Г. ВАГНЕР

ОСНОВЫ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ

2

ИЗДАТЕЛЬСТВО · МИР ·

Harvey M. Wagner

Department of Administrative Science Yale University;
Consultant to McKinsey and Company, Inc.

Principles of Operations Research

With Applications to Managerial Decisions

Prentice-Hall, Inc., Englewood Cliffs,
New Jersey 1969

Г. ВАГНЕР

ОСНОВЫ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ

Том 2

Перевод
с английского
В. Я. Алтаева

Издательство «Мир»
Москва 1973

УДК. 35.073.5

Том посвящен методам динамического, целочисленного и нелинейного программирования. Рассмотрены различные классы динамических моделей (модели управления запасами, модели распределения, модели замен и ряд других) и обсуждены процедуры построения соответствующих алгоритмов оптимизации. Приведен подробный анализ зависимости этих процедур от величины интервала времени, для которого ведется поиск оптимальной стратегии.

Операционным задачам, решение которых можно получить методами целочисленного программирования, посвящена специальная глава. В ней обсуждены также возможности и особенности комбинаторных приемов оптимизации.

Разделы, в которых рассматриваются нелинейные модели, содержат описание всех представляющих практический интерес методов оптимизации, включая разделение переменных, линеаризацию и аппроксимацию решений выпуклыми и вогнутыми функциями. Специально рассмотрен случай нелинейных ограничений. Приведен обобщенный алгоритм нелинейной оптимизации.

Редакция литературы по вопросам новой техники

$\text{В}\ \dfrac{3314\text{-}335}{041(01)\text{-}73}$

Г. ВАГНЕР

Основы исследования операций

Том 2

Редактор *М. ВЕЛИКОВСКИЙ*

Художник *А. Шипов.* Художественный редактор *Ю. Урманчеев*

Технический редактор *Н. Иовлева*

Сдано в набор 26/X 1972 г. Подписано к печати /II 1973 г.
Бумага № 1 $60\times90^1/_{16}$=15,25 бум. л. 30,50 печ. л. Уч.-изд. л. 32,39 Изд. № 20/6479
Ц~~ена 2 р. 54~~ к. Зак. 0734 2 р 60

ИЗДАТЕЛЬСТВО «МИР» Москва, 1-й Рижский пер., 2

Ордена Трудового Красного Знамени Московская типография № 7 «Искра революции» Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли. Москва, К-1, Трехпрудный пер., 9.

ГЛАВА 8

Введение в теорию динамических оптимизационных моделей

8.1. АНАЛИЗ ДИНАМИЧЕСКИХ ПРОЦЕССОВ

Элементы динамики и учет времени играли важнейшую роль в нескольких прикладных задачах исследования операций, рассмотренных в предыдущих главах,— как, скажем, в общей задаче планирования производственной программы (разд. 2.6). Однако ранее основное внимание уделялось рассмотрению алгоритмов, которые позволяют находить решения системы, включающей большое число ограничений. Мы, например, выяснили, каким образом симплексный метод позволяет на каждой итерации получать допустимый план, удовлетворяющий всем ограничениям системы. В том же контексте анализировалась устойчивость решения задачи линейного программирования. В гл. 6 и 7 рассмотрены особенности структуры сетевых моделей; основное внимание здесь было уделено эффективным методам отыскания численных решений задач большой размерности. Единственным важным отступлением от подобного подхода является алгоритм решения задачи отыскания кратчайшего пути на ориентированной ациклической сети (разд. 7.7). В этом алгоритме используется «односторонняя ориентация» сети, и читателю рекомендуется снова внимательно его рассмотреть, прежде чем приступить к изучению разд. 8.2.

В гл. 8—12 решающее значение по-прежнему имеет одновременный учет всех ограничений системы, однако излагаемый здесь материал в основном посвящен динамическим структурным зависимостям оптимизационных моделей. В этих главах рассматриваются только детерминированные задачи, так что в каждой из них процесс решения приводит к однозначному результату. Вероятностные модели исследуются в гл. 17—18. Однако читатель увидит, что аналитические методы решения большинства вероятностных задач являются непосредственным обобщением тех подходов, которые излагаются в гл. 8—12. Соответственно изучение методов решения детерминированных задач оказывается весьма полезным и в качестве подготовки к рассмотрению вероятностных моделей.

В гл. 8—12 основное внимание будет уделено форме и свойствам оптимальных планов. Будут изучены условия, которым должен удовлетворять оптимальный многошаговый процесс принятия решений, и исследовано, каким образом использовать эти условия для нахождения лучшего варианта. Подобный анализ часто называют *динамическим программированием*.

Основным предметом рассмотрения в гл. 8 является анализ устойчивости решений во времени. Например, читатель увидит, что увеличение длительности планового периода может существенно повлиять на правильность текущего выбора. Наряду с этим вопросом будет также изучено влияние начальных условий (начальный уровень запасов) и ограничений (ограниченность производственных мощностей).

Во всех моделях, рассматриваемых в гл. 8—10, плановый период является конечным — например, он может быть равен году или десятилетию. В гл. 11—12 будут обсуждены особенности оптимизации в условиях бесконечного планового периода. (Вопросы дисконтирования во времени и приведения экономических показателей к исходному моменту времени будут излагаться лишь в гл. 11.)

За увеличение глубины анализа динамических процессов приходится платить. Так, в этом случае необходимо радикально сокращать размерность (число ограничений) динамических оптимизационных моделей по сравнению с типичными задачами линейного программирования. Однако предлагаемый подход дает не только ценные дополнительные результаты, но и ряд практических преимуществ. Одно из них уже упоминалось выше: это возможность использования динамического программирования при бесконечном плановом периоде. Другое его преимущество состоит в том, что такой подход позволяет решать ряд задач небольшой размерности, являющихся тем не менее важными частными случаями задач нелинейного программирования. К ним относятся задачи с нелинейной целевой функцией и небольшим числом нелинейных ограничений, а также задачи с целочисленными переменными. В гл. 10 будут приведены конкретные примеры, показывающие, что подобные задачи, даже внешне простые, могут быть сложными и запутанными. Наконец, приведенные численные методы применимы и к ряду проблем принятия решений, по существу не являющихся динамическими, которые, однако, для удобства алгоритмизации можно рассматривать как динамические или преобразовывать в многошаговые.

Для того чтобы читатель видел все излагаемые вопросы в должной перспективе, подчеркивается, что численные решения всех моделей в гл. 8—10 можно получить с помощью ранее изученных алгоритмов (в частности, алгоритма нахождения кратчайшего пути на ориентированной ациклической сети). Важнейшая наша цель — научиться представлять модели в таком виде, в котором выявляются их динамические свойства. Эта цель достигается отчасти благодаря тому, что наш словарь в области исследования операций пополняется новыми терминами, а отчасти благодаря введению удобной системы математических обозначений. Однако читатель поймет, что не существует простых правил, механическое применение которых в любой задаче позволяет выявить ее динамические свойства. Лучшим учителем является *опыт*, поэтому в тексте приводятся многообразные примеры. Многие конкретные оптимизационные задачи удается сформулировать несколькими внешне различными способами, причем, как будет видно из дальнейшего, в каждой из формулировок

лировок в центре внимания находится та или иная из структурных зависимостей.

Значение динамического программирования. Сделанные выше вводные замечания должны облегчить читателю переход к новой точке зрения, отличной от той, которая развивалась в предыдущих главах. Однако до сих пор ничего не говорилось о практической важности изучаемых далее моделей. Велико ли их экономическое значение?

Как было показано в предыдущих главах, модели линейного программирования в большинстве случаев используются в промышленности для принятия крупномасштабных плановых решений в сложных ситуациях. В то же время модели динамического программирования обычно применяются при решении задач значительно меньшего масштаба. Вот некоторые типичные области применения моделей динамического программирования при принятии решений:

Разработка правил управления запасами, устанавливающих момент пополнения запасов и размер пополняющего заказа.

Разработка принципов календарного планирования производства и выравнивания занятости в условиях колеблющегося спроса на продукцию.

Определение необходимого объема запасных частей, гарантирующего эффективное использование дорогостоящего оборудования.

Распределение дефицитных капитальных вложений между возможными новыми направлениями их использования.

Выбор методов проведения рекламной кампании, знакомящей покупателя с продукцией фирмы.

Систематизация методов поиска ценного вида ресурсов.

Составление календарных планов текущего и капитального ремонта сложного оборудования.

Разработка долгосрочных правил замены выбывающих из эксплуатации основных фондов.

Большинство этих применений детально рассматривается в настоящей и последующих главах, так что здесь не требуется дополнительных пояснений.

Процессы принятия решений, к которым относится ряд упомянутых выше моделей, часто относятся к числу *микроэкономических*. Однако во многих реально функционирующих системах еженедельно требуется принимать тысячи таких решений. В связи с этим модели динамического программирования ценны именно тем, что они позволяют принимать миллионы решений на основе стандартного подхода (часто с использованием ЭВМ) при минимальном вмешательстве человека. Без слов понятно, что даже если каждое из таких решений, взятое в отдельности, не является существенным, то в совокупности они могут оказывать большое влияние на прибыли фирм. Многие фирмы, воспользовавшиеся моделями динамического программирования, были неплохо вознаграждены — им удалось добиться сокращения затрат на обслуживание и ремонт или же уровня запасов на 25% и более, причем без какого-либо ухудшения качества обслуживания.

Небольшое наставление. Общей особенностью всех моделей динамического программирования является сведéние задачи принятия решений к получению рекуррентных соотношений. Если читатель никогда не пользовался подобными формализованными методами для решения задач, то связанная с этим система математических обозначений может показаться ему странной и даже вызывающей недоумение. Приводимые ниже советы помогут читателю преодолеть подобные трудности.

Текст рекомендуется прочесть не менее двух раз. При первом чтении следует постараться понять смысл поставленной задачи и хорошо ознакомиться с условными обозначениями; при втором чтении больше внимания целесообразно уделять деталям постановки, в том числе и характеру математических выражений. Необходимо внимательно ознакомиться с численными примерами и проверить правильность расчетов во всех тех случаях, когда это рекомендуется в тексте. Наконец, необходимо проявить терпение и не жалеть времени на изучение материала. Быстро усвоить методы динамического программирования удается далеко не всем, причем дело может медленно продвигаться даже у тех, кто сталкивался с рекуррентными соотношениями прежде. Однако если следовать приведенным советам, то после рассмотрения нескольких примеров в какой-то момент внезапно наступает «озарение», после чего читатель перестает испытывать трудности в понимании смысла рекуррентных соотношений.

8.2. ЗАДАЧА О ДИЛИЖАНСАХ: АЛЛЕГОРИЯ

На аллегорическом примере объясним некоторые важные идеи динамического программирования, а также построим систему условных обозначений динамических моделей. Сама задача не содержит ничего нового — в ней попросту требуется найти кратчайший путь на ориентированной ациклической сети. Однако рассматриваемый в настоящем разделе конкретный пример содержит больше пояснений, чем пример в разд. 7.7, где излагался алгоритм нахождения кратчайшего пути.

Пример. Жил некогда мистер М., который решил отправиться искать счастья в Сан-Франциско. В те дни дилижансы были единственным видом общественного транспорта для поездки из восточных штатов, где проживал мистер М., на Запад. В бюро путешествий ему показали карту Соединенных Штатов (рис. 8.1) с нанесенными на ней дилижансовыми маршрутами, которые обслуживались в то время. Каждый квадрат на карте изображает один из штатов (состояний); для удобства штаты пронумерованы. Заметим, что, какой бы из вариантов пути от штата *1* (Восток) до штата *10* (Запад) мы ни выбрали, он включает 4 дилижансовых маршрута — или 4 «шага».

Поскольку мистеру М. было известно, что путешествие связано с серьезными опасностями для здоровья и жизни, перед отъездом он решил застраховаться. Ставка страхового платежа (иными словами, стоимость, отвечающая принятой стратегии выбора пути) зависела от избираемых дилижансовых маршрутов, и она была тем выше, чем опаснее маршрут. Обозначим через c_{ij} стоимость страхового полиса для переезда из штата i в штат j. Условные численные значения c_{ij} проставлены на рис. 8.1. Цель мистера М.— выбрать

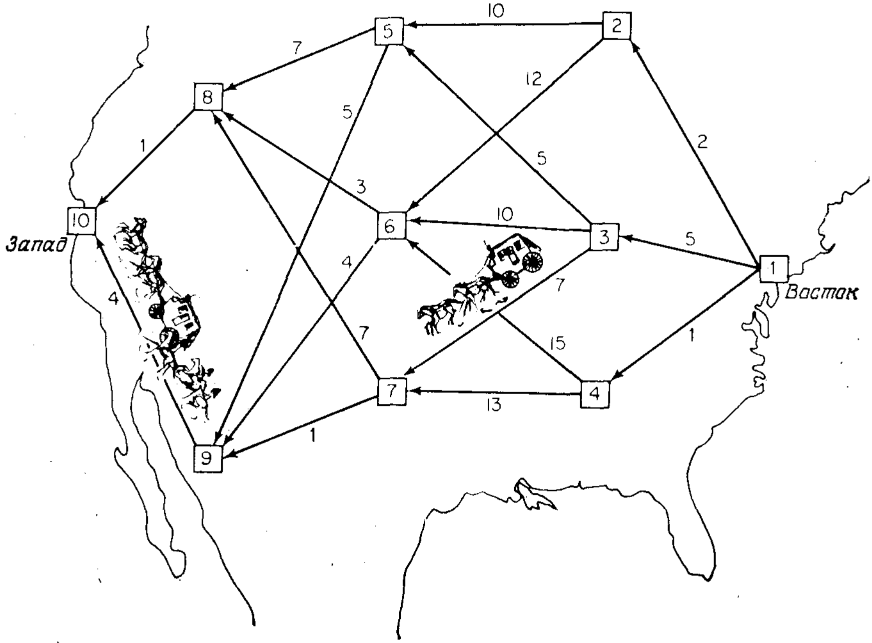


Рис. 8.1. Задача о дилижансах.

такой путь от штата *1* до штата *10*, для которого общая стоимость страхования является минимальной (читателю рекомендуется отыскать оптимальный путь самостоятельно).

Мистер М. начал анализ проблемы с того, что признал существенным сформулированный ниже принцип.

Принцип оптимальности. Оптимальная стратегия обладает тем свойством, что, каков бы ни был путь достижения некоторого штата (или состояния), последующие решения должны принадлежать оптимальной стратегии для части пути, начинающейся с этого состояния.

Таким образом, мистер М. понял, что, например, оптимальный путь из штата *6* не зависит от того, каким маршрутом он прибудет в него. (Вопрос читателю: как найти оптимальный маршрут из штата *6*?) Развивая эту логическую идею далее, мистер М. пришел к выводу, что если бы он знал оптимальные пути из штатов *5*, *6* и *7*, то смог бы достаточно легко определить и оптимальный путь из штата *3* — конечно, если бы он решил ехать через этот штат. В самом деле, потребовалось бы лишь *суммировать* стоимость переезда из штата *3* (буть то c_{35}, c_{36} или c_{37}) с ранее вычисленной стоимостью оптимального пути (соответственно из штатов *5*, *6* или *7*), а затем сравнить полученные суммы и выбрать тот штат, для которого эта сумма минимальна. Аналогичным образом, найдя оптимальные пути из штатов *2*, *3* и *4*, можно определить оптимальный путь из штата *1*. (Вопрос читателю: как это сделать?)

Для того чтобы учесть сформулированный принцип оптимальности и его вычислительный смысл, удобно использовать следующие обозначения:

$f_n(s)$ — стоимость, отвечающая стратегии минимальных затрат для пути от штата (состояния) s, если до конечного штата остается n шагов;

$j_n(s)$ — решение, позволяющее достичь $f_n(s)$.

Многих начинающих чрезвычайно затрудняет подобная система обозначений, которую всегда используют в моделях динамического программирования, и это объясняется тем, что она выглядит весьма сложной. Однако все эти буквы и индексы необходимы и несут важную смысловую нагрузку: f означает, что данное число есть значение целевой функции, s — что это значение зависит от состояния системы, подстрочный индекс n несет динамическую информацию о том, что из состояния s необходимо сделать еще n шагов; наконец, символ j зависит как от шага n, так и от состояния s и соответствует некоторому фиксированному пути.

Читая описание приводимых здесь и далее задач динамического программирования, иногда полезно просто выучить смысл обозначений, как заучивают значения слов при овладении иностранным языком. В моделях линейного программирования нет необходимости в столь сложной системе обозначений, поскольку эти модели решаются «за один шаг». Иное положение в динамическом программировании, где мы приходим к решению за ряд шагов. Будем надеяться, что эти соображения послужат некоторым утешением для читателя.

Возвратимся к задаче, стоящей перед мистером М. Он знает $f_0(10)$:

$$f_0(10) = 0 \text{ для } j_0(10) = \text{Остановка}, \tag{1}$$

поскольку для штата *10* число оставшихся шагов равно нулю, т. е. в этом штате путешествие заканчивается. Затем он видит, что очень легко вычисляются $f_1(8)$ и $f_1(9)$: к $f_0(10)$ попросту надо прибавить соответственно $c_{8,10}$ или $c_{9,10}$. Ободренный достигнутыми успехами, мистер М. решил вычислить $f_2(6)$ — стоимость стратегии минимальных затрат, отвечающей его нахождению в штате *6*, т. е. за два шага до конечного пункта. Он заметил, что если он попадет в штат *6*, то далее можно следовать только по двум путям. Первый из них —

через штат *8*; стоимость соответствующей стратегии определится, если прибавить $c_{6,8}$ к вычисленному ранее значению f_1 (8). Второй путь — через штат *9*; для вычисления его стоимости прибавим $c_{6,9}$ к значению f_1 (9), которое также уже вычислено. Мистер М. понял, что значение f_2 (6) должно равняться меньшей из этих двух сумм. (Вопрос читателю: таким ли способом он определял оптимальный маршрут из штата *6*?)

Здесь мистер М. начал подозревать, что в его подходе можно обнаружить известную методичность, и, разумеется, он был прав. Кратко описываемый метод можно представить в виде так называемого динамического рекуррентного соотношения:

$$f_n(s) = \min_{\substack{\text{по всем } s \\ \text{и } j \text{ на сети}}} [c_{sj} + f_{n-1}(j)], \quad n = 1,\ 2,\ 3,\ 4. \tag{2}$$

Выражение (2) означает, что мистер М. должен вычислить все возможные значения стоимости, отвечающие различным стратегиям, суммируя соответствующую стоимость для очередного шага пути (переезда из штата s в штат j), и стоимость, отвечающую *оптимальной* стратегии выбора пути, от штата j, из которого до конца пути остается только $(n - 1)$ шагов. Вычисленные суммы следует сравнить между собой и выбрать такое j, для которого значение этой суммы минимально. Мистер М. пришел к выводу, что он должен применять выражение (2) следующим образом: вначале вычислить все $f_1(s)$, т. е. $f_1(8)$ и $f_1(9)$; затем в свою очередь для $n = 2$ вычислить $f_2(5)$, $f_2(6)$ и $f_2(7)$; затем определить $f_3(2)$, $f_3(3)$ и $f_3(4)$. Задача будет решена после вычисления $f_4(1)$, так как штат *1* — начало путешествия вдоль пути, включающего 4 шага.

Описанный выше метод решения задач можно назвать «методом самообеспечения». Вычислительный процесс начинается с простых, пожалуй, даже тривиальных расчетов. Полученный результат используется для выполнения дальнейших вычислений, после чего процедура повторяется снова. В принятых условных обозначениях формула (2) означает, что, зная $f_0(s)$, можно вычислить $f_1(s)$, зная $f_1(s)$,— $f_2(s)$ и т. д. Такая вычислительная процедура именуется рекуррентным алгоритмом, а выражение (2) — **рекуррентной формулой или рекуррентным соотношением.**

Вычислительный процесс. Рассмотрим применение рекуррентного алгоритма, выполнив необходимые вычисления для численных значений c_{ij}, проставленных на рис. 8.1.

Для $n = 1$, руководствуясь выражениями (2) и (1), следует выполнить следующие тривиальные вычисления:

$$f_1\ (8) = c_{8\,10} + 0 = 1 \quad \text{для} \quad j_1(8) = 10 \quad (s = 8),$$

$$f_1\ (9) = c_{9\,10} + 0 = 4 \quad \text{для} \quad j_1(9) = 10 \quad (s = 9). \tag{3}$$

Упорядоченная запись остальных вычислений выполняется в таблицах, приведенных на рис. 8.2—8.5.

Для каждого из фиксированных значений числа оставшихся шагов n (а именно для n, равного 1, 2, 3 или 4) построена одна таблица. В ней предусмотрено: по одной строке для каждого состояния, из которого остается сделать n шагов и по одному столбцу для каждого состояния, из которого остается сделать на шаг меньше. Так, в таблице для $n = 1$ (рис. 8.2) имеются две строки, для штатов *8*

$n=1 \quad c_{sj}+f_0(j)$

	$s \backslash j$	Переход в состояние (штат): 10	$j_1(s)$	$f_1(s)$
Состояние (штат)	8	1+0	10	1
	9	4+0	10	4

Рис. 8.2. Задача о дилижансах.

$n=2 \quad c_{sj}+f_1(j)$

	$s \backslash j$	Переход в состояние (штат): 8	9	$j_2(s)$	$f_2(s)$
Исходное состояние (штат)	5	7+1	5+4	8	8
	6	3+1	4+4	8	4
	7	7+1	1+4	9	5

Рис. 8.3. Задача о дилижансах.

и *9*, поскольку мистер М. может выбрать маршрут, проходящий через любой из них. Однако в этой таблице — всего один столбец (для штата *10*), поскольку из штатов *8* и *9* можно попасть только в штат *10*. Для $n = 2$ (рис. 8.3) мистер М. может выбрать штаты *5*, *6* или *7*, поэтому в таблице три строки; но из этих штатов он может попасть только в штаты *8* или *9*, поэтому в таблице два столбца.

Цифры в основных столбцах таблицы (слева от двойной вертикальной черты) представляют собой сумму стоимости очередного шага c_{sj} — перехода из состояния s в состояние j — и стоимости $f_{n-1}(j)$, соответствующей оптимальной стратегии для части пути, начинающейся из состояния j. В каждой строке выбирается наименьшая из этих сумм; она проставлена в первом столбце справа и обозначена $f_n(s)$. Соответствующий оптимальный выбор состояния j обозначен $j_n(s)$ и проставлен во втором столбце справа.

Вычисления для $n = 1$ повторно приведены в таблице рис. 8.2. В данном примере, при $n = 1$, единственным возможным выбором является $j = 10$. Поэтому $j_1(8) = j_1(9) = 10$, что уже было отмечено в выражениях (3).

Вспомним, что при $n = 2$ значение j может равняться только 8 или 9. Для выполнения вычислений при $n = 2$ должны быть известны c_{sj}, а кроме них *лишь* значения $f_1(j)$. Вычисления приведены в таблице рис. 8.3. Отметим, что $f_1(8) = 1$ прибавляется к каждой из величин c_{s8} в столбце для $j = 8$, а $f_1(9) = 4$ — к каждой из c_{s9} в столбце для $j = 9$. Из таблицы видно, что оптимальный маршрут при $n = 2$ из штатов *5* и *6* проходит через штат *8*, а из штата *7* — через *9*.

Вычисления для $n = 3$ приведены в таблице рис. 8.4. Заметим, что здесь две клетки заштрихованы («запрещены»), поскольку из штата *2* нельзя попасть в штат *7*, а из штата *4* — в штат *5*. Еще раз отметим, что из предыдущей таблицы (рис. 8.3) потребовалось взять *только* значения $f_2(j)$. Это важнейший момент во всех методах динамического программирования: *стратегия оптимальных затрат для n*

$n=3 \quad c_{sj}+f_2(j)$

Состояние (штат) s \ j	Переход в состояние (штат) 5	6	7	$j_3(s)$	$f_3(s)$
2	10+8	12+4		6	16
3	5+8	10+4	7+5	7	12
4		15+4	13+5	7	18

Рис. 8.4. Задача о дилижансах.

шагов определяется экономическими последствиями очередного решения и стратегией оптимальных затрат для остальных $(n-1)$ *шагов.*

Завершающие вычисления (для $n=4$) приведены в таблице рис. 8.5. Из этой таблицы видно, что стоимость, отвечающая стратегии минимальных затрат, равна

$$f_4(1) = 17 \quad \text{для} \quad f_4(1) = 3. \tag{4}$$

Какая оптимальная стратегия соответствует этому значению целевой функции? Для того чтобы ответить на этот вопрос, необходимо использовать таблицы следующим образом: начнем с таблицы

$n=4 \quad c_{sj}+f_3(j)$

Состояние (штат) s \ j	Переход в состояние (штат) 2	3	4	$j_4(s)$	$f_4(s)$
1	2+16	5+12	1+18	3	17

Рис. 8.5. Задача о дилижансах.

для $n=4$ (рис. 8.5) и установим, что оптимальный маршрут из штата *1* проходит через штат *3*. Перейдем к таблице для $n=3$ (рис. 8.4), из которого видно, что, если мистер М. следует через штат *3* (вторая строка), оптимальный путь заключается в переезде в штат *7*. Продолжим рассмотрение, перейдя к таблице для $n=2$ (рис. 8.3), согласно которой оптимальный путь из штата *7* проходит через штат *9*. Наконец, из штата *9* мистер М. следует в штат *10*, где путешествие заканчивается. В целом оптимальным является путь *1—3—7—9—10*, затраты по которому составляют $f_4(1) = 5 + 7 + 1 + 4 = 17$.

Очевидно, динамическое программирование здесь более эффективно, чем прямой перебор всех возможных маршрутов, сопровождаемый их оценкой. В данной задаче имеется 14 возможных путей; чтобы для каждого из них определить стоимость, отвечающую выбранной стратегии, придется суммировать четыре (по одному для каждого шага) соответствующие величины c_{ij}. Следовательно, для простого перебора потребуется 42 ($14 \times 3 = 42$) операции сложения. В то же время в рис. 8.3—8.5 сделано всего лишь 16 операций сложения. Преимущество рекуррентного подхода может оказаться огромным при практическом применении, когда полный перебор обычно неосуществим.

Вопросы для повторения. Метод решения аллегорической задачи о дилижансах основан на ряде идей и подходов, которыми мы будем пользоваться и в дальнейшем. Для того чтобы овладеть ими, при изучении каждой новой модели необходимо попытаться самостоятельно ответить на следующие вопросы:

1. Значения каких переменных определяет выбираемая стратегия, иными словами — какие переменные являются «управляющими»?

2. Какова целевая функция, на основе которой определяется оптимальная стратегия?

3. Каким образом в модели и при анализе учитывается многошаговый характер задачи?

4. Что является характеристикой состояния на каждом шаге?

5. Каким образом вводимые ограничения влияют на состояние системы и на определение допустимых значений управляющих переменных?

Если читатель сумеет сформулировать модель в терминах многошагового процесса, он тем самым выполнит первый этап анализа динамических свойств задачи.

8.3. ПРОСТЕЙШАЯ ЗАДАЧА УПРАВЛЕНИЯ ЗАПАСАМИ

В ранее рассмотренных задачах планирования, например в общей задаче планирования производственной программы (разд. 2.6), решения об управлении запасами определялись большой и сложной системой ограничений. В настоящем разделе рамки рассмотрения существенно сужены, что позволяет сосредоточить внимание на небольшом числе важнейших особенностей динамических процессов управления запасами. Модель, к изучению которой мы приступаем, играет такую же — или *почти* такую же — роль в исследовании операций, как законы Ньютона в физике. Хотя рассматриваемая ситуация и является идеализированной, в модели все же учитывается множество важных факторов, влияющих на выбор правил управления запасами.

Следует помнить, что нашей основной целью в дальнейшем явится анализ динамических свойств процессов управления запасами. Поэтому при описании модели не обсуждается строгость или реалистичность исходных предположений; этим вопросам посвящен разд. 8.5. По этой же причине вводимые экономические понятия сопровождаются лишь краткими пояснениями. Более подробно они рассматриваются в гл. 19 и в приложении 2, которые целиком посвящены моделям управления запасами.

Важно отметить, что внедрение модификаций рассматриваемой модели рядом промышленных фирм дало заметный экономический эффект. Разумеется, в таких случаях все необходимые вычислительные операции выполнялись на ЭВМ.

Пример фирмы «Надежный поставщик». Фирма должна разработать календарную программу выпуска некоторого вида изделий на плановый период, состоящий из N отрезков. Предполагается, что для каждого из этих отрезков имеется точный прогноз спроса на выпускаемую продукцию.

Время изготовления партии изделий настолько мало, что им можно пренебречь; соответственно продукция, изготавливаемая в течение отрезка t, может быть использована для полного или частичного покрытия спроса в течение этого отрезка. Для разных отрезков спрос неодинаков; кроме того, на экономические показатели производства влияют размеры изготовляемых партий. Поэтому фирме нередко бывает выгодно изготовлять в течение некоторого месяца продукцию в объеме, превышающем спрос в пределах этого отрезка, и хранить излишки, используя их для удовлетворения последующего спроса. Вместе с тем хранение возникающих при этом запасов связано с определенными затратами. В зависимости от обстоятельств затраты обусловлены такими факторами, как проценты на капитал, взятый взаймы для создания запасов, арендная плата за складские помещения, страховые взносы и расходы по содержанию запасов. Эти затраты необходимо учитывать и при установлении программы выпуска.

Цель фирмы «Надежный поставщик»— разработать такую программу, при которой общая сумма затрат на производство и на содержание запасов минимизируется при условии полного и своевременного удовлетворения спроса на продукцию. Анализ этой задачи начнем с преобразования качественного ее описания в математическую модель.

Построение модели. Введем переменные:

x_t — выпуск продукции в течение отрезка t;

i_t — уровень запасов на конец отрезка t.

Спрос на продукцию для отрезка t обозначим D_t; предполагается, что величины D_t для всех t отображаются неотрицательными целыми числами и что к началу планового периода все D_t известны.

Предположим также, что для каждого отрезка t затраты зависят от выпуска продукции x_t, уровня запасов i_t на конец отрезка t и, кроме того, возможно, от значения t; обозначим затраты на отрезке t $C_t(x_t, i_t)$. Тогда целевую функцию можно записать в следующем виде:

$$\text{Минимизировать } \sum_{t=1}^{N} C_t(x_t, i_t). \tag{1}$$

На значения переменных x_t и i_t наложено несколько ограничений. Во-первых, предполагается целочисленность объемов выпуска:

$$x_t = 0, 1, 2, 3 \ldots \quad (t = 1, 2, \ldots, N). \tag{2}$$

Во-вторых, предполагается, что для администрации фирмы желателен нулевой уровень запасов на конец отрезка N:

$$i_N = 0 \quad \text{(конечный запас равен нулю).} \tag{3}$$

Наконец, в-третьих, ставится условие полного и своевременного удовлетворения спроса в пределах каждого месяца. Выполнение этого условия можно обеспечить, введя два ограничения. Первое из них назовем «балансовым», поскольку в нем утверждается, что

Уровень запасов на конец отрезка $t \equiv$
$\equiv$ (Уровень запасов на начало отрезка t) +
+ (Выпуск продукции на отрезке t) —
— (Спрос на отрезке t).

Если воспользоваться принятыми условными обозначениями, то это ограничение можно будет записать в следующем виде:

$$i_t = i_{t-1} + x_t - D_t$$

или в более удобном для нас виде:

$$i_{t-1} + x_t - i_t = D_t \quad (t = 1, 2 \ldots, N), \tag{4}$$

где i_0 — заданный уровень запасов на начало планового периода.

Согласно второму вводимому ограничению, обеспечивающему своевременное выполнение фирмой «Надежный поставщик» своих обязательств, уровень запасов на начало каждого отрезка и объем выпуска продукции в течение этого отрезка должны быть достаточно велики для того, чтобы уровень запасов на конец отрезка был бы неотрицательным. На самом же деле требуется не только неотрицательность, но и целочисленность уровней запасов (правда, если предположить целочисленность объемов спроса и выпуска продукции, то предположение о целочисленности уровней запасов не создает дополнительных трудностей). Таким образом, требуется, чтобы

$$i_t = 0, 1, 2, 3 \ldots \quad (t = 1, 2, \ldots, N - 1). \tag{5}$$

Отметим, что ограначение (4) является линейным. Если бы все величины затрат $C_t(x_t, i_t)$ линейно зависели от значений переменных, то, как показано ~~ниже~~, полученная модель была бы эквивалентна сетевой модели и для нее легко удалось бы найти решение с помощью методов, описанных в гл. 7. Однако в большинстве практических случаев применения производственных моделей функция затрат нелинейна. Так, для выпуска партии изделий могут потребоваться дорогостоящие подготовительные операции (переналадка), из-за которых затраты на производство первой единицы партии изделий превышают дополнительные затраты на производство остальных единиц. В тех же случаях, когда объем производства в течение

Периоды	x_1	i_1	x_2	i_2	x_3	i_3	x_4	
1	1	−1						$=D_1-i_0$
2		1	1	−1				$=D_2$
3				1	1	−1		$=D_3$
4						1	1	$=D_4$

Рис. 8.6. Матрица ограничений задачи управления запасами.

некоторого периода превышает нормальную мощность производственного участка, дополнительные затраты на единицу изделия могут возрастать из-за использования сверхурочных работ.

Для того чтобы решить задачу при нелинейности каждой из величин $C_t(x_t, i_t)$, сформулируем ее в терминах динамического программирования.

Пусть $N = 4$. Составим балансовые уравнения (4) для $t = 1, 2, 3, 4$. Матрица этой системы ограничений представлена на рис. 8.6.

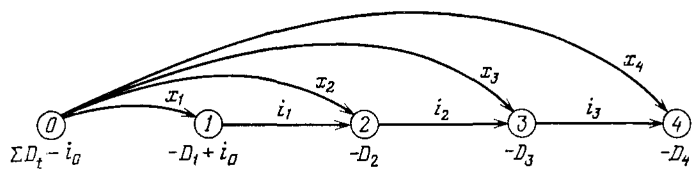


Рис. 8.7. Сеть, отображающая производство продукции и движение запасов.

Построим пятое уравнение, просуммировав четыре уравнения, а затем составим систему из пяти уравнений, содержащую наряду с пятым четыре исходных уравнения, умноженных на −1. Читатель может убедиться в том, что построенная система адекватна сети, изображенной на рис. 8.7.

Динамическая постановка. Вспомним, что в задаче о дилижансах мы строили вычислительный процесс от конечного состояния (сделаны все шаги многошагового процесса) к исходному. Такой же подход будет использован и в настоящей задаче. Здесь конечным состоянием будет начало последнего отрезка планового периода, а исходным — начальный момент первого отрезка (впереди еще N отрезков).

При составлении математической модели удобно использовать систему индексов, при которой подстрочный индекс «1» соответствует конечному, а «N» — *начальному* состоянию. Применим следующие обозначения:

d_n — спрос на продукцию на отрезке n, отстоящем от конца планового периода на n отрезков (включая рассматриваемый);

$c_n(x, j)$ — затраты на отрезке n, связанные с выпуском x единиц продукции и с содержанием запасов, уровень которых *на конец* отрезка равен j единиц.

В этой системе обозначений $d_1 \equiv D_N$ и $d_N \equiv D_1$, а $c_1(x, j) \equiv$ $\equiv C_N(x, j)$.

Пусть $N = 4$, а плановый период начинается с января. Тогда D_1 есть январский спрос, D_4 — апрельский. В модели же используется «обратная система индексов»: январский спрос обозначен d_4, апрельский — d_1. Следовательно, d_2 — мартовский спрос (до конца планового периода — два отрезка, включая март).

Что же определяет состояние системы в начале любого отрезка? Можно считать, что уровень запасов на начало отрезка. Для принятия текущего решения об объеме выпуска не нужно знать, *каким образом* достигнут начальный уровень. Учитывая это обстоятельство, введем следующие обозначения:

$f_n(i)$ — стоимость, отвечающая стратегии минимальных затрат на n оставшихся отрезков при начальном уровне запасов i;

$x_n(i)$ — выпуск, обеспечивающий достижение $f_n(i)$.

Согласно (3), уровень запасов на конец планового периода равен нулю, поэтому можно записать, что

$$f_0(0) = 0 \quad (n = 0). \tag{6}$$

Затем перейдем к $n = 1$. Начальный уровень запасов i может определяться любым неотрицательным целым числом, не большим, чем d_1; вне зависимости от значения i для полного удовлетворения потребности в пределах последнего отрезка объем выпуска должен быть равен $(d_1 - i)$. Следовательно,

$$f_1(i) = c_1(d_1 - i, 0), \quad i = 0, 1, \ldots, d_1. \tag{7}$$

Перейдем к $n = 2$. Заметим, что если начальный уровень запасов равен i, а объем выпуска — x, то общие затраты для двух месяцев составляют

$$c_2(x, i + x - d_2) + f_1(i + x - d_2),$$

причем предполагается, что выбранная стратегия для $n = 1$ была оптимальной. Заметим, что величина $(i + x - d_2)$ есть попросту уровень запасов на *конец* отрезка 2. Величина i может принимать любые неотрицательные целочисленные значения, не превышающие $(d_1 + d_2)$; вопрос к читателю: объясните, почему? При заданном i

целочисленное значение x должно быть не меньше, чем $(d_2 - i)$, что обеспечивает полное удовлетворение потребности на отрезке 2, но не больше, чем $(d_1 + d_2 - i)$, так как конечный запас равен нулю. Оптимальному объему выпуска соответствует такое значение x, при котором минимизируется указанная выше сумма. Выполненный нами анализ ситуации для $n = 2$ можно выразить следующим общим выражением:

$$f_2(i) = \min_x \left[c_2(x,\ i + x - d_2) + f_1(i + x - d_2)\right],$$

где $i = 0, 1, \ldots, d_1 + d_2$, причем для отыскания минимума перебираются все неотрицательные целые значения x, заключенные в пределах $d_2 - i \leqslant x \leqslant d_1 + d_2 - i$.

Как и в задаче о дилижансах, значения $f_3(i)$ можно вычислить, если известны значения $f_2(i)$, и т. д. В конце концов в данной задаче можно вычислить $f_N(i_0)$, где i_0 — уровень запасов на начало планового периода. Общее рекуррентное соотношение записывается в следующем виде:

$$f_n(i) = \min_x \left[c_n(x,\ i + x - d_n) + f_{n-1}(i + x - d_n)\right],$$

$$n = 1,\ 2,\ \ldots,\ N, \qquad (8)$$

где $i = 0, 1, \ldots, d_1 + \ldots + d_n$, причем для отыскания минимума перебираются все неотрицательные целые значения x, заключенные в пределах $d_n - i \leqslant x \leqslant d_1 + d_2 + \ldots + d_n - i$.

Заметим, что, поскольку начальный уровень запасов i рассматривается как переменная величина, полностью характеризующая состояние системы, единственной независимой управляющий переменной в рекуррентном соотношении (8) является x, так как уровень запасов на конец отрезка равен $(i + x - d_n)$. Заметим также, что, поскольку $f_0(0)$ и $f_1(i)$ без труда вычисляются по формулам (6) и (7), можно непосредственно и поочередно вычислить значения $f_2(0)$, $f_2(1), \ldots, f_2(d_1)$, а затем $f_3(0)$, $f_3(1), \ldots, f_3(d_1 + d_2)$. Последовательно переходя ко все большим значениям n, мы дойдем до вычисления $f_{N-1}(0)$, $f_{N-1}(1), \ldots, f_{N-1}(d_1 + d_2 + \ldots + d_{N-1})$ и, наконец, до $f_N(i_0)$.

Для отыскания оптимальной производственной программы определим, какой объем выпуска $x_N(i_0)$ позволяет достичь полученного значения $f_N(i_0)$; соответствующее решение о выпуске является оптимальным решением для начального отрезка планового периода. Уровень запасов на начало следующего отрезка равен $i_0 + x_N(i_0) - d_N$; найдем объем выпуска, позволяющий достичь полученного нами значения $f_{N-1}[i_0 + x_N(i_0) - d_N]$ и т. д. Процесс станет совершенно понятным, когда читатель разберет пример, приведенный в следующем разделе.

Здесь, прежде чем идти дальше, читателю следует удостовериться в том, что ему ясна описываемая постановка задачи в терминах динамического программирования. Процесс принятия решений рассматривается как многошаговый; n — число шагов (в данной задаче число отрезков планового периода) до конца процесса. В иллюстративных целях снова примем $N = 4$, причем эти отрезки соответствуют январю, февралю, марту и апрелю; $n = 1$ относится к апрелю, а $n = 4$ — к январю. В рекуррентном соотношении (8) динамического программирования январский спрос обозначен d_4; аналогичные индексы использованы для целевой функции. До этого момента нет ничего сложного.

Использовано лишь одно новое положение: начальный уровень запасов считается характеристикой состояния системы за n шагов до конца планового периода. Продолжая рассмотрение примера, построенного для четырех месяцев, заметим, что если известен уровень запасов на начало апреля и апрельский спрос, то необходимый объем производства в точности должен быть равен разнице между этими двумя величинами. Такая зависимость отображается уравнением (7). Таким образом, если уровень запасов на начало апреля известен, нахождение оптимального выпуска для этого месяца чрезвычайно просто.

Аналогично этому при известном уровне запасов на начало марта и мартовском спросе необходимый объем выпуска должен быть *не меньше, чем* разность между этими двумя величинами.

В свою очередь принимаемое решение об объеме производства в марте влияет на уровень запасов на начало апреля — этот уровень равен $(i + x - d_2)$. Если последняя величина известна, то можно действовать в апреле оптимальным образом. Однако апрельский выпуск уже был оптимизирован на предыдущем шаге. Поэтому при определении оптимального мартовского объема производства необходимо рассматривать только сумму затрат в марте и соответствующих оптимальных затрат после марта. Вся совокупность этих соображений представлена правой частью рекуррентного соотношения (8) динамического программирования. Те же рассуждения можно повторить для февраля и, наконец, для января.

Рекуррентное соотношение (8) эквивалентно алгоритму нахождения кратчайшего пути на ориентированной ациклической сети. Но какова структура сети, адекватной рассматриваемой модели? Ввиду нелинейности затрат такая сеть не аналогична изображенной на рис. 8.7. В схематическом изображении сети, адекватной соотношению (8) и построенной для реальной задачи, может быть очень много вершин и дуг.

Однако лежащую в ее основе структуру можно охарактеризовать на следующем простом примере.

Пусть

$$d_n = 1 \text{ для } n = 1, 2, 3, 4 \text{ (спрос постоянен во времени)}, \qquad \text{(I)}$$

$$x = 0, 1, 2 \qquad \text{(выпуск ограничен)}. \qquad \text{(II)}$$

Сеть, построенная для этого примера, изображена на рис. 8.8. Вершина, обозначенная индексами (i, n), соответствует возможному значению переменной состояния i за n отрезков до конца планового периода. Пять вершин в левой части сети отображают различные возможности, возникающие при i_0, равном 0, 1, 2, 3 или 4. Наличие

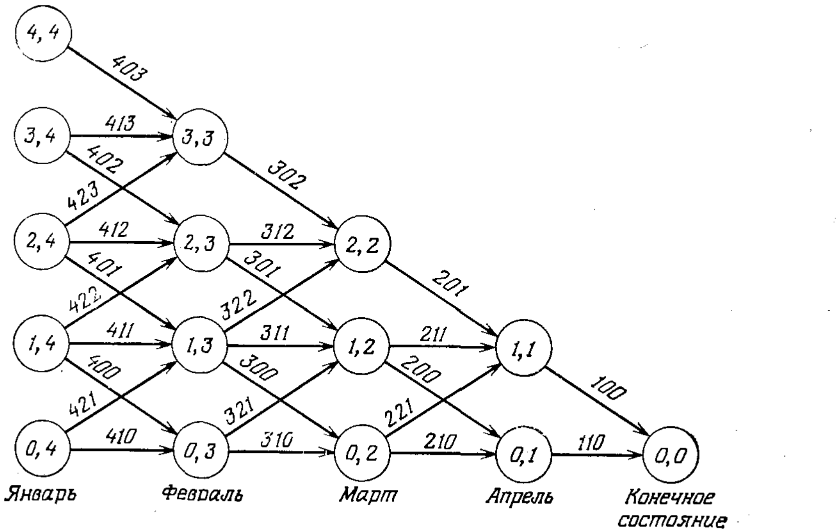


Рис. 8.8. Сеть, отображающая производство продукции и движение запасов.
Индексы вершины: (i, n).
Обозначение на дуге: $c_n(x, j) = nxj$.

только одной вершины в правой части сети обусловлено тем, что конечный запас должен быть нулевым.

Рассмотрим дугу, направленную из вершины (i, n) в вершину $(j, n-1)$. Поскольку $j = i + x - d_n$, то за n отрезков до конца планового периода переменная $x = j - i + d_n$. Каждое допустимое значение переменной отображено одной из дуг, на которой проставлено обозначение «nxj», являющееся сокращением от $c_n(x, j)$. Так, вместо c_4 (0,1) проставлено 401.

Несомненно, схема рис. 8.8 кажется более сложной, чем сеть, изображенная на рис. 8.7. Это объясняется тем, что при нелинейных функциях $C_t(x_t, i_t)$ задача действительно усложняется.

В рекуррентном соотношении (8) последовательность операций *обратна действительной их последовательности во времени.* Это означает, что вычислительный процесс направлен от последнего отрезка планового периода к первому; в примере, где $N = 4$, $f_1(i)$ вычисляется для апреля. Однако можно также разработать и прямой алгоритм, при котором вычислительный процесс направлен от первого отрезка к последнему. В этом случае необходимо задаться некоторым значением исходного уровня запаса i_0. Предположим, что $i_0 = 0$. В случае прямого алгоритма наш подход основывается на вычислении $g_n(i)$:

$g_n(i)$ — минимальные затраты с первого по n-й отрезок при условии, что уровень запасов на *конец* отрезка n от начала планового периода равен i.

При этом

$$g_0(0) = 0, \tag{III}$$

$$g_1(i) = C_1(D_1 + i,\ i)\ (i = 0, 1, \ldots, D_2 + D_3 + \ldots + D_N + i_N), \tag{IV}$$

$$g_n(i) = \min_x [g_{n-1}(i - x + D_n) + C_n(x,\ i)],\ n = 1, 2, \ldots, N, \tag{V}$$

где $i = 0, 1, \ldots, D_{n+1} + \ldots + D_N + i_N$ и для отыскания минимума перебираются все неотрицательные целые значения x, не превышающие $(D_n + i)$. Конечной целью наших вычислений является определение значения $g_N(i_N)$, где i_N — заданный уровень запаса на конец планового периода.

Заметим, что $(i - x + D_n)$ в выражении (V) есть уровень запаса на начало отрезка n. Оптимальное значение x в соотношении (8), позволяющее достичь $f_n(i)$ в течение n последних отрезков планового периода, относится к выпуску первого отрезка. Напротив, оптимальное значение x, позволяющее достичь $g_n(i)$ в соотношении (V), относится к выпуску на отрезке n. Следовательно, после представления $g_n(i)$ для всех n и i в табличном виде мы получим решение, для чего нам придется начать со значения i_N для отрезков N, затем определить соответствующее оптимальное значение x_N, зафиксировать отвечающий ему уровень запасов на начало последнего отрезка, отыскать соответствующее оптимальное значение x_{N-1} и т. д.

8.4. ЧИСЛОВОЙ ПРИМЕР

После того как модель управления запасами сформулирована, читатель подготовлен к решению конкретной задачи, стоящей перед фирмой «Надежный поставщик». Эта задача решается в настоящем разделе. В последующих разделах исследуется вопрос, каким образом увеличение числа отрезков N в плановом периоде влияет на оптимальную стратегию, и показано, что введение ограничения может коренным образом изменить эту стратегию. Данный пример иллюстрирует возможную сложность так называемых комбинаторных задач, даже если их размерность мала.

Справедливости ради нужно отметить, что приведенные ниже численные значения параметров были тщательно подобраны для того, чтобы появились те динамические свойства, которые мы считали желательным продемонстрировать; не следует думать, что подобные свойства присущи процессам управления запасами всегда. Читатель должен понять, что в реальной ситуации только анализ, выполненный методами динамического программирования, позволяет в точности выяснить характер результатов. Это означает, что устойчивость оптимальной программы при изменении длительности планового периода нередко бывает трудно установить *заранее.*

Для упрощения анализа будем считать, что спрос и функция затрат одинаковы для всех отрезков планового периода. Для конкретности примем

$$D_t = 3 \text{ единицам (спрос постоянен во времени).} \tag{1}$$

Предположим также, что затраты равны сумме двух элементов; первый из них относится к производству, а второй определяется стоимостью содержания запасов, которая является линейной функцией объема запасов. Таким образом, для всех отрезков

$$C_t(x_t,\ i_t) = C(x_t) + hi_t, \tag{2}$$

где $C(0) = 0,\ \ C(1) = 15,\ \ C(2) = 17,\ \ C(3) = 19,$

$$C(4) = 21,\ C(5) = 23; \tag{3}$$

$$h = 1. \tag{4}$$

В свою очередь производственные затраты можно рассматривать как сумму условно-постоянных затрат на операции по переналадке (эти затраты равны 13 условным единицам) и пропорциональных затрат (они равны 2 условным единицам на каждую единицу продукции). Поскольку $h = 1$, затраты на содержание запасов численно равны уровню запасов на конец отрезка.

Далее, производственные мощности и складские площади фирмы «Надежный поставщик» ограничены; это вводит в задачу дополнительное усложнение. Примем, что выпуск в течение одного отрезка не может превысить 5 единиц, а уровень запасов на конец отрезка — 4 единицы. Иными словами, для всех отрезков

$$\left.\begin{aligned} x_t &= 0, 1,\ \ldots, 5 \\ \text{и} \qquad i_t &= 0, 1,\ \ldots, 4. \end{aligned}\right\} \tag{5}$$

Заметим, что затраты на переналадку относительно высоки по сравнению с другими элементами; поэтому в оптимальной программе должна появиться тенденция к укрупнению партий. Однако объем выпуска x_t не может превысить 5 единиц, тогда как спрос равен 3; следовательно, в течение одного отрезка уровень запасов не может возрасти более чем на 2 единицы. Таким образом, в течение двух первых отрезков не удается избежать двух переналадок, если исходный запас равен нулю. Вовсе не очевидно, какая программа выпуска окажется оптимальной в случае более длительного планового периода. Ответ на этот вопрос должны дать методы динамического программирования.

Формулировка задачи динамического программирования. При наличии приведенных выше данных об условиях деятельности фирмы

«Надежный поставщик» можно составить динамическое рекуррентное соотношение, отображающее специфику задачи. Напомним, что используются следующие обозначения:

$f_n(i)$ — минимальные затраты в течение n последних отрезков планового периода при начальном уровне запасов i;

$x_n(i)$ — выпуск, позволяющий достичь $f_n(i)$.

Для $n = 1$

$$\left.\begin{aligned} f_1(i) &= C(3-i) \\ x_1(i) &= 3-i \end{aligned}\right\}, \quad i=0,\ 1,\ 2,\ 3, \tag{6}$$

поскольку уровень запасов на конец планового периода равен нулю. В общем виде рекуррентное соотношение можно записать следующим образом:

$$f_n(i) = \min_x [C(x) + 1(i + x - 3) + f_{n-1}(i + x - 3)],$$
$$n = 2, 3, \ldots, \tag{7}$$

где $i = 0, 1, 2, 3, 4$ и для отыскания минимума перебираются все неотрицательные целые значения x, заключенные в пределах $3 - i \leqslant x \leqslant \min(5, 7 - i)$. Ограниченность производственных мощностей, отображаемая первым из условий (5), не позволяет x превысить 5, а ограниченность уровня запасов на *конец* отрезка, отображаемая второй частью (5), не позволяет x превысить $(7 - i)$.

Для того чтобы анализ был содержательным, необходимо располагать всеми значениями функций $f_n(i)$; в связи с этим их вычисление является нашей очередной задачей. Форма расчетных таблиц во многом напоминает таблицы, использованные в задаче о дилижансах. Для каждого шага n построена одна таблица; в ней предусмотрено по одной строке для каждого возможного значения начального уровня запасов i и по одному столбцу — для каждого возможного значения выпуска x. Поскольку спрос на продукцию в пределах каждого отрезка должен быть полностью удовлетворен, а уровень запасов на конец отрезка не может превысить 4 единицы, некоторые клетки в таблицах «запрещены» — они соответствуют недопустимым сочетаниям i и x. Каждое из проставленных в таблице чисел представляет собой сумму затрат для рассматриваемого отрезка n и оптимальных затрат для всех $(n - 1)$ последующих отрезков. В двух правых столбцах таблицы проставлены: минимальная по строке сумма [в столбце $f_n(i)$] и соответствующий ей оптимальный выпуск [в столбце $x_n(i)$].

Значения $f_1(i)$, вычисленные по формуле (6), приведены в таблице рис. 8.9, а значения функции $f_2(i)$ — в таблице рис. 8.10. Рассмотрим структуру последней таблицы более подробно. В ней имеется 5 строк, по одной для каждого допустимого значения i. Клетки, соответствующие некоторым сочетаниям i и x, «запрещены». Так, если $i = 1$, то спрос удастся удовлетворить только при условии $x \geqslant 2$.

Если $i = 4$, то $x \leqslant 2$, иначе нарушится условие нулевого уровня запасов на конец планового периода. Первое из слагаемых в каждой клетке — значение $C(x)$, вычисленное по формуле (3). Второе слагаемое — затраты на содержание запасов, равные уровню запасов на конец отрезка, умноженному на $h = 1$. Так, например, при $i = 3$ и $x = 0$ уровень запасов на конец отрезка также равен нулю; поэтому равно нулю и второе слагаемое в соответствующей клетке. При $i = 3$ и $x = 1$ уровень запасов на конец месяца равен 1; в соответствующей клетке второе слагаемое также равно 1. Аналогичным образом значения вторых слагаемых вычисляются и для других клеток третьей строки ($i = 3$). Наконец, третье слагаемое — это значение $f_1(i + x - 3)$, ранее вычисленное и приведенное в таблице рис. 8.9.

$f_1(i) = C(3 - i)$

	i	$x_1(i)$	$f_1(i)$
	0	3	19
Начальный запас	1	2	17
	2	1	15
	3	0	0

Рис. 8.9. Задача фирмы «Надежный поставщик» ($n = 1$).

Для каждого фиксированного i значение функции $f_2(i)$ представляет собой минимальную из всех сумм в «клетках» данной строки, а $x_2(i)$—соответствующий выпуск. Так, при $i = 1$ и $n = 2$ оптимальный выпуск равен 5 единицам; он позволяет за два месяца достичь затрат, равных 26. Любое другое значение x обусловливает более высокие затраты.

Расчет значений $f_3(i)$ приведен в таблице рис. 8.11. Здесь первое слагаемое равно $[C(x) + 1(i + x - 3)]$, а второе слагаемое есть значение $f_2(i + x - 3)$, взятое из таблицы рис. 8.10. Остальные значения $f_n(i)$ для $n = 4, 5, 6$ представлены в сводной таблице рис. 8.12. Читателю следует проверить, насколько им освоены рекуррентные вычислительные операции метода динамического программирования, построив целиком расчетную таблицу для $f_4(i)$, аналогичную таблице рис. 8.11, и сравнив полученные результататы с теми данными, которые приведены в таблице рис. 8.12. Заметим, что для $n = 4$ оптимальными являются два значения выпуска — 3 единицы и 4 единицы.

8.5 АНАЛИЗ ЧУВСТВИТЕЛЬНОСТИ РЕШЕНИЯ

Численные результаты, приведенные в таблицах разд. 8.4, необходимы для определения оптимальных объемов производства. Наряду с этим в таблицах (рис. 8.9—8.12) содержится множество сведений об устойчивости решения при изменении заданных значений параметров модели, таких, как длительность планового периода или исходный уровень запаса. В связи с этим рассмотрим вопросы чувствительности оптимального решения на том же примере фирмы «Надежный поставщик», для чего воспользуемся приведенными выше результатами.

$C(x)+1(i+x-3)+f_1(i+x-3)$

Выпуск:

Начальный запас i \ x	0	1	2	3	4	5	$x_2(i)$	$f_2(i)$
0				19 + 0 +19	21 + 1 +17	23 + 2 +15	3	38
1			17 + 0 +19	19 + 1 +17	21 + 2 +15	23 + 3 + 0	5	26
2		15 + 0 +19	17 + 1 +17	19 + 2 +15	21 + 3 + 0		4	24
3	0 + 0 +19	15 + 1 +17	17 + 2 +15	19 + 3 + 0			0	19
4	0 + 1 +17	15 + 2 +15	17 + 3 + 0				0	18

Рис. 8.10. Задача фирмы «Надежный поставщик» ($n = 2$).

$[C(x)+1(i+x-3)]+f_2(i+x-3)$

Выпуск:

Начальный запас i \ x	0	1	2	3	4	5	$x_3(i)$	$f_3(i)$
0				19 + 38	22 + 26	25 + 24	4	48
1			17 + 38	20 + 26	23 + 24	26 + 19	5	45
2		15 + 38	18 + 26	21 + 24	24 + 19	27 + 18	4	43
3	0 + 38	16 + 26	19 + 24	22 + 19	25 + 18		0	38
4	1 + 26	17 + 24	20 + 19	23 + 18			0	27

Рис. 8.11. Задача фирмы «Надежный поставщик» ($n = 3$).

Начальный запас	$n=1$		$n=2$		$n=3$		$n=4$		$n=5$		$n=6$	
	$x_1(i)$	$f_1(i)$	$x_2(i)$	$f_2(i)$	$x_3(i)$	$f_3(i)$	$x_4(i)$	$f_4(i)$	$x_5(i)$	$f_5(i)$	$x_6(i)$	$f_6(i)$
0	3	19	3	38	4	48	3, 4	67	5	79	4	96
1	2	17	5	26	5	45	5	64	5	74	5	93
2	1	15	4	24	4	43	5	54	4	72	4	91
3	0	0	0	19	0	38	0	48	0	67	0	79
4			0	18	0	27	0	46	0	65	0	75

Рис. 8.12. Задача фирмы «Надежный поставщик».

Длительность планового периода. Для определенности будем считать, что плановый период начинается в январе. Нас интересует изменение оптимальных месячных объемов выпуска при увеличении числа месяцев N в плановом периоде и, в частности, изменение январского выпуска. Результаты анализа, основанного на данных рис. 8.12, приведены в таблице рис. 8.13; при этом предполагается, что исходный уровень запасов на начало января равен нулю.

Таблица рис. 8.13 построена следующим образом. Январский объем производства (3 единицы) для $N = 1$ взят из первой строки

Длительность планового периода, N	Янв.	Февр.	Март	Апр.	Май	Июнь	Общая сумма затрат	Среднемесячные затраты
1	3						19	19
2	3	3					38	19
3	4	5	0				48	16
4	3 4	4 5	5 0	0 3			67	$16\frac{3}{4}$
5	5	5	0	5	0		79	$15\frac{4}{5}$
6	4	5	0	4	5	0	96	16

Рис. 8.13. Программа выпуска продукции фирмой «Надежный поставщик» ($i_0 = 0$).

таблицы рис. 8.12 при n, равном 1. Январский объем производства (3 единицы) для $N = 2$ взят из той же строки при $n = 2$ и т. д. В случае $N = 6$ январский выпуск равен 4 единицам, а уровень запасов на начало февраля — 1 единице (он равен $i + x - d = 0 + 4 - 3$). Следовательно, февральский выпуск (5 единиц) найдем из второй строки таблицы рис. 8.12 при $n = 5$, поскольку уровень запасов на начало месяца теперь равен 1 единице. В свою очередь это означает, что уровень запасов на начало марта составит 3 единицы ($i + x - d = 1 + 5 - 3 = 3$); поэтому выпуск продукции в марте будет нулевым, как это показано в таблице рис. 8.12 для $i = 3$ и $n = 4$. На основе аналогичных рассуждений определим, что производство в апреле ($n = 3$) должно быть равно 4 единицам, поскольку уровень запасов на начало апреля — нуль ($i + x - d = 3 + 0 - 3 = 0$). При таком выпуске уровень запасов на начало мая будет равен $0 + 4 - 3 = 1$, так что майский выпуск ($n = 2$) составит 5 единиц. Следовательно, в июне оптимален нулевой выпуск, поскольку уровень запасов на начало месяца ($n = 1$) равен $1 + 5 - 3 = 3$ единицам. Читатель может убедиться в том, что минимальная общая сумма затрат для $N = 6$ составит $(21 + 1) + (23 + 3) + (0 + 0) + (21 + 1) + (23 + 3) + (0 + 0) = 96$; эта сумма и проставлена в клетке таблицы рис. 8.12, соответствующей $f_6(0)$.

Анализ оптимальных вариантов производственной программы, приведенных на рис. 8.13, свидетельствует о том, что январский выпуск зависит от длительности планового периода. При возрастании числа месяцев N с 1 до 5 оптимальный январский выпуск возрастает. Однако при $N = 6$ производство в январе должно составить всего лишь 4 единицы (при $N = 5$ эта величина была равна 5 единицам). Таким образом, удлинение планового периода может вызвать как рост, так и сокращение январского объема производства, причем для $N = 4$ имеются две альтернативные оптимальные программы. Из рис. 8.13 ясно, каким образом среднемесячные затраты зависят от N. Отметим, что при увеличении N от 2 до 6 среднемесячные затраты не убывают монотонно, а испытывают колебания.

Оптимальная программа для $N = 5$ заслуживает особого внимания. В этом случае уровень запасов возрастает в январе, феврале и апреле. Таким образом, майский спрос по существу удовлетворяется 2 единицами апрельского выпуска и 1 единицей февральского. При подобных условиях оптимальным оказывается наличие запасов на начало как февраля, так и апреля, хотя в течение *обоих* месяцев фирма несет расходы, связанные с переналадкой.

Исходный уровень запасов. В настоящем подразделе изучается зависимость оптимальной программы от уровня запасов на начало планового периода. Для определения такой зависимости применительно к январскому выпуску рассмотрим таблицу, приведенную на рис. 8.14. Если плановый период включает всего лишь один месяц, каждое увеличение исходного запаса на 1 единицу приводит к снижению на единицу январского выпуска. Однако при $N = 2, 3, 4$ или 6 месяцам увеличение исходного запаса от нуля до 1 приводит не к сокращению, а к росту январского производства. Увеличение же исходного уровня запасов с 1 единицы до 2 единиц приводит к разным результатам в зависимости от значения N: выпуск продукции в январе может сократиться (однако не до объема, соответствующего $i_0 = 0$ при $N = 2$) или остаться прежним (при $N = 4$).

В двух последних столбцах таблицы рис. 8.14 отражено снижение общей суммы затрат при единичном приращении исходного запаса. Рассмотрим, например, длительность планового периода N, равную двум месяцам. При нулевом исходном уровне запасов общая сумма затрат составляет 38 (рис. 8.12). Наличие 1 единицы исходного запаса позволяет снизить затраты до 26, а 2 единиц — до 24. Таким образом, цена первого единичного приращения исходного запаса равна 12, а второго — только 2 (рис. 8.14). Подчеркнем, что эта цена существенно зависит от длительности планового периода, а также от того, рассматривается ли первое или второе единичное приращение исходного запаса. Читатель должен построить оптимальные варианты программы для $N = 4$ и $N = 6$, $i_0 = 1$ и $i_0 = 2$ и выяснить причину наблюдаемых различий в снижении суммы затрат. (Попутно отметим, что при $N = 4$ и $i_0 = 2$ уровень запасов на начало февраля составляет 4 единицы, так что, если бы условиями задачи были бы предусмотрены более строгие ограничения уровня запасов, при которых

Длительность планового периода, N	Январский выпуск			Цена предыдущего единичного приращения исходного запаса	
	$i_0=0$	$i_0=1$	$i_0=2$	$i_0=1$	$i_0=2$
1	3	2	1	2	2
2	3	5	4	12	2
3	4	5	4	3	2
4	3,4	5	5	3	10
5	5	5	4	5	2
6	4	5	4	3	2

Рис. 8.14. Цена единичного приращения исходного запаса фирмы «Надежный поставщик».

должно соблюдаться условие $i < 4$, этот вариант оказался бы запрещенным и общая сумма затрат возросла бы.)

Долгосрочный анализ при «усеченном» плановом периоде. До сих пор изучалась ситуация, при которой длительность планового периода совпадает с длительностью периода существования спроса на данный продукт. Это значит, что варианты программы, приведенные в таблицах на рис. 8.12 и 8.13, относятся к ситуации, при которой $D_t = 0$ для всех $t > N$.

Теперь предположим, что $D_t = 3$ и *после* N месяцев, но *текущие* решения о месячном выпуске определяются только на основе данных для ближайших N месяцев. Например, при $N = 4$ оптимальный вариант январского выпуска выбирается (разумеется, с учетом уровня запасов на начало января) так, *как будто* спрос на продукцию существует только с января по апрель. В феврале процесс повторяется: февральский выпуск определяется с учетом заданного начального запаса, причем так, как если бы плановый период включал четыре месяца (с февраля по май). Такая процедура, иногда называемая **скользящим планированием,** часто применяется в промышленности. Посмотрим теперь, насколько хороши результаты, которые подобная процедура позволяет получить в данном примере.

Начнем с анализа ситуации, в условиях которой величина $D_t=3$ для 10 отрезков, а затем равна нулю. На рис. 8.15 приведены оптимальные варианты программы и соответствующие им значения суммы затрат для разных значений N, причем предполагается, что уровень запасов на начало января равен нулю. Истинный оптимум для $N = 10$ достигается двукратным повторением программы, полученной для $N = 5$ (рис. 8.13). Однако *скользящее планирование не позволяет получить этот оптимальный вариант ни при одном из значений* N. При $N = 4$ возникает несколько вариантов скользящего плана,

Длительность планового периода, N	Янв.	Февр.	Март	Апр.	Май	Июнь	Июль	Авг.	Сент.	Окт.	Общая сумма затрат
1	3	3	3	3	3	3	3	3	3	3	190
2	3	3	3	3	3	3	3	3	3	3	190
3	4	5	0	4	5	0	4	5	0	3	163
4	3 4	3 5	3 0	3 4	3 5	3 0	3 4	4 5	5 0	0 3	181 163
5	5 5	4 4	0 0	5 5	4 4	0 0	3 4	4 5	5 0	0 3	165
6	4 4	5 5	0 0	4 4	5 5	0 0	3 4	4 5	5 0	0 3	163
Оптимальный вариант	5	5	0	5	0	5	5	0	5	0	158

Рис. 8.15. Скользящий план фирмы «Надежный поставщик» для 10-месячного периода ($i_0 = 0$).

поскольку в случае нулевого исходного запаса выпуск может быть принят равным как 3, так и 4 единицам. В таблице рис. 8.15 приведены только два варианта программы, причем затраты на них различны. Наконец, отметим, что при скользящем планировании удлинение планового периода может даже нанести ущерб: затраты при $N = 3$ ниже, чем в случае $N = 5$, хотя ни тот, ни другой вариант программы не оптимален.

Теперь предположим, что величина D_t сохраняет одно и то же значение (3 единицы) в течение бесконечного периода времени и что программа всегда составляется методом скользящего планирования. Тогда (при исходном запасе $i_0 = 0$) производство сводится к постоянно повторяющимся циклам, отображенным на рис. 8.16. Если $i_0 = 0$, то характер функции затрат и наличие различных ограничений приводят к тому, что уровень запасов на конец любого месяца не превышает 2 единиц. Следовательно, если при одинаковых исходных уровнях запасов ($i = 0$, 1 или 2) и при двух различных N получаются одинаковые оптимальные варианты программы, то возникают одинаковые повторяющиеся производственные циклы. В связи с этим в таблице рис. 8.16 учтены только три возможных значения i. Нетрудно заметить, что для $N = 1, 2, \ldots$ существует только семь различных вариантов программы, которые приводят ко всего лишь четырем различным повторяющимся производственным циклам. (Для некоторых значений N существует несколько оптимальных вариантов производственной программы, и соответственно эти значения N повторены необходимое число раз.)

Отметим, что только начиная с $N = 10$ скользящее планирование дает оптимальные варианты программы, обеспечивающие минимальные среднемесячные затраты. Даже при $N = 17$ возникающие повторяющиеся производственные циклы не являются оптимальными,

Длительность планового периода, N*	$x_N(0)$	$x_N(1)$	$x_N(2)$	Повторяющийся производственный цикл					Среднемесячные затраты
1	3	2	1	3					19
2	3	5	4						
4,7	3	5	5						
5, 8, 11, 14, 17	5	5	4	5	4	0			$16\frac{1}{3}$
3, 6, 8, 11, 14, 17	4	5	4	4	5	0			16
4, 7, 9, 12, 13, 14, 16	4	5	5						
10, 13, 14, 15, 16, 18, Оптимальный вариант	5	5	5	5	5	0	5	0	$15\frac{4}{5}$

*Программы для $N \leqslant 6$ приведены в таблице рис. 8.12, а для $N > 6$ разработаны аналогичными методами. Если $i \geqslant 3$, то $x_N(i) = 0$ для любого N.

Рис. 8.16. Скользящий план фирмы «Надежный поставщик» для бесконечного периода ($i_0 = 0$).

и можно показать, что один из альтернативных вариантов для $N = 22$ дает среднемесячные затраты, равные 16 условным единицам. Однако если плановый период достаточно велик ($N \geqslant 18$), оптимальная программа для бесконечного планового периода одновременно является оптимальной и для конечного периода. Это свидетельствует о том, что существует достаточная вероятность ошибиться в оценке необходимой длительности планового периода и такую вероятность можно уменьшить только посредством тщательного анализа.

В гл. 11 и 12 читатель ознакомится с определением оптимальной программы при учете влияния неограниченной длительности планового периода.

Некоторые замечания. В настоящем примере *численные* значения параметров были подобраны специальным образом, однако общая ситуация описана в нем достаточно реалистично. Так, производственные затраты включают условно-постоянные расходы на подготовительные операции, а также пропорциональные (переменные) затраты; затраты на содержание запасов, линейно зависящих от их уровня на конец отрезка; на размеры выпуска и уровень запасов наложены простые ограничения сверху. Тем не менее, например, при изменении длительности планового периода выбор оптимального варианта программы меняется коренным образом. Вряд ли можно установить, насколько часто подобная чрезвычайно высокая чувствительность встречается в реальных ситуациях и насколько серьезны экономические последствия неправильного выбора программы. Однако в данном примере можно понять, что степень и значимость чувствительности решений трудно определить, не выполнив в каждом конкретном случае строгий динамический анализ. Только что изученный подход, основанный на операционных идеях, является фундаментальным методом выполнения подобных исследований.

Известно, что математическая модель нередко является достаточно общей для того, чтобы охватить множество различных реальных ситуаций. Поэтому содержательная оценка модели, приведенной в настоящей главе, должна в первую очередь основываться на анализе системы исходных предположений, а не конкретных условий деятельности фирмы «Надежный поставщик». Перечислим важнейшие из этих предположений.

1. *Прогноз спроса является точным.* Хотя фирме редко удается совершенно точно предсказать спрос на несколько месяцев вперед, размеры ошибки часто достаточно малы и детерминированная модель дает хорошую аппроксимацию действительности. (В подобных обстоятельствах обычно применяется скользящее планирование, описанное выше, однако иногда пересмотр программы при этом осуществляется не ежемесячно, а раз в несколько месяцев.) Если же ошибки прогнозирования существенны, необходим переход к моделям такого типа, который описан в гл. 19 и приложении 2.

2. *Длительность изготовления продукции пренебрежимо мала.* В реальных условиях делается другое предположение, согласно которому можно определять длительность периода производства с пренебрежимо малой ошибкой. Так, пусть изготовление партии изделий длится две недели. Если производственная программа построена с использованием приведенных в настоящей главе рекуррентных соотношений, причем в соответствии с этой программой февральский спрос удовлетворяется февральским выпуском, то на самом деле запуск партии необходимо осуществить на две недели раньше, т. е. во второй половине января.

Другим аспектом данного предположения является возможность определения длительности изготовления партии изделий вне связи с изготовлением других заказов. Если несколько различных видов изделий обрабатываются на одном и том же оборудовании, производственная мощность которого ограниченна, то совокупность программ выпуска, каждая из которых получена с помощью обособленной модели динамического программирования, может оказаться несовместимой.

Модели, приведенные в настоящей главе, часто полезны в тех случаях, когда изделия заказывают у внешнего поставщика, располагающего необходимыми их запасами. Тогда задержка во времени связана с интервалом поставки, а затраты включают не издержки производства, а расходы на закупку изделий.

3. *Затраты по каждому отрезку зависят от текущего выпуска и от уровня запасов на конец отрезка; спрос на каждом отрезке полностью и своевременно удовлетворяется.* Эти два предположения без особого труда можно обобщить на значительно более широкий круг ситуаций. Методы такого обобщения рассматриваются ниже.

Не вводя существенно новых усложнений, можно переписать рекуррентное соотношение динамического программирования (8) в следующем виде:

$$f_n(i) = \min\left[c_n(x, i, i + x - d_n) + f_{n-1}(i + x - d_n)\right]. \qquad \text{(I)}$$

При такой записи соотношение (8) можно распространить на ситуацию, при которой затраты на содержание запасов исчисляются исходя из среднего уровня

$$\frac{i+(i+x-d_n)}{2}. \qquad \text{(II)}$$

Далее можно ввести допущение, согласно которому своевременно неудовлетворенный спрос полностью сохраняется и смещается на последующие отрезки. При таком допущении «балансовое уравнение»

$$i_t = i_{t-1} + x_t - D_t \qquad \text{(III)}$$

по-прежнему остается в силе, но величина i_t может принимать и отрицательные значения, характеризующие накапливающийся неудовлетворенный спрос. Возможно и другое допущение, при котором своевременно неудовлетворенный спрос полностью теряется; в этом случае

$$i_t = \max(i_{t-1} + x_t - D_t,\ 0). \qquad \text{(IV)}$$

Тогда

$$f_n(i) = \min_{x=0,\,1,\,\ldots} \{c_n(x, i, i+x-d_n) + f_{n-1}[\max(i+x-d_n,\ 0)]\}. \qquad \text{(V)}$$

Как в выражении (III), так и в выражении (IV) целевая функция затрат $c_n(x, i, i + x - d_n)$ должна включать штраф за задержку удовлетворения спроса или за полную его потерю. Такой функцией может быть функция (VI):

$$c_n(x, i, i + x - d_n) = c(x) + h[\max(i + x - d_n,\ 0)] - $$
$$- p[\min(i+x-d_n,\ 0)], \qquad \text{(VI)}$$

где $p\ (p > 0)$ представляет собой штраф за единицу спроса, не удовлетворенного в течение того отрезка, когда этот спрос возник.

Очевидно, в случае допущения потери спроса сокращение выручки от реализации будет учитываться в значении p. Эти возможности более подробно рассматриваются в гл. 19 и приложении 2.

В некоторых случаях фирма несет расходы и при переносе выпуска из одного отрезка в другой. Для такой ситуации также можно построить рекуррентное соотношение динамического программирования, однако оно будет более сложным; такое соотношение рассматривается в разд. 9.7.

Если спрос в реальной действительности удается предсказать с достаточно высокой степенью точности, так что будет удобно использовать детерминированные модели управления запасами, то модели, описанные в настоящей главе, вообще говоря, оказываются слишком общими. Как правило, о функции $c_n(x, i, i+x-d_n)$ известна столь детальная информация, что удается получить гораздо более сильные результаты, касающиеся длительности планового периода или характеристик оптимальной стратегии. В результате вычисленные операции по отысканию этой стратегии оказываются намного более простыми, чем рассмотренные в численном примере, изложенном в разд. 8.4. В гл. 9 это положение демонстрируется на примере двух важных частных случаев.

8.6. ПОИСКИ ВОЗМОЖНОСТЕЙ УЛУЧШЕНИЯ ПЛАНА

Если бы читатель отвечал за производство в фирме «Надежный поставщик», его бы весьма обеспокоили полученные выше результаты анализа. Пытаясь выяснить причины повышенной чувствительности оптимальных вариантов производственной программы к изменению параметров, он обратил бы внимание, с одной стороны, на стремление планового отдела избежать высоких затрат на переналадку, что обусловлено характером целевой функции, а с другой — на ограниченность производственных мощностей, допускающих выработку не более 5 единиц в месяц. Вероятно, читатель рассмотрел бы последствия полного отказа от ограничений на объем выпуска и уровень запасов, пользуясь при этом целевой функцией

$$C(x_t) = \begin{cases} 0 & \text{при } x_t = 0, \\ 13+2x_t & \text{при } x_t > 0, \end{cases} \tag{1}$$

значения которой для $x_t = 0, 1, \ldots, 5$ совпадают с ранее заданными. Если при подобных предположениях задаться нулевым значением исходного уровня запасов и по-прежнему принимать $D_t = 3$ единицам, то можно показать, что оптимальный январский выпуск составит 9 единиц; партия этого размера должна изготовляться всякий раз, когда уровень запасов на начало месяца снизится до нуля. Соответствующие минимальные среднемесячные затраты сократятся до $13^1/_3$ — по сравнению с $15^4/_5$, указанными в таблице рис. 8.16.

Допустим, что, стремясь уменьшить отрицательное экономическое влияние ограничения по производственным мощностям, мы открыли возможность повысить месячный выпуск до 6 единиц. Однако соответствующие затраты производства составляют

$$C(6) = 28{,}5,$$

т. е. дополнительные затраты на 6-ю единицу выпускаемой продукции возросли (они равны 28,5 — 23 = 5,5). Очевидно, добавление к числу рассматриваемых вариантов еще одного варианта месячного

Запас на начало, i	$n=1$		$n=2$		$n=3$		$n=4$		$n=5$		$n=6$	
	$x_1(i)$	$f_1(i)$	$x_2(i)$	$f_2(i)$	$x_3(i)$	$f_3(i)$	$x_4(i)$	$f_4(i)$	$x_5(i)$	$f_5(i)$	$x_6(i)$	$f_6(i)$
0	3	19	6	31,5	4	48	6	63	5	79	6	94,5
1	2	17	5	26	5	45	5	57,5	5	74	5	89
2	1	15	4	24	4	43	5	54	4	72	5	85,5
3	0	0	0	19	0	31,5	0	48	0	63	0	79
4			0	18	0	27	0	46	0	58,5	0	75

Рис. 8.17. Результаты для случая возросших производственных мощностей фирмы «Надежный поставщик».

выпуска не может ухудшить результаты, поскольку все ранее имевшиеся варианты сохраняются. Однако окажет ли это добавление большое влияние?

Новые значения $f_n(i)$ и соответствующие им $x_n(i)$ приведены на рис. 8.17. Читателю следует еще раз проверить, насколько хорошо он освоил рекуррентный вычислительный процесс, для чего он должен рассчитать $f_2(i)$ и $f_3(i)$. На рис. 8.18 представлены оптимальные варианты производственной программы для новых условий примера; предполагается что исходный уровень запасов на начало января равен нулю. Сравнив эти результаты с приведенными в таблице рис. 8.13, можно убедиться в том, что варианты программы для $N = 2$, 4 и 6 изменены: теперь в них учтена возможность выпуска 6-й единицы. Вследствие этого минимальные среднемесячные затраты достигаются при длительности планового периода, составляющей два, четыре или шесть месяцев. При N, равном 1, 3 или 5, оптимальный январский выпуск не достигает 6 единиц. Отметим, что приведенная на рис. 8.16 производственная программа для $N = 5$ по-прежнему остается оптимальной и при новых условиях.

Длительность планового периода, N	Янв.	Февр.	Март	Апр.	Май	Июнь	Общая сумма затрат	Среднемесячные затраты
1	3						19	19
2	6	0					31,5	$15\frac{3}{4}$
3	4	5	0				48	16
4	6	0	6	0			63	$15\frac{3}{4}$
5	5	5	0	5	0		79	$15\frac{4}{5}$
6	6	0	6	0	6	0	94,5	$15\frac{3}{4}$

Рис. 8.18. Скользящий план фирмы «Надежный поставщик» для бесконечного периода при возросшей производственной мощности ($i_0=0$).

Влияние величины исходного уровня запасов показано в таблице рис. 8.19. При выборе этого уровня, равным 1 или 2 единицам, оптимальный январский выпуск остается таким же, как на рис. 8.14, однако цена единичных приращений запаса меняется.

Пусть спрос D_t, равный 3 единицам, существует в течение бесконечного времени, причем для разработки программы используется скользящее планирование. Варианты производственной программы для этого случая (при нулевом исходном уровне запасов) приведены

Длительность планового периода, N	Январский выпуск			Цена предыдущего единичного приращения исходного запаса	
	$i_0=0$	$i_0=1$	$i_0=2$	$i_0=1$	$i_0=2$
1	3	2	1	2	2
2	6	5	4	5,5	2
3	4	5	4	3	2
4	6	5	5	5,5	3,5
5	5	5	4	5	2
6	6	5	5	5,5	3,5

Рис. 8.19 Цена единичного приращения исходного запаса фирмы «Надежный поставщик» при возросшей производственной мощности.

в таблице рис. 8.20. Вариант с минимальными среднемесячными затратами возможен при четных значениях N; нечетным значениям N отвечают ранее полученные варианты с более высокими затратами (рис. 8.16).

Рассмотрим более детально различия между оптимальным регулярно повторяющимся производственным циклом для новых и прежних условий (эти циклы приведены в таблицах рис. 8.20 и 8.16). Ранее в течение десяти месяцев выполнялось шесть переналадок,

Длительность планового периода, N	Повторяющийся производственный цикл			Среднемесячные затраты
1	3			19
5	5	4	0	$16\frac{1}{3}$
3	4	5	0	16
2, 4, 6 Оптимальный вариант	6	0		$15\frac{3}{4}$

Рис 8.20. Пересмотренный скользящий план фирмы «Надежный поставщик» для бесконечного периода при возросшей производственной мощности ($i_0 = 0$).

максимальный уровень запасов был равен 4 единицам, причем нулевой (положительный) запас отмечался в восьми месяцах из десяти. При изменившихся условиях партия размером 6 единиц выпускается в каждом нечетном месяце, уровень запасов равен либо 3 единицам, либо нулю. В течение десятимесячного периода выполняется только пять переналадок, максимальный уровень запасов равен 3 единицам, причем запас является нулевым (положительным) всего лишь в пяти месяцах из десяти. Ослабив ограничение, налагаемое на производственные мощности, мы не только снизили среднемесячные затраты, но и получили новую оптимальную программу, которая по своим качественным аспектам является более привлекательной.

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

Упражнения 1—6 связаны с задачей о дилижансах (разд. 8.2).

1. а) Перечислите 14 различных маршрутов из штата *1* (Восток) в штат *10* (Запад) на сети, изображенной на рис. 8.1.

б) Объясните, почему при использовании принципа оптимальности нет необходимости оценивать затраты для всей совокупности перечисленных маршрутов.

2. а) Предположим, что между штатами *7* и *9* не существует дилижансового сообщения. Каким теперь будет оптимальный маршрут из штата *1* в штат *10*?

б) Предположим, что дополнительно введен дилижансовый маршрут, связывающий штаты *3* и *8*. Какой должна быть здесь наименьшая ставка страхового платежа, чтобы мистер М. по-прежнему считал предпочтительным ранее выбранный оптимальный маршрут?

в) Определить диапазон ставок страхового платежа для переезда из штата *1* в штат *3*, в рамках которого ранее выбранный маршрут остается оптимальным. Найти аналогичный диапазон для переезда из штата *3* в штат *7*, а также из штата *2* в штат *6*.

3. Мистер М. полагает, что выбор оптимального маршрута определяется не абсолютными значениями страхового платежа для отдельных участков сети, относящихся к одному и тому же шагу, а разностями между этими значениями. Например, он думает, что выбрал бы тот же оптимальный маршрут, если бы затраты c_{12}, c_{13} и c_{14} изменились на одну и ту же величину, т. е. были бы равны $c_{12} = 2 + \delta$, $c_{13} = 5 + \delta$, $c_{14} = 1 + \delta$, где δ — любая постоянная величина. Прав ли мистер М.? Если нет, объясните причину. Если да, то покажите, чем может оказаться полезным высказанное им соображение.

4. Сестра мистера М., миссис Н., проживает в штате *8*. Найдите оптимальный маршрут из штата *1* в штат *10*, проходящий через штат *8*.

5. Племянник мистера М., мистер П., живет в Сан-Франциско (штат *10*) и хочет отправиться в штат *1*. Предположим, что для каждого из участков сети ставка страхового платежа одинакова при поездке как с Запада на Восток, так и с Востока на Запад.

а) Объясните, почему оптимальный маршрут для мистера П. будет таким же, как для мистера М., но имеющим обратное направление.

б) Выполните рекуррентные вычисления по формуле (2) (разд. 8.2), начав их со штата *1*. (Заметим, что если в настоящем примере мистеру П. предстоит еще один шаг пути ($n = 1$), то он находится либо в штате *2*, либо в штате *3*, либо в штате *4*, а если $n = 4$, то мистер П. находится в штате *10*.)

6. Задача о дилижансах. Найдите оптимальный маршрут из штата *1* в штат *10* для каждого из вариантов задачи, в которых ставки страхового платежа для переезда из штата i в штат j соответственно равны (c_{ij} приведены на рис. 8.1):

а) $c_{ij} + i$ (например, ставки страхового платежа для переезда из штата *1* в штат *2* равны 3, в штат *3* — 6, в штат *4* — 2 и т. д.);

б) $c_{ij} + j$ (например, ставки страхового платежа для переезда из штата *1* в штат *2* равны 4, в штат *3* — 8, в штат *4* — 5 и т. д.);

в) $c_{ij} + j - i$ (например, ставки страхового платежа для переезда из штата *1* в штат *2* равны 3, в штат *3* — 7, в штат *4* — 4 и т. д.);

г) $c_{ij} + i + j$ (например, ставки страхового платежа для переезда из штата *1* в штат *2* равны 5, в штат *3* — 9, в штат *4* — 6 и т.д.).

7. Сравните алгоритм, основанный на формуле (2) из разд. 8.2, с алгоритмом нахождения кратчайшего пути, изложенным в разд. 7.7.

Упражнения 8—20 относятся к модели управления производством и запасами, описанной в разд. 8.3.

8. В каждом из приведенных ниже вариантов задачи известны либо выпуск x_t для отрезка t, либо уровень запасов i_t на конец отрезка t; определите значения неизвестных переменных, будь то i_t или

x_t. Предполагается, что плановый период состоит из 6 отрезков ($N = 6$) и что спрос соответственно равен:
$D_1 = 10, D_2 = 15,\ D_3 = 8,\ D_4 = 25,\ D_5 = 12,\ D_6 = 30$ единицам.

Установите, является ли получаемая программа допустимой, т. е. не отрицательны ли x_t и i_t для всех t (i_0 — уровень запасов на начало отрезка 1).

а) $i_0 = 10,\ \ x_t = 15$ для каждого из отрезков;
б) $i_0 = 5,\ \ x_1 = 20,\ \ x_t = 15$ для $t = 2, 3, \ldots, 6$;
в) $i_0 = 5,\ \ x_t = 15$ для $t = 1, 2, \ldots, 5,\ x_6 = 20$;
г) $i_0 = 1,\ \ x_t = 10$ для $t = 1, 2, 3,\ x_t = 23$, для $t = 4, 5, 6$;
д) $i_0 = 0,\ i_1 = 15,\ i_2 = 20,\ i_3 = 25,\ i_4 = 15,\ i_5 = 5,\ i_6 = 0$;
е) $i_0 = 10,\ i_1 = 15,\ i_2 = 20,\ i_3 = 25,\ i_4 = 15,\ i_5 = 5,\ i_6 = 0$;
ж) $i_0 = 30,\ i_1 = 15,\ i_2 = 20,\ i_3 = 25,\ i_4 = 15,\ i_5 = 5,\ i_6 = 0$;
з) $i_0 = 0,\ i_1 = 10,\ i_2 = 10,\ i_3 = 10,\ i_4 = 10,\ i_5 = 10,\ i_6 = 10$;
и) $i_0 = 35,\ i_1 = 35,\ i_2 = 35,\ i_3 = 35,\ i_4 = 35,\ i_5 = 35,\ i_6 = 0$;
к) $i_0 = 35,\ i_1 = 35,\ i_2 = 35,\ i_3 = 35,\ i_4 = 35,\ i_5 = 10,\ i_6 = 0$;
л) Как изменится программа, если в вариантах д) и ж) — к) $i_6 = 10$?

9. Пусть затраты описываются функцией $C_t\,(x_t,\ i_t)$:

$$C_t\,(x_t,\ i_t) = C\,(x_t) + hi_t,$$

где

$$C\,(x_t) = \begin{cases} 0 & \text{при } x_t = 0, \\ 6 + 10x_t & \text{при } x_t > 0 \end{cases}$$

и $h = 0$. Вычислите общую сумму затрат на производство и содержание запасов для следующих вариантов задачи 8:

а) вариант а), д) вариант е)
б) вариант б), е) вариант з),
в) вариант в), ж) вариант к).
г) вариант д),

10. Пусть известны значения выпуска x_t и уровня запасов, i_t, удовлетворяющие ограничениям (2)—(5) разд. 8.3; напомним, что $C_t\,(x_t,\ i_t)$ есть функция затрат на производство и на содержание запасов, а уровень запасов на конец планового периода $i_N = 0$.

а) Увеличьте x_2 на 1 и уменьшите x_3 на 1, соответствующим образом изменив значения уровней запасов. Объясните, почему измененная программа остается допустимой. Определите изменение общей суммы затрат.

б) Увеличьте x_3 на 1 и уменьшите x_2 на 1, соответствующим образом изменив значения уровней запасов. Объясните, в каких случаях измененная программа может оказаться недопустимой. Укажите, как должен измениться спрос, чтобы полученная программа стала допустимой. Укажите, каковы должны быть эквивалентные изменения уровней запасов. Полагая полученную программу допустимой, определите изменение общей суммы затрат.

в) Увеличьте x_2 на 1 и уменьшите x_4 на 1, изменив соответствующим образом уровни запасов. Объясните, почему измененная программа остается допустимой. Определите изменение общей суммы затрат.

г) Увеличьте уровень запасов i_2 на 1, изменив соответствующим образом объемы выпуска. Объясните, почему в рассмотренных случаях измененная программа может оказаться недопустимой. Укажите, как необходимо изменить спрос для сохранения допустимости новой программы. Полагая полученную программу допустимой, определите изменение общей суммы затрат.

д) Уменьшите уровень запасов i_2 на 1, изменив соответствующим образом объемы выпуска. Объясните, в каких случаях измененная программа может оказаться недопустимой. Укажите условия, выполнение которых необходимо для сохранения допустимости новой программы. Полагая полученную программу допустимой, определите изменение общей суммы затрат.

11. Пусть $N = 6$, составьте уравнение типа уравнения (4) из разд. 8.3 для $t = 1, 2, \ldots, 6$.

а) Представьте систему ограничений модели в виде матрицы, аналогичной изображенной на рис. 8.6.

б) Постройте для данной задачи сеть, аналогичную изображенной на рис. 8.7. (*Указание*: просуммируйте шесть уравнений, перечисленных в п. а), получив в результате этого седьмое уравнение; затем постройте систему из полученного уравнения и из шести исходных уравнений, умноженных на —1.)

в) Предположим, что функция затрат линейна:

$$C_t(x_t, i_t) = C_t x_t + h i_t,$$

где

$$C_1 = 1, \; C_2 = 4, \; C_3 = 3, \; C_4 = 5, \; C_5 = 7, \; C_6 = 4.$$

Приняв $i_0 = 0$, а спрос равным указанному в упражнении 8, найдите оптимальную программу для $h = 0$, $h = 1/2$, $h = 1^1/_2$ и $h = 4$. Укажите альтернативные варианты оптимальной программы, если такие имеются.

12. Объясните устно, почему состояние системы на начало каждого отрезка планового периода полностью характеризуется уровнем запасов. Какие предположения о характере целевой функции и о длительности производства изделий дают возможность столь просто характеризовать состояние системы?

13. Процесс рекуррентных вычислений в динамическом программировании в разд. 8.2 назван «самообеспечением». Объясните устно, каким образом подобный подход использован для решения задачи управления запасами в разд. 8.3.

14. Пусть $N = 6$ и январь является отрезком 1 планового периода. Пусть далее d_n характеризует спрос для месяца, отстоящего на n месяцев от конца планового периода. К какому месяцу относится d_1? d_6? d_5? d_2?

15. Рассмотрите рекуррентное соотношение (8) разд. 8.3. Пусть $i = 0$ за n отрезков до конца планового периода. Каким является наименьшее допустимое значение выпуска x для отрезка n? Пусть вместо этого $i = d_1 + d_2 + \ldots + d_n$ Какова программа выпуска для каждого из оставшихся отрезков?.

16. Рассмотрите рекуррентное соотношение (8) разд. 8.3. Пусть $c_3(x, j) = 5x + 2j$; за 3 отрезка до конца планового периода начальный уровень запасов $i = 4$; уровень запасов на конец каждого отрезка не должен превышать 4 единиц; пусть также $d_3 = 10$. Найдите оптимальную программу производства и соответствующий ей набор уровней запасов на конец отрезка для следующих вариантов задачи:

а) $f_2(0) = 100$, $f_2(1) = 90$, $f_2(2) = 82$, $f_2(3) = 76$, $f_2(4) = 75$;
б) $f_2(0) = 110$, $f_2(1) = 100$, $f_2(2) = 92$, $f_2(3) = 86$, $f_2(4) = 85$;
в) $f_2(j) = 100 - 6j$;
г) $f_2(j) = 100 - 9j$;
д) $f_2(0) = 100$, $f_2(1) = 99$, $f_2(2) = 93$, $f_2(3) = 85$, $f_2(4) = 75$.

17. Напомним, что $x_n(i)$ есть оптимальный выпуск на отрезке, отстоящем на n отрезков от конца планового периода, при начальном уровне запасов i. Пусть $d_n = 2$ для любого n, причем значения $x_n(i)$ равны:

$$\begin{aligned} &x_3(0) = 5, \quad x_2(0) = 4, \quad x_1(0) = 2; \\ &x_3(1) = 4, \quad x_2(1) = 3, \quad x_1(1) = 1; \\ &x_3(2) = 0, \quad x_2(2) = 0, \quad x_1(2) = 0; \\ &x_3(3) = 0, \quad x_2(3) = 0. \end{aligned}$$

Найдите оптимальную производственную программу и соответствующие уровни запасов для вариантов задачи а) — г), если до конца планового периода осталось $n = 3$ отрезкам, а начальный уровень запасов i равен:

а) 0, б) 1 единице, в) 2 единицам, г) 3 единицам.

д) Насколько меняется значение общей суммы затрат при переходе от $i = 2$ к $i = 1$ за $n = 3$ отрезкам от конца планового периода. При переходе от $i = 3$ к $i = 4$? [Для отображения затрат на отрезке, отстоящем на k отрезков от конца планового периода, использовать условное обозначение $c_k(x, j)$.]

18. а) Постройте сеть, аналогичную изображенной на рис. 8.8, для анализа задачи, в которой $d_n = 2$ единицам ($n = 1, 2, 3, 4$).

б) Постройте сеть, аналогичную изображенной на рис. 8.8, для анализа задачи, в которой допустимые значения выпуска x равны 0, 2 или 4 единицам.

в) Объясните, как меняется сеть рис. 8.8, если налагается ограничение, согласно которому уровень запасов на конец каждого из отрезков не должен превышать 1.

г) Объясните, как упрощается сеть рис. 8.8, если известен исходный уровень запасов на начало января, который равен 1.

19. Объясните связь между алгоритмом отыскания кратчайшего пути на сети рис. 8.8 и рекуррентными вычислениями по формуле (8) разд. 8.3.

20. В сети рис. 8.8 измените ориентацию каждой из дуг на обратную. Пусть исходный уровень запасов на начало января равен 1. Объясните, почему отыскание кратчайшего пути от вершины «Конечное состояние» до вершины, соответствующей единичному уровню запасов на начало января, совершенно идентично применению рекуррентного соотношения (IV) в конце разд. 8.3, причем любой из этих методов позволяет получить значение $g_4(0)$.

Упражнения 21—34 относятся к задаче, посвященной фирме «Надежный поставщик» и приведенной в разд. 8.4.

21. Для вариантов а) — в) составьте таблицы, аналогичные таблице рис. 8.11 и дающие значения $f_n(i)$ и $x_n(i)$ для указанных значений n. При построении каждой таблицы использовать значения $f_{n-1}(i)$, приведенные в таблице рис. 8.12.

а) $n = 4$; б) $n = 5$; в) $n = 6$.

22. а) Убедитесь в том, что функция затрат $C(x)$, соответствующая выражению (3) в разд. 8.4, может рассматриваться как функция, включающая затраты на переналадку, равные 13, а также пропорциональные затраты, составляющие 2 на каждую единицу изготовляемой продукции.

б) Объясните, почему выпуск x на любом отрезке должен быть меньше, чем $(3 - i)$, где i — уровень запасов на начало отрезка.

в) Объясните, почему выпуск x на любом из отрезков не должен превышать $(7 - i)$, если уровень запасов на конец отрезка ограничен условием $i \leqslant 4$.

г) Пусть исходный уровень запасов на начало планового периода равен нулю. Вычислите среднемесячные затраты для следующих вариантов производственной программы с месячным выпуском, равным:

3 единицам;

5 единицам в нечетные месяцы и 1 — в четные;

5 единицам, 5 единицам, 2 единицам, 0;

5 единицам, 5 единицам, 0, 2 единицам;

5 единицам, 2 единицам, 5 единицам, 0;

4 единицам в нечетные месяцы и 2 — в четные;

4 единицам, 4 единицам, 4 единицам, 0.

Объясните различия между суммами затрат для этих вариантов.

23. Постройте таблицу, аналогичную изображенной на рис. 8.13, предположив, что исходный уровень запасов на начало января равен:

а) $i_0 = 1$; б) $i_0 = 2$; в) $i_0 = 3$; г) $i_0 = 4$.

д) Построив такую таблицу, проверьте правильность изменения общей суммы затрат, приведенных в таблице рис. 8.14, и дополните эту таблицу результатами, полученными для $i_0 = 3$ и $i_0 = 4$.

24. Постройте таблицу, аналогичную изображенной на рис. 8.12, при условии соблюдения приведенного в а) и б) ограничения на уровень запасов на конец отрезка. (*Указание*: при составлении такой таблицы можно не повторять всех вычислений, разумно используя информацию, которая содержится в таблицах рис. 8.9 — 8.12.)

а) $i \leqslant 5$; б) $i \leqslant 3$.

в) Построив такую таблицу, определите наименьшее значение верхнего предела уровня запасов (3, 4 или 5 единиц), не оказывающего влияния на результаты.

25. Постройте таблицы, аналогичные изображенным на рис. 8.12 и 8.13, для следующих вариантов ограничений производственной мощности:

а) На любом отрезке выпуск x не должен превышать 4 единиц.

б) На любом отрезке выпуск x не должен превышать 6 единиц, причем $C(6) = 25$.

в) Построив таблицы, определите влияние изменений ограничений производственной мощности.

г) Какова оптимальная производственная программа фирмы, если для любого отрезка справедливо ограничение $x \geqslant 1$?

26. Постройте таблицу, аналогичную изображенной на рис. 8.12, при условии, что уровень запасов на конец планового периода равен:

а) 1 единице; б) 2 единицам; в) 3 единицам; г) 4 единицам.

д) Для каждого из вариантов задачи постройте таблицу, аналогичную изображенной на рис. 8.13.

27. Постройте таблицы, аналогичные изображенным на рис. 8.12—8.14, при условии, что для любого отрезка спрос D_t равен:

а) 2 единицам; б) 4 единицам.

в) Объясните, как будет отличаться оптимальная производственная программа при месячном спросе, равном не 3, а 2 или 4 единицам.

28. Постройте таблицы, аналогичные изображенным на рис. 8.12—8.14, для вариантов а) и б) задачи:

а) Элемент функции $C(x)$, которая определяется выражением (3) в разд. 8.4, равен 10 (вместо 15).

б) Затраты на содержание единицы запаса равны 5 (вместо 1).

в) Объясните, какое влияние на выбор оптимальной производственной программы оказывает снижение затрат на переналадку или удорожание содержания запасов.

29. Постройте таблицы, аналогичные изображенным на рис. 8.12 и 8.13, при том условии, что спрос для *последнего* месяца планового периода не равен 3 единицам, а равен цифрам, указанным в вариантах а) и б) (например, при $N = 4$ спрос изменен для апреля). Сравните полученные результаты с производственными программами, приведенными в таблицах рис. 8.12 и 8.13; объясните, как изменится программа, если в январе делается неправильное предположение о том, что спрос последнего месяца планового периода равен 3 единицам вместо значений, указанных в вариантах а) и б):

а) 2 единицы; б) 4 единицы.

30. Спрос задан для десяти месяцев, а производственная программа разрабатывается методом скользящего планирования. Проверьте, будут ли выбраны варианты программы, приведенные в таблице рис. 8.15, если длительность планового периода принята равной:

а) $N = 3$, б) $N = 4$; в) $N = 5$; г) $N = 6$.

31. Спрос задан для бесконечного периода, производственная программа разрабатывается методом скользящего планирования. Проверьте, действительно ли будут выбраны варианты программы и соответствующие им среднемесячные затраты, приведенные в таблице рис. 8.16, если длительность планового периода принята равной:

а) $N = 3$; б) $N = 4$; в) $N = 5$; г) $N = 6$.

32. Пусть функция затрат на содержание запасов (разд. 8.5) модифицирована таким образом, что $h = 1$, но относится к среднему уровню запасов, описываемому формулой (II), которая приведена в конце разд. 8.5. Постройте таблицы, аналогичные изображенным на рис. 8.12 и 8.13, и определите, какое влияние оказывает выбор такого метода исчисления затрат.

33. Предположим, что фирма «Надежный поставщик» отказалась от проводимой ранее политики полного и своевременного удовлетворения спроса. Пусть, далее, функция затрат, включающая штраф за неудовлетворенный спрос, описывается формулой (VI), приведенной в конце разд. 8.5, причем штраф за единицу $p = 10$. Для вариантов а) и б) задачи постройте таблицы, аналогичные изображенным на рис. 8.12 и 8.13, и определите, какое влияние на выбор оптимальной программы окажет введенное изменение правил:

а) Своевременно неудовлетворенный спрос полностью теряется, так что уровень запаса на конец отрезка описывается формулой (IV), приведенной в конце разд. 8.5.

б) Спрос, не удовлетворенный немедленно, полностью сохраняется и должен быть удовлетворен в течение следующего отрезка. Вместе с тем к концу последнего отрезка планового периода не должно существовать неудовлетворенного спроса, так что удовлетворяется весь спрос планового периода. Штраф удерживается на каждом отрезке, спрос по которому своевременно был неудовлетворен.

34. Для каждого из приведенных ниже значений n постройте таблицу, аналогичную изображенной на рис. 8.11 и содержащую $f_n(i)$ и $x_n(i)$. Используйте значения $f_{n-1}(i)$, приведенные в таблице рис. 8.17.

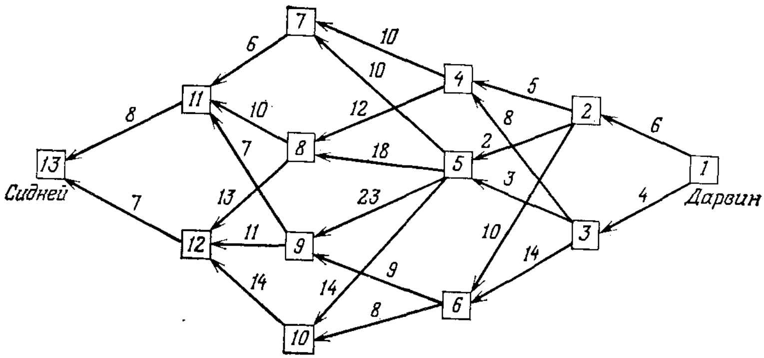
а) $n = 2$; b) $n = 3$; в) $n = 4$; г) $n = 5$; д) $n = 6$.

35. Объясните, как вы понимаете следующие термины:
динамическое программирование;
принцип оптимальности;
рекуррентное соотношение;
рекуррентный алгоритм;
оптимальная стратегия;
шаг;
состояние;
направленность, обратная движению во времени;
скользящий план.

УПРАЖНЕНИЯ НА РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

36. Мистер М. собирается совершить туристическую поездку и хочет отыскать наиболее экономный маршрут от Дарвина (вершина *1*) до Сиднея (вершина *13*) на сети, изображенной на рис. 8.21; затраты для отдельных участков сети, в долларах проставлены на соответствующих дугах. Найдите этот маршрут; вычисления представьте в виде таблиц, аналогичных приведенным на рис. 8.2—8.5.

37. Пусть на сети рис. 8.21 проставленные на дугах цифры характеризуют затраты при поездке как в одном, так и в обратном направлении. Найдите оптимальный маршрут из Сиднея в Дарвин.



Р и с. 8.21.

Представьте расчеты в виде таблиц, аналогичных приведенным на рис. 8.2—8.5.

38. По дороге из Дарвина в Сидней мистер М. хотел бы заехать в район *5*, так как только там можно увидеть самых крупных кенгуру. Найдите маршрут из Дарвина в Сидней, проходящий через район *5*.

Упражнения 39—44 относятся к отдельным частным случаям модели управления запасами, которая описана в разд. 8.3 и характеризуется применением рекуррентного соотношения (8). В каждом из этих упражнений затраты на отрезке t описываются выражением

$$C_t(x_t, i_t) = C_t(x_t) + h_t i_t.$$

Отметим, что отрезком 1 здесь именуется тот месяц, с которого начинается плановый период — например, январь. Для указанных значений длительности планового периода постройте таблицы, аналогичные приведенным на рис. 8.12 и 8.13. (*Примечание*: более эффективные алгоритмы решения задач такого типа описаны в гл. 9.)

39. Пусть $N = 4$, спрос $D_t = 1$, а затраты на содержание запасов для всех отрезков равны $h_t = 0{,}1$. Пусть, далее,

$$C_1(0) = 0,\ C_1(1) = 5,\ C_1(2) = 10,\ C_1(3) = 15,\ C_1(4) = 16;$$

$$C_2(0) = 0,\ C_2(1) = 6,\ C_2(2) = 9,\quad C_2(3) = 12;$$
$$C_3(0) = 0,\ C_3(1) = 5,\ C_3(2) = 7;$$
$$C_4(0) = 0,\ C_4(1) = 3.$$

40. Пусть $N = 5$; $D_1 = 1$, $D_2 = 1$, $D_3 = 2$, $D_4 = 3$, $D_5 = 4$; затраты на содержание запасов для всех отрезков составляют $h_t = 1$. Производственные затраты равны

$$C_t(x_t) = \begin{cases} 0 & \text{при } x_t = 0, \\ s_t + c_t x_t & \text{при } x_t = 1,\ 2,\ 3,\ \ldots, \end{cases}$$

где для всех t $c_t = 10$ и

$$s_1 = 1,\ s_2 = 2,\ s_3 = 4,\ s_4 = 8,\ s_5 = 7.$$

41. Пусть $N = 4$; $D_1 = 1$, $D_2 = 4$, $D_3 = 2$, $D_4 = 2$; $h_t = 0$ для всех месяцев. Производственные затраты описываются функцией, которая по форме идентична приведенной в упражнении 40, причем

$$s_1 = 1,\ s_2 = 12,\ s_3 = 1,\ s_4 = 2;$$
$$c_1 = 3,\ c_2 = 1,\ c_3 = 2,\ c_4 = 1.$$

42. Решите задачу фирмы «С дальным прицелом», рассмотренную в упражнении 28 гл. 6 и в упражнении 48 гл. 7.

43. Фирма может выпускать ограниченное число изделий, оплачивая труд рабочих в течение рабочего дня по обычным расценкам, и увеличивать выпуск (тоже ограниченно), оплачивая работу по сверхурочным расценкам. Численные данные приведены в таблице рис. 8.22. Так, например,

$$C_1(x_1) = \begin{cases} 2x_1 & \text{при } x_1 = 0,\ 1,\ 2,\ 3, \\ 2\cdot 3 + 5(x_1 - 3) = 5x_1 - 9 & \text{при } x_1 = 4,\ 5,\ \ldots,\ 9. \end{cases}$$

		Янв.	Февр.	Март	Апр.	Май	Июнь
Работа в течение рабочего дня	Затраты на единицу	2	4	2	5	2	6
	Мощность	3	1	4	3	1	3
Работа в сверхурочное время	Затраты на единицу	5	6	6	6	3	7
	Мощность	6	3	3	2	0	1
Спрос, D_t		1	2	7	6	0	2

Рис. 8.22.

Различия в значениях затрат для разных отрезков обусловлены специфическими условиями спроса и предложения рабочей силы той квалификации, которая нужна фирме, а также изменением цен на сырье. Предполагаем, что затраты на содержание запасов исчисляются, исходя из уровня запасов на конец каждого отрезка, причем $h_t = 1$ для всех отрезков.

44. Пусть длительность планового периода N равна 3, спрос составляет:

$$D_1 = 3,\ D_2 = 6,\ D_3 = 3,$$

а затраты на содержание запасов не учитываются. Производственные затраты характеризуются приводимыми ниже данными:

	Январь	Февраль	Март
Затраты на переналадку	3	6	20
Пропорциональные затраты (в расчете на 1 изделие) при месячном выпуске:	от 1 до 5 изделий равны 1 от 6 до 10 изделий равны 5 начиная с 11-го изделия, равны 30	от 1 до 5 изделий равны 1 начиная с 6-го изделия, равны 25	за 1-е изделие равны 1 начиная со 2-го изделия, равны 30

Постройте таблицы, аналогичные приведенным на рис. 8.12 и 8.13.

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

45. Рассмотрим модель управления запасами, описанную в разд. 8.3. Пусть $\bar{x}$ — «программный» выпуск продукции для каждого отрезка; при отклонении фактического выпуска x от $\bar{x}$ фирма платит «издержки сглаживания» $v \cdot |x - \bar{x}|$, причем $\bar{x}$ — заранее установленная постоянная величина.

а) Покажите, каким образом следует изменить рекуррентное соотношение (8), для того чтобы учесть эти издержки сглаживания.

б) Пусть $v = 1$, $\bar{x} = 3$, а остальные данные взяты из разд. 8.4. Постройте таблицы, аналогичные изображенным на рис. 8.12 и 8.13, и выясните влияние учета издержек сглаживания.

в) Выполните задание б) при $\bar{x} = 2$.

г) Рассмотрите трудности, возникающие при постановке задачи в терминах динамического программирования, если издержки сглаживания равны $v \cdot |x_t - x_{t-1}|$. (Этот случай описан в гл. 9.)

46. Покажите способы модификации модели управления запасами (разд. 8.3) и рекуррентного соотношения (8), чтобы учесть:

а) порчу запасов, при которой их уровень на начало отрезка t меньше, чем на конец отрезка $(t - 1)$;

б) производственный брак, в результате которого при объеме выпуска x число годных изделий меньше x;

в) возможности «отрицательного спроса», т. е. возврата изделий;

г) возможности «отрицательного выпуска», т. е. использования изделий для внутренних нужд фирмы;

д) зависимости производственных затрат в течение отрезка t от объема выпуска не только на отрезке t, но и на отрезке $(t - 1)$.

47. Покажите способ модификации модели управления запасами разд. 8.3 и рекуррентного соотношения (8), если R_t (y_t) есть общая выручка от реализации y_t изделий в течение отрезка t, причем y_t — управляющая переменная. (Предполагается, что реализация продукции в течение отрезка t не может превышать уровня запасов на конец этого отрезка.)

В упражнении 48 требуется построить модель, использующую рекуррентное соотношение динамического программирования. Дайте определение всем используемым в модели величинам, получившим условные обозначения, и ответьте на пять вопросов, поставленных в конце разд. 8.2. Постройте целевую функцию, соответствующую состоянию за n отрезков до конца планового периода, и в частности при $n = 1$. Объясните, с чего начинаются и чем завершаются вычисления.

48. Рассмотрите случай фирмы «Перманентный рейс», описанный в упражнении 26 гл. 2. Постройте модель (но не решайте задачу) на основе рекуррентного соотношения динамического программирования.

ГЛАВА 9

Динамические оптимизационные модели управления запасами[1])

9.1. ИСПОЛЬЗОВАНИЕ ОСОБЕННОСТЕЙ СТРУКТУРЫ

В настоящей главе продолжается анализ детерминированной модели управления запасами, описанной в разд. 8.3. Будет показано, что при определенных свойствах целевой функции удается характеризовать оптимальную программу значительно полнее. В частности, можно установить ее вид: в свою очередь, обладая такими познаниями, можно разработать упрощенный вычислительный алгоритм построения такой программы.

В гл. 9 нами принято, что целевая функция для любого отрезка t имеет следующий вид:

$$C_t(x_t, i_t) = C_t(x_t) + h_t(i_t), \tag{1}$$

где

$$C_t(x_t) \geqslant 0, \ C_t(0) = 0, \ h_t(i_t) \geqslant 0, \ h_t(0) = 0. \tag{2}$$

Таким образом, для любого отрезка общая сумма затрат представляет собой сумму затрат $C_t(x_t)$ на производство x_t изделий и затрат $h_t(i_t)$, обусловленных наличием запаса i_t единиц на конец отрезка. Балансовые уравнения для любого отрезка имеют вид

$$i_t = i_{t-1} + x_t - D_t, \tag{3}$$

причем постулируется, что $i_0 = 0$, а все D_t — неотрицательные целые числа. Уравнение (3) можно записать и в другом виде:

$$i_t = i_0 + \sum_{k=1}^{t} x_k - \sum_{k=1}^{t} D_k. \tag{4}$$

Наконец, выдвигается требование

$$\text{Неотрицательности и целочисленности } x_t \text{ и } i_t, \tag{5}$$

при выполнении которого спрос удовлетворяется полностью и своевременно.

9.2. ВЫПУКЛЫЕ И ВОГНУТЫЕ ЦЕЛЕВЫЕ ФУНКЦИИ

В последующих разделах делаются некоторые дополнительные предположения о виде функций $C_t(x_t)$ и $h_t(i_t)$. В частности, будем различать два важных класса этих функций.

[1]) В настоящей главе основным предметом рассмотрения является анализ управления запасами, а второстепенным — динамический анализ. Целостность восприятия не уменьшится, если читатель временно опустит эту главу и вернется к ее изучению, перед тем как приступить к гл. 19 и приложению 2 (том 3).

Функция $y(x)$, определенная при целых значениях x, называется **выпуклой**, *если*

$$g(x+1) - g(x) \geqslant g(x) - g(x-1), \tag{1}$$

и **вогнутой**, *если*

$$g(x+1) - g(x) \leqslant g(x) - g(x-1). \tag{2}$$

На рис. 9.1 приведены графики нескольких выпуклых и вогнутых вниз функций.

Если рассматривать $g(x)$ как общую сумму затрат, то функция затрат является выпуклой, если каждая *дополнительная* единица

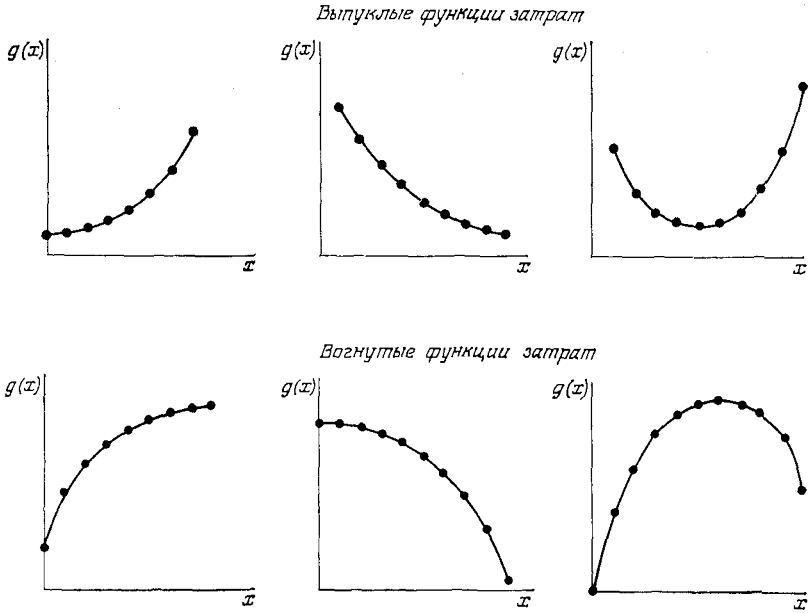


Рис. 9.1.

стóит не меньше предыдущей. Аналогично функция затрат является вогнутой, если каждая *дополнительная* единица стóит не больше предыдущей. Иногда допустимые значения ограничены некоторыми пределами, например можно ввести условия $x = a, a+1, \ldots, b-1, b$ (где $a + 1 \leqslant b - 1$). Тогда приведенные выше определения относятся к значениям x, лежащим в пределах $a + 1 \leqslant x \leqslant b - 1$.

При изучении вида какой-либо конкретной функции удобно переписать условия (1) и (2) соответственно в виде (3) и (4):

$$\frac{g(x+1)+g(x-1)}{2} \geqslant g(x) \quad \text{(функция выпукла)}, \tag{3}$$

$$\frac{g(x+1)+g(x-1)}{2} \leqslant g(x) \quad \text{(функция вогнута)}. \tag{4}$$

Рассмотрим шесть примеров функции $g(x)$; читателю рекомендуется — и будет полезно — начертить график каждой из этих функций. Для того чтобы представить ему такую возможность, в первых четырех примерах зададимся следующими численными значениями параметров: $a = 2$, $b = 1$, $c = 3$.

Пример I. Пусть $g(x) = ax + b$. Тогда

$$\frac{[a(x+1)+b]+[a(x-1)+b]}{2} = ax+b, \tag{5}$$

так что линейная функция $g(x)$ для любых значений a, b и x *одновременно* является *и* выпуклой, *и* вогнутой.

Пример II. Пусть $g(x) = ax^2 + b$. Тогда

$$\frac{[a(x+1)^2+b]+[a(x-1)^2+b]}{2} = ax^2+a+b \begin{cases} \geqslant ax^2+b & \text{при } a \geqslant 0, \\ \leqslant ax^2+b & \text{при } a \leqslant 0. \end{cases} \tag{6}$$

Таким образом, для любых значений x функция $g(x)$ является выпуклой при неотрицательном a и вогнутой — при неположительном a.

Пример III. Пусть

$$g(x) = \begin{cases} ax+b & \text{при } x \geqslant 0, \\ -cx+b & \text{при } x \leqslant 0, \end{cases}$$

где $a \geqslant 0$ и $c \geqslant 0$. Из сравнения с примером I очевидно, что функция $g(x)$ соответствует (3) и является выпуклой как для $x < 0$, так и для $x > 0$. При $x = 0$

$$\frac{[a+b]+[c+b]}{2} = \frac{[a+c]}{2} + b \geqslant b, \tag{7}$$

так что $g(x)$ является выпуклой для всех x.

Пример IV. Пусть

$$g(x) = \begin{cases} 0 & \text{при } x = 0, \\ ax+b & \text{при } x \geqslant 1, \end{cases}$$

где $b \geqslant 0$. Из сравнения с примером I становится ясно, что при $x > 1$ функция $g(x)$ соответствует (4) и является вогнутой. При $x = 1$

$$\frac{[2a+b]+[0]}{2} \leqslant a+b, \tag{8}$$

так что для неотрицательных x функция $g(x)$ является вогнутой.

Пример V. Пусть

$$g(x) = \begin{cases} 0 & \text{при } x = 0, \\ 13+2x & \text{при } x = 1,\ 2,\ 3,\ 4,\ 5, \\ 28{,}5 & \text{при } x = 6. \end{cases}$$

Из сравнения с примером IV понятно, что для $x \leqslant 4$ выполняется условие (4). При $x = 5$

$$\frac{[28{,}5]+[21]}{2} = 24{,}75 > 23, \tag{9}$$

так что выполняется условие (3). Следовательно, $g(x)$ не является ни выпуклой, ни вогнутой.

Пример VI. Читатель продемонстрирует понимание проведенного выше анализа, показав, что кусочно-линейная функция

$$g(x) = \begin{cases} a_1 x + b_1 & \text{при } 0 \leqslant x \leqslant w_1, \\ a_1 w_1 + a_2 (x - w_1) + b_1 & \text{при } w_1 \leqslant x \leqslant w_2, \\ a_1 w_1 + a_2 (w_2 - w_1) + a_3 (x - w_2) + b_1 & \text{при } w_2 \leqslant x \end{cases} \tag{10}$$

является выпуклой в случае $a_1 \leqslant a_2 \leqslant a_3$ и вогнутой в случае $a_1 \geqslant a_2 \geqslant a_3$. (Выполните анализ при $x = w_1$ и $x = w_2$. Начертите графики функции, приняв $a_1 = 2$, $b_1 = 1$, $w_1 = 4$, $a_2 = 3$, $w_2 = 6$, $a_3 = 4$.)

Некоторые из рассмотренных выше частных случаев представляют особый интерес. Так, пример III охватывает и функцию

$$h_t(i_t) = \begin{cases} h_t i_t & \text{при } i_t \geqslant 0, \\ -p_t i_t & \text{при } i_t \leqslant 0, \end{cases} \tag{11}$$

где h_t ($h_t \geqslant 0$) есть стоимость содержания единицы запасов, а p_t ($p_t \geqslant 0$) — штраф за несвоевременное удовлетворение спроса (в расчете на единицу). Пример V является обобщением примера, приведенного в разд. 8.6. Наконец, пример VI охватывает и функцию

$$C_t(x_t) = \begin{cases} r_t x_t & \text{при } 0 \leqslant x_t \leqslant u_t, \\ r_t u_t + 1{,}5 r_t (x_t - u_t) & \text{при } u_t \leqslant x_t \leqslant v_t, \\ r_t u_t + 1{,}5 r_t (v_t - u_t) + 2 r_t (x_t - v_t) & \text{при } v_t \leqslant x_t, \end{cases} \tag{12}$$

где x_t — общее число человеко-часов, предусмотренных программой для отрезка t; r_t — основная ставка за человеко-час работы; u_t — заданная верхняя граница числа человеко-часов на отрезке t, отрабатываемых в основное время; $1{,}5\ r_t$ —«полуторная» ставка за человеко-час сверхурочной работы; v_t — заданная верхняя граница количества сверхурочных часов для отрезка t, а $2r_t$ —«удвоенная» основная часовая ставка.

9.3. МОДЕЛЬ УПРАВЛЕНИЯ ЗАПАСАМИ С ВЫПУКЛОЙ ФУНКЦИЕЙ ЗАТРАТ

В дополнение к допущениям о затратах, запасах и выпуске, сделанным в 9.1, предположим, что:

Функция производственных затрат $C_t(x_t)$ является выпуклой, (1)

Функция затрат на содержание запасов $h_t(i_t)$ является выпуклой. (2)

Случай выпуклости функции затрат иногда называют **случаем убывающей предельной эффективности при увеличении масштаба.** При выпуклости функции затрат можно задаться как ограничением u_t на объем выпуска x_t, так и ограничением b_t на уровень запасов i_t. Просто это можно объяснить так: заменим эти ограничения введением бесконечно больших величин $C_t(x_t)$ для $x_t > u_t$ и бесконечно больших величин $h_t(i_t)$ для $i_t > b_t$. При такой замене выпуклость (1) и (2) сохраняется. Наконец, отметим, что проводимый анализ может быть обобщен на случай разрешенной отсрочки (задержки) в удовлетворении спроса.

Алгоритм вкратце можно описать следующим образом: поочередно рассматривается спрос на каждом из отрезков, начиная с отрезка 1. Каждая пока еще не удовлетворенная единица спроса удовлетворяется самым дешевым из возможных способов, причем эти возможности определяются с учетом использования производственных мощностей, уже предусмотренного в построенной части программы, а также соответствующей схемы движения запасов. Разумеется, возможность получения оптимальной программы с помощью столь простого алгоритма существенно зависит от предположений (1) и (2) о выпуклости целевой функции.

Опишем алгоритм более подробно.

Шаг 1. Пусть p — первый из отрезков, для которого пока еще не удовлетворенный спрос положителен ($D_p > 0$). Для каждого из отрезков $1, 2, \ldots, p$ рассмотрим такое приращение выпуска на единицу по сравнению с уже построенной частью производственной программы, целью которого является удовлетворение единичного спроса на отрезке p.

Шаг 2. Для каждого из p вариантов единичного приращения выпуска на одном из p отрезков вычислим общее увеличение затрат, вызванное приращением выпуска *и* запасов. Выберем вариант с минимальным увеличением затрат и соответствующим образом изменим уже построенную часть программы. Если имеется несколько таких вариантов, выберем вариант с возможно более поздним увеличением выпуска.

Шаг 3. Уменьшим значение пока еще не удовлетворенного спроса D_p на единицу. Проверим, останется ли при этом *хотя бы одно* ненулевое значение D_t. Если останется — перейти к шагу 1; если нет — прекратить вычисления.

Продемонстрируем применение этого алгоритма на примере.

Задача фирмы «Вечная шина». Фирма «Вечная шина» составляет программу выпуска шин на шесть месяцев (с января по июнь). Данные о спросе по месяцам и о совокупном спросе (нарастающим итогом с начала года) соответственно приведены в предпоследней и последней строках таблицы рис. 9.2. Заметим, что *совокупный* спрос за шесть месяцев равен 18 единицам, т. е. совпадает с совокупным спросом фирмы «Надежный поставщик» (разд. 8.4).

В каждом месяце существует возможность выпуска заданного числа шин при нормальном уровне затрат и некоторого дополнительного их числа — при повышенном уровне затрат. Для того чтобы продемонстрировать гибкость алгоритма, примем в данном примере, что как «нормальные» и повышенные затраты, так и верхние границы выпуска при этих затратах неодинаковы для разных месяцев. На рис. 9.2 нормальные и повышенные затраты проставлены в каждой паре клеток, расположенной над диагональю таблицы (верхний показатель — нормальные затраты). Соответствующие верхние границы

	Янв.	Февр.	Март	Апр.	Май	Июнь	Максимальный выпуск	Максимальный совокупный выпуск
Янв.	2	3	4	5	6	7	3	9
	5	6	7	8	9	10	6	
Февр.		4	5	6	7	8	1	13
		6	7	8	9	10	3	
Март			2	3	4	5	4	20
			6	7	8	9	3	
Апр.				5	6	7	3	25
				6	7	8	2	
Май					2	3	1	26
					3	4	0	
Июнь						6	3	30
						7	1	
Спрос D_t	1	2	7	6	0	2		
Совокупный спрос	1	3	10	16	16	18		

Рис. 9.2. Модель с выпуклой функцией затрат фирмы «Вечная шина».

выпуска для каждого месяца приведены во втором столбце справа, а совокупная производственная мощность, т. е. верхние границы выпуска с начала года,— в первом столбце справа.

Так, в январе можно выработать не более 3 единиц при затратах, равных 2 за единицу, и дополнительно до 6 единиц при затратах 5 за единицу. В феврале возможен выпуск 1 единицы при затратах 4 и дополнительно до 3 единиц при затратах 6 за единицу. В марте можно выработать до 4 единиц при затратах 2 и дополнительно до 3 единиц при затратах 6 за каждую и т. д. Таким образом, производственные затраты каждого месяца отображаются функцией вида

$$C_t(x_t) = r_t x_t \quad \text{при } 0 \leqslant x_t \leqslant u_t \text{ (работа в основное время)}, \tag{3}$$

$$C_t(x_t) = r_t u_t + s_t(x_t - u_t) = s_t x_t + (r_t - s_t)u_t$$
$$\text{при } u_t \leqslant x_t \leqslant v_t$$
$$\text{(работа в сверхурочное время)}, \tag{4}$$

причем $r_t < s_t$, так что функция $C_t(x_t)$ является выпуклой. В выражениях (3) и (4) r_t — нормальные, а s_t — повышенные затраты, u_t и v_t — соответствующие верхние границы выпуска (например, для $t = 1$ значения этих параметров равны: $r_1 = 2$, $s_1 = 5$, $u_1 = 3$ единицам, $v_1 = 3 + 6 = 9$ единицам).

В таблице на рис. 9.2 каждому месяцу соответствуют две строки и один столбец; в строках проставляется выпуск продукции в этом месяце, тогда как числа в столбце характеризуют удовлетворение спроса рассматриваемого месяца. Поскольку задержка удовлетворения спроса не разрешена, клетки под диагональю являются «запрещенными». Таким образом, первые две строки таблицы рис. 9.2 показывают, что январский *выпуск* можно направить на удовлетворение спроса любого из шести месяцев, вторые две строки — что февральским *выпуском* можно удовлетворить спрос февраля и любого из последующих месяцев, и т. д. Аналогично июньский *спрос* можно удовлетворить выпуском продукции в любом месяце и т. п. Допустимую программу можно построить, если для каждого месяца соответствующий совокупный возможный выпуск с начала года не меньше совокупного спроса за этот же период. Читателю рекомендуется сравнить числа первого столбца справа и нижней строки таблицы рис. 9.2, чтобы убедиться в выполнении этого условия применительно к рассматриваемой задаче.

Описываемый далее вычислительный алгоритм несколько упрощен по сравнению с трехшаговым алгоритмом, приведенным выше. Такое упрощение оказывается возможным благодаря использованию функций (3) и (4) и линейности затрат на содержание запасов, а именно:

$$h_t(i_t) = h_t i_t \quad \text{(затраты на содержание запасов линейны)}. \tag{5}$$

Для упрощения численного примера положим для всех месяцев $h_t = 1$. Показатели затрат в каждой строке таблицы рис. 9.2 ежемесячно возрастают на 1, что отражает месячные затраты на содержание запасов. Например, если изделие вырабатывается в январе при основном уровне затрат (2 за единицу) и хранится в течение февраля, то общая стоимость единицы становится равной 3 ($2 + 1 = 3$); это число и проставлено в клетке первой строки таблицы и февральского столбца. Если это изделие хранится еще один месяц, затраты возрастают на 1 и становятся равными 4 ($3 + 1 = 4$). Именно это число проставлено в клетке первой строки и мартовского столбца. Попутно отметим, что нет необходимости задаваться условием равенства нулю запаса на конец планового периода ($i_N = 0$), потому что в оптимальной программе выполнение этого условия обеспечивается сделанными допущениями (3), (4) и (5) о виде функции затрат.

Для того чтобы ознакомиться с алгоритмом во всех деталях, читателю следует начертить таблицу рис. 9.2. Заметим, что таблица напоминает матрицу транспортной задачи. Это сходство неслучайно: линейность выражений (3), (4) и (5) позволяет свести задачу к транспортной модели линейного программирования.

Построение оптимальной программы. Пользуясь терминологией транспортной задачи, представим каждую разрешенную клетку таблицы рис. 9.2 как маршрут с соответствующим ему показателем затрат. Алгоритм попросту состоит в поочередном удовлетворении каждой единицы спроса всех месяцев (январь, затем февраль и т. д.) с использованием того из *возможных* маршрутов, затраты по которому минимальны. Если имеются одинаково экономичные варианты, предпочтение отдается возможно более позднему выпуску.

Для того чтобы помочь читателю выполнить необходимые операции на заготовленной им копии таблицы рис. 9.2, вкратце опишем эти операции.

I. Январский спрос (1 единицу) удовлетворим январским выпуском при затратах, равных 2 за единицу.

II. Февральский спрос (2 единицы) удовлетворим январским выпуском при затратах, равных 3 за единицу.

III. Первые 4 единицы мартовского спроса удовлетворим мартовским выпуском при затратах 2 за единицу, следующую единицу — февральским выпуском при затратах 5, последние 2 единицы — мартовским выпуском при затратах 6 за единицу.

IV. Первые 3 единицы апрельского спроса удовлетворим апрельским выпуском при затратах 5 за единицу, следующие 2 единицы — апрельским выпуском в сверхурочное время при затратах 6 за единицу, последнюю единицу — мартовским выпуском в сверхурочное время при затратах 7 за единицу.

V. Майский спрос равен нулю, поэтому в столбце мая не будем проставлять значений выпуска.

VI. Первую единицу июньского спроса удовлетворим майским выпуском при затратах 3 за единицу, вторую — июньским выпуском при затратах 6 за единицу.

Последовательно выполняя операции I — VI, читатель должен соответствующим образом уменьшать неиспользованный остаток производственной мощности (во втором столбце справа таблицы рис. 9.2). Как только этот остаток становится равным нулю, оставшиеся клетки строки больше не используются. Полученную читателем программу следует сравнить с приведенной в таблице рис. 9.3.

Таким образом, оптимальной является следующая программа:

$$x_1 = 3,\quad x_2 = 1,\quad x_3 = 7,\quad x_4 = 5,\quad x_5 = 1,\quad x_6 = 1; \tag{6}$$
$$i_1 = 2,\quad i_2 = 1,\quad i_3 = 1,\quad i_4 = 0,\quad i_5 = 1,\quad i_6 = 0.$$

	Янв.	Февр.	Март	Апр.	Май	Июнь	Максимальный выпуск	Максимальный совокупный выпуск
Янв.	[2] 1	[3] 2	[4]	[5]	[6]	[7]	$\cancel{3}\cancel{2}0$	9
	[5]	[6]	[7]	[8]	[9]	[10]	6	
Февр.		[4]	[5] 1	[6]	[7]	[8]	$\cancel{1}0$	13
		[6]	[7]	[8]	[9]	[10]	3	
Март			[2] 4	[3]	[4]	[5]	$\cancel{4}0$	20
			[6] 2	[7] 1	[8]	[9]	$\cancel{3}\cancel{1}0$	
Апр.				[5] 3	[6]	[7]	$\cancel{3}0$	25
				[6] 2	[7]	[8]	$\cancel{2}0$	
Май					[2]	[3] 1	$\cancel{1}0$	26
					[3]	[4]	0	
Июнь						[6] 1	$\cancel{3}2$	30
						[7]	1	
Спрос D_t	$\cancel{1}0$	$\cancel{2}0$	$\cancel{7}\cancel{3}\cancel{2}0$	$\cancel{6}\cancel{3}\cancel{1}0$	0	$\cancel{2}\cancel{1}0$		
Совокупный спрос	1	3	10	16	16	18		

Рис. 9.3. Производственная программа для модели с выпуклой функцией затрат фирмы «Вечная шина».

Эта программа может быть получена следующим образом: значение x_t исчисляется как использованный выпуск месяца t (который определим по данным второго столбца справа на рис. 9.3), а i_t — как сумма показателей выпуска в клетках, расположенных выше и правее «диагональной» пары клеток месяца $(t+1)$.

Очевидно, этот вычислительный процесс намного проще, чем вычисление $f_n(i)$, выполнявшееся в гл. 8. Он также несколько упрощен по сравнению с описанным выше трехшаговым алгоритмом. В частности, табличное представление данных позволяет на шаге 2 без труда определить наиболее дешевое приращение выпуска с учетом имеющихся возможностей выпуска. Далее, на шаге 3 пока еще не удовлетворенный спрос уменьшается, насколько это возможно, с учетом оставшихся наиболее дешевых возможностей приращения выпуска.

Введение дополнительных ограничений. При рассмотрении сделанных ранее предположений о выпуклости функций (1) и (2) отмечалось, что с помощью трехшагового алгоритма можно учесть как верхние границы u_t выпуска x_t, так и верхние границы b_t запаса i_t. В рассмотренном численном примере соблюдение верхних границ выпуска x_t попросту обеспечивалось рассмотрением только тех значений x_t, которые бы не были слишком большими (уровень запасов не ограничивался). В задаче фирмы «Надежный поставщик» (разд. 8.4) читатель мог убедиться в том, что введение верхних границ иногда оказывает совершенно неожиданное влияние. Однако при выпуклой функции затрат справедлива следующая теорема.

Теорема о влиянии варьирования верхних границ. (а) Оптимальное значение x_p *не* уменьшится при увеличении верхней границы u_p возможного выпуска, а также при увеличении верхней границы b_t уровня запасов i_t, если $t \geqslant p$. (б) Оптимальное значение x_p *не* увеличится при увеличении верхней границы u_t для любого $t \neq p$ или при увеличении верхней границы b_t уровня запасов i_t для $t \leqslant p$.

Для того чтобы убедиться в нетривиальности этой теоремы, достаточно проследить за изменениями оптимальной программы фирмы «Надежный поставщик» (разд. 8.4). При ограничении $x_t \leqslant 5$ для $i_0 = 2$ и $N = 6$ оптимальной являлась программа

$$x_1 = 4, \quad x_2 = 0, \quad x_3 = 3, \quad x_4 = 4, \quad x_5 = 5, \quad x_6 = 0, \tag{7}$$

которую можно построить с помощью таблицы рис. 8.12. Предположим, что только для x_3 верхняя граница увеличена до 6. Тогда оптимальной становится следующая программа, приведенная в таблице рис. 8.17:

$$x_1 = 5, \quad x_2 = 0, \quad x_3 = 6, \quad x_4 = 0, \quad x_5 = 5, \quad x_6 = 0. \tag{8}$$

Следовательно, величина x_1 *возрастает* при увеличении верхнего предела мартовского выпуска, что противоречит части (б) теоремы. (Заметим, что предположение $i_0 = 2$ не оказывает влияния, поскольку точно такой же результат был бы получен при $i_0 = 0$ и $D_1 = 1$.)

Возможность задержки удовлетворения спроса можно учесть, внеся в операции трехшагового алгоритма небольшое число незначительных изменений. Однако объем вычислительных операций при этом существенно возрастает. Пусть неудовлетворенному спросу соответствуют отрицательные значения i_t. Предположим, что функция $h_t(i)$ является выпуклой для всех значений i, т. е. для . . ., —2, —1, 0, 1, 2, ... Изменим операции шага 1 — будем теперь рассматривать единичные приращения выпуска для всех отрезков от 1 до N. Изменим операции шага 2, дополнительно включив в них учет увеличения затрат, вызванного задержкой удовлетворения спроса. Дополнительные вычисления в основном возникают на шаге 2, при расчете изменений затрат для каждого из возможных вариантов изменения уже построенной части программы: после того как определены объемы выпуска в измененной программе, необходимо найти такое распределение этих объемов между месяцами для удовлетворения уже рассмотренного спроса, при котором минимизируются затраты, связанные как с содержанием запасов, так и с задержкой удовлетворения спроса. Операции шага 3 не изменяются.

9.4. АНАЛИЗ ДЛИТЕЛЬНОСТИ ПЛАНОВОГО ПЕРИОДА В МОДЕЛЯХ С ВЫПУКЛОЙ ФУНКЦИЕЙ ЗАТРАТ

При использовании описанного алгоритма для решения приведенной выше задачи фирмы «Вечная шина», являющейся частным случаем модели с выпуклой функцией затрат, становится ясным, что увеличение длительности N планового периода может привести не к сокращению выпуска хотя бы в пределах одного месяца, а разве лишь к его росту. Этот вывод остается справедливым и для общей модели с выпуклой функцией затрат, приведенной в начале разд. 9.3.

Прежде чем сформулировать эту теорему, введем обозначения для совокупного спроса — спроса за p месяцев 1, 2, . . ., p с начала года и для совокупного выпуска — выпуска за этот же период. Обозначим эти величины R_p и X_p:

$$R_p = \sum_{t=1}^{p} D_t \quad \text{и} \quad X_p = \sum_{t=1}^{p} x_t. \tag{1}$$

Предположим, что полной уверенности в действительной величине спроса не имеется, однако можно утверждать, что значение R_p находится в определенных пределах:

$$S_p \leqslant R_p \leqslant T_p, \qquad p = 1, 2, \ldots, N. \tag{2}$$

Обозначим через $X_p(S)$ и $X_p(T)$ производственные программы, оптимальные *для условий* $R_p = S_p$ и $R_p = T_p$ соответственно, т. е. для тех случаев, когда прогнозируемая величина принимает значения, отвечающие нижней и верхней границам. Тогда можно сформулировать следующую теорему о свойствах оптимальной программы:

Теорема о длительности планового периода для модели с выпуклой функцией затрат. (а) При увеличении любого D_t ни одно из оптимальных значений X_p не уменьшится. (б) Для любого отрезка p ($p = 1, 2, \ldots, N$) оптимальная программа удовлетворяет неравенствам $X_p(S) \leqslant X_p \leqslant X_p(T)$.

Часть (б) этой теоремы иллюстрируется графиком, приведенным на рис. 9.4.

Этот результат вполне уместно назвать «теоремой о длительности планового периода», поскольку увеличение длительности с N отрезков до $(N + 1)$ отрезка равносильно увеличению спроса от нуля

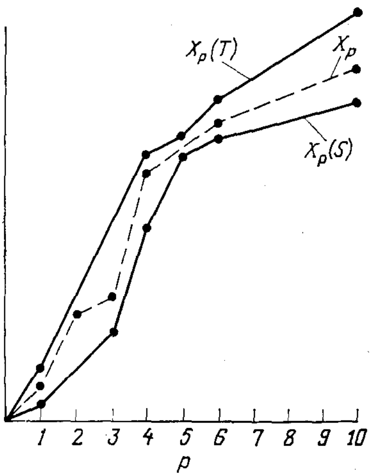


Рис. 9.4. Границы возможных изменений оптимальной программы для модели с выпуклой функцией затрат.

$X_p(S)$ — оптимальная программа для нижней границы прогнозируемого спроса, $X_p(T)$ — то же для верхней границы, X_p — оптимальная программа.

до D_{N+1} при анализе задач с $N + 1$ отрезками. В части (а) теоремы утверждается, что никакое увеличение спроса не приводит к сокращению выпуска ни для одного из отрезков по сравнению с ранее построенной оптимальной программой.

В задаче фирмы «Надежный поставщик» (разд. 8.4) ни одна из частей этой теоремы не применима. Из таблицы рис. 8.13 ясно, что при включении в рассмотрение июньского спроса, т. е. при увеличении N от 5 до 6, январский выпуск сократился с 5 до 4, а выпуск двух первых месяцев, января и февраля,— с 10 до 9.

Важным следствием части (б) этой теоремы является то обстоятельство, что в определенных случаях можно принимать решения об оптимальном объеме производства x_1, располагая лишь очень неточной информацией о более позднем спросе. Такая возможность основана на отыскании значений величин $X_1(S)$ и $X_1(T)$ и определении разницы между ними, если таковая имеется. При небольшом различии значение x_1 можно устанавливать в рамках полученных границ, не опасаясь существенных отклонений от оптимума. Процесс определения программы можно повторить в дальнейшем, при появлении дополнительной информации о спросе. Если же разница между $X_p(S)$ и $X_p(T)$ велика, анализ этих двух вариантов программы помогает выявить, для каких отрезков важно уточнить прогноз спроса.

Один из частных случаев модели с выпуклой функцией затрат представляет особый интерес, поскольку он встречается довольно часто и позволяет получить чрезвычайно сильные и полезные результаты о длительности планового периода. Предположим, что на объем выпуска и на уровень запасов не налагается ограничений, функции производственных затрат для всех отрезков одинаковы, а затраты на содержание запасов пренебрежимо малы, так что

$$C_t(x_t,\ i_t) \equiv C(x_t). \tag{I}$$

Для этого случая откажемся от предположения об *обязательной* целочисленности x_t и сохраним лишь условие его неотрицательности. При этом потребуется определение выпуклости, обобщающее условие (1) разд. 9.2. *Непосредственным* обобщением будет следующее определение: функция $g(x)$ является выпуклой, если для любых w и z ($w \leqslant z$) и для любого p ($0 \leqslant p \leqslant 1$) выполняется условие

$$p\,[g(z) - g(x)] \geqslant (1 - p)\,[g(x) - g(w)], \tag{II}$$

где

$$x = pz + (1 - p)\,w. \tag{III}$$

Перегруппировав в выражении (II) члены и использовав (III), получим

$$pg(z) + (1 - p)\,g(w) \geqslant g\,[pz + (1 - p)\,w] \quad \text{(условие выпуклости)}. \tag{IV}$$

[Приняв $p = 1/2$, $w = x - 1$ и $z = x + 1$, как частный случай (IV) получим условие (3) разд. 9.2.]

Пусть A_p — средний спрос для одного отрезка за первые p отрез ков — равно

$$A_p = \frac{R_p}{p}, \quad p = 1,\ 2,\ \ldots,\ N \tag{V}$$

и пусть A_n — наибольшее значение из A_p (если для нескольких p значения A_p одинаковы, то в качестве отрезка n выберем самый поздний из этих отрезков). Тогда можно доказать, что оптимальным является объем выпуска

$$x_t = A_n, \quad t = 1,\ 2,\ \ldots,\ n, \tag{VI}$$

так что $i_n = 0$. Можно также показать, что для остальных отрезков t ($t > n$) оптимальные значения x_t удовлетворяют условию

$$x_t < A_n, \quad t > n. \tag{VII}$$

Далее, справедлив следующий вывод: совокупность отрезков 1, 2, . . ., n можно рассматривать в качестве отдельного планового периода. Программу выпуска для остальных отрезков, от $(n+1)$ до N, можно строить на основе рассмотрения только этих отрезков, причем остаток на начало отрезка $(n+1)$ должен быть принят равным нулю. Это значит, что расчеты по формуле (V) можно выполнить повторно, исходя из значений B_p:

$$B_p=\frac{R_p-R_n}{p-n}, \quad p=n+1,\ n+2,\ \ldots,\ N. \tag{VIII}$$

Таким образом, второй плановый период начинается с отрезка $(n+1)$ и кончается отрезком q, где q — самое позднее из тех n, при которых B_p принимает наибольшее значение. И здесь B_p является оптимальным объемом выпуска для каждого из отрезков второго планового периода. Понятно, что подобные рассуждения можно повторять до тех пор, пока не будет найдено значение x_N. Одно из следствий полученного результата состоит в том, что если для всех t выполняется условие $D_{t+1} \geqslant D_t$, то при использовании формулы (V) сразу же получаем $n=N$. Напротив, при $D_{t+1} \leqslant D_t$ для всех t оптимальным выпуском для любого отрезка является $x_t=D_t$.

Проделанные выше вычисления соответствуют следующему «графическому» методу построения программы: положим на стол миллиметровую бумагу и воткнем вертикально булавки в точках с координатами $(p,\ R_p)$, где $p=1,\ 2,\ \ldots,\ N$, а также в точке (0, 0). Пропустим нить так, чтобы она проходила выше всех булавок, и завяжем один из ее концов на булавке, воткнутой в точке (0, 0). Второй конец нити туго натянем у булавки с координатами $(N,\ R_N)$; положение натянутой нити покажет оптимальные значения совокупного выпуска X_p.

Если в задаче фирмы «Вечная шина» (рис. 9.2) применимо условие выпуклости (IV), расчеты по формуле (V) позволяют определить следующие значения A_p:

$$A_1=\frac{1}{1}, \quad A_2=\frac{3}{2}, \quad A_3=\frac{10}{3}, \quad A_4=\frac{16}{4} \tag{IX}$$

$$A_5=\frac{16}{5}, \quad A_6=\frac{18}{6}$$

так что $n=4$ и для $t=1$, $t=2$, $t=3$ и $t=4$ получим $x_t={}^{16}/_4$ Затем, по формуле (VIII), вычислим

$$B_5=\frac{16-16}{5-4}=\frac{0}{1} \quad \text{и} \quad B_6=\frac{18-16}{6-4}=\frac{2}{2}, \tag{X}$$

так что $q=6$ и для $t=5$ и $t=6$ получим $x_t={}^2/_2$.

Заметим, что, если значения нескольких D_t изменяются, однако A_n остается наибольшим из всех значений A_p, длительность первого планового периода не изменяется. Так, если любое D_t для $t=n+1$, ...,

. . ., N уменьшается, это не изменит длительности первого планового периода.

Наконец, отметим следующее существенное и удивительное обстоятельство: для построения оптимальной программы выпуска и определения длительности планового периода *не* требуются *численные* значения $C(x_t)$. Достаточно знать, что производственные затраты исчисляются для всех отрезков на основе *одной и той же* функции, что эта функция является выпуклой и что затраты на содержание запасов пренебрежимо малы.

9.5. МОДЕЛЬ УПРАВЛЕНИЯ ПРОИЗВОДСТВОМ И ЗАПАСАМИ С ВОГНУТОЙ ФУНКЦИЕЙ ЗАТРАТ

Модель, рассматриваемая в настоящем разделе, является еще одним частным случаем общей задачи управления запасами, которая приведена в разд. 8.3. Модель настоящего раздела иногда называют **динамической моделью экономически выгодного размера партии;** она заслуживает внимания, потому что отображает часто встречающуюся и практически важную ситуацию. Кроме того, ее сложность является промежуточной между сложностью общей модели разд. 8.3 и модели с выпуклой функцией затрат разд. 9.3.

Дополнительно к допущениям о затратах, запасах и объемах выпуска, сформулированных в разд. 9.1, положим, что

$$\text{Функция } C_t(x_t) \text{ является вогнутой,} \tag{1}$$

$$\text{Функция } h_t(i_t) \text{ является вогнутой,} \tag{2}$$

где $C_t(x_t)$ — производственные затраты, а $h_t(i_t)$ — затраты на содержание запасов. Случай вогнутости функции затрат иногда называют **случаем возрастающей предельной эффективности при увеличении масштаба.** Хотя задержка удовлетворения спроса в данной модели *не* разрешена, проводимый анализ легко модифицировать с учетом допустимости такой задержки (разд. 9.6). Вместе с тем введение верхних границ для x_t и i_t не разрешается, что совершенно противоречит предпосылкам модели с выпуклой функцией затрат. Напомним, что ограничение на объем выпуска можно заменить установлением бесконечно больших затрат $C_t(x_t)$ для значений x_t, превышающих заданную верхнюю границу. Однако при такой замене функция $C_t(x_t)$ перестает быть вогнутой.

Вогнутость функции производственных затрат часто встречается в том случае, когда выпуск продукции связан с необходимостью **затрат на подготовительные операции, или переналадку (единовременные затраты),** после чего выпуску каждой дополнительной единицы продукции соответствуют не меняющиеся пропорциональные затраты:

$$C_t(x_t) = \begin{cases} 0 & \text{при } x_t = 0, \\ s_t + c_t x_t & \text{при } x_t \geqslant 1. \end{cases} \tag{3}$$

Напомним, что именно такой вид имела функция производственных затрат фирмы «Надежный поставщик» (разд. 8.4), где для $x_t \leqslant 5$ затраты на переналадку были равны $s_t = 13$, а пропорциональные затраты составляли $c_t = 2$ на единицу продукции. Хотя $C_t(x_t)$ и является вогнутой для $0 \leqslant x_t \leqslant 5$, эта задача все же *не* соответствует допущениям, сделанным в настоящем разделе, поскольку в нем не разрешается устанавливать верхнюю границу для значений x_t.

Вогнутая функция затрат часто возникает и в другом случае: когда модель используется в качестве модели пополнения запасов путем закупок у внешнего поставщика. В этом случае поставщик при крупных заказах нередко делает **скидки в зависимости от размера заказа.** Для иллюстрации приведем возможные **ступенчатые цены**, возникающие при подобных скидках:

10 долл. за единицу при размере заказа до 12 единиц;
8 долл. за каждую единицу *сверх* 12 при размере заказа от 13 до 144 единиц;
5 долл. за каждую единицу *сверх* 144. (4)

Эти ступенчатые цены описываются следующим математическим выражением:

$$C_t(x_t) = \begin{cases} 10x_t & \text{при} \quad 0 \leqslant x_t \leqslant 12, \\ 120 + 8(x_t - 12) & \text{при} \quad 13 \leqslant x_t \leqslant 144, \\ 120 + 1056 + 5(x_t - 144) & \text{при} \quad 145 \leqslant x_t. \end{cases} \tag{5}$$

Выражение (5) имеет тот же вид, что и выражение (10) в разд. 9.2. Функция $C_t(x_t)$ является вогнутой, поскольку коэффициенты при x_t, а именно 10, 8 и 5, убывают с ростом x_t. Если в (5) включить также затраты на подготовительные операции, функция по-прежнему останется вогнутой. Следовательно, при определении оптимальной программы движения запасов фирма может в данном случае учесть и условно-постоянные затраты — собственные накладные расходы, связанные с оформлением и обработкой заказа.

Структурные результаты. Глубоким обобщением результатов анализа рассмотренной модели является следующая теорема.

Теорема о виде оптимальной программы. Всегда существует программа, обеспечивающая минимальность затрат и обладающая той особенностью, что любой выпуск x_t в ней принимает одно из следующих значений: $0,\ D_t,\ D_t + D_{t+1}, \ldots, D_t + D_{t+1} + \ldots + D_N$.

Из данной теоремы следует, что при отыскании оптимальной программы требуется для каждого t рассматривать только $[1 + N - (t - 1) = N - t + 2]$ возможных значений x_t. Сравним это с результатами, полученными для модели общего вида (разд. 8.3), где выпуск x_t мог принимать любое значение от 0 до $D_t + D_{t+1} + \ldots + D_N$. Важное алгоритмическое следствие сформулированной теоремы состоит в том, что для отыскания оптимальной программы необходимо рассмотреть всего лишь около $0,5\,N^2$ различных вариантов.

Еще одно следствие данной теоремы: всегда существует оптимальная программа, отличающаяся той особенностью, что для любого отрезка, остаток на начало которого строго положителен, выпуск равен нулю. (Читатель должен объяснить, почему это следует из данной теоремы.) При использовании принятых нами условных обозначений данное положение записывается следующим образом: существует оптимальная программа, в которой

$$i_{t-1}x_t = 0, \quad t = 1, 2, \ldots, N. \tag{6}$$

Таким образом, если спрос D_t удовлетворяется за счет запаса i_{t-1} ($i_{t-1} > 0$), то весь этот запас исчерпывается на последнем из отрезков, для которого выпуск был ненулевым. Несомненно, для модели с выпуклой функцией затрат такое утверждение неверно.

Все отрезки, для которых $x_t > 0$, можно называть «точками возобновления». Этот термин отражает то обстоятельство, что начальный уровень запасов для такого отрезка равен нулю и производственный цикл как бы начинается заново.

Остается рассмотреть использование рекуррентных соотношений динамического программирования для эффективного поиска оптимальной программы среди множества допустимых программ, вид которых отвечает сформулированной выше теореме. Прежде чем приступить (в разд. 9.6) к рассмотрению такого алгоритма, изложим основные его идеи словами.

Пусть январь — первый месяц планового периода. Начнем вычислительный процесс с расчета стоимости январского производства при таком объеме выпуска, который достаточен для удовлетворения январского спроса. Затем повторяем расчеты, каждый раз увеличивая длительность планового периода на один месяц и используя полученные ранее результаты. Так, допустим, что найдены оптимальные программы для плановых периодов, включающих: январь и февраль; январь, февраль и март и т. д., до планового периода январь — июнь включительно, причем пусть каждая из этих программ имеет описанный в теореме вид. Теперь же в плановый период требуется включить июль.

Из теоремы о виде оптимальной программы следует, что июльский спрос должен быть *целиком* удовлетворен продукцией, выпущенной только в одном из месяцев, будь то январь, февраль, . . ., июль; всего необходимо рассмотреть семь таких вариантов. Далее, если для удовлетворения июльского спроса используется, например, апрельский выпуск, то этот выпуск должен полностью обеспечить удовлетворение спроса с апреля по июль (объясните, почему это так?). Что же касается оптимальной программы января, февраля и марта, то для этого трехмесячного планового периода она уже найдена ранее. Следовательно, оптимальным вариантом семимесячной программы будет самый дешевый из семи описанных выше вариантов.

Применительно к модели управления запасами с вогнутой функцией затрат теорему о виде оптимальной программы доказать нетрудно. Ранее условие (6) рассматривалось в качестве следствия этой теоремы; теперь используем иной подход — *вначале* докажем справедливость условия (6), а затем покажем, каким образом из этого следует оптимальность программы описанного вида.

Пусть имеется некоторая допустимая программа, в которой хотя бы для одного отрезка $i_{t-1}x_t > 0$, и пусть отрезок p — первый из таких отрезков. Рассмотрим тот более ранний отрезок, в течение которого выпускались изделия, образовавшие запас i_{t-1}, и вычислим приращение затрат, связанное с дополнительным производством 1 единицы продукции и хранением ее до начала отрезка p. Затем сравним эту величину с увеличением затрат при дополнительном производстве 1 единицы на отрезке p.

Если последняя величина меньше, изменим программу, увеличив x_p на i_{p-1} и уменьшив выпуск на более раннем отрезке на это же число единиц. В связи с допущениями (1) и (2) о вогнутости функции затрат общее увеличение затрат на отрезке p не превышает общего снижения затрат, достигнутого благодаря сокращению более раннего выпуска и хранимых запасов. Аналогично этому, если увеличение затрат при дополнительном выпуске единицы на отрезке p является большей величиной, то общая сумма затрат по всему плановому периоду не возрастет при увеличении на x_p выпуска на более раннем отрезке и полном прекращении выпуска на отрезке p. В любом из этих двух случаев построена допустимая программа, в которой условие (6) удовлетворяется при $t = 1, 2, \ldots, p$. Если существуют более поздние отрезки, для которых это условие нарушается, повторим выполнение операции применительно к каждому из них. В конце концов будет получена программа, в которой условие (6) удовлетворяется для всех отрезков, а затраты не больше, чем у исходной программы.

Отсюда следует приведенная выше теорема, поскольку обратное ей утверждение означает существование хотя бы одного отрезка, для которого $0 < i_{t-1} < D_t$. Для допустимости этой программы должно выполняться условие $x_t \geqslant D_t - i_{t-1} > 0$, что противоречит (6).

9.6. АЛГОРИТМ ОПТИМИЗАЦИИ МОДЕЛИ С ВОГНУТОЙ ФУНКЦИЕЙ ЗАТРАТ

Введем следующее условное обозначение:

c_{kj} — общие затраты связанные с производством на отрезке $(k+1)$ объема выпуска, необходимого для удовлетворения спроса на отрезках $(k+1), \ldots, j$, где $k = 0, 1, \ldots, (N-1)$ и $k+1 \leqslant$

$\leqslant j \leqslant N$. Следовательно,

$$c_{kj} = \begin{cases} C_{k+1}(D_{k+1}) & \text{при } j = k+1, \\ C_{k+1}(D_{k+1} + \ldots + D_j) + h_{k+1}(D_{k+2} + \ldots + D_j) + \ldots + h_{j-1}(D_j) & \text{при } j > k+1. \end{cases} \tag{1}$$

Так, если январь есть отрезок $(k + 1)$, а март — отрезок j, то c_{kj} представляет собой сумму, состоящую из затрат на производство в январе объема выпуска, необходимого для удовлетворения январского, февральского и мартовского спроса; затрат на содержание запасов, исчисленных исходя из того их уровня на конец января, который достаточен для удовлетворения февральского и мартовского спроса; затрат на содержание запасов, исчисленных исходя из того их уровня на конец февраля, который достаточен для удовлетворения мартовского спроса.

Задача фирмы «Растущий сбыт». Исходные данные для этой задачи приведены в таблице рис 9.5; длительность планового периода N

Отрезок / $C_t(X_t)$ Выпуск

t \ X_t	1	2	3	4
1	5	10	15	16
2	6	9	12	
3	5	7		
4	3			

$C_t(0) = 0$

Рис. 9.5. Модель с вогнутой функцией затрат фирмы «Растущий сбыт».

c_{kj}

k \ j	1	2	3	4
0	5	10,1	15,3	16,6
1		6	9,1	12,3
2			5	7,1
3				3

c_{kj} — затраты на производство на отрезке $(k+1)$ продукции, удовлетворяющей спрос на отрезках $(k+1), \ldots, j$

Рис. 9.6. Модель с вогнутой функцией затрат фирмы «Растущий сбыт».

равна 4 отрезкам. В дополнение к данным, приведенным на рис. 9.5, $C_t(X_t)$ положим, что для всех отрезков

$$D_t = 1 \quad \text{и} \quad h_t(i_t) = \frac{i_t}{10}. \tag{2}$$

(Читателю рекомендуется попытаться отгадать оптимальную программу для $N = 1$, $N = 2$, $N = 3$ и $N = 4$.)

Приведем несколько примеров исчисления c_{kj} по формуле (1) на основе данных рис. 9.5:

$c_{01} = C_1(1) = 5,$

$c_{02} = C_1(1 + 1) + h_1(1) = 10 + 0{,}1 = 10{,}1,$

$$c_{03} = C_1(1+1+1) + h_1(1+1) + h_2(1) = 15 + 0{,}2 + 0{,}1 = 15{,}3,$$

$$c_{04} = C_1(1+1+1+1) + h_1(1+1+1) + h_2(1+1) + h_2(1) = \\ = 16 + 0{,}3 + 0{,}2 + 0{,}1 = 16{,}6, \quad (3)$$

$$c_{12} = C_2(1) = 6,$$

$$C_{13} = C_2(1+1) + h_2(1) = 9 + 0{,}1 = 9{,}1,$$

$$C_{14} = C_2(1+1+1) + h_2(1+1) + h_3(1) = 12+0{,}2+0{,}1 = 12{,}3.$$

Читатель должен самостоятельно вычислить значения остальных c_{kj}, а именно c_{23}, c_{24} и c_{34}, и сравнить полученные результаты со значениями, приведенными в таблице рис. 9.6.

Введем следующее обозначение:

f_n — минимальная стоимость программы для отрезков от 1 до n при нулевом запасе на конец отрезка n.

Тогда соответствующее рекуррентное соотношение динамического программирования имеет вид

$$f_n = \min_{k=0,\,1,\,\ldots,\,n-1} [f_k + c_{kn}], \quad n = 1,\ 2,\ \ldots,\ N, \quad (4)$$

где $f_0 \equiv 0$. Обозначим k_n такое значение k, которое позволяет достичь f_n. Требуется найти значение f_N.

Выражение в квадратных скобках в правой части (4) представляет собой сумму двух элементов затрат. Первый элемент, т. е. f_k, есть сумма затрат для отрезков от 1 до k, соответствующая оптимальной программе для этого периода. В этой программе запас на конец отрезка равен нулю. Второй элемент, c_{kn}, представляет собой затраты на отрезках от $(k+1)$ до n, соответствующие программе производства, которая позволила бы удовлетворить весь спрос на оставшихся отрезках — при нулевом остатке на начало отрезка $(k+1)$. Значение k_n, при котором сумма $(f_k + c_{kn})$ достигает минимума, определяет оптимальную программу для периода длительностью n отрезков; в плановом периоде от отрезка 1 до отрезка n это значение позволяет указать отрезок $(k_n + 1)$, в котором должно произойти последнее возобновление программы. Поэтому настоящий алгоритм можно рассматривать как способ определения оптимальной последовательности точек возобновления. [В точке возобновления начальный запас всегда равен нулю, поэтому в (4) можно исключить переменную состояния i.] Заметим, что алгоритм построен как «прямой» (разд. 8.3), т. е. вычислительный процесс направлен от первого отрезка к последнему.

Продемонстрируем применение рассмотренного алгоритма на примере, исходные данные для которого приведены в таблице рис 9.5.

$$f_0 = 0,$$

$$f_1 = \min[f_0 + c_{01}] = [0 + 5] = 5 \text{ и } k_1 = 0,$$

$$f_2 = \min[f_0 + c_{02},\ f_1 + c_{12}] =$$

$$
\begin{aligned}
&= \min\,[0 + 10{,}1;\ 5 + 6] = 10{,}1 \text{ и } k_2 = 0,\\
f_3 &= \min\,[f_0 + c_{03},\ f_1 + c_{13},\ f_2 + c_{23}] =\\
&= \min\,[0 + 15{,}3;\ 5 + 9{,}1;\ 10{,}1 + 5] = 14{,}1\\
&\quad \text{и } k_3 = 1, \qquad\qquad (5)\\
f_4 &= \min\,[f_0 + c_{04},\ f_1 + c_{14},\ f_2 + c_{24},\ f_3 + c_{34}] =\\
&= \min\,[0 + 16{,}6;\ 5 + 12{,}3;\ 10{,}1 + 7{,}1;\ 14{,}1 + 3] =\\
&= 16{,}6 \text{ и } k_4 = 0.
\end{aligned}
$$

Эти расчеты показывают, что для планового периода длительностью N отрезков оптимальной программой будет следующая:

$$
\begin{array}{llllll}
x_1 = 1 & & & & \text{при} & N = 1,\\
x_1 = 2, & x_2 = 0 & & & \text{при} & N = 2,\\
x_1 = 1, & x_2 = 2, & x_3 = 0 & & \text{при} & N = 3,\\
x_1 = 4, & x_2 = 0 & x_3 = 0, & x_4 = 0 & \text{при} & N = 4.
\end{array} \qquad (6)
$$

Покажем, как построить оптимальную программу, например, для $N = 3$. В выражении (5) $k_3 = 1$; это означает, что программа для отрезков 1, 2 и 3 явится оптимальной в том случае, если отрезок 1 мы будем рассматривать как отдельный, изолированный плановый период, а на отрезке 2 — выпускать продукцию для удовлетворения спроса как на отрезке 2, так и на отрезке 3. Значение выпуска $x_2 = 2$ равно общему спросу на отрезках 2 и 3. Решение для отрезка 1, рассматриваемого в качестве отдельного планового периода, определяется программой (5) для $N = 1$, согласно которой $x_1 = D_1 = 1$.

Заметим, что при удлинении планового периода оптимальное значение x_1 (январского выпуска) колеблется. Согласно теореме о длительности планового периода, приведенной в разд. 9.5, это явление не свойственно моделям с выпуклой функцией затрат. Однако в моделях с *вогнутой* функцией затрат такие колебания возникают, причем даже в очень простых ситуациях, как это будет показано в разд. 9.7. Таким образом, динамические свойства модели обусловливают более сложный характер изменения результатов, хотя алгоритм для этой модели усложняется ненамного.

Полезно построить сеть, соответствующую рекуррентному соотношению динамического программирования (4). Присвоим вершинам номера от *1* до *5*; c_{kj} — стоимость перемещения единицы по дуге, направленной от вершины (*5* — j) к вершине (*6* — k). Сеть приведена на рис. 9.7; отметим, что по своей структуре она идентична сети, изображенной на рис. 6.10. Вычисления по формуле (5) можно рассматривать как нахождение кратчайшего пути в вершину *5* из всех других вершин; здесь f_n — длина кратчайшего пути из вершины (*5* — n) в вершину *5*. Читателю следует выполнить на сети операции, аналогичные расчетам в соответствии с выражением (5).

Представляется полезным также сравнить (4) с «прямым» рекуррентным соотношением динамического программирования (V), приведенным в разд. 8.3. Отметим, что существует следующее соотношение:

$$f_n' \equiv g_n(0).$$

Таким образом, если известен *вид* оптимальной программы, можно полностью исключить вычислительные операции по определению $g_n(i)$ для всех $i > 0$.

Без особых трудностей можно найти также оптимальную программу для случая разрешенной задержки удовлетворения спроса,

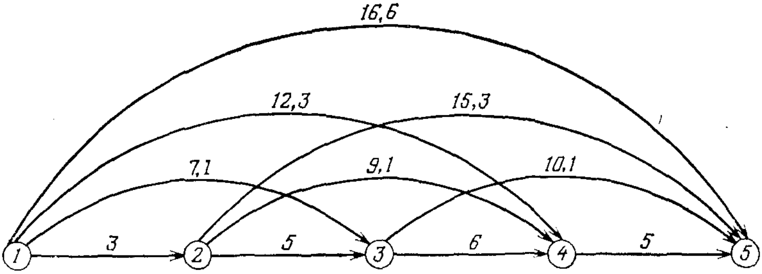


Рис. 9.7. Модель с вогнутой функцией затрат фирмы «Растущий сбыт».

т. е. случая, при котором i_t может быть отрицательным. Допустим, что функция $h_t(i)$ является вогнутой как для $i = 0, 1, 2, \ldots$, так и для $i = 0, -1, -2$; вместе с тем функция $h_t(i)$ *не обязательно* должна быть вогнутой для $i = \ldots, -2, -1, 0, 1, 2, \ldots$. При этом продолжает оставаться справедливым равенство x_t либо нулю, либо общему спросу за ряд последовательных отрезков, включающий отрезок t; однако в данной модели этот ряд отрезков может включать отрезки, не только следующие после отрезка t, но и предшествующие ему. Основная идея состоит в том, что по-прежнему отыскивается оптимальная последовательность точек возобновления, для которых начальный запас равен нулю. Между каждыми двумя такими точками выпуск осуществляется только на одном из отрезков. Определение c_{kj} становится более общим. Теперь излагаем его следующим образом: c_{kj} — минимальные общие затраты, связанные с производством только на одном из отрезков $k+1, \ldots, j$ такого объема выпуска, который необходим для удовлетворения спроса на всех этих отрезках.

Таким образом, вначале на основе предварительных оптимизационных расчетов определяются c_{kj}:

$$c_{kj} = \begin{cases} C_{k+1}(D_{k+1}) & \text{при } j = k+1, \\ \min\limits_{k+1 \leq t \leq j} [C_t(D_{k+1} + \ldots + D_j) + h_{k+1}(i_{k+1}) + \ldots + h_{j-1}(i_{j-1})] & \\ & \text{при } j > k+1. \end{cases}$$

Полученные значения c_{kj} используются в рекуррентном соотношении (4).

9.7. АНАЛИЗ ДЛИТЕЛЬНОСТИ ПЛАНОВОГО ПЕРИОДА В МОДЕЛЯХ С ВОГНУТОЙ ФУНКЦИЕЙ ЗАТРАТ

Как мог заметить читатель, существенное различие между моделями с выпуклой и вогнутой функцией затрат заключается в том, что при увеличении длительности планового периода оптимальное значение x_t может уменьшаться только при вогнутой функции затрат. Такое уменьшение оказывается возможным даже в простейших случаях.

Так, пусть производственные затраты равны сумме затрат на переналадку (единовременные затраты) и пропорциональных затрат:

$$C_t(x_t) = s_t + c_t x_t, \tag{1}$$

где $s_t \geqslant 0$ и $c_t \geqslant 0$. Примем

$$\begin{gathered} s_1 = 1, \quad s_2 = 2, \quad s_3 = 4, \quad s_4 = 8, \quad s_5 = 7; \\ c_t = 0 \quad \text{и} \quad h_t(i_t) = h_t i_t = i_t \quad (h_t = 1) \end{gathered} \tag{2}$$

и положим

$$D_1 = 1, \quad D_2 = 1, \quad D_3 = 2, \quad D_4 = 3, \quad D_5 = 4. \tag{3}$$

Используем описанный в разд. 9.6 алгоритм и убедимся в том, что

$$\begin{gathered} f_1 = 1, \quad f_2 = 2, \quad f_3 = 5, \quad f_4 = 9, \quad f_5 = 16; \\ k_1 = 0, \quad k_2 = 0, \quad k_3 = 1, \quad k_4 = 2, \quad k_5 = 4, \end{gathered} \tag{4}$$

откуда

$$\begin{array}{ll} x_1 = 1 & \text{для } N = 1, \\ x_1 = 2,\ x_2 = 0 & \text{для } N = 2, \\ x_1 = 1,\ x_2 = 3,\ x_3 = 0 & \text{для } N = 3, \\ x_1 = 2,\ x_2 = 0,\ x_3 = 5,\ x_4 = 0 & \text{для } N = 4, \\ x_1 = 2,\ x_2 = 0,\ x_3 = 5,\ x_4 = 0,\ x_5 = 4 & \text{для } N = 5. \end{array} \tag{5}$$

Для того чтобы показать, каким образом из (4) может быть получено (5), рассмотрим, например, случай $N = 4$. Напомним, что $(k_n + 1)$ есть самый поздний отрезок с ненулевым объемом выпуска в том случае, если плановый период длится от отрезка 1 до отрезка n. Поскольку в (4) $k_4 = 2$, отрезки 1 и 2 образуют отдельный плановый период, а выпуск на отрезке 3 покрывает спрос на отрезках 3 и 4. Таким образом, $x_3 = D_3 + D_4 = 2 + 3 = 5$. Для определения выпуска на первых двух отрезках рассмотрим случай $N = 2$. Поскольку $k_2 = 0$, выпуск относится к отрезку 1 и равен спросу на этих двух отрезках, т. е. $x_1 = D_1 + D_2 = 1 + 1 = 2$.

Отметим, что при удлинении планового периода значения как x_1, так и x_2 колеблются; однако справедлива следующая теорема о длительности планового периода.

Теорема о длительности планового периода для модели, учитывающей единовременные затраты. Пусть затраты описываются функцией (1) и $c_t \geqslant c_{t+1}$ для $t = 1, 2, \ldots, N-1$. (а) Если $k_n = n - 1$ при $n \geqslant 2$, то программа, полученная при рассмотрении совокупности отрезков $1, 2, \ldots, n-1$ в качестве отдельного планового периода, всегда является оптимальной. (б) Если $t \geqslant n$, то $k_t \geqslant k_n$.

Чтобы проиллюстрировать часть (а) теоремы, допустим, что при использовании алгоритма применительно к четырем месяцам (с января по апрель) было найдено $k_4 = 3$. Это значит, что при $N \geqslant 4$ апрельский спрос должен удовлетворяться апрельским выпуском. Тогда в части (а) теоремы утверждается, что правильная процедура построения оптимальной программы состоит в выделении первых трех месяцев (января, февраля и марта) в качестве отдельного планового периода, вне зависимости от спроса в мае и в последующих месяцах. Разумеется, при удлинении планового периода будет получена дополнительная информация об оптимальном значении апрельского выпуска. Однако в любом случае $x_4 > 0$, т. е. известно, что в оптимальной программе для $N \geqslant 4$ апрельский выпуск положителен. Заметим, что в примере (4) часть (а) выполняется и для $N = 5$. Следовательно, в (5) значения x_t ($t = 1$, $t = 2$, $t = 3$, $t = 4$), оптимальные при $N = 4$, остаются оптимальными и при $N \geqslant 5$, вне зависимости от значений D_t для $t > 5$.

Чтобы проиллюстрировать часть (б) теоремы, допустим, что при использовании алгоритма применительно к тем же четырем месяцам (с января по апрель) было найдено $k_4 = 2$. Это означает, что при $N = 4$ апрельский спрос должен удовлетворяться мартовским выпуском, а программа для января и февраля определяется исходя из $N = 2$. В части (б) утверждается, что спрос любого месяца, более позднего, чем апрель, удовлетворяется продукцией, выпускаемой *после* февраля — в марте или в другом более позднем месяце. Заметим, что численные данные примера (2) отвечает условиям теоремы и поэтому результаты (4) согласуются с частью (б) теоремы.

Сформулированная теорема позволяет сократить объем вычислительных операций, связанный с использованием алгоритма. Если условия теоремы соблюдаются, рекуррентное соотношение можно записать в следующем виде:

$$f_n = \min_{k=k_{n-1}, \ldots, n-1} [f_k + c_{kn}], \quad n = 1, 2, \ldots, N, \tag{6}$$

что сокращает поиски минимума. Так, в примере (2) можно определить f_5, подставляя в формулу (6) только $k = 2$, $k = 3$ и $k = 4$, поскольку в (4) $k_4 = 2$. При подобном подходе применение соотношения (6) потребует исчисления *не всех* c_{kj} для $k = 0, 1, \ldots, N-1$ и $k+1 \leqslant j \leqslant N$, и целесообразно исчислять c_{kj} только по мере возникновения надобности в них. При этом удобно воспользоваться следующим рекуррентным соотношением:

$$c_{kn} = c_{k,n-1} + c_{k+1}D_n + h_{k+1}(D_n) + \ldots + h_{n-1}(D_n), \quad n > k+1; \tag{7}$$

здесь предполагается линейность затрат на содержание запасов, т. е. $h_t(i_t) = h_t i_t$.

В отличие от теоремы о длительности планового периода для модели с выпуклой функцией затрат часть (а) рассматриваемой теоремы позволяет получить более сильный результат (для планового периода устанавливается определенная программа выпуска), однако положение (б) является более слабым, поскольку, как было видно из (5), при увеличении N x_t может колебаться. Анализ модели может дать дополнительные результаты, если принять, что производственные затраты (1) и затраты на содержание запасов исчисляются для всех отрезков по одним и тем же формулам, при одинаковых значениях параметров, что D_t для всех отрезков одинаковы и что плановый период является бесконечным. Этот случай рассматривается в разд. 12.5.

Заметим, что в теореме предполагается как соответствие функции производственных затрат виду (1), так и выполнение условия $c_t \geqslant c_{t+1}$. Результаты (5) задачи фирмы «Растущий сбыт» (разд. 9.6) свидетельствуют о том, что для модели с вогнутой функцией затрат часть (б) теоремы в общем случае неверна. Приведем еще два примера, показывающих, что теорема *неверна*, если отбросить условие $c_t \geqslant c_{t+1}$ или если все функции $C_t(x_t)$ одинаковы, но не имеют вида (1).

Рассмотрим пример, в котором не выполняются неравенства $c_t \geqslant c_{t+1}$:

$$\begin{array}{llll} s_1 = 1, & s_2 = 12, & s_3 = 1, & s_4 = 2, \\ c_1 = 3, & c_2 = 1, & c_3 = 2, & c_4 = 1, \\ D_1 = 1, & D_2 = 4, & D_3 = 2, & D_4 = 2; \end{array} \tag{I}$$

примем $h_t(i_t) = 0$. (Не сумеет ли читатель отгадать оптимальные программы для $N = 1$? $N = 2$? $N = 3$? $N = 4$?)

Применим алгоритм и убедимся в том, что

$$\begin{array}{llll} f_1 = 4, & f_2 = 16, & f_3 = 21, & f_4 = 24, \\ k_1 = 0, & k_2 = 0, & k_3 = 2, & k_4 = 1, \end{array} \tag{II}$$

что нарушает (б). Какова оптимальная программа для $N = 1$? $N = 2$? $N = 3$? $N = 4$? (Насколько точными оказались предположения читателя?) Наблюдается ли при удлинении планового периода вначале увеличение, а затем сокращение выпуска на каких-либо отрезках?

Теперь рассмотрим пример, в котором $C_t(x_t)$ и $h_t(i_t)$ одинаковы для всех отрезков:

$$C_t(1) = 5, \quad C_t(2) = 10, \quad C_t(3) = 11, \quad h_t(i_t) = i_t \tag{III}$$

и $D_t = 1$ ($t = 1, 2, 3$). Заметим, что $C_t(x_t)$ не соответствует выражению (1). Можно убедиться в том, что

$$\begin{aligned} &f_1 = 5, \quad f_2 = 10, \quad f_3 = 14, \\ &k_1 = 0, \quad k_2 = 1, \quad k_3 = 0 \end{aligned} \tag{IV}$$

и что это также не согласуется с (б).

9.8. МОДЕЛЬ УПРАВЛЕНИЯ ЗАПАСАМИ ПРИ СГЛАЖИВАНИИ ПРОИЗВОДСТВА

В изучавшихся ранее моделях управления запасами осуществлялась экономическая оптимизация, основанная на достижении равновесия между ростом производственных затрат, с одной стороны, и ростом затрат на содержание запасов — с другой. В результате такой оптимизации возникают существенные колебания выпуска в течение установленного планового периода. Даже сам вид оптимальной программы для модели с вогнутой функцией затрат (разд. 9.5) отвечает такому положению, когда изделия заказываются небольшими партиями. Это бывает наиболее удобно, когда фирма выпускает большое число различных видов изделий. Аналогично этому при пополнении запаса путем закупок у внешних поставщиков обычно целесообразно заказывать партии, достаточно большие для удовлетворения спроса в течение нескольких месяцев. Однако в тех случаях, когда анализируемая ситуация относится к разработке программы выпуска одного из основных видов продукции, изготовляемой фирмой, следует учитывать важный экономический фактор, который до сих пор не принимался во внимание. Этот фактор — *стоимость изменения* объема производства.

В принципе такой фактор несложно учесть в общей модели (гл. 8). Воспользуемся теми же, что и раньше, символами x_t и i_t для обозначения объема производства на отрезке t и остатка на конец отрезка соответственно; тогда в модели изменится лишь целевая функция, которая будет иметь следующий вид:

$$\text{Минимизировать } \sum_{t=1}^{N} C_t(x_t, i_t, x_{t-1}). \tag{1}$$

Например, она может иметь вид

$$C_t(x_t, i_t, x_{t-1}) = C_t(x_t, i_t) + v_t(x_t, x_{t-1}), \tag{2}$$

где последняя функция в правой части представляет собой стоимость изменения объема производства при переходе от отрезка ($t - 1$) к отрезку t.

Приведем простой пример **функции издержек сглаживания**

$$v_t(x_t, x_{t-1}) \equiv v_t(x_t - x_{t-1})^2, \quad \text{где } v_t > 0. \tag{3}$$

Еще один простой пример:

$$v_t(x_t, x_{t-1}) \equiv \begin{cases} a_t(x_t - x_{t-1}) & \text{при} \quad x_t - x_{t-1} \geqslant 0, \quad a_t \geqslant 0, \\ b_t(x_{t-1} - x_t) & \text{при} \quad x_t - x_{t-1} \leqslant 0, \quad b_t \geqslant 0. \end{cases} \tag{4}$$

Если сглаживание производства играет существенную роль при составлении программы, то спрос, объем выпуска и уровень запасов часто выражают при анализе в таких «трудовых» единицах измерения, как человеко-неделя. В тех случаях, когда функция (4) считается адекватной, a_t соответствует затратам на увеличение численности работающих, а b_t — на ее сокращение. (Можно принять, что $a_t + b_t > 0$.)

Как следует изменить рекуррентное соотношение динамического программирования для учета этого нового фактора? Вспомним, что в гл. 8 рекуррентное соотношение было записано в следующем виде:

$$f_n(i) = \min_x [c_n(x, i + x - d_n) + f_{n-1}(i + x - d_n)], \quad n = 1, 2, \ldots, N, \tag{5}$$

где $i = 0, 1, \ldots, d_1 + \ldots + d_n$, а минимум отыскивался по всем неотрицательным целым значениям x, находящимся в пределах $d_n - i \leqslant x \leqslant d_1 + \ldots + d_n - i$; запас на конец планового периода предполагался равным нулю. Однако данные об исходном запасе теперь не являются достаточными для того, чтобы полностью характеризовать состояние системы на начало периода: необходимо также знать объем выпуска на предыдущем отрезке. Следовательно, состояние системы определяется двумя этими параметрами. Остальные изменения, которые необходимо внести, очевидны.

Введем следующие условные обозначения:

$f_n(i, y)$ — минимальные затраты для n оставшихся отрезков при начальном уровне запасов i и выпуске на предыдущем отрезке y;

$x_n(i, y)$ — выпуск, позволяющий достичь $f_n(i, y)$.

Заменим (5) новым рекуррентным соотношением:

$$f_n(i, y) = \min_x [c_n(x, i + x - d_n, y) + f_{n-1}(i + x - d_n, x)], \quad n = 1, 2, \ldots, N. \tag{6}$$

Требуется найти значение $f_N(i_0, y_0)$, где y_0 — выпуск на отрезке, предшествующем отрезку 1. Вначале вычисляются значения

$$f_1(i, y) = c_1(d_1 - i, 0, y) \quad \text{для} \quad i = 0, 1, \ldots, d_1 \quad \text{и} \quad y = 0, 1, \ldots, d_1 + d_2. \tag{7}$$

Замечает ли читатель существенные различия между соотношениями (5) и (6)? Имеется одно такое различие. Для вычисления минимума в соответствии с (5) требовалось знать значения $f_n(i)$ для всех возможных значений (i); для вычисления минимума в соответствии с (6) необходимо вычислить $f_n(i, y)$ для возможных *пар* значений i и y. Хотя динамическое программирование и в этом случае обладает огромными преимуществами перед полным перебором, добавление еще одной переменной состояния приводит к существенному росту вычислительных операций, как это видно из приводимого ниже примера. Таким образом, с теоретической точки зрения рекуррентные методы можно обобщить, учитывая с их помощью еще большее число факторов, например выпуск продукции за два предыдущих отрезка или же изменение уровня запасов. Однако это приводит к такому росту вычислений, который в реальных ситуациях быстро лишает подобный подход практического значения. Если же на практике необходимо учитывать большое число зависимостей, приходится либо применять другие методы решения (например, линейное или нелинейное программирование), либо существенным образом использовать выявленные свойства (вид) оптимальной программы.

Задача фирмы «Гарантия». Предположим, что спрос, верхние границы объема выпуска и уровня запасов, а также затраты фирмы «Гарантия» характеризуются следующими данными:

$$D_t = 3 \quad \text{(спрос постоянен)}, \tag{8}$$

$$x_t = 0,\ 1,\ \ldots,\ 5 \text{ и } i_t = 0,\ 1,\ \ldots,\ 4; \quad i_N = 0, \tag{9}$$

$$C_t(x_t,\ i_t,\ x_{t-1}) = C(x_t) + hi_t + v(x_t - x_{t-1})^2$$

(функция затрат для различных отрезков одинакова), (10)
где для всех отрезков

$$C(0) = 0, \quad C(1) = 15, \quad C(2) = 17, \quad C(3) = 19, \quad C(4) = 21, \quad C(5) = 28, \tag{11}$$

$$h = 1 \quad \text{и} \quad v = 1. \tag{12}$$

[Эти показатели полностью совпадают с данными задачи, рассмотренной в разд. 8.4, за исключением третьего члена правой части (10) — *сглаживающей функции*.] Рекуррентное соотношение (6) можно записать в следующем виде:

$$f_n(i,\ y) = \min_x [(C(x) + 1(i + x - 3) + (x - y)^2 + f_{n-1}(i + x - 3,\ x)], \quad n = 1,\ 2,\ \ldots,\ N, \tag{13}$$

где $i = 0, 1, \ldots, 4$, $y = 0, 1, \ldots, 5$, а минимум отыскивается путем перебора всех неотрицательных целых значений x, заключенных в пределах $3 - i \leqslant x \leqslant \min(5,\ 7 - i)$.

Вычисления по формуле (7) для $n = 1$ систематизированы в таблице рис. 9.8. В этой таблице шесть пар столбцов, по одной паре для каждого из возможных значений y, а именно для $0, 1, \ldots, 5$. «Запрещенные» клетки таблицы соответствуют тем сочетаниям i и y, которые оказываются недопустимыми с учетом ограничений (9) на объем выпуска и уровень запасов. Так, при $y = 0$ исходный запас на начало планового периода не может превышать 1, потому что запас на начало предыдущего отрезка не мог быть больше 4, а для удовлетворения спроса на этом отрезке запас приходится уменьшить на 3. Аналогично этому при $y = 5$ исходный запас на начало планового периода должен быть не меньше 2 — также с учетом удовлетво-

Начальный запас

$f_1(i,y)=C(3-i)+[(3-i)-y]^2$

Выпуск на предыдущем отрезке

i	$y=0$		$y=1$		$y=2$		$y=3$		$y=4$		$y=5$	
	$x_1(i,0)$	$f_1(i,0)$	$x_1(i,1)$	$f_1(i,1)$	$x_1(i,2)$	$f_1(i,2)$	$x_1(i,3)$	$f_1(i,3)$	$x_1(i,4)$	$f_1(i,4)$	$x_1(i,5)$	$f_1(i,5)$
0	3	28	3	23	3	20	3	19				
1	2	21	2	18	2	17	2	18	2	21		
2			1	15	1	16	1	19	1	24	1	31
3					0	4	0	9	0	16	0	25

Рис. 9.8. Модель сглаживания выпуска фирмы «Гарантия» ($n = 1$).

рения спроса на предыдущем отрезке, равном 3. Поскольку запас на конец планового периода, согласно принятым допущениям, равен нулю, оптимальный выпуск $x_1(i, y)$ зависит только от значения i, но не от y. Как будет показано ниже, при $n > 1$ $x_n(i, y)$ зависит и от значения y.

Вычисления для $n = 2$ и $y = 2$ систематизированы в таблице рис. 9.9. Читателю рекомендуется проанализировать приведенные

$[C(x)+1(i+x-3)]+(x-y)^2+f_1(i+x-3,x)$

Выпуск на предыдущем отрезке

Начальный запас

i \ x	0	1	2	3	4	5	$x_2(i,2)$	$f_2(i,2)$
0				19+1+19	22+4+21	25+9+31	3	39
1			17+0+20	20+1+18	23+4+24	26+9+25	2	37
2		15+1+23	18+0+17	21+1+19	24+4+16		2	35
3	0+4+28	16+1+18	19+0+16	22+1+9			0,3	32
4								

Рис. 9.9. Модель сглаживания выпуска фирмы «Гарантия» ($n = 2$, $y = 2$).

в таблице результаты и понять, каким образом они получены с помощью рекуррентного соотношения динамического программирования (13). В частности, заметим, что третье слагаемое в каждой из клеток основных столбцов таблицы взято из рис. 9.8. Так, если $i = 0$ и $x = 3$, то 19 есть значение $f_i(i + x - 3, x) = f_1(0, 3)$, записанное в строке $i = 0$ и правом из пары столбцов для $y = 3$ таблицы рис. 9.8. Заметим также, что строка $i = 4$ на рис. 9.9 полностью запрещена, поскольку запас на начало *предыдущего* отрезка

Начальный запас	Выпуск на предыдущем отрезке												(Рис. 8.10)
	$y=0$		$y=1$		$y=2$		$y=3$		$y=4$		$y=5$		
i	$x_2(i,0)$	$f_2(i,0)$	$x_2(i,1)$	$f_2(i,1)$	$x_2(i,2)$	$f_2(i,2)$	$x_2(i,3)$	$f_2(i,3)$	$x_2(i,4)$	$f_2(i,4)$	$x_2(i,5)$	$f_2(i,5)$	$x_2(i)$
0	3	47	3	42	3	39	3	38	3	39	3	42	3
1	2	41	2	38	2	37	2,3	38	3	39	3	42	5
2			2	36	2	35	2	36	2	39	4	41	4
3					0,3	32	3	31	3	32	3	35	0
4							2	25	2	28	2	33	0

Рис. 9.10. Модель сглаживания выпуска фирмы «Гарантия» ($n = 2$).

не мог превышать 4 единицы. При выпуске $y = 2$ и спросе 3 исходный запас на начало планового периода не может быть больше 3 единиц.

Теперь очевидно, что для этой модели объем вычислений *приблизительно* в 6 раз больше, чем для общей модели гл. 8, так как вычисления, аналогичные проделанным в таблице рис. 9.9, выполняются для каждого из возможных значений y, где $y = 0, 1, \ldots, 5$.

Полный набор значений $f_2(i, y)$ и соответствующих им значений $x_2(i, y)$ приведен в таблице рис. 9.10. Читателю следует проверить, понял ли он рекуррентное соотношение (13), самостоятельно вычислив значения $f_2(i, 5)$. Для удобства сравнения в правом столбце таблицы рис. 9.10 проставлены оптимальные объемы выпуска в задаче разд. 8.4, где не учитывают издержки сглаживания. Рассмотрим различие между решениями обеих задач. Заметим, что при $n = 2$ и начальном запасе 1 выпуск 5 единиц не является более оптимальным; теперь оптимален выпуск 3 единиц, если выпуск на предыдущем отрезке был не менее 3, и 2 единиц — в противном случае. Напротив, при начальном запасе 4 оптимальным является выпуск не 0, а 2 единиц.

Влияние, оказываемое введением в модель издержек сглаживания, проиллюстрировано в таблице рис. 9.11. Заметим, что, если исходный запас на начало планового периода i_0 равен 1 или 2, в «сглаженной» программе продукция выпускается как в январе, так и в феврале, тогда как в задаче фирмы «Надежный поставщик» положительным был только январский выпуск. Если i_0 равно 3 или 4, фирма «Гарантия» должна выпускать продукцию в январе, тогда как фирма «Надежный поставщик» выпускала это же число изделий в феврале. Предпочтительность января объясняется тем, что уменьшение выпуска до нуля с последующим его повторным возрастанием приводит в настоящей модели к увеличению затрат.

Исходный запас на начало планового периода, i_0	Фирма „Гарантия"		Фирма „Надежный поставщик"	
	При сглаживании выпуска		Без сглаживания выпуска	
	Янв.	Февр.	Янв.	Февр.
0	3	3	3	3
1	3	2	5	0
2	2	2	4	0
3	3	0	0	3
4	2	0	0	2

Рис. 9.11. Оптимальная производственная программа при длительности N планового периода, равной 2 отрезкам, и выпуске на предыдущем отрезке $x_0 = 4$.

В разд. 6.7 рассматривалась задача разработки производственной программы со сглаживанием занятости, близкая к настоящей модели. Управляющая переменная x_{ij} в задаче сглаживания занятости отображала число бригад, приступавших к работе в начале отрезка i и заканчивавших ее в начале отрезка j. Должно было выполняться следующее ограничение: число бригад, используемых на любом отрезке k ($k = 1, 2, \ldots, n-1$), должно быть не меньше R_k. Затраты, соответствующие x_{ij}, включали заработную плату, набор, перевозку и обучение рабочих. Затраты, не входящие в состав заработной платы, иногда делали экономичным содержание контингента работающих, превышающих минимальную потребность. Таким образом, в задаче ставилась цель разработки такого допустимого графика занятости работающих, при котором минимизируются общие затраты на содержание рабочей силы.

Можно установить следующую связь между такой задачей и моделью сглаживания, приведенной в настоящем разделе. Введем обозначения

$$x_t = \sum_{j=t+1}^{n-1} x_{tj} \quad \text{и} \quad D_t = R_t, \tag{I}$$

где x_t — общее число бригад, нанятых к началу отрезка t (вне зависимости от срока окончания ими работы). При этом выполняемую бригадой работу нельзя переносить из одного отрезка на следующий; если на некотором отрезке k число бригад превышает потребность в них, равную R_k, излишек рабочего времени нельзя сохранить для использования на отрезке $(k+1)$. В связи с этим можно сократить размерность переменной состояния в рекуррентном соотношении (6), исключив i. В результате возникает следующее рекуррентное соотношение динамического программирования:

$$g_n(y) = \min_{x \geq d_n} [g_{n-1}(x) + c_n(x, 0, y)] \quad \text{при} \quad n > 1, \tag{II}$$

$$g_1(y) = \min_{x \geqslant d_1} c_1(x, 0, y) \qquad \text{при} \quad n = 1. \tag{III}$$

С помощью этих соотношений отыскивается значение $g_N(y_0)$.

Отметим различия между соотношением (II) и задачей из разд. 6.7. Последняя обладала известной гибкостью — в том смысле, что издержки сглаживания зависели не только от момента i начала периода работы бригады, но и от длительности работы, т. е. от разности $(j - i)$. Вместе с тем затраты c_{ij} определялись для *каждой* бригады в отдельности, так что общая сумма затрат для некоторого периода являлась линейной функцией числа нанятых бригад.

Что же касается модели настоящего раздела, функция затрат которой имеет вид $C_t(x_t, 0, x_{t-1})$, то она обладает гибкостью в другом отношении. В издержках сглаживания учитывается только изменение уровня занятости для двух смежных отрезков, но не длительность работы каждой из нанимаемых или увольняемых бригад. Однако в данной модели появляется возможность учета нелинейной зависимости затрат от количества используемых бригад. Кроме того, изменение общей суммы издержек сглаживания может нелинейно зависеть от уровней занятости на двух смежных отрезках планового периода.

Если все затраты линейны, то

$$c_{ij} = c_i + c_{i+1} + \ldots + c_{j-1} + a_i + b_j, \tag{IV}$$

где затраты исчисляются для каждой бригады в отдельности, причем c_t — сумма заработной платы на отрезке t, a_t — затраты по найму и b_t — затраты, связанные с увольнением. В этом случае оптимальный график работы может быть найден с помощью любой из этих двух моделей; однако дополнительный анализ соотношения (II) позволяет предложить очень простой алгоритм, базирующийся на подходе, который был использован в модели управления запасами с выпуклой функцией затрат (разд. 9.3).

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

1. Проверьте сделанные в разд. 9.1 утверждения о том, что:

а) соотношения (3) и (4) эквивалентны, истолковав при этом содержание соотношения (4);

б) наложение ограничений $i_t \geqslant 0$, где i_t определяется из (4), обеспечивает своевременное удовлетворение спроса.

2. а) Дайте истолкование определений (3) и (4) разд. 9.2.

б) Начертите графики функций для примеров (I) — (IV) разд. 9.2, приняв для четырех первых примеров $a = 2$, $b = 1$ и $c = 3$, а для двух последних примеров — значения, приведенные в тексте.

в) Начертите график функции для примера (VI), приняв следующие значения параметров: $a_1 = 4$, $b_1 = 1$, $w_1 = 4$, $a_2 = 3$, $w_2 = 6$, $a_3 = 2$.

3. По данным, приведенным в таблице 9.2, составьте функции затрат вида (3) и (4) из разд. 9.3.

Упражнения 4—7 относятся к задаче фирмы «Вечная шина», значения параметров для которой приведены в таблице рис. 9.2. Примените алгоритм для решения задачи с измененными значениями параметров и вычислите приращение общей суммы затрат. Для каждого отрезка t укажите выпуск x_t и остаток на конец отрезка i_t.

4. а) Примите январский спрос равным 2 единицам (вместо 1).

б) Примите майский спрос равным 1 единице (вместо 0).

в) Предположите, что спрос на всех отрезках остается без изменения, кроме отрезка t. До какого предела можно увеличивать D_t, так чтобы по-прежнему существовала возможность построения допустимой программы? (Ответьте на этот вопрос для $t = 1, 2, \ldots, 6$.)

5. а) Примите максимально возможный выпуск января при нормальных затратах равным 4 единицам (вместо 3).

б) Примите максимально возможный выпуск января при нормальных затратах равным 5 единицам (вместо 3).

в) Считайте невозможным производство при повышенных затратах в марте и апреле.

6. Для каждого варианта постановки задачи объясните влияние изменения затрат на решение, приведенное в таблице рис. 9.3.

а) Производственные затраты на единицу продукции для всех отрезков одинаковы (например, 2 на единицу в основное и 5 на единицу в сверхурочное время).

б) Затраты на единицу для основного и сверхурочного времени равны между собой (например, для отрезка 1 возможен выпуск 9 единиц при затратах 2 на единицу, для отрезка 2—4 единиц при затратах 4 на единицу и т. д.).

в) Затраты на содержание запасов для всех отрезков равны $h = 6$ на единицу (вместо 1).

г) Затраты на содержание запасов для всех отрезков равны $h = 0{,}1$ на единицу (вместо 1).

7. Пусть совокупный спрос R_p — см. (1) разд. 9.4 — удовлетворяет неравенствам:

$$1 \leqslant R_1 \leqslant 1, \quad 1 \leqslant R_2 \leqslant 7, \quad 6 \leqslant R_3 \leqslant 14,$$
$$10 \leqslant R_4 \leqslant 20, \quad 10 \leqslant R_5 \leqslant 25, \quad 12 \leqslant R_6 \leqslant 27.$$

Найдите значения $X_p(S)$ и $X_p(T)$, определяемые теоремой о длительности планового периода для модели с выпуклой функцией затрат. Начертите график, аналогичный приведенному в разд. 9.4. Можете ли вы указать оптимальный выпуск для отрезка 1?

8. Рассмотрите частный случай модели с выпуклой функцией затрат, приведенной в конце разд. 9.4. Найдите оптимальную программу при следующих вариантах спроса:

а) $D_1 = 1$, $D_2 = 2$, $D_3 = 7$, $D_5 = 0$, $D_6 = 2$. Примите значение D_4 последовательно равным 5; 4; 3; 2 единицам.

б) $D_1 = 1$, $D_2 = 2$, $D_4 = 6$, $D_5 = 0$, $D_6 = 2$. Примите значение D_3 последовательно равным 11; 12 единицам.

в) $D_1 = 1$, $D_2 = 2$, $D_3 = 7$, $D_4 = 6$, $D_6 = 2$. Примите значение D_5 последовательно равным 1; 2; 4 единицам.

г) $D_3 = 7$, $D_4 = 6$, $D_5 = 0$, $D_6 = 2$. Примите $D_1 = 2$ и $D_2 = 1$, а $D_1 = 3$ и $D_2 = 0$.

9. Предположим, что поставщик предлагает следующую ступенчатую систему цен: 10 долл. за каждую единицу при заказе объемом до 12 единиц; 8 долл. за единицу при заказе объемом от 12 до 144 единиц; 5 долл. за каждую единицу при заказе объемом сверх 144 единиц. Начертите график зависимости общей суммы закупочных затрат от размера заказа. Является ли эта функция вогнутой? Если да, то почему? Останется ли функция затрат вогнутой, если включить в нее затраты на подготовительные операции? Если да, то почему?

10. а) Объясните, почему вид оптимальной программы, построенной для модели управления производством и запасами с вогнутой функцией затрат (разд. 9.5), обусловливает обязательное существование такой программы, в которой выпуск отрезка является нулевым, если запас на начало отрезка положителен.

б) Если такое свойство оптимальной программы считается заданным, вычислите, какое число разных допустимых вариантов программы необходимо рассмотреть. Для определенности положим, что все $D_t > 0$ (напомним также, что запас на начало планового периода $i_0 = 0$).

в) Вычислите, сколько различных значений c_{kj}, определяемых выражением (1) разд. 9.6, необходимо рассчитать для применения алгоритма решения общей модели с вогнутой функцией затрат.

11. Справедливость теоремы о виде оптимальной программы для модели управления производством и запасами с вогнутой функцией затрат доказывается в конце разд. 9.5. Доказательство основано на целесообразности либо увеличения x_p на величину i_{p-1}, либо увеличения на величину x_p выпуска на более раннем периоде. Пусть задана программа, в которой для одного или более отрезков выполняется условие $i_{t-1}x_t > 0$. Будет ли оптимальной измененная программа, полученная из заданной путем таких преобразований? Обоснуйте ваш ответ.

12. Задача фирмы «Растущий сбыт» (разд. 9.6). Рассмотрите пример, условия которого приведены в таблице рис. 9.5. Найдите оптимальную программу, если функция затрат на содержание запасов имеет вид $h_t(i_t) = h_t i_t$, где для всех отрезков

а) $h_t = 1$; б) $h_t = 2$.

13. В модели с вогнутой функцией затрат (разд. 9.5, 9.6) для всех отрезков планового периода $D_t = 1$

$$C_t(x_t) = \begin{cases} 5x_t & \text{при } x_t = 0, 1, 2, 3, \\ 12 + x_t & \text{при } x_t \geqslant 4 \end{cases}$$

и $h_t(i_t) = h_t i_t$. (Заметим, что это упражнение напоминает задачу фирмы «Растущий сбыт» в разд. 9.6, где в таблице рис. 9.5 функция производственных затрат для $t = 1$ полностью совпадает с приведенной в настоящем упражнении.) Найдите оптимальную программу, если для всех отрезков $h_t = 0{,}1$ при длительности планового периода:

а) $N = 4$; б) $N = 5$; в) $N = 6$; г) $N = 7$;

д) $N = 8$, е) $N = 9$;

ж) найдите оптимальную программу для $h_t = 1$ и $N = 9$;

з) то же для $h_t = 2$ и $N = 9$.

14. Выполните упражнение 13, взяв следующую функцию затрат:

$$C_t(x_t) = \begin{cases} 0 & \text{при} \quad x_t = 0, \\ 3 + 2x_t & \text{при} \quad x_t \geqslant 1. \end{cases}$$

15. Выполните упражнение 13, приняв для всех отрезков $D_t = 2$.

16. Выполните упражнение 13, приняв спрос изменяющимся по отрезкам следующим образом:

а) $D_1 = 1,\ D_2 = 2,\ D_3 = 1,\ D_4 = 2,\ \ldots,$

б) $D_1 = 2,\ D_2 = 1,\ D_3 = 2,\ D_4 = 1,\ \ldots.$

17. Выполните упражнение 13, приняв затраты на содержание запасов меняющимися от отрезка к отрезку следующим образом:

а) $h_1 = 1,\ h_2 = 2,\ h_3 = 1,\ h_4 = 2,\ \ldots,$

б) $h_1 = 2,\ h_2 = 1,\ h_3 = 2,\ h_4 = 1,\ \ldots.$

18. В конце разд. 9.6 утверждается, что использование рекуррентного соотношения динамического программирования (4) можно рассматривать как нахождение кратчайшего пути на ациклической сети. Рассмотрите вычисления, проделанные в (5), и повторите эти операции на сети рис. 9.7; покажите связь с нахождением кратчайшего пути на ациклической сети.

19. Пусть разрешена задержка в удовлетворении спроса, однако весь спрос должен быть удовлетворен к концу планового периода, т. е. для $t < N$ допустимы отрицательные значения i_t. Выполните упражнение 13, пп. а) — е), если функция затрат на содержание запасов и штрафа за задержку удовлетворения спроса имеет вид $h_t(i_t) = h_t \mid i_t \mid$, где

а) $h_t = 0{,}1$; б) $h_t = 1$; в) $h_t = 2$.

20. Рассмотрите примеры (1) — (3) разд. 9.7. Проверьте результаты, приведенные в (4) и (5).

21. Сравните сведения о программе, которые дают соответственно теорема о длительности планового периода для модели с выпуклой функцией затрат (разд. 9.4) и теорема о длительности планового периода для модели, учитывающей единовременные затраты (разд. 9.7).

22. Рассмотрите теорему о длительности планового периода для модели, учитывающей единовременные затраты (разд. 9.7); пусть первый месяц планового периода — январь.

а) Если $k_6 = 5$, какие выводы можно сделать об оптимальной программе на основе части (а) теоремы?

б) Если $k_6 = 3$, какие выводы можно сделать об оптимальной программе на основе части (б) теоремы?

23. а) Проверьте справедливость рекуррентного соотношения (7) из разд. 9.7 для случая линейной функции затрат на содержание запасов.

б) Внесите в (7) необходимые изменения для того случая, когда затраты на содержание запасов характеризуются вогнутой функцией общего вида.

24. Проверьте правильность результатов, приведенных в конце разд. 9.7, и объясните, почему в данном случае не выполняется часть (б) теоремы о длительности планового периода для модели, учитывающей единовременные затраты:

а) проверьте (II);

б) проверьте (IV).

25. Задача фирмы «Гарантия» (разд. 9.8). В рекуррентном соотношении (13) динамического программирования показан способ использования квадратичной функции издержек сглаживания.

а) Запишите функцию (13), включив в нее кусочно-линейную функцию издержек сглаживания (4), где для всех отрезков $a_t = b_t = 1$.

б) Используя полученную в п. а) функцию, постройте и заполните таблицы, аналогичные приведенным на рис. 9.8—9.11.

26. Предложите по крайней мере две функции издержек сглаживания, которые могут рассматриваться как приемлемая замена функций (3) и (4) из разд. 9.8.

27. Покажите, какие изменения необходимо внести в рекуррентное соотношение (6) из разд. 9.8 в том случае, когда затраты на отрезке t зависят не только от x_t, i_t и x_{t-2}, но и от x_{t-1}. Приняв существование такой зависимости в задаче фирмы «Гарантия», выявите увеличение объема вычислительных операций по сравнению с теми вычислениями, которые требовались для построения таблиц рис. 9.8—9.10.

28. Объясните, как вы понимаете следующие термины:

выпуклая функция;
вогнутая функция;
убывающая (возрастающая) предельная эффективность при увеличении масштаба производства;
анализ длительности планового периода;
динамическая модель экономически выгодного размера партии;
затраты на подготовительные операции или переналадку (единовременные затраты);
скидки за размер заказа (или ступенчатые цены);
точка возобновления;
издержки сглаживания.

УПРАЖНЕНИЯ НА РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

Упражнения 29—37 относятся к модели управления запасами с выпуклой функцией затрат, описанной в разд. 9.3 (напомним, что исходный запас на начало планового периода $i_0 = 0$).

29. Пусть функции производственных затрат заданы таблицей рис. 9.12, например для отрезка 1 выпуску x_1, равному 12, соответ-

Затраты на выпуск единицы k

Выпуск единицы k	Отрезок 1	Отрезок 2	Отрезок 3	Отрезок 4
$1 \leqslant k \leqslant 5$	1	2	1	4
$6 \leqslant k \leqslant 10$	3	3	6	4
$11 \leqslant k \leqslant 15$	6	5	7	8
$16 \leqslant k \leqslant 20$	10	8	12	9

Рис. 9.12.

ствует общая сумма затрат 32 $(1 \cdot 5 + 3 \cdot 5 + 2 \cdot 6 = 32)$. Пусть, далее, затраты на содержание запасов равны $h_t(i_t) = h_t i_t$, где $h_1 = 1$, $h_2 = 2$ и $h_3 = 1$, а спрос составляет соответственно

$$D_1 = 10, \quad D_2 = 3, \quad D_3 = 17, \quad D_4 = 23.$$

а) Найдите оптимальную производственную программу. Для каждого из отрезков t укажите выпуск x_t и уровень запасов на конец отрезка i_t.

б) Найдите оптимальную производственную программу, приняв для всех отрезков ограничение на уровень запасов $i_t \leqslant 5$.

в) Определите влияние увеличения на единицу спроса на первом отрезке (для которого $D_1 = 11$); на втором отрезке; на третьем отрезке; на четвертом отрезке.

г) Найдите оптимальную производственную программу, приняв для всех отрезков $h_t = 0$; $h_t = 5$.

д) Найдите оптимальную производственную программу, изменив спрос следующим образом: $D_1 = 3$ и $D_2 = 10$; $D_3 = 23$ и $D_4 = 17$; $D_2 = 17$ и $D_3 = 3$.

30. Пусть исходные данные задачи соответствуют принятым в п. а) упражнения 29. Примем, что совокупный спрос R_p, определяемый в соответствии с выражением (1) из разд. 9.4, удовлетворяет неравенствам

$$\sum_{t=1}^{p} D_t - k \leqslant R_p \leqslant \sum_{t=1}^{p} D_t + k.$$

Для каждого из указываемых ниже вариантов задачи найдите значения величин $X_p(S)$ и $X_p(T)$, определенные в теореме о длительности планового периода для модели с выпуклой функцией затрат, вычислите значения соответствующих объемов выпуска и уровней запасов на конец отрезка и начертите график, аналогичный приведенному на рис. 9.4. Объясните, можно ли найти оптимальный выпуск для отрезка 1:

а) $k = 1$; б) $k = 2$; в) $k = 3$.

31. Пусть исходные данные задачи соответствуют п. а) упражнения 29. Примем, что совокупный спрос R_p, определяемый в соответствии с выражением (1) из разд. 9.4, удовлетворяет неравенствам

$$\sum_{t=1} D_t - pk \leqslant R_p \leqslant \sum_{t=1}^{p} D_t + pk.$$

Для каждого из указываемых ниже вариантов задачи найдите значения $X_p(S)$ и $X_p(T)$, определенные в теореме о длительности планового периода для модели с выпуклой функцией затрат, вычислите значения соответствующих объемов выпуска и уровней запасов на конец отрезка и начертите график, аналогичный приведенному на фиг. 9.4. Объясните, можно ли найти оптимальный выпуск для отрезка 1:

а) $k = 1$; б) $k = 2$; в) $k = 3$.

32. Пусть длительность планового периода N равна 3 отрезкам, причем функции производственных затрат и затрат на содержание запасов для всех отрезков имеют вид $C_t(x_t) = x_t^2$ и $h_t(i_t) = h_t i_t$. В каждом из приведенных ниже пунктов упражнения найдите оптимальную производственную программу и укажите уровень запасов на конец каждого из отрезков.

а) Примите для всех отрезков t $h_t = 0$ и $D_t = 6$.

б) Примите для всех отрезков t $h_t = 0$ и $D_1 = 11$, $D_2 = 5$, $D_3 = 2$.

в) Примите для всех отрезков t $h_t = 0$ и $D_1 = 2$, $D_2 = 5$, $D_3 = 11$.

г) Используйте те же данные, что и в п. в), дополнительно наложив на каждый из отрезков ограничение на объем выпуска $x_t \leqslant 5$.

д) Используйте те же данные, что и в п. в), дополнительно ограничив для каждого из отрезков уровень запасов на конец отрезка $i_t \leqslant 3$.

е) Используйте для всех отрезков $h_t = 3$ и те же данные о спросе, что и в п. в). Затем примите для всех отрезков $h_t = 11$. Сравните результаты, полученные для $h_t = 0$, $h_t = 3$ и $h_t = 11$.

ж) Примите для всех отрезков $h_t = 0$, а также используйте те же данные о спросе, что и в п. в), и $C_3(x_3) = 1{,}5\, x_3^2$ Затем примите $h_t = 3$ и, наконец, $h_t = 11$. Сравните полученные результаты с результатами п. е).

33. Примите функции производственных затрат и затрат на содержание запасов такими же, как и в упражнении 32. Пусть совокупный спрос, определяемый в соответствии с (1) из разд. 9.4, удовлетворяет неравенствам

$$1 \leqslant R_1 \leqslant 3, \quad 6 \leqslant R_2 \leqslant 8, \quad 23 \leqslant R_3 \leqslant 29$$

и $h_t = 0$. Найдите значения величин $X_p(S)$ и $X_p(T)$, определенных в теореме о длительности планового периода для модели с выпуклой функцией затрат; вычислите значения соответствующих объемов выпуска и уровней запасов и начертите график, аналогичный приведенному на рис. 9.4. Объясните, можно ли найти оптимальный выпуск для отрезка 1. Выполните те же операции при $h_t = 3$ и $h_t = 11$.

34. Примите, что разрешена задержка в удовлетворении спроса и что функции производственных затрат, затрат на содержание запасов и штрафов за своевременно неудовлетворенный спрос для всех отрезков имеют следующий вид:

$$C_t(x_t) = x_t^2; \quad h_t(i_t) = \begin{cases} h_t i_t & \text{при } i_t \geqslant 0, \\ -p_t i_t & \text{при } i_t \leqslant 0. \end{cases}$$

Для каждого из пп. а) — д) упражнения найдите оптимальную производственную программу и укажите уровень запасов или неудовлетворенный спрос на конец каждого из отрезков (спрос должен быть полностью удовлетворен к концу планового периода).

а) Примите для всех отрезков $h_t = p_t = 0$, а спрос равным соответственно

$$D_1 = 2, \quad D_2 = 5, \quad D_3 = 11.$$

б) Примите для всех отрезков $h_t = p_t = 0$, а спрос равным соответственно

$$D_1 = 11, \quad D_2 = 5, \quad D_3 = 2.$$

в) Примите спрос таким же, как в п. б), а $h_t = 0$ и $p_t = 3$ для всех отрезков. Затем примите для всех отрезков $h_t = 0$ и $p_t = 11$.

г) Примите для всех отрезков $h_t = p_t = 0$, а спрос равным соответственно

$$D_1 = 2, \quad D_2 = 11, \quad D_3 = 5.$$

д) Примите спрос таким же, как в п. д), а $h_t = p_t = 3$ для всех отрезков. Затем для всех отрезков примите $h_t = p_t = 11$ и, наконец, $h_t = 2$ и $p_t = 4$.

35. Пусть длительность планового периода N равна 8, и для всех отрезков функции производственных затрат и затрат на содержание запасов имеют следующий вид: $C_t(x_t) = x_t^2$, $h_t(i_t) = h_t i_t$. Примите спрос равным соответственно

$$D_1 - 4, \quad D_2 = 20, \quad D_3 = 2, \quad D_4 = 0, \quad D_5 = 6, \quad D_6 = 4,$$
$$D_7 = 1, \quad D_8 = 3.$$

Для пп. а) — ж) упражнения найдите оптимальную производственную программу и укажите уровень запасов на конец каждого из отрезков.

а) Примите для всех отрезков $h_t = 0$.

б) Примите для всех отрезков поочередно $h_t = 1$; $h_t = 3$.

в) Примите для всех отрезков $h_t(i_t) = i_t^2$.

г) Примите для всех отрезков $h_t = 0$. Кроме того, примите $C_t(x_t) = 1{,}5\, x_t^2$ поочередно для $t = 2$; $t = 4$; $t = 7$.

д) Примите для всех отрезков $h_t = 0$; длительность планового периода N равна 9. Каким может быть наибольшее значение D_9, при котором оптимальный объем выпуска x_7, найденный для п. а), остается неизменным? То же для x_3? То же для x_1?

е) Примите для всех отрезков $h_t = 1$. Ответьте на вопросы, которые поставлены в п. д).

ж) Примите для всех отрезков $h_t = 3$. Ответьте на вопросы, которые поставлены в п. д).

36. Функции производственных затрат и затрат на содержание запасов остаются такими же, как и в упражнении 35; для всех отрезков $h_t = 0$. Найдите значения величин $X_p(S)$ и $X_p(T)$, определенных в теореме о длительности планового периода для модели с выпуклой функцией затрат, соответствующие значения объемов выпуска и уровней запасов на конец отрезка и начертите график, аналогичный приведенному на рис. 9.4. Объясните, можно ли найти оптимальный выпуск для отрезка 1.

а) $\sum\limits_{t=1}^{p} D_t - k \leqslant R_p \leqslant \sum\limits_{t=1}^{p} D_t + k$. Выполните упражнение для $k = 1$ и $k = 2$.

б) $\sum\limits_{t=1}^{p} D_t - pk \leqslant R_p \leqslant \sum\limits_{t=1}^{p} D_t + pk$. Выполните упражнение для $k = 1$ и $k = 2$.

в) Выполните такой же анализ, если для всех отрезков $h_t = 1$.

37. Разрешена задержка удовлетворения спроса; спрос и функция производственных затрат остаются такими же, как и в упражнении 35; функция затрат на содержание запасов и штрафов за своевременно неудовлетворенный спрос имеет следующий вид:

$$h_t(i_t) = \begin{cases} h_t i_t & \text{при } i_t \geqslant 0, \\ -p_t i_t & \text{при } i_t \leqslant 0. \end{cases}$$

(Спрос должен быть полностью удовлетворен к концу планового периода.) Найдите оптимальную производственную программу для вариантов а) — д) задачи и укажите значение уровня запасов или неудовлетворенного спроса на конец каждого из отрезков. Для всех отрезков:

а) $h_t = p_t = 0$; б) $h_t = p_t = 1$; в) $h_t = p_t = 3$; г) $h_t = 1$ и $p_t = 3$; д) $h_t = 3$ и $p_t = 1$.

Упражнения 38—43 относятся к модели управления запасами с вогнутой функцией затрат, приведенной в разд. 9.5 (напомним, что исходный уровень запасов на начало планового периода $i_0 = 0$).

38. Длительность планового периода N равна 3; функции производственных затрат и затрат на содержание запасов имеют следующий вид:

$$C_1(x_1) = 12 + 8x_1 \quad \text{при} \quad x_1 \geqslant 1,$$
$$C_2(x_2) = 2 + 9x_2 \quad \text{при} \quad x_2 \geqslant 1,$$
$$C_3(x_3) = 5 + 10x_3 \quad \text{при} \quad x_3 \geqslant 1$$

и $h_t(i_t) = h_t i_t$. Спрос равен соответственно $D_1 = 0$, $D_2 = 3$, $D_3 = 20$. Найдите оптимальную производственную программу для пп. а) — в) упражнения и укажите уровень запасов на конец каждого из отрезков. Для всех отрезков:

а) $h_t = 0$; б) $h_t = 1$; в) $h_t = 2$.

39. Длительность планового периода N равна 4, а параметры функций производственных затрат приведены в таблице рис. 9.13;

Затраты на выпуск единицы k

Выпуск единицы k	Отрезок 1	Отрезок 2	Отрезок 3	Отрезок 4
$1 \leqslant k \leqslant 5$	10	8	12	9
$6 \leqslant k \leqslant 10$	6	5	7	8
$11 \leqslant k \leqslant 15$	3	3	6	4
$16 \leqslant k \leqslant 20$	1	2	1	4

Рис. 9.13.

например, для отрезка 1 выпуску $x_1 = 12$ соответствует общая сумма затрат 86 ($10 \cdot 5 + 6 \cdot 5 + 2 \cdot 3 = 86$). Функции затрат на содержание запасов имеют вид $h_t(i_t) = h_t i_t$, где $h_1 = 1$, $h_2 = 2$ и $h_3 = 1$. Спрос равен соответственно

$$D_1 = 10, \quad D_2 = 3, \quad D_3 = 17, \quad D_4 = 23.$$

а) Найдите оптимальную производственную программу. Укажите для всех отрезков объем выпуска x_t и уровень запасов i_t на конец отрезка t.

б) Выявите влияние увеличения на единицу спроса на отрезке 1 (так что $D_1 = 11$); на отрезке 2; на отрезке 3; на отрезке 4.

в) Найдите оптимальную производственную программу, если для всех отрезков $h_t = 0$; $h_t = 5$.

г) Найдите оптимальную производственную программу при следующих изменениях спроса: $D_1 = 3$ и $D_2 = 10$; $D_3 = 23$ и $D_4 = 17$; $D_2 = 17$ и $D_3 = 3$.

д) Найдите оптимальную производственную программу, если для всех отрезков функция затрат на содержание запасов имеет вид $h_t(i_t) = \sqrt{i_t}$.

40. Пусть исходные параметры задачи остаются такими же, как и в упражнении 30, однако разрешена задержка в удовлетворении спроса. Для всех отрезков функция затрат на содержание запасов и штрафов за несвоевременное удовлетворение спроса имеет следующий вид:

$$h_t(i_t) = \begin{cases} h_t i_t & \text{при } i_t \geqslant 0, \\ -p_t i_t & \text{при } i_t \leqslant 0. \end{cases}$$

(К концу планового периода спрос должен быть полностью удовлетворен.) Найдите оптимальную производственную программу для пп. а) — в) упражнения и укажите уровень запасов или неудовлетворенный спрос на конец каждого из отрезков. Для всех отрезков принять:

а) $h_t = p_t = 0$; б) $h_t = p_t = 1$;

в) вместо приведенной выше функции затрат на содержание запасов и штрафов за неудовлетворенный спрос используйте функцию $h_t(i_t) = \sqrt{|i_t|}$.

41. Длительность планового периода N равна 8; для всех отрезков функция производственных затрат имеет вид $C_t(x_t) = \sqrt{x_t}$; спрос равен соответственно:

$$D_1 = 4,\ D_2 = 20,\ D_3 = 2,\ D_4 = 0,\ D_5 = 6,\ D_6 = 4,\ D_7 = 1,\ D_8 = 3.$$

Найдите оптимальную производственную программу для пп. а) — в) задачи и укажите уровень запасов на конец каждого отрезка. Для всех отрезков функция затрат на содержание запасов равна:

а) $h_t(i_t) = 0$; б) $h_t(i_t) = i_t$; в) $h_t(i_t) = \sqrt{i_t}$.

42. Пусть исходные параметры задачи остаются такими же, как в упражнении 41, однако разрешена задержка в удовлетворении спроса. Для всех отрезков функция затрат на содержание запасов и штрафов за несвоевременное удовлетворение спроса имеет следующий вид:

$$h_t(i_t) = \begin{cases} h_t i_t & \text{при } i_t \geqslant 0, \\ -p_t i_t & \text{при } i_t \leqslant 0. \end{cases}$$

(К концу планового периода спрос должен быть полностью удовлетворен.) Найдите оптимальную производственную программу для пп. а) — в) упражнения и укажите уровень запасов или неудовлетворенный спрос на конец каждого из отрезков. Для всех отрезков:

а) $h_t = p_t = 0$; б) $h_t = p_t = 1$;

в) вместо приведенной выше функции затрат на содержание запасов и штрафов за неудовлетворенный спрос используйте функцию $h_t(i_t) = \sqrt{|i_t|}$.

43. Для пп. а) — ж) упражнения укажите оптимальную производственную программу и соответствующие уровни запасов на конец каждого из отрезков при длительности планового периода N, где значения N приведены в условиях для соответствующих пунктов. Для всех отрезков функции производственных затрат и затрат на содержание запасов имеют следующий вид:

$$C_t(x_t) = \begin{cases} 0 & \text{при} \quad x_t = 0, \\ s(t) + c_t x_t & \text{при} \quad x_t \geqslant 1 \end{cases} \quad \text{и} \quad h_t(i_t) = h_t i_t.$$

а) Пусть $N = 10$ и для всех отрезков $s_t = 36$, $c_t = 5$, $h_t = 1$, $D_t = 8$.

б) Используйте данные для п. а), изменив лишь спрос: примите $D_t = 20$ для отрезков 3, 6 и 9 и $D_t = 2$ для всех остальных отрезков.

в) Используйте данные для п. а), изменив лишь спрос: примите $D_t = 20$ для отрезков 2, 5 и 8 и $D_t = 2$ для всех остальных отрезков.

г) Используйте данные для п. а), изменив лишь спрос: примите $D_t = 20$ для отрезков 1, 4, 7 и 10 и $D_t = 2$ для всех остальных отрезков.

д) Пусть $N = 12$, $c_t = 5$, $h_t = 1$; $D_t = 8$ для всех отрезков, $s_t = 100$ для отрезков 4, 5, 6, 10, 11 и 12 и $s_t = 36$ для всех остальных отрезков.

е) Используйте данные для п. а), однако считайте, что функция затрат на содержание запасов имеет вид $h_t(i_t) = \sqrt{i_t}$.

ж) Используйте данные для п. а), однако считайте разрешенной задержку в удовлетворении спроса. Для всех отрезков функция затрат на содержание запасов и штрафов за своевременно неудовлетворенный спрос имеет вид $h_t(i) = |i_t|$.

з) Для пп. а) — д) покажите, каким образом вычислительный алгоритм можно упростить благодаря использованию теоремы о длительности планового периода для модели, учитывающей единовременные затраты (разд. 9.7). Покажите также возможность применения рекуррентного соотношения (7).

44. Рассмотрите задачу, приведенную в упражнении 44 гл. 8. Можно ли получить оптимальную программу, используя:

а) описанный в разд. 9.3 алгоритм для модели управления запасами с выпуклой функцией затрат;

б) описанный в разд. 9.6 алгоритм для модели управления запасами с вогнутой функцией затрат.

в) Дайте истолкование результатам, полученным для пп. а) и б).

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

45. Рассмотрите модель управления запасами, описанную в разд. 9.1. Пусть производственные затраты на отрезке t включают два элемента: затраты на рабочую силу $L_t(x_t)$ и стоимость сырья. Последняя исчисляется следующим образом: на отрезке r фирма может закупать любое количество сырья по цене p_r за единицу (единица измерения количества сырья выбрана таким образом, что она совпадает с расходом сырья на выработку единицы готовой продукции). Цены p_r для всех отрезков известны с начала планового периода. При закупке сырья в начале планового периода и его хранении для последующего использования фирма несет расходы по содержанию запасов; оплата каждой единицы запаса сырья на конец отрезка r составляет H_r. Предложите рациональный метод учета решений о закупке сырья в модели, описанной в разд. 9.3 и 9.5.

46. Для пп. а) и б) изложите способ модификации алгоритма, позволяющий учесть положительный уровень исходного запаса на начало планового периода ($i_0 > 0$). Вначале укажите, какую целесообразно ввести модификацию, в случае когда функции затрат на содержание запасов имеют для всех отрезков вид $h_t(i_t) = h_t i_t$. Затем покажите, как изменяется подход, если не все функции затрат на содержание запасов линейны:

а) в модели управления запасами с выпуклой функцией затрат, описанной в разд. 9.3;

б) в модели управления запасами с вогнутой функцией затрат, описанной в разд. 9.6.

47. а) Предложите модификацию алгоритма, описанного в разд. 9.3, позволяющую учесть наличие верхних границ u_t объемов выпуска x_t и верхних границ уровня запасов b_t на конец отрезка t.

б) Рассмотрите модель с вогнутой функцией затрат, описанную в разд. 9.5, и примите, что после хранения в течение K отрезков хранимая продукция устаревает, так что ее нельзя использовать. Предложите разумное изменение как формулировки теоремы о виде оптимальной программы, так и соответствующего алгоритма.

48. Рассмотрите модель управления запасами, описанную в разд. 9.1, приняв запас на начало планового периода $i_0 = 0$. Если не учитывать условия целочисленности переменных, ограничения (4) и (5) можно представить в следующем виде:

$$\sum_{k=1}^{t} x_k \geqslant \sum_{k=1}^{t} D_k, \qquad t = 1, 2, \ldots, N,$$

причем строгое равенство обязательно соблюдается для $t = N$ и все $x_k \geqslant 0$. Пусть минимизируемая функция затрат имеет вид $C(x_1, x_2, \ldots, x_N; D_1, D_2, \ldots, D_N)$. Предположим, что эта функция является вогнутой по всем управляющим переменным x_k; это означает, что если имеются две допустимые программы, а третья программа строится как выпуклая комбинация двух этих программ (т. е. является их средневзвешенной при любых неотрицательных значениях выбранных весов), то затраты, соответствующие третьей программе, не превышают величину, полученную взвешиванием затрат, отвечающих двум первым программам при тех же весах. Можно показать (разд. 14.9), что оптимальная программа соответствует экстремальной точке системы ограничений. (Вспомним, что если программа соответствует экстремальной точке, ее нельзя представить в виде выпуклой комбинации других допустимых программ.)

а) Проверьте, обладает ли программа, соответствующая экстремальной точке, следующим свойством: если существуют отрезки p и q, $p < q$, такие, что

$$x_p > 0, \quad x_{p+1} = x_{p+2} =, \ldots, = x_{q-1} = 0, \quad x_q > 0,$$

то

$$x_p = \sum_{t=p}^{q-1} D_t.$$

б) Покажите, что а) эквивалентно теореме о виде оптимальной программы, сформулированной в разд. 9.5. Объясните, почему при целых значениях всех D_t оптимальная программа характеризуется целыми значениями управляющих переменных.

в) Покажите, что число допустимых программ, соответствующих экстремальным точкам, равно 2^{N-1} (при $D_1 > 0$). Объясните, как этот результат можно использовать при определении оптимальной программы.

В упражнениях 49—58 необходимо построить модель на основе рекуррентных соотношений динамического программирования. Дайте определения всем используемым условным обозначениям. Предложите адекватную целевую функцию для того случая, когда число отрезков до конца планового периода равно 1, а также рекуррентное соотношение для случая n отрезков до конца планового периода. Объясните, с чего следует начать и чем закончить вычислительный процесс.

49. Рассмотрите модель управления запасами, описанную в разд. 9.1.

а) Добавьте ограничение, согласно которому для всех $t = 1, 2, \ldots, N$ должно выполняться соотношение $(i_{t-1} + i_t)/2 \leqslant B_t$, где все B_t — неотрицательные целые числа.

б) Добавьте ограничение, в соответствии с которым $x_{t-1} + x_t \leqslant v_t$ для всех $t = 1, 2, \ldots, N$, причем все v_t — неотрицательные целые числа, а $x_0 = 0$.

в) Добавьте ограничение, согласно которому $x_{t-1} + x_t \geqslant 1$ для всех $t = 1, 2, \ldots, N$, причем $x_0 = 0$.

50. Задача транспортного обслуживания. Несколько фирм, вырабатывающих пищевые продукты, направляют грузы, предназначенные для продовольственного магазина «Радость хозяйки», на один из складов того города, где расположен магазин. Магазин платит c долл. в сутки за каждые 100 куб. футов складского помещения, которые он использует. Предположим, что владельцы магазина в каждый день t в течение ближайших N дней предполагают получать груз, объем которого равен D_t сотен куб. футов. Предположим также, что ежедневно владельцы магазина могут заказать в транспортной фирме грузовик, который доставит со склада хранящиеся там товары. Заказ грузовика стоит s долл., поэтому владельцы магазина обычно пользуются им не каждый день. В то же время они не допускают, чтобы какой-либо груз хранился на складе более n дней (примем, что грузы поступают на склад по утрам, а грузовик забирает грузы со склада в конце дня; примем также, что плата за арендуемое помещение исчисляется исходя из максимального объема помещения, использованного в течение дня). Покажите метод построения такого графика доставки груза со склада, чтобы он был оптимальным для владельцев магазина.

51. Задача размещения распределительных нефтебаз. Фирма «Супернефтепродукты» контролирует значительную часть торговли бензином в обширном районе, расположенном на полуострове. В районе имеется N городов, соединенных шестиполосной автострадой. Пусть городам присвоены индексы от 1 до N, причем таким образом, что из города j в город k ($j < k$) необходимо проехать по автостраде мимо городов $j+1$, $j+2$, $\ldots$, $k-1$. Фирма «Супернефтепродукты» собирается разместить распределительные нефтебазы в одном или нескольких городах. Годовые расходы по содержанию нефтебазы не зависят от ее величины и для города j составляют s_j. Прогнозируемый годовой спрос на бензин для города j равен D_j, а расстояние между соседними городами j и $(j+1)$ составляет $L_{j,\,j+1}$. Если распределительная нефтебаза, расположенная в городе j, обслуживает город k, то соответствующий годовой расход на перевозку бензина будет равен $h \cdot D_k \cdot L_{jk}$. Фирма стремится найти оптимальную схему размещения нефтебаз. Постройте модель динамического программирования для решения этой задачи:

а) приняв, что обслуживание города k может осуществляться только нефтебазой, расположенной в городе j, причем $j < k$;

б) приняв, что обслуживание города k должно осуществляться той нетебазой, которая ближе всего к этому городу.

52. Задача выбора сортамента продукции. Японская промышленная фирма «Фудзи стил» выпускает строительные балки стандартной длины. Прочность балки зависит от ее веса, и фирма «Фудзи стил» присвоила балкам разного веса индексы j, принимающие значения от 1 до N, где $j = 1$ соответствует самой тяжелой, а $j = N$ — самой легкой балке. В том случае, когда заказчику требуются балки, прочность которых равна k, фирма «Фудзи стил» может, если пожелает, удовлетворить потребность в балках большей прочности j ($j < k$). Фирме «Фудзи стил» приходится решать следующую задачу выбора сортамента продукции. Спрос на балки прочности j равен D_j; весь спрос должен быть полностью удовлетворен. Если фирма «Фудзи стил» принимает решение о выпуске балок j, это требует больших затрат на переналадку, которые равны s_j. Если же фирма удовлетворяет спрос D_j балками k ($k \leqslant j$), это влечет за собой потерю выручки, равную $h\,(w_k - w_j)\,D_j$, где w_k и w_j — вес балок k и j соответственно, а h — цена единицы веса. Построить модель динамического программирования, которая позволила бы фирме «Фудзи стил» найти оптимальную структуру сортамента и объем выпуска балок каждого типа.

53. Рассмотрите модель управления запасами, описанную в разд. 9.1. Предположим, что если выпуск на предыдущем отрезке был нулевым, то выпуск на следующем отрезке связан с «пусковыми затратами». Для конкретности примем, что функция затрат на отрезке t равна

$$C_t(x_t, i_t, x_{t-1}) = C_t(x_t) + h_t(i_t) + v_t(x_{t-1}, x_t),$$

где

$$v_t(x_{t-1}, x_t) = \begin{cases} K_t & \text{при } x_t \geqslant 1 \text{ и } x_{t-1} = 0, \\ 0 & \text{в остальных случаях.} \end{cases}$$

а) Сформулируйте задачу в виде модели динамического программирования. Сравните вычислительные трудности, возникающие при решении рекуррентной модели, с аналогичными трудностями для модели разд. 9.1 (не делая никаких дополнительных допущений о виде функции затрат).

б) Пусть
$$C_t(x_t) = \begin{cases} 0 & \text{при } x_t = 0, \\ s_t + c_t x_t & \text{при } x_t \geqslant 1, \end{cases}$$

причем функция $h_t(i_t)$ является вогнутой. Примем для всех t $s_t \geqslant 0$, $c_t \geqslant c_{t+1}$, спрос $D_t > 0$, запас на начало планового периода $i_0 = 0$. Предположим, что существует оптимальная программа, в которой для всех отрезков $i_{t-1}x_t = 0$. Предложите соответствующую модификацию алгоритма, приведенного в разд. 9.6. Начертите сеть, отыскание кратчайшего пути на которой может служить иллюстрацией предлагаемого подхода.

в) Докажите, что допущения, принятые в п. б), подразумевают существование оптимальной программы, для которой $i_{t-1}x_t = 0$ $(t = 1, 2, \ldots, N)$.

54. Рассмотрите модель управления запасами, описанную в разд. 9.1. Пусть при длительности планового периода, равной N отрезков, выпуск разрешен в течение не более чем K отрезков; это означает, что не более K объемов выпуска x_t $(t = 1, 2, \ldots, N)$ могут быть положительными. Сформулируйте задачу в виде модели динамического программирования. Сравните вычислительные трудности, возникающие при решении рекуррентной модели, с аналогичными трудностями для модели разд. 9.1 (не делая никаких дополнительных допущений о виде функции затрат). Покажите, как изменится решение, если в течение планового периода число отрезков с ненулевым выпуском должно быть в точности равно K.

В упражнениях 55—58 однопродуктовая модель управления запасами, описанная в разд. 9.1, обобщается на случай двух видов запасаемых продуктов (для дальнейшего обобщения модели на большее число продуктов требуется лишь простое расширение получаемых ниже результатов). Обозначим D_{1t} и D_{2t} спрос на отрезке t на продукты 1 и 2 соответственно, x_{1t} и x_{2t} — выпуск этих двух видов продукции, i_{1t} и i_{2t} — уровень их запасов на конец каждого из отрезков. (Все допущения, не сформулированные в явном виде, являются соответствующими обобщениями предположений, сделанных в модели разд. 9.1.)

55. Пусть целевая функция имеет вид

$$\text{Минимизировать} \sum_{t=1}^{N} C_t(x_{1t}, x_{2t}, i_{1t}, i_{2t}),$$

где

$$C_t(x_{1t}, x_{2t}, i_{1t}, i_{2t}) = C_t(x_{1t}, x_{2t}) + h_{1t}(i_{1t}) + h_{2t}(i_{2t}).$$

Постройте модель динамического программирования и сравните вычислительные трудности, возникающие при решении этой модели, с аналогичными трудностями для модели разд. 9.1 (не делая никаких дополнительных допущений о виде функции затрат).

56. Пусть функции производственных затрат в упражнении 55 имеют вид

$$C_t(x_{1t}, x_{2t}) = \begin{cases} 0 & \text{при } x_{1t} = x_{2t} = 0, \\ s_t + C_{1t}x_{1t} + C_{2t}x_{2t} & \text{в остальных случаях,} \end{cases}$$

где $s_t \geqslant 0$, $C_{1t} \geqslant C_{1,\, t+1}$ и $C_{2t} \geqslant C_{2,\, t+1}$, а функции затрат на содержание запасов $h_{1t}(i_{1t})$ и $h_{2t}(i_{2t})$ являются вогнутыми.

а) Примите исходные уровни запасов на начало планового периода $i_{10} = i_{20} = 0$. Предположим, что существует оптимальная программа, в которой для всех отрезков $(i_{1,\, t-1} + i_{2,\, t-1})(x_{1t} + x_{2t}) = 0$. Предложите соответствующую модификацию алгоритма, приведенного в разд. 9.6, которая позволила бы найти оптимальную программу производства обоих видов продукции.

б) Докажите существование оптимальной программы, характер которой соответствует предположению, сделанному в п. а).

57. Пусть функции производственных затрат в упражнении 55 имеют вид

$$C_t(x_{1t}, x_{2t}) = \begin{cases} 0 & \text{при } x_{1t} = x_{2t} = 0, \\ s_t + C_{1t}(x_{1t}) + C_{2t}(x_{2t}) & \text{в остальных случаях,} \end{cases}$$

где $s_t \geqslant 0$, функции $C_{1t}(x_{1t})$ и $C_{2t}(x_{2t})$, а также функции затрат на содержание запасов $h_{1t}(i_{1t})$ и $h_{2t}(i_{2t})$ являются вогнутыми.

а) Считайте, что свойства оптимальной программы, сформулированные в разд. 9.5, сохраняются для каждого из продуктов. Объясните, как можно использовать информацию об этом допущении для упрощения вычислений по рекуррентным формулам в случае упражнения 55.

б) Докажите справедливость сделанного в п. а).

58. Пусть функция производственных затрат в упражнении 55 имеет вид

$$C_t(x_{1t}, x_{2t}) = C_{1t}(x_{1t}) + C_{2t}(x_{2t}),$$

а на объемы выпуска для каждого из отрезков наложены ограничения $a_{1t}x_{1t} + a_{2t}x_{2t} \leqslant b_t$, где a_{1t}, a_{2t}, b_t — неотрицательные целые числа. Предложите для упражнения 55 соответствующую модификацию модели динамического программирования.

ГЛАВА 10

Еще о динамическом программировании

10.1. ВВЕДЕНИЕ

Построение модели, описываемой рекуррентным соотношением динамического программирования, является до некоторой степени искусством. В связи с этим в разд. 10.2—10.6 приводятся разнообразные примеры, позволяющие читателю полнее прочувствовать приемы, используемые при построении таких моделей. Поскольку в гл. 8 были подробно рассмотрены два примера численного решения моделей динамического программирования, в данной главе вычислительным аспектам, связанным с моделями этого класса, не уделяется особого внимания.

В конце главы, в разд. 10.7—10.10, дается общая оценка динамического программирования. В частности, обсуждаются элементы, общие для любых постановок задач динамического программирования, и показывается, что анализ этих элементов позволяет получить ценную информацию для принятия организационных решений. Дается некоторое представление о различиях между теоретическим исследованием моделей динамического программирования и их практическим применением. В заключение кратко освещаются реальные промышленные ситуации, рассмотрение которых на основе использования моделей этого рода принесло ощутимый экономический эффект, вскрываются также причины, содействовавшие получению положительных результатов.

В излагаемых ниже примерах использования моделей сначала приводятся многие второстепенные подробности, а затем дается более схематичное описание и от читателя требуется *тщательное* изучение этих схем, в которые ему следует самостоятельно внести необходимые уточнения. Указания, данные в гл. 8, сохраняют силу. Читателю надлежит проявить терпение и отвести достаточно много времени на проработку разделов этой главы. Нужно предусмотреть по крайней мере двукратное прочтение каждого раздела. Для проверки понимания каждой из приведенных моделей читатель должен дать ответы на следующие вопросы:

1. Какие переменные описывают стратегию (программу), т. е. являются управляемыми переменными?

2. Каков критерий выбора оптимальной стратегии (какова целевая функция)?

3. Как описывается, а затем исследуется задача в виде многошагового процесса?

4. Что характеризует состояние процесса на каждом шаге?

5. Как влияют ограничения на состояние процесса и допустимые значения управляемых переменных?

10.2. МОДЕЛЬ РАСПРЕДЕЛЕНИЯ УСИЛИЙ

Рассмотрим нижеследующий гипотетический, но тем не менее поучительный пример так называемой *модели распределения усилий*. Владелец торговой фирмы «Свежие продукты» должен распределить имеющийся у него недельный запас в количестве N яиц по s магазинам. Из опыта известно, что если направить y_j яиц в магазин j, то прибыль составит $R_j(y_j)$. Владелец фирмы считает, что для максимизации общей прибыли не следует направлять все яйца в один магазин и стремиться найти оптимальное распределение имеющихся в наличии яиц по магазинам.

Эту задачу можно сформулировать следующим образом:

$$\text{Максимизировать} \sum_{j=1}^{s} R_j(y_j) \tag{1}$$

при ограничениях

$$\sum_{j=1}^{s} y_j = N \quad \text{(наличный запас яиц)}, \tag{2}$$

$$y_j = 0,\ 1,\ 2,\ \ldots \quad \text{для любого } j \text{ (учитывая только целые яйца)}. \tag{3}$$

Для преобразования постановки задачи (1) — (3) в задачу динамического программирования примем:

$g_j(n)$ — прибыль при оптимальном распределении n яиц по магазинам $1, 2, \ldots, j$, (4)

$y_j(n)$ — количество яиц, направленных в магазин j и дающих прибыль $g_j(n)$.

Обратим внимание на принятые обозначения[1]. Буква g (goal) определяет значение целевой функции фирмы, т. е. прибыли. Буква n (number) относится к числу яиц, которые нужно распределить. Индекс j (just) обозначает просто номер магазина.

Отметим, что этот пример не относится к разряду динамических задач в смысле многошагового принятия решения. Многошаговое свойство задачи относится, скорее, к *методу* оптимизации, основанному на рассмотрении одного дополнительного магазина на каждом шаге.

Рекуррентное соотношение динамического программирования можно записать в виде

$$g_j(n) = \max_{y_j} [R_j(y_j) + g_{j-1}(n - y_j)], \quad j = 1,\ 2,\ \ldots,\ s, \tag{5}$$

$$g_0(n) \equiv 0, \quad j = 0, \tag{6}$$

где $n = 0,\ 1,\ \ldots,\ N$ и максимум берется только по неотрицательным целочисленным значениям y_j, удовлетворяющим условию $y_j \leqslant n$.

[1]) Эти обозначения соответствуют начальным буквам английских слов, приведенных в скобках.— *Прим. перев.*

При заданных числовых значениях функций прибыли $R_j(y_j)$ и известном значении N вычисления при решении задачи начинаются с определения значений $g_1(0)$, $g_1(1)$, $\ldots$, $g_1(N)$ при $j = 1$. Затем находятся значения $g_2(0)$, $g_2(1)$, $\ldots$, $g_2(N)$. Вычисления продолжаются по той же схеме при последовательном возрастании значений j, пока в конце концов не определяются значения $g_s(N)$. Оптимальное распределение в принципе находится путем осуществляемого в обратном порядке выбора значений y_j, дающих в совокупности $g_s(N)$.

Модель общего вида. Пример торговой фирмы «Свежие продукты» относится к частному случаю **задачи распределения усилий.** Этот пример имеет частный характер в силу того, что единственное ограничение (2) является линейным. Точно такой же динамический подход в той же мере справедлив и в случае, когда единственное ограничение нелинейно. Предположим для определенности, что модель описывается следующими соотношениями:

$$\text{Максимизировать } \sum_{j=1}^{s} R_j(y_j) \tag{7}$$

при ограничениях

$$\sum_{j=1}^{s} H_j(y_j) = N, \tag{8}$$

$$y_j = 0,\ 1,\ 2,\ \ldots \quad \text{при любом } j. \tag{9}$$

Допустим, что каждая функция $H_j(y_j)$ есть неубывающая функция, принимающая целочисленные значения при любом $y_j = 0, 1, 2, \ldots$ и удовлетворяющая условию $H_j(0) = 0$. Для упрощения рассуждений примем, что $H_1(y_1) = y_1$, вследствие чего допустимое решение существует при любом значении N. На каждую величину y_j можно также наложить ограничение сверху, как это будет показано в приводимом ниже числовом примере.

Рекуррентное соотношение динамического программирования, соответствующее задаче (7) — (9), имеет следующий вид:

$$g_j(n) = \max_{y_j} \{R_j(y_j) + g_{j-1}[n - H_j(y_j)]\}, \quad j = 1,\ 2,\ \ldots,\ s, \tag{10}$$

$$g_0(n) \equiv 0, \quad j = 0, \tag{11}$$

где $n = 0, 1, \ldots, N$, а максимум берется только по неотрицательным целочисленным значениям y_j, удовлетворяющим условию $H_j(y_j) \leqslant n$. Отыскивается значение $g_s(N)$. Для выполния вычислений нужно определить по выражению (10) значения каждой функции $g_j(n)$ при $n = 0, 1, \ldots, N$, начиная с $j = 1$ и кончая $j = s$.

Числовой пример. Проиллюстрируем вычислительную схему решения задачи о распределении усилий, задавшись числовыми данными для модели (7) — (9). Примем $s = 3$ и $N = 8$. Значения функций $R_j(y_j)$ и $H_j(y_j)$ приведены в таблице рис. 10.1. Заметим, что

$$\begin{aligned} &R_1(y_1) = 2y_1, \quad R_2(y_2) = 3y_2, \\ &H_1(y_1) = y_1, \quad H_2(y_2) = 2y_2, \quad H_3(y_3) = 3y_3. \end{aligned} \tag{12}$$

Как видно, это линейные функции, но функция $R_3(y_3)$ *нелинейна*. Кроме того, наложим ограничение сверху на все

$$y_j \leqslant 2 \text{ при любом } j. \tag{13}$$

(В этой задаче можно было бы также принять ограничение $y_3 \leqslant 1$, поскольку при $y_3 = 2$ значение целевой функции не возрастает.)

y_j	$R_1(y_1)$	$H_1(y_1)$	$R_2(y_2)$	$H_2(y_2)$	$R_3(y_3)$	$H_3(y_3)$
0	0	0	0	0	0	0
1	2	1	3	2	4,5	3
2	4	2	6	4	4,5	6

Рис. 10.1. Пример задачи о распределении усилий.

Значения $g_j(n)$ и соответствующие им оптимальные решения $y_j(n)$ сведены в таблицу рис. 10.2. Проанализируем, каким образом получены эти величины.

	$j=1$		$j=2$		$j=3$	
n	$y_1(n)$	$g_1(n)$	$y_2(n)$	$g_2(n)$	$y_3(n)$	$g_3(n)$
0	0	0	0	0	0	0
1	1	2	0	2	0	2
2	2	4	0	4	0	4
3			1	5	0	5
4			1	7	0	7
5			2	8	1	8,5
6			2	10	0	10
7					1	11,5
8					1	12,5

Рис. 10.2. Оптимальные стратегии для примера задачи о распределений усилий ($N = 8$).

Вычисления значений $y_1(n)$ и $g_1(n)$ тривиальны. Так, например, если $n = 2$, то из таблицы рис. 10.1 легко видеть, что $H_1(2) = 2$, откуда $y_1(2) = 2$ и соответственно $g_1(2) = R_1(2) = 4$.

Аналогично вычисления значений $y_2(n)$ и $g_2(n)$ при $n = 0, 1$ почти столь же тривиальны, поскольку при этих значениях n ограничение $H_2(y_2) \leqslant n$ может удовлетворяться только при $y_2 = 0$. Следовательно, имеем

$$g_2(0) = 0, \quad y_2(0) = 0, \tag{14}$$

$$g_2(1) = R_2(0) + g_1(1) = 0 + 2 = 2$$

и

$$y_2(1) = 0. \tag{15}$$

Более типичны по своему характеру вычисления значений $y_2(2)$ и $g_2(2)$, поскольку при $n = 2$ оба значения $y_2 = 2$ и $y_2 = 1$, как видно из таблицы рис. 10.1, являются допустимыми. Таким образом, из (10) получаем

$$g_2(2) = \max [R_2(0) + g_1(2),\ \ R_2(1) + g_1(0)] = \tag{16}$$
$$= \max (0 + 4,\ 3 + 0) = 4 \quad \text{и} \quad y_2(2) = 0.$$

Для проверки усвоения вычислительной схемы читателю рекомендуется самостоятельно определить значения остальных элементов таблицы рис. 10.2. (Отметим, что при увеличении n от 0 до 8 оптимальные значения y_3 колеблются.)

Оптимальная стратегия при $N = 8$ находится таким образом. Начиная с $j = 3$, имеем $y_3(8) = 1$. Следовательно, для определения оптимального значения y_2 нужно вычислить $n = 8 - 3y_3(8) = 5$. Проделав это, получаем $y_2(5) = 2$. Тогда новое значение $n = 5 - 2y_2(5) = 1$ и, следовательно, $y_1(1) = 1$. В результате оптимальное распределение определяется следующими значениями переменных: $y_1 = 1$, $y_2 = 2$, $y_3 = 1$. Соответствующее этому распределению оптимальное значение целевой функции $g_3(8) = 12{,}5$.

В частном случае, когда целевая функция (7) *и ограничение* (8) линейны, т. е.

$$R_j(y_j) = R_j y_j \quad \text{и} \quad H_j(y_j) = H_j y_j, \quad y = 1, 2, \ldots, s, \tag{I}$$

можно провести дальнейшие упрощения. Будем, как и прежде, предполагать, что $H_1 = 1$ и что при $j \neq 1$ H_j есть целое число, большее единицы. Вместо (10) и (11) можно применить следующее рекуррентное соотношение:

$$F(n) = \max_j [R_j + F(n - H_j)], \quad n = 1, 2, \ldots, N, \tag{II}$$
$$F(0) \equiv 0, \quad n = 0, \tag{III}$$

где максимизация производится по значениям $j = 1, 2, \ldots, s$, удовлетворяющим также неравенству $H_j \leqslant n$. Целью вычислительного процесса является определение значений величин $F(1)$, $F(2), \ldots, F(N-1)$ и, наконец, $F(N)$, представляющее собой оптимальное значение целевой функции. В данном случае понятия переменных состояния и управляемых переменных совпадают, так что индекс не требуется.

Сравнение (10) и (II) сразу показывает, что если в первом случае необходимо определять значения $g_j(n)$ при $n = 0, 1, \ldots, N$ и $j = 1, 2, \ldots, s$, то во втором нужно только найти $F(n)$ при $n = 0, 1, \ldots, N$. Таким образом, $g_s(n) \equiv F(n)$.

Различие между двумя подходами (10) и (II) наглядно выявляется при рассмотрении модели управления запасами с вогнутой функцией затрат из разд. 9.5. Предположим, что функции затрат и спроса не меняются во времени, т. е. примем, что они стационарны

$$D_t = D, \quad C_t(x_t) = C(x_t) \quad \text{и} \quad h_t(i_t) = h(i_t) \quad \text{для всех интервалов.} \tag{IV}$$

Тогда суммарные производственные затраты c_{kj} на отрезке $k+1$, необходимые для удовлетворения спроса на отрезках $k+1, \ldots, j$, зависят только от *длины* отрезка. Поэтому при $j-k=p$, примем $c_{kj} \equiv R_p = C(pD) + \sum_{m=1}^{p-1} h(mD)$. Для примера, если $N=4$, то

$$\begin{aligned} c_{kj} &= c_{01} = c_{12} = c_{23} = c_{34} \equiv R_1, && j-k=1,\\ c_{kj} &= c_{02} = c_{13} = c_{24} \equiv R_2, && j-k=2,\\ c_{kj} &= c_{33} = c_{14} \equiv R_3, && j-k=3,\\ c_{kj} &= c_{04} \equiv R_4, && j-k=4. \end{aligned} \tag{V}$$

Для этой ситуации из предположения о виде оптимальной стратегии следует, что количество продукции, выпущенной на каждом отрезке, является целым, кратным D, т. е. $1D, 2D, \ldots, ND$. Пусть y_j есть число отдельных включений объема выпуска jD в календарный план, где $j=1, 2, \ldots, N$. Тогда величины y_j нужно определить таким образом, чтобы суммарный выпуск $\sum_{j=1}^{N} (jD)\, y_j$ был равен суммарному спросу ND. Соответствующую оптимизационную модель для планового периода из N отрезков можно записать следующим образом:

$$\text{Минимизировать } \sum_{j=1}^{N} R_j y_j \tag{VI}$$

при ограничениях

$$\sum_{j=1}^{N} j \cdot y_j = N, \quad y_j = 0, 1, \ldots \quad \text{при всех } j, \tag{VII}$$

где обе части (VII) разделены на D. [Таким образом, сумма в левой части (VII) представляет собой просто общую продолжительность календарного плана, которая должна быть равна продолжительности планового периода.]

Заметим, что целевая функция (VI) обеспечивает минимизацию. Вместо сведéния (VI) к задаче максимизации путем изменения знаков R_j можно просто изменить смысл оптимизации в выражении (10). В постановке (VI) и (VII) задача сводится к отысканию числа партий y_j, которые нужно включить в календарный план для удовлетворения потребностей на j отрезках. В этом случае $s=N$, поскольку существует возможность запланировать выпуск количества продукции $N \cdot D$ на отрезке 1, обеспечивающего спрос на всех отрезках планового периода.

В задаче (VI) и (VII) можно привести (10) и (11) к виду

$$g_j(n) = \min_{y_j} [R_j y_j + g_{j-1}(n - j \cdot y_j)], \quad j = 1, 2, \ldots, N, \tag{VIII}$$

$$g_0(n) \equiv 0, \quad j = 0, \tag{IX}$$

где $n = 0, 1, \ldots, N$ и минимизация осуществляется только по неотрицательным целочисленным значениям y_j, удовлетворяющим условию $j \cdot y_j \leqslant n$.

При $n = 1, 2, \ldots, N$ рекуррентное соотношение (II) принимает вид

$$F(n) = \min_{j=1, \ldots, n} [R_j + F(n-j)] \quad \text{и} \quad F(0) \equiv 0. \tag{X}$$

Приняв $f_n \equiv F(n)$ и $k \equiv n - j$, получаем эквивалентное выражение для (X)

$$f_n = \min_{k=0, 1, \ldots, n-1} [f_k + R_{n-k}] \quad \text{и} \quad f_0 \equiv 0, \tag{XI}$$

не отличающееся от рекуррентного соотношения (4) в разд. 9.6 при условии, что выполняется допущение (V) о виде функции затрат.

Для сравнения относительных достоинств постановок (VIII) и (X) примем $N = 4$. Тогда рекуррентное соотношение (X) соответствует отысканию кратчайшего маршрута на ациклической сети,

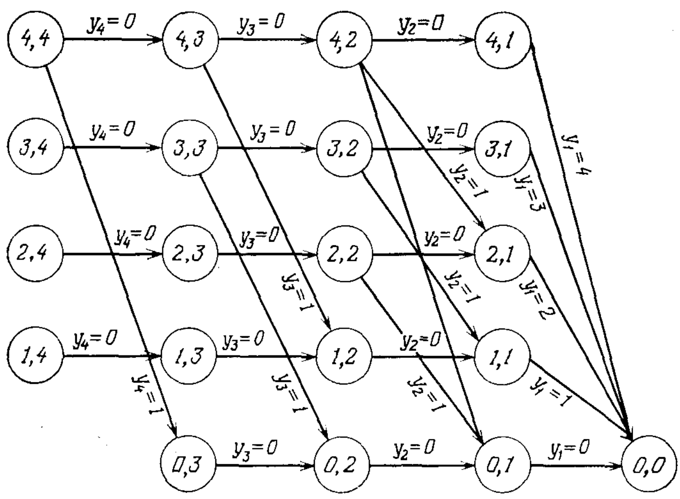


Рис. 10.3. Модель управления запасами с вогнутой функцией затрат

n, j — обозначение узла.

имеющей вид сети, показанной на рис. 9.7 (стр. 70). Рекуррентное соотношение (VIII) соответствует задаче о выборе кратчайшего маршрута на сети, приведенной на рис. 10.3. В этой сети индексы каждого узла (n, j) соответствуют паре значений n и j в выражении (VIII). Каждой дуге, выходящей из узла (n, j), сопоставляется значение y_j, такое, что $j \cdot y_j \leqslant n$. Очевидно, что для решения задачи в постановке (VIII) требуется больший объем вычислений, чем для решения задачи в постановке (X). [Оба рекуррентных соотношения допускают и дальнейшее упрощение, но и после упрощений постановка (X) остается выгодней в вычислительном отношении.]

10.3. РАСПРЕДЕЛЕНИЕ УСИЛИЙ. ДВА ОГРАНИЧЕНИЯ

Фирма «Вечный двигатель» планирует проведение радиорекламной кампании по сбыту своих автомобилей в одном из районов страны. В рекламных радиопередачах в течение двух недель будут рекламироваться новые модели фирмы. В зоне расположено s радиостанций. Для каждой j-й радиостанции отдел исследования сбыта фирмы определил *оценку* функции реакции сбыта, связывающей чистый доход $R_j(y_j)$ (объем продаж за вычетом расходов на рекламу) с выделенной суммой y_j долл. на оплату рекламных передач этой радиостанции. Общая сумма ассигнований на рекламную кампанию составляет N долл. Руководитель отдела рекламы фирмы определил, что число рекламных объявлений по радио в дневное время не должно превышать M. Радиостанция j включает в свою дневную программу $K_j(y_j)$ рекламных объявлений при условии, что фирма выплатит ей общую сумму y_j долл. за рекламные передачи.

Оптимизационную модель задачи, возникшей перед руководителем отдела рекламы, можно записать следующим образом:

$$\text{Максимизировать } \sum_{j=1}^{s} R_j(y_j) \tag{1}$$

при ограничениях

$$\sum_{j=1}^{s} y_j \leqslant N \quad \text{(общая сумма ассигнований)}, \tag{2}$$

$$\sum_{j=1}^{s} K_j(y_j) \leqslant M \quad \text{(число рекламных объявлений в дневное время)}, \tag{3}$$

$$y_j = 0,\ 1,\ 2,\ \ldots \text{ при любом } j. \tag{4}$$

Предположим, что N, M и $K_j(y_j)$ — неотрицательные целые числа.

Поскольку ранее для записи одного ограничения типа (2) в рекуррентном виде требовалась одномерная (скалярная) переменная состояния, теперь можно легко и безошибочно догадаться, что нужна двумерная (векторная) переменная состояния. Соответствующее рекуррентное соотношение имеет вид

$$g_j(n, m) = \max_{y_j} \{R_j(y_j) + g_{j-1}[n - y_j;\ m - K_j(y_j)]\},$$
$$j = 1,\ 2,\ \ldots,\ s, \tag{5}$$

где $n = 0, 1, \ldots, N$ и $m = 0, 1, \ldots, M$, а максимизация производится только по неотрицательным значениям y_j, удовлетворяющим условиям $y_j \leqslant n$ и $K_j(y_j) \leqslant m$. Читателю следует дать определение функции $g_j(n, m)$ в терминах радиорекламной кампании, а также дать ответы на следующие вопросы. В чем существенное различие между (5) и приведенным ранее в разд. 10.2 соотношением (10)? Какой вид принимает рекуррентное соотношение, если добавить к (2) и (3) третье ограничение в виде суммы?

Другим примером только что рассмотренной модели может служить распределение капиталовложений. Ежегодно фирма «Большая энергетика», выпускающая тяжелое энергетическое оборудование, рассматривает значительное число s независимых вариантов капиталовложений в различные виды основных фондов. Для осуществления варианта j требуется выделить K_j долл. Общий объем годовых капиталовложений M строго фиксирован. Ожидаемый доход от капиталовложений по j-му варианту составляет R_j долл. Фирма, естественно, стремится максимизировать общий доход по всем капиталовложениям. Руководители фирмы пришли к выводу, что вследствие ограниченных возможностей организационного управления общее число проектов, реализуемых в рамках любого года, не должно превышать N. Каждый вариант уникален, и нужно решить — принимать ли его на данный год или отвергнуть. Для отображения этого условия примем, что переменная y_j, сопоставляемая j-му варианту, может принимать только значения: $y_j = 0$ (вариант отвергнут) или $y_j = 1$ (вариант принят). Все перечисленные условия можно формализовать в виде следующей задачи:

$$\text{Максимизировать } \sum_{j=1}^{s} R_j y_j \tag{I}$$

при ограничениях

$$\sum_{j=1}^{s} y_j \leqslant N \quad \text{(общее число проектов)}, \tag{II}$$

$$\sum_{j=1}^{s} K_j y_j \leqslant M \quad \text{(общий объем капиталовложений)}, \tag{III}$$

$$y_j = \begin{cases} 1 \\ 0 \end{cases} \text{при любом } j \quad \text{(вариант принят/отвергнут)}. \tag{IV}$$

В этой задаче

$g_j(n, m)$ есть доход от оптимального распределения m долл. капиталовложений по n проектам, выбранным из варианта 1, варианта 2, . . ., варианта j. (V)

10.4. РАСПРЕДЕЛЕНИЕ УСИЛИЙ. ПОГРУЖЕННАЯ ЗАДАЧА

Фирма «Чудесные химикалии», организационная структура которой построена на базе отделений, ежегодно распределяет ассигнования на выполнение научно-исследовательских и опытно-конструкторских работ (НИР — ОКР). Каждое из s отделений представляет руководству фирмы данные трех видов. Информация первой группы относится к проведению поисковых исследований весьма неопределенного характера. Если на исследования такого рода в отделении j выделяется v_j тыс. долл., то оценка ожидаемого долгосрочного дохода равна $P_j(v_j)$ млн. долл. Информация второй группы относится к продукции, по которой поисковые исследования уже завершены и для выпуска которой требуется проведение опытно-конструкторских работ и испытаний. Для таких проектов ассигнования в объеме w_j тыс. долл., согласно имеющейся оценке, дадут в конечном счете доход в размере $Q_j(w_j)$ млн. долл. К третьей группе относится информация, связанная с улучшением качества уже выпускаемой в промышленных масштабах продукции. Затраты x_j тыс. долл., согласно сделанным оценкам, должны принести всего $R_j(x_j)$ млн. долл. дополнительного дохода.

Совет директоров фирмы утверждает общую сумму ассигнований на все проекты НИР—ОКР в размере N тыс. долл., а президент фирмы устанавливает верхний предел L_j ассигнований на НИР—ОКР, которые может получить отделение j. Вице-президент фирмы, ответственный за НИР—ОКР, должен распределить ассигнования между отделениями таким образом, чтобы обеспечивалась максимизация общего дохода фирмы при наложенных ограничениях.

Математическая модель задачи вице-президента описывается следующими соотношениями:

$$\text{Максимизировать } \sum_{j=1}^{s} [P_j(v_j) + Q_j(w_j) + R_j(x_j)] \tag{1}$$

при ограничениях

$$\sum_{j=1}^{s} (v_j + w_j + x_j) \leqslant N \quad \text{(общая сумма ассигнований фирмы на НИР—ОКР)}, \tag{2}$$

$$v_j + w_j + x_j \leqslant L_j, \quad j = 1, 2, \ldots, s \quad \text{(верхний предел ассигнований, выделенных отделению фирмы)}, \tag{3}$$

$$v_j, w_j, x_j \quad \text{(неотрицательные целые при любом } j). \tag{4}$$

Поскольку на все управляемые переменные наложено только одно ограничение (2), а остальные бюджетные и целочисленные ограничения (3) и (4) относятся только к отделению j, можно ожидать, что в данном случае имеет место задача распределения усилий с одним ограничением. Рассуждая таким образом, получаем следующее рекуррентное соотношение:

$$g_j(n) = \max [P_j(v_j) + Q_j(w_j) + R_j(x_j) + g_{j-1}(n - v_j - w_j - x_j)],$$
$$j = 1, 2, \ldots, s, \tag{5}$$

где $n = 0, 1, \ldots, N$ и максимизация производится только по неотрицательным целочисленным значениям v_j, w_j и x_j, удовлетворяющим условию $v_j + w_j + x_j \leqslant \min(L_j, n)$.

Заметим, что в (5) появляется новое для нас затруднение. Если в предыдущих примерах задач о распределении усилий для поиска максимума на каждом шаге нужно было изменять только одну переменную y_j, то теперь приходится решать оптимизационную задачу с тремя переменными. Разумеется, можно было бы попытаться решить задачу методом полного перебора, однако в этом нет необходимости. В рассматриваемом частном примере *на каждом* шаге отыскания максимума можно использовать метод решения задачи о распределении усилий, представив этот пример в следующем виде:

$$\text{Максимизировать } P_j(v_j) + Q_j(w_j) + R_j(x_j) \tag{6}$$

при ограничениях

$$v_j + w_j + x_j \leqslant y, \tag{7}$$

где v_j, w_j и x_j должны быть неотрицательными целыми числами. Необходимо получить решение для каждого значения $y = 0, 1, \ldots$ $\ldots, L_j$.

Чтобы использовать рекуррентный подход к задаче (6) — (7), примем

$$p_j(y) = P_y(y_j), \quad y = 0, 1, \ldots, L_j, \tag{8}$$

$$q_j(y) = \max_{w_i} [Q_j(w_j) + p_j(y - w_j)], \quad y = 0, 1, \ldots, L_j, \tag{9}$$

где максимизация производится только по неотрицательным целым значениям $w_j \leqslant y$, и

$$r_j(y) = \max_{x_j} [R_j(x_j) + q_j(y - x_j)], \quad y = 0, 1, \ldots, L_j, \tag{10}$$

где максимизация осуществляется только по неотрицательным целым значениям $x_j \leqslant y$.

В итоге для каждого j используется $p_j(y)$ из (8) и (9) для отыскания $q_j(y)$, а функции $r_j(y)$ при $y = 0, 1, \ldots, L_j$ определяют по функциям $q_j(y)$ и (10). Далее преобразуется соотношение (5) и находится решение по соотношению

$$g_j(n) = \max_{y} [r_j(y) + g_{j-1}(n - y)], \quad j = 1, 2, \ldots, s, \tag{11}$$

где $n = 0, 1, \ldots, N$ и максимизация осуществляется только по неотрицательным целочисленным значениям y, удовлетворяющим условию $y \leqslant \min(L_j, n)$. Следовательно, для решения этой задачи нужно связать s расчетов распределения усилий с общей моделью распределения усилий.

10.5. ЦЕЛОЧИСЛЕННОЕ ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Приведенные примеры моделей распределения усилий непосредственно обобщаются следующей моделью:

$$\text{Максимизировать } \sum_{j=1}^{s} c_j y_j \tag{1}$$

при ограничениях

$$\sum_{j=1}^{s} a_{ij} y_j = b_i, \quad i = 1, 2, \ldots, k, \tag{2}$$

$$y_j = 0, 1, 2, \ldots \text{ при любом } j, \tag{3}$$

где каждый коэффициент a_{ij} и свободный член b_i являются неотрицательными целыми числами. Рекуррентное соотношение, которое можно применить для этой модели, имеет вид

$$F(n_1, n_2, \ldots, n_k) = \max_j [c_j + F(n_1 - a_{1j}, n_2 - a_{2j}, \ldots, n_k - a_{kj})], \tag{4}$$

где для каждого $i = 1, 2, \ldots, k$ рассматривается $n_i = 0, 1, \ldots$ $\ldots, b_i$ и максимизация производится по каждому значению $j = 1, 2, \ldots, s$, удовлетворяющему условию $a_{ij} \leqslant n_i$ при $i = 1, 2, \ldots, k$.

Этот подход к решению задач ЦЛП, разумеется, совершенно не пригоден с вычислительной точки зрения, когда речь идет о задачах реальной размерности, так как нужно определять значение функции (4) при каждом допустимом состоянии $(n_1, n_2, \ldots, n_k)$. В качестве примера приведем случай, когда каждый коэффициент a_{ij} может принимать только два значения — 0 или 1, а любой свободный член $b_i = 1$. При этих условиях каждая величина n_i может принимать также только два значения — 0 или 1. Следовательно, перебор $(n_1, n_2, \ldots, n_k)$ при каждом n_i, равном 1 или 0, требует просмотра в общей сложности 2^k вариантов. Если $k = 100$, что характеризует задачи линейного программирования не очень большой размерности, то существует более 10^{30} возможных вариантов наборов $(n_1, n_2, \ldots$ $\ldots, n_k)$, для оценки которых требуется объем вычислений, намного превышающий возможности любой ЭВМ. Поэтому многошаговый метод динамического программирования практически нереализуем, если только значение k не является очень небольшим. Рассмотрению моделей ЦЛП будет целиком посвящена гл. 13 книги.

10.6. МОДЕЛЬ ЗАМЕНЫ ОБОРУДОВАНИЯ

В разд. 6.5 на стр. 230 было показано, как можно использовать ациклическую сеть для описания задачи определения стратегии замены оборудования, минимизирующей затраты в рамках планового периода из $(N - 1)$ отрезков. Читателю следует повторить этот материал, прежде чем приступить к проработке данного раздела.

В разд. 6.5 величина c_{ij} представляла собой полные эксплуатационные затраты, начиная с отрезка i и кончая отрезком j, на котором это оборудование заменяется. Хотя пример разд. 6.5 относится к случаю, когда оборудование берется в аренду, модель полностью пригодна и в ситуации, когда фирма приобретает оборудование. В последнем случае величина c_{ij} представляет собой сумму покупной цены и эксплуатационных затрат нового оборудования, за вычетом остаточной стоимости этого оборудования на начало отрезка j.

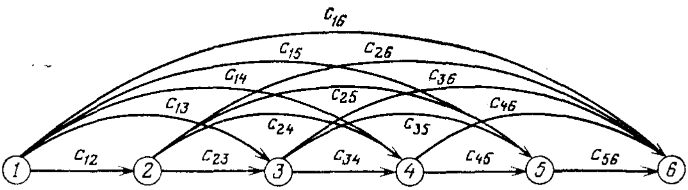


Рис. 10.4. Сеть замены оборудования.

Задача относится к категории задач о восстановлении в том смысле, что определение стратегии замены равносильно составлению списка отрезков, в которых производится закупка нового оборудования. Примем следующее обозначение:

f_n — стратегия замены, минимизирующая затраты на отрезках n, $n+1$, ..., $N-1$, в предположении, что новое оборудование приобретется в начале отрезка n. (1)

Тогда отыскивается значение f_1 и рекуррентное соотношение, соответствующее этому условию, имеет вид

$$f_n = \min_{k=n+1,\ \ldots,\ N} [c_{nk} + f_k], \quad n = N-1,\ N-2,\ \ldots,\ 1, \tag{2}$$

$$f_N \equiv 0. \tag{3}$$

Читателю полезно описать словами выражение (2). Соответствующая ациклическая сеть при $N-1=5$ приведена на рис. 10.4.

К этой задаче можно подойти с несколько иной точки зрения. Предположим, что затраты, отвечающие некоторой стратегии замены, включают две составляющие:

p_{in} — чистая стоимость замены оборудования возраста i на отрезке n новым оборудованием при условии продажи на этом отрезке старого оборудования по его остаточной стоимости; (I)

k_{nt} — стоимость эксплуатации оборудования на отрезке n, возраст которого в конце этого отрезка равен t.

Примем, что

$f_n(i)$ — стратегия, минимизирующая затраты на отрезках n, $n+1$, ..., $N-1$, при условии, что в начале отрезка n возраст оборудования равен i годам. (II)

Отметим, что смысл, придаваемый индексу в (II), в данном примере отличается от смысла, который придавался ему в предыдущих примерах. В самом деле, здесь n относится к отрезку n и всем последующим отрезкам до конца планового периода. (В примерах задач

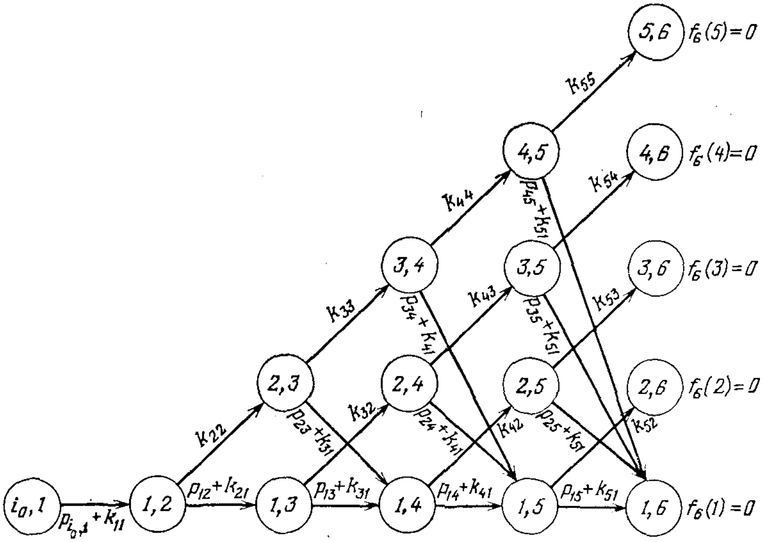


Рис. 10.5. Модель замены оборудования.
i, n — обозначение узла (i — возраст машины; n — номер отрезка)

управления запасами индекс n относится к шагу, на котором до конца планового периода оставалось n отрезков.)

Если оптимальное решение сводится к сохранению оборудования и попадает на отрезок n, то

$$f_n(i) = k_{n\ i+1} + f_{n+1}(i+1), \tag{III}$$

но если оптимальное решение сводится к его замене, то

$$f_n(i) = p_{in} + k_{n1} + f_{n+1}(1). \tag{IV}$$

Таким образом, имеем

$$f_n(i) = \min\,[k_{n\ i+1} + f_{n+1}(i+1),\quad p_{in} + k_{n1} + f_{n+1}(1)],$$
$$n = 1, 2, \ldots, N-1, \tag{V}$$

где

$$f_N(i) \equiv 0. \tag{VI}$$

Мы ищем значение $f_1(i_0)$, где i_0 — возраст оборудования на начало планового периода. Если в это время рассматриваемая единица оборудования отсутствует, то величина $p_{i_0 1}$ представляет собой просто покупную цену нового оборудования и решение *сохранить* оборудование при $n = 1$ не имеет смысла.

Рис. 10.5 представляет собой графическое изображение соотношения (V). В этом примере принято, что $N - 1 = 5$ и что при $n = 1$ оборудование отсутствует. Узлы помечены индексами (i, n), соответствующими индексам, принятым в формуле (V). Читателю следует ответить на вопрос: какие дуги отображают решения о закупке оборудования, и сравнить относительные достоинства (2) и (V), используя рис. 10.4 и 10.5 для обоснования своего мнения.

10.7. СТРУКТУРА МНОГОШАГОВОГО АНАЛИЗА

Рассмотрев различные примеры применения метода динамического программирования, часть которых была тщательно проанализирована, целесообразно дать общую оценку отличительных особенностей этого метода. В заключительных разделах этой главы приводится общая характеристика моделей динамического программирования.

Оптимизационные задачи с ограничениями, охватываемые динамическим программированием, по существу содержательно не отличаются от задач, изложенных до гл. 8 и изучаемых в дальнейшем. Лишь подход к анализу модели, несомненно, является совершенно новым. В основе метода динамического программирования, используемого для решения оптимизационных задач с многими ограничениями и большим числом переменных, лежит разбиение задачи на последовательность шагов, на каждом из которых решается оптимизационная задача меньшей размерности. В отличие от этого метода в большинстве методов линейного и нелинейного программирования предпринимаются попытки решения таких задач путем одновременного учета всех ограничений.

В методе динамического программирования задача сводится к следующему виду:

I) Управляемые переменные и соответствующие ограничения группируются по шагам, и многошаговый процесс принятия решений исследуется в определенной последовательности.

II) Единственная информация о предыдущих шагах, которая используется для выбора оптимальных значений переменных на рассматриваемом шаге, заключена в так называемой переменной состояния, представляющей собой иногда n-мерный вектор.

III) Рассматриваемое решение, принимаемое при заданном текущем состоянии системы, оказывает прогнозируемое влияние на состояние системы на последующем шаге.

IV) Оптимальность текущего решения оценивается в терминах прогнозируемого экономического эффекта для рассматриваемого шага и всех последующих шагов.

В отличие от модели линейного программирования, относящейся к специальному классу математических моделей, которые позволяют получить решение различными методами (среди этих методов самым универсальным является симплексный), динамическое программирование представляет собой особый аналитический метод, применимый к самым различным математическим моделям. Существует каноническая форма записи задачи в терминах динамического программирования, наглядно иллюстрирующая перечисленные в п. I) — IV) структурные свойства.

Обозначим символом s некоторое состояние системы и символом S_n — множество всех возможных состояний на шаге n. Далее примем, что d_n обозначает решение, выбранное на шаге n, а $D_n(s)$ — все *допустимые* значения d_n при условии, что система находится в состоянии s. Наконец, пусть функция $R_n(s, d_n)$ обозначает непосредственный экономический эффект, достигаемый в результате выбора решения d_n при заданном состоянии системы s, а $T_n(s, d_n)$ — преобразованное состояние системы на шаге $n - 1$. Тогда рекуррентное соотношение динамического программирования (РДП) запишется в следующем виде:

$$f_n(s) = \underset{d_n \text{ в } D_n(s)}{\text{optimum}} \{R_n(s_1 d_n) + f_{n-1}[T_n(s_1 d_n)]\} \quad \text{при любом } s \text{ в } S_n,$$

где «optimum» означает максимум или минимум в зависимости от постановки задачи. Все рекуррентные формулы в гл. 8—10 имеют такой вид. Читателю следует рассмотреть все перечисленные выше п. I) — IV) и показать, какое влияние оказывает каждое из указанных в них свойств на вид РДП. Свойство (II) иногда называют *марковским*. Укажите, какой из пунктов содержит *принцип оптимальности*, описанный в гл. 8: оптимальная стратегия (программа) должна обладать таким свойством, что независимо от того, каким образом система оказалась в рассматриваемом конкретном состоянии, последующие решения должны составлять оптимальную стратегию, привязывающуюся к этому состоянию.

Нужно учитывать важную особенность РДП. Для принятия оптимального решения на шаге n не требуется значения *управляемых* переменных, обусловливающих эффект $f_{n-1}[T_n(s, d_n)]$. Такой характер рекуррентных соотношений означает, что на каждом шаге оптимизация проводится лишь по отношению к ограниченному числу управляемых переменных и в последующих вычислениях используются только значения *целевой функции*. Именно эта простота вычислительной схемы обусловливает привлекательность метода динамического программирования. Если пользоваться понятиями линейного программирования, то вычисления, выполняемые при использовании метода динамического программирования, можно рассматривать как последовательное определение значений двойственных переменных. Эта аналогия для примеров, изложенных в данной главе, является совершенно точной в силу разъясняемых ниже причин.

Во всех приведенных примерах управляемые переменные, а также переменные состояния и шага принимали только целочисленные значения. (Задачи такого рода называют задачами **дискретного программирования.**) Кроме того, процесс принятия решений относился к конечному промежутку времени или конечному числу шагов. Наконец, на результаты и переходы из одного состояния в другое не оказывали никакого влияния случайные факторы.

При всех этих допущениях оказалось возможным анализировать РДП в терминах задачи отыскания кратчайшего маршрута на ациклической сети, изложенной в гл. 7. Для построения такой сети каждому возможному значению s и n в рекуррентном соотношении сопоставляется узел с параметрами $[s, n]$. Каждому значению d_n в $D_n(s)$ ставится в соответствие дуга из узла $[s, n]$ в узел $[T_n(s, d_n), n-1]$ с весом $R_n(s, n)$. Состояние s системы на последнем шаге определено (например, «конечный запас должен быть равен нулю»). Поэтому этот узел рассматривается в качестве конечного пункта сети, а функция $f_n(s)$ определяет длину кратчайшего (оптимального) маршрута из узла $[s, n]$ в конечный узел.

При описании сетевой задачи в терминах линейного программирования функция $f_n(s)$ представляет собой значение двойственной переменной в уравнении сохранения потока для узла $[s, n]$. Разумеется, для решения РДП никогда не возникает необходимость графического *изображения* ациклической сети, однако сетевая аналогия позволяет сделать два важных вывода. Во-первых, анализируя сложность соответствующих ациклических сетей, можно оценить эффективность вычислительной схемы различных динамических постановок задач. Хотя и приближенно, но обычно достаточно надежно можно утверждать, что постановка, которой соответствует сеть с минимальным числом дуг, является в этом отношении предпочтительной. (Использование этого признака вполне оправдано, ибо при некоторых постановках требуется выполнение значительного объема вспомогательных вычислений для определения цены прохождения дуги, *прежде чем* применить алгоритм отыскания оптимального маршрута.)

Во-вторых, если модель уже описана каким-либо конкретным рекуррентным соотношением, то для получения решения не нужно находить нового метода. Правда, в задачах динамического программирования, не обладающих свойством целочисленности и не являющихся детермированными, а также не относящихся к конечному промежутку времени, возникают дополнительные затруднения, которые изучаются в гл. 11, 12, 17 и 18.

10.8. СУЩНОСТЬ ДИНАМИЧЕСКИХ ПРОЦЕССОВ

По самой своей природе метод динамического программирования приводит к идее исследования чувствительности решения по отношению к продолжительности планового периода. Рассмотрение различных моделей управления запасами, приведенных в гл. 8 и 9, показывает, что оптимальная стратегия может существенно зависеть от продолжительности планового периода. Анализ этой связи приводит к нескольким важным выводам.

Прежде всего уточняется понятие стратегии. Отыскивая решение рекуррентного уравнения динамического программирования, по существу определяют набор оптимальных решений при каждом допустимом значении переменной состояния на каждом шаге. Таким образом, получают предписание о выборе линии поведения в любой возможной ситуации. Рассмотрим модель календарного планирования производства, чтобы выяснить, к каким последствиям приводит этот результат. Предположим, что выбрана линия поведения, основанная на *допущении*, что функции спроса и затрат не меняются во времени и что плановый период всегда содержит N отрезков, относящихся к последующему временно́му интервалу. Тогда требуется определение оптимального календарного плана для каждого состояния s при условии, что до конца планового периода осталось N отрезков. На каждом отрезке определяется состояние s и выбирается соответствующий календарный план. Существенный вывод здесь заключается в том, что при рассмотрении оставшихся n отрезков единственный фактор, который нужно принимать во внимание,— это текущее состояние системы s, а не предшествующие шаги, которые привели к этому состоянию. Поэтому если принятое решение не приводит к желаемым результатам, то при принятии следующего решения можно «махнуть рукой на прошлое». В гл. 11 и 12 понятие стратегии будет играть важнейшую роль. Кроме того, в этих главах изучается проблема отыскания оптимальной стратегии при условиях, когда допущение о неограниченном плановом периоде непосредственно влияет на выбор аппарата решения задачи.

Еще одним существенным понятием является определение *формы* оптимальной стратегии. При удачном представлении динамического процесса в виде последовательности шагов иногда удается использовать особые свойства ограничений и целевых функций для определения оптимальной формы стратегии. Одним из примеров такого процесса может служить пример замены оборудования, рассмотренный в разд. 10.6. В любой точке восстановления можно разбить промежуток времени на две части: «предшествующую» и «последующую» и рассматривать каждую часть независимо друг от друга.

Два таких понятия, как определение стратегии и форма оптимальной стратегии, тесно связаны с третьим. Теорема о продолжительности планового периода в случае, когда ее можно применить, иногда позволяет получить необходимые данные о влиянии динамических свойств процесса на оптимальную стратегию.

В разд. 9.4 и 9.7 приведены два примера, показывающие, как конкретный вид целевой функции позволяет обнаружить свойства планового периода. В случае выпуклой функции затрат было показано, что единственным следствием увеличения продолжительности планового периода является увеличение значений управляемых переменных. В случае вогнутой функции затрат можно установить длину планового периода, соответствующего оптимальной стратегии, не обращаясь к периоду, предшествующему точке восстановления.

10.9. ВЫЧИСЛИТЕЛЬНЫЕ ВОЗМОЖНОСТИ МЕТОДА ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

При решении задач, изложенных в гл. 8—10, использование метода динамического программирования позволяет достичь весьма существенной экономии по сравнению с методом полного перебора. Эффективность динамического программирования определяется прежде всего размерностью переменной состояния, что является одной из причин, требующих творческого подхода к постановке задач. Неудачная или удачная постановка может решающим образом повлиять на работоспособность вычислительной схемы динамического программирования.

Как показывают приведенные в этой главе примеры, метод динамического программирования можно успешно применять для решения нелинейных целочисленных задач небольшой размерности. Поскольку общей канонической формы представления таких задач (даже для случая дискретных детерминированных моделей с ограниченным плановым периодом) не найдено, не существует общей *эффективной* программы решения на ЭВМ *всех* задач динамического программирования. Тем не менее вид рекуррентного соотношения динамического программирования обычно достаточно прост, в связи с чем составление программы решения конкретной задачи на ЭВМ не вызывает особых затруднений. Правда, при разработке таких программ приходится преодолевать ряд технических трудностей, но они относятся к специфическим особенностям программирования для ЭВМ и поэтому здесь не рассматриваются. Достаточно отметить, что при программировании большинства практических задач, в которых размерность переменной состояния не превышает двух, не возникает непреодолимых препятствий. Если не считать редких случаев (например, модель управления запасами с выпуклой функцией затрат, приведенная в разд. 9.3, алгоритм решения которой чрезвычайно прост), то для отыскания решения задачи динамического программирования требуется применение ЭВМ. При отсутствии более веских причин использование ЭВМ целесообразно уже потому, что освобождает человека от необходимости выполнения однообразных и утомительных вычислений.

В гл. 8—9 вопросы сходимости алгоритмов не возникали, поскольку любая из рассмотренных моделей в сущности сводилась к задаче о выборе кратчайшего маршрута. Однако эти вопросы появятся в двух последующих главах, где изучаются модели с бесконечным плановым периодом и используются методы решения, основанные на последовательных приближениях.

В связи с разнообразием моделей, для решения которых используется метод динамического программирования, не удается найти какого-либо универсального метода анализа чувствительности. Большинство факторов, учитываемых при анализе чувствительности моделей динамического программирования, оказывают влияние на несколько ограничений или параметров целевой функции одновременно. Поэтому для каждой конкретной задачи приходится выводить особые формулы анализа чувствительности.

10.10. ОБЛАСТЬ ПРИМЕНЕНИЯ ДИНАМИЧЕСКОГО ПРОГРАММИРОВАНИЯ

Материал, содержащийся в гл. 8—10, можно разбить на две части. Одна из них относится к углублению знаний читателя относительно явлений, встречающихся в динамических оптимизационных моделях. Другая описывает примеры применения метода динамического программирования для решения важных производственных задач. Наиболее широко этот метод используется для решения задач пополнения запасов, календарного планирования и задач восстановления, включающих замену оборудования. Поскольку в ряде задач динамического программирования фигурируют случайные факторы и рассматривается неограниченный плановый период, приведенные примеры следует считать лишь иллюстрацией к введению в этот метод. В последующих главах содержится еще много примеров иного рода, отвечающих более общей трактовке метода.

В этих заключительных замечаниях стоит повторить соображения, приведенные в начале главы. Практические ограничения, обусловленные размерностью переменной состояния, сужают класс задач, которые удается решать методом динамического программирования. Лишь в довольно редких случаях все существенные факторы экономической задачи достаточно большой размерности удается преобразовать таким образом, что численное решение задачи, сформулированной в терминах динамического программирования, оказывается практически возможным. Это обстоятельство не снижает ценности метода, оно лишь позволяет провести границу между областями практического приложения динамического и линейного программирования.

В частности, многие реальные примеры применения динамического программирования относятся к принятию оперативных решений: когда заказывать изделия или материалы для пополнения запасов, когда налаживать оборудование, когда заменять оборудование и т. п. Большинство таких решений, как правило, относится к компетенции низших уровней руководства. Поэтому практическое использование моделей динамического программирования связано либо с применением ЭВМ с целью принятия оперативных решений в реальном масштабе времени, либо с обеспечением руководителей соответствующих уровней правилами или таблицами, основанными на результатах вычисления оптимальной стратегии. Основными условиями успешного применения метода динамического программирования являются ограниченная сфера влияния отдельного решения, найденного с помощью этого метода, и простота отыскания такого решения. Метод динамического программирования позволяет освободить руководство высших уровней от обязанности оценки огромного количества повседневных решений, каждое из которых не оказывает принципиального влияния на деятельность организации. Вместе с тем этот метод дает возможность руководителю высшего звена эффективно осуществлять стратегическое руководство, так как он сам выбирает модель, учитывающую наиболее важные, по его мнению, экономические и технические факторы, предоставляя право использования этой модели руководителю нижестоящего уровня.

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

1. Предположим, что ограничение (2) в разд. 10.2 распространено на линейное ограничение $\sum_{j=1}^{s} H_j y_j = N$ (где $H_1 = 1$). Измените (5) соответствующим образом и укажите наибольшее значение y_j, которое нужно учитывать при отыскании максимума на шаге j.

2. Предположим, что известно решение задачи (1) — (3) разд. 10.2 при заданном значении N и нужно найти решение при значении $N + 1$. Укажите, какие дополнительные вычисления необходимы, предполагая, что для получения ответа используется рекуррентное соотношение (5).

3. Задача торговой фирмы «Свежие продукты» (разд. 10.2). Предположим, что при отправке яиц в магазин некоторая их доля f_j $(0 < f_j < 1)$ разбивается при перевозке. Измените соответствующим образом рекуррентное соотношение (5).

4. Задача торговой фирмы «Свежие продукты» (разд. 10.2). Предположим, что, пользуясь накопленным опытом, владелец фирмы установил, что функция прибыли $R_j(y_j)$ возрастает при возрастании y_j до определенного значения, а затем убывает при больших значениях y_j (слишком большой запас яиц в магазине приводит к возникновению значительных затрат, связанных с их хранением и сортировкой). Укажите, как следует изменить постановку задачи и получающееся в результате ее решения рекуррентное соотношение динамического программирования, чтобы во все магазины допускалась отправка менее N яиц (если такое изменение вообще необходимо).

5. При описании задачи распределения усилий в разд. 10.2 для простоты предполагалось, что каждая функция $H_j(y_j)$ является неубывающей, целочисленной при $y_j = 0, 1, 2, \ldots$ и удовлетворяет условию $H_j(0) = 0$. Кроме того, было принято допущение $H_1(y_1) = y_1$. Объясните, какие усложнения возникнут, если отказаться от каждого из этих четырех допущений, и как нужно изменить рекуррентную формулу (10) при условии, что в этом действительно имеется необходимость.

Упражнения 6—12 относятся к численным примерам (12) и (13) разд. 10.2 и к решению, приведенному в таблице рис. 10.2.

6. а) Проверьте результаты вычислений, приведенные в таблице рис. 10.2 при $j = 2$ и $n = 3, 4, 5, 6$, а также при $j = 3$ и $n = 0, 1, \ldots, 8$.

б) Укажите все оптимальные решения при $N = 4, 5, 6, 7$.

7. Предположим, что ограничения сверху изменены следующим образом: $y_1 \leqslant 3$, $y_2 \leqslant 3$, $y_3 \leqslant 2$. Найдите все оптимальные решения при: а) $N = 4, 5, 6, 7, 8$; б) $N = 9, 10$.

8. Для упрощения изложения материала трем магазинам были произвольно присвоены индексы (номера) 1, 2, 3 и к принятой нумерации применялось рекуррентное соотношение (10). Если магазины перенумеровать, то вычисления будут выполняться в ином порядке, но окончательные результаты не изменятся. Пусть первые два столбца таблицы рис. 10.1 относятся к магазину А, следующие два — к магазину В и последние два — к магазину С. Рекуррентное соотношение (10) и результаты, приведенные в таблице рис. 10.2, соответствуют следующим условиям: $j = 1$ относится к магазину А, $j = 2$ — к магазину В и $j = 3$ — к магазину С. Примените формулу (10) при следующих комбинациях номеров, присваиваемых магазинам:

а) $j = 1$ — магазин А, $j = 2$ — магазин С, $j = 3$ — магазин В.
б) $j = 1$ — магазин С, $j = 2$ — магазин В, $j = 3$ — магазин А.
в) $j = 1$ — магазин В, $j = 2$ — магазин А, $j = 3$ — магазин С.
г) $j = 1$ — магазин С, $j = 2$ — магазин А, $j = 3$ — магазин В.
д) $j = 1$ — магазин В, $j = 2$ — магазин С, $j = 3$ — магазин А.

9. Предположим, что функция прибыли для $j = 3$ равна $R_3(y_3) = 5y_3$. Используйте рекуррентное соотношение (10) для определения всех оптимальных решений при следующих условиях:

а) $N = 4, 5, 6, 7, 8$; б) $N = 9, 10$.

в) Дайте ответы на вопросы, касающиеся п. а) и б) в случае, когда все ограничения сверху на y_j сняты.

10. Решите задачу п. в) упражнения 9, используя рекуррентное соотношение (II), приведенное в конце разд. 10.2. Сравните объем вычислений, требующийся при использовании формулы (10), с объемом вычислений при использовании формулы (II).

11. По каждому из указанных ниже пунктов запишите задачу определения оптимального размера партии в виде выражений (VI) и (VII), приведенных в конце разд. 10.2. Укажите соответствующее значение каждой величины R_j. Решите задачу, пользуясь рекуррентным соотношением (VIII). Решите также эту задачу, пользуясь формулой (X), и укажите соответствие между формулами (X) и (XI). Используйте числовые данные следующих упражнений гл. 9.

а) Упражнение 13; б) Упражнение 15; в) Упражнение 43, п. е); г) Упражнение 14; д) Упражнение 43, п. а).

12. Предположим, что в примерах (12) — (13) разд. 10.2 введено дополнительное ограничение $y_1 + 3y_2 + y_3 \leqslant 4$, а ограничение (8)

записано в виде неравенства (5). Найдите все оптимальные решения при $N = 4, 5, 6, 7, 8$.

13. При рассмотрении задачи распределения усилий в разд. 10.3 для упрощения рассуждений предполагалось, что ограничения (2) и (3) являются неравенствами и что каждая функция $K_j(y_j)$ неотрицательная и целая. Объясните, какие усложнения возникнут, если снять все эти допущения, и как это повлияет на вычисление решения по рекуррентному соотношению (5).

14. а) Предположим, что ограничение (2) в разд. 10.3 заменено ограничением $\sum_{j=1}^{s} H_j(y_j) \leqslant N$, где значения $H_j(y_j)$ — неотрицательные целые при $y_j = 0, 1, \ldots$. Укажите, каким образом нужно изменить формулу (5) и вычислительную процедуру.

б) Предположим, что в задаче (1) — (4) разд. 10.3 введено дополнительное ограничение $\sum_{j=1}^{s} H_j(y_j) \leqslant P$, где значения $H_j(y_j)$ — неотрицательные целые при $y_j = 0, 1, \ldots$, а P — положительное целое число. Укажите, как следует изменить рекуррентное соотношение (5), и рассмотрите, как увеличивается объем вычислений.

15. Пример фирмы «Чудесные химикалии» (разд. 10.4). Представим, что совет директоров фирмы выбрал нового президента, который считает целесообразным расширение фундаментальных исследований. Пусть y_j тыс. долл. есть размер ассигнований, выделенных отделению j для проведения таких исследований, а $T_j(y_j)$ млн. долл.— максимальная оценка окончательного эффекта, который, по мнению экспертов, получит фирма в результате вложения этой суммы в фундаментальные исследования. Общий фонд ассигнований на фундаментальные исследования выделяется из фондов фирмы, предназначенных на финансирование всех видов НИР—ОКР, а величина y_j составляет часть от ассигнований, выделенных рассматриваемому отделению. Покажите, как нужно изменить постановку задачи (1) — (4), чтобы учесть в ней ассигнования на фундаментальные исследования. Укажите также, какие изменения нужно внести в рекуррентное соотношение (5) и в задачу (6) и (7). Покажите, как следует изменить постановку задачи (8) — (11). (Обязательно определите все принятые обозначения.)

16. Пример фирмы «Чудесные химикалии» (разд. 10.4). Представим, что один из членов совета директоров фирмы является человеком с консервативными взглядами, настаивающим на том, чтобы общая сумма денег, выделяемая на проведение прикладных и поисковых исследований, не превышала V тыс. долл. Покажите, как при этом условии следует изменить постановку задачи (1) — (4). Запишите соответствующее рекуррентное соотношение, измените должным образом (6) и (7), а также (8) — (11). (Обязательно определите все принятые обозначения.)

17. Объясните, как используются линейные свойства целевой функции и ограничений в рекуррентном соотношении (4) разд. 10.5. Рассмотрите, например, вопрос, почему формула (4) была бы непригодной, если бы целевая функция имела вид $\sum_{j=1}^{s} c_j(y_j)$ и по крайней мере одна из функций $c_j(y_j)$ была бы нелинейной.

18. а) Используйте рекуррентное соотношение (4) из разд. 10.5 для решения задачи

$$\text{Максимизировать } 3y_1 + 4y_2 + 6y_3 + 8y_4 + 6y_5$$

при ограничениях

$$1y_1 + 2y_2 + 2y_3 + 3y_4 + 1y_5 \leqslant 3,$$
$$2y_1 + 1y_2 + 2y_2 + 1y_4 + 3y_5 \leqslant 4,$$

где каждое y_j — неотрицательное целое число.

б) Насколько уменьшается значение целевой функции, если значение постоянного члена в правой части первого ограничения равно не 3, а 2? Тот же вопрос для случая, когда постоянный член в правой части второго ограничения равен не 4, а 3. Тот же вопрос для случая, когда в обоих ограничениях постоянные, стоящие в правых частях, уменьшены на 1.

в) Насколько возрастает значение целевой функции, если значение постоянного члена в правой части первого ограничения не 3, а 4? Тот же вопрос для случая, когда постоянный член в правой части второго ограничения равен не 4, а 5. Тот же вопрос для случая, когда в обоих ограничениях постоянные, стоящие в правых частях, увеличены на 1.

19. Задача замены оборудования. В обоих указанных ниже пунктах найдите оптимальную стратегию замены, используя рекуррентное соотношение (V), приведенное в конце разд. 10.6. (Обязательно укажите значения каждой величины p_{in} и k_{nt}.) Примите, что замена должна быть произведена в начале отрезка 1. Сравните работу алгоритма с рекуррентным соотношением (2).

а) Задача фирмы «Таксолюкс», приведенная в п. г) упражнения 15 гл. 6 и упражнения 68 гл. 7.

б) Задача фирмы «Резвая фреза», приведенная в упражнении 41 гл. 6 и упражнении 69 гл. 7.

20. Рассмотрите рекуррентное соотношение динамического программирования, приведенное в разд. 10.7.

а) Укажите, каким образом каждое из свойств (I) — (IV) влияет на вид рекуррентной формулы.

б) Покажите соответствие между каждым символом в РДП и соответствующим символом в рекуррентной формуле (5) разд. 10.2; в рекуррентной формуле (10) разд. 10.2; в рекуррентной формуле (5) разд. 10.3; в рекуррентной формуле (5) разд. 10.4.

в) Начертите ациклическую сеть, соответствующую рекуррентному соотношению (10) разд. 10.2 и числовым данным таблиц рис. 10.1 и 10.2. [Дайте надлежащие обозначения узлам и ценам дуг, укажите смысл всех величин $D_n(s)$ и отобразите все величины $T_n(s, d_n)$.]

21. Объясните, как вы понимаете следующие термины:

задача распределения усилий;
многошаговый анализ;
задача дискретного программирования;
стратегия.

УПРАЖНЕНИЯ НА ПОСТРОЕНИЕ МОДЕЛЕЙ И РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

Во многих из приведенных ниже упражнений от читателя требуется построить модель на основе рекуррентных соотношений динамического программирования. Обязательно дайте определения всех используемых обозначений и ответы на пять вопросов, сформулированных в конце разд. 10.1. Приведите выражение целевой функции для случая, когда до конца планового периода остался один отрезок, а также рекуррентное соотношение для шага n. Объясните, с чего начинаются и где заканчиваются вычисления.

22. Задача о ранце (рюкзаке). Д. готовится к длительному подъему на опасную вершину. В рюкзаке он может нести груз не более W фунтов. Этот груз может включать N видов предметов, причем каждый предмет вида j весит w_j фунтов. Для каждого предмета вида j альпинист рассчитывает число E_j, определяющее ценность предмета этого вида. Так, например, если он кладет в рюкзак пять предметов вида 3 и семь предметов вида 9, то «ценность» такого выбора содержания рюкзака составляет $5E_3 + 7E_9$. Сколько предметов каждого вида целесообразно положить в рюкзак?

а) Сформулируйте задачу в терминах динамического программирования. Объясните, почему принятие решения было бы тривиальным, если бы в рюкзак можно было бы паковать не только целые предметы, но и их произвольные части. Укажите, каково будет оптимальное решение в каждом случае.

б) Измените ответ, полученный в п. а), предположив, что альпинист устанавливает ограничение сверху L_j на число единиц по каждому виду предметов j, которое он кладет в рюкзак. Объясните, почему решение тривиально, если можно брать не только целые предметы, но и их любые части. Укажите алгоритм отыскания решения в каждом случае.

в) Предположим, что альпинист хочет взять самый легкий рюкзак, ценность которого для него не менее E. Покажите, как нужно изменить постановки задач, приведенных в пп. а) и б).

23. Студент К. заканчивает институт. Чтобы получить диплом, он хочет добиться наилучших оценок на выпускных экзаменах. Он делит имеющееся у него в конце недели учебное время на 10 от резков равной длины. Ему надо сдать экзамены по четырем предметам, два из которых он считает легкими, два трудными. По оценке студента, он не получит ни одного выпускного балла, если совсем не будет заниматься легкими предметами, два балла, если отведет на такой предмет один или два отрезка, три балла, если посвятит им три отрезка, и четыре балла, если будет заниматься легким предметом четыре отрезка. Аналогичные оценки для трудных предметов равны нулю, одному, двум и трем баллам. Кроме того, при изучении трудного предмета в течение пяти отрезков он считает, что получит четыре балла. Не вдаваясь в вопрос о том, являются ли эти оценки излишне оптимистическими, как бы вы посоветовали студенту распределить его время, чтобы максимизировать число баллов, которое он может набрать на экзаменах? Постройте соответствующую модель динамического программирования и найдите ее оптимальное решение. Какова оценка в баллах одного дополнительного отрезка времени для подготовки? Каковы потери в баллах при условии, что студент не будет готовиться к экзаменам в течение одного отрезка?

24. Приятельница студента К., положение которого рассматривалось в упр. 23, приехала в город, где он учится, и назначила ему свидание. Студент пересмотрел свое положение и решил, что для получения диплома ему достаточно набрать девять выпускных баллов. Теперь он стремится так распределить свое время, чтобы минимизировать число временны́х отрезков, необходимых для подготовки к экзаменам с гарантией получения не менее девяти выпускных баллов. Сформулируйте стоящую перед студентом задачу в виде модели динамического программирования и найдите ее оптимальное решение. (В случае если имеется более одного оптимального распределения временны́х отрезков, укажите все такие распределения.)

25. Благотворительное общество «Объединенный фонд» должно выделить 10 своих членов для сбора пожертвований на трех фирмах, конторы которых расположены в трех крупных административных зданиях. По оценке директора, выделив y_j человек на здание j, можно добиться получения обязательств на взносы в фонд общества общей суммой $R_j(y_j)$ сотен долл., где $R_j(0) = 0$ и

$$R_1(1) = 5,\quad R_1(2) = 10,\quad R_1(3) = 15,\quad R_1(4) = 25,\quad R_1(5) = 35,$$
$$R_2(1) = 3,\quad R_2(2) = 6,\quad R_2(3) = 12,\quad R_2(4) = 18,\quad R_2(5) = 30,$$
$$R_3(1) = 20,\quad R_3(2) = 35,\quad R_3(3) = 45,\quad R_3(4) = 55,\quad R_3(5) = 60,$$
$$R_1(6) = 50.$$
$$R_2(6) = 55.$$
$$R_3(6) = 65.$$

(Никаких дополнительных пожертвований не удастся получить, если направить в здание 1 более семи человек, в здание 2 — более пяти и в здание 3 — более шести человек.) Сколько членов общества нужно направить в каждое здание? Постройте соответствующую модель динамического программирования и найдите ее оптимальное решение. Покажите, как изменится решение, если всего для сбора взносов направляется 9, 11 или 12 членов общества.

26. Рассмотрите случай общества «Объединенный фонд», описанный в упражнении 25. Предположим, что у директора фонда не хватает людей и он формулирует задачу следующим образом: минимизировать число членов общества, направляемых для сбора взносов, при условии, что в трех зданиях удастся получить обязательства на взносы в благотворительный фонд не менее R сотен долл. Сформулируйте задачу в терминах динамического программирования и найдите оптимальное решение при условиях: а) $R = 80$; б) $R = 90$; в) $R = 100$.

27. Рассмотрите еще раз случай общества «Объединенный фонд», описанный в упражнении 25. Предположим, что директор стремится выделить не более M человеко-дней на сбор пожертвований в трех зданиях. Ему известно, что если направить в любое здание от одного до трех человек, то *всего* для сбора пожертвований потребуется 1 человеко-день. Если же направить в здание более трех человек, то они истратят 2 человеко-дня. Какое количество людей при общей численности в 10 членов общества нужно направить в каждое здание при условиях: а) $M = 3$; б) $M = 4$?

28. Рассмотрите следующую задачу:

$$\text{Максимизировать } R_1(y_1) + R_2(y_2) + R_3(y_3)$$

при ограничениях

$$2y_1 + 3y_2 + 4y_3 \leqslant W$$

$$\text{каждое } y_j = 0, 1, 2, 3,$$

где любое $R_j(0) = 0$ и

$$R_1(1) = 3, \quad R_1(2) = 2, \quad R_1(3) = 8.$$

$$R_2(1) = 5, \quad R_2(2) = 11, \quad R_2(3) = 9.$$

$$R_3(1) = 9, \quad R_3(2) = 14, \quad R_3(3) = 15.$$

а) Найдите оптимальное решение при $W = 8$. Укажите также оптимальные решения при $W = 6$ и $W = 7$ и все отличные от найденных оптимальные решения.

б) Начертите ациклическую сеть, соответствующую использованному рекуррентному соотношению. Проставьте на узлах нужные номера и значения переменных на дугах.

29. Рассмотрите задачу определения оптимального размера партии для случая изделий одного вида, описанную в разд. 1.6. Предположим, что фирма использует N видов изделий и намерена пополнять их запасы, пользуясь правилом выбора оптимального размера партии. Примем, что еженедельно фирма потребляет M_j изделий вида j, что K_j — постоянные затраты на размещение заказа, c_j — покупная цена одного изделия и h_j — недельные затраты на хранение одного изделия. Кроме того, положим, что Q_j — размер партии, пополняющий запас изделий вида j. Средние еженедельные затраты *по всем* видам изделий представляют собой сумму N выражений, каждое из которых имеет вид (12) из разд. 1.6. Предположим, что фирма устанавливает верхний предел L на среднее значение наличного уровня запасов всех изделий, где изделия каждого вида дают составляющую, равную $c_jQ_j/2$. Покажите, как можно использовать постановку задачи в терминах динамического программирования для определения оптимального размера партий по каждому виду изделий.

30. Рассмотрите случай фирмы «Большая энергетика», описанный в конце разд. 10.3. Примем $s = 8$, $K_j = j$, $j = 1, 2, \ldots, 8$ и

$$R_1 = 16, \quad R_2 = 28, \quad R_3 = 45, \quad R_4 = 48,$$
$$R_5 = 65, \quad R_6 = 66, \quad R_7 = 79, \quad R_8 = 80.$$

Найдите оптимальный набор проектов при следующих условиях:

а) Верхний предел числа проектов $N = 4$, верхний предел ассигнований $M = 10$. Укажите решение при $N = 2$ и $N = 3$. Определите также решение при $N = 3$ и $M = 9$.

б) $N = 3$, $M = 11$.

31. Рассмотрите пример фирмы «Чудесные химикалии», описанный в разд. 10.4. Примем $s = 4$, каждое $L_j = 5$, $N = 10$ и

$$P_j(v_j) = a_jP(v_j), \quad \text{где } P(0) = P(1) = P(2) = 0, \quad P(3) = 1,$$
$$P(4) = 8, \quad P(5) = 17;$$
$$Q_j(w_j) = b_jQ(w_j), \quad \text{где } Q(0) = 0, \quad Q(w) = 3 + w, \quad w \geqslant 1;$$
$$R_j(x_j) = c_j, \quad \text{где } R(0) = 0, \quad R(1) = 8, \quad R(2) = 9,$$
$$R(3) = 9{,}5, \quad R(4) = 9{,}75, \quad R(5) = 10.$$

Найдите оптимальное распределение ассигнований при следующих условиях:

а) все $a_j = b_j = c_j = 1$;

б) $a_1 = b_1 = c_1 = 1$, $a_2 = b_2 = c_2 = 2$, $a_3 = b_3 = c_3 = 3$;

в) $a_1 = b_2 = c_3 = 1$, $b_1 = c_2 = a_3 = 2$, $c_1 = a_2 = b_3 = 3$.

г) Определите, как повлияет на решение изменение значения N в каждом из вышеуказанных пунктов (принять $N = 7, 8, 9$).

д) Определите, как повлияет на решение изменение значения N в п. а) [принять $N = 11$].

е) Определите, как повлияет на решение изменение значения всех L_j в пп. а) — в), приняв $L_j = 4$ для всех j.

32. Три брата управляют скотоводческой фермой в штате Аризона, принадлажащей их отцу. В начале планового периода продолжительностью в N лет у фермеров имеется S сотен голов скота. Затраты на уход за s головами скота и их прокорм в течение года n равны $c_n(s)$. Если фермеры решат отправить y_n голов на продажу в конце года n, то они получат $R_n(y_n)$ долл. дохода. Число голов в стаде, оставшемся у фермеров, увеличивается в 1,6 раза в течение следующего года. (Так, например, если в конце первого года ни одна из S голов скота не продана, то в конце второго года число голов в стаде составит 1,6 S.)

Напишите рекуррентное соотношение динамического программирования, с помощью которого можно определить, сколько голов скота выгодно продавать ежегодно в течение N лет.

33. Задача расширения производственной мощности. Фирма «Большой Билл» планирует реализацию программы расширения производственных мощностей, рассчитанную на период в N лет. В начале отрезка 1 производственная мощность фирмы составляет c_0 единиц. По имеющимся оценкам фирма должна иметь мощность не менее R_t единиц на отрезке t, где $t = 1, 2, \ldots, N$. Если фирма увеличит свои производственные мощности на x_t единиц на отрезке t, то связанные с этим затраты составят $K_t(c, x_t)$, где c представляет собой производственную мощность *на начало* отрезка, а x_t принимает только целочисленные значения. Если некоторых уровней x_t достичь невозможно, то соответствующее значение $K_t(c, x_t)$ можно принять произвольно большим. Предположим, что новые мощности вводятся в эксплуатацию настолько близко к началу рассматриваемого отрезка, что их можно использовать для удовлетворения потребностей R_t. Предположим также, что расходы фирмы на эксплуатацию оборудования на отрезке t при выпуске R равны $H_t(C, R)$, где C определяет мощность *в конце* отрезка и $R \leqslant C$. Наконец примем, что при наличии C единиц мощности *на конец* отрезка только $D(C)$ единиц можно использовать в начале следующего отрезка, что обусловлено амортизацией оборудования. Постройте модель расширения производственной мощности на основе рекуррентного соотношения динамического программирования.

34. Акционерное общество «Биржевик» славится своими успехами на бирже, которые обусловлены тем, что маклеры общества способны точно прогнозировать цену на отдельные акции. Несмотря на это, общество все-таки вынуждено прибегнуть к помощи операционистов для определения оптимальной стратегии покупок и продаж. Предположим, что p_t есть цена акций некоторой фирмы на отрезке t и что маклеры «Биржевика» знают цены этих акций в течение периода из T отрезков. Пусть x_t число акций, которые «Биржевик» имеет в конце отрезка t, где $x_t \geqslant 0$. Предположим далее, что в начале отрезка 1 «Биржевику» не принадлежит ни одной акции, котирующейся на бирже. Общество должно выплачивать сумму $C(x_t - x_{t-1})$ при покупке и продаже акций, где $C(0) = 0$. Общество не может заниматься спекуляцией в том смысле, что суммарная стоимость купленных акций не должна превышать общей суммы наличных денег, которой располагает общество (где в начале отрезка 1 эта сумма равна M), а также не имеет права продавать акции, которыми не владеет в начале соответствующего отрезка.

Сформулируйте задачу в терминах динамического программирования при следующих предположениях:

а) Общество «Биржевик» стремится максимизировать общую сумму наличных денег к концу планового периода.

б) «Биржевик» приписывает цену $v_t(d_t)$ сумме наличных денег d_t на отрезке t, изымаемой для выплаты дивидендов, и стремится максимизировать $v_t(d_t)$ в пределах всего планового периода.

35. Диспетчер транспортного агентства должен обеспечить доставку x_t грузов, каждый из которых занимает s_1 единиц объема, и x_2 грузов, каждый из которых занимает s_2 единиц объема. (Таким образом, общий ежедневный объем доставляемых грузов равен $s_1x_1 + s_2x_2$.) В его распоряжении имеется N различных автомашин. Емкость машины i составляет C_i единиц, а стоимость ее эксплуатации — E_i. (Предположим, что $C_1 + \ldots + C_n > s_1x_1 + s_2x_2$ и что имеется допустимое решение.) Диспетчер стремится выбрать автомашины для доставки всех грузов таким образом, чтобы минимизировать эксплуатационные затраты. Сформулируйте проблему принятия решения в виде модели динамического программирования и укажите, каким образом рекуррентная формула позволяет выбирать машины оптимальной емкости и распределять между ними грузы.

36. Задача автомобильной компании «Гигант» (разд. 2.6). Сформулируйте оптимизационную задачу на основе рекуррентного соотношения динамического программирования.

37. Рассмотрите случай спартанской армии, описанный в упр. 31 гл. 2. Сформулируйте оптимизационную задачу, воспользовавшись рекуррентным соотношением динамического программирования.

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

В целом ряде приведенных ниже упражнений от читателя требуется построить модель на основе рекуррентного соотношения динамического программирования. Читателю следует обязательно определить значение всех используемых обозначений и дать ответы на пять вопросов, приведенных в конце разд. 10.1. Запишите выражение оптимизируемой функции для случая, когда до конца планового периода остается один отрезок, и рекуррентную формулу для шага n. Объясните, с чего надо начинать и чем заканчивать вычисления.

38. Задача выбора места или времени производства (разд. 1.6). Рассмотрите модель, описанную в разд. 1.6:

$$\text{Минимизировать } \sum_{j=1}^{N} \frac{x_j^2}{c_j}$$

при ограничениях

$$\sum_{j=1}^{N} x_j = D \quad \text{и} \quad \text{все} \quad x_j \geqslant 0,$$

где все $c_j > 0$. Используйте рекуррентное соотношение динамического программирования, чтобы получить результат (11), т. е. при заданных значениях $x_1, x_2, \ldots, x_{t-1}$ соответствующее оптимальное значение $x_t = c_t(D - x_t - \ldots - x_{t-1})/(c_t + c_N)$.

39. Рассмотрите задачу

$$\text{Максимизировать } \frac{x_1}{c_1}\cdot\frac{x_2}{c_2}\cdots\frac{x_N}{c_N}$$

при ограничениях

$$\sum_{j=1}^{N} x_j = D \quad \text{и} \quad \text{все} \quad x_j \geqslant 0,$$

где все $c_j > 0$. Используйте постановку задачи в терминах динамического программирования для определения вида оптимального решения. Укажите зависимость оптимального значения целевой функции от D.

40. Рассмотрите задачу

$$\text{Максимизировать } \sum_{j=1}^{s} Rx_j$$

при ограничениях

$$\sum_{j=1}^{s} x_j = N, \quad \text{причем все } x_j \text{ — неотрицательные целые.}$$

Заметим, что все функции, входящие в целевой функционал, идентичны.

а) Предположим, что s является степенью 2 (например, 4, 8, 16, 32, ...). Найдите способ вычисления оптимального решения, пользуясь методом динамического программирования, *не требующим* вычислений $g_j(n)$ при *всех* значениях $j = 1, 2, \ldots, s$.

б) Как нужно изменить подход, предложенный в п. а), чтобы его можно было использовать при *любом* значении s?

41. Предположим, что законодательное собрание штата состоит из R членов. Штат административно разделен на s округов. Население округа j составляет p_j человек, $s < R$. При строго пропорциональном представительстве округ j выбирает $R/p_j \equiv r_j$ членов законодательного собрания, однако такое распределение может оказаться нереализуемым, поскольку r_j может быть нецелым числом. Необходимо найти такое распределение представительства y_j от округов j, $j = 1, 2, \ldots, s$, чтобы минимизировать максимальное отклонение y_j от r_j, т. е. минимизировать $[\max(y_1 - r_1, \ldots, y_s - r_s)]$.

а) Постройте модель на основе рекуррентного соотношения динамического программирования.

б) Используйте эту модель для отыскания решения при $R = 4$, $s = 3$, $r_1 = 0{,}4$, $r_2 = 2{,}4$, $r_3 = 1{,}2$. Рассмотрите, имеет ли смысл решение, если учесть содержание задачи.

42. Рассмотрите задачу выбора величин $y_j = 1, 2, \ldots, s$, таких, что достигается

$$\max \{\min [R_1(y_1), R_2(y_2), \ldots, R_s(y_s)]\},$$

при ограничениях

$$\sum_{j=1}^{s} H_j(y_j) \leqslant N, \qquad y_j = 0, 1, 2, \ldots \quad \text{при} \quad \text{любом } j,$$

где все $H_j(y_j)$ — неубывающие целочисленные функции при $y_j = 0, 1, \ldots$, удовлетворяющие условию $H_j(0) = 0$. Найдите рекуррентное соотношение динамического программирования, позволяющее решить эту задачу.

43. Минимаксный маршрут на сети. Студент из Англии, обучающийся в США по обмену, купил подержанный автомобиль и собирается провести лето, путешествуя из Нью-Йорка в Сан-Франциско. Для упрощения условия задачи примем, что он может выбрать маршрут, пользуясь той же картой, которая показана на рис. 8.1. Студента беспокоит возможность отказа двигателя в случае выбора маршрута, который проходит между двумя такими городами, где летом бывает сильная жара. По имеющимся у студента оценкам величина c_{ij} равна максимальной температуре, которая может встретиться при проезде из штата i в штат j. Он стремится выбрать маршрут, минимизирующий максимальную температуру в течение всей поездки.

а) Сформулируйте задачу в терминах динамического программирования и примените рекуррентную формулу к величинам c_{ij}, заданным на рис. 8.1.

б) Запишите рекуррентную формулу в достаточно общем виде, чтобы ее можно было применить к ациклическим сетям, подобным сетям, рассмотренным в разд. 7.7 (иными словами, выразите это соотношение в виде, аналогичном выражениям (1) и (2) разд. 7.7).

44. Следующий за оптимальным маршрут на сети. Рассмотрите задачу отыскания кратчайшего маршрута, изложенную в разделах 7.6 и 7.7. Сформулируйте эту задачу в терминах динамического программирования и найдите следующий по протяженности маршрут из источника в сток. Для упрощения задачи примите, что сеть ациклическая и что кратчайший маршрут является единственным. Если вам удастся предложить более одного метода решения, то укажите все альтернативы. [*Указания*: следующий за оптимальным маршрут должен отличаться от кратчайшего хотя бы одной дугой. Докажите также, что следующий за оптимальным маршрут из узла j в сток должен проходить по некоторой дуге (j, k), а затем либо по самому короткому, либо по следующему за ним маршруту из узла k в сток; кроме того, укажите вытекающее из этого следствие.] Какие трудности возникают при решении этой задачи из сети, содержащей циклы?

45. Рассмотрите задачу отыскания кратчайшего маршрута из источника в сток, изложенную в разд. 7.6 и 7.7. Сформулируйте задачи отыскания кратчайшего маршрута из источника в сток в терминах динамического программирования при условии, что этот маршрут содержит в точности K дуг и в предположении, что по крайней мере один такой маршрут существует. (Может ли этот маршрут содержать контур или цикл?)

46. Задача о ранце. Рассмотрите оптимизационную задачу, приведенную в упражнении 22. Предположите, что для предметов вида k наибольшее значение отношения E_j/w_j равно E_k/w_k. Докажите, что существует некоторое значение величины W^*, такое, что если ограничение по весу равно $W \geqslant W^*$, то имеется оптимальное решение, включающее по крайней мере один предмет вида k.

47. Замена и ремонт оборудования. Рассмотрите модель, приведенную в разд. 10.6. Предположим, что имеется возможность замены либо ремонта оборудования в течение любого отрезка после его приобретения на отрезке 1 и до отрезка N, когда заканчивается полный срок амортизации. Примем далее, что продолжительность ремонта пренебрежимо мала по сравнению с продолжительностью отрезка. Следовательно, машина, отремонтированная в начале отрезка, уже может эксплуатироваться в течение всего этого отрезка. Рассмотрим машину, купленную в начале отрезка n, отремонтированную в начале отрезков $t_1 < t_2 < \ldots < t_p$, где $n < t_1$, и замененную в начале отрезка k, где $k > t_p$. Предположим, что соответствующие затраты равны

$$a_n(n, t_1) + a_n(t_1, t_2) + \ldots + a_n(t_{p-1}, t_p) + b_{nk}(t_p).$$

Величина $a_n(h, j)$, где $n \leqslant h < j$, представляет собой затраты на эксплуатацию машины в течение всех отрезков от h до j плюс затраты на ее ремонт в начале отрезка j при условии, что машина была куплена на отрезке n и отремонтирована в последний раз на отрезке h (либо была куплена при $h = n$). Величина $b_{nk}(t_p)$ соответствует чистым общим затратам, включающим покупку машины на отрезке n плюс эксплуатационные затраты на отрезках от t_p до k, т. е. между последним ремонтом и продажей, за вычетом ее остаточной стоимости на отрезке k. Величина $b_{nk}(n)$ представляет собой затраты на покупку машины и ее эксплуатацию на отрезках от n до k при условии отсутствия ремонтов.

Покажите, как можно использовать рекуррентное соотношение динамического программирования для вычисления значений c_{nk}, которые затем соответствующим образом используются в рекуррентной формуле (2) разд. 10.6.

48. Рассмотрите классическую транспортную задачу, подобную задаче, приведенной в разд. 6.2, где число поставщиков $m = 2$ и суммарная мощность поставщиков равна суммарному спросу потребителей [соотношение (6)]. Предположим, что транспортные затраты на доставку x_{ij} единиц груза из пункта отправления i

в пункт назначения j описываются нелинейной функцией $c_{ij}(x_{ij})$. Примем, что все величины S_i и D_j целые, и наложим такое же ограничение на все x_{ij}.

а) Сформулируйте задачу в терминах рекуррентного соотношения динамического программирования. (*Указание*: для получения одномерной переменной состояния можно использовать равенство общей мощности поставщиков общему спросу потребителей.)

б) Покажите, как изменится постановка задачи п. а) при $m = 3$.

49. Задача владельца ресторана. Рассмотрите случай фирмы «Общепитторг», описанный в упражнении 48 гл. 6. Сформулируйте оптимизационную задачу в терминах рекуррентного соотношения динамического программирования.

50. Задача о раскрое материала. Рассмотрите случай фирмы «Паутинка», которой нужно разрезать большой рулон шириной 70 см на меньшие рулоны шириной 22, 20 и 12 см. Предположим, что при решении модели линейного программирования в рассматриваемое подмножество включены все допустимые варианты. Пусть y_1, y_2 и y_3 — полученные в результате оптимальные значения двойственных переменных для ограничений, отвечающих рулонам шириной 22, 20 и 12 см соответственно. Примените теорему двойственности и рекуррентное соотношение динамического программирования, включающее величину y_i, для проверки оптимальности решения задачи линейного программирования. (Может ли какой-либо вариант, не включенный в подмножество, улучшить найденное решение задачи линейного программирования?)

ГЛАВА 11

Принятие решений при бесконечном плановом периоде

11.1. МОДЕЛИ С БЕСКОНЕЧНЫМ ПЛАНОВЫМ ПЕРИОДОМ

Не вызывает сомнения, что большинство принимаемых решений, если не все, представляют собой часть бесконечной цепи действий. Сделанный ранее выбор влияет на настоящее, принимаемые в данный момент решения влияют на будущее и т. д. В свете такого подхода *все* модели можно рассматривать как элементы **бесконечного планового периода**. Будущее, находящееся за пределами заранее установленного планового периода, в некоторых рассмотренных выше динамических моделях попросту не принималось во внимание; иногда (например, в гл. 9) удается доказать теорему о длительности планового периода, которая показывает, что такая процедура может обеспечить оптимальность текущих решений. В других моделях попытка учесть будущее связана с выбором тех или иных «конечных» условий (таких, например, как установление нижних границ для численности работающих или для производственной мощности). В отличие от этих моделей в настоящей главе длительность планового периода предполагается бесконечной.

Для того чтобы в моделях с бесконечным плановым периодом получить определенные решения, необходимо дополнительное ограничительное допущение, которое, вообще говоря, можно назвать **допущением о стационарности**. В простейшем случае предполагается, что все функции экономического эффекта, выбираемой программы и внешние условия (например, объем спроса) для каждого из отрезков одинаковы. Все приводимые ниже примеры относятся к этому случаю, и необходимо учитывать, что получаемые результаты основаны на таком допущении о стационарности. В гл. 11 и 12 рассматриваются только детерминированные модели; вероятностным моделям посвящены гл. 17 и 18, где постулируется стационарный характер вероятностных закономерностей. При использовании более сложных методов анализа могут быть получены решения и для таких ситуаций, где допущения о стационарности являются не столь сильными, например для случая циклического изменения таких показателей, как сезонный спрос (условный пример приведен в конце гл. 12).

Поскольку в реальной действительности стационарность редко сохраняется в течение длительных периодов времени, то возникает следующий вопрос: имеют ли практическое значение модели, в которых предполагается стационарность в течение бесконечного планового периода? Наш ответ будет утвердительным, что иллюстрирует рассмотрение ситуаций двух типов.

В ситуациях первого типа динамические оптимизационные модели используются для улучшения текущих оперативных решений — скажем, решений о пополнении запасов и разработки графиков производства. В качестве примера рассмотрим фирму, осуществляющую ежедневный контроль своих запасов и заказывающую некоторые изделия раз в несколько недель, когда уровень соответствующего запаса достигнет установленного «критического» значения. Пусть, как это часто бывает в действительности, спрос на данное изделие, а также затраты на его закупку и хранение являются устойчивыми в течение периода, не меньшего 12 месяцев. Тогда для фирмы целесообразно сохранить действующие правила пополнения запасов еще на 3—6 месяцев, а в конце этого срока пересмотреть их на основе нового 12-месячного прогноза. (Такой способ действий оправдан и другими важными соображениями. Если фирма имеет запасы сотен видов различных изделий, частое изменение правил пополнения запасов каждого вида изделий требует слишком больших затрат времени и вызывает отрицательный эффект — связанные с ним издержки могут намного превысить экономию, вызванную улучшением решений.)

Как же должна фирма устанавливать правило пополнения запасов для каждого конкретного вида изделий? Один из возможных подходов заключается в использовании методов, описанных в гл. 8 и 9: для планового периода длительностью 365 дней определяются минимальные затраты $f_N(i)$ и оптимальная программа для исходного уровня запасов i, а затем в течение нескольких первых месяцев применяются правила пополнения запасов, установленные для $N = 365$. Однако модель с бесконечным плановым периодом, основанная на допущении о стационарности, может дать превосходные результаты и, как правило, требует намного меньшего объема вычислений.

В соответствии с этим одно из оснований для использования стационарной модели с бесконечным плановым периодом состоит в том, что при выполнении текущих операций в начале планового периода такой подход является эффективным и относительно простым способом получения оптимальных решений.

К ситуациям второго типа относится использование динамических оптимизационных моделей при принятии повторяющихся важных решений о распределении капитальных вложений, например в случае замены дорогостоящего оборудования. Крупные машинные агрегаты иногда заменяются не чаще чем раз в 15—20 лет. В связи с этим к моменту очередной замены, вероятно, появятся совершенно новые типы оборудования. Каким же образом принимать текущее решение о капиталовложениях? Разумеется, одна из возможных альтернатив состоит в том, что необходимость замены оборудования попросту не учитывается. Как очевидно из следующего далее примера, который взят из реальной действительности, такой подход может ввести в заблуждение и оказаться весьма рискованным.

Одна из фирм пищевой промышленности столкнулась с наличием «узкого места» в производстве. Возможны два пути разрешения периодически возникающих трудностей: либо накопление в течение периода затишья достаточных запасов продукции для покрытия спроса в «часы пик», либо приобретение нового оборудования для увеличения производственных мощностей. Если сравнивать прирост затрат в связи с увеличением запасов только с *начальной* стоимостью дополнительного оборудования, приобретение последнего представляется предпочтительным. Если же учесть и затраты на будущие повторные замены оборудования, его приобретение покажется гораздо менее привлекательным — даже при оптимистической оценке затрат, связанных с последующими заменами.

Таким образом, вторым доводом в пользу применения стационарных моделей с бесконечным плановым периодом является то обстоятельство, что в ситуациях с регулярно повторяющимися во времени капиталовложениями использование подобного подхода может привести к лучшим решениям, чем полной отказ от учета будущего.

Оптимальное поведение при стационарности. Сделав допущение о стационарности, можно интуитивным путем прийти к раскрытию содержания такого понятия, как «принятие текущего решения в условиях бесконечного планового периода». Однако читатель, возможно, удивится, узнав, что интуиция способна оказать лишь ограниченную помощь при выяснении *способа* определения *оптимального* решения.

Для иллюстрации рассмотрим вначале модель динамического программирования с конечным плановым периодом. В начале любого отрезка необходимо знать лишь состояние системы и остающееся число отрезков. Об оптимальности программы можно судить по сумме конечной последовательности эффектов. Теперь перейдем к бесконечному числу отрезков; по определению, «оставшееся число отрезков» здесь всегда одинаково, и любая программа, или стратегия, применяемая в течение всего планового периода, обусловливает бесконечную последовательность значений эффекта. Поскольку для любой стратегии эта последовательность растет с увеличением планового периода, необходим способ *сравнения* стратегий. Разумеется, если одна из стратегий обеспечивает больший эффект, чем другая для *любой* фиксированной длительности планового периода, проблема сравнения не вызывает затруднений. Однако одна и та же стратегия обычно представляется лучшей для одной длительности планового периода и худшей — для другой, так что решение ни в коей мере нельзя считать очевидным. Следовательно, при бесконечном плановом периоде возникают два важных вопроса выбора оптимального решения:

I) Каким критерием воспользоваться для оценки сравнительной предпочтительности различных бесконечных последовательностей значений эффекта?

II) Можно ли в условиях бесконечного планового периода считать оптимальными только стационарные стратегии, т. е. стратегии, зависящие исключительно от состояния системы в настоящий момент?

В настоящей главе внимание сосредоточено на вопросе I). Критически рассматривается несколько часто применяемых критериев оценки бесконечных последовательностей. Будет показан также способ использования этих критериев применительно к простой, но важной модели восстановления. При анализе решения этой модели читатель обнаружит несколько численных методов последовательного приближения, применимых и к обобщенным оптимизационным задачам с бесконечным плановым периодом.

В гл. 12 рассматривается большой класс моделей, для которых ответ на вопрос II) положителен, т. е. всегда можно отыскать такую стационарную стратегию, которая является оптимальной. Для разработки такой стратегии необходимо знать только *текущее* состояние системы, а не последовательность событий, приведших к этому состоянию. Таким образом, мы будем принимать одно и то же решение всякий раз, когда придем к тому же состоянию. В частности, в гл. 12 изучается метод построения оптимальной стационарной стратегии. Построение соответствующих правил или таблиц обеспечивает возможность удобного их применения в рамках динамического программирования для принятия текущих решений.

11.2. ТОНКОСТИ, СВЯЗАННЫЕ С ОЦЕНКОЙ БЕСКОНЕЧНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Возможно, читатель удивится, узнав, что деловые люди, экономисты и математики уже несколько столетий спорят о методах оценки **бесконечных последовательностей значений эффекта.** Приводимые ниже примеры выявляют причины этих споров, и следует четко разъяснить экономический смысл примеров и их важность с точки зрения теории управления.

Опыт показал, что большинство лиц, принимающих решения, не способны на основе интуиции давать непротиворечивые оценки сравнительной предпочтительности различных бесконечных последовательностей значений эффекта. Вследствие этого большинство бизнесменов прибегают к формулам, дающим возможность выполнить хотя бы предварительный отсев или ранжирование альтернативных вариантов. Мы также используем формулы, позволяющие преобразовать бесконечную последовательность значений эффекта в одно число, характеризующее «относительное качество» соответствующего варианта. Однако, прежде чем приступить к этому, мы хотим убедиться в том, что читатель разбирается в некоторых существенных вопросах, имеющих отношение к различным методам выбора одной из бесконечных последовательностей. Только обладая такими познаниями, читатель сможет оценить как сильные стороны, так и ограничения простых в приложении формул.

Приводимое далее обсуждение не может служить исчерпывающим обзором осложнений, возникающих в связи с использованием бесконечных последовательностей, а также предлагаемых способов их разрешения. Мы надеемся лишь на то, что приводимые примеры заставят читателя в дальнейшем проявлять осторожность и воздержаться от некритического принятия, казалось бы, заманчиво простых решений.

В настоящем разделе рассматриваются два вопроса, являющихся основными при динамической постановке оптимизационных проблем:

I) В каких случаях формула оценки является адекватной для сравнения различных стратегий?

II) *Всегда ли* такая формула позволяет свести бесконечную последовательность к одному числу, которое можно использовать в качестве основы для сравнения?

Эти вопросы и рассматриваются далее весьма подробно, поскольку связанные с ними проблемы могут быть чрезвычайно трудными.

Анализируются три «критерия качества». Первым из них является *средний эффект за отрезок.* В практике этот критерий возникает чаще всего в тех случаях, когда экономический эффект измеряется уровнем затрат. Тогда рекомендуется следующее правило выбора: избирать следует вариант, которому отвечают наименьшие средние затраты за отрезок. Второй критерий именуется *интегральным эффектом, дисконтированным к настоящему моменту времени*, или же интегральным дисконтированным эффектом. Как читатель вскоре увидит, при использовании этих двух критериев не всегда выбираются одинаковые стратегии — иногда возникают технические трудности весьма неприятного характера.

Третий критерий именуется *эквивалентным средним эффектом* [1]. Вероятно, его идея окажется новой для читателя; она имеет большое значение в исследовании операций, поскольку обеспечивает математическую увязку двух других упомянутых выше критериев. При использовании этого критерия часто удается определить *вид* оптимальной стратегии; затем, располагая этим результатом, можно вычислить конкретные численные значения — либо среднего дохода за отрезок, либо интегрального дисконтированного эффекта.

Еще одним из критериев является *внутренняя норма эффективности,* которая, к сожалению, часто используется в деловой практике. Здесь этот критерий не рассматривается, потому что такой подход часто приводит к неправильным решениям.

При изучении приведенных ниже численных примеров читатель обнаружит, что большинство необходимых для них алгебраических преобразований уже выполнено. При первом чтении следует без доказательства принять, что формулы являются правильными — экономическое содержание этих примеров не связано с выполняемыми

---

[1]) Этот критерий в известной мере аналогичен широко применяемому в отечественной практике критерию «приведенных затрат».— *Прим. перев.*

в них алгебраическими преобразованиями. Позже читатель может возвратиться к ранее прочитанному материалу и проверить, сумеет ли он вывести формулы самостоятельно.

Полезность денег. Часто приходится слышать, что доллар, которым располагают в настоящий момент, через год будет стоить больше доллара. Почему? Эта формулировка опирается на несколько соображений. Так, иногда обнаруживается, что простая «полезность» или «личная ценность» доллара выше в настоящий момент, чем спустя некоторое время. Предположим, что фирма в течение 25 лет регулярно выплачивает акционерам дивиденды раз в квартал. С большим нежеланием такая фирма пойдет на отказ от очередной выплаты дивидендов, даже обещая выплатить их позже, причем, быть может, с процентами. Замеченная разница между полезностью денежных поступлений и затрат в разные моменты времени и является основой проблемы соизмерения бесконечных последовательностей значений эффекта.

По существу, точно такая же проблема соизмеримости возникает и в моделях с конечным плановым периодом. В условных примерах задач динамического программирования, приведенных в предыдущих главах, различные стратегии, или программы, обусловливали и разные временны́е последовательности прибылей или затрат. Однако трудности соизмерения ранее просто *не принимались во внимание.* Поскольку при конечном плановом периоде всегда можно *суммировать* прибыли или затраты за весь период, имеющаяся целевая функция позволяла исчислить для каждой стратегии одно конечное число, что и упрощало оптимизацию. Однако теперь мы более не можем игнорировать трудности соизмерения, даже если бы мы этого и хотели. Поскольку суммирование осуществляется на бесконечном плановом периоде, общая сумма эффекта для большинства стратегий также будет бесконечной. Нетрудно понять, что при сравнении нескольких бесконечных последовательностей значений эффекта любой безыскусственный подход окажется успешным только в простейших случаях. Так, какой-либо метод сравнения стратегий

	Отрезок						
Стратегия	1	2	3	4	5	6	...
A	3	2	1	3	2	1	...
B	3	1	3	1	3	1	...
C	1	6	—1	1	6	—1	...
D	$2\frac{2}{3}$	2	2	2	2	2	...
E	1	3	1	3	1	3	...
F	1	1	1	1	1	1	...
G	4	0	0	4	0	0	...

Рис. 11.1. Примеры бесконечных последовательностей значений эффекта.

для любой фиксированной конечной длительности планового периода не всегда приносит результаты; продемонстрируем это на следующем примере.

Пусть требуется выбрать один из приносящих прибыль альтернативных вариантов, которые на рис. 11.1 характеризуются показателями эффекта (в данном случае приравненными прибыли) для последовательности временных отрезков, начинающейся с настоящего момента. На первом отрезке варианты A и B дают прибыль 3 единицы, вариант G — 4 единицы и т. д.; на втором отрезке вариант A дает прибыль 2 единицы, вариант B — всего лишь 1 единицу и т. п.

Разумно сразу же исключить вариант F, поскольку вариант A для каждого из отрезков обеспечивает лучшие (или по меньшей мере не худшие) результаты — иными словами, A доминирует над F. Вероятно, можно также обосновать исключение варианта E, поскольку вариант D для любого отрезка обеспечивает большую *совокупную* прибыль, т. е. прибыль с начала планового периода.

Вместе с тем нельзя исключить вариант G. Хотя совокупная прибыль на конец отрезка 2 при этом варианте не столь велика, как аналогичная величина в вариантах A и B, прибыль для отрезка 1 в случае варианта G определенно превышает прибыль всех остальных вариантов. Если требуется максимизировать прибыль текущего отрезка, вариант G является оптимальным.

По какому принципу выбрать лучший из вариантов A, B, C и D? На конец отрезка 2 наиболее эффективным представляется вариант C. Для всех четных отрезков вариант D обеспечивает большую *совокупную* прибыль с начала планового периода, чем вариант B. Для отрезка 3 вариант B обеспечивает совокупную прибыль, равную 7 единицам — по сравнению с 6 единицами для вариантов A и C и $6^2/_3$ единицы для варианта D. Более того, вариант B «опережает» варианты A, C и D по состоянию на конец *каждого* отрезка с индексом $(3 + 6n)$, где $n = 0, 1, 2, \ldots$. Выбор из числа подобных последовательностей в моделях с бесконечным плановым периодом является скорее правилом, чем исключением.

Критерий среднего эффекта за отрезок. Утверждать, что один из вариантов A, B, C, D или G является лучшим, можно *только* на основе дополнительных допущений. Так, можно *дополнительно* постулировать, что для нас имеет одинаковое значение получение единицы эффекта на любом из отрезков. Это означает, что получение эффекта на более раннем отрезке не дает каких-либо преимуществ по сравнению с его получением позже. К какому решению приводит это допущение?

При сделанном допущении целесообразно рассмотреть **средний эффект за отрезок** при бесконечно возрастающем числе отрезков, выбирая вариант с наибольшим средним эффектом. Для варианта A последовательность таких показателей имеет следующий вид: $3/1$, $(3 + 2)/2$, $(3 + 2 + 1)/3, \ldots$; для варианта B: $3/1$, $(3 + 1)/2$,

$(3 + 1 + 3)/3$ и т. д. Результаты подобных вычислений приведены в таблице рис. 11.2. Как показано далее для вариантов A, B, C и D, средний эффект за отрезок стремится к 2, а для варианта G — к $1^1/_3$. Иными словами, при достаточно большом числе отрезков для вариантов A, B, C и D средний эффект будет сколь угодно близок к 2, а для

Стратегия	Отрезок 1	2	3	4	5	6	7	8	9	...
A	3	$2\frac{1}{2}$	2	$2\frac{1}{4}$	$2\frac{1}{5}$	2	$2\frac{1}{7}$	$2\frac{1}{8}$	2	...
B	3	2	$2\frac{1}{3}$	2	$2\frac{1}{5}$	2	$2\frac{1}{7}$	2	$2\frac{1}{9}$	...
C	1	$3\frac{1}{2}$	2	$1\frac{3}{4}$	$2\frac{3}{5}$	2	$1\frac{6}{7}$	$2\frac{3}{8}$	2	...
D	$2\frac{2}{3}$	$2\frac{1}{3}$	$2\frac{2}{9}$	$2\frac{1}{6}$	$2\frac{2}{15}$	$2\frac{1}{9}$	$2\frac{2}{21}$	$2\frac{1}{12}$	$2\frac{2}{27}$	...
G	4	2	$1\frac{1}{3}$	2	$1\frac{3}{5}$	$1\frac{1}{3}$	$1\frac{5}{7}$	$1\frac{1}{2}$	$1\frac{1}{3}$	...

Рис. 11.2. Средний эффект за отрезок.

варианта G — к $1^1/_3$. Следовательно, *если* дополнительно сделать особое допущение об отсутствии временнóй предпочтительности эффекта, варианты A, B, C и D представляются одинаково привлекательными (несмотря на то, что для любого конечного планового периода их эффективность различна), а вариант G — менее привлекательным.

Как читатель может проверить, общие члены последовательностей средних эффектов за отрезок можно представить следующим образом:

$$\text{A:}\quad \frac{3+6n}{1+3n},\ \frac{5+6n}{2+3n},\ \frac{6+6n}{3+3n}\quad n=0,\,1,\,2,\,\ldots; \tag{I}$$

$$\text{B:}\quad \frac{3+4n}{1+2n},\ \frac{4+4n}{2+2n}\quad n=0,\,1,\,2,\,\ldots; \tag{II}$$

$$\text{C:}\quad \frac{1+6n}{1+3n},\ \frac{7+6n}{2+3n},\ \frac{6+6n}{3+3n},\quad n=0,\,1,\,2,\,\ldots; \tag{III}$$

$$\text{D:}\quad 2+\frac{2}{3n},\quad n=1,\,2,\,3,\,\ldots; \tag{IV}$$

$$\text{G:}\quad \frac{4+4n}{1+3n},\ \frac{4+4n}{2+3n},\ \frac{4+4n}{3+3n},\quad n=0,\,1,\,2,\,\ldots\,. \tag{V}$$

Таким образом, при $n \to \infty$ любой член последовательности стремится к 2 для вариантов A, B, C и D и к $1^1/_3$ — для варианта G.

Наиболее очевидным недостатком использования среднего эффекта за отрезок в качестве критерия выбора для бесконечного планового периода является полная нечувствительность этого критерия к эффекту за *конечное* число отрезков. Для иллюстрации предположим, что имеется вариант (будем называть его «альтернативным»), показатели которого полностью идентичны показателям варианта А, за одним лишь исключением, что прибыль на отрезке 1 для альтернативного варианта равна не 3, а 100 единицам. Если пользоваться исключительно критерием среднего эффекта за отрезок, то вариант А и альтернативный вариант представляются одинаково предпочтительными, поскольку для бесконечного периода преимущество альтернативного варианта на первом отрезке оказывается несущественным. Этот критерий обладает и другими недостатками, которые рассматриваются ниже.

Предположим, что *принято решение* использовать критерий среднего эффекта за отрезок; всегда ли можно быть уверенным в том, что данная последовательность имеет определенный предел при возрастании числа отрезков до бесконечности? Пусть, например, две последовательности значений прибыли для вариантов А* и В* получаются путем умножения показателей прибыли первого цикла отрезков на одинаковые для обоих вариантов коэффициенты:

$$\begin{array}{lllll} \text{вариант А*:} & 3, 2, 1, 3, 2, 1; & 6, 4, 2, 6, 4, 2; & 9, 6, 3, 9, 6, 3\ \ldots; \\ \text{вариант В*:} & 3, 1, 3, 1, 3, 1; & 6, 2, 6, 2, 6, 2; & 9, 3, 9, 3, 9, 3\ \ldots. \end{array} \qquad (1)$$

Хотя по аналогии с прежними примерами можно говорить, что обе последовательности одинаково предпочтительны, *правило* сравнения на основе среднего эффекта за отрезок здесь неприменимо, поскольку при бесконечном возрастании числа отрезков бесконечно возрастает и средний эффект за отрезок.

Следовательно, если в качестве критерия оптимальности используется средний эффект за отрезок, необходимо допустить, что для сравниваемых конкретных последовательностей существует единственный и *конечный* предел этого среднего эффекта. Во многих случаях такое допущение является оправданным.

Критерий интегрального дисконтированного эффекта. Другой подход к соизмерению бесконечных последовательностей состоит в исчислении так называемых **интегральных показателей эффекта, дисконтированных к настоящему моменту времени.**

Если для некоторого варианта последовательность значений эффекта имеет вид

$$R_1,\ R_2,\ R_3,\ \ldots,\ R_n,\ \ldots, \qquad (2)$$

то о предпочтительности этого варианта здесь судят по значению интегрального дисконтированного эффекта, равного следующей сумме:

$$R_1+\alpha R_2+\alpha^2 R_3+\ldots+\alpha^{n-1}R_n+\ldots=\sum_{t=1}^{\infty}\alpha^{t-1}R_t, \qquad (3)$$

где i — процентная ставка за отрезок, а $\alpha = [1 + (i/100)]^{-1}$ есть **коэффициент дисконтирования за отрезок.** Чем больше значение i, тем меньше α. Величина нормы процента, которой фирма должна пользоваться при принятии решений, сама по себе является важнейшим вопросом и служит предметом оживленного обсуждения как учеными, так и финансистами-практиками. Здесь предполагается, что для данной фирмы может быть определено адекватное значение i.

Такое предположение можно обосновать учетом внешних условий. Пусть фирма может как брать взаймы, так и ссужать любую требуемую сумму денег при оплате **по правилу сложных процентов** при фиксированной процентной ставке, равной i% за отрезок. (Разумеется, также требуется возмещать и саму полученную ссуду.) Пусть, например, процентная ставка равна 5% в год. Тогда, взяв взаймы 1 долл., через год приходится вернуть $(1 + 0{,}05)$ долл., через два года $(1 + 0{,}05)^2$ долл., через n лет $(1 + 0{,}05)^n$ долл. Аналогично этому 1 долл., который будет получен через n лет, в настоящий момент эквивалентен всего лишь $(1 + 0{,}05)^{-n}$ долл. Если в настоящий момент имеется $(1 + 0{,}05)^{-n}$ долл. и они даны взаймы под сложные проценты при 5% годовых, через n лет ссуда будет возвращена в размере 1 долл.

Вот почему учет внешних условий оправдывает использование критерия интегрального дисконтированного эффекта. Рассмотрим выбор одного из двух вариантов, характеризуемых различными значениями этой величины. Поскольку полезность денежных поступлений и затрат в разные моменты времени может оказаться неодинаковой, для перераспределения эффекта во времени фирма, возможно, на одних отрезках пожелает брать ссуды, а на других — давать взаймы.

Пусть, например, поступление R_2 желательно использовать в настоящий момент, а не спустя один отрезок. Тогда в данный момент мы можем одолжить сумму, равную αR_2, а при получении R_2 оплатить полученную ссуду вместе с процентами из дохода. Аналогично этому можно дать взаймы ссуду на несколько отрезков, а позже получить возвращаемую ссуду вместе со сложными процентами. Обдумав эту ситуацию, читатель поймет, что общая ценность некоторой последовательности эффектов по существу характеризуется числом, показывающим, какую ссуду можно получить *в настоящий момент*, направляя каждый из этих эффектов на погашение ссуды.

Теперь предположим, что выбран вариант, обеспечивающий наибольший интегральный дисконтированный эффект. Согласно сделанному допущению, можно брать и давать взаймы, погашая обязательства ресурсами, которые создаются последовательностью эффектов для этого варианта, так что будут получены *все те* выгоды, которые дают другие варианты; кроме того, от интегрального эффекта дополнительно останется некоторая сумма. Иными словами, при выборе варианта с наибольшим интегральным дисконтированным эффектом можно построить и любую последовательность, соответствующую варианту с меньшим интегральным эффектом; оставшаяся разница представляет собой чистую выгоду. Подводя итоги, можно сказать, что, *какова бы ни была* система предпочтений денег во времени, наилучшим выбором всегда является выбор варианта, которому соответствует наибольший интегральный дисконтированный эффект.

n \ i, %	5	10	20
1	0,952	0,909	0,833
5	0,783	0,621	0,402
10	0,614	0,385	0,161
15	0,481	0,239	0,065
20	0,377	0,149	0,026
40	0,142	0,022	0,001

Рис. 11.3. Коэффициенты дисконтирования $\alpha^n = \left(1+\frac{i}{100}\right)^{-n}$.

Разумеется, допущение о неограниченной возможности брать и давать взаймы весьма редко соответствует действительности, однако часто оно является достаточно хорошим приближением к ней, позволяющим получить адекватные ответы. В пользу критерия интегрального дисконтированного эффекта можно привести и другие доводы. Так, при практическом применении в пользу формулы (3) говорит то обстоятельство, что эффект для отдаленного будущего умножается на малые коэффициенты дисконтирования и соответственно в меньшей степени влияет на принимаемое решение. Из таблицы, приведенной на рис. 11.3, видно, как быстро уменьшаются значения α^n при увеличении как процентной ставки i, так и числа отрезков n. Однако следует помнить, что использование формулы (3) или любой другой формулы может быть оправдано лишь в том случае, если сделаны некоторые специфические допущения о характере предпочтений денег во времени для лица, принимающего решения.

Теперь рассмотрим формулу исчисления интегрального дисконтированного эффекта и проверим, необходимы ли для применимости этого критерия дополнительные допущения. Начнем с проверки того, всегда ли является конечной сумма бесконечного числа членов в формуле (3). Пусть сперва значения эффекта для всех отрезков одинаковы. Тогда интегральный дисконтированный эффект (обозначим его ИДЭ) равен

$$\text{ИДЭ} = R + \alpha R + \alpha^2 R + \alpha^3 R + \ldots, \tag{4}$$

т. е.

$$\text{ИДЭ} = \frac{R}{1-\alpha}, \quad 0 \leqslant \alpha < 1. \tag{5}$$

Коэффициенты при R в формуле (4) попросту являются членами геометрической прогрессии, что и позволяет получить формулу (5). Заметим, что налагается ограничение $\alpha < 1$. При значениях α, близких к единице, интегральный дисконтированный эффект является большим числом — однако числом *конечным*. При $\alpha = 1$ значение интегрального эффекта в формуле (4) бесконечно велико, если $R \neq 0$, а в формуле (5) не определено. Этот случай рассматривается далее.

Теперь исчислим (при $0 \leqslant \alpha < 1$) значения интегрального дисконтированного эффекта для вариантов A, B, C, D и G, экономическая характеристика которых отражена в таблице, приведенной на рис. 11.1.

Вариант A: $$3 + 2\alpha + 1\alpha^2 + 3\alpha^3 + 2\alpha^4 + 1\alpha^5 + \ldots = \tag{6}$$

$$= (3 + 2\alpha + 1\alpha^2)(1 + \alpha^3 + \alpha^6 + \ldots) = \frac{3 + 2\alpha + 1\alpha^2}{1 - \alpha^3}. \tag{7}$$

Вариант B: $$3 + 1\alpha + 3\alpha^2 + 1\alpha^3 + 3\alpha^4 + 1\alpha^5 + \ldots = \tag{8}$$

$$= (3 + 1\alpha)(1 + \alpha^2 + \alpha^4 + \ldots) = \frac{3 + 1\alpha}{1 - \alpha^2}. \tag{9}$$

Вариант C: $$1 + 6\alpha - 1\alpha^2 + 1\alpha^3 + 6\alpha^4 - 1\alpha^5 + \ldots =$$

$$= (1 + 6\alpha - 1\alpha^2)(1 + \alpha^3 + \alpha^6 + \ldots) = \frac{1 + 6\alpha - 1\alpha^2}{1 - \alpha^3}. \tag{10}$$

Вариант D: $$2\frac{2}{3} + 2\alpha + 2\alpha^2 + \ldots = 2\frac{2}{3} + 2\frac{\alpha}{1 - \alpha}. \tag{11}$$

Вариант G: $$4 + 4\alpha^3 + 4\alpha^6 + \ldots = \frac{4}{1 - \alpha^3}. \tag{12}$$

Для сопоставления вариантов A и B можно рассмотреть разность между соответствующими значениями интегрального дисконтированного эффекта:

$$\text{ИДЭ}[A] - \text{ИДЭ}[B] = \frac{3 + 2\alpha + \alpha^2}{1 - \alpha^3} - \frac{3 + 1\alpha}{1 - \alpha^2} =$$

$$= \frac{\alpha}{(1 + \alpha)(1 + \alpha + \alpha^2)} > 0, \tag{13}$$

для всех α, заключенных в пределах $0 < \alpha < 1$.

Таким образом, несмотря на то, что на отрезках с индексами $(3 + 6n)$ при $n = 0, 1, 2, \ldots$ вариант B имеет большую сумму совокупной прибыли, чем вариант A, интегральный дисконтированный эффект для варианта A больше при *всех* α, если $0 < \alpha < 1$. Соответственно на основе этого критерия исключается вариант B. Используя аналогичный подход, убедимся в том, что

$$\text{ИДЭ}[A] - \text{ИДЭ}[C] = \frac{2(1 - \alpha)}{1 + \alpha + \alpha^2}, \tag{14}$$

$$\text{ИДЭ}[A] - \text{ИДЭ}[D] = \frac{(1 - \alpha)(1 + 2\alpha)}{3(1 + \alpha + \alpha^2)}. \tag{15}$$

Какой вывод можно сделать о сравнительной эффективности варианта A, с одной стороны, и вариантов C или D — с другой?

Сравнив варианты B и D получим, что

$$\text{ИДЭ}[\text{B}]-\text{ИДЭ}[\text{D}]=-2\frac{2}{3}+\frac{2\alpha+3}{1+\alpha}. \tag{16}$$

Отсюда следует, что вариант B более предпочтителен, чем D, при $\alpha<1$, и менее предпочтителен при $\alpha>{}^1/_2$. Рассмотрим случай $\alpha={}^1/_2$; согласно (16), оба варианта равноценны. Как показано далее, для любой *конечной* длительности планового периода, составляющей n отрезков, интегральный дисконтированный эффект варианта B строго больше, чем эффект варианта D, при нечетных n ($n = 1, 3, 5, \ldots$) и меньше при четных n ($n = 2, 4, 6, \ldots$). Следовательно, равноценность вариантов B и D возникает *только* при бесконечной длительности планового периода.

Пусть $n=2t$, где $t=0, 1, 2, \ldots$ Тогда интегральный дисконтированный эффект для вариантов B и D при конечной длительности планового периода составляет:
при четном числе отрезков n

$$\left.\begin{aligned}&\text{ИДЭ}[\text{B}\,|\,2t]=(3+1\alpha)\frac{1-\alpha^{2t}}{1-\alpha^2}, && t=1, 2, 3, \ldots,\\ &\text{ИДЭ}[\text{D}\,|\,2t]=2\frac{2}{3}+2\alpha\frac{1-\alpha^{2t-1}}{1-\alpha}, && t=1, 2, 3, \ldots,\end{aligned}\right\} \tag{I}$$

при нечетном числе отрезков $(n+1)$

$$\left.\begin{aligned}&\text{ИДЭ}[\text{B}\,|\,2t+1]=\text{ИДЭ}[\text{B}\,|\,2t]+3\alpha^t, && t=0, 1, 2, \ldots,\\ &\text{ИДЭ}[\text{D}\,|\,2t+1]=\text{ИДЭ}[\text{D}\,|\,2t]+2\alpha^{2t}, && t=0, 1, 2, \ldots.\end{aligned}\right\} \tag{II}$$

Соответственно при $\alpha={}^1/_2$ сопоставление вариантов приводит к следующим результатам:

$$\text{ИДЭ}[\text{B}\,|\,2t]-\text{ИДЭ}[\text{D}\,|\,2t]=-\frac{2}{3}\left(\frac{1}{2}\right)^{2t}<0, \tag{III}$$

$$\text{ИДЭ}[\text{B}\,|\,2t+1]-\text{ИДЭ}[\text{D}\,|\,2t+1]=\frac{1}{3}\left(\frac{1}{2}\right)^{2t}>0. \tag{IV}$$

При сравнении вариантов A и G выясняется, что

$$\text{ИДЭ}[\text{A}]-\text{ИДЭ}[\text{G}]=\frac{3+2\alpha+\alpha^2}{1-\alpha^3}-\frac{4}{1-\alpha^3}=\frac{\alpha^2+2\alpha-1}{1-\alpha^3}>0; \tag{17}$$

эта разность больше нуля при $\alpha>\sqrt{2}-1\approx 0{,}414$.

Таким образом, вариант A оказывается более предпочтительным только при $\alpha>\sqrt{2}-1$, в противном случае предпочтителен вариант G. Если процентная ставка очень высока, так что значение α соответственно мало, то получение при варианте G 4 единиц прибыли на первом отрезке оказывает большее влияние, чем получение более поздних выгод при варианте A. В общем случае, чем меньше α, тем больше значение ранних эффектов, а в пределе при $\alpha=0$ оказывает влияние исключительно лишь прибыль первого отрезка R_1.

До сих пор использование критерия интегрального дисконтированного эффекта в отдельных случаях позволяло отыскать совершенно неожиданные различия между последовательностями, однако сравнение всегда давало определенный ответ, поскольку результатом суммирования по формуле (3) всегда было конечное число. Явилось ли это всего лишь следствием удачного подбора численных значений членов последовательностей? Ответ на этот вопрос будет положительным.

Трудностям, возникающим при использовании критерия среднего эффекта за отрезок, отвечают в той или иной мере аналогичные трудности при использовании критерия интегрального дисконтированного эффекта. (Соответствующие примеры приводятся далее.) В то же время при дисконтировании могут и отпасть трудности, возникающие в ряде случаев приложения критерия среднего эффекта. Так, выше [см. (1)] приводилась последовательность B* с возрастающими значениями эффекта, для которой средний эффект за отрезок не является конечным числом. Между тем для этой последовательности интегральный дисконтированный эффект составляет

$$\text{ИДЭ}\,[\mathrm{B}^*] = (3 + 1\alpha + 3\alpha^2 + 1\alpha^3 + 3\alpha^4 + 1\alpha^5) \times \\ \times [1 + 2(\alpha^6) + 3(\alpha^6)^2 + \ldots], \tag{18}$$

т. е.

$$\text{ИДЭ}\,[\mathrm{B}^*] = (3 + 1\alpha)(1 + \alpha^2 + \alpha^4)\frac{1}{(1-\alpha^6)^2}, \tag{19}$$

что при $0 \leqslant \alpha < 1$ есть конечное число. Можно убедиться в том' что и в случае варианта A*, последовательность значений эффектов для которого приведена в (1), интегральный дисконтированный эффект также является конечным числом.

Критерий эквивалентного среднего эффекта. Прежде чем завершить обсуждение методов оценки бесконечных последовательностей значений эффекта, необходимо рассмотреть еще один подход, связывающий понятия средних и дисконтированных значений.

Этот подход основан на идее построения бесконечной последовательности значений эффектов, для которой значение интегрального дисконтированного эффекта является таким же, как для исходной последовательности. Однако во вновь построенной последовательности значения эффектов (до дисконтирования) на всех отрезках одинаковы; это общее значение можно интерпретировать как *эквивалентный средний эффект* данной последовательности. В частности, пусть интегральный дисконтированный эффект для варианта X при некотором фиксированном значении α составляет $P(\alpha)$. Рассмотрим новую последовательность, в которой недисконтированный эффект для любого из отрезков n равен

$$R_n = (1 - \alpha)\,P(\alpha). \tag{20}$$

Для этой последовательности

$$R_1 + R_2\alpha + R_3\alpha^2 + \ldots =$$
$$= (1-\alpha)\,P(\alpha)\,(1+\alpha+\alpha^2+\ldots) = P(\alpha). \qquad (21)$$

Последовательность, соответствующая (20), характеризуется тем же значением интегрального дисконтированного эффекта, что и вариант X, а $(1-\alpha)\,P(\alpha)$ есть эквивалентный средний эффект. При любом фиксированном значении α $(0 \leqslant \alpha < 1)$ использование критерия эквивалентного среднего эффекта приводит к выбору тех же вариантов, что и критерий интегрального дисконтированного эффекта, потому что показатель эквивалентного среднего эффекта попросту есть показатель интегрального эффекта, умноженный на коэффициент $(1-\alpha)$, одинаковый для всех вариантов.

Применив данный подход к приведенным выше примерам, получим следующие значения эквивалентного среднего эффекта, или сокращенно ЭСЭ:

$$\left.\begin{aligned}
&\text{ЭСЭ}\,[A] = \frac{(1-\alpha)(3+2\alpha+1\alpha^2)}{1-\alpha^3} = \frac{3+2\alpha+1\alpha^2}{1+\alpha+\alpha^2},\\
&\text{ЭСЭ}\,[B] = \frac{(1-\alpha)(3+1\alpha)}{1-\alpha^2} = \frac{3+1\alpha}{1+\alpha},\\
&\text{ЭСЭ}\,[C] = \frac{(1-\alpha)(1+6\alpha-1\alpha^2)}{1-\alpha^3} = \frac{1+6\alpha-1\alpha^2}{1+\alpha+\alpha^2},\\
&\text{ЭСЭ}\,[D] = 2\frac{2}{3}(1-\alpha)+2\alpha,\\
&\text{ЭСЭ}\,[G] = \frac{4(1-\alpha)}{1-\alpha^3} = \frac{4}{1+\alpha+\varkappa^2}.
\end{aligned}\right\} \qquad (22)$$

Важно отметить то обстоятельство, что *во всех случаях, когда средний эффект за отрезок определен, его всегда можно получить в формуле эквивалентного среднего эффекта, устремляя значение* α *к 1 снизу.*

Так, приняв в (22) $\alpha = 1$, получим:

$$\text{ЭСЭ}\,[A] = 2; \quad \text{ЭСЭ}\,[B] = 2; \quad \text{ЭСЭ}\,[C] = 2;$$
$$\text{ЭСЭ}\,[D] = 2; \quad \text{ЭСЭ}\,[G] = \frac{4}{3}. \qquad (23)$$

Эти же значения среднего эффекта за отрезок для вариантов A, B, C, D и G были получены ранее. Иногда эквивалентный средний эффект при $\alpha = 1$ бывает определен, тогда как средний эффект за отрезок не определен (пример этого — вариант H, рассматриваемый ниже). К сожалению, эквивалентный средний эффект *не всегда* определен при $\alpha = 1$, примером чего может служить вариант B*.

При использовании критерия эквивалентного среднего эффекта и при $\alpha = 1$ варианты A, B, C и D будут в равной степени предпочтительными. Действительно ли это так? Ответ на этот вопрос

10—0734

зависит от личного мнения, и ответить на него может только лицо, принимающее решение о выборе. Однако можно привести веские доводы в пользу того, что при $\alpha = 1$ **оптимальным** следует считать вариант A, варианты C и D — **близкими к оптимуму**, а вариант B исключить из рассмотрения. В самом деле, рассматривая (13), (14) и (15), можно прийти к выводу, что при значении α, близком к 1, дисконтированный эффект для варианта A больше, чем для вариантов B, C и D. При $\alpha = 1$ из (14) и (15) следует, что разности между ИДЭ [A] и ИДЭ [C], а также ИДЭ [A] и ИДЭ [D] равны нулю, так что можно говорить, что варианты C или D почти столь же хороши, как и вариант A. Что же касается варианта B, то при $\alpha = 1$ из (13) следует, что ИДЭ [A] — ИДЭ [B] $= {}^1/_6$. Таким образом, несмотря на то, что варианты A и B имеют одинаковые эквивалентные средние эффекты, при α, близком к 1, значения интегрального дисконтированного эффекта для этих вариантов различаются на ${}^1/_6$. Именно по этой причине будет целесообразно исключить вариант B.

В целом отметим, что если значение среднего эффекта за отрезок определено, его можно получить на основе эквивалентного среднего эффекта. Критерий последнего типа часто используется как в настоящей главе, так и в последующих главах; однако не следует забывать о том, что он применим для выбора оптимального варианта при бесконечном плановом периоде лишь при сделанных *допущениях*. (В приведенных выше примерах вариант B обеспечивает больший совокупный эффект по состоянию на конец отрезков 3, 9, 15 и т. д.— через каждые 6 отрезков, однако при использовании критерия эквивалентного среднего эффекта это достоинство варианта совершенно не учитывается.) Кроме того, требуются дополнительные допущения, чтобы обеспечить представление интегрального дисконтированного эффекта для каждого варианта в виде единственного и конечного числа. Наконец, следует помнить и о том, что, даже если при $\alpha = 1$ нескольким вариантам соответствует одно и то же значение эквивалентного среднего эффекта, могут все же найтись веские доводы в пользу предпочтительности одного из этих вариантов.

Приведенные выше условные примеры были простыми, трудности возникали только в случае вариантов A* и B*, описанных выражением (1). Однако нетрудно построить другие примеры, показывающие, почему при рассмотрении бесконечных последовательностей все время наталкиваешься на неожиданности и трудности. Приведем несколько таких патологических примеров.

Пусть последовательность значений прибыли для некоторого варианта H имеет следующий вид:

$$4,\ -4,\ 12,\ -12,\ 20,\ -20,\ 28,\ -28,\ \ldots,\ 4n,\ -4n,\ \ldots,$$

где n — нечетные целые числа.

Какая величина, по мнению читателя, является разумной оценкой среднего эффекта за отрезок для этой колеблющейся последовательности? Для более ясного понимания рассмотрим частичные *суммы* конечного числа первых членов последовательности; они имеют вид

4, 0, 12, 0, 20, 0, 28, 0, . . ., $4n$, 0, . . . (n — нечетные целые числа).

Рассмотрим также последовательность значений *среднего* эффекта за отрезок:

$$4,\ 0,\ 4,\ 0,\ 4,\ 0,\ 4,\ 0,\ \ldots,\ 4,\ 0,\ \ldots .$$

Здесь проблема заключается в том, что хотя все средние показатели для конечного планового периода конечны, они колеблются и не сходятся. Следовательно, в отличие от средних показателей, приведенных в таблице рис. 11.2, поведение которых можно оценить, относительно варианта Н нельзя сказать, что средний эффект сколь угодно близок к некоторой постоянной величине, если число отрезков достаточно велико. При любой фиксированной длительности планового периода средний эффект за отрезок равен *либо* 4, *либо* 0. Можно модифицировать критерий среднего эффекта за отрезок, так чтобы он был применим и к подобным колеблющимся последовательностям. Например, можно выбирать вариант по наибольшей верхней границе (или по наименьшей нижней границе) соответствующей последовательности. Для варианта Н числа 4 и 0 есть верхняя и нижняя границы последовательности значений среднего эффекта за отрезок.

Теперь исчислим для варианта Н интегральный дисконтированный эффект:

$$\text{ИДЭ}[\text{Н}] = 4 - 4\alpha + 12\alpha^2 - 12\alpha^3 + 20\alpha^4 - 20\alpha^5 + \ldots = \frac{4(1+\alpha^2)}{(1-\alpha)(1+\alpha)^2}. \tag{I}$$

При $0 \leqslant \alpha < 1$ значение ИДЭ [Н] определено и равно конечному числу. Соответствующее значение эквивалентного среднего эффекта равно

$$\text{ЭСЭ}[\text{Н}] = \frac{4(1+\alpha^2)}{(1+\alpha)^2}, \tag{II}$$

так что при $\alpha = 1$ значение ЭСЭ [Н] определено и равно 2 единицам. Является ли это среднее значение для рассматриваемой последовательности приемлемым?

Рассмотрим последовательность, общий член которой имеет вид

$$R_n = R^{n-1},$$

где $n = 1, 2, 3, \ldots$, а $R > 1$. Интегральный дисконтированный эффект для этой последовательности равен

$$\text{ИДЭ}[\text{R}] = 1 + \alpha R + (\alpha R)^2 + (\alpha R)^3 + \ldots \tag{III}$$

и неограниченно возрастает при $R \geqslant 1/\alpha$. Для того чтобы руководствоваться здесь критерием интегрального дисконтированного эффекта, необходимо сделать допущение, что прибыль возрастает не так быстро, чтобы превысить влияние дисконтирования.

Наконец, предположим, что α равно фиксированному числу a ($\alpha = a$) и что общий член рассматриваемой последовательности имеет следующий вид:

$$R_n = \left(\frac{-1}{a}\right)^{n-1}, \quad n = 1, 2, 3, \ldots . \tag{IV}$$

Тогда, согласно формуле (6), интегральный дисконтированный эффект равен

$$\text{ИДЭ}\left[-\frac{1}{a}\right] = 1 - 1 + 1 - 1 + 1 - 1 +, \ldots . \tag{V}$$

Несомненно, частичные суммы последовательности (V) ограничены, но чему они равны?

Отвечая на этот вопрос, следует остерегаться, казалось бы, правдоподобных доводов, на самом деле приводящих к нелепостям. Для иллюстрации рассмотрим при $\alpha < a$ интегральный дисконтированный эффект, исчисляемый по формуле (V), и выясним, чтó произойдет, когда α возрастет до a. При $\alpha < a$

$$1 - \frac{\alpha}{a} + \left(\frac{\alpha}{a}\right)^2 - \left(\frac{\alpha}{a}\right)^3 + \ldots = \frac{1}{1 + (\alpha/a)}. \tag{VI}$$

Таким образом, возникает искушение утверждать, что сумма последовательности (VI) равна $^1/_2$, поскольку таким является значение правой части (VI) при $\alpha = a$. Ошибка такого «подхода по аналогии» может быть обнаружена при рассмотрении ряда

$$1 - \left(\frac{\alpha}{a}\right)^2 + \left(\frac{\alpha}{a}\right)^3 - \left(\frac{\alpha}{a}\right)^4 + \left(\frac{\alpha}{a}\right)^5 - \ldots = \frac{1 + (\alpha/a)}{1 + (\alpha/a) + (\alpha/a)^2}, \tag{VII}$$

где $\alpha < a$.

Хотя последовательности (VI) и (VII) при $0 < \alpha < a$ различны, левые их части при $\alpha = a$ идентичны. Однако в (VII) значение правой части при $\alpha = a$ равно $^2/_3$.

Имеется ряд способов разрешения проблемы однозначного определения суммы членов последовательности (1), однако при использовании каждого из этих способов необходимо в явном виде высказывать некоторое допущение о том, как интерпретировать рассматриваемый ряд. [Попутно отметим, что в случае большинства обычно делаемых допущений значение суммы последовательности (V) равно $^1/_2$.]

Выводы. На основе обсуждения, проведенного в настоящем разделе, получены следующие выводы о выборе критерия оптимальности в динамических моделях:

I) метод соизмерения последовательностей значений эффекта должен включать то или иное допущение о предпочтительности денежных поступлений и затрат во времени;

II) метод сведения бесконечных последовательностей к некоторому конечному числу не обязательно применим к любой такой последовательности;

III) даже если с помощью некоторого метода две различные последовательности сводятся к одному и тому же числу, соответствующие варианты не обязательно являются одинаково предпочтительными, если учитывать другие соображения экономического порядка.

Здесь читатель может усомниться — насколько реалистичны приведенные выше условные численные примеры. Разумеется, все они «придуманы», но это еще не причина, чтобы они вызывали недоверие. Как читателя могут уверить специалисты по налоговому обложению и по капиталовложениям, каждой фирме приходится принимать важнейшие стратегические решения, которые по существу сводятся к выбору одной из ряда альтернативных бесконечных последовательностей, *характер* которых аналогичен характеру приведенных примеров. В таких ситуациях — с ними обычно сталкиваются те фирмы, которые быстро расширяются или действуют в условиях постепенно возрастающих затрат,— рассмотренные трудности или их модификации возникают во всей полноте.

Для иллюстрации этого положения обратим внимание на некоторые вопросы, встающие уже при бухгалтерском учете движения денежных средств и основных фондов промышленных фирм. Фирма может выбрать один из нескольких методов исчисления амортизационных отчислений для списания стоимости оборудования, оценки текущих запасов и исчисления налоговой скидки ввиду создания балансового резерва в связи с истощением недр. В каждой из этих ситуаций выбор того или иного метода может ежегодно оказывать существенное влияние на размер отчетной прибыли.

Рассмотрим другую ситуацию — фирма должна принять решение о замене основного оборудования. Иногда такую замену можно отсрочить, производя достаточно большие затраты на ремонт. Если требуются существенные капиталовложения, фирма, вероятно, попробует оценить и варианты, связанные с арендой оборудования. Финансирование затрат также можно осуществлять на основе разных сочетаний таких способов, как увеличение нераспределенной прибыли, увеличение кредиторской задолженности, привлечение капитала посредством выпуска акций. Для того чтобы осуществить выбор, необходимо сравнить все эти варианты в свете их влияния на движение денежных средств фирмы и временну́ю последовательность ее прибылей.

В остальной части настоящей главы и в следующей главе мы не будем рассматривать парадоксы, возникающие при оценке бесконечных последовательностей (за исключением разве лишь отдельных кратких упоминаний), а сосредоточим внимание на вопросах использования критериев выбора и на методах отыскания вариантов, оптимальных при заданном критерии.

11.3. МОДЕЛЬ ЭКСПЛУАТАЦИИ ЛЕСНОГО ХОЗЯЙСТВА

Приводимая в настоящем разделе иллюстративная модель выполняет двоякую цель. С одной стороны, идеи, изложенные в предыдущем разделе, используются для решения конкретной оптимизационной задачи. С другой стороны, анализируется структура широко применимой модели; эта структура рассматривается в последующих разделах настоящей главы.

Фирма «Самосад» намечает засадить лесом новый участок земли. Дерево, сваленное в начале отрезка k (k отсчитывается с момента посадки дерева), по оценке фирмы приносит чистую прибыль N_k единиц ($N_k > 0$). Предположим, что затраты на посадку и содержание леса пренебрежимо малы по сравнению с затратами на валку и транспортировку деревьев в начале отрезка k. Далее, предположим, что все деревья должны быть срублены одновременно. Следовательно, N_k представляет собой выручку от продажи леса за вычетом затрат на валку и транспортировку. Пусть N_k поступает в *начале* отрезка k, а $k = 1$ соответствует текущему отрезку, т. е. отрезку, на котором начинаются лесопосадки.

Однократное принятие решения. Для исчисления интегрального дисконтированного эффекта используются коэффициенты дисконтирования за $(k - 1)$ отрезков, равные α^{k-1}. Таким образом, оптимизационная задача состоит в том, чтобы отыскать

$$\max_{k=1,2,3,\ldots,} \alpha^{k-1}N_k. \tag{1}$$

В выражении (1) предполагается, что после валки леса *не будут* произведены новые посадки. Иными словами, сформулированная выше задача включает однократное принятие решения: в какой момент свалить деревья.

Пусть K есть то значение k, которое является решением (1), и пусть последовательность α^0N_1, α^1N_2, α^2N_3 обладает следующим свойством:

$$\alpha^0N_1 \leqslant \alpha^1N_2 \leqslant \ldots \leqslant \alpha^{K-2}N_{K-1} < \alpha^{K-1}N_K \geqslant$$
$$\geqslant \alpha^KN_{K+1} \geqslant \alpha^{K+1}N_{K+2} \geqslant \ldots, \tag{2}$$

так что значение интегрального дисконтированного эффекта возрастает при увеличении k от 1 до K, а при больших k снова убывает. Тогда неравенства $\alpha^{K-2}N_{K-1} < \alpha^{K-1}N_K$ и $\alpha^{K-1}N_K \geqslant \alpha^KN_{K+1}$ сводятся к условиям (3), имеющим более простой вид:

$$\frac{N_{K-1}}{N_K} < \alpha \leqslant \frac{N_K}{N_{K+1}}. \tag{3}$$

Следовательно, значение K можно найти, вычисляя отношения N_k/N_{k+1} — начав с $k = 1$ и завершив расчеты, как только это отношение станет не меньше чем α. [Интерпретируйте неравенства (2) и критерий прекращения расчетов (3) при $\alpha = 1$.]

Пример. Рассмотрим случай, когда

$$N_k = a - b^k, \tag{4}$$

где $0 < b < 1$ и $a \geqslant b$. (Вычертите график значений N_k при $a = 0{,}75$ и $b = 0{,}5$.) Тогда, как можно проверить посредством небольших алгебраических преобразований, неравенства (3) можно упростить, сведя его к

$$b^{K-1} > \frac{a(1-\alpha)}{1-\alpha b} \geqslant b^K. \tag{5}$$

Здесь возникает новая трудность. Пусть $\alpha = 1$, так что $\frac{a(1-\alpha)}{1-\alpha b}$ равно нулю. Тогда не может существовать *конечного* значения K, удовлетворяющего условиям (5), поскольку $b^k > 0$ для *всех* k. Однако b^k оказывается сколь угодно близким к нулю при достаточно больших k. Таким образом, если N_k имеет вид (4), а $\alpha = 1$, в выражении (1) максимизация не имеет смысла. Читатель должен объяснить, почему при этих обстоятельствах не существует конечный максимум (полезно обратиться к графику значений N_k).

Продемонстрируем вид условий (5) на численном примере. Пусть

$$N_k = 0{,}85 - (0{,}78)^k \quad \text{и} \quad \alpha = 0{,}8. \tag{6}$$

Можно убедиться, что условиям (5) соответствует $K = 4$. Обозначим дисконтированный эффект $\alpha^{k-1}N_k$ как ИДЭ $[k]$; тогда (как читатель может проверить хотя бы для $k = 1$ и $k = 2$)

$$\left.\begin{aligned} &\text{ИДЭ}\,[1] = 0{,}070, \quad \text{ИДЭ}\,[2] = 0{,}193, \quad \text{ИДЭ}\,[3] = 0{,}240,\\ &\text{ИДЭ}\,[4] = 0{,}246, \quad \text{ИДЭ}\,[5] = 0{,}230. \end{aligned}\right\} \tag{7}$$

Эти значения будут использованы в последующих разделах.

Бесконечный плановый период. Теперь рассмотрим, что произойдет, если в следующем после валки леса отрезке снова проводятся лесопосадки, причем такой процесс повторяется бесконечное число раз. Если каждый раз валка леса осуществляется в начале k-го отрезка после посадки, интегральный дисконтированный эффект последовательного ряда прибылей в течение бесконечного планового периода равен

$$\left.\begin{aligned} F(k) &= \text{ИДЭ}\,[k]\,(1 + \alpha^k + \alpha^{2k} + \ldots)\\ F(k) &= \text{ИДЭ}\,[k] + \text{ИДЭ}\,[k]\,(\alpha^k + \alpha^{2k} + \ldots)\\ F(k) &= \text{ИДЭ}\,[k] + \alpha^k\text{ИДЭ}\,[k]\,(1 + \alpha^k + \ldots)\\ F(k) &= \text{ИДЭ}\,[k] + \alpha^k F(k), \end{aligned}\right\} \tag{8}$$

так что

$$F(k) = \frac{\text{ИДЭ}\,[k]}{1-\alpha^k}. \tag{9}$$

Оптимальной для бесконечного планового периода является такая стратегия, при которой максимизируется величина $F(k)$.

Обозначим максимальное значение $F(k)$ через F, причем значение F можно найти из (10):

$$F = \max_{k=1,2,3,\ldots} \frac{\text{ИДЭ}[k]}{1-\alpha^k}, \tag{10}$$

поскольку $F(k)$ определяется по формуле (9). Непосредственным следствием (10) является (11):

$$F \geqslant \frac{\text{ИДЭ}[k]}{1-\alpha^k} \quad \text{для всех } k, \tag{11}$$

или, что эквивалентно этому,

$$F \geqslant \alpha^k F + \text{ИДЭ}[k] \quad \text{для всех } k, \tag{12}$$

причем строгое равенство в (11) и (12) соблюдается только для оптимальной стратегии. В свою очередь это указывает на то, что значение F должно удовлетворять условию

$$F = \max_{k=1,2,3,\ldots} (\alpha^k F + \text{ИДЭ}[k]). \tag{13}$$

Пусть из (10) следует, что оптимальным является $k = k'$, и пусть, далее, последовательность ИДЭ $[1]/(1-\alpha)$, ИДЭ $[2]/(1-\alpha^2)$, ... обладает следующим свойством:

$$\frac{\text{ИДЭ}[1]}{1-\alpha} \leqslant \ldots \frac{\text{ИДЭ}[k']}{1-\alpha^{k'}} \geqslant \frac{\text{ИДЭ}[k'+1]}{1-\alpha^{k'+1}} \geqslant \ldots. \tag{14}$$

Тогда на основе неравенств (14) для $(k'-1)$, k' и $(k'+1)$ можно построить условия, аналогичные (3), а именно

$$\frac{(1+\alpha+\ldots+\alpha^{k'-1})N_{k'-1}}{(1+\alpha+\ldots+\alpha^{k'-2})N_{k'}} < \alpha \leqslant \frac{(1+\alpha+\ldots+\alpha^{k'})N_{k'}}{(1+\alpha+\ldots+\alpha^{k'-1})N_{k'+1}}. \tag{15}$$

Значение k' можно найти, вычисляя величину отношения в правой части условий (15) — начав с $k=1$ и прекратив расчеты, как только это отношение станет не меньше чем α. Для каждого k значение этого отношения будет превышать N_k/N_{k+1}, которое исчислялось при нахождении K на основе (3). Следовательно,

$$k' \leqslant K; \tag{16}$$

это означает, что обычно при бесконечном плановом периоде лес валят чаще — но никак не реже, чем при однократном принятии решения о валке.

Если при использовании (15) $\alpha = 1$, то неравенства можно переписать в следующем виде:

$$\left(\frac{N_{k'-1}}{k'-1}\bigg/\frac{N_{k'}}{k'}\right) < 1 \leqslant \left(\frac{N_{k'}}{k'}\bigg/\frac{N_{k'+1}}{k'+1}\right). \tag{17}$$

Это значит, что стратегия, основанная на k', дает наибольший средний эффект за отрезок.

Для численного примера (6) оптимальным является $k' = 2$, и

$$F(1) = 0{,}350, \quad F(2) = 0{,}536, \quad F(3) = 0{,}492,$$
$$F(4) = 0{,}417, \quad F(5) = 0{,}342, \tag{18}$$

где $F(k)$ определяется выражением (9). Таким образом, если при бесконечном плановом периоде ошибочно использовано значение $K = 4$ (взятое из задачи с однократным принятием решения), то будет получена величина, равная всего лишь $\frac{0{,}417}{0{,}536} = 0{,}78$ от оптимального значения интегрального дисконтированного эффекта.

11.4. МОДЕЛЬ ВОССТАНОВЛЕНИЯ С БЕСКОНЕЧНЫМ ЧИСЛОМ ЭТАПОВ

Читателю уже приходилось иметь дело с некоторыми применениями так называемых **моделей восстановления.** Каждый раз, когда проводится валка леса, в процессе происходит восстановление в том смысле, что фирме «Самосад» необходимо снова принимать решение о сроке следующей валки. Другим примером может служить задача замены оборудования, рассмотренная в разд. 10.6, где период восстановления начинался всякий раз, когда заменялось оборудование. Соответственно управляющими переменными по существу являются последовательные интервалы между моментами замены. Вместо того чтобы рассматривать какой-либо конкретный пример более детально, будем анализировать все такие примеры в свете общей модели.

Предположим, что в процессе принятия решений момент восстановления можно выбирать из N альтернативных вариантов, которым присвоены индексы $k = 1, 2, \ldots, N$. Пусть, далее, если вариант k выбран в момент восстановления t, то следующий момент восстановления наступит на отрезке $(t + k)$; кроме того, введем обозначение:

$$\left.\begin{array}{l} R_k = \text{Затраты варианта } k \text{ при их оценке} \\ \quad \text{в начале его периода восстановления.} \end{array}\right\} \tag{1}$$

Отметим, что (1) включает допущение о стационарности: R_k *не* зависит от того, о каком именно периоде восстановления идет речь. Отметим также, что, поскольку R_k интерпретируется как *затраты*, целью оптимизации является достижение *минимума.*

Конечный плановый период. Введем обозначение:

$$\left.\begin{array}{l} f_n = \text{Интегральные дисконтированные затраты для оптималь-} \\ \quad \text{ной стратегии восстановления, при которой один из} \\ \quad \text{альтернативных вариантов должен быть выбран за} \\ \quad n \text{ отрезков до конца планового периода.} \end{array}\right\} \tag{2}$$

Пусть выбран альтернативный вариант k. При таком выборе немедленно возникают затраты R_k; если предположить, что в следующий момент восстановления, т. е. на отрезке $(n - k)$, также принимается оптимальное решение, то в дальнейшем возникают затраты $\alpha^k f_{n-k}$, где коэффициент α^k используется для дисконтирования будущих затрат к настоящему моменту. Следовательно, за n отрезков до конца планового периода оптимальной является такая стратегия, при которой минимизируется сумма членов, имеющих вид $(\alpha^k f_{n-k} + R_k)$; обозначим соответствующее минимальное значение f_n. Приняв $n \geqslant N$, можно представить f_n с помощью рекуррентного соотношения

$$f_n = \min_{k=1,\,2,\,\ldots,\,N} [\alpha^k f_{n-k} + R_k], \quad f_0 \equiv 0, \quad 0 \leqslant \alpha \leqslant 1. \tag{3}$$

(При $n < N$ минимум отыскивается на ограниченной совокупности k, $k = 1, 2, \ldots, n$.) В действительности, если затраты варианта k рассредоточены по всем k отрезкам, то каждое из R_k зависит также

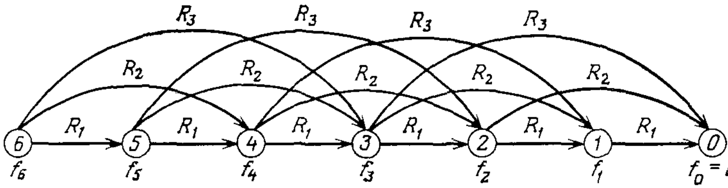


Рис. 11.4. Модель восстановления при конечной длительности планового периода ($n = 6$, $N = 3$, $\alpha = 1$).

и от α; однако мы не будем стремиться отобразить это обстоятельство *в явном виде* и по-прежнему используем сокращенное обозначение R_k вместо $R_k(\alpha)$. На рис. 11.4 приведена сеть, отображающая рекуррентное соотношение (3) при $\alpha = 1$.

Для того чтобы разобраться в операциях, выполняемых на основе соотношения (3), допустим, что для *всех* плановых периодов длительностью n отрезков значение $k = 1$ является оптимальным. Тогда из (3) получаем

$$\begin{aligned} f_n &= \alpha f_{n-1} + R_1 = \alpha\,[\alpha f_{n-2} + R_1] + R_1 = \\ &= \alpha\,[\alpha\,(\alpha f_{n-3} + R_1) + R_1] + R_1 = \\ &= \ldots = R_1 + \alpha R_1 + \ldots + \alpha^{n-1} R_1. \end{aligned} \tag{4}$$

Бесконечный плановый период. Предположим, что плановый период для процесса восстановления является бесконечным. Каждый раз, когда наступает очередной момент восстановления, принятие очередного решения осуществляется в условиях бесконечного планового периода. Можно строго доказать, что существует оптимальная стратегия, которая является стационарной: в каждый из моментов восстановления выбирается *один и тот же* вариант k. Тогда, при $\alpha \neq 1$, соответствующим обобщением (3) является рекуррентное соотношение (5):

$$f = \min_{k=1,\,2,\,\ldots,\,N} [\alpha^k f + R_k], \quad 0 \leqslant \alpha < 1. \tag{5}$$

[Вспомним, что в соотношении (13) разд. 11.3 рассматривался пример аналогичного оптимизационного соотношения — единственным существенным различием является то, что целью оптимизации будет достижение максимума, а не минимума. В разд. 11.3 ИДЭ $[k]$ есть интегральный дисконтированный эффект, который *зависит* от α. Заметим также, что наибольшее допустимое значение k принято априорно равным N.]

Соотношение (5) является примером так называемого **функционального** или **экстремального уравнения.** Неизвестным является значение f, и (5) представляет собой оптимизационное соотношение, которому должно удовлетворять f, если принято, что используется стационарная стратегия. Используя экстремальные уравнения, всегда необходимо определить:

I) имеется ли у данного уравнения конечное решение;

II) если имеется, является ли это решение единственным;

III) если решение единственно, является ли f максимальным значением дисконтированного эффекта для *всех* (не обязательно стационарных) стратегий.

Для того чтобы понять важность этих вопросов, примем $\alpha = 1$, хотя это и противоречит ограничению сверху на значение α, установленному в (5). Если допустить, что все $R_k > 0$, то *ни одно* конечное значение f не удовлетворяет (5); объясните, почему? Однако если вместо этого допустить, что все $R_k = 0$, тогда *любое* конечное значение f удовлетворяет (5); также объясните, почему? Следовательно, функциональное уравнение (5) неприменимо при $\alpha = 1$.

Можно рассматривать (5) как утверждение о том, что f должно удовлетворять условию

$$f \leqslant \alpha^k f + R_k \quad \text{и} \quad f \leqslant \frac{R_k}{1-\alpha^k} \tag{6}$$

для всех k, *причем* (6) должно быть строгим равенством хотя бы для одного значения k. Отсюда следует, что экстремальное уравнение (5) имеет однозначное конечное решение, которое равно

$$f = \min_{k=1,\,2,\ldots,\,N} \left[\frac{R}{1-\alpha^k}\right]. \tag{7}$$

Оптимальная стационарная стратегия соответствует выбору любого альтернативного k, которое позволяет получить оптимальное значение f.

Выражение (7) можно также вывести исходя из соображений стационарности. Поскольку в каждый из очередных моментов восстановления оптимальным является выбор одной и той же альтернативы k, интегральный дисконтированный эффект для этой стратегии равен

$$R_k+\alpha^k R_k+\alpha^{2k}R_k+\alpha^{3k}R_k+\ldots=\frac{R_k}{1-\alpha^k}. \qquad (8)$$

Таким образом, оптимальной является такая стратегия, в которой эта величина минимизируется, как указано в (7).

До сих пор задача с бесконечным плановым периодом решалась в предположении $\alpha \neq 1$, причем было выявлено, что при $\alpha = 1$ соотношение (5) неприменимо. Однако можно распространить анализ и на случай $\alpha = 1$, используя критерий эквивалентного среднего эффекта, предложенный в разд. 11.2. Для этого f заменим g, где $g = (1-\alpha) f$. После простых алгебраических преобразований выражение (7) принимает вид

$$g=\min_{k=1,\,2,\,\ldots,\,N}\left[\frac{R_k}{1+\alpha+\ldots+\alpha^{k-1}}\right]. \qquad (9)$$

Если значение α удовлетворяет условиям $0 \leqslant \alpha < 1$, стратегия, оптимальная согласно (7), является оптимальной и согласно (9), и наоборот. Однако, приняв в (9) $\alpha = 1$, можно убедиться в том, что решением является

$$g=\min_{k=1,\,2,\,\ldots,\,N}\left[\frac{R_k}{k}\right], \quad \alpha=1. \qquad (10)$$

Как и следовало ожидать, альтернатива k является оптимальной, если при ней минимизируются средние затраты за отрезок.

Для того чтобы проверить, как читатель понял изложенный материал, ему предлагается модифицировать (6), (7), (9) и (10) для случая, когда целью оптимизации в выражении (5) является достижение не минимума, а максимума.

Поскольку структура описанной модели восстановления чрезвычайно проста, оптимальную стационарную стратегию легко построить, выполнив расчеты в соответствии с соотношением (7) или (10). Следовательно, для любых практических целей метод решения задачи является тривиальным. Задачи, структура которых имеет более общий вид, нельзя решить столь же легко — для них требуются более сложные методы. В рамках динамического программирования эти методы обычно именуются методами последовательного приближения. Для того чтобы помочь читателю понять основные идеи, лежащие в основе возможных альтернативных подходов, в следующих трех разделах демонстрируется применение таких методов к анализу простой модели восстановления. Это позволяет установить, что для применения подобных методов к задачам более общего вида требуется только изменить систему обозначений.

Одним из возможных направлений использования соотношения (10) является такой важный частный случай, как «модель управления производством и запасами с вогнутой функцией затрат», рассмотренная в разд. 9.5. Пусть функции затрат на производство продукции и содержание запасов являются стационарными и имеют следующий вид:

$$C_t(x_t) = \begin{cases} s + cx_t & \text{при } x_t > 0 \quad \text{(затраты на переналадку и пропорциональные затраты),} \\ 0 & \text{при } x_t = 0, \end{cases} \tag{I}$$

$$h_t(i_t) = hi_t \quad \text{для всех } t \quad \text{(затраты, на содержание запасов линейны),}$$

где $s \geqslant 0$, $c \geqslant 0$ и $h > 0$; x_t — объем выпуска продукции на отрезке t или размер заказа в начале отрезка t, а i_t — уровень запасов на конец отрезка t. Предположим, далее, что спрос является стационарным, так что $D_t = D$ для всех t, и что спрос должен быть полностью и своевременно удовлетворен.

Обозначим Q объем выпуска в момент восстановления, если исходный запас на начало планового периода равен нулю. При конечном плановом периоде, как читателю известно из разд. 9.5,

$$Q = kD, \tag{II}$$

где k — положительное целое число. Предположим, что такой же характер решения сохраняется и для бесконечного планового периода. Тогда оптимизационная задача заключается в том, чтобы найти множитель k, позволяющий минимизировать средние затраты за отрезок.

Заметим, что R_k равно сумме затрат на переналадку, производство продукции и содержание запасов:

$$R_k = s + ckD + hD\,[(k-1) + \ldots + 2 + 1], \tag{III}$$

т. е.

$$R_k = s + ckD + hD\,\frac{(k-1)\,k}{2}. \tag{IV}$$

Следовательно, средние затраты за отрезок составляют

$$\frac{R_k}{k} = \frac{s}{k} + cD + \frac{hD\,(k-1)}{2}. \tag{V}$$

Найти приближенное значение k можно, дифференцируя R_k/k по k и приравняв производную нулю. Отсюда

$$k = \sqrt{\frac{2s}{hD}}, \tag{VI}$$

так что

$$Q = kD = \sqrt{\frac{2sD}{h}}. \tag{VII}$$

Это выражение, определяющее размер Q, часто называют *формулой экономически выгодного размера партии*. Отметим, что (VI) и (VII) являются приближенными потому, что получаемые величины в общем случае не обязательно целочисленны. Вместе с тем, если найденное по формуле (VI) значение k не есть целое число, оптимальную стратегию можно найти, вычисляя средние затраты за отрезок по формуле (V) для двух ближайших к k целых чисел — одного большего и одного меньшего, а затем выбирая лучшее из вычисленных значений.

11.5. МЕТОДЫ ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ

В настоящем разделе начинается рассмотрение численных методов решения экстремальных уравнений, возникающих в моделях динамического программирования с бесконечным плановым периодом. В качестве исходного прототипа используем функциональное уравнение предыдущего раздела:

$$f = \min_{k=1,\,2,\,\ldots,\,N} [\alpha^k f + R_k], \quad 0 \leqslant \alpha < 1. \tag{1}$$

Напомним, что отыскание решения (1) по существу означает отыскание такого значения f, которое удовлетворяет (1); кроме того, хотелось бы найти такое k, которое позволяет достичь этого значения f. Обычно предлагается три подхода к решению задачи.

Первый подход основан на динамическом характере рассматриваемой модели. Он заключается в выяснении — не будет ли стратегия, оптимальная для очень длительного, но конечного планового периода, одновременно и оптимальным решением для бесконечного периода. Второй подход состоит в том, чтобы попытаться угадать некоторое *значение* f. Затем на основе этого значения вычисляется величина в правой части (1) и проверяется выполнение уравнения. Если уравнение не выполняется, результат вычислений рассматривается в качестве скорректированного значения f, и процесс повторяется снова. Третий подход заключается в попытках отгадать *стратегию*, оптимальную для бесконечного периода. Для предполагаемой оптимальной стратегии вычисляется соответствующий интегральный дисконтированный эффект, который и используется в качестве проверяемого значения f; затем производится проверка справедливости уравнения (1) при этом значении. Если уравнение не удовлетворяется, в качестве новой пробы используется та выбранная программа, при которой достигается наименьшее значение правой части (1), и процесс повторяется снова. В настоящем разделе рассматривается первый подход, а два остальных подхода — в следующих разделах.

Во всех этих методах каждую догадку можно толковать как некоторое приближение к решению. Если она удовлетворяет экстремальному уравнению, решение найдено, если нет — следует попытаться снова. Такой итеративный процесс именуется **методом последовательных приближений.**

Бесконечный период как предел конечного периода. Вероятно, наиболее очевидный подход к отысканию такой стратегии, которая удовлетворяет функциональному уравнению (1), заключается в нахождении решения для модели с конечным плановым периодом:

$$f_n = \min_{k=1,\,2,\,\ldots,\,N} [\alpha^k f_{n-k} + R_k], \quad 0 \leqslant \alpha < 1, \tag{2}$$

где n принимает очень большие значения. Примеры, рассмотренные в гл. 8 и 9, заставлют нас проявить особую осторожность в отношении того, можно ли в данном конкретном случае полагаться на применимость такого подхода. Можно ли быть уверенным в том, что при *любом* достаточно большом n получаемое из (2) k_n будет также удовлетворять уравнению (1)? Остается ли одно и то же k оптимальным при переходе ко все большим значениям n? Если длительность планового периода n достаточно велика, является ли стратегия, оптимальная для бесконечного периода, оптимальной и для конечного периода? Очень важно, чтобы в случае рассматриваемой модели восстановления ответы на эти вопросы были положительными.

Теорема о длительности планового периода для модели восстановления. Существует такое конечное значение n^*, что для любого конечного планового периода длительностью n отрезков, $n > n^*$, выполняются следующие соотношения:

$$\text{если } f_n = \alpha^{k_n} f_{n-k_n} + R_{k_n}, \quad \text{то } f = \alpha^{k_n} f + R_{k_n}, \tag{3}$$

$$\text{если } f = \alpha^K f + R_K, \quad \text{то } f_n = \alpha^K f_{n-K} + R_K. \tag{4}$$

Из (3) следует, что любая стратегия k_n, являющаяся оптимальным текущим решением при достаточно большой длительности n планового периода ($n > n^*$), одновременно есть оптимальная стационарная стратегия для бесконечного планового периода. Напротив, (4) является обратным утверждением. Выполнив по определенной схеме расчеты в соответствии с (2), можно найти значение n^*. Подробности этого подхода не имеют отношения к цели настоящего обсуждения и потому опускаются.

11.6. МЕТОД ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ В ПРОСТРАНСТВЕ ФУНКЦИЙ (МЕТОД ИТЕРАЦИЙ ПО КРИТЕРИЮ)

Основная идея в предыдущем методе была связана с нахождением оптимальной стационарной стратегии N для бесконечного планового периода посредством анализа ряда возрастающих значений n. Напротив, идея описываемого ниже метода состоит в последовательном приближении к *значению функции* f в экстремальном уравнении. Соответственно процесс именуется **методом итераций по критерию.**

Пусть f^0 — первоначально выбранное значение f. Метод заключается в построении последовательности приближений f^1, f^2, f^3, ... на основе рекуррентного соотношения

$$f^{n+1} = \min_{k=1,\,2,\,\ldots,\,N} [\alpha^k f^n + R_k], \quad 0 \leqslant \alpha < 1, \tag{1}$$

где f^n — *пробное значение* f на итерации n. [Если целью оптимизации в экстремальном уравнении является достижение максимума, то в выражение (1) вносятся соответствующие изменения.] Ниже приводится пример решения задачи с помощью описываемого метода.

Хотя алгоритм (1) четко сформулирован, возникает три вопроса относительно его применимости:

I) Всегда ли значение f^n стремится к тому f, которое удовлетворяет экстремальному уравнению?

II) Если всегда, то существует ли такое *конечное* n, для которого f^n равно f?

III) Если в соответствии с (1) на двух итерациях подряд выбрано одно и то же k, то является ли оно оптимальным?

Для того чтобы ответить на все эти вопросы, предположим, что все $R_k > 0$. Если принять $f^0 = 0$, то можно доказать, что $f^{n+1} > f^n$, так что совокупность f^n представит собой **монотонно возрастающую последовательность приближений.** Тогда, при достаточно большом n, f^n явится сколь угодно близким к оптимальному значению f. Однако в общем случае *не существует конечного* n, такого, что f^n равно f. Более того, некоторая альтернатива может быть выбрана в качестве значения правой части (1) на двух последовательных итерациях, однако она не обязательно является оптимальной для бесконечного планового периода. (Если решение для модели восстановления отыскивается методом приближения в пространстве функций, то мы никогда не возвращаемся к ранее отвергнутой стратегии.)

Пример. Приводимый пример показывает действие метода приближений при $R_k > 0$. Пусть $N = 5$ и

$$R_1 = 8{,}7;\ R_2 = 12{,}7;\ R_3 = 14{,}7;\ R_4 = 19{,}7;\ R_5 = 28{,}7;\ \alpha = 0{,}8. \tag{2}$$

Можно определить, что решением является

$$f = \min_{k=1,\,2,\,\ldots,\,5}\left[\frac{R_k}{1-0{,}8^k}\right] = \min\,[43{,}50;\, 35{,}28;\, 30{,}00;\, 33{,}39;\, 42{,}84] = 30{,}00, \tag{3}$$

так что $k = 3$ является оптимальным значением.

Вычисления на основе рекуррентного соотношения (1) в пространстве функций дают при $n = 1,\ 2,\ 3$ и $f^0 = 0$:

$$\left.\begin{aligned}
f^1 &= \min_{k=1,\,2,\,\ldots,\,5}[\alpha^k 0 + R_k] = 8{,}7, \quad k = 1;\\
f^2 &= \min\,[0{,}8\cdot 8{,}7 + 8{,}7;\ 0{,}64\cdot 8{,}7 + 12{,}7;\ 0{,}51\cdot 8{,}7 + 14{,}7;\\
&\qquad 0{,}41\cdot 8{,}7 + 19{,}7;\ 0{,}33\cdot 8{,}7 + 28{,}7] = 15{,}66 \text{ для } k = 1;\\
f^3 &= \min\,[0{,}8\cdot 15{,}66 + 8{,}7;\ 0{,}64\cdot 15{,}66 + 12{,}7;\ 0{,}51\cdot 15{,}66 +\\
&\qquad + 14{,}7;\ 0{,}41\cdot 15{,}66 + 19{,}7;\ 0{,}33\cdot 15{,}66 + 28{,}7] =\\
&\qquad = 21{,}23 \text{ для } k = 1.
\end{aligned}\right\} \tag{4}$$

Читателю следует проверить отдельные слагаемые в (4), для того чтобы убедиться, что он полностью понимает сущность вычислительного процесса. Для итераций n, где $n > 3$,

$$\begin{gathered} f^4 = 25{,}53; \quad f^5 = 27{,}57; \quad f^6 = 28{,}76; \quad f^7 = 29{,}37; \\ f^8 = 29{,}68; \quad f^9 = 29{,}84; \quad f^{10} = 29{,}91; \quad f^{11} = 29{,}95; \\ f^{12} = 29{,}97; \quad f^{13} = 29{,}98; \quad f^{14} = 29{,}99; \quad f^{15} = 29{,}99, \end{gathered} \tag{5}$$

причем на всех итерациях $k = 3$. (Вычислите f^{16} и убедитесь, что по-прежнему $k = 3$.)

Замечания. На основе вычислений (4) по материалам приведенного примера видно, что некоторая стратегия ($k = 1$) может быть выбрана на нескольких последовательных итерациях, но при этом она не будет являться оптимальным решением для бесконечного планового периода. Можно так изменить условия примера, что $k = 1$ выбирается в течение сколь угодно большого числа итераций; для этого примем R_1 достаточно близким к 6. Расчеты (5) показывают, что f^n быстро сходится к f, однако f^n *не равно* f ни для какого конечного n.

Заметим, что при $\alpha = 1$ процесс прекращается. Для *каждого* n выбирается такое k, для которого R_k минимально. Однако это значение k обычно не совпадает с решением, минимизирующим средние затраты за отрезок R_k/k.

Метод итераций по стоимости, описываемый уравнением (1), фактически применим при *любых* значениях R_k и *любом* начальном приближении f^0. Однако в общем случае последовательность значений f^n не всегда является монотонной. Альтернативный подход к выбору f^0 *всегда* приводит к **монотонно убывающей последовательности** приближенных значений, т. е. к обязательному выполнению неравенства $f^{n+1} \leqslant f^n$. Идея состоит в том, чтобы угадать оптимальную *стратегию*; пусть f^0 — соответствующие интегральные дисконтированные затраты для этой стратегии.

Если угаданная стратегия оказалась оптимальной при вычислении f^1, то $f^0 = f^1 = f$. Однако если новая стратегия при вычислении f^1 оказалась определенно лучше, то расчеты на основе рекурентного соотношения (1) продолжаются и $f^{n+1} < f^n$. Этот метод описывается ниже.

Сходимость сверху. Рассмотрим пример (2); предположим, что исходной догадкой является $k = 1$. Тогда

$$f^0 = \frac{R}{1-\alpha^1} = 43{,}50, \tag{6}$$

а для $n = 1$

$$\left.\begin{aligned} f^1 = \min\,[&0{,}8 \cdot 43{,}50 + 8{,}7;\ 0{,}64 \cdot 43{,}50 + 12{,}7; \\ &0{,}51 \cdot 43{,}50 + 14{,}7;\ 0{,}41 \cdot 43{,}50 + 19{,}7; \\ &0{,}33 \cdot 43{,}50 + 28{,}7] = 36{,}88 \text{ для } k = 3. \end{aligned}\right\} \tag{7}$$

$$\left.\begin{array}{l}\text{Для последующих итераций } (n > 1)\\ f^2 = 33{,}50,\ f^3 = 31{,}78,\ f^4 = 30{,}90,\ f^5 = 30{,}45,\ f^6 = 30{,}22,\\ f^7 = 30{,}11,\ f^8 = 30{,}05,\ f^9 = 30{,}02,\ f^{10} = 30{,}01,\end{array}\right\} \quad (8)$$

причем на всех этих итерациях $k = 3$. Можно изменить условия примера, приняв R_4 близким к 17,7, так что для сколь угодно большого числа приближений будет выбираться $k = 4$. Проверьте, что если вы начнете вычислительный процесс с $k = 3$, то $f = 30{,}00$ для $k = 3$.

При использовании рекуррентного соотношения (1) f^0 принималось равным нулю лишь ввиду удобства вычислений. Словесное описание процесса последовательных приближений, начатого в этой исходной точке, не приводит к более глубокому пониманию. Однако, если в качестве f^0 принимаются интегральные дисконтированные затраты для выбранной вначале *стратегии*, идея метода проясняется. Значение f^0 в (6) представляет собой интегральные дисконтированные затраты, соответствующие выбору варианта $k = 1$ при бесконечно большом плановом периоде. Предположим, что вместо $k = 1$ первоначально выбирается $k = 3$, а после этого k всегда принимается равным 1. Интегральные дисконтированные затраты для такой стратегии отображены f^1 в (7). Аналогично этому f^2 в (8) по существу представляет собой интегральные дисконтированные затраты для стратегии, в которой первым двум решениям о восстановлении соответствует $k = 3$, а затем принимается $k = 1$. Эти замечания подсказывают нам еще один метод последовательных приближений, рассматриваемый в следующем разделе.

Что́ означает утверждение «метод итераций по критерию на основе рекуррентного соотношения (1) всегда применим»? Необходимо показать, что выполняются три условия. Два из них уже упоминались, а именно то, что экстремальное уравнение имеет конечное решение и что это решение является единственным. Поскольку модель восстановления имеет столь простую структуру, соображения о существовании и об единственности, приведенные в разд. 11.4, являются достаточными. В случае моделей более общего вида эти два момента необходимо доказывать с большей тщательностью.

Третье условие состоит в том, что последовательность f^n сходится к истинному решению при *любых* R_k и исходном приближении f^0. Здесь учитываются следующие соображения. Пусть K есть оптимальная стационарная стратегия для бесконечного планового периода, а k — вариант, выбранный на *данной* итерации n в соответствии с рекуррентным соотношением (1). Тогда

$$f = \alpha^K f + R_K, \tag{I}$$

поскольку K является оптимальным для бесконечного периода, и

$$f^n \leqslant \alpha^K f^{n-1} + R_K, \tag{II}$$

поскольку при k возникает f^n. Таким образом,

$$f^n - f \leqslant \alpha^K (f^{n-1} - f) \leqslant \alpha \mid f^{n-1} - f \mid, \tag{III}$$

так как $0 \leqslant \alpha < 1$.

Аналогично этому

$$f^n = \alpha^k f^{n-1} + R_k, \tag{IV}$$

$$f \leqslant \alpha^k f + R_k, \tag{V}$$

так что

$$f - f^n \leqslant \alpha^k (f - f^{n-1}) \leqslant \alpha \mid f^{n-1} - f \mid. \tag{VI}$$

Совмещая (III) и (VI), получим

$$\mid f^n - f \mid \leqslant \alpha \mid f^{n-1} - f \mid \leqslant \alpha^n \mid f^0 - f \mid. \tag{VII}$$

Поскольку $\alpha < 1$, при достаточно больших n правая часть (VII) является сколь угодно малой. Поэтому f^n в пределе стремится к f. Отметим, что скорость сходимости не меньше, чем у экспоненты.

11.7. МЕТОД ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ В ПРОСТРАНСТВЕ СТРАТЕГИЙ (МЕТОД ИТЕРАЦИЙ ПО СТРАТЕГИЯМ)

Предположим, что при вычислении значения правой части рекуррентного соотношения (1) предыдущего раздела обнаруживается стратегия, определенно лучшая, чем та, которая соответствует f^n. Это означает, что выбор найденной стратегии в случае *непосредственно принимаемого решения* является улучшением по сравнению с выбором ранее рассматривавшейся стратегии. Тогда весьма вероятно — и это действительно оказывается правильным,— что использование новой стратегии в течение всего бесконечного планового периода даст даже лучшие результаты, чем выбор этой стратегии только *в случае непосредственно принимаемого решения*. Следовательно, f^{n+1} можно вычислить как интегральные дисконтированные затраты при многократном повторном выборе новой стратегии. Этот процесс известен под названием **приближения в пространстве стратегий**, или попросту **метода итераций по стратегиям**, поскольку на каждой итерации рассматривается новая проверяемая стационарная стратегия для бесконечного планового периода.

Возникающая последовательность f^n является монотонно убывающей, с n на каждой итерации происходит определенное улучшение; следовательно, мы никогда не возвращаемся к однажды отвергнутой стратегии. Поскольку имеется конечное число N различных стационарных стратегий, расчеты при данном подходе должны завершаться за *конечное* число итераций. Как только какая-либо стратегия остается оптимальной в течение двух итераций подряд, вычисления можно прекратить, причем f^n равно оптимальному f, удовлетворяющему экстремальному уравнению. Как увидит читатель, за получение конечного алгоритма приходится платить увеличением объема расчетов по определению f^{n+1} для новой стратегии на каждой итерации.

Алгоритм имеет следующий вид:

Шаг 1. Выберем произвольную исходную стратегию и примем $n = 0$.

Шаг 2. Для заданной пробной стратегии k вычислим интегральные дисконтированные затраты в течение бесконечного планового периода:

$$f^n = \frac{R_k}{1-\alpha^k}. \qquad (1)$$

Шаг 3. Проверим возможность дальнейших улучшений, вычислив

$$\min_{k=1,\,2,\,\ldots,\,N} [\alpha^k f^n + R_k] = \alpha^{k'} f^n + R_{k'}, \qquad (2)$$

т. е. выбрав k'.

Шаг 4. Прекратим расчеты, если $\alpha^{k'} f^n + R_{k'} = f^n$. В противном случае изменим стратегию на k'. Перейдем от итерации n к итерации $(n + 1)$ и обратимся к выполнению шага 2 на основе новой проверяемой стратегии.

Отметим, что если хорошие пробные значения f получаются непосредственно в процессе аппроксимации в пространстве функций, то при описываемом методе эти значения должны исчисляться дополнительно, по формуле (1). Отметим также, что условие прекращения расчетов на шаге 4 выполняется в том случае, когда k' — тот же вариант, который рассматривался на шаге 2. Иными словами, расчеты прекращаются, если k' не изменяется в течение двух итераций подряд. [Как нужно преобразовать (2), если целью оптимизации на основе экстремального уравнения является максимизация?]

Пример. Для того чтобы продемонстрировать описанный подход, примем

$$R_1 = 8{,}7, \quad R_2 = 12{,}7, \quad R_3 = 14{,}7, \quad R_4 = 19{,}7, \quad R_5 = 28{,}7; \qquad (3)$$

это условия примера (2) в предыдущем разделе.

Как и ранее, пусть исходной предполагаемой стратегией является $k = 1$, так что $f^0 = 43{,}50$. Контрольные расчеты по формуле (2) в точности повторяют расчеты (7) предыдущего раздела. Таким образом, на основе формулы (1) теперь получаем $f^1 = 30{,}00$ для $k = 3$. Вторичное применение проверочной формулы (2) приводит к вычислению:

$$\min\,(0{,}8 \cdot 30{,}00 + 8{,}7;\ 0{,}64 \cdot 30{,}00 + 12{,}7;\ 0{,}51 \cdot 30{,}00 + 14{,}7;$$
$$0{,}41 \cdot 30{,}00 + 19{,}7;\ 0{,}33 \cdot 30{,}00 + 28{,}7) = 30{,}00 \quad \text{для} \quad k' = 3, \qquad (4)$$

так что расчеты прекращаются.

Средний эффект за отрезок. Как обычно, при разработке соответствующего метода для $\alpha = 1$ удобно переформулировать вычислительную процедуру на основе использования эквивалентного среднего эффекта. Аналогом формулы (1) для $\alpha = 1$ является

$$g^n = \frac{R_{k'}}{1+\alpha+\ldots+\alpha^{k'-1}} = \frac{R_{k'}}{k'}. \tag{5}$$

Отметим, что если R_k зависит от α, то значение R_k при $\alpha = 1$ используется для исчисления отношения в правой части (5).

Выражение в левой части проверочного соотношения (2) принимает вид

$$\min_{k=1,\,2,\,\ldots,\,N} [\alpha^k g^n + (1-\alpha) R_k], \tag{6}$$

однако при $\alpha = 1$ выражение в квадратных скобках не зависит от k. Для того чтобы устранить этот недостаток, обратим внимание на следующее обстоятельство. Если при k' достигается минимум функции $g(k)$, то при k' минимизируется и $ag(k) + b$, если $a > 0$. Пусть $g(k)$ — выражение в квадратных скобках (6); $a = (1-\alpha)^{-1}$, $b = -g^n(1-\alpha)^{-1}$. Проделаем необходимые преобразования в (6) и установим, что применим следующий оптимизационный выбор:

$$\min_{k=1,\,2,\ldots,\,N} [-(1+\alpha+\ldots+\alpha^{k-1}) g^n + R_k] =$$
$$= \min_{k=1,\,2,\,\ldots,\,N} [-kg^n + R_k] = -k'g^n + R_{k'}, \quad \alpha = 1, \tag{7}$$

который аналогичен проверочному соотношению (2).

Здесь также приходится использовать значение R_k, соответствующее $\alpha = 1$, если в действительности R_k зависит от α. Читатель должен истолковать критерий (7) при $\alpha = 1$.

В целом метод можно обрисовать следующим образом:

Шаг 1. Выберем произвольную исходную стратегию и примем $n=0$.

Шаг 2. Для заданной пробной стратегии k вычислим соответствующие средние затраты за отрезок:

$$g^n = \frac{R_k}{k}. \tag{8}$$

Шаг 3. Проверим возможность дальнейших улучшений, вычислив

$$\min_{k=1,\,2,\,\ldots,\,N} [-kg^n + R_k] = k'g^n + R_{k'}, \tag{9}$$

т. е. выбрав k'.

Шаг 4. Прекратим расчеты, если $-k'g^n + R_{k'} = 0$. В противном случае изменим стратегию на k'. Перейдем от итерации n к итерации $(n+1)$ и обратимся к выполнению шага 2 на основе новой пробной стратегии.

Для того чтобы понять действие описываемого метода, применим его к примеру (3). Вычисления выполняются в следующей последовательности:

$g^0 = R_1/1 = 8{,}7$ **для** $k = 1$ в качестве исходной программы; (I)

$$\min(-1\cdot 8{,}7 + 8{,}7;\ \ -2\cdot 8{,}7 + 12{,}7;\ \ -3\cdot 8{,}7 + 14{,}7;\ \ -4\cdot 8{,}7 + 19{,}7;\ \ -5\cdot 8{,}7 + 28{,}7) = -15{,}1 \quad \text{для} \quad k' = 4; \tag{II}$$

$$g^1 = R_4/4 = -4{,}925; \tag{III}$$

$$\min_{k=1,2,\ldots,5} (-k\cdot 4{,}925 + R_k) = -0{,}025 \quad \text{для} \quad k' = 3; \tag{IV}$$

$$g^2 = R_3/3 = 4{,}9; \tag{V}$$

$$\min_{k=1,2,\ldots,5} (-k\cdot 4{,}9 + R_k) = 0 \quad \text{для} \quad k' = 3. \tag{VI}$$

Заметим, что один и тот же вариант $k' = 3$ выбирается на двух итерациях подряд; это обусловливает прекращение расчетов на шаге 4, и $k' = 3$ является оптимальным.

Возможно, читателю на самом деле не безразличны два варианта, которые представляются одинаково предпочтительными в соответствии с проверочным соотношением (9). Так, если добавить еще одну альтернативу $k = 6$ с затратами $R_6 = 2R_3 = 29{,}4$ единицы, то для этой альтернативы значение g будет тем же при $\alpha = 1$, что и для альтернативы $k = 3$. Однако при $\alpha < 1$

$$\frac{R_6}{1-\alpha^6} - \frac{R_3}{1-\alpha^3} = \frac{R_3}{1-\alpha^3} > \frac{R_3}{2}.$$

Таким образом, при $0 \leqslant \alpha < 1$ интегральный дисконтированный эффект для $k = 6$ всегда больше, чем для $k = 3$. Следовательно, даже если в соответствии с проверочным соотношением (9) $k = 3$ является оптимальным, действительно предпочтительным может быть $k = 6$.

11.8. ЭКВИВАЛЕНТНАЯ ЗАДАЧА ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

В разд. 10.7 рассматривалась связь между рекуррентными соотношениями динамического программирования и линейным программированием. Вкратце можно отметить, что значения, получаемые на основе динамического программирования, соответствуют длинам кратчайших маршрутов к вершине конечного состояния от остальных вершин ациклической сети. В моделях с бесконечным плановым периодом аналогия с сетью остается в силе, за тем лишь исключением, что оптимальная стационарная стратегия обусловливает бесконечное число прохождений через соответствующий набор дуг. Сетевая интерпретация задачи приведена в гл. 12, однако в настоящем разделе предварительно рассматриваются прямая и двойственная задачи линейного программирования для простой задачи восстановления с бесконечным числом этапов.

Функциональное уравнение

$$f = \min_{k=1,2,\ldots,N} [\alpha^k f + R_k], \quad 0 \leqslant \alpha < 1 \tag{1}$$

можно представить в виде следующей задачи:

$$\text{Максимизировать } f \tag{2}$$

при условиях

$$f \leqslant \alpha^k f + R_k, \quad k = 1,\ 2,\ \ldots,\ N, \tag{3}$$

где на знак f не налагается ограничений. Преобразовав (3), получим ограничения

$$f \leqslant \frac{R_k}{1-\alpha^k}, \quad k = 1,\ 2,\ \ldots,\ N. \tag{4}$$

Задачу удобно преобразовать, построив ее целевую функцию на основе эквивалентного среднего эффекта. Для этой цели, как обычно, примем $g = (1 - \alpha) f$. Тогда задаче (2) и (4) эквивалентна следующая оптимизационная задача:

$$\text{Максимизировать } g \tag{5}$$

при условиях

$$g \leqslant \frac{R_k}{1+\alpha+\ldots+\alpha^{k-1}}, \quad k = 1,\ 2,\ \ldots,\ N, \tag{6}$$

где на знак g не налагается ограничений. Выражение (5) есть целевая функция, а (6) — ограничения двойственной задачи.

Если задача (5) — (6) именуется двойственной, то прямая задача формулируется следующим образом:

$$\text{Минимизировать } \sum_{k=1}^{N} \left(\frac{R_k}{1+\alpha+\ldots+\alpha^{k-1}}\right) x_k \tag{7}$$

при условии

$$\sum_{k=1}^{N} x_k = 1, \tag{8}$$

где $x_k \geqslant 0$. Выражение (7) есть целевая функция, а (8) — ограничение прямой задачи. Отметим, что при $\alpha = 1$ целевая функция принимает такой вид:

$$\text{Минимизировать } \sum_{k=1}^{N} \left(\frac{R_k}{k}\right) x_k. \tag{9}$$

Поскольку в прямой задаче имеется только одно ограничение (8), оптимальное базисное решение включает только одно $x_k = 1$, которое и соответствует лучшему варианту восстановления K; все остальные $x_k = 0$.

11.9. ПОВТОРНОЕ РАССМОТРЕНИЕ ЗАДАЧИ НАХОЖДЕНИЯ КРАТЧАЙШЕГО ПУТИ

В гл. 12 рассматриваются более общие динамические модели, чем задача восстановления, приведенная в предыдущих разделах. Для того чтобы отыскать стратегию, оптимальную для бесконечного планового периода, применяются методы последовательных приближений, описанные в разд. 11.6 и 11.7; это описание должно подготовить читателя к восприятию последующего материала. В качестве дополнительной подготовки читателю полезно снова рассмотреть задачу нахождения кратчайшего пути на сети общего вида.

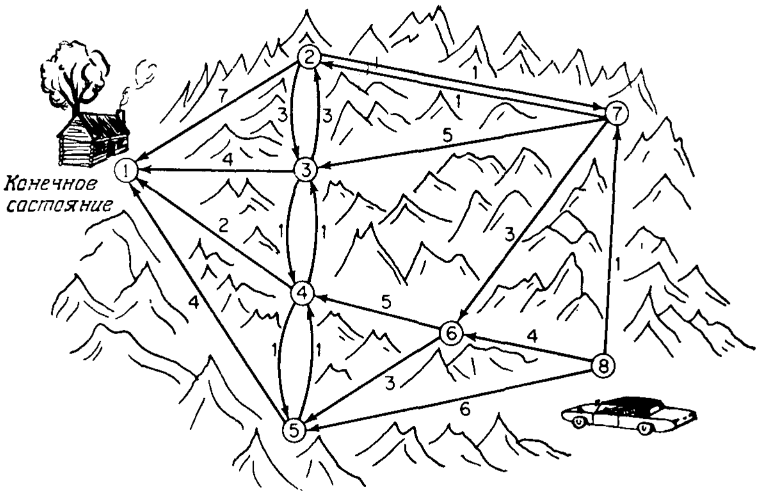


Рис. 11.5. Задача нахождения кратчайшего пути.

Здесь эта задача обобщается с целью включения дисконтирования в рассмотрение проблемы, а также повторного анализа алгоритмов, обращающего особое внимание на изучение идеи последовательных приближений.

Как и ранее, сеть включает p вершин и систему дуг; (i, j) обозначает дугу из вершины i в вершину j, а c_{ij} — затраты на перемещение по дуге (i, j) из вершины i в j. Можно считать, что c_{ij} — интегральные дисконтированные затраты, зависящие от коэффициента дисконтирования α, однако в индексах такая зависимость в явном виде не показана. Вершина конечного состояния обозначена индексом r.

Так, например, рассмотрим сеть рис. 11.5, в точности повторяющую сеть рис. 7.16 (стр. 292). Вершина конечного состояния обозначена индексом *1*. Примем следующую ориентацию во времени для маршрута от вершины k к вершине *1*: вершина k отображает текущее состояние, а вершина *1* — состояние в конце последнего шага. Например, маршрут, проходящий через дуги (7, 2), (2, 3), (3, 4) и (4, 1), отображает текущее решение выехать из вершины *7* в вершину *2*, последующее решение выехать из вершины *2* в вершину *3* и т. д.

Предположим, что для перемещения по каждой дуге требуется один отрезок планового периода. (Более общий случай, при котором для переезда по отдельным дугам требуется более одного отрезка, рассматривается далее.) Пусть коэффициент дисконтирования за один отрезок равен α, причем $0 \leqslant \alpha \leqslant 1$. Обобщим определение y_i, которое в разд. 7.6 обозначало длину кратчайшего пути из вершины i в конечную вершину r. Примем здесь, что

y_i — интегральные дисконтированные затраты, соответствующие перемещению по оптимальному маршруту от вершины i до вершины конечного состояния r, (1)

где «оптимальный» означает маршрут, интегральные дисконтированные затраты для которого минимальны. Таким образом, если первый шаг оптимального маршрута из вершины i в вершину r направлен в вершину j, то

$$y_i = \alpha y_j + c_{ij} \tag{2}$$

и

$$y_i \leqslant \alpha y_k + c_{ik} \quad \text{для всех } k. \tag{3}$$

[Дайте истолкование выражениям (2) и (3).] Поскольку условия (2) и (3) должны выполняться для всех вершин i, отличающихся от конечной вершины r, y_i должны удовлетворять системе экстремальных уравнений

$$y_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha y_j + c_{ij}) \quad \text{для всех } i \neq r, \quad y_r \equiv 0. \tag{4}$$

Иными словами, для каждой вершины i выбирается та из имеющихся в сети дуг (i, j), которая позволяет минимизировать затраты. [Понятно, что при $\alpha = 1$ выражение (4) сводится к тому соотношению, которое было приведено в разд. 7.6.]

Концептуально настоящая задача решения экстремального уравнения в одном отношении более трудна, чем модель восстановления, описанная в предыдущих разделах. В модели восстановления имелась всего лишь одна неизвестная величина f; здесь же имеется $(p - 1)$ неизвестных y_i $(i \neq r)$. Однако в другом отношении модель нахождения кратчайшего пути более проста, поскольку существует вершина конечного состояния, а следовательно, $y_r = 0$. Это значит, что хотя период, необходимый для переезда из некоторой вершины k в вершину r, зависит от длины выбранного маршрута, длительность этого периода всегда конечна. Далее читатель узнает, как методы последовательных приближений могут быть использованы для вычисления всей совокупности y_i.

Метод итераций по критерию. Подходы, рассмотренные при анализе модели восстановления, можно использовать для последовательного получения приближенных значений y_i. *Алгоритм приближений в пространстве функций* или *итераций по критерию* состоит в нахождении

$$y_i^{n+1} = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha y_j^n + c_{ij}) \quad \text{для всех } i \neq r, \ y_r^{n+1} \equiv 0. \tag{5}$$

Вычислительный процесс прекращается на итерации $(n + 1)$, если для каждой вершины i либо $y_i^{n+1} = y_i^n$, либо

$$y_i^n \leqslant \alpha y_i^n + c_{ij} \tag{6}$$

по всем имеющимся в сети дугам (i, j), причем строгое равенство в (6) выполняется хотя бы для одной из дуг (i, j). Удобным исходным предположением о значении y_i для всех i является

$$y_i^0 = 0. \tag{7}$$

Если все $c_{ij} > 0$, то при исходных $y_i^0 = 0$ значения y_i^n монотонно возрастают. Воспользовавшись (7), можно истолковать y_i^n как минимальные интегральные дисконтированные затраты для любого маршрута из вершины i, включающего точно n дуг, при условии что в маршрут не входит вершина конечного состояния; в последнем случае маршрут заканчивается в вершине r. (Каждая из дуг может входить в маршрут более одного раза.)

Всегда ли работает алгоритм итераций по критерию (5) и (7)? Если $\alpha = 1$ и на любом замкнутом циклическом маршруте затраты положительны, то существует такое n^*, что $y_i^n = y_i^{n^*}$ для всех $n \geqslant n^*$, так что вычислительный процесс завершается за конечное число итераций. Это объясняется тем, что по мере увеличения n любой маршрут длиной n дуг в конце концов достигает вершины конечного состояния. Если же $\alpha < 1$, то предположим, что для любой вершины k интегральные дисконтированные затраты на перемещение по каждому из циклических маршрутов, начинающихся и завершающихся в вершине k, будут больше, чем для лучшего, не содержащего циклов маршрута, ведущего от вершины k к вершине конечного состояния. При таком предположении сходимость достигается за конечное число итераций.

Пример. Расчеты с использованием алгоритма итераций по критерию (5) и (7) для примера, изображенного на рис. 11.5, при $\alpha = 1$ приведены в таблице рис. 11.6. Чтобы читатель мог проверить, насколько он усвоил этот метод, ему рекомендуется выполнить расчеты со всеми подробностями, использовав для этого рис. 11.5 (нужно перерисовать сеть на отдельный лист бумаги) и сверяя результаты для каждого n с рис. 11.6. Отметим, что на итерации 9 все $y_i^9 = y_i^8$, так что вычислительный процесс завершен.

Заметим также, что для $n = 0, 1, \ldots, 5$

$$y_2^{n+1} = y_7^n + c_{27} \quad \text{и} \quad y_7^{n+1} = y_2^n + c_{72}. \tag{8}$$

Таким образом, для $n \leqslant 6$ лучший маршрут из вершины *2*, включающий n дуг, содержит цикл, в который входит вершина *7*. Аналогично этому лучший маршрут из вершины *7*, включающий n дуг, содержит цикл, в который входит вершина *2*. Отметим также, что набор входящих в маршрут дуг на итерации $N = 7$ является

$n =$	0	1	2	3	4	5	6	7	8	9
y_1^n	0	0	0	0	0	0	0	0	0	0
y_2^n	0	1	2	3	4	5	6	6	6	6
y_3^n	0	1	2	3	3	3	3	3	3	3
y_4^n	0	1	2	2	2	2	2	2	2	2
y_5^n	0	1	2	3	3	3	3	3	3	3
y_6^n	0	3	4	5	6	6	6	6	6	6
y_7^n	0	1	2	3	4	5	6	7	7	7
y_8^n	0	1	2	3	4	5	6	7	8	8

Рис. 11.6. Приближение снизу в пространстве функций (метод итераций по критерию).

оптимальным, однако y_8^7 не равно истинному значению y_8, в связи с чем расчеты на этой итерации не прекращаются.

Можно применить также метод приближений (5) в пространстве функций, обеспечивающий монотонную сходимость к y_i сверху. В качестве исходного варианта выберем для каждой вершины дугу таким образом, чтобы любая вершина была соединена маршрутом с вершиной конечного состояния. Пусть y_i^0 — соответствующие интегральные дисконтированные затраты для маршрута от вершины i к вершине конечного состояния. Как и выше, существует такое n^*, что $y_i^n = y_i^{n^*}$ для всех $n \geqslant n^*$, так что вычислительный процесс завершается за конечное число итераций. Подобный подход очень тесно связан с алгоритмом нахождения кратчайшего пути, описанным в разд. 7.6. Однако в этом алгоритме не уделяется особого внимания очередности вершин i при вычислении приближенных значений y_i. Здесь же, при *точном* выполнении (5), прежде чем вернуться к какой-то ранее рассмотренной вершине, требуется вычислить новое приближенное значение y_i^{n+1} для *всех* остальных вершин. Следовательно, при использовании (5) сходимость может оказаться не такой быстрой, как в других методах.

$n=$	0	1	2	3
y_1^n	0	0	0	0
y_2^n	6	6	6	6
y_3^n	3	3	3	3
y_4^n	2	2	2	2
y_5^n	3	3	3	3
y_6^n	7	6	6	6
y_7^n	8	7	7	7
y_8^n	9	9	8	8

Рис. 11.7. Приближение сверху в пространстве функций (метод итераций по критерию).

Продемонстрируем метод (5), обеспечивающий сходимость сверху, на примере сети рис. 11.5 при $\alpha=1$. Пусть исходный набор включает дуги (*8*, *7*), (*7*, *3*), (*6*, *4*), (*5*, *4*), (*4*, *1*), (*3*, *4*) и (*2*, *3*). Убедимся в том, что соответствующие значения y_i^0 приведены в таблице рис. 11.7 в столбце, где n равно нулю. Выполнив операции процесса последовательных приближений (5) на сети рис. 11.5, проверим правильность значений y_i^n, проставленных и во всех остальных столбцах таблицы рис. 11.7. При этом заметим, что набор дуг является оптимальным уже при $n = 1$, однако расчеты не прекращаются на следующей ($n = 2$) итерации, поскольку $y_8^2 \neq y_8^1$. Процесс завершается только при $n = 3$, потому что все $y_i^3 = y_i^2$.

Как видно из приведенного примера, при методе аппроксимации в пространстве функций оптимальную стратегию можно найти на итерации n, однако соответствующие значения y_i^n могут быть и не оптимальными. В тех случаях когда приближения осуществляются снизу, y_i^n не обязательно должны соответствовать какому-то конкретному маршруту от вершины i к вершине конечного состояния.

Метод итераций в пространстве стратегий. Перейдем к *приближениям в пространстве стратегий*. На каждой итерации n необходимо исчислить *фактические* значения y_i для проверяемой стратегии. Алгоритм состоит из следующих шагов:

Шаг 1. Выберем некоторый исходный набор (p — 1) дуг таким образом, чтобы каждая вершина была соединена маршрутом с вершиной конечного состояния, и примем $n = 0$.

Шаг 2. Для заданного проверяемого набора дуг вычислим соответствующие y_i^n, причем $y_r^n \equiv 0$.

Шаг 3. Проверим возможность дальнейших улучшений, вычислив

$$\min_{\substack{\text{по всем } (i,j) \\ \text{в сети}}} (\alpha y_i^n + c_{ij}) \equiv Y_i^n \quad \text{для всех } i \neq r. \tag{9}$$

Шаг 4. Прекратим расчеты, если для каждой вершины i выполняется условие $Y_i^n = y_i^n$, поскольку набор дуг является оптимальным. В противном случае изменим входящие в набор дуги для всех вершин k, по которым $Y_k^n < y_k^n$, включив в набор дуги, соответствующие Y_k^n, которые удовлетворяют (9). Перейдем от итерации n к итерации $(n+1)$ и обратимся к шагу 2, использовав новый проверяемый набор дуг.

Для того чтобы проверить, как читатель усвоил алгоритм, следует найти этим методом решение задачи на сети рис. 11.5, начав с того же набора дуг, который был рассмотрен в столбце $n = 0$ таблицы рис. 11.7, и соответственно с тех же значений y_i^0. Убедимся в том, что при $n = 1$ все Y_i^1 равны y_i^1, являющимся истинными значениями y_i. Следовательно, вычислительный процесс завершается при $n = 1$, поскольку оптимальная стратегия найдена.

Сравнивая методы итераций в пространстве функций и в пространстве стратегий, заметим, что, согласно правой части (5) и левой части (9), требуется выполнять одни и те же вычисления. Однако при приближениях в пространстве функций полученное значение (5) используется затем на следующей итерации в качестве нового приближенного значения y_i, тогда как при приближениях в пространстве стратегий на шаге 2 требуется вычислять новые приближенные значения y_i. Именно эти дополнительные вычисления обеспечивают прекращение процесса сразу же по нахождении оптимальной стратегии (более подробно о шаге 2 см. упражнение 58).

Выводы. Методы последовательных приближений — как в пространстве функций, так и в пространстве стратегий — применимы к решению экстремального уравнения, построенного для задачи нахождения кратчайшего пути от каждой вершины k к вершине конечного состояния. Следует обратить внимание на два различия в применении метода итераций по критерию к модели восстановления и к сетевой модели. В модели восстановления имелось только одно неизвестное, а следовательно, только одно экстремальное уравнение, тогда как в сетевой модели имеется $(p-1)$ неизвестных и экстремальных уравнений. Плановый период в модели восстановления является неограниченным, и метод итераций по критерию не обязательно обеспечивает сходимость за конечное число итераций; напротив, при нахождении кратчайшего пути в сетевой модели метод сходится за конечное число итераций. В гл. 12 описанная выше сетевая постановка задачи совмещается с процессом оптимизации в течение бесконечного планового периода.

В приведенном обсуждении было принято, что для перемещения по любой дуге требуется только один отрезок. Однако иногда желательно, чтобы на перемещение по дуге затрачивалось несколько отрезков. Так, в модели восстановления, отображаемой сетью рис. 11.4, перемещение по дуге (i, j) означает, что пройдено $(i - j)$ отрезков. Описанные выше методы легко модифицировать с учетом такого обобщения модели.

Пусть

h_{ij} — число отрезков, необходимое для перемещения из вершины i в вершину j по дуге (i, j). (I)

Тогда каждой дуге (i, j) можно сопоставить коэффициент дисконтирования, равный

$$\alpha_{ij} \equiv (\alpha)^{h_{ij}}, \quad 0 \leqslant \alpha \leqslant 1. \qquad \text{(II)}$$

Экстремальные уравнения, определяющие интегральные дисконтированные показатели y_i, вместо (4) имеют следующий вид:

$$y_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha_{ij} y_i + c_{ij}) \quad \text{для всех } i \neq r \text{ и } y_r = 0; \qquad \text{(III)}$$

точно так же изменяются все остальные формулы в настоящем разделе: везде вместо α пишем α_{ij}.

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

1. Объясните, в каком смысле предполагается стационарность:

а) в численном примере, приведенном в таблице рис. 11.1;

б) в модели эксплуатации лесного хозяйства с бесконечным плановым периодом (разд. 11.3);

в) в модели восстановления с бесконечным плановым периодом (разд. 11.4).

2. В разд. 11.1 утверждается, что при сравнении двух стратегий с бесконечным плановым периодом определение их относительной предпочтительности не вызывает принципиальных трудностей, если при *любой* конечной длительности планового периода одна из стратегий обеспечивает больший совокупный эффект, чем другая. Согласен ли с этим читатель? По каким соображениям? (Имеются ли в приведенном утверждении какие-то неявные допущения?)

3. По данным рис. 11.1 проверьте, что:

а) для любого четного отрезка стратегия D обеспечивает больший совокупный эффект, чем стратегия B;

б) для любого отрезка $(3 + 6n)$, где $n = 0, 1, 2, \ldots$, стратегия B «лучше» стратегий A, C или D;

в) общие члены последовательностей, отображающих средний эффект за отрезок, описываются формулами (I) — (V), которые приведены в начале разд. 11.2;

г) найдите такое значение n^*, что для любого отрезка n, где $n \geqslant n^*$, средний эффект за отрезок для стратегии A заключен в пределах $2 \pm 0{,}01$. Найдите аналогичное значение n^* для стратегий B, C и D. Найдите аналогичное значение n^*, обеспечивающее для стратегии G нахождение среднего эффекта за отрезок в пределах $1^1/_3 \pm \pm 0{,}01$. Как изменится значение n^* для стратегии А, если «допуск» вместо 0,01 будет принят равным 0,001?

4. Рассмотрите задачу выбора одной из двух следующих стратегий: стратегии А (см. таблицу рис. 11.1) и стратегии, отличающейся от нее только тем, что на первом отрезке эффект равен 100 единицам. Предложите одно или несколько *общих правил*, применимых к задаче выбора одной из двух стратегий, для которых значения среднего эффекта за отрезок при бесконечном плановом периоде являются одинаковыми. Всегда ли предложенные вами правила обеспечивают однозначный выбор из числа стратегий, дающих одинаковый средний эффект за отрезок при бесконечном плановом периоде? Обоснуйте ваш ответ.

5. Для стратегии Е (см. таблицу рис. 11.1) вычислите средний эффект за отрезок. Согласны ли вы с приводимым в тексте утверждением о том, что при включении в число возможных альтернатив стратегии D стратегию Е, вероятно, целесообразно исключить из дальнейшего рассмотрения? Предположим, что стратегии D не существует; целесообразно ли по-прежнему исключать стратегию Е из рассмотрения? Приведите соображения, обосновывающие ваши ответы.

6. Пусть рассматриваются варианты а) — з) постановки задачи, при каждом из которых к стратегиям, представленным в таблице рис. 11.1, добавляется соответствующая альтернатива из числа приведенных ниже. Определить, является ли дополнительная стратегия доминирующей по отношению к стратегиям рис. 11.1 или какие-либо из них будут доминирующими по отношению к этой стратегии; вычислите для дополнительной стратегии средний эффект за отрезок.

а) 3, 4, —1, 3, 4, —1, . . .,

б) 6, —1, 1, 6, —1, 1, . . .,

в) 0, 0, 8, 0, 0, 8, 0, 0, . . .,

г) 0, 8, 0, 0, 0, 8, 0, 0, . . .,

д) 4, 0, 4, 0, 4, 0, . . .,

е) 0, 4, 0, 4, 0, 4, . . .,

ж) 2, $2^2/_3$, 2, 2, 2, . . ., 2, . . .,

з) $2+e,\ 2+e^2,\ 2+e^3,\ 2+e^4,\ 2+e^5,\ 2+e^6, \ldots,$

где e — постоянная величина, заключенная между 1 и 0. (*Указание*: ответ зависит от конкретного значения e.)

7. Для каждого варианта постановки задачи в упражнении 6 определите вид общего члена последовательности значений среднего эффекта за отрезок, аналогично тому как это сделано посредством формул (I) — (V) в начале разд. 11.2.

8. Рассмотрите следующую конечную последовательность значений эффекта: $R_1 = 4$, $R_2 = 12$, $R_3 = 24$, $R_4 = 40$; предполагается что коэффициент дисконтирования $\alpha = {}^1/_2$ и что при соответствующей процентной ставке (по формуле сложных процентов) можно брать или давать взаймы любую сумму денег.

а) Чему равен интегральный дисконтированный эффект для этой последовательности? Чему равна соответствующая процентная ставка в формуле сложных процентов?

б) Если мы хотим перераспределить эффект во времени таким образом, что весь эффект будет возникать на отрезке 1, а для остальных отрезков равняться нулю, то какова величина получаемого эффекта на отрезке 1? (Укажите, сколько можно взять или дать взаймы на каждом отрезке, когда будет произведен расчет по этим ссудам, а также каково значение соответствующей процентной ставки.)

в) Ответьте на вопрос б), однако при условии, что весь эффект требуется получить на отрезке 2; на отрезке 3; на отрезке 4.

г) Подробно объясните, почему последовательность $S_1 = 2$, $S_2 = 16$, $S_3 = 20$, $S_4 = 56$ является более предпочтительной, чем последовательность R_j.

д) Подробно объясните, почему последовательность $T_1 = 4$, $T_2 = 8$, $T_3 = 24$, $T_4 = 48$ является менее предпочтительной, чем последовательность R_j.

9. Рассмотрите последовательность значений эффекта в упражнении 8. При каком значении α последовательности R_j и S_j будут одинаково предпочтительны? R_j и T_j? S_j и T_j?

10. Выведите формулу (5) разд. 11.2, определяющую значение интегрального дисконтированного эффекта, и выполните все промежуточные алгебраические преобразования, позволяющие получить (6) — (12).

11. Выведите формулы интегрального дисконтированного эффекта, аналогичные (6) — (12) из разд. 11.2 для последовательностей, приведенных в упражнении 6.

12. а) Выполните промежуточные алгебраические преобразования, необходимые для получения (13) — (17) из разд. 11.2.

б) Какое заключение можно сделать о сравнительной предпочтительности стратегий A и C? Стратегий A и D? Проверьте, действительно ли стратегия B более предпочтительна, чем стратегия D, при $\alpha < {}^1/_2$ и менее предпочтительна при $\alpha > {}^1/_2$.

13. Для каждого из указанных ниже вариантов постановки задачи выведите формулы, аналогичные (13) — (17) из разд. 11.2 для последовательностей, приведенных в таблице рис. 11.1 и в упражнении 6. Укажите диапазоны значений α, на которых предпочтительна та или иная последовательность:

а) стратегия B и стратегия E,

б) стратегия B и стратегия C,

в) стратегия C и стратегия D,

г) стратегия A и стратегия из п. а) упражнения 6, д) стратегия В и стратегия из п. а) упражнения 6,

е) стратегия С и стратегия из п. а) упражнения 6,

ж) стратегия С и стратегия из п. б) упражнения 6,

з) стратегия А и стратегия из п. в) упражнения 6,

и) стратегия D и стратегия из п. з) упражнения 6.

14. Проверьте правильность формул (I) — (IV) в материалах повышенной трудности, приведенных после (16) в разд. 11.2.

15. а) Выполните промежуточные алгебраические преобразования, необходимые для получения формул (18) и (19) разд. 11.2 для стратегии В *.

б) Получите аналогичные результаты для стратегии А* и убедитесь в том, что полученный интегральный дисконтированный эффект при $0 \leqslant \alpha < 1$ является конечным.

в) Выведите для стратегий А* и В* формулу, аналогичную (13) — (17) из разд. 11.2. Укажите диапазоны значений α, на которых предпочтительна та или иная стратегия.

г) Определите интегральный дисконтированный эффект для стратегии С*, представленной следующей последовательностью:

$$1, 6, -1, 1, 6, -1, 2, 12, -2, 2, 12, -2, 3, 18, -3, 3, 18, -3, \ldots.$$

Ответьте на вопросы варианта в) для стратегий В* и С*.

д) Определите интегральный дисконтированный эффект для стратегии D*, представленной следующей последовательностью:

$$2^2/_3,\ 2,\ 2,\ 2,\ 2,\ 2,\quad 4,\ 4,\ 4,\ 4,\ 4,\ 4,\quad 12,\ 12,\ 12,\ 12,\ 12,\ 12,\ \ldots.$$

Ответьте на вопросы варианта в) для стратегий В* и D*.

16. Проверьте правильность (22) из разд. 11.2. Покажите, почему для стратегии В* при $\alpha = 1$ эквивалентный средний эффект не определен.

17. Определите эквивалентный средний эффект для каждой из последовательностей упражнения 6. Вычислите значения при $\alpha = 1$.

18. В тексте разд. 11.2 утверждается, что, хотя в соответствии с критерием эквивалентного среднего эффекта стратегии А, В, С и D (см. таблицу рис. 11.1) при $\alpha = 1$ являются одинаково предпочтительными, решение должно основываться на личном суждении руководителя, а не базироваться на каком-либо научном обосновании. Затем в тексте указывается возможность привести доводы в пользу того, что при $\alpha = 1$ стратегия А является оптимальной, стратегии С и D очень близки к оптимуму, а стратегию В следует исключить из дальнейшего рассмотрения. Рассмотрите затронутые проблемы; укажите, согласны ли вы с приведенными доводами. Обоснуйте высказанную вами точку зрения.

19. а) В тексте разд. 11.3 сделано предположение о том, что последовательность $\alpha^{k-1}N_k$ удовлетворяет неравенствам (2). Рассмотрите трудности, возникающие при отказе от этого упрощающего допущения.

б) Покажите, что из неравенств (2) следует возможность определения оптимального значения K из (3). Всегда ли (3) определяет единственное значение K, если имеются альтернативные оптимальные стратегии? Обоснуйте ваш ответ.

в) Дайте истолкование неравенств (2) и критерия прекращения вычислительного процесса (3) при $\alpha = 1$.

20. а) Вычертите график функции (4) из разд. 11.3 при $a = 0{,}75$ и $b = 0{,}5$.

б) Выполните промежуточные алгебраические преобразования и покажите, что, если N_k описывается (4), критерий (3) прекращения вычислительного процесса упрощается и сводится к (5).

в) Постройте пример такой последовательности N_k, что при фиксированном значении коэффициента дисконтирования $0 < \alpha < 1$ не существует *конечного* значения K, которое было бы оптимальным.

г) Для числового примера (6) разд. 11.3 покажите, что $K = 4$ удовлетворяет (5).

д) Проверьте правильность вычисленных в (7) интегральных дисконтированных показателей для $k = 1$ и $k = 2$.

21. Рассмотрим модель эксплуатации лесного хозяйства (раздел 11.3). Пусть

$$N_k = a + kb, \quad \text{где} \quad a + b \geqslant 0 \quad \text{и} \quad b > 0.$$

а) Определите критерий прекращения вычислительного процесса, аналогичный (5).

б) Примем $\alpha = 2/3$, $a = 3$ и $b = 9$. Чему равно оптимальное значение K? Вычислите дисконтированные показатели $\alpha^{k-1}N_k$ для $k = 1$, $k = 2$, $k = 3$ и $k = 4$.

22. Рассмотрим модель эксплуатации лесного хозяйства (разд. 11.3). Пусть

$$N_k = a + k^2 b, \quad \text{где} \quad a + b \geqslant 0 \quad \text{и} \quad b > 0.$$

а) Определите критерий прекращения вычислительного процесса, аналогичный (5).

б) Примем $\alpha = 2/3$, $a = 3$ и $b = 9$. Чему равно оптимальное значение K?

в) Вычислите дисконтированные показатели $\alpha^{k-1}N_k$ для $k = 1, 2, \ldots, k + 1$, где k — оптимальное значение, полученное в постановке б).

23. а) В тексте разд. 11.3 предполагается, что последовательность ИДЭ $[k]/(1 - \alpha^k)$ удовлетворяет неравенствам (14). Рассмотрите трудности, возникающие при отказе от этого упрощающего допущения.

б) Покажите, почему из (14) следует, что оптимальное значение k^1 можно определить из (15). Всегда ли (15) определяет единственное значение k^1, если имеются альтернативные оптимальные стратегии? Обоснуйте ваш ответ.

в) Объясните, почему отношение в правой части (15) больше, чем N_k/N_{k+1} для любого k.

г) Покажите, что если $\alpha = 1$, неравенства (15) можно представить в виде (17).

24. Рассмотрите численный пример (6) из разд. 11.3.

а) Используйте (15) для проверки того, что при бесконечном плановом периоде $k' = 2$ является оптимальной стратегией.

б) Проверьте правильность приведенных в (18) значений $F(1)$ и $F(2)$.

25. Рассмотрите модель, приведенную в упражнении 21. Определите оптимальную стратегию k' для бесконечного планового периода. Вычислите $F(k'-1)$, $F(k')$ и $F(k'+1)$. Чему равно сокращение интегрального дисконтированного эффекта при использовании вместо k' оптимальной стратегии в условиях однократного принятия решений, т. е. K?

26. Рассмотрите модель, приведенную в упражнении 22. Определите оптимальную стратегию k' для бесконечного планового периода. Вычислите $F(k'-1)$, $F(k')$ и $F(k'+1)$. Чему равно сокращение интегрального дисконтированного эффекта при использовании вместо k' стратегии K, оптимальной в условиях однократного принятия решений?

27. Рассмотрите модель восстановления с бесконечным плановым периодом, приведенную в разд. 11.4.

а) Пусть все $R_k > 0$. Объясните, почему не существует конечного значения f при $\alpha = 1$, удовлетворяющего (5).

б) Пусть все $R_k = 0$. Объясните, почему любое конечное значение f удовлетворяет (5) при $\alpha = 1$.

в) Объясните, почему из (5) следуют неравенства (6), а из (6) в свою очередь — (7).

г) Приведите промежуточные алгебраические преобразования, необходимые для получения правой части (8).

д) Приведите промежуточные алгебраические преобразования, необходимые для получения (9), и покажите, что при $\alpha = 1$ выражение (9) превращается в (10).

е) Покажите, каким образом преобразуются (6), (7), (9) и (10), если целью оптимизации в (5) является достижение максимума.

28. Рассмотрите модель управления запасами с вогнутой функцией затрат, приведенную в конце разд. 11.4.

а) Покажите, что полученное из (VI) k позволяет минимизировать затраты (V).

б) Выведите формулу $R_k(\alpha)$ при коэффициенте дисконтирования α, если $0 < \alpha < 1$ и дисконтируются затраты по каждому отрезку. Какова зависимость оптимального размера партии от значения α? Какова зависимость оптимального размера партии, если такая зависимость имеется, от закупочной цены c? Дайте подробное объяснение.

29. Рассмотрите теорему о длительности планового периода в модели восстановления, приведенной в разд. 11.5. Приняв $K = 5$ и $n^* = 8$, дайте истолкование:

а) формуле (3) при $k_{10} = 4$; при $k_{20} = 4$; при $k_{20} = 6$,

б) формуле (4) при $n = 10$; при $n = 20$,

в) если стратегия 12 оптимальна для всех $n > 22$, верно ли $k_{30} = 12$? Будет ли стратегия 12 оптимальной при бесконечном плановом периоде? [Обоснуйте ваши ответы с учетом (3) и (4).]

г) если стратегия 9 оптимальна при бесконечном плановом периоде, $k_3 = 9$; будет ли оптимальна программа 9 для всех $n > 3$?

30. В каждой из приведенных ниже задач примените метод рассмотрения бесконечности как предела возрастающей длительности конечного планового периода и найдите наименьшее значение n, такое, что стратегия, оптимальная при бесконечном плановом периоде, будет оптимальна и при конечной длительности, равной n, $n + 1$ и $n + 2$. (Напоминаем, что целью оптимизации здесь является достижение максимума.)

а) Условный пример, описываемый выражениями (6) и (18) (модель эксплуатации лесного хозяйства) из разд. 11.3, где $R_k \equiv$ ИДЭ $[k]$.

б) Задача, приведенная в упражнениях (21) и (25).

в) Задача, приведенная в упражнениях (22) и (26).

31. Рассмотрите численный пример (2), использованный для демонстрации алгоритма итераций по критерию (1) в разд. 11.6.

а) Проверьте правильность (4) и, вычислив значения показателей внутри скобок, убедитесь в том, что при $k = 1$ достигается минимум.

б) При известном значении f^{15}, которое приведено в (5), вычислите f^{16} и проверьте, что $k = 3$.

в) Примите $f^0 = 20$ и выполните операции алгоритма итераций по критерию. Завершите вычислительный процесс после $n = 3$.

г) Примите $f^0 = 20$ и выполните операции алгоритма итераций по критерию. Завершите вычислительный процесс после $n = 3$.

д) Пусть $R_1 = 7$. Выполните операции алгоритма итераций по критерию, приняв $f^0 = 0$. При каком n впервые появляется значение $k = 3$? Каково соответствующее значение f^n? Сравните это значение f^n с f^4 в (5).

е) Пусть $R_1 = 10$. Выполните операции алгоритма итераций по критерию, приняв $f^0 = 0$. При каком n впервые появляется значение $f^n = 3$? Каково соответствующее значение f^n? Сравните это значение f^n с f^4 в (5).

ж) Пусть $R_4 = 17{,}9$. Выполните операции алгоритма итераций по критерию, приняв $f^0 = 43{,}50$. При каком n впервые появляется значение $k = 3$? Каково соответствующее значение f^n? Сравните это f^n с f^1 в (7).

з) Предположим, что на каждой итерации n, где $n \geqslant n'$, выбирается одна и та же оптимальная стратегия для бесконечного планового периода. Выведите формулу вычисления f^n при известном значении f^n.

32. Графическое отображение метода итераций по критерию. Для каждого k вычертите график функции $y = \alpha^k f + R_k$, где y — ордината, а f — абсцисса. Для каждого значения f одна (или более) из этих линий будет лежать ниже других. Обведите красным карандашом график кусочно-линейной функции, состоящей из этих «минимальных» отрезков. Затем вычертите график функции $y = f$. Пересечение этой прямой и графика кусочно-линейной функции дает решение — искомое значение f. При использовании метода итераций по критерию мы начинаем с f^0. Проведите в точке $f = f^0$ вертикальную прямую и выберите значение k, отвечающее точке ее пересечения с графиком кусочно-линейной функции. Соответствующее значение этой функции представляет собой следующее пробное значение f, то есть f^1; его можно найти, проведя горизонтальную прямую через точку на графике кусочно-линейной функции до пересечения с прямой $y = f$. По мере повторения выполняемых итераций пробные точки все более и более приближаются к точке пересечения прямой $y = f$ и графика кусочно-линейной функции.

В каждой из приведенных ниже задач вычертите график для каждого значения k, постройте график кусочно-линейной функции, вычертите прямую $y = f$ и найдите решение. Продемонстрируйте реализацию метода итераций по критерию при $f^0 = 0$ и при $f^0 = R_1/(1 - \alpha)$. [Итеративный метод демонстрируйте на графике, не выполняя лежащих в его основе вычислительных операций. Напоминаем, что в задачах б), в) и г) целью оптимизации является достижение максимума.]

а) Численный пример (2) разд. 11.6.

б) Иллюстративный пример, описываемый выражениями (6) и (8) (модель эксплуатации лесного хозяйства) из разд. 11.3.

в) Задача, приведенная в упражнениях 21 и 25.

г) Задача, приведенная в упражнениях 22 и 26.

33. В каждой из указанных ниже задач примените метод итераций по критерию, описанной в разд. 11.7. Обозначьте n' индекс первой итерации, на которой соответствующее k является оптимальным для условий бесконечного планового периода; завершите вычислительный процесс после итерации $(n' + 2)$. Начните в первую очередь с $f^0 = 0$. Затем начните с $f^0 = R_1/(1 - \alpha)$. (Напоминаем, что целью оптимизации является достижение максимума.)

а) Иллюстративный пример, описываемый выражениями (6) и (8) (модель эксплуатации лесного хозяйства) из разд. 11.3, где $R_k \equiv$ ИДЭ $[k]$.

б) Задача, приведенная в упражнениях 21 и 25.

в) Задача, приведенная в упражнениях 22 и 26.

г) Для каждой задачи опишите стратегию, действительно позволяющую получить $f^{n'+2}$. Является ли эта стратегия стационарной? [*Примечание*: такая стратегия может и не существовать.]

д) Для каждой задачи проверьте, действительно ли $|f^n - f| \leqslant \alpha^n |f^0 - f|$ при $n = 1, 2, \ldots, n' + 2$.

34. В каждой из указанных ниже задач примените алгоритм итераций по стратегиям, описанный в разд. 11.7. В качестве исходной стратегии при $n = 0$ используйте сперва $k = 1$, а затем $k = K + 1$, где K — стратегия, оптимальная для бесконечного планового периода.

а) Иллюстративный пример, описываемый выражениями (6) и (8) (модель эксплуатации лесного хозяйства) из разд. 11.3, где $R_k \equiv$ ИДЭ $[k]$.

б) Задача, приведенная в упражнениях 21 и 25.

в) Задача, приведенная в упражнениях 22 и 26.

35. Рассмотрите возможность применения критерия среднего эффекта за отрезок в алгоритме итераций по стратегиям, описанном в разд. 11.7.

а) Покажите, что в терминах эквивалентного среднего эффекта (5) соответствует (1).

б) Аналогично этому покажите, что (6) соответствует (2).

в) Обозначьте через $g(k)$ выражение, заключенное в (6) в квадратные скобки. Обозначьте $a = (1 - \alpha)^{-1}$, $b = -g^n (1 - \alpha)^{-1}$. Постройте $ag(k) + b$ и покажите, что с помощью необходимых алгебраических преобразований получается простое выражение в левой части (7).

г) Дайте истолкование проверочного критерия (7) при $\alpha = 1$.

36. Используйте алгоритм итераций по стратегиям, описанный в разд. 11.7, для нахождения оптимального решения иллюстративного примера (6) — модель эксплуатации лесного хозяйства из разд. 11.3 при $\alpha = 1$. В качестве исходной стратегии примите $k = 1$. (Напоминаем, что целью оптимизации является здесь достижение максимума.)

37. Рассмотрите задачу нахождения кратчайшего пути в разд. 11.9 и дайте истолкование (2) и (3).

38. Примените алгоритм итераций по критерию, описанный в разд. 11.9, для решения примера, отображенного на рис. 11.5, при $\alpha = 1$ и $y_i^0 = 0$.

а) Выполните вычисления со всеми подробностями и проверьте правильность результатов, приведенных в таблице рис. 11.6.

б) Найдите лучший маршрут из n дуг, соединяющих вершину *2* с вершиной конечного состояния при $n \leqslant 6$; то же для вершины *7*.

39. Примените алгоритм итераций по критерию, описанный в разд. 11.9, для решения задачи, изображенной на рис. 11.5, приняв $\alpha = 1$ и выбрав значения y_i^0 соответствующими набору дуг (*8*, *7*), (*7*, *3*), (*6*, *4*), (*5*, *4*), (*4*, *1*), (*3*, *4*) и (*2*, *3*).

а) Проверьте правильность значений y_i^0, приведенных в таблице рис. 11.7.

б) Выполните вычисления со всеми подробностями и проверьте правильность результатов, приведенных в таблице рис. 11.7. [Процесс последовательных приближений (5) выполняйте на рис. 11.5.]

40. Примените алгоритм итераций по критерию, описанный в разд. 11.9, для решения задачи, изображенной на рис. 11.5, приняв $\alpha = 1$ и $y_i^0 = 10$ для всех i $(i \neq r)$.

41. Примените алгоритм итераций по критерию, описанный в разд. 11.9, для решения задачи, изображенной на рис. 11.5, приняв $\alpha = 1$ и выбрав значения y_i^0 соответствующими набору дуг (*8*, *5*), (*7*, *3*), (*6*, *4*), (*5*, *1*), (*4*, *1*), (*3*, *1*) и (*2*, *1*).

42. В каждом из указанных ниже вариантов постановки задачи примените алгоритм итераций по критерию, описанный в разд. 11.9, для решения задачи, изображенной на рис. 11.5, при $\alpha = 0{,}8$. Завершите вычислительный процесс после итерации 9.

а) $y_i^0 = 0$.

б) y_i^0 вычисляются для набора дуг, приведенного в упражнении 39.

в) y_i^0 вычисляются для набора дуг, приведенного в упражнении 41.

43. Примените алгоритм итераций по стратегиям, описанный в разд. 11.9, для решения задачи упражнения 41.

44. Примените алгоритм итераций по стратегиям, описанный в разд. 11.9, для решения задачи упражнения 42:

а) вариант б),

б) вариант в).

45. Объясните, как вы понимаете следующие термины:

бесконечный плановый период;
стационарность;
стационарная стратегия,
бесконечная последовательность значений эффекта;
полезность денежных поступлений и затрат;
средний эффект за отрезок;
процентная ставка и сложные проценты;
коэффициент дисконтирования;
интегральный дисконтированный эффект (или интегральные дисконтированные затраты);
эквивалентный средний эффект;
почти оптимальный;
модель восстановления;
отрезок (или момент) восстановления;
функциональное (или экстремальное) уравнение;
метод последовательных приближений;
метод итераций по критерию (метод последовательных приближений в пространстве функций);
метод итераций по стратегиям (метод последовательных приближений в пространстве стратегий);
монотонно возрастающая (убывающая) последовательность;
сходимость сверху (снизу).

УПРАЖНЕНИЯ НА РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

46. Задача замены оборудования. Рассмотрим задачу фирмы «Резвая фреза», приведенную в упражнении 41 гл. 6. Предположим, что необходимо определить стратегию замены оборудования для бесконечного планового периода. Пусть стоимость новой модели на любом отрезке равна p, число альтернативных стратегий $N = 7$, а остаточная стоимость (продажная цена) заменяемого оборудования и эксплуатационные расходы приведены в упражнении 41 гл. 6 (т. 1). В приведенных ниже вариантах а) и б) найдите оптимальную стратегию замены при $\alpha = 1$.

а) $p = 100$.

б) $p = 90$.

в) Найдите наибольшее и наименьшее значения p, при которых $k = 3$ является оптимальным для $\alpha = 1$.

г) Пусть $\alpha < 1$, закупки и эксплуатационные расходы совершаются в начальный момент отрезка, а остаточная стоимость поступает в конце отрезка. Так, если сверлильный станок приобретают в начале настоящего отрезка и эксплуатируют три отрезка, интегральные дисконтированные затраты для него за эти три отрезка составляют $p + r_1 + \alpha r_2 + \alpha^2 r_3 + \alpha^3 v_3$. Для приведенных выше вариантов а) и б) определите наименьшее значение α, при котором сохраняется оптимальность стратегии, лучшей при $\alpha = 1$; вычислите соответствующую процентную ставку.

д) Если закупочная стоимость увеличена до $(p + d)$, а каждое из значений остаточной стоимости — до $(v_k + d)$, сохранится ли при этом неизменным выбор оптимальной стратегии. Объясните, непременно сформулировав в явном виде все сделанные допущения.

47. Обычно применяемый метод вычисления квадратного корня из положительного числа A представляет собой следующий алгоритм последовательных приближений (предложен Ньютоном): $f^n = 0{,}5 \times [f^{n-1} + (A/f^{n-1})]$. Дайте графическую интерпретацию этого метода. Покажите, что последовательность f^n сходится к истинному значению $f \equiv \sqrt{A}$ при любом исходном $f^0 > 0$. Покажите также, что f^n сходится к f сверху (при $n \geqslant 1$).

48. Рассмотрите модель восстановления с бесконечным плановым периодом, приведенную в разд. 11.4, приняв $N = 5$, $R_1 = 5$, $R_2 = 8$, $R_3 = 10$, $R_4 = 12$, $R_5 = 18$.

а) Определите оптимальную стратегию *для любого* значения α при $0 \leqslant \alpha \leqslant 1$.

б) Пусть $\alpha = 0{,}9$. Примените подход, при котором бесконечность рассматривается как предел возрастающей длительности конечного планового периода (разд. 11.5). Завершите вычислительный процесс если одна и та же стратегия выбирается на пяти итерациях подряд.

в) Пусть $\alpha = 0{,}9$. Примените алгоритм итераций по критерию, описанный в разд. 11.6, приняв сначала $f^0 = 0$, а затем $f^0 = R_1/(1 - \alpha)$. Обозначим n' первую итерацию, на которой k является оптимальным для бесконечного планового периода; завершите вычислительный процесс на итерации $n' = 2$.

г) Используя графическое представление алгоритма, описанное в упражнении 32, выполните на графике итерации варианта в).

д) Пусть $\alpha = 0{,}9$. Примените алгоритм итераций по стратегиям, описанный в разд. 11.7, приняв в качестве исходного $k = 1$.

е) Пусть $\alpha = 1$. Примените алгоритм итераций по стратегиям, описанный в разд. 11.7, приняв в качестве исходного $k = 1$.

49. Для решения каждого из указанных ниже упражнений гл. 7 (т. 1) используйте алгоритм итераций по критерию при нахождении кратчайшего пути, описанный в разд. 11.9, приняв $\alpha = 1$ и y_i^0 соответствующими заданному исходному набору дуг. Затем примените алгоритм итераций по стратегиям, описанный в разд. 11.9, начав с тех же исходных наборов дуг.

а) Упражнение 27, вариант а).
б) Упражнение 28, вариант а).
в) Упражнение 60, вариант а).
г) Упражнение 60, вариант б).
д) Упражнение 61, вариант а).
е) Упражнение 61, вариант б).
ж) Упражнение 62, вариант а).
з) Упражнение 62, вариант б).
и) Упражнение 63, вариант а).
к) Упражнение 63, вариант б).
л) Упражнение 65, вариант а).

50. Для решения каждого из указанных ниже упражнений гл. 7 (т. 1) сначала используйте алгоритм итераций по критерию при нахождении кратчайшего пути, описанный в разд. 11.9, приняв $\alpha = 1$ и выбрав y_i^0 соответствующими заданному исходному набору дуг. Затем примените алгоритм итераций по стратегиям, описанный в разд. 11.9, начав с тех же исходных наборов дуг.

а) Упражнение 27, вариант а), исходный набор дуг (*2*, *1*), (*3*, *1*), (*4*, *1*), (*5*, *1*), (*6*, *4*), (*7*, *6*) и (*8*, *5*).

б) Упражнение 28, вариант а), исходный набор дуг (*1*, *4*), (*3*, *2*), (*4*, *3*), (*5*, *1*), (*6*, *5*), (*7*, *3*) и (*8*, *6*).

в) Упражнение 60, вариант а), исходный набор дуг (*2*, *1*), (*3*, *2*), (*4*, *3*), (*5*, *4*) и (*6*, *5*).

г) Упражнение 60, вариант б), исходный набор дуг (*1*, *2*), (*3*, *2*), (*4*, *3*), (*5*, *4*) и (*6*, *5*).

д) Упражнение 61, вариант а), исходный набор дуг (*2*, *1*), (*3*, *2*), (*4*, *3*), (*5*, *4*) и (*6*, *5*).

е) Упражнение 61, вариант б), исходный набор дуг (*1*, *2*), (*3*, *2*), (*4*, *3*), (*5*, *4*) и (*6*, *5*).

ж) Упражнение 62, вариант а), исходный набор дуг (*2*, *1*), (*3*, *2*), (*4*, *1*), (*5*, *4*) и (*6*, *5*).

з) Упражнение 62, вариант б), исходный набор дуг (*1*, *2*), (*3*, *2*), (*4*, *1*), (*5*, *4*) и (*6*, *5*).

и) Упражнение 63, вариант а), исходный набор дуг (*2*, *1*), (*3*, *2*), (*4*, *3*), (*5*, *4*) и (*6*, *5*).

к) Упражнение 63, вариант б), исходный набор дуг (*1*, *2*), (*3*, *2*), (*4*, *3*), (*5*, *4*) и (*6*, *5*).

л) Упражнение 65, вариант а), исходный набор дуг (*2*, *1*), (*3*, *2*), (*4*, *5*), (*5*, *6*) и (*6*, *3*).

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

Упражнения 51—56 относятся к модели восстановления, описанной в разд. 11.4.

51. а) Покажите, что f, определяемое из (5), представляет собой максимальный дисконтированный эффект для любой (не обязательно стационарной) стратегии.

б) Покажите, что оптимальная стратегия всегда является стационарной. (*Указание*: может оказаться полезным доказательство и использование результата, полученного в упражнении 55.)

52. Докажите теорему о длительности планового периода для модели восстановления, приведенную в разд. 11.5.

53. Задача о ранце. Покажите связь между теоремой о продолжительности планового периода для модели восстановления, приведенной в разд. 11.5, и задачей о ранце, описанной в упражнении 46 гл. 10.

54. Рассмотрите алгоритм итераций по критерию, описанный в разд. 11.6.

а) Покажите, что если все $R_k > 0$, а $f^0 = 0$, то на каждой итерации $f^{n+1} > f^n$.

б) Покажите, что если $f^0 = R_k/(1 - \alpha^k)$ для некоторого выбранного k, то на каждой итерации $f^{n+1} \leqslant f^n$.

в) Покажите, что если f^{n-1}, то $f^n < f^{n+1}$. (Докажите это же утверждение, если в обоих неравенствах знак меняется на противоположный.)

г) Покажите, что если стратегия k выбрана на итерации $(n - 1)$, а затем отвергнута на итерации n, то она больше не будет выбрана.

55. Рассмотрите алгоритм итераций по стратегиям, описанный в разд. 11.7. Предположим, что имеется некоторая стратегия, не обязательно стационарная, и соответствующее ей пробное значение f^n. Покажите, что если стратегия k' удовлетворяет условию $\alpha^{k'}f^n + R_{k'} < f^n$, то использование стратегии k' в качестве стационарной стратегии для бесконечного планового периода позволяет получить $f^{n+1} < f^n$. Дайте объяснение этому явлению.

56. Графическое отображение алгоритма итераций по стратегиям. Разработайте графический подход, аналогичный описанному в упражнении 32 и отображающий вычислительный процесс алгоритма итераций по стратегиям. Примените этот подход к решению численного примера (3) в разд. 11.7. Потребует ли большего числа итераций любая другая исходная стратегия $(k \neq 1)$?

57. Рассмотрите задачу нахождения кратчайшего пути, характеризуемую экстремальными соотношениями (4) в разд. 11.9.

а) Предложите прямой алгоритм (т. е. алгоритм, позволяющий сразу же вычислять истинное значение каждого y_i) для ациклической сети. (*Указание*: модифицируйте подход, описанный в разд. 7.7 т. 1.)

б) Примените предложенный алгоритм для решения задачи мистера М., изображенный на рис. 8.1, приняв для простоты $\alpha = \frac{1}{2}$.

58. Рассмотрите описанный в разд. 11.9 алгоритм итераций по стратегиям для решения задачи нахождения кратчайшего пути.

а) Покажите, что при $\alpha = 1$ на каждой итерации пробный набор дуг на шаге 2 обладает следующим свойством: каждая вершина соединена некоторым маршрутом с вершиной конечного состояния.

б) Постройте пример, позволяющий продемонстрировать, что свойство, которое указано в п. а), не обязательно наблюдается при $\alpha < 1$. Предложите способ вычисления y_i^n для того случая, когда пробный набор дуг включает циклы; дайте объяснение значениям y_i^n, получаемым с помощью предлагаемого метода. (Удостоверьтесь в том, что предлагаемый метод в конечном счете действительно обеспечивает нахождение кратчайшего пути.)

ГЛАВА 12

Методы оптимизации при бесконечном плановом периоде

12.1. ДИСКРЕТНОЕ ДИНАМИЧЕСКОЕ ПРОГРАММИРОВАНИЕ

Теперь читатель подготовлен к тому, чтобы объединить различные изученные им подходы к оптимизации при бесконечном плановом периоде с методами последовательных приближений. При чтении настоящей главы следует стремиться к тому, чтобы:

1) выяснить математические характеристики оптимального решения задачи для бесконечного планового периода;

2) выявить пути нахождения этого решения с помощью методов последовательных приближений.

Полное овладение численными методами является не столь уж важной целью. Более того, практические расчеты выполняются не вручную, а на электронных вычислительных машинах. Однако для эффективного использования полученных решений необходимо знать главные особенности каждой из моделей; это позволит выяснить действительную применимость любого конкретного численного решения.

Модель управления запасами. Начнем с рассмотрения конкретного примера, который позволит нам показать, какой вид оптимизационных рекуррентных соотношений динамического программирования соответствует случаю бесконечного планового периода. В частности, мы рассмотрим стационарную модель управления запасами, близко напоминающую ту модель, которая изучалась ранее, в разд. 8.3. Используем систему обозначений, учитывающую сделанные допущения о стационарности. Управляющей переменной является

$$x = \text{Объем производства},$$

который, согласно нашим требованиям, должен быть неотрицательным целым числом. Обозначим соответствующую функцию производственных затрат для любого отрезка через $C(x)$, а функцию затрат на содержание запасов — через $h(j)$, где j — уровень запасов на конец отрезка. Примем, что для всех отрезков спрос одинаков и равен d (d — положительное целое число), причем он должен быть удовлетворен полностью и своевременно. Как и ранее, положим коэффициент дисконтирования за отрезок равным α $(0 \leqslant \alpha < 1)$.

Введем обозначение:

$f_n(i)$ — интегральный дисконтированный эффект для оптимальной производственной программы при начальном уровне запасов i за n отрезков до конца планового периода.

Тогда рекуррентное соотношение динамического программирования, соответствующее модели с конечным плановым периодом, имеет вид

$$f_n(i) = \min_x [C(x) + h(i + x - d) + \alpha f_{n-1}(i + x - d)], \quad (1)$$

где минимум находится для неотрицательных целых значений при $x \geqslant d - i$.

Для бесконечного планового периода соотношение, аналогичное (1), вероятно, имеет вид

$$f(i) = \min_x [C(x) + h(i + x - d) + \alpha f(i + x - d)], \quad 0 \leqslant \alpha < 1. \quad (2)$$

Здесь $f(i)$ понимается как интегральный дисконтированный эффект для оптимальной производственной программы в случае бесконечного планового периода, если уровень запасов на начало текущего отрезка равен i.

Отметим некоторые различия между выражениями (1) и (2). Важнейшее из них состоит в том, что и в левой, и в правой части соотношения (2) находится *одна и та же* функция $f(i)$. Поэтому выражение (2) называется **функциональным**, или **экстремальным, уравнением.** По существу, это *система* уравнений, где каждому из возможных значений переменной состояния, т. е. начального уровня запасов i, соответствует одно уравнение. Следовательно, (2) представляет собой *функцию* $f(i)$ для всего диапазона значений переменной состояния. Как читатель помнит из рассмотрения простого функционального уравнения, составленного для модели восстановления в разд. 11.4, необходимо определить, имеет ли система уравнений (2) однозначное и конечное решение.

Это можно выяснить, анализируя данную конкретную модель. Однако можно добиться и более существенных успехов, анализируя функциональные уравнения более общей модели, по отношению к которой модель управления запасами является всего лишь частной (полученной путем добавления нескольких слабых ограничений). В процессе подготовки к анализу этой общей модели рассмотрим возможность представления (2) с помощью сетевой оптимизационной модели.

Предположим, что число возможных значений начального уровня запасов является конечным; пусть каждая из вершин сети соответствует одному из этих уровней. Когда мы находимся в вершине i, решение производить количество продукции x можно представить в виде дуги из вершины i в вершину $(i + x - d)$. Соответствующая стоимость перемещения по дуге равна $C(x) + h(i + x - d)$. Вместо $f(i)$ введем в (2) обозначение y_i, так что y_i соответствует интегральному дисконтированному эффекту, отвечающему вершине i. Полученная сеть близко напоминает сеть, приведенную в разд. 11.9, отличаясь от нее лишь тем, что не имеет вершины конечного состояния, так что маршрут, соответствующий значению y_i, по существу является бесконечным.

Характеристика сети. Рассмотрим сеть, включающую p вершин и множество ориентированных дуг, которые соединяют между собой некоторые из вершин. Иными словами, предположим, что из каждой вершины можно переместиться по крайней мере в какую-либо другую вершину. Каждой дуге (i, j) соответствует эффект или затраты c_{ij}, причем время перемещения из i в j равно одному отрезку. Как и выше, коэффициент дисконтирования обозначим α.

Пусть маршрут начинается в некоторой произвольно выбранной вершине i. Предположим, что из вершины i мы направляемся в вершину j и, следовательно, производим затраты c_{ij}. Предположим далее, что из вершины j мы направляемся в вершину k, причем возникают дисконтированные затраты αc_{jk}. Если процесс продолжается неограниченное время, маршрут является бесконечным. Дополнительно добавляются все новые затраты, однако умноженные на все более высокие степени коэффициента дисконтирования α, где $\alpha < 1$. Обозначим через y_i интегральные дисконтированные затраты для *оптимального* бесконечного маршрута, который начинается в вершине i.

Если принятая стратегия является стационарной, то каждый раз, вернувшись к вершине, мы снова выбираем ту же дугу, которая была выбрана при предыдущем заходе в эту вершину. Пусть существует стационарная стратегия, которая является оптимальной; тогда соответствующее y_i удовлетворяет функциональным уравнениям

$$y_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha y_j + c_{ij}) \text{ для всех вершин } i, \ 0 \leqslant \alpha < 1. \tag{3}$$

(Задача нахождения маршрута с минимальными средними затратами за отрезок рассматривается в разд. 12.2.) Заметим, что ни одна из вершин не рассматривается в качестве вершины конечного состояния. В этой модели удобно допустить существование дуг (i, i), поскольку оптимальный маршрут для бесконечного планового периода может включать повторные перемещения из вершины i в нее же. (Так, сеть, которая соответствует модели восстановления с *бесконечным* числом этапов, приведенной в разд. 11.4, включает только одну вершину и N дуг, начинающихся и кончающихся в этой же вершине.)

Сформулировав экстремальные уравнения (3), мы допустили существование оптимальной стационарной стратегии и соответствующего y_i. Однако, для того чтобы использовать (3) на практике, необходимо уметь ответить на следующие вопросы:

I) Всегда ли можно для всех значений y_i отыскать однозначное и конечное решение?

II) Если всегда, то является ли соответствующая стационарная стратегия действительно оптимальной?

Иными словами, прежде чем применять численные методы, необходимо удостовериться, всегда ли имеют решения функциональные уравнения (3). Если такие решения существуют, дуга (i, j), позволяющая достичь минимума для вершины (i), является элементом следующей **стационарной стратегии**: находясь в вершине i, *всегда* выбирайте дугу (i, j). Поскольку число вершин является конечным, в конце концов придется снова и снова перемещаться по одному и тому же набору дуг. Конкретный характер такого **цикла** может зависеть от местонахождения в начальный момент времени. Поэтому второй вопрос заключается в том, нельзя ли улучшить результаты, перейдя к другой стационарной стратегии или же к нестационарной стратегии, т. е. к такой стратегии, при которой учитывается, по какому маршруту мы попали в вершину i?

Ответы как на вопрос I), так и на вопрос II) являются утвердительными. Можно строго показать следующую теорему.

Теорема о стационарной стратегии. Всегда существует однозначно определенное конечное y_i, удовлетворяющее экстремальным уравнениям (3), и соответствующая стационарная стратегия является оптимальной — лучшей из всех допустимых стратегий.

Существование стационарной стратегии, которая является истинным оптимумом, доказывается и для того случая, когда в качестве критерия оптимальности принят минимум эквивалентных средних затрат за отрезок при $\alpha = 1$. Однако в этом случае система (3) не является адекватной системой функциональных уравнений; данный случай рассматривается в следующем разделе.

С помощью почти тривиальных изменений модель можно обобщить на ситуацию, при которой для перемещения по дуге требуется более одного отрезка. Ввиду необходимых изменений системы обозначений формулы кажутся сложными (почти во всех формулах α заменяется на α_{ij}).

В конце разд. 10.7 приводилась достаточно общая каноническая форма модели для конечного планового периода. Если включить коэффициент дисконтирования α, рекуррентное соотношение примет вид

$$f_n(s) = \underset{d_n \text{ в } D_n(s)}{\text{optimum}} \{R_n(s, d_n) + \alpha f_{n-1}[T_n(s, d_n)]\} \quad \begin{matrix}\text{для всех } s \text{ в } S_n,\\ \text{при } f_n(s) \equiv 0,\end{matrix} \qquad \text{(I)}$$

где s — состояние системы, S_n — множество всех возможных состояний на этапе n, d_n — принимаемое решение на этапе n, $D_n(s)$ — множество всех допустимых значений d_n, $R_n(s, d_n)$ — непосредственный экономический эффект, обусловленный решением d_n, если система находится в состоянии s, $T_n(s, d_n)$ — преобразованное состояние системы на следующем этапе.

Предположим, что структура модели является стационарной, так что в (I) можно для каждого n ввести упрощенные обозначения S, $D(s)$, $R(s, d)$ и $T(s, d)$. Допустим, далее, что существует оптимальная стратегия, являющаяся стационарной. Тогда естественным аналогом (I) для бесконечного числа этапов будет

$$\text{optimum}\ \{R(s, d) + \alpha f[T(s, d)]\}\ \text{для всех}\ s\ \text{в}\ S,\ 0 \leqslant \alpha < 1. \qquad (II)$$

Постулируем также, что система может иметь конечное число состояний, которые для удобства пронумеруем, присвоив им индексы $s = 1, 2, \ldots, p$, а также что $D(s)$ включает конечное число решений. При наличии этих двух структурных предположений будем называть модель (II) **дискретной задачей динамического программирования.**

Если $T(s, d) = T(s, d')$ только при $d = d'$, то существует однозначное соответствие между сетевыми экстремальными уравнениями (2) и (II). Каждое состояние s соответствует вершине s, а $f(s) = y_s$. Решение d отображается дугой из вершины s в вершину $T(s, d)$, причем $R(s, d)$ — соответствующая стоимость перемещения по этой дуге.

Рассмотрим сеть, в которой для перемещения по каждой дуге (i, j) требуется время, равное h_{ij} отрезков, причем не все h_{ij} должны быть *обязательно равны одному отрезку. Определим соответствующий* коэффициент дисконтирования следующим образом:

$$\alpha_{ij} \equiv \alpha^{h_{ij}}. \qquad (III)$$

Тогда экстремальные уравнения для этой более общей задачи записываются не в форме системы (3), а в следующем виде:

$$y_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha_{ij} y_j + c_{ij})\ \text{для всех вершин}\ i (0 \leqslant \alpha \leqslant 2). \qquad (IV)$$

12.2. МЕТОДЫ ПОСЛЕДОВАТЕЛЬНЫХ ПРИБЛИЖЕНИЙ

В разд. 11.5, посвященном модели восстановления, были рассмотрены многие важные идеи, лежащие в основе нескольких методов последовательных приближений. Все приведенные ранее соображения в равной мере применимы и здесь, так что, прежде чем читатель перейдет к детальному изучению этих методов, для него может представить интерес краткий обзор упомянутых идей.

Вкратце рассмотрим два подхода. В случае первого из них мы пытаемся угадать неизвестное значение функции y_i и на каждой итерации улучшаем проверяемое приближенное значение. При втором подходе мы пытаемся угадать оптимальную стратегию. Если *при использовании пробной стратегии выполняется экстремальное* уравнение, вычислительный процесс прекращается, если нет, исследуется новая пробная стратегия. По сравнению с первым подходом при втором подходе требуется больший объем вычислительных операций на каждой итерации, однако всегда обеспечивается сходимость за конечное число итераций.

Метод итераций по критерию. Решение функциональных уравнений можно отыскивать **методом последовательных приближений**

в пространстве функций. Подобно тому, как и в гл. 11, будем считать, что y_i^0 — произвольно выбранные значения y_i, и применим метод итераций по критерию:

$$y_i^{n+1} = \min_{\substack{\text{по всем } (i, j) \\ \text{в сети}}} (\alpha y_j^n + c_{ij}) \quad \text{для всех } i. \tag{1}$$

Тогда каждое из y_i^n стремится к некоторому пределу, однако, как было показано для простой модели восстановления в разд. 11.4, сходимость y_i^n *не обязательно* обеспечивается за конечное число итераций. Более того, даже если в течение нескольких итераций (I) выбирается одна и та же стратегия, это еще *не означает*, что она действительно является оптимальной для бесконечного планового периода. По существу даже при сколь угодно большом числе итераций не обязательно существует «сходимость стратегии», т. е. не обязательно существует единственная стратегия, которая оптимальна для *всех* конечных $n \geqslant N$, сколь бы ни велико было N.

Выбрав в качестве начального приближения все значения y_i^0 нулевыми, мы приходим к такому же вычислительному процессу, как и при рассмотрении некоторой фиксированной модели с конечным плановым периодом. В частности, y_i^n отображают интегральные дисконтированные затраты для оптимального маршрута длиной n дуг, начинающегося в вершине i. По аналогии с теоремой о длительности планового периода для модели восстановления, приведенной в разд. 11.5, можно говорить о существовании такого n^*, что для каждого $n > n^*$ *любая* стратегия, являющаяся оптимальным текущим решением в модели с конечным плановым периодом длительностью n отрезков, одновременно есть и оптимальная стационарная стратегия для модели с бесконечным плановым периодом. Более того, при известных c_{ij} и α можно вычислить значение n^*, позволяющее применить данный подход.

Если все $c_{ij} \geqslant 0$ и все $y_i^0 = 0$, значения y_i^n возрастают монотонно. Сходимость сверху обеспечивается для *любого* c_{ij}, если вначале выбрана некоторая пробная стратегия и по приведенной ниже формуле (2) вычислены соответствующие этой стратегии значения y_i^0. Тогда, если прекратить расчеты на итерации m и использовать промежуточную оптимальную стратегию, полученную методом итераций по критерию (I), в качестве стационарной стратегии для модели с бесконечным плановым периодом соответствующие значения интегральных дисконтированных показателей не будут превышать y_i^m в (I).

Доказательство сходимости y_i^n к y_i при произвольно выбранных y_i^0 аналогично доводам, приведенным в разд. 11.6 для модели восстановления с бесконечным числом этапов. Проведем доказательство для сети более общего вида, в которой коэффициент дисконтирования для дуги (i, j) равен α_{ij}.

Пусть для каждой вершины i дуга (i, j^*) оптимальна в модели с бесконечным плановым периодом, а дуга (i, j) оптимальна на итерации n, выполняемой в соответствии с (I). Тогда

$$y_i = \alpha_{ij^*} y_{j^*} + c_{ij^*} \quad \text{для всех } i, \tag{I}$$

$$y_i^n \leqslant \alpha_{ij^*} y_{j^*}^{n-1} + c_{ij^*} \quad \text{для всех } i, \tag{II}$$

так что

$$y_i^n - y_i \leqslant \alpha_{ij^*} (y_{j^*}^{n-1} - y_{j^*}) \leqslant \alpha d_{n-1}, \tag{III}$$

где
$$d_{n-1} \equiv \max_k | y_k^{n-1} - y_k |.$$

Аналогично этому

$$y_i^n = \alpha_{ij} y_j^{n-1} + c_{ij}, \tag{IV}$$

$$y_i \leqslant \alpha_{ij} y_j + c_{ij}, \tag{V}$$

так что

$$y_i - y_i^n \leqslant \alpha_{ij} (y_j - y_j^{n-1}) \leqslant \alpha d_{n-1}. \tag{VI}$$

Следовательно,

$$| y_i^n - y_i | \leqslant \alpha d_{n-1} \leqslant \alpha^n d_0. \tag{VII}$$

Поскольку $0 \leqslant \alpha < 1$, отсюда следует сходимость.

Пусть все $c_{ij} = 0$ и все $y_i^0 = 0$. Тогда монотонное возрастание y_i^n можно установить на основе следующих *индуктивных* соображений. Для всех i при $n = 1$

$$y_i^1 = \min_{\substack{\text{по всем } (i, j) \\ \text{в сети}}} (\alpha_{ij} 0 + c_{ij}) \geqslant 0 = y_i^0, \tag{VIII}$$

так что $y_i^1 \geqslant y_i^0$. Предположим, что для $n = 1, 2, \ldots, m$ и для всех i $y_i^n \geqslant y_i^{n-1}$. Пусть на итерации m дуга (i, j) является оптимальной. Тогда

$$y_i^{m+1} = \alpha_{ij} y_j^m + c_{ij} \geqslant \alpha_{ij} y_j^{m-1} + c_{ij} \geqslant \min_{\substack{\text{по всем } (i, j) \\ \text{в сети}}} (\alpha_{ij} y_j^{m-1} + c_{ij}) = y_i^m \tag{IX}$$

и, следовательно, $y_i^n \geqslant y_i^{n-1}$ при $n = m + 1$.

Аналогично этому если величины y_i^0 вычисляются в соответствии с первоначально выбранной стратегией, то сходимость сверху можно доказать по индукции. Для $n = 1$ и для всех i

$$y_i' = \min_{\substack{\text{по всем } (i, j) \\ \text{в сети}}} (\alpha_{ij} y_j^0 + c_{ij}) \leqslant \alpha_{ik} y_k^0 + c_{ik} = y_i^0, \tag{X}$$

где (i, k) — дуга, соответствующая начальной стратегии, так что $y_i^1 \leqslant y_i^0$. Предположим, что для $n = 1, 2, \ldots, m$ и для всех i $y_i^n \leqslant y_i^{n-1}$. Пусть на итерации m дуга (i, j) является оптимальной. Тогда

$$y_i^{m+1} \leqslant \alpha_{ij} y_j^m + c_{ij} \leqslant \alpha_{ij} y_j^{m-1} + c_{ij} = y_i^m \tag{XI}$$

и, следовательно, $y_i^n \leqslant y_i^{n-1}$ при $n = m + 1$.

Приводимый ниже пример, основанный на сетевом представлении выбора между вариантами B и D (которые приведены в таблице рис. 11.1), позволяет при $\alpha = 0{,}5$ продемонстрировать сходимость y_i^n без соответствующей сходимости стратегий. Пусть сеть имеет 4 вершины, однако несколько возможных решений будет только в вершине *1*:

$$\begin{array}{llll} \text{вершина } 1: & c_{12} = 2^2/_3, \quad c_{13} = 3; & \text{вершина } 2: & c_{22} = 2; \\ \text{вершина } 3: & c_{34} = 1; & \text{вершина } 4: & c_{43} = 3, \end{array} \tag{I}$$

и пусть $\alpha = 0{,}5$. Начертим сеть, соответствующую (I). Убедимся также в том, что

$$y_1 = 4\frac{2}{3}, \quad y_2 = 4, \quad y_3 = 3\frac{1}{3}, \quad y_4 = 4\frac{2}{3} \tag{II}$$

удовлетворяют функциональным уравнениям и что *обе* дуги (*1*, *2*) и (*1*, *3*) являются оптимальными.

Реализацию алгоритма итераций в пространстве функций начнем, приняв все $y_i^0 = 0$. Тогда на итерации n, где $n = 1, 2, 3$, получим:

$$y_1^1 = \min\,(0{,}5\times0+2\frac{2}{3};\ 0{,}5\times0+3) = 2\frac{2}{3} \text{ для дуги } (1, 2); \tag{III}$$

$$y_2^1 = 0{,}5\times0+2=2;\ y_3^1=0{,}5\times0+1=1;\ y_4^1=0{,}5\times0+3=3;$$

$$y_2^1 = \min\,(0{,}5\times2+2\frac{2}{3};\ 0{,}5\times1+3) = 3\frac{1}{2} \text{ для дуги } (1, 3); \tag{IV}$$

$$y_2^2=0{,}5\times2+2=3;\quad y_3^2=0{,}5\times3+1=2\frac{1}{2};\ y_4^2=0{,}5\times1+3=3\frac{1}{2};$$

$$y_1^3=\min\,(0{,}5\times3+2\frac{2}{3};\ 0{,}5\times2\frac{1}{2}+3) = 4\frac{1}{6} \text{ для дуги } (1, 2); \tag{V}$$

$$y_2^3=0{,}5\times3+2=3\frac{1}{2};\ y_3^2=0{,}5\times3\frac{1}{2}+1=2\frac{3}{4};\ y_4^3=0{,}5\times2\frac{1}{2}+3=4\frac{1}{4}.$$

При увеличении n значения y_i^n стремятся к значениям y_i в (II), однако при нечетных n всегда позволяет достичь минимума дуга (*1*,*2*), а при четных — дуга (*1*, *3*). Следовательно, «сходимость стратегий» отсутствует.

Метод итераций по стратегиям. Экстремальные уравнения (I) можно решать с помощью **алгоритма последовательных приближений в пространстве стратегий,** который имеет следующий вид.

Шаг 1. Выберем произвольную исходную стратегию и примем $n = 0$.

Шаг 2. Для заданной пробной стратегии вычислим значения y_i^n в соответствии с уравнениями расчета «стоимости вершин»:

$$y_i^n = \alpha y_j^n + c_{ij} \quad \text{или} \quad y_i^n - \alpha y_j^n = c_{ij} \tag{2}$$

для всех вершин i (формула расчета «стоимости вершин»), где дуга (i, j) отображает решение для вершины i, соответствующее конкретной пробной стратегии.

Шаг 3. Проверим возможность дальнейших улучшений, вычислив

$$\min_{\substack{\text{по всем } (i,\, j)\\ \text{в сети}}} (\alpha y_j^n + c_{ij}) \text{ для всех } i \text{ (проверка возможности улучшения).} \quad (3)$$

Шаг 4. Прекратим расчеты, если $Y_i^n = y_i^n$ для всех i. В противном случае изменим стратегию для каждой вершины k, для которой $Y_k^n < y_k^n$, выбрав дугу, позволяющую достичь в (3) значения Y_k^n. Перейдем от итерации n к итерации $(n + 1)$ и обратимся к выполнению шага 2 на основе новой пробной стратегии.

Техника и логика метода итераций по стратегиям. Отметим, что в формуле расчета стоимости (2) требуется решить систему линейных уравнений. Поскольку имеется по одному уравнению и по одной переменной y_i^n для каждой вершины i и поскольку $0 \leqslant \alpha < 1$, система (2) всегда имеет единственное решение. Применяя формулу (2), читатель обнаружит, что удобно вначале вычислять значения y_i^n, входящих в цикл, а лишь затем — остальных переменных, выражая их через ранее вычисленные переменные. Вычисления прекращаются в тот момент, когда для всех i выполняется условие $Y_i^n = y_i^n$, поскольку в этом случае значения y_i^n удовлетворяют экстремальным уравнениям.

Если в (3) $i = k$, величину $(\alpha y_j^n + c_{kj})$ можно интерпретировать как интегральные дисконтированные затраты на перемещение из вершины k в вершину j; *затем* используется *полная* стратегия, проверяемая на итерации n и включающая дугу, *ранее* зафиксированную для вершины k в составе пробной стратегии, проверяемой на итерации n. Если при этом для некоторого k выяснится, что при выборе дуги (k, j) $Y_k^n < y_k^n$, то, вероятно, решение можно дополнительно улучшить, выбирая эту дугу *всякий раз*, когда мы окажемся в вершине k. Новое значение y_k^{n+1} отображает именно такую пересмотренную стратегию.

Таким образом, можно сделать важный вывод: если при использовании некоторой стратегии *постоянно* выбирается дуга (k, j), то можно проверить, нельзя ли улучшить решение, используя такой критерий, при котором дуга (k, j) выбирается только в качестве *текущего* решения! (Этот вывод следует прочесть повторно, потому что полученный глубокий результат вовсе не очевиден.)

Метод последовательных приближений в пространстве стратегий обладает следующими свойствами:

I) $y_i^{n+1} \leqslant y_i^n$ для любой вершины i и $y_k^{n+1} < y_k^n$, если $Y_k^n < y_k^n$.

II) Алгоритм является конечным, т. е расчеты выполняются за конечное число итераций.

III) Стратегия, позволяющая по завершению расчетов достичь Y_i^n, является оптимальной.

Как будет более четко показано в разд. 12.5, метод последовательных приближений в пространстве стратегий обладает сформулированными выше свойствами потому, что он весьма близок к симплексному методу. На каждой итерации текущая *стратегия* является допустимой; однако соответствующие значения переменных y_i^n могут и не быть допустимыми в том смысле, что они могут не удовлетворять экстремальным уравнениям. Правило прекращения расчетов представляет собой проверку допустимости. Доказательство сходимости метода итераций по стратегиям к оптимальному решению за конечное число итераций весьма напоминает доказательство того, что при симплексном методе число итераций до прекращения расчетов является конечным. Суть доказательства состоит в том, что ввиду I) мы никогда не возвращаемся к ранее проверенной стратегии. Поскольку число возможных стратегий является конечным, в конце концов итеративный процесс прекращается.

При сравнении методов последовательных приближений в пространстве функций и в пространстве стратегий заметим, что одни и те же расчеты выполняются для нахождения значений правых частей в (4) и левых частей в (3). В методе итераций по критерию полученные на некоторой итерации значения используются на следующей итерации в качестве новых приближенных значений y_i. Напротив, новые приближенные значения y_i в методе итераций по стратегиям вычисляются с помощью дополнительных расчетов, а именно в результате решения уравнений расчета «стоимости вершин» (2) на следующей итерации. Эти дополнительные вычисления есть цена, которую мы платим за обеспечение сходимости за конечное число итераций.

Метод итераций по стратегиям демонстрируется на численных примерах в разд. 12.4.

В сети, где для дуги (i, j) коэффициент дисконтирования равен α_{ij}, замена α на α_{ij} необходима лишь в формулах (1) — (3).

12.3. МИНИМИЗАЦИЯ СРЕДНЕГО ЭФФЕКТА ЗА ОТРЕЗОК

Нельзя ожидать, что экстремальные уравнения относительно y_i в разд. 12.1 окажутся пригодными при $\alpha = 1$, поскольку в этом случае суммарный эффект для бесконечного планового периода может быть бесконечно большим (например, если все $c_{ij} > 0$). Для того чтобы построить требуемое функциональное уравнение, перепишем все формулы с использованием эквивалентного среднего эффекта за отрезок, а затем примем $\alpha = 1$.

Начнем с предположения, что нам удалось вычислить эквивалентный средний эффект за отрезок для каждого из возможных на данной сети циклов и $\bar{c}$ соответствует наименьшему из значений эффекта при $\alpha = 1$. Разумеется, при решении практической задачи нежелательно вычислять все значения среднего эффекта, поэтому, как правило, истинное значение $\bar{c}$ остается неизвестным до завершения оптимизации. Однако в приводимых далее алгебраических преобразованиях $\bar{c}$ следует рассматривать лишь как символ, обозначающий такую величину, которая, несомненно, существует.

Для того чтобы избежать необходимости рассмотрения многих частных случаев, дополнительно примем, что сеть обладает одним свойством: *предполагается, что существует маршрут, точнее, направленный маршрут, ведущий от любой вершины i к одной (по крайней мере) из вершин, входящих в цикл, который позволяет достичь $\bar{c}$.* (Это условие удовлетворяется, например, в том случае, если существует направленный маршрут от любой вершины i к любой вершине j.) Подобный подход разъясняется ниже; он может быть модифицирован таким образом, что позволит рассматривать и сети, не обладающие указанным свойством.

Построим функциональные уравнения, включающие $\bar{c}$. Как читатель увидит далее, эти уравнения включают и другие необходимые для расчетов величины, а именно w_i. Следует предупредить, что приводимые ниже соображения не отвечают полностью требованиям строгого математического анализа; они приводятся здесь лишь для объяснения конечных результатов, которые сами по себе могут показаться странными. Поэтому если читатель нетерпелив, он может опустить эвристические соображения и перейти непосредственно к экстремальным уравнениям (5).

Величину эквивалентного среднего эффекта для каждой вершины i, равную $(1-\alpha)\,y_i$, удобно связать с $\bar{c}$ посредством следующего соотношения, определяющего w_i:

$$(1-\alpha)\,y_i \equiv (1-\alpha)\,w_i + \bar{c}, \quad 0 \leqslant \alpha < 1 \tag{1}$$

(покажите, какой смысл вложен в величину w_i). Таким образом,

$$y_i \equiv w_i + \frac{\bar{c}}{1-\alpha}, \quad 0 \leqslant \alpha < 1. \tag{2}$$

Можно интерпретировать w_i как разность между y_i и $\bar{c}/(1-\alpha)$, причем последняя величина есть интегральный дисконтированный эффект от получения $\bar{c}$ на каждом отрезке в течение бесконечного планового периода. [Уже сама система обозначений является неточной, поскольку и y_i, и w_i зависят от α, так что правильнее было бы использовать обозначения $y_i(\alpha)$ и $w_i(\alpha)$.]

Тогда экстремальное уравнение, приведенное в разд. 12.1, можно переписать в следующем виде:

$$w_i + \frac{\bar{c}}{1-\alpha} = \min_{\substack{\text{по всем } (i,\,j) \\ \text{в сети}}} \left[\alpha\left(w_j + \frac{\bar{c}}{1-\alpha}\right) + c_{ij}\right] \quad \text{для всех вершин } i. \tag{3}$$

Поскольку $\bar{c}$ есть величина постоянная, слагаемые, включающие $\bar{c}$, можно объединить либо перенести одно из таких слагаемых в правую часть (3) из левой, либо в левую часть из правой. Мы воспользуемся первой из этих возможных альтернатив, поскольку она допускает обобщение на случай различных коэффициентов дисконтирования α_{ij} для каждой дуги (i, j). Таким образом, как нетрудно проверить, выражение (3) можно представить в следующем виде:

$$w_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha w_j + c_{ij} - \bar{c}). \tag{4}$$

При $\alpha = 1$ выражение (4) упрощается:

$$w_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j + c_{ij} - \bar{c}) \quad \text{для всех вершин } i. \tag{5}$$

Если перенести постоянную $\bar{c}$ в левую часть, получим эквивалентное выражение

$$w_i + \bar{c} = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j + c_{ij}) \quad \text{для всех вершин } i. \tag{6}$$

(При точной системе обозначений w_i в (5) и (6) принимается равным предельному значению $w_i(\alpha)$ при стремлении α к 1 снизу.)

До сих пор полагалось, что преобразование приведенных ранее экстремальных уравнений, соответствующих $\alpha < 1$, в экстремальные уравнения (5), где $\alpha = 1$, является оправданным. При таком преобразовании предполагается существование стационарной стратегии, оптимальной в том смысле, что она включает цикл с минимальными эквивалентными средними затратами за отрезок. Такое предположение оправдано, поскольку можно доказать следующую теорему.

Т е о р е м а о с т а ц и о н а р н о й с т р а т е г и и. Всегда существуют единственное конечное $\bar{c}$ и конечные w_j, удовлетворяющие экстремальным уравнениям (5), причем соответствующая стационарная стратегия включает цикл, для которого эквивалентные затраты за отрезок являются минимальными по всему множеству возможных стратегий.

Как и в случае $\alpha < 1$, дуга, которой соответствует минимум (5) или (6), отображает решение, которое следует принимать, находясь в вершине i. Такая стратегия должна включать не менее одного цикла. Стратегия соответственно представляется в виде маршрута, начинающегося в рассматриваемой вершине i и ведущего к циклу, для которого эквивалентные средние затраты за отрезок являются минимальными. (Разумеется, в цикл должна входить вершина i.) Если оптимальная стационарная стратегия включает более чем один цикл, то для всех циклов эквивалентные средние затраты за отрезок должны быть одинаковыми. Это следует из сделанного ранее допущения, что всегда можно отыскать маршрут из любой вершины i к вершине, входящей в цикл, где достигается $\bar{c}$. Поскольку $\bar{c}$ связано с циклом, а величины c_{ij} стационарны, значение $\bar{c}$ одновременно является средними затратами за отрезок, минимальными для всех возможных циклов. Это значит, что подход на основе «эквивалентного среднего эффекта за отрезок» позволяет найти стратегию, оптимальную по критерию «средний эффект за отрезок».

Как упоминалось выше, величины w_i представляют интерес главным образом с вычислительной точки зрения — для того чтобы вычислить $\bar{c}$ и найти оптимальную стратегию, необходимо определить значения w_i. Однако существует несколько интерпретаций экономического смысла w_i. Вот одна из них. Из (2) очевидно, что для любой пары i и j

$$y_i - y_j \equiv w_i - w_j, \quad 0 \leqslant \alpha < 1.$$

Таким образом, предельное значение *разности* между интегральными дисконтированными показателями для вершин i и j при стремлении α к 1 равно $w_i - w_j$. Следовательно, $w_i - w_j$ можно рассматривать как *приращение* затрат, обусловленное нахождением в вершине i, а не в вершине j при $\alpha = 1$.

Метод итераций по стратегиям. При $\alpha = 1$ для решения (5) или (6) может быть использован метод последовательных приближений в пространстве стратегий. Для удобства изложения будем предполагать, что *на каждой итерации проверяемая стратегия включает один цикл.* Дополнительные указания, требуемые для расчетов при наличии на некоторой итерации более одного цикла, приводятся в конце настоящего раздела.

Если найдены значения w_i, удовлетворяющие экстремальным уравнениям (5) и (6), добавление одной и той же константы к этим значениям дает новые значения, также удовлетворяющие таким уравнениям. Таким образом, значения w_i не являются единственными, и в приводимом ниже алгоритме удобно принять одно из них равным нулю. Для определенности примем $w_1 = 0$.

Алгоритм включает следующие операции:

Шаг 1. Выберем произвольную исходную стратегию и примем $w_1 = 0$.

Шаг 2. Для заданной пробной стратегии вычислим значения w_i^n и $\bar{c}^n$ в соответствии с уравнениями расчета «стоимости вершин»

$$w_i^n - w_j^n + \bar{c}^n = c_{ij} \quad \text{для всех вершин } i \text{ (формулы расчета «стоимости вершин»)}, \qquad (7)$$
$$w_1^n = 0,$$

где дуга (i, j) отображает решение для вершины i, соответствующее конкретной пробной стратегии.

Шаг 3. Проверим возможность дальнейшего улучшения, вычислив

$$\min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j^n + c_{ij} - \bar{c}^n) \equiv W_i^n \text{ (показатели возможного улучшения)}. \qquad (8)$$

Шаг 4. Прекратим расчеты, если $W_i^n = w_i^n$ для всех i. В противном случае изменим стратегию для каждой такой вершины k, для которой $W_k^n < w_k^n$, выбрав дугу, позволяющую достичь в (8) значения W_k^n. Перейдем от итерации n к итерации $(n + 1)$ и обратимся к выполнению шага 2 на основе новой проверяемой стратегии.

Техника решения. Заметим, что в (7) требуется решить систему линейных уравнений. Если проверяемая стратегия включает один цикл, уравнения (7) всегда имеют единственное решение. Отметим также, что, если просуммировать уравнения (7) для всех вершин, входящих в цикл, w_i^n взаимно уничтожаются, так что в результате получим

$$\bar{c}^n = \sum_{\substack{\text{по всем } (i, j) \\ \text{в цикле}}} c_{ij}/\text{число дуг в цикле.} \tag{9}$$

Мы провели формулу расчета показателей возможного улучшения (8), основанную на экстремальных уравнениях (5). Следует отметить, что слагаемое $\bar{c}^n$ в (8) не влияет на результат минимизации. Поэтому операция вычитания не обязательна при нахождении минимума, она требуется лишь при исчислении результирующего значения W_i^n. Эту формулу удобно представить в виде (8), поскольку она может быть обобщена на случай различной длительности перемещения по отдельным дугам.

Приняв должные меры предосторожности, чтобы избежать зацикливания, аналогичного возможному зацикливанию при использовании симплексного метода, получаем алгоритм, обладающий следующими свойствами:

I) На каждой итерации $\bar{c}^n \leqslant \bar{c}^{n-1}$.

II) Вычисления прекращаются после конечного числа итераций.

III) Оптимальной по завершению расчетов является стратегия, позволяющая достичь W_i^n.

Пусть h_{ij} — число отрезков, необходимое для перемещения по дуге (i, j), и $\alpha_{ij} \equiv \alpha^{h_{ij}}$. Тогда (3), (4) и (5) можно представить в следующем виде:

$$w_i + \frac{\bar{c}}{1-\alpha} - \min_{\substack{\text{по всем } (i, j) \\ \text{в сети}}} \left[\alpha_{ij}\left(w_j + \frac{\bar{c}}{1-\alpha}\right) + c_{ij}\right], \tag{I}$$

$$w_i = \min_{\substack{\text{по всем } (i, j) \\ \text{в сети}}} [\alpha_{ij} w_j + c_{ij} - (1 + \alpha + \ldots + \alpha^{h_{ij}-1})\bar{c}], \tag{II}$$

$$w_i = \min_{\substack{\text{по всем } (i, j) \\ \text{в сети}}} (w_j + c_{ij} - h_{ij}\bar{c}). \tag{III}$$

При учете коэффициентов h_{ij} изменяется соответственно значение $\bar{c}^n$ в (7) и (8), а знаменатель в (9) становится равным сумме h_{ij} по дугам (i, j), входящим в цикл.

Даже если исходная стратегия включает только один цикл, стратегии, проверяемые на последующих итерациях, могут включать два или более цикла. В этом случае необходима незначительная модификация алгоритма. Поскольку в любой пробной стратегии каждая из вершин i либо входит в один из циклов, либо находится на маршруте к одному из циклов, множество всех вершин можно разбить на непересекающиеся группы, по одной группе для каждого цикла. Пусть $\bar{c}_i^n$ обозначает средний эффект за отрезок для цикла, с которым связана вершина i [мы будем по-прежнему использовать систему обозначений, принятую для сети общего типа, в которой h_{ij} — число отрезков, необходимое для перемещения по дуге (i, j)].

Тогда, если дана пробная стратегия, необходимо решить систему

$$w_i^n - w_j^n + h_{ij}\bar{c}_j^n = c_{ij} \quad \text{для всех вершин } i, \tag{I}$$

где все $\bar{c}_j^n$, связанные с одним и тем же циклом, считаются одинаковыми, причем в каждом из циклов произвольно выбирается одна из вершин i, для которой принимается $w_i^n = 0$.

[Так, например, если один из циклов включает три вершины 1, 2 и 3, то дополнительно к (I) строятся уравнения $\bar{c}_1^n = \bar{c}_2^n = \bar{c}_3^n$ и, скажем, $w_1^n = 0$.]

Решив систему (I), найдем

$$\min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\bar{c}_j^n) \equiv \bar{C}_i^0 \quad \text{для всех } i. \tag{II}$$

Если для некоторого k выполняется условие $\bar{C}_k^n < \bar{c}_k^n$, то соответствующим образом изменим стратегию на основе (II). Если при этом новая стратегия по-прежнему включает более одного цикла, снова решим систему (I), в противном случае возвратимся к (7). Если же условие $\bar{C}_k = \bar{c}_k^n$ выполняется для всех k, то $\bar{c}_k^n$ одинаковы также для всех k; приняв $\bar{c}^n$ равным этому общему значению, выполним, как и прежде, расчет показателей возможного улучшения по формуле (8).

Можно глубже понять выражения (5), связав их с моделью нахождения кратчайшего пути на сети общего типа. Введем обозначение:

$$\bar{c}_{ij} = c_{ij} - h_{ij}\bar{c}. \tag{III}$$

Тогда (5) можно представить в виде

$$w_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j + \bar{c}_{ij}) \quad \text{для всех } i \text{ при } w_1 = 0. \tag{IV}$$

Существует одно различие между (IV) и обычной моделью нахождения кратчайшего пути (например, в таком ее виде, в каком она изложена в разд. 11.9): хотя $w_1 = 0$, вершина *1* в действительности не является вершиной конечного состояния. Стратегия вполне может включать дугу из вершины *1* к вершине j.

Пусть в угаданной стратегии $\bar{c}$ есть минимальные эквивалентные средние затраты за отрезок, соответствующие некоторому циклу. Тогда сумма $\bar{c}_{ij}$ по дугам этого цикла равна нулю. Если для каждого из остальных возможных циклов на сети сумма $\bar{c}_{ij}$ является неотрицательной, предлагаемая стратегия является оптимальной. В этом случае нахождение кратчайшего пути методом последовательных приближений в пространстве функций приводит к решению (IV). Если же, однако, существует цикл, для которого средние затраты, исчисленные в текущих $\bar{c}_{ij}$, являются отрицательными, то имеется возможность улучшения, и алгоритм нахождения кратчайшего пути, изложенный в разд. 7.6 и 11.9, оказывается непригодным. Выполняя операции алгоритма, мы раньше или позже обнаружим, что повторно пересматриваем w_i, относящиеся к тому циклу, который представляется в качестве возможного улучшения.

Таким образом, возникает следующий «смешанный» алгоритм. Выберем стационарную стратегию и вычислим результирующее пробное значение $\bar{i}$. Применим алгоритм итераций по критерию (5), приведенный в разд. 11.9, где $\bar{c}_{ij}$ исчисляются по формуле (III). Примем $w_i^0 = \infty$ для всех $i \neq 1$. Если при расчетах обнаруживается цикл, для которого соответствующая сумма $\bar{c}_{ij}$ отрицательна, должным образом изменим стратегию, повторно вычислим $\bar{c}$ для этого улучшающего цикла, пересчитаем $\bar{c}_{ij}$ и повторно выполним операции. Как только вычисления по алгоритму нахождения кратчайшего пути завершатся получением значений w_i, мы найдем, что соответствующее $\bar{c}$ оптимально.

12.4. ДЕМОНСТРАЦИЯ МЕТОДА ИТЕРАЦИЙ ПО СТРАТЕГИЯМ НА ЧИСЛОВЫХ ПРИМЕРАХ

Модели, приведенные в настоящей главе, являются сложными, и рассмотрение нескольких примеров поможет овладеть методами оптимизации. Именно эту цель преследуют иллюстративные примеры, помещенные в разд. 12.4 и 12.5.

В примерах 1 и 2 демонстрируются подробные расчеты, выполненные методом итераций по стратегиям, при $0 \leqslant \alpha < 1$. В примере 3 при $\alpha = 1$ имеется более одной оптимальной стратегии, и вычислительный процесс завершается *без* окончательного выбора одной из этих стратегий, которые являются оптимальными также и при $0 \leqslant \alpha \leqslant 1$. Метод итераций по стратегиям можно дополнить указанием, согласно которому всегда выбирается стратегия, оптимальная не только при $\alpha = 1$, но и при всех α, близких к 1.

В примерах 4, 5 и 6 при $\alpha = 1$ оптимальная стратегия для *любой* конечной длительности планового периода n зависит от четности n. Вместе с тем в примерах 4 и 5 показано, что алгоритм итераций по стратегиям может определить оптимальную стратегию однозначно. В примере 6 аналогично примеру 3 имеется несколько оптимальных стратегий; формальными методами, возможно, и не удастся установить различие между стратегией, оптимальной только при $\alpha = 1$, и стратегией, оптимальной при $0 \leqslant \alpha \leqslant 1$. (В примере 7 на первой итерации должна быть выбрана стратегия с двумя циклами, хотя исходная стратегия включала только один цикл; в связи с этим применяются формулы (I) и (II) предыдущего раздела.)

На этих примерах читатель поймет, почему необходимо глубоко разбираться в экономическом смысле формул и алгоритмов. Механическое их использование может привести к численно правдоподобным, однако явно непригодным ответам.

Полезно переписать формулы расчета «стоимости вершин» (2) и показателей возможного улучшения (3) из разд. 12.2 при $\alpha < 1$, а также соответствующие формулы (7) и (8) из разд. 12.3 при $\alpha = 1$.

Пример 1. Рассмотрим сеть, изображенную на рис. 12.1. На дугах проставлены соответствующие значения c_{ij}. Для того чтобы начать расчеты методом итераций по стратегиям, необходимо

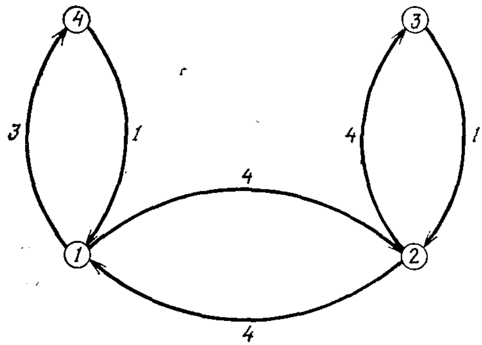


Рис. 12.1. Пример 1.

выбрать исходную стратегию, в которой для каждой вершины имеется выходящая из нее дуга. Пусть такой стратегией являются:

вершина *1* — дуга (*1*, *4*); вершина *2* — дуга (*2*, *1*); (1)

вершина *3* — дуга (*3*, *2*); вершина *4* — дуга (*4*, *1*).

Таким образом, данная стратегия включает единственный цикл — дуги (*1*, *4*) и (*4*, *1*).

Применив формулы расчета «стоимости вершин» (2) из разд. 12.2, решим следующую систему уравнений, линейных относительно y_i^0:

$$\begin{aligned} y_1^0 &= \alpha y_4^0 + c_{14}; \\ y_2^0 &= \alpha y_1^0 + c_{21}; \\ y_3^0 &= \alpha y_2^0 + c_{32}; \\ y_4^0 &= \alpha y_1^0 + c_{41}. \end{aligned} \qquad (2)$$

которые можно представить в следующем виде:

$$\begin{aligned} y_1^0 \qquad\qquad\quad -\alpha y_4^0 &= 3; \\ -\alpha y_1^0 + \quad y_2^0 \qquad\qquad &= 4; \\ -\alpha y_2^0 + y_3^0 \qquad &= 1; \\ -\alpha y_1^0 \qquad\qquad\quad + y_4^0 &= 1. \end{aligned} \tag{3}$$

В конкретных задачах значение α фиксируется на определенном уровне. Однако полезно выполнить здесь расчеты, не устанавливая конкретного значения α и выявляя его влияние на результаты. Решим первое и четвертое уравнения относительно y_1^0 и y_4^0 (это уравнения для вершин, входящих в цикл), а затем остальные уравнения. Можно убедиться в том, что:

$$\begin{aligned} &y_1^0 = \frac{3+1\alpha}{1-\alpha^2}; \\ &y_2^0 = \alpha y_1^0 + 4 = \frac{4+3\alpha-3\alpha^2}{1-\alpha^2}; \\ &y_3^0 = \alpha y_2^0 + 1 = \frac{1+4\alpha+2\alpha^2-3\alpha^3}{1-\alpha^3}; \\ &y_4^0 = \frac{1+3\alpha}{1-\alpha^2}. \end{aligned} \tag{4}$$

Достаточно одного взгляда на рис. 12.1, чтобы увидеть, что возможность выбора решений существует только для вершин *1* и *2* — из вершин *3* и *4* выходит всего лишь по одной дуге. При $0 \leqslant \alpha < 1$, согласно формуле расчета показателей возможности улучшения (7) из разд. 12.2, для вершины *1* имеем

$$\begin{aligned} &\min(\alpha y_2^0 + c_{12},\ \alpha y_4^0 + c_{14}) = \\ &= \min\left(\frac{4+4\alpha-\alpha^2-3\alpha^3}{1-\alpha^2},\ \frac{3+1\alpha}{1-\alpha^2}\right) = \\ &= \frac{3+1\alpha}{1-\alpha^2} \equiv Y_1^0 \text{ для дуги } (1,\ 4) \text{ при } 0 \leqslant \alpha < 1. \end{aligned} \tag{5}$$

Аналогично этому для вершины *2*

$$\begin{aligned} &\min(\alpha y_1^0 + c_{21},\ \alpha y_3^0 + c_{23}) = \\ &= \min\left(\frac{4+3\alpha-3\alpha^2}{1-\alpha^2},\ \frac{4+\alpha+2\alpha^3-3\alpha^4}{1-\alpha^2}\right) = \\ &= \frac{4+3\alpha-3\alpha^2}{1-\alpha^2} \equiv Y_2^0 \quad \text{для дуги } (2,\ 1) \end{aligned} \tag{6}$$

при $\frac{2}{3} \leqslant \alpha < 1$.

Поскольку для каждой вершины i выполняется условие $Y_i^0 = y_i^0$, исходная стратегия оптимальна, если только $\frac{2}{3} \leqslant \alpha < 1$. Если же

$\alpha < \frac{2}{3}$, тогда минимум в (6) достигается на дуге (*2*, *3*), и это решение должно на следующей итерации заменить дугу (*2*, *1*). Читателю предлагается объяснить, почему при $\alpha < 2/3$ дуга (*2*, *3*) является лучшим решением.

Убедимся в том, что показатель возможного улучшения для дуги (*2*, *3*) в (6) действительно представляет собой интегральные дисконтированные затраты на перемещение из вершины *2* в вершину *3*, из *3* в *2*, затем в вершину *1* и, наконец, по петле, состоящей из вершин *1* и *4* и соединяющих их дуг. Однако если дуга (*2*, *3*) действительно используется вместо (*2*, *1*), то интегральные дисконтированные затраты исчисляются на перемещение из вершины *2* в вершину *3* и обратно по циклу бесконечное число раз. Таким образом, можно определить, предпочтителен ли повторяющийся цикл, включающий вершины *2* и *3*, проверив, улучшается ли решение при однократном прохождении этого цикла.

Из рис. 12.1 видно также, что только стратегия (1) оптимальна при $\alpha = 1$. Цикл с минимальными средними затратами состоит из дуг (*1*, *4*) и (*4*, *1*), так что $\bar{c} = \frac{3+1}{2} = 2$.

Пример 2. Рассмотрим сеть с параметрами, приведенными на рис. 12.1, однако примем $c_{23} = 3$. Поскольку исходная стратегия (1) не включает эту дугу, проверяемые значения y_i^0 соответствуют (4). Аналогично этому расчеты показателей возможного улучшения по формуле (5) дают те же результаты. Вместе с тем

$$\alpha y_3^0 + c_{23} = \frac{3+\alpha+\alpha^2+2\alpha^3-3\alpha^4}{1-\alpha^2} \tag{7}$$

и в результате при $0 \ll \alpha < 1$ получим

$$\alpha y_3^0 + c_{23} \equiv Y_2^0 \quad \text{для дуги } (2, 3). \tag{8}$$

Таким образом, дуга (*2*, *3*) заменяет в проверяемой стратегии дугу (*2*, *1*). Это в свою очередь значит, что при $n = 1$ второе уравнение в системе (3) заменяется на

$$-\alpha y_3^1 + y_2^1 = 3 \tag{9}$$

и новое решение имеет следующий вид:

$$y_1^1 = y_2^1 = \frac{3+1\alpha}{1-\alpha^2},$$
$$y_3^1 = y_4^1 = \frac{1+3\alpha}{1-\alpha^2}. \tag{10}$$

Теперь показатели возможного улучшения для вершин *1* и *2* при $0 \leqslant \alpha < 1$ одинаковы:

$$\min\left(\frac{4+3\alpha-3\alpha^2}{1-\alpha^2}, \frac{3+1\alpha}{1-\alpha^2}\right) = \frac{3+1\alpha}{1-\alpha^2} \equiv Y_1^1 = Y_2^1$$
$$\text{для вершин } 1 \text{ и } 2. \tag{11}$$

Вычисления прекращаются, поскольку для всех i $Y_i^1 = y_i^1$.

Пример 3. Продолжим рассмотрение приведенного выше примера, приняв $\alpha = 1$. Из рис. 12.1 (с внесенным изменением $c_{23} = 3$) видно, что циклы вершин *1* и *4* и вершин *2* и *3* в равной мере оптимальны. Начав с исходной стратегии (1) и применив формулы расчета «стоимости вершин» (7) из разд. 12.3, получим систему уравнений:

$$\begin{aligned} w_1^0 - w_4^0 &= c_{14} - \bar{c}^0; \\ w_2^0 - w_1^0 &= c_{21} - \bar{c}^0; \\ w_3^0 - w_2^0 &= c_{32} - \bar{c}^0; \\ w_4^0 - w_1^0 &= c_{41} - \bar{c}^0; \\ w_1^0 &= 0. \end{aligned} \tag{12}$$

Можно убедиться в том, что система (12) имеет следующее решение:

$$\bar{c}^0 = 0, \quad w_1^0 = 0, \quad w_2^0 = 2, \quad w_3^0 = 1, \quad w_4^0 = -1. \tag{13}$$

Расчет показателей возможного улучшения по формуле (8) из разд. 12.3 дает следующие результаты:

$$\min(w_2^0 + c_{12} - \bar{c}^0,\ w_4^0 + c_{14} - \bar{c}^0) = \quad \text{(для вершины } 1)$$
$$= \min(2 + 4 - 2,\ -1 + 3 - 2) = 0 \equiv W_1^0 \text{ для дуги } (1, 4); \tag{14}$$
$$\min(w^0 + c_{21} - \bar{c}^0,\ w_3^0 + c_{23} - \bar{c}^0) = \quad \text{(для вершины } 2)$$
$$= \min(0 + 4 - 2,\ 1 + 3 - 2) = 2 \equiv W_2^0 \text{ для обеих дуг.} \tag{15}$$

Расчеты прекращаются, поскольку для всех i $W_i^0 = w_i^0$.

Отметим, что хотя исходная стратегия (1), включающая единственный цикл из вершин *1* и *4*, согласно вычисленным показателям возможного улучшения, является оптимальной, предпочтительной, вероятно, будет все же альтернативная оптимальная стратегия, включающая цикл из вершин *2* и *3*. Это объясняется тем, что дуга (*2*, *3*) оптимальна не только при $\alpha = 1$, но и является *единственной* оптимальной дугой при $0 \leqslant \alpha < 1$, как это подтверждается (8).

Пусть

$$\bar{c}_{ij} = c_{ij} - \bar{c} = c_{ij} - 2. \tag{I}$$

На рис. 12.2 изображена сеть, для которой

$$w_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j + \bar{c}_{ij}) \quad \text{и} \quad w_1 = 0. \tag{II}$$

Заметим, что суммарная «длина» оптимального цикла равна нулю, тогда как для любого другого цикла она строго положительна.

Пример 4. Рассмотрим сеть, изображенную на рис. 12.3. Вначале примем $c_{12} = 1{,}5$. При изучении сети быстро обнаруживается, что при $\alpha = 1$ дуга (*2*, *3*) является лучшим решением, чем дуга (*2*, *1*). Если снова и снова в течение ряда отрезков начинать из вершины *2*, то цикл, включающий вершины *2* и *3*, обусловливает меньшие совокупные затраты, чем цикл из вершин *2* и *1*.

Наряду с этим не столь ясно, действительно ли при $\alpha = 1$ дуга (*1*, *3*) лучше дуги (*1*, *2*). Для выяснения причины этого вычислите, проходя отрезок за отрезком, совокупные затраты с начала планового периода для маршрута, начинающегося в вершине *1*, идущего в вершину *2*, а затем по циклу, включающему вершины *2* и *3*.

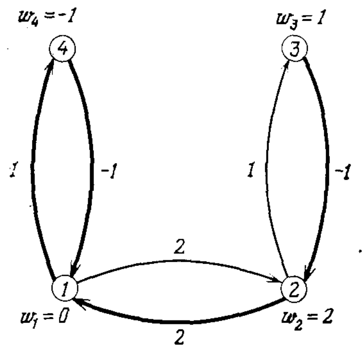


Рис. 12.2. Пример 3 (на дугах проставлены $\bar{c}_{ij}$).

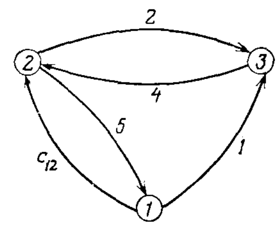


Рис. 12.3. Пример 4.

Сравните эту сумму с аналогичными затратами для маршрута, идущего в вершину *3* из вершины *1*. Читатель обнаружит, что дуга (*1*, *3*) более предпочтительна при нечетном числе отрезков (1, 3, 5, . . . ,) в плановом периоде, а дуга (*1*, *2*) — при четном.

Что же происходит при использовании алгоритма итераций в пространстве стратегий? Пусть исходной стратегией является следующая:

$$\text{вершина } 1 \text{ — дуга } (1, 3); \text{ вершина } 2 \text{ — дуга } (2, 3);$$
$$\text{вершина } 3 \text{ — дуга } (3, 2). \tag{16}$$

Найдем значения y_i^0, решив систему уравнений:

$$\begin{aligned} y_1^0 &= \alpha y_3^0 + c_{13}; \\ y_2^0 &= \alpha y_3^0 + c_{23}; \\ y_3^0 &= \alpha y_2^0 + c_{32}. \end{aligned} \tag{17}$$

Нетрудно убедиться в том, что решение имеет следующий вид:

$$y_1^0 = \frac{1 + 4\alpha + \alpha^2}{1 - \alpha^2};$$

$$y_2^0=\frac{2+4\alpha}{1-\alpha^2};\qquad(18)$$
$$y_3^0=\frac{4+2\alpha}{1-\alpha^2}.$$

Расчет показателей возможного улучшения для вершины *1* дает

$$\min(\alpha y_2^0+c_{12},\ \alpha y_3^0+c_{13})=\qquad\text{(для вершины } 1)$$
$$=\min\left(\frac{1{,}5+2\alpha+2{,}5\alpha^2}{1-\alpha^2},\ \frac{1+4\alpha+\alpha^2}{1-\alpha^2}\right)\equiv Y_1^0\quad\text{для дуги } (1,\ 2)\qquad(19)$$

при $\frac{1}{3}\leqslant\alpha<1$. Таким образом, при больших значениях α дуга (*1*, *2*) лучше дуги (*1*, *3*), так что стратегию (16) необходимо заменить. Новая оптимальная стратегия имеет следующий вид:

$$y_1^1=Y_1^0,\quad y_2^1=y_2^0,\quad y_3^1=y_3^0.\qquad(20)$$

Теперь применим алгоритм итераций по стратегиям при $\alpha=1$ и убедимся в том, что приведенное выше решение, включающее дугу (*1*, *2*), является единственным оптимумом.

Пример 5. Продолжим рассмотрение предыдущего примера, однако примем $c_{12}=2$. И здесь при конечном плановом периоде дуга (*1*, *3*) является строго предпочтительной в случае нечетного числа отрезков и строго худшей в случае четного их числа. Если в качестве исходной стратегии выбрать (16), результаты (17) и (18) остаются в силе.

Показатель возможного улучшения для вершины *1* при $0\leqslant\leqslant\alpha<1$ равен

$$\min_{(1,\,2)\,(1,\,3)}\left(\frac{2+2\alpha+2\alpha^2}{1-\alpha^2},\ \frac{1+4\alpha+\alpha^2}{1-\alpha^2}\right)\equiv Y_1^0\quad\text{для дуги } (1,\ 3).\qquad(21)$$

Следовательно, дуга (*1*, *3*) оптимальна при всех значениях α, меньших чем 1.

Пример 6. Продолжим рассмотрение предыдущего примера, однако примем $\alpha=1$. Как можно было предположить, дуга (*1*, *3*) остается оптимальной. Тем не менее, согласно показателям возможного улучшения, дуга (*1*, *2*) также является оптимальной. Здесь необходимо решить систему уравнений:

$$\begin{aligned} w_1^0-w_2^0&=c_{12}-\bar{c}^0;\\ w_2^0-w_3^0&=c_{23}-\bar{c}^0;\\ w_3^0-w_2^0&=c_{32}-\bar{c}^0;\\ w_1^0&=0,\end{aligned}\qquad(22)$$

решением которой является

$$\bar{c}^0=3,\quad w_1^0=0,\quad w_2^0=1,\quad w_3^0=2.\qquad(23)$$

Можно убедиться в том, что при использовании дуги (*1*, *3*) вместо (*1*, *2*) мы получим точно такие же результаты.

Показатели возможного улучшения для вершины *1* имеют следующий вид:

$$\min(w_2^0 + c_{12} - \bar{c}^0,\ w_3^0 + c_{13} - \bar{c}^0) = \qquad \text{(для вершины } 1)$$
$$= \min(1 + 2 - 3,\ 2 + 1 - 3) - 0 \equiv W_1^0 \quad \text{для дуг } (1, 2) \text{ и } (1, 3), \tag{24}$$

так что проверяемая стратегия, включающая дугу (*1*, *2*), удовлетворяет проверке на оптимальность.

Пример 7. Рассмотрим сеть, изображенную на рис. 12.4, при $\alpha = 1$. Пусть исходная стратегия имеет вид:

$$\begin{array}{ll} \text{вершина } 1 - \text{дуга } (1, 4); & \text{вершина } 2 - \text{дуга } (2, 4); \\ \text{вершина } 3 - \text{дуга } (3, 2); & \text{вершина } 4 - \text{дуга } (4, 1). \end{array} \tag{III}$$

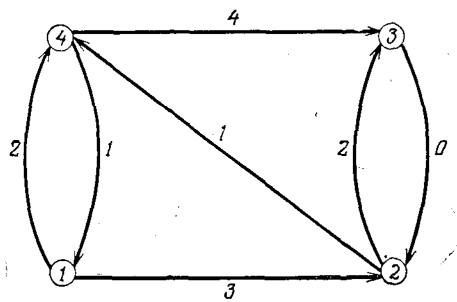


Рис. 12.4. Пример 7.

Таким образом, единственный цикл включает вершины *1* и *4*. Этой стратегии соответствует решение

$$\bar{c}^0 = \frac{3}{2}, \quad w_4^0 = 0, \quad w_2^0 = -1, \quad w_3^0 = -\frac{5}{2}, \quad w_4^0 = -\frac{1}{2}. \tag{IV}$$

Можно проверить, что показатели возможного улучшения равны:

$$\begin{array}{ll} W_1^0 = 0 = w_1^2 & \text{для дуги } (1, 4); \\ W_2^0 = -2 < w_2^0 & \text{для дуги } (2, 3); \\ W_3^0 = -\frac{5}{2} = w_3^0 & \text{для дуги } (3, 2); \\ W_4^0 = -\frac{1}{2} = w_4^0 & \text{для дуги } (4, 1). \end{array} \tag{V}$$

Поэтому стратегия, соответствующая (V), является улучшением. Заметим, что в новой стратегии имеется два цикла, так что на следующей итерации приходится использовать (I) и (II) из разд. 12.3.

Соответствующим путем решаются следующие уравнения расчета стоимости:

$$
\begin{aligned}
&w_1^1 - w_4^1 + \bar{c}_4^1 = 2; \quad \bar{c}_1^1 = \bar{c}_4^1; \quad w_1^1 = 0;\\
&w_2^1 - w_3^1 + \bar{c}_3^1 = 2; \quad \bar{c}_2^1 = \bar{c}_3^1; \quad w_2^1 = 0;\\
&w_3^1 - w_2^1 + \bar{c}_2^1 = 0;\\
&w_4^1 - w_1^1 + \bar{c}_1^1 = 1.
\end{aligned}
\tag{VI}
$$

Можно проверить, что

$$\bar{c}_1^1 = \bar{c}_4^1 = \frac{3}{2}, \quad \bar{c}_2^1 = \bar{c}_3^1 = 1 \tag{VII}$$

и что

$$
\begin{aligned}
&\bar{C}_1^1 = 1 < \bar{c}_1^1 \quad \text{для дуги } (1,\ 2);\\
&\bar{C}_2^1 = 1 = \bar{c}_2^1 \quad \text{для дуги } (2,\ 3);\\
&\bar{C}_3^1 = 1 = \bar{c}_3^1 \quad \text{для дуги } (3,\ 2);\\
&\bar{C}_4^1 = 1 < \bar{c}_5^1 \quad \text{для дуги } (4,\ 3).
\end{aligned}
\tag{VIII}
$$

В новой стратегии, представленной (VIII), имеется только один цикл. Читателю предлагается выполнить еще одну итерацию и убедиться в том, что стратегия (VIII) действительно оптимальна.

Выводы. В настоящем разделе демонстрировалось применение метода последовательных приближений в пространстве стратегий для решения нескольких примеров. Было показано, как для случая $0 \leqslant \alpha < 1$ критерий соблюдения условий оптимальности Y_i^n характеризует возможность улучшения стратегии; была также показана зависимость этой величины от значения α. Для случая $\alpha = 1$ метод позволяет найти не менее одного цикла, однако критерий проверки оптимальности не является настолько чувствительным, чтобы выявить различия между альтернативными оптимальными стратегиями, которые обусловливают существенно разные интегральные дисконтированные затраты при значениях α, близких к 1. Как упоминалось выше, алгоритм можно изменить так, что этот недостаток будет устранен. Поскольку необходимые модификации выходят за рамки настоящего изложения, они здесь не приводятся.

12.5. ПРОСТАЯ МОДЕЛЬ УПРАВЛЕНИЯ ЗАПАСАМИ

В приведенных выше примерах рассматриваются простые сети, позволяющие сосредоточить внимание на технике расчетов по методу последовательных приближений. В настоящем разделе этот метод применяется к динамической задаче, имеющей практическую значимость, а именно к простой детерминированной модели управления запасами, которая вначале изучалась в разд. 8.3, а затем была обобщена в разд. 12.1. Воспользуемся теми же данными, что и в примере фирмы «Надежный поставщик», приведенном в разд. 8.7:

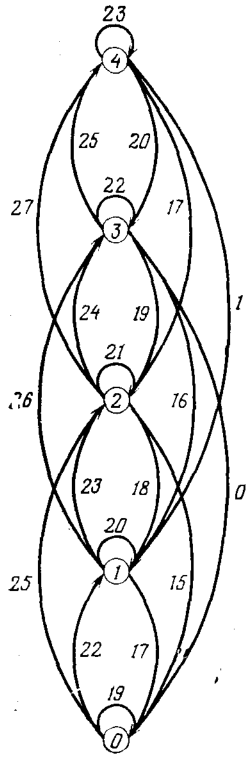


Рис. 12.5. Модель управления запасами (при бесконечном плановом периоде) фирмы «Надежный поставщик». Уровень запасов для вершины i равен i; объем выпуска на дуге (i, j) равен $j - i + 3$.

$$\begin{aligned} &\text{Спрос:} \qquad D = 3 \text{ единицам}, \\ &\text{Объем выпуска:} \quad x \leqslant 5 \text{ единицам}, \end{aligned} \qquad (1)$$

Уровень запасов на конец отрезка: $j \leqslant 4$ единицам;

затраты на производство и хранение продукции

$$C(x, j) = C(x) + h \cdot j, \qquad (2)$$

где x и j — неотрицательные целые числа и

$$\begin{gathered} C(0) = 0, \quad C(1) = 15, \quad C(2) = 17, \\ C(3) = 19, \quad C(4) = 21, \quad C(5) = 23, \\ h = 1 \end{gathered} \qquad (3)$$

для всех отрезков в течение бесконечного планового периода.

Переменной состояния является уровень запасов на *начало* отрезка, обозначенный i, где $i = 0, 1, \ldots, 4$. Таким образом, стационарная стратегия представляет собой правило установления объема выпуска, позволяющего получить заданное i. На рис. 8.16 оптимальная стратегия $x_\infty(i)$ для бесконечного планового периода представлялась следующей:

$$\begin{gathered} x_\infty(0) = x_\infty(1) = x_\infty(2) = 5, \\ x_\infty(3) = x_\infty(4) = 0. \end{gathered} \qquad (4)$$

Теперь (4) необходимо подтвердить методом итераций по стратегиям.

Сеть, отображающая условия рассматриваемого примера, изображена на рис. 12.5. Каждому уровню запасов i на начало отрезка соответствует некоторая вершина; при заданном i каждому допустимому объему выпуска x соответствует дуга. Так, если уровень запасов на начало отрезка равен i, объем выпуска составляет x, а спрос равен 3 единицам, то запас на начало следующего отрезка составляет $j \equiv i + x - 3 \geqslant 0$. Следовательно, дуга (i, j) отображает производство $(j - i + 3)$ единиц. В связи с наличием в (1) ограничений на объем выпуска и уровень запасов некоторые сочетания i и j не являются допустимыми. Так, например, при $i = 0$ должно быть $j \leqslant 2$, поскольку $x \leqslant 5$. Затраты для дуг, приведенные в рис. 12.6, определяются следующим выражением:

$$c_{ij} = C(j - i + 3) + h \cdot j. \tag{5}$$

Предположим, что в качестве критерия оптимальности выбран минимум средних затрат за отрезок. Тогда решаемое функциональное уравнение имеет вид

$$w_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j + c_{ij} - \bar{c}) \quad \text{для всех } i. \tag{6}$$

Пусть исходная стратегия состоит в том, чтобы производить недостающее для удовлетворения спроса количество продукции при $i \leqslant 3$, так чтобы уровень запасов на конец отрезка был равен

$i \backslash j$	0	1	2	3	4
0	19 + 0	21 + 1	23 + 2		
1	17 + 0	19 + 1	21 + 2	23 + 3	
2	15 + 0	17 + 1	19 + 2	21 + 3	23 + 4
3	0 + 0	15 + 1	17 + 2	19 + 3	21 + 4
4		0 + 1	15 + 2	17 + 3	19 + 4

Рис. 12.6. Модель управления запасами фирмы «Надежный поставщик».
Затраты для дуг определяются выражением
$c_{ij} = C(j - i + 3) + h \cdot j$.

$j = 0$, и вообще не выпускать продукцию, если на данном отрезке $i = 4$, так что $j = 1$. На сети эта стратегия представляется следующим образом:

$$\begin{aligned}
&\text{вершина } 0 \text{ — дуга } (0, 0), \quad x = 3;\\
&\text{вершина } 1 \text{ — дуга } (1, 0), \quad x = 2;\\
&\text{вершина } 2 \text{ — дуга } (2, 0), \quad x = 1;\\
&\text{вершина } 3 \text{ — дуга } (3, 0), \quad x = 0;\\
&\text{вершина } 4 \text{ — дуга } (4, 1), \quad x = 0.
\end{aligned} \tag{7}$$

Например, находясь в вершине 1 ($i = 1$), мы выпускаем 2 единицы продукции, так что к концу отрезка спрос 3 единиц уменьшает запас до нуля. Если начертить сеть, включающую только эти дуги, то можно заметить, что выбранная стратегия в конечном счете приводит к стационарному выпуску 3 единиц продукции в течение каждого отрезка.

Используем алгоритм итераций по стратегиям и будем находить значения w_i^0 и $\bar{c}^0$ с помощью системы уравнений расчета стоимости:

$$
\begin{aligned}
w_0^0 - w_0^0 + \bar{c}^0 &= 19,\\
w_1^0 - w_0^0 + \bar{c}^0 &= 17,\\
w_2^0 - w_0^0 + \bar{c}^0 &= 15,\\
w_3^0 - w_0^0 + \bar{c}^0 &= 0,\\
w_4^0 - w_1^0 + \bar{c}^0 &= 1,\\
W_0^0 &= 0,
\end{aligned} \tag{8}
$$

решением которой являются

$$\bar{c}^0 = 19,\ w_0^0 = 0,\ w_1^0 = -2,\ w_2^0 = -4,\ w_3^0 = -19,\ w_4^0 = -20. \tag{9}$$

Можно убедиться в том, что операция нахождения

$$\min_{\substack{\text{по всем } (i,\, j)\\ \text{в сети}}} (w_j^n + c_{ij} - \bar{c}^n) \equiv W_i^n \tag{10}$$

при $n = 0$ дает решение

$$
\begin{aligned}
W_0^0 &= 0 && \text{для дуги } (0,\ 0);\\
W_1^0 &= -12 && \text{для дуги } (1,\ 3);\\
W_2^0 &= -14 && \text{для дуги } (2,\ 3);\\
W_3^0 &= -19 && \text{для дуги } (3,\ 0);\\
W_4^0 &= -20 && \text{для дуги } (4,\ 1).
\end{aligned} \tag{11}
$$

Поскольку для $i = 1$ и $i = 2$ $W_i^0 < w_i^0$, выполнение операции на этом не прекращается. Читателю предлагается вычертить сеть, отвечающую новой стратегии, которая была бы основана на (11).

Результаты, полученные в ходе ряда последовательных итераций, приведены в таблице рис. 12.7. Читателю полезно самостоятельно повторить расчеты, используя данные таблицы рис. 12.7 в качестве контроля. Следует также начертить сети, изображающие каждую из пробных стратегий. Отметим, что, рассматривая эту таблицу, можно сделать следующие выводы:

I) От $n = 0$ до $n = 1$ значение $\bar{c}^n$ не улучшается.

II) Хотя при $n = 2$ средние затраты за отрезок улучшаются ($\bar{c}^2 < \bar{c}^1$), однако относительные стоимости возрастают $w_i^2 > W_i^1$ для всех i и $w_i^2 > w_i^1$ для всех $i \neq 0$.

III) Стратегии для разных итераций имеют следующий вид:

$$x = 5 \text{ для } i = 1,\ 2,\quad x = 4 \text{ для } i = 0,\ 3,\quad x = 0 \text{ для } i = 4,$$

$$\text{что соответствует производственному циклу } (5,\ 4,\ 0) \text{ при } n = 2; \tag{12}$$

$$x = 4 \text{ для } i = 1,\quad x = 5 \text{ для } i = 0,\ 2,\quad x = 0 \text{ для } i = 3,4,$$

$$\text{что соответствует производственному циклу } (4,\ 5,\ 0) \text{ при } n = 3; \tag{13}$$

Метод итераций по стратегиям

Вершина i	Дуга (i, j)	w_i^0	W_i^0	Дуга (i, j)	w_i^1	W_i^1	Дуга (i, j)	w_i^2	W_i^2	Дуга (i, j)	w_i^3	W_i^3	Дуга (i, j)	w_i^4	W_i^4	Дуга (i, j)	w_i^5	W_i^5	Дуга (i, j)	w_i^6	W_i^6
0	(0,0)	0	0	(0,0)	0	-9	(0,1)	0	$-3\frac{2}{3}$	(0,2)	0	0	(0,2)	0	-1	(0,1)	0	-1	(0,2)	0	0
1	(1,0)	-2	-12	(1,3)	-12	-12	(1,3)	$-4\frac{2}{3}$	$-5\frac{2}{3}$	(1,2)	-2	-8	(1,3)	$-6\frac{2}{3}$	$-6\frac{2}{3}$	(1,3)	-6	-6	(1,3)	$-5\frac{3}{5}$	$-5\frac{3}{5}$
2	(2,0)	-4	-14	(2,3)	-14	-22	(2,4)	$-11\frac{1}{3}$	$-11\frac{1}{3}$	(2,4)	-8	-10	(2,3)	$-8\frac{2}{3}$	$-11\frac{1}{3}$	(2,4)	-10	-10	(2,4)	$-9\frac{1}{5}$	$-9\frac{1}{5}$
3	(3,0)	-19	-19	(3,0)	-19	-24	(3,4)	$-13\frac{1}{3}$	$-17\frac{1}{3}$	(3,0)	-17	-17	(3,0)	$-16\frac{1}{3}$	$-16\frac{1}{3}$	(3,0)	-16	-16	(3,0)	$-15\frac{4}{5}$	$-15\frac{4}{5}$
4	(4,1)	-20	-20	(4,1)	-30	-30	(4,1)	-21	-21	(4,1)	-18	-18	(4,1)	-22	-22	(4,1)	-21	-21	(4,1)	$-20\frac{2}{5}$	$-20\frac{2}{5}$
$\bar{c}^n$	19			19			$17\frac{1}{3}$			17			$16\frac{1}{3}$			16			$15\frac{4}{5}$		

Рис. 12.7. Модель управления запасами фирмы «Надежный поставщик».

$$x = 5 \text{ для } i = 0{,}1,\ x = 4 \text{ для } i = 2,\ x = 0 \text{ для } i = 3{,}4,$$
$$\text{что соответствует производственному циклу} \quad (5, 4, 0) \text{ при } n = 4; \tag{14}$$
$$x = 4 \text{ для } i = 0,\ x = 5 \text{ для } i = 1{,}2,\ x = 0 \text{ для } i = 3{,}4,$$
$$\text{что соответствует производственному циклу} \quad (4, 5, 0) \text{ при } n = 5; \tag{15}$$
$$x = 5 \text{ для } i = 0, 1, 2,\ x = 0 \text{ для } i = 3, 4,$$
$$\text{что соответствует производственному циклу} \quad (5, 5, 0, 5, 0) \text{ при } n = 6. \tag{16}$$

Таким образом, в ходе выполняемых итераций анализируются два цикла (5, 4, 0) и (4, 5, 0) при все более низких уровнях запасов, причем этот процесс продолжается до тех пор, пока не будет достигнута сходимость к оптимальной стратегии (16), которая подтверждает (4).

Для того чтобы проверить, как вы усвоили материал, ослабим ограничение, налагаемое на объем выпуска, и будем считать допустимым $x = 6$; этому объему соответствуют затраты $C(6) = 28{,}5$. Выполните одну итерацию алгоритма итераций по стратегиям и *проверьте*, что оптимальная стратегия имеет вид

$$\begin{array}{ll} \text{вершина } \mathit{0} \text{ — дуга } (\mathit{0}, \mathit{3}); & \text{вершина } \mathit{1} \text{ — дуга } (\mathit{1}, \mathit{3}); \\ \text{вершина } \mathit{2} \text{ — дуга } (\mathit{2}, \mathit{3}); & \text{вершина } \mathit{3} \text{ — дуга } (\mathit{3}, \mathit{0}), \\ \text{вершина } \mathit{4} \text{ — дуга } (\mathit{4}, \mathit{1}), & \end{array} \tag{I}$$

чему соответствует

$$\begin{array}{ll} x = 6 \text{ для } i = 0, & x = 5 \text{ для } i = 1, \\ x = 4 \text{ для } i = 2, & x = 0 \text{ для } i = 3, 4, \end{array} \tag{II}$$
$$\text{производственный цикл } (6, 0)$$

и $\bar{c} = 15{,}75$.

12.6. ПОДХОД НА ОСНОВЕ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

В настоящем разделе разъясняется постановка дискретной задачи динамического программирования для бесконечного планового периода в виде модели линейного программирования. Выяснение связи между двумя этими задачами помогает понять, почему алгоритм итераций по стратегиям сходится за конечное число итераций: он в сильной степени напоминает симплексный метод. Кроме того, читатель сумеет представлять задачу динамического программирования рассмотренного вида в такой форме, которая позволит применять стандартные машинные программы, предназначенные для решения задач линейного программирования. Подобные преобразования важны при решении практических задач, поскольку на всех больших ЭВМ имеются весьма совершенные программы для реализации алгоритмов линейного программирования. Дополнительным преимуществом является и то обстоятельство, что возникает возможность использовать хорошо разработанные методы анализа чувствительности.

Минимизация интегральных дисконтированных затрат. Рассмотрим случай $0 \leqslant \alpha < 1$ и соответствующие функциональные уравнения

$$y_i = \min_{\substack{\text{по всем } (i,j) \\ \text{в сети}}} (\alpha y_j + c_{ij}) \quad \text{для всех } i. \tag{1}$$

Эти уравнения означают, что каждая величина y_i должна удовлетворять линейным ограничениям (ограничения двойственной задачи по дугам):

$$y_i \leqslant \alpha y_j + c_{ij} \text{ или же } y_i - \alpha y_j \leqslant c_{ij} \text{ для каждой дуги } (i, j) \text{ в сети.} \tag{2}$$

На знак y_i ограничений не налагается; вместе с тем система неравенств (2) не полностью адекватна экстремальным уравнениям (1), поскольку значения y_i, удовлетворяющие (2), могут и не удовлетворять (1). Так, например, если все $c_{ij} > 0$, то $y_i = 0$ удовлетворяют неравенствам (2), но не уравнениям (1). Для обеспечения адекватности по каждому i одно из неравенств (2) должно быть *равенством* хотя бы на одной дуге (i, j). Сколь бы удивительным это ни казалось, однако истинные оптимальные значения *всех* y_i достигаются в результате использования линейной целевой функции (целевой функции двойственной задачи):

$$\text{Максимизировать } \sum_{k=1}^{p} r_k y_k, \tag{3}$$

где все $r_k > 0$, но в остальном могут принимать произвольные значения, а p — наибольший номер вершины на сети.

Назовем совокупность ограничений (2) и целевую функцию (3) двойственной задачей; тогда прямая задача формулируется следующим образом:

$$\text{Минимизировать } \sum_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} \sum c_{ij} x_{ij} \tag{4}$$

при условиях

$$\sum_{\substack{\text{по всем } (k,\, j) \\ \text{в сети}}} x_{kj} - \sum_{\substack{\text{по всем } (i,\, k) \\ \text{в сети}}} \alpha x_{ik} = r_k \quad (\text{для всех вершин } k), \tag{5}$$

$$x_{ij} \geqslant 0 \quad (\text{для каждой дуги } (i, j) \text{ в сети}). \tag{6}$$

Здесь (4) — целевая функция, а (5) — ограничения прямой задачи по вершинам.

Как и ранее, символ двойного суммирования в (4) означает, что суммирование осуществляется по всем дугам (i, j) сети. Наличие в (5) только одного знака суммирования для каждой фиксированной

Дуга	y_1	y_2	y_3	y_4	
(1,2)	1	$-\alpha$			$\leqslant c_{12}$
(1,4)	1			$-\alpha$	$\leqslant c_{14}$
(2,1)	$-\alpha$	1	$-\alpha$		$\leqslant c_{21}$
(2,3)		1			$\leqslant c_{23}$
(3,2)		$-\alpha$	1		$\leqslant c_{32}$
(4,1)	$-\alpha$			1	$\leqslant c_{41}$
	r_1	r_2	r_3	r_4	Максимизировать

На знак y_k не налагается ограничений

Рис. 12.8. Двойственная задача линейного программирования для примера 1, приведенного на рис. 12.1.

вершины k сети означает, что суммирование осуществляется по всем дугам (k, j) и (i, k) для данного значения k. Двойственная и прямая задачи линейного программирования для сети примера 1 разд. 12.4

Вершина	x_{12}	x_{14}	x_{21}	x_{23}	x_{32}	x_{41}	
1	1	1	$-\alpha$			$-\alpha$	$= r_1$
2	$-\alpha$		1	1	$-\alpha$		$= r_2$
3				$-\alpha$	1		$= r_3$
4		$-\alpha$				1	$= r_4$
	c_{12}	c_{14}	c_{21}	c_{23}	c_{32}	c_{41}	Минимизировать

$x_{ij} \geqslant 0$

Рис. 12.9. Прямая задача линейного программирования для примера 1, приведенного на рис. 12.1.

(рис. 12.1) записаны в виде таблицы на рис. 12.8 и 12.9 (прямая задача весьма близка к обобщенной сетевой задаче разд. 6.9).

Симплексный метод и метод итераций по стратегиям. Предположим, что для решения прямой задачи (4) — (6) используется симплексный метод. Поскольку $r_k > 0$ для всех вершин k, по крайней мере одно $x_{jk} > 0$. На каждой итерации симплексного метода пробное решение представляет собой опорный план, включающий p переменных, где p — число вершин. Следовательно, в таком плане должно быть точно по *одной* переменной x_{kj} для каждой вершины k.

Иными словами, опорный план в прямой модели аналогичен пробной стратегии в двойственной модели.

Когда в опорный план вводится новая переменная, то для сохранения его базисности из него *должна* быть исключена одна из базисных переменных. Таким образом, на каждой итерации симплексного метода новая дуга вводится в план, а одна из базисных дуг исключается. После изменения плана он проверяется на соблюдение условий оптимальности с помощью расчетов, равноценных проверке выполнения всех неравенств (2) при текущих значениях двойственных переменных. Последние совпадают с y_k^n, используемыми для проверки текущей стратегии, так что обычный критерий проверки оптимальности плана в симплексном методе соответствует правилу улучшения стратегии при алгоритме последовательных приближений в пространстве стратегий.

Теперь можно увидеть, в чем заключается единственное реальное различие между подходом к решению прямой задачи линейного программирования на основе симплексного метода и методом итераций по стратегиям, примененным к решению двойственной задачи. В последнем алгоритме на каждой итерации просматриваются *все* вершины и одновременно делаются *все* возможные улучшения. Соответственно в алгоритме итераций по стратегиям в базис проверяемого решения можно в течение одной итерации вводить несколько переменных (дуг). Напротив, в симплексном методе на каждой итерации вводится только одна переменная, после чего проверяется достижение оптимальности. Теорией еще не установлено, можно ли считать метод с многократным введением переменных перед каждой проверкой оптимальности более эффективным с вычислительной точки зрения. Относительные преимущества каждого из двух подходов в значительной степени зависят от особенностей машинной программы для данной конкретной ЭВМ.

Связь между двумя подходами обусловливает также возможность доказательства сходимости метода последовательных приближений в пространстве стратегий к оптимальному решению за конечное число итераций. Требуется лишь незначительно изменить доводы относительно достижения оптимума и конечности симплексного алгоритма.

С помощью приводимых ниже соображений можно строго показать, что возможен *произвольный* выбор положительной r_k в двойственной целевой функции (3) и ограничений прямой задачи (5). Предположим, что двойственная задача решена для некоторого набора r_k, и пусть известны значения x_{ij}, соответствующие этой стратегии. Согласно теореме двойственности (гл. 5), x_{ij} также существуют, причем они конечны. Для конкретности пусть соответствующий вектор базисных переменных есть x_B. Значения элементов вектора x_B можно найти из ограничений прямой задачи (5) для вершин сети:

$$(I - A)\, x_B = r, \tag{I}$$

где I — единичная матрица размерностью $p \times p$, A — матрица коэффициентов, r — вектор-столбец r_k. Уравнение (I) можно записать в следующем виде:

$$x_B = r + Ax_B; \tag{II}$$

повторно подставляя значение x_B в правую часть (II), получим

$$x_B = r + A(r + Ax_B) = (I + A + A^2 + \ldots) r. \tag{III}$$

Поскольку x_B является конечным и $r > 0$, сумма $(I + A + A^2 + \ldots)$ должна сходиться к матрице, которую обозначим $(I - A)^{-1}$. Элементы каждой из матриц, входящих в рассматриваемую сумму, являются неотрицательными; следовательно, $(I - A)^{-1}$ есть неотрицательная матрица, диагональные элементы которой имеют значения, не меньшие единицы.

Поэтому если вместо r использовать $r^* > 0$, то

$$x_B^* = (I + A + A^2 + \ldots) r^* \geqslant r^* > 0, \tag{IV}$$

так что предыдущий базис остается допустимым (и невырожденным). Такая производимая в правой части замена оставляет значения показателей достижения оптимальности (значения двойственных переменных) базиса B неизменными; это и позволяет прийти к требуемому выводу о сохранении оптимальности стратегии x_B.

Минимизация эквивалентных средних затрат за отрезок. При $\alpha = 1$ функциональное уравнение динамического программирования имеет следующий вид:

$$w_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j + c_{ij} - \bar{c}) \quad \text{для всех } i, \tag{7}$$

где, как и ранее, предполагается, что от каждой вершины i существует маршрут к одной (по крайней мере) из вершин, связанной с тем циклом, где достигается $\bar{c}$. Это предположение можно опустить, изменив соответствующим образом экстремальные уравнения.

Как было показано на численных примерах в разд. 12.4, значения w_i не определяются однозначно. Более того, хотя метод итераций по стратегиям и позволяет найти стратегию, содержащую цикл с минимальными эквивалентными средними затратами за отрезок, этот метод не позволяет выяснить ту из альтернативных оптимальных стратегий, которая оптимальна и при значениях α, близких к 1. Аналогичному подходу на основе линейного программирования присущ тот же недостаток. Наряду с этим существуют и практические методы нахождения стратегии, оптимальной и при значениях α, близких к 1.

Оптимальное значение $\bar{c}$ и связанную с ним стратегию можно найти, решив двойственную задачу

$$\text{Максимизировать } \bar{c} \tag{8}$$

при условиях

$$w_i - w_j + \bar{c} \leqslant c_{ij} \quad \text{для каждой дуги } (i, j) \text{ в сети}, \tag{9}$$

где на знаки величин w_i и $\bar{c}$ не налагается ограничений. Здесь (8) — целевая функция, а (9) — ограничения двойственной задачи по дугам. Соответствующая прямая задача имеет следующий вид:

$$\text{Минимизировать} \sum_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} \sum c_{ij} v_{ij} \tag{10}$$

при условиях

$$\sum_{\substack{\text{по всем } (k,\, j) \\ \text{в сети}}} v_{kj} - \sum_{\substack{\text{по всем } (i,\, k) \\ \text{в сети}}} v_{ik} = 0 \quad (\text{для каждой вершины } k), \tag{11}$$

$$\sum_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} \sum v_{ij} = 1, \tag{12}$$

$$v_{ij} \geqslant 0 \quad (\text{для каждой дуги } (i, j) \text{ в сети}). \tag{13}$$

Здесь (10) — целевая функция прямой задачи, (11) — ее ограничения по вершинам, (12) — ограничение нормализации во времени, которое именуется **нормирующим уравнением** и исключает возможность получения решения $v_{ij} = 0$. В результате величины v_{ij}, как правило, *не* являются целочисленными, и положительные v_{ij} в оптимальном решении связаны с циклом, обеспечивающим минимум эквивалентных средних затрат. Величина v_{ij} представляет собой долю тех отрезков, в течение которых, при наличии в оптимальной стратегии единственного цикла, осуществляется перемещение по дуге (i, j).

Если суммировать уравнения (11) по всем вершинам, слагаемые в левой части взаимно уничтожаются, так что возникает избыточность числа ограничений и в процессе вычислений можно отбросить одно уравнение. Это явление имело место и в сетевых моделях, рассмотренных в гл. 6 и 7; оно соответствует тому обстоятельству, что в (9) величины w_i не определяются однозначно — если некоторый набор w_i удовлетворяет неравенствам (9), то им удовлетворяет и набор значений (w_i + постоянная).

Пример. Для сети, изображенной на рис. 12.1, двойственная и прямая задачи приведены на рис. 12.10 и 12.11. Читателю следует сравнить эти модели с теми, которые приводились выше в форме таблиц на рис. 12.8 и 12.9, где $0 \leqslant \alpha < 1$. Можно убедиться в том, что решение примера 3

$$w_1 = 0, \quad w_2 = 2, \quad w_3 = 1, \quad w_4 = -1, \quad \bar{c} = 2 \tag{14}$$

является допустимым для задачи рис. 12.10 и что в результате этого оба допустимых решения

$$v_{14} = v_{41} = \frac{1}{2}, \quad \text{все прочие } v_{ij} = 0, \tag{15}$$

$$v_{14} = v_{41} = v_{23} = v_{32} = \frac{1}{4}, \quad \text{все прочие } v_{ij} = 0 \tag{16}$$

Дуга	w_1	w_2	w_3	w_4	$\bar{c}$	
(1, 2)	1	−1			1	$\leqslant c_{12}$
(1, 4)	1			−1	1	$\leqslant c_{12}$
(2, 1)	−1	1			1	$\leqslant c_{21}$
(2, 3)		1	−1		1	$\leqslant c_{23}$
(3, 2)		−1	1		1	$\leqslant c_{32}$
(4, 1)	−1			1	1	$\leqslant c_{41}$
						Максимизировать

На знак w_k не налагается ограничений

Рис. 12.10. Двойственная задача линейного программирования для примера 1, приведенного на рис. 12.1.

являются оптимальными для задачи, приведенной на рис. 12.11. (Если уравнение для вершины *1* исключить как избыточное, то в базис решения (15) входят и $v_{21} = v_{32} = 0$.)

Вершина	k_{12}	k_{14}	v_{21}	v_{23}	v_{32}	v_{41}	
1	1	1	−1			−1	=0
2	−1		1	1	−1		=0
3				−1	1		=0
4		−1				1	=0
В среднем за цикл	1	1	1	1	1	1	1
	c_{12}	c_{14}	c_{21}	c_{23}	c_{32}	c_{41}	Минимизировать

$v_{ij} \geqslant 0$

Рис. 12.11. Прямая задача линейного программирования для примера 1, приведенного на рис. 12.1.

Простые модификации приведенных выше постановок достаточны для того, чтобы обобщить их и на ситуации, при которых h_{ij} — число отрезков, необходимое для перемещения по дуге (i, j) и $\alpha_{ij} \equiv \equiv \alpha^{h_{ij}}$. В случае минимизации интегральных дисконтированных затрат мы попросту заменяем α на α_{ij} всюду, где α встречается в формулах. В случае минимизации эквивалентных средних затрат за отрезок требуется лишь поместить коэффициент h_i перед $\bar{c}$ в экстремальных уравнениях (7) и в соответствующих им ограничениях (9) двойственной задачи по дугам. Аналогично этому в выражении ограничения нормализации во времени (12) также ставится коэффициент h_{ij} перед v_{ij}. При такой постановке $h_{ij}v_{ij}$ есть доля тех отрезков, в которых производится перемещение по дуге (i, j), если в оптимальной стратегии имеется единственный цикл.

12.7. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

В настоящей главе методы последовательных приближений были применены для решения экстремальных уравнений в моделях динамического программирования с бесконечным плановым периодом, которые удовлетворяют следующим допущениям:

I) исходы решений являются детерминированными;

II) состояние системы *исследуется* в дискретные моменты времени;

III) как управляющие переменные, так и переменные состояния являются дискретными и имеют конечное число возможных значений;

IV) параметры системы являются стационарными.

В гл. 16—18 изучаются модели, для которых допущение I) ослаблено и включает вероятностные исходы. Случайные события приводят к необходимости распространения оптимизационных процессов на новый тип поведения. Однако методы решения этих моделей немногим отличаются от методов, излагаемых в гл. 8—12.

Как читатель увидит в моделях теории массового обслуживания (гл. 20 и приложение III), некоторые реальные ситуации полезно рассматривать так, как будто решения могут приниматься в любой момент времени, а не только в дискретные моменты, предусмотренные допущением II). Например, управляющий магазина самообслуживания, увидев в «час пик», что очередь покупателей к кассирам слишком велика, может в любой момент увеличить число работающих касс. Очевидно, он не обязан принимать подобное решение только в определенные моменты времени, например разделенные пятиминутными интервалами. В случае таких ситуаций различные изученные читателем методы последовательных приближений можно адаптировать с целью разработки реализуемых оптимизационных алгоритмов.

Допущение III), относящееся к дискретному характеру переменных и к конечному числу их значений, нередко вводится по соображениям удобства как анализа, так и вычислений. Однако часто реальная система столь же хорошо моделируется при предположениях о непрерывности и даже неограниченности управляющих переменных и переменных состояния. Для иллюстрации рассмотрим простую модель управления запасами, обобщенную в разд. 12.1. В этой модели предполагалось, что объем выпуска продукции x должен быть целочисленным. Поскольку спрос отображается целым числом единиц, допущение о целочисленности объема производства влечет за собой целочисленность уровня запасов. Предположим вместо этого, что объем выпуска рассматривается как скорость (выпуск в течение единицы времени — отрезка), так что постулат о его целочисленности отбрасывается. Тогда рекуррентное соотношение для бесконечного планового периода имеет вид

$$f(i) = \min_x [C(x) + h(i + x - d) + \alpha f(i + x - d)], \qquad (1)$$

где $-\infty \leqslant i \leqslant +\infty$ и x — любое неотрицательное действительное число, не меньшее чем $(i - d)$.

При анализе такой модели возникают некоторые ранее не встречавшиеся технические детали. Так, например, если предположить, что функция $f(i)$ существует и единственна, потребуется все же выяснить, является ли $f(i)$ и непрерывной функцией. Аналогично этому потребуется определить, является ли функция $x_\infty(i)$, которая определяет оптимальную стратегию, однозначной и непрерывной.

Если в выражении типа (1) используются непрерывные переменные, иногда удается, применив аппарат математического анализа, получить в явном виде простую формулу для решения $f(i)$ и формулу принятия решений $x_\infty(i)$. Такие результаты обладают существенными достоинствами. Не только упрощаются вычисления, необходимые для получения конкретных ответов, но и в большей мере выявляется структура оптимальной стратегии и ее зависимость от экономических параметров. Методы решения задач с непрерывными переменными представляют наибольший интерес для операциониста-практика, в связи с чем здесь они более подробно не рассматриваются.

Ослабление допущения о стационарности. Отметим, каким образом допущение IV) было использовано в сетевой модели разд. 12.1. Гипотеза о стационарности заключалась в том, что множество имеющихся в сети дуг (i, j) и соответствующих им стоимостных показателей c_{ij} остается неизменным для всех отрезков. Однако, надлежащим образом определив переменные состояния, можно обобщить модели на большее число ситуаций, чем кажется с первого взгляда. Ниже приводится пример циклического варьирования значений параметров.

Рассмотрим простую модель управления запасами, описанную в разд. 12.5. В ней предполагалось, что для любого отрезка $D_t = 3$. Вместо этого предположим, что спрос может быть равным как 3, так и 2 единицам. Тогда число вершин, отображающих значения переменной состояния $i = 0, 1, \ldots, 4$, т. е. уровни запасов на начало отрезка, уже не ограничено пятью. Теперь имеется уровень запасов на начало отрезка при спросе за отрезок 3 единицы, который обозначим i' и примем равным $i' = 0, 1, \ldots, 4$, и уровень запасов $i'' = 0, 1, \ldots, 4$ при спросе 2 единицы. Таким образом, новая сеть состоит из двух групп по пять вершин на каждой. Допустимое решение о выпуске x единиц продукции в отрезке со спросом $D_t = 3$ отображается дугой из вершины i' в вершину $j' = i' + x - 3$. Аналогично этому решение о производстве x единиц при $D_t = 2$

единицы соответствует дуге из вершины i'' в вершину $j'' = i'' + x - 2$. При помощи такого способа можно также отображать колебание затрат на производство и хранение продукции от отрезка к отрезку.

Использование особенностей структуры. В модели восстановления, приведенной в разд. 11.4, показана возможность упрощения численного решения функционального уравнения при наличии информации о виде оптимальной стратегии. Другие такие примеры будут приведены в гл. 18 и приложении II (том III). Предваряя эти примеры, рассмотрим одну из операций метода итераций по стоимости, а именно нахождение

$$\min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha y^n + c_{ij}). \tag{2}$$

В некоторых задачах информация об экономических параметрах и о возможных решениях может свидетельствовать о том, что если дуга (i, j^*) оптимальна при $i = 0$, то эта дуга оптимальна и при всех $i \leqslant j^*$. Тогда если известно, что решением является $i = 0$, то может исчезнуть необходимость нахождения решений для многих других значений i.

В качестве другого иллюстративного примера рассмотрим

$$g_i^n(j) \equiv \alpha y_j^n + c_{ij} \quad \text{для всех } i \tag{I}$$

и предположим, что $j = 1, 2, \ldots, N_i$. Если модель позволяет установить вогнутость $g_i^n(j)$ на этом множестве j, то решением (2) попросту является либо $j = 1$, либо $j = N_i$. Таким образом, для каждого i в (2) потребуется проверять только два значения j. Если же функция $g_i^n(j)$ является выпуклой, возможно другое упрощение, при котором обычно приходится проверять более двух значений j, но нет необходимости рассматривать все N_i.

Эффективный метод поиска минимума выпуклой функции основан на использовании так называемых **чисел Фибоначчи**. По существу, метод применим для любой функции $g(i)$, причем такой, что $g(i)$ строго убывает при возрастании j в пределах $1 \leqslant j \leqslant j^*$ и строго возрастает при увеличении j, если $j > j^*$. Такие функции именуются **дитоническими**. Если $g(j)$ есть дитоническая функция, то $-g(j)$ называется **одновершинной** или унимодальной функцией.

Последовательность чисел Фибоначчи имеет следующий вид:

$$1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ 89,\ 144,\ \ldots, \tag{II}$$

так что в этой последовательности каждое последующее число есть сумма двух предыдущих. Пусть F_n — член n последовательности (II). Тогда, согласно так называемому «плану поиска с использованием чисел Фибоначчи», минимум функции $g(j)$, где $j = 1, 2, \ldots$ $\ldots, (F_n - 1)$, определяется с помощью $(n - 1)$ вычислений значения функции.

Так, например, предположим, что $F_n - 1 = 20$, т. е. $n = 7$. Тогда требуется вычислить только $n - 1 = 6$ значений. Начнем с числа Фибоначчи F_{n-1} и со следующего меньшего числа F_{n-2}. В данном примере $F_6 = 13$ и $F_5 = 8$. Вычислим и сравним $g(13)$ и $g(8)$. На основе этого сравнения можно выполнить еще одно вычисление с использованием очередного меньшего числа Фибоначчи, т. е. 5. В данном случае, если оказалось, что $g(8) \leqslant g(13)$, интервал поиска минимума функции $g(j)$ сокращается до $j = 1, 2, \ldots, 12$. При таком результате сравнения необходимо вычислить и сравнить $g(5)$ и $g(8)$. Если же, напротив, $g(8) > g(13)$, то интервал поиска сократился до $j = 9, 10, \ldots, 20$. При таком интервале следует вычислить значения функции для $j = 13$ $(j - 8 = 5)$ и $j = 16$ $(j - 8 = 8)$. Таким образом, в любом случае для очередного сравнения необходимо вычислить $g(j)$ только для одного нового значения j. Процесс продолжается указанным способом и завершается нахождением минимума $g(j)$ после проверки шести значений j.

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

1. Рассмотрите модель управления запасами, описанную в разд. 12.1. Пусть возможными значениями уровня запасов на начало отрезка i являются $i = 0, 1, \ldots, 7$, а спрос на отрезке равен $d = 4$.

а) Сколько различных значений $f(i)$ необходимо определить в модели (2)? Напишите экстремальное уравнение для каждого из этих значений и укажите, какие значения x следует рассматривать при минимизации.

б) Предположим, что оптимальная стратегия производства заключается в том, чтобы при $i \leqslant 5$ выпускать $11 - i$ единиц продукции, а при $i > 5$ прекращать выпуск. Объясните, каким образом можно вычислить значения $f(i)$ (составьте соответствующие экстремальные уравнения).

в) Вычертите сеть, отображающую всю оптимизационную задачу, и укажите маршрут, соответствующий оптимальной стратегии для варианта б) постановки задачи.

2. Рассмотрите сетевую постановку, описанную в разд. 12.1.

а) Объясните, почему значения y_i, соответствующие интергальным дисконтированным затратам для оптимального бесконечного маршрута, который начинается в вершине i, являются конечными.

б) Объясните, почему все y_i непременно удовлетворяют экстремальным уравнениям (3).

3. Рассмотрите сетевую постановку, описанную в разд. 12.1. Пусть в сети имеется шесть вершин, и все вершины попарно соединены дугами в обоих направлениях.

а) Сколько должно удовлетворяться экстремальных уравнений (3)? Напишите каждое из этих уравнений, приведя все их члены, которые должны учитываться при минимизации.

б) Пусть в оптимальный набор дуг входят дуги (*1*, *2*), (*2*, *3*), (*3*, *4*), (*4*, *5*), (*5*, *6*) и (*6*, *1*). Укажите, какую систему линейных уравнений необходимо решить для нахождения соответствующих значений y_i.

4. В алгоритме итераций по критерию, приведенном в разд. 12.2, не обязательно будет существовать только одна стратегия, оптимальная при принятии текущих решений для *всех* конечных $n \geqslant N$ — вне зависимости от того, сколь велико значение N. В тексте раздела утверждается, что существует такое n^*, что для всех $n > n^*$ любая стратегия, оптимальная при принятии текущих решений для конечного планового периода длительностью n отрезков, одновременно является оптимальной стационарной стратегией для модели с бесконечным плановым периодом. Противоречат ли эти утверждения друг другу? Объясните, почему.

5. Рассмотрите алгоритм итераций по критерию, приведенный в разд. 12.2. Предположим, что значения y_i^0 вычисляются на основе некоторой первоначально выбранной стратегии и что вычислительный процесс прекращается на итерации m. Опишите стратегию, позволяющую получить значения y_i^m. Можно ли считать, что эта стратегия будет обязательно стационарной?

6. Рассмотрите иллюстративный пример (I), приведенный после описания алгоритма итераций по критерию в разд. 12.2.

а) Начертите сеть, отображающую (I).

б) Проверьте, что значения (II) удовлетворяют функциональным уравнениям для сети и что оптимальными являются как дуга (*1*, *2*), так и дуга (*1*, *3*).

в) Если значения y_i^3 в (V) заданы, вычислите y_i^4 и y_i^5 и убедитесь в том, что дуга (*1*, *3*) лучше при $n = 4$, а дуга (*1*, *2*) лучше при $n = 5$.

7. В каждом из приведенных ниже вариантов постановки задачи вначале вычислите значения y_i с учетом анализа, проделанного в разд. 12.4. Затем примените алгоритм итераций по критерию, описанный в разд. 12.2, начав с $y_i^0 = 0$. Если вы обозначите через n' номер первой из итераций, на которой общая стратегия согласуется с оптимальной текущей стратегией, то прекратите расчеты на итерации $n' + 3$. Нанесите на график значения y_1^n и y_2^n.

а) Пример 1 на рис. 12.1 при $\alpha = 0{,}8$.

б) Пример 1 на рис. 12.1 при $\alpha = 0{,}5$.

в) Пример 4 на рис. 12.3 ($c_{12} = 1{,}5$) при $\alpha = 0{,}5$.

г) Пример 4 на рис. 12.3 ($c_{12} = 1{,}5$) при $\alpha = \frac{1}{4}$.

д) В каждом варианте убедитесь в том, что для всех i

$$|y_i^n - y_i| \leqslant \alpha d_{n-1}, \quad \text{где } d_{n-1} = \max_k |y_k^{n-1} - y_k|.$$

8. В каждом из приведенных ниже вариантов постановки задачи примените алгоритм итераций по критерию, приведенный в разд. 12.2,

вычисляя y_i^0 на основе указанного для этого варианта исходного набора дуг (исходной стратегии). Если вы обозначите через n' номер первой из итераций, на которой общая стратегия согласуется с оптимальной текущей стратегией, то прекратите расчеты на итерации $n'+3$. Нанесите на график значения y_1^n и y_2^n.

а) Пример 1 на рис. 12.1 при $\alpha = 0{,}8$; исходная стратегия включает дуги *(1, 4)*, *(4, 1)*, *(2, 3)* и *(3, 2)*.

б) Пример 1 на рис. 12.1 при $\alpha = \frac{1}{2}$; исходная стратегия включает дуги *(1, 4)*, *(4, 1)*, *(2, 1)* и *(3, 2)*.

в) Пример 4 на рис. 12.3 ($c_{12} = 1{,}5$) при $\alpha = \frac{1}{2}$; исходная стратегия включает дуги *(1, 3)*, *(3, 2)* и *(2, 3)*.

г) Пример 4 на рис. 12.3 ($c_{12} = 1{,}5$) при $\alpha = \frac{1}{4}$; исходная стратегия включает дуги *(1, 2)*, *(2, 3)* и *(3, 2)*.

д) В каждом варианте убедитесь в том, что для всех i

$$|y_i^n - y_i| \leqslant \alpha d_{n-1}, \quad \text{где } d_{n-1} \equiv \max_k |y_k^{n-1} - y_k|.$$

9. Рассмотрите сеть с вершинами *1* и *2* и дугами *(1, 1)*, *(2, 2)*, *(1, 2)* и *(2, 1)*, причем $c_{11} = c_{22} = 50$ и $c_{12} = c_{21} = 10$; $\alpha = \frac{1}{2}$. Изучите сеть и вычислите y_1 и y_2. Примените алгоритм итераций по критерию, приведенный в разд. 12.2, приняв $y_1^0 = 0$ и $y_2^0 = 40$; выполните итерации до $n = 5$. Нанесите на график y_1^n и y_2^n. Изменяются ли величины y_i^n монотонно?

10. Рассмотрите сетевую постановку, описанную в разд. 12.1. Предположим, что для каждой вершины i выбрана дуга (i, j).

а) Объясните, почему в выбранный набор дуг входит по крайней мере один цикл. Если имеются два цикла, объясните, почему циклы разъединены (т. е. почему вершина не может одновременно входить в несколько циклов).

б) Рассмотрите формулу расчета «стоимости вершин» (2) в разд. 12.2. Пусть выбранная стратегия включает только один цикл, а именно дуги *(1, 2)*, *(2, 3)* и *(3, 1)*. Составьте соответствующие уравнения (2). Покажите, что решением этой системы линейных уравнений являются:

$$y_1^0 = c_{12} + \alpha c_{23} + \alpha^2 c_{31} + \alpha^3 c_{12} + \alpha^4 c_{23} + \alpha^5 c_{31} + \ldots,$$
$$y_2^0 = c_{23} + \alpha c_{31} + \alpha^2 c_{12} + \alpha^3 c_{23} + \alpha^4 c_{31} + \alpha^5 c_{12} + \ldots,$$
$$y_3^0 = c_{31} + \alpha c_{12} + \alpha^2 c_{23} + \alpha^3 c_{31} + \alpha^4 c_{12} + \alpha^5 c_{23} + \ldots.$$

в) Пусть применяется алгоритм итераций по стратегиям, где значения y_i^0 приняты в соответствии с вариантом б). Далее, пусть на шагах 3 и 4 обнаруживается, что $Y_1^0 = \alpha y_3^0 + c_{13}$, $Y_1^0 < y_1^0$, а для всех остальных $Y_i^0 = y_i^0$. Покажите, что $y_1^1 < y_1^0$ и что как следствие этого для любой вершины i $y_i^1 \leqslant y_i^0$.

11. В разд. 12.3 для простоты предполагается, что из каждой вершины i существует маршрут к одной (по крайней мере) из вершин цикла, позволяющего достичь минимальных эквивалентных средних затрат за отрезок ($\alpha = 1$).

а) Какие осложнения возникают, если отбросить это допущение? (*Указание*: рассмотрите сеть, стоимости на дугах которой равны $c_{12} = c_{21} = c_{13} = 1$, $c_{34} = c_{43} = 2$.)

б) Объясните, почему при $0 \leqslant \alpha < 1$ не возникает необходимость в таком допущении.

12. а) Каков смысл величины w_i в выражениях (1) и (2) разд. 12.3?

б) Приведите промежуточные алгебраические преобразования, позволяющие из (3) получить (4).

в) Перенесите выражение из правой части (3) в левую; затем покажите, что при $\alpha = 1$ результат эквивалентен (6).

г) Предположим, что разность $w_i - w_j$ интерпретируется как такое приращение затрат по сравнению с затратами при оптимальной стратегии, которое обусловлено нахождением в вершине i, а не в вершине j. С учетом этой интерпретации опишите словами экстремальные уравнения (6).

д) Пусть даны значения w_i, удовлетворяющие экстремальным уравнениям (5) или (6). Объясните, почему прибавление к этим значениям константы позволяет получить новые значения w_i, также удовлетворяющие уравнениям (5) или (6). (Продемонстрируйте правильность ваших соображений на сети, включающей три вершины с попарным соединением их дугами в обоих направлениях.)

13. Рассмотрите алгоритм итераций по стратегиям, описанный в разд. 12.3, при $\alpha = 1$.

а) Пусть сеть включает вершины $1, 2, \ldots, p$. Сколько уравнений и сколько неизвестных будет иметь система линейных уравнений (7)?

б) Пусть пробная стратегия включает цикл (1, 2), (2, 3) и (3, 1). Составьте соответствующие уравнения (7) и покажите, что $\bar{c}^n = (c_{12} + c_{23} + c_{31})/3$, как это утверждается в (9).

в) Пусть дуга (i, j) входит в оптимальную стратегию. Чему равна разность $w_i - w_j$? Пусть дуга (j, k) также входит в эту же оптимальную стратегию. Чему равна разность $w_i - w_k$?

г) Измените описание шагов алгоритма итераций по стратегиям, который соответствует представлению экстремальных уравнений в виде (6).

д) Измените описание алгоритма итераций по стратегиям, если целью является максимизация среднего эффекта за отрезок.

14. Рассмотрите сетевую модель, в которой для перемещения по дуге (i, j) требуется h_{ij} отрезков, причем не все h_{ij} равны 1. Этот случай обсуждается в конце разд. 12.3.

а) Убедитесь в том, что (3), (4) и (5) действительно преобразуются в (I), (II) и (III).

б) Убедитесь в том, что $w_i - w_j$ по-прежнему может интерпретироваться как приращение затрат, обусловленное нахождением в вершине i, а не в вершине j, при $\alpha = 1$. С использованием этой интерпретации опишите словами экстремальные уравнения (III).

15. а) Обратившись к разд. 12.4, убедитесь в том, что решением (3) действительно является (4).

б) Проверкой алгебраических действий, выполненных в (5) и (6), убедитесь в том, что дуга (*1*, *4*) действительно позволяет достичь минимума в (5) при $0 \leqslant \alpha < 1$, а дуга (*2*, *1*) — минимума в (6) при $\frac{2}{3} \leqslant \alpha < 1$. Объясните, почему дуга (*2*, *3*) является лучшим решением при значениях α, меньших чем $\frac{2}{3}$.

в) Покажите, какие изменения необходимо внести в выражения (2) — (6) в том случае, когда в исходной стратегии вместо дуги (*2*, *1*) используется дуга (*2*, *3*). Придем ли мы к тем же выводам, что в варианте б)?

г) Убедитесь в том, что показатель возможности улучшения (6) для дуги (*2*, *3*) представляет собой интегральные дисконтированные затраты, обусловленные перемещением из вершины *2* в *3*, возвращением в *2*, затем перемещением в *1* и, наконец, движением по циклу между вершинами *1* и *4*. Сравните значение этого показателя со значением y_2^0 в варианте в).

16. а) В разд. 12.4 проверьте алгебраические действия в (7) — (11) и убедитесь в том, что при $0 \leqslant \alpha < 1$ минимум соответствует (11).

б) Проверьте, что при $\alpha = 1$ формулы расчета «стоимости вершин» имеют вид (12), а их решением является (13). Проверьте также расчеты в (14) и (15). Обсудите приведенное в тексте утверждение, что предпочтительней все же является цикл, включающий вершины *2* и *3*.

в) Предположим, что используется стратегия (1), которой при $0 \leqslant \alpha < 1$ соответствуют интегральные дисконтированные затраты (4). Рассмотрите разности $y_2 - y_1$, $y_3 - y_1$ и $y_4 - y_1$ и примите $\alpha = 1$. Покажите, что результирующие величины совпадают с $w_2 - w_1$, $w_3 - w_1$ и $w_4 - w_1$ в (13).

17. Рассмотрите иллюстративный пример на рис. 12.3 при $c_{12} = 1{,}5$ (пример 4).

а) Покажите, что если используются дуги (*2*, *3*) и (*3*, *2*), то выбор дуги (*1*, *2*) более предпочтителен, чем выбор дуги (*1*, *2*) для конечного планового периода при нечетном числе отрезков, и менее предпочтителен — при четном.

б) Проверьте, что в разд. 12.4 выражение (18) является решением (17). Проверьте алгебраическое действие в (19), а также то, что дуга (*1*, *2*) позволяет достичь минимума при $\frac{1}{3} \leqslant \alpha < 1$.

в) Покажите, как необходимо изменить выражения (17) — (19), в том случае если вместо дуги (*1*, *3*) используется дуга (*1*, *2*). Проверьте правильность (20).

г) Выполните расчеты алгоритма итераций по стратегиям при $\alpha = 1$. Проверьте, что решение, включающее дугу (*1*, *2*), является единственным оптимумом.

д) Рассмотрите стратегию, включающую дуги (*1*, *2*), (*2*, *3*) и (*3*, *2*). Примите $0 \leqslant \alpha < 1$; найдите значения y_1, y_2 и y_3, представив их в виде, аналогичном (18). Рассмотрите разности $y_2 - y_1$ и $y_3 - y_1$ и примите $\alpha = 1$. Покажите, что результирующие величины совпадают с $w_2 - w_1$ и $w_3 - w_1$ в (23).

18. Рассмотрите иллюстративный пример на рис. 12.3 при $c_{12} = 2$ (пример 5). Проверьте (21), а также и то, что дуга (*1*, *3*) оптимальна при $0 \leqslant \alpha < 1$.

19. Рассмотрите иллюстративный пример на рис. 12.3 при $c_{12} = 2$ и $\alpha = 1$ (пример 6).

а) Проверьте, что формулы (уравнения) расчета «стоимости вершин» при использовании дуги (*1*, *2*) имеют вид (22), а их решением является (23).

б) Составьте и решите аналогичные уравнения расчета «стоимости вершин» при использовании дуги (*1*, *3*).

в) Проверьте, что при использовании алгоритма итераций по стратегиям как дуга (*1*, *2*), так и дуга (*1*, *3*) являются оптимальными по показателю возможности улучшения.

20. Рассмотрите пример на рис. 12.4 при $\alpha = 1$ (пример 7), обсуждавшийся в конце разд. 12.4.

а) Составьте уравнение расчета «стоимости вершин» для стратегии (III) и проверьте, что их решением является (IV). Проверьте также значения показателей возможного улучшения, приведенные в (V).

б) Покажите, что стратегия (V) включает два цикла. Проверьте, что решением системы уравнений (VI) является (VIII). Проверьте также значения показателей возможного улучшения (VIII).

в) Завершите выполнение расчетов согласно алгоритму итераций по стратегиям и убедитесь в том, что стратегия (VIII) является оптимальной.

Упражнения 21—26 связаны с задачей фирмы «Надежный поставщик», приведенной в разд. 12.5. Все номера выражений относятся к этому же разделу.

21. а) Пусть $0 \leqslant \alpha < 1$. Используя систему обозначений, соответствующую (2) из разд. 12.1, запишите в полностью развернутом виде все функциональные уравнения, которые необходимо удовлетворить.

б) Пусть $\alpha = 1$. Введем обозначения $g(i) \equiv w_i$ и $\bar{g} \equiv \bar{c}$ взамен использованных в (6) из разд. 12.5. Запишите функциональные уравнения в этой системе обозначений и таким образом представьте экстремальные уравнения в виде, аналогичном (2) из разд. 12.1.

22. В каждом из приведенных ниже вариантов задачи начертите сеть, которая содержит только дуги, соответствующие стратегиям а) (7) и (11);

б) (12) — (16).

23. а) Проверьте, что (8) представляют собой уравнения расчета «стоимости вершин», соответствующие стратегии (7), и что (9) является решением (8). Приведите со всеми подробностями расчеты (10) и проверьте, что их результаты имеют вид (11).

б) Выполните со всеми подробностями остальные расчеты согласно алгоритму итераций по стратегиям и проверьте результаты, приведенные в таблице рис. 12.7. В частности, составьте и решите уравнения расчета «стоимости вершин» и сравните показатели возможного улучшения.

в) Пусть дана оптимальная стратегия (4); приняв $\alpha < 1$, найдите соответствующие значения $y_0, y_1, \ldots, y_4$. Рассмотрите разности $y_1 - y_0$, $y_2 - y_0$, $y_3 - y_0$, $y_4 - y_0$ и примите $\alpha = 1$. Покажите, что полученные величины совпадают с величинами $w_1 - w_0$, $w_2 - w_0$, $w_3 - w_0$ и $w_4 - w_0$ для оптимальной стратегии.

24. Примените алгоритм итераций по стратегиям ($\alpha = 1$), выбрав в качестве исходной стратегии набор дуг (*0*, *1*), (*1*, *0*), (*2*, *1*), (*3*, *2*) и (*4*, *3*).

25. Примените алгоритм итераций по стратегиям, определив y_i^0 в соответствии со стратегией (7). На каждой итерации укажите производственный цикл, обусловливаемый проверяемой стратегией. Примите:

а) $\alpha = 0{,}9$;

б) $\alpha = \frac{1}{2}$;

в) $\alpha = 0$.

г) Опишите влияние значения α на выбор оптимальной стратегии.

д) Каково наименьшее значение α и чему равна соответствующая процентная ставка, при которой стратегия (4) все еще остается оптимальной?

26. Примените алгоритм итераций по стоимости, приняв вначале $y_i^0 = 0$, а затем определяя для y_i^0 те значения, которые отвечают стратегии (7). Пусть n' обозначает номер первой итерации, на которой общая стратегия согласуется с оптимальной текущей стратегией для упражнения 25; прекратите расчеты на итерации $n' + 3$.

Примите:

а) $\alpha = 0{,}9$; б) $\alpha = \frac{1}{2}$; в) $\alpha = 0$.

27. Рассмотрите численный пример, приведенный в конце раздела 12.5.

а) Примите $\alpha = 1$. Проверьте, что стратегия (1) оптимальна, если разрешен объем выпуска $x = 6$ при $C(6) = 28{,}5$. Проверьте также, что стратегия (1) обусловливает производственную программу (II).

б) Определите оптимальную стратегию при $\alpha = \frac{1}{2}$ и сравните ее со стратегией (1).

28. Рассмотрите числовой пример, приведенный в разд. 12.5. В каждый из приведенных ниже вариантов постановки задачи внесите необходимые изменения, кроме того, измените сеть рис. 12.5 и определите оптимальную стратегию с помощью алгоритма итераций по стратегиям при $\alpha = 1$.

а) Спрос для каждого отрезка равен 2 единицам (вместо 3).

б) Спрос для каждого отрезка равен 4 единицам (вместо 3).

в) Остаток на конец отрезка не должен превышать 3 единиц (вместо 4).

г) Объем выпуска в пределах отрезка не может превышать 4 единиц (вместо 5).

д) Разрешены только объемы выпуска, равные $x = 0, 2, 4$ единицам.

е) Объем выпуска на каждом отрезке должен быть не меньше 1 единицы ($x \geqslant 1$).

ж) Условно-постоянные затраты на операции по переналадке в (3) составляют 10 единиц (вместо 13).

з) Стоимость содержания запасов равна $h = 5$ (вместо 1).

и) Затраты на содержание запасов при ставке h исчисляются исходя из среднего уровня запасов на отрезке, а не из уровня запасов на конец отрезка.

29. Рассмотрите иллюстративный пример модели управления запасами, приведенной в разд. 12.5. Пусть удовлетворение спроса разрешается задерживать на срок не более одного отрезка, однако удовлетворение задержанного спроса не должно превышать 3 единиц.

а) Соответствующим образом измените сеть рис. 12.5.

б) Пусть штраф за задержку удовлетворения единичного спроса на срок в один отрезок равен 10 единицам. Приведите соответствующие стоимости на дугах сети для варианта а).

30. В каждом из приведенных ниже вариантов постановки задачи составьте прямую и двойственную модели линейного программирования, описанные в разд. 12.6. Сначала примите $0 \leqslant \alpha < 1$, так чтобы условия задач были аналогичны приведенным на рис. 12.8 и 12.9. Затем примите $\alpha = 1$, так чтобы условия задач были аналогичны приведенным на рис. 12.10 и 12.11.

а) Пример на рис. 12.3 при $c_{12} = 1{,}5$ (пример 4 разд. 12.4). Приведите оптимальные решения прямой и двойственной задач линейного программирования при $\alpha = 1$.

б) Пример на рис. 12.4 (пример 7 разд. 12.4).

в) Модель управления запасами разд. 12.5 (рис. 12.5).

г) Приведите оптимальные решения прямой и двойственной задач линейного программирования для варианта в) при $\alpha = 0{,}9$; $\alpha = \frac{1}{2}$; $\alpha = 0$ (упражнение 25).

31. Специальный материал об «ослаблении стационарности» в разд. 12.7 содержит описание сети, отображающей модель управления запасами с переменным спросом.

а) Начертите описанную сеть. Пусть функции затрат соответствуют приведенным в разд. 12.5; попытайтесь угадать оптимальную стратегию, вычислите соответствующие средние затраты ($\alpha = 1$) и проверьте показатели возможного улучшения в алгоритме итераций по стратегиям, для того чтобы определить, действительно ли предлагаемая стратегия является оптимальной. (*Указание*: при выборе стратегии обратите пристальное внимание на характер переменного спроса и на функции затрат.)

б) Пусть переменный спрос принимает поочередно значения 3 и 1 вместо 3 и 2. Ответьте на те же вопросы, что и в варианте а).

в) Если вам не удалось угадать оптимальные стратегии в вариантах а) и б) определения этих стратегий, используйте алгоритм итераций по стратегиям.

32. Поиск с использованием чисел Фибоначчи. Эффективный метод поиска минимального значения дитонической функции описан в конце разд. 12.7. Примените этот метод для нахождения приведенных ниже функций:

а) $g(j) = (j-7)^2$ для $j = 0, 1, \ldots, 50$.

б) $g(j) = 2j + (j-7)^2$ для $j = 0, 1, \ldots, 50$.

в) $g(j) = -2j + (j-7)^2$ для $j = 0, 1, \ldots, 50$.

г) $$g(j) = \begin{cases} 100 - j & \text{для } j = 0, 1, \ldots, 10, \\ 90 - 3j & \text{для } j = 11, 12, \ldots, 20, \\ 30 + 5j & \text{для } j = 21, 22, \ldots, 30, \\ 180 + j & \text{для } j = 31, 32, \ldots, 40. \end{cases}$$

33. Объясните, как вы понимаете следующие термины:
функциональные, или экстремальные, уравнения;
стационарная стратегия;
цикл;
задача дискретного динамического программирования;
метод последовательных приближений;
метод итераций по критерию (метод последовательных приближений в пространстве функций);
метод итераций по стратегиям (метод последовательных приближений в пространстве стратегий);
монотонно возрастающий (убывающий);
сходимость сверху (снизу);
формула расчета «стоимости вершин»;
нормирующее уравнение;
числа Фибоначчи;
дитоническая функция;
одновершинная, или унимодальная, функция.

УПРАЖНЕНИЯ НА ПОСТАНОВКУ ЗАДАЧ И РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

34. Рассмотрите сеть, изображенную на рис. 12.12; наличие стрелок на обоих концах дуги означает, что перемещение по дуге разрешено в обоих направлениях и что затраты при этом соответствуют проставленному на дуге числу.

а) Пусть $\alpha = 1$. Решите соответствующие экстремальные уравнения (5) из разд. 12.3, используя метод итераций по стратегиям.

б) Примените метод итераций по стратегиям для нахождения y_i и соответствующей стратегии при $\alpha = 0{,}9$; $\alpha = \frac{1}{2}$; $\alpha = 0$. Используйте оптимальное решение варианта а) в качестве исходной стратегии.

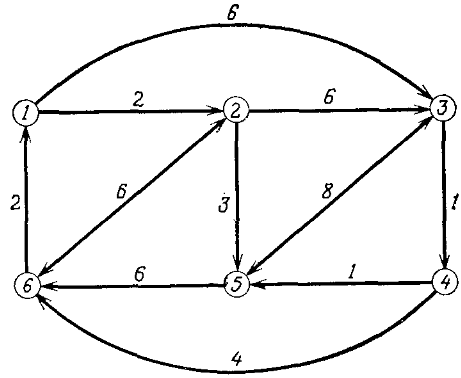


Рис. 12.12.

в) Примите $\alpha = 0{,}9$. Примените алгоритм итераций по критерию, начав с $y_i^0 = 0$. Если n обозначает номер первой итерации, на которой общая стратегия согласуется с оптимальной текущей стратегией, прекратите расчеты на итерации $n' + 3$.

35. Рассмотрите сеть, изображенную на рис. 12.13; наличие стрелок на обоих концах дуги означает, что перемещение по дуге разрешено в обоих направлениях и что затраты при этом соответствуют проставленному на дуге числу. Ответьте на вопросы, поставленные в упражнении 34.

В упражнениях 36—40 требуется изменить постановку задач по сравнению с упражнениями, приведенными в предыдущих главах. Как и ранее, дайте определения всем используемым обозначениям.

36. В каждом из приведенных ниже вариантов задачи принято, что затраты и эффект дисконтируются к текущему отрезку, причем α — коэффициент дисконтирования за отрезок, $0 \leqslant \alpha < 1$. Постройте соответствующую рекуррентную формулу динамического программирования для конечного планового периода, аналогичную формуле (1) разд. 12.1.

а) Задача расширения производственных мощностей в упражнении 33 гл. 10.

б) Задача общества «Биржевик» в упражнении 34 гл. 10.

в) Задача сглаживания занятости автомобильной компании «Гигант» в упражнении 36 гл. 10.

г) Задача разработки производственной программы в упражнении 39 гл. 10, где принято, что в течение отрезка j выпуск продукции составляет x_j. При заданных значениях $x_1, x_2, \ldots, x_{t-1}$ укажите оптимальное значение x_t.

37. В каждом из приведенных ниже вариантов постановки задачи примите, что все параметры являются стационарными, плановый

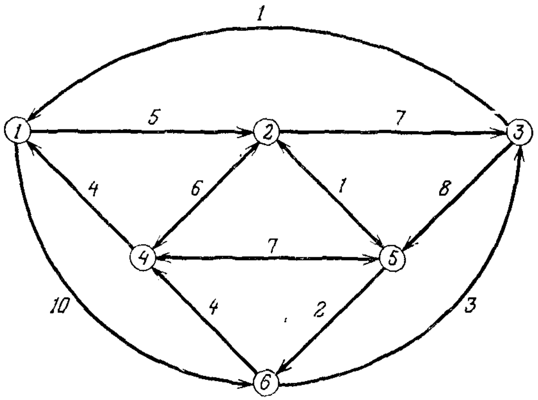


Рис. 12.13.

период — бесконечным, затраты и эффект дисконтируются к текущему отрезку, α — коэффициент дисконтирования за отрезок, $0 \leqslant \alpha < 1$. Постройте требуемую систему экстремальных уравнений, аналогичных (2) из разд. 12.1.

а) Модель сглаживания производства в упражнении 45 гл. 8.

б) Модель сглаживания производства в разд. 9.8.

в) Модель построения производственной программы в упражнении 53 гл. 9.

г) Задача о продаже скота в упражнении 32 гл. 10.

38. Рассмотрите модель сглаживания производства в упражнении 45 гл. 8. Используйте параметры модели в иллюстративном примере разд. 12.5; пусть «программный» объем выпуска продукции составляет $\bar{x} = 3$ единицы, а «издержки сглаживания» $v = 1$.

а) Примените алгоритм итераций по стратегиям для нахождения оптимальной стратегии при $\alpha = 1$.

б) Примените алгоритм итераций по стратегиям для нахождения оптимальной стратегии при $\alpha = 0{,}9$; $\alpha = 0{,}5$; $\alpha = 0$. В качестве исходной проверяемой стратегии примите оптимальную стратегию варианта а). Опишите влияние, которое значение α оказывает на выбор оптимальной стратегии.

в) Примите $\alpha = 0{,}9$. Примените алгоритм итераций по критерию, начав с $y_i^0 = 0$. Если n' обозначает номер первой итерации, на которой общая стратегия согласуется с оптимальной текущей стратегией, прекратите расчеты на итерации $n' + 3$.

39. Рассмотрите модель управления запасами, аналогичную приведенной в разд. 12.5. По-прежнему предполагается, что спрос $D = 3$, объем выпуска $x \leqslant 5$, уровень запасов на конец отрезка $j \leqslant 4$. Затраты на производство и содержание запасов описываются функцией $C(x) + hj$, где $h = 1$, а

$$C(x) = \begin{cases} 5x & \text{при } x = 0,\ 1,\ 2,\ 3, \\ 12 + x & \text{при } x \geqslant 4. \end{cases}$$

а) Примените алгоритм итераций по стратегиям для нахождения оптимальной стратегии при $\alpha = 1$.

б) Примените алгоритм итераций по стратегиям для нахождения оптимальной стратегии при $\alpha = 0{,}9$; $\alpha = 0{,}5$; $\alpha = 0$. В качестве исходной стратегии используйте оптимальное решение варианта а). Охарактеризуйте влияние выбора α на оптимальную стратегию.

в) Примите $\alpha = 0{,}9$. Примените алгоритм итераций по критерию, начав с $y_i^0 = 0$. Если n' обозначает номер первой итерации, на которой общая стратегия согласуется с оптимальной текущей стратегией, прекратите расчеты на итерации $n' + 3$.

г) Повторите варианты а), б) и в) при $h = 2$.

д) Повторите варианты а), б) и в) при спросе $D = 2$. (*Примечание*: аналогичные вопросы для конечного планового периода поставлены в упражнениях 13 и 15 гл. 9.)

40. Ответьте на вопросы, поставленные в упражнении 39, приняв, что функция затрат производства имеет вид

$$C(x) = \begin{cases} 0 & \text{при } x = 0, \\ 3 + 2x & \text{при } x = 1. \end{cases}$$

(*Примечание*: для конечного планового периода этот вопрос поставлен в упражнении 14 гл. 9.)

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

41. Рассмотрите сеть, аналогичную описанной в разд. 12.1, причем коэффициент дисконтирования удовлетворяет условию $0 \leqslant \leqslant \alpha < 1$.

а) Пусть задана пробная стратегия (не обязательно стационарная) и соответствующие интегральные дисконтированные показатели y_i^n. Предположим, что имеется хотя бы одна (а возможно, и несколько) вершин k, для которой

$$\min_{\substack{\text{по всем } (k,\, j) \\ \text{в сети}}} (\alpha y_j^n + c_{kj}) < y_k^n.$$

Рассмотрите *стационарную* стратегию, в которую включены дуги, позволяющие достичь минимума $(\alpha y_j^n + c_{ij})$ для каждой вершины i.

Покажите, что эта новая пробная стратегия позволяет достичь такого значения y_j^{n+1}, что $y_j^{n+1} \leqslant y_j^n$ для каждой вершины j, причем для вершины k это неравенство является строгим.

б) Предположим, что заключение, сделанное в варианте а), является верным; тогда покажите, что существует стационарная стратегия, которая лучше всех других возможных стратегий.

42. Рассмотрите сеть, аналогичную описанной в разд. 12.1, причем коэффициент дисконтирования α удовлетворяет условию $0 \leqslant \alpha < 1$. Пусть для каждой переменной y_i существует нижний предел L_i и верхний предел U_i, такие, что $L_i \leqslant y_i \leqslant U_i$. Покажите, что использование дуги (i, k) не приводит к оптимуму, если только

$$\alpha L_k + c_{ik} > \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\alpha U_j + c_{ij}).$$

Подробно разъясните, как можно использовать это неравенство для ускорения сходимости методов последовательных приближений.

43. Рассмотрите доказательство сходимости алгоритма итераций по критерию при $0 \leqslant \alpha < 1$, приведенное в материалах повышенной трудности в конце разд. 12.2. Измените доказательство так, чтобы показать, что алгоритм итераций по стратегиям обеспечивает скорость сходимости, по крайней мере не меньшую, чем скорость геометрической прогрессии (VII).

44. Рассмотрите двойственную модель линейного программирования при $\alpha = 1$, а именно (8) и (9) из разд. 12.6, а также соответствующую прямую модель (10) — (13).

а) Покажите, что мы ничего не упускаем, приняв, что все $w_i \geqslant 0$.

б) Покажите, что мы ничего не упускаем, приняв, что в прямой модели ограничения имеют вид неравенств типа «не меньше».

в) Покажите, что если все $c_{ij} \geqslant 0$, то без потери строгости можно предположить, что $\bar{c} \geqslant 0$ и что нормирующее ограничение (12) является неравенством типа «не меньше».

45. Предположим, что мы отказываемся от допущения, сделанного в разд. 12.3 и говорящего, что от любой вершины i имеется маршрут по крайней мере к одной из вершин, принадлежащей тому циклу, который позволяет достичь $\bar{c}$. Пусть $\bar{c}_i$ — минимальные средние затраты за отрезок, достижимые на тех маршрутах, которые начинаются в вершине i. Тогда аналогично экстремальным уравнениям (5) должны удовлетворяться следующие экстремальные уравнения:

$$w_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (w_j + c_{ij} + \bar{c}_j) \quad \text{для всех вершин } i,$$

$$\bar{c}_i = \min_{\substack{\text{по всем } (i,\, j) \\ \text{в сети}}} (\bar{c}_j) \quad \text{для всех вершин } i.$$

а) Придумайте эвристические доводы, обосновывающие эти экстремальные уравнения.

б) Покажите, каким образом эти уравнения приводят к шагам (I) и (II) алгоритма, приведенного в конце разд. 12.3.

в) Постройте соответствующие двойственную и прямую модели линейного программирования, аналогичные (8) — (13) из разд. 12.3. Соответствующая двойственная целевая функция имеет вид максимизации выражения $\sum_{k=1}^{p} r_k \bar{c}_k$, где каждое $r_k > 0$, однако в остальном может выбираться произвольно.

г) Объясните, почему эти экстремальные уравнения упрощаются до вида (5), если сделано особое допущение, приведенное в разд. 12.3.

46. Рассмотрите задачу полицейского управлеиия города Шум-гам-Сити, приведенную в упражнении 25 гл. 2, а в более общей постановке — в упражнении 27 гл. 5. Рассмотрите также результаты, полученные в упражнении 54 гл. 6, в которых утверждается, что если число отрезков является нечетным, то модель не эквивалентна сетевой задаче (следовательно, применение линейного программирования может и не давать целочисленного решения). Выясните, каким образом при любой длительности периода T ставить задачу в виде системы экстремальных уравнений, аналогичных (5) или (6) из разд. 12.3 для сетевой модели. В частности:

а) Продемонстрируйте построение требуемой сети. Присвойте индексы вершинам таким образом, чтобы они характеризовали состояние системы на каждом отрезке. Присвойте индексы дугам таким образом, чтобы был ясен характер соответствующего решения, и покажите, как учитываются ограничения задачи.

б) Напишите экстремальные уравнения, соответствующие результатам, полученным в варианте а), и покажите, почему решение поставленных вопросов позволяет получить оптимальную стратегию для исходной модели.

в) Модифицируйте ответы, полученные в вариантах а) и б), применительно к ситуации, в которой приступающий к работе в данной системе на отрезке t остается в ней и в течение отрезков $t+1$ и $t+2$, так что ограничения имеют вид $y_{t-2} + y_{t-1} + y_t \geqslant r_t$.

47. Рассмотрите следующий «смешанный» алгоритм. Примените алгоритм итераций по критерию в течение N итераций подряд. На итерации N используйте пробную стратегию и примените шаг 2 метода итераций по стратегиям. Прекратите расчеты, если экстремальные уравнения выполняются. В противном случае повторите вычисления, приняв в качестве y_i^{N+1} значения, полученные по формулам расчета «стоимости вершин» в алгоритме итераций по стратегиям. Объясните, почему этот «смешанный» алгоритм сходится за конечное число итераций.

ГЛАВА 13

Модели целочисленного программирования и комбинаторные модели

13.1. ПОИСК ФИЛОСОФСКОГО КАМНЯ

В этой главе, а также и в двух последующих исследуются модели математического программирования, в которых ослаблены два допущения: делимости и аддитивности (разд. 2.2). Напомним, что из указанных допущений вытекает линейность как целевой функции, так и ограничений и, кроме того, при этих условиях переменные могут принимать дробные значения, скажем 2,5 или $^{10}/_3$. Здесь снимается **допущение делимости** и рассматриваются задачи, в которых некоторые или все переменные могут принимать только целочисленные значения. В дальнейшем в гл. 14 и 15 снимается также допущение аддитивности.

Рассмотрим следующую модель:

$$\text{Оптимизировать } \sum_{j=1}^{n} c_j x_j \tag{1}$$

при ограничениях

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

$$x_j \geqslant 0, \qquad j = 1, 2, \ldots, n, \tag{3}$$

$$x_j \text{ — целые}, \quad j = 1, 2, \ldots, p \ (\leqslant n). \tag{4}$$

Задачи оптимизации такого класса называются задачами целочисленного (или **диофантового**, или **дискретного**) программирования. Если $p = n$, т. е. все переменные должны быть целыми числами, то модель определяет полностью целочисленную задачу. В противном случае, т. е. когда $p < n$, имеем частично целочисленную задачу.

В зависимости от конкретного содержания задачи оптимизация целевой функции (1) может иметь смысл максимизации или минимизации. (В рассматриваемых ниже алгоритмах и примерах всегда указывается характер искомого оптимума целевой функции.) Кроме того, любая задача целочисленного программирования может включать ограничения в виде неравенств ($\geqslant$) и равенств. [Линейные ограничения (2) можно считать заданными в канонической форме; во всех тех случаях, когда в исходной записи задачи они не соответствуют (2), следует применять преобразования, приведенные в гл. 3.]

Задача целочисленного программирования отличается от любой задачи линейного программирования условиями дискретности (4). В случае когда линейные ограничения (2) соответствуют некоторой сетевой модели, существует оптимальное решение задачи (1), (2) и (3), удовлетворяющее ограничениям целочисленности (4). Этот результат уже известен из теоремы о непрерывности для потоков в сетях. Однако в общем случае условие (4) накладывает дополнительные ограничения, вследствие которых максимальное значение целевой функции задачи целочисленного программирования оказывается меньше значения целевой функции соответствующей задачи линейного программирования.

Значение задач целочисленного программирования. Выше было показано, что, как правило, практическое использование моделей математического программирования связано с принятием плановых решений в сложных ситуациях. Часто встречаются условия, когда модели планирования содержат целочисленные переменные.

1. *Использование оборудования.* Переменными x_j можно обозначить единицы оборудования, которые должны функционировать в течение планового периода, описываемого моделью. Если каждая единица оборудования имеет большую мощность и высокую стоимость (например, автоматический винторезный станок, океанский танкер или 150-дюймовая бумагорезная машина), то дробное значение x_j, скажем $^{10}/_3$, может и не иметь смысла (т. е. оказаться нереализуемым) в рамках реальной задачи принятия решения. В таком случае на значения x_j приходится наложить требование целочисленности.

2. *Затраты на подготовку производства.* Может возникнуть необходимость рассмотрения операции, выполнение которой связано с так называемыми постоянными затратами (или затратами на *подготовку* производства) C_j всегда, когда соответствующее значение $x_j > 0$, причем C_j не зависит от фактического значения x_j. Так, например, если x_j определяет часовой показатель использования объема доменной печи на металлургическом заводе, то C_j соответствует затратам на задувку домны и на ее нагрев до заданной температуры. В разд. 13.2 будет показано, как учесть постоянные затраты в модели математического программирования за счет введения целочисленных переменных.

3. *Размеры партий.* При разработке некоторых производственных планов на значения x_j могут накладываться ограничения вида $x_j = 0$ или $x_j \geqslant L_j$. Так, например, величина x_j может представлять собой количество определенных изделий, которое нужно выпустить в течение периода t, а L_j — минимально возможный размер партии этих изделий. Подобное условие является примером ограничения вида «или — или», и, как будет показано в разд. 13.2, его можно формально ввести в модель, используя целочисленные переменные.

4. *Решения типа «да — нет».* Может возникнуть необходимость определения ситуаций «или — или» иного характера. С этой целью на переменные x_j накладываются ограничения $x_j = 1$ или $x_j = 0$, соответствующие решениям «принять» или «отвергнуть» либо «да»

или «нет». В качестве примера можно принять, что $x_j = 1$ соответствует решению строить новое промышленное предприятие, или организовать сбыт продукции в новом районе, или приобрести новое предприятие, или продать имеющееся в распоряжении фирмы оборудование. Довольно часто альтернативы такого рода относятся к разряду решений о *распределении капиталовложений*, поскольку их реализация связана с крупными затратами денежных средств и ресурсов. Это можно считать основной причиной, объясняющей, почему целочисленное программирование играет столь важную роль в организационных решениях. Оптимальное решение задачи распределения капиталовложений может принести фирме гораздо большую прибыль, чем приближенное или угаданное. Так, например, может оказаться, что фирме, производящей цемент и имеющей 25 предприятий, удастся значительно повысить получаемую прибыль, сократив число своих предприятий до 20 или менее *при условии*, что такое мероприятие предусматривается оптимальным планом. Сокращение накладных расходов при уменьшении числа предприятий может с избытком компенсировать рост транспортных затрат, если размещение оставшихся предприятий также подчиняется оптимальному плану.

В ряде других задач принятия решений может потребоваться построение моделей целочисленного программирования. Один класс таких задач связан с решениями, определяющими выбор упорядочений, календарных планов (расписаний) и маршрутов. Примером задачи, принадлежащей к этому классу, может служить *задача коммивояжера*. В этой задаче нужно отыскать кратчайший маршрут коммивояжера, который должен побывать в каждом из n городов, причем маршрут начинается и заканчивается городом 1. В качестве другого примера задач указанного класса можно взять задачу составления *расписания обработки деталей на станках*. Наиболее простым вариантом этой задачи является случай, когда каждая из n деталей должна обрабатываться на каждом из k станков. Предположим, что деталь нельзя обрабатывать на станке j, пока не закончена ее обработка на станке $j - 1$. Предположим, далее, что время обработки каждой детали на каждом станке строго фиксировано. Тогда оптимальная последовательность определяется расписанием обработки, минимизирующим общую продолжительность обработки всех деталей на всех станках. Другими примерами задач упорядочения, выбора маршрута и календарного планирования являются задачи *баланса сборочных линий*, *сетевого планирования при ограниченных ресурсах* (разд. 6.6), *предупредительного ремонта при ограничениях на рабочую силу* и *диспетчирования автотранспорта*.

Несмотря на большое внимание, уделяемое в литературе по исследованию операций задачам упорядочения, календарного планирования и выбора маршрута, которые отображаются моделями целочисленного программирования, до сих пор эти модели находят лишь ограниченное применение на практике. Одна из причин, объясняющих такое положение, заключается в том, что в моделях целочисленного программирования часто не учитываются некоторые важные условия, характерные для реальных задач календарного планирования. Другая причина — вычислительные трудности, оказавшиеся настолько значительными, что затраты, связанные с получением решений для моделей реальной размерности, превышают эффект, достигаемый за счет реализации найденного решения.

Задачи упорядочения, календарного планирования и выбора маршрута являются частными случаями *комбинаторных* задач.

Комбинаторная оптимизационная задача состоит в отыскании среди конечного множества альтернатив одной, которой отвечает экстремальное значение принятой целевой функции. Так, например, в задаче коммивояжера при n городах это конечное множество содержит $(n-1)!$ различных допустимых маршрутов, начинающихся и заканчивающихся в городе 1. В самой простой модификации задачи обработки деталей конечное множество альтернатив включает $(n!)^k$ допустимых последовательностей обработки n деталей на k станках.

При достаточной изобретательности всегда можно сформулировать нетривиальную постановку комбинаторной задачи оптимизации в терминах целочисленного программирования. Некоторые математические приемы, приводящие к таким постановкам, изложены в разд. 13.2.

Довольно часто задачи целочисленного программирования имеют очень большую размерность (т. е. в них фигурирует огромное число переменных и ограничений) и поэтому их постановка не имеет практического смысла. Однако, как будет показано, существует несколько алгоритмических методов решения задач целочисленного программирования, которые можно непосредственно применять к комбинаторным моделям без их предварительного преобразования в целочисленные модели.

Поиск длиною в жизнь. Как решить, что сложнее: задачи целочисленного или линейного программирования? Пожалуй, большинство ответит, что задачи целочисленного программирования проще. В защиту этого мнения говорят, что дихотомия, например $x_j = 0$ или $x_j = 1$, содержит гораздо меньше альтернатив, чем непрерывный интервал $0 \leqslant x_j \leqslant 1$. Исходя из этого делают вывод, что поиск оптимума в первом случае должен быть проще. Однако, к сожалению, на самом деле справедливо обратное. В моделях математического программирования вычисления значительно проще в случае непрерывных переменных, чем в случае дискретных. Это весьма убедительно демонстрирует следующий простой пример, который приводится ниже.

Рассмотрим модель

$$\text{Максимизировать } 21x_1 + 11x_2 \qquad (5)$$

при ограничениях

$$7x_1 + 4x_2 \leqslant 13, \tag{6}$$

$$x_1 \text{ и } x_2 \text{ — неотрицательные целые.} \tag{7}$$

Почти сразу можно определить, что единственным оптимальным решением является $x_1 = 0$, $x_2 = 3$. Ясно, что поскольку размерность этой задачи чрезвычайно мала, то для отыскания решения и его проверки можно воспользоваться самыми различными методами (включая метод динамического программирования, изложенный в разд. 10.2 применительно к задаче «распределения усилий»). Однако, поскольку здесь ставится цель продемонстрировать относительную сложность решения задач целочисленного программирования средней и большой размерности, сразу же исключим из рассмотрения любую схему, подходящую для решения некоторой частной задачи, а также общие алгоритмы, позволяющие находить решения лишь для задач малой размерности.

Существует один подход к отысканию целочисленных решений, который часто предлагают начинающие. Этот подход сводится к тому, что сначала не учитываются условия целочисленности и определяется оптимальное решение обычной задачи линейного программирования. Если полученное решение удовлетворяет ограничениям целочисленности, то оно является оптимальным решением и исходной целочисленной задачи (покажите, что это утверждение справед ливо). В противном случае предлагается переходить к целочисленному решению с помощью округления решения обычной задачи линейного программирования до целых чисел.

Предположим, что такой подход применяется к примеру (5) — (7). Оптимальное дробное решение соответствующей задачи линейного программирования равно $x_1 = {}^{13}/_7$ и $x_2 = 0$. Очевидное округленное решение равно $x_1 = 2$ и $x_2 = 0$, что является недопустимым. Решение с округлением в «меньшую сторону» будет $x_1 = 1$ и $x_2 = 0$. Это решение допустимо, но оно очень далеко от оптимального. Трудно представить себе даже возможность существования общего, имеющего практическую ценность метода округления дробного решения для задач средней и большой размерности, который давал бы оптимальное целочисленное решение.

Если некоторые коэффициенты a_{ij} в линейных ограничениях (2) в какой-либо конкретной модели отрицательны, то сама задача округления обычного дробного решения до *допустимого* целочисленного может оказаться достаточно сложной. Таким образом, хотя с помощью метода округления в некоторых случаях и удается отыскивать решения целочисленных задач, нельзя рассчитывать, что этот метод можно использовать как универсальный.

Еще более неприятная особенность, свойственная задачам целочисленного программирования, заключается в том, что нет простого способа, позволяющего определить, является ли данное допустимое решение оптимальным. В этом одно из важных отличительных различий между задачами целочисленного и линейного программирования. Чтобы показать это различие, предположим, что в приведенной выше задаче (5) — (7) нужно проверить, является ли решение $x_1 = x_2 = 1$ оптимальным. С этой целью нужно проверить, соответствует ли это решение какому-либо локальному оптимуму в том смысле, что значение целевой функции не улучшается в любой соседней целочисленной точке (или в **узле решетки**) $x_1 = 1 + d$ и $x_2 = 1 + e$, где d, $e = -1$; 0; 1. Покажите, что допустимыми соседними точками в данном случае являются точки ($x_1 = x_2 = 0$; $x_1 = 0$ и $x_2 = 1$; $x_1 = 0$ и $x_2 = 2$; $x_1 = 1$ и $x_2 = 0$). Покажите также, что решение $x_1 = x_2 = 1$ на самом деле лучше *всех* этих решений, но тем не менее *не является* оптимальным. Следовательно, некоторая точка может быть локально оптимальной по отношению к соседним точкам решетки, но все-таки не соответствовать глобальному оптимуму. (Заметим, что такая проверка на локальный оптимум с помощью перебора для задач большой размерности в сущности не годится, ибо может существовать огромное число соседних точек. В самом деле, если задача описывается полностью целочисленной моделью и на каждую переменную x_j наложено ограничение $x_j = 0$ или $x_j = 1$, то поиск всех соседних точек любого допустимого решения равносилен перебору всех решений исходной задачи.)

Сейчас уже пора задать вопрос, сводится ли наилучший общий метод решения к простому перебору всех допустимых решений и последующему выбору наилучшего из них. Если имеется всего несколько допустимых решений, то такой метод действительно проще реализовать, чем любой из рассмотренных далее в этой главе алгоритмов. В ряде случаев, встретившихся на практике, именно такой метод и применялся (часто многие допустимые решения сразу исключаются из рассмотрения как «явно» неоптимальные). Однако, как правило, метод полного перебора оказывается неработоспособным. Это объясняется тем, что число допустимых решений не всегда конечно, но даже тогда, когда условие конечности не нарушается, это число обычно огромно. Рассмотрим, например, случай отыскания оптимального решения задачи целочисленного программирования, содержащей 100 переменных, каждая из которых может принимать всего два значения — 0 и 1. Время перебора всех 2^{100} возможных решений на самой быстродействующей ЭВМ намного превышает продолжительность человеческой жизни.

Практически эффективные алгоритмы. Из всего изложенного выше очевидно, что общий алгоритм решения задач целочисленного программирования должен исключать необходимость *явного* перебора всех допустимых альтернатив. Требуются методы, обеспечивающие *частичный* перебор сравнительно небольшого числа допустимых вариантов и неявный перебор всех остальных. Напомним, что симплексный метод, применяемый для решения обычных задач линейного программирования, обладает именно такими характеристиками. Этот метод предусматривает оценку лишь небольшого числа всех допустимых базисных решений. (Отметим попутно, что удовлетворительная математическая теория, объясняющая *причину* столь малого числа базисных решений при практическом применении симплексного метода, до сих пор не разработана.) Аналогично в рекуррентных соотношениях, используемых в методе динамического программирования, применяют *принцип оптимальности*, позволяющий устранить необходимость перебора всех допустимых решений. Эффективность этих методов оптимизации, основанных на частичном переборе, оправдывает постановку вопроса о поиске аналогичных подходов к решению задач целочисленного программирования.

Разработка таких алгоритмов для целочисленных задач представляет собой область интенсивных исследований специалистов по прикладной математике. Пока еще преждевременно выделять какие-либо алгоритмы, утверждая, что они столь же эффективны для целочисленных задач, как симплексный алгоритм для обычных задач линейного программирования. По существу, рассматривая любой из предложенных до сих пор алгоритмов, легко построить примеры целочисленных задач, для решения которых с помощью этих алгоритмов требуется объем вычислений, эквивалентный полному перебору. Поэтому практическую эффективность любого из алгоритмов нужно в конечном счете оценивать эмпирически, выделяя классы задач, для которых рассматриваемый алгоритм оказывается наиболее выгодным.

В различных подходах к целочисленным задачам, которые на данном этапе представляются перспективными, складываются несколько общих принципов, имеющих фундаментальное значение. Эти идеи будут изложены в последующих разделах, начиная с раздела 13.3. Примечательно, что в большинстве алгоритмов в качестве элементарных блоков используются модели линейного программирования и понятия динамического программирования. Опыт успешного решения реальных целочисленных задач быстро накапливается. Если операционистам удастся усовершенствовать алгоритмы их решения, то при условии использования быстродействующих ЭВМ этим ученым действительно можно будет приписать честь открытия философского камня.

13.2. ПОСТАНОВКИ ЗАДАЧ ЦЕЛОЧИСЛЕННОГО ПРОГРАММИРОВАНИЯ

При оценке различных моделей планирования, изложенных в гл. 2, легко представить себе ситуации, в которых определенные переменные должны принимать только целочисленные значения. Имеется бесчисленное множество других примеров, относящихся прежде всего к задачам распределения ресурсов в течение короткого периода времени (суток, недели, месяца), которые также можно отобразить в виде моделей целочисленного линейного программирования. Иллюстрацией таких примеров могут служить задачи выбора автомашин различных габаритов для доставки заданных количеств определенных грузов; задачи календарного планирования работы бумагоделательных машин и прокатных станов при различных номенклатурах продукции, заданной требованиями заказчика; задачи распределения крупных заказов по предприятиям фирмы, отличающимся производственной мощностью, эффективностью и издержками производства. Зная методы построения моделей линейного программирования, легко сформулировать такие задачи распределения ресурсов в виде моделей целочисленного программирования. Поэтому в данной главе подобные задачи более подробно не рассматриваются. Тем не менее было бы неверным, если бы у читателя создалось впечатление, что задачи этого класса составляют основное содержание проблемы принятия организационных решений.

Напротив, наиболее существенными являются задачи, в которых рассматриваются вопросы постоянных затрат: задачи двоичного (дихотомического) выбора (типа «да — нет» или «принято — отвергнуто») и сложные задачи распределения капиталовложений. Построение моделей целочисленного программирования, отображающих подобные ситуации, до некоторой степени являются искусством. В связи с этим ниже поясняются некоторые типовые приемы, которые полезны для решения этой задачи.

Для расширения кругозора читателя приводится также пример постановки комбинаторной задачи в виде модели целочисленного программирования. В качестве такого примера выбрана задача коммивояжера. Существенным моментом в этом примере является демонстрация связи между комбинаторными задачами и задачами целочисленного программирования. Хотя целочисленные постановки часто не имеют непосредственно практической ценности, ибо даже небольшая комбинаторная задача обычно превращается в целочисленную модель огромной размерности, эта связь позволяет выяснить соотношение между идеями, лежащими в основе важнейших алгоритмов решения целочисленных задач.

Размещение предприятий: пример с постоянными затратами. В гл. 6 изучалась классическая транспортная задача

$$\text{Минимизировать} \quad \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \tag{1}$$

при ограничениях

$$\sum_{j=1}^{n} x_{ij} \leqslant S_i; \quad i=1,\ 2,\ \ldots,\ m \ \text{(поставка)}, \tag{2}$$

$$\sum_{i=1}^{m} x_{ij} \geqslant D_j, \quad j=1,\ 2,\ \ldots,\ n \ \text{(спрос)}, \tag{3}$$

$$x_{ij} \geqslant 0 \quad \text{при любом } i \text{ и } j. \tag{4}$$

Вспомним, что одна из интерпретаций этой задачи сводилась к тому, что имеется m поставщиков (пунктов отправления), располагающих товарами, которые нужно доставить n потребителям (в n пунктов назначения). Поставщик i может отправить потребителю не более S_i изделий, а потребителю j требуется не менее D_j изделий. Затраты на транспортировку единицы груза из пункта i в пункт j составляют c_{ij}. В задаче нужно выбрать план перевозок, минимизирующий общие транспортные затраты.

В ряде практических применений часть этой задачи заключается в том, чтобы определить, каких именно из m возможных поставщиков целесообразно использовать. Поставщиками, как правило, являются предприятия или склады, работа которых сопряжена с накладными расходами, включаемыми в затраты на транспорт при определении прибыли фирмы. Обычно необходимо сопоставить выигрыш (экономию), достигаемый за счет использования меньшего числа поставщиков (предприятий и складов), с потерями, обусловленными повышением собственно транспортных расходов.

Рассмотрим в качестве примера фирму «Прогресс», намеренную расширить сферу своей деятельности, включив в нее обслуживания потребителей, проживающих в штатах, которые лежат к западу от Скалистых гор. Чтобы обеспечить решение этой проблемы, фирма «Прогресс» планирует открыть несколько новых сборочных заводов. Предположим, что пункт i представляет собой одну из возможных точек размещения нового завода. Предположим также, что производственная мощность этого завода равна S_i, а постоянные затраты, связанные с его эксплуатацией, равны $F_i \geqslant 0$ независимо от фактического объема выпуска. Существует всего m возможных пунктов размещения новых предприятий, однако постоянные затраты, связанные с их эксплуатацией, настолько значительны, что открывать их во всех этих пунктах было бы нерационально. Примем, что величина $c_{ij} \geqslant 0$ представляет собой совокупные производственные и транспортные затраты, которые необходимы для доставки одного изделия из пункта изготовления i в пункт сбыта j. Кроме того, примем, что существуют постоянные затраты $F_{ij} \geqslant 0$, связанные с обеспечением транспортировки грузов по маршруту из пункта i в пункт j. Значение величины F_{ij} не зависит от значения $x_{ij} > 0$, но при $x_{ij} = 0$ она также равна нулю.

Для того чтобы построить математическую модель, введем целочисленные управляемые переменные:

$$y_i = \begin{cases} 1, & \text{если выбран пункт размещения } i, \\ 0 & \text{в противном случае,} \end{cases} \qquad i = 1, 2, \ldots, m;$$

$$z_{ij} = \begin{cases} 1, & \text{если маршрут из пункта } i \\ & \text{в район } j \text{ используется,} \\ 0 & \text{в противном случае} \end{cases} \qquad \text{для всех } i \text{ и } j.$$

Тогда вместо целевой функции (1) в обычном виде получаем целевую функцию следующего вида:

$$\text{Минимизировать} \quad \sum_{i=1}^{m} F_i y_i + \sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij} x_{ij} + \sum_{i=1}^{m} \sum_{j=1}^{n} F_{ij} z_{ij}. \tag{5}$$

Вместо ограничения (2) на мощности поставщиков вводим ограничения на пропускные способности:

$$\sum_{j=1}^{n} x_{ij} - S_i y_i \leqslant 0, \quad i = 1, 2, \ldots, m \text{ (поставки)}. \tag{6}$$

Если пункт размещения i не выбран, т. е. $y_i = 0$, то ограничение (6) обеспечивает недопустимость неравенства $x_{ij} > 0$. Ограничения (3) по спросу потребителей и требование (4) неотрицательности переменных остаются без изменений. Для некоторых алгоритмов решения задач целочисленного программирования достаточно, кроме того, указать, что все переменные y_i и z_{ij} должны быть равны либо 0, либо 1. Однако существуют также алгоритмы, для которых эти ограничения должны быть выражены в развернутой форме:

$$y_i \leqslant 1, \; z_{ij} \leqslant 1 \quad \text{при любом } i \text{ и } j, \tag{7}$$

$$y_i z_{ij} \text{ — неотрицательные целые при любом } i \text{ и } j. \tag{8}$$

[Объясните, почему условия (7) и (8) совместно гарантируют, что любая величина y_i или z_{ij} равна либо 0, либо 1.]

Построение модели еще не закончено, ибо не установлена зависимость между значениями z_{ij} и x_{ij}. Нужно добиться того, чтобы условие $x_{ij} > 0$ выполнялось только в случае, когда $z_{ij} = 1$. Это достигается с помощью линейных ограничений

$$x_{ij} - U_{ij} z_{ij} \leqslant 0 \quad \text{при любом } i \text{ и } j, \tag{9}$$

где U_{ij} — коэффициент, имеющий достаточно большое значение. Например, можно принять $U_{ij} = \min(S_i, D_j)$. Покажите, что если $z_{ij} = 0$, то из условия неотрицательности (4) и условия (9) следует, что значение x_{ij} должно быть равно нулю, а если $z_{ij} = 1$, то условие (9) при наличии нового ограничения (6) на мощности поставщиков и ограничения (3) на спрос потребителей является избыточным. [Убедитесь также в том, что в оптимальном решении никогда не требуется, чтобы выполнялось условие $z_{ij} = 1$ и $x_{ij} = 0$.]

Легко построить примеры, в которых обычная задача линейного программирования, определяемая (5) — (7), (9), имеет оптимальное решение с дробными значениями переменных. Поэтому ограничение (8), требующее целочисленных значений y_i и z_{ij}, весьма существенно. Как и в обычной транспортной задаче, необходимо только поставить условие неотрицательности всех x_{ij}, поскольку свойство целочисленности этой задачи сохраняется при условии целочисленности y_i и z_{ij}.

Условие (9) обобщается на другие случаи. При наличии постоянных затрат $C_j \geqslant 0$, сопоставляемых обычной переменной x_j в модели линейного программирования, для постановки целочисленной задачи нужно ввести двоичную (булеву) переменную $\left(z_j = \begin{cases} 1 \\ 0 \end{cases}\right)$ и ограничение $(x_j - Uz_j \leqslant 0)$, где значение U выбирается достаточно большим. Заметим, что при этом модель усложняется, поскольку добавляются одна целочисленная переменная и по линейному ограничению на каждую переменную, входящую в постоянные затраты.

Размеры партий. Пример с альтернативными ограничениями. Предположим, что x_j есть количество изделий, которое выпускается в течение некоторого планового периода. При этом возможны две ситуации: $x_j \geqslant L_j$, где L_j — наименьший допустимый размер партии, либо $x_j = 0$ в течение всего рассматриваемого периода. Предположим, что можно задать достаточно большое число U_j, такое, что ограничение $(x_j \leqslant U_j)$ в оптимальном решении наверняка выполняется. Тогда дихотомию ($x_j = 0$ или $x_j \geqslant L_j$) можно выразить, введя булеву переменную ($y_j = 0$ или $y_j = 1$) и два линейных ограничения

$$x_j - U_j y_j \leqslant 0, \tag{10}$$

$$x_j - L_j y_j \geqslant 0. \tag{11}$$

При $y_j = 0$ из ограничений (10) и (11) следует, что изделия не выпускаются ($x_j = 0$). При $y_j = 1$ ограничение (10) теряет смысл, а из ограничения (11) вытекает заданное условие на размер партии.

Изложенный подход можно обобщить, распространив его на случаи, когда решение должно удовлетворять по крайней мере k из p ограничений:

$$\sum_{j=1}^{n} a_{ij} x_j \leqslant b_i, \quad i = 1, 2, \ldots, p. \tag{12}$$

Введем p булевых переменных $y_i = 0$ или 1, $i = 1, 2, \ldots, p$ и наложим $1 + p$ линейных ограничений:

$$\sum_{i=1}^{p} y_i \geqslant k, \tag{13}$$

$$\sum_{j=1}^{m} a_{ij} x_j \leqslant b_i y_i + U_i (1 - y_i) \tag{14}$$

или

$$\sum_{j=1}^{n} a_{ij} x_j - (b_i - U_i) y_i \leqslant U_i, \quad i = 1, 2, \ldots, p,$$

где значение U_i выбирается столь большим, чтобы условие $\left(\sum_{j=1}^{n} a_{ij} x_j \leqslant U_i\right)$ обязательно выполнялось в оптимальном решении.

Объясните, почему ограничения (13) и (14), а также условие, что любая величина y_i равна либо 0, либо 1, являются корректной постановкой задачи. [На самом деле ограничение (13) можно записать в виде равенства. Почему?]

Распределение капиталовложений. Пример с взаимозависимыми альтернативами. Пожалуй, наиболее важными организационными решениями, приводящими к задачам целочисленного программирования, являются альтернативы «да — нет». Альтернативы такого рода часто возникают в стратегических моделях планирования, относящихся к нескольким плановым интервалам. Обычно значение переменной x_j определяет выбор конкретного проекта (операции или варианта капиталовложений), где

$$x_j = \begin{cases} 1, & \text{если проект } j \text{ выполняется}, \\ 0 & \text{в противном случае.} \end{cases}$$

Соответствующий коэффициент a_{ij} в ограничении часто выражает количество лимитирующего ресурса, скажем наличных денег, требующихся для выполнения проекта j на интервале i, где общее количество этого ресурса, используемое на каждом интервале, ограничено. Ограничения подобного рода аналогичны ограничениям, встречающимся в обычных моделях линейного программирования, предназначенных для решения задач долгосрочного стратегического планирования. Поэтому конкретные примеры таких задач здесь не приводятся.

Однако вследствие того, что переменная x_j является двоичной, ее можно использовать для отображения ограничений комбинаторного типа, часто характеризующих задачи распределения капиталовложений. Предположим, к примеру, что нужно наложить ограничение на решение, потребовав, чтобы оно допускало выбор не более k из первых p проектов. (Возможно, что для реализации каждого проекта требуется назначение ведущего инженера, а имеется возможность выделить всего k инженеров такой квалификации.) Тогда это условие можно выразить с помощью линейного ограничения

$$\sum_{j=1}^{p} x_j \leqslant k. \tag{15}$$

Если записать ограничение (15) в виде равенства, то будут приняты к выполнению в точности k проектов. Особый случай наличия p взаимоисключающих проектов можно учесть, приняв в (15) $k = 1$.

Этот случай может встретиться, если первые p проектов на самом деле относятся к одной альтернативе, а переменная x_j указывает, должна ли эта альтернатива быть реализована на интервале j или нет.

Если, допустим, можно принять вариант x_2 только при условии, что принят вариант x_1, то линейное ограничение

$$-x_1 + x_2 \leqslant 0 \tag{16}$$

на две булевы переменные обеспечивает выполнение этого условия, в чем читателю следует убедиться самостоятельно. Предположим теперь, что альтернативы x_1 и x_2 являются взаимоисключающими и что x_3 можно выполнять только при условии, когда принят либо x_1, либо x_2. Покажите, что такая зависимость выражается линейными ограничениями

$$x_1 + x_2 \leqslant 1, \tag{17}$$

$$-x_1 - x_2 + x_3 \leqslant 0. \tag{18}$$

Теперь уже очевидно, что при достаточной изобретательности можно самые разнообразные комбинаторные зависимости между проектами выразить с помощью линейных ограничений, накладываемых на булевы переменные.

Задача коммивояжера. Коммивояжеру нужно побывать в каждом из n городов, начиная и заканчивая свой маршрут городом 1. Он не имеет права дважды заезжать ни в один из городов. Обозначим через $c_{ij} \geqslant 0$ расстояние между городами i и j, приняв $c_{ij} = \infty$, если прямого маршрута между городами i и j не существует. В некоторых случаях $c_{ij} \neq c_{ji}$. Оптимизационная задача заключается в отыскании кратчайшего **цикла**. (Применение этой модели для решения задачи календарного планирования производства было описано в разд. 6.5.)

Предложено несколько математических постановок этой задачи. В приводимой ниже постановке используется относительно небольшое число переменных. Определим булевы переменные следующим образом:

$$x_{ij} = \begin{cases} 1, & \text{если цикл включает переезд из города } i \\ & \text{в город } j, \\ 0 & \text{в противном случае.} \end{cases} \qquad \text{для всех } i \text{ и } j$$

Целевая функция записывается в следующем виде:

$$\text{Минимизировать } \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij}x_{ij}, \text{ где } c_{ii} = \infty, \ i = 1, 2, \ldots, n. \tag{19}$$

Переменные должны удовлетворять ограничениям

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots, n \text{ (отъезд)}, \tag{20}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad j = 1, 2, \ldots, n \text{ (прибытие)}, \tag{21}$$

$$x_{ij} \text{ — неотрицательные целые при любом } i \text{ и } j. \tag{22}$$

(Условие $c_{ii} = \infty$ принимается для того, чтобы исключить возможность появления в оптимальном решении значений $x_{ii} = 1$, не имеющих смысла. С другой стороны, можно просто исключить переменные x_{ii} из условий задачи.) Ограничения (20) — (22) обеспечивают равенство каждой переменной x_{ij} либо 0, либо 1. Соотношения (20) требуют, чтобы цикл включал в точности один выезд из каждого города, аналогично соотношения (21) требуют в точности одного прибытия в каждый город. Если бы задача полностью описывалась условиями (19) — (22), то она была бы эквивалентна задаче о назначениях, рассмотренной в разд. 6.4.

К сожалению, хотя переменные x_{ij} в любом цикле должны удовлетворять указанным выше ограничениям, допустимое при этих ограничениях решение не обязательно является циклом. В частности, допустимое решение, удовлетворяющее условиям (20), (21) и (22), может включать два или более несвязанных между собой цикла или **подцикла**, например $x_{12} = x_{23} = x_{31} = 1$ и $x_{45} = x_{56} = \ldots$ $\ldots = x_{n4} = 1$. (Покажите, почему такое допустимое решение *не является* для коммивояжера циклом.) Поэтому необходимо ввести дополнительные ограничения, обеспечивающие получение решения в виде цикла.

Существует много способов задания линейных ограничений на целочисленные переменные, исключающих возможность образования подциклов. Так, например, ограничение $x_{23} + x_{32} \leqslant 1$ исключает возможность образования подцикла между городами 2 и 3, аналогично ограничение $x_{23} + x_{34} + x_{42} \leqslant 2$ исключает возможность образования подцикла между городами 2, 3 и 4. Однако можно применить изящный способ задания линейных ограничений, исключающий возникновение *всех* подциклов.

Введем переменные u_i, $i = 2, 3, \ldots, n$, и наложим на них следующие $(n-1)^2 - (n-1)$ условий:

$$u_i - u_j + nx_{ij} \leqslant n - 1, \quad \begin{array}{l} i = 2, 3, \ldots, n, \\ j = 2, 3, \ldots, n \ (i \neq j). \end{array} \tag{23}$$

Чтобы устранить возникновение подциклов, на переменные u_i не требуется накладывать никаких дополнительных ограничений, однако наложение условия неотрицательности и целочисленности не может принести никакого ущерба. В условиях (20) — (23) содержится $n^2 - n + 2$ линейных ограничения и $n^2 + n - 1$ целочисленных переменных, из которых n переменных x_{ii} должны быть равны нулю. Очевидно, что механизм действия ограничений (23) пока не ясен. Его обоснование приводится ниже.

Для демонстрации эффективности ограничений (23) нужно доказать два утверждения: а) эти ограничения действительно исключают все подциклы; б) ни один полный цикл не исключается. Эти два утверждения рассматриваются последовательно.

Прежде всего ограничения (20) — (23) исключают возможность появления всех подциклов между городами 2, 3, . . ., n. Чтобы убедиться в этом, рассмотрим любой цикл между k из этих городов, который определяется значениями $x_{ij} = 1$ для k переменных. Добавим в (23) k ограничений, соответствующих переменным, образующим подцикл. Такому общему ограничению должно также удовлетворять любое решение, допустимое по условию (23). Читателю следует убедиться в том, что для каждого из k городов в полученной сумме фигурируют величины u_i и $(-u_i)$. Следовательно, в общем ограничении нет ни одной величины u_i. Однако это недопустимо для *всех* $x_{ij} = 1$ в этом ограничении, ибо в таком случае сумма в левой части неравенства, равная nk, будет больше суммы в правой части неравенства, равной $(n - 1)\,k$. Следовательно, условие (23) исключает возможность появления любого подцикла.

Это доказательство легко проиллюстрировать на примере, рассмотрев подцикл между городами 2, 3, 4, 2, которому отвечают следующие значения переменных: $x_{23} = x_{34} = x_{42} = 1$. Соответствующие ограничения в (23) при $k = 3$ имеют вид

$$\begin{aligned} u_2 - u_3 + nx_{23} &\leqslant n - 1,\\ u_3 - u_4 + nx_{34} &\leqslant n - 1,\\ u_4 - u_2 + nx_{42} &\leqslant n - 1. \end{aligned} \tag{I}$$

Сложив эти ограничения, получим общее ограничение

$$(u_2 - u_3 + u_3 - u_4 + u_4 - u_2) + n\,(x_{23} + x_{34} + x_{42}) \leqslant (n - 1)\,3. \tag{II}$$

Убедитесь в том, что ограничение (II) исключает возможность $x_{23} = x_{34} = x_{42} = 1$.

Покажем теперь, что ограничения (20) — (23) не приводят к устранению любого полного цикла. Для этого достаточно установить, что существуют значения u_i, удовлетворяющие ограничению (23), для любого заданного цикла. Рассмотрим такой цикл, который по определению заключается в посещении каждого города 2, 3, . . ., n один и только один раз. Пусть t_i есть положение в таком цикле, когда коммивояжер попадает в город i, причем для города 1 примем $t_1 = 1$. Тогда в цикле город 1 — город 3 — город 5 . . . значения будут равны: $t_1 = 1$, $t_3 = 2$, $t_5 = 3$, ... При этом условии $u_i = t_i$, $i = 2, 3, \ldots, n$, есть допустимое множество значений t_i. Чтобы убедиться в этом, предположим, что в данном цикле $x_{ij} = 1$, так что $t_j = t_i + 1$. Тогда ограничение для этого x_{ij} в (23) удовлетворяется, поскольку

$$t_i - (t_i + 1) + n\,(1) \leqslant n - 1. \tag{III}$$

Предположим теперь, что $x_{ij} = 0$. Тогда соответствующее неравенство в (23) принимает вид $(u_i - u_j \leqslant n - 1)$. Это неравенство должно выполняться, поскольку $u_i < n$ и $u_j > 1$.

Иные методы постановки задач целочисленного программирования. Предположим, что переменная x должна принимать только определенные значения, скажем $X_1, X_2, \ldots, X_p$. Добиться выполнения этого условия можно, исключив переменную x и введя вместо нее соответствующим образом p булевых переменных w_k. Заменим, например, x везде, где эта переменная фигурирует в модели, выражением

$$X_1 w_1 + X_2 w_2 + \ldots + X_p w_p \tag{24}$$

и наложим на переменные w_k следующие ограничения:

$$w_1 + w_2 + \ldots + w_p = 1, \tag{25}$$

$$\text{все } w_k \text{ — целые.} \tag{26}$$

Далее, если в модели имеется также нелинейная функция $c(x)$, то, используя (24) — (26), можно заменить функцию $c(x)$ выражением

$$c(X_1)\, w_1 + c(X_2)\, w_2 + \ldots + c(X_p)\, w_p, \tag{27}$$

линейным по w_k.

Описанный подход можно обобщить, распространив его на модели, содержащие непрерывные кусочно-линейные функции $c(x)$ неотрицательной переменной x. Модели такого вида рассматриваются более подробно в разд. 14.7, так что читатель может отложить разбор приводимого ниже пояснения, пока не проработает материал следующей главы.

Предположим для простоты, что значения переменной x изменяются на интервале $0 \leqslant x \leqslant X$ и что значения $0 \equiv X_1 < X_2 < \ldots < X_p \equiv X$ являются конечными точками линейных отрезков, т. е. функции $c(x)$ линейны при $X_k \leqslant x \leqslant X_{k+1}$, где $k = 1, 2, \ldots, p-1$. Введем прежде всего p неотрицательных весовых коэффициентов w_k, заменим переменную x в модели выражением (24) и наложим условие (25), где все $w_k \geqslant 0$ (в отличие от (26)), и заменим $c(x)$ выражением (27).

Читателю следует убедиться в том, что эти замены допустимы при условии, что решение содержит только одну переменную $w_k = 1$ или не более двух *соседних* переменных w_k и w_{k+1}, где $w_k + w_{k+1} = 1$. Для того чтобы выполнялось это свойство, получившее название *свойства соседних весов*, введем $(p-1)$ булевых переменных z_k и наложим $(p+1)$ ограничений

$$\begin{array}{llll}
w_1 - z_1 & & & \leqslant 0, \\
w_2 - z_1 - z_2 & & & \leqslant 0, \\
w_3 & - z_2 - z_3 & & \leqslant 0, \\
\cdot & & & \cdot \\
\cdot & & & \cdot \\
\cdot & & & \cdot \\
w_{p-1} & & - z_{p-2} - z_{p-1} & \leqslant 0, \\
w_p & & - z_{p-1} & \leqslant 0,
\end{array} \tag{28}$$

$$z_1 + z_2 + z_3 + \ldots + z_{p-2} + z_{p-1} = 1, \tag{29}$$

а также ограничения

$$\text{все } z_k \text{ — целочисленные.} \qquad (30)$$

Читателю рекомендуется проверить, что ограничения (29) и (30) обеспечивают равенство единице в точности одной переменной z_k, гарантируя тем самым, что в (28) не более двух весов w_k и w_{k+1} могут быть положительными, причем их сумма в соответствии с (25) должна быть равна 1.

Модификации изложенного метода позволяют рассматривать кусочно-линейные функции, имеющие разрывы вида скачка в точках X_k.

13.3. ОБЩИЕ СВЕДЕНИЯ О МЕТОДАХ РЕШЕНИЯ ЗАДАЧ ЦЕЛОЧИСЛЕННОГО ПРОГРАММИРОВАНИЯ

В настоящее время известны два основных подхода к отысканию точного оптимального решения задач целочисленного программирования.

1. *Методы отсекающих плоскостей* (*методы отсечения*). Предложено несколько вариантов этого подхода к решению целочисленных задач. Один из них, излагаемый в следующем разделе, предназначен для полностью целочисленной модели. Исходным моментом является оптимальное решение соответствующей задачи линейного программирования, полученной в результате отбрасывания условий целочисленности. На каждой итерации добавляется линейное ограничение, удовлетворяющее целочисленному решению исходной задачи, но исключающее текущее нецелочисленное решение. Вычислительный процесс прекращается, как только будет достигнуто любое целочисленное решение. Сходимость обеспечивается за конечное, но иногда очень большое число итераций.

2. *Методы возврата*. В этой группе методов также имеются различные модификации. Первый метод, названный методом «ветвей и границ», изложен в разд. 13.5 и предназначен для решения частично целочисленных задач. Как и в методе отсечения, решение задачи начинается с отыскания оптимального решения соответствующей регулярной задачи линейного программирования. Затем формируется *семейство* связанных, но различных задач линейного программирования. В разд. 13.6 на примере задачи коммивояжера рассматривается несколько вариантов этого подхода. Наконец в разд. 13.7 излагается алгоритм (неявного) частичного перебора, применяющийся здесь только к задачам, содержащим булевы целочисленные переменные. Благодаря такой особой структуре задачи выислительные процедуры существенно упрощаются. Термин «возврат» определяет специфический способ формирования и решения последовательности задач. Смысл этого термина станет ясным после прочтения разд. 13.5.

Существуют другие подходы к решению целочисленных задач, которые хотя и не гарантируют отыскания оптимального решения, тем не менее в ряде частных случаев позволяют его находить, а нередко дают решения, близкие к оптимальному.

Один из таких подходов заключается в использовании случайной выборки допустимых решений с последующим улучшением каждого попавшего в выборку решения в случае, когда возможность улучшения соответствующего решения удается достаточно просто обнаружить. Если, к примеру, применить этот подход при решении задачи коммивояжера с n городами, то нужно взять случайную выборку перестановок городов 2, 3, . . ., n из общего числа $(n - 1)!$ перестановок, а затем сравнить длину соответствующих циклов. Другой подход, получивший название *эвристического программирования*, сводится к применению как простых, так и достаточно сложных правил, основанных на некоторых рациональных соображениях, а также метода проб и ошибок для отыскания допустимого и по возможности близкого к оптимальному решения. Для задачи коммивояжера простейшим правилом такого рода является переезд из города j в ближайший город, который еще не посещался. Метод проб и ошибок в этом случае в известной мере сводится к оценке эффекта поочередной замены одного города другим в последовательности объезда городов в пробном цикле с точки зрения уменьшения общего расстояния (общей длины цикла).

В ряде практических ситуаций такие методы отыскания решений оказались достаточно эффективными. Однако, поскольку сведения об этих методах не приведены в стройную систему и они часто применяются сообразно с конкретной ситуацией, в дальнейшем мы их не будем рассматривать.

Еще несколько замечаний. В соответствии с общим замыслом учебника, в котором раскрываются основные идеи решения операционных задач, вычислительные схемы решения задач целочисленного программирования излагаются в упрощенном виде. Технические детали, имеющие существенное значение для разработки эффективных программ для ЭВМ, опущены. Эти аспекты, играющие решающую роль для программирования ЭВМ, представляют интерес прежде всего для специалистов соответствующего профиля, и поэтому в данной главе не рассматриваются. Однако один из моментов, связанных с построением вычислительных схем, заслуживает того, чтобы остановиться на нем особо.

Как уже известно читателю, чтобы можно было пользоваться алгоритмом во всех случаях, когда он годится для решения задачи, требуется его однозначное описание. Это объясняется тем, что на практике вычисления производит ЭВМ, программа которой должна быть точно задана. Разумеется, это требование относится ко всем алгоритмам, уже изложенным ранее. Тем не менее именно здесь следует вновь обратить внимание на это обстоятельство, ибо описание большинства алгоритмов решения задач целочисленного программирования включает шаги, допускающие возможность произвольного выбора на каждой итерации.

Любопытно, что, применяя эти алгоритмы к решению задач небольшой размерности, часто удается использовать специальные свойства модели и сделать обоснованный выбор, интуитивно ведущий к получению оптимального решения. В результате большинство алгоритмов в случае их применения к решению таких «игрушечных» задач выглядят весьма эффективными. Однако при программировании этих алгоритмов для ЭВМ выбор осуществляется во всех случаях согласно заранее заданным правилам без вмешательства интуиции человека. Опыт показал, что объем вычислений, связанных с реализацией алгоритмов, существенно зависит от метода решения задачи произвольного выбора, который в принципе отличается от метода реализации симплексного алгоритма. При программировании этого алгоритма различные варианты *критерия I* (выбора переменных, вводимых в неоптимальное базисное решение) приводят к примерно одинаковому объему вычислений. При изучении каждого из изложенных ниже алгоритмов следует обратить особое внимание на шаги, допускающие возможность произвольного выбора.

13.4. АЛГОРИТМЫ ОТСЕЧЕНИЯ (МЕТОД ЦЕЛОЧИСЛЕННЫХ ФОРМ)

Нижеследующий элементарный пример поясняет основную идею алгоритма. Рассмотрим задачу

$$\text{Максимизировать } 5x_1 + 1x_2 \tag{1}$$

при ограничениях

$$2x_1 + 1x_2 = 3, \tag{2}$$

$$x_1 \text{ и } x_2 \text{ — неотрицательные целые.} \tag{3}$$

Как легко видеть, $x_1 = {}^3/_2$ есть оптимальное решение задачи линейного программирования (1) и (2) в случае, когда условие (3) ослаблено и имеет вид $x_1 \geqslant 0$, $x_2 \geqslant 0$. Решение этой задачи недопустимо по условию (3), поскольку оно является дробным. Заметим, однако, что из (2) и (3) следует ограничение в виде линейного неравенства

$$x_1 \leqslant 1. \tag{4}$$

Читателю следует объяснить, откуда вытекает это следствие, а также показать, что оптимальное решение задачи линейного программирования (1), (2) и (4) при $x_1 \geqslant 0$ и $x_2 \geqslant 0$ равно $x_1 = 1$ и $x_2 = 1$. Подчеркнем, что это решение *целочисленно*. Следовательно, полученное решение задачи линейного программироваиия является оптимальным для задачи (1) — (3), поскольку оно допустимо по условию (3).

По существу, оптимальное решение задачи целочисленного программирования было найдено путем построения и решения расширенной задачи *линейного* программирования, содержащей линейные ограничения, вытекающие из условий исходной целочисленной задачи. В данном случае удалось чрезвычайно просто обнаружить ограничение (4), позволившее реализовать этот метод на элементарном примере. Излагаемый ниже алгоритм обеспечивает систематический поиск таких ограничений для *любой* задачи целочисленного программирования. Следовательно, основная идея построения такого алгоритма состоит в замене целочисленной задачи соответствующим образом расширенной задачей линейного программирования.

Заметим, что в рассмотренном примере оптимальное решение целочисленной задачи содержит обе переменные с положительным знаком, хотя имеется лишь одно линейное ограничение (2). Это обстоятельство подчеркивает особенность, отличающую модели линейного и целочисленного программирования. Для задачи линейного программирования, включающей m ограничений и n переменных, всегда существует оптимальное решение, содержащее не более m строго положительных переменных.

В оптимальном решении целочисленной задачи такой же размерности может оказаться необходимым, чтобы все целочисленные переменные были строго положительными. Именно в силу этого требования, применяя метод отсечения, нужно увеличивать размерность исходной задачи, добавляя дополнительные ограничения.

Предложено несколько вариантов алгоритмов, основанных на методе отсечения. Здесь излагается вариант, используемый для решения полностью целочисленных задач, в которых все переменные принимают только целые значения. Алгоритм можно легко обобщить на частично целочисленные задачи. Достоинством выбранного варианта является простота его описания. Другие варианты, более сложные, но в то же время и более эффективные, представляют интерес для узких специалистов. От рассмотренного варианта они отличаются тем, что построены на основе результатов дальнейшего развития идей, излагаемых ниже.

Сущность алгоритмов, основанных на методе отсечения, легко уяснить, обратившись к геометрическим представлениям в пространстве решений (разд. 3.4). Определим **выпуклую оболочку** множества допустимых целочисленных точек (решений) как минимальное выпуклое множество, содержащее все эти точки (или, что то же самое, как множество точек евклидового n-мерного пространства, состоящего из всех **выпуклых комбинаций** вида $[wx_1 + (1-w)\,y_1,\ wx_2 + (1-w)\,y_2,\ \ldots,\ wx_n + (1-w)\,y_n]$, где $0 \leqslant w \leqslant 1$ и $(x_1, x_2, \ldots, x_n)$ и $(y_1, y_2, \ldots, y_n)]$ — допустимые целочисленные решения). Предположим, что оптимальное значение линейной целевой функции является конечным. Тогда легко показать, что целевая функция достигает оптимального значения в одной из вершин этой выпуклой оболочки. Такая экстремальная точка представляет собой одно из допустимых целочисленных решений.

В свою очередь выпуклую оболочку можно представить конечным множеством линейных ограничений. В алгоритме отсечения вначале рассматривается выпуклое множество, определенное линейными ограничениями и условиями неотрицательности переменных исходной задачи, и отыскивается экстремальная точка этого множества. Если такое решение оказывается нецелочисленным, то добавляют ограничение, отсекающее текущую экстремальную точку и уменьшающее «объем» выпуклого множества. Однако новое ограничение не отсекает ни одной экстремальной точки выпуклой оболочки, принадлежащей допустимым целочисленным решениям. В конце концов вводится такое число дополнительных ограничений, что экстремальная точка усеченного выпуклого множества совпадает с экстремальной точкой выпуклой оболочки. Геометрическая интерпретация работы алгоритма будет пояснена на примере, приведенном в этом разделе.

Подробное изложение алгоритма. Допустим, что полностью целочисленная задача имеет следующий вид:

$$\text{Максимизировать} \quad \sum_{j=1}^{n} c_j x_j \tag{5}$$

при ограничениях

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = 1, 2, \ldots, m, \tag{6}$$

$$\text{все } x_j \text{ — неотрицательные целые.} \tag{7}$$

Предположим, что при ограничениях (6) и (7) существует допустимое решение и что в оптимальном решении целевая функция принимает конечное значение. Прежде чем сформулировать конкретные шаги алгоритма, покажем в общем виде, как построить дополнительные линейные ограничения, которым должно удовлетворять любое решение, допустимое по условиям (6) и (7).

Рассмотрим любое линейное уравнение, которое можно получить с помощью алгебраических операций над линейными ограничениями (6), например путем сложения двух или более уравнений или умножения уравнения на отличную от нуля постоянную. Тогда, если некоторое решение удовлетворяет условию (6), оно должно также удовлетворять и новому уравнению. Примем, что выражение

$$\sum_{j=1}^{n} a_j x_j = b \tag{8}$$

определяет именно такое новое ограничение. Возможно, одна или несколько величин a_j и b являются дробными, скажем 2,5 или $^{10}/_3$. Обозначим символом $[d]$ **целую часть** d, т. е. наибольшее целое число, меньшее или равное действительному числу d. Например,

$$[2{,}5]=2, \qquad \left[\frac{10}{3}\right]=3, \qquad [4]=4,$$
$$[-2{,}5]=-3, \qquad \left[-\frac{10}{3}\right]=-4, \qquad [-4]=-4. \tag{9}$$

Поскольку на величины x_j наложено ограничение (7), откуда следует, что они должны быть неотрицательными целыми, любые значения x_j, удовлетворяющие ограничению (6), а следовательно, и (8), должны, очевидно, удовлетворять более слабому ограничению

$$\sum_{j=1}^{n}[a_j]\,x_j \leqslant b. \tag{10}$$

(Читателю следует доказать справедливость этого утверждения.) Далее, поскольку сумма в левой части (10) должна быть целочисленной, нужно доказать, что (10) можно усилить следующим образом:

$$\sum_{j=1}^{n}[a_j]\,x_j \leqslant [b]. \tag{11}$$

Наконец, (11) можно преобразовать в равенство, добавив свободную переменную

$$\sum_{j=1}^{n}[a_j]\,x_j + x = [b] \quad \text{(целые части)}, \tag{12}$$

т. е. на x наложено условие неотрицательности и целочисленности [Объясните, почему из (11) следует, что величина x должна быть целой.] Таким образом, если к условию (6) добавить линейное ограничение (12), задача все равно останется полностью целочисленной. В результате получено ограничение (12), в котором величины a_j и b получены с помощью алгебраических операций над ограничениями (6) и которому удовлетворяет любое допустимое по условиям (6) и (7) решение.

Для упрощения словесного описания алгоритма выполним элементарные преобразования над (8) и (12). Определим значения f_j и f тождествами

$$[a_j] + f_j \equiv a_j \quad \text{и} \quad [b] + f \equiv b, \tag{13}$$

такими, что $0 \leqslant f_j < 1$ и $0 \leqslant f < 1$. Эти величины называются **дробными частями** a_j и b. (Для примера дробная часть величин 2,5 и —2,5 равна 0,5, величины $^{10}/_3$ равна $^1/_3$ и величины $-^{10}/_3$ равна $^2/_3$.) Вычитая (8) из (12), получим ограничение, состоящее только из дробных частей

$$\sum_{j=1}^{n}(-f_j)\,x_j + x = -f \quad \text{(дробные части)}. \tag{14}$$

В приводимом ниже алгоритме вместо (12) используется ограничение (14).

Рассмотрим пример, показывающий, как работают новые ограничения (12) и (14). В приведенном выше примере умножим обе части (2) на $^1/_2$, в результате чего получим

$$1x_1+\frac{1}{2}x_2=\frac{3}{2}. \tag{15}$$

Теперь нужно показать, что применение к этому выражению формулы (12) дает

$$1x_1+0x_2+x=1, \tag{16}$$

что эквивалентно ограничению (4). Применение формулы (14) дает

$$-\frac{1}{2}x_2+x=-\frac{1}{2}, \tag{17}$$

откуда следует, что $x_2 \geqslant 1$.

Алгоритм отсечения состоит из следующих шагов.

Шаг 1. Найти оптимальное решение задачи линейного программирования с целевой функцией (5) и линейными ограничениями (6), не учитывая условия целочисленности (7), но требуя, чтобы все $x_j \geqslant 0$.

Шаг 2. Прекратить вычисления, если текущее решение задачи линейного программирования является целочисленным. В противном случае выбрать какую-нибудь дробную базисную переменную. Составить ограничение (14) из уравнения, содержащего эту базисную переменную в текущем оптимальном решении задачи линейного программирования.

Шаг 3. Добавить к исходной задаче линейного программирования это новое ограничение, найти новое оптимальное решение задачи с дополнительным ограничением и вернуться к шагу 2.

Прежде чем проиллюстрировать работу алгоритма на числовом примере, рассмотрим более подробно перечисленные шаги. Предположим, что на некоторой итерации текущее оптимальное решение задачи линейного программирования получено с помощью симплексного алгоритма, описанного в разд. 4.4. Допустим, что x_k есть базисная переменная в строке i на последней итерации:

$$x_k+\sum_{\substack{j \text{ — неба-} \\ \text{зисные}}} t_{ij}x_j=b \quad \text{(строка } i\text{)}, \tag{18}$$

где b имеет дробное значение. Следовательно, $x_k = b$ не удовлетворяет условию целочисленности. Тогда на шаге 2 ограничение (14), образованное из (18), записывается в такой форме:

$$\sum_{\substack{j \text{ — неба-} \\ \text{зисные}}} (-f_{ij})x_j+x=-f \quad \text{(отсечение)}, \tag{19}$$

где f_{ij} — дробная часть t_{ij}, а f — дробная часть b. Заметим, что после первой итерации сумма небазисных переменных в (19) может включать свободные переменные, обозначенные в (14) символом x, которые были введены на шаге 2 предыдущей итерации. Подобное включение вполне правомерно, ибо на каждую такую свободную переменную наложено условие целочисленности.

Заметим также, что текущее решение задачи линейного программирования не удовлетворяет новому ограничению (19) в том смысле, что значение $x = -f$ строго отрицательно, поскольку сумма в левой части, содержащая только небазисные переменные, в текущем решении равна нулю. Когда на шаге 3 добавляется так называемое отсечение (19) наряду с условием $x \geqslant 0$, текущее решение задачи линейного программирования становится недопустимым или, выражаясь геометрическим языком, отсекается. Это означает, что после оптимизации на шаге 3 значение целевой функции не может возрасти, а может лишь уменьшиться. На практике оптимизация обычно осуществляется с помощью двойственного симплексного алгоритма (разд. 5.8), который применяется к модели, расширенной за счет условия (19), после добавления величины x к текущему набору базисных переменных.

Выбор дробной базисной переменной на шаге 2 является достаточно произвольным. Однако при построении строгого доказательства сходимости алгоритма за конечное число итераций нельзя считать этот выбор полностью произвольным. С практической точки зрения сходимость часто повышается при выборе переменной с наибольшей дробной частью (т. е. переменной, дающей максимальное значение f). Можно также сформировать несколько подобных ограничений для различных базисных переменных с дробной частью и ввести их на шаге 3, однако реальные данные не подтверждают, что этот путь дает какое-либо сокращение объема вычислений.

Если ввести некоторую модификацию в процедуру, предусмотренную шагом 3, то можно обеспечить ограничение бесконечного роста размерности расширенной задачи линейного программирования. Эта модификация заключается в том, что во всех случаях, когда текущее оптимальное базисное решение включает свободную переменную x, введенную на предыдущей итерации, из текущего решения, еще до того, как производится оптимизация, исключается уравнение, содержащее x. Из этой модификации следует, что размерность расширенной задачи никогда не превышает числа n линейных уравнений. Как только в задаче оказывается n уравнений, введение любого нового ограничения должно приводить к тому, что в базисном решении будет содержаться одна или более свободных переменных, определяемых условием (14). Важным следствием, вытекающим из всех приведенных соображений, является то, что вычислительные трудности, связанные с решением задачи целочисленного программирования, в значительной степени определяются числом переменных, входящих в задачу.

Теоретически доказано, что алгоритм сходится к оптимуму за конечное число итераций. В конце этого раздела обсуждаются практические аспекты реализации вычислительной схемы алгоритма. Однако отметим сразу, что алгоритм не гарантирует получения допустимого целочисленного решения до самой последней итерации и в этом смысле относится к категории двойственных. Вследствие этого при преждевременном прекращении вычислений нельзя получить допустимого решения. (Существуют алгоритмы, базирующиеся на методе отсечения, которые относятся к разряду прямых в том смысле, что на каждой итерации получаемое решение является допустимым и целочисленным, но эти алгоритмы в данной книге не рассматриваются, хотя упражнение 69 дает общее представление о сущности заложенных в них идей.)

Пример. Рассмотрим задачу

$$\text{Максимизировать } 21x_1 + 11x_2 \tag{20}$$

при ограничениях

$$7x_1 + 4x_2 + x_3 = 13, \tag{21}$$

$$x_1,\ x_2,\ x_3 \text{ — неотрицательные целые.} \tag{22}$$

(Величину x_3 можно рассматривать как дополняющую переменную, поскольку она не входит в целевую функцию.) В соответствии с принятой схемой выразим целевую функцию линейным уравнением

$$x_0 - 21x_1 - 11x_2 = 0, \tag{23}$$

учитывая, что в этом примере на величину x_0 можно наложить условие целочисленности. Прежде чем двигаться дальше, читателю следует найти оптимальное решение.

Нужно также показать, что, выполнив шаг 1, получаем следующее решение задачи линейного программирования:

$$\begin{aligned} &x_0 \quad + x_2 \quad + 3x_3 = 39 \quad \text{(строка 0)},\\ &x_1 + \frac{4}{7}x_2 + \frac{1}{7}x_3 = 1\frac{6}{7} \quad \text{(строка 1)}. \end{aligned} \tag{24}$$

Поскольку согласно шагу 2 значение x_1 является дробным, нужно добавить ограничение (14), образованное из строки 1. Покажите, что это ограничение выражается следующим образом:

$$-\frac{4}{7}x_2 - \frac{1}{7}x_3 + x_4 = -\frac{6}{7} \quad \text{(строка 2)}, \tag{25}$$

где x_4 — целочисленная дополняющая переменная.

Переходя к шагу 3, получаем задачу линейного программирования (24), дополненную (25). Покажите, что новое оптимальное решение выражается следующими уравнениями:

$$x_0 \quad + 2\frac{3}{4}x_3 + 1\frac{3}{4}x_4 = 37\frac{1}{2} \quad \text{(строка 0)},$$

$$\begin{aligned} &x_1 && + \quad x_4 = 1 \quad &&\text{(строка 1)}, \\ &x_2 + \frac{1}{4}x_3 && -1\frac{3}{4}x_4 = 1\frac{1}{2} \quad &&\text{(строка 2)}. \end{aligned} \tag{26}$$

Возвращаясь к шагу 2, убедитесь в том, что новое ограничение, образованное из строки 2, имеет вид

$$-\frac{1}{4}x_3 - \frac{1}{4}x_4 + x_5 = -\frac{1}{2} \qquad \text{(строка 3)} \tag{27}$$

и что на шаге 3 оптимальное решение задачи линейного программирования (26) — (27) описывается уравнениями:

$$\begin{aligned} x_0 + x_1 & & + 11x_5 &= 33 \quad \text{(строка 0)}, \\ -x_1 & + x_3 & - \quad x_5 &= 1 \quad \text{(строка 1)}, \\ 2x_1 + x_2 & & + \quad x_5 &= 3 \quad \text{(строка 2)}, \\ x_1 & + x_4 & &= 1 \quad \text{(строка 3)}. \end{aligned} \tag{28}$$

При следующем возвращении к шагу 2 итерации заканчиваются оптимальным целочисленным решением $x_1 = 0$, $x_2 = 3$ и $x_3 = 1$, что дает в результате $x_0 = 33$. [Если бы решение (28) оказалось не целочисленным, пришлось бы до выполнения следующего цикла оптимизации исключить соотношение из строки 3, так как величина x_4 вновь вошла в базис.]

Можно глубже уяснить излагаемый алгоритм, рассмотрев, каким образом дополнительные ограничения (25) и (27) можно выразить в виде неравенств через переменные исходной задачи x_1 и x_2. При этом легко представить эти неравенства графически, используя геометрическую *интерпретацию пространства решений* (разд. 3.4), что позволит наглядно показать, как они отсекают соответствующее решение задачи линейного программирования.

Ограничение (21) можно записать в виде неравенства

$$7x_1 + 4x_2 \leqslant 13, \tag{I}$$

опустив дополнительную переменную x_3. Исключая x_3 из (25), пользуясь (21) и опуская дополнительную переменную x_4 из полученного результата, имеем неравенство

$$x_1 \leqslant 1. \tag{II}$$

Исключая x_3 из (27) с помощью (21), затем исключая x_4 из полученного результата с помощью (25) и опуская дополнительную переменную x_5, имеем неравенство

$$2x_1 + x_2 \leqslant 3. \tag{III}$$

Покажите, что оптимальное неотрицательное решение задачи линейного программирования (20), (I), (II), (III) совпадает с решением (28).

Прибегая к геометрическому представлению пространства решений, нужно начинать с определения множества допустимых целочисленных решений, вытекающих из (I), а также с выделения области допустимых решений задачи линейного программирования (покажите, что $x_1 = 1^6/_7$ есть оптимальное решение задачи линейного программирования). Добавив, далее, условие (II), заметим, что это ограничение не исключает ни одного допустимого целочисленного решения, но отсекает текущее решение задачи линейного программирования. [Покажите, что решение $x_1 = 1$ и $x_3 = 1^1/_2$ является оптимальным при добавлении условия (II).]

Вводя, наконец, условие (III), заметим вновь, что это новое ограничение не исключает ни одного допустимого целочисленного решения, отсекая в то же время текущее решение задачи линейного программирования, и дает новое оптимальное решение, которое теперь является целочисленным.

Заключительные замечания. Практический опыт реализации вычислительных схем сложных вариантов алгоритмов, построенных на методе отсечения, ни в коей мере нельзя признать успешным во всех случаях. В частности, эти алгоритмы оказались мало эффективными при решении комбинаторных задач, представленных в целочисленной постановке. Возникли также трудности при решении задач, в которых величины a_{ij} и b_i имеют большие значения. Однако алгоритмы этого класса хорошо зарекомендовали себя при решении задач, в которых оптимальное решение задачи линейного программирования уже содержит целочисленные значения многих переменных.

Наибольшие вычислительные трудности возникают при отыскании оптимального решения на шаге 3. После выполнения нескольких итераций на шаге 3 могут появиться многочисленные альтернативные оптимальные решения задачи линейного программирования. Вследствие этого целевая функция остается постоянной (обычно принимая неоптимальное значение) в течение многих итераций, пока альтернативные решения постепенно не будут исключены в силу ограничений, формируемых на шаге 2. (Такое зацикливание иногда называют **сплошной вырожденностью**.) К сожалению, это явление часто возникает в задачах средней и большой размерности. Имеется также много примеров задач малой размерности (не более 10 переменных и уравнений), при решении которых для достижения сходимости потребуются тысячи итераций.

Число итераций, необходимых для получения решения, по-видимому, существенно зависит также от конкретного вида постановки задачи. В качестве примера такой чувствительности решения, получившей название чувствительности к форме постановки, можно указать на случай, когда введение, по всей видимости, избыточных ограничений может значительно повысить скорость сходимости (допустим, ограничения $x_1 \leqslant 1$ и $x_2 \leqslant 1$ для задачи с ограничением $x_1 + x_2 \leqslant 1$).

Мы упустили один момент, на котором стоит особо остановиться. Пожалуй, может вызвать удивление столь простой вывод ограничения (12). Если же задуматься относительно формирования аналогичных отсекающих ограничений, то легко обнаружить существование многочисленных возможностей систематического вывода таких ограничений. Однако отсечение (12) отличается от множества других отсечений (ограничений) тем, что можно строго доказать, что результирующий алгоритм обеспечивает сходимость к оптимальному решению за конечное число итераций. (Можно показать, что некоторые отличные от рассмотренного варианта методы отсечения не обладают свойствами сходимости при их приложении к определенным классам задач.) Вследствие того что вариант метода отсечения, в котором используется ограничение (12), оказался эффективным с вычислительной точки зрения лишь в отдельных частных случаях, доказанное теоретически свойство сходимости этого алгоритма, разумеется, не может служить оправданием для его универсального применения. Впрочем, и другие варианты не могут претендовать на более высокую эффективность даже в тех случаях, когда установлена сходимость.

13.5. МЕТОД ВЕТВЕЙ И ГРАНИЦ

Этот метод можно применять для решения как полностью, так и частично целочисленных задач дискретного программирования. Предположим для определенности, что рассматриваемая модель имеет следующий вид:

$$\text{Максимизировать } \sum_{j=1}^{n} c_j x_j \tag{1}$$

при ограничениях

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

$$x_j \text{ — целые}, \quad j = 1,\ 2,\ \ldots,\ p\ (\leqslant n), \tag{3}$$

$$x_j \geqslant 0, \qquad j = p+1,\ \ldots,\ n. \tag{4}$$

Кроме того, допустим, что для каждой целочисленной переменной можно задать верхнюю и нижнюю границы, в пределах которых безусловно содержатся ее оптимальные значения

$$L_j \leqslant x_j \leqslant U_j, \quad j = 1,\ 2,\ \ldots,\ p. \tag{5}$$

Обычно $L_j = 0$, но это условие не обязательно. (Как было показано в разд. 3.3, предположение, что все $L_j = 0$, не приводит к потере общности. Однако, пользуясь общим символом L_j, алгоритм проще описать.)

Идея метода ветвей и границ основана на следующем элементарном факте. Рассмотрим любую переменную x_j и примем, что I есть некоторое целое число, где $L_j \leqslant I \leqslant U_j - 1$. Тогда оптимальное решение задачи (1) — (5) будет удовлетворять либо линейному ограничению

$$x_j \geqslant I + 1, \tag{6}$$

либо линейному ограничению

$$x_j \leqslant I. \tag{7}$$

Для иллюстрации возможности использования этой дихотомии предположим, что условие целочисленности (3) не учитывается, и найдем оптимальное решение задачи линейного программирования (1), (2), (4) и (5), в котором, к примеру, $x_1 = 1^2/_3$. Далее поставим и решим еще две задачи линейного программирования, каждая из которых содержит условия (1), (2) и (4). Однако в одной из этих задач условие (5) при $j = 1$ записывается в виде $2 \leqslant x_1 \leqslant U_1$, а в другой — в виде $L_1 \leqslant x_1 \leqslant 1$. Предположим, далее, что каждая из этих задач имеет оптимальное решение, удовлетворяющее условию целочисленности (3). Тогда решение, доставляющее большее значение целевой функции, на самом деле является оптимальным решением исходной целочисленной задачи. (Покажите справедливость этого утверждения.) Обычно одна (или обе) из таких задач не имеет оптимального решения, удовлетворяющего (3), вследствие чего могут потребоваться дополнительные вычисления. Излагаемый ниже алгоритм определяет систематическую схему применения (6) и (7), обеспечивающую в конечном счете получение оптимального решения.

Некоторые пояснения. Ниже дается формальное описание метода, а его вычислительные аспекты обсуждаются в конце раздела. Метод ветвей и границ основан на решении некоторого *множества* задач линейного программирования. Как было показано в предыдущем разделе, оптимальное решение задачи целочисленного программирования не обязательно должно быть базисным решением, удовлетворяющим m линейным ограничениям (2). Иначе говоря, оптимальное решение может содержать более m переменных, имеющих строго положительное значение. Границы (5) на каждую переменную x_j служат для расширения задач линейного программирования в такой степени, чтобы в случае необходимости в оптимальное решение входили все переменные. Такое расширение является свидетельством того, что трудоемкость вычислений определяется прежде всего числом целочисленных переменных, содержащихся в задаче.

Алгоритм. На каждой итерации (обозначим ее t) имеется нижняя граница (оценка), скажем x_0^t, оптимального значения целевой функции. Для упрощения изложения предположим, что на первой итерации значение x_0^1 либо *строго* меньше оптимального значения, либо равно значению целевой функции, соответствующему фиксированному допустимому решению. При отсутствии какой-либо дополнительной информации об особенностях рассматриваемой задачи в самом неблагоприятном случае можно принять $x_0^1 = -\infty$. Помимо нижней границы имеется основной список задач линейного программирования, подлежащих решению. Единственное отличие этих задач друг от друга связано с изменением условий (5). На итерации 1 основной список содержит всего одну задачу (1), (2), (4) и (5).

На произвольной итерации t выполняются следующие шаги.

Шаг 1. Прекратить вычисления, если основной список пуст. В противном случае выбрать одну задачу линейного программирования из основного списка.

Шаг 2. Решить выбранную задачу. Если она не имеет допустимого решения или если полученное в результате оптимальное значение целевой функции меньше или равно x_0^t, то принять $x_0^{t+1} = x_0^t$ и вернуться к шагу 1. В противном случае перейти к шагу 3.

Шаг 3. Если полученное оптимальное решение задачи линейного программирования удовлетворяет целочисленным ограничениям, то зафиксировать его, принять x_0^{t+1} равным соответствующему оптимальному значению целевой функции и вернуться к шагу 1. В противном случае перейти к шагу 4.

Шаг 4. Выбрать любую переменную x_j, $j = 1, 2, \ldots, p$, не имеющую целого значения в полученном оптимальном решении выбранной задачи линейного программирования. Пусть b_j соответствует этому значению, а $[b_j]$ определяет наибольшее целое число, меньшее или равное b_j. Добавить две задачи линейного программирования в основной список. Эти две задачи идентичны задаче, выбранной на шаге 1, за исключением того, что в одной из них нижняя граница x_j заменена на $[b_j] + 1$, а в другой верхняя граница x_j заменена на $[b_j]$. Принять $x_0^{t+1} = x_0^t$ и вернуться к шагу 1.

При остановке алгоритма в случае, когда фиксируется допустимое решение, дающее x_0^t, это решение оптимально, в противном случае допустимого решения не существует. Как показывает приводимый ниже пример, можно получить целочисленное решение, не дойдя до последней итерации, однако при этом не известно, является ли оно действительно оптимальным. По этой причине алгоритм можно назвать *приближенно двойственным.*

Пример. Рассматриваемый пример поясняет подробности изложенного алгоритма. Возьмем задачу

$$\text{Максимизировать } 3x_1 + 3x_2 + 13x_3 \tag{8}$$

при ограничениях
$$-3x_1 + 6x_2 + 7x_3 \leqslant 8,$$
$$6x_1 - 3x_2 + 7x_3 \leqslant 8, \tag{9}$$

где каждая переменная x_j должна быть неотрицательным целым. Предположим, что заданы следующие границы на каждую переменную:

$$0 \leqslant x_j \leqslant 5, \quad j = 1, 2, 3. \tag{10}$$

Как принято, обозначим x_0^t значение целевой функции. (Найдите оптимальное решение простым подбором.)

На итерации 1 примем нижнюю границу $x_0^1 = 0$, поскольку при всех $x_j = 0$ имеем допустимое решение. Основной список содержит лишь одну задачу линейного программирования (8), (9) и (10), которую назовем задачей 1. Выберем эту задачу на шаге 1 и найдем ее оптимальное решение на шаге 2:

$$x_0 = 16, \quad x_1 = x_2 = 2\,\frac{2}{3}, \quad x_3 = 0 \quad \text{(задача 1).} \tag{11}$$

Поскольку решение не является целочисленным, перейдем от шага 3 к шагу 4 и выберем x_1. Внесем, далее, в основной список

$$\begin{gathered}\text{задачу 2: ограничения (9)}\\ 3 \leqslant x_1 \leqslant 5, \quad 0 \leqslant x_2 \leqslant 5, \quad 0 \leqslant x_3 \leqslant 5,\\ \text{задачу 3: ограничения (9)}\\ 0 \leqslant x_1 \leqslant 2, \quad 0 \leqslant x_2 \leqslant 5, \quad 0 \leqslant x_3 \leqslant 5.\end{gathered} \tag{12}$$

Возвращаясь к шагу 1 при $x_0^2 = x_0^1 = 0$, выберем задачу 2. Читатель может убедиться в том, что на шаге 2 обнаруживается отсутствие допустимого решения задачи 2. Поэтому примем $x_0^3 = x_0^2 = 0$ и вернемся к шагу 1.

Выберем теперь задачу 3 и получим на шаге 2 оптимальное решение

$$x_0 = 15\,\frac{5}{7}, \quad x_1 = x_2 = 2, \quad x_3 = \frac{2}{7} \qquad \text{(задача 3),} \tag{13}$$

не являющееся целочисленным. Поэтому перейдем от шага 3 к шагу 4, где выберем x_3 и внесем в основной список

$$\begin{gathered}\text{задачу 4: ограничения (9)}\\ 0 \leqslant x_1 \leqslant 2, \quad 0 \leqslant x_2 \leqslant 5, \quad 1 \leqslant x_3 \leqslant 5,\\ \text{задачу 5: ограничения (9)}\\ 0 \leqslant x_1 \leqslant 2, \quad 0 \leqslant x_2 \leqslant 5, \quad 0 \leqslant x_3 \leqslant 0.\end{gathered} \tag{14}$$

Заметим, что задачи 4 и 5 отличаются от задачи 3 только границами x_3.

Возвращаясь к шагу 1 при $x_0^4 = 0$, выберем задачу 4, оптимальное решение которой

$$x_0 = 15, \quad x_1 = x_2 = \frac{1}{3}, \quad x_3 = 1 \quad \text{(задача 4).} \tag{15}$$

Таким образом, приходим к шагу 4. Предположим, что выбирается x_2, вследствие чего получаем

$$\begin{gathered}\text{задачу 6: ограничения (9)}\\ 0 \leqslant x_1 \leqslant 2, \quad 1 \leqslant x_2 \leqslant 5, \quad 1 \leqslant x_3 \leqslant 5,\\ \text{задачу 7: ограничения (9)}\\ 0 \leqslant x_1 \leqslant 2, \quad 0 \leqslant x_2 \leqslant 0, \quad 1 \leqslant x_3 \leqslant 5\end{gathered} \tag{16}$$

Заметим вновь, что задачи 6 и 7 отличаются от задачи 4 только границами x_2.

Возвращаясь к шагу 1 при $x_0^5 = 0$, выберем задачу 6. Читателю следует убедиться в том, что задачи 5 и 7 остаются в основном списке. На шаге 2 обнаруживается, что задача 6 не имеет допустимого решения и поэтому нужно вернуться к шагу 1, приняв $x_0^6 = 0$. Выберем теперь задачу 7, оптимальное решение которой

$$x_0 = 14\frac{6}{7}, \quad x_1 = x_2 = 0, \quad x_3 = 1\frac{1}{7} \qquad \text{(задача 7)}. \tag{17}$$

Поскольку значение x_3 дробное, на шаге 4 получаем

$$\begin{gathered} \text{задачу 8: ограничения (9)} \\ 0 \leqslant x_1 \leqslant 2, \quad 0 \leqslant x_2 \leqslant 0, \quad 2 \leqslant x_3 \leqslant 5, \\ \text{задачу 9: ограничения (9)} \\ 0 \leqslant x_1 \leqslant 2, \quad 0 \leqslant x_2 \leqslant 0, \quad 1 \leqslant x_3 \leqslant 1. \end{gathered} \tag{18}$$

(Покажите, что выбор задачи 8 на итерации 7 не дает допустимого решения на шаге 2.) Выбор задачи 9 на итерации 8 приводит к решению

$$x_0 = 13, \quad x_1 = x_2 = 0, \quad x_3 = 1 \qquad \text{(задача 9)} \tag{19}$$

на шаге 2. Поэтому на шаге 3 фиксируется (19) и принимается $x_0^9 = 13$.

Возвращаясь к шагу 1, находим, что в основном списке осталась только задача 5. Оптимальное решение этой задачи линейного программирования равно

$$x_0 = 13, \quad x_1 = 2, \quad x_2 = 2\frac{1}{3}, \quad x_3 = 0 \qquad \text{(задача 5)}. \tag{20}$$

Поскольку значение x_0 в (20) равно x_0^9, возвращаемся к шагу 1 и прекращаем вычисления, так как основной список пуст. Оптимальным решением задачи целочисленного программирования является (19), зафиксированное на итерации 8.

Отметим, что возможность произвольного выбора возникает при реализации алгоритма в двух местах: выбор задачи, вычеркиваемой из основного списка на шаге 1, и выбор переменной, по которой формируются дополнительные задачи на шаге 4. Число итераций, необходимых для решения задачи, может существенно изменяться в зависимости от того, каким образом на самом деле производится выбор указанных альтернатив. (Так, например, выбор задачи 4 вместо задачи 5 на итерации 4 оказывается более выгодным.) Для повышения эффективности выбора были разработаны вспомогательные вычислительные процедуры, позволяющие оценивать «качество» выбора, но они здесь не рассматриваются, так как представляют интерес главным образом для узких специалистов. Точно так же выбор более удачной границы x_0^t уменьшает вероятность того, что после шага 2 придется перейти к шагам 3 и 4.

Краткое повторение. Ход итераций можно представить графически блок-схемой в виде дерева, показанной на рис. 13.1. Отметим, что каждая вершина дерева отображает задачу в основном списке, а каждая ветвь инцидентна одной из задач, добавляемой в основной список на шаге 4. Связь с графическим представлением объясняет использование понятия «ветвь» в названии метода, а понятие «граница» (оценка) появилось в этом названии благодаря испытанию, выполняемому на шаге 2. (Метод можно было бы также назвать **методом обрыва ветвей**.) Заметим, что ветвь, ведущая к задаче 5, обрывается, несмотря на то, что оптимальное решение этой задачи не является целочисленным. Это объясняется тем, что при достижении задачи 5 на $t = 9$ уже получается допустимое решение, при котором целевая функция принимает значение 13. Поэтому дальнейшее ветвление при наложении более сильных ограничений уже не может улучшить полученное ранее допустимое решение.

Теперь становится очевидным, почему метод получил название **метода возврата**. При обрыве некоторых ветвей (в рассмотренном примере это ветви, заканчивающиеся задачами 2, 6, 8 и 9) нужно вернуться обратно, просматривая соответствующие ветви с целью отыскания любой нерешенной задачи, содержащейся в блок-схеме дерева.

Существует несколько приемов, облегчающих вычисление оптимального решения на шаге 2. Так, например, применяя преобразование, рассмотренное в разд. 3.3, можно произвести замену каждой целочисленной переменной таким образом, что нижняя оценка будет равна нулю, а верхняя — величине $L_j + U_j$. В результате получается модель линейного программирования с m полностью развитыми линейными ограничениями, соответствующими (20), и p неотрицательными, ограниченными сверху переменными. В качестве исходного базиса при этих условиях можно выбрать оптимальные значения переменных, полученные при решении предыдущей задачи, которая породила рассматриваемую задачу на шаге 4. Такой исходный базис вследствие изменения границ более не приводит к неотрицательному решению, однако он удовлетворяет условиям, требующимся для применения двойственного симплексного алгоритма, изложенного в разд. 5.8. (Можно также использовать алгоритм с двусторонне ограниченными (ограниченными сверху) переменными, приведенный в разд. 5.10.)

Заключительные замечания. Метод ветвей и границ наряду с методом отсечения с вычислительной точки зрения обладает существенными достоинствами. Дело в том, что алгоритмы, построенные на этих методах, сравнительно легко программируются для ЭВМ и поэтому указанные алгоритмы реализуются на любой итерации без вмешательства человека. Метод ветвей и границ эффективен при

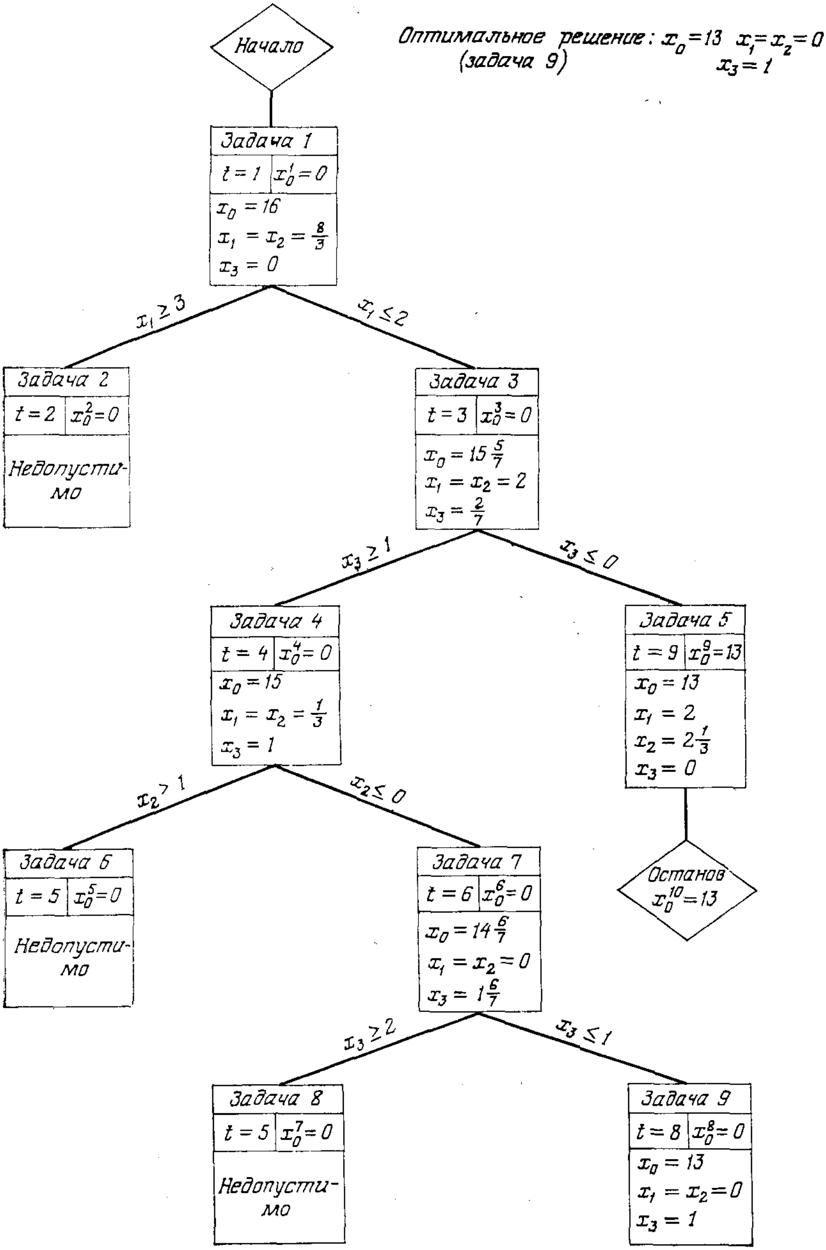


Рис. 13.1. Пример блок-схемы алгоритма, основанного на методе ветвей и границ.

решении задач, содержащих небольшое число целочисленных переменных. Однако если число таких переменных велико или решение задачи линейного программирования далеко от оптимального решения целочисленной задачи (как это имело место в рассмотренном примере), то число итераций, необходимых для получения решения, может оказаться слишком большим и для реализации алгоритма потребуются неоправданные с практической точки зрения затраты машинного времени.

Разумеется, рассмотренный метод можно применить для решения комбинаторных задач, содержащих только целочисленные булевы переменные. В этом случае любая величина $L_j = 0$, а $U_j = 1$ и две задачи, добавляемые на шаге 4, характеризуются альтернативами $x_j = 0$ и $x_j = 1$. Эффективность метода ветвей и границ часто удается повысить, используя специфические структурные особенности конкретной модели. В оставшейся части этой главы демонстрируются различные приемы использования такой возможности.

13.6. ЗАДАЧИ КОММИВОЯЖЕРА

Покажем теперь, как целесообразно модифицировать метод ветвей и границ, изложенный в предыдущем разделе, чтобы иметь возможность использовать особые структурные свойства комбинаторных задач. С этой целью выбрана задача коммивояжера, на примере которой показаны три различные модификации метода ветвей и границ. Многие задачи с булевыми переменными в постановках, которые приведены в разд. 13.2, можно решить с помощью одной или нескольких из этих модификаций, внеся в эти постановки необходимые изменения. Практическим результатом использования специфических структурных свойств конкретной модели является возможность решения задач довольно большой размерности. Так, например, задача коммивояжера, содержащая до 40 городов, была решена излагаемыми ниже методами при сравнительно небольших затратах машинного времени.

Задачу коммивояжера удобно записывать в постановке, рассмотренной в разд. 13.2:

$$\text{Минимизировать} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij}x_{ij}, \quad \text{где } c_{ii} = \infty, \quad i = 1, 2, \ldots, n, \tag{1}$$

при ограничениях

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i = 1, 2, \ldots, n \quad \text{(отъезд)}, \tag{2}$$

$$\sum_{i=1}^{n} x_{ij} = 1, \quad j = 1, 2, \ldots, n \quad \text{(прибытие)}, \tag{3}$$

$$x_{ij} \text{ — неотрицательные целые при всех } i \text{ и } j, \tag{4}$$

$$\text{решение есть цикл.} \tag{5}$$

Напомним, что $x_{ij} = 1$ означает, что коммивояжер переезжает из города i непосредственно в город j, где $c_{ij} \geqslant 0$ — соответствующее расстояние между городами. Если исключить условие (5), то (1) — (4) представляет собой задачу о назначениях (разд. 6.4), которую можно решить любым из методов, рассмотренных в приложении I (том I). (Отметим, однако, что в данном случае целью является

Расстояния c_{ij}

Из города \ В город	1	2	3	4	5
1	∞	10	25	25	10
2	1	∞	10	15	2
3	8	9	∞	20	10
4	14	10	24	∞	15
5	10	8	25	27	∞

Оптимальное назначение: $c_{15} + c_{23} + c_{34} + c_{42} + c_{51} = 60$

Рис. 13.2. *Условия задачи коммивояжера.*

«минимизация», в то время как в предыдущем разделе рассматривалась задача «максимизации». Излагаемые ниже модификации метода ветвей и границ отличаются вполне очевидными деталями от приведенной ранее модификации, что вызвано изменением содержания целевой функции.)

Три различные модификации метода ветвей и границ рассматриваются на примере конкретной задачи, исходные данные которой сведены в таблицу рис. 13.2. Подчеркнем, что единственное решение соответствующей задачи о назначениях $x_{15} = x_{51} = x_{23} = x_{34} = x_{42} = 1$ не удовлетворяет условию (5), так как оно содержит два цикла. Поэтому суммарное расстояние, определяемое этим решением и равное 60, строго меньше длины оптимального цикла. Существует всего $(n - 1)! = 4! = 24$ возможных циклов. (Читателю, прежде чем идти дальше, следует попытаться выявить оптимальный цикл.)

Метод исключения подциклов. Из трех рассматриваемых в дальнейшем модификаций метода ветвей и границ данная в наименьшей степени отличается от изложенного ранее алгоритма. В начале итерации t известна *верхняя* граница (оценка) x_0^t оптимального значения целевой функции. Можно принять x_0^1 равным достаточно большому числу, скажем сумме $(c_{12} + c_{23} + \ldots + c_{n1})$, соответствующей циклу, включающему проезд из города 1 в город 2... в город n, в город 1. Кроме того, имеется основной список, содержащий ряд задач о назначениях. Все эти задачи имеют вид (1) — (4), но отличаются друг от друга тем, что в них различные величины c_{ij} равны ∞. На итерации 1 основной список состоит из задачи о назначениях (1) — (4), которую назовем задачей 1. Определим подцикл как цикл, содержащий не все n городов, скажем $x_{15} = x_{51} = 1$ или $x_{23} = x_{34} = x_{42} = 1$.

На итерации t выполняются следующие шаги.

Шаг 1. Прекратить вычисления, если основной список пуст. В противном случае выбрать одну задачу, вычеркнув ее из основного списка.

Шаг 2. Решить выбранную задачу о назначениях. Если оптимальное значение целевой функции (которое может быть равным ∞) больше или равно x_0^t, то принять $x_0^{t+1} = x_0^t$ и вернуться к шагу 1. В противном случае перейти к шагу 3.

Шаг 3. Если полученное оптимальное решение выбранной задачи о назначениях является циклом, то зафиксировать его, принять x_0^{t+1} равным соответствующему оптимальному значению целевой функции и вернуться к шагу 1. В противном случае перейти к шагу 4.

Шаг 4. Остановиться в полученном оптимальном решении выбранной задачи о назначениях на подцикле, содержащем минимальное число городов. Каждому $x_{ij} = 1$ в выбранном подцикле поставить в соответствие задачу о назначениях, внеся ее в основной список и приняв соответствующее значение $c_{ij} = \infty$. Оставить все остальные коэффициенты теми же, что в задаче, выбранной на шаге 1. Принять $x_0^2 = x_0^1 = 65$ и вернуться к шагу 1.

Разбор примера, приведенного в таблице рис. 13.2, позволяет уяснить детали этого алгоритма. Примем на итерации 1 $x_0^t = 65$, что соответствует циклу город 1 — город 2,. . ., город 5 — город 1. Основной список содержит лишь одну задачу о назначениях, в которой значения c_{ij} соответствуют указанным в таблице рис. 13.2. Поэтому на шаге 1 нужно выбрать эту задачу. Находим, что оптимальное решение на шаге 2 дает значение целевой функции, равное 60, что отражено в таблице рис. 13.2 и приводит к шагу 3. Поскольку полученное оптимальное решение содержит два подцикла, переходим к шагу 4. Выбираем подцикл $x_{15} = x_{51} = 1$, так как в него входит на один город меньше, чем в подцикл $x_{23} = x_{34} = x_{42} = 1$. Вносим две задачи в основной список, полагая в одной из них $c_{15} = \infty$, а в другой — $c_{51} = \infty$ и оставляя все остальные значения c_{ij} без изменений. Принимаем $x_0^2 = x_0^1 = 65$.

Обосновать шаг 4 совсем несложно. Поскольку цикл $x_{15} = x_{51} = 1$ недопустим, то в оптимальном цикле либо $x_{15} = 0$, либо $x_{51} = 0$ (либо обе эти переменные равны нулю). Задачи, введенные на шаге 4, допускают любую из этих возможностей. Аналогичный ход рассуждений справедлив для любого подцикла. (Использование подциклов как источников формирования новых задач на шаге 4 является отличительной особенностью этой модификации по сравнению с алгоритмом, рассмотренным в предыдущем разделе. В этом варианте несколько задач, включенных в основной список, могут оказаться тождественными. Такое совпадение легко обнаружить при ручном выполнении вычислений, решая задачу малой размерности. При обнаружении избыточных задач их просто исключают из дальнейшего рассмотрения. Однако при реализации этой модификации алгоритма на ЭВМ может оказаться предпочтительным решение избыточных задач, а не проверка повторений.)

Возвращаясь к шагу 1, выберем задачу, в которой $c_{15} = \infty$, назвав ее задачей 2. Оптимальное значение целевой функции соответствующей задачи о назначениях равно 65. Это значение представляет собой нижнюю границу для любого цикла, исключающего маршрут

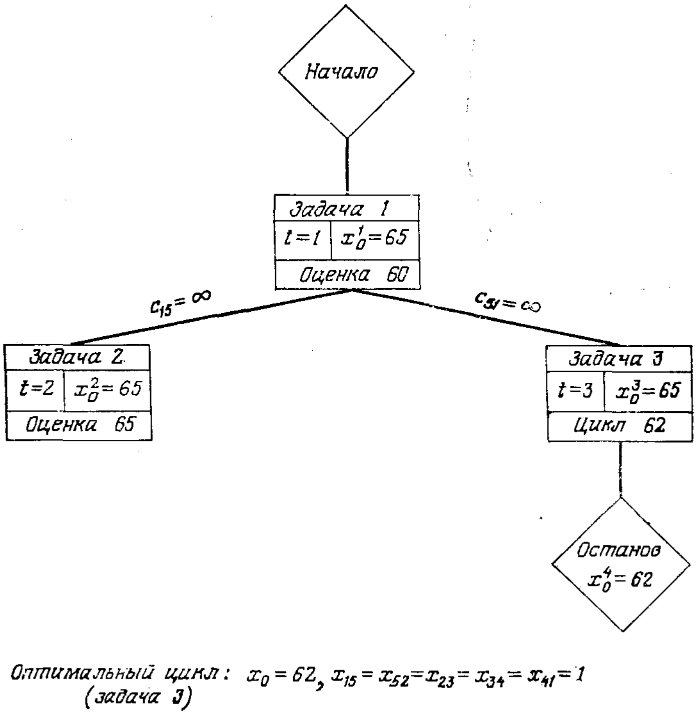


Рис. 13.3. Блок-схема алгоритма исключения подциклов.

из города 1 в город 5. Поскольку цикл длиной 65 уже имеется, нет надобности далее рассматривать эту задачу. Таким образом, в соответствии с шагом 2 примем $x_0^3 = x_0^2 = 65$ и вернемся к шагу 1. Выберем оставшуюся задачу, в которой $c_{51} = \infty$, и назовем ее задачей 3. Теперь мы получим оптимальное значение целевой функции в задаче о назначениях, равное 62, и переходим к шагу 3. Далее находим соответствующее оптимальное решение:

$$\text{Минимальное расстояние} = 62,\ x_{15} = x_{52} = x_{23} = x_{3\cdot} = x_{41} = 1, \tag{6}$$

которое является циклом. Поэтому фиксируем (6), полагаем $x_0^4 = 62$ и возвращаемся к шагу 1, обнаруживая, что основной список пуст, вследствие чего вычисления прекращаются. Дерево, соответствующее рассмотренному примеру, приведено на рис. 13.3.

Метод задания маршрутов. Изложенный выше алгоритм позволяет ограничить число просматриваемых ветвлений, что достигается за счет вычисления эффективной нижней оценки (границы) целевой функции для любого цикла, порождаемого каждой задачей. Однако для получения этой оценки приходится отыскивать оптимальное решение задачи о назначениях. Покажем теперь, как можно вычислять оценки более простым способом. Однако цена, которую приходится платить за такое упрощение, определяется тем, что в этом алгоритме нужно исследовать большее число ветвей соответствующего дерева.

В начале любой итерации t известна верхняя оценка x_0^t оптимального значения целевой функции. Значение x_0^1 можно определить, исходя из тех же соображений, что излагались выше. Кроме того, имеется основной список задач, в котором некоторое подмножество значений c_{ij} изменено и принято равным ∞, а подмножество $x_{ij} = 1$. Среди значений $x_{ij} = 1$ отсутствуют наборы, образующие подциклы. На итерации 1 основной список включает две задачи: в одной из них значение выбранного c_{ij} изменено на ∞, в другой — соответствующая переменная $x_{ij} = 1$, а $c_{ji} = \infty$. (Полагая $c_{ji} = \infty$, запрещают тем самым $x_{ji} = 1$, что приводило бы к образованию подцикла город i — город j — город i.)

Положение каждой задачи, входящей в основной список, удобно определять следующим образом. Возьмем матрицу c_{ij}, подобную приведенной на рис. 13.2. Если $x_{kh} = 1$, вычеркнем k-ю строку и h-й столбец. Остальные c_{ij} (из которых некоторые равны ∞) соответствуют еще не назначенным переменным x_{ij}. Нижнюю оценку оптимального значения целевой функции для любого цикла, содержащего заданное подмножество $x_{ij} = 1$, можно вычислить различными способами. В целом чем больше нижняя оценка, тем меньшее число ветвей приходится исследовать.

Один простой, но достаточно эффективный метод вычисления этой оценки основан на том, что расстояние должно быть по крайней мере равно сумме c_{ij} при $x_{ij} = 1$ плюс сумма наименьших c_{ij} в каждой из невычеркнутых строк. Можно (и должно) увеличить эту оценку в еще большей степени, вычитая минимальный коэффициент c_{ij} в каждой невычеркнутой строке из всех остальных c_{ij} этой строки. Таким образом, получаются так называемые *уменьшенные расстояния*. Далее, к указанной выше сумме добавляется минимальное из уменьшенных расстояний по каждому из невычеркнутых столбцов. Пример, иллюстрирующий эту процедуру, приведен в таблице рис. 13.4. Заметим, что значение величины x_{23} было принято равным единице. В верхней таблице указаны минимальные расстояния по каждой строке, исключая вторую. Каждое из этих расстояний вычитается

$x_{23}=1$ $(c_{23}=10)$

Из города \ В город	1	2	3	4	5	Минимум по строкам
1	∞	10		25	10	10
2						
3	8	∞		20	10	8
4	14	10		∞	15	10
5	10	8		27	∞	8

Уменьшенные расстояния

Из города \ В город	1	2	3	4	5
1	∞	0		15	0
2					
3	0	∞		12	2
4	4	0		∞	5
5	2	0		19	∞
Минимум по столбцам	0	0		12	0

Оценка: $c_{23}+\Sigma$(Минимум по строкам)$+\Sigma$(Минимум по столбцам)$=$
$=10+10+8+10+8+0+0+12+0=58$

Рис. 13.4. Метод определения оценок в алгоритме задания маршрутов.

из каждого элемента в соответствующей строке, что дает нижнюю таблицу.

(На самом деле, можно найти нижнюю оценку, решая задачу о назначениях, образованную из величин c_{ij}, которые принадлежат невычеркнутым строкам и столбцам, добавляя полученное оптимальное значение целевой функции к сумме коэффициентов c_{ij}, для которых заданное подмножество $x_{ij}=1$. Такой подход, пожалуй, более трудоемок с вычислительной точки зрения, чем только что описанная модификация методов ветвей и границ или изложенный простой прием вычисления оценок.) На итерации t выполняются следующие шаги.

Шаг 1. Прекратить вычисления, если основной список пуст. В противном случае выбрать одну задачу и вычеркнуть ее из основного списка.

Шаг 2. Определить нижнюю оценку целевой функции для любого цикла, порождаемого выбранной задачей. Если нижняя оценка больше или равна x_0^t, то принять $x_0^{t+1} = x_0^t$ и вернуться к шагу 1. В противном случае перейти к шагу 3.

Шаг 3. Если текущее решение определяет цикл, то зафиксировать его, принять x_0^{t+1} равным соответствующему значению целевой функции и вернуться к шагу 1. В противном случае перейти к шагу 4.

Шаг 4. При наличии возможности выбрать переменную x_{hk}, не входящую в текущее решение, такую, что $c_{hk} < \infty$ при условии, что $x_{hk} = 1$ не приводит к образованию подцикла на переменных, уже вошедших в решение. При таком выборе внести в основной список две задачи. Каждую из этих задач принять идентичной задаче, выбранной на шаге 1, за исключением лишь того, что в одну из них ввести изменение $c_{hk} = \infty$, а в другую — условие $x_{hk} = 1$ и изменение $c_{kh} = \infty$. Принять $x_0^{t+t} = x_0^t$ и вернуться к шагу 1.

Отметим два существенных отличия данного варианта алгоритма ветвей и границ от предыдущей. В данном варианте на шаге 2 вычисляется нижняя оценка для выбранной задачи, но не отыскивается ее оптимальное решение. Кроме того, на шаге 4 в основной список могут вноситься две задачи или не добавляться ни одной, если не существует переменной x_{hk}, удовлетворяющей указанным условиям. В предыдущей модификации на шаге 4 образуется от 2 до $n/2$ новых задач, вносимых в основной список.

Применение рассмотренного алгоритма к примеру рис. 13.2 дает дерево, изображенное на рис. 13.5. Очевидно, что исследуемые ветви этого дерева определяются выбором задач из основного списка на шаге 1 и выбором x_{hk} на шаге 4. Правило выбора на шаге 1 предусматривает исключение из основного списка задачи, имеющей наибольший номер. Этот номер присваивается на шаге 4 задаче, в которой $x_{hk} = 1$. Оптимальный цикл для рассматриваемого примера был получен на итерации $t = 8$ в результате решения задачи 12.

Часто число ветвей удается уменьшить за счет еще нескольких изменений на шаге 4 коэффициентов c_{ij} в новой задаче, для которой $x_{hk} = 1$. Изменение $c_{kh} = \infty$, запрещающее подцикл между городами $h - k - h$, уже было указано. Однако можно также принять и другие коэффициенты c_{ij} равными ∞, чтобы запретить более длинные подциклы. Так, например, если $x_{12} = x_{23} = 1$ входят в текущее решение и выбирается x_{34}, то можно положить $c_{41} = c_{42} = \infty$. Если этот прием используется на всех итерациях, то нет надобности вводить на шаге 4 условие «$x_{hk} = 1$ не ведет к образованию подцикла», поскольку ограничение $c_{hk} < \infty$ уже запрещает такую возможность.

Метод частичных циклов. Главной причиной изложения этой модификации метода ветвей и границ является ознакомление с под-

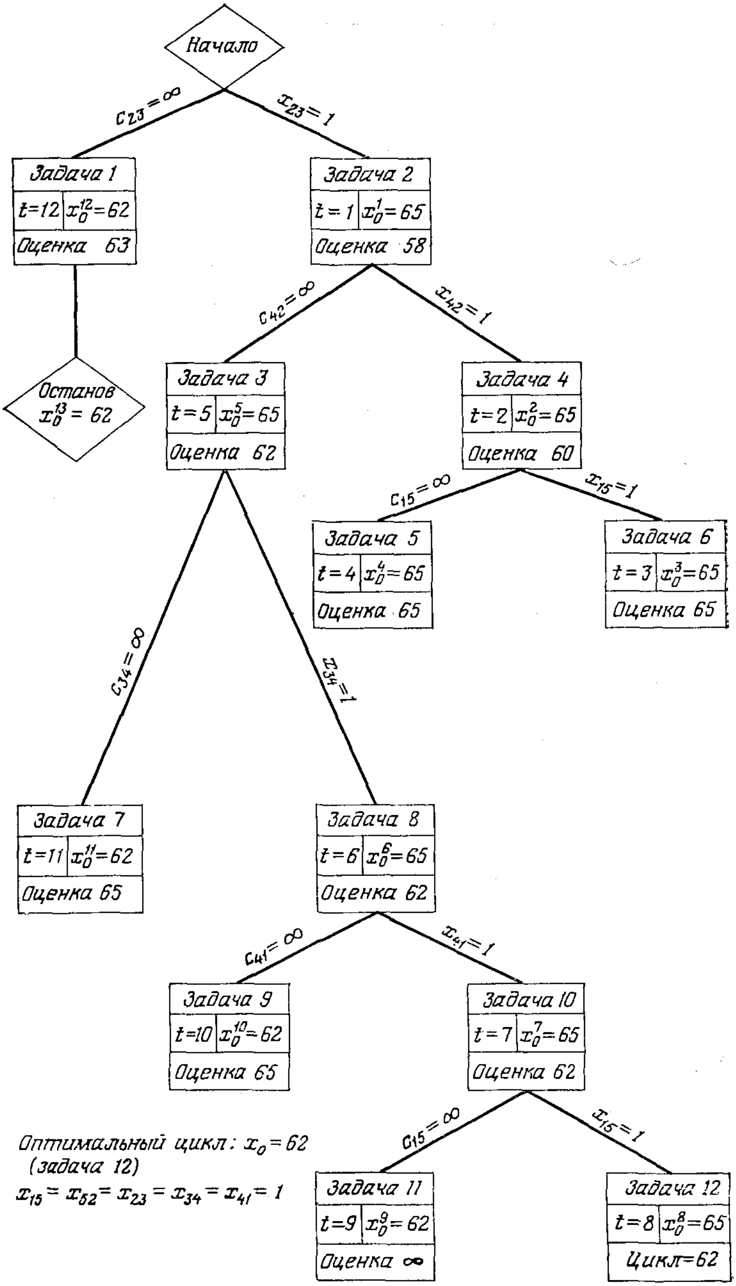


Рис. 13.5. Блок-схема алгоритма задания маршрутов.

ходом, используемым для решения самых различных комбинаторных задач, например задач упорядочения (в которых $x_{ij} = 1$ соответствует размещению i-го элемента на j-м месте некоторой последовательности). Для этого потребуется использование понятия *частичный цикл*, которое определяется как последовательность, содержащая менее n различных городов и начинающаяся с города 1 (например, город 1 — город i — город j — город k, где $n > 4$).

В начале любой итерации t известна верхняя оценка x_0^t оптимального значения целевой функции. Значение x_0^1 определяется общепринятым способом. Кроме того, задан основной список задач, содержащий некоторое подмножество $x_{ij} = 1$, определяющее частичный цикл, и подмножество значений c_{ij}, принятых в результате пересмотра равными ∞. На итерации 1 основной список содержит $n - 1$ задач, по одной для каждого $x_{1j} = 1$ и пересмотренных $c_{j1} = \infty$, $j = 2, 3, \ldots, n$. Для вычисления нижней оценки оптимального значения целевой функции, соответствующей циклу, который является дополнением частичного цикла, можно применить тот же метод, что и в алгоритме *задания маршрутов*. (С другой стороны, можно определять оптимальное решение задачи о назначениях, включив в эту задачу коэффициенты c_{ij}, принадлежащие строкам и столбцам, не связанным с подмножеством $x_{ij} = 1$, которые входят в частичный цикл.)

Рассматриваемая модификация отличается от метода задания маршрутов только шагом 4.

Шаг 4. По каждому городу k, который не входит в частичный цикл задачи, выбранной на шаге 1, внести дополнительную задачу в основной список, расширив частичный цикл из города j, являющегося последним городом, включенным в частичный цикл, до города k, изменив при этом c_{kj} на ∞. Принять $x_0^{t+1} = x_0^t$ и вернуться к шагу 1.

В этом алгоритме в случае, когда частичный цикл содержит m городов (включая город 1), на шаге 4 в основной список включается $n - m$ задач. Отметим, что на шаге 4 отсутствует произвольный выбор. (Избыточные вычисления в данном алгоритме могут иметь место, если в основной список входят несколько задач, содержащих в своих частичных циклах одни и те же m городов, например город 1 — город 2 — город 3 и город 1 — город 3 — город 2. В этом случае вычисления оценок на шаге 2 повторяются.)

Применение этого алгоритма к примеру рис. 13.2 дает дерево, показанное на рис. 13.6. В этом случае анализируемые ветви определяются выбором задачи из основного списка на шаге 1. (Для решения примера была использована наиболее удобная процедура.)

Заключительные замечания. Как легко заметить, анализируя результаты неудачного выбора на шагах 1 и 4, в рассмотренных методах вычислительная трудоемкость определяется прежде всего качеством полученного решения для задачи произвольного выбора. Пользуясь алгоритмами, основанными на методе ветвей и границ,

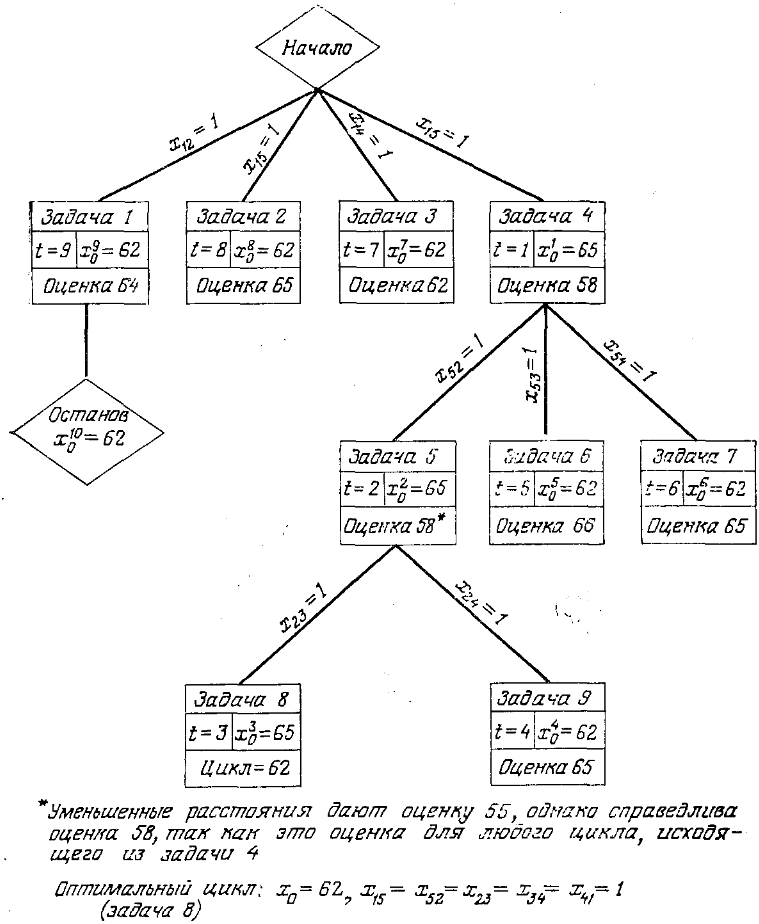


Рис. 13.6. Блок-схема алгоритма частичных циклов.

почти всегда стоит выполнять дополнительные вычисления на шаге 2 с целью определения хорошей оценки оптимального значения целевой функции. Эти «затраты» компенсируются за счет того, что уменьшается число ветвей, подвергаемых анализу. Кроме того, на шаге 2 обычно можно использовать результаты вычислений, уже выполненных при решении задачи, породившей рассматриваемую. Практическая эффективность применения метода ветвей и границ для решения реальных комбинаторных задач зависит прежде всего от учета специфики конкретной задачи, позволяющей реализовать указанные рекомендации.

13.7. МЕТОД ЧАСТИЧНОГО (НЕЯВНОГО) ПЕРЕБОРА

Многие важные задачи целочисленного программирования можно описать следующим образом:

$$\text{Максимизировать } \sum_{j=1}^{n} c_j x_j \tag{1}$$

при ограничениях

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

где условия целочисленности сведены к

$$x_j = \begin{cases} 0, \\ 1, \end{cases} \quad j = 1, 2, \ldots, n. \tag{3}$$

Предположим, что любой коэффициент c_j есть целое число (этого всегда можно добиться, выбрав правильный масштаб целевой функции при условии, что исходные значения коэффициентов заданы рациональными числами).

Модели распределения капиталовложений часто можно представить в виде (1) — (3). Кроме того, многие полностью целочисленные задачи можно преобразовать таким образом, чтобы каждая переменная удовлетворяла условиям (3). Предположим, к примеру, что переменная x имеет определимую верхнюю оценку U. Тогда x во всех выражениях можно заменить эквивалентным **двоичным представлением**

$$x \equiv 1w_1 + 2w_2 + 4w_3 + \ldots + 2^{k-1}w_k, \tag{4}$$

где $w_j = 1$ или 0 и значение k выбирается так, чтобы выполнялось условие $2^k - 1 \geqslant U$. Так, если известно заранее, что $x_1 \leqslant 6$ и в задачу входит ограничение $3x_1 + x_2 \leqslant 25$, то, приняв $x_1 = 1w_1 + 2w_2 + 4w_3$, это ограничение можно записать в виде $3w_1 + 6w_2 + 12w_3 + x_2 \leqslant 25$. (Заметим, что в этом примере при преобразовании (4) теряется некоторая информация. Это связано со следующим обстоятельством: формально допускается возможность $w_1 = w_2 = w_3 = 1$, что соответствует максимальному значению x_1, равному 7, в то время как известно, что значение x_1 не может быть больше 6. Эту потерю легко устранить, используя некоторый коэффициент, который достаточно мал по сравнению 2^{k-1}. В рассмотренном примере в качестве коэффициента при w_3 вместо 4 можно принять 3. Читателю следует сформулировать общее правило выбора коэффициента при w_k, обеспечивающее такое преобразование, при котором исходная информация о верхней оценке переменной не теряется.)

Напомним, что в случае применения алгоритма ветвей и границ, приведенного в разд. 13.4, к решению задачи (1) — (3) приходится решать последовательность задач линейного программирования.

В излагаемом ниже алгоритме используется условие (3), вследствие чего вычисления ограничиваются операциями сложения (и вычитания). По этой причине такой метод иногда называют **аддитивным алгоритмом.**

Если учитывать лишь условие (3), то существует 2^n возможных выборов значений $(x_1, x_2, \ldots, x_n)$. Разумеется, многие из них недопустимы из-за линейных ограничений (2) и лишь очень небольшое число из них являются оптимальными. Рассмотрим некоторое подмножество x_j, в котором каждому x_j поставлено в соответствие определенное числовое значение (либо нуль, либо единица). Такое подмножество называется **частичным решением**. Переменные x_j, не входящие в частичное решение, носят название **свободных переменных.** Любой конкретный выбор числовых значений свободных переменных называется **дополнением** соответствующего частичного решения. Если частичное решение содержит s переменных, то существует 2^{n-s} дополнений. В рассматриваемом алгоритме каждая задача основного списка соответствует *частичному решению* [не обязательно допустимому по условиям (21)], а возможные дополнения порождают ветви дерева.

Нижеследующий пример поясняет понятия, введенные в предыдущем абзаце. Предположим, что $n = 5$. Тогда $(x_2 = 1,\ x_4 = 0)$ есть одно частичное решение, а соответствующими свободными переменными являются x_1, x_3 и x_5. Полагая $x_1 = 0$, $x_3 = 1$ и $x_5 = 1$, получаем одно из дополнений этого частичного решения. В общей сложности существует $2^{5-2} = 8$ дополнений частичного решения $(x_2 = 1,\ x_4 = 0)$.

Допустим, что известна нижняя оценка оптимального значения целевой функции и что зафиксировано допустимое решение, дающее эту нижнюю оценку. Тогда при условии, что получено некоторое частичное решение, нет необходимости дальнейшего ветвления, если каким-либо образом можно показать, что не существует допустимого дополнения, которое имеет значение целевой функции и превышает текущую нижнюю оценку. В этом случае говорят, что частичное решение **прозондировано**. При зондировании частичного решения, содержащего s переменных, неявным образом перебирается 2^{n-s} возможных «назначений», которые удовлетворяют условию, выраженному (3).

Существует тесная связь между основной идеей рассматриваемого метода и принципом оптимальности динамического программирования. Соответствующую формулировку указанного принципа применительно к аддитивному алгоритму можно выразить следующим образом: при заданном частичном решении значения остальных переменных должны выбираться так, чтобы дополнение этого решения было оптимальным.

При отсутствии таких значений остальных переменных, которые дают допустимое решение, или в случае, когда полученные оптимальные значения этих переменных приводят к худшему решению, чем найдено ранее, не существует оптимального решения, содержащего заданное частичное решение.

На любой итерации t известна нижняя оценка x_0^t оптимального значения целевой функции. Значение x_0^1 можно выбрать точно так же, как в методе ветвей и границ (разд. 13.2). Кроме того, имеется основной список задач, в котором каждой задаче соответствует определенное частичное решение. На итерации 1 основной список содержит две задачи, полученные в результате выбора x^k, причем принимается, что частичное решение одной задачи $x_k = 0$, а другой $x_k = 1$.

Прежде чем перейти к изложению алгоритма, полезно сделать несколько замечаний. Существует несколько способов, которые можно применить, чтобы установить, имеется ли какое-либо допустимое дополнение, дающее значение целевой функции, превышающее текущую нижнюю оценку. Один из таких простых способов иллюстрируется двумя приводимыми ниже примерами. Если задача содержит ограничение

$$-1x_1 + 2x_2 - 6x_3 + 4x_4 - 1x_5 \leqslant 0, \tag{5}$$

то не существует дополнения частичного решения ($x_1 = 1$, $x_3 = 0$, $x_4 = 1$), допустимого по условию (5). (Объясните, почему это так.) Предположим, что целевая функция описывается выражением

$$2x_1 + 1x_2 + 5x_3 + 3x_4 - 1x_5 \tag{6}$$

и что $x_0^t = 7$. Тогда, чтобы улучшить значение x_0^t, дополнение должно удовлетворять ограничению

$$2x_1 + 1x_2 + 5x_3 + 3x_4 - 1x_5 \geqslant 8$$
$$\text{или} \quad -2x_1 - 1x_2 - 5x_3 - 3x_4 + 1x_5 \leqslant -8. \tag{7}$$

Ни одно дополнение частичного решения ($x_2 = 1$, $x_3 = 0$, $x_5 = 1$) не является допустимым по условию (7). (Объясните, почему это так.)

Смысл рассмотренных примеров заключается в следующем. При заданном частичном решении представим линейные ограничения в виде

$$\sum_{\substack{\text{по всем}\\ \text{свободным}\\ \text{переменным}}} a_{ij}x_j \leqslant b_i - \sum_{\substack{\text{по переменным}\\ \text{частичного}\\ \text{решения}}} a_{ij}x_j, \quad i = 0, 1, 2, \ldots, m, \tag{8}$$

где каждой переменной x_j в частичном решении поставлено в соответствие определенное значение, входящее под знак суммы в правой части неравенства (8), а коэффициенты в ограничении при $i = 0$ равны $a_{0j} = -c_j$ и $b_0 = -x_0^t - 1$. Тогда допустимого дополнения, которому соответствует значение целевой функции, превосходящее нижнюю оценку, не существует, если

$$\sum_{\substack{\text{по всем}\\ \text{свободным}\\ \text{переменным}}} \min(a_{ij}, 0) > b_i - \sum_{\substack{\text{по переменным}\\ \text{частичного}\\ \text{решения}}} a_{ij}x_j \quad \text{при любом } i. \tag{9}$$

Член в левой части неравенства (9) представляет собой просто сумму всех отрицательных коэффициентов при свободных переменных. Если эта сумма больше правой части неравенства, то, даже полагая $x_j = 1$, для каждой свободной переменной, где $a_{ij} < 0$, невозможно выполнить i-е ограничение (8). [Для проверки того, насколько хорошо вы понимаете это утверждение, следует применить (9) к примерам (5) и (7).]

Другим важным моментом является то, что при заданном частичном решении иногда удается определить, какое значение должна иметь свободная переменная при *любом* допустимом дополнении в случае, когда значение целевой функции превосходит текущую нижнюю оценку. Справедливость этого утверждения можно показать на примере, рассмотрев задачу, содержащую ограничение

$$1x_1 - 1x_2 + 2x_3 - 1x_4 \leqslant 0. \qquad (10)$$

Читателю следует убедиться в том, что при частичном решении $(x_1 = 1,\ x_2 = 1)$ в любом дополнении должно выполняться условие $x_3 = 0$, при котором оно допустимо по ограничению (10). Этот пример позволяет сделать вывод о возможности выполнения проверки, аналогичной (9). В частности, для любой свободной переменной x_k в случае, когда

$$\sum_{\substack{\text{по всем}\\ \text{свободным}\\ \text{переменным}}} \min(a_{ij},\ 0) + |a_{ik}| > b_i - \sum_{\substack{\text{по пере-}\\ \text{менным час-}\\ \text{тичного ре-}\\ \text{шения}}} a_{ij}x_j \quad \text{при любом } i, \qquad (11)$$

$x_k = 0$, если $a_{ik} > 0$, и $x_k = 1$, если $a_{ik} < 0$. [Проверки (9) и (11) можно также применить к составным ограничениям, образованным путем добавления положительных комбинаций ограничений (2), а именно

$$\sum_{j=1}^{n} \Big(\sum_{i=1}^{m} y_i a_{ij}\Big) x_j \leqslant \sum_{i=1}^{m} y_i b_i,$$

где $y_i > 0$. Чтобы показать, как можно использовать такие **составные** или **замещенные** ограничения, рассмотрим условия $-x_1 - x_2 - x_3 \leqslant -1$, $\quad 2x_2 - 2x_4 + 2x_5 \leqslant 0$, $\quad 2x_3 + 2x_4 - 2x_5 \leqslant 0$. При частичном решении $x_1 = 0$ применение проверки (9) к каждому ограничению не указывает на недопустимость дополнения. Однако легко убедиться в том, что применение (9) к сумме этих ограничений показывает, что допустимого дополнения не существует. Аналогично при $x_1 = 1$ применение (11) к каждому ограничению не дает информации о значениях свободных переменных. Однако проверка (11) суммы ограничений показывает, что в любом допустимом дополнении $x_2 = x_3 = 0$.]

Перейдем теперь к описанию алгоритма. На итерации t выполняются следующие шаги.

Шаг 1. Прекратить вычисления, если основной список пуст. В противном случае выбрать задачу из основного списка и вычеркнуть ее из него.

Шаг 2. Если можно найти свободные переменные, которые должны иметь определенные значения при любом допустимом дополнении, когда значение целевой функции превосходит x_0^t, то соответствующим образом расширить выбранное частичное решение. Если можно установить, что не существует допустимого дополнения, у которого значение целевой функции превосходит x_0^t, то положить $x_0^{t+1} = x_0^t$ и вернуться к шагу 1. В противном случае перейти к шагу 3.

Шаг 3. Если (расширенное) частичное решение является полным (т. е. содержит все n переменных), зафиксировать его, принять x_0^{t+1} равным соответствующему значению целевой функции и вернуться к шагу 1. В противном случае перейти к шагу 4.

Шаг 4. Выбрать любую свободную переменную x_k, не входящую в (расширенное) частичное решение. Внести две задачи в основной список. В одной из них положить $x_k = 0$ (в расширенном) частичном решении, в другой принять $x_k = 1$. Положить $x_0^{t+1} = x_0^t$ и вернуться к шагу 1.

При прекращении выполнения операций алгоритма в случае, когда зафиксировано допустимое решение, дающее x_0^t, это решение является оптимальным. В противном случае допустимого решения не существует. Алгоритм должен обеспечивать сходимость за конéчное число итераций.

Пример. Рассмотрим задачу

$$\text{Максимизировать } 3x_1 + 6x_2 + 3x_3 + 6x_4 + 13x_5 \tag{12}$$

при ограничениях

$$-3x_1 - 6x_2 + 6x_3 + 12x_4 + 7x_5 \leqslant 8, \quad i = 1, \tag{13}$$

$$6x_1 + 12x_2 - 3x_3 - 6x_4 + 7x_5 \leqslant 8, \quad i = 2, \tag{14}$$

$$x_j = \begin{cases} 0, \\ 1, \end{cases} \quad j = 1, 2, \ldots, 5. \tag{15}$$

Это пример из разд. 13.5, измененный соответствующим образом, т. е. включающий только булевы переменные. Примем $x_0^1 = 0$, что соответствует допустимому решению, в котором все $x_j = 0$. Предположим, что в основной список входит задача 1 с частичным решением $x_5 = 0$ и задача 2 с частичным решением $x_5 = 0$.

Вычеркнем на шаге 1 задачу 1. Допустимое дополнение, имеющее значение целевой функции, превосходящее x_0^1, должно удовлетворять условиям (13) — (15), а также условию

$$-3x_1 - 6x_2 - 3x_3 - 6x_4 - 13x_5 \leqslant -1, \quad i = 0. \tag{16}$$

В первой части шага 2 попытаемся найти какие-либо свободные переменные, которые должны иметь определенные значения, чтобы удовлетворять ограничениям (13) — (16) при условии $x_5 = 1$. Следует показать, что применение проверки (11) при $i = 1$ и $i = 2$ дает $x_4 = x_2 = 0$. Далее, при условии $x_5 = 1$, $x_2 = x_4 = 0$ вторичное применение (11) при $i = 1$ и $i = 2$ дает $x_3 = x_1 = 0$. Таким образом, получено расширенное решение, содержащее все переменные. Во второй части шага 2 проверяется допустимость по $i = 0, 1, 2$, формально оцениваемая путем применения проверки (9). [Поскольку свободных переменных не осталось, расчет сводится к проверке допустимого расширенного решения по условиям (13) — (16). Следует показать, что оно допустимо.]

Далее переходим к шагу 3 и фиксируем дополнение

$$\text{Целевая функция} = 13, \quad x_1 = x_2 = x_3 = x_4 = 0, \quad x_5 = 1. \quad (17)$$

Полагаем $x_0^2 = 13$ и возвращаемся к шагу 1.

Теперь вычеркнем задачу 2, где $x_5 = 0$. Ограничение (16) изменяется следующим образом:

$$-3x_1 - 6x_2 - 3x_3 - 6x_4 - 13x_5 \leqslant -14, \quad i = 0. \quad (18)$$

Применение (11) при $i = 0, 1, 2$ не приводит к расширению частичного решения так же, как применение (9) при $i = 0, 1, 2$ не исключает возможности появления допустимого дополнения со значением целевой функции, превосходящим 13. Следовательно, переходим к шагу 3, а затем к шагу 4, поскольку частичное решение не является полным. На шаге 4 выбираем переменную x_1 и вносим в основной список задачу 3, в которой $x_5 = 0$, $x_1 = 1$, и задачу 4, в которой $x_5 = 0$, $x_1 = 0$.

Возвращаясь к шагу 1 при $x_0^3 = 13$, вычеркиваем задачу 3. Нужно показать, что применение проверки (11) при $i = 0, 1$ не определяет обязательных значений свободных переменных. Переходя к $i = 2$, следует убедиться в том, что проверка (11) дает $x_2 = 0$ для любого допустимого дополнения. Расширяя соответствующим образом решение и повторно применяя (11) при $i = 0$, следует показать что допустимого дополнения не существует. Поэтому шаг 2 заканчивается и нужно вернуться к шагу 1, приняв $x_0^4 = 13$.

После выбора задачи 4, в которой $x_1 = x_5 = 0$, проверка (11) при $i = 0$ показывает, что для любого допустимого дополнения $x_3 = 1$. Применяя (11) при $i = 1, 2$, находим, что $x_4 = 0$ и $x_2 = 0$. Далее, в результате проверки (9) при $i = 0$ получаем, что не существует допустимого дополнения со значением целевой функции, превосходящим x_0^4. Возвращаясь к шагу 1, прекращаем дальнейшие вычисления, так как основной список пуст. Оптимальное значение целевой функции равно 13.

Блок-схема алгоритма применительно к рассмотренному примеру показана на рис. 13.7. Как легко видеть из задачи 3, всякий раз, когда проверка (11) приводит к расширению частичного решения, более выгодным может оказаться повторение этой проверки по всем ограничениям ($i = 0, 1, \ldots, m$) на основе использования вновь полученного частичного решения.

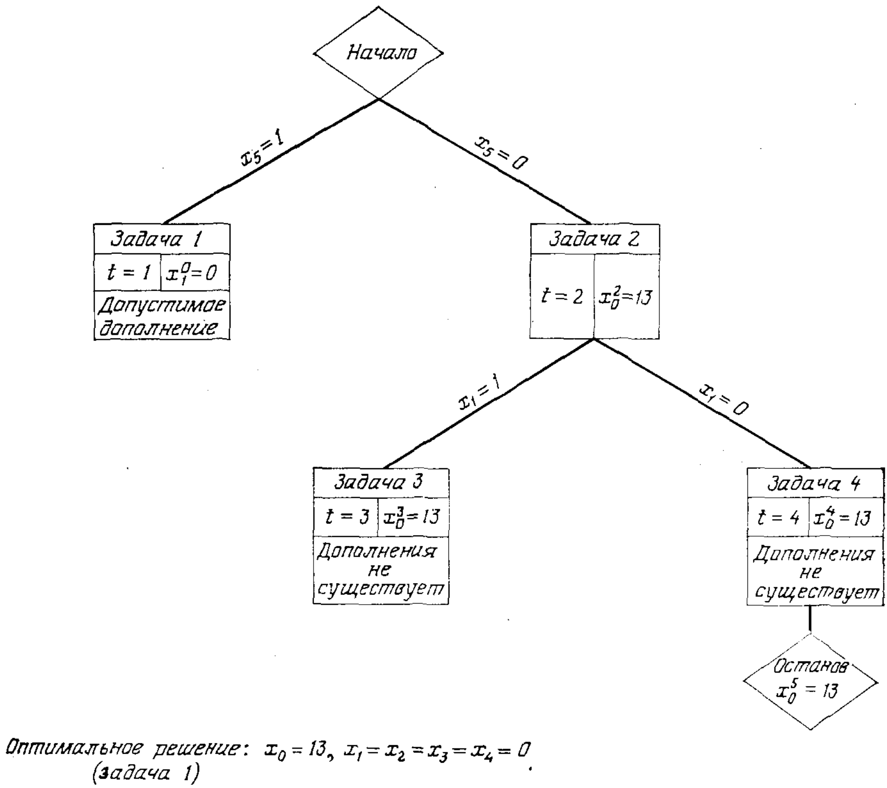


Рис. 13.7. Блок-схема аддитивного алгоритма (алгоритма частичного перебора).

Заключительные замечания. Следует обратить внимание на два основных различия между изложенным методом и методом ветвей и границ, рассмотренным в разд. 13.4. Во-первых, в аддитивном алгоритме требуется выполнение только операций сложения и вычитания (не нужно ни умножения, ни деления). Выбор на шагах 1 и 4 может основываться на информации, полученной из оптимального решения задачи линейного программирования (1), (2) и ограничений $0 \leqslant x_j \leqslant 1$.

Во-вторых, каждое частичное решение удовлетворяет условиям целочисленности, но в отличие от метода, основанного на решении задач линейного программирования, может не удовлетворять линейным неравенствам (2). Применяя удачные правила выбора на шагах 1 и 4, с помощью аддитивного алгоритма можно найти допустимое по всем ограничениям и близкое к оптимальному решение на начальной итерации.

Если прекратить вычисления непосредственно перед тем, как алгоритм сходится (т. е. еще тогда, когда основной список еще не пуст), то можно установить, какая часть из 2^n возможных решений подверглась перебору (явному или неявному). Для определения этого показателя после шагов 2 и 3 всякий раз возвращаются к шагу 1, подсчитывая значение дроби $(^1/_2)^s$, где величина s равна числу переменных, содержащихся в частичном решении на шаге 1. В приведенном выше примере после завершения шага 3 по задаче 1 перебирается $(^1/_2)^1 = {^1/_2}$ из всех 2^5 вариантов решения. Далее, выполнив шаг 2 по задаче 3, осуществляют перебор еще $[(^1/_2)^2 = {^1/_4}]$ вариантов, так что всего уже перебрано $^3/_4$ вариантов. Оставшаяся часть $[(^1/_2)^2 = {^1/_4}]$ вариантов перебирается после решения задачи 4.

Как уже отмечалось несколько раз в этой главе, вычислительная трудоемкость алгоритмов тесно связана с количеством целочисленных переменных, входящих в задачу. В данном алгоритме объем вычислений определяется прежде всего числом задач, входящих в основной список. Заметим, что дополнительные ограничения (особенно составные) на самом деле могут оказаться полезными с точки зрения уменьшения объема вычислений Хотя нужно выполнить большее число проверок на шаге 2, гораздо чаще удается возвращаться непосредственно к шагу 1 без добавления новых задач в основной список.

Практический опыт реализации этого алгоритма (при условии его надлежащей модификации путем добавления сложных составных ограничений и при разработке эффективной программы для быстродействующей ЭВМ) показывает, что с его помощью можно решать некоторые практические задачи, содержащие до 100 переменных и до 50 ограничений. Такая размерность задач выглядит весьма скромной по сравнению с имеющейся возможностью решения задач линейного программирования, включающих несколько сотен переменных и ограничений. Однако наблюдается бурный прогресс в разработке методов решения целочисленных задач. Поэтому в ближайшем будущем можно ожидать, что удастся решать целочисленные задачи, также содержащие несколько сотен переменных.

Эффективные программы реализации аддитивного алгоритма на ЭВМ включают ряд оригинальных вариантов, в том числе упрощенное представление частичного решения, а также экономичный способ отображения основного списка. Кроме того, в этих программах обычно используются более эффективные проверки на шаге 2, чем (9) и (11), наряду с более рациональными правилами выбора на шагах 1 и 4, что позволяет избежать излишних ветвлений. Точно так же операции на шаге 3 иногда расширяются, для чего в них включается проверка получения допустимого решения, производимая после того, как каждая свободная переменная $x_j = 1$, если $c_j > 0$. В случае положительного результата проверки такое дополнение представляет собой наилучшее допустимое дополнение и его нужно зафиксировать в соответствии с описанием алгоритма.

Алгоритм частичного перебора можно модифицировать для решения частично целочисленных задач, в которых только переменные $x_j = 1$, $j = 1, 2, \ldots, p$ $(<n)$, должны быть равны нулю или единице, а на остальные переменные наложено ограничение $0 \leqslant x_j \leqslant 1$, $j = p+1, \ldots, n$. Частичное решение должно включать только целочисленные переменные, однако к свободным переменным относятся любые переменные, не вошедшие в частичное решение. Проверки (9) и (11) сохраняются, если не считать того, что проверка (11) применяется только к свободным целочисленным переменным. Кроме того, $b_0 = -x_0^t$ вместо $-x_0^t - 1$. Наконец, описание шага 5 изменяется следующим образом.

Шаг 3. Если (расширенное) частичное решение содержит все p целочисленные переменные, то нужно решить полученную задачу линейного программирования, отыскав тем самым соответствующие оптимальные значения остальных свободных переменных (при этом ограничения задачи линейного программирования формируются с учетом выбранных значений целочисленных переменных, вошедших в частичное решение). Если допустимого решения не существует, то принять $x_0^{t+1} = x_0^t$ и вернуться к шагу 1. В противном случае необходимо зафиксировать полученное оптимальное решение, принять x_0^{t+1} равным соответствующему оптимальному значению целевой функции и вернуться к шагу 1. Если (расширенное) частичное решение содержит менее p целочисленных переменных, перейти к шагу 4.

Алгоритмы, изложенные в данном разделе и в разд. 13.4, легко модифицировать для решения (по крайней мере в принципе) некоторых оптимизационных задач, содержащих непрерывные и целочисленные переменные наряду с нелинейной целевой функцией и нелинейными ограничениями. Рассмотрим в качестве примера задачу, где функции в (I) и (II) могут быть нелинейными.

$$\text{Максимизировать } c(x_1, x_2, \ldots, x_n) \qquad \text{(I)}$$

при ограничениях

$$a_i(x_1, x_2, \ldots, x_n) \leqslant 0, \qquad i = 1, 2, \ldots, m, \qquad \text{(II)}$$

$$x_j \text{ — целые и } L_j \leqslant x_j \leqslant U_j, \qquad j = 1, 2, \ldots, p, \ (\leqslant n), \qquad \text{(III)}$$

$$x_j \geqslant 0, \quad j = p+1, \ldots, n, \quad \text{если } p < n. \qquad \text{(IV)}$$

В гл. 14 и 15 будут рассматриваться способы решения таких задач, в которые не вводятся условия целочисленности. Если нелинейные функции удовлетворяют определенным требованиям, сформулированным для различных методов решения, приведенных в гл. 14 и 15, то целочисленные ограничения можно учесть непосредственно с помощью простой модификации метода, изложенного в разд. 13.4. В частности, можно найти оптимальное решение задачи (I) — (IV), применяя алгоритмы, описанные в гл. 14 и 15, и решая некоторую последовательность задач нелинейного программирования, каждая из которых содержит только непрерывные переменные, но отличается одна от другой границами целочисленных переменных.

Предположим теперь, что условия (III) и (IV) изменены следующим образом:

$$x_j = \begin{cases} 0 \\ 1 \end{cases} \text{при} \quad j = 1, 2, \ldots, n \tag{V}$$

и что любую функцию $a_i(x_1, x_2, x_n)$ можно записать в виде

$$a_i(x_1, x_2, \ldots, x_n) \equiv f_i(x_1, x_2, \ldots, x_n) - g_i(x_1, x_2, \ldots, x_n), \tag{VI}$$

где все $f_i(x_1, x_2, \ldots, x_n)$ и $g_i(x_1, x_2, \ldots, x_n)$ являются неубывающими монотонными функциями (т. е. $f_i(x_1, x_2, \ldots, x_n) \geqslant \geqslant f_i(y_1, y_2, \ldots, y_n)$, если $x_j \geqslant y_j$ при любом j; аналогично выражается это условие и для функций g_i). Тогда алгоритм, изложенный в этом разделе, можно непосредственно применить для решения рассматриваемой задачи. Некоторых изменений требуют лишь проверки на шаге 2.

Положим, например, что F_i есть значение функции $f_i(x_1, x_2, \ldots, x_n)$, где каждая переменная в частичном решении имеет определенное выбранное значение, а каждая свободная переменная принята равной нулю. Аналогично положим, что G_i есть значение функции $g_i(x_1, x_2, \ldots, x_n)$, где каждая переменная в частичном решении имеет определенное выбранное значение, а каждая свободная переменная принята равной нулю. Следует показать, что допустимого дополнения не существует, если

$$F_i - G_i > 0 \quad \text{при любом } i, \tag{VII}$$

где $i = 1, 2, \ldots, m$. Эта проверка сводится к (9) в случае, когда функция $a_i(x_1, x_2, \ldots, x_n)$ линейна. Если целевую функцию $c(x_1, x_2, \ldots, x_n)$ также можно представить в виде разности двух неубывающих монотонных функций

$$c(x_1, x_2, \ldots, x_n) \equiv -f_0(x_1, x_2, \ldots, x_n) + g_0(x_1, x_2, \ldots, x_n), \tag{VIII}$$

то проверку (VII) можно расширить, включив в нее условие

$$F_0 - G_0 + (x_0^t + 1) > 0, \quad i = 0. \tag{IX}$$

Пусть $F_{ik}(x_k)$ представляет собой функцию, полученную из $f_i(x_1, x_2, \ldots, x_n)$ следующим образом: каждая переменная в частичном решении имеет определенное выбранное значение, а каждая свободная переменная, кроме x_k, равна единице. Следует показать, что любая свободная переменная x_k дает допустимое дополнение при следующих условиях:

$$\left.\begin{array}{lll} \text{если } F_{ik}(0) - G_{ik}(0) > 0, & \text{то} & x_k = 1 \\ \text{если } F_{ik}(1) - G_{ik}(1) > 0, & \text{то} & x_k = 0 \end{array}\right\} \text{при любом } i. \tag{X}$$

Если выполняется условие (VIII), то (X) можно расширить, включив $i = 0$, когда к левой части добавляется константа $(x_0^t + 1)$ по аналогии с (IX). Проверка (X) упрощается, сводясь к (11) в случае, когда функция $a_i(x_1, x_2, \ldots, x_n)$ является линейной.

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

1. а) Вычислите количество циклов в задаче коммивояжера при числе городов, равном 3; то же, но при равном 5.

б) Рассмотрите задачу коммивояжера с пятью городами, но допустите, что коммивояжер должен вернуться в город 1, побывав в двух других городах, а затем посетить два оставшихся города. Определите число возможных последовательностей.

в) Определите число возможных последовательностей обработки трех деталей на одном станке, трех деталей на двух станках, двух деталей на трех станках, двух деталей на четырех станках. Покажите, как изменится результат, если детали должны обрабатываться в одном и том же порядке на каждом станке.

г) Рассмотрите задачу выбора номенклатуры продукции на одном интервале времени, в которой каждой операции x_j, $j = 1, 2, \ldots, n$, соответствуют затраты K_j на подготовку производства. Подсчитайте, сколько различных вариантов подготовки производства существует (предполагая, что равенство всех x_j нулю является одной из возможностей).

д) Рассмотрите ограничение $x_1 + x_2 + x_3 + \ldots + x_n = N$, где любая величина x_j должна быть неотрицательным целым. Определите число различных допустимых решений при следующих условиях: $n = 6$, $N = 1$; $n = 7$, $N = 1$; $n = 6$, $N = 2$; $n = 6$, $N = 6$; $n = 7$, $N = 6$; $n = 6$, $N = 6$ и ограничение задано в виде неравенства $\leqslant$; $n = 6$, $N = 1$ и ограничение задано в виде неравенства $\leqslant$; $n = 6$, $N = 2$ и ограничение задано в виде неравенства $\leqslant$.

е) Рассмотрите задачу распределения капиталовложений, в которой фигурируют 3 группы проектов. В группе 1 имеется восемь проектов и нужно выбрать два, в группе 2 — десять проектов и нужно выбрать четыре, в группе 3 — семь проектов и нужно выбрать пять. Определите общее число возможных комбинаций выбора проектов. Покажите, как изменится это число, если заданное число проектов, выбираемых в каждой группе, является лишь верхней оценкой допустимого выбора.

ж) Фирме нужно составить расписание движения трех автомашин, обслуживающих восемь магазинов. Каждая машина может останавливаться у одного или нескольких магазинов, но ни в один магазин не допускается прибытия более одной машины. Определите число различных расписаний обслуживания магазинов.

з) Фирма выполняет предупредительный ремонт 4 крупных объектов. Работу на объекте 1 можно начинать в любую из трех недель, на объекте 2 — в любую из шести недель, на объекте 3 —

в любую из двух недель и на объекте 4 — в любую из пяти недель. Подсчитайте общее число различных календарных планов ремонта (т. е. «недель, в которых начинается работа», на каждом из четырех объектов).

и) Фирма разрабатывает стратегию организации сбыта трех новых видов продукции в четырех районах. Стратегия строго «последовательна» в том смысле, что сначала выбирается один район, а затем порядок выпуска на рынок продукции каждого вида (по одному виду через каждые две недели), далее выбирается следующий район и порядок выпуска на рынок всех видов продукции и т. д. Подсчитайте общее число различных стратегий организации сбыта новой продукции.

к) Фирма планирует строительство шести новых предприятий в предварительно выбранных пунктах. Строительство всего этого комплекса рассчитано на пять лет. Допускается возможность строительства всех предприятий как в течение первого года, так и в течение пятого года. Более правдоподобно предположить, что какие-то предприятия будут строиться в течение каждого года. Определите число возможных вариантов строительства этих предприятий при этом условии. Фирму интересует также число предприятий, строящихся ежегодно (в дополнение к числу предприятий, строившихся в предыдущем году). Определите все возможные варианты строительства предприятий в рамках пятилетнего интервала времени (например, шесть в течение первого года и ни одного в последующие годы или шесть в течение пятого года и ни одного в предыдущие и т. д.).

2. Рассмотрите задачу

$$\text{Максимизировать} \quad 3x_1 + 6x_2 + x_3$$

при ограничениях

$$x_1 + 2x_2 + 2x_3 \leqslant 2\,\frac{2}{3},$$

$$x_1 + 2x_2 + 3x_3 \leqslant 2\,\frac{1}{2},$$

$$\text{любое } x_j \text{ — неотрицательное целое.}$$

а) Каково оптимальное решение, если снять условие целочисленности на каждую переменную x_j? Существует ли более одного оптимального решения?

б) Определите путем простого подбора оптимальное целочисленное решение. Округляется ли ответ, полученный в п. а) до этого решения?

в) Представьте задачу графически в пространстве решений, приняв $x_3 = 0$ (т. е. начертите ограничения на x_1 и x_2). Укажите все допустимые целочисленные решения.

3. Предположим, что фирма имеет пять предприятий, снабжающих 100 складов. Если фирма эксплуатирует все предприятия и склады, то перед ней возникает задача минимизации затрат, имеющая вид классической транспортной задачи, описанной в начале гл. 6.

а) Допустим, однако, что общая производственная мощность предприятий значительно превышает потребности складов и что при работе каждого предприятия i приходится нести крупные ежегодные условно-постоянные расходы в размере P_i. Предложите *практически* оправданный способ определения целесообразности остановки некоторых предприятий (если такая мера действительно целесообразна), минимизирующий сумму ежегодных условно-постоянных и транспортных затрат.

б) Предположим теперь, что фирма хочет закрыть некоторые склады (при этом спрос соответствующих потребителей будут удовлетворять оставшиеся склады). Таким образом, фирма экономит на условно-постоянных затратах эксплуатации некоторых складов, но при этом возрастают транспортные расходы. Будет ли при этом решение, предложенное в п. а), эффективным? Дайте необходимые пояснения.

4. Задача фирмы «Прогресс» (разд. 13.2). Рассмотрите следующие исходные данные:

$$m = 3, \quad n = 4, \quad S_1 = 40, \quad S_2 = 50, \quad S_3 = 60,$$
$$D_1 = 10, \quad D_2 = 15, \quad D_3 = 20, \quad D_4 = 25.$$

а) Выпишите подробно все линейные ограничения задачи.

б) Объясните, почему ограничения (7) и (8) совместно гарантируют равенство каждой из величин y_i и z_{ij} либо 0, либо 1.

в) Объясните, почему с точки зрения оптимальности решения не существенно, чтобы выполнялись условия $z_{ij} = 1$ и $x_{ij} = 0$.

г) Предположим, что все $F_i = 100$, $c_{ij} = 1$, $F_{ij} = 0$. Каково оптимальное решение, если снять условие целочисленности? Существуют ли альтернативные (различные) оптимальные решения?

д) Объясните, почему величина $U_{ij} = \min(S_i, D_j)$ в (9) достаточно велика.

5. Рассмотрите задачу размещения предприятий, описанную в разд. 13.2.

а) Если каждая величина $F_i = 0$, то следует ли в оптимальном решении положить каждую величину $y_i = 1$? Дайте необходимые пояснения.

б) Предположим, что все $y_i = 1$ и что, кроме того, общие поставки S_i равны суммарному спросу D_j. Будет ли в оптимальном решении обязательно использовано $m + n - 1$ маршрутов? Если все $F_{ij} = F$ (некоторая константа), то совпадает ли оптимальное решение с оптимальным решением классической транспортной задачи, в которой величины F_{ij} отсутствуют ($F = 0$)? (В случае положительного ответа на этот вопрос дайте необходимое обоснование, в случае отрицательного — приведите небольшой пример, подтверждающий справедливость такого ответа.)

6. Задача с постоянными платежами. Рассмотрим задачу

$$\text{Минимизировать } \sum_{j=1}^{n} c_j(x_j)$$

при ограничениях

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

$$\text{любое } x_j \geqslant 0,$$

где
$$c_j(x_j)=\begin{cases} 0 & \text{при} \quad x_j=0, \\ K_j+c_jx_j & \text{при} \quad x_j>0, \end{cases}$$

а любое $K_j \geqslant 0$. Величины K_j называются постоянными платежами.

а) Покажите, как сформулировать эту задачу в виде частично целочисленной задачи линейного программирования.

б) Обсудите затруднение, которое может возникнуть, если целью является не минимизация, а максимизация (или если не все $K_j \geqslant 0$). (*Указание*: насколько малым может быть значение x_j при условии, что это величина положительная?)

в) Предположим, что все $K_j = K > 0$. Сформулируйте достаточное условие того, чтобы оптимальное решение задачи совпадало с оптимальным решением задачи линейного программирования, в которой величины K_j отсутствуют (все $K_j = 0$).

г) Покажите, как следует изменить это условие, если ограничения имеют вид

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

где
$$a_{ij}(x_j)=\begin{cases} 0 & \text{при} \quad x_j=0, \\ A_{ij}+a_{ij}x_j & \text{при} \quad x_j>0, \end{cases}$$

а любое $A_{ij} \geqslant 0$.

д) Обсудите затруднение, которое может возникнуть в п. г) в том случае, когда некоторая величина $A_{ij} < 0$ или линейное ограничение имеет вид неравенства $\geqslant$.

7. Рассмотрите задачу выбора номенклатуры продукции, подобную задаче, описанной в разд. 2.2, имеющую вид

$$\text{Максимизировать } \sum_{j=1}^{n} c_jx_j$$

при ограничениях

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

$$\text{любое } x_j \geqslant 0,$$

где любое $c_j \geqslant 0$ и $n \geqslant m$.

а) Предположим, что фирма ставит перед собой цель выпускать не более k различных видов продуктов. Покажите, как выразить это условие через целочисленные переменные.

б) Какие затруднения возникают, если фирма стремится выпускать по крайней мере k различных видов продуктов. (*Указание*: насколько мало может быть значение величины x_j, если она положительна?)

8. Представление задачи в пространстве решений. В каждом из указанных ниже пунктов дайте чертеж области допустимых решений для указанных различных наборов ограничений, предполагая $x_1 \geqslant 0$, $x_2 \geqslant 0$.

а) Либо $x_1 + 2x_2 \leqslant 1$, либо $2x_1 + x_2 \leqslant 1$, либо оба эти ограничения совместно.

б) Либо $x_1 + x_2 \leqslant 1$, либо $x_1 + x_2 \geqslant 2$, либо оба эти ограничения совместно.

в) Либо пара ограничений $x_1 \leqslant 1$ и $x_2 \leqslant 1$, либо пара $x_1 \geqslant 1$ и $x_2 \geqslant 1$, либо обе эти пары совместно.

г) Должно выполняться по крайней мере одно из следующих ограничений: $x_1 \leqslant 1$, $x_2 \leqslant 1$ и $x_1 + x_2 \leqslant 2$. Должны выполняться по крайней мере два из этих ограничений. Должны выполняться все три ограничения.

9. Рассмотрите подход к представлению альтернативных ограничений, описанный в разд. 13.2 [выражения (12) — (14)]. В каждом из указанных ниже пунктов подробно запишите (13) и (14). (*Указание*: анализируя ограничения, вычислите значение U_i, примите значение U_i равным минимуму, при котором получается нужный результат.)

а) $$x_1 + 2x_2 + x_3 \leqslant 8,$$
$$4x_1 + x_2 + x_3 \leqslant 20,$$
$$x_1 + x_2 + 3x_3 \leqslant 15.$$

Покажите, как записываются эти условия при $k = 1$ и $k = 2$.

б) Возьмите те же ограничения, что и в п. а), добавив к ним $x_2 - x_4 \leqslant 0$. Покажите, как записываются эти условия при $k = 3$.

в) Объясните подробно, почему полученные в пп. а) и б) представления наряду с условием, что любая величина y_i равна либо 0, либо 1, дают решение задачи (т. е. обеспечивают выполнение по крайней мере k из p ограничений).

г) Объясните, почему выражение (13) можно представить также в виде равенства.

д) Обсудите трудность, которая возникает в случае, когда нужно обеспечить, чтобы выполнялось *в точности* k из p линейных ограничений (12).

10. В каждом из указанных ниже пунктов сформулируйте задачу в терминах целочисленного программирования.

а) Максимизировать $\sum\limits_{j=1}^{n} c_j x_j$

при ограничениях

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

$$\text{любое } x_j \geqslant 0$$

и выполняются либо три ограничения $x_1 \leqslant 1$, $x_2 \leqslant 1$ и $x_1 + x_2 \leqslant 1{,}5$, либо три ограничения $x_1 \geqslant 1$, $x_2 \geqslant 1$ и $x_1 + x_2 \geqslant 2{,}5$.

б) Максимизировать $[\max(x_1 + x_2, 2x_1 + x_2, x_1 + 2x_2)]$ при ограничениях

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

$$\text{любое } x_j \geqslant 0$$

(т. е. найти значения x_j, удовлетворяющие ограничениям и обеспечивающие максимальное возможное значение трем указанным линейным функциям).

в) Каким образом следует отыскивать числовое решение задач, приведенных в пп. а) и б)?

11. В упражнении 1 этой главы, п. к), описана фирма, планирующая строительство шести новых предприятий в заранее выбранных пунктах при условии, что все эти предприятия должны быть введены в эксплуатацию в течение пяти лет.

а) Сформулируйте задачу выбора, пользуясь линейными ограничениями, включающими целочисленные переменные. (Обязательно точно определите введенные символы и объясните смысл каждого ограничения.) В каждом из указанных ниже пунктов покажите, как можно выразить сформулированное условие в целочисленном виде. Рассмотрите каждый пункт по отдельности.

б) Предприятия 1, 2 и 3 должны быть введены в строй не позднее третьего года.

в) Можно вводить в строй не более двух предприятий в течение любого года.

г) Можно вводить в строй не более трех предприятий в течение первых двух лет и не более пяти предприятий в течение первых четырех лет.

д) Нужно ввести в эксплуатацию в точности три предприятия в течение первых двух лет и в точности пять предприятий в течение первых четырех лет.

е) Предприятие 2 нельзя вводить в строй до ввода предприятия 1 (оба предприятия можно ввести в одном и том же году).

ж) Предприятия 4, 5 и 6 нельзя вводить в строй до ввода предприятий 1, 2 и 3 (все шесть предприятий можно ввести в одном и том же году).

з) Предприятия 1, 2, 3 должны быть введены в строй до ввода предприятий 4, 5, 6 или наоборот (все шесть предприятий нельзя вводить в одном и том же году).

и) Предприятия 1 и 2 должны быть введены в эксплуатацию в одном и том же году. Аналогично предприятия 3, 4 и 5 нужно ввести в одном и том же году.

к) Если предприятия 1 и 2 введены в эксплуатацию в одном и том же году, то в этот год нельзя вводить никаких других предприятий.

л) В первом году фирма вводит в эксплуатацию либо предприятия 1 и 2, либо 3 и 4. (Предполагается, что один из этих вариантов должен обязательно иметь место.)

м) В течение первых двух лет фирма вводит в эксплуатацию либо предприятия 1 и 2, либо 3 и 4. (Предполагается, что один из этих вариантов должен обязательно иметь место).

н) Предприятие 1 должно быть введено в строй в течение первого года только при условии, что введено предприятие 2 или 3, но не оба они вместе.

о) Предположим, что для строительства предприятия 1 возможен выбор двух пунктов (вместо одного уже выбранного). Покажите, как нужно изменить предшествующие постановки задач, чтобы учесть в них возможность такого выбора.

п) Предположим, что вместо шести пунктов в распоряжении фирмы имеется всего пять, причем фирма может строить в одном определенном пункте либо предприятие 1, либо 3. Покажите, как следует изменить предшествующие постановки, чтобы учесть в них возможность такого выбора. (Может возникнуть необходимость изменения допущений, принятых в некоторых вышестоящих пунктах. В таких случаях эти изменения следует оговорить особо.)

12. Рассмотрим задачу коммивояжера с $n = 4$ городами.

а) Используя постановку задачи (19) — (23) из разд. 13.2, выпишите подробно все линейные ограничения.

б) Покажите, что цикл город 1 — город 3 — город 2 — город 4 — город 1 приводит к значениям x_{ij} и u_i, удовлетворяющим всем ограничениям п. а).

в) Покажите, что подцикл город 3 — город 2 — город 3 не может приводить к значениям x_{ij} и u_i, удовлетворяющим всем ограничениям п. а). Решите такую же задачу для подцикла город 1 — город 2 — город 3 — город 1.

13. Рассмотрим задачу

$$\text{Максимизировать } 3x + 7y$$

при ограничениях

$$2x + y \leqslant 25,$$
$$x + 2y \leqslant 6,$$

где $y \geqslant 0$, а x может принимать только значения 0, 1, 4, 6.

а) Сформулируйте эквивалентную задачу в терминах целочисленного программирования.

б) Предположим, что вместо члена $3x$ целевая функция содержит член $3x^2$. Измените соответствующим образом ответ п. а).

14. Рассмотрим задачу с линейными ограничениями из упражнения 13, в которой $x \geqslant 0$ и $y \geqslant 0$, но предположим, что целевая функция сводится к максимизации $c(x) + 7y$, где

$$c(x) = \begin{cases} 4x & \text{при } 0 \leqslant x \leqslant 2, \\ 14 - 3x & \text{при } 2 \leqslant x \leqslant 3, \\ 5 & \text{при } 3 \leqslant x \leqslant 3{,}5, \\ -2 + 4x & \text{при } 3{,}5 \leqslant x \leqslant 5, \\ -22 + 8x & \text{при } 5 \leqslant x \leqslant 6. \end{cases}$$

а) Начертите график функции $c(x)$.

б) Сформулируйте эквивалентную задачу в терминах целочисленного программирования, используя постановку (24) — (30), приведенную в разд. 13.2.

в) Объясните, почему ограничения (29) и (30) обеспечивают равенство единице в точности одной величины z_k, гарантируя тем самым, что в выражении (28) будет не более двух весовых положительных коэффициентов и что в соответствии с (25) их сумма должна быть равна 1. Используйте постановку задачи, данную в п. б), для обоснования ваших рассуждений.

г) Не учитывая ограничений (28) — (30), определите оптимальное решение. Является ли это решение действительно оптимальным для исходной нелинейной задачи? Дайте необходимые пояснения.

д) Покажите, как изменится постановка задачи п. б), если на интервале $0 \leqslant x \leqslant 2$ функция $c(x)$ вместо $4x$ имеет выражение $4x - 2$. (*Указание*: прежде чем дать аналитический ответ, начертите график функции.)

15. Обозначим, как в разд. 13.4, символом $[d]$ наибольшее целое, меньшее или равное действительному числу d.

а) Каковы значения выражений

$$[12{,}8], \quad [12], \quad [-12{,}8] \quad [-12]$$

$$[0{,}8], \quad [0], \quad [-0{,}8]$$

$$\left[3\frac{3}{5}\right], \quad [3], \quad \left[\frac{3}{5}\right] \quad \left[-\frac{3}{5}\right], \quad [-3], \quad \left[-3\frac{3}{5}\right].$$

б) Приведите значения дробных частей выражений, стоящих в скобках в п. а).

16. а) Объясните подробно, на основании чего можно записать неравенства (10) и (11) и равенство (12) из разд. 13.4, а также принять условие, что величина x в (12) должна быть целой.

б) Приведите выкладки, обосновывающие утверждение, что вычитание выражения (8) из (12) дает (14).

17. Рассмотрим числовой пример из разд. 13.4. Предположим, что постоянная в правой части выражения (2) равна не 3, а 5.

а) Каково оптимальное решение задачи линейного программирования (пренебрегая условиями целочисленности)?

б) Используйте идею метода отсечения (не применяя формального алгоритма) для отыскания оптимального целочисленного решения.

18. Рассмотрим задачу

$$\text{Максимизировать} \quad x_1 + 2x_2$$

при ограничениях

$$5x_1 + 7x_2 \leqslant 21,$$

$$-x_1 + 3x_2 \leqslant 8,$$

$$x_1 \text{ и } x_2 \text{ — неотрицательные целые.}$$

а) Каково оптимальное решение задачи линейного программирования (если пренебречь условиями целочисленности)?

б) Используйте идею метода отсечения (не применяя формального алгоритма) для определения оптимального целочисленного решения.

19. При решении числового примера (20) — (22) из разд. 13.4 с помощью алгоритма отсечения покажите, что справедливы следующие утверждения:

а) ограничение (14), примененное к строке 1 в (24), приводит к ограничению (25);

б) задача линейного программирования (24), расширенная с помощью (25), имеет оптимальное решение, приведенное в (26);

в) отсечение, построенное из строки 2 в (26), определяется (27);

г) оптимальное решение задачи линейного программирования (26) и (27) определяется (28);

д) дайте обоснование справедливости утверждения, что решение (28) оптимально для исходной задачи (20), (21) и (22).

20. Примените алгоритм отсечения для решения числового примера (20) — (22) из разд. 13.4, приняв следующие условия:

а) постоянная в правой части выражения (21) равна не 13, а 15;

б) коэффициент при x_1 в (21) вместо 7 равен 6.

21. В конце разд. 13.4 показано, как выражать отсечения через переменные исходной задачи и как графически отображать работу алгоритма. Опишите отсечения через такие переменные и дайте соответствующие графические изображения для обоих пунктов упражнения 20.

22. Исследуйте применение алгоритма ветвей и границ к примеру (8), (9) из разд. 13.5. Проследите подробно все вычисления, выполняемые на каждой итерации. В частности, покажите, что решения в (11), (13), (15), (17), (19) и (20) оптимальны. Проверьте, как формируются ограничения в каждой из девяти задач. Покажите, что задачи 2, 6 и 8 не имеют допустимого решения. Объясните подробно правила, позволяющие прекратить вычисления после решения задачи 5, и изложите основания для утверждения, что (19) является оптимальным решением.

23. Рассмотрите пример, приведенный в разд. 13.5. Предположим, что на итерации 4 вместо задачи 4 выбрана задача 5. Выполните вычисления с помощью алгоритма ветвей и границ при таком выборе. Постройте дерево работы алгоритма, аналогичное дереву, показанному на рис. 13.1.

24. Рассмотрите пример, приведенный в разд. 13.5, изменив в нем постоянные в правой части обоих выражений (9) с 8 на 22. Примените алгоритм ветвей и границ для решения задачи. Постройте дерево работы алгоритма, аналогичное дереву, показанному на рис. 13.1.

25. Рассмотрите, как решена задача коммивояжера *методом исключения подциклов* в разд. 13.6. Предположим, что на первой итерации вводятся три задачи, которым соответствует подцикл $x_{23} = x_{34} = x_{42} = 1$ *вместо* подцикла $x_{15} = x_{51} = 1$. Выполните оставшиеся вычисления и постройте дерево работы алгоритма, аналогичное дереву, изображенному на рис. 13.3.

26. Рассмотрите, как решена задача коммивояжера методом задания маршрутов в разд. 13.6. Правило выбора на шаге 1 предусматривало выбор из оставшихся в списке задач задачи с наибольшим номером. Поэтому на итерации 1 была выбрана задача 2. Реализуйте тот же алгоритм, применив на шаге 1 правило выбора задачи с наименьшим номером из числа оставшихся в списке. (Следовательно, на итерации 1 начните с задачи 1.) На шаге 4 примените прежнее условие присвоения наибольшего номера задаче, в которой $x_{hk} = 1$. Постройте дерево работы алгоритма, аналогичное рис. 13.5.

27. Примените алгоритм задания маршрутов к задаче коммивояжера, приведенной в качестве примера в разд. 13.6, положив, что в задаче 1 $c_{15} = \infty$, а в задаче 2 $x_{15} = 1$. (Воспользуйтесь теми же самыми правилами выбора на шаге 1 и нумерации на шаге 4, которые приведены на рис. 13.5.) Постройте дерево работы алгоритма, аналогичное дереву, изображенному на рис. 13.5.

28. Выполните задание упражнения 27, приняв, что задачи 1 и 2 относятся к c_{51} и $x_{51} = 1$ соответственно.

29. Рассмотрите, как решена задача коммивояжера в разделе 13.6 с помощью алгоритма частичных циклов. При условии что цикл проходит через все города, можно в качестве «начального пункта» взять любой город. Примите, что цикл начинается в городе 2, и примените алгоритм. Постройте дерево работы алгоритма, аналогичное дереву, изображенному на рис. 13.6.

30. Предположим, что целочисленная переменная x в оптимальном решении принимает значение в интервале $0 \leqslant x \leqslant U$, где U — целое. Укажите способ представления величины x, аналогичный (4) из разд. 13.7 и такой, что максимально возможное значение x в этом представлении в точности равно U.

31. Рассмотрите задачу целочисленного программирования, подобную задаче, описанной в начале разд. 13.7. Предположим, что одно из линейных ограничений (2) задано выражением

$$x_1 - x_2 + 7x_3 - 10x_4 + 3x_5 - 6x_6 < 0.$$

В каждом из указанных ниже пунктов примените проверки (9) и (11) для установления существования допустимого дополнения к указанному частичному решению, а также для определения того, что одна или более переменных должны принимать строго определенные значения

а) $x_4 = x_6 = 0$,
б) $x_1 = x_3 = 1$ и $x_4 = 0$,
в) $x_1 = x_3 = 1$ и $x_6 = 0$,
г) $x_1 = x_3 = x_5 = 1$,
д) $x_1 = x_2 = x_5 = 1$ и $x_6 = 0$,
е) $x_3 = 1$ и $x_2 = x_6 = 0$.

32. Рассмотрите задачу целочисленного программирования, аналогичную описанной в начале 13.7. Предположим, что в частичном решении значения переменных $x_6, x_7, \ldots, x_n$ фиксированы, а оставшиеся члены в максимизируемой целевой функции $2x_1 - 3x_2 + 6x_3 - 7x_4 + 10x_5$. Примем также, что $x_0^t = 7$. В каждом из указанных ниже пунктов примените проверки (9) и (11) для выяснения того, может ли заданное расширение частичного решения привести к полному решению, обеспечивающему улучшение значения целевой функции, и должны ли одна или более переменных принять определенные значения, чтобы улучшить значение $x_0^t = 7$.

а) $x_4 = 1$,
б) $x_5 = 0$,
в) $x_1 = x_2 = 1$.

33. Рассмотрите применение алгоритма частичного перебора к примеру (12) — (15) из разд. 13.7. Проверьте все промежуточные вычисления на каждом шаге. Покажите подробно все проверки.

34. Рассмотрите пример (12) — (15), приведенный в разд. 13.7. Примените алгоритм частичного перебора в случае, когда задача 1 является частичным решением $x_1 = 1$, а задача 2 — частичным решением $x_1 = 0$. Положите $x_0^1 = 0$, что соответствует всем $x_j = 0$. Постройте дерево работы алгоритма, аналогичное дереву, изображенному на рис. 13.7.

35. Объясните, как вы понимаете термины:

допущение делимости;
задача целочисленного (диофантового, дискретного) программирования;
полностью целочисленная задача;
частично целочисленная задача;
комбинаторная оптимизационная задача;
алгоритм возврата;
выпуклая оболочка;
выпуклая комбинация;
целая часть;
дробная часть;
отсечение;
сплошная вырожденность;
алгоритм ветвей и границ;
алгоритм частичного (неявного)

перебора;
аддитивный алгоритм;
двоичное представление;
частичное решение;
свободные переменные;
дополнение;
прозондированное частичное решение;
составные (замещающие) ограничения;
точка (вершина) решетки;
булева (двоичная) переменная;
альтернативные ограничения;
альтернативы «да — нет» («принято — отвергнуто»);
задача коммивояжера;
циклы;
подциклы и частичные циклы;
алгоритм отсечения (метод целочисленных форм).

УПРАЖНЕНИЯ НА РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

36. Задача о ранце. Рассмотрим следующий вариант задачи о ранце, описанный в упражнениях 22 и 46 гл. 10:

$$\text{Максимизировать } 60x_1 + 60x_2 + 40x_3 + \\ + 10x_4 + 20x_5 + 10x_6 + 3x_7$$

при ограничениях

$$3x_1 + 5x_2 + 4x_3 + 1x_4 + 4x_5 + 3x_6 + 1x_7 \leqslant 10,$$
$$\text{все } x_j = 0,\ 1.$$

а) Решите задачу с помощью алгоритма отсечения, приведенного в разд. 13.4.

б) Решите задачу с помощью алгоритма ветвей и границ, приведенного в разд. 13.5.

в) Решите задачу с помощью алгоритма неполного перебора, приведенного в разд. 13.7.

г) Решите задачу с помощью рекуррентного метода динамического программирования.

37. Рассмотрим доводы, изложенные в разд. 13.4, которые устанавливают справедливость выражений (11) и (12), определяющих отсечение при заданном уравнении (8).

а) Выведите выражение для отсечения, аналогичное (11) и (12) в случае, когда уравнение (8) умножается на константу ($p \neq 0$).

б) Предположим, что (8) имеет вид $2\frac{1}{2}x_1 + 3\frac{1}{4}x_2 = 6\frac{5}{16}$. Какое неравенство для отсечения получается в следующих случаях: $p = 1$, $p = 1/2$, $p = 2$, $p = 2\frac{1}{2}$, $p = 3\frac{1}{4}$? Являются ли все эти отсечения одинаково сильными? Дайте необходимые пояснения.

в) Пусть символ $\langle d\rangle$ обозначает наименьшее целое, превосходящее или равное по значению действительному числу d. Покажите, что из (8) можно вывести соотношение

$$\sum_{j=1}^{n} \langle a_j\rangle x_j \geqslant \langle b\rangle.$$

38. Примените алгоритм отсечения, приведенный в разд. 13.4, для решения нижеследующих задач, предполагая, что все переменные должны быть неотрицательными целыми.

а) Задача «двух картошек», приведенная в разд. 1.6. (*Указание*: положите коэффициенты при дополняющих переменных равными 0, 1, обеспечивая тем самым целочисленность этих переменных.)

б) Задача фирмы «Мультиконвейер», описанная в разд. 2.2 и решенная симплексным методом в разд. 4.4.

в) Задача фирмы «Большая энергетика», описанная в конце разд. 10.3. (Используйте данные упражнения 30 из гл. 10.)

г) Задачи из упражнения 10, пп. б), в), д), е) гл. 3.

д) Задача из упражнения 20, п. а) гл. 4.

е) Пример (8) — (9) из разд. 13.5.

ж) Максимизировать $x_1 + x_2$ при ограничениях

$$4x_1 - x_2 \leqslant 10,$$
$$2x_1 + 5x_2 \leqslant 10,$$
$$4x_1 - 3x_2 \leqslant 6.$$

з) Максимизировать $10x_1 - x_2$ при ограничениях

$$-25x_1 + 22x_2 \leqslant 33,$$
$$20x_1 - 5x_2 \leqslant 24.$$

и) Выразите отсечения в пп. а), г), е) и ж) через переменные исходной задачи, пользуясь методом, изложенным в конце раздела 13.4. Покажите работу алгоритма в графическом виде в пространстве решений.

39. Примените алгоритм ветвей и границ, изложенный в разд. 13.5, для решения нижеследующих задач, предполагая, что все переменные должны быть неотрицательными целыми. Постройте дерево работы алгоритма, аналогичное дереву, изображенному на рис. 13.1. Для решения задач большой размерности можно использовать алгоритм с переменными, ограниченными сверху (с двусторонним ограничением), описанный в разд. 5.10.

а) Задача «двух картошек», приведенная в разд. 1.6.

б) Задача фирмы «Мультиконвейер», описанная в разд. 2.2 и решенная в разд. 4.4 симплексным методом.

в) Задача, приведенная в упражнении 18, п. а) гл. 10.

г) Задача, приведенная в упражнении 18 гл. 10, в которой постоянная в правой части первого уравнения равна не 3, а 2.

д) Задача, приведенная в упражнении 18 гл. 10, в которой постоянная в правой части второго уравнения равна не 4, а 3; не 4, а 5.

е) Задача, приведенная в упражнении 18 гл. 10, в которой постоянная в правой части первого уравнения равна не 3, а 2, а в правой части второго уравнения — не 4, а 3.

ж) Задача фирмы «Большая энергетика», описанная в конце разд. 10.3. (Используйте данные из упражнения 30 гл. 10.)

з) Задача, приведенная в упражнении 20, п. а) гл. 4.

и) Пример (20) — (22), приведенный в разд. 13.4.

40. Примените алгоритм частичного перебора, описанный в разделе 13.7, для решения нижеследующих задач. Если задача не выражена в двоичных (булевых) переменных, преобразуйте ее к этому виду, прежде чем пользоваться алгоритмом. Постройте дерево работы алгоритма, аналогичное дереву, изображенному на рис. 13.7.

а) Модифицированная задача фирмы «Мультиконвейер», описанная в разд. 2.2 и решенная симплексным методом в разд. 4.4:

Максимизировать $4x_1 + 28x_2 + 5x_3 + 9x_4 + 63x_5 + 11x_6$
при ограничениях

$$1x_1 + 7x_2 + 1x_3 + 1x_4 + 7x_5 + 1x_6 \leqslant 15,$$
$$7x_1 + 49x_2 + 5x_3 + 3x_4 + 21x_5 + 2x_6 \leqslant 120,$$
$$3x_1 + 21x_2 + 5x_3 + 10x_4 + 70x_5 + 15x_6 \leqslant 100,$$
$$\text{все } x_j = 0,\ 1.$$

б) Задача, приведенная в упражнении 18, п. а) гл. 10.

в) Та же задача, в которой постоянная в правой части первого уравнения равна не 3, а 2.

г) Та же задача, в которой постоянная в правой части второго уравнения равна 4, а не 3; равна 5, а не 4.

д) Та же задача, в которой постоянная в правой части первого уравнения равна 2, а не 3, а в правой части второго уравнения равна 4, а не 3.

е) Задача фирмы «Мультиконвейер», описанная в конце раздела 10.3. (Воспользуйтесь данными упражнения 30 гл. 10.)

ж) Задача, приведенная в упражнении 20, п. а) гл. 4.

з) Пример (20) — (22) из разд. 13.4.

и) Задача, приведенная в упражнении 38, п. ж) этой главы.

к) Задача, приведенная в упражнении 38, п. з) этой главы. (Принять $0 \leqslant x_j \leqslant 4$.)

л) Максимизировать $-4x_1 + 5x_2 + x_3 - 3x_4 + x_5$ при ограничениях

$$-1x_1 + 2x_2 \qquad - 1x_4 - 1x_5 \leqslant -2,$$
$$-4x_1 + 5x_2 + 1x_3 - 3x_4 + 1x_5 \leqslant -2,$$
$$2x_1 - 3x_2 - 2x_3 + 6x_4 - 2x_5 \leqslant 1,$$
$$\text{все } x_j = 0,\ 1.$$

41. Рассмотрите задачу коммивояжера со следующими исходными данными:

$$c_{12} = 20, \quad c_{13} = 4, \quad c_{14} = 10,$$

$$
\begin{array}{lll}
c_{23} = 5 & & c_{25} = 10, \\
 & c_{34} = 6, & c_{35} = 6, \\
 & & c_{45} = 20,
\end{array}
$$

где $c_{ij} = c_{ji}$ и маршрут между городами i и j отсутствует, если соответствующее значение c_{ij} не приведено в условии. Решите эту задачу следующими методами:

а) исключения подциклов,

б) помеченных маршрутов,

в) частичных циклов.

42. Примените алгоритм частичного перебора для решения следующей нелинейной задачи с двоичными переменными, воспользовавшись методом, описанным в конце разд. 13.7:

$$
\text{Максимизировать} \quad 6x_1 - 2x_1^2 + x_1x_2 - x_1x_2^2 + 2x_1x_3 - 5x_2^3 - \\ -x_2x_3^3 + 4x_3^4 + 10x_4
$$

при ограничениях

$$
\begin{aligned}
3x_1^2 - 3x_1x_2 + x_2^2 &\leqslant 5, \\
x_2 + 4x_3^3 + x_3^2x_4 - 6x_4 &\geqslant 3, \\
7x_1^3 - x_3^2 + 3x_4 &\leqslant 10, \\
\text{все} \quad x_j &= 0,\ 1.
\end{aligned}
$$

УПРАЖНЕНИЯ НА ПОСТРОЕНИЕ МОДЕЛЕЙ

В упражнениях 43—62 нужно построить модели целочисленного программирования. Обязательно дайте точное определение всех используемых символов, приведите обоснование каждого ограничения и запишите целевую функцию. Если в упражнении имеются числовые данные, обязательно полностью используйте их при записи модели.

43. Отдел стипендий и оказания материальной помощи студентам одного из университетов готовит предложения относительно назначения стипендий на следующий год. В качестве будущих стипендиатов выбрано n студентов, причем студенту i предполагается назначить стипендию не менее M_i долл., $i = 1, 2, \ldots, n$. В распоряжении отдела имеется s различных стипендий. Стипендия j дает стипендиату a_j долл. Отдел может оказаться вынужденным назначить отдельным студентам по несколько стипендий, чтобы они получали не менее чем по M_i долл., но руководство отдела не имеет права уменьшить размер каждой стипендии ниже фиксированной суммы a_j. Если стипендия j на рассматриваемый год вообще не назначается, то к величине a_j прибавляются проценты, и эта стипендия может быть назначена в увеличенном размере в следующем году.

а) Постройте модель назначения стипендий, максимизирующую сумму сэкономленных денег при условии, что каждый студент получит по крайней мере M_i долл.

б) Покажите, как нужно изменить ответ п. а), если каждому студенту i не разрешается получать более двух различных стипендий, а общая сумма его стипендиальных доходов не может превышать $1{,}1M_i$.

44. Администрация штата объявила торги на n строительных подрядов для n фирм. Ни с одной фирмой не заключается более одного контракта, в связи с чем задачу можно рассматривать как задачу о назначениях, описанную в разд. 6.4. По политическим соображениям чиновники администрации стремятся не заключать более N крупных контрактов с фирмами, расположенными за пределами штата. Обозначим через $1, 2, \ldots, s$ крупные контракты, а через $1, 2, \ldots, t$ — фирмы, расположенные за пределами штата. Целью является минимизация общих затрат при указанном условии.

а) Сформулируйте оптимизационную задачу.

б) Покажите, что при введении дополнительного условия в задачу о назначениях (3) — (4) из разд. 6.4 и отыскании решения задачи линейного программирования (в которой целочисленные ограничения не учитываются) допустимое назначение получается не всегда. (*Указание*: рассмотрите следующий случай: $n = 3$, $s = t = 2$, $N = 1$, все $c_{ii} = 0$, $c_{13} = c_{31} = 1$, а все остальные $c_{ij} = 2$.)

45. Многоиндексная задача. Губернатор штата (см. предыдущее упражнение) включает в правила заключения контрактов дополнительное условие, предусматривающее, что в каждый трехмесячный период начиная через полгода с момента подписания контрактов и кончая сроком $6 + 3n$ месяцев с этого момента будет завершаться строительство одного объекта. В соответствии с этим условием каждая фирма представила новый проект контракта, в котором стоимость строительства определяется заданным сроком его завершения. Сформулируйте постановку этой оптимизационной задачи.

46. Рассмотрите классическую транспортную задачу, аналогичную описанной в разд. 6.2. Введем дополнительные условия, заключающиеся в том, что ни один поставщик не может обслуживать более M различных потребителей и ни один потребитель не может получать товары более чем от N различных поставщиков. Предположим, что при этих условиях существует допустимое решение. Сформулируйте постановку такой оптимизационной задачи.

47. Студент первого курса одного университета решил в течение первого семестра прослушать лекции по пяти предметам, обозначенным A, B, C, D и E. Каждый предмет читают в четырех группах, занимающихся в различное время. Обозначим через $A1$ группу 1, в которой читают предмет A, а через t_{A1} — начало занятий по этому предмету в данной группе. Для простоты предположим, что каждый предмет читают ежедневно, а занятия в группах начинаются в 8, 9, . . ., 16 час. На выбор студента оказывает влияние время проведения занятий и репутация преподавателя. Пусть, например, символ P_{A1} обозначает предпочтение, которое отдает студент предмету A в группе 1. К сожалению, он не может выбрать самое привлекательное, с его точки зрения, распределение предметов по группам в связи с перекрытием занятий во времени.

а) Постройте модель выбора допустимого расписания, максимизирующего сумму оценок предпочтения студента.

б) Укажите, как следует изменить постановку в п. а), чтобы студент имел часовой перерыв на обед в 12 или 13 час.

в) Укажите, как нужно изменить постановку задачи, если цель студента — максимизировать непрерывную последовательность часов свободного времени начиная с 8 час. до первого занятия или в конце дня.

48. Рассмотрите задачу, описанную в упражнении 35 гл. 10. Дайте постановку этой задачи в терминах целочисленного программирования.

49. Задача расширения производственных мощностей. Фирма «Агрохимик» имеет три предприятия. В начале интервала 1 производственная мощность каждого предприятия составляет 10 единиц. К концу интервала 3 предприятие 1 должно иметь мощность 11, предприятие 2 — мощность 12 и предприятие — 3 мощность 13. Эти показатели выбраны с таким расчетом, чтобы к концу интервала 3 мощность каждого предприятия была достаточной для удовлетворения спроса в районе сбыта, обслуживаемом соответствующим предприятием. Мощность каждого предприятия можно увеличивать в течение любого интервала только на целое число единиц. Для упрощения задачи предположим, что дополнительная мощность, вводимая на каком-либо интервале, может быть использована для удовлетворения спроса на этом же интервале. Примем также, что достигнутая мощность является постоянной, т. е. что отклонений от достигнутой мощности в любую сторону не наблюдается. Таким образом, существует три возможных варианта расширения мощности предприятия 1 от уровня 10 в начале интервала 1 до уровня 11 в конце интервала 3:

I) ввод единицы мощности на интервале 1, II) ввод единицы мощности на интервале 2, III) ввод единицы мощности на интервале 3. (Покажите, что для предприятия 2 существует шесть допустимых вариантов увеличения мощности с 10 единиц в начале интервала 1 до 12 единиц к концу интервала 3.) Обозначим через K_{it} (p, q) затраты на расширение мощности предприятия i на p единиц в течение интервала t при условии, что его мощность на начало этого интервала составляла q единиц.

Положим, что D_{it} есть спрос в районе сбыта, обслуживаемом предприятием i в течение интервала t. Если мощности предприятия не хватает для удовлетворения спроса D_{it}, то фирма вынуждена осуществлять поставки продукции в этот район с одного или обоих предприятий, обслуживающих другие районы. Примем, что величина c_{ijt} определяет затраты на перевозку единицы продукции с предприятия j на предприятие i в течение интервала t.

Постройте модель определения стратегии расширения производственной мощности каждого предприятия, минимизирующей общие затраты и обеспечивающей в то же время полное удовлетворение спроса в течение всего рассматриваемого планового периода.

50. Фирма «Фил О'Троп» должна составить календарный план (расписание) предупредительного ремонта пяти своих крупных агрегатов, рассчитанный на восемь недель. Предположим, что агрегатам присвоены номера *1*, *2*, *3*, *4*, *5*. Для ремонта агрегата 1 в течение первой недели требуется затрата четырех единиц ресурса (допустим, человеко-недель), в течение второй недели — шести единиц и в течение третьей — трех единиц. Ремонт можно начинать не раньше начала недели 1 и не позже начала недели 4. Аналогичные данные по ремонту остальных установок приведены в таблице рис. 13.8.

Агрегат	Первая неделя ремонта	Вторая неделя ремонта	Третья неделя ремонта	Раннее начало	Позднее начало
1	4	6	3	1	4
2	3	2	5	1	3
3	7	1	1	2	5
4	1	3	6	2	6
5	8	9	2	3	5

Рис. 13.8.

В каждом из указанных ниже пунктов постройте соответствующую оптимизационную модель.

а) Предположим, что максимальное количество рабочей силы, имеющееся в течение недели t, равно t_t. Найдите допустимый календарный план (т. е. определите номер недели, в начале которой следует приступить к ремонту каждого агрегата).

б) Предположим, что фирма стремится минимизировать суммарные недельные колебания в использовании рабочей силы. [Так, например, в случае наиболее раннего начала ремонта каждого агрегата еженедельное использование трудовых ресурсов составит 7, 16, 20, 16, 2, 0, 0, 0, а сумма недельных колебаний будет равна $(16-7)+(20-16)+(20-16)+(16-2)+(2-0)+(0-0)+$ $+(0-0)=33$.]

в) Предположим, что фирма стремится минимизировать максимальное количество трудовых ресурсов, используемых в течение любой недели, на протяжении восьминедельного периода.

г) Предположим, что фирма стремится минимизировать максимальные недельные колебания в использовании трудовых ресурсов.

(Если ремонт всех агрегатов начинается в наиболее ранние возможные сроки, то максимальное колебание составит $16 - 2 = 14$ между четвертой и пятой неделей.)

д) Покажите, как следует изменить ответ п. а) в следующих случаях: ремонт агрегата 3 нельзя начинать до начала ремонта агрегата 1; ремонт агрегата 4 должен начаться в ту же неделю, что и ремонт агрегата 3 (следовательно, «позднее начало» ремонта этого агрегата теперь уже неделя 5, а не 6); ремонт агрегатов 4 и 5 нельзя начинать в одну и ту же неделю; ремонт агрегата 5 нельзя начинать, пока не закончен ремонт агрегата 1 (это означает, что если, например, ремонт агрегата 1 начинается в начале недели 1, то ремонт агрегата 5 не может начаться ранее недели 4).

51. Фирма «Пальчики оближешь» снабжает своими изделиями сеть небольших ресторанов, отпускающих обеды на дом. Фирма стремится обеспечить каждому владельцу такого домового ресторана достаточную прибыль. Рассматривается возможность размещения n новых точек подобного типа в крупном микрорайоне. По имеющейся оценке, размещение ресторана в пункте j обеспечит получение приемлемого дохода R_j при условии, что в радиусе 5 миль от него аналогичная торговая точка отсутствует. Примем $d_{ij} = 1$, если пункты i и j размещения домовых ресторанов находятся в пределах радиуса 5 миль, и $d_{ij} = 0$ в противном случае. Фирма рассчитала все возможные d_{ij} и стремится выбрать схему размещения новых домовых ресторанов, обеспечивающую максимизацию общего дохода.

а) Сформулируйте эту задачу в виде математической оптимизационной модели.

б) Покажите, как изменить постановку задачи п. а), если фирма снабжает также сеть закусочных и может использовать каждый из n пунктов либо в качестве домового ресторана, либо в качестве закусочной. Две закусочные не должны находиться друг от друга ближе чем в 10 милях.

52. Задача о покрытии. Фирма «Вечная память» хранит различную информацию, записанную на магнитных лентах. Фирма должна составить статистические данные по m различным демографическим характеристикам (например, таким, как распределение населения по возрастным группам, распределение доходов, обеспеченность жильем и т. п. по выбранным восьми городским районам). Предположим, что вся требуемая информация хранится в n массивах. Примем, что T_j ($j = 1, 2, \ldots, n$) — время поиска ЭВМ нужной информации в массиве j и что время поиска не зависит от числа характеристик, выбираемых из массива. Некоторые из m видов данных записаны более чем в одном из n массивов, т. е. различные массивы содержат информацию, относящуюся к одной и той же характеристике. Положим $a_{ij} = 1$, если данные по i-й характеристике записаны в j-м массиве, и $a_{ij} = 0$ в противном случае. Так, например, равенство $a_{13} = a_{18} = a_{19} = 1$ указывает, что данные, относящиеся к первой характеристике, записаны в массивах 3, 8 и 9.

а) Постройте модель целочисленного программирования для определения того, в каких из n массивов производить поиск, чтобы выбрать данные, относящиеся ко всем требуемым статистическим характеристикам, за минимальное время.

б) Объясните, как следует изменить модель, если время поиска в массиве j равно T_j в случае, когда такой поиск производится, плюс t_{ij} в случае, когда в результате поиска получают характеристику i.

53. Задача о доставке грузов (задача о покрытии). Фирма «Автопегас» должна доставить грузы пяти своим клиентам в течение рассматриваемого дня. Клиенту A нужно доставить груз весом в 1 единицу, клиенту B — в 2 единицы, клиенту C — в 3 единицы, клиенту D — в 5 единиц и клиенту E — в 8 единиц. Фирма располагает четырьмя автомашинами следующей грузоподъемности: машина 1 — 2 единицы, машина 2 — 6 единиц, машина 3 — 8 единиц, машина 4 — 8 единиц. Стоимость эксплуатации автомашины j составляет c_j. Предположим, что одна машина не может доставлять груз обоим клиентам A и C, аналогично одна машина не может использоваться для доставки груза обоим клиентам B и D.

а) Постройте модель целочисленного программирования для определения такого назначения автомашин для доставки всех грузов, при котором минимизируются суммарные затраты.

б) Покажите, как нужно изменить эту модель, если при доставке груза клиенту i производятся дополнительные затраты c_{ij}.

в) Покажите, как изменится модель, если автомашина не может в течение дня сделать более двух рейсов.

г) Разъясните, какое влияние оказывает на модель введение дополнительных ограничений на возможные маршруты движения автомашин.

54. Одно небольшое предприятие производит четыре вида изделий. Обозначим через D_{kt} потребность в изделиях k-го вида на интервале t, где $k = 1, 2, 3, 4$ и $t = 1, 2, \ldots, T$, предполагая, что любое D_{kt} — целое. Пусть x_{kt} обозначает уровень (объем) производства изделий вида k на интервале t. Примем, что на каждом интервале весь спрос должен удовлетворяться, однако в течение какого-либо интервала можно выпускать продукцию на склад и использовать ее для удовлетворения спроса на последующих интервалах. Обозначим через $c_{kt}x_{kt}$ производственные затраты, соответствующие уровню x_{kt}, а через h_k — стоимость содержания запасов изделий вида k, хранящихся на складе к концу рассматриваемого интервала. (Примем, что в начале интервала 1 уровень запасов равен нулю.) Пока что условия задачи не отличаются от условий модели, описанной в разд. 8.3. Однако вводится дополнительное ограничение, заключающееся в том, что в течение любого интервала можно выпускать не более одного вида изделий.

а) Постройте оптимизационную модель, учитывающую это производственное ограничение.

б) Сформулируйте задачу в терминах рекуррентного соотношения динамического программирования.

55. Покажите, как можно описать переменную структуру затрат в модели управления запасами, приведенной в упражнении 44 гл. 8, с помощью аппарата целочисленного программирования.

56. Покажите, как можно описать переменную структуру затрат в модели управления запасами, приведенной в упражнении 9 гл. 9, с помощью аппарата целочисленного программирования.

57. Рассмотрим модель целочисленного программирования, содержащую только булевы переменные.

а) Предположим, что модель включает член $x_1^2 x_5 x_9^3$ (входящий в целевую функцию или в одно из ограничений). Покажите, как можно выразить этот член с помощью булевой переменной y при условии, что вводятся ограничения $x_1 + x_5 + x_9 - y \leqslant 2$ и $-x_1 - x_5 - x_9 + 3y \leqslant 0$.

б) Как обобщается метод, использованный в п. а), на случай $(x_1)^{p_1}, (x_2)^{p_2}, \ldots, (x_n)^{p_n}$, где любое $p_j \geqslant 0$? [Обязательно используйте условия, когда в п. а) $p_j = 0$.]

в) Покажите, что x_j^p, где $p > 0$, можно выразить через x_j.

г) Примените результаты, полученные в предыдущих пунктах в упражнении 42 этой главы.

58. Квадратичная задача о назначениях. Консультационная фирма «Айрсон» дает своим клиентам рекомендации относительно выполнения проектов компоновки промышленных предприятий и учреждений. Характерной задачей, возникающей при разработке такого проекта, является выбор размещения оборудования на новом предприятии, разбитом на n различных производственных участков. На каждом участке должен быть размещен один из различных производственных агрегатов. Если, например, один агрегат типа 1 размещается на участке 3, а один агрегат типа 2 — на участке 5, то такой компоновке соответствуют затраты $c_{13;\,25}$, обусловленные транспортировкой материалов между агрегатами 1 и 2, перемещением обслуживающего персонала между ними и расстоянием между участками 3 и 5. (Поскольку имеется всего n^2 *попарных* комбинаций размещения агрегатов по участкам, общее количество коэффициентов затрат $c_{ij,\,kl}$ составляет n^4.) Постройте модель размещения агрегатов по участкам, минимизирующую затраты. (*Указание*: целевая функция модели является квадратичной.)

59. Существует целочисленная постановка задачи коммивояжера, основанная на использовании булевых переменных x_{ijk}. Значение $x_{ijk} = 1$ определяет, что k есть положение в цикле, когда коммивояжер находится в городе i и оттуда переезжает в город j. В любом цикле для города 1 принимается $k = 1$. Запишите подробно все ограничения и целевую функцию в такой постановке для задачи, включающей четыре города. (*Указание*: коммивояжер в ходе своего объезда должен выезжать из каждого города. Если в положении k он находится в городе i, откуда переезжает в город j, то, попадая

в город j, он оказывается в положении $k+1$.) Обязательно поясните, почему в этой постановке исключена возможность образования подциклов.

60. Календарное планирование методом критического пути. Рассмотрите задачу фирмы «Блитцстрой», описанную в разд. 6.6, с условиями, приведенными в таблице рис. 6.13. Предположим, что для выполнения каждой операции требуется заданное количество дефицитного ресурса, например рабочего времени, в течение всего процесса осуществления этой операции. Так, например, при условии, что продолжительность операции A составляет t_A единиц времени, предположим, что в течение каждой недели выполнения этой операции нужно r_A человек. Предположим также, что R_p есть максимальное количество рабочих, которых можно использовать для выполнения всех производственных операций на протяжении недели p.

а) Постройте математическую оптимизационную модель задачи предполагая, что целью, как и прежде, является завершение комплекса в минимально возможное время. (*Указание*: введите булевы переменные, определяющие неделю начала операции. Выразите срок начала операции через эти переменные и перепишите соответствующим образом ограничения следования (2) — (6), приведенные в разделе 6.6.) Объясните, как изменится постановка задачи, если интенсивность использования рассматриваемого ресурса меняется еженедельно в течение всего срока выполнения операции.

б) Покажите, как изменится модель п. а), если учитывается второе ограничение, накладываемое на ресурс иного вида. Предположим, например, что на протяжении всего срока выполнения операции A еженедельно требуются затраты s_A единиц этого ресурса и что S_p есть максимальное количество данного ресурса, имеющееся в наличии для выполнения всех операций в течение недели p.

61. Календарное планирование методом критического пути. Рассмотрим вновь пример фирмы «Блитцстрой», описанный в разд. 6.6, условия которого приведены в таблице рис. 6.13. Предположим, что существует несколько альтернативных вариантов реализации всего комплекса. В частности, вместо операции D можно выполнить либо операцию G, либо операцию H. Если выполняется операция G, то операции C непосредственно предшествуют операции A и G, операции D — операция B и операции E — операция B. Остальные данные, приведенные в таблице рис. 6.13, не меняются. Если же принимается решение выполнять операцию H, то эта операция не имеет непосредственно предшествующих, операции E непосредственно предшествует операция B, операции F — операции C, E и H. Остальные данные, приведенные в таблице рис. 6.13, остаются без изменений. Измените постановку задачи (1) — (6), учтя в ней возможность выбора между операциями D, G и H. Целью здесь, как и прежде, является минимизация продолжительности реализации комплекса.

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

62. Баланс сборочной линии. Сборочная линия состоит из ряда рабочих мест, на каждом из которых можно в процессе сборки изделия выполнять одну или более операций. Предположим, что всего нужно выполнить шесть операций и что изделие проходит по сборочной линии, двигаясь от рабочего места 1 к рабочему месту 2, . . . вплоть до последнего рабочего места. На каждую операцию затрачивается t_i мин. Суммарное время выполнения всех операций на одном рабочем месте не должно превышать величины T, называемой продолжительностью цикла линии. Примем в рассматриваемом примере $T = 10$. Операции должны выполняться в определенном порядке. Отношение следования между операциями и их продолжительности заданы таблицей рис. 13.9.

Операция i	Непосредственно предшествующие операции	Продолжительность операции t_i
1	—	5
2	—	6
3	1	2
4	2, 3	5
5	1, 2	4
6	4	3

Рис. 13.9

Оптимизационная задача заключается в минимизации числа рабочих мест при условии соблюдения заданной продолжительности цикла и сохранения заданного отношения следования. Постройте математическую оптимизационную модель этой задачи. (*Указание*: принять x_{ij} в качестве булевой переменной, равной 1, если операция i выполняется на рабочем месте j. При осуществлении операций на каждом рабочем месте должен выполняться заданный порядок следования. Продолжительность выполнения всех операций на любом рабочем месте не должна превышать T мин.)

63. Задача оперативно-календарного планирования. Предположим, что имеется три детали и нужно определить порядок их обработки на n станках. Каждая деталь должна сначала обрабатываться на станке 1, затем на станке 2 и т. д. вплоть до станка n. Последовательность обработки деталей на каждом станке может быть различной. Обозначим через t_{ij} продолжительность обработки детали i на станке j и предположим, что любая величина t_{ij} — целая. Целью является минимизация продолжительности обработки всех деталей на всех станках.

а) Постройте целочисленную модель задачи. Составьте подробную модель при $n = 4$. (*Указание*: один из подходов сводится к тому, чтобы принять в качестве переменной x_{ij} момент начала обработки детали i на станке j. При таком подходе необходимо гарантировать, чтобы ни на одном станке не обрабатывалось одновременно более одной детали и чтобы обработка детали на станке $j + 1$ не начиналась до окончания ее обработки на станке j. Другой подход заключается в введении булевых переменных x_{ijk}, посредством которых определяется включение детали i на k-е место в последовательности операций, выполняемых станком j.)

б) Укажите, возможны ли какие-либо упрощения в постановке задачи, если принимается условие, что все детали обрабатываются на каждом станке в одной и той же последовательности. (Можно показать, что такой календарный план является оптимальным в случае двух или трех станков. Для случая двух станков предложен простой алгоритм отыскания оптимального решения.)

64. Задача оперативно-календарного планирования. Примите в упражнении 63, что все детали должны обрабатываться в одной и той же последовательности на каждом из n станков. Дайте метод решения задачи, аналогичный методу *частичных циклов*, изложенному в разд. 13.6. [*Указание*: при заданном частичном включении деталей на места 1, 2, . . ., k в последовательности их обработки нижнюю оценку общей продолжительности обработки всех деталей можно выразить через момент окончания k-й операции на каждом станке и сумму продолжительностей операций, выполняемых на каждом станке и еще не включенных в календарный план (расписание).]

65. Задача о ранце. Рассмотрите упражнение 36 этой главы и постройте алгоритм решения, аналогичный *алгоритму задания маршрутов*, описанному в разд. 13.6. (*Указание*: при заданном наборе переменных, имеющих определенные значения, верхнюю оценку значения целевой функции с точки зрения оптимального выбора значений остальных переменных можно найти, решив соответствующую задачу линейного программирования.)

66. Задача с фиксированными платежами. Рассмотрите модель, описанную в упражнении 6 этой главы. Постройте алгоритм решения таких задач, аналогичный *алгоритми задания маршрутов*, приведенному в разд. 13.6.

67. Квадратичная задача о назначениях. Вернемся вновь к примеру консультационной фирмы «Айрсон», описанному в упражнении 59 этой главы. Рассмотрим все возможные попарные сочетания агрегатов: (агрегат 1, агрегат 2), (агрегат 1, агрегат 3), . . ., (агрегат 1, агрегат n), (агрегат 2, агрегат 3), (агрегат 2, агрегат 4), . . ., (агрегат n — 1, агрегат n). Аналогично рассмотрим все возможные попарные сочетания рабочих мест (место 1, место 2), (место 1, место 3), . . ., (место n — 1, место n). Покажите, что существует всего n (n — 1)/2 попарных сочетаний агрегатов и столько же попарных сочетаний рабочих мест.

Представим себе таблицу, имеющую n (n — 1)/2 строк, по одной для каждой пары агрегатов, и столько же столбцов, по одному для каждой пары рабочих мест. Предположим, что структура затрат такова, что при монтаже агрегата 1 на рабочем месте 3 и агрегата 2 на рабочем месте 5 или монтаже агрегата 1 на рабочем месте 5 и агрегата 2 на рабочем месте 3 затраты одинаковы. Иными словами, для определения величины затрат нужно знать только одно: какая пара агрегатов (агрегат 1, агрегат 2) должна быть размещена на соответствующей паре рабочих мест (рабочее место 3, рабочее место 5). Поясните, почему такое допущение о затратах может иметь практический смысл.

а) Предположим, что затраты, соответствующие попарному сочетанию агрегатов с попарным сочетанием рабочих мест, сведены в квадратную матрицу. Дает ли решение задачи о назначении при этих коэффициентах затрат ответ для задачи о размещении оборудования? (*Указание*: примите $n = 4$ и постройте квадратную матрицу, приведя в ней названия строк и столбцов. Выясните, далее, всегда ли решение задачи о назначениях является допустимым для задачи размещения оборудования.)

б) Исходя из выводов, сделанных в п. а), предложите способы модификации нескольких алгоритмов решения задачи коммивояжера, изложенных в разд. 13.6, которые позволяют отыскивать решение задачи о размещении оборудования.

68. Прямой алгоритм отсечения. Метод отсечения, описанный в разд. 13.4, можно определить как «двойственный» в том смысле, что постоянно выполняются двойственные ограничения (после того, как с помощью симплексного метода найдено оптимальное решение задачи линейного программирования без учета требований целочисленности). Последовательно получаемые решения не являются допустимыми по условиям прямой задачи до последней итерации, поскольку лишь на этой итерации удовлетворяются условия целочисленности. При использовании прямого алгоритма требования целочисленности, а также линейные ограничения выполняются постоянно. Прямой алгоритм, сходимость которого доказана, подробно в этой книге не рассматривается. Однако некоторые принципиальные идеи, лежащие в его основе, можно показать на конкретном примере.

Предположим, что имеется исходное, целочисленное, но не оптимальное решение, полученное с помощью прямого алгоритма. Применим обычный *критерий I* симплексного метода для выбора переменной, вводимой в базис на следующей итерации. Далее применим *критерий II* симплексного метода для определения максимального значения, при котором выбранная переменная может войти в базис при одновременном сохранении у всех остальных базисных переменных неотрицательных значений. Обозначим это значение выбранной переменной через $r \geqslant 0$. Если обычный ведущий элемент не равен 1, то строится отсечение, используемое в качестве строки, в которой производится элементарное преобразование. Это отсечение должно обладать следующими свойствами: I) ведущий элемент для выбранной переменной равен 1, вследствие чего все коэффициенты и базисные переменные после элементарного преобразования остаются целочисленными; II) константа в правой части не превосходит величины r, вследствие чего базисные переменные остаются неотрицательными. Рассмотрим способ отыскания такого отсечения.

Вычислим отношения по *критерию II* и возьмем теперь целую часть каждого отношения. Выберем любую строку, для которой целая часть отношения меньше или равна r. (Если имеется несколько таких строк, то выбирается строка с наибольшей целой частью.) В качестве вспомогательной операции разделим выбранную строку на коэффициент вводимой в базис переменной, а затем построим отсечение, определяемое выражением (12) из разд. 13.4. Дополним уравнения этим отсечением, выполним элементарное преобразование и вернемся к *критерию I*.

(Если $r < 1$, то константа в правой части отсечения обязательно равна 0. Следовательно, на такой итерации значение целевой функции не улучшается. Для обеспечения сходимости метод нужно модифицировать так, чтобы исключалась возможность образования бесконечной последовательности отсечений, имеющих нулевые правые части. Здесь эта модификация не излагается.)

Продемонстрируем работу алгоритма на следующем, несколько измененном примере фирмы «Мультиконвейер» (разд. 2.2 и 4.4):

$$\begin{aligned} x_0 - 4x_1 - 9x_3 \qquad\qquad &= 0 \quad \text{(строка 0)},\\ 1x_1 + 1x_3 + x_5 \qquad\quad &= 12 \quad \text{(строка 1)},\\ 3x_1 + 10x_3 \qquad + x_7 &= 85 \quad \text{(строка 2)}. \end{aligned}$$

(Покажите, что здесь имеются следующие отличия от исходного варианта, рассмотренного в разд. 4.4: переменные x_4 и x_6 опущены, переменная x_2 принята равной 3, а затем исключена, второе ограничение также исключено.)

На первой итерации для ввода в базис выберем переменную x_3. Выполнив вычисления, предусмотренные *критерием II*, найдем, что $r = (85/10) = 8{,}5$. Повторим вычисления отношений, используя целые части, в результате чего получим значения $[12/1]$ и $[85/10]$. Поскольку только $[85/10] \leqslant r$, нужно записать отсечение по уравнению в строке 2. В качестве вспомогательного вычисления разделим строки 2 на 10, получив в результате

$$\frac{3}{10}x_1 + x_3 + \frac{1}{10}x_7 = \frac{35}{10},$$

что приводит к отсечению

$$x_3 + x_8 = 8 \quad \text{(строка 3)},$$

где x_8 — дополняющая переменная отсечения. Добавим строку 3 и выполним элементарное преобразование по x_3.

а) Получите оптимальное решение задачи с помощью рассмотренного алгоритма. [*Указание*: необходимо найти еще три отсечения. Значение целевой функции возрастает с 0 до 72 (после элементарного преобразования по строке 3), затем до 76 и до 83, оставаясь равным 83 на последней итерации. Следует ввести дополняющие переменные x_9 и x_{10}, для получения сходимости необходимо выполнить элементарные преобразования по переменным x_8 и x_9. При элементарном преобразовании по переменной x_9 выбирается строка 1, что приводит к построению отсечения. Поясните, почему это так.]

б) Примените алгоритм для решения исходного варианта задачи фирмы «Мультиконвейер».

69. Теоретико-групповой алгоритм. Эффективный метод решения задач целочисленного программирования основан на использовании тех свойств этих задач, которые рассматриваются в разделе математики, называемом теорией групп. Рассмотрение такого подхода в деталях выходит за рамки этой книги, но его сущность можно пояснить на конкретном примере. Обратимся вновь к задаче фирмы «Мультиконвейер», описанной в разд. 2.2, решение которой симплексным методом приведено в разд. 4.4. Рассмотрим уравнения (7) на последней итерации. Пусть алгебраическое выражение одного из уравнений (7) имеет вид $\sum_{j=1}^{7} A_i x_i = B$. Тогда можно записать

$$A_i \equiv [A_i] + f_i,$$

где $0 \leqslant f_i < 1$ и $B \equiv [B] + f$, где в свою очередь $0 \leqslant f < 1$. Иными словами, f_i есть дробная часть A_i, а f — дробная часть B. Теперь это уравнение можно представить следующим образом:

$$\sum_{i=1}^{7} [A_i]\, x_i + \sum_{i=1}^{7} f_i x_i = [B] + f.$$

Объясните, почему для получения любого допустимого целочисленного решения необходимо, чтобы выражение $\sum_{i=1}^{7} f_i x_i - f$ было целым. Покажите, что если x_i — базисная переменная, то $f_i = 0$. Объясните, исходя из этого, почему переменная x_i должна удовлетворять условию $\sum_{i=1}^{7} f_i x_i = f \pmod 1$, т. е. для обеспечения допустимости решения левая часть должна отличаться от правой на целое число.

Покажите, обратившись к уравнениям (7), что таким образом должны выполняться условия

$$\frac{5}{7} x_2 + \frac{2}{7} x_4 + \frac{3}{7} x_5 + \frac{6}{7} x_7 \equiv \frac{1}{7} \pmod 1,$$

$$\frac{1}{7} x_2 + \frac{6}{7} x_4 + \frac{2}{7} x_5 + \frac{4}{7} x_7 \equiv \frac{3}{7} \pmod 1,$$

$$\frac{2}{7} x_2 + \frac{5}{7} x_4 + \frac{4}{7} x_5 + \frac{1}{7} x_7 \equiv \frac{6}{7} \pmod 1.$$

Покажите, что, умножив обе части этих уравнений на 7, получим эквивалентные выражения

$$\begin{aligned} 5x_2 + 2x_4 + 3x_5 + 6x_7 &\equiv 1 \pmod 7, \\ 1x_2 + 6x_4 + 2x_5 + 4x_7 &\equiv 3 \pmod 7, \\ 2x_2 + 5x_4 + 4x_5 + 1x_7 &\equiv 6 \pmod 7. \end{aligned} \tag{M}$$

Покажите, что при подстановке допустимого решения системы уравнений (M) в (7) получаем целые значения базисных переменных, однако эти значения не обязательно являются неотрицательными.

Из системы (7) известно, что значение целевой функции будет равно ${}^{695}/_7 - {}^3/_7 x_2 - {}^{11}/_7 x_4 - {}^{13}/_7 x_5 - {}^5/_7 x_7$. Поэтому отыскиваются допустимые значения небазисных переменных, минимизирующие выражение $3x_2 + 11x_4 + 13x_5 + 5x_7$ (чтобы избавиться от дробей, все выражение целевой функции умножено на 7).

Покажите, что при соотношениях по модулю 7, заданных в (M), $x_2 = 2$ эквивалентно $x_5 = 1$; $x_5 = 2$ эквивалентно $x_7 = 1$; $x_5 = x_7 = 1$ эквивалентно $x_4 = 1$. В основе этих утверждений лежит принцип, заключающийся в том, что имеется группа из семи столбцов, в которой каждый столбец содержит по элементу, соответствующему каждой строке; такому элементу ставится в соответствие любой набор целых значений x, входящих в систему (M). В рассматриваемом примере можно построить такие столбцы, взяв любой столбец коэффициентов, допустим при x_2 $\left[\text{в (M) это} \begin{pmatrix} 5 \\ 1 \\ 2 \end{pmatrix}\right]$, и вычислить столбцы: $\begin{pmatrix} 5x_2 \\ 1x_2 \\ 2x_2 \end{pmatrix}$, $x_2 = 0, 1, \ldots, 7$, производя все операции по модулю 7. Покажите, что в итоге получим

0	5	3	1	6	4	2
0	1	2	3	4	5	6
0	2	4	6	1	3	5
1	**2**	**3**	**4**	**5**	**6**	**7**

где числа, набранные полужирным шрифтом, служат идентификаторами столбцов. Отметим, что **4** определяет правую часть (M).

Минимизацию рассмотренной выше целевой функции при ограничениях (M) можно свести к задаче выбора кратчайшего маршрута. Будем считать, что узлами являются **1**, **2**, . . ., **7**, соответствующие семи столбцам чисел. Сопоставим все $x_i = 0$ в левой части (M) узлу **1**. Поставим, далее, в соответствие каждому $x_i = 1$ дугу, идущую из узла **1** в другой узел. Так, например, $x_2 = 1$ дает дугу, идущую от узла **1** в узел **2**, с коэффициентом затрат (ценой), равным 3; $x_4 = 1$ — дугу из узла **1** в узел **7** с ценой 11; $x_5 = 1$ — дугу из узла **1** в узел **3** с ценой 13. Какую дугу дает $x_7 = 1$?

По этому же принципу проводятся по четыре дуги из всех остальных узлов, кроме узла **4**, соответствующего правой части (M).

(При этом нет необходимости проводить дуги, ведущие обратно в узел **1**.) Так, например, дуга с ценой 3 из узла **2** в узел **3** соответствует единице x_2, дуга с ценой 13 из узла **2** в узел **4** соответствует единице x_5. Узлы и несколько дуг показаны на рис. 13.10. Завершите построение сети и решите оптимизационную задачу отыскания кратчайшего маршрута из узла **1** в узел **4** на этой сети.

Маршрут **1—5—2—4** соответствует условиям $x_7 = 2$ и $x_5 = 1$. Покажите, что кратчайший маршрут в сети соответствует условию

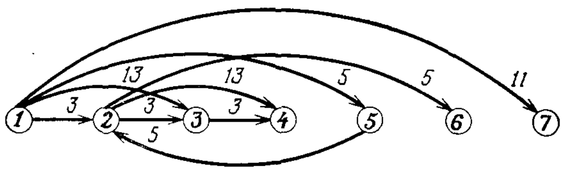


Рис. 13.10.

$x_2 = 3$. Какие значения принимают базисные переменные x_i в (7) при $x_2 = 3$? Объясните, почему это и есть оптимальное решение исходной задачи.

а) Примените алгоритм для решения задачи (20) — (22), описанной в разд. 13.4.

б) Задача, указанная в п. а), является примером задачи о ранце, приведенной в упражнениях 22 и 46 гл. 10. Используйте результаты, полученные в п. а), для определения характерных особенностей оптимального решения при любом возможном значении постоянной в правой части ограничения.

в) Предположим, что в задачу фирмы «Мультиконвейер» дополнительно введена переменная x_8 с коэффициентами a_1, a_2, a_3 в столбцах 1, 2, 3 соответственно, где каждый a_i — целое число. Покажите, что столбец дробных частей, которые ставятся в соответствие x_8 в (7), будет после умножения на 7 представлять собой один из столбцов **1**, **2**, . . ., **7**.

г) Выясните связь между рассмотренным алгоритмом и сетевым представлением задачи, приведенными в этом упражнении, с одной стороны, и алгоритмом и сетевым представлением задачи, определяемыми рекуррентным соотношением (II) в конце разд. 10.2,— с другой.

71. Решение задачи коммивояжера можно выразить в виде рекуррентного соотношения динамического программирования. Отметим, что в любой точке рассматриваемого цикла, в которой коммивояжер принимает решение о выборе следующего города, соответствующую переменную состояния можно рассматривать как город, где он находится в данный момент, и *множество* городов, в которых он уже побывал (однако это *не относится* к порядку посещения городов). Сформулируйте рекуррентное соотношение динамического программирования исходя из этого условия. Оцените, является ли такой подход реализуемым с точки зрения получения числовых решений.

72. В материале, изложенном в конце разд. 13.2, указан метод, позволяющий включать кусочно-линейную функцию $c(x)$ в задачу целочисленного программирования. С этой целью используются соотношения (24), (25), (27) — (30). Такой метод является разновидностью подхода, рассмотренного в разд. 14.7. Иную постановку можно дать, исходя из метода, изложенного в разд. 14.7, а именно (I) — (VII). Покажите, как можно на основе этого метода построить такую модель целочисленного программирования, которая будет обладать *свойством соседних весов*.

ГЛАВА 14

Оптимизация при нелинейной целевой функции

14.1. ВВЕДЕНИЕ В НЕЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Напомним, что любую оптимизационную задачу линейного программирования можно сформулировать следующим образом:

$$\text{Максимизировать } \sum_{j=1}^{n} c_j x_j \tag{1}$$

при условиях

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

$$x_{bj} \geqslant 0, \quad j=1, 2, \ldots, n. \tag{3}$$

Читателю уже приходилось сталкиваться с несколькими нелинейными вариантами этой модели, например с задачами динамического программирования в гл. 8—10 или же с комбинаторными моделями в гл. 13, где некоторые из x_j должны принимать только целочисленные значения. К **нелинейному программированию** обычно относят задачи следующего типа:

$$\text{Максимизировать } c(x_1, x_2, \ldots, x_n) \tag{4}$$

при условиях

$$a_i(x_1, x_2, \ldots, x_n) \leqslant 0, \quad i=1, 2, \ldots, m, \tag{5}$$

где как $c(x_1, \ldots, x_n)$, так и $a_i(x_n, \ldots, x_n)$ представляют собой действительные, нелинейные, однако регулярные функции n действительных переменных. По существу (4) — (5) можно рассматривать как каноническую форму задачи нелинейного программирования, поскольку методы, изложенные в гл. 3, позволяют интерпретировать оптимизацию как «максимизацию», а ограничения — как неравенства, знак которых определяется (5). Ограничения неотрицательности переменной x_j также могут быть включены в (5) в виде $-x_j \leqslant 0$. Обычно в «задачах нелинейного программирования» не выдвигается требование целочисленности переменных.

Здесь следует оговориться, что в (5) всегда можно включить нелинейные ограничения, благодаря которым в допустимые решения модели будут входить только целочисленные значения тех или иных переменных. Пусть, например, x_j должен принимать только значения 0 или 1. Выполнение этого условия целочисленности можно обеспечить, наложив ограничения $-x_j \leqslant 0$, $x_j - 1 \leqslant 0$ и $x_j(1-x_j) \leqslant 0$. В настоящей и следующей главах такие приемы не рассматриваются. В конце разд. 13.6 кратко освещено решение задач нелинейного программирования с целочисленными переменными.

Необходимость в методах нелинейного программирования. Даже самого краткого обзора простых, но все же поучительных моделей линейного программирования, приведенных в гл. 2, будет достаточно для того, чтобы вызвать у читателя сомнения в действительной адекватности строго линейных моделей многим реальным ситуациям. Легко может создаться впечатление, что при линейном подходе попросту игнорируются такие явления, как: эффективность или неэффективность укрупнения операций в многопродуктовых моделях; отсутствие аддитивности объемных показателей при составлении химических смесей; влияние объема реализации на цену реализации, а следовательно, на выручку от реализации. Если у читателя и создалось такое впечатление, приведенное ниже обсуждение позволит устранить возникшее недоразумение.

Имеется много данных об очень успешном применении моделей линейного программирования в условиях нелинейности. Поскольку любая модель неизбежно оказывается лишь приближением к реальности, возникает важный вопрос: «*в каких случаях* линеаризованный вариант является адекватным отображением нелинейного явления?» Следовательно, читатель должен научиться отличать условия, в которых непосредственное применение линейного программирования приемлемо, от условий, где оно неприемлемо.

Если, занимаясь конкретным приложением, мы располагаем некоторыми познаниями относительно допустимого диапазона значений переменных в оптимальном решении, то, как правило, можно наложить соответствующие ограничения и получить достаточно надежное линейное приближение. Не удается построить достаточно хорошее линейное приближение только в тех случаях, когда существует широкий диапазон допустимых решений и мы не имеем представления о характере оптимального решения. Эти утверждения являются до некоторой степени тавтологией, поскольку они в первую очередь относятся к нашему умению строить модель, ее целевую функцию и ограничения. Однако в большинстве случаев практического применения читатель обнаружит, что вопрос об адекватности линеаризованного варианта решается проще, чем это можно было бы предположить из приведенных выше намеренно неопределенных замечаний. Далее следуют два примера, на основе которых можно судить о вероятной неадекватности линеаризованного варианта в некоторых реальных ситуациях.

Рассмотрим фирму, находящуюся на начальном этапе разработки модели перспективного планирования в масштабах всей фирмы. Специалист по управлению обычно знает, что даже опытному бизнесмену трудно дать точный и детальный прогноз оптимальных объемов производства фирмы и распространения ее контроля на рынки сбыта на последующие 10 и более лет. В самом деле, применение администратором такой модели в основном обусловлено именно тем, что он понимает, как легко ошибиться при использовании только лишь интуитивных соображений в стремлении оценить влияние экономических факторов в последующие периоды. Если затраты производства и выручка от реализации зависят от объема операций нелинейно, линеаризованные догадки могут оказаться недостаточными для получения надежных ответов. (Вместе с тем при многократном применении линейная модель может оказаться применимой, если только параметры модели значительно не меняются во времени.)

Второй пример относится к фирме, составляющей производственную программу с помощью динамической многопродуктовой модели, отображающей существенные затраты времени на наладку станков, ограниченную мощность отдельных групп оборудования, колеблющийся спрос. Обычно в таких случаях *сущность* оптимизационной задачи состоит в варьировании различных нелинейных факторов, влияющих на принимаемые решения, относительно производственной программы. Если только специалист, разрабатывающий программу, не имеет очень хорошего представления о характере оптимального решения, любая простая линеаризация задачи, вероятно, приведет к нарушению фундаментальных принципов оптимизации.

Однако даже если в данном конкретном случае может потребоваться построение нелинейной модели, иногда удается использовать метод решения, лишь немногим отличающийся от *метода решения* для линейной модели. Наряду с этим при существенной нелинейности в связи с ее спецификой или влиянием ее на характер модели приходится применять методы оптимизации, гораздо более сложные, чем симплексный алгоритм. Уяснением этих вопросов мы займемся ниже, а также в следующей главе.

Значение нелинейных моделей для управления. В настоящее время применение математического программирования в преобладающем большинстве реальных ситуаций сводится к моделям линейной аппроксимации, а не к нелинейным моделям в явном их виде. Однако значимость нелинейного программирования и его использования постоянно возрастает. Это обусловлено быстро растущими познаниями руководителей и специалистов по исследованию операций в части использования математических моделей, предназначенных для подготовки решений, а также все большей доступностью машинных программ решения нелинейных задач большой размерности.

В большинстве случаев нелинейности, которые необходимо отобразить в моделях, относятся к одной из двух категорий:

I) эмпирически наблюдаемые соотношения, такие, как непропорциональные изменения затрат, выхода продукции, показателей качества;

II) структурно полученные соотношения, к которым относятся постулируемые физические явления, а также выведенные математически или установленные руководством правила поведения.

Очевидно, четкое разграничение этих двух категорий невозможно, поскольку при наличии достаточных данных можно вывести структурное соотношение, лежащее в основе эмпирически наблюдаемого явления.

Примером I) может служить тот случай, когда на предприятии в течение ряда лет прирост выпуска продукции отстает от роста затрат труда, тогда как темпы роста количества отходов его обгоняют. Примером II) является фирма, которая должна оплатить счет за электроэнергию в случае, когда расчеты ведутся по нелинейной формуле, учитывающей как среднесуточный расход, так и «пиковую» потребность в энергии. В данном случае фирма получает сведения о нелинейном характере затрат из договора о ставках оплаты, заключенного с компанией, обеспечивающей энергоснабжение.

Нелинейность «встраивается» в модели программирования и в других случаях, например, в следующих:

1. *Приготовление бензиновых смесей.* В модели приготовления бензина определенного состава из отдельных фракций, полученных в результате перегонки нефти, обычно имеется нелинейное ограничение на октановое число смеси, поскольку эта характеристика качества нелинейно зависит от количества добавляемого к смеси тетраэтилового свинца.

2. *Управление производственным процессом.* В модели металлургического завода значение переменной, характеризующей температуру в доменной печи, может быть описано функцией от других переменных, соответствующих количеству потребного тепла и временным показателям процесса. В свою очередь каждая из этих переменных входит в другие ограничения, а также в целевую функцию.

3. *Выручка от реализации продукции.* Спрос на продукцию компании может существенно зависеть от цен реализации: чем ниже цена продукта, тем больше объем реализации, несмотря на аналогичное снижение цен, производимое конкурентами. Следовательно, выручка от реализации продукции не изменяется пропорционально цене, и это обстоятельство должно быть отражено в целевой функции многопродуктовой модели с помощью нелинейного слагаемого. Для иллюстрации примем, что $x(p)$ есть объем реализации, зависящий от цены p; тогда выручка от реализации равна $p \cdot x(p)$. Пусть на представляющем для нас интерес интервале изменения p функция объема реализации от цены линейна, т. е. имеет вид $x(p) = ap + b$. Тогда слагаемые в целевой функции, относящиеся к выручке от реализации, являются квадратичными относительно управляющей переменной p и имеют вид $(ap^2 + bp)$.

4. *Размер многопродуктового заказа.* В моделях управления запасами, приведенных в гл. 8 и 9, число продуктов было равно одному. Однако нередко оптовый покупатель пополняет свои запасы, заказывая у одного и того же поставщика одновременно несколько видов продукции. Тем самым достигается экономия на транспортных затратах, расходах по оформлению документации и скидке на размер заказа, предоставляемой поставщиком. Эта ситуация может рас-

сматриваться на основе использования модели математического программирования большой размерности, в которой затраты на пополнение запасов являются нелинейной функцией нескольких переменных — размеров заказов отдельных продуктов.

5. *Уровень страховых запасов.* В большинстве моделей математического программирования, используемых для общефирменного планирования, длительность отрезков планового периода редко составляет менее трех месяцев и часто превышает год и более. В таких динамических, «многопериодных» моделях обычно предусматривается условие наличия страховых запасов, которые должны выполнять роль компенсатора колебаний еженедельного объема реализации. В этих моделях применяется, в частности, следующий подход: уровень страхового запаса предполагается зависимым как от прогнозированного объема реализации, так и от степени использования производственных мощностей, обусловленной этим прогнозом. Так, например, пусть c — максимально возможный недельный объем производства рассматриваемого продукта, s — прогнозированный *средненедельный* объем реализации этого продукта и $n \cdot s$ — уровень страхового запаса продукта, где n — число недель, зависимое от коэффициента использования производственных мощностей s/c. Для примера предположим, что администрация приняла следующую формулу расчета n: $n = m + f\,(s/c)$. Тогда уровень страхового запаса представляет собой квадратичную функцию прогнозированного средненедельного уровня реализации, имеющую следующий вид: $[ms + (f/c)\,s^2]$. Этот уровень может входить как в ряд ограничений модели, так и в целевую функцию.

6. *Распределение усилий.* В гл. 10 приводилось несколько примеров так называемой задачи распределения усилий. Предполагается целочисленность управляющих переменных, а число ограничений сводится до одного или двух, так чтобы подход динамического программирования оказался реализуемым с вычислительной точки зрения. Если исключить условие целочисленности переменных, то для решения задачи распределения усилий может быть также применена нелинейная модель. При этом возможно использование методов нелинейного программирования, реализуемых с вычислительной точки зрения даже в том случае, когда число ограничений намного превышает два.

7. *Вероятностные элементы.* Нелинейность часто возникает в том случае, когда коэффициенты модели математического программирования рассматриваются как случайные величины. Вкратце рассмотрим два примера.

Вначале остановимся на обычной транспортной модели, аналогичной тем, которые были описаны в разд. 2.7 и 6.2. Предположим, однако, что спрос D_j каждого потребителя j представлен в виде вероятностного распределения и что действительные значения всех D_j становятся известными только *после* отгрузки продукции. Если отгрузка потребителю j составляет x_j, остаток продукта у потребителя после удовлетворения спроса равен $(x_j - D_j)$, где отрицательное значение разности соответствует неудовлетворенному спросу. Пусть $c_j(x_j - D_j)$ — затраты, соответствующие остатку $(x_j - D_j)$. Объясните, почему эта *функция* затрат нелинейна. Таким образом, в данной модели целевой функцией является минимизация суммы транспортных затрат и *ожидаемых* значений $c_j(x_j - D_j)$. Соответственно целевая функция нелинейно зависит от значений управляющих переменных x_j. Эта модель подробно рассматривается в разд. 16.6.

Изложенный пример можно формулировать как транспортную модель, в которой ряд правых частей ограничений являются случайными величинами. Теперь рассмотрим общую модель линейного программирования, в которой правые части b_i отдельных линейных неравенств (2) есть случайные величины. Предположим, например, что модель включает два ограничения (2), т. е. $i = 1, 2$, и что распределения параметров b_i являются независимыми; $G_i(b)$ — вероятность, что значение случайной величины b_i не меньше чем b. Предположим также, что требуется выбрать значения x_j таким образом, чтобы совместная вероятность удовлетворения обоих ограничений была бы не меньше чем β:

$$P\left[\sum_{j=1}^{n} a_{1j}x_j \leqslant b_1\right]\cdot P\left[\sum_{j=1}^{n} a_{2j}x_j \leqslant b_2\right] \geqslant \beta \quad (0 < \beta \leqslant 1). \tag{6}$$

Тогда, как читатель убедится в разд. 16.5, ограничения, эквивалентные (6), можно записать в следующем виде:

$$-y_i + \sum_{j=1}^{n} a_{ij}x_j = 0, \quad i = 1, 2, \tag{7}$$

$$G_1(y_1)\cdot G_2(y_2) \geqslant \beta, \tag{8}$$

где наличие в (8) произведения двух величин обусловливает наложение нелинейных ограничений на y_1 и y_2.

8. *Выбор портфеля ценных бумаг.* Специалисты по финансовому анализу, работающие в банках и страховых обществах, уделяли немало внимания разработке математических моделей, помогающих определить наилучший набор акций, облигаций и других ценных бумаг на выделенную сумму. Такими моделями учитывается оценка как ожидаемой прибыли от приобретения предлагаемого набора бумаг, так и связанного с этим риска или вероятностного колебания действительных значений прибыли. Детальное обсуждение методов построения таких моделей выходит за рамки настоящей главы, однако ниже приводится простой вариант подобной модели, позволяющий осознать удобство постановки задачи в виде модели нелинейного программирования.

Пусть x_j — *доля* имеющихся средств, выделяемая для приобретения ценных бумаг j. Предположим, что прибыль к концу планового периода на каждый доллар, вложенный в бумаги j, характеризуется двумя показателями: a_j — фактическая прибыль (случайная величина), α_j — соответствующая *ожидаемая* прибыль. Предположим, далее, что задано условие: ожидаемая прибыль на доллар инвестиций должна быть для всего набора ценных бумаг не ниже заданной величины b. Тогда ограничения этой модели имеют следующий вид:

$$\sum_{j=1}^{n} x_j = 1 \quad \text{и} \quad x_j \geqslant 0, \quad j = 1, 2, \ldots, n, \tag{9}$$

$$\sum_{j=1}^{n} \alpha_j x_j \geqslant b, \tag{10}$$

где левая часть (10) отображает ожидаемую прибыль на доллар инвестированного капитала, поскольку ожидаемое значение суммы случайных величин равно сумме ожидаемых значений этих величин.

Риск учитывается с помощью целевой функции. Пусть x_j выбираются таким образом, чтобы минимизировать дисперсию *фактической* прибыли при выполнении условий (9) и (10). Тогда целевая функция представляет собой квадратичную форму:

$$\text{Минимизировать} \sum_{i=1}^{n} \sum_{j=1}^{n} \sigma_{ij} x_i x_j, \tag{11}$$

где

$$\sigma_{ij} \equiv E\,[a_i - \alpha_i] \cdot [a_j - \alpha_j] \tag{12}$$

обозначает ковариацию прибыли для ценных бумаг вида i и вида j.

В более сложных вариантах модели могут вводиться дополнительные ограничения, относящиеся к структуре набора ценных бумаг; модели могут быть динамическими и использовать другие показатели риска. Часто вычислительный процесс решения таких моделей включает тщательный анализ чувствительности, при котором выявляются ожидаемые значения прибыли, компенсирующей тот или иной риск. Не приходится и говорить, что подобные модели следует применять с осторожностью, поскольку значения параметров α_j и σ_{ij} исчисляются на основе ретроспективных данных и могут оказаться ошибочными в случае их применения к будущему.

Выводы. Как видно из сказанного, множество разнообразных обстоятельств приводит к нелинейной формулировке ограничений или целевых функций задач математического программирования. Естественно, за введение нелинейных зависимостей в модели математического программирования приходится платить дополнительную цену. Если число нелинейностей невелико или они несущественны, увеличение объема вычислений может оказаться незначительным. Тогда единственное неудобство заключается в необходимости найти машинную программу, позволяющую решить построенную нелинейную модель. В противном случае дополнительные усложнения могут привести к серьезному увеличению объема вычислительных операций, необходимых для нахождения решения, так что иногда приходится сокращать число ограничений и переменных модели.

В любой конкретной ситуации необходимо в должной перспективе рассмотреть требуемую размерность и сложность модели и оценить ее предполагаемое влияние на принимаемые решения. Так, например, модель нелинейного программирования средних размеров может оказаться достаточной для отыскания области, включающей оптимальное решение. После определения этой области можно использовать более детальную модель линейного программирования, имеющую большую размерность, причем необходимая аппроксимация параметров этой модели основывается на полученном решении нелинейной модели. Такой двухэтапный подход особенно полезен тогда, когда модели применяются для стратегического планирования, а не для принятия текущих решений.

14.2. НАПРАВЛЕННОСТЬ ПОДХОДА И КРУГ ОХВАТЫВАЕМЫХ ВОПРОСОВ

Как и в других разделах книги, в настоящей и следующей главах мы сосредоточим наше внимание на принципиальных идеях. Ниже излагаются необходимые основы для последующего обсуждения техники расчетов.

Выбор метода. Теория нелинейного программирования пока еще находится в процессе развития, а накопленный опыт применения конкретных алгоритмов ограничен. Следовательно, мы лишены возможности выделить два-три подхода, назвать их исключительно успешными и заявить, что они столь же важны для нелинейного программирования, как симплексный метод для линейного программирования. Практика показала, что, хотя от некоторых предложенных методов можно уже отказаться как от неудачных, все же, по-видимому, никогда не удастся отобрать два-три метода, которые являлись бы лучшими для решения *любых* нелинейных задач математического программирования. Поэтому приводимые в настоящей главе алгоритмы следует рассматривать не более как результаты ограниченного отбора методов решения.

Систематическое сопоставление методов осложняется тем обстоятельством, что какой-либо алгоритм нелинейного программирования может оказаться чрезвычайно эффективным для решения задач одного типа и совершенно неудачным для решения других задач. Более того, успешность использования алгоритма при решении конкретной задачи может существенно зависеть от постановки самой задачи. Так, например, изменение масштабов измерения тех или иных переменных может кардинальным образом повлиять на скорость сходимости алгоритма! (Это явление иногда называют чувствительностью к форме задачи.) Наконец, при сравнительной оценке алгоритмов необходимо сопоставлять объемы вычислений для методов, требующих лишь нескольких итераций, хотя на каждой из них выполняется большой объем расчетов, и методов со значительным числом итераций, где, однако, отдельная итерация требует небольшого числа вычислительных операций. Результаты подобных сравнений отчасти зависят и от показателей используемой ЭВМ. По этой причине в книге не даются сравнительные оценки алгоритмов.

В отличие от методов линейного и динамического программирования большинство алгоритмов нелинейного программирования, которые нетрудно обосновать и описать, столь мало эффективны, что практически их нельзя применять. В то же время, к сожалению, алгоритмы, которые оказались наиболее мощными на практике, слишком сложны, чтобы включать их описание в настоящую книгу, являющуюся вводным курсом. Поэтому, стремясь дать объективное и в то же время доступное для начинающих изложение методов нелинейного программирования на основе одних элементарных понятий, мы выбрали методы, которые легко объяснить и которые в то же время достаточно эффективны, если их соответствующим образом модифицировать для использования на ЭВМ. Однако возможность излагать подобные модифицированные методы, которые крайне необходимы для успешного решения практических задач, мы оставляем авторам трудов, целиком посвященных численным методам. Кроме того, здесь описываются лишь методы, эффективные при решении задач средней и большой размерности, и почти не уделяется внимания алгоритмам, которые, возможно, применимы для моделей, включающих малое число ограничений и переменных.

Вычислительные трудности. Предложим читателю вкратце возвратиться к обсуждению алгоритмического метода, приведенного в разд. 4.3. В этом разделе отмечалось, что алгоритм определяется такими четырьмя взаимосвязанными характеристиками, как:

полнота;
область применения;
сходимость;
вычислительные требования.

Большинство этих характеристик значительно труднее выявить при анализе методов нелинейного программирования, чем алгоритмов линейного программирования. Этими вопросами занимаются математики, изучающие теорию нелинейного программирования. Однако, не говоря уже о теоретических проблемах, при использовании методов нелинейного программирования в конкретной ситуации возникают вычислительные трудности.

Предположим, используется алгоритм, который с исчерпывающей полнотой, шаг за шагом *описывает* выполняемые операции. Однако может оказаться, что сходимость алгоритма обеспечивается при выполнении некоторых математических условий — необходимых или достаточных или необходимых и достаточных, которым должны удовлетворять целевая функция и ограничения модели. Проверить соблюдение таких условий в случае практического применения бывает слишком трудно. Поэтому не исключено, что определить пригодность алгоритма для решения данной конкретной задачи можно будет, лишь попытавшись ее решать.

Далее, даже если сходимость алгоритма для некоторого легко распознаваемого класса задач является доказанной, это все же не позволит дать хорошую оценку числа итераций, необходимых для получения пригодного решения.

Более того, иногда оказывается удобным использовать алгоритм, *с теоретической точки зрения* весьма сомнительный, например не всегда обеспечивающий сходимость к истинному решению. Однако на практике алгоритм может давать хорошие результаты, часто сходясь к истинному оптимуму за относительно небольшое число итераций.

Исследования в области разработки методов нелинейного программирования ведутся в двух направлениях. В области теории исследования сосредоточены на установлении слабых и, как можно надеяться, проверяемых условий, налагаемых на целевую функцию и ограничения, которые являются достаточными для обеспечения *гарантированной* сходимости алгоритма к оптимуму вне зависимости от исходной точки расчетов. Исследования в области вычислительных аспектов сосредоточены на выяснении того, насколько хорошо данный алгоритм решает конкретные задачи. В обоих этих направлениях сущность затруднений состоит в том, что большинство методов нелинейного программирования *не* всегда обеспечивает сходимость за конечное число итераций (если только исходная точка случайно не оказалась точкой оптимума). Некоторый алгоритм может обеспечивать монотонное улучшение значения целевой функции при переходе от одной итерации к другой. Но даже если сходимость алгоритма известна, на практике решения, получаемые на отдельных итерациях, могут резко изменяться, хотя значения целевой функции при этом улучшаются. Кроме того, если только не доказана теоретическая правильность метода, возможна сходимость к решению, *отличающемуся* от оптимума. (Это демонстрируется на примере, приведенном в разд. 14.4.)

Рассматриваемые вопросы. В нескольких следующих разделах закладываются основы для объяснения того, как найти оптимальное значение функции одной переменной, а затем и функции нескольких не подчиненных ограничениям переменных. Как читатель убедится, для того чтобы учесть включение ограничений, большинство фундаментальных идей, лежащих в основе методов отыскания безусловного оптимума, нуждается лишь в надлежащей обработке.

В остальной части настоящей главы рассматриваются главным образом методы решения оптимизационных моделей с линейными ограничениями и специальными, но важными видами нелинейных целевых функций. Несколько алгоритмов, предназначенных для решения моделей нелинейного программирования более общего вида, описано в гл. 15. Следует отметить, что во многих методах решения задач на условный экстремум, приведенных в этих двух главах, в качестве части вычислительного процесса используются решения некоторых вспомогательных задач линейного программирования. В гл. 15 излагается также математическая теория нелинейного программирования, понятия которой соответствуют понятию двойственности в линейном программировании.

14.3. ОПТИМИЗАЦИЯ НЕЛИНЕЙНОЙ ФУНКЦИИ ОДНОЙ ПЕРЕМЕННОЙ

Имеются две причины изучения методов оптимизации функции одной переменной. Первая из них заключается в том, что на одномерных задачах можно демонстрировать многие трудности, с которыми приходится сталкиваться при решении нелинейных оптимизационных задач. Нет необходимости для разъяснения этих трудностей обращаться к многомерным задачам, иногда менее понятным. Вторая причина состоит в том, что во многих методах решения нелинейных оптимизационных задач со многими переменными на том или ином шаге, по существу, используются алгоритмы оптимизации для одной переменной. Следовательно, для применения методов, предназначенных для решения оптимизационных задач общего вида, требуется знакомство с методами оптимизации одномерных моделей.

Начнем с того, что в рамках всего последующего обсуждения ограничимся рассмотрением «максимизации» в смысле оптимизации. [Если в реальной задаче требуется минимизировать целевую функцию $f(x)$, то можно переформулировать модель, максимизируя $-f(x)$.]

Сначала рассмотрим задачу максимизации *линейной* функции одной переменной $c(x) \equiv (c_0 + c_1x)$. Ответ здесь настолько тривиален, что он даже не обсуждался в главах, посвященных линейному программированию, однако в данном разделе он используется как отправная точка для рассмотрения оптимизационных методов в нелинейных ситуациях. Если на x не налагается никаких ограничений, то при любом $c_1 \neq 0$ значение целевой функции может быть сделано сколь угодно большим. (Объясните, каким образом.) Если же x может принимать только ограниченные значения, т. е. $a_1 \leqslant x \leqslant a_2$, то $x = a_1$ оптимально при $c_1 \leqslant 0$, а $x = a_2$ — при $c_1 \geqslant 0$. Следовательно, оптимальным всегда является решение, соответствующее **экстремальной точке.** Однако, если мы откажемся от допущения о линейности функции $c(x)$, становятся возможными и другие ситуации; некоторые из них отображены на рис. 14.1.

Одна из таких возможностей состоит в том, что $c(x)$ достигает максимального значения при конечном значении x, хотя на x не налагается никаких ограничений. На рис. 14.1 этому соответствует пример 2, где оптимум достигается при $x = 2$. Однако, даже если значение x заключено в пределах сегмента $[a_1, a_2]$, решение для экстремальной точки может и не быть оптимумом. Так, если величина x должна находиться в пределах сегмента $[0, 3]$, то оптимальное значение x в примере 2 *не* соответствует ни одной из экстремальных точек этого сегмента ($x = 0$ или $x = 3$), а, как отмечалось выше, расположено внутри сегмента ($x = 2$).

В примере 3, как и в примере 1, значение $c(x)$ неограниченно возрастает, если только на x не наложено ограничений. Однако в отличие от примера 1 сколь угодно большого значения $c(x)$ можно

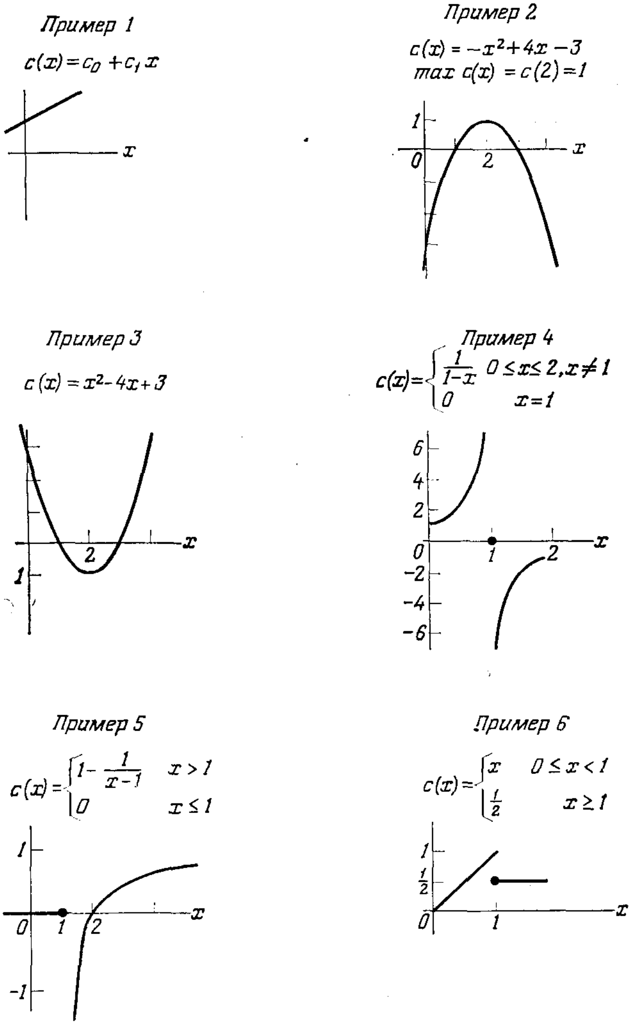


Рис. 14.1. Примеры нелинейных функций.

здесь достичь *как* при достаточно большом, *так и* при достаточно малом значении x. В примере 4 показано, что возможно также достижение сколь угодно больших значений $c(x)$, даже если значение x ограничено, скажем, пределами сегмента $[0, 1]$. Здесь в точке $x = 1$ имеет место разрыв функции $c(x)$; если x стремится к 1 снизу, $c(x)$ неограниченно возрастает.

Наконец, $c(x)$ может быть ограничено сверху, и все же может отсутствовать *конечное* значение x, при котором $c(x)$ достигает максимума. Это характерно для примера 5, где величина x не ограничена. Заметим, что для всех конечных значений x $c(x)$ меньше 1, однако $c(x)$ можно сколь угодно приблизить к 1, приняв величину x достаточно большой. При этом 1 именуется **верхней гранью** функции $c(x)$. В качестве второго примера возьмем пример 6. До каких пределов может здесь возрасти $c(x)$?

Подведем итоги. Функция $c(x)$ может неограниченно возрастать, даже если значения x ограничены; функция $c(x)$, ограниченная сверху, не обязательно достигает максимума, причем вне зависимости от того, ограничены ли значения x. При любом исчерпывающем рассмотрении алгоритмических подходов к максимизации нелинейной функции одной переменной необходимо учитывать все эти различные возможности. Однако в соответствии с целями настоящей главы удобно предположить, что:

I) x лежит в пределах конечного сегмента $I \equiv [a_1, a_2]$, т. е. $a_1 \leqslant x \leqslant a_2$, где $-\infty < a_1 < a_2 < \infty$;

II) функция $c(x)$ непрерывна и ограничена сверху для всех x на сегменте I.

Важным следствием этих двух допущений является то, что $c(x)$ достигает максимума при значении x, находящемся на сегменте I.

Для того чтобы читатель мог проверить, как им усвоен материал, он должен объяснить, почему при допущении II) сегмент I не включает $x = 1$ в примере 4, или $x \geqslant 1$ в примере 5, или $x \leqslant 1$ в примере 6. На рис. 14.1 следует также обратить внимание на то, каким образом форма функции $c(x)$ и характер сегмента I влияют на определение оптимального значения x. Так, например, в примере 3 оптимум *всегда* достигается при одном из экстремальных значений x, т. е. при $x = a_1$ или $x = a_2$; в то же время в примере 2 такое экстремальное значение x соответствует оптимуму *только* в том случае, если $x = 2$ не принадлежит множеству внутренних точек сегмента I. Эти замечания заставляют предположить, что алгоритм оптимизации $c(x)$ при x, принадлежащем I, должен использовать характеристики вида функции $c(x)$.

Сначала рассмотрим, что произойдет, если функция $c(x)$ является монотонной на сегменте I. В этом случае $c(x)$ всегда достигает максимума по x на I в одной из экстремальных точек сегмента, т. е. либо при $x = a_1$, либо при $x = a_2$. То же справедливо, если $c(x)$ является **выпуклой** функцией x на I, т. е. для любых x_1 и x_2 на I, где $x_1 < x_2$, и для любого p, $0 \leqslant p \leqslant 1$, функция $c(x)$ удовлетворяет условию

$$pc(x_1) + (1-p)\,c(x_2) \geqslant c[px_1 + (1-p)\,x_2] \quad \text{(выпуклая).} \qquad (1)$$

Проверьте, что в примере (3) $c(x)$ является выпуклой функцией и поэтому всегда достигает максимума в экстремальной точке I.

В более общем случае $c(x)$ всегда достигает максимального значения в экстремальной точке I, если $c(x)$ есть **квазивыпуклая** функция на I, т. е. если для любых x_1 и x_2 на I, где $x_1 < x_2$, и для любого p, $0 \leqslant p \leqslant 1$ выполняется условие

$$\max [c(x_1),\ c(x_2)] \geqslant c[px_1 + (1-p)\,x_2] \quad \text{(квазивыпуклая).} \qquad (1)$$

Это условие выполняется, если $c(x)$ является либо монотонной, либо выпуклой функцией. Приведем два примера квазивыпуклой функции, не являющейся ни монотонной, ни выпуклой; это модификации функции $c(x)$ из примера 3. Первый пример имеет вид: $c(x) = x^2 - 4x + 3$ при $x \leqslant 3$ и $c(x) \equiv 0$ при $x > 3$. Второй пример: $c(x) \equiv x^2 - 4x + 3$ при $x \neq 2$ и $c(2) \equiv -2$. Начертите графики этих двух функций и проверьте выполнение условия (1).

В том случае если форма $c(x)$ не обеспечивает оптимальность решения, соответствующего экстремальной точке, поиски оптимального значения x следует продолжать во внутренней части сегмента I. Для этого можно разбить сегмент I на части «сеткой» равноудаленных точек и вычислить значение $c(x)$ в каждой из этих точек. Если функция $c(x)$ не является регулярной, то наилучшее из найденных значений x может быть весьма далеким от истинного оптимума, поскольку в этом случае поиск напоминает попытку «найти иголку в стоге сена». Однако, если функция *является* регулярной, т. е. при малом варьировании x изменяется в ограниченных пределах, описанный метод «сетки» обычно можно усовершенствовать, используя так называемые *адаптивные* (или *последовательные*) *методы поиска*. Вместо того чтобы установить совокупность проверяемых точек x заранее, используется информация относительно уже проверенных значений x; на этой основе определяется следующее проверяемое значение x. Это значит, что если имеется ограниченная информация относительно значений $c(x)$, полученная на основе ранее выполненных проверок, то удается сделать разумный вывод относительно возможного диапазона нахождения оптимального значения x. Эта идея поясняется ниже.

Последовательный поиск. Рассмотрим функцию $c(x)$, график которой изображен на рис. 14.2. Максимальное значение $c(x)$ достигается при $x = f$. Как нетрудно представить, для функций такого типа адаптивный процесс поиска максимума $c(x)$ вряд ли окажется эффективным: он может привести к хорошим результатам лишь случайно. Для иллюстрации предположим, что первое проверяемое значение x меньше a. Тогда возможно, что каждое последующее значение x будет все больше и больше приближаться к a, но не будет превышать a, поскольку $c(a)$ есть **локальный максимум.** Далее, предположим, что первое проверяемое значение x находится на сегменте $[d,\ e]$. Тогда, по-видимому, последовательный поиск закончится сразу же, поскольку как при небольшом увеличении, так и при небольшом уменьшении значения x функция $c(x)$ не меняет своей величины и ее улучшение не представляется возможным. Адаптивные методы, которые излагаются ниже, не могут гарантировать, что в конечном счете область поиска сузится до небольшого интервала вокруг f (рис. 14.2). Однако рассмотрение функции $c(x)$ на сегменте $[0, g]$ по существу является более общей ситуацией, чем это действительно необходимо для достижения непосредственно поставленной цели, а именно разработки адаптивных методов поиска оптимума

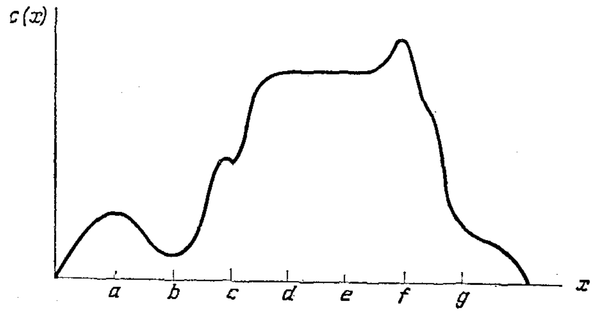


Рис. 14.2. Пример функции, обладающей многими локальными оптимумами.

функций одной переменной, которые в дальнейшем будут использованы при решении многомерных нелинейных задач оптимизации. По этой причине и вводится дополнительный постулат относительно формы функции $c(x)$.

В частности, рассмотрим так называемую **унимодальную** функцию, обладающую следующими свойствами: если x есть единственное оптимальное значение x на сегменте I, то с возрастанием x $c(x)$ строго возрастает при $x \leqslant \bar{x}$ и строго убывает при $x \geqslant \bar{x}$. Если максимальное значение $c(x)$ на I, которое обозначим $\bar{c}$, имеет место не для одной точки, а для некоторого интервала оптимальных значений x, то $c(x)$ строго возрастает до достижения $\bar{c}$, сохраняет это значение вдоль всего интервала оптимальных x, а затем строго убывает. Формальное и более точное определение такой функции имеет следующий вид: функция $c(x)$ является **унимодальной** на сегменте I, если для любых x_1 и x_2 на I, где $x_1 < x_2$ и $c(x_1) < c(\bar{x}) \equiv \bar{c}$, функция $c(x)$ удовлетворяет условиям

$$
\left.\begin{array}{ll}
c(x_1) < c(x_2) \quad \text{при} \quad x_1 < x_2 < \bar{x}, & \\
 & \text{(унимодальная).} \qquad (2) \\
c(x_1) > c(x_2) \quad \text{при} \quad \bar{x} < x_1 < x_2 &
\end{array}\right.
$$

Функция является унимодальной, если она **вогнута**, т. е. если для любых x_1 и x_2 на I, где $x_1 < x_2$, и для любого p, $0 \leqslant p \leqslant 1$, функция $c(x)$ удовлетворяет условию

$$pc(x_1) + (1 - p)\,c(x_2) \leqslant c\,[px_1 + (1 - p)\,x_2] \quad \text{(вогнутая).} \qquad (3)$$

Если функция $c(x)$ имеет регулярную вторую производную d^2c/dx^2, то эквивалентным условием вогнутости является $d^2c/dx^2 \leqslant 0$ для всех x. Если $d^2c/dx^2 < 0$ для всех x, то $c(x)$ **строго вогнута.** Здесь и далее дополнительно к (I) и (II) вводится условие:

$$c(x) \text{ унимодальна по } x \text{ на сегменте } I. \qquad \text{(III)}$$

Нетрудно проверить, что введение условия (III) исключает возможность использования в качестве сегмента I таких сегментов (рис. 14.2), как $[b, g]$, однако условию (III) отвечают сегменты $[b, e]$ и $[e, g]$. Теперь покажем, каким образом введение условия (III) позволяет применить адаптивный метод поиска для нахождения интервала заданной величины, содержащего то значение x, при котором достигается максимум $c(x)$.

Пусть известно, что оптимальное значение x находится на сегменте $[A_1, A_2]$ и вычислены значения $c(x)$ в двух точках $x_{\text{лев}}$ и $x_{\text{пр}}$ этого сегмента, причем $x_{\text{лев}} < x_{\text{пр}}$ (первая точка расположена «левее», а вторая —«правее»). Если $c(x_{\text{лев}}) \geqslant c(x_{\text{пр}})$, то $c(x)$ не может возрастать при $x > x_{\text{пр}}$, так что оптимальное $\bar{x}$ *должно* находиться на сегменте $[A_1, x_{\text{пр}}]$. Начертите график и объясните, почему. Если же $c(x_{\text{лев}}) \geqslant c(x_{\text{пр}})$, то $c(\bar{x})$ не может возрастать при $x > x_{\text{пр}}$, так что оптимальное значение $\bar{x}$ *должно* находиться на сегменте $[A_1, x_{\text{пр}}]$. (Объясните, почему.) В любом случае сравнение $c(x_{\text{лев}})$ и $c(x_{\text{пр}})$ позволило сократить длину сегмента, на котором, как нам известно, должно находиться оптимальное значение x. Как нетрудно себе представить, имеется множество различных алгоритмов последовательного поиска, которые можно разработать с учетом этого свойства унимодальной функции.

Описываемый ниже метод, иногда именуемый **методом золотого сечения,** обладает такими достоинствами, как простота, так и достаточно высокая эффективность. (Этот метод весьма близок к *методу Фибоначчи.*) Введем обозначение:

$$r \equiv 0{,}5\,(\sqrt{5} - 1) = 0{,}618033, \; \ldots \qquad (4)$$

Значение этой константы станет ясным после того, как будет изложен алгоритм поиска максимального значения функции $c(x)$, где x находится на сегменте $I = [a_1, a_2]$.

Шаг 1. Примем $A_1 = a_1$, $A_2 = a_2$, $H = A_2 - A_1$, $x_{\text{лев}} = A_1 + r^2H$ и $x_{\text{пр}} = A_1 + rH$.

Шаг 2. Сравним $c(x_{\text{лев}})$ и $c(x_{\text{пр}})$. В зависимости от результатов сравнения перейдем к шагу 3 или к шагу 4.

Шаг 3. Если $c(x_{\text{лев}}) \geqslant c(x_{\text{пр}})$, примем $A_2 = x_{\text{пр}}$ и $H = x_{\text{пр}} - A_1$. Прекратим расчеты, если H достаточно мало. В противном случае в качестве нового $x_{\text{пр}}$ выберем предыдущее $x_{\text{лев}}$, а в качестве нового $x_{\text{лев}}$ — точку $(A_1 + r^2H)$. Возвратимся к выполнению шага 2.

Шаг 4. Если $c(x_{\text{лев}}) < c(x_{\text{пр}})$, примем $A_1 = x_{\text{лев}}$ и $H = A_2 - x_{\text{лев}}$. Прекратим расчеты, если H достаточно мало. В противном случае в качестве нового $x_{\text{лев}}$ выберем предыдущее $x_{\text{пр}}$, а в качестве нового $x_{\text{пр}}$ — точку $(A_1 + rH)$. Возвратимся к выполнению шага 2.

При заданном в (4) значении r алгоритм обладает тремя важными свойствами. Во-первых, после каждой итерации новое значение H в $1/r$ раз меньше предыдущего значения H. Следовательно, после n итераций оптимальное значение x находится на известном сегменте, длина которого равна $r^n (a_2 - a_1)$. Таблица значений r^n для $n \leqslant 10$ приведена на рис. 14.3.

n	r^n	$(0,5)^{n+1}$
1	0,6180	0,25
2	0,3820	0,125
3	0,2361	0,0625
4	0,1459	0,0313
5	0,0902	0,0156
6	0,0557	0,0078
7	0,0344	0,0039
8	0,0213	0,0020
9	0,0132	0,0010
10	0,0081	0,0005

Рис. 14.3. Показатели сокращения длины сегмента при использовании методов золотого сечения и дихотомии.

Во-вторых, после каждой итерации $x_{\text{лев}}$ и $x_{\text{пр}}$ находятся на одинаковом расстоянии от ближайших к ним концов сегмента, т. е. $x_{\text{лев}} - A_1 = A_2 - x_{\text{пр}}$.

В-третьих, при выполнении шага 2 (на всех итерациях, кроме первой) новой является только одна из точек $x_{\text{лев}}$ и $x_{\text{пр}}$, так что необходимо дополнительно вычислить только одно значение $c(x)$.

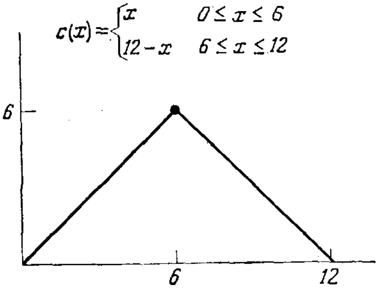
Пример применения алгоритма приведен на рис. 14.4; значение $x = 6$ является единственным оптимальным решением. Исходный сегмент имеет вид $[a_1, a_2] = [0, 10]$. Заметим, что после итерации 2 сегмент сузился до $[A_1, A_2] = [3{,}82; 7{,}64]$. Пробное значение $x_{\text{лев}} = 5{,}28$ превышает A_1 на 1,46 ($5{,}28 - 3{,}82 = 1{,}46$), а пробное значение $x_{\text{пр}} = 6{,}18$, которое уже рассматривалось на итерации 1, меньше A_2 на 1,46 ($7{,}64 - 6{,}18 = 1{,}46$). Покажите, что после завершения итерации 4 новое значение $x_{\text{пр}}$, для которого на шаге 2 следует вычислить значение $c(x)$, равно 6,74. Поскольку $a_2 - a_1 = 10$, длина сегмента $[A_1, A_2]$ после итерации n составляет $10r^n$, где значения r^n приведены в таблице рис. 14.3.

Отметим, что адаптивный процесс обеспечивает только получение информации о сегменте, на котором находится оптимальное значение $\bar{x}$. Лишь при удачном совпадении это значение удается найти за конечное число итераций. По существу, если расчеты выполняются с заданным числом знаков, метод может вообще не дать *точного* ответа о значении $\bar{x}$. Например, если $c(x) = -0{,}25x^4 + 2x$ и $[1, 2]$, оптимальное значение $\bar{x}$ является единственным, оно равно $\bar{x} = \sqrt[3]{2} = 1{,}2599, \ldots$. Однако это число нельзя точно отобразить конечным числом десятичных знаков.

Более того, данный метод не дает информации о максимальном значении $c(x)$. (Если известно, что поиск будет прекращен, как только длина сегмента H окажется меньше некоторой заданной величины, всегда можно построить функцию, для которой максимальное значение $c(x)$ сколь угодно больше, чем наилучшее значение $c(x)$, полученное во время поиска.) Однако предположим, что известно о вогнутости $c(x)$ и что эта функция имеет регулярные производные в точках $x = A_1$ и $x = A_2$, где $A_1 < A_2$. В этом случае *можно* вычислить верхнюю грань

$$c(x) \leqslant \left[\frac{dc(A_1)}{dx} - \frac{dc(A_2)}{dx}\right]^{-1}\left[c(A_2) - c(A_1) + A_1\frac{dc(A_1)}{dx} - A_2\frac{dc(A_2)}{dx}\right] \quad \text{(вогнутая)}$$

для любых x на I. Для того чтобы проверить, правильно ли вы понимаете приведенное выше выражение, убедитесь в том, что в примере на рис. 14.5 эта грань равна 6, т. е. $c(x) \leqslant 6$, на любой из итераций.



Итерация n	Шаг	Расчет
0	1	$A_1=0,\ A_2=10,\ H=10,\ x_{\text{лев}}=3{,}82,\ x_{\text{пр}}=6{,}18$
1	2 4	$c(3{,}82)=3{,}82,\ c(6{,}18)=5{,}82$ $A_1=3{,}82,\ (A_2=10),\ H=10-3{,}82=6{,}18,$ $x_{\text{лев}}=6{,}18,\ x_{\text{пр}}=3{,}82+(0{,}618)\,6{,}18=7{,}64$
2	2 3	$c(6{,}18)=5{,}82,\ c(7{,}64)=4{,}36$ $(A_1=3{,}82),\ A_2=7{,}64,\ H=7{,}64-3{,}82=3{,}82,$ $x_{\text{пр}}=6{,}18,\ x_{\text{лев}}=3{,}82+(0{,}382)\,3{,}82=5{,}28$
3	2 4	$c(5{,}28)=5{,}28,\ c(6{,}18)=5{,}82$ $A_1=5{,}28,\ (A_2=7{,}64),\ H=7{,}64-5{,}28=2{,}36$

Рис. 14.4. Пример прменения метода золотого сечения.

Другой метод адаптивного поиска, именуемый **методом дихотомии** или **методом деления пополам**, можно построить таким образом, что в случае его применения будет использоваться информация о значениях производной функции $c(x)$ для отличных от оптимума значений x. При данном методе учитывается то обстоятельство, что производная унимодальной функции $c(x)$ не меняет знака больше чем один раз. Пусть $c(x)$ не является обычной монотонно возрастающей функцией. Это значит, что при $x = a_1$ производная $c(x)$ является неотрицательной при увеличении x и остается неотрицательной до

Итерация n	Шаг	Расчет
0	1	$A_1=0,\ A_2=10,\ H=10$
1	2 4	$x_{\text{прб}}=5,\ dc/dx=1>0$ $A_1=5,\ (A_2=10),\ H=10-5=5$
2	2 3	$x_{\text{прб}}=5+0{,}5\,(5)=7{,}5,\ dc/dx=-1<0$ $(A_1=5),\ A_2=7{,}5,\ H=2{,}5$
3	2 3	$x_{\text{прб}}=5+0{,}5\,(2{,}5)=6{,}25,\ dc/dx<0$ $(A_1=5),\ A_2=6{,}25,\ H=1{,}25$
4	2 3	$x_{\text{прб}}=5+0{,}5\,(1{,}25)=5{,}625,\ dc/dx>0$ $A_1=5{,}625,\ (A_2=6{,}25),\ H=0{,}625$

Рис. 14.5. Пример применения метода дихотомии.

тех пор, пока $c(x)$ не достигнет максимального значения; затем, при дальнейшем увеличении x, производная неположительна. Следовательно, в точке с оптимальным значением $\bar{x}$ производная меняет знак; при рассматриваемом подходе повторно сужается сегмент, на котором происходит смена знака. (Метод применим и для монотонных функций.)

Для применения указанного метода необходимо вывести формулу производной dc/dx и дополнительно *предположить*, что форма $c(x)$ такова, что dc/dx не только регулярна, но и равна нулю лишь при оптимальном значении x. (Нетрудно построить унимодальные функции на конечном сегменте I, не удовлетворяющие сделанным допущениям о виде функции. Такими функциями будут, например, кусочно-линейные функции при неоптимальных значениях x и функции, имеющие либо точку перегиба, либо точку с бесконечным значением dc/dx.) На каждой итерации при использовании метода дихотомии длина сегмента, на котором находится оптимальное значение $\bar{x}$, сокращается в два раза. После n исчислений значения производной dc/dx длина сегмента составляет $0{,}5^{n+1}\,(a_2 - a_1)$, тогда как после n расчетов значения $c(x)$ метод золотого сечения позволяет сократить сегмент только в $r^n\,(a_2 - a_1)$ раз. Таким образом, коэффициент $0{,}5^{n+1}$ следует сравнивать с коэффициентом r^n, значения которого приведены на рис. 14.3. Вместе с тем в случае использования метода дихотомии возможны значительные вычислительные трудности, если производная dc/dx оказывается более сложным выражением, чем $c(x)$, например когда $c(x) = x^2e^x$.

Идея алгоритма заключается в том, что если для некоторого пробного $x_{\text{прб}}$ $dc/dx > 0$, то существует оптимальное значение $\bar{x}$, такое, что $\bar{x} > x_{\text{прб}}$. Аналогично этому если $dc/dx < 0$, то оптимальное значение $\bar{x} < x_{\text{прб}}$. Опишем предлагаемый алгоритм.

Шаг 1. Пусть $A_1 = a_1$, $A_2 = a_2$ и $H = A_2 - A_1$.

Шаг 2. Примем $x_{\text{прб}} = A_1 + 0{,}5H$. Вычислим значение dc/dx в точке $x_{\text{прб}}$. Прекратим расчеты, если $dc/dx = 0$, поскольку в этом случае $x_{\text{прб}}$ является точкой оптимума. В противном случае в зависимости от знака производной перейдем к шагу 3 или к шагу 4.

Шаг 3. При $dc/dx < 0$ примем $A_2 = x_{\text{прб}}$ и $H = x_{\text{прб}} - A_1$. Прекратим расчеты, если H достаточно мало. В противном случае возвратимся к выполнению шага 2.

Шаг 4. При $dc/dx > 0$ примем $A_1 = x_{\text{прб}}$ и $H = A_2 - x_{\text{прб}}$. Прекратим расчеты, если H достаточно мало. В противном случае возвратимся к шагу 2.

Применение метода дихотомии демонстрируется в таблице рис. 14.5 на решении того же примера, что и на рис. 14.4. После четвертого исчисления dc/dx длина сегмента, где находится оптимальное значение $\bar{x}$, сведено к 0,625 — в сравнении с 2,36 после четырех исчислений $c(x)$ при использовании метода золотого сечения.

14.4. МАКСИМИЗАЦИЯ НЕЛИНЕЙНОЙ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ БЕЗ ОГРАНИЧЕНИЙ

В настоящем и в следующем разделах рассматривается максимизация функций действительных переменных $c(x_1, x_2, \ldots, x_n)$, где каждая из переменных x_j может принимать любые действительные значения, положительные или отрицательные. (Это допущение относительно часто выражают следующими словами: множество значений $x_1, x_2, \ldots, x_n$ содержится в *евклидовом n-мерном пространстве*, которое сокращенно обозначается E^n; или, что эквивалентно, каждая из x_j лежит на действительной оси, сокращенно R^1.) Изучение этой задачи обусловлено двумя причинами. Во-первых, анализ многомерных нелинейных задач на безусловный максимум создает условия для анализа задач на условный максимум. Алгоритмические трудности, которые приходится преодолевать здесь, присущи и задачам с ограничениями, а изложенные ниже методы могут быть должным образом модифицированы при наложении ограничений. Во-вторых, задачу на условный максимум часто удается решить, преобразовав ее в задачу без ограничений. Алгоритм, основанный на таком подходе, рассматривается в разд. 15.9.

Для того чтобы *обеспечить* применимость алгоритмов решения, как и в предыдущих разделах, нам приходится делать известные допущения относительно целевой функции. Выражаясь не совсем строго, мы постулируем, что функция $c(x_1, x_2, \ldots, x_n)$ является гладкой и достигает конечного максимума при конечных значениях $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$. Сокращенно обозначая набор значений $x_1, x_2, \ldots$ $\ldots, x_n$ символом x, а выражение $c(x_1, x_2, \ldots, x_n)$ — символом $c(x)$, сделанные допущения можно записать более строго:

I) При всех значениях x функция $c(x)$ однозначна и конечна.

II) При всех значениях x каждая частная производная $\partial c(x)/\partial x_j$ однозначна, конечна и непрерывна, так что функция $c(x)$ также непрерывна.

III) Функция $c(x)$ обладает конечным максимумом $\bar{c}$.

IV) При любых возможных значениях $c(x)$, например c, существует такое конечное число M_c, что всякая величина $|x_j| \leqslant M_c$, если $c(x) \geqslant c$.

Читателю полезно выписать эти четыре допущения на отдельном листе, поскольку мы будем неоднократно возвращаться к ним на протяжении всей главы.

При наличии первых трех допущений из (IV) следует, что $c(x)$ принимает максимальное значение c при конечных значениях x_1, $x_2, \ldots, x_n$. Фактически же из (IV) следует еще более сильное свойство: если задано любое возможное значение c функции $c(x)$, то значения $c(x)$, превышающие c, могут соответствовать лишь x_j, находящимся в пределах конечных сегментов, т. е. $-M_c \leqslant x_j \leqslant M_c$ для каждого j (причем $M_c < \infty$ и M_c зависит от c).

Для того чтобы почувствовать силу сделанных допущений, рассмотрим пример 5, приведенный на рис. 14.1. Изображенная здесь функция $c(x)$ исключается, во-первых, из-за допущения II), поскольку dc/dx в точке $x = 1$ разрывна. Но даже если изменить $c(x)$ таким образом, что dc/dx станет непрерывной, причем $c(x) < 1$ для всех x, этот пример по-прежнему исключается из-за допущения III), поскольку функция $c(x)$ не имеет максимума. [Хотя для всех x $c(x) \leqslant 1$, однако не существует такого конечного x, что $c(x) = 1$.] Более того, даже если функцию $c(x)$ дополнительно изменить так, чтобы она обладала конечным максимумом, например при $x = \frac{1}{2}$ $\bar{c} = 3$, пример 5 не удовлетворяет допущению (IV). Для выяснения причины этого предположим, что в (IV) $c = \frac{1}{3}$, тогда $c(x) \geqslant \frac{1}{3}$ при всех $x \geqslant 2{,}5$. Следовательно, не существует значения M_c, постулированного в допущении IV).

Используя математическую терминологию, можно сформулировать эти четыре допущения более кратко. Во-первых, допущение II) можно переформулировать, сказав, что функция $c(x)$ должна быть *непрерывно дифференцируемой*; из этого следует допущение I) и непрерывность $c(x)$. Во-вторых, поставим условие, что для любого x' множество x, таких, что $c(x) \geqslant c(x')$, является *компактным* (замкнутым и ограниченным); из этого следует допущение IV). Поскольку непрерывная функция имеет конечный максимум на компактном множестве, два кратко сформулированных постулата имеют следствием допущение III).

Пользуясь дифференциальным исчислением, можно сформулировать следующее положение.

Необходимое условие существования максимума. При допущениях I) — III) функция $c(x)$ имеет максимум в $\bar{x}$ только в том случае, если $\partial c(\bar{x})/\partial x_j = 0$ для всех $j = 1, 2, \ldots, n$.

Справедливость этого положения нетрудно понять. Пусть имеется переменная x_j, такая, что $\partial c(\bar{x})/\partial x_j > 0$. Тогда $c(x)$ можно увеличить, незначительно увеличив значение x_j по сравнению с $\bar{x}_j$. Аналогично этому, если $\partial c(\bar{x})/\partial x_j < 0$, $c(x)$ можно увеличить, незначительно уменьшив x_j по сравнению с $\bar{x}_j$.

Каковы вычислительные следствия необходимого условия? Сразу же можно сделать вывод, что некоторый набор значений $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ не соответствует максимуму $c(x)$, если только *все* $\partial c(\bar{x})\, dx_j$ не равны нулю. Однако, к сожалению, необходимое условие *недостаточно* для достижения максимума, если не будут наложены дополнительные ограничения на вид функции $c(x)$: $\bar{x}$ может и не соответствовать максимуму $c(x)$, даже если все $\partial c(\bar{x})/\partial x_j = 0$. Кривая, изображенная на рис. 14.2, показывает причину этого. Производная $dc/dx = 0$ в точках a, b, c, d, e, а также в f, которая является *единственным* глобальным максимумом.

На мгновение представим себе, что справедливо следующее утверждение: для заданной функции $c(x)$ необходимое условие одновременно является достаточным. Значение $\bar{x}$ оптимально тогда и только тогда, когда все $\partial c(\bar{x})/\partial x_j = 0$; значит ли это, что нахождение $\bar{x}$ становится тривиальным? Вероятно, нет, потому что обычно условия $\partial c/\partial x_j = 0$, $j = 1, 2, \ldots, n$ приводят к необходимости решения системы n *нелинейных* уравнений, а это может оказаться не менее трудной вычислительной задачей, чем нахождение максимума $c(x)$. (В самом деле, часто используемый метод решения системы нелинейных уравнений $a_i(x) = 0$ для $i = 1, 2, \ldots, m$ состоит в решении вместо этого соответствующей оптимизационной задачи: минимизировать $\sum_{i=1}^{m} [a_i(x)]^2$.) Заметим, однако, что если $c(x)$ есть квадратичная функция, то условия $\partial c/\partial x_j = 0$ для $j = 1, 2, \ldots, n$ приводят к системе n *линейных* уравнений, которая обычно легко поддается решению. В более общем случае оптимальное значение $\bar{x}$ отыскивается с помощью итеративного процесса описанного ниже типа.

Алгоритмический метод. Многие вычислительные методы максимизации $c(x)$ можно представить в стандартном виде.

Шаг 1. Выберем в качестве исходной произвольную пробную точку x^0.

Шаг 2. Прекратим расчеты, если в пробной точке x^k $\partial c/\partial x_j = 0$ для $j = 1, 2, \ldots, n$. В противном случае определим значения y_j^k для $j = 1, 2, \ldots, n$ и перейдем к шагу 3.

Шаг 3. Определим новую пробную точку

$$x_j^{k+1} = x_j^k + y_j^k \text{ для } j = 1, 2, \ldots, n. \tag{1}$$

Возвратимся к выполнению шага 2, заменив x^k на x^{k+1}.

Для большинства нелинейных целевых функций итеративный процесс не позволяет найти такое значение x^k, что все $\partial c/\partial x_j = 0$. Следовательно, по критерию окончания расчетов, указанному в шаге 2, выполнение алгоритма никогда не завершается; расчеты приходится прекращать, когда $c(x^k)$ близко к оптимуму. Для того чтобы обеспечить конечность числа итераций, можно, например, установить предельное их число или же завершать расчеты, когда все значения $|\partial c/\partial x_j|$ окажутся достаточно малы.

В различных алгоритмах используются разные методы выбора y_j^k на шаге 2. Многие из этих методов основаны на следующей идее. Во-первых, выбирается так называемое **направление** d_j^k для $j = 1, 2, \ldots, n$; этот выбор обычно основан на информации о поведении $c(x)$ вблизи проверяемой точки x^k, т. е. на *локальных* свойствах $c(x)$. Во-вторых, выбирается **длина шага** t^k, основанная на информации о поведении $c(x)$ за пределами окрестности точки x^k при движении в выбранном направлении $d_1^k, d_2^k, \ldots, d_n^k$. Наконец, направление и длина шага объединяются в величинах

$$y_j^k = t^k d_j^k, \quad j = 1, 2, \ldots, n. \tag{2}$$

Визуальное представление процесса поможет читателю запомнить смысл этих терминов.

Пусть требуется максимизировать функцию $c(x_1, x_2) \equiv -(x_1 - 3)^2 - 4(x_2 - 2)^2$. Исследуя $c(x)$, можно установить, что единственным оптимальным решением является $\bar{x}_1 = 3$, $\bar{x}_2 = 2$, при которых $c(\bar{x}) = 0$. [Поскольку $c(x)$ является квадратичной функцией, оптимальное значение x можно вычислить непосредственно из необходимых условий $\partial c/\partial x_j = 0$ для $j = 1, 2$. Тем не менее мы используем квадратичный пример для демонстрации основных идей алгоритма.] Линии уровня $c(x_1, x_2)$ имеют вид эллипсов и изображены на рис. 14.6. Начнем с шага 1 и примем $x_1^0 = x_2^0 = 0$, а значит, $c(x^0) = -25$. Поскольку

$$\frac{\partial c}{\partial x_1} = -2(x_1 - 3) \quad \text{и} \quad \frac{\partial c}{\partial x_2} = -8(x_2 - 2), \tag{3}$$

а обе эти производные в точке x^0 строго положительны, результаты, полученные на шаге 2, указывают на возможность дальнейшего улучшения.

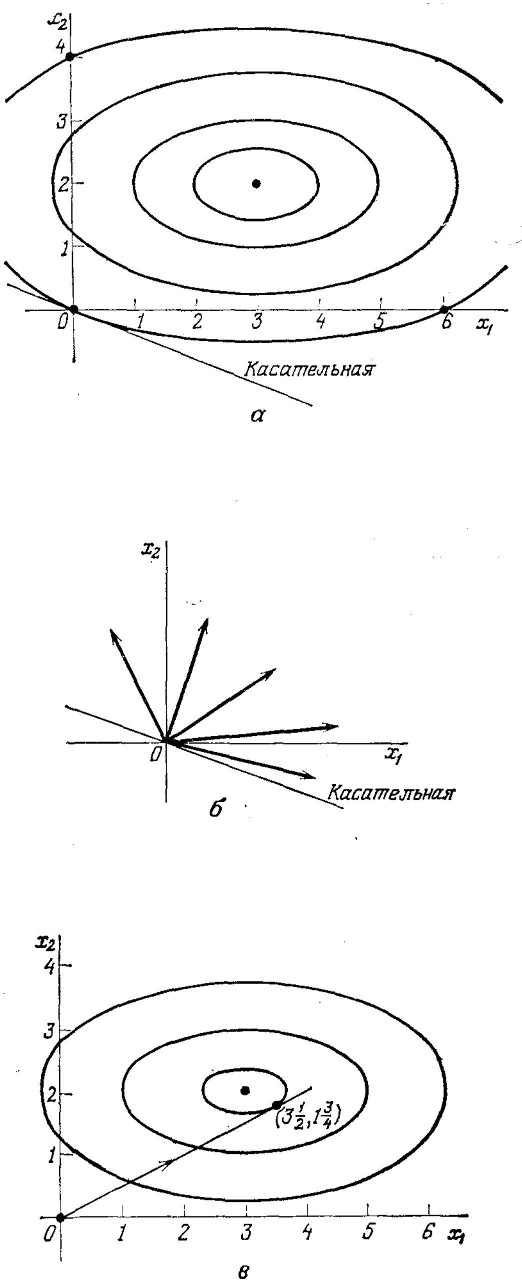


Рис. 14.6. *а* — линии уровня; *б* — направления; *в* — оптимальная длина шага.

В процессе подготовки к выбору направления d_1 и d_2 проведем касательную к линии уровня $c(x)$ в точке x^0, описываемую уравнением

$$\frac{\partial c}{\partial x_1}(x_1 - x_1^0) + \frac{\partial c}{\partial x_2}(x_2 - x_2^0) = 0. \tag{4}$$

Проверьте, что при $x_1^0 = x_2^0 = 0$ уравнение (4) имеет следующий вид:

$$6x_1 + 16x_2 = 0 \quad \text{или} \quad x_2 = -\frac{3}{8}x_1; \tag{5}$$

эта касательная нанесена на рис. 14.6, *а*. Направление $(d_1 d_2)$ можно показать на графике в виде вектора, выходящего из исходной точки. Некоторые возможные направления изображены на рис. 14.6, *б*. Значение $c(x)$ возрастает по сравнению с $c(x) = -25$ при движении вдоль *любого* вектора направления (d_1, d_2), выходящего из x^0 и лежащего выше касательной, *при условии*, что длина шага, сделанного в этом направлении, не слишком велика. Какое бы направление мы ни выбрали, при слишком большом шаге $c(x)$ в конце концов уменьшается.

Если нам повезло и мы остановились на d_1^0 и d_2^0, пропорциональных 3 и 2 соответственно, например выбрав $d_1^0 = 1$ и $d_2^0 = \frac{2}{3}$, то можно получить оптимальное решение, если и длина шага окажется выбранной правильно. Однако при максимизации нелинейной функции $c(x)$ произвольного вида в общем случае не имеется возможности обеспечить выбор такого направления, которое всего лишь за одну итерацию приведет к оптимальному решению. Предположим, что в данном примере мы выбрали $d_1^0 = 2$ и $d_2^0 = 1$, как это показано на рис. 14.6, *в*. Если мы хотим максимально увеличить значение $c(x)$, двигаясь в этом направлении, то следует продолжить вектор ($d_1^0 = 2$, $d_2^0 = 1$) до той точки, в которой он окажется касательной к некоторой линии уровня $c(x_1, x_2)$. (Объясните, почему.) Это показано на рис. 14.6, *в*, где вектор касается линии уровня в точке $x_1 = 3\frac{1}{2}$, $x_2 = 1\frac{3}{4}$, так что формулы (1) и (2) приобретают следующий вид:

$$\begin{aligned} x_1^1 &= 0 + 1\tfrac{3}{4}\cdot 2 = 3\tfrac{1}{2}, \\ x_2^1 &= 0 + 1\tfrac{3}{4}\cdot 1 = 1\tfrac{3}{4}, \end{aligned} \tag{6}$$

где $t^0 = 1\frac{3}{4}$ есть оптимальный размер шага. В новой точке, характеризуемой (6), $c(x^1) = -\frac{1}{2}$.

Если заданы направления d_j^k для $j = 1, 2, \ldots, n$, то проблема выбора соответствующей **оптимальной длины шага** является всего лишь одномерной оптимизационной задачей:

$$\underset{\text{при } t \geqslant 0}{\text{Максимизировать}}\ c(x_1^k + td_1^k,\ x_2^k + td_2^k,\ \ldots,\ x_n^k + td_n^k) = c(x_1^k + t^k d_1^k,\ x_2^k + t^k d_2^k,\ \ldots,\ x_n^k + t^k d_n^k) = c(x^{k+1}). \qquad (7)$$

Иными словами, каждый из аргументов функции $c(x)$ попросту имеет вид $(x_j^k + td_j^k)$, или, в краткой записи, $(x^k + td^k)$; следовательно, функция $c(x^k + td^k)$ зависит от *единственной* переменной t. Оптимальной длиной шага t^k является такое значение t, которое максимизирует $c(x^k + td^k)$, что символически представлено в (7); новая точка x — это $(x^k + t^k d^k)$.

Продемонстрируем это на примере; пусть $x_1^0 = x_2^0 = 0$ и выбраны $d_1^0 = 1$ и $d_2^0 = 2$. Тогда в соответствии с (7) необходимо максимизировать

$$c(0 + t\cdot 1,\ 0 + t\cdot 2) = -(t-3)^2 - 4(2t-2)^2. \qquad (8)$$

Поскольку выражение в правой части (8) есть вогнутая квадратичная функция t, оптимальное значение t можно найти аналитически, приравняв производную нулю. Проверьте, что это дает $t^0 = 1\frac{2}{9}$. Для более сложных функций $c(x)$ нахождение *оптимальной длины шага* может оказаться нелегкой задачей, если только форма функции $c(x)$ не даст возможность использовать методы, описанные в разд. 14.3. К этим вопросам мы возвратимся в следующем разделе, где будет рассматриваться случай максимизации вогнутой функции $c(x)$.

Теперь можно указать на основное различие между линейными и нелинейными оптимизационными задачами. В линейной задаче значение целевой функции изменяется на c_i при изменении x_i на единицу вне зависимости от значения x_i и от значений всех прочих x_j. В нелинейной задаче изменение значения целевой функции при изменении x_i на единицу может зависеть от значений x_i и всех остальных x_j. Именно эта зависимость является источником трудностей при выборе направления на шаге 2.

Рассмотрим процесс, где на каждой итерации выбирается как направление, позволяющее строго улучшить значение $c(x)$, так и соответствующая оптимальная длина шага. Как показывает приводимый ниже пример, хотя при таком процессе $c(x)$ стремится к некоторому пределу, этот предел не обязательно должен представлять собой максимальное значение $c(x)$.

Пусть $c(x) \equiv -(x_1)^2 - (x_2)^2$; это строго вогнутая функция, единственный максимум которой достигается при $x_1 = x_2 = 0$. Пусть, далее, x^0 — точка на окружности с центром в начале координат, а радиус окружности равен 2. При $k = 0, 1, 2, \ldots$ будем выбирать точку x^{k+1} таким образом, чтобы прямая через точки x^k и x^{k+1} была касательной к окружности радиуса $(k+2)/(k+1)$. Можно убедиться в том, что значение $c(x)$ на каждой итерации улучшается и что в данном процессе длина шага является оптимальной, поскольку прямая — касательная к линии уровня. При возрастании k пробные точки лежат на спирали, все больше и больше приближающейся к окружности с центром в начале координат и с единичным радиусом. Таким образом, значения $c(x)$ на последовательных итерациях строго возрастают и в пределе стремятся к $-(1)^2-(1)^2=-2$, тогда как истинное значение максимума $c(x)$ в точке начала координат равно нулю. Как видно из этого примера, направления на шаге 3 должны выбираться с известной осторожностью, даже если функция $c(x)$ регулярна.

14.5. МЕТОД СКОРЕЙШЕГО ПОДЪЕМА

Как отмечалось выше, большинство правил выбора направления d_j^k для $j=1, 2, \ldots, n$ являются **«недальновидными»** в том смысле, что в них используется информация относительно $c(x)$, относящаяся лишь к текущей пробной точке x^k. Ниже описывается хорошо известный вычислительный процесс такого типа, применяемый на шаге 2.

Метод наискорейшего подъема с оптимальной длиной шага. а) Примем все $d_j^k=\partial c/\partial x_j$ в пробной точке x^k. б) Найдем $t^k \geqslant 0$, при котором максимизируется $c(x_1^k+td_1^k, \ldots, x_n^k+td_n^k)$. в) На шаге 2 примем что, все $y_j^k=t^k d_j^k$.

Этот подход впервые был предложен в 1847 г. Коши. Идея использовать в качестве направления набор частных производных, обычно именуемый **градиентом** функции $c(x)$ в точке x^k и сокращенно обозначаемый символом $\nabla c(x)$, является интуитивно привлекательной. Нестрого говоря, градиент представляет собой направление скорейшего возрастания $c(x)$ в окрестности точки x^k. Тем не менее данный подход не может обеспечить быструю сходимость к оптимальному решению, поскольку, как и для всех «недальновидно» определенных направлений, рост значений функции $c(x)$ может замедлиться, как только мы сдвинемся по выбранному направлению за пределы непосредственной окрестности точки x^k.

Так называемая **производная по направлению** функции $c(x)$, взятая в точке x по направлению $d_1, d_2, \ldots, d_n$, определяется как предел выражения

$$\frac{c(x_1+hd_1, \ldots, x_n+hd_n)-c(x_1, \ldots, x_n)}{h\sqrt{\sum\limits_{j=1}^{n}(d_j)^2}} \tag{I}$$

при стремлении к нулю положительного h. Если учесть допущения о гладкости $c(x)$, приведенные в разд. 14,4, этот предел равен

$$\left(\sum_{j=1}^{n}\frac{\partial c}{\partial x_j}d_j\right)\Big/\sqrt{\sum_{j=1}^{n}(d_j)^2}. \tag{II}$$

Если d_j нормированы таким образом, что значение знаменателя (II) равно 1, то производная по направлению достигает максимума при d_j, пропорциональных $\partial c/\partial x_j$. Иными словами, скорость прироста значения $c(x)$ в расчете на единицу евклидова расстояния является наибольшей при движении в направлении градиента. Именно в этом смысле и говорится, что градиент есть направление наискорейшего подъема.

Примеры. В качестве первого примера снова рассмотрим функцию $c(x) \equiv -(x_1-3)^2 - 4(x_2-2)^2$. Из правила скорейшего подъема следует, что при $x_1^0 = x_2^0 = 0$ $d_1^0 = 6$ и $d_2^0 = 16$; это соответствует приведенным в (3) разд. 14.4 значениям частных производных. При оптимальной длине шага должно максимизироваться значение функции $[-(6t-3)^2 - 4(16t-2)^2]$, что дает $t^0 = \frac{73}{530}$. В результате получаем

$$\begin{aligned} x_1^1 &= 0 + \frac{73}{530}\cdot 6 = \frac{219}{265} \approx 0{,}83; \\ x_2^1 &= 0 + \frac{73}{530}\cdot 16 = \frac{584}{265} \approx 2{,}20, \end{aligned} \tag{1}$$

откуда $c(x') = -4{,}89$; эта точка показана на рис. 14.7. (Заметим, что движение по градиенту не привело к столь же большому росту $c(x)$, как движение по направлению в разд. 14.4, приведшему к результату (6).) Продолжая расчеты в течение еще одной итерации, получим направление для точки x^1:

$$\begin{aligned} \frac{\partial c}{\partial x_1} &= -2\left(\frac{219}{265} - 3\right) = \frac{1152}{265}; \\ \frac{\partial c}{\partial x_2} &= -8\left(\frac{584}{265} - 2\right) = -\frac{432}{265}. \end{aligned} \tag{2}$$

Отметим, что на рис. 14.7 это направление является **ортогональным**, т. е. перпендикулярным к предыдущему направлению. Это объясняется тем, что прежний градиент одновременно являлся касательной к линии уровня, проходящей через точку x^1, поскольку был использован шаг оптимальной длины, а новый градиент по самой своей сущности перпендикулярен к этой касательной в точке x^1. Это свойство ортогональности выполняется для всех попарно взятых последовательных градиентов и математически может быть выражено в следующем виде:

$$\sum_{j=1}^{n} \frac{\partial c(x^2)}{\partial x_j} \cdot \frac{\partial c(x^{k+1})}{\partial x_j} = 0. \tag{3}$$

В качестве второго примера рассмотрим максимизацию функции

$$c(x) \equiv 6x_1 + 4x_2 + 2x_3 - 3x_1^2 - 2x_2^2 - \frac{1}{3}x_3^2, \tag{4}$$

частные производные которой имеют вид

$$\frac{\partial c}{\partial x_1}=6(1-x_1),\quad \frac{\partial c}{\partial x_2}=4(1-x_2),\quad \frac{\partial c}{\partial x_3}=2\left(1-\frac{1}{3}x_3\right). \tag{5}$$

Согласно необходимому условию существования максимума, приведенному в разд. 14.4, $\bar{x}_1=1$, $\bar{x}_2=1$ и $\bar{x}_3=3$, что действительно

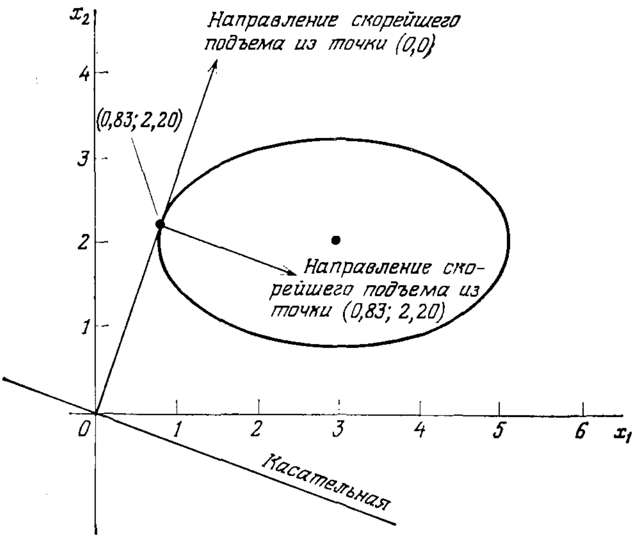


Рис. 14.7. Применение метода скорейшего подъема.

является глобальным оптимумом. (Этот же пример повторно используется в следующих далее разделах, так что читателю для удобства пользования следует выписать (4), (5) и $\bar{x}$ на отдельном листе.) При *любом* заданном наборе d_j и длине шага t выражение (4) становится квадратичной функцией t:

$$6(x_1+td_1)+4(x_2+td_2)+2(x_3+td_3)-$$
$$-3(x_1+td_1)^2-2(x_2+td_2)^2-\frac{1}{3}(x_3+td_3)^2. \tag{6}$$

Далее, в данном примере можно показать, что выражение (6) есть вогнутая функция t и что оптимальную длину шага можно найти аналитически, приравняв производную нулю и решив полученное уравнение относительно t. В результате этих расчетов оптимальное значение t оказывается равным

$$\frac{6(1-x_1)d_1+4(1-x_2)d_2+2\left(1-\frac{1}{3}x_3\right)d_3}{6(d_1)^2+4(d_2)^2+\frac{2}{3}(d_3)^2}. \tag{7}$$

При условии $d_j = \partial c/\partial x_j$, выполняемом в методе скорейшего подъема, оптимальное значение t никогда не является отрицательным.

Результаты применения метода скорейшего подъема приведены на рис. 14.8, где для исходной точки все $x_j^0 = 0$. Обратите внимание на то, как последовательные значения x_1^k и x_2^k колеблются относительно оптимальных значений x_1 и x_2. Обратите внимание также и на то,

Итерация k	$c\,(x_1^k,\ x_2^k,\ x_3^k)$	x_1^k	x_2^k	x_3^k	Оптимальное значение t^k
0	0	0	0	0	0,1981
1	5,547	1,189	0,729	0,396	0,4001
2	6,544	0,736	1,125	1,091	0,2554
5	7,646	1,088	0,999	2,004	0,3642
6	7,778	0,895	1,000	2,246	0,2[illegible]51
10	7,965	0,959	1,000	2,702	0,2551
11	7,978	1,022	0,999	2,753	0,3642
18	7,999	0,994	1,000	2,954	0,2551
19	7,999	1,003	1,000	2,961	0,3642
20	7,999	0,996	1,000	2,971	—
Оптимальные значения	8	1	1	3	0

Рис. 14.8. Пример применения метода скорейшего подъема с оптимальной длиной шага.

$$c\,(x) = 6x_1 + 4x_2 + 2x_3 - 3x_1^2 - 2x_2^2 - \frac{1}{3}\,x_3^2$$

насколько быстро возрастает значение целевой функции на нескольких первых итерациях; после итерации 5 полученное значение $c\,(x)$ отличается от истинного максимума менее чем на 5%. В противоположность этому для получения значений x_j^k, которые для всех j менее чем на 5% отличаются от соответствующих оптимальных значений, требуется количество итераций, превышающее указанную величину более чем вдвое.

Специфика данного примера может быть использована для того, чтобы показать, почему движение в направлении градиента *не всегда* обусловливает быструю сходимость. Предположим, что при выборе направления мы используем не градиент, а подход, основанный на «поочередном изменении переменных». Поскольку в рассмотренном примере в $c\,(x)$ не входят слагаемые, включающие произведения x_j для разных j, как, например, x_1x_2, точное оптимальное решение достигается за три итерации (вне зависимости от порядка выбора переменных). В то же время в данном примере метод скорейшего подъема с исходной точкой $x_j^0 = 0$ для всех j обеспечивает сходимость лишь в пределе.

Модификация метода скорейшего подъема. Предыдущий алгоритм называется методом *первого порядка*, поскольку в нем используются только частные производные первого порядка. Если его не модифицировать, метод скорейшего подъема обычно обеспечивает медленную сходимость после первых итераций, а также обладает другими неприятными особенностями. Так, например, свойство ортогональности (3) обусловливает **зигзагообразность** движения, продемонстрированную на рис. 14.7. Существует несколько численных методов, позволяющих модифицировать метод скорейшего подъема таким образом, чтобы избежать подобных зигзагов, что ускоряет сходимость.

Один из таких методов, названный **методом Ньютона — Рафсона,** использует направление, установленное исходя из квадратичной аппроксимации целевой функции. В этом методе вторые частные производные используются для построения средневзвешенной первых частных производных; метод относится к категории так называемых **модифицированных градиентных методов.**

В частности, d_j^k находятся на основе решения системы линейных уравнений

$$\sum_{j=1}^{n} \frac{\partial^2 c}{\partial x_i \, \partial x_j} d_j = -\frac{\partial c}{\partial x_i}, \quad i = 1,\ 2,\ \ldots,\ m, \tag{8}$$

где предполагается, что $c(x)$ таково, что (8) всегда имеет регулярное решение в каждой пробной точке x^k. Этот метод второго порядка в случае его применения к квадратичной функции $c(x)$ позволяет при любом x^0 за одну итерацию попасть в стационарную точку. Если же принять все $x_j^0 = 0$, то решение системы (8) эквивалентно решению системы $\partial c/\partial x_j = 0$ для $j = 1,\ 2,\ \ldots,\ n$.

При другом методе ускоренного поиска, именуемом **методом субрелаксации,** длина шага принимается меньшей, чем оптимальная. Множество полезных модификаций метода скорейшего подъема слишком велико, а вопрос является слишком специальным, чтобы рассматривать его более подробно в настоящем контексте.

Достопримечательным свойством градиентного подхода является то, что скорость сходимости может зависеть от постановки задачи. Так, например, в приведенном выше примере, где рассматривалась функция $c(x) = -(x_1 - 3)^2 - 4(x_2 - 2)^2$, положим $z = 2x_2 - 2$, так что $x_2 = 0{,}5(z + 2)$. Тогда $c(x_1, z) = -(x_1 - 3)^2 - (z - 2)^2$, и оптимальны $x_1 = 3$ и $z = 2$. Далее, $\partial c/\partial x_1 = -2(x_1 - 3)$ и $\partial c/\partial z = -2(z - 2)$. Приняв $x_1^0 = z^0 = 0$, получим направления скорейшего подъема $d_1^0 = 6$ и $d_2^0 = 4$; следовательно, $t^0 = \frac{1}{2}$ и x_1^1 и z' являются оптимальными. Таким образом, выполненная замена переменных позволила добиться того, что метод скорейшего подъема обеспечивает сходимость за одну итерацию.

Сходимость алгоритма скорейшего подъема. Приведенные выше примеры функции $c(x)$ были достаточно общими для того, чтобы продемонстрировать последовательность расчетов по методу скорейшего подъема с шагом оптимальной длины. Однако конкретное поведение процесса, характерное для этих примеров, может оказаться обманчивым, поскольку они были квадратичными и удовлетворяли необходимому условию существования максимума в единственной точке. Можно задать вопрос: как этот алгоритм проявит себя применительно к такой функции, которая приведена на рис. 14.2? Можно показать, что если $c(x)$ удовлетворяет допущениям I) — IV) разд. 14.4, то для произвольного исходного проверяемого значения x^0 выполняются следующие условия:

А) Вся последовательность $c(x^k)$ возрастает, стремясь к пределу c^*.

Б) По крайней мере некоторая подпоследовательность x^k сходится к точке x^*.

В) В точке x^* целевая функция принимает значение $c(x^*) = c^*$ и необходимое условие существования максимума удовлетворяется, иными словами, все $\partial c/\partial x_j = 0$.

Следовательно, не только последовательность $c(x^k)$ является строго возрастающей, но при достаточно больших k величина $c(x^k)$ подходит к предельному значению c^* сколь угодно близко. Однако в общем случае не существует такого *конечного* k, что $c(x^k) = c^*$ или же $x^k = x^*$; в нестрогой формулировке **сходимость достигается лишь в бесконечности.**

К сожалению, три перечисленных выше свойства не означают, что вся последовательность x^k сходится к такой точке $\bar{x}$, для которой значение $c(\bar{x})$ является максимальным. Во-первых, c^* не обязательно должно оказаться максимальным значением $c(x)$. Во-вторых, на рис. 14.2 x^* может быть такой точкой, как a, c (если $x^0 = c$), d или e. (В многомерной задаче x^* может быть седловой точкой.) Следовательно, x^* может быть всего лишь так называемой **стационарной точкой,** т. е. точкой, удовлетворяющей необходимому условию существования максимума.

Свойство Б) можно пояснить на двумерном примере. В нем утверждается следующее: существует такая точка x^*, что если начертить окружность с центром в x^* и принять радиус сколь угодно малым, то в пределах этой окружности будет лежать бесконечное множество точек x^k. Однако то обстоятельство, что, скажем, x^{1000} лежит внутри окружности, *не обеспечивает* нахождения внутри этой окружности следующей точки x^{1001}. Хотя *вся* последовательность x^k не обязательно должна сходиться к x^*, в практических задачах она обычно все же сходится к ней.

Если на функцию $c(x)$ не наложить дополнительных ограничений, сформулированные выше свойства сходимости — лучшее, чего можно добиться. Если не имеется достаточных сведений о виде некоторой конкретной функции $c(x)$, то обычно алгоритм применяется несколько раз, причем каждый раз начинают из новой исходной точки x^0, удаленной и от ранее использованных x^0, и от ранее полученных решений x^*.

Свойства А) и Б) можно доказать на основе положения, хорошо известного в математическом анализе, а именно что всякая бесконечная последовательность имеет в компактном пространстве сходящуюся подпоследовательность. (Это положение часто называют теоремой Больцано — Вейерштрасса.) Из допущений I), III) и IV) разд. 14.4 следует, что множество точек $[x, c(x)]$ в E^{n+1}, для которого $c(x) \geqslant c(x^0)$, является компактным; следовательно, имеется подпоследовательность, сходящаяся к $[x^*, c^*]$. Поскольку $c(x^{k+1}) \geqslant \geqslant c(x^k)$, вся последовательность $c(x^k)$ в пределе стремится к c^*. Таким образом, свойствами А) и Б) обладает любой алгоритм, при использовании которого значение $c(x)$ улучшается на каждой итерации. Конкретные особенности метода скорейшего подъема с *оптимальной* длиной шага используются при доказательстве свойства В).

Обозначим X^i общий член последовательности, сходящейся к x^*. Тогда в силу непрерывности $c(x)$

$$c(x^*) = \lim c(X^i) = c^*. \tag{I}$$

Сокращенно обозначив $c[X_1^i + t^i(\partial c/\partial x_1), \ldots, X_n^i + t^i(\partial c/\partial x_n)]$ символом $c[X^i + t^i \nabla c(X^i)]$, получим

$$c(X^{i+1}) \geqslant c[X^i + t^i \nabla c(X^i)] \geqslant c[X^i + t\nabla c(X^i)] \quad \text{для} \quad t \geqslant 0, \tag{II}$$

где неравенство (I) следует из процесса «скорейшего подъема», а (II)— из процесса выбора «оптимальной длины шага». Перейдя к пределам в обеих частях (II) при одновременном использовании допущения о непрерывности каждой $\partial c/\partial x_j$, а также (I), получим

$$c(x^*) \geqslant c[X^* + t\nabla c(x^*)] \quad \text{для всех} \quad t \geqslant 0. \tag{III}$$

Далее, (III) в свою очередь означает, что $\nabla c(x^*) = 0$, т. е. в x^* каждая $\partial c/\partial x_j = 0$, потому что в противном случае существует положительное значение t, такое, что правая часть (III) строго больше, чем $c(x^*)$, а это противоречит неравенству (III). Следовательно, свойство В) доказано.

Вогнутые целевые функции. Предположим, что на функцию $c(x)$ наложено дополнительное ограничение, являющееся допущением относительно ее вида. Функция многих переменных $c(x)$ именуется **вогнутой**, если для любых двух точек x и y, $x \neq y$, и для любого p, $0 \leqslant p \leqslant 1$, выполняется условие

$$pc(x) + (1-p)c(y) \leqslant$$
$$\leqslant c[px_1 + (1-p)y_1, \ldots, px_n + (1-p)y_n] \quad \text{(вогнутая).} \tag{9}$$

Эта функция именуется **строго вогнутой**, если для $0 < p < 1$ неравенство (9) является строгим. Используя геометрические представления, можно сказать, что функция называется вогнутой, если отрезок прямой, соединяющий любые две точки на кривой этой функции, расположен на кривой или под ней, и строго вогнутой если отрезок всегда расположен под кривой (за исключением точек x и y). При другом, эквивалентном определении функция считается вогнутой, если соответствующая ей поверхность расположена на касательной к ней гиперплоскости или под ней. В примере 2 (рис. 14.1) показана строго вогнутая функция; если бы ее видоизменить так, чтобы она включала отрезок прямой, то она была бы просто вогнутой. На рис. 14.6 также изображена строго вогнутая функция. Теперь мы можем сформулировать следующее положение:

Достаточное условие существования максимума вогнутой функции. При справедливости допущений I) — IV) разд. 14.4 и при вогнутости функции $c(x)$, если все $\partial c/\partial x_j = 0$ в точке $\bar{x}$, то $c(\bar{x})$ есть максимальное значение $c(x)$. Если же $c(x)$ строго вогнута, то $\bar{x}$ — единственный максимум.

Следовательно, *локальный* максимум вогнутой функции одновременно является глобальным максимумом. Отсюда вытекает важное следствие: если применяется метод скорейшего подъема с шагом оптимальной длины, то последовательность $c(x^k)$ в пределе возрастает до максимального значения $c(x)$; если же функция строго вогнута, *вся* последовательность x^k сходится к единственному оптимальному решению $\bar{x}$. Еще одно существенное с вычислительной точки зрения следствие состоит в том, что выбор оптимальной длины шага может основываться на методе поиска, описанном в разд. 14.3. Это объясняется тем, что если $c(x)$ вогнута, то $c(x_1 + td_1, \ldots, x_n + td_n)$ есть вогнутая функция t. На каждой итерации необходимо устанавливать верхнюю грань $\bar{t}$ оптимальной длины шага, а затем отыскивать лучшее значение t на сегменте $[0, \bar{t}]$.

Если расчеты по алгоритму прекращаются на пробном значении x^k, а $c(x)$ является вогнутой, то можно определить верхнюю грань максимального значения $c(x)$. Данный подход основан на том обстоятельстве, что гиперплоскость, касательная к поверхности $c(x)$ в точке x^k, для любого x дает большие значения, чем $c(x)$, а именно:

$$c(x) \leqslant c(x^k) + \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j}(x_j - x_j^k) \quad \text{(вогнутая)} \tag{I}$$

для любого x.

Для того чтобы продемонстрировать эту идею, предположим, что мы можем утверждать, насколько $\bar{x}$ может быть удален от x^k. Для конкретности примем, что для каждого j $|\bar{x}_j^k - x_j^k| \leqslant e$, где $e > 0$. Тогда максимальное значение $c(x)$ не может превышать $c(x^k) + e\sum_{j=1}^{n} |\partial c(x^k)/\partial x_j|$. Поправочный коэффициент мал, если e и градиент в точке x^k близки к нулю.

При изучении метода скорейшего подъема в том виде, как он представлен на рис. 14.6 и 14.7, можно прийти к выводу, что обеспечивалось совпадение локального максимума с глобальным формой линий уровня. В частности, множество точек x, таких, что $c(x) \geqslant c$, являлось выпуклым; в свою очередь этим обеспечивается, что линии уровня для $c(x) = c$ при возрастающих c концентричны, т. е. «вложены» друг в друга. Отсутствуют «впадины» или «ловушки», мешающие последовательности x^k сходиться к оптимуму $\bar{x}$. Легко показать, что вогнутые функции всегда обладают указанным свойством — линии уровня в них концентричны. Однако это характерно и для некоторых других функций. [Рассмотрим, например, функцию $c(x) = -x^2/(1+x^2)$, имеющую форму колокола и, следовательно, не вогнутую. Множество точек x, таких, что $c(x) \geqslant c$, является выпуклым, точнее представляет собой сегмент с центром в $x = 0$. По мере увеличения c длина сегментов уменьшается, и они оказываются концентричными.]

Таким образом, предположим, что постулируется только свойство выпуклости концентрических линий уровня $c(x)$. Функция, обладающая таким свойством, именуется **квазивогнутой** и математически определяется условием, согласно которому для любых двух точек x и y $x \neq y$ и для любого p, $0 \leqslant p \leqslant 1$,

$$\min\left[c(x),\ c(y)\right] \leqslant c\left[px_1 + (1-p)\,y_1,\ \ldots,\ px_n + (1-p)\,y_n\right] \quad \text{(квазивогнутая).} \qquad \text{(II)}$$

(Вогнутая функция всегда одновременно является квазивогнутой.) Данное допущение не вполне достаточно для обеспечения того, чтобы последовательность $c(x^k)$ при использовании метода скорейшего подъема стремилась к максимальному значению. Допущение квазивогнутости (II) не исключает той возможности, что предел подпоследовательности x^k, а именно x^*, представляет собой точку перегиба. Однако если, кроме того, допустить, что $c(x)$ *не* имеет точек перегиба, или же если можно убедиться в том, что x^* *не* является точкой перегиба (значение $c(x)$ не возрастает при движении в любом направлении от x^*), то $c(x^*)$ действительно есть максимальное значение.

Попутно отметим, что если предполагается квазивогнутость и если $c(x)$ не имеет точек перегиба, то для нахождения оптимальной длины шага можно также использовать методы последовательного поиска в одномерном пространстве, изложенные в разд. 14.3.

14.6. КВАДРАТИЧНОЕ ПРОГРАММИРОВАНИЕ

В настоящем разделе мы приступаем к рассмотрению методов оптимизации нелинейной целевой функции в условиях, когда на переменные налагаются ограничения. Начнем с, вероятно, простейшего нелинейного обобщения модели линейного программирования.

Задача **квадратичного программирования** включает целевую функцию, составленную из линейных и квадратичных слагаемых, и систему линейных ограничений. Ниже рассматривается следующая форма записи задачи:

$$\text{Максимизировать } \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} \sum_{k=1} c_{jk} x_j x_k \tag{1}$$

при условиях

$$\begin{aligned} &\sum_{j=1}^{n} a_{ij} x_j + s_i = b_i, && i = 1, 2, \ldots, m, \\ &x_j \geqslant 0, && j = 1, 2, \ldots, n, \\ &s_i \geqslant 0, && i = 1, 2, \ldots, m, \end{aligned} \tag{2}$$

где s_i представляют собой дополняющие переменные, превращающие неравенства в уравнения. Для того чтобы не рассматривать все возникающие варианты, примем, что все $b_i \geqslant 0$. Предположим также, что $c(x)$ на всем множестве допустимых решений достигает конечного максимума $\bar{c}$, где $c(x)$ обозначает целевую функцию (1). При модифицированных вариантах формы записи требуется решать задачи, где не налагается ограничений на знак b_i либо отсутствуют в явном виде дополняющие переменные, либо возможны бесконечные оптимальные решения. Эти модификации просты, однако в настоящей книге они не рассматриваются.

В разд. 14.1 приводилось несколько иллюстративных примеров («выручка от реализации продукции», «уровень страховых запасов», «выбор портфеля ценных бумаг»), где при построении модели приходилось использовать квадратичные слагаемые. Практическая важность таких ситуаций в известной мере оправдывает то, что весь настоящий раздел посвящен методам решения задач квадратичного программирования. Однако нужно признать, что эти случаи — не наиболее частые примеры нелинейных проблем. Следовательно, основным доводом для подробного обсуждения данного частного случая является то обстоятельство, что эта тема представляет собой очередной логический шаг в овладении идеями и методами нелинейного программирования. Особенно важно, чтобы читатель обратил внимание на следующие моменты:

I) Как наложение ограничений влияет на оптимальное решение?

II) Как учитывается система ограничений при использовании градиентных методов оптимизации?

III) Как квадратичный характер задачи дает возможность использовать линеаризованные алгоритмы и, в частности, каким образом модифицируется симплексный метод при его преобразовании в алгоритм квадратичного программирования, обеспечивающий сходимость к истинному оптимуму за *конечное* число итераций?

Методы квадратичного программирования важны и потому, что они иногда используются в качестве подпрограмм в более общих

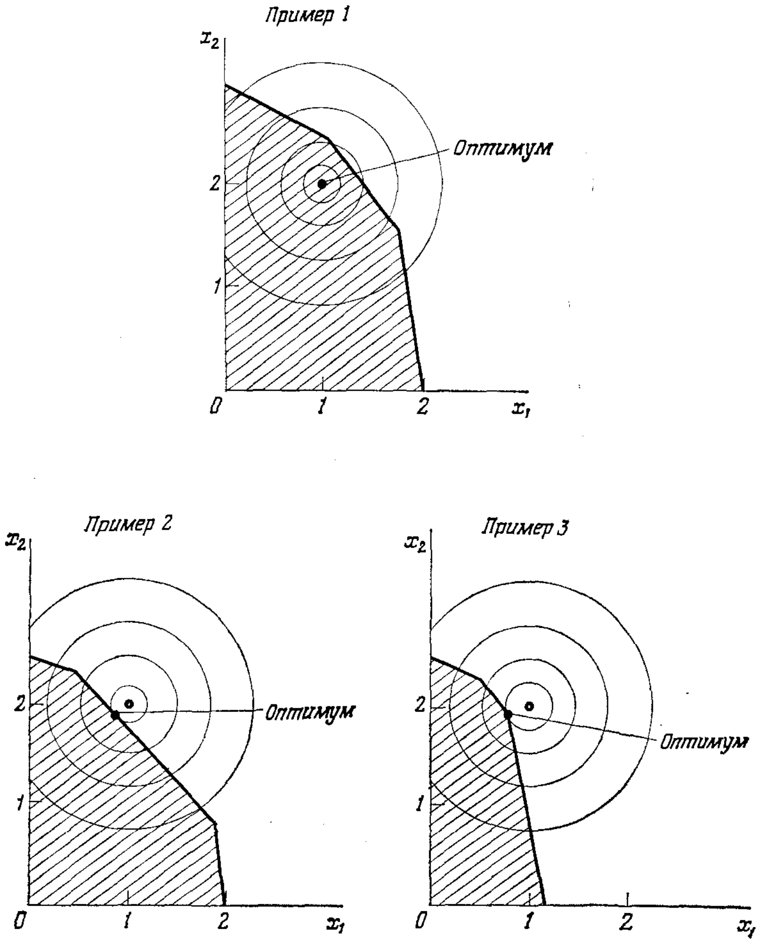


Рис. 14.9. Решения задачи на условный оптимум для функции $c(x) = -(x_1 - 1)^2 - (x_2 - 2)^2 + 5 = 2x_1 + 4x_2 - x_1^2 - x_2^2$. Пример 1 — оптимум во внутренней точке. Пример 2 — оптимум в граничной точке. Пример 3 — оптимум в вершине.

алгоритмах нелинейного программирования: так, в пробной точке нелинейная целевая функция аппроксимируется квадратичной функцией, которая в свою очередь оптимизируется с целью определения направления улучшения. Этот прием относится к категории усложненных численных методов ускорения сходимости и в настоящем тексте не рассматривается.

При сделанных допущениях о b_i и $c(x)$ возможны три типа оптимальных решений, как это показано на рис. 14.9. Из примера 1 видно, что оптимальное решение может находиться внутри области, определяемой наложенными ограничениями; в данном примере единственным оптимальным решением является **внутренняя точка области** $\bar{x}_1 = 1$, $\bar{x}_2 = 2$. В примере 2 показано, что оптимальная точка может находиться на **границе области** ограничений, но не обязательно в вершине или в экстремальной точке. Пример 3 соответствует тому типу решения, которое возникает в случае линейной целевой функции, а именно тогда, когда базисное оптимальное решение соответствует *вершине* области допустимых решений. Поскольку пример 3 здесь не является единственной возможностью, в алгоритмах нелинейного программирования должны учитываться в полном противоречии с методами линейного программирования небазисные решения.

Предварительные замечания. Прежде чем рассматривать собственно алгоритм, целесообразно сделать несколько замечаний о характере квадратичной целевой функции. Вначале предложим удобный способ представления этой функции; такой способ легко уяснить на примере. Пусть

$$c(x) \equiv 2x_1 + 5x_2 - \frac{1}{4}x_1^2 + 1x_1x_2 - 3x_1x_3 - 4x_2^2. \tag{3}$$

Заметим, что $c(x)$ представлена в (3) в таком виде, что имеется не более одного члена, содержащего любую фиксированную пару переменных. Так, если, скажем, вначале в квадратичной функции имелись члены $-5x_1x_3$ и $+2x_3x_1$, то эти два члена, каждый из которых содержит одну и ту же пару переменных x_1 и x_3, приводятся к виду $-3x_1x_2$. Это позволяет записать $c(x)$ в следующем виде:

$$\begin{aligned} c(x) = {} & \left(0 + 1x_1 + \frac{5}{2}x_2 + 0x_3\right)\cdot 1 + \\ & + \left(1 - \frac{1}{4}x_1 + \frac{1}{2}x_2 - \frac{3}{2}x_3\right)\cdot x_1 + \\ & + \left(\frac{5}{2} + \frac{1}{2}x_1 - 4x_2 + 0x_3\right)\cdot x_2 + \\ & + \left(0 - \frac{3}{2}x_1 + 0x_2 + 0x_3\right)\cdot x_3. \end{aligned} \tag{4}$$

Можно убедиться в том, что выражение (4) соответствует выражению (3), если выполнены операции умножения (раскрытия скобок) и приведения подобных членов. Отметим, что диагональные члены (4), а именно $-\frac{1}{4}$, -4 и 0, являются коэффициентами при квадратах переменных в (3), а коэффициенты, не находящиеся на главной диагонали (4), равны половине соответствующих коэффициентов в (3). Отметим также, что матрица коэффициентов в (4) симметрична — коэффициент в строке i и столбце j равен коэффициенту в строке j

и столбце i. Теперь рассмотрим частные производные $c(x)$ из (3):

$$\frac{\partial c}{\partial x_1}=2-\frac{1}{2}x_1+1x_2-3x_3\equiv 2\left(1-\frac{1}{4}x_1+\frac{1}{2}x_2-\frac{3}{2}x_3\right),$$
$$\frac{\partial c}{\partial x_2}=5+1x_1-8x_2\equiv 2\left(\frac{5}{2}+\frac{1}{2}x_1-4x_2\right), \qquad (5)$$
$$\frac{\partial c}{\partial x_3}=-3x_1\equiv 2\left(-\frac{3}{2}x_1\right).$$

Из их рассмотрения можно сделать вывод, что частная производная $c(x)$ по x_j равна удвоенному выражению в скобках, которое в (4) умножалось на соответствующее x_j. Положительное значение частной производной показывает, что при увеличении соответствующей переменной целевая функция возрастает. Отметим, что частные производные являются линейными функциями x_j. Именно это обстоятельство позволяет ограничиться незначительным изменением симплексного алгоритма при его использовании в квадратичном программировании (см. далее).

В процессе использования квадратичного алгоритма потребуется ввести новые переменные, связанные с выражениями в скобках в (4). По причинам, которые выяснятся позже, назовем их **свободными переменными**. Для иллюстрации предположим, что необходимо ввести свободную переменную — назовем ее u — для выражения в скобках, умноженного в (4) на x_1. Построим следующее выражение, определяющее

$$\frac{1}{4}u=1-\frac{1}{4}x_1+\frac{1}{2}x_2-\frac{3}{2}x_3. \qquad (6)$$

Условимся коэффициент при u брать таким же, как коэффициент при x_1 в (6), взятый со знаком минус. Переставив слагаемые в (6), можно представить x_1 в следующем виде:

$$x_1=4-u+2x_2-6x_3. \qquad (7)$$

Применяя рассматриваемый алгоритм, потребуется также исключать из квадратичной формы переменную x_j и заменять ее линейным выражением относительно остальных переменных. Например, может потребоваться исключить из (4) x_1 и подставить вместо него выражение, стоящее в правой части (7). Если мы хотим выполнять эти операции систематически, то, во-первых, исключим x_1 из выражения в скобках:

$$\left(4-1u+\frac{9}{2}x_2-6x_3\right)\cdot 1+$$
$$+\left(0+\frac{1}{4}u+0x_2+0x_3\right)\cdot x_1+$$
$$+\left(\frac{9}{2}-\frac{1}{2}u-3x_2-3x_3\right)\cdot x_2+$$
$$+\left(-6+\frac{3}{2}u-3x_2+9x_3\right)\cdot x_3. \qquad (I)$$

Во-вторых, подставим правую часть (7) вместо x_1 в (I); затем умножим второе выражение в скобках в (I) на каждое из слагаемых правой части (7), после чего выполним приведение подобных членов полученного результата и других выражений в скобках в (I). Например, при умножении второго выражения в скобках в (I) на третье слагаемое правой части (7) получим $\left(0 + \frac{1}{4}u + 0x_2 + 0x_3\right)\cdot 2x_2$; это выражение следует сложить с $\left(\frac{9}{2} - \frac{1}{2}u - 3x_2 - 3x_3\right)\cdot x_2$. В результате получим

$$\begin{aligned}&\left(4+0u+\frac{9}{2}x_2-6x_3\right)\cdot 1+\\&+\left(0-\frac{1}{4}u+0x_2-0x_3\right)\cdot u+\\&+\left(\frac{9}{2}+0u-3x_2-3x_3\right)\cdot x_2+\\&+(-6+0u-3x_2+9x_3)\cdot x_3.\end{aligned} \tag{II}$$

Эта форма также является симметричной. Для того чтобы проверить усвоение изложенного, читателю следует самостоятельно выполнить операции, приводящие к (I) и (II).

Не случайно коэффициенты при u (за исключением диагональных) равны нулю, а все остальные коэффициенты остаются такими же, как в (I). Подобное упрощение *всегда* возникает, если при подстановке используется выражение, определявшее выше свободную переменную; следовательно, в таком случае нет необходимости выполнять арифметические операции шага 2. Если же, например, вместо x_1 подставлялось бы выражение $(1 + 4x_2 - 4x_3)$, то пришлось бы выполнять умножение, предписываемое на шаге 2.

Квадратичный симплексный алгоритм. Алгоритм включает следующие операции:

Шаг 1. Примем $s_i = b_i \geqslant 0$ для $i = 1, 2, \ldots, m$ в качестве исходного допустимого базисного решения.

Шаг 2. Определим направления локального улучшения на основе текущего решения. Прекратим расчеты, если возможностей улучшения не имеется. В противном случае найдем оптимальную длину шага для выбранного направления и перейдем к шагу 3.

Шаг 3. Соответствующим образом вычислим новое решение. Возвратимся к шагу 2.

Характер алгоритма на любой итерации можно вкратце описать следующим образом. Имеется система линейных ограничений, полученная из (2), а также некоторые дополнительные ограничения, возникшие в процессе выполнения предыдущих итераций. На шаге 1 имеется только m ограничений (2), однако на последующих итерациях общее число ограничений может доходить до $m + n$. Этой системе ограничений соответствует набор базисных переменных, включающих как x_j, так и дополняющие переменные s_i. Небазисные переменные состоят из остальных x_j и s_i, а также из свободных переменных, введенных на предыдущих итерациях.

Подробные указания по определению направления улучшения на шаге 2 можно сформулировать следующим образом:

Критерий I квадратичного симплексного алгоритма. а) Введите в решение любую свободную переменную, если соответствующая ей частная производная целевой функции не равна нулю. б) Если все такие производные, соответствующие свободным переменным, равны нулю, то выберите для введения в решение такую переменную x_j или s_i, для которой соответствующая частная производная целевой функции является положительной и наибольшей по значению. в) Прекратите расчеты, если все частные производные для свободных переменных равны нулю, а для других небазисных переменных неположительны.

Если не считать изменения формулировок, вызванного нелинейным характером задачи, *критерий I квадратичного симплексного алгоритма* основан на том же подходе, что и *критерий I симплексного алгоритма*. В обоих случаях улучшение возникает благодаря введению в решение *одной* небазисной переменной.

Вычисление оптимальной длины шага напоминает *критерий II симплексного алгоритма*. Отчасти оно является проверкой максимально возможного значения выбранной небазисной переменной, при котором не нарушается допустимость решения, т. е. выполнение ограничений. Однако здесь также определяется значение переменной, выше которого начинается уменьшение значения целевой функции. Следовательно, оптимальная длина шага есть меньшая из этих двух величин. Правило ее определения формулируется таким образом:

Критерий II квадратичного симплексного алгоритма. Введите в решение выбранную небазисную переменную с таким ее значением, при котором улучшение значения целевой функции является наибольшим, а допустимость решения не нарушается.

Подробное описание этого критерия демонстрируется на приводимом ниже примере.

Пример. Рассмотрим задачу:

$$\text{Максимизировать } 6x_1+4x_2+2x_3-3x_1^2-2x_2^2-\frac{1}{3}x_3^2 \qquad (8)$$

при условиях

$$x_1+2x_2+x_3+s=4 \quad \text{(строка 1)}; \qquad (9)$$

все $x_j \geqslant 0$ и $s \geqslant 0$. [Читателю полезно выписать (8) и (9) для удобства дальнейшего обращения к ним.] Напомним, что та же функция $c(x)$ анализировалась в разд. 14.5, где единственным безусловным максимумом было решение $x_1=1$, $x_2=1$, $x_3=3$. Можно показать, что при учете (9) эта точка более не является допустимой. Пусть $s=4$ есть исходное допустимое решение, выбранное на шаге 1;

поскольку s есть дополняющая переменная, коэффициент при s в (8) равен нулю. Воспользовавшись предложенным приемом представления целевой функции, запишем (8) в следующем виде:

$$\begin{aligned} &(+3x_1 - 2x_2 + 1x_3)\cdot 1 + \\ &+(3 - 3x_1 \qquad\qquad\qquad)\cdot x_1 + \\ &+(2 \qquad - 2x_2 \qquad\quad)\cdot x_2 + \\ &+\left(1 \qquad\qquad - \frac{1}{3}x_3\right)\cdot x_3, \end{aligned} \tag{10}$$

где слагаемые с нулевыми коэффициентами не проставлены. Напомним, что ведущее слагаемое каждого выражения в скобках равно половине частной производной по соответствующей небазисной переменной. Поэтому, согласно части б) *критерия I*, необходимо ввести в решение x_1 (объясните, почему).

Для того чтобы определить значение x_1, сначала применим подход, использованный в *критерии II симплексного алгоритма* для определения максимального значения x_1, которое не нарушает допустимости решения, т. е. соответствия линейным ограничениям. Для этого вычисляются частные от деления текущих правых частей на коэффициенты вектора вводимой переменной. В настоящем примере для сохранения допустимости решения необходима длина шага

$$x_1 \leqslant \frac{4}{1} = 4. \tag{11}$$

Значение x_1, *максимально* улучшающее целевую функцию $c(x)$ *при условии, что все остальные небазисные переменные сохраняют свои прежние значения*, является решением уравнения

$$\frac{\partial c}{\partial x_1} = 0 \quad \text{или, что равносильно,} \quad \frac{1}{2}\frac{\partial c}{\partial x_1} = 0, \tag{12}$$

где частная производная вычислена для текущей пробной точки. В соответствии с (12) $3 - 3x_1 = 0$, откуда длина шага, улучшающая текущее значение целевой функции, должна быть не более

$$x_1 \leqslant 1. \tag{13}$$

Поскольку неравенство (13) является более жестким, чем (11), то, согласно *критерию II квадратичного симплексного алгоритма*, переменная x_1 должна быть введена в решение со значением 1. Тем самым завершается шаг 2, и мы переходим к шагу 3.

В следующем пробном решении можно достичь $x_1 = 1$, дополнительно добавив уравнение, определяющее свободную переменную, которая связана с x_1:

$$3u_1 = 3 - 3x_1 \quad \text{или, что равносильно,} \quad x_1 + u_1 = 1. \tag{14}$$

В текущем решении $u_1 = 1$, поскольку $x_1 = 0$. Однако, выводя u_1 из базиса, так чтобы его значение стало нулевым, мы увеличиваем x_1 до 1. Подводя итоги, можно сказать, что к (9) добавляется ограничение (14), затем x_1 вводится в решение и слагаемое с x_1 из (14) делается ведущим элементом. Получаем следующую систему ограничений:

$$\begin{aligned} 2x_2 + x_3 + s - u_1 &= 3 \quad \text{(строка 1)}, \\ x_1 \qquad\qquad + u_1 &= 1 \quad \text{(строка 2)}. \end{aligned} \tag{15}$$

Исключим x_1 из целевой функции (10), выполнив замену переменных согласно (14). В результате этого получаем

$$\begin{aligned} &(3 \qquad + 2x_2 \quad + 1x_3)\cdot 1 + \\ &+(\ -3u_1 \qquad\qquad\quad)\cdot u_1 + \\ &+(2 \qquad -2x_2 \qquad\quad)\cdot x_2 + \\ &+\left(1 \qquad\qquad\quad -\frac{1}{3}x_3\right)\cdot x_3. \end{aligned} \tag{16}$$

Каждое выражение в скобках по-прежнему равно половине частной производной по соответствующей небазисной переменной, однако здесь учитывается и то обстоятельство, что изменение этой переменной влияет на значение целевой функции ввиду соответствующего изменения уровней базисных переменных. Иногда это значение называют **приведенной частной производной.** Первая итерация завершена. Величина 3, расположенная в левом верхнем углу (16), равна значению целевой функции при текущем пробном решении.

Возвращаясь к шагу 2, можно увидеть, что решение улучшится, если в него ввести x_2. Ограничение, обеспечивающее сохранение допустимости в соответствии с (15), имеет вид $x_2 \leqslant \frac{3}{2} = 1{,}5$. Можно убедиться в том, что наибольшее увеличение целевой функции дает $x_2 = 1$. Следовательно, для выполнения шага 3 введем еще одну свободную переменную:

$$2u_2 = 2 - 2x_2 \quad \text{или, что равносильно,} \quad x_2 + u_2 = 1. \tag{17}$$

Добавляя (17) к (15), вводя в решение x_2 и делая ведущим элементом слагаемое с x_2 из (17), получим

$$\begin{aligned} x_3 + s - u_1 - 2u_2 &= 1 \quad \text{(строка 1)}, \\ x_1 \qquad + u_1 \qquad\quad &= 1 \quad \text{(строка 2)}, \\ x_2 \qquad\qquad\quad + u_2 &= 1 \quad \text{(строка 3)}. \end{aligned} \tag{18}$$

Затем используя (17) для исключения x_2 из (16), получаем следующую приведенную целевую функцию:

$$\begin{aligned} &(5 \qquad\qquad\qquad + 1x_3)\cdot 1 + \\ &+(\ -3u_1 \qquad\qquad\quad)\cdot u_1 + \\ &+(\qquad -2u_2 \qquad\quad)\cdot u_2 + \\ &+\left(\qquad\qquad\quad -\frac{1}{3}x_3\right)\cdot x_3. \end{aligned} \tag{19}$$

Заметим, что и в (16), и в (19) первые слагаемые выражений в скобках, умножаемых на свободные переменные u_1 и u_2, равны нулю. Следовательно, для этих пробных решений частные производные преобразованных целевых функций (16) и (19) по каждой из свободных переменных равны нулю.

Теперь установим, что, согласно части б) *критерия I*, в решение должна быть введена переменная x_3. Определив длину шага, выясним, что в данном случае более жестким ограничением является условие необходимости соблюдения допустимости решения, основывающееся на (18). Поэтому на шаге 3 этой итерации выполняется обычная смена базиса — переменная x_3 вводится в базисное решение, переменная s выводится из него. Поскольку x_3 входит только в строку 1 системы (18), нет необходимости в изменении коэффициентов матрицы ограничений. (Если бы в исходной задаче было бы не одно ограничение, а большее их число, то пришлось бы, как обычно, преобразовывать матрицу с использованием ведущего элемента.) Исключим x_3 из (19), подставив значение этой переменной из строки 1 системы (18), в результате чего получим

$$\begin{aligned}
&\left(6\frac{2}{3}+\frac{2}{3}u_1+\frac{4}{3}u_2-\frac{2}{3}s\right)\cdot 1+\\
&+\left(\frac{2}{3}-\frac{10}{3}u_1-\frac{2}{3}u_2+\frac{1}{3}s\right)\cdot u_1+\\
&+\left(\frac{4}{3}-\frac{2}{3}u_1-\frac{10}{3}u_2+\frac{2}{3}s\right)\cdot u_2+\\
&+\left(-\frac{2}{3}+\frac{1}{3}u_1+\frac{2}{3}u_2-\frac{1}{3}s\right)\cdot s.
\end{aligned}\tag{20}$$

Текущее пробное решение имеет следующий вид:

$$x_1=1,\ x_2=1,\ x_3=1,\tag{21}$$

так что все три переменные приняли положительные значения. Является ли это решение оптимальным? Согласно части а) приведенного выше *критерия I* — оно не оптимально, и вот почему. Уравнения (14) и (17), связанные с переменными u_1 и u_2, были добавлены в качестве искусственных ограничений, обеспечивающих выбор оптимальной длины шага согласно *критерию II*. Нет необходимости в том, чтобы u_1 и u_2 *обязательно* были равны нулю. В самом деле, на знаки u_1 и u_2 не налагается ограничений — именно поэтому они именуются *свободными* переменными. Соответственно если частная производная целевой функции по свободной переменной не равна нулю, то решение можно улучшить, двигаясь в направлении, определяемом знаком этой частной производной. Как видно из (20), частные производные по u_1 и u_2 не равны нулю. Для введения в решение выберем u_1.

Поскольку частная производная по u_1 положительна, значение u_1 должно быть увеличено. Процесс нахождения нового значения u_1

является точно таким же, как и на предыдущих итерациях. Согласно условиям допустимости, u_1 не может превышать 1. [Если бы значение частной производной было бы отрицательным, следовало бы рассмотреть систему ограничений (18) для того, чтобы установить, насколько большим по абсолютной величине можно принять u_1, не нарушая ограничений. Это равносильно проверке, насколько большим может быть принято $(-u_1)$.] Значение u_1, позволяющее достичь максимального улучшения целевой функции, *если все остальные небазисные переменные сохраняют прежний уровень*, равно $u_1 = {}^2/_{10}$, что является решением уравнения, составленного для текущего пробного решения:

$$\frac{1}{2}\frac{\partial c}{\partial u_2} = \left(\frac{2}{3} - \frac{10}{3}u_1\right) = 0. \tag{22}$$

Для введения в следующее решение переменной u_1, со значением $\frac{2}{10}$, необходима новая свободная переменная из

$$\frac{10}{3}u_3 = \frac{2}{3} - \frac{10}{3}u_1 - \frac{2}{3}u_2 + \frac{1}{3}s \tag{23}$$

или

$$-\frac{1}{10}s + 1u_1 + \frac{1}{5}u_2 + 1u_3 = \frac{1}{5}. \tag{24}$$

Теперь используем (24) для исключения u_1 из (18):

$$\begin{aligned}
x_3 + \frac{9}{10}s - \frac{9}{5}u_2 + u_3 &= \frac{6}{5} \quad \text{(строка 1)},\\
x_1 \qquad + \frac{1}{10}s - \frac{1}{5}u_2 - u_3 &= \frac{4}{5} \quad \text{(строка 2)},\\
x_2 \qquad + \quad u_2 \quad &= 1 \quad \text{(строка 3)},
\end{aligned} \tag{25}$$

а также из (20):

$$\begin{aligned}
&\left(6\frac{4}{5} \qquad + \frac{6}{5}u_2 - \frac{3}{5}s\right)\cdot 1 +\\
&+\left(\qquad -\frac{10}{3}u_3 \qquad\qquad \right)\cdot u_3 +\\
&+\left(\frac{6}{5} \qquad -\frac{16}{5}u_2 + \frac{3}{5}s\right)\cdot u_2 +\\
&+\left(-\frac{3}{5} \qquad + \frac{3}{5}u_2 - \frac{3}{10}s\right)\cdot s.
\end{aligned} \tag{26}$$

Поскольку u_1 в (24) является базисной переменной, а на ее знак не налагается ограничений, она не может снова стать небазисной переменной. Таким образом, на последующих итерациях (24) можно отбросить, так что система ограничений представится в виде (25).

В общем случае, как только свободная переменная войдет в решение, соответствующее ограничение можно отбросить — вот почему общее число ограничений не может превысить $m+n$. (Это объясняется тем, что в наихудшем случае, когда все переменные x_j и s_i являются базисными, так что система включает $m+n$ ограничений, небазисными могут быть только свободные переменные. После того как одна из них будет выбрана для введения в базис, возможна одна из двух альтернативных ситуаций. Либо одна из x_j или s_i становится небазисной переменной — тогда можно отбросить строку, в которой выбранная свободная переменная стала базисной, и в этом случае общее число ограничений сокращается на единицу. Либо свободная переменная, соответствующая вновь наложенному ограничению, становится небазисной, так что это ограничение можно отбросить после того, как элемент с выбранной свободной переменной в этом ограничении будет использован в качестве ведущего элемента симплексной итерации. В этом случае общее число ограничений остается неизменным.)

Возвратимся к шагу 2. Следует ввести в базис переменную u_2, поскольку в (26) ей соответствует положительная частная производная. Расчеты длины шага показывают, что необходима еще одна свободная переменная:

$$\frac{16}{5}u_4=\frac{6}{5}-\frac{16}{5}u_2+\frac{3}{5}s. \tag{27}$$

Исключение u_2 из (25) дает следующие результаты:

$$\begin{aligned}
x_3+\frac{9}{16}s+u_3+\frac{9}{5}u_4&=\frac{15}{8} &&\text{(строка 1)},\\
x_1\qquad+\frac{1}{16}s-u_3+\frac{1}{5}u_4&=\frac{7}{8} &&\text{(строка 2)},\\
x_2\qquad+\frac{3}{16}s\qquad-\qquad u_4&=\frac{5}{8} &&\text{(строка 3)},
\end{aligned} \tag{28}$$

а исключение u_2 из (26):

$$\begin{aligned}
&\left(7\frac{1}{4}\qquad\qquad\qquad\qquad-\frac{3}{8}s\right)\cdot 1+\\
+&\left(\qquad-\frac{10}{3}u_3\qquad\qquad\qquad\right)\cdot u_3+\\
+&\left(\qquad\qquad\qquad-\frac{16}{5}u_4\qquad\right)\cdot u_4+\\
+&\left(-\frac{3}{8}\qquad\qquad\qquad\qquad-\frac{1}{6}s\right)\cdot s.
\end{aligned} \tag{29}$$

На этот раз после шага 2 оказывается, что итерации можно прекратить. (Объясните, почему?) Текущее решение имеет вид

$$x_1=\frac{7}{8}=0{,}875,\quad x_2=\frac{5}{8}=0{,}625,\quad x_3=\frac{15}{8}=1{,}875. \tag{30}$$

Это решение является глобальным оптимумом, причем соответствующее значение целевой функции равно 7,25.

Обзор результатов. Снова рассмотрим описанный подход и попытаемся выявить заложенные в него основные идеи. В частности, отметим следующие свойства алгоритма:

А. Как и в симплексном алгоритме, улучшение обусловливается введением в решение на каждой итерации только одной небазисной переменной.

Б. Как и в симплексном алгоритме, значение вводимой переменной не может превысить уровня, при котором значение любой переменной, входящей в текущий базис, уменьшается до нуля.

В. Если значение целевой функции при движении в выбранном направлении достигает максимума раньше того момента, когда хотя бы одна базисная переменная уменьшится до нуля, к (2) добавляется дополнительное ограничение.

Г. Выражение, входящее в дополнительное ограничение, пропорционально частной производной целевой функции по вводимой переменной, а следовательно, является линейным, поскольку целевая функция квадратична.

Д. Такие ограничения используются как вспомогательные, чтобы облегчить обеспечение оптимальной длины шага в выбранном направлении. В дальнейшем они смогут расширить базис, так что в оптимальном решении положительные значения могут иметь более чем m переменных.

Следует предупредить, что алгоритм не был изложен в виде, наиболее удобном для составления машинной программы, поскольку соответствующая детализация выходила бы за рамки книги, которая носит характер введения. Практическая необходимость использования ЭВМ для нахождения решения подтверждается приведенным примером: хотя модель включала всего лишь одно линейное ограничение и условия неотрицательности, для нахождения оптимального решения потребовался значительный объем вычислений. (Таким образом, применять этот алгоритм вручную можно лишь для решения примеров очень малой размерности.)

Сходимость. Можно построить достаточно простое доказательство (здесь оно не приводится) того, что для любой квадратичной целевой функции и системы линейных ограничений метод обеспечивает сходимость за конечное число итераций. (Для этого алгоритм должен быть несколько изменен, так чтобы он включал меры против зацикливания, упоминавшиеся применительно к симплексному методу в разд. 4.7.) Однако до сих пор не утверждалось, что настоящий алгоритм всегда приводит к получению оптимального окончательного решения. Этот вопрос рассматривается ниже.

Как следовало ожидать с учетом соображений, изложенных в разд. 14.4 и 14.5, если не сделаны дополнительные допущения о виде целевой функции, можно говорить лишь о том, что окончательное решение является *относительной* стационарной точкой. Точно так же необходимое условие существования максимума (разд. 14.5) можно обобщить на случай условного оптимума; *если квадратичная* *функция вогнута, то окончательное решение является глобальным оптимумом на множестве всех допустимых решений.* (Разумеется, достаточное условие глобальной оптимальности не является здесь необходимым: квадратичная целевая функция может и не быть вогнутой, и все же окончательное решение может оказаться истинным оптимумом.) Квадратичная функция (1) вогнута тогда и только тогда, когда

$$\sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk} x_j x_k \leqslant 0 \quad \text{для всех} \quad x_j \text{ и } x_k, \quad j, k = 1, 2, \ldots, n, \qquad (31)$$

В математических терминах в (31) констатируется, что **квадратичная форма** является **отрицательно полуопределенной.**

В большинстве руководств по линейной алгебре и теории матриц приводятся простые формулы проверки того, удовлетворяет ли квадратичная форма условию (31). Такие формулы легко применить, если n мало; для больших n они могут быть запрограммированы на ЭВМ. На практике сама постановка модели обычно позволяет получить ответ. Так, в задаче выбора портфеля ценных бумаг (разд. 14.1) цель состоит в минимизации квадратичной формы — выражения (11), в котором коэффициенты соответствуют ковариациям. В наиболее полных учебниках статистики можно найти доказательство того, что такая форма является **положительно полуопределенной**, т. е. что знак неравенства (31) меняется на обратный. Следовательно, умножив эту квадратичную форму на —1 и переходя к оптимизации в смысле «максимизация», мы автоматически обеспечиваем вогнутость.

Другие алгоритмы. Имеется ряд других алгоритмов решения задач квадратичного программирования. Особо важен алгоритм, основанный на использовании двойственных переменных; он излагается в разд. 15.8. Этот алгоритм также сходится за конечное число итераций. Можно, кроме того, применять методы, описанные в остальных разделах настоящей и следующей глав, однако в полном противоречии с описанным выше алгоритмом они не обязательно сходятся к стационарной точке за *конечное* число итераций. Это будет показано на примере, приводимом в настоящей главе.

14.7 СЕПАРАБЕЛЬНОЕ ПРОГРАММИРОВАНИЕ

В предыдущем разделе было показано, что для решения задач квадратичного программирования требуется лишь незначительная модификация симплексного метода. При этом использовалось то обстоятельство, что частные производные квадратичной целевой функции линейны. В настоящем разделе мы переходим к изложению другого подхода к решению нелинейных задач, связанного с иной модификацией симплексного алгоритма. Метод применим к задачам, в которых все нелинейные функции сепарабельны (этот термин разъясняется ниже). Идея заключается в построении задачи на условный оптимум, которая является линейной *аппроксимацией* исходной задачи. При аппроксимации размер модели возрастает, однако, поскольку в качестве метода решения используется вариант симплексного метода, данный алгоритм имеет значительную практическую ценность. При данном подходе в одинаковой мере возможна аппроксимация как нелинейной целевой функции, так и нелинейных ограничений; оба эти случая рассматриваются ниже.

Сепарабельные целевые функции. Предположим, что целевую функцию можно представить в виде суммы n функций, по одной для каждой из переменных x_j:

$$c(x) \equiv \sum_{j=1}^{N} c_j(x_j); \tag{1}$$

такая функция называется **сепарабельной.** В данном контексте не будем конкретизировать, необходимо ли максимизировать или минимизировать функцию $c(x)$. Для простоты изложения метод описывается только для одной переменной x_1 и соответствующей функции $c_1(x)$; совершенно аналогичные построения применимы для каждой из остальных переменных. Кроме того, поскольку рассматриваемые ниже операции производятся только над одной переменной x_1, целесообразно опустить подстрочный индекс и говорить просто о переменной x.

Для простоты изложения предположим также, что, согласно имеющимся сведениям, x должен находиться на сегменте $0 \leqslant x \leqslant X$. Для линеаризации выберем некоторую «сетку» значений x, построенную следующим образом: $0 \equiv X_1 < X_2 < \ldots < X_p \equiv X$. Затем любое значение x будем выражать в виде средневзвешенной X_k:

$$x = X_1 w_1 + X_2 w_2 + \ldots + X_p w_p, \tag{2}$$

где веса w_k удовлетворяют условию

$$\sum_{k=1}^{p} w_k = 1 \quad \text{и} \quad w_k = 0 \quad k = 1, 2, \ldots, p. \tag{3}$$

При построении «сетки» всегда следует руководствоваться следующим удобным правилом: если на некотором интервале значений x_1 функция $c_1(x_1)$ линейна, то этому интервалу должно соответствовать не более двух точек сетки.

Для построения «аппроксимированной» модели заменим x, где бы эта переменная ни встречалась в ограничениях, на правую часть (2), включим в модель ограничения (3), а в целевой функции (1) заменим $a(x_1)$ ее **кусочно-линейной аппроксимацией:**

$$\sum_{k=1}^{p} c(X_k) w_k. \tag{4}$$

Для того чтобы продемонстрировать соответствующие построения, рассмотрим выпуклую функцию $c_1(x_1) = x_1^2$ и интервал изменения $0 \leqslant x_1 \leqslant 10$ (рис. 14.10).

Пусть в данном примере точки сетки выбраны следующими: 0, 1, 4, 9, 10. Тогда (2) имеет вид

$$x \equiv 0w_1 + 1w_2 + 4w_3 + 9w_4 + 10w_5, \tag{5}$$

а (4) представляется следующим образом:

$$0w_1 + 1^2w_2 + 4^2w_3 + 9^2w_4 + 10^2w_5. \tag{6}$$

Пусть $x = 6$. Тогда выражению (5) соответствуют веса $w_3 = 3/5$, $w_4 = 2/5$, остальные $w_k = 0$. При этом приближенное значение (6) равно $210/5 = 42$, тогда как истинное значение $c_1(6) = 36$. Предположим, далее, что на $x_1 \equiv x$ наложено ограничение

$$5x - 2x_2 \leqslant 1. \tag{7}$$

Подставив выражение для x из (5) в (7), перепишем это ограничение в виде

$$5w_2 + 20w_3 + 45w_4 + 50w_5 - 2x_2 \leqslant 1. \tag{8}$$

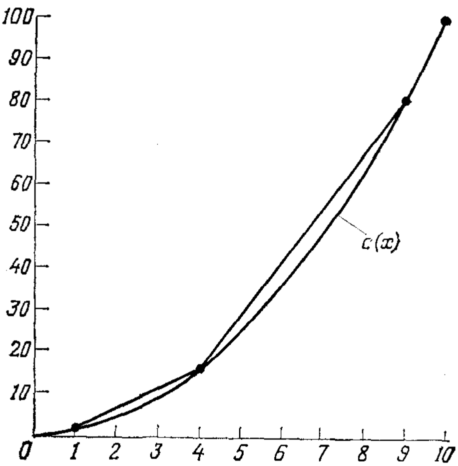


Рис. 14.10. Линейная аппроксимация функции $c(x) = x^2$.

При анализе численного примера *предполагалось*, что в случае $x = 6$ следует использовать веса w_3 и w_4, поскольку 6 расположено между точками сетки $X_3 = 4$ и $X_4 = 9$. Однако веса $w_1 = 0{,}4$ и $w_5 = 0{,}6$ также дают в (5) значение 6, только в этом случае приближенное значение $c_1(6)$ в (6) равно $10^2 \cdot 0{,}6 = 60$. Следовательно, чтобы (4) действительно представляло бы собой *желательную* кусочно-линейную аппроксимацию $c_1(x_1)$, важно, чтобы в решении использовались положительные значения либо одного, либо как максимум двух весов для *смежных* точек сетки w_k и w_{k+1}. Можно без труда сформулировать следующее свойство.

Свойство весов смежных точек. Если все ограничения являются линейными, $c(x_1, x_2, \ldots, x_n)$ должна максимизироваться, а $c_1(x_1)$ есть вогнутая функция, то любое оптимальное решение аппроксимирующей модели, построенное с использованием (2), (3) и (4), включает один вес w_k или не более двух весов w_k и w_{k+1} с положительными значениями для смежных точек. То же свойство решения сохраняется и в том случае, когда $c(x)$ минимизируется, а $c_1(x_1)$ выпукла.

Важным следствием этого свойства является следующее обстоятельство: если поставленные условия удовлетворяются для всех $c_j(x_j)$, а все ограничения, налагаемые на x_j, линейны, то преобразованная модель представляет собой задачу линейного программирования и может быть решена с помощью обычного симплексного метода, описанного в гл. 4. Если при этом исходная модель включает m линейных неравенств, то расширенная модель состоит из $m+n$ ограничений, где n дополнительных ограничений имеют вид (3). Следовательно, оптимальное базисное решение может включать все n переменных x_j с положительными значениями (наряду с m дополняющими переменными). Кроме того, возникающее решение является аппроксимацией глобального оптимального решения исходной задачи.

Однако, если поставленные условия не удовлетворяются, например если $c(x)$ максимизируется, а $c_1(x_1)$ *выпукла*, решение аппроксимационной модели обычно не обладает свойством весов смежных точек. Эту трудность можно устранить, модифицировав в деталях *критерий I симплексного метода*, так чтобы данное свойство выполнялось в *обязательном* порядке. Это выражается в виде следующего конкретного условия: если на некоторой итерации вес w_k входит в текущий базис, то может рассматриваться вопрос о вводе в решение только смежных весов w_{k-1} и w_{k+1}. Если же в текущий базис входят два смежных веса w_k и w_{k+1}, то на данной итерации не может рассматриваться вопрос о вводе других весов, относящихся к этой же переменной. Эта модификация называется **правилом ограниченного ввода в базис**. Если поставленные выше условия не удовлетворяются и используется правило ограниченного ввода в базис, *получаемое решение есть не более чем приближение к локальному оптимальному решению исходной задачи.* (Альтернативой использования правила ограниченного ввода в базис является применение модели дискретного программирования, описанной в конце разд. 13.2. Хотя последний подход обеспечивает приближение к глобальному оптимальному решению исходной задачи в большей мере, он обычно менее привлекателен, чем метод ограниченного ввода в базис, поскольку связан с большими вычислительными трудностями.)

Можно построить патологические примеры, показывающие, что при использовании правила ограниченного ввода в базис симплексный алгоритм может даже не давать локального оптимума аппроксимирующей задачи. Такая ситуация возникает, если в текущий базис входят два смежных веса w_k и w_{k+1}, причем, вводя в базис w_{k+2} и исключая w_k, мы можем улучшить значение целевой функции. Все же такие необычайные ситуации встречаются редко.

Вместо метода средневзвешенных, при котором в системе ограничений должны быть добавлены ограничения (3), можно использовать другой подход, связанный с увеличением числа неотрицательных, ограниченных сверху переменных. Данную идею легко уяснить на основе приведенного выше примера. Вместо (2) представим x в виде суммы

$$x \equiv y_1 + y_2 + \ldots + y_{p-1}, \qquad \text{(I)}$$

где

$$0 \leqslant y_k = X_{k+1} - X_k, \quad k = 1, 2, \ldots, p-1.$$

В данном примере вместо (5) имеем

$$x \equiv y_1 + y_2 + y_3 + y_4, \qquad \text{(II)}$$

где условия неотрицательности и ограниченности переменных сверху представляются в виде

$$0 \leqslant y_1 \leqslant 1 - 0 = 1, \quad 0 \leqslant y_2 \leqslant 4 - 1 = 3,$$
$$0 \leqslant y_3 \leqslant 5, \quad 0 \leqslant y_4 \leqslant 1. \qquad \text{(III)}$$

Теперь вместо (4) используется

$$\sum_{k=1}^{d-1} C_k y_k, \qquad \text{(IV)}$$

где

$$C_k \equiv \frac{c_1(X_{k+1}) - c_1(X_k)}{X_{k+1} - X_k}. \qquad \text{(V)}$$

Таким образом, C_k есть тангенс угла наклона линейного отрезка, аппроксимирующего $c_1(x_1)$ между точками X_k и X_{k+1}. В настоящем примере (IV) имеет такой вид:

$$\frac{(1^2-0)}{1-0} y_1 + \frac{(4^2-1^2)}{4-1} y_2 + \frac{(9^2-4^2)}{9-4} y_3 + \frac{(10^2-9^2)}{10-9} y_4 =$$
$$= 1y_1 + 5y_2 + 13y_3 + 19y_4. \qquad \text{(VI)}$$

При $x = 6$ представлению (II) соответствует следующий набор значений y_k: $y_1 = 1$, $y_2 = 3$, $y_3 = 2$ и $y_4 = 0$. В (VI) это дает в качестве приближения к $c_1(6) = 36$ значение целевой функции, равное 42. Использовав (II), перепишем (7) в виде

$$5y_1 + 5y_2 + 5y_3 + 5y_4 - 2x_2 \leqslant 1. \qquad \text{(VII)}$$

Если все ограничения модели линейны, можно воспользоваться симплексным алгоритмом с двусторонними ограничениями на переменные, изложенным в разд. 5.10.

При такой постановке свойство весов смежных точек выражается аналогичной формулировкой: y_{k+1} может принимать положительные значения только в том случае, если значение y_k равно его верхнему пределу, т. е. $X_{k+1} - X_k$. Для того чтобы понять, почему в случае соблюдения тех же условий, что и ранее, решение при данной постановке сохраняет описанное свойство, рассмотрим численный пример. Заметим, что перед всеми переменными y_k в (VII) стоят одинаковые коэффициенты, так что они оказывают одинаковое влияние на удовлетворение ограничений. Однако выпуклость c_1x_1 обусловливает монотонное возрастание коэффициентов в (VI). Следовательно, если задача состоит в минимизации $c(x)$, то, прежде чем в пробное решение войдет y_{k+1}, переменная y_k должна принять максимально возможное значение.

При подобной постановке задачи, когда все ограничения линейны, симплексный алгоритм с двусторонними ограничениями на переменные также может быть модифицирован с учетом правила ограниченного ввода в базис, обусловливающего обязательное выполнение свойства весов смежных точек. На любой итерации y_{k+1} может быть введено в решение только в том случае, если y_k входит в базис, причем с максимально допустимым значением, равным $X_{k+1} - X_k$; y_k может быть исключена из базиса, только если y_{k+1} не входит в базис и имеет нулевое значение.

Сепарабельные ограничения. Пусть оптимизационная задача включает сепарабельные ограничения

$$\sum_{j=1}^{n} a_j(x_j) \leqslant b^k \tag{9}$$

Здесь могут быть использованы изложенные выше методы линеаризации. Снова рассмотрим только одну переменную x_1 и соответствующую функцию $a_1(x_1)$. Пусть линеаризация основана на представлении (2) — (3). Выражение

$$\sum_{k=1}^{p} a_1(X_k)\, w_k, \tag{10}$$

представляющее собой кусочно-линейную аппроксимацию, подставим в (9) вместо $a_1(x_1)$.

Если все ограничения сепарабельны, как (9), и каждая из составляющих их функций выпукла, то использование преобразования (10) приводит к оптимальному решению, обладающему свойством весов смежных точек. Такое решение является приближением к глобальному оптимуму исходной задачи. Если хотя бы одна из составляющих функций не выпукла, то для обеспечения этого свойства, возможно, придется добавить к *критерию I симплексного метода* правило ограниченного ввода в базис, как это было сделано выше. В результате возможно получение локального, а не обязательно глобального оптимального решения исходной задачи.

Примеры. Рассмотрим иллюстративный пример задачи квадратичного программирования, приведенный в разд. 14.6:

$$\text{Максимизировать } (6x_1 - 3x_1^2) + (4x_2 - 2x_2^2) + \left(2x_3 - \frac{1}{3}x_3^2\right) \tag{11}$$

при условиях

$$x_1 + 2x_2 + x_3 \leqslant 4 \quad \text{и все} \quad x_j \geqslant 0. \tag{12}$$

Слагаемые $c(x)$ сгруппированы в виде (11) для того, чтобы был очевиден сепарабельный характер функции. Можно проверить, что каждая из функций, заключенных в скобки, вогнута. Пусть сетка точек для x_1 и x_2 имеет вид (0; 0,4; 0,7; 1), а для x_3 — (0; 1; 1,5; 2,3), так что

$$\begin{aligned} x_1 &\equiv 0u_1 + 0{,}4u_2 + 0{,}7u_3 + 1u_4, \\ x_2 &\equiv 0v_1 + 0{,}4v_2 + 0{,}7v_3 + 1v_4, \\ x_3 &\equiv 0w_1 + 1w_2 + 1{,}5w_3 + 2w_4 + 3w_5. \end{aligned} \tag{13}$$

Тогда ограничения сепарабельной модели можно записать в виде

$$\begin{gathered} 0{,}4u_2 + 0{,}7u_3 + 1u_4 + 0{,}8v_2 + 1{,}4v_3 + 2v_4 + \\ + 1w_2 + 1{,}5w_3 + 2w_4 + 3w_5 \leqslant 4, \\ \sum_{k=1}^{4} u_k = 1, \quad \sum_{k=1}^{4} v_k = 1, \quad \sum_{k=1}^{4} w_k = 1, \\ \text{все } u_k, \quad v_k, \quad w_k \geqslant 0. \end{gathered} \tag{14}$$

Целевая функция сепарабельной модели имеет вид (убедитесь в этом самостоятельно):

$$\text{Максимизировать} \quad 1{,}92u_2 + \ldots + 1{,}28v_2 + \ldots \\ \ldots + 1{,}6667w_2 + 3w_5. \tag{15}$$

Оптимальным решением является следующий набор переменных: $u_4 = 1$, $v_3 = 1$, $w_3 = 0{,}8$, $w_5 = 0{,}2$, которому соответствует значение целевой функции 7,15. Это эквивалентно значениям $x_1 = 1$, $x_2 = 0{,}7$, $x_3 = 0{,}8 \cdot 1{,}5 + 0{,}2 \cdot 2 = 1{,}6$ и значению целевой функции (11), равному 7,167. В то же время точное оптимальное решение имеет вид $x_1 = 0{,}875$, $x_2 = 0{,}625$, $x_3 = 1{,}875$ со значением (11), равным 7,25.

Рассмотрим «обратную» задачу максимизации

$$-x_1 - 2x_2 - x_3 \tag{16}$$

при условиях

$$-c(x) \leqslant -7{,}25 \quad \text{и все} \quad x_j \geqslant 0. \tag{17}$$

Оптимальная точка x остается той же, что и для (11) и (12). (Объясните, почему.) Используя ту же сетку точек и заменяя переменные на основе (13), получаем следующее оптимальное решение сепарабельной аппроксимирующей задачи: $u_4 = 1$, $v_3 = 1$, $w_3 = 0{,}568$, $w_4 = 0{,}432$. Это соответствует значениям $x_1 = 1$, $x_2 = 0{,}7$, $x_3 = 1{,}716$, причем значение (16) равно $-4{,}116$, что эквивалентно значению $c(x)$, равному 7,270.

Достижение сепарабельности. Подведем итоги. Пусть оптимизационная задача в целом может быть сформулирована следующим образом:

$$\text{Максимизировать} \quad \sum_{j=1}^{n} c_j(x_j) \tag{18}$$

при условиях

$$\sum_{j=1}^{n} a_{ij}(x_j) \leqslant b_i, \quad i=1,\ 2,\ \ldots,\ m, \tag{19}$$

где (19) может также включать ограничения неотрицательности переменных. Тогда описанные выше методы позволяют преобразовать модель в такую задачу, где все нелинейные функции аппроксимируются с помощью кусочно-линейных функций. Если все без исключения $c_j(x_j)$ вогнуты, а все $a_{ij}(x_j)$ выпуклы, то решение, полученное с помощью симплексного алгоритма, будет глобальным оптимумом для аппроксимирующей модели и, можно надеяться, хорошим приближением к глобальному оптимальному решению исходной задачи. Если же эти условия не удовлетворяются, то, модифицируя симплексный метод, можно отыскать хотя бы локальный оптимум аппроксимирующей задачи.

Важность этого подхода зависит от двух факторов. Один из них — возможность сформулировать нелинейную задачу таким образом, что целевая функция и ограничения включают только сепарабельные функции; этот аспект обсуждается ниже. Второй фактор — вычислительные трудности, возникающие при данном методе; этот аспект обсуждается несколько позже.

Удивительно большое число нелинейных выражений можно привести к сепарабельному виду, вводя дополнительные переменные и ограничения. В некоторой степени эти методы являются нестандартными — их можно даже скорее назвать «специальными приемами»,— и, следовательно, их нельзя классифицировать систематически. Однако на двух примерах можно показать характер используемых подходов.

Пусть либо в целевой функции, либо в ограничениях, либо в тех и других функциях содержатся произведения двух выражений:

$$f(x_1,\ \ldots,\ x_n)\cdot g(x_1,\ \ldots,\ x_n). \tag{20}$$

Простой пример этого — произведение $x_1 \cdot x_2$. Во-первых, для достижения сепарабельности заменим всюду произведение (20) одной переменной w. Во-вторых, введем еще две переменные y и z, связав их двумя определяющими ограничениями:

$$f(x_1,\ \ldots,\ x_n) - (y+z) = 0 \quad \text{и} \quad g(x_1,\ \ldots,\ x_n) - (y-z) = 0. \tag{21}$$

Если сами функции $f(x_1,\ \ldots,\ x_n)$ и $g(x_1,\ \ldots,\ x_n)$ являются сепарабельными, то ограничения (21) сепарабельны по *всем* переменным; если же нет, то необходимо применить дополнительные преобразования для превращения $f(x_1,\ \ldots,\ x_n)$ и $g(x_1,\ \ldots,\ x_n)$ в сепарабельные функции. В третьих, для завершения процесса наложим сепарабельное ограничение, связывающее w с y и z:

$$w - (y^2 - z^2) = 0. \tag{22}$$

Это обеспечивает справедливость уравнения $w \equiv f(x_1, \ldots, x_n) \times g(x_1, \ldots, x_n)$. (Объясните, почему.) Заметим, что функция $-y^2$ в (22) *не* является выпуклой. Короче говоря, произведение $f \cdot g$ (20) повсюду исключается, вместо него подставляется w, а к модели добавляются ограничения (21) и (22). Можно убедиться в том, что если (20) попросту имеет вид $x_1 \cdot x_2$, то для преобразования потребуются три дополнительных ограничения и три дополнительные переменные.

В качестве второго примера рассмотрим модель, которая включает выражение

$$[h(x_1, \ldots, x_n)]^{f(x_1, \ldots, x_n)}, \tag{23}$$

где $h(x_1, \ldots, x_n) > 0$ для всех допустимых x_j. Здесь вместо (23) подставим функцию e^w, а к модели добавим определяющее ограничение

$$w - f(x_1, \ldots, x_n) \cdot \ln h(x_1, \ldots, x_n) = 0. \tag{24}$$

Хотя функция (24) сама по себе не является сепарабельной, ее можно сделать такой с помощью того же метода, который был использован для достижения сепарабельности выражения (20), приняв $g \equiv \ln h$.

Вычислительные аспекты. Как читатель мог убедиться, преобразование модели нелинейного программирования в аппроксимирующую модель, включающую только сепарабельные функции, увеличивает размерность модели в двух направлениях. Во-первых, преобразование к сепарабельному виду заставляет вводить новые ограничения и переменные. Во-вторых, последующая линеаризация увеличивает число ограничений и переменных в еще большей степени. Если исходная задача содержит всего лишь несколько нелинейностей, этот подход вполне применим. Если же модель существенно нелинейна, то данный метод может и не оказаться столь же эффективным, как алгоритмы, приводимые в следующей главе.

Если характер целевой функции и ограничений таков, что обеспечивается свойство весов смежных точек, то можно применить обычный симплексный алгоритм, программы которого имеются на всех ЭВМ. Модифицированный алгоритм, обеспечивающий наличие этого свойства в обязательном порядке, имеется на большинстве мощных ЭВМ.

Разумеется, в случае применения настоящего метода к решению практической задачи необходимо уделять внимание вопросам построения сетки для каждой аппроксимации. Вообще говоря, желательно такое построение сетки, которое позволяет отыскать решение, близкое к оптимальному решению исходной задачи. Хотя для этого необходима сетка густо расположенных точек, ее использованию препятствует такой фактор, как большая размерность возникающих аппроксимирующих задач. Дальнейшее обсуждение этого вопроса выходит за рамки настоящей книги, однако необходимо отметить, что альтернативный алгоритм использования сепарабельности структуры излагается в разд. 15.10. При таком методе сетка не устанавливается *заранее*, однако на каждой итерации приходится выполнять больший объем вычислительных операций.

14.8. НЕПОСРЕДСТВЕННАЯ ЛИНЕАРИЗАЦИЯ

В отдельных случаях удается линеаризовать нелинейную целевую функцию $c(x_1, \ldots, x_n)$, преобразуя модель в эквивалентную задачу с линейной целевой функцией и линейными ограничениями, а затем решая ее симплексным методом. Разумеется, преобразованная задача является моделью линейного программирования, если на переменные налагаются линейные ограничения. Ниже рассматриваются три важных аспекта непосредственной линеаризации.

Минимизация суммы абсолютных значений отклонений. Рассмотрим следующую *минимизируемую* нелинейную целевую функцию:

$$c(x) \equiv \sum_{k=1}^{p} \Big| \sum_{j=1}^{n} c_{kj}x_j - f_k \Big|. \qquad (1)$$

Для выполнения необходимых преобразований дополним модель p линейными ограничениями:

$$\sum_{j=1}^{n} c_{kj}x_j - f_k = y_k - z_k, \quad k = 1, 2, \ldots, p, \qquad (2)$$

которые можно записать в следующем виде:

$$\sum_{j=1}^{n} c_{kj}x_j - y_k + z_k = f_k, \quad k = 1, 2, \ldots, p, \qquad (3)$$

где $$y_k = 0 \quad \text{и} \quad z_k = 0, \quad k = 1, 2, \ldots, p. \qquad (4)$$

Соответствующая *минимизируемая* линейная целевая функция преобразованной модели имеет вид

$$\sum_{k=1}^{p} (y_k + z_k). \qquad (5)$$

Идея преобразования состоит в том, что если выражение в левой части (2) положительно, то этому значению равно y_k, и $z_k = 0$, если отрицательно — этому значению равно $(-z_k)$, а $y_k = 0$. Оптимизация целевой функции (5) обеспечивает равенство нулю либо y_k, либо z_k, либо и y_k, и z_k. Таким образом, модель линейного программирования состоит из целевой функции (5), ограничений (3) и (4), а также всех прочих ограничений, которым должны удовлетворять x_j (например таких, как условия неотрицательности).

Достижение минимакса целевой функции. Теперь рассмотрим следующую *минимизируемую* целевую функцию:

$$c(x) = \max\Big(\sum_{j=1}^{n} c_{1j}x_j - f_1, \ \ldots, \ \sum_{j=1}^{n} c_{pj}x_j - f_p\Big), \qquad (6)$$

Для выполнения необходимых преобразований дополним модель p линейными ограничениями

$$\sum_{j=1}^{p} c_{kj}x_j - f_k \leqslant y, \quad k=1,\ 2,\ \ldots,\ p, \tag{7}$$

которые можно записать в следующем виде:

$$\sum_{j=1}^{n} c_{kj}x_j - y \leqslant f_k, \quad k=1,\ 2,\ \ldots,\ p. \tag{8}$$

Здесь y — переменная, на знак которой не налагается ограничений. Линейная целевая функция преобразованной модели имеет вид

$$\text{Минимизировать } y. \tag{9}$$

Предположим, что $c(x)$ в (6) представляется следующим образом:

$$c(x) \equiv \max\left(\left|\sum_{j=1}^{n} c_{1j}x_j - f_1\right|,\ \ldots,\ \left|\sum_{j=1}^{n} c_{pj}x_j - f_p\right|\right). \tag{10}$$

Заметим, что

$$\left|\sum_{j=1}^{n} c_{ij}x_j - f_i\right| = \max\left(\sum_{j=1}^{n} c_{ij}x_j - f_i,\ f_i - \sum_{j=1}^{n} c_{ij}x_j\right).$$

Следовательно, в дополнение к (8) и (9) модель максимизации целевой функции (10) должна включать ограничения

$$-\sum_{j=1}^{n} c_{kj}x_j - y \leqslant -f_k, \quad k=1,\ 2,\ \ldots,\ p \tag{11}$$

и $y \geqslant 0$. Целевую функцию (10) иногда называют **критерием Чебышева:** в ней минимизируется наибольшее абсолютное отклонение.

Как (6) и (10), так и целевая функция (1), минимизирующая сумму абсолютных значений отклонений, используются для выравнивания кривых и в задачах регрессионного анализа. Вычисляется p наборов показателей коэффициентов c_{kj} для n независимых переменных, где $p > n$, а f_k — соответствующие значения зависимой переменной. Тогда x_j — коэффициенты регрессии, значения которых необходимо выбрать для выравнивания кривой.

Задача дробно-линейного программирования. Пусть максимизируемая функция имеет вид

$$c(x) \equiv \frac{c_0 + \sum_{j=1}^{n} c_j x_j}{f_0 + \sum_{j=1}^{n} f_j x_j}. \tag{12}$$

(Эту модель обычно называют **моделью дробно-линейного программирования,** а иногда — **моделью гиперболического программирования.**) Для того чтобы при объяснениях избежать необходимости рассмотрения множества различных возможных вариантов, предположим, что на x_j налагаются такие ограничения, при которых знаменатель в (12) строго положителен для всех допустимых значений x_j, а также что максимум $c(x)$ является конечным.

Для того чтобы преобразовать задачу дробно-линейного программирования в модель линейного программирования, введем переменную r, определив ее следующим образом:

$$r \equiv \left(f_0 + \sum_{j=1}^{n} f_j x_j\right)^{-1}. \tag{13}$$

Тогда (12) можно записать в следующем виде:

$$c_0 r + \sum_{j=1}^{n} c_j x_j r. \tag{14}$$

Согласно сделанным допущениям, $r > 0$ при всех допустимых значениях x_j. Произведем замену переменных:

$$y_j \equiv r x_j. \tag{15}$$

Преобразованная модель имеет вид

$$\text{Максимизировать } c_0 r + \sum_{j=1}^{n} c_j y_j, \tag{16}$$

где в соответствии с (13) переменные r и y удовлетворяют линейным ограничениям

$$f_0 r + \sum_{j=1}^{n} f_j y_j = 1 \tag{17}$$

и $r > 0$. Отметим, что замена переменных x_j на основе (15) должна быть выполнена и во всех ограничениях, налагаемых на x_j. Так, например, если имеются дополнительные линейные ограничения

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \qquad i = 1, 2, \ldots, m, \tag{18}$$

то преобразованные ограничения можно построить, умножив каждое из уравнений (18) на r, что дает

$$\sum_{j=1}^{n} a_{ij} y_j - b_i r = 0, \quad i = 1, 2, \ldots, m. \tag{19}$$

Модель (12) применяется в некоторых задачах планирования производства, где возникают отходы. Так, пусть знаменатель (12) характеризует общее количество используемого сырья, а числитель — количество полезно используемого сырья (за вычетом отходов); тогда максимизируемое отношение — доля полезно используемого сырья.

14.9. МАКСИМИЗАЦИЯ ВЫПУКЛОЙ ЦЕЛЕВОЙ ФУНКЦИИ

Рассмотрим случай, при котором все ограничения линейны, а *максимизируемая* целевая функция $c(x)$ выпукла. О такой целевой функции иногда говорят, что она отражает экономичность укрупнения. По-прежнему предположим, что $c(x)$ регулярна и обладает конечным максимумом $\bar{c}$ на множестве допустимых значений x. Вместе с тем нет необходимости делать допущение, что все частные производные $\partial c/\partial x_j$ определены и непрерывны при всех значениях x, удовлетворяющих ограничениям. Допущение о выпуклости само по себе обеспечивает непрерывность $c(x)$ во *внутренней* части области допустимых решений; на *границах* этой области функция $c(x)$ может быть разрывной. Так, например, $c(x)$ может включать условно-постоянные затраты на операции по переналадке, которые возникают только при строго положительных x_j. Можно доказать следующую теорему.

Теорема о виде оптимального решения. Если все ограничения линейны и существует базисное допустимое решение, а функция $c(x)$ выпукла, регулярна и на множестве допустимых значений x имеет конечный максимум, то существует *базисное* решение, на котором $c(x)$ достигает максимума.

Этот результат можно использовать для нахождения оптимального решения, если число линейных ограничений мало или если модель обладает специальной структурой. Ниже приводятся примеры.

Рассмотрим модель распределения усилий, обсуждавшуюся в разд. 10.2; в частности, пусть модель имеет вид

$$\text{Максимизировать } R(y_1, y_2, \ldots, y_3) \tag{1}$$

при выполнении одного линейного ограничения

$$\sum_{j=1}^{n} H_j y_j = N, \tag{2}$$

где все $y_j = 0$ и являются непрерывными переменными. Предположим, что $R(y)$ является выпуклой функцией и что все $H_j > 0$. Тогда в сформулированной выше теореме утверждается, что существует оптимальная стратегия, для которой положительно значение только одной y_j. Оптимальную y_j можно найти, сравнивая значения $R(0, 0, \ldots, N/H_j, \ldots, 0)$ при $j = 1, 2, \ldots, s$. Если имеется m линейных ограничений, оптимальное решение можно найти, сравнивая значение целевой функции для всех возможных базисных решений, которых существует не более C_n^m, т. е. не более числа сочетаний из n по m. Если m мало или n невелико по сравнению с m, такой перебор осуществим на мощной ЭВМ. Например, при $m = 2$ необходимо рассмотреть всего $n(n-1)/2$ возможных вариантов.

Второй пример — модель управления запасами с детерминированным спросом и вогнутой функцией затрат, приведенная в разд. 9.5.

Если умножить целевую функцию на -1 и тем самым изменить направление оптимизации, то мы получим модель, которая соответствует предположениям о максимизации выпуклой целевой функции при линейных ограничениях.

Специальная структура системы линейных ограничений этой детерминированной модели управления запасами может быть использована для нахождения всех базисных допустимых решений. Пусть, например, спрос для каждого отрезка положителен, а исходный запас на начало планового периода равен нулю. Тогда для планового периода, содержащего N отрезков, имеется всего лишь 2^{N-1} базисных допустимых производственных программ, что меньше, чем C_{2N}^{N}. Каждая из этих программ соответствует некоторой схеме распределения по времени отрезков со строго положительным объемом выпуска (выпуск для отрезка 1 всегда положителен) и удовлетворяет теореме о виде оптимальной программы (разд. 9.5). Полный перебор можно реализовать, если совокупная сумма затрат есть вогнутая функция общего вида от всех переменных, характеризующих объем производства и уровень запасов; он легко выполним на мощной ЭВМ при $N \leqslant 12$. Однако если целевая функция к тому же сепарабельна по отрезкам, то применим алгоритм из разд. 9.6, так что объем вычислительных операций сокращается до рассмотрения $0{,}5N^2$ возможных вариантов.

Сформулированная выше теорема о виде оптимального решения справедлива для случая линейной, а следовательно выпуклой, $c(x)$. Поскольку линейная функция одновременно и вогнута, базисное решение, которое является локальным оптимумом, должно быть и глобальным оптимумом. Эти соображения служат еще одним доказательством того, что окончательное решение, полученное с помощью симплексного метода, есть глобальный оптимум.

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

1. Укажите несколько случаев использования математического программирования при принятии управленческих решений, когда следует ожидать наличие нелинейности в целевой функции или ограничениях. При описании существа нелинейности будьте как можно более конкретны. Укажите, являются ли соотношения эмпирически наблюдаемыми, полученными структурно или объединяют оба вида зависимостей. Возможна ли линейная аппроксимация? Перечислите трудности, которые могут возникнуть при получении данных, позволяющих численно характеризовать нелинейность.

2. Объясните, почему обычно труднее решить задачу нелинейного программирования, включающую n переменных и m ограничений, чем задачу линейного программирования той же размерности.

3. Покажите, как следует изменить задачу (6) — (8) разд. 14.1, если имеется также третье ограничение ($i = 3$) со случайным b_i, причем:

а) совместная вероятность удовлетворения всех трех ограничений не меньше чем β;

б) совместная вероятность удовлетворения ограничений $i = 1$ и $i = 2$ составляет β_1, ограничений $i = 2$ и $i = 3$ равна β_2, $i = 3$ и $i = 1$ равна β_3.

в) Покажите, как преобразовать (8) в эквивалентное ограничение, содержащее сумму двух нелинейных функций (при этом обязательно укажите, какие дополнительные допущения необходимы для выполнения преобразований).

4. Рассмотрите, какие бы вы предприняли действия в следующей ситуации: применяя итеративный алгоритм для решения нелинейной задачи (4) — (5) из разд. 14.1, получено допустимое решение x_1^k, x_2^k, . . ., x_n^k, значение целевой функции $c\,(x_1^k, \ldots, x_n^2)$ для которого близко к максимуму (об этом можно судить по известной верхней грани максимального значения целевой функции), однако имеются существенные основания полагать, что значения нескольких x_j^k далеки от оптимума.

5. Из допущений I) и II) разд. 14.3 следует, что функция $c\,(x)$ одной переменной x достигает максимума при значении x, находящемся на сегменте I. В каждом из приведенных вариантов постановки задачи определите, достигнет ли $c\,(x)$ максимума на этом сегменте, а если нет, то связано ли это с невыполнением допущения I), или допущения II), или обоих этих допущений вместе взятых.

а) $c\,(x) = 7 \quad$ при $\quad 3 \leqslant x < 8.$

б) $c\,(x) = \begin{cases} x & \text{при} \quad 0 \leqslant x < \frac{1}{2}, \\ x-1 & \text{при} \quad \frac{1}{2} \leqslant x \leqslant 1. \end{cases}$

в) $c\,(x) = \left|\frac{1}{x-2}\right| \quad$ при $\quad 1 \leqslant x \leqslant 3.$

г) $c\,(x) = \begin{cases} 0, & \text{если } x \text{ рациональное число,} \\ x & \text{в противном случае} \end{cases} \quad$ при $\quad 0 \leqslant x \leqslant 1.$

6. В каждом приведенном ниже варианте задачи укажите значения (если такие существуют) параметров a и b, при которых функция $c\,(x)$ выпукла на сегменте $[0,\ 1]$, т. е. $0 \leqslant x \leqslant 1$; вогнута; монотонна; унимодальна.

а) $c\,(x) = ax + b.$

б) $c\,(x) = ax^b.$

в) $c\,(x) = a^{bx}.$

г) $c\,(x) = a \log bx.$

д) Проделайте то же для $-1 \leqslant x \leqslant 0.$

е) Проделайте то же для $-1 \leqslant x \leqslant 1.$

7. В каждом из приведенных ниже вариантов постановки задачи либо докажите правильность сделанного утверждения, либо постройте опровергающий пример. (Можете считать, что функции являются функциями одной переменной x.)

а) Сумма двух выпуклых функций выпукла.

б) Сумма двух вогнутых функций вогнута.

в) Сумма двух монотонных функций монотонна.

г) Сумма двух унимодальных функций унимодальна.

д) Сумма выпуклой и вогнутой функций либо выпукла, либо вогнута.

е) Сумма унимодальной и выпуклой функций либо унимодальна, либо выпукла.

ж) Разность между двумя выпуклыми функциями может быть выпуклой, вогнутой, ни выпуклой и ни вогнутой, одновременно выпуклой и вогнутой.

з) Если $c(x)$ имеет регулярную вторую производную, то функция выпукла при $d^2c/dx^2 = 0$ для всех x.

и) Функция $c(x)$ вогнута тогда и только тогда, когда $-c(x)$ выпукла.

8. В каждом из приведенных ниже вариантов постановки задачи примените метод золотого сечения к решению примера, приведенного на рис. 14.4, приняв:

а) $5 \leqslant x \leqslant 10$;

б) $2 \leqslant x \leqslant 8$;

в) $c(x) = 18 - 2x$ при $6 \leqslant x \leqslant 12$, причем исходным сегментом для поиска является $[0, 10]$.

9. В каждом из приведенных ниже вариантов постановки задачи примените для поиска максимума метод золотого сечения; прекратите расчеты после $n = 3$.

а) Пример 2 на рис. 14.1, на сегменте $1 \leqslant x \leqslant 4$.

б) Пример 2 на рис. 14.1, на сегменте $0 \leqslant x \leqslant 3$.

в) Пример 3 на рис. 14.1, на сегменте $2 \leqslant x \leqslant 4$.

г) Пример 3 на рис. 14.1, на сегменте $0 \leqslant x \leqslant 2$.

д) Постройте верхнюю грань (II) разд. 14.3 для вариантов а) и б). Покажите, почему (II) невозможно построить для вариантов в) и г).

е) Решите варианты а) и б) методом дихотомии.

10. После описания в разд. 14.3 метода золотого сечения приводятся три свойства алгоритма, основывающиеся на определении r в (4). Докажите, что алгоритм действительно обладает этими свойствами.

11. Выведите формулу верхней грани (II), приведенную в конце разд. 14.3.

12. Рассмотрите общий алгоритм нахождения безусловного максимума, описанный в разд. 14.4.

а) Объясните, какие трудности могут возникнуть при использовании шага фиксированной, а не оптимальной длины.

б) Используйте пример, приведенный на рис. 14.6, для демонстрации следующего подхода к выбору направления. Пусть на первой итерации направление параллельно оси x_1, т. е. $d_1^0 = 1$, все остальные $d_j^0 = 0$. На следующей итерации направление параллельно оси x_2, т. е. $d_2^1 = 1$, все остальные $d_j^1 = 0$. Продолжим такое движение, возвращаясь в конце концов снова к $d_1^k = 1$ и т. д. (Этот процесс называется *покоординатным подъемом.*) На всех итерациях используйте, шаг оптимальной длины. За сколько итераций достигается сходимость? Поверните оси эллипсов, изображенных на рис. 14.6, на 45°. Снова примените метод и исследуйте сходимость.

в) С помощью двумерного чертежа, аналогичного рис. 14.6, разъясните, почему оптимальная длина шага определяется условием касания к линии уровня. Попытайтесь начертить соответствующее трехмерное изображение (третье измерение характеризует значения целевой функции).

13. а) Покажите, что если $c(x_1, x_2, \ldots, x_n)$ есть квадратичная функция, то выражение, максимизация которого необходима для нахождения оптимальной длины шага, также является квадратичной функцией длины шага.

б) Пусть дана квадратичная функция $c(x)$ одной переменной, *строго* вогнутая для всех значений x. Выведите формулу безусловного максимума $c(x)$ и обоснуйте вывод.

14. Рассмотрите численный пример (4) — (5) разд. 14.5.

а) Проверьте правильность (6).

б) Если (6) дано, выполните промежуточные преобразования, приводящие к определению оптимальной длины шага по формуле (7).

в) Проверьте результаты, приведенные в таблице рис. 14.8, для $k = 0$, 1 и 2.

15. Рассмотрите численный пример (4) и (5) разд. 14.5. Примените метод скорейшего подъема с оптимальной длиной шага для каждой из указанных ниже исходных точек. Прекратите расчеты после выполнения всех операций для $k = 3$.

а) $x_1^0 = x_2^0 = 0$ и $x_3^0 = 3$.

б) $x_1^0 = 0$, $x_2^1 = 1$ и $x_3^0 = 3$.

в) $x_1^0 = \frac{1}{3}$, $x_2^0 = 0$ и $x_3^0 = -\frac{5}{3}$.

г) $x_1^0 = 2$, $x_2^0 = 2$ и $x_3^0 = 6$.

16. Рассмотрите метод Ньютона — Рафсона выбора направлений, а именно выбор на основе системы (8), приведенной в разд. 14.5. Покажите, что принятие каждой $x_j^0 = 0$ и решение системы (8) эквивалентно решению системы $\partial c/\partial x_j = 0$ для каждого j, если $c(x)$ квадратична.

17. Объясните, что́ происходит, если, например, размер шага принят равным 0,9 оптимальной его длины, а для определения направлений используется метод скорейшего подъема. Для иллюстративных целей воспользуйтесь линиями уровня, аналогичными изображенным на рис. 14.5. Объясните, почему так называемый *метод субрелаксации* может обеспечить более быструю сходимость, чем подход, основанный на использовании оптимального значения шага.

18. а) Покажите, что если $c(x)$ строго вогнута при всех $x = (x_1, x_2, \ldots, x_n)$, причем $c(\bar{x})$ есть максимальное значение, то не существует другого $x^* \neq \bar{x}$, такого, что $c(x^*) = c(\bar{x})$, т. е. имеется единственная точка, в которой $c(x)$ достигает максимума.

б) Покажите, что если $c(x)$ есть вогнутая функция, то и $c(x_1 + td_1, \ldots, x_n + td_n)$ является вогнутой функцией t.

19. Рассмотрите функцию $c(x)$ одной переменной и предположите, что она удовлетворяет допущениям I), II) и III) из разд. 14.3.

а) Является ли условие $dc/dx = 0$ в точке x^* необходимым, или достаточным, или и необходимым и достаточным, для того чтобы $c(x^*)$ было значением безусловного максимума?

б) Предложите условие, которому должно удовлетворять x^*, необходимое и достаточное для того, чтобы $c(x^*)$ было максимальным значением. (В своем ответе будьте осторожны, поскольку условие $dc/dx = 0$ может соблюдаться на некотором интервале значений x.)

20. В каждом из приведенных ниже вариантов постановки задачи запишите функцию $c(x)$ в виде, аналогичном (4) разд. 14.6. Вычислите все первые частные производные. Приведите определяющее соотношение, аналогичное (7), для свободной переменной u, соответствующей x_2.

а) $3x_1 - 2x_2 + x_3 + 7x_1^2 + 3x_1x_2 - 8x_2x_3 + 20x_2^2$.

б) $5x_1 + 9x_3 + 4x_1x_2 - 2x_1x_3 + 6x_2^2 + x_2x_3 - 3x_3^2 + x_3x_1 + 6x_3x_2$.

в) $3(-4x_1 + 2x_2 + 3x_3 - 5)^2$.

г) $(2x_1 + 3x_2 + 1)^2 + 2(4x_1 - x_3 + 2)^2 + 3(-5x_2 + 5x_3)^2$.

д) В каждом из приведенных выше выражений исключите переменную x_2, произведя замену переменных на основе соотношения для свободной переменной u.

е) В каждом из приведенных выше выражений исключите переменную x_2, произведя подстановку на основе соотношения $1 - x_1 - x_3$.

21. Рассмотрите функцию $c(x)$, приведенную в (3) разд. 14.6, и ее частные производные (5). Пусть пробные значения всех x_j равны нулю.

а) Улучшится ли значение целевой функции, если ввести в решение x_1? x_2? x_3? Опишите изменение значений целевой функции, если x_1 возрастает, а все остальные x_j остаются равными нулю; то же, если возрастает x_2.

б) При использовании *критерия II квадратичного симплексного метода* значение некоторой выбранной небазисной переменной x_j может быть доведено до сколь угодно большой величины без нарушения ограничений. Может также оказаться, что $\partial c/\partial x_j > 0$ для всех неотрицательных значений x_j (если все остальные небазисные переменные остаются равными нулю). Что произойдет, если оба эти случая возникнут одновременно? Какое допущение, сделанное в начале разд. 14.6, исключает возможность одновременного возникновения таких двух случаев?

22. Рассмотрите пример (8) — (9) разд. 14.6, решенный квадратичным симплексным алгоритмом.

а) Приведите все промежуточные вычисления и проверьте результаты (10) — (30).

б) Дойдя до (20), введите в решение вместо u_1 переменную u_2 и продолжайте расчеты, пока не будет получено оптимальное решение.

23. Рассмотрите пример (8) — (9) разд. 14.6. Примените квадратичный симплексный алгоритм, изменив условия примера так, как это указано в приводимых ниже вариантах постановки задачи.

а) Правая часть (9) равна 8 (вместо 4).

б) Правая часть (9) равна 3 (вместо 4).

в) Правая часть (9) равна 2 (вместо 4).

г) Добавьте ограничение $x_1 + x_2 \leqslant 1$.

д) Добавьте ограничение $x_2 + x_3 \leqslant 2$.

е) Добавьте ограничение $-2x_1 + 2x_3 \leqslant 1$.

ж) Целевая функция имеет вид $6x_1 + 4x_2 + 2x_3 - \frac{1}{3} x_1^2 - 2x_2^2 - 3x_3^2$.

з) Целевая функция имеет вид $2x_1 + 6x_2 + 4x_3 - 3x_1^2 - 2x_2^2 - \frac{1}{3} x_3^2$.

и) Добавьте к целевой функции слагаемые $-x_1x_2 - 2x_2x_3$.

24. Примените квадратичный симплексный алгоритм, описанный в разд. 14.6, для решения примеров, изображенных на рис. **14.9**, для которых линейные ограничения соответствуют трем вариантам постановки задачи, приведенным ниже. Нанесите проверяемые на каждой итерации решения на рисунок, аналогичный рис. **14.9**.

а) Вариант 1:
$$\begin{aligned} x_1 + 2x_2 &\leqslant 6, \\ 8x_1 + 6x_2 &\leqslant 23, \\ 6x_1 + x_2 &\leqslant 12. \end{aligned}$$

б) Вариант 2:
$$\begin{aligned} x_1 + 2x_2 &\leqslant 5, \\ 16x_1 + 12x_2 &\leqslant 37, \\ 6x_1 + x_2 &\leqslant 12. \end{aligned}$$

в) Вариант 3:
$$\begin{aligned} x_1 + 2x_2 &\leqslant 5, \\ 16x_1 + 12x_2 &\leqslant 37, \\ 24x_1 + 4x_2 &\leqslant 27. \end{aligned}$$

25. Рассмотрите пример сепарабельной функции $c_1(x_1) = x_1^2$, приведенный в разд. 14.7. Запишите формулы, аналогичные (5) — (8), если $0 \leqslant x_1 \leqslant 20$, а сетка включает точки 0, 5, 8, 10, 15 и 20. Какие веса при такой аппроксимации соответствуют $x_1 = 6$ и насколько близко к действительному аппроксимированное значение целевой функции? Ответьте на эти же вопросы для $x_1 = 9$; 12; 17.

26. Пусть $c_1(x_1) = \sqrt{x_1}$. Объясните, как построить кусочно-линейную аппроксимацию для $0 \leqslant x_1 \leqslant 10$ при точках сетки, совпадающих с 0, 1, 4, 9, 10. Покажите, в каком виде представляется линейное ограничение $3x_1 + 4x_2 \geqslant 12$ после замены переменных. Какие веса при этой аппроксимации соответствуют $x_1 = 6$?

27. а) Приведите обоснования свойства весов смежных точек, сформулированного в разд. 14.7. (*Указание*: используйте пример (5) — (8) для иллюстрации приводимых соображений; помните, что x^2 — выпуклая функция.)

б) Пусть требуется максимизировать $c(x)$, имеющую вид (1), причем $c_1(x_1) = x_1^2$, а для всех прочих $c_j(x_j) = c_j x_j$; используется аппроксимация (5) — (6). Пусть, далее, истинным оптимальным решением является $x_1 = 8$. Какие трудности могут возникнуть, если не воспользоваться правилом ограниченного ввода в базис?

в) Приведите разумные доводы в пользу того, что если все ограничения сепарабельны, как в (9), а каждый их элемент является выпуклой функцией, то использование преобразования типа (10) приводит к оптимальному решению, которое обладает свойством весов смежных точек.

28. Рассмотрите альтернативный метод представления сепарабельной функции, описанный и продемонстрированный на примере (I) — (VII) в разд. 14.7.

а) Запишите формулы, аналогичные (VI) и (VII), если $0 \leqslant x_1 \leqslant 20$, при точках сетки, совпадающих с 0, 5, 8, 10, 20. Какие значения y_k соответствуют $x_1 = 6$ и насколько близко к действительному аппроксимированное значение целевой функции? Ответьте на эти же вопросы для $x_1 = 9$; $x_1 = 12$; $x_1 = 17$.

б) Пусть $c_1(x_1) = \sqrt{x_1}$. Объясните, как построить кусочно-линейную аппроксимацию для $0 \leqslant x_1 \leqslant 10$ при точках сетки, совпадающих с 0, 1, 4, 9, 10. Покажите, в каком виде представляется линейное ограничение $3x_1 + 4x_2 = 12$ после замены переменных. Какие значения y_k соответствуют $x_1 = 6$?

29. Рассмотрите пример (11) — (12) разд. 14.7.

а) Полностью напишите целевую функцию (15). Убедитесь в том, что приведенное решение действительно оптимально.

б) Запишите формулы, аналогичные (13) — (15), если сетка для x_1 и x_2 имеет вид (0; 0,3; 0,6; 0,8; 1), а для $x_3 = 0$; 1,2; 1,4; 1,6; 1,8; 3.

в) Найдите оптимальное решение для аппроксимированной задачи варианта б). (*Указание*: используйте имеющуюся информацию об истинном оптимальном решении.) Покажите, что этот ответ соответствует исходной задаче.

г) Рассмотрите «обратную» задачу (16) — (17). Найдите оптимальное решение для сепарабельной аппроксимации, используя сетки, приведенные в варианте б). Покажите, что этот ответ соответствует исходной задаче.

30. Преобразуйте каждый приведенный выше вариант задачи к сепарабельному виду.

а) Максимизировать $x_1x_2 + x_2x_3 + x_1x_2x_3$ при условиях

$$x_1^2 + x_1 - 2x_1x_2x_3 \leqslant 0,$$
$$2x_2^2 + 3x_2 + 6x_2x_3 \leqslant 12,$$
$$x_3^2 + 4x_1x_3 \leqslant 0,$$
$$\text{все } x_j \geqslant 0.$$

б) Максимизировать $(x_1 + 2x_2 + 3x_3)(x_1 + x_3)$ при условиях

$$(x_1 - x_2 + 4x_3)^2 + 3(2 + x_1x_2)^{x_1x_3+x_2} \leqslant 52,$$
$$x_1x_2 \geqslant 1,$$
$$\text{все } x_j \geqslant 0.$$

в) Максимизировать $5(2 + x_1x_2 + x_3)(x_3/x_4)^{x_1-x_2}$ при условиях

$$x_1x_2x_3x_4 \leqslant 16,$$
$$x_1 \geqslant 0, \quad x_2 = 0, \quad x_3 = 2, \quad x_4 \geqslant 2.$$

г) Максимизировать $4x_1x_2x_3$ при условиях

$$(5 + 3x_1)^{x_1+2x_2^2}\left(\frac{2^{x_3}}{x_3}\right) \leqslant 100,$$
$$x_1 \geqslant 0, \quad x_2 \geqslant 0, \quad x_3 \geqslant 1.$$

д) Пусть после достижения в каждом из вариантов сепарабельности мы аппроксимируем каждую нелинейную сепарабельную функцию кусочно-линейной функцией, используя сетку с 10 точками (для каждой функции). Определите размерность соответствующей задачи линейного программирования. Приведите расчеты со всеми подробностями.

31. Рассмотрите три целевые функции (1), (6) и (10) разд. 14.8. Пусть сущность задачи заключается в построении прямой регрессии; c_{k1} — k-е значение независимой переменной, c_{k2} — «коэффициент» точки пересечения, f_k — наблюденное значение, соответствующее c_{k1}. Таким образом, x_1 есть тангенс угла наклона выравненной прямой регрессии, а x_2 — точка пересечения. Для конкретности примем $c_{k1} = k$ и $c_{k2} = 1$ для $k = 1, 2, \ldots, 10$ и

$$f_1 = 87, \quad f_2 = 100, \quad f_3 = 107, \quad f_4 = 98, \quad f_5 = 88,$$
$$f_6 = 108, \quad f_7 = 104, \quad f_8 = 113, \quad f_9 = 116, \quad f_{10} = 107.$$

а) Если на x_1 и x_2 не налагается никаких ограничений, сформулируйте со всеми подробностями эквивалентную задачу линейного программирования, где целью ставится минимизация функции $c(x)$, заданной соответственно выражениями (1), (6) или (10).

б) Укажите, как изменится постановка для варианта б), если дополнительно налагаются ограничения $\frac{1}{2} \leqslant x_1 \leqslant 2$ и $x_2 \geqslant 2$.

в) Постройте двойственную задачу линейного программирования, соответствующую вашему ответу для варианта а).

г) Покажите, как формулируется оптимизационная модель, имеющая целью минимизацию суммы квадратов значений отклонений искомой прямой от f_k. Примите, что на x_1 и x_2 не налагается ограничений; затем наложите ограничения, указанные в варианте б).

32. Рассмотрите задачу:

$$\text{Минимизировать } \left[\max\left(x-2,\ -x,\ -\frac{1}{4}x-\frac{1}{2}\right)\right].$$

а) Нанесите на график каждую из линейных функций, покажите кусочно-линейный график максимума трех функций и укажите местонахождение минимума.

б) Сформулируйте задачу, используя подход, аналогичный (8) и (9) из разд. 14.8. Затем начертите график, показывающий область допустимых значений x и y; укажите оптимальное решение.

в) Ответьте на те же вопросы, что в вариантах а) и б), добавив также ограничения (11), для целевой функции

$$\text{Минимизировать } \left[\max\left(|x-2|,\ |-x|\left|-\frac{1}{4}x-\frac{1}{2}\right|\right)\right].$$

Если вы получаете отличающиеся ответы, объясните почему.

33. Для каждого из приведенных ниже вариантов постановки задачи примите $0 \leqslant x \leqslant 1$; максимизируемая функция $c(x)$ указана в соответствующем варианте. Начертите график функции $c(x)$ для допустимых значений x. Постройте эквивалентную задачу линейного программирования, используя подход (13) — (19) из разд. 14.8. Начертите рисунок, показывающий область допустимых решений для этой эквивалентной задачи; укажите оптимальное решение.

$$\text{а) } \frac{2+3x}{4-x}; \quad \text{б) } \frac{2-3x}{4-x}; \quad \text{в) } \frac{2-2x}{4+x}; \quad \text{г) } \frac{2+2x}{4+x}.$$

34. Преобразуйте следующую задачу в эквивалентную модель линейного программирования:

$$\text{Максимизировать } \frac{-3+2x_1+4x_2-5x_3}{6+3x_1-x_2}$$

при условиях

$$x_1 - x_2 \geqslant 0,$$
$$7x_1 + 9x_2 + 10x_3 \leqslant 30,$$
$$x_1 = 0, \quad x_2 \leqslant 1, \quad x_3 \geqslant 0.$$

35. Продумайте доказательство теоремы о виде оптимального решения, приведенного в разд. 14.9. (*Указание*: для иллюстрации используйте рис. 14.9, где максимизируется $-c(x)$.)

36. Найдите оптимальное решение задачи:

$$\text{Максимизировать } (4y_1 + 3y_2 - y_3 + 1)^2$$

при условиях

$$3y_1 + 4y_2 + 2y_3 = 12,$$
$$\text{все } y_j \geqslant 0.$$

Обоснуйте свой ответ.

37. В разд. 14.9 рассматриваются вычислительные аспекты того обстоятельства, что нам известна возможность нахождения оптимума только в базисной или экстремальной точке. Варианты постановки а) и б) относятся к численным оценкам, упомянутым в тексте.

а) Вычислите максимальное число базисных допустимых решений задачи с m линейными ограничениями и n переменными, где $m = 2, 3, 4$, а $n = 6, 7, 8$ (всего для 9 попарных сочетаний m и n).

б) Вычислите 2^{N-1} и C_{2N}^{N} для $N = 1, 2, 3, 4, 5, 6$.

38. Объясните, как вы понимаете следующие термины:

нелинейное программирование;
экстремальное решение;
внутренняя точка;
граничная точка;
безусловная максимизация (минимизация);
верхняя грань;
разрывность;
непрерывная функция;
ограниченная функция;
выпуклая функция;
вогнутая функция;
строго выпуклая (вогнутая) функция;
квазивыпуклая (квазивогнутая) функция;
унимодальная функция;
адаптивный (последовательный) поиск;
локальный максимум (минимум);
глобальный максимум (минимум);
точка перегиба;
метод золотого сечения;
метод дихотомии, или деление пополам;
необходимое и (или) достаточное условие;
направление;
длина шага (оптимальная длина шага);
линии уровня;
метод скорейшего подъема;
«недальновидное» правило;
градиент;
производная по направлению;
сходимость в бесконечности;
сходящаяся последовательность (подпоследовательность);
ортогональный;

зигзагообразное движение;
метод Ньютона — Рафсона;
модифицированный градиентный метод;
метод субрелаксации;
стационарная точка;
задача квадратичного программирования;
квадратичный симплексный алгоритм;
свободная переменная;
приведенная частная производная;
квадратичная форма;
отрицательно (положительно) полуопределенная форма;
сепарабельная функция;
кусочно-линейная аппроксимация;
свойство весов смежных точек;
правило ограниченного ввода в базис;
непосредственная линеаризация;
критерий Чебышева;
дробно-линейное (гиперболическое) программирование.

УПРАЖНЕНИЯ НА ПОСТАНОВКУ ЗАДАЧ И РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

39. Рассмотрите квадратичный симплексный метод, описанный в разд. 14.6.

а) Предложите способ применения алгоритма, если не все $b_i = 0$ и не все s_i выписаны в явном виде в ограничениях (2).

б) Рассмотрите операции алгоритма, если не имеется ни линейных ограничений, ни ограничений неотрицательности. Дополните ваши рассуждения иллюстрациями на примере целевой функции (8).

в) Поясните операции алгоритма, пользуясь терминами разд. 14.4, а именно исходя из выбора направлений и длины шага. Убедитесь в том, что вам в явном виде известны направления, выбираемые на каждой итерации. Укажите также, является ли длина шага оптимальной. [*Указание*: поясните ответ на примере (8) — (9) разд. 14.6.]

40. Рассмотрите целевую функцию

$$\text{Максимизировать } 6x_1 - 2x_1^2 + 2x_1x_2 - 2x_2^2.$$

а) Решите задачу методом скорейшего подъема, выбрав в качестве исходного решения $x_1^0 = x_2^0 = 0$. Начертите решения, получаемые на каждой итерации, на двумерном рисунке. Прекратите расчеты после $k = 5$ итераций.

б) Ответьте на вопросы варианта а), начав с $x_1^0 = x_2^0 = 10$.

в) Наложите ограничения $x_1 + x_2 \leqslant 2$, $x_1 = 0$ и $x_2 = 0$. Решите задачу квадратичным симплексным методом. Нанесите решения, получаемые на каждой итерации, на рисунок, изображающий область допустимых решений.

г) Ответьте на вопросы варианта в), приняв $x_1 + x_2 \leqslant 3$.

д) Ответьте на вопросы варианта в), приняв $x_1 + x_2 \leqslant \frac{1}{2}$.

е) Ответьте на вопросы варианта в), приняв, что линейное ограничение имеет вид $2x_1 + x_2 \leqslant 2$.

ж) Ответьте на вопросы варианта в), приняв, что линейное ограничение имеет вид $x_1 + 2x_2 \leqslant 2$.

з) Решите задачу безусловной максимизации, используя квадратичный симплексный метод.

и) Ответьте на вопросы варианта в), дополнительно добавив ограничение $x_1 \leqslant 1$.

к) Ответьте на вопросы варианта в), дополнительно добавив ограничение $x_1 \leqslant 1\frac{7}{12}$.

л) Ответьте на вопросы варианта в), дополнительно добавив ограничение $x_2 \leqslant \frac{1}{6}$.

м) Ответьте на вопросы варианта в), дополнительно добавив ограничение $x_2 \leqslant \frac{5}{12}$.

41. Рассмотрите целевую функцию

$$\text{Максимизировать } 36x_1 + 36x_2 - 5x_1^2 - 8x_1x_2 - 5x_2^2.$$

а) Решите задачу методом скорейшего подъема, начав с $x_1^0 = x_2^0 = 0$. Нанесите решения, получаемые на каждой итерации, на двумерный рисунок. Прекратите расчеты после $k = 5$ итераций.

б) Ответьте на вопросы варианта а), начав с $x_1^0 = x_2^0 = 3$.

в) Наложите ограничения $x_1 + x_2 \leqslant 2$, $x_1 \geqslant 0$ и $x_2 \geqslant 0$. Решите задачу квадратичным симплексным методом. Нанесите решения, получаемые на каждой итерации, на рисунок, изображающий область допустимых решений.

г) Ответьте на вопросы варианта в), приняв, что линейное ограничение имеет вид $x_1 + 3x_2 \leqslant 2$.

д) Ответьте на вопросы варианта в), дополнительно добавив ограничение $x_1 \leqslant \frac{1}{2}$.

е) Наложите ограничения $x_1 - x_2 \leqslant 3$, $-x_1 + x_2 \leqslant 3$, $x_1 \geqslant 0$ и $x_2 \geqslant 0$. Решите задачу квадратичным симплексным методом. Нанесите решения, получаемые на каждой итерации, на рисунок, изображающий область допустимых решений.

42. Рассмотрите целевую функцию

$$\text{Максимизировать } 16x_1 + 16x_2 + 40x_3 - 4x_1^2 + 6x_1x_2 - 6x_1x_3 - 11x_2^2 + 8x_2x_3 - 11x_3^2.$$

а) Решите задачу методом скорейшего подъема, начав с $x_1^0 = x_2^0 = x_3^0 = 0$. Прекратите расчеты после $k = 5$ итераций.

б) Ответьте на вопросы варианта а), начав с $x_1^0 = x_2^0 = x_3^0 = 3$.

в) Наложите ограничения $x_1 + x_2 + x_3 \leqslant 4$ и все $x_j \geqslant 0$. Решите задачу квадратичным симплексным методом.

г) Решите задачу на безусловный максимум, используя квадратичный симплексный метод.

43. Рассмотрите целевую функцию

$$\text{Максимизировать } 40x_1 + 10x_2 + 30x_3 - 6x_1^2 - 8x_1x_2 - 4x_1x_3 - 11x_2^2 + 14x_2x_3 - 9x_3^2.$$

а) Решите задачу методом скорейшего подъема, начав с $x_1^0 = x_2^0 = x_3^0 = 0$. Прекратите расчеты после $k = 5$ итераций.

б) Ответьте на вопросы варианта а), начав с $x_0^2 = x_2^0 = x_3^0 = 5$.

в) Ответьте на вопросы варианта а), начав с $x_1^0 = 2$, $x_2^0 = 3$, $x_3^0 = 1$.

г) Наложите ограничения $x_1 + x_2 + x_3 \leqslant 3$ и все $x_j \geqslant 0$. Решите задачу квадратичным симплексным методом.

д) Решите задачу на безусловный максимум, используя квадратичный симплексный метод.

44. Рассмотрите оптимизационную модель математического программирования, включающую n неотрицательных переменных, линейную целевую функцию, $m - 1$ линейных ограничений и одно квадратичное ограничение

$$\sum_{j=1}^{n} a_j x_j + \sum_{j=1}^{n} \sum_{k=1}^{n} a_{jk} x_j x_k \geqslant b,$$

левая часть которого вогнута. Предложите вычислительную схему нахождения численного решения данной задачи, использующую алгоритмы линейного и квадратичного программирования.

45. Рассмотрите задачу

$$\text{Максимизировать } 2x_1^2 + x_2^2$$

при условиях

$$2x_1 + x_2 \leqslant 2, \quad x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

а) Решите задачу на основе ее анализа.

б) Примените квадратичный симплексный алгоритм. Сперва начните с исходного базисного решения, включающего только дополняющие переменные. Затем в качестве исходного базисного решения примите $x_1 = 1$. Рассмотрите особенности применения алгоритма в данной задаче.

в) Преобразуйте модель в аппроксимированную задачу, используя приемы сведе́ния к сепарабельному виду, изложенные в разделе 14.7. Пусть сетка для x_1 имеет вид $\left(0, \frac{1}{2}, \frac{3}{4}, 1\right)$, а для x_2 — $\left(0, \frac{1}{2}, 1, \frac{3}{2}, 2\right)$. Для нахождения решения примените симплексный метод. Обсудите результаты. (Пришлось ли использовать правило ограниченного ввода в базис?)

46. Рассмотрите задачу

$$\text{Максимизировать } 6x_1 + 8x_2 - x_1^2 - x_2^2$$

при условиях

$$4x_1^2 + x_2^2 \leqslant 16,$$

$$3x_1 + 5x_2 \leqslant 15,$$

$$x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

а) Преобразуйте модель в аппроксимированную задачу, используя приемы сведе́ния к сепарабельному виду, изложенные в разд. 14.7. Пусть сетка для x_1 имеет вид $\left(0,\ \frac{1}{2},\ 1,\ \frac{3}{2},\ 2\right)$, а для $x_2\,(0,\ 1,\ 2,\ 3,\ 4)$.

б) Найдите оптимальное решение аппроксимированной задачи варианта а). Начертите область допустимых решений и нанесите как решения, получаемые на каждой из итераций, так и истинное оптимальное решение.

в) Измените знак всех коэффициентов целевой функции, а также знак неравенства в первом ограничении. Соответственно измените задачу варианта а). Для нахождения решения используйте симплексный метод. (Пришлось ли использовать правило ограниченного ввода в базис?) Начертите область допустимых решений и нанесите как решения, получаемые на каждой из итераций, так и истинное оптимальное решение.

47. Задача о двух картошках (разд. 1.6). Напомним, что имеются два поставщика картофеля, причем из сырья, поставляемого каждым из них, возможен выпуск трех видов продукции, однако процент выхода каждого из видов продукции для разных поставщиков неодинаков. Предположим, что можно, при определенных затратах, несколько изменить проценты выхода. Пусть f_1, f_2 и f_3 — выход (в долях единицы) каждого из трех продуктов на единицу веса картофеля — сырья, полученного от поставщика 1, а g_1, g_2 и g_3 — аналогичные показатели для поставщика 2. Пусть каждое из значений f_i и g_i может изменяться в пределах $\pm 10\%$ от значений, приведенных на рис. 1.1, а $c_1\,(f_1,\ f_2,\ f_3)$ и $c_2\,(g_1,\ g_2,\ g_3)$ — *затраты*, связанные с достижением соответствующих выходов.

а) Сформулируйте соответствующую оптимизационную модель.

б) Примените приемы, описанные в разд. 14.7, для приведения задачи к сепарабельному виду. Обсудите вопрос, целесообразно ли для получения численного решения использовать правило ограниченного ввода в базис.

48. Некоторая фирма желает применить научный подход для объективного установления уровней заработной платы. Структуру ставок предполагается основывать на следующей системе балльных оценок. Отдел нормирования оценивает каждый вид работы по трем различным характеристикам — требуемая квалификация, степень ответственности, необходимый трудовой стаж. Пусть для первой характеристики установлено пять уровней, обозначенных $A1$, $A2$, $A3$, $A4$ и $A5$, причем $A1$ соответствует низшей, а $A5$ — высшей квалификации. Аналогично этому для второй характеристики установлено четыре уровня: $B1$, $B2$, $B3$ и $B4$, где $B1$ — низший, а $B4$ — высший уровень. Наконец, для третьей характеристики предусмотрено пять уровней — от $C1$ до $C5$. Фирма желает присвоить каждому уровню любой из характеристик оценку в денежных единицах, так чтобы сумма оценок могла служить основой для установления средней заработной платы по данной работе. Так, например, пусть некоторому виду работы соответствуют уровни $A2$, $B3$ и $C5$. Если уровню $A2$ присвоена оценка 300 единиц, $B3$ — 400 единиц и $C5$ — 250 единиц, то в результате возникает *средняя* ставка $300 + 400 + 250 = 950$ единиц.

Предположим, что фирма определила уровни характеристик по P видам работ и под давлением профсоюза соглашается на 10%-ное повышение ставок по сравнению с существующими по каждому виду работ. Пусть a_p, b_p и c_p — определившиеся уровни по трем характеристикам для работы p, а s_p — 110% от существующей в настоящее время *средней* ставки. Фирма желает присвоить оценки уровням $A1, A2, \ldots, A5, B1, \ldots, B4, C1, \ldots, C4$, так чтобы минимизировать сумму квадратов разностей между возникающей в результате этого новой *средней* ставкой и s_p при $p = 1, 2, \ldots, P$. Сформулируйте соответствующую задачу нелинейного программирования. Примите, что разница в оценке между смежными уровнями одной и той же характеристики должна быть не меньше 20 денежных единиц (например, оценка уровня $A2$ должна не менее чем на 20 единиц превышать оценку уровня $A1$). Примите также, что оценка каждого из низших уровней $A1$, $B1$ и $C1$ должна быть не менее 25 единиц. Построение ограничений для отдельных работ продемонстрируйте на следующих примерах: для работы 1 определены уровни $A3$, $B1$ и $C4$, а $s_1 = 10\,000$ единиц; для работы 2 определены уровни $A2$, $B3$ и $C5$, тогда как $s_2 = 12\,000$ единиц.

49. Рассмотрите регрессионную модель $y = f(x) + e$, где y — зависимая переменная, x — независимая переменная, e — ошибка (отклонение). Пусть имеется p пар наблюденных значений независимой и зависимой переменных, которые обозначим x_k и y_k, $k=1, 2, \ldots, p$; по этим данным необходимо определить функцию $f(x)$. Не предполагается какой-либо конкретный вид функции, например что функция является линейной или квадратичной. Для простоты предположим, что все значения x_k различны, т. е. что $x_i \neq x_j$ при $i \neq j$, а также что x_k упорядочены таким образом, что $x_1 < x_2 <, \ldots, < x_p$. Пусть f_k — значение $f(x_k)$ согласно предлагаемому уравнению регрессии. (Заметьте, что знак f_k может быть любым.)

а) Составьте систему уравнений, по одному для каждого k, определяющих отклонение e_k — разность между расчетным значением f_k и наблюденным значением зависимой переменной y_k. (Заметьте, что знак e_k может быть любым.)

б) Если на значения f_k не налагается других ограничений, укажите, какими должны быть их оптимальные значения, минимизирующие сумму $\sum_{k=1}^{p} e_k^2$.

в) Объясните, какие ограничения следует наложить на f_k, чтобы функция $f(x)$ была монотонно возрастающей; монотонно убывающей; возрастала только при $x \leqslant x_4$.

г) Объясните, какие ограничения следует наложить на f_k, чтобы функция $f(x)$ была выпуклой; вогнутой.

д) Объясните, какие ограничения следует наложить на f_k, чтобы разность между расчетными значениями $f(x_i)$ и $f(x_{i+1})$ составляла не более d; $d > 0$; не менее d.

е) Укажите, как следует изменить приведенную выше постановку, если имеется несколько одинаковых x_k.

ж) Обсудите тип и размерность модели, которая получается при минимизации суммы абсолютных значений отклонений, т. е. $\sum_{k=1}^{p} |e_k|$; при минимизации максимального абсолютного значения отклонения.

50. а) Рассмотрите задачу «минимизации суммы абсолютных значений отклонений», определяемую целевой функцией (1) из разд. 14.8. (Предполагается, что на знаки x_j не налагается никаких ограничений.) Составьте двойственную задачу линейного программирования. Покажите, как преобразовать двойственную задачу в задачу, включающую n линейных ограничений и p неотрицательных ограниченных сверху переменных.

б) Рассмотрите задачу «минимизации максимального абсолютного значения отклонения», определяемую целевой функцией (10) из разд. 14.8. (Предполагается, что на знаки x_j не налагается ограничений.) Составьте двойственную задачу линейного программирования.

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

51. а) Рассмотрите функцию $g(c)$, определяемую следующим образом:

$$g(c_1, c_2, \ldots, c_n) \equiv \max \sum_{j=1}^{n} c_j x_j$$

при условиях

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

$$x_j \geqslant 0, \quad j = 1, 2, \ldots, m.$$

Таким образом, $g(c)$ есть оптимальное значение целевой функции задачи линейного программирования, причем коэффициенты этой целевой функции равны $c \equiv (c_1, c_2, \ldots, c_n)$. Определите, является ли $g(c)$ выпуклой или вогнутой функцией.

б) Рассмотрите функцию $h(b)$, определяемую следующим образом:

$$h(b_1, b_2, \ldots, b_m) \equiv \max \sum_{j=1}^{n} c_j x_j,$$

где на x_j налагаются те же ограничения, что и в варианте а). Таким образом, $h(b)$ есть оптимальное значение целевой функции задачи линейного программирования с вектором правых частей ограничений $b \equiv (b_1, b_2, \ldots, b_m)$. Определите, является ли $h(b)$ выпуклой или вогнутой функцией. Означает ли полученный вами результат, что оптимальное значение двойственной переменной i завышает или занижает оценку выгоды, достигаемой благодаря наличию дополнительной единицы дефицитного ресурса, к которому относится ограничение i?

52. Докажите достаточное условие существования максимума вогнутой функции, сформулированное в разд. 14.5.

53. Пусть $c(x_1, x_2, \ldots, x_n)$ есть вогнутая функция. Рассмотрите две точки $w = (w_1, w_2, \ldots, w_n)$ и $y \equiv (y_1, y_2, \ldots, y_n)$. Покажите, что функция $c(t) \equiv c[tw + (1-t)y]$ вогнута по t, где $0 \leqslant t \leqslant 1$. Каким будет соответствующий результат, если функция $c(x_1, x_2, \ldots, x_n)$ выпукла?

54. В разд. 14.4 необходимое условие существования максимума функции $c(x_1, x_2, \ldots, x_n)$ формулируется с использованием первых частных производных. Пусть $c(x)$ обладает непрерывными вторыми частными производными. Покажите, что если, кроме того, $c(\bar{x})$ есть максимальное значение, то квадратичная форма

$$\sum_{j=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 c(\bar{x})}{\partial x_j \, \partial x_k} x_j x_k$$

должна быть отрицательно полуопределенной. [В руководствах по линейной алгебре это условие иногда формулируется следующим образом: «гессиан функции $c(x)$ для точки $\bar{x}$ должен быть отрицательно полуопределенным».]

55. Рассмотрите модель сепарабельного программирования, приведенную в разд. 14.7. Пусть минимизируемая целевая функция включает $c_1(x_1) = x_1^2$, где $0 \leqslant x_1 \leqslant 10$.

а) Объясните, каким образом значение $c_1(x_1)$ может быть аппроксимировано с помощью y, где на y налагается ограничение

$$y \geqslant \max(p_1 x_1 + q_1, \; p_2 x_2 + q_2, \; \ldots, \; p_k x_k + q_k).$$

Воспользуемся рис. 14.10; пусть первому линейному отрезку соответствуют $p_1 = 1$ и $q_1 = 0$. Каковы значения остальных p_i и q_i для остальных отрезков на рис. 14.10? Сравните преимущества такого метода аппроксимации с аппроксимацией на основе (5) — (18).

б) Обсудите, какими должны быть характеристики функции $c(x)$ одной переменной x для того, чтобы был применим аппроксимирующий подход варианта а).

в) Пусть оптимизация понимается в смысле «максимизации». Применим ли подход согласно варианту а)? Объясните почему. Дайте постановку, аналогичную использованной в варианте а), применимую в случае вогнутости функции $c(x)$.

г) Предложите, каким образом можно обобщить подход, использованный в варианте а), на случай выпуклой функции $c(x_1, x_2, \ldots, x_n)$ многих переменных.

56. Рассмотрите функцию $c(x)$ одной переменной x, где $0 \leqslant x \leqslant U$. Предположим, что требуется аппроксимировать $c(x)$ непрерывной кусочно-линейной функцией, состоящей из n отрезков. Примите, что вершины аппроксимирующей ломаной лежат на $c(x)$, т. е. что в каждой вершине аппроксимирующая ломаная имеет ту же ординату, что кривая $c(x)$, и что вершины могут соответствовать только значениям $x = 0, 1, 2, \ldots, U$, где U — целое число, намного большее, чем n.

а) Постройте задачу динамического программирования, имеющую целью найти такую аппроксимацию, при которой минимизируется максимальное абсолютное значение отклонения аппроксимированной величины от $c(x)$ при $x = 0, 1, 2, \ldots, U$.

б) Покажите, как изменить задачу в варианте а), если минимизируется сумма абсолютных значений отклонений аппроксимированных величин от $c(x)$ для $x = 0, 1, 2, \ldots, U$.

57. Рассмотрите задачу дробно-линейного программирования и эквивалентную ей задачу линейного программирования (12) — (19) из разд. 14.9.

а) Предположим, что с помощью симплексного метода найдено оптимальное решение эквивалентной задачи линейного программирования. Пусть x_k — небазисная переменная, которая не входит в оптимальное решение, имея положительное значение. Объясните, как найти интервал значений c_k, в пределах которого решение остается оптимальным; то же для f_k; то же для a_{ik}. [*Указание*: пусть w_i — двойственная переменная, соответствующая уравнению (19), а w_0 — двойственная переменная, соответствующая уравнению (17).]

б) Примите $f_0 > 0$. Покажите, как можно использовать (17) для исключения переменной r, так чтобы эквивалентная задача линейного программирования включала только m ограничений. Покажите, как изменяется ответ, если $f_0 = 0$.

в) Разработайте алгоритм, при котором m ограничений сохраняются в своем первоначальном виде, используется *критерий II* симплексного метода для выбора исключаемой из базиса переменной, операции по преобразованию матрицы выполняются, как обычно, на основе ведущего элемента, однако выбор вводимой переменной осуществляется следующим образом: исключаются частные производные $\partial c(x)/\partial x_j$ для текущего базисного решения по каждой небазисной переменной и в базис вводится та переменная, для которой частная производная положительна и имеет наибольшее значение. Отличается ли этот подход (а если отличается, то чем) от применения симплексного алгоритма для решения эквивалентной задачи линейного программирования (16), 17) и (19)?

ГЛАВА 15

Усовершенствованные методы нелинейного программирования

15.1. КРУПНОШАГОВЫЕ МЕТОДЫ

В предыдущей главе были приведены методы решения некоторых важных частных задач нелинейного программирования. В каждом из этих методов использовалась какая-либо специфическая особенность целевой функции (квадратичной или сепарабельной) или ограничений (линейных или сепарабельных). В настоящей главе изучаются методы решения задач нелинейного программирования более общего вида.

При использовании описываемых ниже методов вид целевой функции $c(x)$ не может быть совершенно произвольным. Она должна обладать хотя бы свойством гладкости, наличие которого предполагалось при рассмотрении случая неограниченной максимизации в разделе 14.4. В первых четырех разделах настоящей главы все ограничения считаются линейными. Затем, начиная с разд. 15.5, рассматриваются модели, которые включают нелинейные ограничения. Эти ограничения также должны удовлетворять некоторым требованиям к гладкости и виду функций.

Для того чтобы избежать недоразумений, вновь опишем оптимизационную модель, которую мы исследовали, а также приведем постулируемые свойства $c(x)$. В разд. 15.2—15.4 рассматривается задача

$$\text{Максимизировать } c(x_1, x_2, \ldots, x_n) \tag{1}$$

при условиях

$$\begin{aligned} &\sum_{j=1}^{n} a_{ij}x_j + s_i = b_i, \quad i = 1, 2, \ldots, m, \\ &x_j \geqslant 0, \quad j = 1, 2, \ldots, n, \\ &s_i \geqslant 0, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{2}$$

Для того чтобы избежать детального обсуждения всех возможных случаев, примем, что ограничения (2) удовлетворяют следующим условиям:

I) Все $b_i \geqslant 0$, следовательно, существует допустимое базисное решение при заданной системе ограничений, например все $x_j = 0$ и все $s_i = b_i$.

II) Ограничения (2) являются достаточно жесткими, так что максимум выражения $c_1x_1 + c_2x_2 + \ldots + c_nx_n$ является конечным на всем множестве допустимых x при *произвольно* выбранных значениях коэффициентов c_j.

Условие I) гарантирует, что модель имеет хотя бы одно допустимое решение, причем нахождение такого решения будет тривиальным. Если отказаться от условия I), алгоритмы необходимо обобщить, включив в них приемы нахождения допустимого решения, коль скоро оно существует.

Условие II) иногда называют **постулатом регуляризации.** При описании приводимых далее алгоритмов оно выполняет две функции. Во-первых, устраняет необходимость рассмотрения неограниченных решений, возникающих на промежуточных этапах алгоритма, когда $c(x)$ аппроксимируется линейной функцией. Во-вторых, в сочетании с условием гладкости $c(x)$ условие II) исключает возможность существования неограниченного оптимального решения исходной задачи (1) — (2). [Одним из способов обеспечить выполнение условия II) является включение в (2) ограничения, согласно которому сумма всех переменных не может превысить некоторую большую, но конечную верхнюю грань.] Если отказаться от условия II), то алгоритм должен быть модифицирован с учетом возможности получения неограниченного решения на промежуточных итерациях.

Предположим, что функция $c(x)$ удовлетворяет также следующим условиям гладкости и вида функции:

III) Функция $c(x)$ однозначна и конечна при любом x, удовлетворяющем ограничениям (2).

IV) Все частные производные $\partial c(x)/\partial x_j$ однозначны, конечны и непрерывны; следовательно, $c(x)$ непрерывна при любом x, удовлетворяющем ограничениям (2).

V) На множестве всех значений x, удовлетворяющих ограничениям (2), функция $c(x)$ имеет конечный максимум $\bar{c}$.

Если сравнить условия III) — V) настоящего раздела с допущениями I) — III) разд. 14.4, то оказывается, что условия III) — V) применимы только к тем x, которые являются допустимыми при налагаемых ограничениях. Условие достижения максимума $c(x)$ при конечных значениях x следует из I) — V). Условия III) — V) можно ослабить, однако это необходимо компенсировать увеличением числа случаев, которые необходимо рассмотреть в процессе выполнения операций алгоритма.

Предварительные замечания. Заметим, что модель (1) — (2) в качестве частного случая включает задачу квадратичного программирования, рассмотренную в разд. 14.6. Поэтому в модели (1) — (2) могут возникнуть оптимальные решения тех же типов, что и в указанной задаче. В частности, оптимальное решение может соответствовать внутренней точке области допустимых решений, граничной, но не экстремальной точке и, наконец, экстремальной точке. Поскольку оптимум не обязательно должен находиться в вершине этой области, оптимальное решение может включать более чем m переменных x_j со строго положительными значениями. Далее, все дополняющие переменные s_i могут принимать положительные значения, что эквивалентно нахождению оптимума во внутренней части области допустимых решений. Читателю следует постараться понять, каким образом каждый из приводимых далее алгоритмов создает возможность включения в оптимальное решение со строго положительными значениями всех n переменных x_j и всех m переменных s_i.

Как и в случае отыскания безусловного максимума, алгоритмы не всегда сходятся к глобальному оптимальному решению; к исключениям относятся случаи, когда функция $c(x)$ имеет особый вид, например относится к вогнутым функциям. При этом обычно не обеспечивается сходимость за конечное число итераций, даже если $c(x)$ стремится к некоторому предельному значению.

Многие алгоритмы настоящей главы удобно излагать, пользуясь такой же формой записи, какая была приведена в разд. 14.4 и 14.5 при рассмотрении алгоритма «скорейшего подъема с оптимальной длиной шага», поэтому, прежде чем перейти к чтению дальнейшего материала, рекомендуется снова прочесть указанные разделы. В общем случае крупношаговый метод включает следующие операции:

Шаг 1. Выбираем произвольную исходную допустимую точку x^0.

Шаг. 2. а) На итерации k определяем направления d_j^k ($j = 1, 2, \ldots, n$). б) Определяем длину шага t^k, позволяющую максимизировать значение $c(x)$ по всем *допустимым* точкам, лежащим в выбранном направлении от точки x^k. Прекращаем расчеты при $t^k = 0$. В противном случае переходим к шагу 3.

Шаг 3. Определяем новое проверяемое решение

$$x_j^{k+1} = x_j^k + t^k d_j^k, \quad j = 1, 2, \ldots, n. \tag{3}$$

Возвращаемся к шагу 2, заменив x^k на x^{k+1}.

Наименование **крупношаговый метод** вызвано тем, что на шаге 2 используется шаг оптимальной длины. Если мы вновь обратимся к алгоритму скорейшего подъема с оптимальной длиной шага, то можно заметить три важных отличия в формулировке операций шага 2. Во-первых, условие равенства нулю всех $\partial c/\partial x_j$ более не используется в качестве критерия прекращения расчетов. Это объясняется тем, что в стационарной точке задачи на *условный* максимум (будем называть ее относительной стационарной точкой) все частные производные не обязательно должны быть равны нулю. Во-вторых, оптимальную длину шага t^k не удается определить только на основе $c(x)$; t^k следует выбирать таким образом, чтобы точка x^{k+1} удовлетворяла ограничениям задачи. При выполнении допущения о регуляризации II), условие допустимости означает, что t^k всегда конечно. Результирующее значение t^k можно назвать **оптимальной допустимой длиной шага.** В-третьих, на шаге 2 не все направления обязательно определяются по значениям $\partial c/\partial x_j$, вычисленным для текущей пробной (проверяемой) точки. Используя пример, легко отыскать причину того, что значения частных производных не являются адекватным отображением направлений.

Рассмотрим задачу квадратичного программирования, ранее исследованную в разд. 14.6.

$$\text{Задача P:} \begin{cases} \text{максимизировать } c(x) \equiv 6x_1 + 4x_2 + 2x_3 - \\ \qquad\qquad - 3x_1^2 - 2x_2^2 - \frac{1}{3}x_3^2 & (4) \\ \text{при условиях} \quad x_1 + 2x_2 + x_3 + s = 4, & (5) \end{cases}$$

где все $x_j \geqslant 0$ и $s \geqslant 0$. Пусть проверяемая точка определяется как $x_j^k = 1$ для всех j, так что (5) удовлетворяется при $s = 0$; значение $c(x^k)$ равно 6,67. Предположим, что, как и при использовании метода скорейшего подъема, мы принимаем каждое $d_j^k = \partial c/\partial x_j$. Это значит, что $d_1^k = d_2^k = 0$ и $d_3^k = \frac{4}{3}$. К сожалению, в этом направлении от x^k не имеется *ни одной допустимой* точки. Вместе с тем проверяемая точка *не* является локальным оптимумом. Так, рассмотрим направление $d_1^k = -1$, $d_2^k = 0$ и $d_3^k = 1$, которое лежит в гиперплоскости (5). Можно убедиться в том, что при длине шага $t = 0{,}1$ значение целевой функции улучшается: $c(x^{k+1}) = 6{,}76$.

Подводя итоги, можно сказать, что при проверке условий прекращения расчетов, а также при выборе длины шага и направления следует учитывать возможные последствия того, что проверяемая точка либо движется к границе области допустимых решений и доходит до нее, либо лежит на этой границе.

В следующих трех разделах обсуждается несколько крупношаговых методов, использующих линейность ограничений. Читателю следует обратить внимание на детали шага 2.

Нетрудно увидеть, что нелинейные алгоритмы содержат множество деталей и для решения даже небольших условных примеров требуется немалый объем вычислительных операций. Для того чтобы облегчить усвоение материала, читателю рекомендуется выписывать важнейшие уравнения и другие данные. Если воспользоваться этими рекомендациями, можно будет устранить потребность в многократном возврате к прочитанному ранее материалу.

Задача P, описываемая (4) и (5), неоднократно используется в настоящей главе для демонстрации эффективности алгоритмов. Соответственно (4) и (5) следует выписать для справок. Единственное оптимальное решение этой задачи имеет вид

$$\bar{x}_1 = \frac{7}{8} = 0{,}875; \quad \bar{x}_2 = \frac{5}{8} = 0{,}625; \quad \bar{x}_3 = 1\,\frac{7}{8} = 1{,}875; \quad \bar{c} = 7{,}25; \qquad (6)$$

выпишите также и это решение.

15.2. МЕТОД ВЫПУКЛЫХ КОМБИНАЦИЙ

Идея настоящего алгоритма заключается в том, чтобы на шаге 2 определять направление с помощью оптимизационной задачи линейного программирования, используя линейную аппроксимацию $c(x)$

и линейные ограничения задачи. Рассмотрим допустимую проверяемую точку x^k. Если выполняется условие IV) из разд. 15.1 о частных производных функции $c(x)$, значение $c(x)$ в точке z, находящейся в окрестности x^k, может быть аппроксимировано разложением в ряд Тейлора с использованием первых производных:

$$c(z) \approx c(x^k) + \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot (z_j - x_j^k). \tag{1}$$

Предположим, что нам удалось найти допустимую точку z, в которой максимизируется выражение

$$\sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot z_j \tag{2}$$

при условии выполнения линейных ограничений, налагаемых на переменные. Пусть эта точка $\bar{z}$ и пусть

$$\sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot \bar{z}_j > \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot x_j^k. \tag{3}$$

Тогда выражение под знаком суммы в правой части (1) строго положительно. По всей вероятности, $\bar{z}$ *не* находится в непосредственной окрестности точки x^k, так что аппроксимация (1) является плохой. В самом деле, $c(\bar{z})$ может быть даже меньшим, чем $c(x^k)$. Однако ввиду непрерывности $c(x)$ и ее частных производных существует некоторая **выпуклая комбинация**, или **средневзвешенная**, x^k и $\bar{z}$, которая позволяет получить увеличение $c(x)$ по сравнению с $c(x^k)$. Иными словами, существует такое t, $0 < t \leqslant 1$, что, приняв

$$x_j^{k+1} = (1 - t)\, x_j^k + t\bar{z}_j, \quad j = 1, 2, \ldots, n, \tag{4}$$

получим

$$c(x^{k+1}) > c(x^k). \tag{5}$$

Поскольку все ограничения линейны, выпуклая комбинация x^k и z является допустимой для любого t, значение которого находится в пределах $0 < t \leqslant 1$.

Определим каждое направление следующим образом:

$$d_j^k \equiv \bar{z}_j - x_j^k, \quad j = 1, 2, \ldots, n; \tag{6}$$

тогда (4) можно записать в стандартном виде

$$x_j^{k+1} = x_j^k + t d_j^k, \quad j = 1, 2, \ldots, n, \quad 0 < t \leqslant 1, \tag{7}$$

что читатель может проверить самостоятельно. [Преобразуйте операции суммирования в (3) так, чтобы представить это неравенство в терминах направлений.]

Итак, вкратце повторим полученные результаты. Шаг 2 можно детально описать следующим образом:

I) Определим значения всех частных производных $\partial c/\partial x_j$ в проверяемой точке x^k.

II) Решим задачу линейного программирования, описываемую целевой функцией (2) и линейными ограничениями, где во всех ограничениях x_j заменяются на z_j. Определим направления d_j^k в соответствии с (6).

III) Если неравенство (3) не выполняется, прекратим расчеты.

IV) В противном случае определим такую длину шага t^k, при которой максимизируется значение $c(x^k + td^k)$ при $0 < t \leqslant 1$.

Если оптимальное решение $\bar{z}$ задачи линейного программирования на итерации k позволяет достичь равенства между двумя суммами в левой и правой частях (3), то текущая проверяемая точка x^k является относительной стационарной точкой. Нередко это означает, что в непосредственной окрестности x^k не существует допустимой точки, в которой $c(x)$ превышает $c(x^k)$. (Однако может случиться, что x^k есть не более чем точка перегиба.) Вместе с тем обычно не существует такой итерации k, на которой левая и правая части (3) равны между собой. Тем не менее для любой произвольной исходной проверяемой точки x^0 метод всегда обеспечивает сходимость к пределу в следующем смысле:

А) Вся последовательность $c(x^k)$ монотонно возрастает, стремясь к некоторому пределу.

Б) По крайней мере подпоследовательность x^k сходится к некоторой допустимой точке $\bar{x}$.

В) Точка $\bar{x}$ является относительной стационарной точкой; это значит, что в $\bar{x}$ либо *все* частные производные равны нулю, либо некоторые частные производные не равны нулю, однако в допустимой окрестности $\bar{x}$ не представляется возможным улучшить значение $c(x)$.

Рассмотрим важный частный случай вогнутости $c(x)$. Напомним определение: функция многих переменных $c(x)$ является **вогнутой**, если для любых двух точек x и y, $x \neq y$, и для любого p, $0 \leqslant p \leqslant 1$, выполняется условие

$$pc(x) + (1-p)\,c(y) \leqslant c\,[px_1 + (1-p)\,y_1, \ldots$$
$$\ldots, px_n + (1-p)\,y_n];$$

функция $c(x)$ является **строго вогнутой**, если для $0 < p < 1$ неравенство является строгим.

Если функция $c(x)$ вогнута, точка $\bar{x}$ всегда представляет собой глобальный оптимум, а $c(\bar{x})$ действительно является максимальным значением $c(x)$. Более того, в этом случае можно найти оптимальную длину шага t^k, воспользовавшись методом поиска в одномерном пространстве, который описан в разд. 14.3.

Следует отметить некоторые важные особенности метода выпуклых комбинаций.

I) Только случайно каждое из направлений (6) может оказаться равным соответствующей частной производной функции $c(x)$ в точке x^k. В общем случае не обязательно, чтобы каждое d_j^k было равно направлению скорейшего подъема в проверяемой точке.

II) Поскольку $\bar{z}$ — оптимальное решение задачи линейного программирования, имеющей m ограничений, в этом решении содержится не более чем m строго положительных переменных. Тем не менее если $x_j^k \neq \bar{z}_j$, то x_j^{k+1} будет в соответствии с (4) отличаться от x_j^k. Иными словами, при переходе к следующей итерации могут измениться значения *всех* x_j.

III) Если система линейных ограничений обладает специальной структурой, например если она образует сетевую модель, то на шаге 2 при нахождении оптимального решения задачи линейного программирования это обстоятельство можно использовать.

IV) Если целевая функция характеризуется лишь незначительной нелинейностью, так что большинство переменных x_j входит в целевую функцию линейно, с коэффициентами c_j, то в (2) коэффициенты при этих переменных также равны c_j. Это упрощает вычисления на шаге 2.

V) Если $c(x)$ вогнута, то на каждой итерации можно вычислить верхнюю грань оптимального значения $c(x)$. Это свойство особенно полезно, поскольку верхнюю грань можно использовать для определения момента прекращения итераций в связи с достижением достаточно хорошего решения. Расчет верхней грани основан на том обстоятельстве, что при разложении $c(x)$ в ряд Тейлора с использованием первых производных, т. е. при разложении, аналогичном I), значение $c(x)$ завышается, если эта функция вогнута. Пусть $\bar{x}$ — оптимальное решение; тогда

$$c(\bar{x}) \leqslant c(x^k) + \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot (\bar{x}_j - x_j^k) \leqslant c(x^k) + \\ + \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot (\bar{z}_j - x_j^k), \qquad (8)$$

где первое неравенство следует из условия вогнутости, а второе — из факта оптимизации при целевой функции (2). Таким образом, верхняя грань представляет собой выражение в правой части (8). Поскольку эта грань не обязательно уменьшается на каждой итерации, следует запоминать наименьшее значение верхней грани, полученное на предыдущих итерациях.

Сходимость алгоритма можно ускорить с помощью разложения в ряд Тейлора с использованием вторых производных, что на шаге 2 приводит к задаче не линейного, а квадратичного программирования.

Соответствующим образом изменяется и неравенство (3): в него дополнительно включаются слагаемые, связанные со вторыми частными производными.

Пример. Для того чтобы продемонстрировать эффективность данного метода, рассмотрим задачу P, приведенную в конце предыдущего раздела. Соответствующая этой задаче линеаризованная целевая функция (2) имеет вид

$$[6(1-x_1^k)]\,z_1+[4(1-x_2^k)]\,z_2+\left[2\left(1-\frac{1}{3}x_3^k\right)\right]z_3. \qquad (9)$$

При наличии единственного линейного ограничения задача линейного программирования имеет только следующие базисные допустимые решения:

$$\begin{aligned}
&\text{точка 1:}\quad z_1=4,\quad z_2=0,\quad z_3=0;\\
&\text{точка 2:}\quad z_1=0,\quad z_2=2,\quad z_3=0;\\
&\text{точка 3:}\quad z_1=0,\quad z_2=0,\quad z_3=4;\\
&\text{точка 4:}\quad z_1=0,\quad z_2=0,\quad z_3=0.
\end{aligned} \qquad (10)$$

Чтобы читатель мог проверить, насколько хорошо он усвоил изложенный выше материал, ему рекомендуется разработать простое правило, позволяющее определить, какая из этих точек оптимальна, если значения x_j^k в (9) известны.

Формула определения длины шага (7) из разд. 14.5 также может быть здесь использована, для чего требуется лишь незначительная ее модификация:

$$\text{Оптимальное } t=\min\left[\frac{6(1-x_1)d_1+4(1-x_2)d_2+2\left(1-\frac{1}{3}x_3\right)d_3}{6(d_1)^2+4(d_2)^2+\frac{2}{3}(d_3)^2},\ 1\right], \qquad (11)$$

где надстрочный индекс k для краткости опущен.

Напомним, что единственное оптимальное решение имеет вид

$$\bar{x}_1=\frac{7}{8}=0{,}875;\quad \bar{x}_2=\frac{5}{8}=0{,}625;\quad \bar{x}_3=1\frac{7}{8}=1{,}875;\quad c(\bar{x})=7{,}25. \qquad (12)$$

Можно проверить, что если в качестве проверяемой x^k принять $\bar{x}$, то целевая функция (9) принимает вид $\frac{6}{8}z_1+\frac{12}{8}z_2+\frac{6}{8}z_3$. В результате каждая из первых трех точек в (10) оказывается оптимальной для соответствующей задачи линейного программирования и две величины, представляющие собой левую и правую части контрольного неравенства (3), равны между собой; это значит, что $\bar{x}$ является оптимумом.

Предположим, что в качестве исходной точки для этого примера при использовании метода выпуклых комбинаций принято решение, в котором все $x_j^0 = 0$. Проверьте, что при линеаризованной целевой функции (9) точка 1 в (10) является оптимальной. В результате получаем, что $d_1^1 = 1 - 0 = 1$ и $d_2^1 = d_3^1 = 0$; согласно (11), соответствующая оптимальная длина шага $t^1 = 0{,}25$. Тогда на основании формулы (7) для следующей проверяемой точки $x_1^1 = 0 + \frac{1}{4} \cdot 4 = 1$ $x_2^1 = x_3^1 = 0$. Результаты дальнейших итераций приведены в таблице рис. 15.1.

$$c(x) \equiv 6x_1 + 4x_2 + 2x_3 - 3x_1^2 - 2x_2^2 - \frac{1}{3} x_3^2,$$

$$x_1 + 2x_2 + x_3 \leqslant 4$$

Итерация k	$c(x^k)$	x_1^k	x_2^k	x_3^k	Длина шага t^k	Верхняя грань	Экстремальная точка
0	0	0	0	0	0,25	24	1
1	3	1	0	0	0,48	11	3
2	4,92	0,52	0	1,92	0,12	13,56	1
3	5,417	0,920	0	1,699	0,26	11,50	2
4	6,222	0,677	0,529	1,250	0,06	10,21	1
5	6,338	0,871	0,498	1,177	0,16	8,09	3
6	6,480	0,730	0,418	1,633	0,04	9,32	1
10	6,717	0,768	0,484	1,670	0,03	8,73	1
11	6,748	0,868	0,469	1,618	0,07	7,82	2
18	6,907	0,806	0,532	1,725	0,02	8,16	1
19	6,919	0,868	0,521	1,691	0,04	7,59	2
20	6,934	0,830	0,587	1,617	—	—	—
Оптимальные значения	7,25	0,875	0,625	1,875			

Рис. 15.1. Метод выпуклых комбинаций.

Заметим, что на последующих итерациях значения $c(x)$ вначале быстро возрастают, но после пятой итерации рост сильно замедляется. Такое поведение типично для метода, использующего только первые производные. Следует также отметить, что на каждой четной итерации оптимальным решением задачи линейного программирования является точка 1 в (10), тогда как на нечетных итерациях оптимальны попеременно точки 2 и 3. Это чередование обусловливает затухающие колебания значений x_j^k. Наконец, отметим, что проверяемые x^k находятся во внутренней части области допустимых решений, определяемой линейным неравенством, хотя оптимальное решение соответствует превращению этого неравенства в равенство.

Для сравнения рассмотрим результаты применения алгоритма в том случае, если x^0 — одна из остальных экстремальных точек в (10) или же точка, для которой все $x_j^0 = 1$ (эта точка также в точности соответствует превращению линейного ограничения в равенство). Скорость сходимости $c(x^k)$ характеризуется данными, приведенными в таблице рис. 15.2. Заметьте, насколько быстрее $c(x^k)$

Итерация k	Точка 4 $x_j^0=1$	Точка 2	Точки 1 и 3	$x_j^0=1$
1	3,000	4,571	6,000	7,167
2	4,920	7,020	7,143	7,219
3	5,417	7,187	7,228	7,236
4	6,222	7,224	7,244	7,244
5	6,338	7,239	7,248	7,247
10	6,717	7,249	7,249	7,249
20	6,934	7,249	7,249	7,249

Оптимальное значение $c(x) = 7{,}25$

Рис. 15.2. Скорость сходимости значения $c(x)$.

приближается к максимуму при движении из этих исходных точек. После десятой итерации не только $c(x^k)$ почти равна оптимальному значению, но и соответствующие x_j^k отличаются от оптимальных $\bar{x}_j$ менее чем на 0,03.

Сравним результаты в другом случае. Ослабим ограничение, приняв правую часть равной 6 или 8. В этих двух случаях ограничение не является существенным — в том смысле, что *безусловное* оптимальное решение ($x_1 = 1$, $x_2 = 1$, $x_3 = 3$) допустимо; значение целевой функции для него составляет 8. Если правая часть ограничения равна 6 и мы начинаем со всех $x_j^0 = 0$, проверяемое решение для $k = 20$ имеет вид: $x_1^{20} = 0{,}9162$, $x_2^{20} = 0{,}8307$, $x_3^{20} = 2{,}3585$, $c(x^{20}) = 7{,}7844$. И здесь сходимость после пятой итерации чрезвычайно замедляется. При правой части, равной 8, результаты после итерации 20 намного лучше: $x_1^{20} = 0{,}9851$, $x_2^{20} = 0{,}9485$, $x_3^{20} = 2{,}7668$, что дает $c(x^{20}) = 7{,}9759$.

15.3. СИМПЛЕКСНЫЙ МЕТОД ВОГНУТОГО ПРОГРАММИРОВАНИЯ

Как видно из самого названия, настоящий алгоритм позволяет получить оптимальное решение, если функция $c(x)$ вогнута; он основан на обычном симплексном методе, в который внесены некоторые незначительные изменения. По существу, здесь сочетаются идеи симплексного метода с двусторонними ограничениями на переменные (разд. 5.10) и квадратичного симплексного метода (разд. 14.6). Как в методе с двусторонними ограничениями, в настоящем методе используется система только лишь из m ограничений. Исходные «структурные» переменные делятся на базисные и небазисные, причем некоторые небазисные переменные могут все же входить в решение с положительными значениями. Как в квадратичном симплексном методе, рассматривается результат замены на каждой итерации только одной переменной. Если значение выбранной небазисной переменной может быть изменено так, что значение целевой функции увеличивается, то это изменение осуществляется с соблюдением условий сохранения допустимости, причем с обеспечением максимального прироста значения целевой функции. Наиболее важным отличием этого подхода от изложенного выше метода выпуклых комбинаций является то, что при настоящем методе на каждой итерации меняют значения не более $(m+1)$ переменных, а именно переменные текущего базиса и выбранная небазисная переменная. Если $c(x)$ не вогнута, метод позволяет получить относительную стационарную точку.

Пусть x^k — проверяемая допустимая точка на итерации k. Для простоты изложения предположим, что переменные пронумерованы таким образом, что в базис на итерации k входят $x_1, x_2, \ldots, x_m$, а $x_{m+1}, \ldots, x_N$ являются небазисными. Обратите внимание на то, что вводится обозначение $N \equiv m+n$, так что дополняющие переменные обозначаются здесь тем же символом x. Пусть на итерации k система линейных ограничений представлена в таком же виде, как при применении симплексного метода:

$$\begin{array}{llllll}
x_1 & & + t_{1,m+1}x_{m+1} + \ldots + t_{1N}x_N & = \bar{b}_1 & \text{(строка 1)}, & \\
& x_2 & + t_{2,m+1}x_{m+1} + \ldots + t_{2N}x_N & = \bar{b}_2 & \text{(строка 2)} & (1)\\
& & \vdots \qquad\qquad \vdots \qquad\quad \vdots & \vdots & & \\
& & x_m + t_{m,m+1}x_{m+1} + \ldots + t_{mN}x_N & = \bar{b}_m & \text{(строка } m), &
\end{array}$$

где для простоты записи опущен надстрочный индекс k. Поскольку при данном алгоритме некоторые из небазисных переменных могут иметь строго положительные значения, $\bar{b}_i$ может быть отрицательным. Соответствующее условие неотрицательности для точки x^k можно записать следующим образом:

$$x_i = \bar{b}_i - \sum_{j=m+1}^{N} t_{ij}x_j \geqslant 0, \quad i = 1, 2, \ldots, m. \tag{2}$$

Соображения, позволяющие обосновать применимость целевой функции в изложенном выше методе выпуклых комбинаций, остаются в силе и здесь. Проиллюстрируем все это на конкретном примере. Рассмотрим функцию

$$\sum_{j=1}^{N} \frac{\partial c(x^k)}{\partial x_j} \cdot x_j \tag{3}$$

или, представляя ее в том же виде, как в симплексном методе,

$$x_0 - \frac{\partial c(x^k)}{\partial x_1}\cdot x_1 - \ldots - \frac{\partial c(x^k)}{\partial x_m}\cdot x_m -$$
$$- \frac{\partial c(x^k)}{\partial x_{m+1}}\cdot x_{m+1} - \ldots - \frac{\partial c(x^k)}{\partial x_N}\cdot x_N = 0 \quad \text{(строка 0)}. \qquad (4)$$

Если бы здесь применялся симплексный метод, то коэффициенты строки 0 для любой переменной, входящей в текущий базис, были бы равны нулю. Для того чтобы и в данном методе обеспечить выполнение аналогичного условия, исключим $x_1, \ldots, x_m$ из (4) на основе (1). Говоря конкретно, в случае каждого $i = 1, 2, \ldots, m$ умножим строку i на $\partial c(x^k)/\partial x_i$ и прибавим полученное произведение к (4). В результате этого получим

$$x_0 + 0x_1 + \ldots + 0x_m + t_{0,\, m+1}x_{m+1} + \ldots$$
$$\ldots + t_{0N}x_N = \bar{b}_0 \quad \text{(строка 0)}, \qquad (5)$$

где благодаря использованию процесса исключения

$$t_{0j} \equiv \sum_{i=1}^{m} \frac{\partial c(x^k)}{\partial x_i}\cdot t_{ij} - \frac{\partial c(x^k)}{\partial x_j}, \quad j = m+1, \ldots, N. \qquad (6)$$

[Если бы $c(x)$ была линейной функцией, то каждая ее частная производная по базисной переменной x_i, проставленная под знаком суммы в (6), была бы равна коэффициенту c_i при этой переменной в исходной целевой функции.] Значение t_{0j} именуется **приведенной (или относительной) частной производной.** Как упоминалось выше, надстрочный индекс k в переменных t_{0j} опущен для простоты записи. Поскольку константа $\bar{b}_0$ в правой части (5) не играет роли при дальнейших расчетах, условно обозначим ее символом $\bar{0}$.

Как и в симплексном методе, значение t_{0j} может быть истолковано следующим образом.

Интерпретация коэффициентов строки 0. Каждый коэффициент при небазисной переменной отображает *скорость* возрастания (для отрицательных коэффициентов) или убывания (для положительных коэффициентов) значения функции $c(x)$ при увеличении соответствующей небазисной переменной по сравнению с ее текущим уровнем x_j^k.

Здесь следует сделать несколько замечаний по поводу содержательного и численного значений этого истолкования.

Пусть на итерации k все $t_{0j} \geqslant 0$, т. е. все t_{0j} —«положительные коэффициенты». Является ли решение оптимальным? Да, является, если все небазисные переменные имеют нулевые значения, потому что в этом случае небольшое возрастание любой небазисной переменной x_j приведет к уменьшению $c(x)$ по сравнению с $c(x^k)$. Соответственно x^k — относительная стационарная точка, а следовательно, и оптимальная точка, так как $c(x)$ вогнута. Однако если существует небазисная переменная x_j, имеющая положительное значение, и если $t_{0j}x_j^k > 0$, то, возможно, решение удастся улучшить, *уменьшив* значение x_j^k. Если удовлетворяются оба условия

$$\text{все } t_{0j} \geqslant 0 \quad \textit{и} \quad \text{все } t_{0j}x_j^k = 0, \tag{7}$$

то x^k является истинным оптимумом.

Предположим, что условия (7) полностью не выполнены: имеется либо отрицательный коэффициент t_{0j}, либо положительное произведение $t_{0j}x_j^k$, либо то и другое одновременно. Тогда необходимо выбрать небазисную переменную и изменить ее значение, стремясь улучшить текущее значение целевой функции. Для этого, во-первых, определим, как и при использовании *критерия I* симплексного метода, наименьший из коэффициентов t_{0j}:

$$\min(t_{01}, \ldots, t_{0N}) = t_{0p} \equiv t_{\min}. \tag{8}$$

Очевидно, что $t_{\min} \leqslant 0$ и если условие (7) не удовлетворяется, то $t_{\min}$ может иметь и отрицательное значение. Во-вторых, определим наибольшее значение $t_{0j}x_j^k$:

$$\max(t_{01}, x_1^k, \ldots, t_{0N}x_N^k) = t_{0q}x_q^k \equiv t_{\max}. \tag{9}$$

Если (7) не удовлетворяется, $t_{\max} \geqslant 0$ и может иметь положительное значение. При условии, что (7) не выполняется, можно считать, что либо $t_{\min}$, либо $t_{\max}$, либо та и другая величины одновременно являются ненулевыми. Можно показать, что справедливо следующее правило выбора изменяемой небазисной переменной.

Правило выбора изменяемой небазисной переменной. Если условие (7) не выполняется при $t_{\min} < -t_{\max}$, то следует увеличить значение переменной x_p; в противном случае следует уменьшить значение переменной x_q.

После выбора изменяемой небазисной переменной необходимо вычислить соответствующие направления d_j. Все $d_j = 0$, *за исключением* тех j, которые соответствуют базисным переменным и выбранной небазисной переменной. Сначала рассмотрим случай выбора x_p. Здесь можно установить следующее правило на основе простой модификации *критерия II* симплексного метода.

Правило выбора направлений при увеличении x_p. а) Вычислите отношения x_i, определенные по формуле (2), к коэффициентам t_{ip} (не рассматривайте при этом те отношения, в которых знаменатель является нулевым или отрицательным). б) Обозначьте через r значение наименьшего из этих отношений. в) Примите $d_p = r$ и $d_i = -t_{ip}r$ $(i = 1, 2, \ldots, m)$.

Теперь перейдем к случаю выбора x_q. Здесь правило строится на основе простой модификации *критерия II* для симплексного метода с двусторонними ограничениями на переменные.

Правило выбора направлений при уменьшении x_q. а) Вычислите отношения $-x_i$, определенные по формуле (2), к коэффициентам t_{ip} (не рассматривайте при этом отношения, в которых знаменатель является нулевым или положительным). б) Обозначьте R значение наименьшего из этих отношений и $r = \min(R, x_q^2)$. в) Примите $d_q = -r$ и $d_i = t_{ip}r$ $(i = 1, 2, \ldots, m)$.

Отметим, что в обоих случаях $r \geqslant 0$. (Если $r = 0$, то *все* $d_j = 0$, однако расчеты прекращать все же не следует.)

Наконец, после того как определены направления, необходимо вычислить оптимальную длину шага t, тем самым завершая расчеты по шагу 2. Операции этого шага остаются в точности такими же, как в методе выпуклых комбинаций.

На шаге 3 новая проверяемая точка x^{k+1} вычисляется как обычно. Если оптимальная длина шага t^k лежит в интервале $0 < t^k < 1$, возвратимся к шагу 2 и повторим процесс, пересмотрев значения t_{0j} в (6) для новой проверяемой точки x^{k+1}. (Это всегда возможно, если переменная x_q уменьшается, а $r < R$.) Однако в случае оптимальной длины шага $t^k = 1$ (что эквивалентно $r = 0$); прежде чем возвратиться к шагу 2, необходимо выполнить смену базиса, а именно ввести в базис переменную x_p или x_q (ту из них, которая определяется правилом выбора изменяемой небазисной переменной) и исключить из базиса ту переменную, для которой достигается значение отношения, соответствующее r.

Теперь приведем общее описание произвольно выбранной итерации k алгоритма. Текущая проверяемая точка x^k определяется набором базисных переменных и набором небазисных переменных, причем некоторые из последних могут иметь положительные значения. Итерация включает следующие операции:

I) Вычисление приведенных частных производных по формуле (6) и выполнение проверки на оптимальность согласно (7). Если возможно улучшение, следует выбрать изменяемую небазисную переменную.

II) Определение направления изменения текущих базисных переменных и выбранной небазисной переменной; эти направления вычисляются таким образом, чтобы при оптимальной длине шага $t^k = 1$ решение оставалось допустимым.

III) Производимый после определения направлений расчет длины шага t^k, максимизирующей значение целевой функции при движении в выбранном направлении от точки x^k, т. е. максимизация $c(x^k + td)$ при $0 < t \leqslant 1$.

IV) Выполняемое соответствующим образом смещение проверяемой точки, которое производится тогда, когда оптимальная длина шага $t^k = 1$; для завершения итерации k производится смена базиса.

При использовании настоящего алгоритма гарантируется $c(x^{k+1}) \geqslant c(x^k)$. Однако для доказательства того, что $c(x^k)$ в пределе стремится к максимальному значению $c(x)$, если только $c(x)$ вогнута, необходимо принять дополнительные меры предосторожности против зацикливания. (Это явление здесь не рассматривается, однако если читатель захочет, он может повторно изучить его применительно к симплексному методу, обратившись к разд. 4.7.)

Настоящий алгоритм обладает теми же двумя полезными свойствами, что и алгоритм выпуклых комбинаций. Во-первых, при замене базиса можно использовать специальную структуру системы линейных ограничений, если она имеется; во-вторых, вычисления целевой функции (6) упрощаются, если нелинейность $c(x)$ незначительна. Повторим также замечание, сделанное в начале раздела: от метода выпуклых комбинаций настоящий алгоритм принципиально отличается тем, что в методе выпуклых комбинаций все направления могут быть ненулевыми, а в симплексном методе вогнутого программирования ненулевыми могут оказаться не более $m+1$ направлений.

Пример. Для демонстрации того, как реализуются операции алгоритма, снова рассмотрим задачу P из разд. 15.1. Поскольку мы хотим привести несколько итераций алгоритма, откажемся от индексации базисных переменных от 1 до m, принятой в (1) настоящего раздела. Линеаризованная целевая функция (4) принимает в этом случае вид

$$x_0-[6(1-x_1^k)]\,x_1-[4(1-x_2^k)]\,x_2-$$
$$-\left[2\left(1-\frac{1}{3}x_3^k\right)\right]x_3=\overline{0}\quad\text{(строка 0).}\tag{10}$$

В качестве исходного решения примем все $x_j^0=0$ и $s=4$, что приводит к

$$x_0-6x_1-4x_2-2x_3=\overline{0}\quad\text{(строка 0).}\tag{11}$$

Начнем с увеличения переменной x_1. Определим направления $d_1=4$, $d_2=d_3=0$ и $d_4=-4$, где d_4 — направление для дополняющей переменной. Здесь применима формула определения длины шага (11) из разд. 15.2, согласно которой $t^0=0{,}25$. Следовательно, очередная проверяемая точка имеет вид ($x_1^1=1$, $x_2^1=x_3^1=0$, $s^1=3$); для нее $c(x^1)=3$. Поскольку $t^0\leqslant 1$, смены базиса не требуется.

На итерации $k=1$ целевая функция (10) преобразуется в следующую:

$$x_0-4x_2-2x_3=\overline{0}\quad\text{(строка 0).}\tag{12}$$

На этот раз будем увеличивать переменную x_2. Поскольку

$$s_1=4-x_1^1-2x_2^1-x_3^1=3,\tag{13}$$

можно убедиться в том, что $d_1=0$, $d_2=\frac{3}{2}$, $d_3=0$ и $d_4=-3$, оптимальная длина шага t^1 равна $\frac{2}{3}$. Это дает точку ($x_1^2=1$, $x_2^2=1$, $x_3^2=0$, $s^2=1$), для которой $c(x^2)=5$.

При $k = 2$ целевая функция (10) принимает вид

$$x_0 - 2x_3 = \overline{0} \quad \text{(строка 0)}, \tag{14}$$

так что для увеличения выбирается переменная x_3. Поскольку

$$s^2 = 4 - 1 \cdot 1 - 2 \cdot 1 - 1 \cdot 0 = 1, \tag{15}$$

можно убедиться в том, что $d_1 = 0$, $d_2 = 0$, $d_3 = 1$ и $d_4 = -1$. Теперь оптимальная длина шага $t^2 = 1$, так что новая проверяемая точка имеет вид ($x_1^3 = 1$, $x_2^3 = 1$, $x_3^3 = 1$, $s^3 = 0$); для этой точки $c(x^3) = 6{,}67$. Здесь необходима смена базиса; введем в базис x_3 и исключим s. Используя линейное ограничение для исключения x_3 из (10), получим следующую строку 0:

$$\begin{aligned} x_0 &+ \left[1 \cdot 2\left(1 - \frac{1}{3}x_3^k\right) - 6\left(1 - x_1^k\right)\right]x_1 + \\ &+ \left[2 \cdot 2\left(1 - \frac{1}{3}x_3^k\right) - 4\left(1 - x_2^k\right)\right]x_2 + \\ &+ \left[1 \cdot 2\left(1 - \frac{1}{3}x_3^k\right)\right]s = \overline{0} \quad \text{(строка 0)}. \end{aligned} \tag{16}$$

Заметим, что в (16) выгодность изменения значений x_1, x_2 или s зависит от частной производной по x_3 и от коэффициентов при всех x_j и s в линейном ограничении. Убедимся в том, что для проверяемой точки x^3 (16) имеет вид

$$x_0 + \frac{4}{3}x_1 + \frac{8}{3}x_2 + \frac{4}{3}s = \overline{0} \quad \text{(строка 0)}. \tag{17}$$

Все коэффициенты $t_{0j} \geqslant 0$; однако $t_{01}x_1^3 = \frac{4}{3}$ и $t_{02}x_2^3 = \frac{8}{3}$, так что прекращать расчеты не следует. Пусть для уменьшения выбрана

$$c(x) \equiv 6x_1 + 4x_2 + 2x_3 - 3x_1^2 - 2x_2^2 - \frac{1}{3}x_3^2, \quad x_1 + 2x_2 + x_3 + s = 4$$

Итерация k	$c(x^k)$	x_1^k	x_2^k	x_3^k	s^k	Длина шага t^k
1	3	1	0	0	3	0,2500
2	5	1	1	0	1	0,6667
3	6,667	1	1	1	0	0,4000
4	7,2	1	0,6	1,8	0	0,12
5	7,248	0,88	0,6	1,92	0	0,025
6	7,2499	0,88	0,624	1,872	0	0,005
7	7,2499	0,8752	0,624	1,8768	0	0,0010
8	7,2499	0,8752	0,6250	1,8745	0	0,0002
Оптимальные значения	7,25	0,875	0,625	1,875	0	

Рис. 15.3. Симплексный метод вогнутого программирования.

переменная x_2. Можно убедиться в том, что $d_1 = 0$, $d_2 = -1$, $d_3 = 2$, $d_4 = 0$ и оптимальная длина шага $t = 0{,}4$. Это дает точку ($x_1^4 = 1$; $x_2^4 = 0{,}6$; $x_3^4 = 1{,}8$; $s^4 = 0$), значение $c(x^4)$ для которой равно 7,2. Результаты следующих итераций приведены в таблице рис. 15.3. В качестве базисной переменной остается x_3. При $k = 5$ значение x_1 уменьшается; при $k = 6$ снова увеличивается значение x_2. После этого изменения значений становятся пренебрежимо малы. Очевидно, в этом условном примере применяемый метод обеспечивает быструю сходимость как к оптимальному значению $c(x)$, так и к оптимальным значениям каждого из x_j.

15.4. ДРУГИЕ ПОДХОДЫ

Имеются десятки методов оптимизации нелинейной целевой функции при условии выполнения линейных ограничений. Во второй части настоящей главы приведено несколько таких методов. В данном разделе вкратце описываются два метода, играющие важную роль в литературе по нелинейному программированию. Подобные методы можно видоизменить так, что их можно будет использовать для решения задач с нелинейными ограничениями, однако необходимые для этого изменения в книге не излагаются.

Один из методов именуется **методом секущих плоскостей**; в нем используется линейная верхняя грань (8) из разд. 15.2. Метод применим в случае вогнутой функции $c(x)$. На любой итерации k имеется текущая проверяемая допустимая точка x^k, а также система *линейных* ограничений, включающая ограничения исходной задачи и некоторые «отсекающие» ограничения, которые описываются ниже. Задача дополняется переменной x_0, входящей во все «отсекающие» ограничения. Переменную x_0 необходимо максимизировать, так что на итерации k задача решается с помощью алгоритма линейного программирования. Получаемое оптимальное значение x_0 используется в качестве очередной верхней грани оптимального значения $c(x)$.

Для нахождения x^{k+1} к имеющейся системе ограничений добавляется еще одно «отсекающее» линейное ограничение

$$x_0 - \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot x_j \leqslant c(x^k) - \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot x_j^k. \qquad (1)$$

Этому ограничению *не* удовлетворяют x_0^k и x^k (если они не являются оптимальными), поскольку x_0^k всегда больше, чем $c(x^k)$. Соответственно благодаря ограничению (1) текущая проверяемая точка отсекается от рассматриваемой допустимой области.

Отсюда следует, что $x_0^{k+1} \leqslant x_0^k$. Далее, можно доказать, что последовательность x_0^k сходится к оптимальному значению $c(x)$ и что имеется подпоследовательность x^k, сходящаяся к оптимальному решению. Однако соответствующая последовательность $c(x^k)$ *не* *обязательно* должна возрастать монотонно. Следовательно, прекращая итерации еще до достижения сходимости, необходимо выбрать ту из проверяемых точек, которая обеспечивала наибольшее из полученных до сих пор значений $c(x)$; этой точкой может и не быть x^k последней итерации. Результат применения алгоритма для решения задачи P приведен в таблице рис. 15.4. Здесь следует обратить

$$c(x) \equiv 6x_1+4x_2+2x_3-3x_1^2-2x_2^2-\frac{1}{3}x_3^2,$$

$$x_1+2x_2+x_3 \leqslant 4$$

Итерация k	$c(x^k)$	x_0^k	x_1^k	x_2^k	x_3^k
0	0	—	0	0	0
1	−24	24	4	0	0
2	2	16	2	1	0
3	5,908	12,667	1,167	0	2,833
4	6,066	11,022	0,756	1,233	0,778
5	5,685	6,887	0,222	0,511	2,756
8	6,849	7,530	1,437	0,855	0,853
9	6,869	7,390	1,036	0,892	1,180
10	6,702	7,373	1,281	0,464	1,791
15	7,249	7,264	0,868	0,639	1,853
Оптимальные значения	7,25	7,25	0,875	0,625	1,875

Рис. 15.4. Метод секущих плоскостей.

особое внимание на «блуждание» x^k при последовательных итерациях.

Еще один из алгоритмов называется **методом проецируемых градиентов.** В отличие от других описанных выше алгоритмов здесь на шаге 2 не используются методы линейного программирования. Метод проецируемых градиентов ближе к методу скорейшего подъема: в обоих этих методах сходны способы определения направления на основе частных производных $c(x)$ в проверяемой точке x^k, а также длины шага, при которой $c(x)$ максимизируется по всем *допустимым* точкам, лежащим в выбранном направлении от точки x^k. Компоненты направления вычисляются путем проектирования градиента функции $c(x)$ на ту часть границы области, которая определяется ограничениями, реализуемыми в виде равенств в точке x^k (дополняющие переменные этих ограничений принимают нулевое значение). [На самом деле вычислительный процесс несколько более сложен, однако при настоящем изложении его детали не важны.] При вычислении длины шага сначала отыскивается наибольшее значение t, напри-

$$c(x) \equiv 6x_1 + 4x_2 + 2x_2 - 3x_1^2 - 2x_2^2 - \frac{1}{3}x_3^2,$$

$$x_1 + 2x_2 + x_3 \leqslant 4$$

Итерация k	$c(x^k)$	x_1^k	x_2^k	x_3^k	Длина шага t^k
0	0	0	0	0	0,1981
1	5,547	1,189	0,793	0,396	0,3667
2	6,537	0,774	1,097	1,033	0,2749
3	6,966	1,060	0,817	1,306	0,3075
4	7,137	0,835	0,813	1,539	0,2749
5	7,205	0,949	0,701	1,648	0,3075
6	7,232	0,859	0,700	1,741	0,2749
10	7,249	0,872	0,637	1,854	0,2749
15	7,249	0,876	0,626	1,873	0,3075
Оптимальные значения	7,25	0,875	0,625	1,875	—

Рис. 15.5. Метод проецируемых градиентов.

мер T, при котором точка $(x^k + td)$ остается допустимой, а затем в одномерном пространстве осуществляется поиск максимума функции $c(x^k + td)$ на сегменте $[0, 1]$, причем используется метод, изложенный в разд. 14.3.

Результаты вычислений приводятся для задачи Р. Пусть проверяемая точка x^k лежит *на* плоскости, соответствующей линейному уравнению относительно x_1, x_2 и x_3 (т. е. дополняющая переменная s равна нулю). Рассмотрим *произвольную* систему направлений D_1, D_2, D_3. Метод позволяет преобразовать эти направления в d_1, d_2, d_3 так, что получаемая точка x^{k+1} остается на этой же плоскости. Для выполнения этого условия система d_j должна удовлетворять ограничению

$$1d_1 + 2d_2 + 1d_3 = 0. \tag{2}$$

Объясните, почему должно удовлетворяться это ограничение. Соответствующее преобразование, или **проецирование**, D_j на d_j отображается, как можно показать, системой

$$\begin{aligned} d_1 &= \frac{5}{6}D_1 - \frac{2}{6}D_2 - \frac{1}{6}D_3, \\ d_2 &= -\frac{2}{6}D_1 + \frac{2}{6}D_2 - \frac{2}{6}D_3, \\ d_3 &= -\frac{1}{6}D_1 - \frac{2}{6}D_2 + \frac{5}{6}D_3. \end{aligned} \tag{3}$$

Можно убедиться в том, что при подстановке правых частей (3) в левую часть (2) результат равен нулю, как это и должно быть. В методе проецируемых градиентов для получения значений D_j используются частные производные $c(x)$ в точке x^k. Говоря нестрого, при использовании данного метода осуществляется скорейший подъем по тем линиям уровня $c(x)$, которые лежат на плоскостях ограничений, являющихся существенными в точке x^k.

Результаты ряда итераций процесса решения задачи Р приведены в таблице рис. 15.5. На второй итерации проверяемая точка x^k уже лежит на плоскости ограничения. Хороший выбор исходной точки, лежащей на плоскости ограничения, может ускорить сходимость. Так, например, если в качестве исходной точки используется любая из первых трех экстремальных точек (10) из разд. 15.2 или же принимается, что все $x_j^0 = 1$, то значение $c(x^5)$ отличается от оптимального значения 7,25 не более чем на 0,04.

15.5. ОПТИМИЗАЦИЯ ПРИ НЕЛИНЕЙНЫХ ОГРАНИЧЕНИЯХ

Начиная с настоящего раздела, мы будем изучать методы решения оптимизационных задач, включающих нелинейные ограничения. Для определенности положим, что модель формулируется следующим образом:

$$\text{Максимизировать } c(x_1, x_2, \ldots, x_n) \tag{1}$$

при условиях

$$a_i(x_1, x_2, \ldots, x_n) \leqslant 0, \quad i = 1, 2, \ldots, m, \tag{2}$$
$$x_j \geqslant 0, \quad j = 1, 2, \ldots, n.$$

Любое из ограничений может быть линейным, так что с формальной точки зрения задача (1) — (2) включает в качестве частных случаев задачи линейного программирования модели гл. 14 и предыдущих разделов настоящей главы. Условие неотрицательности *всех* переменных вводится только для удобства изложения; на самом деле все приводимые далее алгоритмы чрезвычайно просто модифицировать с таким расчетом, чтобы некоторые x_j могли иметь любой знак. Многие методы построения задач, описанные в разд. 3.2, относятся к нелинейным функциям, так что (1) — (2) можно рассматривать как каноническую форму задачи нелинейного программирования. Читателю рекомендуется переписать задачу (1) — (2) так, чтобы к ней было удобно обращаться при изучении остальной части главы.

Поскольку наша цель заключается в разъяснении принципиальных идей, лежащих в основе различных методов решения, сделаем некоторые упрощающие допущения относительно вида $c(x)$ и $a_i(x)$. Однако большинство методов применимо и при менее сильных допущениях. Сначала рассмотрим допущения о функциях ограничений $a_i(x)$, а затем — о целевой функции $c(x)$.

Область допустимых решений. Допущения о каждой из нелинейных функций $a_i(x)$ относятся к виду и гладкости этой функции. Прежде чем перейти к непосредственному изложению, дадим определение выпуклой функции. Действительнозначная функция $a(x)$ называется выпуклой, если для любых двух точек x и y, $x \neq y$, и для любого p, $0 \leqslant p \leqslant 1$, выполняется условие

$$pa(x_1, \ldots, x_n) + (1-p)\, a(y_1, \ldots, y_n) \geqslant$$
$$\geqslant a[px_1 + (1-p)\, y_1, \ldots, px_n + (1-p)\, y_n] \quad \text{(выпуклая);} \quad (3)$$

функция именуется строго выпуклой, если для $0 < p < 1$ неравенство (3) является строгим. [Отметим, что если $-a(x)$ вогнута, то $a(x)$ выпукла.]

С этим определением связано другое свойство выпуклой функции: для любых двух точек x и y

$$a(y) \geqslant a(x) + \sum_{j=1}^{n} \frac{\partial a(x)}{\partial x_j}(y_j - x_j) \quad \textbf{(выпуклая).} \quad (4)$$

Выражению (4) можно дать следующую геометрическую интерпретацию. Пусть x — некоторая заданная точка, такая, что $a(x) = 0$. Тогда, приравняв нулю выражение под знаком суммирования в правой части (4), получим гиперплоскость, касательную к поверхности уровня $a(y) = 0$. В более общем виде можно сказать, что правая часть (4) представляет собой **линейную опору** функции в в заданной точке x, и неравенство (4) можно интерпретировать, считая, что оно определяет функцию, лежащую целиком над любой из своих линейных опор. [Пусть $a(x) \equiv x_1^2 - x_2$; в качестве координат выбранной точки примем $x_1 = 2$ и $x_2 = 4$. Начертим линию уровня $a(y) = y_1^2 - y_2 = 0$. Прямая, которую получим, приравняв нулю выражение под знаком суммирования в правой части (4), имеет вид $4(y_1 - 2) - (y_2 - 4) = 0$; это и есть линейная опора в данной точке.]

В оставшейся части настоящей главы будем считать (если не оговаривается обратное), что $a_i(x)$ в (2) удовлетворяют следующим допущениям о виде и гладкости:

I) Каждая из $a_i(x)$ однозначно определена, конечна и выпукла при *всех* значениях $(x_1, x_2, \ldots, x_n)$.

II) Каждая частная производная $\partial a_i(x)/\partial x_j$ непрерывна при всех x, удовлетворяющих ограничениям (2).

[Отметим, что хотя из условия I) следует непрерывность функции $a_i(x)$, необходимо более сильное условие II) о непрерывности первых частных производных.]

Дополнительно постулируем следующее условие.

Характеристика системы ограничений. Существует по крайней мере одна точка $x^0 \geqslant 0$, такая, что все $a_i(x^0) < 0$.

Говоря нестрого, можно сказать следующее: характеристика системы ограничений утверждает, что существует такое допустимое решение, которое лежит строго во внутренней части области, ограниченной (2). (Как читатель увидит позже, такая точка необходима для того, чтобы начать расчеты, используя несколько изложенных ниже алгоритмов.) Полезно выписать условия I) — II) и характеристику системы ограничений, так как мы неоднократно к ним возвращаемся в течение настоящей главы.

Важным следствием I) является то, что все множество точек, удовлетворяющих ограничениям (2), образует так называемое **выпуклое множество.** Это значит, что если две любые точки x и y, $x \neq y$, являются допустимыми с учетом ограничений (2), то допустимой является также и точка, представляющая собой их выпуклую, или средневзвешенную, комбинацию $[px + (1 - p)\, y]$, где $0 \leqslant p \leqslant 1$. Это следует из (3): поскольку $a_i(x) \leqslant 0$ и $a_i(y) \leqslant 0$ для каждого i, то $a_i[px_1 + (1 - p)\, y_1, \ldots, px_n + (1 - p)\, y_n] \leqslant 0$, где $0 \leqslant p \leqslant 1$. Можно показать, что такие же соображения применимы и к ограничениям неотрицательности. Характеристика системы ограничений обеспечивает наличие в выпуклом множестве допустимых решений полномерной внутренней части, включающей бесконечное множество точек.

Не столь очевидны два других следствия условий I) и II) и характеристики системы ограничений. Во-первых, если точка z допустима с учетом системы ограничений (2), причем для некоторого i $a_i(z) = 0$, то градиент функции $a_i(x)$ в точке z не может быть нулевым; это значит, что существует по крайней мере одно j, такое, что $\partial a_i(z)/\partial x_j \neq 0$. Во-вторых, модель, по существу, не содержит ограничений, представляющих собой уравнения, если только эти ограничения не являются линейными; в последнем случае каждое такое ограничение можно включить в (2) в виде пары неравенств, отказавшись для этой пары от принятой выше характеристики системы ограничений.

Во многих ситуациях выпуклость $a_i(x)$ удается установить, выяснив, что эту функцию можно описать следующим образом:

А) Как сумму выпуклых функций, например $3x^2 - 6x$.

Б) Как отрицательную вогнутую функцию, например $-(4y - 2y^2)$.

В) Как монотонно возрастающую выпуклую функцию выпуклой функции, например a^{x^2} при $a > 0$, поскольку $a(z) \equiv a^z$ является монотонно возрастающей и выпуклой при $a > 0$, а $z \equiv x^2$ выпукла.

Г) Как монотонно убывающую выпуклую функцию вогнутой функции, например a^{y^2} при $a > 0$, поскольку $a(z) \equiv a^{-z}$ является монотонно убывающей и выпуклой при $a > 0$, а $z = -y^2$ вогнута.

Д) Как выпуклую функцию линейной функции, например $a^{(x-y)^2}$ при $a > 0$, поскольку $a(z) \equiv a^{z^2}$ является выпуклой при $a > 0$, а $z \equiv x - y$ линейна.

Следовательно, в соответствии с А) функция $3x^2 - 6x - 4y + 2y^2 + a^{x^2} + a^{y^2} + a^{(x-y)^2} + 10$ является выпуклой при $a > 0$.

Кроме того, если $a(x)$ есть функция одной переменной x и имеет непрерывную вторую производную, то $a(x)$ выпукла тогда и только тогда, когда $d^2a(x)/dx^2 \geqslant 0$ при всех x. (Многомерным аналогом этого неравенства является утверждение, что матрица Гессе вторых частных производных должна быть положительно полуопределенной.)

Целевая функция. Предполагается, что функция $c(x)$ также должна удовлетворять некоторым допущениям о виде и гладкости. Будем считать по-прежнему справедливыми допущения III) — V) из разд. 15.1 и добавим еще одно условие. Объединим все эти допущения.

III) Функция $c(x)$ однозначна и конечна при любом x, удовлетворяющем ограничениям (2).

IV) Все частные производные $\partial c(x)/\partial x_j$ однозначны, конечны и непрерывны при любом значении x, удовлетворяющем ограничениям (2).

V) На множестве всех значений x, удовлетворяющих ограничениям (2), функция $c(x)$ имеет конечный максимум $\bar{c}$.

VI) Функция $c(x)$ вогнута при всех значениях x, удовлетворяющих ограничениям (2).

[Допущение VI) также следует выписать для дальнейших справок при чтении оставшейся части главы.]

Допущения I) — VI) и характеристика системы ограничений в совокупности гарантируют, что:

А) существует по крайней мере одно допустимое решение $\bar{x}$, такое, что $c(\bar{x}) = \bar{c}$;

Б) если $c(x)$ строго вогнута, то существует единственное оптимальное решение $\bar{x}$;

В) если x — стационарная точка в условиях ограничений, то x является глобальным оптимумом.

Эти три следствия справедливы и при несколько более слабых допущениях о $a_i(x)$ и $c(x)$. Однако для следствия В) существенно необходима выпуклость области допустимых решений, поскольку в противном случае локальный оптимум не обязательно одновременно является глобальным, даже если $c(x)$ удовлетворяет условиям III) — VI).

Новые трудности. Алгоритмы решения задач с нелинейными ограничениями имеют много общего с алгоритмами для задач, где ограничения линейны. В частности, каждая итерация начинается с рассмотрения допустимой проверяемой точки x^k и проверки ее оптимальности. Если возможно дальнейшее улучшение значения функции $c(x)$, то определяется направление движения и отыскивается новая проверяемая точка с большим значением $c(x)$. Однако при определении направлений в случае нелинейных ограничений возникают новые трудности. (Прежде чем продолжить чтение, может оказаться полезным более глубокое рассмотрение того, каким образом в методах, излагаемых в разд. 15.2—15.4, линейность ограничений используется для определения *допустимых* направлений улучшения.)

Пусть проверяемая точка находится на границе существенно нелинейного ограничения, как это показано на рис. 15.6. Из рисунка очевидно, что проверяемую точку целесообразно двигать *вдоль* границы области допустимых решений. Однако из-за нелинейности границы такой подход может оказаться чрезвычайно трудным с точки

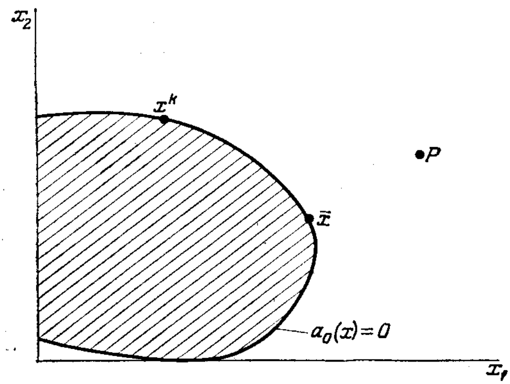


Рис. 15.6. Задача нахождения допустимых направлений: найти допустимую точку x, расположенную как можно ближе к P.

зрения его вычислительной реализации. Существует несколько альтернативных способов нахождения допустимых направлений движения, которые описываются в последующих разделах. Различия между этими способами можно в общем виде охарактеризовать следующим образом.

Один из способов состоит в том, что если проверяемая точка находится на нелинейной границе, то рассматриваются только те направления, которые ведут во внутреннюю часть области допустимых решений. Из рис. 15.6 видно, что, если только проверяемая точка не является истинным оптимумом, направление движения, обеспечивающего улучшение значения целевой функции, *всегда* существует. Этот подход используется в разд. 15.6. Другой способ — не допустить выхода проверяемой точки на границу области; этот подход рассматривается в разд. 15.9. Третий способ заключается в использовании сетки допустимых точек и в аппроксимации границы области допустимых решений посредством ее линеаризации в этих точках. Этому методу посвящен разд. 15.10. При всех указанных подходах значения $c(x)$ на последовательных итерациях монотонно сходятся к оптимальному значению $\bar{c}$. По крайней мере подпоследовательность проверяемых точек сходится к оптимальному решению, а если оно является единственным, вся последовательность сходится к этой точке.

Применение нескольких алгоритмов демонстрируется на решении задачи P из разд. 15.1, а также задачи Q:

$$\text{Задача Q}\begin{cases}\text{Максимизировать } c(x) \equiv -x_1 - 2x_2 - x_3 \\ \text{при условиях } -6x_1 - 4x_2 - 2x_3 + 3x_1^2 + 2x_2^2 + \\ \qquad\qquad + \frac{1}{3}x_3^2 + 7{,}25 \leqslant 0, \\ \text{все } x_j = 0,\end{cases}$$

имеющей то же единственное оптимальное решение, что и задача P, при $\bar{c} = -4$. Задачу Q также следует выписать для последующих справок.

Повторим предупреждение, сделанное в гл. 14; при сравнении разных алгоритмов, основанном на приводимых далее численных результатах, следует быть чрезвычайно осторожным в выводах, поскольку задачи P и Q весьма специфичны — квадратичные формы в них не содержат попарных произведений различных переменных. Более того, из этих результатов не видно, что́ происходит при использовании различных приемов, ускоряющих сходимость. Однако эти примеры еще раз подтверждают справедливость сделанного ранее замечания о том, что нелинейные задачи, как правило, труднее решать, чем линейные.

15.6. МЕТОД ДОПУСТИМЫХ НАПРАВЛЕНИЙ

Настоящий алгоритм является обобщением метода выпуклых комбинаций, изложенного в разд. 15.2. Основной особенностью метода допустимых направлений является способ построения задачи линейного программирования, позволяющей установить на шаге 2 направления улучшения. Как и метод выпуклых комбинаций, данный алгоритм сходится к относительной стационарной точке, если отказаться от допущения VI) о вогнутости функции $c(x)$.

В конце предыдущего раздела было показано, что принципиальные трудности в отыскании **допустимых направлений** возникают в тех случаях, когда проверяемая точка x^k находится на нелинейной границе области допустимых решений. В случае используемого ниже подхода эти трудности преодолеваются путем линейной аппроксимации нелинейных ограничений в точке x^k с последующим выбором направлений, идущих вовнутрь области, т. е. от тех нелинейных ограничений, которые в точке x^k являются существенными. Ради удобства объяснения сформулируем идеи, опираясь на такие величины, как направления d_j, а для простоты записи опустим у них надстрочный индекс k, обозначающий, что рассматриваемые значения d_j получены на итерации k.

При выполнении допущений о непрерывности $c(x)$ текущее значение этой функции улучшится при движении из точки x^k

в направлениях d_j, если только

$$\sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot d_j > 0. \tag{1}$$

Следовательно, (1) является критерием проверки, показывающей, действительно ли некоторый конкретный набор направлений может обусловить увеличение значения целевой функции.

Удобно также отказаться от введения в явном виде различий между $a_j(x)$ и условиями неотрицательности; в связи с этим мы представим всю систему ограничений следующим образом:

$$A_i(x) \leqslant 0, \quad i = 1, 2, \ldots, M, \tag{2}$$

где $M = m + n$.

Используем линейную аппроксимацию первого порядка функции $A_i(x^k + d)$:

$$A_i(x_1^k + d_1,\ x_2^k + d_2, \ldots, x_n^k + d_n) \approx$$
$$\approx A_i(x^k) + \sum_{j=1}^{n} \frac{\partial A_i(x^k)}{\partial x_j} \cdot d_j. \tag{3}$$

Напомним сделанное в предыдущем разделе утверждение о том, что допущения относительно $a_i(x)$ обеспечивают наличие под знаком суммирования хотя бы одной ненулевой частной производной.

Предположим, что точка x^k находится «с внутренней стороны» границы, определяемой ограничением i, так что $A_i(x^k) < 0$. Тогда, при всех $d_j = 0$, правая часть (3) строго отрицательна. Можно проверить, что в этом случае точка $(x + td)$ будет удовлетворять ограничению i при *любом* наборе направлений d, если только t, $t > 0$, достаточно мало.

Предположим, далее, что x^k находится на границе, определяемой ограничением i, так что $A_i(x^k) = 0$. Тогда существует такое значение t, $t > 0$, что точка $(x^k + td)$ удовлетворяет ограничению i только в том случае, если выбранные направления d_j обусловливают *строгую* отрицательность правой части (3). Это объясняется тем, что, приравняв нулю правую часть (3), мы получаем направления, лежащие на линейной опоре или касательной гиперплоскости ограничения $A_i(x)$ в точке x^k. Следовательно, если при некоторых направлениях d_j правая часть (3) принимает нулевое или положительное значение, то эти направления лежат на линейной опоре или над ней и новая проверяемая точка $(x^k + td)$ не будет удовлетворять ограничению i при любом $t > 0$; объясните почему.

Подведем итоги. Направление d_j ведут к новой *допустимой* проверяемой точке $(x^k + td)$, где значение $c(x)$ возрастает, если (1) строго положительно, а правая часть (3) строго отрицательна. Если таких направлений не существует, то x^k — относительная стационарная точка.

Для нахождения допустимых направлений, позволяющих улучшить значение $c(x)$, если такие направления существуют в точке x^k, можно использовать модель линейного программирования. В этой модели переменными являются $x_0, d_1, \ldots, d_n$, а сама она имеет следующий вид:

$$\text{Максимизировать } x_0 \tag{4}$$

при условиях

$$x_0 \leqslant \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot d_j, \text{ или, что эквивалентно,}$$
$$x_0 - \sum_{j=1}^{n} \frac{\partial c(x^k)}{\partial x_j} \cdot d_j \leqslant 0, \tag{5}$$

$$A_i(x^k) + \sum_{j=1}^{n} \frac{\partial A_i(x^k)}{\partial x_j} d_j \leqslant -x_0, \text{ или, что эквивалентно,} \tag{6}$$
$$x_0 + \sum_{j=1}^{n} \frac{\partial A_i(x^k)}{\partial x_j} \cdot d_j \leqslant A_i(x^k), \qquad i=1, 2, \ldots, M.$$

На знаки x_0 и всех d_j не налагается никаких ограничений. (7)

Объясните, почему задачу (4) — (6) можно использовать для нахождения направлений, удовлетворяющих критерию проверки (1) и придающих правой части (3) строго отрицательные значения, если x^k не есть относительная стационарная точка. (Обратите внимание на то, что оптимальное значение x_0 никогда не может быть отрицательным, поскольку одним из допустимых решений является равенство нулю всех d_j и x_0.)

С вычислительной точки зрения задача (4) — (7) не совсем полна. Оптимизация может привести к неограниченному решению. Для того чтобы обойти эту трудность, удобно добавить линейные ограничения регуляризации. Ограничения неотрицательности переменных, входящие в $A_i(x)$, обеспечивают с учетом условий (6), что при $x_0 \geqslant 0$ значения d_j имеют нижнюю грань, равную $-x_j^k$; следовательно, дополнительные ограничения регуляризации должны лишь гарантировать невозможность больших d_j. Здесь применимы многие приемы, причем наиболее простой из них имеет вид

$$d_j \leqslant b \quad \text{для всех } j, \tag{8}$$

где b — произвольно выбранная положительная константа (например, 1). Используя (8) и произведя замену переменных ($z_j \equiv d_j + x_j^k$), можно преобразовать задачу (5) — (7) в модель линейного программирования, имеющую всего лишь $1 + m$ ограничений и n неотрицательных, ограниченных сверху переменных. Здесь одно ограничение соответствует (5), а m ограничений — тем соотношениям $a_i(x)$, которые входят в (6); каждое из z_j ($j = 1, 2, \ldots, n$) удовлетворяет условию $0 \leqslant z_j \leqslant b + x_j^k$. Следовательно, для решения преобразованной задачи применим симплексный алгоритм с двусторонними ограничениями (разд. 5.10).

Шаг 2 включает следующие операции:

I) Вычислим все значения частных производных $\partial c/\partial x_j$ в проверяемой точке x^k.

II) Решим задачу линейного программирования (4) — (8).

III) Прекратим расчеты, если оптимальное значение переменной x_0 равно нулю.

IV) В противном случае найдем такую длину шага t, при которой $c(x^k + td)$ максимизируется по всем *допустимым* точкам, лежащим в направлении d от точки x^k.

Если на операции III) оптимальное значение $x_0 = 0$, то x^k — относительная стационарная точка. На операции IV) оптимальная допустимая длина шага всегда конечна, если выполняется условие, что на множестве всех значений x, удовлетворяющих ограничениям, функция $c(x)$ имеет конечный максимум $\bar{c}$.

На самом же деле нет необходимости определять оптимальное значение x_0 на *каждой* итерации. Если мы ведем поиск оптимума регулярно, например на каждой десятой итерации, на всех остальных итерациях требуется лишь находить допустимое решение, при котором значение x_0 положительно.

Если функция $c(x)$ линейна, значение верхней грани b в (8) должно быть принято большим, поскольку (5) в точности отражает влияние приращений на $c(x)$. Если ограничение $A_i(x)$ линейно, то в соответствующее ограничение (6) можно не включать x_0. (Такое упрощение применимо и к условиям неотрицательности $-x_j \leqslant 0$.) Подобная модификация позволяет выбирать направления таким образом, что они будут лежать на гранях тех линейных ограничений, которые для текущей проверяемой точки являются существенными. В предельном случае, когда все $A_i(x)$ линейны, так что величина x_0 не включена ни в одно из ограничений, за исключением (5), задача приобретает почти такой же вид, как при методе выпуклых комбинаций, изложенном в разд. 15.2. Принципиальным отличием рассматриваемого подхода является лишь необходимость в операции регуляризации. Если выполняется условие разд. 15.2, согласно которому исходные линейные ограничения удовлетворяют предположению о регуляризации, то нет необходимости налагать дополнительные ограничивающие условия. Однако здесь необходима дополнительная регуляризация, поскольку цель задачи — определение направлений. Можно также разработать процесс получения верхней грани оптимального значения $c(x)$, аналогично тому как это было сделано в (8) из разд. 15.2 применительно к методу выпуклых комбинаций.

При данном подходе используются только первые частные производные функции $c(x)$; его можно модифицировать с учетом использования и вторых производных. Тогда для вычисления направлений применим алгоритм квадратичного программирования.

Пример. Используем описываемый метод для решения задачи Q. Ограничение (5) имеет вид

$$x_0 - d_1 - 2d_2 - d_3 \leqslant 0, \tag{9}$$

а ограничения (6)

$$x_0 + [-6(1-x_1^k)]\,d_1 + [-4(1-x_2^k)]\,d_2 + \left[-2\left(1-\frac{1}{3}x_3^k\right)\right]d_3 \leqslant 6x_1^k + 4x_2^k + 2x_3^k - 3(x_1^k)^2 - 2(x_2^k)^2 - \frac{1}{3}(x_3^k)^2 - 7{,}25, \tag{10}$$

$$x_0 - d_j \leqslant x_j^k, \qquad j = 1, 2, 3. \tag{11}$$

Получаемая последовательность проверяемых точек характеризуется данными, приведенными в таблице рис. 15.7. Задача решалась

Итерация k	$c(x^k)$	x_1^k	x_2^k	x_3^k	d_1^k	d_2^k	d_3^k	Длина шага t^k
0	−4	1	1	3	−1	−1	−1	0,37
1	−4,51	0,627	0,627	2,627	1	−0,67	−1	0,29
2	−4,17	0,918	0,460	2,336	−1	0,71	−1	0,14
3	−4,09	0,781	0,557	2,199	1	−0,21	−1	0,12
4	−4,05	0,900	0,532	2,080	−1	−0,83	−1	0,07
5	−4,02	0,826	0,594	2,006	1	0,10	−1	0,06
9	−4,00	0,869	0,625	1,881	1	0	−1	0,01
10	−4,00	0,875	0,625	1,875	—	—	—	—
Оптимальные значения	−4	0,875	0,625	1,875				

Рис. 15.7. Метод допустимых направлений — решение задачи Q.

с использованием в (8) величины $b = 1$. Дальнейшее упрощение процесса было достигнуто благодаря тому, что вместо (11) налагались ограничения регуляризации $-d_j \leqslant 1$. При $k \geqslant 1$ проверяемая точка лежит на границе, соответствующей квадратичному ограничению. Решение при $k = 10$ согласуется с оптимальным решением с точностью до четырех десятичных знаков.

15.7. ТЕОРЕТИЧЕСКИЕ СВОЙСТВА ОПТИМАЛЬНОГО РЕШЕНИЯ

Прежде чем перейти к дальнейшему рассмотрению алгоритмов решения задачи нелинейного программирования, сформулированной в начале разд. 15.5, полезно проанализировать некоторые теоретические основы. В настоящем разделе приводится несколько тесно связанных между собой формулировок условий, которые как необходимы, так и достаточны для того, чтобы точка была оптимальным решением. Такие результаты полезны по крайней мере в трех отношениях, поскольку эти условия:

1) позволяют разработать численные методы проверки оптимальности предлагаемого решения;

2) дают информацию о чувствительности решения;

3) приводят к алгоритмическим методам нахождения оптимального решения.

Каждое из этих направлений использования демонстрируется ниже. Таким образом, цель настоящего обсуждения — позволить читателю глубже понять сущность как оптимальных решений, так и оптимизационных методов.

В предыдущем разделе было отмечено, что существование направлений улучшения в проверяемой точке x^k можно установить, решая следующую задачу линейного программирования:

$$\text{Максимизировать } x_0 \tag{1}$$

при условиях
$$x_0 - \sum_{j=1}^{m} \frac{\partial c(x^k)}{\partial x_j} \cdot d_j \leqslant 0, \tag{2}$$

$$x_0 + \sum_{j=1}^{n} \frac{\partial A_i(x^k)}{\partial x_j} \cdot d_j \leqslant -A_i(x^k), \quad i = 1, 2, \ldots, M, \tag{3}$$

$$\text{где} \qquad M = m + n.$$

На знаки x_0 и всех d_j ограничений не налагается. (4)

Для достижения поставленных нами здесь целей нет необходимости дополнительно вводить условия регуляризации для каждого d_j, поэтому такие ограничения, как (8) из предыдущего раздела, в модель не включаются. При условии, что для $c(x)$ и $a_i(x)$ справедливы допущения I) — VI) из разд. 15.5, точка x^k является истинным оптимумом в том и только том случае, когда оптимальное значение x_0 в задаче (1) — (4) равно нулю.

Теперь предположим, что $\bar{x}$ — оптимальная точка, и пусть в (2) и (3) $x^k = \bar{x}$, так что $x_0 = 0$ является оптимумом для задачи (1) — (4). Рассмотрим соответствующую двойственную задачу линейного программирования (разд. 5.4):

$$\text{Минимизировать } -\sum_{i=1}^{M} A_i(\bar{x})\, y_i \tag{5}$$

при условиях
$$\sum_{i=0}^{M} y_i = 1, \tag{6}$$

$$-\frac{\partial c(\bar{x})}{\partial x_j} \cdot y_0 + \sum_{i=1}^{M} \frac{\partial A_i(\bar{x})}{\partial x_j} \cdot y_i = 0, \quad j = 1, 2, \ldots, n, \tag{7}$$

$$y_i \geqslant 0, \qquad i = 1, 2, \ldots, M. \tag{8}$$

Теорема двойственности обеспечивает существование оптимальных значений $\bar{y}_i$, а также равенство нулю оптимального значения выражения (5), при условии что $\bar{x}$ — оптимальная точка. Следовательно, после перестановки и упрощения членов (7) можно сделать следующее утверждение.

Условия оптимальности. Если для $a_i(x)$ и $c(x)$ справедлива характеристика системы ограничений и допущения I) — VI), приведенные в разд. 15.5, точка $\bar{x}$ является оптимальной в том и только том случае, когда

$$A_i(\bar{x}) \leqslant 0, \quad i = 1, 2, \ldots, M, \tag{9}$$

где $M = m + n$, и существуют такие $\bar{u}_i \geqslant 0$ $(i = 1, 2, \ldots, M)$, что

$$A_i(\bar{x}) \cdot \bar{u}_i = 0, \quad i = 1, 2, \ldots, M, \tag{10}$$

$$\frac{\partial c(\bar{x})}{\partial x_j} - \sum_{i=1}^{M} \frac{\partial A_i(\bar{x})}{\partial x_j} \cdot \bar{u}_i = 0, \quad j = 1, 2, \ldots, n, \tag{11}$$

где $\bar{u}_i = \bar{y}_i/\bar{y}_0$ и $\bar{y}_i$ удовлетворяют условиям (5) — (8). Ограничения (10) выражают условия дополнительной нежесткости.

Соотношения (10) — (11) часто называют **условиями Куна — Таккера**, по именам математиков, доказавших их справедливость.

Неравенства (9) включают условия неотрицательности каждой из переменных, следовательно, эти неравенства попросту означают, что точка $\bar{x}$ должна быть допустимой. Легко показать, что ограничения (10) можно получить делением выражения (5) на $\bar{y}_0 > 0$, замечая при этом, что результирующая сумма в точке оптимума по-прежнему равна нулю; последняя означает, что нулю равен каждый член суммы в отдельности, поскольку все $A_i(\bar{x}) \leqslant 0$ и $\bar{u}_i \geqslant 0$. Наконец, нетрудно проверить, что ограничения (11) можно получить, разделив (7) на $-\bar{y}_0$. Следовательно, условия оптимальности можно вывести исходя из того, что задача линейного программирования (1) — (4) применима для проверки оптимальности точки $\bar{x}$. Если примем $A_i(x) \equiv a_i(x)$ для $i = 1, 2, \ldots, m$, а $A_{m+j}(x) \equiv -x_j$ для $j = 1, 2, \ldots, n$, то можно показать, что из (10) и (11) следует

$$a_i(\bar{x}) \cdot \bar{u}_i = 0, \quad i = 1, 2, \ldots, m, \tag{12}$$

$$\left[\frac{\partial c(\bar{x})}{\partial x_j} - \sum_{i=1}^{m} \frac{\partial a_i(\bar{x})}{\partial x_j} \cdot \bar{u}_i\right] \cdot \bar{x}_j = 0, \quad j = 1, 2, \ldots, n, \tag{13}$$

$$\frac{\partial c(\bar{x})}{\partial x_j} - \sum_{i=1}^{m} \frac{\partial a_i(\bar{x})}{\partial x_j} \cdot \bar{u}_i \leqslant 0, \quad j = 1, 2, \ldots, n. \tag{14}$$

Заметим, что в (12) при $a_i(\bar{x}) < 0$ равно нулю $\bar{u}_i$; при $\bar{x}_j > 0$ в (13) равно 0 соответствующее выражение в (14). [Для проверки того, насколько хорошо вы поняли этот материал, покажите, что если $c(x)$ и все $a_i(x)$ линейны, то (9), (12), (13) и (14) сводятся к результатам, содержащимся в теореме о дополнительной нежесткости, которая приведена в разд. 5.3.]

Условия (9), (10) и (11) остаются *достаточными* для оптимальности решения $\bar{x}$ и при несколько более слабых допущениях относительно вида $c(x)$ и $a_i(x)$. Говоря нестрого, функции $A_i(x)$ должны быть такими, чтобы определяемая ими область допустимых решений являлась выпуклой; при одновременном учете вида $c(x)$ это позволяет заключить, что локальный оптимум одновременно является и глобальным оптимумом.

Вместо того чтобы исключать y_0, мы можем сформулировать условия оптимальности (10) — (11) следующим образом: существуют $\bar{y}_i \geqslant 0$ $(i = 0, 1, \ldots, M)$, причем *не все* $\bar{y}_i$ равны нулю, так что

$$A_i(\bar{x}) \cdot \bar{y} = 0, \quad i = 1, 2, \ldots, M, \tag{I}$$

$$\frac{\partial c(\bar{x})}{\partial x_j} \cdot \bar{y}_0 - \sum_{i=1}^{M} \frac{\partial A_i(\bar{x})}{\partial x_j} \cdot \bar{y}_i = 0, \quad j = 1, 2, \ldots, n. \tag{II}$$

[Соотношения (I) и (II) иногда называют **условиями Джона.**] То обстоятельство, что не все $\bar{y}_i$ равны нулю, является следствием из (6). Эта формулировка очень важна, поскольку при очень общих допущениях условия (9), (I) и (II) *необходимы* для того, чтобы оптимальному решению соответствовала точка $\bar{x}$. Единственное требуемое допущение заключается в том, чтобы все функции $a_i(x)$ и $c(x)$ имели непрерывные первые частные производные для всех значений $(x_1, x_2, \ldots, x_n)$; при выполнении этого они могут иметь любой вид. Чтобы при столь слабых ограничениях условия (9), (10) и (11) оставались также и *необходимыми*, следует ввести какое-то дополнительное допущение (как, например, характеристику системы ограничений, приведенную в разд. 15.5), которое обеспечило бы выполнение неравенства $\bar{y}_0 \geqslant 0$.

Условия оптимальности полезны во всех трех отношениях, указанных в начале настоящего раздела. Ниже будет показано, как их можно использовать для проверки оптимальности или нахождения оптимального решения задачи небольшой размерности. Далее демонстрируется важность этих условий для проверки устойчивости и для разработки алгоритмов решения задач большой размерности.

Рассмотрим задачу Р из разд. 15.1; пусть ограничения представлены в следующем виде:

$$\begin{aligned} A_i(x) &\equiv -x_i \leqslant 0, \qquad i = 1, 2, 3, \\ A_4(x) &\equiv x_1 + 2x_2 + x_3 - 4 \leqslant 0. \end{aligned} \tag{15}$$

После упрощений (10) представляются в виде

$$x_1u_1 = 0, \quad x_2u_2 = 0, \quad x_3u_3 = 0, \qquad (16)$$
$$(x_1 + 2x_2 + x_3 - 4)\,u_4 = 0,$$

а (11), также после упрощений,

$$6\,(1 - x_1) = -u_1 + u_4,$$
$$4\,(1 - x_2) = -u_2 + 2u_4, \qquad (17)$$
$$2\left(1 - \frac{1}{3}x_3\right) = -u_3 + u_4.$$

Пусть мы предположили, что в оптимальном решении все $x_j > 0$ и, следовательно, исходя из (16) $u_1 = u_2 = u_3 = 0$. Тогда можно выразить каждую переменную x_j через u_4 в соответствии с (17):

$$x_1 = 1 - \frac{u_4}{6}, \quad x_2 = 1 - \frac{u_4}{2}, \quad x_3 = 3 - \frac{3u_4}{2}. \qquad (18)$$

Подставив соотношения (18) в последнее ограничение (16), после упрощения получим

$$(3 - 4u_4)\,u_4 = 0, \qquad (19)$$

так что

$$u_4 = \frac{3}{4} \quad \text{и} \quad x_1 = \frac{7}{8}, \quad x_2 = \frac{5}{8}, \quad x_3 = \frac{15}{8}, \quad c(x) = 7{,}25. \qquad (20)$$

Следовательно, точка x в (20) оптимальна, поскольку она является допустимой при учете ограничений (15) и поскольку имеется соответствующий набор $u_i \geqslant 0$, удовлетворяющий в точке x условиям (16) и (17). Чтобы проверить, насколько хорошо вы поняли это, покажите, что из предположений $x_1 > 0$ и $x_2 = x_3 = 0$ следует, что не существует $u_i \geqslant 0$, удовлетворяющих условиям (16) и (17).

Двойственная задача линейного программирования (5) — (8) является источником информации о чувствительности решения. Однако вместо использования этого источника мы рассмотрим в следующих разделах важную альтернативную формулировку условий оптимальности, непосредственно позволяющую осуществить требуемый анализ. Более того, данный подход удобен для объяснения алгоритмических следствий изложенных выше условий оптимальности; эти вычислительные аспекты более детально рассматриваются далее в настоящей главе.

Функция Лагранжа и ее седловая точка. Необходимые и достаточные условия оптимальности проверяемого решения $\bar{x}$ могут быть также сформулированы с помощью так называемой **функции Лагранжа.** Пусть

$$L(x, u) \equiv c(x) - \sum_{i=1}^{M} A_i(x) \cdot n_i. \qquad (21)$$

Говорят, что функция Лагранжа имеет седловую точку, соответствующую паре векторов $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ и $(\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_M)$, где каждое значение $u_i \geqslant 0$, если

$$L(x, \bar{u}) \leqslant L(\bar{x}, \bar{u}) \leqslant L(\bar{x}, u) \tag{22}$$

при любых $(x_1, x_2, \ldots, x_n)$ и неотрицательных $(u_1, u_2, \ldots, u_M)$. В настоящем контексте переменные u_i именуются **множителями Лагранжа.** Говоря нестрого, значение функции $L(x, \bar{u})$ уменьшается по мере удаления x от $\bar{x}$, а значение функции $L(\bar{x}, u)$ возрастает по мере удаления u от $\bar{u}$.

Можно показать, что пара $\bar{x}$ и $\bar{u} \geqslant 0$, удовлетворяющих условиям оптимальности (9), (10) и (11), является также седловой точкой $L(x, u)$, и наоборот. Следовательно, используя подход Лагранжа, можно сформулировать следующие условия.

Условия оптимальности Лагранжа. Если для функций $a_i(x)$ и $c(x)$ выполняются условия, определяемые характеристикой системы ограничений и допущениями I) — VI), приведенными в разд. 15.5, то точка $\bar{x}$ будет оптимальна в том и только том случае, если существует точка $\bar{u}$, $\bar{u} \geqslant 0$, такая, что пара $(\bar{x}, \bar{u})$ является седловой точкой функции $L(x, u)$. При этом для седловой точки сумма в правой части (21) равна нулю.

[Интересно отметить, что *достаточное* условие оптимальности $\bar{x}$ в том случае, когда пара $(\bar{x}, \bar{u})$ есть седловая точка, сохраняет свою силу и тогда, когда мы откажемся от всех сделанных в разд. 15.5 допущений относительно вида и гладкости функций $a_i(x)$ и $c(x)$.]

Заметим, что функция Лагранжа $L(x, u)$ вогнута по x при $u \geqslant 0$ и линейна относительно u. Связь между ранее высказанными соображениями и условиями оптимальности Лагранжа можно установить следующим образом. В соответствии с (14) в седловой точке первые частные производные функции $L(x, u)$ по каждому из x_j должны быть неположительными; объясните, почему это так. Далее, в соответствии с (13) либо $\bar{x}_j$, либо первая частная производная $L(x, u)$ по x_j (либо обе эти величины одновременно) должна быть равна нулю, потому что в противном случае значение $L(x, u)$ можно увеличить, изменяя $\bar{x}_j$. Аналогично этому, в соответствии с (9) первая частная производная $L(x, u)$ по каждому u_i должна быть неотрицательной. При этом в соответствии с (12) либо $\bar{u}_i$, либо первая частная производная $L(x, u)$ по u_i (либо обе эти величины одновременно) должна быть равна нулю; объясните, почему это так.

Поскольку в седловой точке $L(x, u)$ равна оптимальному значению $c(x)$, чувствительность последнего к изменениям параметров можно установить, вычисляя влияние этих изменений на $L(x, u)$. Так, например, если ограничение i заменить на $A_i(x) \leqslant e$, где e — очень малое положительное число, то оптимальное значение $c(x)$

возрастет приблизительно на $e\bar{u}_i$ (за исключением тех случаев, когда при бесконечно малых значениях e величины $\bar{x}$ и $\bar{u}$ не являются непрерывными функциями e). Например, в (20) $u = \frac{3}{4}$ для задачи Р. Если четвертое ограничение (15) заменить на $A_4(x) \leqslant e$, то значение $c(x)$ при бесконечно малом e увеличится по сравнению с 7,25 приблизительно на $\frac{3}{4}e$.

Функцию Лагранжа можно также использовать для *непосредственного* нахождения оптимального решения. Имеется много методов выполнения этого; один из них детально рассмотрен в разд. 15.9, а два других вкратце описаны ниже.

Пусть модель включает только одно нелинейное ограничение, например $a_1(x)$, а все остальные ограничения линейны. Заменим целевую функцию задачи на $[c(x) - a_1(x)u]$ и исключим $a_1(x)$ из системы ограничений. Начав с $u = 0$, решим модель, которая теперь имеет нелинейную целевую функцию и только линейные ограничения; для такой задачи применимы методы, изложенные в разд. 15.2, 15.3 и 15.4. Если при $u = 0$ получаем решение, в котором $a_1(x) \leqslant 0$, то это решение оптимально. В противном случае будем увеличивать значение u до тех пор, пока не начнет выполняться условие $a_1(x) = 0$. При этом мы, по существу, находим седловую точку функции Лагранжа (21), и соответствующее решение является оптимальным.

Такой же подход применим и к решению некоторых задач динамического программирования, как, например, к задаче распределения ресурсов, приведенной в разд. 10.2. Например, задачу с двумя ограничениями можно свести к задаче с одним ограничением. Соответственно для решения модифицированной задачи может быть использовано рекуррентное соотношение динамического программирования для одной переменной состояния. Однако если значения управляющих переменных должны быть целочисленными, то окончательное решение, найденное с помощью указанного подхода, может оказаться не более чем приближенным решением исходной задачи.

Другим вычислительным применением функции Лагранжа являются так называемые *мелкошаговые градиентные* алгоритмы. При использовании таких методов строится система дифференциальных уравнений, основанных на первых частных производных функции Лагранжа по каждому из x_j и u_i. При определенных допущениях траектории переменных, определяемых дифференциальными уравнениями, сходятся к оптимальному решению.

Двойственность. Используя функцию Лагранжа (21), можно сформулировать следующую оптимизационную задачу, которая называется *двойственной*: найти x и u, $u \geqslant 0$, такие, что

$$\text{Минимизируется } L(x, u) \tag{23}$$

при условиях

$$\frac{\partial c(x)}{\partial x_j} - \sum_{i=1}^{M} \frac{\partial A_i(x)}{\partial x_j} \cdot u_i = 0, \quad j = 1, 2, \ldots, n. \tag{24}$$

Можно доказать следующую теорему.

Теорема двойственности нелинейного программирования. Пусть для функций $a_i(x)$ и $c(x)$ справедливы условия, определяемые характеристикой системы ограничений и допущениями I) — VI), приведенными в разд. 15.5. Тогда:

I) Если $\bar{x}$ — оптимальное решение задачи нелинейного программирования, то существует $\bar{u} \geqslant 0$, такая, что пара $\bar{x}$ и $\bar{u}$ является решением двойственной задачи (23) — (24). Далее, если x является допустимым при учете условий исходной задачи, а x и u — допустимыми при учете условий двойственной задачи, то $c(x) \leqslant L(x, u)$; равенство соблюдается для оптимальной пары $\bar{x}$ и $\bar{u}$.

II) Обратно, если пара $\bar{x}$ и $\bar{u}$ является решением двойственной задачи (23) — (24), причем $A_i(\bar{x}) \leqslant 0$ для каждого i, а также если $c(x)$ строго вогнута, либо существует такое k, что $A_k(x)$ строго выпукла и $\bar{u}_k > 0$, то $\bar{x}$ есть решение исходной задачи нелинейного программирования. При этом $c(\bar{x}) = L(\bar{x}, \bar{u})$.

Обратите внимание на то, что целевая функция (23) рассматриваемой двойственной задачи имеет не такой простой вид, как в двойственной задаче линейного программирования. В частности, в $L(x, u)$ в качестве составной части входит целевая функция прямой задачи $c(x)$. Теорема имеет несколько важных областей использования. Одна из них, которая в тексте не рассматривается ввиду ее сложности, связана с двойственной постановкой задачи квадратичного программирования, позволяющей построить несколько изящных алгоритмов. Другое направление использования описывается в разд. 15.9.

15.8. ВОЗВРАЩЕНИЕ К КВАДРАТИЧНОМУ ПРОГРАММИРОВАНИЮ

В разд. 14.6 рассматривалась следующая задача квадратичного программирования:

$$\text{Максимизировать } c(x) \equiv \sum_{j=1}^{n} c_j x_j + \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk} x_j x_k \tag{1}$$

при условиях

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

$$x_j \geqslant 0, \quad j = 1, 2, \ldots, n. \tag{3}$$

В настоящем разделе примем, что коэффициенты c_{jk} определены так, что для всех k и j $c_{jk} = c_{kj}$; это допущение не является существенным, поскольку любую квадратичную форму можно записать в предложенном виде (например, $8x_1x_2 - 5x_2x_1 \equiv 1{,}5x_1x_2 + 1{,}5x_2x_1$). (В задачу разд. 14.6 были включены дополняющие переменные $s_i \geqslant 0$; они будут введены и здесь, но несколько ниже.) Для простоты примем, что все $b_i \geqslant 0$. Постулируем также, что квадратичная форма под знаком двойного суммирования в правой части (1) представляет собой вогнутую функцию. (С математической точки зрения это равносильно утверждению, что квадратичная форма является отрицательно полуопределенной.)

Теперь используем необходимые и достаточные условия оптимальности решения, представленные в предыдущем разделе выражениями (10) — (11). Для этого построим систему

$$A_i(x) \equiv \sum_{j=1}^{n} a_{ij}x_j - b_i, \quad i = 1, 2, \ldots, m, \tag{4}$$

$$A_{m+j}(x) \equiv -x_j, \quad j = 1, 2, \ldots, n.$$

Как нетрудно проверить,

$$\frac{\partial c(x)}{\partial x_j} = c_j + 2\sum_{k=1}^{m} c_{jk}x_k, \quad j = 1, 2, \ldots, n. \tag{5}$$

Тогда из (10) предыдущего раздела следует

$$\left(\sum_{j=1}^{n} a_{ij}x_j - b_i\right)\cdot u_i = 0, \quad i = 1, 2, \ldots, m, \tag{6}$$

$$(-x_j)\cdot u_{m+j} = 0, \quad j = 1, 2, \ldots, n, \tag{7}$$

а из (11)

$$c_j + 2\sum_{j=1}^{n} c_{jk}x_k - \sum_{i=1}^{m} a_{ij}u_i + u_{m+j} = 0, \quad j = 1, 2, \ldots, n. \tag{8}$$

Эти выражения можно упростить, введя некоторые дополнительные обозначения. Пусть $s_i \geqslant 0$ — дополняющая переменная ограничения i в (2), так что

$$\sum_{j=1}^{n} a_{ij}x_j + s_i = b_i, \quad i = 1, 2, \ldots, m, \tag{9}$$

и пусть $w_j = u_{m+j}$ для $j = 1, 2, \ldots, n$. Тогда, после упрощений, (6) и (7) принимают вид

$$s_iu_i = 0, \quad i = 1, 2, \ldots, m, \tag{10}$$

$$x_jw_j = 0, \quad j = 1, 2, \ldots, n. \tag{11}$$

Поскольку s_i, u_i, x_j и w_j должны быть неотрицательными, (10) и (11) можно заменить одним ограничением дополнительной нежесткости

$$\sum_{i=1}^{m} s_iu_i + \sum_{j=1}^{n} x_jw_j = 0. \tag{12}$$

Используя только что введенные переменные w_j, можно записать (8) в следующем виде:

$$2\sum_{k=1}^{n} c_{jk}x_k - \sum_{i=1}^{m} a_{ij}u_i + w_j = -c_j, \quad j = 1, 2, \ldots, n. \tag{13}$$

Одним из следствий условий оптимальности, приведенных в разд. 15.7, является то, что если удается найти такие неотрицательные значения x_j, s_i, u_i и w_i, при которых удовлетворяются (9), (12) и (13), то эти значения x_j представляют собой оптимальное решение задачи квадратичного программирования. Отметим, что система (9) и (13) включает $m + n$ *линейных* ограничений и 2 $(m + n)$ неотрицательных переменных. Разумеется, (12) не является линейным ограничением, но оно имеет столь специфический характер, что можно разработать целый ряд методов, позволяющих полностью использовать линейность остальных необходимых и достаточных условий оптимальности решения. Один из таких алгоритмов излагается ниже; он основан на довольно простых модификациях симплексного метода и обеспечивает сходимость за конечное число итераций.

Модифицированный симплексный алгоритм. Прежде чем перейти к описанию собственно алгоритма, определим некоторые удобные для нас термины. Назовем x_j и s_j прямыми переменными, а w_j и u_i — двойственными переменными. Назовем также *взаимодополняющими* переменными любую пару x_j и w_j или аналогично этому любую пару s_i и u_i. В настоящем методе система линейных ограничений состоит из (9) и (13), причем на каждой итерации имеется базисное решение, включающее $m + n$ переменных. Каждая входящая в базис прямая переменная неотрицательна; как следствие исходные ограничения (2) всегда удовлетворяются. Как и в симплексном методе, итерации завершаются, когда не остается отрицательных двойственных переменных.

Базисное решение на любой итерации всегда принадлежит к одному из двух возможных типов. Первая возможность заключается в том, что для любого j или i в базис входит только одна из взаимодополняющих переменных. Такое решение именуется *стандартным базисом*, оно удовлетворяет ограничению дополнительной нежесткости (12). Вторая возможность состоит в том, что решение включает *базисную пару* взаимодополняющих переменных, в результате чего имеется также и *небазисная пара* взаимодополняющих переменных. Такое решение называется *нестандартным базисом*. (Некоторые авторы применяют термины *взаимодополняющий базис* и *почти взаимодополняющий базис*.) Принципиальное отличие данного метода от симплексного состоит в том, что всякий раз, когда базис оказывается нестандартным, выбор переменной, вводимой в базис, осуществляется таким образом, что восстанавливается выполнимость условия дополнительной нежесткости (12).

Путем простых алгебраических преобразований, используя ограничения (9) и (13), получаем следующее выражение для $c(x)$, справедливое при выполнении условия дополнительной нежесткости (12), т. е. в случае стандартного базиса:

$$c(x)=\frac{1}{2}\left(\sum_{j=1}^{n} c_j x_j+\sum_{i=1}^{m} b_i u_i\right). \tag{14}$$

Введя обозначение

$$x_0 \equiv 2c(x), \tag{15}$$

можно дополнить систему ограничений *линейным* соотношением

$$x_0-\sum_{j=1}^{n} c_j x_j-\sum_{i=1}^{m} b_i u_i=0 \tag{16}$$

так, чтобы переменная $2c(x)$ вошла в стандартный базис.

Объяснять конкретные правила реализации алгоритма мы будем на задаче P, приведенной в разд. 15.1. Система линейных ограничений, состоящих из (16), (9) и (13), имеет следующий вид:

$$\begin{array}{llllllll}
x_0-6x_1-4x_2-2x_3-4u_1 & & & & & = & 0 & \text{(строка 0)},\\
\quad x_1+2x_2+x_3 & +s_1 & & & & = & 4 & \text{(строка 1)},\\
-6x_1 & -u_1 & +w_1 & & & = & -6 & \text{(строка 2)},\\
\quad +4x_2 & -2u_1 & & +w_2 & & = & -4 & \text{(строка 3)},\\
\quad -\frac{2}{3}x_3-u_1 & & & & +w_3 & = & -2 & \text{(строка 4)}.
\end{array} \tag{17}$$

Пусть исходное решение включает переменные x_0, s_1, w_1, w_2, w_3; это — стандартный базис. Затем применим следующий критерий.

Модифицированный критерий I симплексного метода. а) При стандартном базисе следует прекратить расчеты, как только все двойственные переменные примут неотрицательные значения. При наличии отрицательных двойственных переменных нужно выбрать для ввода в базис небазисную прямую переменную, которая является взаимодополняющей для отрицательной базисной двойственной переменной с наибольшим абсолютным значением. б) При нестандартном базисе следует выбрать для ввода в базис двойственную переменную из небазисной взаимодополняющей пары.

Поскольку в системе (17) отрицательной двойственной переменной с наибольшим абсолютным значением является w_1, выберем для ввода в базис x_1.

При определении переменной, исключаемой из базиса, желательно сохранить стандартный характер базиса, если он имелся, или создать его, если его не было. Поскольку, однако, данный метод всегда обеспечивает соблюдение допустимости в отношении прямых переменных, не всегда возможно сохранить или создать стандартный базис. Эти соображения учитываются в следующем критерии выбора переменной, исключаемой из базиса.

Модифицированный критерий II симплексного метода. а) Необходимо вычислить отношения текущих правых частей к коэффициентам столбца вводимой переменной для тех строк, которые соответствуют находящимся в базисе прямым переменным (не следует рассматривать при этом те отношения, в которых знаменатель является нулевым или отрицательным). Значение наименьшего из этих отношений обозначается через r; пусть оно оказалось в строке k, где находится некоторая прямая базисная переменная. б) При стандартном базисе нужно рассмотреть отношение текущей правой части к коэффициенту вводимой прямой переменной, расположенному в строке, соответствующей взаимодополняющей двойственной переменной. Если это отношение положительно и не превышает r, требуется исключить из базиса взаимодополняющую двойственную переменную. В противном случае следует исключить из базиса прямую переменную из строки k. в) При нестандартном базисе нужно рассмотреть отношение текущей правой части к коэффициенту вводимой переменной в строке, соответствующей двойственной переменной базисной взаимодополняющей пары. Если это отношение положительно и не превышает r, необходимо исключить эту двойственную переменную из базиса. В противном случае требуется исключить прямую базисную переменную из строки k.

Поскольку вводится x_1, а текущее решение (17) представляет собой стандартный базис, согласно части а) получаем $r = 4/1$; из части б) следует выбрать отношение $-6/-6$. Следовательно, из базиса выводится взаимодополняющая двойственная переменная w_1 и следующий базис снова является стандартным. Операция преобразования матрицы на основе использования ведущего элемента выполняется так же, как в обычном симплексном методе.

После ввода в базис переменной x_1 и исключения w_1 система ограничений приобретает следующий вид:

$$\begin{array}{llllllll}
x_0 & -4x_2 - 2x_3 & -3u_1 & & -w_1 & & = 6 & \text{(строка 0)},\\
 & 2x_2 + x_3 & -\frac{1}{6}u_1 & +s_1 & +\frac{1}{6}w_1 & & = 3 & \text{(строка 1)},\\
x_1 & & +\frac{1}{6}u_1 & & -\frac{1}{6}w_1 & & = 1 & \text{(строка 2)},\\
 & -4x_2 & -2u_1 & & & +w_2 & = -4 & \text{(строка 3)},\\
 & \quad -\frac{2}{3}x_3 & -u_1 & & & +w_3 & = -2 & \text{(строка 4)}.
\end{array} \qquad (18)$$

Теперь, согласно *критерию I*, в базис следует ввести переменную x_2, поскольку в (18) w_2 является отрицательной двойственной переменной с наибольшим абсолютным значением. Следовательно, согласно *критерию II*, $r = 3/2$, а отношение в части б) равно $-4/-4$, так что из базиса исключается переменная w_2. И здесь следующее решение представляет собой стандартный базис; выполняется так называемое **взаимодополняющее преобразование матрицы на основе использования ведущего элемента,** поскольку в базис вводится прямая переменная, а из базиса исключается соответствующая взаимодополняющая двойственная переменная.

После этой операции система ограничений приобретает такой вид:

$$
\begin{array}{lll}
x_0 \quad - 2x_3 - u_1 - w_1 - w_2 & = 10 & \text{(строка 0)},\\
x_3 - \frac{7}{6}u_1 + s_1 + \frac{1}{6}w_1 + \frac{1}{2}w_2 & = 1 & \text{(строка 1)},\\
x_1 \quad + \frac{1}{6}u_1 \quad - \frac{1}{6}w_1 & = 1 & \text{(строка 2)}, \quad (19)\\
x_2 \quad + \frac{1}{2}u_1 \quad - \frac{1}{4}w_2 & = 1 & \text{(строка 3)},\\
-\frac{2}{3}x_3 - u_1 \quad + w_3 & = -2 & \text{(строка 4)}.
\end{array}
$$

Теперь, согласно *критерию I*, должна вводиться переменная x_3. Далее, из строки 1 $r = 1/1$, а отношение для части б) *критерия II* равно $-2/-\frac{2}{3}$, так что из базиса исключается s_1. Соответственно следующее решение представляет собой нестандартный базис.

После преобразования матрицы на основе использования ведущего элемента получаем

$$
\begin{array}{lll}
x_0 \quad -\frac{10}{3}u_1 + 2s_1 - \frac{2}{3}w_1 & = 12 & \text{(строка 0)},\\
x_3 - \frac{7}{6}u_1 + s_1 + \frac{1}{6}w_1 + \frac{1}{2}w_2 & = 1 & \text{(строка 1)},\\
x_1 \quad + \frac{1}{6}u_1 \quad - \frac{1}{6}w_2 & = 1 & \text{(строка 2)}, \quad (20)\\
x_2 \quad + \frac{1}{2}u_1 \quad - \frac{1}{4}w_2 & = 1 & \text{(строка 3)},\\
-\frac{16}{9}u_1 + \frac{2}{3}s_1 + \frac{1}{9}w_1 + \frac{1}{3}w_2 + w_3 & = -\frac{4}{3} & \text{(строка 4)}.
\end{array}
$$

Базисными переменными в (20) являются x_0, x_1, x_2, x_3, w_3; небазисную пару взаимодополняющих переменных составляют s_1 и u_1. Соответственно, согласно части б) *критерия I*, в базис должна быть включена переменная u_1. Далее, $r = \min\left(1\Big/\frac{1}{6}, 1\Big/\frac{1}{2}\right) = 2$, а отношение для части в) равно $\left(-\frac{4}{3}/-\frac{16}{9}\right) = \frac{3}{4}$. Таким образом, из базиса исключается w_3, и тем самым следующее решение снова представляет собой стандартный базис.

После преобразования система ограничений приобретает следующий вид:

$$
\begin{array}{llr}
x_0 \qquad + \frac{3}{4} s_1 - \frac{7}{8} w_1 - \frac{5}{8} w_2 - \frac{15}{8} w_3 = 14\frac{1}{2} & \text{(строка 0)}, & \\
x_3 \quad + \frac{9}{16} s_1 + \frac{3}{32} w_1 + \frac{9}{32} w_2 - \frac{21}{32} w_3 = 1\frac{7}{8} & \text{(строка 1)}, & \\
x_1 \qquad + \frac{1}{16} s_1 - \frac{5}{32} w_1 + \frac{1}{32} w_2 + \frac{3}{32} w_3 = \frac{7}{8} & \text{(строка 2)}, & (21) \\
x_2 \qquad + \frac{3}{10} s_1 + \frac{1}{32} w_1 - \frac{5}{32} w_2 + \frac{9}{32} w_3 = \frac{5}{8} & \text{(строка 3)}, & \\
u_1 - \frac{3}{8} s_1 - \frac{1}{16} w_1 - \frac{3}{16} w_2 - \frac{9}{16} w_3 = \frac{3}{4} & \text{(строка 4)}. &
\end{array}
$$

Расчеты прекращаются, поскольку текущее решение представляет собой стандартный базис с неотрицательными значениями всех двойственных переменных. Оптимальное значение $c(x)$ равно $0{,}5x_0 = 7{,}25$.

Полное сравнение вычислительных достоинств настоящего алгоритма и алгоритма, описанного в разд. 14.6, выходит за рамки данной книги, поскольку оно связано с рядом деталей составления эффективной машинной программы. Однако вкратце можно сказать, что, по-видимому, алгоритм настоящего раздела обеспечивает сходимость за меньшее число итераций; за это, однако, приходится платить тем, что система ограничений все время включает $m + n$ уравнений. Алгоритм, приведенный в разд. 14.6, имеет то преимущество, что он всегда обеспечивает сходимость к относительной стационарной точке, даже если целевая функция не вогнута; напротив, для квадратичной целевой функции произвольного вида настоящий алгоритм не всегда обеспечивает сходимость.

Ряд других имеющихся алгоритмов квадратичного программирования близок к модифицированному симплексному методу в том отношении, что в них также уделяется большое внимание взаимодополняющему преобразованию матрицы ограничений; кроме того, и они не обеспечивают сходимость за конечное число итераций.

15.9. МЕТОД ШТРАФНОЙ ФУНКЦИИ

В разд. 15.7 было показано, что задачу нелинейного программирования можно сформулировать в терминах нахождения седловой точки функции Лагранжа. Из такой постановки ведет свое начало и метод штрафной функции. В этом методе ряд ограничений (или вся система ограничения) включается в состав целевой функции. Имеется несколько способов реализации данного подхода; мы выбрали описываемый ниже алгоритм ввиду его успешной практической реализации. По-прежнему предполагается, что задача нелинейного программирования удовлетворяет как допущениям I) — VI), так и характеристике системы ограничений (разд. 15.5).

Начнем объяснение с того, что покажем способ включения одного ограничения $a_i(x)$ в целевую функцию. Рассмотрение этого простого случая поможет обосновать общий подход, который рассматривается далее. В качестве модифицированной максимизируемой целевой функции примем следующую:

$$c(x \mid r) \equiv c(x) + r\,[1/a_i(x)], \tag{1}$$

где $r > 0$ — параметр. Для тех x, при которых $a_i(x) < 0$, функция $[1/a_i(x)]$ вогнута; поэтому $c(x \mid r)$ вогнута для точек x, находящихся строго во внутренней части области, определенной ограничением $a_i(x) \leqslant 0$. Для допустимой точки x, расположенной вблизи границы $a_i(x) = 0$, величина $r\,[1/a_i(x)]$ становится очень большим по абсолютному значению отрицательным числом. Соответственно точка, в которой $c(x \mid r)$ максимизируется при условии выполнения всех ограничений, никак не может находиться *на* границе $a_i(x) = 0$. Однако если значение параметра r достаточно близко к нулю, то такая точка может оказаться близкой к оптимуму исходной задачи.

Для того чтобы применить подход с использованием штрафной функции, несколько ограничений (или вся их система) включается в целевую функцию аналогично (1). Находим допустимое решение, оптимальное для модифицированной целевой функции при заданном исходном значении r. Затем последовательно будем уменьшать значение r ($r > 0$), каждый раз находя новое решение модифицированной задачи. Расчеты прекращаются, когда выражение, содержащее r, станет оказывать пренебрежимо малое влияние на $c(x \mid r)$; получаемое решение можно считать приближенным оптимумом. Если исходная модель включает ограничение $f_i(x) \leqslant 0$, причем можно предполагать, что оптимальное решение действительно лежит на границе $f_i(x) = 0$, то можно ввести функцию $a_i(x) \equiv f_i(x) - e$, где e — малое положительное число. Однако, если выбрать e слишком большим, такой подход может привести нас в точку, где $f_i(x) > 0$.

Почему такой подход привлекателен с вычислительной точки зрения? Поскольку проверяемая точка никогда не попадает на нелинейную границу, на каждой итерации можно вычислить *направления* улучшения, *как в случае*, когда оптимизационный процесс проводится для отыскания безусловного максимума целевой функции (т. е. максимума при отсутствии ограничений). Поэтому данный метод можно рассматривать как алгоритм последовательного нахождения безусловного максимума. Вычислив набор направлений, мы затем определяем оптимальную *допустимую* длину шага. При оптимизации помогает то обстоятельство, что во внутренней части области допустимых решений модифицированная целевая функция строго вогнута. При практическом применении алгоритма мы можем для каждого r найти только приближение к оптимальному значению $c(x \mid r)$, поскольку, как правило, точный оптимум невозможно найти за конечное число итераций. Детали метода полностью описаны ниже.

В приводимом описании алгоритма все ограничения, в том числе и условия неотрицательности переменных, включаются в модифицированную целевую функцию. Однако следует подчеркнуть, что если в системе ограничений имеется всего лишь несколько нелинейных функций, а остальные линейны, то метод штрафной функции можно сочетать с любым алгоритмом оптимизации нелинейной функции при условии выполнения линейных ограничений. В частности, можно модифицировать целевую функцию, включая в нее только нелинейные ограничения; тогда применим алгоритм, аналогичный описанным в разд. 15.2—15.4, при последовательно убывающих значениях r, $r > 0$. Каждое получаемое решение является допустимым решением исходной задачи, а при достаточно малом r — приближением к оптимуму.

Алгоритм последовательного нахождения безусловного максимума. Предположим, что в целевую функцию включены все ограничения, в том числе и условия неотрицательности $x_j \geqslant 0$. Пусть ограничения имели вид $A_i(x) \leqslant 0$ для $i = 1, 2, \ldots, M$, где $M = m + n$. Тогда модифицированная задача заключается в максимизации

$$c(x \mid r) \equiv c(x) + r\left[\sum_{i=1}^{M} \frac{1}{A_i(x)}\right], \quad r > 0, \tag{2}$$

по всем допустимым значениям x. Выражение под знаком суммирования в правой части (2) включает и члены вида $(1/-x_j)$, так что $c(x \mid r)$ строго вогнута и имеет единственное оптимальное решение для каждого значения r. Выражение в квадратных скобках в правой части (2) отрицательно при всех допустимых значениях x, так что $c(x \mid r) < c(x)$.

Алгоритм включает следующие операции.

Шаг 1. Выберем исходное значение $r^1 > 0$ и исходную проверяемую точку x^0, такую, что каждая $A_i(x^0) < 0$. Примем $k = 0$ и $n = 1$.

Шаг 2. Если x^k очень близка к точке максимума функции $c(x \mid r^n)$, перейдем к шагу 4. В противном случае определим направления d_j^k для $j = 1, 2, \ldots, n$. Выберем длину шага t^k, позволяющую максимизировать $c(x^k + td^k \mid r)$, где точка $(x^k + t^k d^k)$ должна удовлетворять всем ограничениям. Перейдем к шагу 3.

Шаг 3. Определим новую проверяемую точку

$$x_j^{k+1} = x_j^k + t^k d_j^k, \quad j = 1, 2, \ldots, n. \tag{3}$$

Перейдем от k к $k + 1$ и возвратимся к шагу 2.

Шаг 4. Прекратим расчеты, если точка x^k близка к оптимуму. В противном случае зададимся новым значением r^{n+1}, где $0 < r^{n+1} < r^n$. Перейдем от n к $n + 1$ и возвратимся к шагу 2.

Эффективные способы выполнения шага 1 представляют интерес главным образом для специалистов по исследованию операций, и поэтому здесь они подробно не обсуждаются. Однако необходимо остановиться на выборе значения r^1. Если r^1 слишком велико, то оптимальное решение задачи максимизации $c(x \mid r^1)$ будет, вероятно, лежать весьма далеко от оптимального решения исходной задачи. С другой стороны, если r^1 слишком мало, то нахождение приближенного максимума функции $c(x \mid r^1)$ обычно требует значительного числа итераций. При выборе значения r^1 таким образом разумно избегать как чрезмерного завышения, так и чрезмерного занижения r^1.

На шаге 2 можно использовать любой метод выбора направлений, эффективный при решении задач на безусловный максимум. Поскольку функция $c(x \mid r)$ строго вогнута для всех допустимых x, условия

$$\frac{\partial c(x \mid r)}{\partial x_j} = \frac{\partial c(x)}{\partial x_j} - r\sum_{i=1}^{M}\left\{\frac{\frac{\partial A_i(x)}{\partial x_j}}{[A_i(x)]^2}\right\} = 0 \text{ для всех } j \qquad (4)$$

являются как необходимыми, так и достаточными для того, чтобы при допустимом x значение $c(x \mid r)$ достигало максимума. Если направления выбираются в соответствии с методом скорейшего подъема, описанным в разд. 14.5, то каждая величина $d_j = dc(x \mid r)/dx_j$. (Практики считают полезным использовать методы второго порядка, в которых учитываются вторые частные производные.)

На шаге 4 хорошие результаты показало применение следующей формулы уменьшения значения r: $r^{n+1} = f \cdot r^n$, где $0 < f < 1$. Конкретное значение f, по-видимому, не оказывает решающего влияния, однако при малом f оптимизация $c(x \mid r^n)$ потребует для каждого n большего числа итераций.

Настоящий метод обеспечивает увеличение значения $c(x \mid r)$ для каждой новой проверяемой точки, и можно доказать, что по мере приближения r к нулю максимальное значение $c(x \mid r)$ сходится к оптимальному значению $\bar{c}$ исходной задачи. Более того, если по достижении шага 4 точка x^k действительно соответствует максимуму $c(x \mid r^n)$, то полученное $c(x^k)$ по крайней мере столь же велико, как на предыдущем шаге 4, а обычно — строго больше. Аналогично этому выражение в квадратных скобках в правой части (2) никогда не превышает значения, достигнутого на предыдущем шаге 4, и верхняя грань $\bar{c}$ равна

$$\bar{c} \leqslant c(x^k) - r^n\left[\sum_{i=1}^{M}\frac{1}{A_i(x^k)}\right]. \qquad (5)$$

Как отмечалось выше, x^k практически не достигает на шаге 4 истинного оптимума $c(x \mid r^n)$. Поэтому при использовании алгоритма отмеченные свойства наблюдаются только в том случае, если целевая функция $c(x^k \mid r^n)$ достаточно близка к своему максимальному значению. С учетом этого предупреждения можно сравнивать значение выражения в правой части (5) со значением $c(x^k)$, чтобы на основе этого сравнения принимать решение о прекращении расчетов на шаге 4. В заключение сделаем еще одно предупреждение относительно практического применения метода: при вычислениях следует быть крайне осторожным, поскольку в случае малых значений r и знаменателей выражения в квадратных скобках (2) результаты оказываются чрезвычайно чувствительными к округлению.

Справедливость (5) можно обосновать результатами, полученными при рассмотрении проблемы двойственности в разд. 15.7. Пусть x — истинно оптимальное решение для целевой функции $c(x \mid r^n)$. Тогда в соответствии с (4) точка x должна удовлетворять условию

$$\frac{\partial c(x)}{\partial x_j} - r^n \left\{ \sum_{i=1}^{M} \frac{\dfrac{\partial A_i(x)}{\partial x_j}}{[A_i(x)]^2} \right\} = 0 \text{ для всех } j. \tag{I}$$

Примем

$$u_i \equiv \frac{r^n}{[A_i(x)]^2} \geqslant 0, \quad i = 1, 2, \ldots, M, \tag{II}$$

так что x и u удовлетворяют (24) из разд. 15.7. Следовательно, исходя из теоремы двойственности нелинейного программирования,

$$\bar{c} \leqslant L(x, u) = c(x) - r^n \left[\sum_{i=1}^{M} \frac{1}{A_i(x)} \right], \tag{III}$$

что согласуется с (5).

Примеры. Результаты решения задач P и Q из разд. 15.1 и 15.5 приведены в таблицах рис. 15.8 и 15.9. Для каждого r^n приближенное оптимальное значение $c(x \mid r^n)$ было найдено с помощью метода скорейшего подъема, т. е. в качестве направлений принимались первые частные производные функций $c(x \mid r^n)$ в точке x^k. Приведенные в таблицах значения k соответствуют тем x^k, которые на шаге 2 считались приближенным оптимумом. Отметим, что наибольшее улучшение значения целевой функции $c(x)$ происходило при максимизации $c(x \mid r^1)$.

Если при решении задачи P использовать ту же исходную точку x^0, но принять $f = 0{,}75$, сходимость несколько замедляется. В этом случае при $n = 10$ $r^{10} = 0{,}0027$; соответствующее приближенное оптимальное решение для максимизируемой функции $c(x \mid r^{10})$ есть $x^{33} = (0{,}8813,\ 0{,}6372,\ 1{,}7873)$ и для него $c(x^{33}) = 7{,}2043$ и $c(x^{33} \mid r^{10}) = 7{,}1473$. Полученные результаты сопоставимы с результатами для $n = 5$ и $k = 27$ в таблице рис. 15.8.

$f=0,5$

n	r^n	k	$c(k^k)$	x_1^k	x_2^k	x_3^k	Верхняя грань	$c(x^k \mid r^n)$
	—	0	1,147	0,1	0,1	0,1	—	—
1	0,0365	14	7,091	0,854	0,595	1,755	7,397	6,785
2	0,0182	19	7,136	0,872	0,606	1,770	7,323	6,950
3	0,0091	22	7,168	0,877	0,619	1,779	7,285	7,051
4	0,0046	25	7,189	0,881	0,628	1,786	7,262	7,115
5	0,0023	27	7,203	0,883	0,634	1,790	—	7,157
10	0,0001	34	7,239	0,890	0,650	1,801	—	7,232
Оптимальные значения	—	—	7,239	8,90	6,50	1,801	—	7,232

Рис. 15.8. Алгоритм последовательного нахождения безусловного максимума — решение задачи P.

Если начать с лучшей исходной точки $x^0 = (0,5\ 0,5,\ 0,5)$, для которой $c(x^0) = 4,6667$, однако с большего $r^1 = 0,6068$ (при $f = 0,5$), то получаются удивительные результаты. Хотя исходная проверяемая точка и лучше, чем та, которая использована в таблице рис. 15.8, проверяемое решение при $n = 5$ и $k = 44$ имеет вид $x^{44} = (0,8748,\ 0,6246,\ 1,6646)$, причем для него $c(x^{44}) = 7,0767$

$f=0,5$

n	r^n	k	$c(x^k)$	x_1^k	x_2^k	x_3^k	Верхняя грань	$c(x^k \mid r^n)$
	—	0	−6	1	1	3	—	—
1	0,3	15	−4,679	0,935	7,776	2,193	−3,120	−6,239
2	0,15	23	−4,474	0,913	0,724	2,113	−3,553	−5,396
3	0,075	29	−4,332	0,897	0,685	2,064	−3,774	−4,890
4	0,0375	31	−4,240	0,884	0,656	2,043	−3,899	−4,581
5	0,0188	33	−4,171	0,874	0,634	2,028	−3,955	−4,387
6	0,094	36	−4,119	0,868	0,618	2,015	−3,975	−4,263
7	0,0047	39	−4,085	0,866	0,609	2,001	−3,989	−4,182
Оптимальные значения	—	—	−4,00	0,875	0,625	1,875	—	−4,00

Рис. 15.9. Алгоритм последовательного нахождения безусловного максимума — решение задачи Q.

и $c(x^{44} \mid r^5 = 0{,}0379) = 6{,}7704$. При большем значении r^1 требуется гораздо больше итераций для того, чтобы получить результаты, сравнимые с данными таблицы рис. 15.8 для $n = 1$ и $k = 14$.

Если правую часть линейного ограничения принять равной не 4, а 4,1, то при $n = 3$ и $k = 33$ получаем проверяемое решение $x^{33} = (0{,}8869,\ 0{,}6584,\ 1{,}7775)$, для него $c(x^{33}) = 7{,}2301$. Эта точка немногим отличается от точки x^{34} из таблицы рис. 15.8.

При решении задачи Q использование r^1, равного 0,1 или 0,2, пренебрежимо мало влияет на значение k, необходимое для достижения результатов, которые сравнимы с решением при $n = 7$ и $k = 39$, приведенным в таблице рис. 15.9.

15.10. ОБОБЩЕННЫЙ АЛГОРИТМ ПРОГРАММИРОВАНИЯ

Этот метод имеет много общего с методом сепарабельного программирования, изложенным в разд. 14.7, поскольку в нем тоже используется как линейная аппроксимация на основе сетки точек, так и симплексный метод — для отыскания решения соответствующей задачи линейного программирования. Однако обобщенный алгоритм программирования более усложнен в том отношении, что дает возможность по ходу итеративного процесса улучшать сетку. Данный подход может быть реализован самыми разными способами; поэтому приводимое ниже описание следует рассматривать как не более чем пояснение общей идеи. По-прежнему предполагается, что нелинейные функции удовлетворяют допущениям I) — VI) и характеристике системы ограничений, приведенным в разд. 15.5.

Начнем с того, что всю систему нелинейных ограничений и ограничений неотрицательности переменных разделим на две группы. Пусть первая группа включает R ограничений, которые запишем в следующей форме:

$$A_i(x_1, x_2, \ldots, x_n) \leqslant 0, \quad i = 1, 2, \ldots, R; \tag{1}$$

некоторые из этих ограничений могут представлять собой условия неотрицательности. Не исключено, что в (1) входят все ограничения задачи; тогда $R = m + n$, и вторая группа является пустой. Если же $R < m + n$, то область допустимых точек, удовлетворяющих ограничениям второй группы, выпукла и включает все свои граничные точки, поскольку на эту систему распространяется допущение I) из разд. 15.5. Для удобства введем символ

$S \equiv$ множество всех точек, удовлетворяющих тем ограничениям, которые не включены в (1).

(Если $R = m + n$, S представляет собой все евклидово n-мерное пространство.) Следовательно, оптимизационную задачу можно сформулировать следующим образом:

$$\text{Максимизировать } c(x_1, x_2, \ldots, x_n) \tag{2}$$

при том условии, что точка x будет удовлетворять ограничениям (1) и лежать в области S.

Рассмотрим итерацию k алгоритма. Пусть текущая аппроксимирующая сетка имеет вид $X^0, X^1, \ldots, X^T$, причем каждая из этих точек удовлетворяет ограничениям S, по крайней мере одна допустима относительно (1), а для каждого нелинейного $A_i(x)$ неравенство является строгим. (Вспомним, что в характеристике системы ограничений, приведенной в разд. 15.5, мы постулировали существование такой точки.) Располагая этой сеткой, мы решаем так называемую **суженную координирующую задачу**, представляющую собой линейную аппроксимацию исходной модели:

$$\text{Максимизировать} \quad \sum_{t=0}^{T} c(X^T)\, w_t \tag{3}$$

при условиях

$$\sum_{t=0}^{T} A_i(X^t)\, w_t \leqslant 0, \quad i = 1, 2, \ldots, R, \tag{4}$$

$$\sum_{t=0}^{T} w_t = 1, \quad \text{все} \quad w_t \geqslant 0. \tag{5}$$

Термин «суженная» относится к тому обстоятельству, что задача (3) — (5) строится не для всех возможных точек X^t, допустимых относительно S, а лишь для подмножества $S+1$ точек сетки. Каждое w_t является весом соответствующей точки X^t. Суженная координирующая задача — задача линейного программирования, так что она имеет базисное допустимое решение и конечное максимальное значение c^k; объясните, почему это так. Поскольку $A_i(x)$ — выпуклые функции, а множество S выпукло, нетрудно показать, что любое допустимое решение системы ограничений (4) — (5) представляет собой точку с координатами

$$x_j = \sum_{t=0}^{T} X_j^t\, w_t \quad \text{для всех } j, \tag{6}$$

которая является допустимым решением исходной задачи.

Теперь рассмотрим задачу, двойственную задаче (3) — (5), а именно:

$$\text{Минимизировать } y_{R+1} \tag{7}$$

при условиях

$$\sum_{i=1}^{R} A_i(X^t)\, y_i + y_{R+1} \geqslant c(X^t), \quad t = 0, 1, \ldots, T, \tag{8}$$

$$y_i \geqslant 0, \quad i = 1, 2, \ldots, R, \tag{9}$$

на знак y_{R+1} не налагается никаких ограничений.

Как показано ниже, обращение к двойственной задаче позволяет получить критерий проверки, т. е. показать, действительно ли оптимальное решение суженной координирующей задачи одновременно является и оптимальным решением исходной задачи.

Пусть точка w_k оптимальна для суженной координирующей задачи (3) — (5); соответствующую точку x^k найдем, исчислив средневзвешенные x_j по формуле (6). Пусть y^k — соответствующее оптимальное решение двойственной задачи. Легко показать, что если

$$\sum_{i=1}^{R} A_i(x)\, y_i^k + y_{R+1}^k \geqslant c(x), \text{ или, что эквивалентно,}$$

$$y_{R+1}^k \geqslant c(x) - \sum_{i=1}^{R} A_i(x)\, y_i^k, \tag{10}$$

для *всех* x в S, то x^k является истинным оптимумом исходной задачи. Однако если в S существует X^{T+1}, такая, что

$$y_{R+1}^k < c(X^{T+1}) - \sum_{i=1}^{R} A_i(X^{T+1})\, y_i^k, \tag{11}$$

то добавление этой новой точки к суженной координирующей задаче может привести к улучшению решения. Выяснение того, существует ли такая точка X^{T+1}, называется **подзадачей**. Эффективность обобщенного алгоритма программирования существенно зависит от того, насколько трудно решать подзадачу. Этот аспект вопроса рассматривается несколько более подробно далее в настоящем же разделе.

Подведем итоги. Итерация k требует выполнения следующих операций.

Шаг 1. Найдем оптимальное решение суженной координирующей задачи. Соответствующую проверяемую точку x^k определим по формуле (6).

Шаг 2. Выясним, существует ли в S такая точка X^{T+1}, для которой выполняется неравенство (11). Если нет, прекратим расчеты; x^k является истинным оптимумом. Если такая точка существует, перейдем к шагу 3.

Шаг 3. Добавим точку X^{T+1} к суженной координирующей задаче, введя новый вес w_{T+1} с коэффициентами $c(X^{T+1})$ в (3), $A_i(X^{T+1})$ в (4) и 1 в (5). Перейдем от k к $k+1$ и возвратимся к шагу 1.

Можно доказать, что оптимальные значения c^k целевой функции суженной координирующей задачи монотонно возрастают, сходясь к оптимальному значению $\bar{c}$ целевой функции исходной задачи. Поскольку $c(x)$ вогнута, значение $c(x^k)$ по крайней мере столь же велико, как и c^k.

Решение подзадачи. Нахождение X^{T+1}, удовлетворяющего неравенству (11), само по себе может рассматриваться как оптимизационная задача: найти x в S, при котором

$$\text{Максимизируется } c(x) - \sum_{i=1}^{R} A_i(x)\, y_i^k. \tag{12}$$

Для простоты изложения в течение всей оставшейся части настоящего раздела будем полагать, что максимум (12) является конечным. (Если это предположение не соответствует действительности, возможна простая модификация метода.) Можно убедиться в том, что (12) — вогнутая функция, поскольку $y_i^k \geqslant 0$ для всех $i \geqslant R$. Если в группу R включена вся система ограничений исходной задачи, то (12) решается как задача на безусловный максимум.

Если X^{T+1} — решение (12), то можно утверждать, что

$$\bar{c} \leqslant c^k + \left[c\left(X^{T+1}\right) - \sum_{i=1}^{R} A_i\left(X^{T+1}\right) y_i^k - y_{R+1}^k\right], \tag{13}$$

где, как и выше, c^k — максимальное значение целевой функции суженной координирующей задачи. Далее, расчеты алгоритма можно прекратить на итерации $k+1$, если $w_{T+1}=1$ является оптимальным значением; объясните, почему это так. Однако следует отметить, что значение X^{T+1} обычно не является допустимым для всех R ограничений; соответственно $w_{T+1} = 1$ не будет допустимым в суженной координирующей задаче. Таким образом, в общем случае алгоритм не обеспечивает сходимости за конечное число итераций; расчеты можно прекратить, как только текущее значение $c\left(x^k\right)$ окажется достаточно близким к наименьшему из полученных значений выражения, стоящего в правой части (13).

Заметим, что выражение (12) имеет такой же вид, как функция Лагранжа (21) в разд. 15.7. В самом деле, если R включает все $m+n$ ограничений, обобщенный алгоритм программирования можно рассматривать как систематизированный поиск седловой точки функции Лагранжа. Говоря нестрого, при использовании этого алгоритма попеременно отыскивается улучшенная проверяемая точка при заданных проверяемых значениях множителей Лагранжа и пересматриваются значения множителей Лагранжа в соответствии с текущим проверяемым решением.

Сепарабельная задача. Данный алгоритм весьма эффективен применительно к моделям с сепарабельной целевой функцией и сепарабельными ограничениями:

$$\text{Максимизировать } \sum_{j=1}^{n} c_j\left(x_j\right) \tag{14}$$

$$\text{при условиях } \sum_{j=1}^{n} a_{ij}\left(x_j\right) - b_i \leqslant 0, \quad i=1,\ 2,\ \ldots,\quad m \equiv R, \tag{15}$$

где S теперь включает только условия неотрицательности переменных. Здесь каждая $c_j\left(x_j\right)$ вогнута и каждая $a_{ij}\left(x_j\right)$ выпукла.

В этом случае подзадача (12) распадается на n отдельных оптимизационных задач:

$$\underset{x_j \geqslant 0}{\text{Максимизировать}} \sum_{j=1}^{n} c_j\left(x_j\right) - \sum_{i=1}^{m} a_{ij}\left(x_j\right) y_i^k, \quad j=1,\ 2,\ \ldots,\ n. \tag{16}$$

В некоторых практических случаях вид функций (16) может оказаться настолько простым, что формулу вычисления X_j^{T+1} удается получить, приравняв производную нулю. В худшем случае для нахождения X_j^{T+1} применимы методы поиска в одномерном пространстве, как, например, в случае, приведенном в разд. 14.3.

Существует более усложненный вариант алгоритма решения сепарабельных задач, при котором сходимость обычно ускоряется, однако за счет увеличения размерности суженной координирующей задачи. По существу, это та же постановка, которая была приведена для задач сепарабельного программирования в разд. 14.7. Веса вводятся для *каждого* x_j, и для каждого набора весов включаются суммы весов, аналогичные (5). Подзадача (16) остается без изменения, однако применяется критерий (11) и двойственная переменная используется для построения суммирующего уравнения весов, определяющих x_j.

Примеры. Задачи P и Q из разд. 15.1 и 15.5 удовлетворяют условиям сепарабельности (14) и (15). Выпишите задачи P и Q, представив в явном виде все $c_j(x_j)$ и $a_{1j}(x_j)$. Проверьте, что в задаче P должны решаться $n = 3$ отдельных подзадач, имеющих вид.

$$\begin{aligned}
&\text{Максимизировать } 6x_1 - 3x_1^2 - x_1 y_1^k \quad \text{для } x_1 \geqslant 0,\\
&\text{Максимизировать } 4x_2 - 2x_2^2 - 2x_2 y_1^k \quad \text{для } x_2 \geqslant 0,\\
&\text{Максимизировать } 2x_3 - \frac{1}{3} x_3^2 - x_3 y_2^k \quad \text{для } x_3 \geqslant 0.
\end{aligned}$$

Решения этих задач можно получить, приравняв производную каждого выражения из (17) нулю; получим

$$X_1^{T+1} = 1 - \frac{y_1^k}{6}, \quad X_2^{T+1} = 1 - \frac{y_1^k}{2}, \quad X_3^{T+1} = 3 - \frac{3y_1^k}{2}, \qquad (18)$$

если только y_1^k достаточно мало для того, чтобы эти выражения были неотрицательными. Если же какое-либо из выражений (18), для некоторого j, принимает отрицательное значение, то положим $X_j^{T+1} = 0$.

Аналогичный анализ задачи Q позволяет получить вместо (18) следующие результаты:

$$X_1^{T+1} = 1 - \frac{1}{6y_1^k}, \quad X_2^{T+1} = 1 - \frac{1}{2y_1^k}, \quad X_3^{T+1} = 3 - \frac{3}{2y_1^k}. \qquad (19)$$

Результаты применения алгоритма приведены в таблицах рис. 15.10 и 15.11. Отметим (рис. 15.10), что при $k = 2$ проверяемое решение x^2 оптимально, хотя для подтверждения этого обстоятельства потребовалось еще несколько итераций. Отметим также, что верхняя грань, получаемая на основе (13), *не* уменьшается монотонно. Интересно отметить, что применительно к задаче Q алгоритм дает для каждого k значение y_1^k, обратное y_1^k в таблице рис. 15.10.

t	X^t_1	X^t_2	X^t_3	$Q(X^t)$	$L(X^t)$	k	x^k_1	x^k_2	x^k_3	Оптимальные веса		c^4	$c(x^k)$	Верхняя грань	y^k_1
0	0	0	0	0	0	—	—	—	—	—		—	—	—	—
1	1	1	3	8	6	1	0,667	0,667	2	$w_0=0,33$	$w_1=0,67$	5,33	7,11	7,70	1,33
2	0,778	0,333	1	5,63	2,44	2	0,875	0,625	1,875	$w_1=0,44$	$w_2=0,56$	6,67	7,25	7,26	0,67
3	0,889	0,667	2	7,41	4,22	3	0,875	0,625	1,875	$w_2=0,125$	$w_3=0,875$	7,19	7,25	7,34	1,00
4	0,833	0,5	1,5	6,67	3,33	4	0,675	0,625	1,875	$w_3=0,75$	$w_4=0,25$	7,22	7,25	7,26	0,83
5	0,861	0,583	1,75	7,07	3,78	5	0,875	0,625	1,875	$w_3=0,5$	$w_3=0,5$	7,24	7,25	7,25	0,75
6	0,875	0,625	1,875	7,25	4,00	6	0,875	0,625	1,875	$w_3=1$		7,25	7,25	—	—
Оптимальные значения	—	—	—	—	—	—	0,875	0,625	1,875	—		—	7,25	—	—

$$Q(x) \equiv 6x_1+4x_2+2x_3-3x_1^2-2x_1^2-\frac{1}{3}\times\frac{2}{3}$$

$$L(x) \equiv x_1+2x_2+x_3$$

Рис. 15.10. Обобщенный алгоритм программирования — решение задачи P.

t	k	x_1^k	x_2^k	x_3^k	Оптимальные веса	$Q(x^k)$	$c^k = c(x^k)$	Верхняя грань	y_1^k
0	—	—	—	—	—	—	—	—	—
1	1	0,906	0,906	2,719	$w_0=0,91$; $w_1=0,9$	−7,93	−5,44	−3,66	0,75
2	2	0,933	0,789	2,367	$w_1=0,68$; $w_2=0,32$	−7,76	−4,88	−3,99	1,50
3	3	0,879	0,637	1,911	$w_2=0,09$; $w_3=0,91$	−7,30	−4,06	−3,91	1,00
4	4	0,877	0,631	1,894	$w_3=0,79$; $w_4=0,21$	−7,27	−4,03	−3,99	1,20
5	5	0,876	0,627	1,882	$w_3=0,53$; $w_5=0,47$	−7,26	−4,01	−3,99	1,33
6	6	0,875	0,625	1,875	$w_6=1$	−7,25	−4	—	—
ОЗ	—	0,875	0,625	1,875	—	−7,25	−4	—	—

$$Q(x) \equiv -6x_1-4x_2-2x_3+3x_1^2+2x_2^2+\frac{1}{3}x_3^2$$

ОЗ — оптимальные значения.

Рис. 15.11. Обобщенный алгоритм программирования — решение задачи Q.

В результате, согласно (18) и (19), алгоритм приводит к совершенно такой же последовательности X^t, как и в задаче P. (Следовательно, эти X^t не повторяются в таблице рис. 15.11.) Однако в задаче Q оптимальное решение x^k не удается отыскать до итерации $k = 6$.

15.11. ДЕКОМПОЗИЦИЯ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ

Важным частным случаем сепарабельной модели (14) — (15), рассмотренной в предыдущем разделе, является обычная задача линейного программирования.

В этом случае разделим систему m ограничений на суженную координирующую задачу, включающую R ($R < m$) линейных ограничений, и подзадачу, куда входят остальные ($m - R$) линейных ограничений, а также условия неотрицательности переменных. Такая декомпозиция (разложение) задачи нередко целесообразна по двум причинам. Во-первых, размерность задачи, определяемая числом ограничений m, может оказаться слишком большой для той ЭВМ, на которой выполняются расчеты; расчленение задачи на меньшие части снижает требования к мощности ЭВМ. Во-вторых, $m - R$ ограничений, входящие в S, могут иметь специфическую структуру, особенности которой удобно использовать при вычислениях, необходимых для решения подзадачи; например, они могут представлять собой уравнения баланса потоков по вершинам в сетевой задаче. Поскольку имеются разные способы разложения задачи линейного программирования, приводимое ниже изложение является скорее пояснением, чем исчерпывающим рассмотрением подхода.

Пусть модель имеет следующий вид:

$$\text{Максимизировать } \sum_{j=1}^{n} c_j x_j \tag{1}$$

при условиях

$$A_i(x) \equiv \sum_{j=1}^{n} a_{ij}x_j - b_i \leqslant 0, \quad i = 1, 2, \ldots, R, \tag{2}$$

где x должны также удовлетворять условиям

$$\sum_{j=1}^{n} a_{ij}x_j - b_j \leqslant 0, \quad i = R+1, \ldots, m, \tag{3}$$

$$\text{и все } x_j \geqslant 0.$$

Суженная координирующая задача является точно такой же, как задача (3) — (5) из предыдущего раздела; следовательно, она включает $R+1$ линейных ограничений. Соответствующая подзадача заключается в нахождении значений x_j, удовлетворяющих (3) и максимизирующих

$$\sum_{j=1}^{n} c_j x_j - \sum_{i=1}^{R} \Big(\sum_{j=1}^{n} a_{ij}x_j\Big)\cdot y_i^k = \sum_{j=1}^{n} \Big(c_j - \sum_{i=1}^{R} a_{ij}y_i^k\Big)\cdot x_j, \tag{4}$$

где y_i^k — двойственные переменные, связанные с решением суженной координирующей задачи на итерации k. Отметим, что выражение (4) есть линейная функция x_j и, следовательно, подзадача представляет собой обычную задачу линейного программирования. Если существует решение $x_j = X_j$, такое, что

$$y_{k+1}^k < \sum_{j=1}^{n} \Big(c_j - \sum_{i=1}^{R} a_{ij}y_i^k\Big)\cdot X_j + \sum_{i=1}^{R} b_i y_i^k, \tag{5}$$

то итеративный процесс продолжается, причем X вводится в суженную координирующую задачу в качестве новой точки сетки.

Применительно к линейным задачам обобщенный алгоритм нелинейного программирования часто называют **алгоритмом декомпозиции, алгоритмом разложения** или **блочным алгоритмом.** Он обеспечивает сходимость к оптимальному решению за конечное число итераций. Более того, в конце шага 1 из суженной координирующей задачи можно исключить те точки сетки, которые соответствуют небазисным переменным. Следовательно, в начале любой итерации координирующая задача не должна включать более чем $R+1$ переменных (весов). Если подзадача решается симплексным алгоритмом, то для каждой точки X^t сетки положительные значения принимают не более $m-R$ переменных. Однако, поскольку оптимальное решение, соответствующее базисным переменным суженной координирующей задачи, является средневзвешенным этих X^t, а именно $\sum_{t=0}^{T} X_j^t w_t$ для каждого j, в такое решение может входить *больше* чем m переменных x_j с положительными значениями (если только задача в целом не имеет единственного оптимального решения). Это важное различие между симплексным методом при его использовании для решения задачи в целом, с одной стороны, и алгоритмом декомпозиции — с другой.

Приводимый ниже пример позволяет продемонстрировать применение метода, а также показать небольшие модификации подхода. Рассмотрим задачу

$$\text{Максимизировать } x_1 + x_2 + x_3 + x_4 \tag{6}$$

при условиях

$$\begin{aligned} x_1 + 2x_2 + 2x_3 \qquad\quad &\leqslant 3 \quad \text{(строка 1)},\\ -x_1 \qquad + \ x_3 + 2x_4 &\leqslant 2 \quad \text{(строка 2)},\\ x_3 + \ x_4 &\leqslant 3 \quad \text{(строка 3)},\\ x_3 - \ x_4 &\leqslant 1 \quad \text{(строка 4)}\\ \text{и все } x_j &\geqslant 0. \end{aligned} \tag{7}$$

Поскольку x_1 и x_2 входят только в два первых ограничения, мы включим их *в явном виде* в следующую суженную координирующую задачу:

$$\text{Максимизировать } x_1 + x_2 + \sum_{t=0}^{T} [X_3^t + X_4^t]\, w_t \tag{8}$$

при условиях

$$\begin{aligned} x_1 + 2x_2 + \sum_{t=0}^{T} [2X_3^t]\, w_t &\leqslant 3,\\ -x_1 \qquad + \sum_{t=0}^{T} [X_3^t + 2X_4^t]\, w_t &\leqslant 2,\\ \sum_{t=0}^{T} w_t &= 1,\\ x_1,\ x_2 \text{ и все } w_t &\geqslant 0. \end{aligned} \tag{9}$$

Отметим, что при данной постановке постоянные $b_1 = 3$ и $b_2 = 2$ включены в правую часть (9).

Пусть подзадача включает переменные x_3 и x_4, два последних линейных ограничения исходной задачи и условия неотрицательности переменных. Следовательно, суммирование в (4) выполняется только для x_3 и x_4. Позадача имеет вид

$$\begin{aligned} \text{Максимизировать } [1 - (2y_1^k + 1y_2^k)]\, x_3 + \\ + [1 - (0y_1^k + 2y_2^k)]\, x_4 \end{aligned} \tag{10}$$

при условии выполнения ограничений определяемых строками 3 и 4 системы (7), а также $x_3 \geqslant 0$, $x_4 \geqslant 0$.

Область допустимых значений x_3 и x_4 для подзадачи показана на рис. 15.12; заметим, что имеются четыре вершины, обозначенные A, B, C и D, которые отображают возможные базисные решения подзадачи на каждой итерации. Если существует такое решение подзадачи X_3 и X_4, что

$$y_{M+1}^k < [1 - (2y_1^k + 1y_2^k)]\, X_3 + [1 - (0y_1^k + 2y_2^k)]\, X_4, \tag{11}$$

то итеративный процесс продолжается. Отметим, что член $\sum_{i=1}^{R} b_i y_i^k$ *не* появляется в правой части (11), поскольку константы b_1 и b_2 перенесены в правую часть суженной координирующей задачи (9).

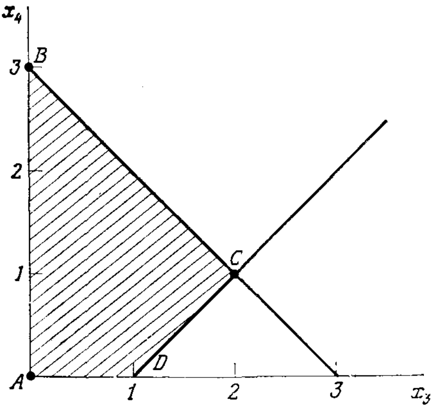


Рис. 15.12. Область допустимых решений для подзадачи.

Пусть исходное проверяемое решение имеет вид ($X_3^0 = 0$, $X_4^0 = 0$), что на рис. 15.12 соответствует точке A. Оптимальным решением суженной координирующей задачи является ($x_1^1 = 3$, $w_0^1 = 1$), а соответствующими двойственными переменными являются ($y_1^1 = 1$, $y_2^1 = 0$, $y_3^1 = 0$).

Получаем следующую целевую функцию подзадачи:

$$\text{Максимизировать } -x_3 + x_4. \tag{12}$$

Используя рис. 15.12, можно убедиться, что оптимальному решению этой подзадачи отвечает вершина B, которой соответствует ($X_3^1 = 0$, $X_4^1 = 3$). Поскольку для этой точки значение целевой функции (12) равно 3, а $y_3^1 = 0$ для текущего решения, точку B необходимо ввести в суженную координирующую задачу. Определите, как введение этой точки изменит (8) и (9).

Оптимальное решение очередной суженной координирующей задачи имеет вид $\left(x_1^2 = 3,\ w_0^2 = \frac{1}{6},\ w_1^2 = \frac{5}{6}\right)$, а соответствующие двойственные переменные равны $\left(y_1^2 = \frac{3}{2},\ y_2^2 = \frac{1}{2},\ y_3^2 = 0\right)$.

Целевая функция подзадачи приобретает вид

$$\text{Максимизировать } -\frac{5}{2}\, x_3. \tag{13}$$

Теперь оптимальным решением становится точка A, которая уже включена в оптимальное решение суженной координирующей задачи. Следовательно, эта точка преобразует (11) в равенство, и текущее проверяемое решение является истинным оптимумом. Соответствующие значения x_3 и x_4 равны

$$\begin{aligned} x_3 &= X_3^0 w_0^2 + X_3^1 x_1^2 = 0\cdot\frac{1}{6} + 0\cdot\frac{5}{6} = 0, \\ x_4 &= X_4^0 w_0^2 + X_4^1 w_1^2 = 0\cdot\frac{1}{6} + 3\cdot\frac{5}{6} = \frac{5}{2}. \end{aligned} \tag{14}$$

На крупных ЭВМ имеется ряд программ декомпозиции, предназначенных для решения задач линейного программирования. Обычно применение таких методов требует значительно большего объема вычислений, чем использование для решения той же задачи симплексного метода. Следовательно, подобный подход эффективен только в тех случаях, когда исходная задача слишком велика для того, чтобы ее можно было непосредственно решить с помощью симплексного алгоритма, учитывая ограниченные возможности ЭВМ, или же тогда, когда удается использовать специфические особенности структуры ограничений подзадачи.

КОНТРОЛЬНЫЕ УПРАЖНЕНИЯ

1. Рассмотрите задачу максимизации целевой функции $-(x_1 - 1)^2 - (x_2 - 1)^2$ при условиях $x_1 + x_2 \leqslant 1$, $x_1 \geqslant 0$ и $x_2 \geqslant 0$.

а) Начертите область допустимых решений, покажите оптимальное решение, а также оптимальное решение задачи на безусловный максимум.

б) Пусть проверяемым решением является точка $x_1^0 = 1$, $x_2^0 = 0$. Покажите на рисунке направление скорейшего подъема. Имеются ли в этом направлении от проверяемой точки допустимые точки?

в) Покажите графически, каким образом отыскивается оптимальное решение при методе выпуклых комбинаций, изложенном в разд. 15.2. (Укажите значения x^k, $\bar{z}$, а также линейной целевой функции для $k = 1, 2, 3, 4$.)

г) Проделайте то же для исходной проверяемой точки $x_1^0 = \frac{1}{4}$, $x_2^0 = \frac{3}{4}$.

2. Рассмотрите задачу максимизации целевой функции $-(x_1 - 3)^2 - (x_2 - 3)^2$ при условиях $0 \leqslant x_1 \leqslant 6$ и $0 \leqslant x_2 \leqslant 2$.

а) Начертите область допустимых решений, укажите оптимальное решение, а также оптимальное решение задачи на безусловный их максимум.

б) Покажите на рисунке, каким образом находится оптимальное решение при методе выпуклых комбинаций, изложенном в разд. 15.2, начав с исходной точки $x_1^0 = x_2^0 = 0$ (для итераций k, $k = 1, 2, 3, 4$ укажите значения x^k и соответствующих $\bar{z}$). Покажите, каким образом $c(\bar{z})$ может оказаться меньше, чем $c(x^k)$, а также почему должна существовать выпуклая комбинация точек x^k и $\bar{z}$, а именно x^{k+1}, такая, что $c(x^{k+1}) > c(x^k)$. Объясните, почему фактически используемые направления не являются направлениями скорейшего подъема.

в) Какие возникнут изменения, если для исходной точки $x_1^0 = 3$, $x_2^0 = 0$? $x_1^0 = 0$, $x_2^0 = 1$? $x_1^0 = 0$, $x_2^0 = 2$?

3. Рассмотрите задачу из упражнения 2, припяв $0 \leqslant x_2 \leqslant 4$.

а) Начертите область допустимых решений и укажите оптимальное решение.

б) Покажите на рисунке, каким образом находится оптимальное решение при методе выпуклых комбинаций, изложенном в разд. 15.2, начав с исходной точки $x_1^0 = x_2^0 = 0$ (для итераций k, $k = 1, 2, 3, 4$ укажите значения x^k и соответствующих $\bar{z}$).

4. а) Дайте алгебраическое доказательство сделанного в разд. 15.2 утверждения о том, что выпуклая комбинация двух допустимых решений x^k и $\bar{z}$ также является допустимой при любом значении t в интервале $0 < t \leqslant 1$.

б) Покажите, что если использовать направления, определенные в соответствии с (6), то (4) можно записать в виде (7).

в) Преобразуйте операции суммирования в (3) таким образом, чтобы представить это неравенство с помощью характеристик направлений. Дайте истолкование полученного неравенства.

г) Объясните, в какой своей части метод выпуклых комбинаций основывается на допущениях II), IV) и V) из разд. 15.1.

5. а) Выведите первое неравенство в соотношении (8) из разд. 15.2 для верхней грани.

б) Выведите второе неравенство соотношения (8).

в) Объясните, почему из выполнения неравенств (8) следует, что если для $\bar{z}$ обеспечивается равенство сумм в левой и правой частях (3), то соответствующая величина x^k является истинным оптимумом.

6. Рассмотрите задачу максимизации $(x_1 - 3)^2 + (x_2 - 3)^2$ при условиях

$$x_1 + x_2 \leqslant 7, \quad -x_1 + 3x_2 \leqslant 9, \quad 3x_1 - x_2 \leqslant 9,$$
$$x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

а) Начертите область допустимых решений и укажите оптимальное решение.

б) Предположим, что применяется метод выпуклых комбинаций и что в качестве исходной точки принята точка $x_1^0 = 3$, $x_2^0 = 4$. Определите «поведение» решений.

в) Выполните вариант б), начав с исходной точки $x_1^0 = x_2^0 = 3\frac{1}{2}$.

г) Выполните вариант б), начав с исходной точки $x_1^0 = x_2^0 = 3$.

д) Выполните вариант б), начав с исходной точки $x_1^0 = 3$, $x_2^0 = 0$.

7. Рассмотрите задачу максимизации целевой функции $18x_1 + 18x_2 - 5x_1^2 - 8x_1x_2 - 5x_2^2$ при условиях $x_1 + x_2 \leqslant 1$, $x_1 \geqslant 0$ и $x_2 \geqslant 0$. В каждом из приведенных ниже вариантов постановки задачи выполните одну итерацию метода выпуклых комбинаций для нахождения $\bar{z}$, соответствующего указанной точке x^0. Затем вычислите $c(x^0)$, $c(\bar{z})$ и верхнюю грань (8) из разд. 15.2.

а) $x_1^0 = x_2^0 = 0$; г) $x_1^0 = \frac{1}{4}$; $x_2^0 = \frac{1}{4}$;

б) $x_1^0 = 1$, $x_2^0 = 0$; д) $x_1^0 = \frac{1}{3}$; $x_2^0 = \frac{2}{3}$;

в) $x_1^0 = \frac{1}{4}$, $x_2^0 = \frac{3}{4}$; е) $x_1^0 = x_2^0 = \frac{1}{2}$.

Является ли это решение оптимальным? Объясните, почему это так.

8. Рассмотрите использование метода выпуклых комбинаций для решения примера (9) — (12) из разд. 15.2. Выполните со всеми подробностями расчеты для $k = 0, 1, \ldots, 5$ и проверьте правильность результатов, приведенных в таблице рис. 15.1.

9. Рассмотрите использование симплексного метода вогнутого программирования для решения примера (10) — (17) из разд. 15.2. Выполните со всеми подробностями расчеты для $k = 1, 2, \ldots, 5$ и проверьте правильность результатов, приведенных в таблице рис. 15.3.

10. В каждом из приведенных ниже вариантов постановки задачи используйте симплексный метод вогнутого программирования, описанный в разд. 15.2, для решения задачи P со внесенными в нее изменениями условий, которые указываются в соответствующих пунктах (эти изменения вносились также в упражнение 23 из гл. 14). Прекратите итеративный процесс после $k = 4$ (если сходимость не будет достигнута ранее).

а) Правая часть (5) равна 8 (вместо 4).
б) Правая часть (5) равна 3 (вместо 4).
в) Правая часть (5) равна 2 (вместо 4).
г) Добавьте ограничение $x_1 + x_2 \leqslant 1$.
д) Добавьте ограничение $x_2 + x_3 \leqslant 2$.
е) Добавьте ограничение $-2x_1 + 2x_3 \leqslant 1$.
ж) Целевая функция имеет вид $6x_1 + 4x_2 + 2x_3 - \frac{1}{3}x_1^2 - x_2^2 - 3x_3^2$.
з) Целевая функция имеет вид $2x_1 + 6x_2 + 4x_3 - 3x_1^2 - 2x_2^2 - \frac{1}{3}x_3^2$.
и) Добавьте к целевой функции слагаемые $-x_1x_2 - 2x_2x_3$.

11. В каждом из приведенных ниже вариантов постановки задачи примените симплексный метод вогнутого программирования, описанный в разд. 15.3. Прекратите итеративный процесс после $k = 4$ (если сходимость не будет достигнута ранее). Нанесите результаты расчетов для каждой итерации на рисунок области допустимых решений.

а) Упражнение 1. б) Упражнение 2. в) Упражнение 3. г) Упражнение 7. д) Максимизировать целевую функцию $10x_2 - 6x_1^2 - 19x_2^2 + 16x_1x_2$ при условиях

$$3x_1 + x_2 \leqslant 3, \; x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

е) Измените на обратные знаки коэффициентов целевой функции варианта д) и примите в качестве исходного базисного решения $x_1 = 3$. Можете ли вы получить глобальный оптимум? Дайте обоснование вашему ответу.

12. В разд. 15.3 утверждалось, что в случае использования метода выпуклых комбинаций на некоторой итерации все направления могут быть ненулевыми, тогда как при использовании симплексного метода вогнутого программирования ненулевыми могут быть не более чем $m + 1$ направлений. Дайте подробное обоснование этому утверждению.

13. Рассмотрите метод секущих плоскостей, описанный в разд. 15.4, применительно к решению следующей задачи:

Максимизировать $-x^2 + 4x$

при условиях $0 \leqslant x \leqslant 5$.

а) Найдите оптимальное значение x. Начертите график целевой функции при $0 \leqslant x \leqslant 5$.

б) Примите $x^0 = 0$. Продемонстрируйте ход выполнения операций алгоритма для $k = 1, 2, 3, 4$, начертив область допустимых решений для переменных x и x_0 и показав секущую плоскость, добавляемую на каждой итерации, а также значения x^k и x_0^k. С помощью этого рисунка объясните, почему данный алгоритм называется методом секущих плоскостей.

в) Измените на обратные знаки коэффициентов целевой функции варианта а) и объясните, почему данный метод неприменим, если максимизируемая целевая функция выпукла.

14. Рассмотрите функцию $a(x) \equiv x_1^2 - x_2$.

а) Начертите линию уровня $a(x) = 0$.

б) Примите $x_1 = 2$ и $x_2 = 4$ и напишите в развернутом виде неравенство (4) из разд. 15.5. Проверьте, что функция, полученная путем приравнивания нулю выражения под знаком суммирования в правой части (4), имеет вид $4(y_1 - 2)$ и $(y_2 - 4) = 0$ и является линейной опорой линии уровня варианта а) в указанной точке.

в) Выполните аналогичный анализ для $a(x) \equiv 9x_1^2 - x_2$ и точки

$$x_1 = \frac{2}{3}, \quad x_2 = 4.$$

15. Для каждого из приведенных ниже вариантов постановки задачи определите, удовлетворяет ли функция $a(x)$ допущениям I) и II) и характеристике системы ограничений, приведенным в разд. 15.5. Определите допустимые значения x. Проверьте, действительно ли $da(z)/dx \neq 0$ в точке z, где $a(z) = 0$.

а) $x^2 - 4 \leqslant 0$;	е) $(x^2 - 4)^2 \leqslant 0$;
б) $x^2 + 4 \leqslant 0$;	ж) $-(x^2 - 4)^3 \leqslant 0$;
в) $-x^2 - 4 \leqslant 0$;	з) $(x^2 - 4)^3 \leqslant 0$;
г) $-x^2 + 4 \leqslant 0$;	и) $-(4 - x^2)^3 \leqslant 0$;
д) $-(x^2 - 4)^2 \leqslant 0$;	к) $(4 - x^2)^3 \leqslant 0$.

16. Для каждого из приведенных ниже вариантов постановки задачи определите, удовлетворяет ли функция $a(x)$ допущениям I) и II) и характеристике системы ограничений, приведенным в разд. 15.5. Найдите допустимые значения x_1 и x_2. Проверьте, действительно ли $\partial a(z)/\partial x_j \neq 0$ по крайней мере для одного j в точке z, где $a(z) = 0$.

а) $x_1 + x_2 - 1 \leqslant 0$; г) $(x_1 + x_2 - 1)^2 \leqslant 0$;

б) $-x_1 - x_2 + 1 \leqslant 0$; д) $-(x_1 + x_2 - 1)^3 \leqslant 0$;

в) $-(x_1 + x_2 - 1)^2 \leqslant 0$; е) $(x_1 + x_2 - 1)^3 \leqslant 0$.

17. Для каждого из приведенных ниже вариантов постановки задачи начертите область допустимых решений. (Примите $x_1 \geqslant 0$ и $x_2 \geqslant 0$.) Проверьте, выполняется ли характеристика системы ограничений.

а) $x_1^2 + (x_2 - 1)^2 - 1 \leqslant 0$ и $(x_1 - 1)^2 + x_2^2 - 1 \leqslant 0$;

б) $x_1^2 + (x_2 - 1)^2 - 1 \leqslant 0$ и $x_1^2 + x_2^2 - 1 \leqslant 0$;

в) $x_1^2 + (x_2 - 2)^2 - 1 \leqslant 0$ и $x_1^2 + x_2^2 - 1 \leqslant 0$;

г) $x_1^2 + x_2^2 - 4 \leqslant 0$, $(x_1 - 3)^2 + x_2^2 - 4 \leqslant 0$ и

$(x_1 - \frac{3}{2})^2 + (x_2 - 3)^2 - 4 \leqslant 0$;

д) $x_1^2 + x_2^2 - 1 \leqslant 0$ и $x_1 + x_2 - \frac{1}{2} \leqslant 0$;

е) $x_1^2 + x_2^2 - 1 \leqslant 0$ и $-x_1 - x_2 + \frac{1}{2} \leqslant 0$;

ж) $x_1^2 + x_2^2 - 8 \leqslant 0$ и $x_1 + x_2 - 3 \leqslant 0$;

з) $x_1^2 + x_2^2 - 8 \leqslant 0$, $-x_1 - x_2 + 3 \leqslant 0$ $x_1 + x_2 - 3{,}5 \leqslant 0$;

и) $x_1^2 + x_2^2 - 8 \leqslant 0$ и $-x_1 - x_2 + 4 \leqslant 0$;

к) $-e^{2x_1+3x_2} + 10 \leqslant 0$ и $2x_1 + 3x_2 - 10 \leqslant 0$;

л) $-x_1x_2 + 1 \leqslant 0$ и $x_1 + x_2 - 2 \leqslant 0$.

м) Измените на обратный знак первого неравенства в варианте а). Покажите, почему область допустимых решений перестает быть выпуклой.

18. Приведите промежуточные алгебраические преобразования, показывающие, что если каждое из ограничений $a_i(x)$ представляет собой выпуклую функцию, удовлетворяющую условию (3) из разд. 15.5, то область допустимых решений является выпуклой.

19. В разд. 15.5 указано пять способов проверки того, является ли выпуклой некоторая функция $a_i(x)$. Продемонстрируйте справедливость этих выводов.

20. Для каждого из приведенных ниже вариантов постановки задачи установите, является ли выпуклой указанная функция. Если функция выпукла, обоснуйте ваш вывод, основываясь на одном из свойств А) — Д), перечисленных в разд. 15.5. Если же функция не выпукла, приведите численный пример, свидетельствующий о том, что определение выпуклой функции (3) не удовлетворяется.

а) $e^{\log x}$ при $x > 0$;

б) $-x_1 - x_2 + x_1^4 + x_2^6$ при любых значениях x_1 и x_2;

в) $x_1 x_2$ при $x_1 \geqslant 0$ и $x_2 \geqslant 0$;

г) $-\log(2x_1 + 3x_2 + 4) + 5(x_1 - 6x_2)^2$ при $x_1 \geqslant 0$ и $x_2 \geqslant 0$;

д) $(x_1 - 2)^3 - (x_2 - 2)^3$ при $x_1 \geqslant 0$ и $x_2 \geqslant 0$;

е) $(2 - x_1)^3 + (2 - x_2)^3$ при $x_1 \geqslant 0$ и $x_2 \geqslant 0$.

21. Рассмотрите оптимизационную задачу нелинейного программирования, удовлетворяющую допущениям, сделанным в разд. 15.5. Пусть проверяемая точка расположена на границе области допустимых решений, однако не является истинным оптимумом. Объясните, почему должно существовать направление улучшения, идущее внутрь области допустимых решений. (*Указание*: для обоснования ваших соображений используйте рисунок, аналогичный рис. 15.6.)

22. а) Если точка x^k находится во внутренней части области, определяемой ограничением i, объясните, почему точка $(x^k + td)$ удовлетворяет этому ограничению при любом наборе направлений d, если только $t > 0$ достаточно мало.

б) Рассмотрите точку x^k на рис. 15.6. Начертите линейную опору (касательную гиперплоскость) нелинейной функции в точке x^k. Укажите, при каких направлениях правая часть (3) *строго* отрицательна. Должно ли такое направление пересекать границу, определяемую нелинейной функцией? Объясните, почему для любого такого направления имеется точка вблизи x^k, лежащая во внутренней части области допустимых решений.

23. а) Объясните, почему оптимизационная модель (4) — (6) из разд. 15.6 может быть использована для нахождения направлений, удовлетворяющих критерию проверки (1) и придающих правым частям (3) строго отрицательные значения, если проверяемая точка x^k не является относительной стационарной точкой.

б) Покажите, что условия неотрицательности переменных, включенные в $A_i(x)$, при $x_0 \geqslant 0$ обеспечивают с учетом (6) наличие нижней грани d_j, равной $-x_j^k$.

24. а) Проверьте, что оптимизационная модель (4) — (6) из разд. 15.6 применительно к задаче Q, приведенной в конце разд. 15.5, принимает вид (9) — (11).

б) Запишите в развернутом виде и решите оптимизационную задачу для $k = 0$, начав с проверяемой точки $x_1^0 = x_2^0 = 1$, $x_3^0 = 3$, принятой в качестве исходной в таблице рис. 15.7. Убедитесь в том, что после итерации получена точка x^1.

в) Напишите в развернутом виде и решите оптимизационную задачу, приняв в качестве x^k истинно оптимальное решение (приведенное внизу таблицы рис. 15.7).

25. а) Запишите задачу P, представленную в виде (4) — (5) из разд. 15.1, как оптимизационную задачу (4) — (8) из разд. 15.6.

б) Сравните модель варианта а) с постановкой, использованной для демонстрации метода выпуклых комбинаций в разд. 15.2. Укажите влияние выбора того или иного значения b в (8).

в) Покажите, каким образом преобразовать полученный вами ответ на вопросы варианта а) в модель, включающую два ограничения и три переменных с двусторонними ограничениями.

г) В варианте а) примите $b = 1$, а в качестве исходной точки $x_1^0 = x_2^0 = x_3^0 = 0$. Завершите расчеты при $k = 0$ и получите новую проверяемую точку для $k = 1$. Выполните то же для $b = 5$.

д) Запишите модель варианта а) в развернутом виде и решите возникающую оптимизационную задачу для случая, когда проверяемая точка является истинным оптимумом.

26. Примите $m = 3$, $n = 5$ и $A_{3+j}(x) \equiv -x_j$ для $j = 1, 2, \ldots, 5$. Запишите в развернутом виде модель линейного программирования (1) — (4) из разд. 15.7 и двойственную ей модель (5) — (8). Напишите также в развернутом виде условия оптимальности (9) — (11).

27. Для каждого из приведенных ниже вариантов постановки задачи запишите в развернутом виде условия оптимальности (9) — (11).

а) Задача P, представленная в виде (4) — (5) из разд. 15.1. Проверьте правильность (16) и (17).

б) Задача Q, приведенная в конце разд. 15.5. Найдите также для оптимального $\bar{x}$ значения $\bar{u}_i$, удовлетворяющие этим условиям.

в) Покажите для задач P и Q, что из предположений $x_1 > 0$ и $x_2 = x_3 = 0$ следует отсутствие $u_i \geqslant 0$, удовлетворяющих условиям оптимальности.

28. Рассмотрите условия оптимальности из разд. 15.7.

а) Проверьте, что (10) можно получить, разделив (5) на $\bar{y}_0 > 0$. Дайте обоснование недостающим звеньям доказательства.

б) Выведите условия (11) из уравнений (7). Дайте обоснование вашему подходу.

в) Покажите, почему из (10) и (11) следует (12), (13) и (14).

г) Предположите, что $c(x)$ и все $a_i(x)$ линейны. Покажите, что (9), (12) — (14) сводятся к результатам, содержащимся в теореме о дополнительной нежесткости из разд. 5.4.

29. Предположим, что требуется проверить, является ли проверяемая допустимая точка x^k оптимальным решением задачи нелинейного программирования. Обсудите вопрос возможности применения условий оптимальности из разд. 15.7 вместо модели линейного программирования (1) — (4). Опишите связь между двумя подходами.

30. В каждом из приведенных ниже вариантов постановки задачи найдите, путем анализа, оптимальное решение. Проверьте, удовлетворяются ли условия оптимальности из разд. 15.7, т. е. найдите значения $\bar{u}_i \geqslant 0$.

а) Максимизируйте $-(x-4)^2$ при $1 \leqslant x \leqslant 3$. Удовлетворяются ли условия оптимальности при $x = 1$? при $x = 3$? Поясните.

б) Максимизируйте $(x-4)^2$ при $1 \leqslant x \leqslant 6$. Удовлетворяются ли условия оптимальности при $x = 6$? при $x = 4$? Поясните.

в) Максимизируйте x при условиях $(x^2-4)^2 - 25 \leqslant 0$.

г) Максимизируйте x при условиях $(x^2-4)^2 \leqslant 0$. Объясните разницу в ответах на вопросы вариантов в) и г).

д) Покажите, что ответы для вариантов а) и в) соответствуют седловой точке функции Лагранжа (21).

е) Покажите, что функции Лагранжа для вариантов б) и г) не имеют седловых точек.

ж) Максимизируйте $(x-4)^3$ при $1 \leqslant x \leqslant 6$. Удовлетворяются ли условия оптимальности при $x = 1$? при $x = 4$? при $x = 6$? Поясните.

31. Рассмотрите функцию Лагранжа $L(x, u)$, определяемую выражением (21) из разд. 15.7. Запишите это выражение в развернутом виде для случая, когда $c(x)$ и все $a_i(x)$ являются линейными функциями. Обсудите связь между условиями оптимальности Лагранжа для этого случая и теоремой двойственности для задач линейного программирования.

32. В модели квадратичного программирования, приведенной в разд. 15.8, примите $m = 2$ и $n = 3$. Запишите в развернутом виде выражения (1) — (4) и приведите промежуточные преобразования, необходимые для получения (5) — (14).

33. Рассмотрите решение задачи Р с помощью модифицированного симплексного алгоритма, приведенное в разд. 15.8. Выполните во всех деталях численные операции, проверьте результаты всех шагов алгоритма и убедитесь в правильности результатов (18) — (21).

34. В каждом из приведенных ниже вариантов постановки задачи примените модифицированный симплексный алгоритм, описанный в разд. 15.8, для решения задачи Р, условия которой изменены так, как это указано ниже (см. также упражнение 10).

а) Правая часть (5) равна 8 (вместо 4).

б) Правая часть (5) равна 3 (вместо 4).

в) Правая часть (5) равна 2 (вместо 4).

г) Добавьте ограничение $x_1 + x_2 \leqslant 1$.

д) Добавьте ограничение $x_2 + x_3 \leqslant 2$.

е) Добавьте ограничение $-2x_1 + 2x_3 \leqslant 1$.

ж) Целевая функция имеет вид $6x_1 + 4x_2 + 2x_3 - \frac{1}{3}x_1^2 - 2x_2^2 - 3x_3^2$.

з) Целевая функция имеет вид $2x_1 + 6x_2 + 4x_3 - 3x_1^2 - 2x_2^2 - \frac{1}{3}x_3^2$.

и) Добавьте к целевой функции слагаемые $-x_1x_2 - 2x_2x_3$.

35. а) Пусть $a_i(x)$ — выпуклая функция. Покажите, что $[1/a_i(x)]$ — вогнутая функция при x, таких, что $a_i(x) < 0$.

б) Проверьте правильность первого равенства в (4) из разд. 15.9.

в) Дайте детальное обоснование (III) и (5) из разд. 15.9.

36. Рассмотрите задачу

Максимизировать $-x^2 + 4x$

при условии $x \leqslant 1$.

а) Запишите штрафную функцию (2) из разд. 15.9.

б) Выполните дифференцирование в (4) и постройте графическое отображение x, максимизирующего $c(x \mid r)$ при заданном значении r. [*Указание*: можно так преобразовать (4), чтобы представить его в виде равенства между линейной и гиперболической функциями; в свою очередь это равенство графически отображается пересечением двух кривых.] Покажите, как изменяется решение при стремлении r к нулю. Является ли предельное значение x истинным оптимумом?

в) Пусть ограничение имеет вид $x \leqslant 5$. Как влияет это ограничение на решение при заданном r и на предельное решение при r, стремящемся к нулю?

37. Рассмотрите обобщенную модель программирования из разд. 15.10.

а) Объясните, почему суженная координирующая задача имеет базисное допустимое решение и конечное максимальное значение целевой функции c^k.

б) Покажите, что любое допустимое решение, удовлетворяющее условиям (4) и (5), представляет собой точку с координатами (6), которая является допустимым решением исходной задачи.

в) Примем в задаче (3) — (5) $R = 2$ и $T = 3$. Запишите в развернутом виде двойственную задачу (7) — (9).

г) Покажите, что если неравенство (10) соблюдается для всех x в S, то x^k есть истинный оптимум для исходной задачи, как это утверждается в разд. 15.10.

д) Покажите, что целевая функция (12) подзадачи вогнута.

е) Дайте обоснование наличия верхней грани для $\bar{c}$, определяемой (13). [*Указание*: поскольку X^{T+1} может оказаться любым из X^t, где $t = 0, 1, \ldots, T$, неравенство «не больше» в (11) заведомо выполняется.]

ж) Объясните, почему итеративный процесс можно завершить на следующей итерации после того, как достигнуто $w_{T+1} = 1$.

38. а) Рассмотрите решение задач P и Q с помощью обобщенного алгоритма программирования, приведенного в разд. 15.10. Покажите, посредством каких промежуточных преобразований обосновывается получение с помощью этого подхода (17) — (19).

б) Доведите решение задачи P с помощью этого алгоритма (таблица рис. 15.10) до получения x^k для $k = 2$. (Обратите внимание на то, что $X_1^0 = X_2^0 = X_3^0 = 0$.)

в) Напишите в развернутом виде суженную координирующую задачу для задачи P, включив в нее только X^3 и X^5 из таблицы рис. 15.10, а также дополняющую переменную первого ограничения. Проверьте, что оптимальными весами являются $w_3 = w_5 = 0{,}5$, как это показано в таблице рис. 15.10, и завершите следующую (конечную) итерацию.

39. Рассмотрите пример (6) — (7) из разд. 15.11, демонстрирующий применение алгоритма декомпозиции. Проверьте правильность выражений (8) — (11) и выполните детальные расчеты, необходимые для получения результатов (12) — (14).

40. Рассмотрите пример (6) — (7) из разд. 15.11. Замените целевую функцию на $x_1 + x_2 + 9x_3 + 3x_4$ и примите правую часть строки 4 равной 0 (вместо 1). Решите задачу с помощью алгоритма декомпозиции.

41. Рассмотрите задачу фирмы «Мультиконвейер», приведенную в разд. 2.2 и решенную симплексным методом в разд. 4.4. Решите эту задачу алгоритмом декомпозиции, включив в подзадачу первое ограничение ($x_1 + x_2 + x_3 + x_4 \leqslant 15$) и условия неотрицательности переменных.

42. Объясните, как вы понимаете следующие термины:

крупношаговый метод;
постулат регуляризации;
оптимальная допустимая длина шага;
относительная стационарная точка;
выпуклая (средневзвешенная) комбинация;
вогнутая функция (строго вогнутая функция);
выпуклая функция (строго выпуклая функция);
метод первого (второго) порядка;
метод выпуклых комбинаций;
симплексный метод вогнутого программирования;
приведенная (или относительная) частная производная;
метод секущих плоскостей;
метод проецируемых градиентов;
проецирование;
линейная опора (касательная гиперплоскость);
характеристика системы ограничений;
выпуклое множество;
допустимое направление;
метод допустимых направлений;
условия Куна — Таккера;
условия Джона;
функция Лагранжа;
множители Лагранжа;
седловая точка;
взаимодополняющее преобразование матрицы;
метод штрафной функции;
обобщенный алгоритм программирования;
суженная координирующая задача;
подзадача;
алгоритм декомпозиции.

УПРАЖНЕНИЯ НА РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

43. Рассмотрите следующую задачу, которая была приведена также в варианте в) упражнения 24 гл. 14 и соответствует примеру 3 рис. 14.9:

$$\text{Максимизировать } -(x_1-1)^2-(x_2-2)^2+5$$

при условиях

$$x_1+2x_2\leqslant 5,$$
$$16x_1+12x_2\leqslant 37,$$
$$24x_1+4x_2\leqslant 27,$$
$$x_1\geqslant 0 \text{ и } x_2\geqslant 0.$$

В каждом из приведенных ниже вариантов постановки решите задачу указанным методом. Начертите область допустимых решений и нанесите на рисунок результаты, полученные для каждой итерации. (При использовании тех методов, которые не обеспечивают сходимость за конечное число итераций, прекратите расчеты после $k=5$.)

а) Метод выпуклых комбинаций, описанный в разд. 15.2.

б) Симплексный метод выпуклого программирования, описанный в разд. 15.3.

в) Модифицированный симплексный алгоритм, описанный в разд. 15.8.

г) Обобщенный алгоритм программирования для сепарабельных задач, описанный в разд. 15.10.

д) Используйте для получения оптимального решения условия оптимальности, приведенные в разд. 15.7.

44. Ответьте на вопросы, поставленные в упражнении 43, заменив лишь ограничения (они были приведены в варианте б) упражнения 24 гл. 14 и соответствуют примеру 2 рис. 14.9).

$$x_1+2x_2\leqslant 5,$$
$$16x_1+12x_2\leqslant 37,$$
$$6x_1+x_2\leqslant 12,$$
$$x_1\geqslant 0 \quad \text{и} \quad x_2\geqslant 0.$$

45. Ответьте на вопросы, поставленные в упражнении 43, заменив лишь ограничения (они были приведены в варианте а) упражнения 24 гл. 14 и соответствуют примеру 1 рис. 14.9).

$$x_1+2x_2\leqslant 6,$$
$$8x_1+6x_2\leqslant 23,$$
$$6x_1+x_2\leqslant 12,$$
$$x_1\geqslant 0 \quad \text{и} \quad x_2\geqslant 0.$$

46. Рассмотрите обычную транспортную задачу, аналогичную приведенной в разд. 6.2, где ресурсы у поставщиков равны соответственно $S_1 = 4$ и $S_2 = 6$, а спрос составляет $D_1 = 5$, $D_2 = 2$ и $D_3 = 3$. Пусть минимизируемая функция затрат имеет вид

$$4\left(x_{11}^2 + x_{12}^2 + x_{22}^2 + x_{23}^2\right) + 3x_{13}^2 + 2x_{21}^2.$$

Для каждого из приведенных ниже вариантов постановки задачи объясните, каким образом в алгоритме используется специальная структура системы ограничений. Является ли оптимальное решение целочисленным? (Ограничьтесь значением $k = 5$.)

а) Решите задачу методом выпуклых комбинаций.

б) Решите задачу симплексным методом вогнутого программирования.

47. Рассмотрите следующую задачу (она приводилась также в упражнении 46 гл. 14):

$$\text{Максимизировать } 6x_1 + 8x_2 - x_1^2 - x_2^2$$

при условиях

$$4x_1^2 + x_2^2 \leqslant 16,$$
$$3x_1 + 5x_2 \leqslant 15,$$
$$x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

В каждом из приведенных ниже вариантов постановки решите задачу указанным методом. Начертите область допустимых решений и нанесите на нее результаты, полученные для каждой итерации. (При тех методах, которые не обеспечивают сходимость за конечное число итераций, прекратите расчеты после $k = 5$.)

а) Метод допустимых направлений, описанный в разд. 15.6.

б) Модифицированный симплексный метод, описанный в разделе 15.8.

в) Обобщенный алгоритм программирования для сепарабельных задач, описанный в разд. 15.10.

г) Используйте для получения оптимального решения условия оптимальности, приведенные в разд. 15.7.

д) Замените целевую функцию на следующую: максимизировать $(x_1 - 1)^3 + (x_2 - 1)^3$. Примените метод допустимых направлений, начав с исходной точки $x_1^0 = x_2^0 = 0$ и приняв $b = 1$. Сможете ли вы получить оптимальное решение? Что произойдет, если принять $b = 2$? Для точки $x_1 = x_2 = 1$ примените условия оптимальности из разд. 15.7 и обсудите значение полученных результатов.

48. Рассмотрите задачу

$$\text{Максимизировать } x_1 + x_2$$

при условиях

$$x_1^2 + (x_2 - 4)^2 - 16 \leqslant 0,$$
$$(x_1 - 4)^2 + x_2^2 - 16 \leqslant 0,$$
$$x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

В каждом из приведенных ниже вариантов постановки решите задачу указанным методом. Начертите область допустимых решений и нанесите на нее результаты, полученные на каждой итерации. (Прекратите расчеты после значения $k = 5$.)

а) Метод допустимых направлений, описанный в разд. 15.6.

б) Обобщенный алгоритм программирования для сепарабельных задач, описанный в разд. 15.10.

в) Используйте для нахождения оптимального решения условия оптимальности, приведенные в разд. 15.7.

г) Замените целевую функцию на следующую

$$\text{Максимизировать } (x_1 - 1)^3 + (x_2 - 1)^3.$$

Примените метод допустимых направлений, начав с исходной точки $x_1^0 = x_2^0 = 0$ и приняв $b = 1$. Сможете ли вы получить оптимальное решение? Что произойдет, если принять $b = 2$? Для точки $x_1 = x_2 = 1$ примените условия оптимальности из разд. 15.7 и обсудите значение полученных результатов.

49. Ответьте на вопросы, поставленные в вариантах а), б) и в) упражнения 48, заменив лишь целевую функцию на следующую:

$$\text{Максимизировать } -26x_1 + 50x_2 - 3x_1^2 - 9x_2^2 + 10x_1x_2.$$

50. Покажите, как решить задачу упражнения 2, используя обобщенный алгоритм программирования, описанный в разд. 15.10, причем в R включается вся система ограничений. (Ограничьтесь значением $k = 5$.) Нанесите результаты, полученные на каждой из итераций, на рисунок, изображающий область допустимых решений.

51. Используйте алгоритм декомпозиции, описанный в разд. 15.11, для решения задач с двусторонними ограничениями на переменные, которые приведены в упражнении 28 гл. 5. Пусть в подзадачу входят ограничения сверху на переменные и условия неотрицательности переменных.

52. Используйте алгоритм декомпозиции, описанный в разд. 15.11, для решения транспортной задачи, приведенной в разд. 7.3 (рис. 7.2). Пусть в R входят ограничения на ресурсы поставщиков, а в S — ограничения спроса и условия неотрицательности переменных.

УПРАЖНЕНИЯ НА ПОСТАНОВКУ ЗАДАЧ И РАЗВИТИЕ ВЫЧИСЛИТЕЛЬНЫХ НАВЫКОВ

53. Рассмотрите оптимизационную модель нелинейного программирования, аналогичную приведенной в разд. 15.5. Покажите, как преобразовать эту модель в эквивалентную ей задачу с линейной целевой функцией. (*Указание*: введите еще одну переменную и еще одно ограничение, связывающее новую переменную с целевой функцией исходной задачи.)

54. Рассмотрите задачу, включающую нелинейное ограничение

$$\sum_{j=1}^{n} a_j x_j - K\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j\right]^{1/2} - b \geqslant 0,$$

где $k > 0$ и выражение в квадратных скобках неотрицательно для всех x_k. Приведите эквивалентную постановку, включающую два ограничения, из которых одно линейно и одно нелинейно, причем нелинейная функция $a(x)$ является выпуклой.

55. Рассмотрите задачу максимизации нелинейной функции $c(x)$ при условии выполнения линейных ограничений, аналогичную задаче (1) — (2) из разд. 15.1.

а) Запишите условия оптимальности и их следствия (9) — (14) из разд. 15.7. [*Указание*: сначала исключите дополняющие переменные s_i из (2) разд. 15.1.]

б) Используйте ответ, полученный для варианта а), применительно к задаче

$$\text{Максимизировать } -(x_1-4)^2-(x_2-4)^2$$

при условиях

$$\begin{aligned} x_1 + 3x_2 &\leqslant 12,\\ 3x_1 + x_2 &\leqslant 12,\\ -x_1 - x_2 &\leqslant -1,\\ -x_1 + x_2 &\leqslant 3,\\ x_1 \geqslant 0 \quad \text{и} \quad x_2 &\geqslant 0. \end{aligned}$$

в) Начертите область допустимых решений для задачи варианта б). Укажите оптимальное решение. С помощью рисунка интерпретируйте условия (13) и (14). {*Указание*: частные производные $[\partial c(\bar{x})/\partial x_1,\ \partial c(\bar{x})/\partial x_2]$ можно представить как направления скорейшего подъема от точки $\bar{x}$. Для каждого ограничения i, удовлетворяемого как равенство, $[\partial a_i(\bar{x})/\partial x_1,\ \partial a_i(x)\,\partial x_2]$ можно представить в качестве направления, ортогонального (перпендикулярного) этому ограничению в точке $\bar{x}$.}

56. Рассмотрите задачу

$$\text{Максимизировать } x_1 + x_2$$

при условиях

$$x_1^2 + (x_2-1)^2 - 1 \leqslant 0, \quad x_1^2 + x_2^2 - 1 \leqslant 0, \quad x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

а) Начертите область допустимых решений и укажите на ней оптимальное решение.

б) Запишите условия оптимальности и их следствия (9) — (14) из разд. 15.7.

в) Дайте геометрическую интерпретацию ответа на вопрос, поставленный в варианте б). {*Указание*: можно представить $[\partial c(\bar{x})/\partial x_1, \partial c(\bar{x})/\partial x_2]$ в качестве направления скорейшего подъема из точки $\bar{x}$. Для каждого ограничения i, удовлетворяемого как равенство, $[\partial a_i(\bar{x})/\partial x_1, \partial a_i(\bar{x})/\partial x_2]$ можно представить в качестве направления, ортогонального (перпендикулярного) линейной опоре этого ограничения в точке $\bar{x}$.}

57. Рассмотрите задачу

$$\text{Максимизировать } c_1x_1 + x_2$$

при условиях

$$x_1^2 + x_2^2 \leqslant 25, \quad x_1 - x_2 \leqslant 1, \quad x_1 \geqslant 0 \quad \text{и} \quad x_2 \geqslant 0.$$

Используйте условия оптимальности из разд. 15.7 для получения диапазона значений c_1, на котором решение $x_1 = 4$ и $x_2 = 3$ является оптимальным. Нанесите полученные результаты на чертеж области допустимых решений.

58. Задача «места и времени» (разд. 1.6). Определите вид оптимального решения этой задачи, применив условия оптимальности из разд. 15.7. Проделайте то же для задачи

$$\text{Максимизировать } (x_1 + x_2 + \ldots + x_N)$$

при условиях

$$\sum_{j=1}^{N} \frac{x_j^2}{c_j} \leqslant C \quad \text{и все} \quad x_j \geqslant 0.$$

59. Рассмотрите задачу

$$\text{Максимизировать } \sum_{j=1}^{n} c_j(x_j)$$

при условиях

$$\sum_{j=1}^{n} a_jx_j \leqslant b \quad \text{и все} \quad x_j \geqslant 0,$$

где каждая функция $c_j(x_j)$ строго вогнута по всем x_j и имеет непрерывную первую производную.

а) Запишите условия оптимальности и их следствия (9) — (14) из разд. 15.7.

б) Покажите, что если b увеличить до $b + e$ $(e > 0)$, то новое оптимальное значение каждой x_j по крайней мере столь же велико, как и в решении для b.

60. Рассмотрите задачу управления производством и запасами в течение планового периода из трех отрезков с выпуклой функцией затрат, приведенную в упражнении 32 гл. 9. Откажитесь от допущения о необходимости целочисленности всех x_t.

а) Для решения каждого из вариантов этого упражнения используйте условия оптимальности из разд. 15.7. Определите влияние на затраты бесконечно малых изменений каждого показателя спроса D_t.

б) Предположите, что для отрезка 1 функция затрат производства имеет вид $C_1(x_1) = c_1 x_1^2$ (причем в упражнении 32 гл. 9 $c_1 = 1$). Используйте условия оптимальности для получения диапазона значений c_1, на котором решения вариантов а) — в) и е) упражнения 32 гл. 9 остаются оптимальными.

61. Фирме на отрезке t требуется штат работающих численностью D_t ($t = 1, 2, \ldots, n$). Исходная численность работающих на начало отрезка 1 равна $x_0 = 0$. Обозначим через x_t число нанимаемых ($x_t \geqslant 0$) или увольняемых ($x_t < 0$) в начальный момент отрезка t. Значения x_t должны удовлетворять ограничениям

$$\sum_{j=1}^{t} x_j \geqslant D_t, \quad t = 1, 2, \ldots, n,$$

где знак суммирования в левой части неравенства позволяет отобразить наличие работающих на отрезке t. Пусть для отрезка t затраты найма или увольнения составляют $c_1(x_t)$, а стоимость простоя равна $c_2\left(\sum_{j=1}^{t} x_j - D_t\right)$. Предположим, что эти две функции строго выпуклы и имеют непрерывные первые производные, причем производная $c_2(x)$ при $x = 0$ неотрицательна. Пусть значение $c_1(x)$ неограниченно возрастает при увеличении $|x|$, а $c_2(x)$ — при увеличении x. Целевая функция имеет вид

$$\text{Максимизировать} \quad -\sum_{t=1}^{n} \left[c_1(x_t) + c_2\left(\sum_{j=1}^{t} x_j - D_t\right)\right].$$

а) Запишите условия оптимальности (9) — (11) из разд. 15.7.

б) Покажите, что если существует отрезок t, $t < n$, для которого оптимальный вариант заключается в наличии некоторого избытка работающих, так что $\sum_{j=1}^{t} x_j > D_t$, то оптимальным решением будет принять на работу столько человек на следующем отрезке, чтобы $x_{t+1} > x_t$.

62. Рассмотрите двойственную теорему нелинейного программирования, приведенную в разд. 15.7. Составьте двойственную задачу (23) — (24), если все ограничения линейны. Покажите, как задачу (23) — (24) можно свести к обычной двойственной задаче линейного программирования.

63. При методе допустимых направлений, изложенном в разделе 15.6, относительная стационарная точка будет достигнута в том случае, когда не существует допустимого направления, обеспечивающего улучшение решения, если оценивать это направление с помощью первых частных производных; в разд. 15.6 это условие выражается (1). Покажите, что если $c(x)$ вогнута, а x^k — относительная стационарная точка, то x^k — глобальный оптимум. [*Указание*: можно использовать свойство вогнутой функции, согласно которому для любой точки x и для любого направления d $c(x+d) \leqslant \leqslant c(x) + \sum_{j=1}^{n} \frac{\partial c(x)}{\partial x_j} \cdot d_j$.]

64. Пусть $A_i(x)$ — линейное ограничение, а x^k — точка, удовлетворяющая ограничению $A_i(x^k) \leqslant 0$. Покажите, что направления d_j, удовлетворяющие (6) из разд. 15.6 при исключенной переменной x_0, являются допустимыми.

65. Рассмотрите задачу

$$\text{Максимизировать } c_1 x_1$$

при условиях

$$a_{11}x_1 - x_2 \leqslant 0,$$
$$a_{21}x_1 - x_3 \leqslant 0,$$
$$-\log a_{32}(x_2+1) - \log a_{33}(x_3+1) - b_3 \leqslant 0,$$
$$\text{все } x_j \geqslant 0,$$

где все $a_{ij} > 0$. Рассмотрите решение задачи с помощью обобщенного алгоритма программирования, изложенного в разд. 15.10. Сравните трудность решения суженной координирующей задачи и подзадачи в двух случаях — когда нелинейное ограничение включено и когда не включено в S.

66. Рассмотрите использование обобщенного алгоритма программирования для сепарабельных задач, описанное в разд. 15.10. Обсудите влияние включения условий неотрицательности переменных $x_j \geqslant 0$ в явном виде в суженную координирующую задачу и исключения их из подзадачи. [*Указание*: продемонстрируйте ваш ответ на задаче Р.]

67. Задача «двух картошек» (разд. 1.6). В упражнении 47 гл. 14 эта задача была обобщена на случай возможного изменения коэффициентов выхода продуктов. Покажите, как сформулировать и решить задачу упражнения 47 из гл. 14 с помощью обобщенного алгоритма программирования, если предполагается линейность функций *затрат* $c_1(f_1, f_2, f_3)$ и $c_2(g_1, g_2, g_3)$.

ДОПОЛНИТЕЛЬНЫЕ УПРАЖНЕНИЯ

68. а) Пусть $a(x)$ является выпуклой функцией, где $x \equiv \equiv (x_1, x_2, \ldots, x_n)$. Рассмотрите точки $y_1, y_2, \ldots, y_s$ и веса $w_1, w_2, \ldots, w_s$, где каждый вес $w_i \geqslant 0$ и $\sum_{i=1}^{s} w_i = 1$. Покажите, что

$$\sum_{i=1}^{s} a(y_i) w_i \geqslant a\left(\sum_{i=1}^{s} y_i w_i\right),$$

где знак суммирования в правой части обозначает средневзвешенную (по каждой координате) точек $y_1, \ldots, y_s$. Как формулируется соответствующая теорема для вогнутой функции?

б) Пусть каждая $a_i(x)$ для $i = 1, 2, \ldots, s$ является выпуклой функцией, где $x \equiv (x_1, x_2, \ldots, x_n)$. Покажите, что $A(x) = \max[a_1(x), a_2(x), \ldots, a_3(x)]$ — выпуклая функция. Как формулируется соответствующая теорема, если каждая $a_1(x)$ — вогнутая функция?

69. а) Покажите, что непрерывно дифференцируемая функция $a(x)$ удовлетворяет неравенству (4) из разд. 15.5 в том и только том случае, когда эта функция выпукла.

б) Покажите, что функция $a(x)$, имеющая непрерывные вторые частные производные, выпукла в том и только том случае, если квадратичная форма

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 a(x)}{\partial x_i \, \partial x_j} d_i d_j$$

является положительно полуопределенной в любой точке x.

в) Покажите, как изменятся полученные результаты, если $a(x)$ вогнута.

70. Пусть неравенству (1) из разд. 15.6 удовлетворяет набор направлений d_j (не обязательно допустимых). Докажите, что существует точка $(x^k + td)$, где $t > 0$, такая, что $c(x^k + td) > c(x^k)$.

71. Рассмотрите оптимизационную модель нелинейного программирования, аналогичную приведенной в разд. 15.5. Пусть $c(x_1, x_2, \ldots, x_n)$ является *строго выпуклой* функцией. Докажите, что оптимальное значение $\bar{c}$ возможно лишь в экстремальной точке области допустимых решений. (*Напоминание*: точка x является экстремальной, если ее нельзя представить в виде выпуклой комбинации двух других допустимых точек.)

72. Рассмотрите условия оптимальности Лагранжа, приведенные в разд. 15.7. Покажите, что точка $\bar{x}$ является оптимальной, если существует точка $\bar{u} \geqslant 0$, такая, что пара $(\bar{x}, \bar{u})$ представляет собой седловую точку функции $L(x, u)$.

73. Пусть $c(x_1, x_2, \ldots, x_n)$ и $a_i(x_1, x_2, \ldots, x_n)$ для $i = 1, 2, \ldots, m$ — действительные функции произвольного вида, определенные на произвольной области допустимых значений переменных $x_1, x_2, \ldots, x_n$. Пусть, далее, $\bar{u}_i \geqslant 0$ для $i = 1, 2, \ldots, m$ есть заданные величины. Предположим, что $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ — оптимальное решение задачи

$$\text{Максимизировать } c(x) - \sum_{i=1}^{m} a_i(x)\bar{u}_i,$$

где x принадлежит R. Покажите, что $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ одновременно является оптимальным решением задачи

Максимизировать $c(x)$

при условиях $a_i(x) - a_i(\bar{x}) \leqslant 0, \quad i = 1, 2, \ldots, m,$

где x также принадлежит R.

74. Рассмотрите задачу максимизации нелинейной функции $c(x)$ при условии выполнения линейных ограничений, аналогичную задаче (1) — (2) из разд. 15.1. Измените постановку задачи с целью включения ограничений, являющихся уравнениями, в которых отсутствуют s_i, а также переменных, на знак которых не налагается ограничений. Выведите условия оптимальности и их следствия, аналогичные (9) — (14) из разд. 15.7.

75. В случае применения метода штрафной функции из разд. 15.9 используется особое выражение, умножаемое на параметр r, а именно сумма вида $[1/A_i(x)]$. Рассмотрите более общую модель, в которой на r умножается сумма вида $G_i[A_i(x)]$.

а) Обсудите свойства, которыми должна обладать функция G_i, для того чтобы ее можно было использовать в качестве штрафной функции. Предложите одну или несколько функций G_i, обладающих такими свойствами, и покажите получаемые выражения для $\partial c(x \mid r)/\partial x_j$. (*Указание*: рассмотрите логарифмическую функцию.)

б) Разработайте штрафную функцию, применимую в том случае, если мы заставляем r стремиться к $-\infty$.

в) Разработайте штрафную функцию, применимую и в случае ограничений, имеющих вид уравнений $a_i(x) = 0$.

76. Рассмотрите обобщенный алгоритм программирования, приведенный в разд. 15.10. Докажите, что оптимальные значения c^k целевой функции суженной координирующей задачи монотонно возрастают, стремясь в пределе к оптимальному значению c целевой функции исходной задачи.

77. Рассмотрите оптимизационную модель нелинейного программирования, аналогичную приведенной в разд. 15.5. Согласно характеристике системы ограничений, предполагается существование точки x^0, такой, что каждая $a_i(x^0) < 0$.

а) Предположите, что требуется найти точку z, такую, что каждая $a_i(z) \leqslant 0$. Обсудите возможности нахождения z. [*Указание*: начните с произвольной точки y, удовлетворяющей некоторым, хотя и не всем, ограничениям: например, пусть $a_1(y) > 0$. Постройте оптимизационную задачу, где $a_1(x)$ включена в целевую функцию, решение которой позволяет получить точку z, такую, что $a_1(z) \leqslant 0$, причем сохраняется допустимость z с учетом всех тех ограничений, которым удовлетворяла y. Предложите также альтернативный подход с использованием искусственных переменных.]

б) Предложите один или несколько способов изменения ответов на вопросы, поставленные в варианте а) в том случае, если требуется получить точку x^0, для которой каждая $a_i(x^0) < 0$.

ЛИТЕРАТУРА

Литература к главам 8 и 10

КНИГИ

1. *Beckmann M. J.*, Dynamic Programming of Economic Decisions, Springer-Verlag, Berlin, 1968.
2. *Bellman R. E.*, Dynamic Programming, Princeton University Press, 1957, русский перевод: *Беллман Р.*, Динамическое программирование, ИЛ, 1960.
3. *Bellman R. E.*, *Dreyfus S. E.*, Applied Dynamic Programming, Princeton University Press, 1962; русский перевод: *Беллман Р.*, Прикладные задачи динамического программирования, изд-во "Наука", 1965.
4. *Nemhauser G. L.*, Introduction to Dynamic Programming, Wiley 1966.

СТАТЬИ

5. *Bellman R.*, *Glicksberg I.*, *Gross O.*, The Theory of Dynamic Programming as Applied to a Smoothing Problem, *SIAM J. on Applied Mathematics*, 2, 82—88 (1954).
6. *Breiman L.*, Stopping-Rule Problems, 284—319 in Applied Combinatorial Mathematics, Edwin F. Beckenbach (ed.), Wiley, 1964.
7. *Greenberg H. J.*, The Use of Branching in Dynamic Programming for Parametric Analysis, *Operations Research*, **15**, 976—977 (1967).
8. *Karush W.*, On a Class of Minimum-Cost Problems, *Managements Science*, 4, 136—153 (1958).
9. *Karush W.*, A General Algorithm for the Optimal Distribution of Effort, *Management Science*, 9, 50—72 (1962).
10. *Karush W.*, *Vazsonyi A.*, Mathematical Programming and Employment Scheduling, *Naval Research Logistics Quarterly*, **4**, 297—320 (1957).
11. *Marshak T. A.*, *Yahav J. A.*, The Sequential Selection of Approaches to a Task, *Management Science*, **12**, 627—647 (1966).
12. *McCall J. J.*, The Economics of Information and Optimal Stopping Rules, *J. of Business*, **38**, 300—317 (1965).
13. *McNaughton R.*, Scheduling with Deadlines and Loss Functions, *Management Science*, **6**, 1—12 (1959).
14. *Mitten L. G.*, Composition Principles for Synthesis of Optimal Multi-Stage Processes, *Operations Research*, **12**, 610—619 (1964).
15. *Randolph P.*, An Optimal Stopping Rule, *Operations Research*, **15**, 562—564 (1967).
16. *Veinott A. F., Jr.*, *Wagner H. M.*, Optimal Capacity Scheduling — I, *Operations Research*, **10**, 518—532 (1962).
17. *Veinott A. F., Jr.*, *Wagner H. M.*, Optimal Capacity Scheduling — II, *Operations Research*, **10**, 533—546 (1962).
18. *White L. S.*, The Analysis of a Simple Class of Multistage Inspection Plans, *Management Science*, **12**, 685—693 (1966).
19. *Wong P. J.*, *Luenberger D. G.*, Reducing the Memory Requirements of Dynamic Programming, *Operations Research*, **16**, 1115—1125 (1968).
20. *Ying C. C.*, Learning by Doing — An Adaptive Approach to Multiperiod Decisions, *Operations Research*, **15**, 797—812 (1967).

Литература к главе 9

КНИГИ

1. *Arrow K. J.*, *Karlin S.*, *Scarf H. E.* (eds.), Studies in the Mathematical Theory of Inventory and Production, Stanford University Press, 1958.

2. *Buchan J.*, *Koenigsberg E.*, Scientific Inventory Management, Prentice-Hall, 1963.
3. *Buffa E. S.*, Production-Inventory Systems: Planning and Control, Irwin, 1968.
4. *Elmaghraby S. E.*, The Design of Production Systems, Reinhold, 1966.
5. *Hadley G.*, *Whitin T. M.*, Analysis of Inventory Systems, Prentice-Hall, 1963; русский перевод: *Хедли Дж.*, *Уайтин Т.*, Анализ систем управления запасами, изд-во "Наука", 1969.
6. *Holt C. C.*, *Modigliani F.*, *Muth J. F.*, *Simon H. A.*, Planning Production, Inventories, and Work Force, Prentice-Hall, 1960.
7. *Naddor E.*, Inventory Systems, Wiley, 1966.
8. *Scarf H. E.*, *Dorothy G.*, *Shelly M. W.* (eds.), Multistage Inventory Models and Techniques, Stanford University Press, 1963.
9. *Starr M. K.*, Production Management Systems and Synthesis, Prentice-Hall, 1964.
10. *Starr M. K.*, *Miller D. W.*, Inventory Control: Theory and Practice, Prentice-Hall, 1962.

СТАТЬИ

11. *Charnes A.*, *Cooper W. W.*, *Mellon B.*, A Model for Optimizing Production by Reference to Cost Surrogates, *Econometrica*, **23**, 307—323 (1955).
12. *Eppen G. D.*, *Gould E. J.*, A Lagrangian Application to Production Models, *Operations Research*, **16**, 819—829 (1955).
13. *Eppen G. D.*, *Gould F. J.*, *Pashigian B. P.*, Extensions of the Planning Horizon Theorem in the Dynamic Lot Size Model, *Management Science*, **15**, 268—277 (1969).
14. *Fabrycky W. J.*, *Banks J.*, A Hierarchy of Deterministic Procurement Inventory Systems, *Operations Research*, **14**, 888—901 (1966).
15. *Hoffman A. J.*, *Jacobs W.*, Smooth Patterns of Production, *Management Science*, **1**, 86—91 (1954).
16. *Hu T. C.*, *Prager W.*, Network Analysis of Production Smoothing, *Naval Research Logistics Quarterly*, **6**, 17—23 (1959).
17. *Hwang C. L.*, *Fan L. T.*, *Erickson L. E.*, Optimum Production Planning by the Maximum Principle, *Management Science*, **13**, 751—755 (1967).
18. *Johnson S. M.*, Sequential Production Planning over Time at Minimum Cost, *Management Science*, **3**, 435—437 (1957).
19. *Klein M.*, Some Production Planning Problems, *Naval Research Logistics Quarterly*, **4**, 269—286 (1957).
20. *Klein M.*, On Production Smoothing, *Management Science*, **7**, 286—293 (1961).
21. *Kortanek K. O.*, *Sodaro D.*, *Soyster A. L.*, Multi-Product Production Scheduling via Extreme Point Properties of Linear Programming, *Naval Research Logistics Quarterly*, **15**, 287—300 (1968).
22. *Lippman S. A.*, *Rolfe A. J.*, *Wagner H. M.*, *Yuan J. S. C.*, Optimal Production Scheduling and Employment Smoothing with Deterministic Demands, *Management Science*, **14**, 127—158 (1967).
23. *Lippman S. A.*, *Rolfe A. J.*, *Wagner H. M.*, *Yuan J. S. C.*, Algorithms for Optimal Production Scheduling and Employment Smoothing, *Operations Research*, **15**, 1011—1029 (1967).
24. *Manne A. S.*, Programming of Economic Lot Sizes, *Management Science*, **4**, 115—135 (1958).
25. *Manne A. S.*, *Veinott A. F.*, Jr., Optimal Plant Size with Arbitrary Increasing Time Paths of Demand, 178—192 in Investments for Capacity Expansion, Size, Location, and Time-Phasing, Alan S. Manne (ed.), Massachusetts Institute of Technology Press, 1967.
26. *Nemhauser G. L.*, *Nuttle H. L. W.*, A Quantitative Approach to Employment Planning, *Management Science*, **11**, B155—B165 (1965).
27. *Silver E. A.*, A Tutorial on Production Smoothing and Work Force Balancing, *Operations Research*, **15**, 985—1010 (1967).

28. *Sobel M. J.*, Smoothing Start-Up and Shut-Down Costs in Sequential Production, *Operations Research*, **17**, 133—144 (1969).
29. *Veinott A. F.*, Jr., Production Planning with Convex Costs: a Parametric Study, *Management Science*, **10**, 441—460 (1964).
30. *Wagner H. M.*, A Postscript to "Dynamic Problems in the Theory of the Firm", *Naval Research Logistics Quarterly*, 7, 7—12 (1960).
31. *Wagner H. M.*, *Whitin T. M.*, Dynamic Problems in the Theory of the Firm, *Naval Research Logistics Quarterly*, 5, 53—74 (1958).
32. *Wagner H. M.*, *Whitin T. M.*, Dynamic Version of the Economic Lot Size Model, *Management Science*, 5, 89—96 (1958).
33. *Zabel E.*, Some Generalizations of an Inventory Planning Horizon Theorem, *Management Science*, **10**, 465—471 (1964).
34. *Zangwill W. I.*, A Deterministic Multi-Period Production Scheduling Model with Backlogging, *Management Science*, **13**, 105—119 (1966).
35. *Zangwill W. I.*, A Deterministic Multiproduct, Multifacility Production and Inventory Model, *Operations Research*, **14**, 486—507 (1966).
36. *Zangwill W. I.*, Production Smoothing of Economic Lot Sizes with Non-Decresing Requirements, *Management Science*, **13**, 191—209 (1966).

Литература к главам 11 и 12

СТАТЬИ

1. *Blackwell D.*, Discrete Dynamic Programming, *Annals Mathematical Statistics*, 33, 719—726 (1962).
2. *Blackwell D.*, Discounted Dynamic Programming, *Annals Mathematical Statistics*, **36**, 226—235 (1965).
3. *Denardo E. V.*, Contraction Mappings in the Theory Underlying Dynamic Programming, *SIAM Review*, **9**, 165—177 (1967).
4. *Denardo E. V.*, *Miller B. L.*, An Optimality Condition for Discrete Dynamic Programming with no Discounting, *Annals Mathematical Statistics*, 39, 1220—1227 (1968).
5. *Miller B. L.*, Finite State Continuous Time Markov Decision Processes with a Finite Planning Horizon, *SIAM J. on Control*, **6**, 266—280 (1968).
6. *Shapiro J. F.*, Shortest Route Methods for Finite State Space Deterministic Dynamic Programming Problems, *SIAM J. Appl. Math.*, **16**, 1232—1250 (1968).
7. *Strauch R.*, Negative Dynamic Programming, *Annals Mathematical Statistics*, 37, 871—890 (1966).
8. *Veinott A. F.*, Jr., On Finding Optimal Policies in Discrete Dynamic Programming With no Discounting, *Annals Mathematical Statistics*, 37, 1284—1294 (1966).

Литература к главе 13

КНИГИ

1. *Hu T. C.*, Integer Programming and Network Flows, Addison-Wesley, 1969.
2. *Muth J. F.*, *Thompson G. L.*, Industrial Scheduling, Prentice-Hall, 1963.
3. *Tonge F. M.*, A Heuristic Program for Assembly Line Balancing, Prentice-Hall, 1961.
4. *Weingartner M. H.*, Mathematical Programming and the Analysis of Capital Budgeting Problems, Prentice-Hall, 1960.

СТАТЬИ

5. *Alcaly R. E., Klevorick A. K.*, A Note on the Dual Prices of Integer Programs, *Econometrica*, **34**, 206—214 (1966).
6. *Allen S. G.*, Computation for the Redistribution Model with Set-up Charge, *Management Science*, 8, 482—489 (1962).
7. *Balas E.*, An Additive Algorithm for Solving Linear Programs with Zero-One Variables, *Operations Research*, **13**, 517—546 (1965).
8. *Balas E.*, Discrete Programming by the Filter Method, *Operations Research*, 15, 915—957 (1967).
9. *Balinski M. L.*, Integer Programming: Methods, Uses, Computation, *Management Science*, **12**, 253—313 (1965).
10. *Balinski M. L., Quandt R. E.*, On an Integer Program for a Delivery Problem, *Operations Research*, **12**, 300—304 (1964).
11. *Bellmore M., Nemhauser G. L.*, The Travelling Salesman Problem: A Survey, *Operations Research*, **16**, 538—558 (1968).
12. *Burt O. R., Harris C. C.*, Jr., Apportionment of the U.S. House of Representatives: A Minimum Range, Integer Solution, Allocation Problem, *Operations Research*, **11**, 648—652 (1963).
13. *Cooper L., Drebes C.*, An Approximate Solution Method for the Fixed Charge Problem, Naval Research Logistics Quarterly, **14**, 101—113 (1967).
14. *Cabot A. V., Hurter A. V.*, An Approach to Zero-One Integer Programming, *Operations Research*, **16**, 1206—1211 (1968).
15. *Crowston W., Thompson G. L.*, Decision CPM: A Method for Simultaneous Planning, Scheduling, and Control of Projects, *Operations Research*, 15, 407—426 (1967).
16. *Dalton R. E., Lewellyn R. W.*, An Extension of the Gomory Mixed-Integer Algorithm to Mixed-Discrete Variables, *Management Science*, **12**, 569—575 (1966).
17. *Dantzig G. B., Fulkerson D. R., Johnson S. M.*, On a Linear Programming, Combinatorial Approach to the Travelling Salesman Problem, *Operations Research*, 7, 58—66 (1959).
18. *D'Esopo D. A., Lefkowitz B.*, Note on an Integer Linear Programming Model for Determining a Minimum Embarkation Fleet, *Naval Research Logistics Quarterly*, **11**, 7, 9—82 (1964).
19. Dwyer P. S., Use of Completely Reduced Matrices in Solving Transportation Problems with Fixed Charges, *Naval Research Logistics Quarterly*, **13**, 289—313 (1966).
20. *Efroymson M. A., Ray T. L.*, A Brahch-Bound Algorithm For Plant Location *Operations Research*, **14**, 361—368 (1966).
21. *Elmaghraby S. E.*, The Machine Sequencing Problem — Review and Extensions, *Naval Research Logistics Quarterly*, **15**, 205—232 (1968).
22. *Freeman R. J.*, Computational Experience with a "Balasian" Integer Programming Algorithm, *Operations Research*, **14**, 935—941 (1966).
23. *Gavett J. W., Plyter N. V.*, The Optimal Assignments of Facilities to Locations by Branch and Bound, *Operations Research*, **14**, 210—232 (1966).
24. *Geoffrion A. M.*, Integer Programming by Implicit Enumeration and Balas' Method, *SIAM Review*, **9**, 178—190 (1967).
25. *Giglio R. J., Wagner H. M.*, Approximate Solutions to the Three-Mechine Scheduling Problem, *Operations Research*, **12**, 305—324 (1964).
26. *Glassey C. R.*, Minimum Change-Over Scheduling of Several Products on One Machine, *Operations Research*, **16**, 342—352 (1968).
27. *Glover F.*, A Multiphase-Dual Algorithm for the Zero-One Integer Programming Problem, *Operations Research*, **13**, 879—919 (1965).
28. *Glover F.*, Stronger Cuts in Integer Programming, *Operations Research*, 15, 117—1177 (1967).
29. *Glover F.*, A New Foudnation for a Simplified Primal Integer Programming Algorithm, *Operations Research*, **16**, 727—740 (1968).

30. *Glover F.*, Surrogate Constraints, *Operations Research*, **16**, 741—749 (1968).
31. *Glover F.*, A Note on Linear Programming and Integer Feasibility, *Operations Research* **16**, 1212—1216 (1968).
32. *Gomory R E.*, An Algorithm for Integer Solutions to Linear Programs, 269—302 in Recent Advances in Mathematical Programming, Robert L. Graves and Philip Wolfe (eds.) J. McGraw-Hill, 1963.
33. *Gomory R. E.*, On the Relation Between Integer and Noninteger Solutions to Linear Programs, *National Academy of Sciences*, **53**, 260—265 (1965).
34. *Gomory R. E., Baumol W. J.*, Integer Programming and Pricing, *Econometrica*, **28**, 521—550 (1960).
35. *Graves G. W.*, *Whinston A.*, A New Approach to Discrete Mathematical Programming, *Management Science*, **15**, 177—190 (1968).
36. *Gutjahr A. L.*, *Nemhauser G. L.*, An Algorithm for the Line Balancing Problem, *Management Science*, **11**, 308—315 (1964).
37. *Haldi J.*, *Isaacson L. M.*, A Computer Code for Integer Solutions to Linear Problems, *Operations Research*, 13, 946—959 (1965).
38. *Held M.*, *Karp R. M.*, A Dynamic Programming Approach to Sequencing Problems, *SIAM J. on Applied Mathematics*, **10**, 196—210 (1962).
39. *Held M.*, *Karp R. M.*, *Shareshian R.*, Assembly-Line Balancing-Dynamic Programming with Precedence Constraints, *Operations Research*, **11**, 442—459 (1963).
40. *Hillier F. S.*, *Connors M. M.*, Quadratic Assignment Problem Algorithms and the Location of Indivisible Facilities, *Management Science*, **13**, 42—57 (1966).
41. *Hirsch W. M.*, *Dantzig G. B.*, The Fixed Charge Problem, *Naval Research Logistics Quarterly*, **15**, 413—424 (1968).
42. *Ignall E.*, *Schrage L. E.*, Application of the Branch and Bound Technique to Some Flow-Shop Scheduling Problems, *Operations Research*, **13**, 400—412 (1965).
43. *Karg R. L.*, *Thompson G. L.*, A Heuristic Approach to Solving Travelling Salesman Problems, *Management Science*, 10, 225—248 (1964).
44. *Kolesar P. J.*, A Branch and Bound Algorithm for the Knapsack Problem, *Management Science*, **13**, 723—735 (1967).
45. *Koopmans T. C.*, *Beckmann M.*, Assignment Problems and the Location of Econimic Activities, *Econometrica*, **25**, 53—76 (1957).
46. *Kotranek K. O.*, *Maxwell W. L.*, On a Class of Combinatorial Optimizers for Multi-Product Single-Machine Scheduling, *Management Science*, **15**, 239—248 (1969).
47. *Kuehn A. A.*, *Hamburger M. J.*, A Heuristic Program for Locating Warehouses, *Management Science*, 9, 643—666 (1963).
48. *Kuhn H. W.*, *Baumol W. J.*, An Approximative Algorithm for the Fixed-Charges Transportation Problem, *Naval Research Logistics Quarterly*, **9**, 1—15 (1962).
49. *Land A. H.*, *Doig A. G.*, An Automatic Method of Solving Discrete Programming Problems, *Econometrica*, **28**, 497—520 (1960).
50. *Lawler E. L.*, The Quadratic Assignment Problem, *Management Science*, **9**, 586—599 (1963).
51. *Lawler E. L.*, *Wood D. E.*, Branch-and-Bound Methods: A Survey, *Operations Research*, **14**, 699—719 (1966).
52. *Lemke C. E.*, *Spielberg K.*, Direct Search Algorithms for Zero-One and Mixed-Integer Programming, *Operations Research*, **15**, 892—914 (1967).
53. *Levy F. K.*, *Thompson G. L.*, *Wiest J. D.*, Multiship. Multishop, Work-load-Smoothing Program, *Naval Research Logistics Quarterly*, 9, 37—44 (1962).
54. *Little J. D. C.*, *Murty K. G.*, *Sweeney D. W.*, *Karel C.*, An Algorithm for the Travelling Salesman Problem, *Operations Research*, **11**, 972—898 (1963).
55. *Manne A. S.*, On the Job-Shop Scheduling Problem, *Operations Research*, **8**, 219—223 (1960).
56. *Manne A. S.*, Plant Location under Economies-of-Scale — Decentralization and Computation. *Management Science*, **11**, 213—235 (1964).

57. *Mao J. C. T., Wallingford B. A.*, An Extension of Lawler and Bell's Method of Discrete Optimization with Examples from Capital Budgeting, *Management Science*, **15**, B51—B60 (1968).
58. *Markowitz H. M., Manne A. S.*, On the Solution of Discrete Programming Problems, *Econometrica*, 25, 84—110 (1957).
59. *Mizukami K.*, Optimum Redundancy for Maximum System Reliability by the Method of Convex and Integer Programming, *Operations Research*, **16**, 392—406 (1968).
60. *Moore J. M.*, An *n* Job, One Machine Sequencing Algorithm for Minimizing the Number of Late Jobs, *Management Science*, 15, 102—109 (1968).
61. *Murty K. Q.*, Solving the Fixed Charge Problem by Ranking the Extreme Points, *Operations Research*, 16, 268—279 (1968).
62. *Nemhauser G. L., Ullmann Z.*, A Note on the Generalized Lagrange Multiplier Solution to an Integer Programming Problem, *Operations Research*, 16, 450—453 (1968).
63. *Pierce J. F.*, Application of Combinatorial Programming to a Class of All-Zero-One Integer Programming Problems, *Management Science*, **15**, 191—209 (1968).
64. *Reiter S., Sherman G.*, Discrete Optimizing, *SIAM J. on Applied Mathematics*, **13**, 864—889 (1965).
65. *Shapiro J. F.*, Dynamic Programming Algorithms for the Integer Programming Problem I: The Integer Programming Problem Viewed as a Knapsack Type Problem, *Operations Research*, **16**, 103—121 (1968).
66. *Shapiro J. F.*, Group Theoretic Algorithms for the Integer Programming Problem II: Extension to a General Algorithm, *Operations Research*, **16**, 928—947 (1968).
67. *Smith R. D., Dudek R. A.*, A General Algorithm for Solution of the *n*-Job, *M*-Machine Sequencing Problem of the Flow Shop, *Operations Research*, **15**, 71—82 (1967).
68. *Spielberg K.*, Algorithms for the Simple Plant-Location Problem with Some Side Conditions, *Operations Research*, **17**, 85—111 (1969).
69. *Tillman F. A., Liittschwager J. M.*, Integer Programming Formulation to Constrained Reliability Problems, *Management Science*, 13, 887—899 (1967).
70. *Tonge F. M.*, Assembly Line Balancing Using Probabilistic Combinations of Heuristics, *Management Science*, **11**, 727—735 (1965).
71. *Vergin R. C., Rogers J. D.*, An Algorithm and Computational Procedure for Locating Economic Facilities, *Management Science*, **13**, B240—B254 (1967).
72. *Wagner H. M.*, An Integer Linear-Programming Model for Machine Scheduling, *Naval Research Logistics Quarterly*, **6**, 131—140 (1959).
73. *Wagner H. M., Giglio R. J., Glaser R. G.*, Preventive Maintenance Scheduling by Mathematical Programming, *Management Science*, **10**, 316—334 (1964).
74. *Watters L.*, Reduction of Integer Polynomial Programming Problems to Zero-One Linear Programming Problems, *Operations Research*, **15**, 1171—1174 (1967).
75. *Weingartner H. M.*, Capital Budgeting of Interrelated Projects: Survey and Synthesis, *Management Science*, **12**, 485—516 (1966).
76. *Weingartner H. M., Ness D. N.*, Methods for the Solution of the Multi-Dimensional O/I Knapsack Problem, *Operations Research*, **15**, 83—103 (1967).
77. *Wiest J. D.*, A Heuristic Model for Scheduling Large Projects with Limited Resources, *Management Science*, **13**, B359—B377 (1967).
78. *Wilson R. B.*, Stronger Cuts in Gomory's All-Integer Integer Programming Algorithm, *Operations Research*, 15, 155—157 (1967).
79. *Young R. D.*, A Simplified Primal (All-Integer) Integer Programming Algorithm, *Operations Research*, **16**, 750—782 (1968).
80. *Zionts S.*, An Algorithm for the Solution of Mixed Integer Programming Problems, *Management Science*, **15**, 113—116 (1968).

Литература к главам 14 и 15

КНИГИ

1. *Abadie J.* (ed.), Nonlinear Programming, Wiley, 1967.
2. *Arrow K. J., Hurwicz L. Uzawa H.*, Studies in Linear and Non-Linear Programming, Stanford University Press, 1958; русский перевод: *Эрроу К., Гурвиц Л., Удзава Х.*, Исследования по линейному и нелинейному программированию, ИЛ, 1962.
3. *Bracken J., McCormick G. P.*, Selected Application of Nonlinear Programming, Wiley, 1968.
4. *Duffin R. J., Peterson E. L., Zener C.*, Geometric Programming: Theory and Application, Wiley, 1967.
5. *Fiacco A. V., McCormick G. P.*, Nonlinear Programming, Sequential Unconstrained Minimization Techniques, Wiley, 1968.
6. *Hadley G.*, Nonlinear and Dynamic Programming, Addison-Wesley, 1964; русский перевод: *Хедли Дж.*, Нелинейное и динамическое программирование, изд-во "Мир", 1967.
7. *Kunzi H. P., Krelle W., Oettli W.*, Nonlinear Programming, Blaisdell, 1966.
8. *Mangasarian O. L.*, Nonlinear Programming, McGraw-Hill, 1969.
9. *Markowitz H. M.*, Portfolio Selection. Efficient Diversification of Investments, Wiley, 1959.
10. *Saaty T. L., Bram J.*, Nonlinear Mathematics, McGraw-Hill 1964.
11. *Wilde D. J.*, Optimum Seeking Methods, Prentice-Hall, 1964; русский перевод: *Уайлд Дж.*, Методы поиска экстремума, изд-во "Наука", 1967.
12. *Wilde D. J., Beightler C. S.*, Foundations of Optimization, Prentice-Hall, 1967.
13. *Zangwill W. I.*, Non-Linear Programming — An Unified Approach, Prentice-Hall, 1969.
14. *Zoutendijk G.*, Methods of Feasible Directions, Elsevier, 1960.

СТАТЬИ

15. *Arrow K. J., Enthoven A. C.*, Quasi-Concave Programming, *Econometrica* 29, 779—800 (1961).
16. *Baumol W. J., Bushnell R. C.*, Error Produced by Linearization in Mathematical Programming, *Econometrica*, 35, 447—471 (1967).
17. *Baumol W. J., Wolfe P.*, A Warehouse-Location Problem, *Operations Research*, 6, 252—263 (1958).
18. *Beale E. M. L.*, On Quadratic Programming, *Naval Research Logistios Quarterly*, 6, 227—243 (1959).
19. *Beale E. M. L.*, Note on a Comparison of Two Methods of Quadratic Programming, *Operations Research*, 14, 442—443 (1966).
20. *Bernholtz B.*, A New Derivation of the Kuhn-Tucker Conditions, *Operations Research*, 12, 295—299 (1964).
21. *Boot J. C. G.*, On Sensitivity Analysis in Convex Quadratic Programming Problems, *Operations Research*, 11, 771—786 (1963).
22. *Brooks R., Geoffrion A. M.*, Finding Everetts' Lagrange Multipliers by Linear Programming, *Operations Research*, 14, 1149—1153 (1966).
23. *Charnes A., Cooper W. W.*, Nonlinear Network Flows and Convex Programming Over Incidence Matrices, *Naval Research Logistics Quarterly*, 5, 231—240 (1958).
24. *Charnes A., Cooper W. W.*, Programming with Linear Fractional Functionals, *Naval Research Logistics Quarterly*, 9, 181—186 (1962).
25. *Cottle R. W.*, Nonlinear Programs with Positively Bounded Jacobians, *SIAM J. on Applied Mathematics*, 14, 147—158 (1966).
26. *Dantzig G. B.*, Linear Control Processes and Mathematical Programming, *SIAM J. on Control*, 4, 56—60, (1966).
27. *Dinkelbach W.*, On Nonlinear Fractional Programming, *Management Science*, 13, 492—498 (1967).

28. *Dorn W. S.*, Self Dual Quadratic Programs, *SIAM J. on Applied Mathematics*, 9, 51—54 (1961).
29. *Dorn W. S.*, Non-Linear Programming — A Survey, *Management Science*, 9, 171—208 (1963).
30. *Duffin R. J.*, *Peterson E. L.*, Duality Theory for Geometric Programming, *SIAM J. on Applied Mathematics*, 14, 1307—1349 (1966).
31. *Everett H.*, Generalized LaGrange Multiplier Method for Solving Problems of Optimum Allocation of Resources, *Operations Research*, 11, 399—417 (1963).
32. *Frank M.*, *Wolfe P.*, An Algorithm for Quadratic Programming, *Naval Research Logistics Quarterly*, 3, 95—110 (1956).
33. *Fromovitz S.*, Non-Linear Programming with Randomization, *Management Science*, 11, 831—846 (1965).
34. *Gaschütz G. K.*, *Ahrens J. H.*, Suboptimal Algorithms for the Quadratic Assignment Problem, *Naval Research Logistics Quarterly*, 15, 49—62 (1968).
35. *Geoffrion A. M.*, Strictly Concave Parametric Programming, Part I: Basic Theory, *Management Science*, 13, 244—253 (1966).
36. *Geoffrion A. M.*, Strictly Concave Parametric Programming, Part II: Additional Theory and Computational Considerations, *Management Science*, 13, 359—370 (1967).
37. *Geoffrion A. M.*, Reducing Concave Programs with Some Linear Constraints, *SIAM J. on Applied Mathematics*, 15, 653—664 (1967).
38. *Geoffrion A. M.*, Solving Bicriterion Mathematical Problems, *Operations Research*, 15, 39—54 (1967).
39. *Graves R. L.*, A Principal Pivoting Simplex Algorithm for Linear and Quadratic Programming, *Operations Research*, 15, 482—494 (1967).
40. *Hartley H. O.*, Non-Linear Programming by the Simplex Method, *Econometrica*, 29, 223—237 (1961).
41. *Hartley H. O.*, *Hocking R. R.*, Convex Programming by Tangential Approximation, *Management Science*, 9, 600—612 (1963).
42. *Hu T. C.*, Minimum-Cost Flows in Convex-Cost Networks, *Naval Research Logistics Quarterly*, 13, 1—9 (1966).
43. *Huard P.*, Dual Programs, в книге "Recent Advances in Mathematical Programming", R.L. Graves and Ph. Wolfe eds. McGraw-Hill, 1963, 55—62.
44. *Hwang C. L.*, *Fan L. T.*, A Discrete Version of Pontryagin's Maximum Principle, *Operations Research*, 15, 139—146 (1967).
45. *John F.*, Extremum Problems with Inequalities as Subsidiary Conditions, Studies and Essays' (Courant Anniversary Volume). N. Y., 1948, 187—204.
46. *Joksch H. C.*, Programming with Fractional Linear Objective Functions, *Naval Research Logistics Quarterly*, 11, 197—204 (1964).
47. *Kelly J. E.*, Jr., The Cutting-Plane Method for Solving Convex Programs, *SIAM J. on Applied Mathematics*, 8, 703—712 (1960).
48. *Kiefer J.*, Optimum Sequential Search and Approximation Methods under Minimum Regularity Assumptions, *SIAM J. of Applied Mathematics*, 5, 105—136 (1957).
49. *Kortanek K. O.*, *Evans J. P.*, Pseudo-Concave Programming and Lagrange Regularity, *Operations Research*, 15, 882—891 (1967).
50. *Kriebel C. H.*, Coefficient Estimation in Quadratic Programming Models, *Management Science*, 13, B473—B486 (1967).
51. *Kuhn H. W.*, *Tucker A. W*, Nonlinear Programming, 481—492, in Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Jerzy Neyman (ed.), University of California Press, 1950.
52. *Lemke C. E.*, A Method of Solution for Quadratic Problems, *Management Science*, 8, 442—453 (1962).
53. *Luenberger D. G.*, Quasi-Convex Programming, *SIAM J. on Applied Mathematics*, 16, 1090—1100 (1968).
54. *Mao C. T.*, *Särndal C. E.*, A Decision Theory Approach to Portfolio Selection, *Management Science*, 12, B323—B333 (1966).
55. *Martos B.*, Hyperbolic Programming, *Naval Research Logistics Quarterly*, 11, 135—155 (1964).

56. *McCormick G. P.*, Anti-Zig-Zagging by Bending, *Management Science*, **15**, 315—320 (1969).
57. *Miller C. E.*, The Simplex Method for Local Separable Programming, 89—100, in Recent Advances in Mathematical Programming, Robert L. Graves and Philip Wolfe (eds.), McGraw-Hill, 1963.
58. *Moore J. H.*, *Whinston A. B.*, Experimental Methods in Quadratic Programming, *Management Science*, **13**, 58—76 (1968).
59. *Pierskalla W. P.*, Mathematical Programming with Increasing Constraint Functions, *Management Science*, 15, 416—425 (1969).
60. *Rosen J. B.*, The Gradient Projection Method for Nonlinear Programming, Part I, Linear Cosntraints, *SIAM J. on Applied Mathematics*, **8**, 181—217.
61. *Rosen J. B.*, The Gradient Projection Method for Nonlinear Programming, Part II, Nonlinear Constraints, *SIAM J. on Applied Mathematics*, 9, 514—532 (1961).
62. *Rosen J. B.*, *Ornea J. C.*, Solution of Nonlinear Programming Problems by Partitioning, *Management Science*, **10**, 160—173 (1963).
63. *Roy A. D.*, Safety First in the Holding of Assets, *Econometrica*, **20**, 431—449 (1952).
64. *Sanders J. L.*, A Nonlinear Decomposition Principle, *Operations Research*, 13, 266—271 (1965).
65. *Scarf H. E.*, The Approximation of Fixed Points of a Continuous Mapping, *SIAM J. on Applied Mathematics*, **15**, 1328—1343 (1967).
66. *Sengupta J. K.*, *Tintner G.*, *Millham C.*, On Some Theorems of Stochastic Linear Programming with Applications, *Management Science*, **10**, 143—159 (1963).
67. *Sharpe W. F.*, A Simplified Model for Portfolio Analysis, *Management Science*, 9, 277—293 (1963).
68. *Theil H.*, van de *Panne C.*, Quadratic Programming as an Extension of Classical Quadratic Maximization, *Management Science*, **7**, 1—20 (1960).
69. *Topkis D. M.*, *Veinott A. F.*, Jr., On the Convergence of Some Feasible Direction Algorithms for Nonlinear Programming, *SIAM*, *J. on Control*, 5, 268—279 (1967).
70. *Tucker A. W.*, Linear and Nonlinear Programming, *Operations Research*, **5**, 244—257 (1957).
71. van de *Panne C.*, *Whinston A.*, Simplicial Methods for Quadratic Programming, *Naval Research Logistics Quarterly*, **11**, 273—302 (1964).
72. *van de Panne C.*, *Whinston A.*, A Comparison of Two Methods of Quadratic Programming, *Operations Research*, **14**, 422—441 (1966).
73. *Veinott A. F.*, Jr., The Supporting Hyperplane Method for Unimodal Programming, *Operations Research*, 15, 147—152 (1967).
74. *Wagner H. M.*, Linear Programming and Regression Analysis, *Journal of the American Statistical Association*, **54**, 206—212 (1959).
75. *Wagner H. M.*, Non-linear Regression with Minimal Assumptions, *Journal of the American Statistical Association*, **57**, 572—578 (1962).
76. *Whinston A.*, Conjugate Functions and Dual Programs, *Naval Research Logistics Quarterly*, **12**, 315—322 (1965).
77. *Whinston A.*, The Bounded Variable Problem — An Application of the Dual Method for Quadratic Programming, *Naval Research Logistics Quarterly*, **12**, 173—179 (1965).
78. *Wolfe P.*, The Simplex Method for Quadratic Programming, *Econometrica*, **27**, 382—398 (1959).
79. *Wolfe P.*, A Duality Theorem for Nonlinear Programming, *Quarterlly of Applied Mathematics*, 19, 239—244 (1961).
80. *Wolfe P.*, Accelerating the Cutting Plane Method for Non-Linear Programming, *SIAM J. on Applied Mathematics*, 9, 481—488 (1961).
81. *Zionts S.*, Programming with Linear Fractional Functionals, *Naval Reserach Logistics Quarterly*, 15, 449—451 (1968).
82. *Zoutendijk G.*, Nonlinear Programming: A Numerical Survey, *SIAM J. on Control*, **4**, 194—210 (1966).

Оглавление