел Q1, то дуга типа S, ведущая к узлу Q1, выбирается автоматически. Поскольку эта дуга содержит условие CUR.TIME>100?, транзакция не может перейти на узел Q1 до тех пор, пока время имитации (переменная CUR.TIME) не превысит 100 единиц. Такая же логика работает, когда в соответствии с правилом LBC(Q1,Q2) будет выбран узел Q2. Таким образом, в течение первых 100 единиц времени будет выполняться переход по дуге типа L в узел Q3.

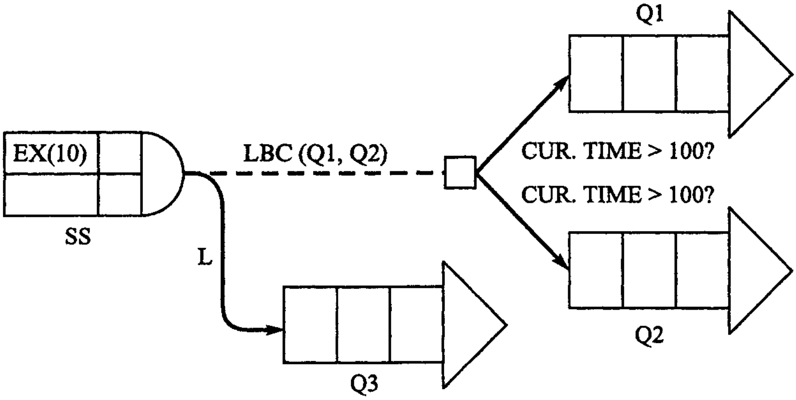


Рис. Б.10

Пример Б.6–2. (Модель загрузки автобусов)

Автобусы прибывают на остановку с интервалом времени EX(30) минут. Количество пассажиров в автобусе обычно от 20 до 50 человек (равномерное распределение). Обычно от 2 до 6 пассажиров (равномерное распределение) покидают автобус на остановке, заходит в автобус от 1 до 8 пассажиров (также равномерное распределение). Высадка одного пассажира занимает UN(3,8) секунд, а посадка — UN(4,7) секунд.

Сетевая модель, описывающая процесс загрузки–разгрузки автобуса на остановке, показана на рис. Б.11. Приведем код SIMNET II, представляющий данную модель.

```
$PROJECT;bus model;5/10/90;Taha:
$DIMENSION;ENTITY(50):
$BEGIN:
       BUS  *S;EX(30)*60:
             *B;UNLD;;I=1,
                   n_on=INT(UN(20,51)),
                   n_off=INT(UN(2,7))%:
       UNLD *A;UN(3,8):
             *B;UNLD/1;I<n_off?;I=I+1%:
             *B;LOAD/L;;J=1,
                   n_wait=INT(UN(1,9)),
                   n_empty=50-n_on+n_off,
                   IF,n_wait>=n_empty,THEN,n_board=n_emp
                 ty,
                   ELSE,n_board=n_wait,ENDIF%:
       LOAD *A;UN(4,7):
             *B;LOAD/1;J<n_board?;J=J+1%:
             *B;TERM/L:
$END:
$RUN-LENGTH=48000:
$STOP:
```

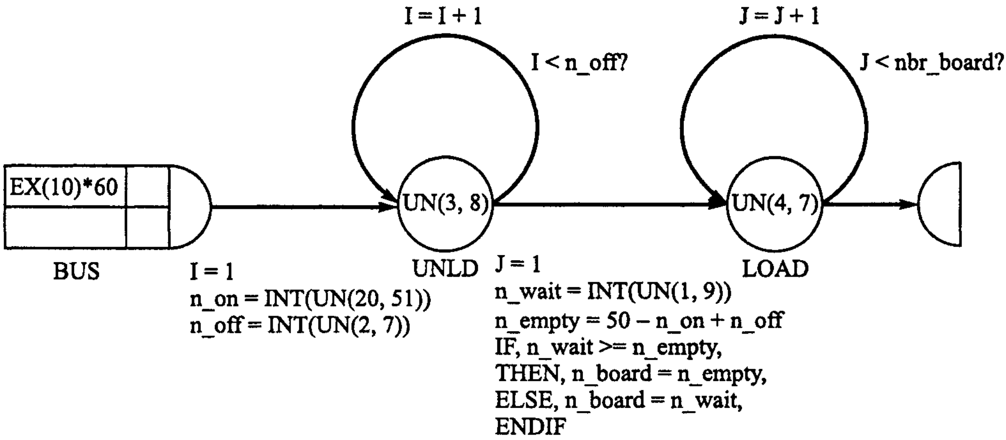


Рис. Б.11

Отметим, что число пассажиров в автобусе, а также количество пассажиров, покидающих автобус и заходящих в него на остановке, определяются как реализация случайных величин, имеющих равномерное *дискретное (целочисленное)* распределение на интервале (a, b), и вычисляются как INT(UN(a, b + 1)), где INT — функция, вычисляющая целую часть своего аргумента (см. табл. Б.1).

Дуга типа A, исходящая из источника BUS, вычисляет количество пассажиров в автобусе (n_on), прибывающем на остановку, и число n_off — количество пассажиров, выходящих из автобуса. Петля вокруг дополнительного узла UNLD имитирует процесс высадки пассажиров. Из этого узла исходит две дуги. Транзакция возвращается на вход узла UNLD по возвратной дуге (дуга с условием UNLD/1), если индекс I меньше n_off. Во время каждого прохода транзакции по этой дуге значение индекса I возрастает на 1. После того как цикл будет выполнен n_off раз, транзакция по дуге типа L (выбор последней дуги) переходит к дополнительному узлу LOAD, который имитирует посадку пассажиров в автобус. Количество пассажиров, которое зайдет в автобус (n_board), определяется как минимум двух чисел: n_empty — количество свободных мест в автобусе и n_wait — количество пассажиров, ожидающих на остановке. Отметим, что в данном случае n_board можно вычислить по следующей формуле.

```
n_board = MIN(INT(UN(1,9)),50-n_on+n_off)
```

В данной модели для вычисления использован оператор IF–ENDIF.

Упражнения Б.6,а

1. Колонны из четырех пустых грузовиков прибывают в порт каждые EX(3) часа и спустя UN(15,20) минут достигают дока, где происходит их погрузка. В доке в каждый момент времени может загружаться только один грузовик. Время погрузки одного грузовика составляет EX(30) минут. Загруженные грузовики покидают док также колонной. Постройте модель погрузки и имитируйте ее работу в течение 100 часов. Определите время, необходимое для загрузки всех грузовиков одной колонны. Также подсчитайте процент использования погрузочных мощностей.

2. Банк открывается в 8:00 и закрывается в 4:30 по полудню. Клиенты прибывают в банк каждые EX(5) минут. Банк имеет одного кассира внутри банка, который начинает работу в 8:00, и одного оператора, который обслуживает клиентов прямо в автомобилях. Последний начинает работу в 9:00. Время обслуживания одного клиента внутри банка составляет EX(4) минут, а в автомобилях — EX(6) минут. Клиенты, которые обслуживаются внутри банка, тратят примерно 3 минуты на парковку автомобиля. Сымитируйте данную ситуацию в течение 5 дней и определите средние времена ожидания обслуживания клиентов внутри банка и на автомобилях. Также определите процент загруженности кассира и оператора.

Б.7. Статистические переменные

Поле 4 (F4) в описании дуг используется для вычисления следующих двух типов определяемых пользователем статистических функций (третий тип функций, называемый RUN.END, здесь не представлен, см. [1]).

1. **Переменные наблюдения** (OBS.BASED), значения которых равны средним каких-либо наблюдений, т.е. равны сумме наблюдаемых значений, деленной на количество наблюдений.
2. **Временные переменные** (TIME.BASED), значения которых равны средним каких-либо наблюдений, зависящих от времени, т.е. равны площади под кривой, соответствующей изменениям переменной во времени, деленной на временной интервал наблюдения за переменной.

Но прежде, чем использовать статистическую переменную в описании дуг, она должна быть определена, т.е. должны быть заданы ее имя, тип, значение и данные для гистограммы. Для этого используется следующий оператор, который размещается выше заголовка $BEGIN.

```
$VARIABLES: Имя_переменной;Тип;Набл_значение;Данные_для_гистограммы
```

Имя переменной определяется пользователем. Наблюдаемые значения для переменных приведены в табл. Б.7. Спецификация данных для гистограмм здесь не приводится; она описана в книге [1].

Таблица Б.7

Статистические переменные

Наблюдаемые значения	Описание
Для переменных всех типов (OBS.BASED и TIME.BASED)	
Выражение	Любое допустимое математическое выражение SIMNET II
Только для переменных типа OBS.BASED	
TRANSIT(#)	Значение разности CUR.TIME – A(#), где # — любое выражение SIMNET II (> 0), при необходимости округляемое до целого
BET.ARVL	Интервал времени между последовательными появлениями транзакций
ARVL.TIME	Время появления транзакций
FIRST	Время появления первой транзакции

Наблюдаемое значение TRANSIT(#) очень удобно для вычисления интервала времени, необходимого для перехода транзакции от одного узла сети к любому другому. В этом случае обычно значение текущего времени CUR.TIME присваивается атрибуту A(#) начального (для отслеживаемой транзакции) узла сети. Когда транзакция достигает конечного узла, время перехода от начального узла к конечному вычисляется путем вычитания из значения текущего времени CUR.TIME значения атрибута A(#).

Приведем в качестве примера несколько определений статистических переменных.

```
$VARIABLES:   SYS TIME;OBS.BASED;TRANSIT(1):
                INV LEVEL;TIME.BASED;I:
                T BET BLKGE;OBS.BASED;BET.ARVL:
```

Для того чтобы можно было собрать статистические данные для переменных типа OBS.BASED, имена этих переменных должны помещаться в поле F4 соответствующих дуг с использованием следующего формата.

```
имя_пеменной_1,имя_переменной_2,...%
```

Запись в этом поле обязательно должна заканчиваться знаком процентов %. В SIMNET II работа с переменными типа TIME.BASED автоматизирована, поэтому их имена нет необходимости помещать в поле F4 дуг.

Пример Б.7–1. (Модель мойки автомобилей)

Автомобили прибывают на мойку, имеющую только один бокс, каждые EX(10) минут. Они становятся в одну очередь, которая вмещает не более пяти машин. Если очередь полна, то автомобиль уезжает. Время мойки одного автомобиля составляет UN(10,15) минут. Нам необходимо оценить две статистики: время ожидания автомобилем начала мойки и интервал времени между последовательными отъездами "немытых" автомобилей.

На рис. Б.12 показана сетевая модель, имитирующая описанную ситуацию. Приведем код модели, а также фрагмент выходного отчета об имитации, содержащего значения статистических переменных (BET BALKS — переменная, отслеживающая время между последовательными отъездами "упущенных" (balk) автомобилей).

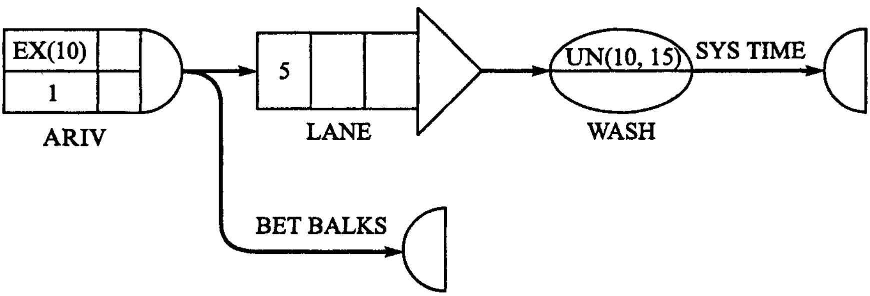


Рис. Б.12

```
$PROJECT;Car Wash Model;4/8/90;Taha:
$DIMENSION;ENTITY(20),A(1):
$VARIABLES:SYS TIME;OBS.BASED;TRANSIT(1):
           BET BALKS;OBS.BASED;BET.ARVL:
```

```
$BEGIN:
        ARIV *S;EX(10);;1:    !Присваивание A(1)=CUR.TIME
             *B;LANE:
             !Постановка в очередь, если она не полна
             *B;TERM/L;/V/BET BALKS%:
             !Отъезд авто из мойки, если очередь полна
        LANE *Q;5:            !Емкость очереди 5 авто
        WASH *F;;UN(10,15):   !Мойка автомобиля
             *B;TERM;/V/SYS TIME%: !Вычисление SYS TIME
$END:
$RUN-LENGTH=480:
$STOP:
```

			VARIABLES			
	UPDATES	AV.VALUE	STD DEV	MIN VALUE	MAX VALUE	LAST UPDATE
(O)SYS TIME	38	41.61	21.75	12.33	77.74	73.42
(O)BET BALKS	9	17.15	21.40	.68	71.29	1.21

Когда источник ARIV создает транзакцию, автоматически происходит присваивание A(1)=CUR.TIME (поскольку поле 3 описания источника имеет значение 1). Первая дуга из источника ARIV ведет транзакцию в очередь LANE. Если эта очередь полная, то транзакция по дуге типа L переходит в пункт TERM, т.е. удаляется из системы. Это соответствует тому, что "немытый" автомобиль покидает автомойку. Статистическая переменная BET BALKS имеет тип наблюдения BET.ARVL и вычисляется на дуге, ведущей в пункт TERM. Эта переменная автоматически отслеживает время между отъездами "немытых" автомобилей. Для транзакций, завершивших обслуживание в узле WASH (Мойка), вычисляется значение переменной SYS TIME как длина интервала времени между созданием транзакции в источнике ARIV и ее удалением из системы.

Б.8. Логические переключатели

Очереди — это буферные устройства между другими узлами, где транзакции могут находиться неопределенное время. Вместе с тем поступающая на вход очереди транзакция может "перескочить" через нее, если она пуста и следующий узел доступен. Кроме того, узел средства обслуживания автоматически получает транзакцию из предшествующей очереди, как только у него освобождаются сервисы. Таким образом, поступление транзакций на вход очереди и выход их из нее управляются состоянием *других* узлов сети, а не самой очередью.

Но в некоторых ситуациях необходимо управлять транзакциями непосредственно в очереди. Например, постановку оборудования на ремонт можно смоделировать как "блокировку" пути из предшествующей очереди, когда транзакция не может поступить на обслуживание последующим узлом, имитирующим оборудование, которое подлежит ремонту. После завершения ремонта можно “восстановить” оборудование путем востребования из очереди ожидающей транзакции. Логические переключатели предназначены для такого управления транзакциями в очередях.

Логический переключатель определяется с помощью оператора $SWITCHES (Переключатели) следующим образом.

```
$SWITCHES:   имя_переключателя;нач_сост;очередь1,очередь2,...:
```

Имя переключателя определяется пользователем. Начальное состояние переключателя (определяется до начала имитации) — это одно из состояний ON (Включено) или OFF (Выключено). Список имен очередей соответствует тем очередям, которые управляются данным переключателем. Например, оператор

```
$SWITCHES:   SW;ON;Q1,Q2:
```

показывает, что переключатель SW в первоначальном состоянии включен и управляет очередями Q1 и Q2.

Для управления самими логическими переключателями применяются два специальных оператора присваивания

```
SWITCH имя_переключателя=ON
SWITCH имя_переключателя=OFF
```

Эти операторы применяются двумя различными способами.

1. В виде условия, которое проверяет текущее состояние переключателя.
2. Как операторы присваивания, которые изменяют текущее состояние переключателей (из состояния ON в состояние OFF и наоборот).

В первом случае условие, проверяющее состояние переключателя, включается (возможно с помощью логических операторов AND и OR) в запись второго поля описания дуги или в оператор IF-ENDIF. Во втором случае оператор присваивания входит в третье поле оператора описания дуги.

Проверку состояния переключателя можно рассматривать просто как двоичное условие (0 или 1), которое выполняется или нет. Но использование операторов присваивания значений переключателям позволяет управлять очередями. Выполнение оператора присваивания

```
SWITCH имя_переключателя=ON
```

автоматически задерживает выход транзакций из всех очередей, перечисленных в списке оператора определения переключателя. Конечно, если очередь пуста или последующий (после очереди) узел занят либо условия на дуге не позволяют перемещение транзакции, то выполнение оператора присваивания SWITCH никакого эффекта не окажет.

Следует запомнить, что задержка транзакций в управляемых очередях происходит при включенном (ON) состоянии переключателя только тогда, когда дуга, где задан данный переключатель, становится активной. Выполнение оператора присваивания `SWITCH имя_переключателя=OFF` изменяет состояние переключателя, но не оказывает влияния на контролируемые очереди.

Пример Б.8–1. (Профилактика оборудования)

Детали для обработки на неком станке поступают каждые EX(11) минут. Время обработки одной детали составляет EX(12) минут. После 8 часов работы станок останавливается для профилактических работ, которые длятся UN(15,20) минут.

Приведем код модели, описывающей данную ситуацию (графическая схема модели показана на рис. Б.13).

```
$PROJECT;Maintenance Model;4/10/90;Taha:
$DIMENSION;ENTITY(30):
$SWITCHES:  SW;ON;QJOBS:
$BEGIN:
       SS    *S;/L/LIM=1:    !Модель обслуживания станка
       DELAY *A;480:
              *B;MAINT;;SW=OFF%:
       MAINT *A;UN(15,20):
              *B;DELAY;;SW=ON%:
       ARIV  *S;EX(11):      !Модель работы станка
       QJOBS *Q:
              *B;MACH/1;SW=ON?:
       MACH  *F;;EX(12);goto-TERM:
$END:
$RUN-LENGTH=1000:
$STOP:
```

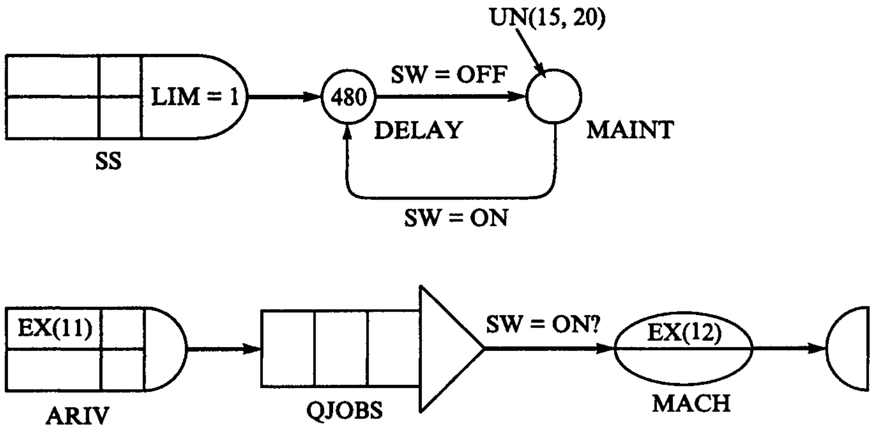


Рис. Б.13

Отметим, что модель разбита на два *отдельных* сегмента. Первый моделирует цикл профилактики станка, второй соответствует обработке деталей на станке. Условие SW=ON? на дуге, идущей от очереди QJOBS (очередь деталей, поступивших на обработку) к узлу средства обслуживания MACH (станок), управляет выходом очереди QJOBS. Транзакции могут покинуть эту очередь, если только переключатель SW включен.

Состоянием переключателя управляет сегмент модели, имитирующий профилактический цикл станка. В этом сегменте источник SS продуцирует только *одну* транзакцию, которая поступает на дополнительный узел DELAY (Задержка), где она ожидает 8 часов, соответствующих времени работы станка MACH между профилактическими работами. Когда транзакция покидает узел DELAY, станок MACH должен стать на профилактику. Эта транзакция при переходе по дуге от узла DELAY к узлу MAINT (Профилактика) выполняет присваивание SW = OFF, останавливающее поток транзакций от очереди QJOBS к узлу MACH. Транзакция, поступившая на узел MAINT, остается в нем период времени, равный времени профилактического обслуживания станка. Когда транзакция покидает этот узел, она опять поступает на вход узла DELAY, чтобы повторить профилактический 8-часовой цикл обслуживания станка. На дуге, ведущей к узлу DELAY, выполняется присваивание SW = ON, которое открывает ожидающим в очереди QJOBS транзакциям путь к узлу MACH. Если во время присваивания SW = ON очередь QJOBS пуста, то происходит разблокирование этой очереди и вновь пришедшая на вход очереди транзакция просто пройдет ее без задержки и сразу поступит на обработку в узел MACH.

Упражнения Б.8,а

1. В модели профилактических работ из примера Б.8–1 профилактика начинается сразу, как только истечет 480 минут после начала работы станка (когда транзакция покинет узел DELAY). Но вполне вероятно, что станок (узел MACH) в этот момент будет занят обработкой очередной детали. Поэтому его профилактика должна начаться тогда, когда он завершит обработку этой детали. Измените модель в соответствии с данным условием.
2. Измените модель профилактических работ из примера Б.8–1 таким образом, чтобы профилактика начиналась не после 8 часов работы станка, а после того, как на нем будут обработаны 50 деталей.

Б.9. Специальные операторы присваивания

В предыдущем разделе мы использовали специальные операторы присваивания `имя_переключателя=ON` или `OFF`, с помощью которых можно управлять состоянием переключателей. Кроме этих операторов присваивания, в SIMNET II существуют другие специальные операторы присваивания, которые позволяют выполнять следующие функции.

1. Активизировать и деактивизировать источники транзакций.
2. Собирать значения статистических переменных.
3. Управлять транзакциями в очередях и средствах обслуживания.
4. Управлять параметрами очередей.
5. Выбирать входные данные из файлов.
6. Управлять атрибутами.
7. Управлять временем имитации модели.
8. Подсоединять внешние файлы READ и WRITE.

В этом приложении мы рассмотрим только первые три возможности из всех перечисленных, остальные описаны в книге [1].

Б.9.1. Активизация и деактивизация источников

В определении узла источника (раздел Б.2.1) пятое поле (LIM=) используется для задания количества генерируемых транзакций или времени генерации транзакций. Эти ограничения, заданные до начала выполнения имитации, нельзя изменить во время выполнения имитации модели. Но с помощью специальных операторов присваивания можно мгновенно *приостановить* (suspend) или *возобновить* (resume) работу источника. Для этого используются следующие операторы присваивания.

```
SUSPEND=имя_источника
RESUME=имя_источника
```

Выполнение этих операторов "навсегда" (во время текущей имитации) аннулирует значения в поле F2 (время генерации первой транзакции) и в поле F5 (ограничения на количество генерируемых транзакций или на время их генерирования) в коде оператора описания источника.

Пример Б.9–1. (Производственная линия с поломками)

С автоматического конвейера каждые UN(1,2) минуты сходит одна единица продукции, которая поступает далее на участок контроля. Проверка одной единицы продукции занимает 1.5 минуты. Примерно каждые EX(120) минут на конвейере происходит сбой (поломка), устранение которого требует UN(5,10) минут.

Приведем код модели, описывающей данную ситуацию (графическая схема модели показана на рис. Б.14).

```
$PROJECT;Production Line Breakdown;4/14/90;Taha:
$DIMENSION;ENTITY(50),A(1):
$VARIABLES:  SYS TIME;;TRANSIT(1):
$BEGIN:
       P_LINE    *S;UN(1,2);;1:
                      !Получение продукции с конвейера
       Q_INSPECT *Q:              !Ожидание проверки
       F_INSPECT *F;;1.5:         !Проверка
                  *B;TERM;/v/SYS TIME%:
                                  !Вычисление SYS TIME
       S_BREAK   *S;EX(120);120:  !Первая поломка
                  *B;REPAIR;;SUSPEND=P_LINE%:
                                  !Остановка конвейера
       REPAIR    *S;UN(5,10):     !Ремонт конвейера
                  *B;TERM;;RESUME=P_LINE%:
                      !Восстановление работы конвейера
$END:
$RUN-LENGTH=480:
$STOP:
```

На рис. Б.14 видно, что в данном случае модель состоит из двух отдельных сегментов, где первый моделирует нормальную работу конвейера (точнее, процесс проверки изделий, получаемых с конвейера), а второй — ситуацию поломки конвейера. Транзакция, соответствующая поломке конвейера, генерируется источником S_BREAK, причем первая транзакция появится не ранее чем через 120 минут после начала имитации. После генерации такая транзакция переходит в дополнительный узел REPAIR (Ремонт), причем при переходе по соответствующей дуге происходит присваивание `SUSPEND=P_LINE` и конвейер (узел-источник P_LINE) сразу останавливается. После проведения ремонта (узел REPAIR) на следующей дуге происходит присвоение `RESUME=P_LINE`, следовательно, возобновляется генерация транзакций источником P_LINE.

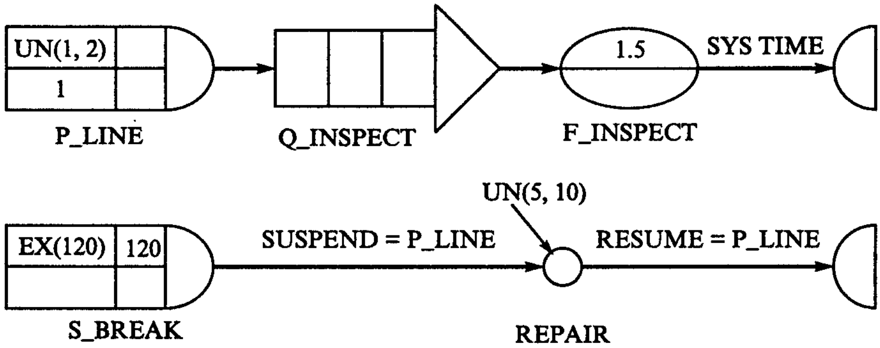


Рис. Б.14

Б.9.2. Сбор значений статистических переменных

Мы видели в разделе Б.7, что наблюдаемые значения статистических переменных типа OBS.BASED можно собрать с помощью списка имен этих переменных в четвертом поле оператора описания соответствующих дуг. Такого же результата можно достичь посредством оператора

```
COLLECT=имя_переменной
```

Этот оператор часто применяется тогда, когда надо собрать статистическую информацию в зависимости от некоторых условий, например, в операторе IF-ENDIF. Применение этого оператора будет проиллюстрировано далее в примере Б.9–3.

Б.9.3. Операторы управления транзакциями

Операторы этой группы (табл. Б.8) позволяют менять местами, удалять, добавлять, копировать, заменять транзакции и определять их местоположение в очередях и средствах обслуживания. Будем называть узел в правой части оператора присваивания *донором*, а узел в левой части — *реципиентом*. Например, в операторе `LAST(QQ)=1(WW)` очередь QQ является реципиентом, а WW — донором. Этот оператор перемещает первую транзакцию очереди WW на последнюю позицию в очереди QQ. Существенно, что выражения, обозначенные как a и b в табл. Б.8, могут быть произвольными математическими выражениями языка SIMNET II. Например, допустимым будет выражение

```
((J+1)**2+K)(Q1)=(MAX(I+J,K-M))(Q2).
```

SIMNET II при необходимости самостоятельно округляет вычисленные значения до целых.

ТАБЛИЦА Б.6

Операторы присваивания, управляющие транзакциями

Оператор[a]	Описание
Только для очередей	
a(Q1) = *b*(Q2)	Перемещает *b*-ю транзакцию очереди Q2 в *a*-ю позицию очереди Q1
a(Q1) = ALL(Q2)	Перемещает *все* транзакции очереди Q2 в очередь Q1 и помещает их начиная с *a*-й позиции
a(Q1) = TRANS	Помещает копию транзакции, проходящую в текущий момент по данной дуге, в *a*-ю позицию очереди Q1
a(Q1) = DEL	Удаляет (разрушает) транзакцию в *a*-й позиции очереди Q1
ALL(Q1) = DEL	Удаляет все содержимое очереди Q1 (эта очередь становится пустой)
INS(Q1) = *b*(Q2)	Передает *b*-ю транзакцию очереди Q2 на вход очереди Q1 (эта транзакция будет помещена в очередь Q1 в соответствии с ее принципом организации)
INS(Q1) = ALL(Q2)	Передает *все* транзакции очереди Q2 на вход очереди Q1 (транзакции будут помещены в очередь Q1 в соответствии с ее принципом организации)
INS(Q1) = TRANS	Передает транзакцию, проходящую в текущий момент по данной дуге, на вход очереди Q1 (транзакция будет помещена в очередь Q1 в соответствии с ее принципом организации)
Только для средств обслуживания	
a(F1) = REL	Немедленно освобождает от обслуживания *a*-ю транзакцию в узле обслуживания F1
ALL(F1) = REL	Немедленно освобождает от обслуживания все транзакции в узле обслуживания F1
Для очередей и средств обслуживания	
a(Q1 или F1) = REP	Заменяет атрибуты *a*-й транзакции в Q1 или F1 на атрибуты текущей транзакции
COPY = *b*(Q1 или F1)	Заменяет атрибуты транзакции, проходящей в текущий момент по данной дуге, на атрибуты *b*-й транзакции, находящейся в Q1 или F1

[a] Здесь *a* и *b* — произвольные математические выражения языка SIMNET II или служебное слово LAST. Очереди Q1 и Q2 могут быть одной и той же очередью, если в этой очереди необходимо переупорядочить транзакции.

При использовании данных операторов присваивания следует руководствоваться следующими правилами.

1. Операторы из раздела "Только для очередей" (за исключением тех, которые содержат опцию ALL) являются *активными* в том смысле, что очередь–реципиент автоматически принимает поступающую транзакцию, причем в тот момент, когда выполняется оператор присваивания. В случае использования операторов с опцией ALL при необходимости следует применять логические переключатели (см. раздел Б.8), чтобы обеспечить обработку всех транзакций.

2. Автоматическое перемещение транзакций *не выполняется*, если реализация оператора присваивания, управляющего транзакциями, происходит *внутри* другого аналогичного оператора присваивания. Вложенность этих операторов происходит тогда, когда перемещение транзакций является результатом выполнения оператора присваивания, управляющего транзакциями, что, в свою очередь, вызвано выполнением *другого* оператора присваивания. Например, если выполнение оператора a(Q1) = b(Q2) привело к выходу транзакции из очереди Q1, в результате чего должен выполниться оператор a(Q3) = b(Q4), то SIMNET II осуществит перемещение транзакции в очередь Q3, но это *не приведет* к выходу очередной транзакции из очереди Q3. Чтобы в данной ситуации осуществить выход из очереди Q3, следует использовать логические переключатели.
3. Если узел-донор пуст, то никаких действий не происходит.
4. Если очередь-реципиент имеет *конечную емкость* и в момент выполнения оператора присваивания она полна, то происходит остановка процесса имитации по ошибке. Исключением из этого правила является ситуация, когда донор и реципиент являются одной и той же очередью (т.е. происходит реорганизация транзакций внутри одной очереди).

Пример Б.9–2. (Модель банка)

Изменим модель банка из примера Б.5–1, наложив дополнительное условие: если в полосе, ведущей к пункту обслуживания, не больше двух автомобилей, то автомобиль из более длинной очереди переезжает в более короткую.

Ниже представлены операторы, описывающие эту ситуацию; соответствующая сетевая модель графически показана на рис. Б.15. Положим, что автомобиль переходит из более длинной очереди в более короткую после того, как будет обслужен автомобиль из более короткой очереди. Пункты обслуживания WL и WR ведут к одному дополнительному узлу AX. На дуге, выходящей из узла AX, вычисляется значение переменной DIFF, равной разности между количествами автомобилей на правой и левой полосах движения. Если DIFF > 1, то транзакция LAST(QR) перемещается в положение LAST(QL). Если DIFF < –1, то происходит обратное перемещение.

```
$PROJECT;Bank Model;6 June 1990;Taha:
$DIMENSION;ENTITY(50),A(2):
$VARIABLES: Sys Time(1-2);OBS.BASED;TRANSIT(1):
$BEGIN:
  CARS   *S;EX(5);;1/S/LBC(QL+WL,QR+WR):
  QL     *Q;3:
  WL     *F;;UN(3,4):
         *B;AX;;A(2)=1%:
  QR     *Q;3:
  WR     *F;;UN(3,4):
         *B;AX;;A(2)=2%:
  AX     *A:
         *B;TERM;
             /a/DIFF=LEN(QR)+LEN(WR)-(LEN(QL)+LEN(WL));
             IF,DIFF>1,THEN,LAST(QL)=LAST(QR),ENDIF;
```

```
              IF,DIFF<-1,THEN,LAST(QR)=LAST(QL),ENDIF%;
              /v/Sys Time(A(2))%:
$END:
$RUN-LENGTH=480:
$STOP
```

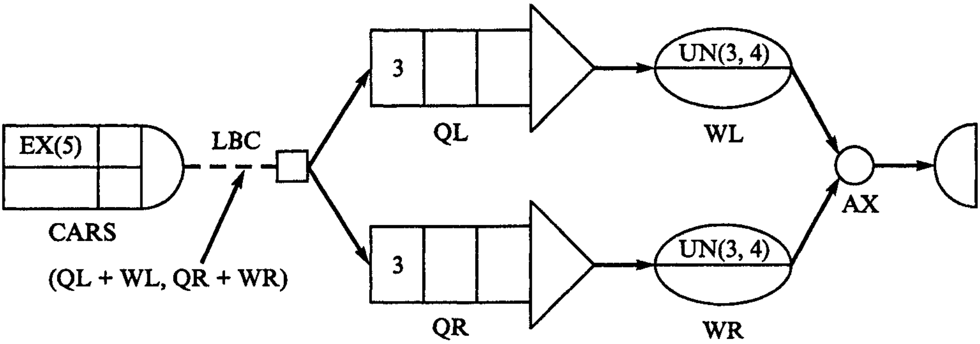


Рис. Б.15

Пример Б.9–3. (Модель канала передачи)

Сообщения на входе простого канала передачи появляются каждые UN(7,8) секунд, при этом процесс передачи сообщения занимает примерно UN(6,8) секунд. Примерно раз в UN(600,650) секунд в канале передачи происходит сбой, тогда сообщение, которое находилось в процессе передачи на момент сбоя, необходимо передать заново. Восстановление канала передачи после сбоя обычно занимает около 30 секунд.

Приведем код имитационной модели.

```
$PROJECT;Transmission Channel;8 June 1990;Taha:
$DIMENSION;ENTITY(50),A(2):
$VARIABLES: Sys Time;;TRANSIT(1):
                  PRCNT ABORTED;RUN.END;N/COUNT(CHNL)*100:
$SWITCHES: SW;;QMSG:
$BEGIN:
  ARIV     *S;UN(7,8);;1:               !Генерация сообщений
  QMSG     *Q:                          !Ожидание в очереди
           *B;CHNL/1;SW=ON?:            !SW управляет QMSG
  CHNL     *F;;UN(6,8):                 !Передача сообщения
           *B;TERM;;                    !Вычисления
             IF,A(2)=-1,THEN,A(2)=0,N=N+1,1(QMSG)=TRANS,
             ELSE,COLLECT=SYS TIME,ENDIF%:
  START    *S;/L/LIM=1:                 !Сбой
  FAIL     *A;UN(600,650):              !Вычисление момента сбоя
           *B;RESET;                    !Вычисления при сбое
             SW=OFF;                    !Блокировка QMSG
             COPY=1(CHNL);
             A(2)=-1;
             1(CHNL)=REP;
```

```
                    1(CHNL)=REL%:
        RESET *A;30:                    !Вычисление длительности сбоя
              *B;AIL;;SW=ON%:           !Разблокировка QMSG
      $END:
      $RUN-LENGTH=9000:
      $STOP
```

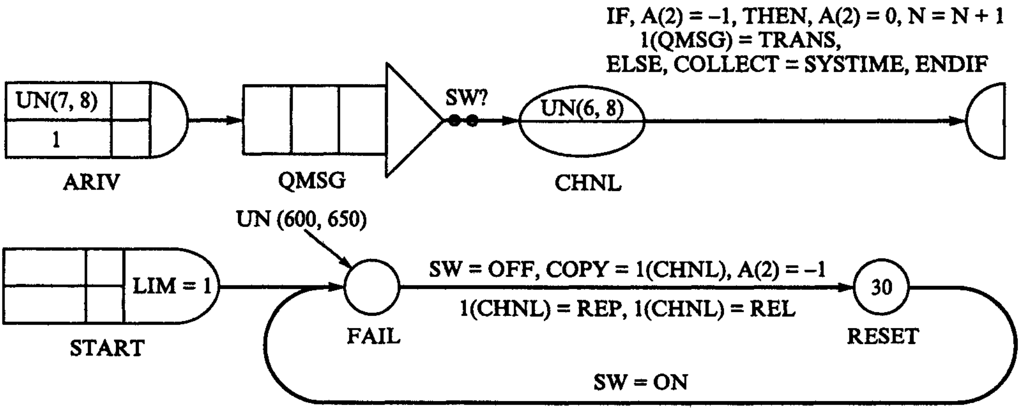


Рис. Б.16

Данная модель, как показано на рис. Б.16, состоит из двух сегментов. Первый сегмент представляет модель передачи сообщений по каналу, второй — моделирует ситуацию сбоя в канале передачи. Когда в канале происходит сбой, соответствующая транзакция[1], покидая дополнительный узел FAIL (Сбой), устанавливает переключатель SW в состояние OFF и тем самым запрещает транзакциям (имитирующим сообщения) покидать очередь QMSG (Сообщения). Сообщение, находящееся в момент сбоя в процессе передачи в узле обслуживания CHNL (Канал), вследствие выполнения оператора `1(CHNL)=REL` освобождается от обслуживания. Это сообщение еще *до выхода* из узла CHNL помечается атрибутом A(2) = –1. Для этого сначала выполняется оператор присваивания `COPY=1(CHNL)`, копирующий атрибуты транзакции–сообщения атрибутам транзакции, проходящей по дуге FAIL→RESET. Затем происходит присваивание A(2) = –1 и далее выполняется оператор `1(CHNL)=REP`, заменяющий значение атрибута A(2) на –1 у транзакции, находящейся внутри узла CHNL, но оставляющий без изменения атрибут A(1). Если в момент сбоя узел CHNL был пустым, то никаких действий не происходит.

Покидающая узел CHNL транзакция имеет атрибут A(2) = –1, т.е. помечена как сбойное сообщение. Эта транзакция, проходя по дуге, выходящей из узла CHNL, инициирует выполнение операторов `A(2) = 0` и `1(QMSG) = TRANS`, которые помещают сбойное сообщение в заголовок очереди QMSG, изменив при этом атрибут A(2) на нуль. Далее происходит восстановление канала передачи (выход транзакции из узла RESET (Восстановление)), при этом переключатель SW переводится во включенное состояние (оператор `SW = ON`), что разрешает передачу сообщений через узел CHNL.

[1] Транзакцию, инициирующую сбой, генерирует источник START. — *Прим. ред.*

Для нахождения времени, затрачиваемого на передачу сообщений, статистическую переменную SYS TIME следует вычислять только тогда, когда атрибут A(2) очередной транзакции имеет нулевое значение. Это достигается путем использования оператора `COLLECT=SYS TIME` внутри условного оператора IF-ENDIF, который проверяет условие A(2) = 0 всех транзакций, покидающих узел CHNL.

Упражнения Б.9,а

1. С помощью операторов присваивания SUSPEND и RESUME смоделируйте источник транзакций, которые будет генерировать транзакции только по запросу.
2. Банк ежедневно открывается в 8:00 и закрывается в 16:30. Интервал времени между приходами клиентов в банк подчиняется экспоненциальному распределению EX(7) (минуты). Клиентов обслуживают два банковских кассира, время обслуживания одного клиента составляет EX(10) минут. В 16:30 банк закрывается, но клиенты, оставшиеся внутри банка, должны быть обслужены. Оцените максимальное, минимальное и среднее время обслуживания клиентов, оставшихся в банке после 16:30.
3. Детали последовательно обрабатываются на двух станках: MACH1 и MACH2. Они поступают на обработку с интервалом времени EX(0.0588) минут. Время обработки на первом и втором станках составляет соответственно 0.04 и 0.0357 минут. Вспомогательная площадь возле станка может вместить 50 деталей для станка MACH1 и 40 — для станка MACH2. На станке MACH2 каждые EX(8) минут необходимо заменять режущий инструмент, на что уходит примерно UN(1,3) минуты. Постройте модель и выполните ее имитацию в течение 30 минут для определения коэффициента загрузки станков.

Б.10. Начальные данные

В SIMNET II модели могут иметь шесть типов начальных данных.

1. Начальное содержимое узлов очередей и средств обслуживания.
2. Плотности вероятностей дискретных распределений.
3. Таблично задаваемые функции.
4. Значения массивов.
5. Значения переменных (неиндексированных).
6. Функции и математические выражения.

Эти данные содержатся в модели в специальных форматах, зависящих от типа данных, и используются в течение одной имитационной сессии.

Б.10.1. Начальное содержимое очередей и средств обслуживания

Для задания начального содержимого узлов очередей и средств обслуживания используется следующий формат.

```
$INITIAL-ENTRIES: i-j/имя_узла_1/атрибуты_первой_транзакции;
                                                    •
                                                    •
                                                    •
    атрибуты_последней_транзакции:
                                              •
                                              •
                                              •
                                    (содержимое других узлов)
                                    •
                                    •
                                    •
                        (повторение j - i раз)
```

Здесь i и j — целочисленные значения, определяющие порядковые номера обращений к данным узлам; только тогда будут использованы эти начальные данные. Формат записи данных для каждой транзакции следующий:

```
(d)A(1),A(2),...,A(n)
```

В этой записи d обозначает число дубликатов списка A(1),A(2),...,A(n), принимаемых узлом транзакций, n — количество атрибутов, определенных с помощью оператора $DIMENTION (Размерность). Каждый список атрибутов для одной транзакции должен заканчиваться точкой с запятой, а последний список узла — двоеточием. Нулевые значения атрибутов принимаются по умолчанию, поэтому их можно опускать (обозначая их местоположение запятыми).

Приведем пример оператора $INITIAL-ENTRIES, задающего начальное содержимое для очередей Q1 и Q2 и средства обслуживания F1. Здесь предполагается, что транзакции имеют *два* атрибута.

```
$INITIAL-   1-1/Q1/(2)11,22;12,23;  !Первые 3 транзакции для Q1
ENTRIES:
                   -11,;            !4-я транзакция (-11,0)
                   ;                !5-я транзакция (0,0)
                   :                !6-я транзакция (0,0)
                F1/-11,-22:         !1 транзакция для F1
            2-3/Q2/10,20:           !Значения для 2 и 3
                                    !обращений
                F1/110,220:         !к F1 и Q2
```

Хотя последовательные значения для одного узла можно записать в одну строку, приведенный формат более удобочитаем.

Применение оператора $INITIAL-ENTRIES показано в примере Б.10–1 в конце раздела Б.10.4.

Б.10.2. Задание плотностей вероятностей

Эмпирические дискретные распределения в SIMNET II задаются в следующем формате.

```
$DISCRETE-PDFS: i-j/N1/x11,p11;...;xin,pin:
                                      •
                                      •
                                      •
                            (Задание других плотностей)
                           •
                           •
                           •
                           (Повторение j - i раз)
```

Здесь

N — количество точек (x, p), задающих дискретную плотность вероятности,
x — значение случайной переменной,
p — вероятность, с которой случайная величина принимает значение x.

Приведем пример задания дискретных распределений (т.е. дискретных плотностей вероятностей).

```
$DISCRETE-PDFS: 1-1/3/1,.1;2,.4;3,.5:         !Функция 1
                           2/0,.6;1,.4:       !Функция 2
                     2-3/3/2,.3;3,.5;4,.2:    !Функция 1
```

Случайное число, генерируемое SIMNET II в соответствии с заданными распределениями, обозначается как DI(a, RS), где a — номер плотности вероятностей для данного этапа имитации, RS — порядковое число, определяющее последовательность случайных чисел (по умолчанию RS = 1). Как параметр a, так и RS могут быть любыми математическими выражениями SIMNET II. Если параметр a положителен, то для генерирования случайного числа используется *дискретная* плотность вероятности, определенная оператором $DISCRETE-PDFS. Если же параметр a отрицателен, тогда при вычислении случайного числа используется *кусочно-линейная* плотность вероятности, определенная на основе данных, приведенных в операторе $DISCRETE-PDFS. Случайное число DI(a, RS) можно использовать в любых математических выражениях SIMNET II.

Б.10.3. Таблично-заданные функции

Эти функции используются для задания значений зависимой переменной y как функции независимой переменной x. Оператор задания таких функций подобен оператору $DISCRETE-PDFS и имеет следующий вид.

```
$TABLE-LOOKUPS: i-j/N1/x11,y11;...;x1n,y1n:
                                      •
                                      •
                                      •
                            (Задание других функций)
                           •
                           •
                           •
                           (Повторение j - i раз)
```

Приведем пример задания таких функций.

```
$TABLE-LOOKUPS: 1-1/4/1,2;3,5;6,7;7,9:        !Функция 1
                           3/0,0;1,4;2,11:    !Функция 2
```

Значения таких функций обозначаются как TL(n,x), где n — номер функции, а x — аргумент функции. Если n — отрицательное число, то значение TL(n,x) получается путем линейной интерполяции данных, представленных в операторе $TABLE-LOOKUPS. Параметры n и x могут быть любыми математическими выражениями SIMNET II. При необходимости значение выражения, представляющее параметр n, автоматически округляется до ближайшего целого. Значение x должно быть из области определения таблично-заданной функции. Значение TL(n,x) может входить в любые математические выражения.

В предыдущем примере TL(1,3) = 5, тогда как выражение TL(1,4) приводит к ошибке, поскольку функция 1 не определена при $x = 4$. Вместе с тем для второй функции имеем значение TL(–2,1.5) = 7.5, которое получено линейной интерполяцией значений функции при $x = 4$ и $x = 11$.

Б.10.4. Задание элементов массивов

Значения элементов массивов, определенных оператором $DIMENSION, задаются с помощью следующего оператора.

```
$ARRAYS:
          массив_1; i-j/список значений:
                               •
                               •
                               •
                            (Повторение j - i раз)
                         •
                         •
                         •
                      (Задание других массивов)
```

В SIMNET II существует два формата списка значений элементов массива.

1. *Явный*, когда каждый элемент матрицы идентифицируется своими индексами.
2. *Неявный*, когда элементы матрицы идентифицируются местоположением в списке значений.

В следующем примере проиллюстрировано применение этих форматов для задания значений двух массивов BC(2) и YZ(3, 2), определенных с помощью оператора $DIMENSION (BC является одномерным массивом, состоящим из двух элементов, массив YZ — двухмерным, состоящим из 3 столбцов и 2 строк).

```
$ARRAYS:
   BC;1-1/NS/11,22:  !Неявное задание: BC(1)=11, BC(2)=22
       2-4/2,33:     !Явное задание: BC(2)=33
   YZ;1-2/1,2,88;3,1,99:
                     !Явное задание: YZ(1,2)=88, YZ(3,1)=99
       3-3/NS/11,22,33:
       !Неявное задание: YZ(1,1)=11, YZ(1,2)=22, YZ=(2,1)=33
```

Когда используется опция NS, указывающая на неявное задание, присваивание значений элементам массива происходит по столбцам сверху вниз. Явное задание удобно тогда, когда надо задать значения только некоторым элементам массива. Отметим, что если при неявном задании значения каких-либо элементов не указаны, они по умолчанию считаются равными нулю. Например, в последнем задании массива YZ значения элементов YZ(2,2), YZ(3,1) и YZ(3,2) равны нулю.

Пример Б.10–1. (Модель перевозки гравия)

Этот пример иллюстрирует применение операторов задания начальных данных $INITIAL-ENTRIES и $ARRAYS.

Компания использует три 20-тонных и два 30-тонных грузовика для доставки гравия потребителям. Спрос на гравий очень высок, что позволяет избегать простоев грузовиков. 20-тонный грузовик загружается за 10 минут, а 30-тонный — за 15 минут. Время, которое затрачивается на доставку гравия потребителю и обратно, равно UN(30,60) минут. Мы хотим путем имитации этой ситуации оценить общий тоннаж грузов, перевозимых грузовиками каждого типа за 24 часа работы.

Приведем код модели; ее графическое представление показано на рис. Б.17.

```
$PROJECT;Gravel Hauling;9 June 1990;Taha:
$DIMENSION;ENTITY(50),A(2),Tonnage(2),Load_time(2):
$ATTRIBUTES:TYPE,CAPACITY:
$SWITCHES:SW;;TRKS:
$VARIABLES: Tonnage 1;RUN.TIME;Tonnage(1):
                 Tonnage 2;RUN.TIME;Tonnage(2):
$BEGIN:
  ARIV   *S;/L/LIM=1:
         *B;TERM;;SW=ON%:       !Активизация очереди TRKS
  TRKS   *Q:                    !См. $INITIAL-ENTRIES
  LOADS  *F;;Load_time(TYPE):
  TRIP   *A;UN(30,60):          !Доставка груза потребителю
         *B;TRKS;;              !Вычисление тоннажа
                Tonnage(TYPE)=Tonnage(TYPE)+CAPACITY%;
$END:
$RUN-LENGTH=1440:
$INITIAL-ENTRIES: 1-3/TRKS/(3)1,20;  !3 20-т. грузовика
                            (2)2,30:  !2 30-т. грузовика
$ARRAYS:
     Load_time;1-3/NS/10,15: !Время загрузки грузовиков
$STOP
```

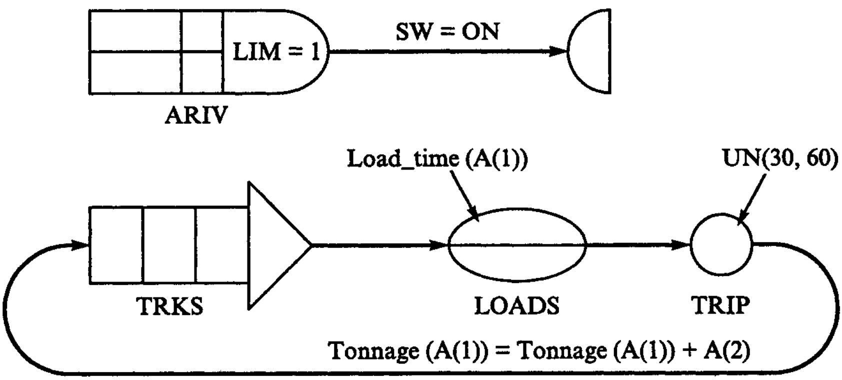


Рис. Б.17

Оператор $INITIAL-ENTRIES помещает три 20-тонных и два 30-тонных грузовика в очередь TRKS (Грузовики). Каждая транзакция, соответствующая этим грузовикам, имеет два атрибута: A(1) равен 1 или 2 в зависимости от типа грузовика (20- или 30-тонный), атрибут A(2) имеет значение грузоподъемности грузовика. Мы также используем имена TYPE (Тип) и CAPACITY (Грузоподъемность) для обозначения этих атрибутов.

Времена загрузки грузовиков (время обслуживания в узле LOADS (Погрузка)) приведены как значения массива Load_time (Время загрузки): Load_time(1) = 10, Load_time(2) = 15 соответственно для 20- и 30-тонных грузовиков. Поскольку атрибут A(1) соответствует типу грузовика, в модели используется выражение Load_time(A(1)) (точнее, Load_time(TYPE)) для определения времени загрузки. После загрузки грузовик следует к потребителю и обратно (дополнительный узел TRIP), после чего возвращается в очередь TRKS. Общий тоннаж использованных грузовиков каждого типа записывается в массив Tonnage(A(1)) и будет распечатан в виде значения переменной RUN.TIME. (Переменная RUN.TIME получает конечное значение только по окончании сеанса имитации.)

Для инициализации движения грузовиков используется источник ARIV, который продуцирует только одну транзакцию. Эта транзакция устанавливает переключатель SW в состояние ON (Включено), тем самым открывая выход очереди TRKS.

Упражнение Б.10,а

1. Пассажиры подходят к автобусной остановке в случайные моменты времени. Количество свободных мест в автобусе, прибывающем на остановку, также является случайной величиной. Время посадки одного пассажира в автобус составляет 7 секунд. Пусть время выполнения автобусного маршрута равно 30 минут (т.е. время между последовательными появлениями на остановке одного и того же автобуса составляет 30 минут). Предположим, что в настоящее время на маршруте находится только один автобус. Пассажиры согласны ждать автобус на остановке не более 20 минут, после чего уходят.

   Время между последовательными приходами пассажиров на остановку имеет следующую эмпирическую дискретную плотность вероятности.

Время (минуты)	2.0	2.5	3.0	3.5	4.0
Относительная частота	0.2	0.24	0.28	0.18	0.10

   Количество свободных мест в автобусе может быть 7, 8 или 9 с равными вероятностями. Определите процент пассажиров, не дождавшихся автобуса при времени имитации 480 минут.

Б.11. Заключение

В системе моделирования SIMNET II применяется всего четыре типа узлов, которые просты в изучении и использовании. Несмотря на свою простую структуру, язык SIMNET II является мощным инструментом для построения достаточно сложных имитационных моделей. Применение *специальных операторов присваивания* совместно со структурами IF-ENDIF значительно расширяет возможности языка как средства моделирования.

Полная версия системы SIMNET II включает средства отладки моделей и получения глобальных статистических результатов. Полная версия языка также включает такие важные элементы моделей, как ресурсы, согласования и сборки, а также средства, позволяющие моделировать периодически повторяющиеся события (см. [1]).

Литература

1. Taha H. *Simulation with SIMNET II*, 2nd ed., SimTec Inc., Fayetteville, AR, 1995.

Инсталляция и выполнение программ TORA и SIMNET II

Обе программы, TORA и SIMNET II, выполняются на IBM PC или совместимых компьютерах под управлением операционной системы MS-DOS 3.2 или выше.[1] Программа TORA также может выполняться с гибкого диска. Программа SIMNET II должна запускаться с жесткого диска. Обе программы требуют не более 512 Кбайт оперативной памяти.

В.1. Инсталляция и выполнение

Обе программы находятся на Web-узле Издательского дома "Вильямс" по адресу: www.williamspublishing.com и имеют два каталога — Tora и Simnet. С помощью средств своего Web-обозревателя следует переписать эти каталоги на жесткий диск своего компьютера.

Инструкции по выполнению этих программ приведены в ASCII-файлах README.DOC (каждая программа имеет свой файл README). Программа TORA управляется с помощью понятного и простого меню и не требует специального "Руководства пользователя". Введение в систему SIMNET II приведено в Приложении Б. Обе системы имеют одинаковую рабочую среду, позволяющую редактировать и выполнять модели, созданные с помощью TORA и SIMNET II. Примеры решения задач посредством программы TORA широко использованы в данной книге.

[1] Конечно, эти программы выполняются и под управлением операционных систем Windows 9X в отдельном окне DOS. — *Прим. ред.*

Статистические таблицы

ТАБЛИЦА

Функция нормального распределения $F(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-t^2/2}dt$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998									
4.0	0.99997									
5.0	0.9999997									
6.0	0.999999999									

Источник: Miller I. and Freund J. *Probability and Statistics for Engineers*, Prentice Hall, Upper Saddle River, N.J., 1985.

Таблица Г.2

Процентные точки распределения Стьюдента $t_{\alpha,\nu}$

ν	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	ν
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16

Окончание табл. Г.2

ν	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	ν
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
∞	1.282	1.645	1.960	2.326	2.576	∞

Таблица Г.3

Процентные точки распределения χ^2_ν

ν	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	ν
1	0.0000393	0.000157	0.000982	0.00393	3.841	5.024	6.635	7.879	1
2	0.0100	0.0201	0.0506	0.103	5.991	7.378	9.210	10.597	2
3	0.0717	0.115	0.216	0.352	7.815	9.348	11.345	12.838	3
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860	4
5	0.412	0.554	0.831	1.145	11.070	12.832	15.056	16.750	5
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548	6
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278	7
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955	8
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589	9
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188	10
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757	11
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300	12
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819	13
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319	14
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801	15
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267	16
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718	17

Окончание табл. Г.3

ν	$\alpha = 0.995$	$\alpha = 0.99$	$\alpha = 0.975$	$\alpha = 0.95$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$	ν
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156	18
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582	19
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997	20
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401	21
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796	22
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181	23
24	9.886	10.856	12.401	13.484	36.415	39.364	42.980	45.558	24
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928	25
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290	26
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645	27
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993	28
29	13.121	14.256	16.047	17.708	42.557	45.772	49.588	52.336	29
30	13.787	14.953	16.791	18.493	43.773	46.979	50.892	53.672	30

Эта таблица с разрешения администрации *Biometrika* основана на табл. 8 из *Biometrika Tables for Statisticians*, Vol. 1.

Ответы к упражнениям[1]

Глава 2

Упражнения 2.2,a

1. a) $-x_1 + x_2 \geq 1$.
   e) $0.5x_1 - 0.5x_2 \geq 0$.
3. $s_1 = 4$ тонны, $s_2 = 0$ тонн.

Упражнения 2.3,a

3. a) Решение: (2.5, 1.75), $z = \$19.50$.
   b) Решение: (2, 2), $z = \$18$.
   c) Решение: (1, 2), $z = \$13$.
   d) Решение: (6, 0), $z = \$30$.
   e) Решения нет.
5. На игру отводится 4 часа, на учебу — 6 часов, $z = 14$.

Упражнения 2.3,b

3. $x_1 = x_2 = z = 0$. Решение не имеет смысла.

Упражнения 2.3,c

1. $s_1 = 2$ тонны, $s_2 = 1$ тонна.
3. a) Использовать 40 минут дополнительного времени, $z = \$1517.50$.
   b) Использовать дополнительное время не рекомендуется, $z = \$1513.50$.

Упражнения 2.4,a

1. a) $-1 \leq c_1/c_2 \leq 1.5$.
   b) $-1 \leq c_2/c_1 \leq 5/3$.
   c) $-1 \leq c_1/c_2 < \infty$, если $c_1 = 0$, и $-1/3 \leq c_2/c_1 < \infty$, если $c_2 = 0$.
3. a) Необходимо ежедневно иметь 100 банок колы A1 и 400 банок — B&K, $z = \$33$.
   b) $-\infty \leq c_1/c_2 \leq 1$.

[1] Автором представлены ответы не ко всем упражнениям, а преимущественно к упражнениям с нечетными номерами, но также далеко не ко всем. — *Прим. ред.*

5. a) Следует вложить \$66 667 в кредитование частных клиентов и \$133 333 в кредитование покупок автомобилей, z = \$20 067. Коэффициент возврата равен 10.033%.
   c) Решение остается оптимальным.

7. a) Следует ежедневно производить 500 упаковок сока и 6000 упаковок пасты, z = \$6300.
   b) $0 \le c_1/c_2 \le 3$.

Упражнения 2.4,b

1. a) 100 шляп первого типа и 200 шляп второго, z = \$1800.
   b) Стоимость возрастания производства на одну шляпу второго типа составляет \$4 в интервале (200, 500).
   c) Стоимость возрастания предельного спроса на одну шляпу первого типа составляет \$0 в интервале (100, 200).
   d) Стоимость возрастания предельного спроса на одну шляпу второго типа составляет \$1 в интервале (100, 400).

3. a) x_1 = 52.94 единицы, x_2 = 14.11 единицы, z = \$148.14.
   b) Стоимости единицы возрастания количества времени выполнения процессов 1, 2 и 3 составляют соответственно \$0.1294, \$0.1176 и \$0. Поэтому наибольший приоритет имеет процесс 1, следующий по значению приоритет у процесса 2.

5. a) 5128 тонн угля C1 и 10256 тонн угля C2, z = 153 846 (фунт/час).
   b) Ослабление на 1 фунт в час ограничения на количество выбрасываемых аэрозольных частиц приведет к увеличению на 7696.2 (фунт/час) количества вырабатываемой электроэнергии.

7. a) 480 рубашек и 840 блузок еженедельно.
   b) Необходимо увеличить количество рабочих на участке раскроя до 92 человек, швей — до 134 человек и на участке пакетирования — до 15 человек.
   c) Стоимости одного часа рабочего времени, затрачиваемого на раскрой, пошив и пакетирование, составляют соответственно 0.0247, 0.02867 и 0 долларов.

9. a) 50.88 штук радиоприемников HiFi-1 и 31.68 штук радиоприемников HiFi-2.
   b) Стоимости одного процента уменьшения времени профилактических работ для сборочных линий 1, 2 и 3 равны 0, 3 и 0 соответственно.

Упражнения 2.5,a

1. a) Двойственная цена равна \$0.5471 на 1 фунт увеличения пищевой добавки. Один фунт увеличения пищевой добавки состоит из 0.588 фунта кукурузной муки и 0.412 фунта соевой муки. Соответствующее возрастание значения целевой функции составит $0.3 \times 0.588 + 0.9 \times 0.412$ = \$0.5472.
   b) D_1 = 100 фунтов, требуется 529.41 фунта кукурузной муки и 370.59 фунта соевой муки, стоимость равна \$492.36.
   c) d_1 = \$0.1, d_2 = \$0.15, решение останется оптимальным.
   d) $-0.7 \le (0.3 + d_1)/(0.9 + d_2) \le 1$.

Упражнения 2.6,а

1. a) $x_1 + x_2 + x_3 + (x_4 + x_5) \le 12$ и $(x_4 + x_5) \ge 4.8$.

   b) D_2 = \$7.2 млн. *Способ 1.* Двойственная цена равна –0.0084 долл. на \$1 инвестиций при $0 \le D_2 \le 7.2$, новое значение целевой функции z = \$0.936 млн. *Способ 2.* Имея значение D_2 = \$7.2 млн, вычисляем величины сельскохозяйственных и коммерческих кредитов, а затем значение целевой функции z = \$0.936 млн.

   c) Значения D_1 = \$8 млн и D_2 = \$4.2 млн приводят к допустимому решению: \$11 млн вкладывается в кредитование жилья, \$9 млн составляют коммерческие кредиты, что даст \$1.6524 млн дохода.

3. a) (i) $D_2 = 4$, минимальное количество автобусов равно 26. (ii) $D_3 = -3$, минимальное количество автобусов — 23.

   b) $D_1 = 2$, $D_2 = 1$, $D_3 = 2$, $D_4 = 0$, $D_5 = 3$ и $D_6 = 2$. Новое решение: $x_1 = 6$, $x_2 = 12$, $x_4 = 9$ и $x_5 = 6$, всего 33 автобуса.

5. a) Введем обозначения: x_{11} — количество обычного бензина, вырабатываемого на основе бензинового полуфабриката, x_{12} — количество высококачественного бензина, вырабатываемого на основе бензинового полуфабриката, x_{21} — количество обычного бензина, вырабатываемого на основе дистиллята, x_{12} — количество высококачественного бензина, вырабатываемого на основе дистиллята (единицы измерения — баррель в день). Математическая модель:

   $$\text{Максимизировать } z = 7.7x_{11} + 12.3x_{12} + 5.2x_{21} + 10.4x_{22}$$

   при выполнении условий

   $$5(x_{11} + x_{12} + x_{21} + x_{22}) \le 600\,000,$$
   $$x_{21} + x_{22} \le 40\,000,$$
   $$x_{11} + x_{21} \le 80\,000,$$
   $$x_{12} + x_{22} \le 50\,000,$$
   $$x_{ij} \ge 0 \text{ для всех } i \text{ и } j.$$

   Оптимальное решение: x_{11} = 70 000 баррелей в день, x_{12} = 50 000 баррелей в день, доход равен \$1 154 000.[1]

   b) D_1 = 50 000 баррелей в день, двойственная цена равна \$1.54 за баррель при выполнении неравенств $-350\,000 \le D_1 \le 50\,000$. Увеличение дохода составляет \$77 000 в день. Одноразовое вложение (\$3 500 000) плюс ежедневные издержки на поддержание новой производительности перегонной колонны (\$15 000 ежедневно) будут превышены за счет увеличения прибыли в течение 57 дней. Рекомендуется реализовать данную возможность.

7. a) Пусть x_i — выполненная часть проекта i. Решение: $x_1 = x_2 = x_3 = x_4 = 1$, $x_5 = 0.8413$, z = \$116 061.10.

   b) В формулировку задачи добавляется новое ограничение $x_2 \le x_6$. Решение: $x_5 = 0.0317$, остальные переменные равны 1, z = \$113 677.80.

---

[1] Здесь и далее, если не указаны значения каких-либо переменных, они равны нулю. — *Прим. ред*

c) Каждое дополнительное вложение в \$1000 приносит \$2888.90 дохода, что составляет 188.89% прибыли.

d) Пусть s_i — сумма, оставшаяся в конце года i. Преобразуем все ограничения задачи в равенства, используя для этого переменную s_1 в первом ограничении, разность $s_2 - s_1$ во втором ограничении, $s_3 - s_2$ — в третьем и $s_4 - s_3$ — в четвертом. Оптимальное решение зависит от всех шести проектов с z = \$131 300.

e) Решение такое же, как и в предыдущем пункте. Займы на третьем и четвертом годах составляют \$11 100, процент прибыли равен 137.3%.

9. a) Обозначим через x_i сумму инвестиций в проект i, $i = 1, 2, 3, 4$, через y_j — сумму денег, положенную в банк в j-м году, $j = 1, 2, ..., 5$.

   Решение: в первый год вложить \$10 000 в проект 2, во второй год инвестировать \$6000 в проект 3, в третий положить \$6800 в банк, в четвертый положить в банк \$33 642. На начало 5-го года получим y_5 = \$53 628.74.

   b) Доходность инвестиций составляет 536.61.

   c) Сумма, которая будет получена в конце 5-го года, уменьшится на \$3 730.90.

   d) $D_1 = 0$, $D_2 = D_3 = D_4 = D_5$ = \$1000. Решение: x_2 = \$10 000, x_3 = \$5000, y_3 = \$5000, y_4 = \$28 825 и w = \$46 998.64.

11. a) Следует производить 28 единиц изделия 1 и 148 единиц изделия 2. Доход составит \$5 280.

   b) Не более \$6.

   c) Стоимость машинного времени не должна превышать \$20 на одну единицу изделия 3.

   d) $10 \le d_2 \le 22.50$.

   e) $-13.33 \le D_2 \le 445$.

13. a) Уровень долевого участия внешних партнеров составит 0.7113 в первом проекте, во втором проекте они не участвуют.

   b) Нет.

   c) Процентные ставки: $i_1 = 0.02$, $i_2 = 0.025$, $i_3 = 0.171$, $i_4 = 0.02$.

   d) Искомая двойственная цена = 1.1945 – 1.025 × 1.02 = \$0.149.

15. a) Следует вложить \$100 000 в проект A в первом году и \$170 000 в проект B во втором году.

   b) Один доллар инвестиций приносит \$5.10 в конце срока инвестирования.

17. a) Следует производить 1800 тонн сплава A и 1000 тонн сплава B.

   b) Для производства сплава A требуется 1000 тонн руды первого типа и 3000 тонн руды третьего типа, для производства сплава B необходимо 2000 тонн руды второго типа.

   c) Наиболее неблагоприятны для оптимального решения ограничения, которым соответствуют отрицательные двойственные цены.

   d) Максимальные допустимые цены на руду трех типов составляют \$90, \$110 и \$30 соответственно.

Глава 3

Упражнения 3.2,a

1. Максимизировать $z = 2x_1^+ - 2x_1^- + 3x_2 + 5x_3$

   при ограничениях

$$-x_1^+ + x_1^- - x_2 + x_3 = 5,$$
$$-6x_1^+ + 6x_1^- + 7x_2 - 9x_3 - x_4 = 4,$$
$$x_1^+ - x_1^- + x_2 + 4x_3 + x_5 = 10,$$

все переменные неотрицательные.

3. Минимизировать $z = y$

   при ограничениях

$$-y + x_1 - x_2 + 3x_3 + s_1 = 0,$$
$$-y - x_1 + x_2 - 3x_3 + s_2 = 0,$$
$$-y - x_1 + 3x_2 - x_3 + s_3 = 0,$$
$$-y + x_1 - 3x_2 + x_3 + s_4 = 0,$$

все переменные неотрицательные.

Упражнения 3.2,b

1. Базисное решение $(x_1, x_4) = (0, 4)$ допустимо, решение $(x_1, x_5) = (-6, 14/3)$ недопустимо, решение $(x_2, x_3) = (0, 2)$ допустимо.
3. a) Оптимальное решение: $x_1 = 8, x_2 = 0, x_3 = 3, x_4 = 0, z = 31$.
   b) Альтернативные оптимальные решения: $(x_1, x_2, x_3, x_4) = (4, 0, 0, 0)$ и $(x_1, x_2, x_3, x_4) = (0, 2, 0, 0)$, $z = 4$.

Упражнения 3.2,c

1. Максимизировать $z = 2x_1 + 5x_2 + 3x_3 - 15s_1^- - 10s_2^-$

   при ограничениях

$$2x_1 + x_2 + 2x_3 + s_1^+ - s_1^- = 80,$$
$$x_1 + x_2 + 0.5x_3 + s_2^+ - s_2^- = 65,$$

все переменные неотрицательные.

   Оптимальное решение: $x_1 = 0, x_2 = 60, x_3 = 10, z = 330$.

Упражнения 3.3,a

3. a) (1) x_8, (2) x_3, (3) x_1, (4) Нет, (5) x_3.
   b) (1) 20, (2) –8, (3) 0, (4) ∞, (5) 0.

5. a) Задача эквивалентна следующей: максимизировать $z = x_1$ при ограничениях $5x_1 \le 4$, $6x_1 \le 8$, $3x_1 \le 3$. Ограничения удовлетворяются при $x_1 \le 4/5$. Последнее число совпадает с минимальным отношением, которое получается, если рассматривать переменную x_1 в качестве вводимой в базис, а переменные x_2 и x_4 составляют начальное базисное решение. Следовательно, максимальное значение целевой функции равно $z = 0.8$.

7. a) Оптимум достигается в точке $E = (5/2, 2)$, $z = 39/2$.
   b) $A \to G \to F \to E$.
   c) Вводимая переменная x_1; значения отношений: 2, 3, 5; $A \to B$; $\Delta z = 8$.
   d) Вводимая переменная x_2; значения отношений: 1, 2, 3; $A \to G$; $\Delta z = 4$.

9. Точка A: базисные переменные (s_1, s_2, s_3, s_4), небазисные переменные (x_1, x_2, x_3).
   Точка J: базисные переменные (x_1, s_2, x_2, x_3), небазисные переменные (s_1, s_2, s_4).

Упражнения 3.4,a

3. a) $z - (5 - 2M)x_1 - (6 + 3M)x_2 = -3M$.
   c) $z - (3 - 4M)x_1 - (6 - 8M)x_2 - Ms_5 = 5M$.

5. Начальное уравнение для z-строки: $z - x_1 - 12x_2 + 0x_3 + 0x_4 = -8$.

7. Начальное уравнение для z-строки: $z + (2 - 2M)x_1 + (1 + M)x_2 + 0x_3 + 0R = 9 - 4M$.

Упражнения 3.4,b

1. a) Сумма значений искусственных переменных — это мера "недопустимости" решения.
   b) Сумма значений искусственных переменных минимизируется в задачах ЛП любого типа.

3. a) На первом этапе после 4 итераций имеем $x_1 = 45/7$ и $x_2 = 4/7$. На втором этапе после двух итераций приходим к такому же решению.

5. a) На первом этапе после одной итерации получаем $x_2 = 2$ и $R_2 = 0$.

7. Если переменные x_1, x_3 или x_5 принимают положительные значения, тогда целевая функция в конце первого этапа обязательно принимает положительное значение (проверьте!). Это указывает на то, что задача не имеет допустимого решения.

Упражнения 3.5,a

1. a) $A \to B \to C \to D$.
   b) В точке A — одна итерация, в точке B — одна итерация, в C — три, в D — тоже одна.

Упражнения 3.5,b

1. Альтернативные оптимальные базисные решения: $(x_1, x_2, x_3) =$ (0, 0, 10/3), (0, 5, 0) и (1, 4, 1/3). $\bar{x}_1 = \alpha_3$, $\bar{x}_2 = 5\alpha_2 + 4\alpha_3$, $\bar{x}_3 = (10/3)\alpha_1 + (1/3)\alpha_3$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$, все $\alpha_i \ge 0$.

3. В оптимальной симплекс-таблице дополнительная переменная, ассоциируемая с третьим ограничением, является базисной и имеет нулевое значение (вырожденность). Дополнительная переменная, ассоциируемая со вторым ограничением, является небазисной и имеет нулевой коэффициент в уравнении z-строки (альтернативное решение).

Упражнения 3.5,d

1. Это требование удовлетворить нельзя — можно произвести не более 275 единиц изделий.

Глава 4

Упражнения 4.2,a

1. Максимизировать $w = 10y_1 + 8y_2$ при ограничениях $y_1 + 2y_2 \le 5$, $2y_1 - y_2 \le 12$, $y_1 + 3y_2 \le 4$, $y_1 \le 0$, y_2 не имеет ограничения в знаке.
5. c) Минимизировать $w = 5y_1 + 3y_2$ при ограничениях $y_1 - y_2 = 5$, $2y_1 + 5y_2 + y_3 \ge 6$, y_1 не имеет ограничения в знаке, $y_2 \le 0$, $y_3 \ge 0$.

Упражнения 4.3,a

1. Из уравнений $2y_1 - y_2 - 12 = 0$ и $y_1 + 3y_2 - 4 = 3/5$ (порождаемых соотношениями двойственности) получаем $y_1 = 29/5$ и $y_2 = -2/5$.
3. a) Минимизировать $w = 3y_1 + 4y_2$ при ограничениях $y_1 + 2y_2 \ge 1$, $2y_1 - y_2 \ge 5$, $y_1 \ge 3$, y_2 не имеет ограничения в знаке.
   b) $y_1 = 3, y_2 = -1, w = 5$.
5. $y_1 = 4, y_2 = 0, w = 16$.
9. a) Недопустимое решение.
   b) Допустимое, но не оптимальное решение.
   c) Допустимое и оптимальное решение.

Упражнения 4.4,a

1. a) –1500 долл.
   b) –750 долл.
   c) 0 долл.
3. Компания не может платить более \$11.88 за 1 кв. метр кожи и \$21.25 за 1 час дополнительных работ.

Упражнения 4.4,b

1. Время выполнения первой операции необходимо сократить не менее чем на 50%.
3. Изделия типа PP3 и PP4 не производятся. Уменьшение дохода в случае увеличения их производства на единицу составляет соответственно \$0.1429 и \$1.1429.

Упражнения 4.5,a

1. a) a) Нет, поскольку в двойственном симплекс-методе промежуточные решения будут недопустимыми, пока не будет достигнуто оптимальное решение.
   b) $L \to I \to F$.
3. a) Оптимальное решение: $x_1 = 3/4$, $x_2 = 1/4$, $z = 7/2$. Вычисления начинаются в точке (0, 0), проходят через точку (2/3, 0), заканчиваются в точке (3/4, 1/4).
5. a) Вводится искусственное ограничение $x_3 \le M$. Оптимальное решение: $x_1 = 56/9$, $x_2 = 26/3$, $x_3 = 14/9$, $z = 28/9$.

Упражнения 4.6,a

3. a) $\mathbf{X}_B = (3,\ 15)^T$, $z_1 - c_1 = 0$, $z_3 - c_3 = 2$. Базисное решение оптимально и допустимо.
   b) Базисное решение оптимально, но недопустимо.
5. a) Базисное решение $\mathbf{X}_B = (x_2, x_1)^T = (18/5,\ 14/5)^T$, при котором $z = 57.2$.
   b) Решение оптимально, поскольку $z_3 - c_3 = 3/5$, $z_4 - c_4 = 29/5$.
7. a) $b_1 = 30$, $b_2 = 40$.
   b) $y_1 = 5$, $y_2 = 0$.
   c) $a = 23$, $b = 5$, $c = -10$.

Упражнения 4.7,a

1. Новое решение: $(x_1, x_2, x_3) = (0, 95, 240)$, $z = \$1390$.
3. $x_1 = 10/3$, $x_2 = 2$, $z = 74/4$.

Упражнения 4.7,b

1. a) $-20 \le D_2 \le 400$, $D_3 \ge -20$.
   b) Стоимость одной минуты рабочего времени второй операции составляет \$2, третьей операции — \$0.
   c) Новое решение: $x_1 = 0$, $x_2 = 90$, $x_3 = 250$. Значение целевой функции возрастет на $40 \times \$2 = \$80$.
   d) Новое решение: $x_1 = 0$, $x_2 = 100$, $x_3 = 230$. Изменение значения целевой функции составит $30 \times \$0 = \$0$.
   e) Новое решение: $x_1 = 0$, $x_2 = 90$, $x_3 = 230$, $z = \$1330$.
3. a) Доход за 1 час рабочего времени составляет $60 \times \$1 = \$60$, что превышает стоимость одного часа дополнительного времени (= \$50). Данное заключение справедливо, если количество времени, выделяемого для первой сборочной операции, находится в интервале от 3.833 до 7.333 часов. Это означает, что в данном случае можно использовать только 10 минут дополнительного времени. Ситуацию, когда дополнительное время для первой операции превысит 10 минут, необходимо рассматривать отдельно.
   b) Стоимость двух часов сверхурочной работы на второй сборочной операции составляет \$110. Дополнительный доход от двух часов сверхурочной работы равен $120 \times \$2 = \$240$. Следовательно, экономически выгодно использовать на второй сборочной операции дополнительное время.

c) Дополнительное время на второй сборочной операции использовать не целесообразно, так как стоимость ресурса здесь равна нулю.

d) Новое оптимальное решение: $x_1 = 0$, $x_2 = 105$, $x_3 = 230$. Значение целевой функции увеличится на \$10.

e) В данной ситуации доход будет уменьшен на \$30, уменьшение стоимости составит \$7.50. Поэтому уменьшение фонда рабочего времени для второй операции экономически нецелесообразно.

Упражнения 4.7,c

1. a) Новое ограничение избыточно.

   b) Новое решение: $x_1 = 0$, $x_2 = 88$, $x_3 = 230$, $z = \$1326$.

Упражнения 4.7,d

1. a) Решение остается оптимальным.

   b) Новое решение: $x_1 = 10$, $x_2 = 102.5$, $x_3 = 215$, $z = \$860$.

3. a) $d_1 \leq 4$.

   b) $2 \leq d_2 \leq 8$.

   c) $d_3 \geq -8/3$.

5. a) $-3 \leq d_1 \leq 1$.

   b) $-2/3 \leq d_2 \leq 6$.

7. a) $4 - 0.5d_1 - 0.25d_2 + 1.5d_3 \geq 0$, $\quad 1 + 0.5d_2 \geq 0$, $\quad 2 - 0.25d_2 + 0.5d_3 \geq 0$.

9. a) Решение не изменится.

   b) Новое решение: $x_1 = 20$, $x_2 = x_3 = 0$, $z = \$1400$.

   c) Решение не изменится.

11. a) $\$6 \leq c_1 \leq \$26$.

    b) Новое решение: $x_1 = 0$, $x_2 = 165$, $x_3 = 10$, $z = \$4105$.

Упражнения 4.7,e

1. 42.86%.

3. a) Производство новой модели (модели пожарной машины) экономически не выгодно.

   b) Следует произвести 100 единиц модели пожарной машины.

Глава 5

Упражнения 5.1,a

1. a) Ложно.

   b) Истинно.

   c) Истинно.

3. Спрос распределительных центров в Денвере и Майами не будет удовлетворен соответственно на 150 и 50 автомобилей.

5. Оптимальная схема поставок: из Лос-Анджелеса в Денвер 1000 автомобилей, из Детройта в Денвер — 1100 автомобилей, из Детройта в Майами — 200 автомобилей и из Нового Орлеана в Майами — 1200 машин. Распределительный центр в Денвере недополучит 200 автомобилей. Общая стоимость перевозок равна $333 200.

7. c) Город 1 восполнит недостаток электроэнергии в 22.5 млн кВт/ч из альтернативной электросети за $22 500.

9. b) Бензохранилище 1 получит 4 млн галлонов бензина от завода 1. Бензохранилище 2 получит 2 млн галлонов бензина от завода 1, 5 млн галлонов бензина — от завода 2 и 1 млн галлонов бензина — от завода 3. Бензохранилище 3 получит 7 млн галлонов бензина от завода 3. Общая стоимость транспортировки равна $243 000.

11. b) Бензохранилище 1 получит 4 млн галлонов бензина от завода 1. Бензохранилище 2 получит 2 млн галлонов от завода 1, 5 млн галлонов бензина — от завода 2 и 1 млн галлонов бензина — от завода 3. Бензохранилище 3 получит 4 млн галлонов бензина от завода 3. Завод 3 имеет избыток 3 млн галлонов бензина. Общая стоимость транспортировки равна $207 000.

Упражнения 5.2,a

1. Оптимальный план: в первый месяц следует произвести 50 единиц продукции; во второй месяц — 180 единиц продукции, из которых 50 единиц предназначены для погашения задолжности первого месяца; в третий месяц — 280 единиц продукции, из которых 70 единиц предназначены для погашения задолжности второго месяца и 30 — для удовлетворения заявки четвертого месяца; в четвертый — 270 единиц продукции (для покрытия заявки четвертого месяца). Общая стоимость произведенной продукции равна $31 461.

3. Понедельник: покупается 24 полотна, 12 полотен отправляется на ночную заточку и 12 — на 2-дневную. Вторник: 6 полотен — на ночную заточку и 6 — на 2-дневную. Среда: покупается 8 полотен, 8 полотен отправляется на ночную заточку и 6 — на 2-дневную. Четверг: 12 полотен — на ночную заточку и 8 — на 2-дневную. Пятница: 8 полотен отправляется на ночную заточку и 10 — в "остаток". Суббота: 14 полотен — на ночную заточку. Воскресенье: 22 полотна — в "остаток". Общая стоимость полотен равна $840.

5. Оптимальное решение: в первый месяц следует произвести 500 тонн продукции, из которых 100 тонн предназначены для покрытия спроса второго месяца; во второй месяц — 600 тонн продукции, из которых 200 тонн предназначено на покрытие спроса третьего месяца и 180 тонн — на покрытие спроса четвертого; в третий месяц следует произвести 200 тонн продукции; в четвертый также будет произведено 200 тонн продукции. Общая стоимость производства равна $190 040.

7. Первый месяц: покупка 200 деталей, 12 отработанных деталей отправляется в местную мастерскую, 188 — в центральную. Второй месяц: покупка 180 деталей, 148 отработанных деталей отправляется в местную мастерскую, 32 — в централь-

ную. Третий месяц: покупка 140 деталей, 10 отработанных деталей отправляется в местную мастерскую, 290 — в центральную. Четвертый месяц: 198 отработанных деталей отправляются в местную мастерскую. Пятый месяц: 198 отработанных деталей отправляется в "остаток". Шестой месяц: 290 отработанных деталей отправляется в "остаток".

Упражнения 5.3,a

1. a) Метод северо-западного угла: $x_{11} = 5$, $x_{12} = 1$, $x_{22} = 4$, $x_{23} = 3$, $x_{33} = 7$, стоимость = \$42. Метод наименьшей стоимости: $x_{11} = 5$, $x_{13} = 1$, $x_{22} = 5$, $x_{23} = 2$, $x_{33} = 7$, стоимость = \$37. Метод Фогеля: такое же начальное решение, как и в методе наименьшей стоимости.

Упражнения 5.3,b

1. a) После трех итераций имеем $x_{11} = 1$, $x_{13} = 5$, $x_{21} = 4$, $x_{22} = 5$, $x_{33} = 5$, стоимость равна \$33.

3. $x_{12} = 10$, $x_{21} = 20$, $x_{22} = 10$, $x_{23} = 50$, $x_{31} = 5$, $x_{32} = 10$, стоимость равна \$515.

5. a) \$1475.

   b) $c_{12} \geq \$3$, $c_{13} \geq \$8$, $c_{23} \geq \$13$, $c_{31} \geq \$7$.

Упражнения 5.3,c

1. $u_1 = 0$, $u_2 = 5$, $u_3 = 7$, $v_1 = -3$, $v_2 = 2$, $v_3 = 4$, $v_4 = 11$, целевая функция = \$435.

Упражнения 5.4,a

1. a) 1–5, 2–3, 3–2, 4–4, 5–1, стоимость = \$21.

3. Новый рабочий должен заменить третьего рабочего.

5. Билет 1: отправка из Далласа 3 июня, возвращение 28 июня. Билет 2: отправка из Атланты 7 июня, возвращение 10 июня. Билет 3: отправка из Атланты 12 июня, возвращение 17 июня. Билет 4: отправка из Атланты 21 июня, возвращение 25 июня. Стоимость билетов \$1180. Задача имеет альтернативные решения.

Упражнения 5.5,a

1. a) $x_{13} = 100$, $x_{24} = 200$, $x_{35} = 100$, $x_{45} = 200$, $x_{56} = 150$, стоимость равна \$2250. Задача имеет альтернативное оптимальное решение.

3. a) a) $x_{15} = 900$, $x_{24} = 1400$, $x_{34} = 1000$, $x_{45} = 1300$, $x_{46} = 1100$, $x_{57} = 1000$, $x_{58} = 1200$, стоимость равна \$8640.

   b) $x_{15} = 900$, $x_{24} = 1400$, $x_{34} = 1000$, $x_{45} = 1060$, $x_{46} = 1100$, $x_{57} = 760$, $x_{58} = 1200$, стоимость равна \$8016. Дилер 7 недополучит 240 автомобилей.

5. $x_{12} = 50\,000$, $x_{37} = 60\,000$, $x_{75} = 60\,000$, $x_{54} = 20\,000$, $x_{56} = 40\,000$, $x_{62} = 40\,000$, стоимость равна 2 660 000 (галлон × миля).

Глава 6

Упражнения 6.2,a

1. a) 1 – 3 – 4 – 2.
   b) 1 – 5 – 4 – 3 – 1.
   c) 1 – 3 – 4 – 5 – 1.
   d) {1 – 3, 3 – 4, 3 – 5}.
   e) {1 – 3, 3 – 4, 4 – 2, 2 – 5}.

Упражнения 6.3,a

3. (SE–LA), (LA–CH), (CH–NY), (NY–DC), (DE–DA), (DA–CH).
5. (1–2), (2–3), (3–4), (4–6), (1–5), (5–7), (5–9), (9–8).

Упражнения 6.4,a

1. Замена автомобилей должна начаться в 2003 году, новые автомобили будут эксплуатироваться до 2006 года.
3. Минимальное время равно 106 секунд.

Упражнения 6.4,b ·

1. a) 1–3–6–8, 1–2–3–6–8, 1–3–5–6–8 или 1–2–3–5–6–8, длина всех маршрутов равна 8.
   c) 4–5–6–8 или 4–6–8, длина обоих маршрутов равна 8.
3. Замена автомобилей должна начаться в 2003 году, новые автомобили будут эксплуатироваться до 2006 года. Общая стоимость равна $9800.
5. План производства: в первый месяц производится 100 единиц продукции, во второй — 140 единиц, в третий — 390 единиц. Общая стоимость производства равна $7296.

Упражнения 6.4,c

1. a) 5–4–2–1, длина маршрута равна 12.
   d) 5–4–2, длина маршрута равна 9.
3. Связь между районами 1 и 2: маршрут 1–3–2, длина маршрута = 500 миль.
   Связь между районами 1 и 4: маршрут 1–3–2–4, длина маршрута = 700 миль.
   Связь между районами 1 и 5: маршрут 1–3–5, длина маршрута = 800 миль.

Упражнения 6.5,a

1. Разрез: (1, 2), (1, 4), (3, 4), (3, 5), пропускная способность разреза равна 60.
   Разрез: (1, 2), (1, 3), (4, 3), (4, 5), пропускная способность разреза равна 75.

Упражнения 6.5,b

1. a) Величина неиспользованных пропускных способностей через дугу (2, 3) равна 40, через дугу (2, 5) — 10, через дугу (4, 3) — 5, через остальные дуги равна нулю.

b) Величины потоков, проходящих через узлы 2, 3 и 4, равны соответственно 20, 30 и 20 единиц.

c) Нет, поскольку в этой сети "узким местом" являются дуги, исходящие из узла 1.

3. Максимальный поток равен 25.

5. Максимальный поток равен 100, через насосную станцию 4 проходит 30 млн баррелей в день, через насосную станцию 5 — 40 млн, а через насосную станцию 6 — 60.

7. Максимальный поток равен 170.

9. Максимальное количество произведенных игрушек равно 800.

Упражнения 6.6,b

1. a) Минимизировать $z = x_{12} + 5x_{13} + 3x_{24} + 4x_{32} + 6x_{34}$

при ограничениях

$$x_{12} + x_{13} = 50,$$
$$-x_{12} + x_{24} - x_{32} = -40,$$
$$-x_{13} + x_{32} + x_{34} = 20,$$
$$-x_{24} - x_{34} = -30,$$
$$30 \le x_{13} \le 40,\ x_{24} \ge 10,\ x_{32} \ge 10.$$

b) Минимизировать $z = x_{12} + 5x_{13} + 3x_{24} + 4x_{32} + 6x_{34}$

при ограничениях

$$x_{12} + x_{13} = 20,$$
$$-x_{12} + x_{24} - x_{32} = -40,$$
$$-x_{13} + x_{32} + x_{34} = 40,$$
$$-x_{24} - x_{34} = -20,$$
$$x_{13} \le 10.$$

2. $x_{13} = 20,\ x_{15} = 80,\ x_{23} = 20,\ x_{45} = 90.$

Упражнения 6.6,c

1. Следует произвести 210 единиц продукции на первом этапе и 220 единиц — на третьем. Общая стоимость производства составит \$9 895.

3. Необходимо произвести 110 единиц продукции на первом этапе, 95 — на втором этапе, 125 — на третьем и 100 единиц — на четвертом. Следует перенести 10 единиц продукции, произведенной на первом этапе, в счет второго этапа, 25 единиц продукции, произведенной на третьем этапе, в счет четвертого этапа, при этом 5 единиц продукции, произведенной на третьем этапе, возвращается в счет недопоставки второго периода. Общая стоимость производства составит \$10 177.50.

5. Школа 1 принимает 450 учащихся из второй общины национальных меньшинств и 1000 из первой "обычной" общины. Школа 2 принимает 500 учащихся из первой общины национальных меньшинств, 300 человек из третьей общины националь-

ных меньшинств и 1000 из второй "обычной" общины. Значение целевой функции, определяемой как произведение количества учащихся на расстояние от их местожительства до школы, равно 24 300.

7. $x_{16} = 8$, $x_{24} = 10$, $x_{34} = 6$, $x_{35} = 6$, $x_{36} = 6$, стоимость = \$146.

Упражнения 6.7,b

1. 1–3–4–5–6–7, длительность проекта = 19.
3. *A–C–D–F–G–H–J–L–N–S–T*, длительность проекта = 38 дней.
4. Существует два критических пути: *A–C–E–F–J–L–*(*M* или *N*)*–P–Q–S–T–U*, длительность проекта = 22.1 дня.

Упражнения 6.7,c

1. $\square_i + D_{ij}$ и $\Delta_j - D_{ij}$.
3. a) 10.
   b) 5.
   c) 0.
5. a) B–F, длительность проекта = 45 дней.
   b) Процесс A: $TF = 12$, $FF = 0$. Процесс C: $TF = FF = 12$. Процесс D: $TF = 20$, $FF = 0$. Процесс E: $TF = 11$, $FF = 0$. Процесс G: $TF = FF = 11$. Процесс H: $TF = FF = 31$. Процессы A, D и E помечены "красными флажками".
   c) Начала процессов C, D, G и H задерживаются на 5 дней. Процесс E не может начаться раньше 20-го дня.
   d) Две единицы этого оборудования.

Глава 7

Упражнения 7.2,a

1. Максимизировать $z = (2, 3, 0, 0, -M)\,(x_1, x_2, ..., x_6)^T$
   при ограничениях

$$\begin{bmatrix} 1 & 2 & 0 & | & 1 & 0 & 0 \\ 1 & 2 & 0 & | & 0 & 1 & 0 \\ 5 & -2 & -1 & | & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_6 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix}.$$

Упражнения 7.2,c

1. Уравнение $\mathbf{P}_1 x_1 + \mathbf{P}_2 x_2 = \mathbf{b}$ имеет (1) единственное решение, если векторы $\mathbf{P}_1$ и $\mathbf{P}_2$ независимы; (2) бесконечно много решений, если векторы $\mathbf{P}_1$, $\mathbf{P}_2$ и $\mathbf{b}$ линейно зависимы; (3) не имеет решения, когда векторы $\mathbf{P}_1$ и $\mathbf{P}_2$ зависимы, а вектор $\mathbf{b}$ независим.
3. a) Истинно.
   b) Истинно.
   c) Истинно.

Упражнения 7.3,a

3. $(x_1, x_2) = \lambda_1(0, 0) + \lambda_2(2, 0) + \lambda_3(0, 2) = (2\lambda_2 + 2\lambda_3)$, $\lambda_1 + \lambda_2 + \lambda_3 = 1$, $\lambda_i \geq 0$, $i = 1, 2, 3$.

Упражнения 7.3,b

1. b) Одна крайняя точка, у которой координата x_3 положительна, остальные координаты равны нулю.
   c) Нет, поскольку переменная x_4 должна иметь отрицательное значение.

Упражнения 7.4,a

1. Поскольку $\det(\mathbf{P}_1, \mathbf{P}_2) = -6$, набор векторов $(\mathbf{P}_1, \mathbf{P}_2)$ образует базис. Так как $\det(\mathbf{P}_2, \mathbf{P}_3) = 0$, этот набор векторов не составляет базис. Набор векторов $(\mathbf{P}_3, \mathbf{P}_4)$ также образует базис, поскольку $\det(\mathbf{P}_3, \mathbf{P}_4) = 1$.

3. Поскольку $\det(\mathbf{P}_2, \mathbf{P}_3) = 0$, набор векторов $(\mathbf{P}_2, \mathbf{P}_3)$ не составляет базис и, следовательно, не может соответствовать никакой крайней точке пространства решений.

5.

Базис	x_1	x_2	x_3	x_4	x_5	Решение
z	0	0	–2/5	–1/5	0	12/5
x_1	1	0	–3/5	1/5	0	3/5
x_2	0	1	4/5	–3/5	0	6/5
x_5	0	0	–1	1	1	0

7. a) Необходимо исключить вектор $\mathbf{P}_1$.
   b) Да.

13. $n - m$.

15. Вырожденность решения связана с минимумом отношения, определяемого при вычислении вводимой в базис переменной. Вырожденность решения повторится на следующей итерации, если в симплекс-таблице в столбце вводимой переменной есть положительный коэффициент, соответствующий нулевой базисной переменной. Если в этом столбце все коэффициенты отрицательные или нулевые, то на следующей итерации вырожденность исчезнет.

17. a) Новое значение переменной x_j = (старое значение переменной x_j)/β.
    b) Новое значение переменной x_j = (α/β) × (старое значение переменной x_j).

Упражнения 7.5,a

1. $$\begin{bmatrix} 1 & -\frac{3}{4} & -\frac{1}{4} \\ 1 & \frac{1}{2} & -\frac{1}{2} \\ -1 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

3. Если ведущий элемент, в соответствии с которым строится вектор ξ, не равен нулю, тогда **B** является базисом.

Упражнения 7.5,b

1. a) После выполнения первой итерации имеем следующую симплекс-таблицу.

Базис	x_1	x_2	x_3	x_4	x_5	x_6	Решение
z	0	$-1/3$	0	$2/3$	0	0	8
x_3		$4/3$					2
x_1		$2/3$					4
x_5		$5/3$					5
x_6		1					2

3. Оптимальное решение: $x_1 = x_2 = x_6 = 0$, $x_3 = 2$, $x_4 = 10$, $x_5 = 8$, $z = -78$.

Упражнения 7:5,c

3. a) Оптимальное решение: $x_1 = 0$, $x_2 = 3/2$, $x_3 = 1$, $z = 6$.
   b) Оптимальное решение: $x_1 = 4$, $x_2 = 7/4$, $x_3 = 0$, $z = 83/4$.

Упражнения 7.5,d

1. a) $(x_1, x_2) = \alpha_1(0, 0) + \alpha_2(0, 1) + \alpha_3(1, 2) + \alpha_4(2, 2) + \alpha_5(10/3, 4/3) + \alpha_6(4, 0)$, $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 = 1$, все $\alpha_i \geq 0$.
   b) Пространства допустимых решений не существует.
   c) $(x_1, x_2) = \alpha_1(0, 0) + \alpha_2(10, 0) + \alpha_3(20, 10) + \alpha_4(20, M) + \alpha_5(0, M)$, M — достаточно большое положительное число, $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 = 1$, все $\alpha_i \geq 0$.

3. Подпространства: $(x_1, x_2) = (9/2\alpha_2 + 4\alpha_3,\ \alpha_3 + 2\alpha_4)$, $(x_3, x_4) = (4\beta_2 + 9\beta_3,\ \beta_3 + 10\beta_4)$. Оптимальное решение получается при $\alpha_3 = 1$ и $\beta_3 = 1$, при этом имеем $x_1 = 4$, $x_2 = 1$, $x_3 = 9$, $x_4 = 1$, $z = 54$.

5. Оптимальное решение: $x_1 = 2$, $x_2 = 8$, $x_3 = 0$, $x_4 = 12$, $x_5 = 28$, $x_6 = 0$, $z = 156$.

7. Оптимальное решение: $y_1 = 0$, $y_2 = 2$, $y_3 = 0$, $y_4 = 5$, $y_5 = 0$, $z = 44$.

Упражнения 7.6,b

3. a) Минимизировать $z = 2y_1 + 5y_2$ при ограничениях $2y_1 + y_2 \geq 5$, $-y_1 + 2y_2 \geq 12$, $3y_1 + y_2 \geq 4$, y_1 произвольного знака, $y_2 \geq 0$.
   b) i) Решение двойственной задачи не является допустимым, поэтому базис ($\mathbf{P}_4$, $\mathbf{P}_3$) не порождает оптимального решения. ii) Решение двойственной задачи не является допустимым, поэтому базис ($\mathbf{P}_2$, $\mathbf{P}_3$) не порождает оптимального решения. iii) Решение двойственной задачи является допустимым, поэтому базис ($\mathbf{P}_1$, $\mathbf{P}_2$) порождает оптимальное решение. iv) Решение двойственной задачи не является допустимым, поэтому базис ($\mathbf{P}_1$, $\mathbf{P}_4$) не порождает оптимального решения.

5. $z = 34$.

7. Минимизировать $w = \mathbf{Yb}$ при ограничениях $\mathbf{YA} = \mathbf{C}$, переменные $\mathbf{Y}$ произвольного знака.

Упражнения 7.7,a

1. $-2/7 \le t \le 1$.
3. $x_1 = 2/5, x_2 = 9/5, x_3 = 0, z = (17 - 29t)/5$.
5. $t_1 = 1$.

Упражнения 7.7,b

1. a) $t_1 = 10$, $\mathbf{B}_1 = (\mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4)$.
   b) $t_1 = 5$, $\mathbf{B}_1 = (\mathbf{P}_5, \mathbf{P}_3, \mathbf{P}_6)$.
3. Для $0 \le t \le 3/8$ имеем $x_1 = (3 + 7t)/5$, $x_2 = (6 - 11t)/5$, $z = (21 - 11t)/5$. Для $3/8 \le t \le 2/5$ имеем $x_1 = -3 + 11t$, $x_2 = 6 - 15t$, $z = 6 + 3t$.

Глава 8

Упражнения 8.2,a

1. Вводится еще одна целевая функция: минимизировать s_5^- при ограничении $55x_u + 3.5x_p - 0.0675x_б + s_5^+ - s_5^- = 0$.
3. Обозначим через x_1 количество первокурсников–жителей данного штата, через x_2 — число первокурсников–жителей других штатов, через x_3 — количество первокурсников–неамериканцев. Частные целевые функции:

$$\text{минимизировать } s_1^+,$$
$$\text{минимизировать } s_2^+,$$
$$\text{минимизировать } s_3^+,$$
$$\text{минимизировать } s_4^-,$$
$$\text{минимизировать } s_5^-,$$

при ограничениях

$$x_1 + x_2 + x_3 + s_1^+ - s_1^- = 1200,$$
$$2x_1 + x_2 - 2x_3 + s_2^+ - s_2^- = 0,$$
$$-x_1 - x_2 + 0.9x_3 + s_3^+ - s_3^- = 0,$$
$$0.2x_2 + (7/9)x_3 + s_4^+ - s_4^- = 0,$$
$$0.2x_1 - 0.8x_2 + 0.2x_3 + s_5^+ - s_5^- = 0,$$

все переменные неотрицательны.

5. Обозначим через x_j количество партий изделий, изготовленных в j-ю смену, $j = 1, 2, 3$. Получаем задачу: минимизировать $s_1^+ + s_1^-$ при ограничениях $-100x_1 + 40x_2 - 80x_3 + s_1^+ - s_1^- = 0$, $4 \le x_1 \le 5$, $10 \le x_2 \le 20$, $3 \le x_3 \le 5$.

7. Обозначим через x_1 и x_2 количество произведенных изделий 1 и 2 соответственно. Частные целевые функции:

$$\text{минимизировать } s_1^+,$$
$$\text{минимизировать } s_2^+,$$
$$\text{минимизировать } s_3^-,$$
$$\text{минимизировать } s_4^-,$$

при ограничениях

$$x_1 + s_1^+ - s_1^- = 80,$$
$$x_2 + s_2^+ - s_2^- = 60,$$
$$5x_1 + 3x_2 + s_3^+ - s_3^- = 480,$$
$$6x_1 + 2x_2 + s_4^+ - s_4^- = 480,$$

все переменные неотрицательны.

9. Обозначим через (x, y) координаты местоположения выбираемого дома. Частные целевые функции:

$$\text{минимизировать } s_1^-,$$
$$\text{минимизировать } s_2^+,$$
$$\text{минимизировать } s_3^-,$$

при ограничениях

$$[(x-1)^2 + (y-1)^2]^{1/2} + s_1^+ - s_1^- = 0.25,$$
$$[(x-20)^2 + (y-15)^2]^{1/2} + s_2^+ - s_2^- = 10,$$
$$[(x-4)^2 + (y-7)^2]^{1/2} + s_3^+ - s_3^- = 1,$$

все переменные неотрицательны.

Упражнения 8.3,а

1. $x_н = 0.0201$, $x_р = 0.0457$, $x_о = 0.0582$, $x_б = 2$ цента, $s_5^- = 1.45$, все остальные s_i равны нулю. Сумма налога на бензин составляет \$1.45 млн вместо желаемых \$1.6 млн.

3. На первый курс университета будет принято 840 жителей данного штата, 240 жителей других штатов и 120 неамериканцев.

5. $x_1 = 4$ партии изделий, $x_2 = 16$ партий, $x_3 = 3$ партии. Поскольку $s_1^+ = s_1^- = 0$, производство колес и сидений сбалансировано.

7. Для производства 80 единиц первого изделия и 60 единиц второго необходимо использование сверхурочных работ: 100 минут для первой операции и 120 минут — для второй.

9. $y = 0.8571 + 1.0714x_1 + 2.881x_2 - 0.9048x_3$.

Упражнения 8.3,b

1. Оптимизация целевой функции $G_1 = s_1^+$ дает $x_1 = 2.5$ минуты, $x_2 = 3.75$ минуты, $s_1^+ = 5$ и $s_2^- = 0$. Цель G_1 не выполнена на 5 млн человек, но цель G_2 выполняется автоматически.

3. a) $x_1 = 5$ концертов, $x_2 = 2.5$ выставки, $s_3^+ = 175$. Цель G_3 не достигнута.

   b) $x_1 = 3.6$ концертов, $x_2 = 3.2$ выставки, $s_1^+ = 280$. Цель G_1 не достигнута.

Глава 9

Упражнения 9.2,a

1. Новые ограничения: $x_1 \le x_5$ и $x_3 \le x_5$. Решение: $x_2 = x_3 = x_5 = 1$, $z = 90$.

3. Обозначим через x_{ij} количество бутылок типа i, полученных индивидуумом j, при этом $i = 1$ (полная бутылка), 2 (заполненная наполовину), 3 (пустая), $j = 1, 2, 3$. Возможное решение: $x_{11} = 3$, $x_{21} = 1$, $x_{31} = 3$, $x_{12} = 3$, $x_{22} = 1$, $x_{32} = 3$, $x_{13} = 1$, $x_{23} = 5$, $x_{33} = 1$. Задача имеет и другие решения.

5. Переменные задачи x_{ij} обозначают количество яблок, проданных сыном i за цену j. Решение: Джим продал 42 яблока по цене 1/7 долл. и 8 яблок по цене 3 долл., Билл продал 21 яблоко по цене 1/7 долл. и 9 яблок по цене 3 долл., Джон продал 10 яблок по цене 3 долл. Каждый из детей заработал по 30 долл.

7. Решение: AFT, TVA, ADV, OSF, KEN, суммарная метка равна 167.

9. На первой стороне кассеты записываются песни 1, 2, 4 и 8, на второй — 3, 5, 6 и 7. Емкость кассеты должна быть не менее 28 минут на каждой стороне.

Упражнения 9.2,b

1. $x_1 = 600$, $x_2 = 200$, $x_3 = 1200$, $z = \$9800$.

3. $x_{11} = 1200$, $x_{13} = 600$, $x_{22} = 1400$, $x_{32} = 300$, $x_{33} = 1000$.

Упражнения 9.2,c

1. Пусть переменные x_{ij} равны 1, если проект i предшествует проекту j, и равны 0 в противном случае. Количество сотрудников, входящих в комнату для совещаний и выходящих из нее, а также работающих над соответствующими проектами, показано в следующей таблице.

	1	2	3	4	5	6
1	—	4	4	6	6	5
2	4	—	6	4	6	3
3	4	6	—	4	8	7
4	6	4	4	—	6	5
5	6	6	8	6	—	5
6	5	3	7	5	5	—

Упражнения 9.2,d

1. Следует выбрать маршруты 5 и 6, общая длина маршрутов равна 104 мили.
3. Две станции, одна из которых обслуживает населенные пункты 1, 3 и 5, а вторая — 2, 4 и 6.

Упражнения 9.2,e

1. a)

5	6	4
4	5	6
6	4	5

b)

6	7	2
1	5	9
8	3	4

3. Производство следует организовать во втором цехе, при этом $x_1 = 26$, $x_2 = 3$, $x_3 = 0$.
5. Пусть $y = 0$, если $x_3 = 0$, и $y = 1$, когда $x_3 > 5$. Математическая модель:

$$\text{Максимизировать } z = 25x_1 + 30x_2 + 45x_3$$

при выполнении условий

$$3x_1 + 4x_2 + 5x_3 \le 100,$$
$$4x_1 + 3x_2 + 6x_3 \le 100,$$
$$x_3 - My \le 0,$$
$$-x_3 + My \le M - 5.$$

Решение: $x_1 = 0$, $x_2 = 11$, $x_3 = 11$, $y = 1$.

7. Необходимо заменить b_i на $b_i + My_i$, где $y_i = 0$ или 1 и $y_1 + y_2 + \ldots + y_m = 1$.

Упражнения 9.3,a

3. a) $x_1 = 2.5$, $x_2 = 2$ (также $x_1 = 1$, $x_2 = 3$), $z = 11$.
   c) $x_1 = 4.1667$, $x_2 = 1$, $z = 5.1667$.
5. b) $x_1 = x_2 = 0$, $x_3 = 10$, $z = 50$.

Упражнения 9.3,b

1. a) $x_1 = x_2 = x_3 = 1$, $z = 7$.
   c) Допустимого решения нет.
3. $x_1 = x_2 = x_3 = x_4 = 1$, $x_5 = 0$, $z = 95$.

Упражнения 9.3,c

3. a) Да, поскольку это отсечение проходит через целочисленные точки (допустимые и недопустимые) и при этом не исключает ни одной допустимой целочисленной точки.
   b) Нет, поскольку исключает допустимые целочисленные точки.

Упражнения 9.3,d

1. Отсечение I: $x_2 \le 3$. Отсечение II: $x_1 + x_2 \le 7$.
3. b) Целочисленное решение: $x_1 = 5$, $x_2 = 2$, $x_3 = 2$, $z = 23$. Решение, полученное путем округления оптимального непрерывного решения, — $x_1 = 5$, $x_2 = 3$, $x_3 = 3$, $z = 27$. Последнее решение недопустимо, поскольку не удовлетворяет второму ограничению.

Глава 10

Упражнения 10.2,a

1. Маршрут 1→3→5→7, длина маршрута = 21 миля.

Упражнения 10.3,a

1. Маршрут 1→3→5→7, длина маршрута = 21 миля.
3. Маршрут 1→2→3→5→7, длина маршрута = 17 миль.

Упражнения 10.4,a

1. $(m_1, m_2, m_3) = (0, 3, 0)$, доход равен 47.
3. Турист должен взять одну единицу еды, 2 комплекта средств первой помощи и 3 предмета одежды. Общая сумма предпочтений равна 26.
5. Под помидоры отводится 2 ряда, под бобы — 1 ряд и под кукурузу — также 1 ряд. Сумма предпочтений равна 30.
7. Следует вложить средства в предвыборную кампанию в избирательных участках 1, 2 и 3; число избирателей здесь равно 92 000.
9. $y_1 = y_2 = ... = y_n = c/n$, $z = (c/n)^n$.

Упражнения 10.4,b

1. a) Количество рабочих, занятых от первой до пятой недели, равно соответственно 6, 5, 3, 6 и 8 человек.
3. На каждую из четырех недель следует арендовать 7, 4, 8 и 8 автомобилей соответственно; общая сумма аренды равна \$6940.

Упражнения 10.4,c

1. a) *C*→*З*→*C*→*C*, доход = \$72 800.
3. *C*→*З*→*C*→*C*→*C*, доход = \$6020.

Упражнения 10.4,d

1. Первый год: инвестировать \$5000 в первый банк. Второй год: инвестировать \$4090 во второй банк. Третий год: \$3090 в первый банк. Четвертый год: инвестировать \$2065 в любой банк.

Упражнения 10.5,а

1. а) $0 \le x_1 \le 7/3$, $x_2 = 3 - (2/7)x_1$.

Глава 11

Упражнения 11.3,а

1. а) $y^* = 346.4$ единиц, $t_0^* = 11.55$ дней, затраты/день = \$17.30.
3. а) Необходимо сделать заказ на 200 единиц продукции в том случае, когда уровень запаса опустится до 150 единиц.
   b) Приблизительно 91 заказ.
5. Необходимо 100 поддонов каждые 20 дней.

Упражнения 11.3,b

1. Отелю следует воспользоваться скидкой, поскольку стоимость стирки партии из 1800 полотенец составит \$414, а партии из 2500 — \$356.94.
3. Да, для заказа партии в 150 единиц.

Упражнения 11.3,с

1. $y_1 = 10.85$, $y_2 = 16.9$, $y_3 = 22.36$.
3. $y_1 = 141.42$, $y_2 = 100$, $y_3 = 67.1$, $y_4 = 63.24$.

Упражнения 11.4,а

1. а) 500 единиц комплектующих в кварталы 1, 4, 7 и 10.
   b) 200 единиц комплектующих в кварталы –1, 2, 5, 8 и 300 единиц в кварталы 1, 4, 7 и 10.

Упражнения 11.4,b

1. Первый месяц: 90 единиц продукции произведено в обычном режиме, 50 единиц — за счет сверхурочных работ. Второй месяц: 100 единиц продукции в обычном режиме, 60 единиц — за счет сверхурочных. Третий месяц: 120 единиц в обычном режиме, 80 единиц — за счет сверхурочных. Четвертый месяц: 110 единиц в обычном режиме и 50 — за счет сверхурочных.
3. Первый этап: 100 единиц продукции произведено в обычном режиме, 50 — за счет сверхурочных работ и 23 единицы — на условиях субподряда. Второй этап: 40 единиц продукции произведено в обычном режиме, 60 единиц — за счет сверхурочных работ и 80 — на условиях субподряда. Третий этап: 90 единиц продукции в обычном режиме, 80 единиц — за счет сверхурочных работ и 70 — на условиях субподряда. Четвертый этап: 60 единиц продукции в обычном режиме и 50 единиц — за счет сверхурочных работ. Пятый этап: 60 единиц продукции в обычном режиме, 50 — за счет сверхурочных работ и 83 — на условиях субподряда.

Упражнения 11.4,c

1. a) Нет.
   b) i) $1 \le x_2 \le 6, 0 \le x_3 \le 4, 0 \le z_1 \le 5, 1 \le z_2 \le 5, 0 \le z \le 4$.

Упражнения 11.4,d

1. В первый период нет заказа, во второй — заказ на 112 единиц, в третий нет заказа, в четвертый период заказ на 67 единиц; стоимость = $632.
3. Заказ на 50 единиц в первый период и на 260 единиц во второй период; стоимость = $3 190.
5. Заказ на 150 единиц на первом этапе, на 120 единиц на втором этапе, на 110 единиц на четвертом, на 90 единиц на шестом этапе, на 310 единиц на седьмом, заказ на 190 единиц на девятом этапе; стоимость = $7 090.

Упражнения 11.4,e

1. В январе необходимо произвести 210 единиц изделия, 255 единиц изделия в апреле, 210 — в июле и 165 — в октябре.

Глава 12

Упражнения 12.2,a

1. a) 0.15, 0.25.
   b) 0.6.

Упражнения 12.2,b

1. События E и F являются несовместными. $P\{E + F\} = 2/3$.

Упражнения 12.2,c

1. 0.4.
3. 5/32.
5. a) 0.125.
   b) 0.6.

Упражнения 12.3,a

1. a) $p(x) = x/15, x = 1, 2, ..., 5$.
   b) 0.4.
3. $P\{\text{потребность} \ge 1100\} = 0.3$.

Упражнения 12.4,a

1. 2/3 марки.

Упражнения 12.4,b

1. Математическое ожидание = 2.67 марки, дисперсия = 1.56.

Упражнения 12.4,c

1. a) Для x_1 и x_2 $p_1 = 0.4$, $p_2 = 0.2$, $p_3 = 0.4$.
   b) Нет.
   c) 4.
   d) 0.
   e) 48.8.

Упражнения 12.5,a

1. $(0.5)^{10}$.
3. Вероятность этого события равна 0.0547 при условии, что все предсказания будут случайными (т.е. все исходы будут равновероятными).

Упражнения 12.5,b

1. 0.8646.

Упражнения 12.5,c

1. 0.3678.

Упражнения 12.5,d

1. a) $P\{x \geq 26\} = 0.0228$.
   b) $P\{x \leq 17\} = 0.0062$.
3. $P\{x_1 - x_2 \geq 0\} = 0.24$.

Глава 13

Упражнения 13.2,a

1. $y_{25}^* = 60.33$ единицы.
3. При $n = 3$ получим $y_{1990}^* = 1791.3$ посетителей на автомобилях и 938.33 посетителей, воспользовавшихся воздушным транспортом.
5. Населенный пункт 1: при $n = 5$ получим $y_{1995}^* = 238.6$ слушателей первого семестра, 1 260.2 слушателей второго семестра и 117 — третьего семестра.

Упражнения 13.3,a

1. $y_{25}^* = 59.63$ единицы.
3. $y_{1990}^* = \$26.27$ млн.

Упражнения 13.4,a

1. $y^* = 39.23 + 1.262x$, $r = 0.394$, $y^*_{25} = 70.77$ единицы.

3. $y^* = 20.6 + 0.873x$, $r = 0.991$, $y^*_{1990} = \$30.2$ млн.

Глава 14

Упражнения 14.2,a

1. $w_A = 0.44214$, $w_B = 0.25184$, $w_C = 0.30602$.

Упражнения 14.2,b

1. $w_S = 0.331$, $w_J = 0.292$, $w_M = 0.377$, $CR_A = 0.028$, $CR_{A_c} = 0.428$, $CR_{A_o} = 0.558$, $CR_{A_p} = 0.17$, только матрица A является достаточно согласованной.

3. $w_P = 0.502$, $w_H = 0.498$, $CR_A = 0.0072 < 0.1$.

5. $w_Ф = 0.34$, $w_M = 0.66$.

Упражнения 14.3,a

1. $MV(B) = \$380$, $MV(H) = -\$50$.

3. MV(простой) = 7.2%, MV(специальный) = 13.5%, MV(глобальный) = 11.7%.

5. MV(с рекламой) = \$625 000, MV(без рекламы) = \$360 000.

7. P{исход 1} = P{исход 2} = 1/12, P{исход 3} = 0.5, P{исход 4} = 1/3. Для всех шести возможных выборов $MV \le -\$0.53$. Устойчивый выигрыш в этой игре невозможен.

9. MV(крупное предприятие) = \$3 250 000, MV(малое предприятие) = \$900 000.

11. MV(крупное предприятие) = \$3 300 000, MV(малое предприятие) = \$1 667 600.

Упражнения 14.3,b

1. Оптимальная длина цикла равна 8 месяцам, MV(месячная стоимость ремонта) = \$397.50.

3. В первый день следует заказать 200 булочек, во второй — столько, сколько было реализовано в первый день, MV(доход) = \$186.84.

7. В данном случае критерии предельного уровня можно записать в виде неравенств $(\ln L - L/200) \ge 4.089$ и $(\ln L - L/100) \ge 3.505$, которые выполняются при $100 \le L \le 150$.

Упражнения 14.3,c

1. P{не будет дождя при данном прогнозе} = 0.07.

3. MV(биржа) = \$2491.46, MV(депозит) = \$800.

5. MV(растениеводство) = \$2477.89, MV(пастбище) = \$7500.

7. a) $P\{a_1 \mid s_1\} = 0.96$, $P\{a_2 \mid s_1\} = 0.04$, $P\{a_1 \mid s_2\} = 0.85$, $P\{a_2 \mid s_2\} = 0.15$, $P\{a_1 \mid s_3\} = 0.57$, $P\{a_2 \mid s_3\} = 0.43$.

   b) Если в результате проверки оба изделия являются доброкачественными, то партия изделий отправляется потребителю A, издержки = \$116.44. Если в результате проверки только одно изделие доброкачественно, то партия изделий также отправляется потребителю A, издержки = \$216.86. Если в результате проверки оба изделия бракованны, то партия изделий и в этом случае отправляется потребителю A, издержки = \$471.30.

Упражнения 14.3,d

1. a) Нет смысла, так как MV(возврат) = \$5.

   b) $U(\$10) = 100$, $U(\$x) = 0$, $x \neq 10$.

   c) Да, поскольку $U(\$10) = 100$, а $U(\$5) = 0$.

3. a) $p = 0.667$.

   c) Ожидаемая полезность инвестиции в предприятие I = 80, а ожидаемая полезность инвестиции в предприятие II = 84. Поэтому следует делать инвестиции в предприятие II.

Упражнения 14.4,a

1. a) По всем критериям следует выбрать альтернативу a_3.

   b) Все критерии указывают или на альтернативу a_2 или a_3.

3. Все критерии указывают на третий станок.

Упражнения 14.5,a

1. a) Стратегии 2 и 3.

   b) Стратегии 1 и 3.

3. a) Обозначим через v цену игры. Тогда $2 < v < 4$.

   b) $0 < v < 7$.

   c) $2 < v < 3$.

   d) $-1 < v < 0$.

Упражнения 14.5,b

1. $(x_1, x_2) = (0.5, 0.5)$, $(y_1, y_2) = (0.5, 0.5)$, $v = 0$.

3. a) $(x_1, x_2) = (0.5, 0.5)$, $(y_1, y_2, y_3) = (0, 0.65, 0.35)$, $v = 0.5$.

   b) $(x_1, x_2, x_3) = (0.25, 0.75, 0)$, $(y_1, y_2) = (0.75, 0.25)$, $v = 5.75$.

Упражнения 14.5,c

3. a) UA: $(x_1, x_2, x_3, x_4) = (0, 0.5, 0, 0.5)$, DU: $(y_1, y_2, y_3, y_4) = (0.14, 0.34, 0.27, 0.25)$, $v = 0.5$.

   b) UA

5. Оба игрока применяют стратегии (1, 2) и (2, 1) с вероятностями 0.571 и 0.429 соответственно.

Глава 15

Упражнения 15.2,а

1. Вращение 1: продолжать вращение колеса. Вращение 2: продолжать вращение, если на первом вращении колесо остановилось на числе из интервала [1, 6], иначе закончить игру. Вращение 3: продолжать вращение, если на втором вращении колесо остановилось на числе из интервала [1, 6], иначе закончить игру. Вращение 4: продолжать вращение, если на третьем вращении колесо остановилось на числе из интервала [1, 5], иначе закончить игру. Вращение 5: продолжать вращение, если на четвертом вращении колесо остановилось на числе из интервала [1, 4], иначе закончить игру. Вращение 6: окончание игры.

Упражнения 15.3,а

1. Первый год: инвестировать все $10 000. Второй год: инвестировать все накопленные деньги. Третий год: воздержаться от инвестиций. Четвертый год: инвестировать все накопленные деньги.

3. Первый год: произвести два суперкомпьютера. Второй год: произвести два суперкомпьютера плюс число суперкомпьютеров, определяемое вложениями за счет продажи суперкомпьютеров, произведенных в предыдущем году. Третий год: так же, как и во втором году. Четвертый год: если вложения за счет продажи суперкомпьютеров, произведенных в предыдущем году, меньше $1 млн, то производство следует остановить; если есть задолженность в один суперкомпьютер за третий год, то нужно произвести один суперкомпьютер.

Упражнения 15.4,а

3. Игра 1: ставка = $0 (т.е. следует пропустить эту игру). Игра 2: ставка = $1. Игра 3: ставка = $1, если был выигрыш в предыдущей игре, иначе третью игру пропустить. Максимальная вероятность получить в этой игре $4 равна 0.0625.

Глава 16

Упражнения 16.2,а

1. a) Заказ в 1000 единиц следует делать тогда, когда уровень запаса опустится до 537 единиц.

3. a) 0.0023.

   b) Заказ в 300 единиц кинопленки необходимо сделать в том случае, когда уровень запаса опустится до 70 единиц.

Упражнения 16.2,b

1. a) 3.125 заказов.

   b) $312.50 в месяц.

c) \$408.
d) \$2.0397.
e) 0.06.

3. $y^* = 316.85$ галлонов, $R^* = 58.73$ галлонов.

Упражнения 16.3,a

3. $19 \leq p \leq 35.7$.
5. 230 копий.
7. Приблизительно 39 пальто.
9. $y^* = 5.5$ единиц.

Упражнения 16.3,b

1. Если $x < 3.52$, то заказ = $8 - x$, иначе заказ не делать.
3. Если $x < 136$, то заказ = $144 - x$, иначе заказ не делать.

Упражнения 16.4,a

1. Этап 1: если $x < 9.02$, то заказ = $9.02 - x$, иначе заказ не делать.
   Этап 2: если $x < 7.5$, то заказ = $7.5 - x$, иначе заказ не делать.

Глава 17

Упражнения 17.1,a

1. a) 71%.
   b) Для того чтобы время ожидания не превышало 3 минут, необходимо не менее пяти кассиров. Для достижения 90% эффективности достаточно двух кассиров.

Упражнения 17.2,a

1. a) Самолет и взлетно-посадочная полоса.
   b) Такси и поездка пассажира.
5. a) Нет.
   g) Есть возможность перехода из очереди в очередь.

Упражнения 17.3,a

1. a) Средний интервал времени между поступлениями заявок = 1/(интенсивностью поступления заявок в единицу времени).
   b) i) $\lambda = 6$ заявок в час, средний интервал времени между поступлениями заявок = 1/6 часа.
   c) i) $\mu = 5$ обслуживаний в час, среднее время обслуживания = 0.2 часа.

3. a) $f(t) = 20 \exp(-20t)$, $t \geq 0$.
   b) 0.00674.
   c) 0.632.
   d) 11.67.

Упражнения 17.3,b

1. a) Ложно.
   b) Истинно.
3. a) 0.6886.
   b) 0.1738.
   c) 0.1376.
5. 75.97 центов.
7. a) 0.4866.
   b) 0.948 минуты.
9. a) $f(t) = (1/10) \exp(-t/10)$.
   b) $f(t) = (1/7) \exp(-t/7)$.

Упражнения 17.4,a

1. 0.5595.
3. a) 10 заявок.
   b) 0.3679.
   c) 0.6321.
   d) 0.002479.
5. a) 7.5 полок.
   b) 0.

Упражнения 17.4,b

3. a) 0.4787.
   b) Приблизительно 25 билетов.
5. 0.37116.
7. 0.00005.

Упражнения 17.5,a

1. a) (15/55, 24/55, 16/55).
   b) 2.02 кассы.
   c) 0.98 кассы.
3. a) 0.4445.
   b) 0.5555.
5. a) 0.00129.
   b) 0.2273.

c) 2.9768 столов.
d) 0.2796.

7. a) $5.5p_1 = 10p_0$, $10p_0 + 6p_2 = 14.5p_1$, $9p_1 + 6.5p_3 = 14p_2$, $8p_2 + 7p_4 = 13.5p_3$.
b) $p_0 = 0.08888$, $p_1 = 0.1614$, $p_2 = 0.2422$, $p_3 = 0.2982$, $p_4 = 0.2981$.

Упражнения 17.6,a

1. a) 0.1917 автомобилей.
b) 0.53264 часа.
c) 1.0104 автомобилей.

Упражнения 17.6,b

1. a) 66.67%.
b) 0.6667.
c) 0.961.
d) $s \geq 11$.
3. a) 0.225 случаев.
b) 37.5%.
c) 2.4 недели.
5. a) 0.4.
b) 0.9.
c) 2.25.
d) 0.0036.

Упражнения 17.6,c

1. 0.659.

3. $\mu \geq 10$.

5. $37.95.

Упражнения 17.6,d

1. a) 0.3654.
b) 0.207 часа.
c) 3.212 свободных мест.
d) 0.048.
e) Приблизительно 10 автомобилей в час.

3. Не более двух стульев.

5. a) 0.00002.
b) 0.00007.

7. a) 0.962.
b) 0.19 клиентов в час.
c) 1.286 клиентов.
d) 0.1424 часа.

Упражнения 17.6,e

1. a) 0.711.
   b) 0.596.
   c) 0.4 автомобиля для $c = 2$ и 0.8 — для $c = 4$.
   d) Пять и более автомобилей.
3. a) $c \geq 2$.
   b) $c \geq 4$.
   c) $c \geq 1$.
5. Не менее 10 мест.
7. a) 0.65772.
   b) 0.0662 часа.
   c) 3.29 программы.
   d) 2.1%.
   e) 0.667 ЭВМ.

Упражнения 17.6,f

1. a) 1.3 автомобиля.
   b) 0.04468.
3. a) 0.1677 механиков.
   b) 3.354 клиентов.
   c) 0.9441.
   d) 0.10559.
   e) 6.7081 механизмов.
   f) 0.9441.

Упражнения 17.6,g

1. a) 0.
   b) 1.
   c) 0.63923.
   d) $39.780.

Упражнения 17.6,h

1. b) 2.01 механиков.
   c) 0.10779.
   d) 0.34492.
5. a) 1.2077 сотрудников.
   b) 0.22972.

Упражнения 17.7,a

1. 33.3%.

3. a) 4.2 заказа.
   b) 17 дней.
   c) 2.136 дней.
5. a) 0.9395 изделий.
   b) 0.278.
   c) 74.78 минут.

Упражнения 17.9,a

1. b) Модель 3 имеет наименьшую стоимость, равную \$631.20.
3. Сканер А стоит \$5.53 в час, сканер В — \$4.01 в час.
5. 92.65 работ в неделю.

Упражнения 17.9,b

1. Три служащих со стоимостью \$159.77 в час.
3. a) Выиграет \$5 920 в месяц.
   b) Дополнительный выигрыш в \$2 880.

Упражнения 17.9,c

1. a) Пять механиков.
   b) Шесть механиков.

Глава 18

Упражнения 18.2,a

1. a) 90 кв. см.
   b) $f(x) = 1/8$ при $-7 \le x \le 1$, $f(y) = 1/8$ при $-2 \le y \le 6$. $x = -7 + 8R_1$, $y = -2 + 9R_2$.
3. a) Если $0 \le R \le 0.5$, то игрок А получает \$10, если же $0.5 \le R \le 1$, то игрок Б получает \$10.
   b) $-2.03 \le \mu \le 3.63$.
   c) 0.

Упражнения 18.3,a

1. a) Дискретный тип.
   b) Непрерывный тип.
   c) Дискретный тип.

Упражнения 18.4,a

1. Событие A1: поступает работа из первого источника. Событие A2: поступает работа из второго источника. Событие D: завершение выполнения работы на машине.

3. События A1 и A2: прибывают автомобили по линиям 1 и 2 соответственно. Событие A3: прибывший автомобиль уезжает в поисках другого банка для обслуживания. События D1 и D2: обслуживание автомобилей на линиях 1 и 2 соответственно.

Упражнения 18.4,b

1. 0, 0.015, 0.295, 0.458.

5. В первом случае выполняется ремонт, если значение случайного числа R удовлетворяет неравенствам $0 \le R \le 0.2$. Во втором случае ремонт будет выполнен, если $0.8 \le R \le 1$.

7. $x = a + [R(b-a)(c-a)]^{1/2}$, если $0 \le R < (b-a)/(c-a)$ и
   $x = c - [(1-R)(c-b)(c-a)]^{1/2}$, если $(b-a)/(c-a) \le R \le 1$.

9. $x = 0$, если $0 \le R < 1 - p$, в противном случае значение x равно наибольшему целому, не превышающему величины $\ln(1-R)/\ln(p)$.

Упражнения 18.4,c

1. 0.803 часа.
3. Метод сверток дает $y = 8.1094$, метод Бокса–Мюллера — $y = 7.5365$.
7. $n = 4$.

Упражнения 18.4,d

3. $g(x) = 0.707$ при $0 \le x \le \pi/2$. $h(x) = 0.637$ при $0 \le x \le \pi/2$.

Упражнения 18.5,a

1. $R_4 = 0.0769$, $R_5 = 0.2307$, $R_6 = 0.6154$, $R_7 = 0.0769$, $R_8 = 0.2307$.

Упражнения 18.6,a

1. Средняя длина очереди = 1.11, среднее время ожидания в очереди = 11.59, средняя занятость парикмахера = 0.841.

5. a) 1 парикмахер.
   b) 35 минут.
   c) 10 минут.

Упражнения 18.7,a

1. 3.32 временных единиц.
3. $0.305 \le \mu \le 1.577$.

Упражнения 18.8,a

1. $\overline{z} = 1.81$, $s_z = 0.595$.
3. $\overline{z} = -0.149$, $s_z = 0.67$.

Глава 19

Упражнения 19.2,a

1. Первый и второй годы: рекламную кампанию следует проводить только при снижении доходов. Третий год: рекламная кампания не проводится.
5. Заказать 2 холодильника при полном отсутствии их в магазине, в противном случае заказ не делать.

Упражнения 19.3,a

1. Рекламная кампания проводится всегда при состоянии 1.

Глава 20

Упражнения 20.2,a

1. a) Нет экстремальных точек.
   b) Минимум в точке $x = 0$.
   c) Точка перегиба при $x = 0$, минимум при $x = 0.63$ и максимум при $x = -0.63$.
3. Минимум достигается только в точке (1, 2, 0).

Упражнения 20.3,a

1. a) $\partial_c f = -0.046$ и $\partial_c f = -0.04618$. Да, влияние линейной аппроксимации менее существенно при уменьшении величины ∂x_2.
   b) $\partial x_1 = -0.283\partial x_2$, $\partial x_3 = 0.25\partial x_2$.
   c) $\partial x_1 = -0.0283$.
   d) $\partial_c f = -0.46$.

Упражнения 20.3,b

1. a) Нет.
3. $\partial f = 2\partial C^{(2-n)/n}$.
5. $\partial f = 0.472$.

Упражнения 20.3,c

1. Максимум достигается в точке (12/5, 9/5, 0, 0).

Упражнения 20.3,d

5. $\lambda_1 \leq 0$, λ_2 не имеет ограничения в знаке, $\lambda_3 \geq 0$. Необходимые условия выполняются, если функции f и g_1 вогнутые, g_2 — линейная, а g_3 — выпуклая.

Глава 21

Упражнения 21.1,a

3. Максимальное число итераций = $1.44 \ln[-1 + (b - a)/\Delta]$.

Упражнения 21.1,b

1. $(x_1, x_2) = (1/3, 1/3)$.

Упражнения 21.2,a

3. Следует использовать подстановку $\ln y = \ln x_1 + \ln x_2 + \ln x_3$.

Упражнения 21.2,b

1. $x_1 = 1$, $x_2 = 0$, $z = 4$.

Упражнения 21.2,c

1. Здесь при $x_j > 0$ необходимые условия не выполняются. Целевая функция имеет инфимум при $x = 0$.
3. $(x_1, x_2) = (1.39, 1.13)$.

Предметный указатель

И

К

Л

М

С

Т

У

Ф

Ц

Э

Я

Научно-популярное издание

Хэмди А. Таха

Введение в исследование операций, 6-е издание

Литературный редактор *Е.Д. Давидян*
Верстка *И.В. Родюк*
Художественный редактор *В.Г. Павлютин*
Технический редактор *Г.Н. Горобец*
Корректоры *Л.А. Гордиенко, Л.В. Коровкина, О.В. Мишутина*

Издательский дом "Вильямс".
101509, Москва, ул. Лесная, д. 43, стр. 1.
Изд. лиц. ЛР № 090230 от 23.06.99
Госкомитета РФ по печати.

Подписано в печать 10.05.2001. Формат 70×100/16.
Гарнитура Times. Печать офсетная.
Уч.-изд. л. 47,3. Усл. печ. л. 61,49. Тираж 5000 экз. Заказ № 164[illegible].

Отпечатано с диапозитивов в ФГУП "Печатный двор"
Министерства РФ по делам печати,
телерадиовещания и средств массовых коммуникаций.
197110, Санкт-Петербург, Чкаловский пр., 15.