ЦЕЛОЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ И СВЯЗАННЫЕ С НИМИ ЭКСТРЕМАЛЬНЫЕ ПРОБЛЕМЫ

THOMAS L. SAATY
University of Pennsylvania

OPTIMIZATION IN INTEGERS AND RELATED EXTREMAL PROBLEMS

From a course given at the University of California, Los Angeles, and at the George Washington University

McGraw-Hill Book Company
New York St. Louis San Francisco London
Sydney Toronto Mexico
Panama
1970

Т. СААТИ

ЦЕЛОЧИСЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ И СВЯЗАННЫЕ С НИМИ ЭКСТРЕМАЛЬНЫЕ ПРОБЛЕМЫ

Перевод с английского
В. Н. ВЕСЕЛОВА

Под редакцией
И. А. УШАКОВА

Издательство «Мир»
МОСКВА 1973

УДК 62.505

В книге просто, но в то же время со всей необходимой математической строгостью изложены вопросы целочисленной оптимизации. Рассмотрены проблемы оптимизации, возникающие при анализе диофантовых уравнений. Описан ряд задач геометрической оптимизации (раскрашивание графа, реализация графа с минимальным числом пересечений, наиболее плотная упаковка). Отдельная глава посвящена непосредственно целочисленному программированию. Изложение материала сопровождается большим числом интересных примеров и упражнений. В конце каждой главы приводится список литературы по затрагиваемым вопросам.

Книга является хорошим пособием для преподавателей, аспирантов и студентов технических вузов и университетов по специальностям: исследование операций, системотехника и прикладная математика, а также представляет большой интерес для инженеров и математиков, сталкивающихся в своей деятельности с решением различных задач оптимизации в целых числах.

Редакция литературы по новой технике

С 3314-153/041(01)-73

Предисловие редактора русского издания

Дискретный математический анализ является одним из древнейших и в то же время одним из современнейших направлений математики. С одной стороны, сюда относятся и одна из самых интригующих проблем математики — решение диофантовых уравнений, названных так по имени древнегреческого математика Диофанта, жившего в третьем веке нашей эры, и задачи, связанные с числами Фибоначчи, которые также имеют многовековую историю. (Заметим, что теорема Ферма также имеет непосредственное отношение к диофантовым уравнениям.) С другой стороны, с этим направлением математики неразрывно связаны получившие большое развитие в последнее время алгоритмические методы, важной ветвью которых является математическое программирование. Наблюдаемый всплеск в развитии этих алгоритмических методов связан в первую очередь с разработкой электронных вычислительных машин, позволяющих доводить решения в алгоритмической форме до практических результатов. Можно сказать, что дискретная математика и в первую очередь методы дискретной оптимизации переживают в наши дни поистине второе рождение.

Предлагаемая вниманию читателей книга Томаса Саати представляет собой своеобразную коллекцию задач и методов оптимизации в целых числах, составленную на основе курсов лекций, прочитанных автором в Калифорнийском университете Лос-Анджелеса и в университете Дж. Вашингтона. Автор — крупный американский математик, известный советским читателям по его книгам «Математические методы исследования операций», Воениздат, 1962, и «Элементы теории массового обслуживания», изд-во «Сов. радио», 1965. Эти монографии пользуются большим успехом у наших инженеров и научных работников, поскольку в них на строгом, но в то же время доступном уровне изложены современные модели и методы исследования операций. Не случайно и к американскому изданию этой книги Саати предпослал эпиграф: «Я люблю обе стороны математики: чистую — как возвышенный уход от реальности, прикладную — как страстное стремление к жизни». Такое отношение к математике пронизывает почти все работы автора, но, возможно, оно более всего проявляется именно в этой книге.

Книга «Целочисленные методы оптимизации и связанные с ними экстремальные проблемы» является в определенном смысле уникаль-

ной: ни в отечественной, ни в зарубежной научно-технической литературе нет работ, в которых воедино сводились бы разнообразные вопросы целочисленной оптимизации. Правда, следует заметить, что в настоящее время нет цельной теории этой ветви математики, что привело автора к своеобразной композиции и манере изложения материала. Весь материал разбит на несколько смысловых групп, в каждой из которых приводится много интересных иллюстративных примеров и красивых математических этюдов. В каком-то смысле эту книгу можно сравнить с живописью Матисса: те же неожиданно броские фрагменты, то же калейдоскопическое чередование их, столь же скупое и подчас умышленно сжатое изложение материала и в то же время такое же ощущение композиционной и логической цельности, остающееся от всей работы.

В книге можно найти множество интересных идей, а также увидеть многие вопросы в новом и даже неожиданном освещении. Например, хорошо описано здесь так называемое псевдобулево программирование. Много внимания автор уделил систематическому изложению методов геометрической оптимизации. По духу и стилю изложения это не учебник по методам дискретной оптимизации, а очень хорошее пособие для инженеров и научных работников, занимающихся математическими вопросами исследования операций и технической кибернетики, которые прочтут эту книгу не только с пользой, но и с большим удовольствием.

В заключение следует отметить плодотворное сотрудничество Саати с нами в процессе работы над русским переводом его книги. Он написал предисловие к русскому изданию, прислал ряд исправлений, а также коренным образом переработал один из разделов, посвященных геометрической оптимизации.

И. Ушаков

Предисловие автора к русскому изданию

Один крупный английский математик высказал мысль, что он не понимает желания людей познакомиться со всеми последними достижениями математики, поскольку эта наука возникла несколько тысяч лет назад, будет развиваться еще тысячи лет и ничего страшного не произойдет, если что-то не станет известным сразу же.

Однако существуют области, в которых математика добилась таких существенных результатов, что нельзя ждать тысячи лет, пока ими заинтересуются. Такой областью является целочисленная оптимизация, которая в настоящее время является основным математическим аппаратом исследования операций.

Во всем мире проводятся глубокие исследования в области целочисленного программирования. Со времени публикации этой книги в 1970 г. появилось несколько монографий, посвященных различным вопросам этой области вычислительной математики. Однако круг тем, затрагиваемых в данной книге, намного шире.

Для тех, кому нравится прослеживать математические доказательства, изложение материала может показаться слишком нестрогим. Эта книга больше подходит для тех, кто желает ознакомиться с проблемой в целом.

Т. Саати

Предисловие

В современной жизни почти нет сфер деятельности, в которых математика не играла бы важной роли. Некоторые из них сами порождают математические идеи. Пожалуй, чаще всего в прикладных задачах используются фундаментальные понятия математики — понятия максимума и минимума.

Человек является прирожденным максимизатором и минимизатором. Короче говоря, его можно назвать оптимизатором. Он занимается оптимизацией, потому что ему необходимо экономить свои ограниченные запасы энергии, способности и ресурсы. При построении теорий он оптимизирует путем использования упрощений, а также посредством поиска регулярности и симметрии. Человек также проявляет свою склонность к оптимизации путем построения изящных математических доказательств. Он оптимизирует, чтобы сократить продолжительность работы. Поиск максимумов и минимумов неотъемлем от существа человека с его поисками красоты и совершенства и с его неудержимым стремлением к рационализму. Кажется, что сама природа требует от человека выбирать наилучшую стратегию, которая максимизирует его выигрыш, постоянно думать о том, как сделать этот выигрыш побольше при относительно малых затратах.

Три короткие истории служат иллюстрацией того, что стратегия оптимизации в задачах с ограничениями не является ни чистой максимизацией, ни чистой минимизацией, а скорее некоторым оптимумом или смесью того и другого.

Первая история о трех швеях, ни одна из которых не хотела отстать от подруг. Первая вдела в свою иглу очень длинную нитку, которая скоро запуталась в шитье, и работа остановилась. Вторая швея заметила ошибку первой и взяла очень короткую нитку, но скоро ей пришлось вдевать в свою иглу новую нитку, и она тоже потеряла время. Третья «благоразумно» отмерила нитку длиной от носа до кончиков пальцев вытянутой руки. Она избежала затруднений, с которыми столкнулись ее подруги.

Во второй истории рассказывается об умирающем богаче, который завещал каждому из своих сыновей столько земли, сколько тот сможет обойти, выйдя из определенного места на восходе солнца и вернувшись туда же на закате. Тот, кто не успевал вернуться до заката, должен был потерять свою долю наследства. Один сын

ушел очень далеко и не смог вернуться до наступления темноты. Другой сын, не желая терять своей доли, сделал короткий путь и вернулся к полудню. Только третий сын оказался не слишком жадным, не слишком осторожным и воспользовался всем днем, чтобы отмерить себе участок земли.

В третьей истории рассказывается, как однажды некоего математика попросили оценить количество бобов в большом контейнере неправильной формы, объем которого он сумел оценить, мысленно преобразовав его в цилиндр. Затем он принял, что бобы имеют форму сферы, измерил их и оценил диаметр; применив свои знания о наиболее плотной упаковке сфер в трехмерном пространстве, он получил ответ, очень близкий к действительности, и затем скорректировал его с учетом того, что бобы имеют не сферическую форму. Этот математик получил премию за хорошее решение.

Под оптимумом иногда понимается равновесие между «не слишком много» и «не слишком мало». В других случаях он иногда понимается как «самое лучшее» или «нет ничего лучше». Можно ассоциировать с оптимумом понятие «самое худшее» или «нет ничего хуже».

Иногда оптимум называют максимум максиморумом, если рассматривается лучший из лучших исходов, или минимум миниморумом — в случае худшего из худших исходов. Конечно, все эти понятия имеют строгую математическую интерпретацию. Например, если непрерывная функция на замкнутом ограниченном множестве в евклидовом пространстве имеет несколько локальных максимумов, то наибольший из них иногда называется оптимумом, или максимум максиморумом.

Некоторые простые задачи, для которых надо показать, что решение существует, могут быть вложены в соответствующую схему оптимизации, для которой можно найти ответ. Возьмем, например, задачу о покрытии обычной шахматной доски, в которой квадраты в нижнем левом и в правом верхнем углах исключены, пластинками домино. (Предполагается, что каждая пластинка покрывает два соседних квадрата доски целиком, а сами пластинки не перекрываются.) Эту задачу можно сформулировать в терминах максимального количества пластинок домино, необходимых для покрытия как можно большего числа квадратов доски. При такой формулировке задачу оптимизации труднее решить, но она содержит ответ на предыдущую задачу. Иногда оказывается возможным трудную задачу максимизации вложить в более простую схему. Например, необходимые условия существования максимумов и минимумов в общем случае связаны с вопросом разрешимости уравнения или системы уравнений. В математическом анализе — это алгебраические уравнения, в вариационном исчислении — дифференциальные уравнения.

Дополнительные возможности, возникающие при замене одной задачи на другую, могут позволить получить решение исходной задачи. Так, в задаче о шахматной доске, поставленной в виде задачи о максимизации, число пластинок домино не может превышать 30.

Действительно, ясно, что каждая пластинка должна покрывать два смежных квадрата разного цвета, а на доске осталось на два белых квадрата больше, чем черных (вспомним, что два черных квадрата исключены), и, следовательно, доску нельзя покрыть 31 пластинкой.

Во многих отношениях достоин сожаления тот факт, что мы привыкли отождествлять все задачи с непрерывными задачами аналитической геометрии. Каждая задача, в которой фигурируют поверхности, мысленно ассоциируется с непрерывной задачей, при этом теряется полнота представления различных сторон дискретных задач. Мы должны переосмыслить этот старинный и вводящий в заблуждение подход, имея в виду две схемы, предназначенные для решения дискретных задач: старую и новую.

Дискретная математика в отличие от непрерывной не имеет единой теории, и поэтому здесь исследования, естественно, сводятся к изучению частных случаев. Такой подход развивает широкий взгляд на проблему в целом и способствует возникновению идей, оказывающихся полезными при исследовании различных проблем дискретной математики.

Задачу оптимизации можно представить и в геометрической и в алгебраической форме. Геометрический подход можно проиллюстрировать на примере задачи об упаковке максимального количества идентичных сфер в параллелепипед. Чтобы получить представление о задаче, можно построить физическую модель упаковки апельсинов в ящик и искать пути улучшения упаковки. Можно также изучать влияние формы ящика на число упакованных апельсинов, если ящик не велик по размеру. Горизонтальный разрез через центры апельсинов в правильно определенном слое (если такой слой существует) приводит к идее об упаковке кругов в прямоугольнике. Алгебраические задачи связаны со строгой формулировкой, в рамках которой должна максимизироваться или минимизироваться функция, возможно, при наличии ограничений, задаваемых в виде уравнений или неравенств. При целочисленной оптимизации необходимо, чтобы переменные были целыми числами, часто неотрицательными. Иногда требуется, чтобы максимальное или минимальное значение также было целым числом. Желательно было бы свести все задачи оптимизации к формулировкам в алгебраической форме. Однако для некоторых задач сделать это труднее, чем дать другое решение с использованием геометрических рассуждений.

Существуют две большие новые области оптимизации, развитие которых имеет решающее значение для большого числа теоретических и прикладных задач. Во-первых, это оптимизация над дискретными множествами, что является предметом изложения данной книги. Вторая область, привлекающая внимание исследователей,— это стохастическая оптимизация, в которой допускается использование выражений со случайными величинами. Здесь читатель сможет найти ответы для сложных социальных и политических задач.

Книга предназначена для студентов старших курсов, специализирующихся по математическим и техническим дисциплинам, социальным наукам и исследованию операций, которые имеют хорошую подготовку по математическому анализу и линейной алгебре. Некоторые теоремы формулируются без доказательств. Цель книги в том, чтобы пробудить у студентов интерес к различным аспектам излагаемого здесь предмета. Дополнительный материал читатель может найти в источниках, указанных в списке литературы.

Эта книга, объем которой дает возможность изучить ее в течение одного семестра, является первой книгой такого рода. Ввиду недостатка места обычно приводится только одно доказательство, достаточное для понимания материала и для того, чтобы дать студентам представление о типах доказательств, встречающихся в этой области.

При решении задач целочисленной оптимизации возникает сложная гамма чувств от возбуждения до сомнения и иногда даже до разочарования. В непрерывном случае можно найти компромиссную постановку задачи и получить хорошее приближенное решение. В дискретном случае требования гораздо сильнее и они не столь гибки. Здесь из всего пространства исследователь должен выбрать сравнительно немного точек и работать с ними. Поэтому, если область оптимизации менее богата точками и имеет бо́льшие ограничения, трудности умножаются. Иногда поиск оптимума является не более чем навязчивой идеей, так как полезность полученных результатов не оправдывает затраченных усилий. Часто бывает достаточно построить хорошие верхнюю и нижнюю границы точного решения.

Существуют задачи, для которых можно доказать, что некоторый максимум в качестве априорной верхней границы не может быть превышен. В таком случае следует доказать отдельно, что максимум действительно достигается. Например, используя правильные многоугольники и тот факт, что по меньшей мере три многоугольника должны сходиться в вершине многогранника, можно доказать, что существует самое большее (максимальное число) пять правильных выпуклых многоугольников. Однако, чтобы доказать, что их ровно пять, надо давать отдельное доказательство.

Другие интересные вопросы в задачах определения максимумов и минимумов — это теоремы *существования* и *единственности*, *описание* свойств решения, *построение* его, *сходимость* используемых алгоритмов и *приближение* к истинному значению желаемого решения.

Эта книга не предназначена для того, чтобы дать исчерпывающее изложение всех дискретных задач оптимизации. Наша более скромная цель состоит в том, чтобы возбудить и стимулировать у читателей интерес к элементарным методам и идеям дискретной оптимизации и смежным задачам.

Томас Л. Саати

Глава 1

Основные понятия: примеры задач и методов

1.1. Введение

При рассмотрении максимумов и минимумов на множествах всегда имеется в виду некоторое соотношение между элементами множеств, которое позволяет их сравнивать. Для примера возьмем множество, элементами которого являются три действительных числа $E = \{a, b, c\}$ с обычным отношением порядка. Если $a < b < c$, то a является *минимумом*, или наименьшим элементом E, тогда как c является наибольшим элементом, или *максимумом*. В множестве E не существует элемента, меньшего чем a, и также не существует элемента, большего чем c. Заметим, что b не является ни минимумом, ни максимумом, так как имеется элемент, который превосходит его, а также существует другой элемент, меньший b.

Нам известны и другие виды порядка, кроме соотношений «меньше, чем» и «больше, чем», существующих между числами. Так, например, совокупность множеств можно упорядочить в соответствии с включением множеств. В этом случае максимальным элементом будет наибольшее множество в совокупности, а минимальным элементом — наименьшее множество в совокупности. Например, в последовательности из пяти концентрических окружностей внешнюю окружность можно считать наибольшей, а внутреннюю — наименьшей, потому что первая окружность содержит все остальные, а наименьшая содержится во всех других.

Понятия максимума и минимума тесно связаны с понятиями нижней и верхней границ. Верхней границей множества чисел E является число c, такое, что ни один элемент из E не превосходит c. Если, например, $E = \{a, b\}$, $a < b$, то c будет верхней границей, если $a \leqslant c$ и $b \leqslant c$. Точной верхней границей множества действительных чисел является верхняя граница, которая не превосходит никакой другой верхней границы или мажорируется любой другой верхней границей. Можно принять, что множество верхних границ E имеет наименьший элемент, так как оно ограничено снизу (элементом b). Этот наименьший элемент может принадлежать или не принадлежать множеству E. Если среди элементов множества имеется максимум, то он и является точной верхней границей множества. Множество E действительных чисел имеет максимум, если в нем содержится число c, такое, что $x \leqslant c$ при всех $x \in E$. Таким образом,

максимум является точной верхней границей, но обратное неверно. Если точная верхняя граница принадлежит множеству, то она является максимумом. Точной верхней границей множества $\{1-1/n\}$, $n = 1, 2, \ldots$, в пределе является 1, но это множество не содержит этот элемент и, следовательно, не имеет максимума. В конечном множестве целых чисел точная верхняя граница и максимум совпадают. Аналогично определяются точные нижние границы и минимумы.

Из вышеизложенных рассуждений очевидно, что при изучении максимумов и минимумов необходимо ввести какое-то упорядочение. Но это можно сделать только для подмножества всего множества. Например, множество комплексных чисел, т. е. чисел вида $a + bi$, где a и b — действительные числа и $i = \sqrt{-1}$, нельзя упорядочить простым использованием соотношения $\leqslant$ без дополнительных условий на соответствующие действительные и мнимые части упорядоченных пар. Однако действительные числа, которые составляют подмножество множества комплексных чисел, могут быть совершенно упорядочены в соответствии с соотношением $\leqslant$. Вообще говоря, не очевидно, можно ли установить отношение порядка между элементами n-мерного евклидова пространства E_n так, чтобы получилось совершенное упорядочение пространства. Однако ситуация не является совершенно безнадежной. Лексикографическое упорядочение [1]) представляет собой упорядочение обычного типа, при котором точки $(x_1, \ldots, x_n)$ пространства E_n, координатами которых служат действительные числа, располагаются в соответствии с величиной первой координаты x_1; если первые координаты равны, то в соответствии с величиной второй координаты x_2 и т. д. Меньшим элементом (меньшей точкой) является тот, в котором раньше появляется меньшая координата. Так, например, (1, 7, 2) меньше, чем (1, 9, 0), потому что первые координаты равны, а вторая координата второго элемента 9 превышает 7. Естественно, что такое искусственное упорядочение должно гармонировать с частным случаем E_1, а именно с пространством действительных чисел, для которых отношение порядка уже хорошо определено.

Однако можно вводить упорядочение внутри каждого из нескольких подмножеств множества, не распространяя его на все множество. Например, можно рассмотреть множество, элементы которого доминируют (превосходят или предшествуют) один над другим в соответствии с правилом, определяемым несвязным направленным графом, стрелки которого указывают направление подчинения от управляющего к управляемому. На вопрос, кто является начальником подобной совокупности двух организаций, нельзя дать ответ, так как неизвестно отношение подчиненности. Такое множество является

---

[1]) Примером лексикографического упорядочения может служить алфавитный порядок слов в обычных словарях.— *Прим. ред.*

только частично упорядоченным, но не совершенно упорядоченным [1]).

В большинстве задач на максимумы и минимумы отыскивается точка множества, которая дает максимум или минимум функции, определенной на множестве и принимающей действительные значения. Это, вообще говоря, оправдано только тем, что множество действительных чисел является совершенно упорядоченным. Если бы отображение осуществлялось на множество векторов, то возникала бы задача упорядочения и максимумы и минимумы потребовали бы специального исследования.

Задачи оптимизации, которые будут рассматриваться здесь, связаны с отысканием максимумов или минимумов функций, определенных на множествах точек действительной прямой, плоскости и в более общем случае E_n, координатами которых служат целые действительные числа. Множества, на которых в этом случае может быть определена функция, называются *дискретными*, потому что их точки изолированы друг от друга. Таким образом, если в качестве окрестностей в основном пространстве (т. е. в пространстве, в которое вложено рассматриваемое множество) выбраны сферы, то около каждой точки можно построить такую сферу, что она не будет содержать никаких других точек данного множества. Однако понятия оптимизации можно использовать и при рассмотрении множеств, которые являются частично дискретными. В этом случае вокруг некоторых точек можно построить сферы в основном пространстве, в которых не содержится никаких других точек множества.

Будем предполагать, что функция, которую надо максимизировать или минимизировать, принимает действительные значения. В некоторых задачах будем считать, что функция принимает только целочисленные значения, в других не будем делать это предположение. Однако может оказаться, что максимум функции должен быть целым числом.

---

[1]) Возможны и другие схемы частичного упорядочения множеств, например когда устанавливается отношение порядка между отдельными подмножествами, но отсутствует отношение порядка для элементов внутри каждого из подмножеств. Чисто формальным примером может служить упорядочение комплексных чисел по их модулю. В этом случае можно упорядочить комплексные числа по принципу $a_1 + ib_1 > a_2 + ib_2$, если $a_1^2 + b_1^2 > a_2^2 + b_2^2$, но при этом подмножество комплексных чисел $a + ib$ с равными модулями (множество точек окружности на комплексной плоскости с радиусом $\sqrt{a^2 + b^2}$ и с центром в нуле) упорядочить не удается. В качестве примера, близкого к тому, который приведен автором, можно рассмотреть иерархическую структуру организации, где на каждом уровне функционирует совет равноправных членов, каждый из которых доминирует над членом нижестоящего совета и в свою очередь подчиняется любому члену вышестоящего совета.— *Прим. ред.*

Далее в этой главе приводятся определения, некоторые полезные идеи и теоремы и дается краткое изложение материала, касающегося рассматриваемых здесь вопросов для непрерывного случая, когда можно использовать методы математического анализа. В конце главы приводятся примеры задач дискретной оптимизации.

Остальная часть книги разделена на две части. В гл. 2 и 3 рассматриваются задачи, которым дается геометрическая или физическая формулировка. В гл. 4 и 5 используется чисто алгебраический подход. В гл. 4 обсуждаются некоторые элементарные диофантовы задачи. В гл. 5 излагается целочисленное программирование и кратко обсуждаются некоторые полезные алгоритмы.

Для того чтобы можно было перейти к обсуждению и использованию этих идей, нужно дать некоторые определения, но так, чтобы изложение не стало скучным, так как большинство определений можно найти в стандартных широко распространенных учебниках. Конечно, наша цель состоит в том, чтобы разработать методы отыскания максимумов и минимумов на дискретных множествах, где классические методы анализа обычно не могут быть использованы. Некоторые задачи даются в геометрической постановке, другие — в алгебраической. Было бы желательно разработать стандартный подход, но в настоящее время это трудно сделать, потому что при алгебраическом подходе часто искажается естественная постановка задачи.

На протяжении всей книги символ $[x]$ означает наибольшее целое число, которое не превосходит x. При различении случаев использования этого символа и простых квадратных скобок не должно возникать никаких затруднений.

1.2. Элементарные определения и полезные теоремы

Порядок

Отношение является первичным понятием, которое указывает на соответствие между элементами множеств.

Бинарное отношение R во множестве E можно рассматривать как множество упорядоченных пар. Рассмотрим упорядоченную пару (x, y), в которой x занимает первую позицию, а y — вторую позицию. (Этим определяется упорядочение в паре x, y.) Говорят, что элемент y соответствует элементу x при отношении R. Таким образом, записываем yRx. Так как множество всех упорядоченных пар образует декартово множество $E \times E$, то отношение, определенное на E, является подмножеством $E \times E$. Множество первых элементов упорядоченных пар, определенных отношением R, называется *областью определения*, а множество вторых элементов называется *множеством принимаемых значений*, или *множеством образов*. Оба эти множества являются подмножествами множества E, на котором определено R.

Так же как и отношение, можно определить функцию как соответствие между двумя множествами, между множествами точек

определения и принимаемых значений, так что каждой точке из множества определения соответствует одна точка множества принимаемых значений.

Важным классом бинарных отношений являются отношения эквивалентности.

Определение. Отношение эквивалентности на множестве E определяется следующими свойствами:

1. xRx для любого $x \in E$ (рефлексивность).
2. Если xRy, то yRx для любых $x, y \in E$ (симметричность).
3. Если xRy, а yRz, то xRz для любых $x, y, z \in E$ (транзитивность).

Примерами отношения эквивалентности являются конгруэнтность и равенство.

Определение. Множество называется частично упорядоченным, если на множестве E задано отношение на элементах E, такое, что для $x, y, z \in E$:

1. $x \leqslant x$ (рефлексивность).
2. Из $x \leqslant y$, $y \leqslant z$ следует $x \leqslant z$ (транзитивность).
3. Из $x \leqslant y$, $y \leqslant x$ следует $x = y$ (антисимметричность).

Определения частичного упорядочения и цепей будут использоваться при обсуждении локальных максимумов и минимумов.

Определение. Говорят, что множество E с отношением $\leqslant$, определенным между некоторыми его элементами, образует частично упорядоченную систему.

Возможно, что для некоторой пары элементов $\bar{x}$, $\bar{y}$ ни $\bar{x} \leqslant \bar{y}$, ни $\bar{y} < \bar{x}$.

Определение. Множество E называется совершенно упорядоченным, просто упорядоченным или упорядоченным, если, кроме того, имеет место следующее:

4. Для каждой пары элементов $x, y \in E$ или $x \leqslant y$, или $y \leqslant x$.

Часто совершенно упорядоченное подмножество частично упорядоченного множества называется *цепью*, если свойство 4 выполняется для любых двух элементов подмножества. Цепь называется максимальной (или связной), если она имеет вид $x_0 < x_1 < \ldots < x_n$, где x_i следует за (или покрывает) x_{i-1} при всех i.

Определение. Упорядоченное множество E называется вполне упорядоченным, если каждое непустое подмножество содержит наименьший элемент.

Положительные числа образуют вполне упорядоченное множество, а действительные нет.

Замечание. Заметим, что хорошо известный принцип математической индукции применим только к вполне упорядоченным множествам. Если такое упорядочение интерпретируется соответствующим

образом, то оказывается возможным применение принципа индукции для вполне упорядоченных множеств. Об индукции на частично упорядоченных множествах мало что известно.

Границы, максимумы и супремумы

Определение. Если E — множество действительных чисел, то верхней границей E является число y, такое, что для каждого x из E $x \leqslant y$. Точной верхней границей E называется верхняя граница, которая не превосходит любую другую верхнюю границу.

Целые числа представляют собой пример множества, которое не имеет верхней границы. Множество может иметь только одну точную верхнюю границу, потому что если y_1 и y_2 — две такие границы, то $y_1 \leqslant y_2$, $y_2 \leqslant y_1$ и, следовательно, $y_1 = y_2$.

Определение. Максимумом множества E действительных чисел является элемент E, который в то же время является верхней границей E. Максимум E, если он существует, является единственным и служит точной верхней границей E.

Определение. Целочисленной решеткой называется частично упорядоченная система, в которой каждые два элемента x, y имеют точную нижнюю границу и точную верхнюю границу.

Совокупность S подмножеств множества E является совокупностью *конечного* характера, если для каждого подмножества e из E $e \in S$ тогда и только тогда, когда $e_1 \in S$, где e_1 — любое конечное подмножество e.

Принцип максимума. Совокупность подмножеств S множества E конечного характера имеет максимальный член при частичном упорядочении S путем включения множеств.

Многие задачи оптимизации решаются в точках целочисленной решетки E_n, т. е. в точках, все координаты которых — целые числа. Цель состоит в том, чтобы найти точку целочисленной решетки E_n, которая дает максимум или минимум данной функции, при наличии дополнительных ограничений в виде равенств или неравенств, также определенных в точках целочисленной решетки.

Если положить, что $\max(x_1, \ldots, x_n)$ означает максимум множества действительных чисел $\{x_1, \ldots, x_n\}$, а $|x|$ представляет абсолютную величину, определенную соотношением

$$|x| = \begin{cases} x, & \text{если } x \geqslant 0, \\ -x, & \text{если } x < 0, \end{cases}$$

то можно (хотя и сложно) выразить максимум множества действительных чисел в замкнутой форме. Действительно, не только максимумы и минимумы, но также и промежуточные величины могут быть представлены в замкнутой форме.

Теорема 1.1. $max\,(x, y) = {}^1/_2\,(|\,x - y\,| + x + y)$.

Доказательство. Если $x \geqslant y$, то $x - y \geqslant 0$ и $|\,x - y\,| = x - y$, откуда следует искомый результат, так как x является максимумом. Если $x \leqslant y$, то y является максимумом; поскольку $x - y \leqslant 0$, получаем $|\,x - y\,| = -x + y$, и правая часть равна y.

Теорема 1.2. $max\,(x, y, z) = {}^1/_4\,(|\,2z - |\,x - y\,| - x - y\,| + 2z + |\,x - y\,| + x + y)$.

Доказательство. $\max\,(x, y, z) = \max\,[z, \max\,(x, y)]$, и, применяя предыдущий результат, получаем

$$\max\,(x, y, z) = \frac{1}{2}\left(\left|z - \frac{|\,x-y\,|+x+y}{2}\right| + z + \frac{|\,x-y\,|+x+y}{2}\right),$$

откуда следует искомый результат.

Упражнение 1.1. Получите соответствующие выражения для минимумов.

Упражнение 1.2. Покажите, что справедливо равенство $\max\,(x_1, \ldots, x_n) = \max \times \{\max\,[\max\,(\ldots, x_{n-1})]\,x_n\}$ и запишите выражение для $\max\,(x_1, x_2, x_3, x_4)$. Покажите, что $\min\,(x_1, \ldots, x_n) = -\max\,(-x_1, \ldots, -x_n)$.

Пусть $\mathrm{int}\,(x, y, z)$ означает второе по величине число из трех чисел; получаем [16] следующую теорему:

Теорема 1.3. $int\,(x, y, z) = {}^1/_4\,(2x + 2y + |\,|\,x - y\,| - x - y + 2z\,| - |\,|\,x - y\,| + x + y - 2z\,|)$.

Доказательство.

$$\begin{aligned}\mathrm{int}\,(x, y, z) &= -[\max\,(x, y, z) + \min\,(x, y, z) - x - y - z] = \\ &= [x + y + z - \max\,(x, y, z) - \min\,(x, y, z)] = \\ &= [x + y + z - \max\,(x, y, z) + \max\,(-x, -y, -z)],\end{aligned}$$

откуда следует результат.

Если обозначить через $M_n^k\,(x_1, \ldots, x_n)$ k-е по величине число из n чисел $(x_1, \ldots, x_n)$, то, например, $M_3^2\,(x_1, x_2, x_3) = \mathrm{int}\,(x_1, x_2, x_3)$.

Теорема 1.4. *При* $x_i \neq x_j$, $i, j = 1, \ldots, n$, $2 \leqslant k \leqslant n - 1$,

$$M_n^k\,(x_1, \ldots, x_n) = M_3^2\,[M_{n-1}^k\,(x_1, \ldots, x_{n-1}),$$
$$M_{n-1}^{k-1}\,(x_1, \ldots, x_{n-1}),\ x_n].$$

Доказательство.

$$M_n^k = \begin{cases} M_{n-1}^k, & \text{если } M_{n-1}^k > x_n, \\ M_{n-1}^{k-1}, & \text{если } M_{n-1}^{k-1} < x_n, \\ x_n, & \text{если } M_{n-1}^{k-1} > x_n > M_{n-1}^k. \end{cases}$$

Можно без труда доказать эти неравенства относительно x_n. Получаем

$$M_n^k = \begin{cases} M_{n-1}^k, & \text{если } M_{n-1}^{k-1} > M_{n-1}^k > x_n, \\ M_{n-1}^{k-1}, & \text{если } x_n > M_{n-1}^{k-1} > M_{n-1}^k, \\ x_n, & \text{если } M_{n-1}^{k-1} > x_n > M_{n-1}^k. \end{cases}$$

Таким образом, делаем вывод, что

$$M_n^k = M_3^2(M_{n-1}^k,\ M_{n-1}^{k-1},\ x_n).$$

Применение. Какими должны быть размеры ковров, которые можно уложить на квадратный пол? Другими словами, какие условия, наложенные на величины сторон прямоугольника a и b, гарантируют, что его можно поместить внутри единичного квадрата?

Решение [20а]. Ковер может быть уложен внутри квадрата, если $\max(a, b) \leqslant 1$. Если $\max(a, b) > 1$ и ковер можно разложить, то его можно разложить и в положении, симметричном относительно диагонали квадрата, причем должно выполняться соотношение $a + b \leqslant \sqrt{2}$. Эти два условия можно объединить:

$$\min\left[\max(a, b),\ \frac{a+b}{\sqrt{2}}\right] \leqslant 1,$$

или, используя выражения

$$\max(x, y) = \frac{x+y+|x-y|}{2}$$

и

$$\min(x, y) = \frac{x+y-|x-y|}{2},$$

получаем

$$(1+\sqrt{2})(a+b)+|a-b|-|(1-\sqrt{2})(a+b)+|a-b|| \leqslant 4.$$

Примеры, иллюстрирующие использование границ

Часто встречаются задачи, в формулировке которых используются такие выражения, как «самое большее», «не более чем», «верхняя граница», «ограниченный сверху», «точная верхняя граница», «не менее чем», «по меньшей мере», и т. д.

Верхние и нижние границы полезны при оценивании числа решений уравнений и неравенств и при получении информации о предельных значениях, принимаемых функцией. Ввиду отсутствия формальных методов решения сложных задач оптимизации может возникнуть необходимость в проверке различных значений для отыскания оптимума. Верхняя граница числа возможных решений помогает оценить работу, необходимую для отыскания оптимума.

Полезно иметь оценку порядка величины числа точек, составляющих множество ограничений. Классической иллюстрацией анализа такого типа является следующий простой пример.

Требуется найти число целочисленных решений [6] $x^2 + y^2 \leqslant n$, где n — данное целое число. Эту оценку легко вычислить при малых значениях n. Действительно, сначала можно перечислить все решения. Существует одно решение при $n = 0$, пять при $n = 1$, девять при $n = 2$ и т. д. (их легко перечислить).

Точки с целочисленными координатами распределены на плоскости равномерно. Единичный квадрат соответствует в точности одной такой точке. Площадь круга приблизительно равна числу единичных квадратов, которые он содержит, и, следовательно, числу решений неравенства, потому что они лежат в вершинах квадратов.

Теорема 1.5. *Число целочисленных решений N неравенства*

$$x^2 + y^2 \leqslant n,$$

где n — целое число, приближенно вычисляется по формуле $N \approx \pi n$ с ошибкой $|N - \pi n| < 2\pi(\sqrt{2n} + 1)$.

Доказательство. Рассмотрим все точки декартовой плоскости с целочисленными координатами. Эти точки располагаются в вершинах единичных квадратов. Поставим в соответствие каждому единичному квадрату правую верхнюю вершину. Пусть S_n — площадь соответствующих квадратов, C_n^1 и C_n^2 — две окружности, концентрические с C_n, имеющие радиусы $\sqrt{n} - \sqrt{2}$ и $\sqrt{n} + \sqrt{2}$ соответственно. Площадь S_n полностью заключена в C_n^2 и содержит C_n^1. Так как площадь S_n равна N, получаем

$$\pi(\sqrt{n} - \sqrt{2})^2 < N < \pi(\sqrt{n} + \sqrt{2})^2.$$

Это дает

$$N \approx \pi n \text{ и } |N - \pi n| < 2\pi(\sqrt{2n} + 1),$$

что и завершает доказательство.

Упражнение 1.3. Покажите при помощи геометрических соображений, аналогичных изложенным выше, что число целочисленных решений неравенства

$$x^2 + y^2 + z^2 \leqslant n$$

приблизительно равно $^4/_3\pi n^{3/2}$.

Приведем теперь второй пример, иллюстрирующий использование границ.

Теорема 1.6. *Достаточным условием того, что по меньшей мере два ящика содержат одинаковое число объектов среди B ящиков, в которых содержится всего N объектов, является* [7]

$$N < \frac{B(B-1)}{2}.$$

Доказательство. Если никакие два ящика не содержат одинакового числа объектов, то полное число объектов N равно по меньшей мере

$$0+1+2+\ldots+(B-1)=\frac{B(B-1)}{2}$$

Таким образом, для того чтобы хотя бы два ящика содержали одинаковое количество объектов, достаточно, чтобы имело место неравенство

$$N<\frac{B(B-1)}{2}.$$

Определение. Числом Фибоначчи [1]) F_n является наибольшее целое число, ближайшее к

$$\frac{[(1+\sqrt{5})/2]^n}{\sqrt{5}}.$$

Теорема 1.7. *Количество чисел Фибоначчи, не превосходящих данное положительное число N, равно наибольшему целому числу, меньшему чем* [3]

$$\frac{\log[(N+1/2)\sqrt{5}]}{\log[(1+\sqrt{5})/2]}-1.$$

Доказательство. Соотношение $F_n \leqslant N$ выполняется тогда и только тогда, когда

$$\frac{[(1+\sqrt{5})/2]^n}{\sqrt{5}}<N+\frac{1}{2},$$

т. е. тогда и только тогда, когда

$$n<\frac{\log[(N+1/2)\sqrt{5}]}{\log[(1+\sqrt{5})/2]}.$$

Так как $F_1=F_2=1$, получаем искомый результат.

Следующие две теоремы дают еще два примера использования границ.

Теорема 1.8. *Величина* $\sum_{i=1}^{n}\varepsilon_i a_i$ *принимает по меньшей мере* C_{n+1}^2+1 *различных значений, где* $0<a_1<a_2<\ldots<a_n$ *и* $\varepsilon_i=\pm 1$.

Доказательство [20]. Наименьшее значение равно $C\equiv\sum_{i=1}^{n}(-a_i)$. Отправляясь от него, строим все другие возможные значения. Таким

[1]) Чаще числа Фибоначчи F_n определяют из рекуррентного соотношения, приведенного на стр. 41. Числа Фибоначчи имеют ряд очень интересных свойств и находят различные приложения в комбинаторике и других разделах математики. Подробнее см. [29*].— *Прим. ред.*

образом,

$$C < C + 2a_1 < C + 2a_2 < \ldots < C + 2a_n < C + 2a_n + \\ + 2a_1 < \ldots < C + 2a_n + 2a_{n-1} < C + 2a_n + 2a_{n-1} + \\ + 2a_1 < \ldots < C + 2\left(\sum_{i=1}^{n} a_i\right) = \sum_{i=1}^{n} a_i.$$

Следовательно, число различных значений суммы не менее чем

$$1 + n + (n-1) + (n-2) + \ldots + 2 + 1 = 1 + \frac{n(n+1)}{2}.$$

Упражнение 1.4. Докажите предыдущий результат по индукции. (Начните с $S = \sum_{i=1}^{k-1} a_i \geqslant \sum_{i=1}^{k-1} \varepsilon_i a_i$ и предположите, что последняя сумма имеет $(C_k^2 + 1)$ различных величин. Рассмотрите различные значения $a_k + S$, $a_k + S - a_{k-1}$, $a_k + S - a_{k-2}$, . . ., $a_k + S - a_1$. Таким образом, существует не менее чем $k + C_k^2 + 1 = C_{k+1}^2 + 1$ различных величин.)

Теорема 1.9. *Если* $\sum_{i=1}^{n} a_i = A$, a_i — *неотрицательные целые числа, то*

$$\sum_{i=1}^{n-1} a_i a_{i+1} \leqslant \frac{A^2}{4}.$$

Доказательство [19]. Если $a_k = \max(a_1, \ldots, a_n)$, то

$$\sum_{i=1}^{n-1} a_i a_{i+1} = \sum_{i=1}^{k-1} a_i a_{i+1} + \\ + \sum_{i=k}^{n-1} a_i a_{i+1} \leqslant a_k \sum_{i=1}^{k-1} a_i + a_k \sum_{i=k}^{n-1} a_{i+1} = \\ = a_k (A - a_k) = \frac{A^2}{4} - \left(\frac{A}{2} - a_k\right)^2 \leqslant \frac{A^2}{4}.$$

Равенство имеет место тогда и только тогда, когда $a_k = A/2$.

Обратимся теперь к изучению максимумов и минимумов функций, которые отображают множества в E_n на прямую E_1. Иногда эти множества состоят из точек целочисленной решетки, которые можно отобразить на множество целых чисел в E_1. Во всяком случае, существование ограничений означает, что функция определена на подмножестве, а не на всем пространстве.

1.3. Максимумы и минимумы функций, определенных на n-мерном евклидовом пространстве E_n

Непрерывный случай

Пусть $f(x_1, \ldots, x_n)$ — дважды дифференцируемая функция с непрерывными вторыми производными, которая отображает область D евклидова пространства E_n [пространство $(x_1, \ldots, x_n)$ наборов из n действительных чисел x_i, $i = 1, \ldots, n$, с метрикой $(\sum_{i=1}^{n} x_i^2)^{1/2}$] в множество действительных чисел. Пусть $f_{x_i}(x)$ и $f_{x_i x_j}(x)$, $i, j = 1, \ldots, n$, означают первые и вторые частные производные f.

Определение. Точка $x^0 \equiv (x_1^0, \ldots, x_n^0) \in D$ является точкой абсолютного максимума f, если $f(x) \leqslant f(x^0)$ при всех $x \in D$, $x \equiv (x_1, \ldots, x_n)$.

Определение. Точка $x^0 \equiv (x_1^0, \ldots, x_n^0) \in D$ является точкой относительного или локального максимума f, если существует $\varepsilon > 0$, такое, что f имеет абсолютный максимум в x^0 для всех $x \in D$, которые удовлетворяют условию

$$|x_i - x_i^0| < \varepsilon, \quad i = 1, \ldots, n.$$

Аналогично можно определить абсолютный и локальный минимум [для которого выполняется условие $f(x) \geqslant f(x_0)$].

Упражнение 1.5. Докажите, что максимум $f(x)$ достигается в той же точке, что и минимум функции $-f(x)$.

Определение. Значение функции в точке максимума или в точке минимума называется *экстремальным значением*, или *экстремумом*. В некоторых работах это значение называется оптимумом (например, минимальное значение функции стоимости или максимальное значение функции выпуска продукции).

Определение. Точка $x^0 \in D$ является точкой экстремума f, если в этой точке достигается максимум или минимум (абсолютный или относительный).

Определение. Точка $x^0 \in D$, в которой имеет место равенство $f_{x_i}(x^0) = 0$, $i = 1, \ldots, n$, является критической, или стационарной, точкой f.

Часто при исследовании экстремальных точек и экстремальных значений требуется анализ критических точек.

Определение. Функция $f(x_1, \ldots, x_n; y_1, \ldots, y_m)$ имеет седловую точку в $(x^0, y^0) \equiv (x_1^0, \ldots, x_n^0; y_1^0, \ldots, y_m^0)$ тогда и только

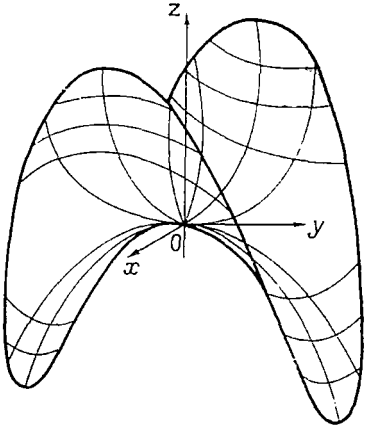
тогда, когда

$$f(x, y^0) \leqslant f(x^0, y^0) \leqslant f(x^0, y).$$

Смотри фиг. 1.1.

Определение. Якобианом системы функций $f_i(x_1, \ldots, x_n)$, $i = 1, \ldots, n$, называется определитель матрицы $\{\partial f_i/\partial x_j\}$, составленной из частных производных первого порядка.

Определение. Гессианом функции $f(x)$ называется матрица $\{f_{x_i x_j}(x)\}$, составленная из частных производных второго порядка.



Фиг. 1.1. Гиперболический параболоид $z = x^2 - y^2$. В малой окрестности точки (0, 0), которая является седловой точкой, имеются точки, в которых значение функции z больше, чем в точке (0, 0), и точки, в которых ее значение меньше, чем в точке (0, 0). Таким образом, начало координат не является ни максимумом, ни минимумом, а седловой точкой.

Эти понятия понадобятся в дальнейшем, например, для того, чтобы определить понятие выпуклости, которое помогает оптимизировать некоторые дискретные функции путем вложения.

Замечание. Если $f(x)$ имеет максимум в точке $x = (x_1, \ldots, x_n)$, то $\max_{x_1}\max_{x_2} \ldots \max_{x_n} f(x) = \max_{x_1, \ldots, x_n} f(x)$. Это соотношение справедливо при любой перестановке максимумов в левой части. Докажем это в случае двух переменных. Получаем

$$f(x, y) \leqslant \max_{x, y} f(x, y),$$

откуда следует

$$\max_x \max_y f(x, y) \leqslant \max_{x, y} f(x, y).$$

Чтобы доказать противоположное неравенство, заметим, что

$$f(x, y) \leqslant \max_y f(x, y) = f[x, y^*(x)] \equiv g(x),$$

$$\max_{x, y} f(x, y) \leqslant \max_{x, y} g(x) = \max_x g(x) = \max_x \max_y f(x, y),$$

так как $g(x)$ не зависит от y. Доказательство общего случая предоставляется читателям в качестве упражнения. Последующее пред-

ставляет собой небольшое отступление с целью распространения некоторых из этих идей на вариационное исчисление.

Определение. *Функционалом* называется отображение, которое ставит в соответствие каждому элементу абстрактного пространства действительное число.

Обычная функция, принимающая действительные значения, тоже является примером функционала. Определенный интеграл Римана, который ставит в соответствие интегрируемой функции (например, непрерывной функции) действительное число, представляет собой другой пример функционала.

Определение. Относительным экстремумом (максимумом или минимумом) функционала называется наименьшее значение функционала среди значений, полученных по кривым в данной окрестности. Если экстремум достигается на кривой $y_0(x)$ среди всех кривых, для которых $|y(x) - y_0(x)|$ мало (близость нулевого порядка), то он называется *сильным*. Если имеет место близость первого порядка, т. е. если $|y(x) - y_0(x)|$ и $|y'(x) - y_0'(x)|$ малы, то экстремум называется *слабым*. Таким образом, сильный экстремум одновременно является слабым, но обратное неверно.

В качестве примера уже приводилась последовательность $\{1 - 1/n\}$, где n пробегает все положительные целые значения. Мы говорили, что она не имеет максимума. Однако можно ввести понятие, которое в некоторых случаях может заменить понятие максимума.

Определение. Если f — принимающая действительные значения функция, определенная на множестве E (которое может быть дискретным), то $\sup_{x \in E} f(x)$ означает точную верхнюю границу (или супремум) множества F всех значений $f(x)$, т. е. множества F всех чисел y, таких, что $y = f(x)$ для некоторого $x \in E$. Аналогично $\inf_{x \in E} f(x)$ означает точную нижнюю границу (или инфинум) множества F всех значений $f(x)$. (Заметим, что $\sup_n \{1 - 1/n\} = 1$.)

Легко видеть, что если указанные границы существуют, то

$$\sup_{x \in E} [-f(x)] = -\inf_{x \in E} f(x)$$

и

$$\inf_{x \in E} [-f(x)] = -\sup_{x \in E} f(x).$$

Ниже часто будет встречаться случай, когда D является подмножеством точек целочисленной решетки в E_n.

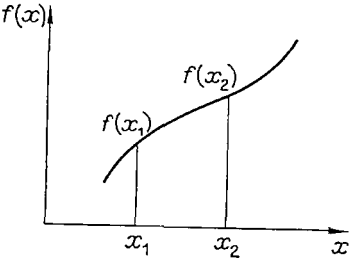
Определение. Точка $x \in D \subset E_n$, $x = (x_1, \ldots, x_n)$, называется точкой целочисленной решетки, если x_j — целое число, $j = 1, \ldots, n$. Совокупность точек целочисленной решетки называется *единичной решеткой* в E_n.

Замечание. Бывают также случаи, когда область D, на которой определена функция f, может состоять из точек, у которых лишь некоторые координаты — целые числа. Иногда f принимает только целые значения или даже только положительные целые значения. Область D обычно определяется при помощи системы ограничений, которые аналитически задаются в виде уравнений и неравенств типа $g_i(x) \leqslant 0$, $i = 1, \ldots, m$. Эта область представляет собой объединение двух множеств (внутренние точки и граничные точки). Граница D представляет собой поверхность, определенную уравнениями $g_i(x) = 0$, $i = 1, \ldots, m$. Каждое ограничивающее неравенство может быть сведено к уравнению путем введения неотрицательной «вспомогательной» переменной. Полученная в результате система определяет новую область в E_{n+m}. Например, ограничение $x + y \leqslant 0$ в E_2 путем введения вспомогательной переменной $z \geqslant 0$ превращается в $x + y + z = 0$ в верхнем полупространстве E_3.

Специальные функции

Введем теперь понятия *монотонности* и *выпуклости*, которые иногда используются при рассмотрении выражений, которые надо оптимизировать, и ограничений, задающих их область определения.

Определение. Функция $f(x)$ от одной переменной x называется монотонно возрастающей (убывающей), если $f(x_1) \leqslant (\geqslant) f(x_2)$ при $x_1 \leqslant x_2$. Она строго возрастает, если имеет место строгое неравенство (фиг. 1.2).



Фиг. 1.2.

Это определение можно распространить на функции от нескольких переменных, требуя монотонности по каждой переменной. Для монотонной всюду дифференцируемой функции $df(x)/dx \geqslant 0$, если она возрастающая, и $df(x)/dx \leqslant 0$, если она убывающая.

Замечание. Если $f(x)$ — дифференцируемая функция, то стационарные точки $f(x)$ (точки, в которых первая производная обращается в нуль), в которых $f(x)$ не равна нулю, совпадают со стационарными точками $\log f(x)$. Это верно потому, что $d/dx \log f(x) = f'(x)/f(x)$ и последнее выражение равно нулю, если $f'(x)$ равно нулю при условии, что $f(x)$ не обращается в нуль в этой точке x.

Упражнение 1.6. Пусть a_i, b_i, $i = 1, \ldots, n$, — положительные числа. Покажите, что

$$f(x) = \prod_{i=1}^{n} [xa_i + (1-x)b_i], \quad x \in [0, 1],$$

достигает максимума или при $x = 0$ или при $x = 1$ тогда и только тогда, когда

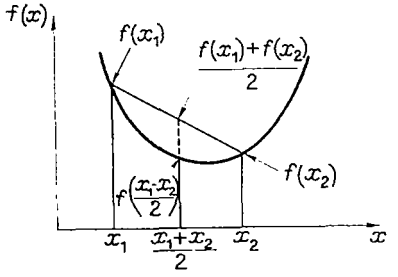
$$\left(\sum_{i=1}^{n} \frac{a_i - b_i}{a_i}\right)\left(\sum_{i=1}^{n} \frac{a_i - b_i}{b_i}\right) \geqslant 0.$$

Указание. Рассмотрите функцию $\log f(x)$ и наложите на нее условие монотонности, т. е. или $f'(x)/f(x) \geqslant 0$, или $f'(x)/f(x) \leqslant 0$. Обе граничные точки должны удовлетворять этому условию, какое бы из двух последних соотношений ни имело место; поэтому подставляем $x = 0$ и $x = 1$ в каждом случае. В обоих случаях получается записанное выше условие.

Определение. Функция $f(x_1, \ldots, x_n)$ называется выпуклой, если

$$f[\theta x + (1-\theta)\, y] \leqslant \theta f(x) + (1-\theta)\, f(y),$$

где $0 \leqslant \theta \leqslant 1$, $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$. Функция является вогнутой, если в определении выпуклой функции знак $\leqslant$ заменен на знак $\geqslant$. Выпуклость и вогнутость называются строгими, если имеет место строгое неравенство. (На фиг. 1.3 приведен пример выпуклой функции в E_2 с $\theta = {}^1/_2$.)



Фиг. 1.3.

Выпуклость и вогнутость дифференцируемой функции можно проверить при помощи определителей главных миноров гессиана функции f. Они должны быть неотрицательными для выпуклых функций и неположительными для вогнутых функций. В случае функции от одной переменной $f(x)$ получаем условие выпуклости $d^2f/dx^2 \geqslant 0$ и условие вогнутости $d^2f/dx^2 \leqslant 0$. Для функции от двух переменных $f(x, y)$ получаем условия выпуклости

$$f_{xx} \geqslant 0,\ f_{yy} \geqslant 0 \text{ и } f_{xx}f_{yy} - f_{xy}^2 \geqslant 0.$$

Упражнение 1.7. Получите условия, при которых $f(x, y)$ будет вогнутой функцией. Заметьте, что $(-f)$ должна быть выпуклой функцией.

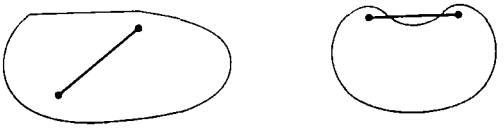
Упражнение 1.8. Покажите, что функция

$$f(x, y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(x^2 + y^2)\right],$$

которая представляет собой плотность двумерного нормального распределения, является вогнутой функцией над кругом $x^2 + y^2 \leqslant \sigma^2$.

Определение. Множество E называется выпуклым, если для всех $x \in E$ и $y \in E$ имеет место $\theta x + (1-\theta)\, y \in E$, где $0 \leqslant \theta \leqslant 1$.

Таким образом, если E содержит две точки x и y, то оно содержит и весь отрезок прямой, соединяющий их. Множество, изображенное



Фиг. 1.4.

на фиг. 1.4, *а*, является выпуклым, а множество на фиг. 1.4, *б* не является выпуклым, потому что часть сегмента, изображенного здесь, лежит вне множества.

Определение. *Выпуклой оболочкой множества* E *называется* наименьшее выпуклое множество, содержащее E.

Выпуклая оболочка области, имеющей форму звезды, получается путем соединения вершин звезды отрезками прямой.

При изучении максимумов и минимумов на дискретных множествах иногда полезно бывает вложить область определения в выпуклое множество, чтобы воспользоваться алгоритмами, разработанными для этого случая, и попытаться получить таким путем решения дискретной задачи.

Легко доказать, что если $x = (x_1, \ldots, x_n)$ и $f_i(x)$, $i = 1, \ldots, m$, — выпуклые (вогнутые) функции на выпуклом множестве E в E_n, то функция

$$f(x) = \sum_{i=1}^{m} f_i(x)$$

также является выпуклой (вогнутой) функцией на E. Чтобы доказать выпуклость f, заметим, что если $x^{(1)} = (x_1^{(1)}, \ldots, x_n^{(1)})$, $x^{(2)} = (x_1^{(2)}, \ldots, x_n^{(2)})$ принадлежат E и $0 \leqslant \theta \leqslant 1$, то

$$f[\theta x^{(1)} + (1-\theta)x^{(2)}] = \sum_{i=1}^{m} f_i[\theta x^{(1)} + (1-\theta)x^{(2)}] \leqslant$$

$$\leqslant \sum_{i=1}^{m} [\theta f_i(x^{(1)}) + (1-\theta) f_i(x^{(2)})] =$$

$$= \theta \sum_{i=1}^{m} f_i(x^{(1)}) + (1-\theta) \sum_{i=1}^{m} f_i(x^{(2)}) = \theta f(x^{(1)}) + (1-\theta) f(x^{(2)}).$$

Это полезный результат. Если дана функция, которая представляет собой сумму нескольких членов, каждый из которых является выпуклой функцией на некотором выпуклом множестве, то можно сделать вывод, что эта функция выпуклая. Если требуется получить целочисленный минимум, то можно затем применить методы, которые будут обсуждаться позже.

Упражнение 1.9. Покажите, что соотношение вида $g(x_1, \ldots, x_n) \leqslant 0$, где g — выпуклая функция, определяет выпуклое множество E, т. е. если $x^{(1)} = (x_1^{(1)}, \ldots, x_n^{(1)}) \in E$ и $x^{(2)} = (x_1^{(2)}, \ldots, x_n^{(2)}) \in E$, то также $\theta x^{(1)} + (1 - \theta) x^{(2)} \in E$, $0 \leqslant \theta \leqslant 1$.

Упражнение 1.10. Покажите, что пересечение выпуклых множеств представляет собой выпуклое множество. Затем установите, что если $g_i(x_1, \ldots, x_n) \leqslant 0$, $i = 1, \ldots, m$, где все g_i — выпуклые функции, то они в совокупности определяют выпуклое множество.

Замечание. При каких условиях функция имеет один минимум? Примером [20б] условия, которому должна удовлетворять функция, чтобы минимум был единственным, является строгая квазивыпуклость; т. е. если $f(x) \leqslant f(x^0)$, то $f(\theta x + (1 - \theta) x^0) < f(x^0)$ для $0 \leqslant \theta \leqslant 1$. Таким образом, выпуклая функция обладает свойством, что каждый локальный минимум является глобальным минимумом. Этот глобальный минимум является единственным, если имеет место строгая выпуклость или (более слабое условие) строгая квазивыпуклость.

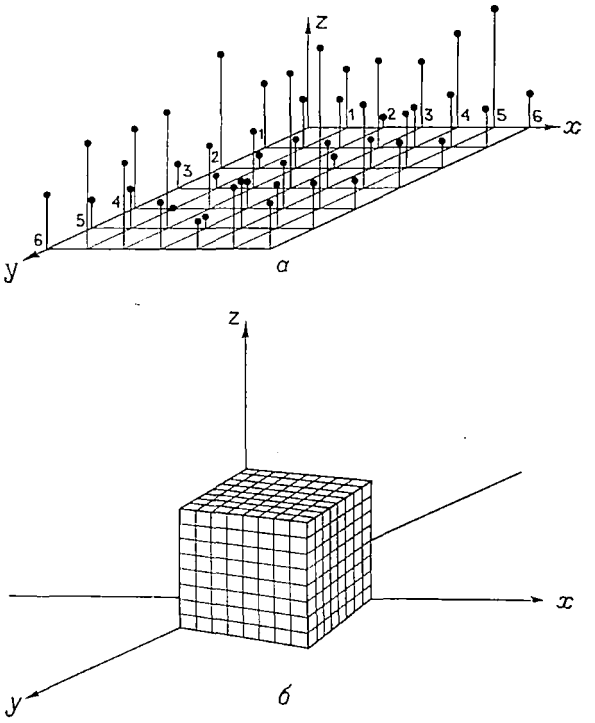
Дискретный случай

Приведенные определения глобального максимума и минимума функции $f(x_1, \ldots, x_n)$ сохраняются, если x_j, $j = 1, \ldots, n$, принимают только целые значения. Однако локальные максимумы и минимумы требуется определять более тщательно. Напомним, что условие обращения в нуль производной дифференцируемой функции от одной переменной в точке максимума x получено из соотношений

$$f(x) - f(x + \Delta x) \geqslant 0,$$

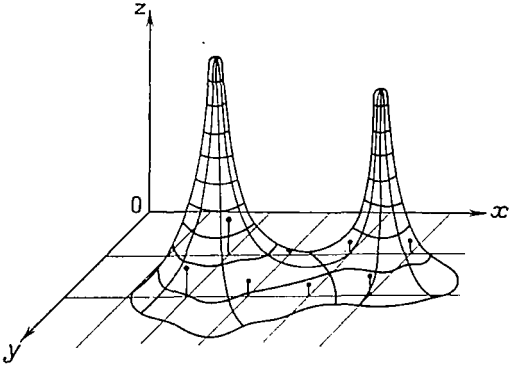
$$f(x) - f(x - \Delta x) \geqslant 0, \qquad \Delta x > 0.$$

Разделив первое неравенство на Δx, затем умножив на -1 и переходя к пределу при Δx, стремящемся к нулю, находим $df/dx \leqslant 0$. Если умножить второе неравенство на -1, затем разделить на $-\Delta x$ и перейти к пределу относительно $-\Delta x$, то получится $df/dx \geqslant 0$. Комбинируя эти два условия, находим, что в точке максимума $df/dx = 0$. Если x принимает дискретные значения, то нельзя придавать Δx произвольно малые значения. Если x принимает только целочисленные значения, то наименьшее значение, возможное для Δx, это $\Delta x = \pm 1$. Задача теперь состоит в том, чтобы посмотреть, когда следует пользоваться таким подходом. (На фиг. 1.5, *а* изображена функция, определенная в некоторых точках целочисленной решетки плоскости. Углы маленьких кубов на фиг. 1.5, *б* представляют собой точки целочисленной решетки E_3. На фиг. 1.6 показано, во что превращается непрерывная функция, если рассматривать только ее значения над точками целочисленной решетки.)



Ф и г. 1.5.

а — пример функции, определенной в точках решетки плоскости; б — иллюстрация трехмерной решетки, точками которой служат вершины маленьких кубов.



Ф и г. 1.6. От непрерывной функции останется несколько точек, если рассматривать только ее значения в точках решетки. Ее максимумы и минимумы могут быть безвозвратно потеряны.

Локальные максимумы и минимумы на дискретных множествах

Если функция определена на дискретном множестве и множество является частично упорядоченным, то вообще бессмысленно говорить о локальном максимуме и минимуме. Однако часто существуют неявные предположения о природе рассматриваемой функции, которые делают осмысленным понятие локального оптимума даже для дискретного множества. Это имеет место, когда функцию можно распространить на большее недискретное пространство, в которое вложено данное дискретное пространство, и класс всех расширений, сохраняющих природу функции, таков, что любые два члена этого класса эквивалентны в следующем смысле: они имеют локальные минимумы и локальные максимумы в дискретном пространстве в одних и тех же местах. В этом заключается причина, по которой иногда применимы методы анализа, с помощью которых исследуется расширение функции. Ниже приводится пример такого специального случая.

Предположим, что дана функция $f(x_1, \ldots, x_n)$, определенная в точках решетки E_n, и предположим, что природа функции f такова, что она разлагается на монотонные компоненты. Таким образом, $f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i)$. В этом случае применимо все то, что говорилось о вложении и расширении (на все E_n), и можно воспользоваться методами анализа, после чего для получения искомого ответа надо взять ближайшее целочисленное значение.

Имеют ли смысл локальные оптимумы в общем случае?

Заметим, что в определении локального максимума (или минимума) в случае непрерывной функции, определенной на E_n, предполагается возможным выбор $\varepsilon > 0$. В E_n мы имеем метрику и, следовательно, можно говорить о выборе положительного ε. Метрикой называется функция с действительными значениями, определенная для каждой пары элементов E_n. Будем обозначать метрику через $d(x, y)$; она удовлетворяет условию $d(x, y) = 0$ тогда и только тогда, когда $x = y$, $d(x, y) = d(y, x)$, $d(x, z) \leqslant d(x, y) + d(y, z)$. Метрика может не существовать на произвольном множестве, но понятие окрестности сохраняется. На множестве определена топология, если существует система подмножеств (называемых *открытыми множествами*, например открытые сферы в E_n), такая, что конечные пересечения и произвольные объединения элементов системы, все пространство и нулевое множество принадлежат этой системе. Для произвольного множества с топологией (называемого *топологическим пространством*) окрестностью точки называется всякое множество, которое содержит открытое множество, включающее данную точку. Близость в том смысле, в каком она понимается в метрических пространствах, может не существовать.

Вернемся к максимизации и минимизации. В общем случае если функция определена на произвольном множестве E, то немногое можно сделать для определения локального оптимума f, если только для E не задана некоторая структура, используя которую можно определить топологию. Получить такую структуру можно двумя путями: вложением и частичным упорядочением. Например, если множество точек представляет собой множество в E_n (т. е. оно вложено в топологическое метрическое пространство E_n), то естественный путь определения топологии на E (называемой *естественной топологией*) состоит в том, чтобы использовать относительную топологию, индуцированную на E стандартной топологией E_n. Открытыми множествами E являются все те множества, которые представляют собой пересечения E с открытыми множествами (открытые сферы, определенные метрикой) топологии на E_n. (Если множество дискретное, то каждая точка будет открытой.)

Также можно предположить, что существует частичное упорядочение на E. Тогда можно ввести естественную топологию на E путем рассмотрения всех максимальных цепей. Предположим, что мы имеем упорядоченное множество, состоящее из бесконечной последовательности $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n \leqslant \ldots \leqslant a$. Топология здесь является порядковой топологией, где окрестность a представляет собой множество всех точек x, таких, что $x \geqslant a_n$. Таким образом, f имеет локальный максимум в точке a тогда и только тогда, когда существует n, такое, что $f(x) \geqslant f(a_k)$ при всех $k \geqslant n$. Вопрос теперь заключается в том, как ввести топологию на частично упорядоченном множестве. Естественно, желательно было бы определить топологию так, чтобы окрестности для совершенно упорядоченных подмножеств были бы теми же, что и окрестности, только что данные для таких множеств. Это требование дает единственную топологию для любого данного частично упорядоченного множества. Если формально изложить то, что сейчас было сказано, то топология *определяется* следующим образом: если P — частично упорядоченное множество и S — подмножество P, то S будет открытым тогда и только тогда, когда пересечение S с каждой максимальной цепью в P является открытым в смысле порядковой топологии на цепи.

Изложенный в этом разделе метод является прямой противоположностью методу вложения в евклидово пространство, который обсуждался выше. При вложении подмножество получает свою естественную топологию от всего пространства. Здесь в частично упорядоченном множестве топология вводится таким образом, что совершенно упорядоченные подмножества сохраняют свою естественную топологию.

При определении топологии на цепи полезно ввести понятия *предбазиса* и *базиса*. Предбазисом называется совокупность открытых множеств, каждое из которых состоит из всех элементов, которые являются последующими или предшествующими для данного элемента; кроме того, вся цепь и пустое множество являются открытыми.

Базис состоит из всех конечных пересечений элементов предбазиса. Это дает топологию, которую теперь можно использовать для определения локального максимума функции в обычном смысле, а именно с помощью ранее данного определения открытого множества.

Для примера предположим, что функция определена на конечном частично упорядоченном множестве точек $P = (a, b, c, d, e, f)$, в котором цепями являются $(a \leqslant b \leqslant c;\ d \leqslant e;\ f)$. Очевидно, что максимальные цепи — это $P_1 = (a, b, c)$, $P_2 = (d, e)$, $P_3 = (f)$. Если $S = (a, c, d)$, то $S \cap P_1 = (a, c)$, $S \cap P_2 = (d)$. Чтобы увидеть, являются ли эти пересечения открытыми, заметим, что цепь (a, b, c) открытая и a — открытое множество, потому что это множество точек, предшествующих и последующих b. Объединение $a \cup c$ открытое, так как a и c открытые и, следовательно, $S \cap P_1$ — открытое множество. Аналогично подмножество $P_2 = (d)$ открытое. Заметим, что если топология определена на (f), то (f) должно быть открытым множеством.

Возвращаясь к топологии частично упорядоченного множества, заметим, что если порядок P не может быть описан конечным числом предложений, то определить естественную топологию на P невозможно, потому что для этого потребовалось бы более чем конечное число предложений и невозможно было бы определить окрестности и локальные максимумы.

Естественная топология на всяком конечном частично упорядоченном множестве предполагает, что каждая точка открытая и, следовательно, каждое подмножество открытое. Так как любая точка, в которой определена функция, открытая, то она является своей окрестностью. Поэтому значение функции в каждой точке является локальным минимумом и локальным максимумом. Таким образом, если функция определена на дискретном множестве точек и неявно *не предполагается распространять ее*, то говорить о локальных максимумах и минимумах практически бесполезно. Однако если подразумевается, что функция будет распространена, например, с точек координатной решетки на все евклидово пространство, то определение локальных максимумов и минимумов может представлять практический интерес.

Во многих задачах возможно как распространение функции с последующим использованием методов математического анализа, так и сужение области определения функции до заданного частично упорядоченного множества (топология вводится, как изложено выше). Для примера рассмотрим функцию $f(x_1, \ldots, x_n)$, определенную в точках решетки E_n. Вместо того чтобы продолжать f на все пространство E_n, можно ограничиться точками решетки и использовать индуцированную порядковую топологию; это приводит к следующему методу. Для того чтобы имел место локальный максимум в точке x_j при $x_j = x_j^0$ (т. е. максимум относительно цепи из точек решетки на линии $x_i = \text{const}$, $i \neq j$), можно потребовать выполнения соот-

ношения

$$f(x_1, \ldots, x_j^0, \ldots, x_n) \geqslant f(x_1, \ldots, x_j^0 \pm 1, \ldots, x_n), \quad j = 1, \ldots, n.$$

В более общем случае, чтобы имел место локальный максимум в точке $x = (x_1^0, \ldots, x_n^0)$, должно выполняться более сильное требование:

$$f(x_1^0, \ldots, x_n^0) \geqslant f(x_1^0 + \varepsilon_1, \ldots, x_n^0 + \varepsilon_n),$$

где $\varepsilon_j = 0, 1, -1, j = 1, \ldots, n$.

Таким образом, аргумент f в правой части в действительности принимает 3^n возможных значений [из которых $(x_1^0 + 0, \ldots, x_n^0 + 0)$ не является новым], и значение f в точке x должно превосходить значения f во всех ближайших соседних точках решетки. Аналогично можно дать определение локального минимума.

Приведенное выше условие локального максимума легко выразить для функции одной переменной. Если ввести оператор разности Δ (аналогично производной), который определяется соотношениями

$$\Delta f(x) = f(x+1) - f(x),$$
$$\Delta^i f(x) = \Delta^{i-1} \Delta f(x) = \Delta^{i-1} [f(x+1) - f(x)],$$

то относительный, или локальный, максимум в точке x_0 должен удовлетворять следующим необходимым условиям:

$$f(x_0) \geqslant f(x_0 - 1), \quad \text{т. е.} \quad \Delta f(x_0 - 1) \geqslant 0,$$
$$f(x_0) \geqslant f(x_0 + 1), \quad \text{т. е.} \quad \Delta f(x_0) \leqslant 0.$$

Эти два условия дают необходимое условие максимума

$$\Delta f(x_0) \leqslant 0 \leqslant \Delta f(x_0 - 1).$$

В аналогичном условии для локального минимума все неравенства изменены на обратные.

Достаточное условие абсолютного максимума в точке x_0 задается соотношением

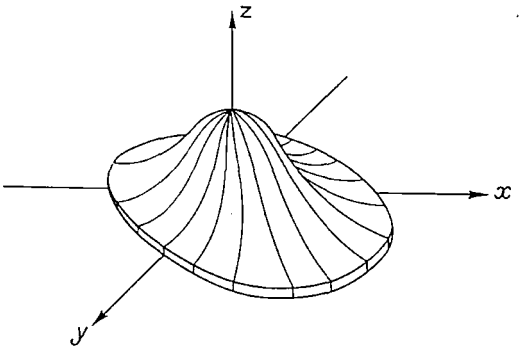
$$\Delta^2 f(x_0) \leqslant 0.$$

В случае абсолютного минимума неравенство направлено в обратную сторону. Достаточное условие абсолютного максимума в точке $x = x_0$ состоит в том, что функция $f(x)$ должна монотонно возрастать при $x \leqslant x_0$ [т. е. $f(x) \leqslant f(x_0)$, $x \leqslant x_0$] и монотонно убывать при $x \geqslant x_0$ [т. е. $f(x) \geqslant f(x_0)$, $x \geqslant x_0$]. В качестве примера можно привести точки, изображающие значения вогнутой функции, определенной на целых числах.

Пример. Рассмотрим функцию нормального распределения (фиг. 1.7)

$$f(m, n) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{1}{2\sigma^2}(m^2 + n^2)\right]$$

со средним в начале координат, определенную на всем множестве точек решетки плоскости. Определим, имеет ли она хотя бы одну



Фиг. 1.7.

точку максимума. Запишем $3^2 - 1$ неравенств

$$f(m, n) \geqslant f(m + \varepsilon_1, n + \varepsilon_2).$$

Прологарифмируем обе части. Так как логарифм — монотонно возрастающая функция, все неравенства сохраняются. Получим следующие соотношения:

$$\begin{aligned}
&\varepsilon_1 = 1, \quad \varepsilon_2 = 0, \quad x \geqslant -\tfrac{1}{2},\\
&\varepsilon_1 = 0, \quad \varepsilon_2 = 1, \quad y \geqslant -\tfrac{1}{2},\\
&\varepsilon_1 = -1, \quad \varepsilon_2 = 0, \quad x \leqslant \tfrac{1}{2},\\
&\varepsilon_1 = 0, \quad \varepsilon_2 = -1, \quad y \leqslant \tfrac{1}{2},\\
&\varepsilon_1 = 1, \quad \varepsilon_2 = 1, \quad x + y \geqslant -1,\\
&\varepsilon_1 = -1, \quad \varepsilon_2 = -1, \quad x + y \leqslant 1,\\
&\varepsilon_1 = 1, \quad \varepsilon_2 = -1, \quad x - y \geqslant -1,\\
&\varepsilon_1 = -1, \quad \varepsilon_2 = 1, \quad y - x \geqslant -1.
\end{aligned}$$

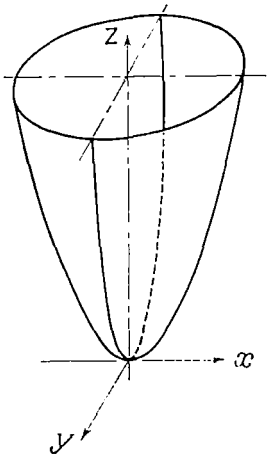
Первые четыре соотношения ограничивают значение (x, y) квадратом с вершинами $(^1/_2, {}^1/_2)$, $(^1/_2, -^1/_2)$, $(-^1/_2, {}^1/_2)$ и $(-^1/_2, -^1/_2)$. Точка $(0, 0)$ — единственная точка решетки, которая лежит в этом квадрате и в которой, следовательно, достигается локальный максимум. В этом случае последние четыре соотношения являются избыточными. В общем случае из них можно получить полезную информацию, особенно если они нелинейны по x и y.

Упражнение 1.11. В последнем примере докажите, что (0, 0) — глобальный максимум.

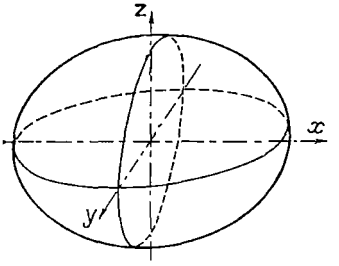
Упражнение 1.12. Покажите, что эллиптический параболоид (фиг. 1.8)

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=z$$

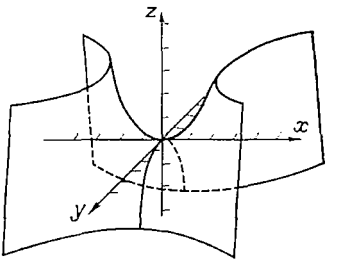
имеет в точках решетки единственный минимум в начале координат.



Фиг. 1.8.



Фиг. 1.9.



Фиг. 1.10.

Упражнение 1.13. Покажите, что эллипсоид (фиг. 1.9)

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$$

имеет одну точку максимума и одну точку минимума в точках решетки. Они находятся в начале координат.

Упражнение 1.14. Покажите, что гиперболический параболоид (фиг. 1.10)

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=z$$

не имеет ни максимумов, ни минимумов в точках решетки плоскости.

1.4. Классификация алгебраических задач

Выбор метода оптимизации зависит от того, какая информация об оптимизируемой функции имеется в распоряжении. Методы оптимизации можно классифицировать следующим образом:

I. Случай функции от одной или нескольких переменных, на которую не наложены ограничения.

A. Функция не задана в замкнутой форме.

1. Даны все значения функции (если функция принимает конечное число значений).

2. Оптимум ищется путем выбора отдельных значений функции. Таким образом, поиск оптимума заключается в определении лучших путей выбора отдельных значений с целью получения наилучшей оценки оптимального значения. Если известно, что функция унимодальна (т. е. имеет один горб), последовательность выбираемых значений может быть определена лучше, чем в случае мультимодальной функции. Чтобы при этих условиях оценить оптимум, можно воспользоваться методом поверхностей отклика и другими экспериментальными методами.

Б. Функция задана в замкнутой форме.

II. Случай функции, на которую наложены ограничения.

Для того чтобы можно было разработать эффективные методы решения, и функция и ограничения должны быть заранее описаны аналитически. Ограничения (которые для целей данного исследования должны быть заданы в алгебраической форме) могут быть заданы в виде равенств (диофантовых уравнений или систем уравнений в случае дискретной оптимизации) или неравенств. В дискретных задачах методы анализа помогают получить оценки оптимума, однако еще надо доказать, что они дают искомый ответ. Однако методы математического программирования позволяют непосредственно решать задачи оптимизации при наличии ограничений, в которых требуется получить целочисленные решения. Часто оптимум ищется на множестве положительных или неотрицательных значений аргумента.

Во многих геометрических задачах оптимизации сначала обычно определяется значение оптимума на основе информации, полученной из постановки задачи. Затем дается доказательство того, что это значение является искомым оптимумом. В некоторых геометрических задачах полезную роль при отыскании оптимума играет симметрия.

Методы изучения

Если требуется оптимизировать выражение при наличии ограничений, причем все они заданы аналитически в замкнутой форме, то, чтобы получить решение, обращаются к стандартным методам анализа, методу множителей Лагранжа или программированию. Когда функция известна не полностью, но часть ее значений может быть определена экспериментально, подход к отысканию оптимума становится более сложным и менее определенным. В общем случае целочисленное решение получить трудно и перечисленные здесь методы, вообще говоря, не могут быть использованы. Если возможно, то для получения целочисленных решений в качестве стандартной процедуры используется целочисленное программирование.

Случай IA

Функция не задана в замкнутой форме: даются все значения функции.

Напомним, что по теоремам 1.1 и 1.2 можно, хотя и сложно, дать замкнутое выражение для максимума из n действительных чисел $\alpha_1, \ldots, \alpha_n$. Практически максимум легче получить путем систематического сравнения.

Теорема 1.10. *Чтобы определить* $\max(\alpha_1, \ldots, \alpha_n)$, *где* $\alpha_1, \ldots, \alpha_n$ — *действительные числа, требуется в точности* $(n-1)$ *сравнений.*

Доказательство. Сравним α_1 последовательно с $\alpha_2, \ldots, \alpha_n$. Если $\alpha_1 \geqslant \alpha_i$, $i = 2, \ldots, n$, то $\max(\alpha_1, \ldots, \alpha_n) = \alpha_1$ и используется $(n-1)$ сравнений. Если $\alpha_1 \geqslant \alpha_i$, $i = 2, \ldots, k$, но $\alpha_1 \leqslant \alpha_{k+1}$, то $\alpha_1, \ldots, \alpha_k$ исключаются и процесс повторяется, начиная с α_{k+1}. Продолжая таким образом, приходим к выводу, что должно быть проведено $(n-1)$ сравнений.

Экспериментальный метод поиска максимума [11, 13, 28] [1])

Предположим, что при поиске максимума мы имеем возможность выбрать в точности n значений данной дискретной функции одной переменной $f(x)$. Начнем с выбора двух значений. При этом естественно продолжать поиск вблизи большего из двух значений. *Таким образом, выбор третьей точки производится возле большего значения.* Необходима схема, по которой можно было бы решить, где (слева или справа) и на каком расстоянии от исходной точки следует выбирать последующие значения.

Предположим, что интервал неопределенности после выбора $(n-1)$ точек имеет длину I_{n-1}, т. е. можно сказать, что наибольшее выбранное значение $f(x)$ лежит в I_{n-1}. Требуется расположить последнюю точку относительно предыдущей так, чтобы длина I_n следующего интервала неопределенности стала как можно меньшей.

Будем считать, что функция $y = f(x)$ унимодальная. Предположим, что точки на оси x выбираются, начиная с x_1 и кончая x_n. Точки x_k, $k = 1, \ldots, n$, не обязательно должны быть расположены в порядке возрастания. Заметим, что если, например, выбираются три точки $x_1 > x_2 > x_3$ на единичном интервале, на котором ищется максимум $f(x)$, и если соответствующие значения функции равны y_1, y_2, y_3, то

если $y_1 > y_2 > y_3$, максимум лежит в $[0, x_2)$,
если $y_1 < y_2 > y_3$, максимум лежит в (x_1, x_3),
если $y_1 < y_2 < y_3$, максимум лежит в $(x_2, 1]$.

[1]) В работе [28, стр. 47—54] дается подробное изложение предлагаемого метода.— *Прим. ред.*

Желательно было бы сделать интервал, содержащий максимум, как можно более тесным, так чтобы после выбора последней точки оценка максимума была бы как можно более близкой к истинному значению. Заметим, что для каждого выборочного плана $(x_1, \ldots, x_n)$ существует индекс α, которому соответствует некоторая величина x_α, при которой $y(x_\alpha) \equiv y_\alpha$ является максимумом.

Пусть $I_{n,\alpha}(x)$ — длина интервала, в котором находится максимум. Таким образом, $I_{n,\alpha}(x) = x_{\alpha+1} - x_{\alpha-1}$ и содержит внутри себя x_α. Пусть $L_n(x)$ — самый длинный из этих интервалов при любом выборе x. Поэтому

$$L_n(x) = \max\{I_{n,\alpha}(x)\} \quad \text{по всем } x.$$

С теоретической точки зрения каждой произвольной выборке объема n соответствует интервал, в котором находится максимум функции. Рассмотрим наибольший из этих интервалов по всем выборкам. Тогда задача формулируется как задача отыскания «наилучшего» выбора x, который обозначим через $\bar{x}$, такого, чтобы величина длины $\bar{L}_n \equiv L_n(\bar{x})$ была минимальной. В этом случае

$$\bar{L}_n = \min_x \max\{I_{n,\alpha}(x)\}.$$

Без потери общности примем, что функция определена на единичном интервале $[0, 1]$. Эвристически, чтобы определить оптимальную процедуру выбора, рассмотрим произвольный интервал $\bar{L}_j$, конечные точки которого представляют собой пару точек x_k, одна (или обе) из которых может быть конечной точкой интервала $[0, 1]$. Одна из этих конечных точек должна принадлежать интервалу $\bar{L}_{j-1}$. В общем алгоритме отыскания x рассуждения, использованные при выборе положения следующей точки, которая располагается внутри $\bar{L}_{j-1}$ и порождает, таким образом, $\bar{L}_j$, приводят к тому, что эта точка должна быть симметрично расположена относительно другой конечной точки $\bar{L}_{j-1}$, что даст $\bar{L}_{j+1}$. Причина этого заключается в том, что при отсутствии конкретных значений функции в двух конечных точках нет оснований отдавать предпочтение одному значению перед другим и, следовательно, надо выбирать две новые точки. Таким образом, две конечные точки $\bar{L}_{j-1}$ должны быть симметричны относительно наибольшего наблюдавшегося до сих пор значения функции. Это дает

$$\bar{L}_{j-1} = \bar{L}_j + \bar{L}_{j+1}. \tag{1.1}$$

Имеем следующие два случая:

1. Число выбираемых точек n задано. В этом случае естественно выбирать n-ю, или конечную, точку [1]), в середине $\bar{L}_{n-1}$. Поэтому

[1]) В работе [28] показано, что начальную точку следует выбирать на расстоянии $\bar{L}_2 = F_{n-2}/F_n$ от одного конца исходного интервала.— *Прим. ред.*

$\overline{L}_{n-1} = 2\overline{L}_n$. Комбинируя этот результат с уравнением (1.1), получаем

$$\overline{L}_{n-2} = \overline{L}_{n-1} + \overline{L}_n = 3\overline{L}_n,$$
$$\overline{L}_{n-3} = \overline{L}_{n-2} + \overline{L}_{n-1} = 5\overline{L}_n.$$

Заметим, что коэффициенты $\overline{L}_n$ представляют собой последовательность чисел Фибоначчи 1, 2, 3, 5, 8, . . ., которые получаются из рекуррентного соотношения

$$F_k = F_{k-1} + F_{k-2}, \quad k \geqslant 1,$$

где

$$F_0 = F_1 = 1.$$

Таким образом, предыдущие итерации можно, вообще говоря, представить в виде

$$\overline{L}_{n-k} = F_{k+1}\overline{L}_n.$$

Так как

$$\overline{L}_1 = 1 = F_n\overline{L}_n,$$

получаем

$$\overline{L}_n = \frac{1}{F_n}.$$

2. Число выбираемых точек n заранее не задано. Теперь нельзя воспользоваться соотношением $\overline{L}_{n-1} = 2\overline{L}_n$. Вместо этого попытаемся сделать постоянным отношение длин последовательных интервалов. Это дает

$$\frac{\overline{L}_{j-1}}{\overline{L}_j} = \frac{\overline{L}_j}{\overline{L}_{j+1}} = c. \tag{1.2}$$

Разделив обе части уравнения (1.1) на $\overline{L}_{j+1}$, получим

$$\frac{\overline{L}_{j-1}}{\overline{L}_{j+1}} = \frac{\overline{L}_j}{\overline{L}_{j+1}} + 1. \tag{1.3}$$

Подставляя выражение $\overline{L}_j^2 = \overline{L}_{j-1}\overline{L}_{j+1}$ вместо $\overline{L}_j$ в среднее соотношение (1.2), находим

$$\left(\frac{\overline{L}_{j-1}}{\overline{L}_{j+1}}\right)^{1/2} = c.$$

Таким образом, соотношение (1.3) превращается в

$$c^2 = c + 1$$

и $c = (1 + \sqrt{5})/2$ — единственный положительный корень. Тогда $\overline{L}_1 = 1$, и, следовательно, получаем из (1.2) $\overline{L}_2 = 1/c$ и далее последовательно $\overline{L}_n = 1/c^{n-1}$. Уайлд [28] называет этот второй метод «поиском посредством золотого сечения».

Можно показать, что

$$F_n = \frac{c^{n+1} - (-c)^{-(n+1)}}{\sqrt{5}} \approx \frac{c^{n+1}}{\sqrt{5}}.$$

Используя это приближение, можно сравнить величины $\bar{L}_n$ для случаев, когда n задано заранее и не задано. Вычисляя отношение, получаем

$$\frac{c^{n+1}}{\sqrt{5}\, c^{n-1}} = \frac{c^2}{\sqrt{5}} = 1{,}1708.$$

Обобщения на нелинейное целочисленное программирование изложены в работе [17а].

Наискорейший подъем [30*] вдоль поверхности, построенной по экспериментальным точкам

Проводились исследования [1, 2] статистических методов определения точки оптимума функции от нескольких переменных, не заданной в замкнутой форме, при отсутствии ограничений. Ниже будет кратко описана эта процедура.

Предположим, что взаимодействия нескольких факторов (x_1, ..., x_n) в эксперименте описываются неизвестной функцией $f(x_1, \ldots, x_n)$. Предполагая, что эта функция имеет оптимум, например максимум, требуется определить или оценить этот оптимум путем выбора точек экспериментов и построения поверхности по этим точкам, не пытаясь полностью найти $f(x_1, \ldots, x_n)$, так как обычно это очень длительный и дорогостоящий процесс.

Построение поверхности производится последовательно, чтобы можно было определить направление, в котором можно достичь максимума; следуя вдоль пути наискорейшего подъема, где-то надо остановиться и повторить процесс.

Одним из наиболее популярных методов является метод поочередного изменения факторов. Сохраняя значения всех контролируемых переменных, кроме одной, на определенном уровне и изменяя эту одну переменную, находят максимум; затем значение этой переменной, при котором достигается максимум, фиксируют и изменяют значение другой переменной и т. д.

Более надежная процедура, приводящая к максимуму, состоит в последовательном продвижении в направлении наискорейшего подъема; при этом в малых областях производится аппроксимация плоскостями или поверхностями второго порядка, что может быть оправдано разложением в ряды. При помощи метода наименьших квадратов требуется наилучшим образом провести плоскость на малом участке экспериментальной области. Направление наискорейшего подъема определяем, придавая переменным изменения, равные направляющим косинусам, которые в свою очередь пропорциональны коэффициентам аппроксимирующей плоскости. Затем вдоль этого пути находятся экспериментальные значения функции, до тех пор

пока они не начнут убывать. Повторение процедуры вокруг этой точки может показать, достигнут ли максимум или следует искать новый путь подъема. Если аппроксимирующая поверхность оказывается плоской, это не значит, что достигнут максимум. Поверхность может иметь гребень или может встретиться седловая точка. Необходимо дальнейшее исследование, которое зависит от того, какого типа решение практической задачи ищется.

Иногда для аппроксимации можно использовать линейные и квадратичные члены разложения в ряд, а затем для определения коэффициентов применить метод наименьших квадратов. При увеличении числа коэффициентов для их определения требуется большее число выборочных точек, так как для того, чтобы построить аппроксимирующую функцию, содержащую N констант, необходимо по крайней мере N наблюдений. Сведение к канонической форме путем переноса или ортогонального вращения облегчает определение пути наискорейшего подъема через корни соответствующего характеристического уравнения.

Случай IБ и случай II

Если $f(x_1, \ldots, x_n)$ дважды дифференцируема по x_j, $j = 1, \ldots, n$, то необходимое условие оптимума имеет вид [22]

$$\frac{\partial f}{\partial x_j} \equiv 0, \quad j = 1, \ldots, n.$$

Достаточные условия можно выразить через главные диагональные определители матрицы, составленной из производных второго порядка, вычисленных в точке, которая служит решением приведенной выше системы. Все эти определители должны быть положительными. (Для случая двух переменных в условиях выпуклости нужно заменить знак $\geqslant$ на $>$.)

Если требуется оптимизировать $f(x_1, \ldots, x_n)$ при наличии ограничений

$$g_i(x_1, \ldots, x_n) \leqslant 0, \quad i = 1, \ldots, m,$$

составляется лагранжиан

$$F(x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_m) = f - \sum_{i=1}^{m} \lambda_i g_i.$$

Параметры $\lambda_1, \ldots, \lambda_m$ называются множителями Лагранжа.

Например, лагранжиан функции $xyz = \max$ при ограничении $x^2 + y^2 + z^2 = r^2$ имеет вид

$$F(x, y, z, \lambda) = xyz - \lambda(x^2 + y^2 + z^2 - r^2).$$

1. Необходимыми условиями максимума в случае $g_i = 0$, $i = 1, \ldots, m$, являются

$$\frac{\partial F}{\partial x_j} = 0, \quad j = 1, \ldots, n,$$

$$\frac{\partial F}{\partial \lambda_i} = 0, \quad i = 1, \ldots, m.$$

Простейшие условия достаточности для специальных случаев будут обсуждаться ниже.

2. Случай $g_i \leqslant 0$, $i = 1, \ldots, m$, встречается часто, особенно если существуют дополнительные требования «неотрицательности»

$$x = (x_1, \ldots, x_n) \geqslant 0,$$

т. е.

$$x_j \geqslant 0, \quad j = 1, \ldots, n.$$

Эти условия можно записать в виде

$$-g_{m+1} \equiv -x_1 \leqslant 0, \ \ldots, \ -g_{m+n} \equiv -x_n \leqslant 0.$$

Предположим, что следующее условие выполняется в любой точке x^0 на границе при любом достаточно малом приращении dx относительно точки x^0:

$$\nabla g_i(x^0)\, dx < 0, \quad i = 1, \ldots, m,$$

если только

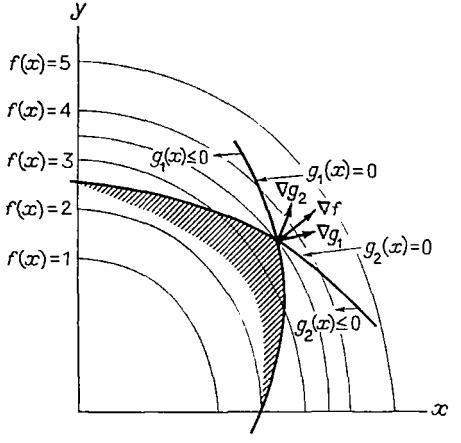
$$g_i(x^0) = 0, \quad \nabla \equiv \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right),$$

и $dx_j \geqslant 0$, если $x_j^0 = 0$. Тогда dx лежит в допустимой области. Эти условия предназначены для исключения вырожденных ситуаций, таких, которые могут встречаться в точке заострения. Пусть $G(x)$ — вектор-столбец g_i, $i = 1, \ldots, m$. Имеет место следующая теорема.

Теорема 1.11 (Куна — Таккера). *Необходимое условие того, что $f(x)$ достигает максимума в граничной точке $\bar{x}$ множества $G(x) \leqslant 0$ при $\bar{x} \geqslant 0$, состоит в том, что существуют $\lambda \geqslant 0$ и $\mu \geqslant 0$, такие, что* [17]

$$\nabla f(\bar{x}) = \lambda \nabla G(\bar{x}) - \mu,$$

где если $\bar{x}_j > 0$, то $\mu_j = 0$, и если $g_i(\bar{x}) < 0$, то $\lambda_i = 0$ (фиг. 1.11).



Фиг. 1.11. Оптимум на границе.

Заметим, что если $\bar{x}$ находится внутри области, определенной ограничениями, то $\bar{x} > 0$, $G(\bar{x}) < 0$ и, следовательно, $\mu = 0$, $\lambda = 0$ и условия сводятся к тому, что требуется обращение в нуль производной f.

Мы уже видели, что необходимые условия оптимизации функции при наличии ограничений в виде равенств в предположении дифференцируемости функции приводят к решению системы уравнений. Иногда можно, наоборот, от решения системы уравнений перейти к задаче оптимизации.

Возьмем функцию двух переменных $F(x, y)$, линейную по y. Можно считать y множителем Лагранжа и рассматривать $F(x, y)$ как лагранжиан в задаче оптимизации, в которой требуется оптимизировать функцию, которая представляет собой часть $F(x, y)$, не содержащую y. В этом случае коэффициент при y приравнивается к нулю как ограничение. Конечно, можно прийти к такой функции, интегрируя систему алгебраических уравнений, которую можно рассматривать как необходимое условие оптимума, полученное путем дифференцирования лагранжиана. Эту идею можно обобщить на случай нескольких переменных, по некоторым из которых функция линейна.

Пример. Следующий пример может показаться слишком простым, но он приведен только для того, чтобы проиллюстрировать идею. Рассмотрим систему нелинейных уравнений

$$2x_i + y_i(\sin x_i + 1) = 0, \quad i = 1, \ldots, n.$$

Мы хотим выяснить, имеет ли эта система действительное решение по x_i, $i = 1, \ldots, n$, при некоторых y_i, $i = 1, \ldots, n$. Вычислив лагранжиан путем интегрирования по x_i и суммируя по i, получим

$$F(x_1, \ldots, x_n; y_1, \ldots, y_n) = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i(\cos x_i - x_i + c_i),$$

где c_i, $i = 1, \ldots, n$,— постоянные интегрирования, которые надо определить при заданных условиях. Лагранжиан можно было записать, отправляясь от задачи оптимизации: требуется найти

$$\max \sum_{i=1}^{n} x_i^2$$

при ограничениях

$$\cos x_i = x_i + c_i, \quad i = 1, \ldots, n.$$

Эта система имеет действительные ограниченные решения, и поэтому максимум существует. Следовательно, исходная система имеет действительное решение при некоторых значениях y_i, $i = 1, \ldots, n$.

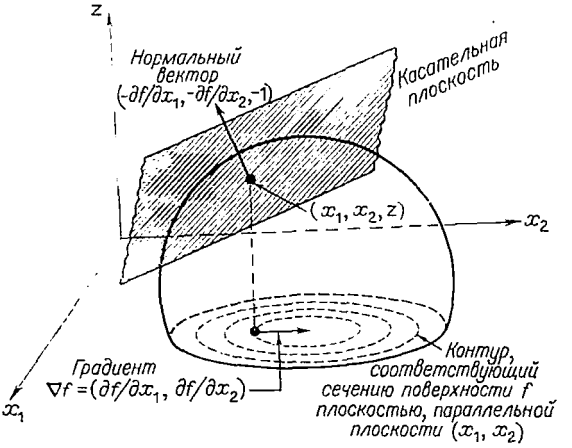
Градиент и градиентный метод

Градиентом функции $z = f(x_1, \ldots, x_n)$ называется вектор

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right).$$

Заметим, что прямая, перпендикулярная к касательной плоскости f в точке, представляет собой вектор в $(n+1)$-мерном пространстве, тогда как вектор ∇f находится в n-мерном пространстве, так как это вектор с n компонентами (фиг. 1.12). Градиент ∇f указывает в окрестности данной точки направление наискорейшего возрастания

функции или наискорейшего подъема вдоль контуров поверхности $z = f(x_1, \ldots, x_n)$, и, следовательно, $-\nabla f$ указывает направление наискорейшего убывания или наискорейшего спуска. Чтобы убедиться в этом, рассмотрим в окрестности точки поверхности произвольное направление $dx = (dx_1, \ldots, dx_n)$. Мы хотим выбрать dx так,



Фиг. 1.12.

чтобы оно указывало направление наискорейшего увеличения f в окрестности точки. Во-первых, заметим, что полное изменение f определяется ее полной производной

$$df = \frac{\partial f}{\partial x_1} dx_1 + \ldots + \frac{\partial f}{\partial x_n} dx_n = \nabla f \, dx = |\nabla f| \, |dx| \cos\theta.$$

Таким образом, величина df будет наибольшей, если $\theta = 0$ и, следовательно, направление dx совпадает с направлением градиента. Поэтому градиент указывает направление наискорейшего увеличения f.

Градиент можно использовать в итеративном методе (известном как *градиентный метод*) для отыскания максимума или минимума функции. Таким образом, если записать

$$x^{(k)} = x^{(k-1)} - \lambda \nabla f(x^{(k-1)}),$$

то можно надеяться на то, что при соответствующем выборе λ величина $x^{(k)}$ будет ближе к оптимуму, чем $x^{(k-1)}$. Исходная точка $x^{(0)}$ выбирается произвольно. При соответствующем выборе λ и $x^{(0)}$ можно доказать сходимость. Минимизация $f(x_1, x_2) = x_1^2 + x_2^2 - 2x_1x_2$ при $x^{(0)} = (x_1^{(0)}, x_2^{(0)}) = (1, 0)$ дает $\nabla f = (2x_1 - 2x_2, 2x_2 - 2x_1)$, $x^{(1)} = x^{(0)} - \lambda \nabla f(x^{(0)}) = (1, 0) - \lambda(2, -2) = (1 - 2\lambda, 2\lambda)$. Если подставить $x^{(1)}$ в f, то последняя станет функцией λ. Затем минимизируем ее как функцию λ при помощи соотношения $df/d\lambda = 0$.

Это дает $\lambda = {}^1/_4$ и точку $x^{(1)} = ({}^1/_2, {}^1/_2)$, в которой, как можно показать, используя вышеупомянутые условия достаточности, достигается минимум.

Упражнение 1.15. Требуется определить экстремумы (максимумы и минимумы) функции

$$f(x) = (x-2)^2, \quad 0 \leqslant x \leqslant 3,$$

и доказать, что $x = 0$ является глобальным максимумом, $x = 3$ — локальным максимумом, а $x = 2$ — глобальным минимумом.

Начертите график и дайте интерпретацию результатов.

Упражнение 1.16. Докажите, что функция

$$f(x_1, x_2) = \frac{x_1^2 + x_2^2}{2}, \quad 0 \leqslant x_1 \leqslant 1, \quad 0 \leqslant x_2 \leqslant 1,$$

имеет

а) глобальный минимум при $x_1 = x_2 = 0$;

б) глобальный максимум при $x_1 = x_2 = 1$.

Докажите, что $x_1 = 1$, $x_2 = 0$ и симметрично $x_1 = 0$, $x_2 = 1$ не являются ни максимумами, ни минимумами.

Начертите график и дайте интерпретацию результатов.

Упражнение 1.17. Найдите максимум функции xy при ограничении $x^2 + y^2 = 25$.

Упражнение 1.18. Найдите максимум функции xyz при ограничении

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} - 1 = 0.$$

Упражнение 1.19. Определите размеры прямоугольной коробки с вырезанной верхней поверхностью, которая имеет максимальный объем, если площадь поверхности равна 108 см^2.

Упражнение 1.20. Найдите максимальное и минимальное значения функции

$$f(x, y) = 2x^2 + 2xy + 3y^2$$

в области $x^2 + y^2 \leqslant 1$, $x + y \geqslant 0$.

Упражнение 1.21. Требуется максимизировать

$$f(x, y) = \int_0^x e^{-t^2/2}\,dt + \int_0^y e^{-t^2/2}\,dt$$

при ограничениях $g(x, y) \equiv x + y \leqslant c$, $c > 0$, $x \geqslant 0$, $y \geqslant 0$.

Указание. Сначала покажите, что максимум находится не внутри области. Докажите, что f — вогнутая функция. (В соответствии с критерием выпуклости покажите, что определители главных миноров матрицы, составленной из производных второго порядка, неотрицательны при всех x и y. Чтобы доказать вогнутость функции, надо умножить ее на -1 и проверить выпуклость.) Следовательно, максимум находится на границе. Примените теорему Куна — Таккера.

Упражнение 1.22 (Джошуа Уиллард Гиббс). Пусть

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i) \quad \text{и} \quad \sum_{i=1}^{n} x_i = C, \quad x_i \geqslant 0.$$

Используя теорему 1.11, докажите, что необходимое условие того, что в точке $(x_1^0, \ldots, x_n^0)$ достигается максимум f при указанных ограничениях, заключается в существовании действительного числа λ, такого, что

$$f_i'(x_i^0) \equiv \frac{\partial f_i(x_i)}{\partial x_i}\Big|_{x_i = x_i^0} = \lambda, \quad \text{если } x_i^0 > 0,$$

$$f_i'(x_i^0) \equiv \frac{\partial f_i(x_i)}{\partial x_i}\Big|_{x_i = x_i^0} \leqslant \lambda, \quad \text{если } x_i^0 = 0.$$

Докажите, что это условие является также достаточным, если $f_i(x_i)$, $i = 1, \ldots, n$,— вогнутые функции, т. е. $f_i''(x_i) \leqslant 0$. После того как λ определено и в предположении вогнутости функций докажите, что достаточное условие задается соотношениями

$f_i'(0) > \lambda$ тогда и только тогда, когда $x_i^0 > 0$,

$f_i'(0) \leqslant \lambda$ тогда и только тогда, когда $x_i^0 = 0$.

Запишите необходимое условие минимума в данной задаче. В предположении выпуклости функций дайте достаточные условия.

Упражнение 1.23. Используя условия выпуклости из упражнения 1.22, покажите, что имеет место выпуклость функций, и минимизируйте выражение

$$\sum_{i=1}^{n} a_i e^{-b_i x_i}, \quad a_i, \ b_i > 0,$$

при ограничении

$$\sum_{i=1}^{n} x_i = 1, \quad x_i \geqslant 0.$$

Упражнение 1.24. Воспользуйтесь градиентным методом для получения точки минимума эллипсоида

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1.$$

Указание. Решите уравнение относительно z и воспользуйтесь нижней поверхностью.

Упражнение 1.25. Покажите, что задача решения системы $f_i(x_1, \ldots, x_n) = 0$, $i = 1, \ldots, n$, эквивалентна задаче отыскания точки $(x_1, \ldots, x_n)$, в которой достигается минимум выражения $\sum_{i=1}^{n} f_i^2$. Очевидно, что минимальное значение выражения равно нулю.

Метод Ньютона

При использовании метода множителей Лагранжа требуется решать системы из $(m+n)$ уравнений $f_i = 0$, $i = 1, \ldots, m+n$, с $(m+n)$ неизвестными. Как указывалось в упражнении 1.25, чтобы решить эти уравнения, их можно возвести в квадрат, сложить и при помощи градиентного метода достичь минимума, равного нулю. С другой стороны, их можно решить методом Ньютона. В этом случае задается произвольная начальная точка

$$x^{(0)} = (x_1^{(0)}, \ldots, x_n^{(0)}; \lambda_1^{(0)}, \ldots, \lambda_m^{(0)}) \equiv (x_1^{(0)}, \ldots, x_{m+n}^{(0)})$$

и получается выражение

$$x^{(k)} = x^{(k-1)} - [f_1(x^{(k-1)}), \ldots, f_{m+n}(x^{(k-1)})]\, J^{-1}(x^{(k-1)})$$

с помощью правила Крамера, которое применяется для решения линейной системы, полученной путем разложения в ряд каждой функции f_i относительно $x^{(k-1)}$ и отбрасывания всех членов, кроме линейных. (Член $J^{-1}(x)$ представляет собой матрицу, обратную матрице, составленной из производных первого порядка.) Таким образом,

$$f_i(x) = f_i(x^{(k-1)}) + \sum_{j=1}^{m+n} \frac{\partial f_i}{\partial x_j}\bigg|_{x^{(k-1)}} (x_j - x_j^{(k-1)}), \quad i = 1, \ldots, m+n$$

Заметим, что $f_i(x)$ равна нулю. Решая относительно x_j и используя это значение в качестве нового приближения $x_j^{(k)}$, получаем вышеприведенное выражение. (Относительно сходимости см. [22].)

Упражнение 1.26. Рассмотрите систему

$$2x^3 - y^2 - 1 = 0,$$
$$xy^3 - y - 4 = 0.$$

Начиная с $x^{(0)} = (1, 2; 1, 7)$, покажите, что метод Ньютона дает $x^{(1)} = (1, 23; 1, 66)$.

1.5. Примеры дискретной оптимизации функций в замкнутой форме: критерий достаточности

Проиллюстрируем некоторые методы решения дискретных задач. Этот раздел, вероятно, следовало бы включить в гл. 4, но он помещен здесь с целью введения некоторых понятий.

Теорема 1.12. *Величина* $\sum_{i,j=1}^{n} a_i b_j$ *достигает максимума, если выбирать* $i = j$, *где последовательности* a_i *и* b_j *заданы и удовлетворяют условиям*

$$a_1 > a_2 > \ldots > a_n > 0, \quad b_1 > b_2 > \ldots > b_n > 0.$$

Замечание. В указанную сумму каждое значение из последовательностей a_i и b_j входит только один раз.

Доказательство [27]. Проведем доказательство по индукции. При $n = 1$ очевидно, что утверждение справедливо. При $n = 2$ получаем

$$(a_1 b_1 + a_2 b_2) - (a_1 b_2 + a_2 b_1) = (a_1 - a_2)(b_1 - b_2) > 0.$$

Отсюда

$$a_1 b_1 + a_2 b_2 > a_1 b_2 + a_2 b_1.$$

Предположим, что теорема справедлива при $i = 1, \ldots, k$, $j = 1, \ldots, k$; покажем, что утверждение верно и при $k + 1$. Можно записать $a_i = a_{k+1} + p_i$, $b_i = b_{k+1} + q_i$, где $p_i > 0$, $q_i > 0$, $p_{k+1} = q_{k+1} = 0$. Подстановка дает

$$\sum_{i,j=1}^{k+1} a_i b_j = a_{k+1} b_{k+1} + \sum_{i=1}^{k+1} b_{k+1} p_i + \sum_{j=1}^{k+1} a_{k+1} q_j + \sum_{i,k=1}^{k+1} p_i q_j.$$

Первые три выражения в правой части являются константами, и, следовательно, они не зависят от того, равны i и j или нет. По предположению

$$a_{k+1} + p_i = a_i > a_{i+1} = a_{k+1} + p_{i+1}$$

и, следовательно, $p_i > p_{i+1}$. Аналогично $q_i > q_{i+1}$. Теперь если $\sum_{i,j=1}^{k+1} p_i q_j$ содержит член $p_{k+1} q_{k+1}$ (который равен нулю), то

$$\sum_{i,j=1}^{k+1} p_i q_j = \sum_{i,j=1}^{k} p_i q_j$$

достигает максимума при $i = j$. Рассмотрим теперь случай, когда появляются два члена, равные нулю: $p_{k+1} q_s$, $1 \leqslant s \leqslant k$, и $q_{k+1} p_t$, $1 \leqslant t \leqslant k$. Тогда сумму можно переписать следующим образом:

$$\sum_{i,j=1}^{k+1} p_i q_j = \sum_{i,j=1}^{k} p_i q_j - p_t q_s.$$

По индукции сумма в правой части достигает максимума, если $i = j$, и так как $p_t q_s > 0$, получаем

$$\sum_{i=1}^{k+1} p_i q_i = \sum_{i=1}^{k} p_i q_i > \sum_{i,j=1}^{k} p_i q_i - p_t q_s = \sum_{i,j=1}^{k+1} p_i q_j.$$

Таким образом, сумма становится максимальной, если $i = j$. Этим завершается доказательство.

Упражнение 1.27. Мужчина имеет $n \equiv 0 \bmod 5$ рубашек, и каждый рабочий день он надевает новую рубашку (всего за неделю уходит пять штук). Число рубашек, которые могут быть постираны за минимальную цену, кратно пяти. Он может отправлять рубашки в прачечную и забирать их только по субботам. Из прачечной можно получить рубашки только через календарную неделю. Каким образом он должен отправлять рубашку в прачечную, чтобы минимизировать число поездок в прачечную? Проанализируйте решение. Является ли решение задачи единственным? Дайте доказательства.

Упражнение 1.28. Покажите, что [15]

$$\sum_{m=1}^{M}\sum_{n=1}^{N}\min(m, n) = \frac{N(N+1)(3M-N+1)}{6}.$$

Обратите внимание, что

$$\sum_{m=1}^{M}\sum_{n=1}^{N}\min(m, n) = \sum_{n=1}^{N}\sum_{m=1}^{n} m + \sum_{m=n+1}^{M} n,$$

откуда следует результат.

В некоторых задачах оптимизации нелегко получить оптимум непосредственно. Вместо этого используются рассуждения, с помощью которых находят нижнюю оценку оптимума. С помощью других рассуждений получается верхняя оценка. Если эти две оценки совпадают, то это и есть искомый максимум; в противном случае максимум находится в интервале между нижней и верхней оценками.

Иногда методами математического анализа определяется предполагаемое значение оптимума дискретной функции. Таким образом, решение задачи начинается с вложения задачи максимизации $f(x_1, \ldots, x_n)$, где x_i принимают целочисленные значения, в более общую или расширенную задачу, в которой все x_i — действительные числа.

Теорема 1.13. *Положительное целое m, которое минимизирует функцию*

$$f(m, n) = 2m + 3 + \frac{n}{m},$$

представляет собой целое число, ближайшее к $\sqrt{n/2}$.

Доказательство 1. Используя определение локального минимума для дискретного случая, получаем

$$f(m, n) - f(m+1, n) = -2 + \frac{n}{m} - \frac{n}{m+1} \leqslant 0,$$

$$f(m, n) - f(m-1, n) = 2 + \frac{n}{m} - \frac{n}{m-1} \leqslant 0,$$

откуда следует

$$m(m-1) \leqslant \frac{n}{2} \leqslant m(m+1)$$

или

$$\left| \frac{n}{2m} - m \right| \leqslant 1.$$

Чтобы найти m, которое удовлетворяет этому неравенству, положим $(n/2m) - m = 0$, откуда имеем $m = \sqrt{n/2}$ и, так как m — целое число, выбираем $m = [\sqrt{n/2}]$.

Доказательство 2. Считая f непрерывной и дифференцируемой по m функцией, имеем

$$\frac{df}{dm} = \frac{m(4m+3) - (2m^2+3m+n)}{m^2} = \frac{4m^2+3m-2m^2-3m-n}{m^2} = 0,$$

или

$$2m^2 - n = 0.$$

Таким образом, $m = \sqrt{n/2}$.

Этим обосновывается выбор в качестве m целого числа, ближайшего к $\sqrt{n/2}$. Обозначим это целое число через $[\sqrt{n/2}]$. Подставляя его в f, получаем

$$2\left[\sqrt{\frac{n}{2}}\right] + 3 + \frac{n}{[\sqrt{n/2}]}.$$

Теперь мы доказываем, что

$$f\left(\left[\sqrt{\frac{n}{2}}\right], n\right) \leqslant f(m, n) \text{ при всех } m.$$

Запишем

$$2\left[\sqrt{\frac{n}{2}}\right] + 3 + \frac{n}{[\sqrt{n/2}]} \leqslant 2m + 3 + \frac{n}{m},$$

или

$$2\left[\sqrt{\frac{n}{2}}\right] + \frac{n}{[\sqrt{n/2}]} \leqslant 2m + \frac{n}{m} \text{ при всех } m.$$

После упрощений получаем соотношение

$$2\left\{\left[\sqrt{\frac{n}{2}}\right] - m\right\} \leqslant n\left\{\frac{[\sqrt{n/2}] - m}{m[\sqrt{n/2}]}\right\},$$

которое, как было показано, справедливо. Теперь если $m < [\sqrt{n/2}]$, то мы должны показать, что

$$2 \leqslant \frac{n}{m[\sqrt{n/2}]},$$

или

$$2m\left[\sqrt{\frac{n}{2}}\right]\leqslant n,$$

или

$$m\left[\sqrt{\frac{n}{2}}\right]\leqslant\frac{n}{2}=\sqrt{\frac{n}{2}}\sqrt{\frac{n}{2}},$$

или

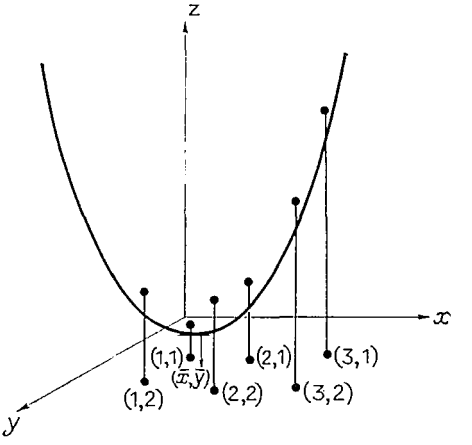
$$m\left(\sqrt{\frac{n}{2}}+\theta\right)\leqslant\sqrt{\frac{n}{2}}\sqrt{\frac{n}{2}},\qquad|\theta|\leqslant\frac{1}{2}.$$

Так как это соотношение должно выполняться при любом $-1/2\leqslant\theta\leqslant 1/2$, то должно иметь место неравенство $m\leqslant[\sqrt{n/2}]$. Если $m>[\sqrt{n/2}]$, то аналогичными рассуждениями можно доказать, что

$$2\geqslant\frac{n}{m\,[\sqrt{n/2}]}.$$

Таким образом, величина $m=[\sqrt{n/2}]$ дает минимум f.

Упражнение 1.29. Используя критерий со второй производной, покажите, что $f(m, n)$, рассматриваемая как непрерывная и дифференцируемая функция переменной m, является выпуклой функцией по m. Исходя из этого, докажите, что минимум f, рассматриваемой как дискретная функция переменной m, достигается в одной из двух точек, соседних с точкой m, которая дает минимум $f(m, n)$.



Фиг. 1.13.

Критерий достаточности

На фиг. 1.13 изображена выпуклая функция $z=f(x, y)$, определенная в точках решетки плоскости. Точка $(\bar{x}, \bar{y})$ дает минимум соответствующей непрерывной функции. Для отыскания минимумов следует сравнить значения функции в четырех соседних точках. В пространстве E_n точку, в которой достигается минимум в непрерывном случае, окружают 2^n точек решетки, и все они должны быть проверены, если не использовать какие-то соображения, которые дают возможность опустить часть точек. Критерий, изложенный ниже, позволяет это сделать.

Ниже дается условие достаточности для подхода [4б], основанного на использовании множителей Лагранжа, в случае максимизации $f(x_1, \ldots, x_n)$ при ограничениях

$$g_i(x_1, \ldots, x_n) \leqslant 0, \quad i = 1, \ldots, m.$$

Оно не включает предположений о дифференцируемости и никаких условий на область определения D, которая может быть дискретной. Используя лагранжиан $F(x; \lambda) \equiv F(x_1, \ldots, x_n; \lambda_1, \ldots, \lambda_m)$, докажем следующую теорему:

Теорема 1.14 (Эверетт). *Пусть даны* $\bar{\lambda}_i \geqslant 0$, $i = 1, \ldots, m$, *и вектор* $\bar{x}$, *который максимизирует* $F(x; \lambda)$ *при всех* $x \in D$, тогда $\bar{x}$ максимизирует $f(x)$ среди всех $x \in D$, которые удовлетворяют условиям

$$g_i(x) \leqslant g_i(\bar{x}), \quad i = 1, \ldots, m.$$

Замечание. Эта теорема утверждает, что в точке $\bar{x}$ достигается условный максимум, если результирующее значение каждой функции $g_i(x)$, $i = 1, \ldots, m$, вычисленное в точке $\bar{x}$, представляет собой константу, которая ограничивает сверху соответствующую $g_i(x)$.

Доказательство. По предположению

$$F(\bar{x}; \bar{\lambda}) \geqslant F(x; \bar{\lambda}).$$

Таким образом,

$$f(\bar{x}) \geqslant f(x) + \sum_{i=1}^{m} \bar{\lambda}_i [g_i(\bar{x}) - g_i(x)]$$

для всех $x \in D$. Так как второй член в правой части неотрицателен, получаем $f(\bar{x}) \geqslant f(x)$, и доказательство закончено.

Эту теорему можно обобщить на случай, когда лагранжиан имеет вид $f(x) - G[g_1(x), \ldots, g_m(x); \lambda_1, \ldots, \lambda_m]$ и где из $g_i(x^1) \leqslant g_i(x^2)$, $i = 1, \ldots, m$, следует, что

$$G[g_1(x^1), \ldots, g_m(x^1); \lambda_1, \ldots, \lambda_m] \leqslant G[g_1(x^2), \ldots, g_m(x^2); \lambda_1, \ldots, \lambda_m]$$

для всех λ_i, $i = 1, \ldots, m$. Это соотношение показывает, что G должна быть монотонной функцией на направленном множестве ограничивающих векторов, частично упорядоченном путем включения множеств. Один вектор включает другой, если каждая его компонента превосходит соответствующую компоненту другого вектора.

Докажем теперь, что решение, которое дает значение лагранжиана $F(x; \lambda)$, близкое к максимуму, должно также давать значение $f(x)$, близкое к условному максимуму, при наличии ограничений $g_i(x) = g_i(\bar{x})$, $i = 1, \ldots, m$.

Теорема 1.15. *Если*

$$F(\bar{x}; \bar{\lambda}) \geqslant F(x; \bar{\lambda}) - \varepsilon \quad \textit{при малом } \varepsilon > 0,$$

то

$$f(\bar{x}) \geqslant f(x) - \varepsilon$$

для

$$g_i(x) = g_i(\bar{x}), \quad i = 1, \ldots, m.$$

Доказательство. Доказательство идентично предыдущему доказательству, но в нем используется ε.

В общем случае различные значения $\lambda = (\lambda_1, \ldots, \lambda_m)$ дают различные максимумы. Однако полезно исследовать влияние изменений величины λ или ее некоторых компонент на решение.

Теорема 1.16. *Если* $\bar{x}^1$ *и* $\bar{x}^2$ *— два решения, соответствующие* $\bar{\lambda}_1$ *и* $\bar{\lambda}_2$ *соответственно, и если* $g_i(\bar{x}^1) = g_i(\bar{x}^2)$, $i \neq k$, *и* $g_k(\bar{x}^1) > g_k(\bar{x}^2)$, *то компоненты* $\bar{\lambda}_k^1$ *и* $\bar{\lambda}_k^2$ *удовлетворяют соотношению*

$$\bar{\lambda}_k^2 \geqslant \frac{f(\bar{x}^1) - f(\bar{x}^2)}{g_k(\bar{x}^1) - g_k(\bar{x}^2)} \geqslant \bar{\lambda}_k^1.$$

Доказательство.

$$f(\bar{x}^1) \geqslant f(x) + \bar{\lambda}_k^1 [g_k(\bar{x}^1) - g_k(x)] + \sum_{i \neq k} \bar{\lambda}_k^1 [g_k(\bar{x}^1) - g_k(x)].$$

При подстановке $\bar{x}^2$ вместо x это соотношение по-прежнему имеет место; используя гипотезу при $i = k$, получаем

$$\frac{f(\bar{x}^1) - f(\bar{x}^2)}{g_k(\bar{x}^1) - g_k(\bar{x}^2)} \geqslant \bar{\lambda}_k^1.$$

Заменяя везде $\bar{x}^1$ на $\bar{x}^2$, получаем другую часть неравенства, что завершает доказательство.

Пример. Пусть дана система, состоящая из m подсистем, соединенных последовательно. Каждая i-я подсистема состоит из n_i элементов, соединенных параллельно (фиг. 1.14). Вероятность, что

Фиг. 1.14.

элемент i-й подсистемы функционирует, равна p_i, а стоимость его равна c_i. Система может функционировать только при условии, что в каждой подсистеме исправен хотя бы один элемент. Требуется

построить кривую, которая показывает связь между максимально достижимой надежностью системы и минимальной ее стоимостью.

Так как вероятность того, что элемент i-й подсистемы исправен, равна p_i, вероятность того, что он неисправен, равна $1 - p_i$; вероятность того, что все элементы i-й подсистемы неисправны, равна $(1 - p_i)^{n_i}$. Следовательно, $[1 - (1 - p_i)^{n_i}]$ — вероятность того, что хотя бы один элемент i-й подсистемы исправен. Так как подсистемы соединены последовательно, то, чтобы получить вероятность $P(n_1, \ldots, n_m)$ того, что система исправна, надо взять произведение указанных вероятностей; в результате получается

$$P(n_1, \ldots, n_m) = \prod_{i=1}^{m} [1 - (1 - p_i)^{n_i}].$$

Стоимость системы равна $C = \sum_{i=1}^{m} c_i n_i$. Можно решать или задачу максимизации P при заданном значении C или задачу минимизации C при заданном значении P.

Теперь P является вогнутой функцией n_i, $i = 1, \ldots, m$. Чтобы убедиться в этом, заметим, что функция $[1 - (1 - p_i)^{n_i}]$ монотонно не убывает по n_i. Следовательно, произведение P является монотонно неубывающей функцией $n = (n_1, \ldots, n_m)$. Далее если рассматривать P как непрерывную функцию n, то она будет вогнутой. Следовательно, можно сначала воспользоваться методами математического анализа для определения максимума функции P, рассматривая ее как непрерывную функцию своего аргумента, и затем искать ближайшую точку решетки (которая будет одной из 2^m возможных вершин куба, лежащих в точках решетки, которые окружают вычисленное значение P). В одной из этих точек достигается максимум. Однако 2^m может быть очень большим числом, и тогда желательно найти лучшую процедуру достижения максимума. Этого можно добиться при помощи соответствующего выбора параметра лагранжиана. Напомним, что если предполагается дифференцируемость, то необходимое условие максимизации функции одновременно является и необходимым условием максимизации ее логарифма. Поэтому возьмем $\log P$, прибавим к нему $\lambda \sum_{i=1}^{m} c_i n_i$, положим

$$\frac{\partial \log P}{\partial n_i} = 0$$

и разрешим это уравнение относительно n_i. Это дает

$$n_i = \frac{\log \{1/[1 - \log(1 - p_i)/\lambda c_i]\}}{\log(1 - p_i)}.$$

Зафиксировав значение λ, получим по этой формуле оценку для n_i, $i = 1, \ldots, m$, а затем возьмем в качестве пробного решения $[n_i]$ одно из двух ближайших к n_i целых чисел при каждом i.

Как более эффективное средство используем сформулированную выше теорему достаточности (теорема 1.14). Сначала приближенно выберем значение λ и получим решение $\bar{n}_i$, которое является максимумом для всех n_i, удовлетворяющих условию

$$\sum c_i n_i \leqslant \sum c_i \bar{n}_i = C.$$

Затем выберем другое значение λ и улучшим ограничивающее условие, при этом получим еще лучшее значение максимума. Продолжая этот процесс, найдем точки кривой надежность — стоимость, которая дает возможность принимать решения в зависимости от соотношения между стоимостью и надежностью. Выбирая λ достаточно малым, так что только одно значение n_i будет изменяться при переходе к следующему пробному значению, можно установить, на какую подсистему выгоднее всего расходовать средства. Заметим, что если каждый раз при выборе нового значения λ изменяются несколько величин n_i, то трудно решить, в какой подсистеме выгоднее всего увеличивать n_i с точки зрения достижения максимума. В задачах с большим числом переменных процедура с использованием пробных значений более экономична, чем вычисление значений P в каждой вершине куба.

Замечание. Что касается предыдущего примера, можно показать, что существует соотношение двойственности между максимизацией

$$\prod_{i=1}^{m} f_i(n_i) \text{ при ограничении } \sum_{i=1}^{m} C_i(n_i) = C$$

и минимизацией

$$\sum_{i=1}^{m} C_i(n_i) \text{ при ограничении } \prod_{i=1}^{m} f_i(n_i) = P,$$

причем последнее соотношение изучать легче, так как его целевая функция представляет собой сумму.

Упражнение 1.30. Рассмотрим задачу отыскания неотрицательных целых чисел x_i, $i = 1, \ldots, n$, которые максимизируют функцию

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} \log \frac{a_i x_i}{a_i + x_i}$$

при условии

$$\prod_{i=1}^{n} x_i = C.$$

Обладает ли функция f свойствами, которые позволяют выбирать целочисленные величины, ближайшие к решению, полученному при помощи дифференцирования? Можно ли использовать метод множителей Лагранжа? Объясните свой подход к решению этой задачи.

Что произошло бы, если бы производилась минимизация $\prod_{i=1}^{n} x_i$ при условии $f(x_1, \ldots, x_n) = C$, т. е. поменялись бы ролями f и ограничивающее уравнение?

1.6. Асимптотические результаты

Иногда необходимо и полезно получать решение задачи при предположении, что одна или несколько переменных могут принимать произвольно большие значения. Такой подход позволяет сравнивать предельные решения дискретной задачи и ее непрерывного расширения, если есть такое расширение. Использование асимптотических выражений проиллюстрируем задачей о почтальоне, который должен разносить почту по обеим сторонам единственной улицы поселка. Если домов немного и они расположены далеко друг от друга, он может ходить по улице вдоль и поперек, пытаясь минимизировать суммарное пройденное расстояние. Однако по мере возрастания числа домов по обеим сторонам улицы, очевидно, для почтальона выгоднее сначала обслужить все дома на одной стороне улицы, а затем перейти на другую сторону. Его стратегия в конце концов зависит только от расположения домов. Одним из интересных упражнений является выбор кратчайшего пути для нескольких типичных конфигураций. В этой задаче асимптотический случай дает хорошую основу для использования интуиции, которая подсказывает, что кратчайший путь равен сумме удвоенной длины улицы и ее ширины. Эта величина во всех других случаях является верхней границей для пути, который проходит почтальон. В данной книге будет дано несколько примеров, иллюстрирующих получение и использование асимптотических результатов.

Следующие обозначения иногда оказываются полезными в задачах, связанных с асимптотическими разложениями и другими предельными соотношениями. Выражение $f(x) = O(g(x))$ при $x \to x_0$ означает, что существует положительная величина A, такая, что $|f(x)| \leqslant Ag(x)$ при $x \to x_0$. Если ничего не говорится о соотношении $x \to x_0$, то запись O означает, что $|f(x)| \leqslant Ag(x)$ при всех x. Например, $f(x) = O(1)$ означает, что $f(x)$ — ограниченная функция. Имеют место соотношения

$$\sin x = O(|x|), \quad (x+1)^2 = O(1) \quad \text{при} \quad x \to 0,$$
$$\sin x = O(1), \quad (x+1)^2 = O(x^2) \quad \text{при} \quad x \to \infty.$$

Выражение $f(x) = o(g(x))$ означает, что $f(x)/g(x) \to 0$ при $x \to x_0$. Так, например,

$$\sin x = o(x^2), \quad x - 1 = o(x^2) \quad \text{при} \quad x \to \infty.$$

Выражение $o(1)$ означает, что функция стремится к нулю при $x \to x_0$. Выражение $f(x) \sim g(x)$ означает, что $f(x)/g(x) \to 1$ при $x \to x_0$.

Пример [15а]. Предположим, что дано положительное целое число n и требуется определить N — максимальное число непересекающихся пар положительных чисел с различными суммами, которые все меньше n. Например, при $n = 10$ получаем три пары (1, 8), (2, 6) и (3, 4). Ни один из членов каждой пары не встречается ни в какой другой паре, и, следовательно, они не пересекаются.

Решение. Поскольку имеется N пар, получаем $2N$ чисел, сумма которых не должна превосходить $1 + 2 + \ldots + 2N = N(2N+1)$. Из постановки задачи следует, что эта сумма не превышает

$$(n-1) + (n-2) + \ldots + (n-N) = \frac{N}{2}(2n - N - 1).$$

Эти два соотношения дают $N(2N+1) \leqslant (N/2)(2n - N - 1)$, или $N \leqslant (2n-3)/5$. Теперь если k — наибольшая целая часть $n/5$, то пары $(1, 4k)$, $(3, 4k-1)$, $(5, 4k-2)$, $\ldots$, $(2k-1, 3k+1)$ и $(2, 3k)$, $(4, 3k-1)$, $(6, 3k-2)$, $\ldots$, $(2k-2, 2k+2)$ не пересекаются и имеют различные суммы. Сумма любой пары не превосходит

$$2k - 1 + 3k + 1 = 5k < n.$$

Так как имеется $(2k-1)$ пар, удовлетворяющих условиям задачи, приходим к выводу, что

$$N \geqslant 2k - 1 \geqslant 2\left(\frac{n}{5} - 1\right) - 1 = \frac{2n}{5} - 3.$$

Это дает

$$\frac{2n}{5} - 3 \leqslant N \leqslant \frac{2n-3}{5}.$$

Теперь при n, стремящемся к бесконечности, N ведет себя как $2n/5$. Таким образом, можно записать

$$N = \frac{2n}{5} + O(1),$$

где $O(1)$ означает ограниченную функцию, которая в данном случае является константой. Можно записать более сильное соотношение $N \sim 2n/5$ при $n \to \infty$, так как отношение этих двух величин стремится к 1 при $n \to \infty$.

1.7. Примеры задач

Существуют два общих подхода к постановке задач оптимизации, особенно в целых числах. При первом из них задача формулируется при помощи общих качественных выражений, как в случае физических или геометрических задач. Например, требуется найти максимальное число пластинок домино, каждая из которых покрывает два квадрата, необходимое для покрытия шахматной доски, из которой удалены квадраты в левом нижнем и правом верхнем углах. Для краткости будем классифицировать задачи, которые включают пространственные соотношения между объектами, как

геометрические. Иногда такую задачу можно сформулировать алгебраически, иногда нет. Второй подход к постановке задач оптимизации — алгебраический. Здесь соотношения выражаются через алгебраические переменные. Между этими двумя подходами трудно провести четкую грань. Вообще говоря, алгебраическая формулировка является более предпочтительной, но часто геометрическая формулировка обеспечивает более глубокое понимание задачи.

Задачи в геометрической постановке

Задача о делении линии. Требуется разделить линию длиной N единиц на n отрезков, где $n < N$, так, чтобы длина каждого отрезка равнялась целому числу единиц, а произведение длин отрезков было максимальным. В алгебраической формулировке, если $x_1, \ldots, x_n$ — длины n отрезков, то в задаче требуется выбрать положительные целые числа x_i, $i = 1, \ldots, n$, которые максимизируют $\prod_{i=1}^{n} x_i$ при ограничении $\sum_{i=1}^{n} x_i = N$. (В гл. 4 дается решение этой задачи.)

Задача о расположении монет. Какое максимальное число идентичных монет диаметром 1 см можно уложить, покрывая поверхность квадратного стола со стороной 1 м, чтобы никакие две монеты не перекрывались?

Задача о кокосовых орехах. Однажды n матросов набрали мешок орехов и решили поделить их утром. Ночью один из матросов проснулся и решил забрать свою долю. Он отложил d_1 орехов, так чтобы остаток делился на n, и забрал свою долю. Потом проснулся другой матрос, отложил d_2 орехов из оставшейся кучи, поделил остаток на n частей и забрал свою долю и т. д. В конце концов оказалось, что $(m - 1)$ матросов забрали ночью свою долю, причем матрос i, $i = 1, \ldots, m - 1$, откладывал d_i орехов, чтобы остаток делился на n. Утром все n матросов отложили d_m орехов из того, что осталось, и разделили остаток между собой поровну. Какое минимальное число орехов первоначально должно быть в мешке, чтобы в результате каждый матрос получил хотя бы по одному ореху [14]? (См. решение этой задачи в гл. 3.)

Задача о бомбардировщике. Пусть имеется n целей, разбросанных на большой территории. Каждой цели приписывается одно из трех возможных значений ценности 1, 2, 3; большее число соответствует большей ценности. Самолет с фиксированным запасом горючего и максимальной бомбовой нагрузкой 3200 кг должен разбомбить ряд целей за один вылет и вернуться на базу. Имеются три типа бомб: по 200 кг, которые могут разрушать цели с ценностью 1; по 400 кг, которые могут разрушать цели с ценностями 1 и 2, и по 800 кг, которые могут разрушать цели с ценностями 1, 2 или 3. Самолет

может нести не более 12 бомб. Расход горючего пропорционален весу самолета, который зависит от веса неистраченного бомбового груза и количества имеющегося горючего. Считается, что атакованная цель полностью разрушается. Оптимальный выбор целей определяет порядок, в котором загружаются, а потом сбрасываются бомбы. Бомбы какого типа и какой порядок загрузки обеспечивают разрушение целей с наибольшей суммарной ценностью?

Задача о фальшивой монете. Пусть даны рычажные весы и множество монет, среди которых имеется одна фальшивая монета. В поисках фальшивой монеты на обе чашки весов кладут одинаковое количество монет. Требуется также определить, легче или тяжелее фальшивая монета, чем настоящая. При помощи фиксированного числа взвешиваний n нужно найти максимальное число монет, среди которых еще можно выделить фальшивую, и определить, легче или тяжелее она, чем настоящая [5]. (Обсуждение этой задачи см. в гл. 3.)

Задача о четырех цветах. Чему равно наименьшее число цветов, с помощью которых можно раскрасить области произвольной плоской карты так, чтобы никакие две области, имеющие общую границу (не просто точку), не были раскрашены в один цвет? Чему равно наибольшее число областей, которые можно раскрасить четырьмя цветами? (Алгебраическую формулировку задачи см. в гл. 5, а более подробное обсуждение — в гл. 2.)

Задача о наименьшем числе пересечений. Полный граф с n вершинами, начерченный на плоскости, представляет собой граф, полученный путем соединения простой кривой каждой пары из n вершин. Среди все возможных реализаций графа найдите ту, которая имеет наименьшее число пересечений простых кривых в точках, отличных от n данных вершин. Две простые кривые могут пересекаться не более чем в одной точке, которая не является вершиной. (Обсуждение этой задачи см. в гл. 2.)

Задачи в алгебраической постановке

Задача о распределении рабочей силы [26]. Пусть имеется m рабочих специальностей, n видов работ и заданы средние значения c_{ij} производительности рабочего i-й специальности при выполнении j-й работы. Требуется задать назначение x_{ij} i-го рабочего на j-ю работу (величина x_{ij} равна нулю или единице) для всех i и j так, чтобы суммарная производительность $\sum_{j=1}^{n}\sum_{i=1}^{m} c_{ij}x_{ij}$ всех рабочих была бы максимальна при ограничениях

$$\sum_{j=1}^{n} x_{ij} \leqslant a_i, \qquad i = 1, \ldots, m,$$

$$\sum_{i=1}^{m} x_{ij} \leqslant b_j, \qquad j = 1, \ldots, n,$$

где a_i означает число рабочих i-й специальности, а b_j — число работ типа j.

Транспортная задача. Предположим, что однородный продукт отправляют из m пунктов в n пунктов назначения, причем из каждого пункта отправления вывозится определенное количество продукта, а в каждый пункт назначения ввозится тоже определенное количество продукта, так что суммарные количества ввозимого и вывозимого продуктов равны. Пусть a_i — количество продукта, вывозимого из пункта отправления i ($i = 1, \ldots, m$), а b_j — количество продукта, ввозимого в пункт назначения j ($j = 1, \ldots, n$); это известные величины. Тогда

$$\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j,$$

где $a_i,\ b_j \geqslant 0$ при всех $i,\ j$.

Пусть x_{ij} — неизвестное количество продукта, который надо перевозить из пункта i в пункт назначения j. Тогда

$$\sum_{j=1}^{n} x_{ij} = a_i, \qquad i = 1, \ldots, m,$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \qquad j = 1, \ldots, n,$$

$$x_{ij} \geqslant 0 \quad \text{при всех} \quad i \text{ и } j.$$

Пусть c_{ij} — стоимость перевозки единицы продукта из пункта i в пункт j. Эти величины также заданы. Задача состоит в том, чтобы найти x_{ij}, которые удовлетворяют указанным выше ограничениям и минимизируют функцию

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

(Решение примера см. в гл. 5.)

Задача о поставщике. Поставщику известно, что столовой, которую он договорился обслуживать в течение n дней, требуется r_j ($\geqslant 0$) свежих салфеток в течение j-го дня, где $j = 1, \ldots, n$. Стирка обычно занимает p дней, т. е. грязная салфетка, посланная в прачечную сразу же после того, как ее использовали на j-й день, снова будет пригодна к использованию на $(j + p)$-й день. Однако прачечная за более высокую цену может возвращать салфетки через $q < p$ дней (p и q — целые числа). Если у поставщика нет на руках или в прачечной пригодных к использованию салфеток, то, чтобы удовлетворить потребности столовой, он может купить их по a центов за штуку. Стирка одной салфетки стоит b центов, а срочная стирка — c центов. Как должен действовать поставщик, чтобы удовлетворить потребности столовой и минимизировать свои расходы за n дней [9]?

Задача о закупках [23]. Каждый корабль предназначен для выполнения определенных функций с использованием электронной аппаратуры. Для выполнения кораблем функции j разрешается получить не более b_j единиц аппаратуры, и, хотя каждая единица аппаратуры в принципе может выполнять несколько функций, предполагается, что, будучи установленной, она выполняет только одну из них. Имеется m типов аппаратуры, и аппаратуру i-го типа надо распределять, учитывая, что в распоряжении имеется a_i единиц аппаратуры. Задача состоит в том, чтобы распределить по кораблям имеющуюся в распоряжении аппаратуру всех типов наилучшим образом.

Предположим, что мы ввели множество величин c_{ij}, представляющих собой ценность (измеренную в приемлемых единицах) от использования аппаратуры типа i для выполнения функции j. Эти величины определяются таким образом, что $c_{ij} > c_{hh}$ означает, что использование одной единицы аппаратуры i-го типа для выполнения функции j (назовем это распределением ij) предпочтительнее, чем использование аппаратуры k-го типа для выполнения функции h (распределение kh). Кроме того, $c_{ij} = wc_{hh}$ означает, что распределение i, j в w раз выгоднее, чем распределение k, h.

При условии, что заданы суммарные затраты M и стоимости единицы аппаратуры каждого типа p_i, требуется определить, как надо распределить деньги для закупки аппаратуры, чтобы эффект от использования закупленного оборудования был оптимальным. В хорошем плане закупок, конечно, должна быть учтена аппаратура, которая хранится на складе или заказана, и принято во внимание то, как предполагается использовать закупленную аппаратуру на кораблях. Поэтому закупки и распределение аппаратуры следует рассматривать совместно. Задача формулируется как задача линейного программирования следующим образом: требуется найти

$$\max \sum_i \sum_j c_{ij} x_{ij}$$

при ограничениях

$$\sum_j x_{ij} \leqslant a_i + y_i,$$

$$\sum_i x_{ij} \leqslant b_j,$$

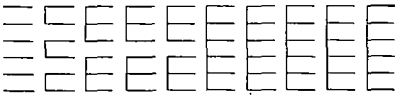
$$\sum_i p_i y_i \leqslant M,$$

где $x_{ij}, y_i \geqslant 0$ — неотрицательные целые числа, $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$. Величины y_i означают количество аппаратуры каждого типа, которое должно быть закуплено.

Построение наилучших схем ступенчатого включения в телефонии [24, 25]. Правило построения наилучших схем ступенчатого включения при формировании телефонных пучков было предложено

Дж. Ф. О'Деллом в 1927 г., и оно принято в телефонных службах во всем мире.

Схема подключения состоит из n групп, каждая из которых имеет по k контактов, а всего образовано R пучков. Контакты в каждой группе пронумерованы 1, 2, . . ., k. Контакты с одинаковыми порядковыми номерами, но принадлежащие разным группам, могут образовывать индивидуальные линии, пучки, доступные только для части нагрузки, и пучки, доступные всем вызовам (фиг. 1.15).



Фиг. 1.15. Диаграмма при $n = 6$, $k = 10$, $R = 20$.

В этой структуре имеется в виду, что $k \leqslant R \leqslant kn$, где предельные случаи $R = k$ и $R = kn$ соответствуют пучкам, доступным всем вызовам, и наличию только индивидуальных линий.

Чтобы практически построить наилучшую схему подключения при заданных параметрах n, k и R, число групп n разлагается на множители, так что получается q сомножителей f_i ($i = 1, \ldots, q$), расположенных в порядке возрастания:

$$1 = f_1 < f_2 < \ldots < f_{q-1} < f_q = n.$$

Все номера контактов разбиваются на q классов в соответствии со значениями f_i; все контакты с одинаковым номером, отнесенные к i-му классу, объединяются в пучки по f_i контактов ($i = 1, \ldots, q$). Следовательно, все контакты с одинаковым номером, отнесенные к i-му классу, разбиваются на число пучков (в каждом по f_i контактов), равное n/f_i.

Пусть x_i — число различных номеров контактов, из которых образуются f_i-кратные пучки. Следовательно, полное число пучков равно

$$\sum_{i=1}^{q} x_i \frac{n}{f_i} = R, \qquad \sum_{i=1}^{q} x_i = k,$$

где $0 \leqslant x_i \leqslant k$, $k = i, \ldots, q$. Согласно правилу наилучшего построения схемы подключения, оптимальное разбиение будет при минимальной сумме модулей последовательных разностей

$$D = |x_1 - x_2| + |x_2 - x_3| + \ldots + |x_{q-1} - x_q|.$$

О'Делл опубликовал таблицы оптимальных решений (при данных k, n и R), но в них ничего не было сказано о возможности неединственных решений. Существование и единственность оптимальных решений рассматривал Сиски (см. гл. 4).

Задача распределения летных экипажей. Авиакомпания хочет минимизировать расходы, связанные с назначением экипажей на беспосадочные полеты между двумя городами. Задача состоит в том, чтобы найти оптимальное расписание назначений. Матрицу платежей можно представить в виде

	Экипаж 1	Экипаж 2	Экипаж m
Рейс 1	1	0	. . .
Рейс 2	0	1	. . .
Рейс 3	1	0	. . .
. . . . .			
Рейс n	1	0	. . .

Здесь 1 и 0 относятся к случаям, когда команда выполняет данный рейс и не выполняет. Расходы при назначении экипажа i на n рейсов равны c_j, $j = 1, \ldots, m$. Поскольку каждый рейс выполняет одна команда, сумма элементов в каждой строке должна быть равна единице. Задачи такого типа решались методом секущих плоскостей. Согласно данным Шиппера, время типичного решения с матрицей, состоящей из 350 строк и 3000 столбцов, на IBM 7094 составляет ~ 45 мин.

Задача о графике движения танкеров [4]. Задачу определения минимального числа танкеров, необходимых для того, чтобы выдержать определенный график, можно рассматривать как задачу линейного программирования типа транспортной задачи.

Задачу составления графика можно сформулировать следующим образом:

		Пункты разгрузки 1	2	...	j
Пункты заполнения	1	$t_{11}^1 t_{11}^2 \ldots t_{11}^{k_{11}}$	$t_{12}^1 t_{12}^2 \ldots t_{12}^{k_{12}}$	...	$t_{1j}^1 t_{1j}^2 \ldots t_{1j}^{k_{1j}}$
	2				
	.	. . . . .	. . . . .	...	. . . . .
	p	$t_{p1}^1 t_{p1}^2 \ldots t_{p1}^{k_{p1}}$	$t_{p2}^1 t_{p2}^2 \ldots t_{p2}^{k_{p2}}$	...	$t_{pj}^1 t_{pj}^2 \ldots t_{pj}^{k_{pj}}$

Здесь p — пункт заполнения, j — пункт разгрузки, k_{pj} — последний элемент в клетке pj, t_{pj} — момент времени, в который танкер должен быть полностью заполнен в пункте p и отправлен в пункт j. Количество величин t_{pj} конечно. Задаются два массива положительных чисел a_{pj} и b_{pj}, где a_{pj} — время прохождения загруженного танкера от p до j, а b_{pj} — время прохождения незагруженного танкера от j к p.

Задача состоит в том, чтобы расположить числа t_{pj} в S последовательностей, где каждая последовательность представляет собой

график для одного танкера так, чтобы выполнялись следующие условия:

1. Каждая последовательность монотонно возрастает.

2. Если $t^{k_1}_{p_1 j_1} < t^{k_2}_{p_2 j_2}$ — последовательные числа в одной из S последовательностей, то $t^{k_2}_{p_2 j_2} - t^{k_1}_{p_1 j_1} \geqslant a_{p_1 j_1} + b_{p_2 j_2}$, что означает, что сумма времени, требуемого для заполнения танкера в p_1 и перехода к j_1, и времени, требуемого для разгрузки в j_1 и перехода к пункту p_2, не может быть больше, чем разность двух моментов отправления танкера из пунктов заполнения.

3. Число последовательностей S должно быть минимальным.

Задачу о танкерах можно переформулировать как задачу линейного программирования следующим образом: для удобства примем, что t^k_{pj}, a_{pj} и b_{pj} — положительные числа. Определим

$$T^k_{pj} = (t^k_{pj} + a_{pj})$$

как время, когда танкер, заполненный в p, прибывает в j. Пусть $n_{\alpha p}$ — число танкеров, загружающихся в p в момент α, $N_{\beta j}$ — число танкеров, находящихся в пункте j в момент β, и $X_{\alpha p \beta j}$ — число танкеров, направляющихся в момент β из пункта j в пункт заполнения p, α — время прибытия. Тогда при любом графике будут выполняться следующие неравенства:

$$\begin{aligned} \sum_{\alpha,\, p} X_{\alpha p \beta j} &\leqslant N_{\beta j}, \\ \sum_{\beta,\, j} X_{\alpha p \beta j} &\leqslant n_{\alpha p}, \end{aligned} \tag{1.4}$$

$$X_{\alpha p \beta j} \geqslant 0, \tag{1.5}$$

где $b_{pj} > \alpha - \beta$ означает, что $X_{\alpha p \beta j} = 0$. Неравенства (1.4) можно превратить в систему равенств, которые соответствуют задаче типа транспортной, путем введения неотрицательных вспомогательных переменных $X_{\alpha p}$, $Y_{\beta j}$ и $Z = \sum\limits_{\alpha,p} \sum\limits_{\beta,j} X_{\alpha p \beta j}$. Неравенства (1.4) теперь можно переписать в виде

$$\begin{aligned} \sum_{\alpha,\, p} X_{\alpha p \beta j} + Y_{\beta j} &= N_{\beta j}, \quad Y_{\beta j} \geqslant 0, \\ \sum_{\beta,\, j} X_{\alpha p \beta j} + X_{\alpha p} &= n_{\alpha p}, \quad X_{\alpha p} \geqslant 0, \\ \sum_{\alpha,\, p} X_{\alpha p} + Z &= \sum_{\alpha,\, p} n_{\alpha p}, \quad Z \geqslant 0, \\ \sum_{\beta,\, j} Y_{\beta j} + Z &= \sum_{\beta,\, j} N_{\beta j}. \end{aligned} \tag{1.6}$$

Таким образом, любой график приводит к целочисленному решению неравенств (1.5) и уравнений (1.6). График можно также составить, отправляясь от целочисленного решения (1.5) или (1.6). Задачу

составления графика движения танкеров можно свести к задаче минимизации $\sum_{\alpha p} X_{\alpha p}$ на множестве целочисленных решений (1.5) и (1.6).

График можно составить, исходя из целочисленного решения, следующим образом: $X_{\alpha p}$ — число танкеров, начинающих свой индивидуальный график в момент α в пункте p; всего будет перегруппировано $X_{\alpha p}$ последовательностей, первым членом которых будет $t^k_{pj} = \alpha$. Исключим один такой член $t^{k_0}_{p_0 j_0} = \alpha_0$ из t; пусть $\beta_0 = T^{k_0}_{p_0 j_0} = \alpha_0 + \alpha_{p_0 j_0}$. Поскольку $N_{\beta_0 j_0} > 0$, хотя бы одна из переменных $X_{\alpha p \beta_0 j_0}$, $Y_{\beta_0 j_0}$ имеет положительное значение. Выберем одно из них.

Случай 1. Если выбрано значение $X_{\alpha_1 p_1 \beta_0 j_0} > 0$, то некоторое $t^{k_1}_{p_1 j_1} = n_{\alpha_1 p_1} > 0$. Сделаем α_1 вторым членом последовательности. Вычтем α_1 из t и уменьшим $X_{\alpha_1 p_1 \beta_0 j_0}$, $N_{\beta_0 j_0}$, $n_{\alpha_1 p_1}$ на единицу.

Случай 2. Если выбрано значение $Y_{\beta_0 j_0} > 0$, то последовательность оканчивается на α_0. Уменьшим $Y_{\beta_0 j_0}$, $N_{\beta_0 j_0}$ на единицу.

В случае 1 пусть $\beta_1 = T^{k_1}_{p_1 j_1} = \alpha_1 + \alpha_{p_1 j_1}$ и рассмотрим значения переменных $X_{\alpha p \beta_1 j_1}$, $Y_{\beta_1 j_1}$; одна из них должна быть положительной. Перейдем к случаю 1 или 2, но теперь α_1 будет играть роль α_0. Повторение этой процедуры в конце концов должно закончиться выбором некоторого $Y_{\beta_h j_h} > 0$, и, таким образом, одна из последовательностей будет закончена. Другие последовательности можно получить таким же способом.

Задача о связи через спутники. Рассмотрим спутник связи типа Телестар, для которого известны отрезки времени, когда он доступен для комплексов средств связи. (Задача о стационарном спутнике типа Синком является более простым частным случаем данной модели [21].)

Задача состоит в том, чтобы составить расписание сеансов связи между парами станций через спутник для обмена информацией. Каждая такая пара считается направлением связи. Время, в течение которого спутник на каждой из своих орбит может поддерживать связь, задано. На спутниках установлено заданное число дуплексных (двусторонних) передатчиков, причем каждый передатчик имеет заданную емкость канала. Требования, предъявляемые каким-то направлением связи, задаются в виде числа передатчиков и количества времени. Станция может использовать спутник для связи с другой станцией с помощью любого числа спутниковых передатчиков. Однако в то же время остальные передатчики могут быть использованы для нужд других направлений связи. За спутником следит радиолокационная антенна, и причем одна антенна не может следить за двумя спутниками одновременно. Если на некотором направлении связи может быть использована дополнительная аппаратура для слежения, то связь по этому направлению можно поддерживать более чем через один спутник. При необходимости поддерживать

связь в дальнейшем дополнительная антенна готовится к слежению за следующим по расписанию спутником, который попадет в зону видимости радиолокатора. Мы считаем, что каждое направление связи располагает достаточным числом антенн и каждый спутник имеет достаточное число приемно-передающих устройств для того, чтобы выполнялись требования программы сеансов связи. (Эта задача будет сформулирована в гл. 5 как задача целочисленной оптимизации.)

ЛИТЕРАТУРА [1])

1. Box G. E. P., Wilson K. B., On the Experimental Attainment of Optimum Conditions, *J. Roy. Stat. Soc.*, ser B., 13 (1951).
2. Box G. E. P., The Exploration and Exploitation of Response Surfaces, *Biometrics*. 10 (1954).
3. Cloud J. D., Jackson W. D., Number of Fibonacci Numbers Not Exceeding N., *Am. Math. Monthly*, 798 (Sept. 1964).
4. Dantzig G. B., Fulkerson D. R., Minimizing the Number of Tankers to Meet a Fixed Schedule, *Naval Res. Log. Quart.*, 1, № 3 (Sept. 1954).

4а. Dorn W. S., Lagrange Multipliers and Inequalities, *Operations Res.*, 9, № 1, 95 (1961).

4б. Everett, III, Hugh, Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources, *Operations Res.*, 11, № 3, 399 (1963); русский перевод: сб. «Оптимальные задачи надежности», под ред. Ушакова И. А., изд-во «Сов. радио», 1967.

5. Fine N. J., The Generalized Coin Problem, *Am. Math. Monthly*, 489 (Oct. 1947).
6. Гельфанд И. М. и др., Метод координат, изд-во «Наука», 1964.
7. Golomb S. W., Klamkin M. S., N Objects in B Boxes, *Am. Math. Monthly*, 60, 552 (Oct. 1953).
8. Hadwiger H., Debrunner H., Combinatorial Geometry in the Plane, Holt, Rinehart and Winston, Inc., N. Y., 1964; русский перевод: Хадвигер Г. и Дебруннер Г., Комбинаторная геометрия плоскости, изд-во «Наука», 1965.
9. Jacobs W. W., The Caterer Problem, *Naval Res. Log. Quart.*, 1, 154 (1954).
10. Johnson R. A., Relating to a Problem in Minima, *Am. Math. Monthly*, 24, 243 (1917).
11. Johnson S. M., Optimal Search for a Maximum is Fibonaccian, Rand Corp. Rept. P-856. 1956.
12. Kaplan S., Solution of the Lorie-Savage and Similar Integer Programming Problems by the Generalized Lagrange Multiplier Method, *Operations Res.*, 14, № 6, 1130 (1966).
13. Kiefer J., Sequential Minimax Search for a Maximum, *Proc. Am. Math. Soc.*, 4, 502 (1953).
14. Kirchner R. B., The Generalized Coconut Problem, *Am. Math. Monthly*, 516 (June—July 1960).
15. Klamkin M. S., Quickie Number 264, *Math. Mag.*, 175 (1960).

15а. Klamkin M. S., Newman D. J., Some Combinatorial Problems of Arithmetic, *Math. Mag.*, 42, 53 (March 1969).

16. Konhauser J. D. E., Brown J. L., Chale D., Greatest of Three, *Math. Mag.*, 187 (May—June 1962).
17. Kuhn H. W., Tucker A. W., Non-linear Programming, in: Neyman J. (ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Univ. Calif. Press, Berkeley, Calif., 1951.

---

[1]) Звездочкой отмечены работы, добавленные при переводе.

17а. Krolak D. P., Further Extensions of Fibonaccian Search to Nonlinear Programming Problems, *SIAM J. Control*, 6, № 2 (1968).
18. Lennes N. J., Note on Maxima and Minima by Algebraic Methods, *Am. Math. Monthly*, 17, 9 (1910).
19. MacDonald J. E., Jr., Cohen J., An Inequality, *Am. Math. Monthly*, 914 (Oct. 1964).
20. Moser L., Toskey B. R., Minimum Number of Distinct Values Assumed by a Sum, *Am. Math. Monthly*, 670 (June—July 1963).
20а. Newman D. J., Silverman D. L., Rectangular Rugs in a Square Room, *Am. Math. Monthly*, 209 (Feb. 1964).
20б. Ponstein J., Seven Types of Convexity, *SIAM Rev.*, 9, № 1 (1967).
21. Saaty T. L., Suzuki G., A Nonlinear Programming Model in Optimum Communication Satellite Use, *SIAM Rev.*, 7, 403 (July 1965).
22. Saaty T. L., Bram J., Nonlinear Mathematics, McGraw-Hill, N. Y., 1964.
23. Suzuki G., Procurement and Allocation of Naval Electornic Equipments, *Naval Res. Log. Quart.*, 4, № 1, 1 (March 1957).
24. Syski R., Algebraic Properties of Optimum Gradings, 3d Intern. Teletraffic Cong., Paris, 1961.
25. Syski R., Introduction to Congestion Theory, Oliver and Boyd Ltd., London, 1960.
26. Votaw D. F., Methods of Solving Some Personnel-Classification Problems, *Psychometrika*, 17, № 3, 255 (1952).
27. Wagstaff R., The Maximum of $\sum_{i,j=1}^{n} a_i b_j$, *Am. Math. Monthly*, 46 (Jan.—Feb. 1964).
28. Wilde D. J., Optimum Seeking Methods, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1964; русский перевод: Уайлд Д. Дж., Методы поиска экстремума, изд-во «Наука», 1967.
29*. Виленкин Н. Я., Комбинаторика, изд-во «Наука», 1969.
30*. Воробьев Н. Н., Числа Фибоначчи, изд-во «Наука», 1969.
31*. Растригин Л. А., Статистические методы поиска, изд-во «Наука», 1968.
32*. Риордан Дж., Введение в комбинаторный анализ, ИЛ, 1963.

Глава 2

Методы геометрической оптимизации

2.1. Введение

Экстремальные задачи, или задачи оптимизации, можно рассматривать абстрактно в терминах множеств и преобразований на множествах. Обычно задача состоит в том, чтобы найти для некоторой области, которая подвергается преобразованию, максимальный элемент в множестве принимаемых значений. Последнее часто представляет собой множество действительных чисел. Иногда ставится обратная задача: при условии, что заданы некоторые ограничения на множество принимаемых значений, требуется найти в некотором смысле максимальное или минимальное свойство области определения функции, благодаря которому ее можно отобразить во множество принимаемых значений. Естественно, все преобразования можно использовать для отображения множества в себя или в свои подмножества.

В табл. 2.1 указаны некоторые общие темы, где встречаются геометрические экстремальные задачи. На протяжении всей главы в начале каждого раздела будет даваться краткое обсуждение таких задач.

Наша цель здесь состоит в том, чтобы изучить задачи, в которых требуется дискретная оптимизация, т. е. вопросы типа: существует не меньше и не больше (минимум или максимум) объектов такого-то вида, и указать методы определения этих значений (точные, приближенные или асимптотические). Таким образом, в общем случае будем искать верхние и нижние границы, а также минимальные и максимальные значения функции, заданной на соответствующим образом определенном множестве и принимающей действительные значения. Нет необходимости говорить, что такая функция редко задается в явном виде. При помощи соответствующей процедуры оценивания можно добиться точного ответа. (См., например, задачу об упаковке кругов на плоскости, разд. 2.8.)

В разд. 2.2 будет приведено несколько примеров, в которых для получения оптимума используются соображения симметрии. Кроме специального использования симметрии или представления задачи (если возможно) в виде стандартной алгебраической задачи, нет ни одного прямого метода формулировки и решения геометрических задач, в которых требуется проведение оптимизации. Анализ разнообразных примеров является сам по себе мощным средством решения вновь возникающих задач.

Таблица 2.1

Оптимизация в дискретной геометрии

Не все обширные темы, приведенные в табл. 2.1, могут быть достаточно полно рассмотрены в одной главе. Поскольку наша цель состоит в многостороннем изложении предмета, будем использовать интересные примеры. Для тех, кто хочет глубже изучить этот предмет, приводится большой список литературы.

2.2. Симметрия и оптимизация

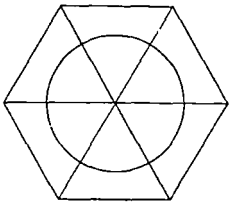
Существование симметрии — это основное условие для решения многих задач, которые иначе было бы трудно решить. Иногда несимметричную задачу можно представить как часть более широкой задачи, обладающей свойством симметрии, что приводит к решению несимметричной задачи. Решение исходной задачи можно получить из решения более широкой задачи, которую можно рассматривать как ее обобщение. Вероятно, полезно помнить девиз: *чтобы максимизировать или минимизировать, надо симметрировать.*

Пример. Рассмотрим игру на прямоугольной доске. Два игрока по очереди укладывают на доску монеты так, чтобы они не перекрывались. Тот, кто последним положит монету на еще свободное место, побеждает. Как надо играть?

Решение. Очевидно, что тот, кто делает первый ход и помещает свою монету в центр, выиграет. Он добьется этого, помещая каждую свою монету симметрично соответствующим монетам соперника, т. е. на линии, проходящей через центр доски, и последнюю монету противника на том же расстоянии с другой стороны от центра доски. Конечно, использование других видов симметрии тоже приводит его к победе.

Упражнение 2.1. Назовите еще две стратегии, использующие симметрию, которые приводят к победе в этой игре.

Пример. При условии, что задан равносторонний треугольник рассмотрим кривую, которая начинается в некоторой точке на одной стороне треугольника и оканчивается в некоторой точке на другой стороне. Требуется определить форму кривой наименьшей длины, которая делит площадь треугольника на две равные части [85].



Фиг. 2.1.

Решение. Предположим, что мы провели кривую, которая делит треугольник на две равные части. Отобразим треугольник симметрично относительно одной из двух сторон, содержащей конечную точку этой кривой. Отобразим образ относительно стороны, содержащей другую конечную точку, затем отобразим этот образ и т. д., пока не получим шестиугольник. Кривая превратится в замкнутую кривую внутри шестиугольника (фиг. 2.1). Эта кривая делит площадь

шестиугольника на две равные части. Кривая наименьшей длины охватывающая половину площади шестиугольника, представляет собой окружность с тем же центром, что и у шестиугольника. Таким образом, искомая кривая является дугой окружности (см. теорему 2.6 об этом свойстве окружности.)

Упражнение 2.2. Определите форму простой замкнутой кривой минимальной длины, заключенной внутри равностороннего треугольника, которая охватывает половину площади треугольника. Воспользуйтесь соображениями непрерывности относительно площади, которую ограничивает кривая.

Другим примером использования симметрии при оптимизации является задача справедливого дележа (см. гл. 3), в которой требуется разделить пирог на n частей так, чтобы каждый участник получил, по его мнению, справедливую долю. В действительности это тоже пример оптимизации, хотя на первый взгляд здесь нет ни максимизации, ни минимизации.

Пример. (Не слишком много и не слишком мало.) Пусть имеются чашка с кофе и чашка с молоком, в каждую из которых налито одинаковое количество жидкости. Столовую ложку молока переливают в чашку с кофе и после тщательного перемешивания, чтобы сделать смесь однородной, столовая ложка этой смеси переливается обратно в чашку с молоком. Покажите, что в чашке с кофе содержится столько же молока, сколько и кофе в чашке с молоком.

Решение. Можно дать аналитическое доказательство, в котором будет фигурировать доля молока в чашке с кофе (та же доля будет и в столовой ложке) и доля кофе в чашке с молоком.

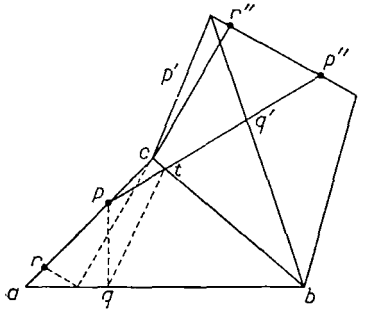
Можно дать следующее простое доказательство, не зависящее от однородности перемешивания, в котором используются соображения симметрии. Поскольку в обеих чашках имеется одинаковое количество жидкости, то количество молока, добавленного в кофе, должно быть равно количеству кофе, добавленного в чашку с молоком.

Пример. (Применение принципа отражения [43, 44].) Предположим, что дан многоугольник и некоторая точка на одной из его сторон. Требуется найти кратчайший путь, который начинается в данной точке, касается некоторых или всех других сторон и затем возвращается в исходную точку. Как найти этот путь?

Решение. Ответ можно получить не путем последовательного опускания перпендикуляров, а путем зеркального отражения многоугольника относительно первой из сторон, которой должна коснуться траектория после исходной точки, затем отражения образа относительно второй из сторон, которой должна коснуться траектория, и т. д. вплоть до последней из сторон, которой должна коснуться траектория.

Если рассмотреть фигуру, полученную из исходного многоугольника и всех его отражений, и пометить образ исходной точки в послед-

нем образе многоугольника, то можно получить ответ. Проведем прямую линию, соединяющую исходную точку с ее последним образом. Если линия пересекает каждое отражение многоугольника, можно взять отражение этой линии в исходном многоугольнике и получить тем самым искомый ответ. Если оказывается, что линия не пересекает некоторый образ многоугольника или часть линии проходит вне конфигурации, сегменты кратчайшей линии проводятся соответствующим образом от исходной точки к ее конечному образу.



Ф и г. 2.2.

Такую же процедуру можно использовать для проведения кратчайшего пути, касающегося различных граней многогранника, путем соответствующих отражений многогранника относительно граней в требуемой последовательности. На фиг. 2.2 треугольник abc отображен сначала относительно bc, затем относительно образа ab. Сегмент pp'' целиком лежит внутри фигуры, и его отражение в abc дает искомый кратчайший путь, который определяет точку t на стороне исходного треугольника. Если в качестве исходной точки взять r, то сегмент rr'' будет лежать вне фигуры. В этом случае кратчайший путь будет состоять из отрезка rc, а также из отражения отрезка cr''.

Упражнение 2.3. Припишите координаты вершинам рассмотренного выше треугольника и проведите вычисления для получения решения задачи в виде формул для того, чтобы убедиться, насколько это утомительно.

Упражнение 2.4. Рассмотрите задачу отыскания кратчайшего пути, который начинается в данной точке многоугольника и касается заданного числа сторон, не возвращаясь в исходную точку.

Упражнение 2.5 [54а]. Как найти кратчайшее расстояние между двумя точками на поверхности многогранника (определение в разделе 2.3), конуса и цилиндра? (Для цилиндра решением является дуга винтовой линии, которая обвивает цилиндр и проходит через данные точки.)

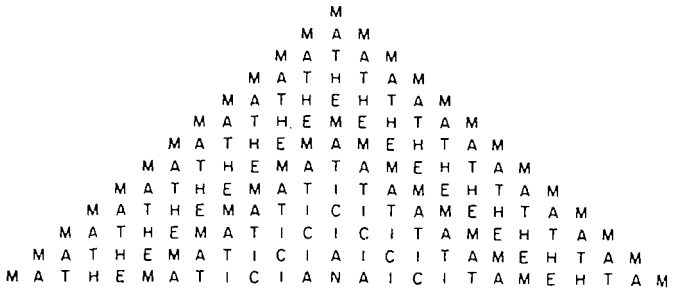
Указание: надо расплющить фигуры на плоскости.

Симметрия при подсчете

Следующий элементарный пример хорошо иллюстрирует использование симметрии при подсчете.

Пример. Найдите число путей на фиг. 2.3, которые образуют слово «MATHEMATICIAN» [66].

Решение 1. Можно пересчитать пути, ведущие «назад» от буквы N. При подсчете в левой части таблицы, включая центральный столбец, на каждом шагу назад имеются два выбора. Таким образом, это



Фиг. 2.3.

дает 2^{12} путей. Удваивая это число и вычитая центральный столбец, учтенный дважды, получаем $2^{13} - 1 = 8191$ путь.

Решение 2. Каждый путь должен начинаться с буквы M, лежащей на границе, и оканчиваться в единственной букве N. Пути, целиком лежащие в левой части таблицы (включая вертикальный путь), однозначно соответствуют словам из двенадцати букв, каждое из которых выбирается из пары (H, V), где H означает *горизонтальный*, а V — *вертикальный*. Количество таких слов равно $2^{12} = 4096$. Аналогично число путей, лежащих целиком в правой части таблицы (исключая вертикальный путь), равно $2^{12} - 1 = 4095$. Таким образом, общее число разных путей равно 8191, поскольку каждый путь лежит в одной из половин таблицы.

Пример. Элементарный пример использования симметрии дан Гауссом для вычисления суммы n чисел, образующих арифметическую прогрессию. Так,

$$1 + 2 + 3 + \ldots + n$$

плюс (та же сумма, записанная в обратном порядке)

$$n + (n - 1) + (n - 2) + \ldots + 1$$

дает после сложения каждого члена нижней суммы со стоящим непосредственно над ним членом верхней суммы

$$(n + 1) + (n + 1) + \ldots + (n + 1) = n(n + 1),$$

откуда следует, что сама сумма равна $n(n + 1)/2$.

Симметрию можно использовать для преобразования несимметричной выпуклой фигуры в симметричную. Обычно это преобразование делается относительно точки, линии, плоскости и т. д. На плоскости симметричное преобразование относительно линии или фикси-

рованной оси симметрии L заменяет выпуклую фигуру новой следующим образом: каждая хорда C (отрезок линии, концы которого лежат на выпуклой граничной кривой) фигуры, перпендикулярная к L, перемещается вдоль своего продолжения в новое положение, симметричное относительно L. В результате появляется фигура, симметричная относительно L.

Для симметризации выпуклой фигуры относительно точки часто используется следующий метод: проводится полоса, определенная двумя параллельными линиями, касательными к границе фигуры, и затем проводится новая полоса, перпендикулярная к ней, граничные линии которой расположены относительно точки, взятой в качестве центра симметрии, на таком же расстоянии, как граничные линии исходной полосы. Если повторять этот процесс для бесконечного числа возможных пар параллельных линий, то их пересечения образуют выпуклую фигуру, симметричную относительно точки. (Другие понятия симметрии будут встречаться на протяжении этой главы в различных контекстах.)

2.3. Многоугольники и многогранники

В этом разделе будут рассматриваться главным образом многоугольники и многогранники (эти понятия будут часто встречаться в этой главе). В нескольких местах, особенно в связи с графами и картами, воспользуемся формулой Эйлера. Эти сведения будут использованы также для построения границ в итерациях задач линейной оптимизации, в которых множество ограничений представляет собой многогранник. Многие вопросы оптимизации, связанные с многогранниками, будут рассмотрены в разделах, соответствующих некоторым темам, указанным в табл. 2.1.

Определение. Многоугольник — это конечное множество отрезков, такое, что в каждой конечной точке каждого отрезка встречаются в точности два отрезка, и ни одно подмножество этих отрезков не обладает таким свойством. Отрезки называются *сторонами* или *ребрами* многоугольника, а конечные точки — *вершинами*. В плоском многоугольнике каждой вершине можно поставить в соответствие угол, описывающий изменения в направлении движения точки вдоль ребер, которая возвращается в исходное положение.

Определение. Многоугольник называется правильным, если все его ребра и углы равны между собой.

Определение. Многогранником в трехмерном пространстве называется конечное множество многоугольников, такое, что каждое ребро одного многоугольника принадлежит в точности другому многоугольнику, и никакое подмножество многоугольников не обладает этим свойством. Многоугольники являются гранями многогранника; их ребра и вершины представляют собой ребра и вершины многогранника.

Каждой вершине многогранника можно поставить в соответствие фигуру (которая в общем случае представляет собой асимметричный многоугольник), образованную линиями, соединяющими средние точки каждой пары ребер граней, примыкающих к этой вершине. Многогранник называется правильным, если все такие фигуры при вершинах и все грани многогранника являются правильными многоугольниками.

Определение. Множество точек в n-мерном пространстве, координаты которых удовлетворяют линейному уравнению

$$a_1x_1 + \ldots + a_nx_n = b,$$

где не все a_i, $i = 1, \ldots, n$, равны нулю, a_i и b — действительные числа, называется *гиперплоскостью*, или $(n-1)$-*мерным подпространством*

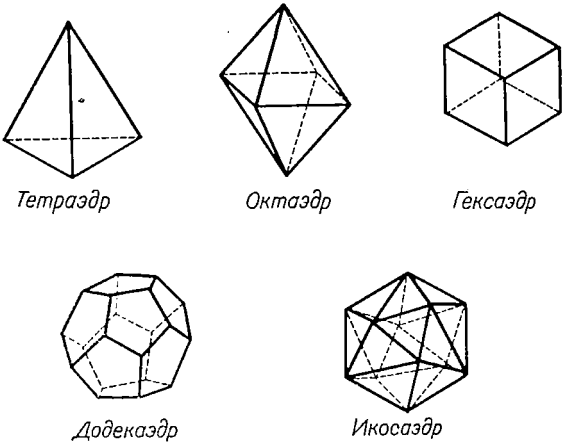
Определение. Выпуклым многогранником в n-мерном пространстве называется выпуклая оболочка конечного множества точек, не все из которых принадлежат одной гиперплоскости.

Определение. Симплексом называется выпуклая оболочка $(n+1)$ точек в n-мерном пространстве, не все из которых принадлежат одной гиперплоскости.

Мы уже упоминали (в предисловии) об идеях, приводящих к следующей теореме.

Теорема 2.1. *В трехмерном пространстве существует не более пяти правильных выпуклых многогранников.*

На фиг. 2.4 показано, что существует не менее пяти таких многогранников.



Фиг. 2.4. Пять правильных выпуклых многогранников в трехмерном пространстве.

Упражнение 2.6. Докажите эту теорему, рассматривая правильные треугольники, квадраты и пятиугольники, которые имеют общие вершины. Сколько таких многоугольников можно свести в одной такой точке без наложений? Можно ли использовать меньшее число многоугольников в каждой вершине правильного многогранника?

Замечание. Существуют четыре невыпуклых правильных многогранника. Грани каждого такого многогранника представляют собой правильные треугольники.

Некоторые свойства выпуклых многогранников полезны при анализе задач оптимизации. Ограничения, существующие в задаче, могут привести к такому выпуклому множеству. Например, в линейном программировании множество S, определяемое ограничениями

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

$$x_j \geqslant 0, \quad j = 1, \ldots, n,$$

является пересечением полупространств. Это выпуклое многогранное множество. (См. также гл. 5.)

Упражнение 2.7. Покажите, что пересечение полупространств с линейными границами представляет собой выпуклое множество, границей которого является многогранник, возможно открытый.

Соотношение Эйлера

Рассмотрим теперь некоторые ограничения на соотношения между вершинами V, ребрами E и гранями F замкнутого многогранника в трехмерном пространстве. Легко показать, что соотношение

$$V - E + F = 2$$

всегда имеет место для такого многогранника.

Упражнение 2.8. Докажите это соотношение, вырезая из многогранника некоторую грань, расплющивая его на плоскости так, чтобы не перекрывались оставшиеся грани, и замечая, что $V - E + F$ не меняется, если две вершины в области связать ребром или добавить новую вершину и связать ее с каждой из двух существующих вершин некоторого ребра. Операции, обратные к вышеуказанным, также не меняют $V - E + F$. Сведите фигуру на плоскости к треугольнику, для которого $V - E + F = 1$. С учетом вырезанной грани правая часть соотношения равна 2 для многогранника.

Соотношение Эйлера определяет функциональную зависимость числа граней какой-то одной размерности от числа граней двух других размерностей. Заметим, что для построения многогранника в вершине должны сходиться не менее трех ребер. Так как каждое

ребро соединяет две вершины,

$$3V \leqslant 2E,$$

т. е. удвоенное число ребер не менее утроенного числа вершин. Эта граница точная и не может быть улучшена. Равенство имеет место в случае тетраэдра, где $V = 4$ и $E = 6$. Так как $V = E + 2 - F$, получаем $3(E + 2 - F) \leqslant 2E$, или

$$E \leqslant 3F - 6.$$

Аналогично каждая грань многогранника ограничена тремя ребрами, и каждое ребро служит границей двух граней. Это дает

$$3F \leqslant 2E.$$

Снова получаем

$$E \leqslant 3V - 6.$$

Наконец, из $3V \leqslant 2E$ и $E \leqslant 3F - 6$ имеем

$$V \leqslant 2F - 4.$$

Аналогично

$$F \leqslant 2V - 4.$$

Таким образом, получаем 6 точных границ и одно соотношение в виде равенства между гранями многогранника.

Многогранник должен иметь не менее чем одну грань с не более чем пятью ребрами. В противном случае если каждая грань имеет не менее шести ребер, то $6F \leqslant 2E$ и ввиду $E \leqslant 3F - 6$ получаем $-6 \geqslant 0$, что невозможно. Если многогранник имеет менее двенадцати граней, то *по крайней мере* одна грань имеет менее пяти ребер. В противном случае $5F \leqslant 2E$ и ввиду $E \leqslant 3F - 6$ получаем $F \geqslant 12$, что невозможно.

Если F_i, $i = 0, 1, \ldots, n - 1$, обозначает число i-мерных граней многогранника в n-мерном пространстве, то формула Эйлера допускает следующее обобщение:

$$\sum_{i=0}^{n-1} (-1)^i F_i = 1 - (-1)^n, \quad n = 1, 2, \ldots .$$

Одним из методов, используемых при решении задач линейного программирования с n переменными, является симплекс-метод (см. гл. 5). Это систематический метод поиска вершины многогранного множества в n-мерном пространстве, в которой достигается оптимум линейной формы. Таким образом, верхняя граница для числа вершин может служить полезной оценкой максимального количества времени, которое может понадобиться для решения такой задачи. Так как число ограничений определяет число граней многогранного множества, полезно выразить верхнюю границу числа вершин F_0 через число граней наивысшей размерности F_{n-1} [32]. Частично доказанная (для некоторых n), частично гипотетическая граница выражается через биномиальные коэффициенты (числа

сочетаний):

$$F_0 \leqslant \binom{F_{n-1} - \left[\frac{n+1}{2}\right]}{F_{n-1} - n} + \binom{F_{n-1} - \left[\frac{n+2}{2}\right]}{F_{n-1} - n}.$$

Напомним, что здесь $[x]$ — наибольшее целое значение, не превосходящее x, а

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

Задача. Найдите твердое тело в форме многогранника, сделанное из однородного материала, с наименьшим числом граней, которое устойчиво только на одной грани, т. е. оно может покоиться, не падая под действием силы тяжести, только на одной из своих граней. Голдберг нашел такое тело с 21 гранями. Грубо говоря, это цилиндр, в котором со стороны каждого основания сделаны срезы под острым углом, а его цилиндрическая поверхность аккуратно обточена по всей длине, так что образованы 19 многоугольников, постоянно убывающих по величине. Пример еще не был опубликован к моменту написания этой работы.

2.4. Разбиения или разложения

В этом разделе будет дано несколько элементарных примеров геометрических задач, в которых встречаются разбиения. Цель состоит в том, чтобы либо получить при разбиении наибольшее число частей, либо найти минимальное число разбиений, дающих заданное число частей. Широкий класс задач, которые кратко описаны здесь, не является классом задач оптимизации. В них определяются условия, при которых из частей одной конфигурации можно составить другую конфигурацию.

В формулировках таких задач оптимизации часто предполагается поиск экстремума, хотя прямо об этом не говорится. Мы будем изменять формулировки задач так, чтобы они непосредственно включали требование оптимизации. Для примера рассмотрим следующую задачу: «На сколько частей трехмерное пространство разбивается n произвольно расположенными плоскостями? Любые три из этих плоскостей имеют одну общую точку, и никакие четыре или большее число плоскостей не имеют общих точек». Эту задачу можно сформулировать и в следующем виде [50а]:

Теорема 2.2. *Максимальное число частей, на которое n произвольно расположенных плоскостей делят трехмерное пространство, равно* $(n + 1)\,(n^2 - n + 6)/6$.

Нам понадобятся две леммы. Доказательство первой леммы очевидно.

Лемма 2.1. *Прямая линия делится n точками на $(n+1)$ частей.*

Упражнение 2.9. Проведите доказательство этой леммы.

Лемма 2.2. *Максимальное число частей, на которое n линий делят плоскость, равно* $(n^2+n+2)/2$.

Доказательство. Если x_n — максимальное число частей, на которые n линий делят плоскость, и если x_{n+1} — максимум для $(n+1)$ линий, то вследствие того, что каждая линия пересекает n остальных линий в n точках, по лемме 2.1 линия делится на $(n+1)$ частей остальными n линиями. Поскольку каждая из этих частей должна быть границей двух частей плоскости, одна из которых образована n линиями, а другая, новая, образована путем добавления $(n+1)$-й линии, получаем $(n+1)$ новых частей плоскости. Это дает

$$x_{n+1} = x_n + n + 1,$$

откуда

$$\begin{aligned} x_n &= x_{n-1} + n, \\ x_{n-1} &= x_{n-2} + n - 1, \\ &\cdots\cdots\cdots\cdots \\ x_2 &= x_1 + 2. \end{aligned}$$

Очевидно, что $x_1 = 2$. Поэтому

$$x_n = x_1 + n + (n-1) + \ldots + 2 = 1 + \sum_{i=1}^{n} i = 1 + \frac{n(n+1)}{2} = \frac{n^2+n+2}{2}.$$

Доказательство теоремы 2.2 Пусть y_n — максимальное число частей, на которые n плоскостей делят трехмерное пространство, и пусть y_{n+1} — соответствующая величина для $(n+1)$ плоскостей. Любая плоскость пересекает остальные n плоскостей по n линиям. По лемме 2.2 эти линии делят плоскость на $(n^2+n+2)/2$ частей, каждая из которых является границей старой и новой частей пространства. Это дает дополнительно $(n^2+n+2)/2$ частей пространства. Поэтому

$$y_{n+1} = y_n + \frac{n^2+n+2}{2}$$

и, поскольку $y_1 = 2$, получаем путем подстановок и упрощений

$$y_n = \frac{(n+1)(n^2-n+6)}{6}.$$

Упражнение 2.10. Докажите, что прямая линия делится на $(2n+1)$ частей n окружностями (т. е. n парами точек) и что n пар точек делят окружность на $2n$ частей.

Упражнение 2.11. Докажите, что плоскость и поверхность сферы каждая делятся n попарно пересекающимися окружностями на (n^2-n+2) частей.

Упражнение 2.12. Докажите, что пространство делится n сферами максимум на $n(n^2 - 3n + 8)/3$ частей. Вместо требования максимума можно потребовать, чтобы все сферы попарно пересекались.

Упражнение 2.13. В трехмерном пространстве произвольно расположены шесть точек (никакие три не лежат на одной прямой, и никакие четыре не лежат в одной плоскости). Каждый из пятнадцати отрезков прямой, соединяющих их попарно, закрашивается или в красный, или в голубой цвет. Докажите, что найдется треугольник, все стороны которого закрашены в один цвет.

Альтернативная формулировка. Докажите, что в группе из любых шести человек найдутся трое людей, которые или все знакомы между собой, или все незнакомы.

Рассмотрим n точек на плоскости, никакие три из которых не коллинеарны. Пусть каждая из этих точек соответствует одному из n индивидуумов. Если какие-то два человека знакомы, то соответствующая пара точек соединяется отрезком прямой, в противном случае не соединяется. Полным треугольником назовем треугольник, все пары вершин которого соединены между собой, что указывает на то, что каждая пара соответствует людям, знакомым между собой. Пустой треугольник — это множество из трех несвязанных точек, что указывает на отсутствие знакомства между соответствующими лицами. Имеет место следующая теорема:

Теорема 2.3 (Гудмен). *Пусть E и F — числа пустых и полных треугольников соответственно. Тогда при любом взаимном расположении n точек* [31]

$$E + F \geqslant \begin{cases} \dfrac{u(u-1)(u-2)}{3}, & \text{если } n \geqslant 2u, \\ \dfrac{2u(u-1)(4u+1)}{3}, & \text{если } n \geqslant 4u+1, \\ \dfrac{2u(u+1)(4u-1)}{3}, & \text{если } n = 4u+3, \end{cases}$$

где u — неотрицательное целое число. Эта нижняя граница является точной для любого положительного целого n.

Эта теорема еще не обобщена на случай полных и пустых четырехугольников и фигур с бо́льшим числом сторон.

Задача о кубе. Предположим, что требуется разделить куб со стороной p см на p^3 маленьких кубов со стороной 1 см при помощи ножа. Отрезанные куски можно располагать рядом в любом удобном порядке перед следующим разрезанием. Каково наименьшее число разрезов? Заметим, что если $p = 2$, то необходимо 3 разреза. Если $p = 3$, то необходимо шесть разрезов. Если $p = 4$, можно сделать один разрез вертикально вдоль средней линии верхней грани и затем составить два куска для выполнения следующего разреза; при

этом получаются четыре плиты каждая размером 4^2 и т. д. Всего необходимо шесть разрезов [89].

Эта задача является частным случаем следующей теоремы:

Теорема 2.4. *Наименьшее число плоских разрезов прямоугольного параллелепипеда со сторонами p, q и r, необходимое для разрезания его на pqr единичных кубов, равно $\alpha + \beta + \gamma$, где α, β и γ — наименьшие целые числа, такие, что*

$$2^{\alpha-1} \leqslant p \leqslant 2^{\alpha}, \quad 2^{\beta-1} \leqslant q \leqslant 2^{\beta}, \quad 2^{\gamma-1} \leqslant r \leqslant 2^{\gamma}.$$

Доказательство. Рассматривая аналогичные разрезы прямой линии, замечаем, что необходимо не менее n разрезов для того, чтобы получить 2^n различных отрезков. Для этого разрезаем отрезок пополам, составляем две части, затем снова разрезаем пополам и составляем вместе полученные части и т. д. Если $p = 2^{\alpha}$, $q = 2^{\beta}$, $r = 2^{\gamma}$, то, чтобы получить $pqr = 2^{\alpha+\beta+\gamma}$ частей, необходимо не менее $(\alpha + \beta + \gamma)$ разрезов. Если какой-то член из p, q, r не представляется в виде степени 2, параллелепипед можно рассматривать как часть большего параллелепипеда, стороны которого равны ближайшей степени 2, т. е., конечно, α, β и γ. Поэтому для данного параллелепипеда требуется не более $(\alpha + \beta + \gamma)$ разрезов. Чтобы показать, что это минимальное число, будем рассуждать по индукции. Допустим, что утверждение теоремы справедливо для любого параллелепипеда, меньшего, чем данный. Предположим, что первый оптимальный разрез данного параллелепипеда делит p на две части p_1 и p_2, как можно более близкие друг к другу. Тогда минимальное число разрезов [обозначим его через $A(x, y, z)$ для параллелепипеда со сторонами x, y, z] удовлетворяет следующему соотношению (не менее):

$$A(p, q, r) \geqslant 1 + \max\{A(p_1, q, r), A(p_2, q, r)\} =$$
$$= 1 + A\left(\left[\frac{p}{2}\right], q, r\right) = 1 + [(\alpha - 1) + \beta + \gamma] = \alpha + \beta + \gamma,$$

где $[p/2]$ — ближайшее целое число, большее $p/2$. Так как $A(p, q, r)$ не превосходит и не меньше, чем $\alpha + \beta + \gamma$, имеет место равенство. Этим заканчивается доказательство.

Упражнение 2.14. Тривиально получите результат для случая разрезания прямоугольников.

Упражнение 2.15. Попробуйте обобщить задачу на случай четырехмерного куба.

Равносоставленность [7]

Многоугольник разрезается (разбивается) прямой линией на две части. Каждая часть тоже может быть аналогично разрезана. Многогранник разрезается на две части плоскостью.

Если фигуру можно разрезать и затем из ее частей составить новую фигуру, то говорят, что эти две фигуры равносоставлены.

Согласно теореме Больяй — Гервина, два многоугольника с равными площадями равносоставлены.

Чтобы построить фигуру, центрально симметричную относительно другой фигуры, выберем точку, называемую *центром*, вне данной фигуры и проведем линию из каждой точки фигуры через центр. Определим точку на продолжении прямой линии за центр, расстояние от которой до центра равно расстоянию от исходной точки до центра. Совокупность таких точек дает фигуру, центрально симметричную относительно исходной фигуры. Согласно теореме Хадвигера — Глюра, любые два многоугольника с равными площадями можно разбить таким образом, чтобы каждая часть одного многоугольника могла быть получена из соответствующей части другого при помощи параллельного переноса и центрально симметричного отображения.

Для того чтобы можно было установить равносоставленность выпуклого многоугольника с квадратом с помощью только параллельных переносов, необходимо и достаточно, чтобы этот многоугольник был центрально симметричен.

Двугранный угол многогранника — это угол, образованный двумя гранями внутри многогранника при их общем ребре. Угол измеряется между двумя прямыми линиями, по которым грани пересекаются с плоскостью, перпендикулярной этому ребру.

Пусть α_i, $i = 1, \ldots, p$, обозначают радианные меры двугранных углов многогранника P, $d_1, \ldots, d_p$ — длины соответствующих ребер. Введем функцию $f(P) = \sum_{i=1}^{p} d_i f(\alpha_i)$, где $f(\alpha_i)$ удовлетворяет специальному аддитивному соотношению $\sum_{i=1}^{p} k_i f(\alpha_i) = 0$, если только выполняется равенство $\sum_{i=1}^{p} k_i \alpha_i = 0$ (т. е. если α_i линейно зависимы при некотором выборе ненулевых целых чисел k_i, $i = 1, \ldots, p$). Функция $f(P)$ называется инвариантом P. Она зависит от P и от выбора $f(\alpha_i)$, $i = 1, \ldots, p$. Хадвигер доказал, что два многогранника P и Q с соответствующими двугранными углами α_i ($i = 1, \ldots, p$) и β_j ($j = 1, \ldots, q$) не равновелики, если существует функция f, аддитивная относительно $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \pi$, которая удовлетворяет соотношениям $f(\pi) = 0$, $f(P) \neq f(Q)$. Эту теорему можно обобщить на случай многогранника в n-мерном пространстве, в котором двугранные углы определяются для каждой из $(n-2)$-мерных граней. Каждая такая грань представляет собой пересечение в точности двух $(n-1)$-мерных граней.

Теорема 2.5 (Ден). *Куб и правильный тетраэдр, имеющие одинаковый объем, не равносоставлены* [7].

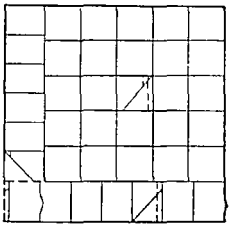
Замечание. Существуют неправильные тетраэдры, которые можно разрезать таким способом и собрать из частей куб.

Доказательство. Двугранный угол куба равен $\pi/2$. Обозначим двугранный угол правильного тетраэдра через φ. Чтобы воспользоваться теоремой Хадвигера, нужно взять углы $\pi/2$, φ, π. Если $k_1\pi/2 + k_2\varphi + k_3\pi = 0$, где $k_i = 1, 2, 3$ — целые числа, то также и $(k_1 + 2k_3)\,\pi + 2k_2\varphi = 0$, и мы получаем линейную зависимость между π и φ. Однако известно, что для тетраэдра $\cos\varphi = {}^1/_3$, и, следовательно, можно показать, что π и φ несоизмеримы (т. е. их отношение не является рациональным числом). Поэтому $k_1 + 2k_3 = 0$, $k_2 = 0$ и после подстановки в верхнее соотношение получим $k_3\pi + (-2k_2)\,\pi/2 = 0$; другой линейной зависимости между $\pi/2$, φ и π не существует. Пусть $f(\pi) = f(\pi/2) = 0$, $f(\varphi) = 1$. Таким образом, функция f аддитивна, так как $k_3 f(\pi) + (-2k_3) f(\pi/2) = 0$. Наконец, нужно показать, что f (куб) $\neq f$ (тетраэдр). Пусть длина каждой из двенадцати сторон куба равна d. Тогда f (куб) $= 12df(\pi/2) = 0$. Если b — длина каждой из сторон тетраэдра, то f (тетраэдр) $= 6bf(\varphi) \neq 0$. Этим заканчивается доказательство.

Теорему можно обобщить на случай n-мерного пространства. В этом случае используется тот факт, что угол A, для которого $\cos A = 1/n$, несоизмерим с π.

Задача. Пусть дано n равных единичных квадратов. Разобьем каждый из них, так же как и раньше, прямыми разрезами на $p(n)$ частей так, что $np(n)$ частей можно собрать в квадрат со стороной $\sqrt{n}$. Исследуйте $k(n)$, минимальные значения $p(n)$, и покажите, что $k(n) \leqslant 5$ при всех n.

Решение [26]. Пусть $[\sqrt{n}]$ — целая часть $\sqrt{n}$. Тогда, если $\sqrt{n} - [\sqrt{n}] > {}^1/_2$, в качестве ребра базисного квадрата выбираем $[\sqrt{n}]$, а если $\sqrt{n} - [\sqrt{n}] < {}^1/_2$, выбираем $[\sqrt{n}] - 1$. Если базисный квадрат помещается в углу квадрата со стороной $\sqrt{n}$, как показано на фиг. 2.5, он будет ограничен L-образной областью шириной w, которая ограничена условиями ${}^3/_2 > w > {}^1/_2$. При помощи одного разреза эту L-образную область можно превратить в прямоугольник шириной w.

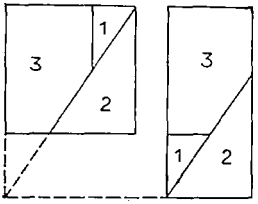


Фиг. 2.5.

Заполним базисный квадрат необходимым числом данных единичных квадратов. Остальные единичные квадраты следует теперь преобразовать путем рассечения на прямоугольники шириной w. Так как ${}^3/_2 > w > {}^1/_2$, для этого преобразования никогда не потребуется разрезания более чем на три части, как показано сплошными линиями. (См. отдельное изображение этого построения.) Пунктирные линии, по которым осуществляется разрезание в прямоугольнике шириной w на части, из которых можно составить L-образную

область, будут при соответствующей ориентации сторон основного прямоугольника рассекать один из основных прямоугольников самое большее на пять частей. Все другие основные прямоугольники и квадраты должны быть разрезаны на такие же пять частей.

Разрезание единичного квадрата. Единичный квадрат рассекается на три части, из которых можно составить прямоугольник шириной w, где $^3/_2 > w \geqslant {}^1/_2$, как показано на фиг. 2.6. Верхний край части 1 никогда не превышает половины длины прямоугольника по горизонтали. Поэтому вертикальный разрез, делящий этот прямоугольник в любом искомом отношении, всегда можно выбрать так, что он будет проходить через части 2 и 3 и не затрагивать часть 1. Поэтому получится только пять частей.



Фиг. 2.6.

При $1 \geqslant w \geqslant {}^1/_2$ разрезы осуществляются таким же образом, за исключением того, что w является малой стороной прямоугольника. Горизонтальный разрез, делящий прямоугольник в любом искомом отношении, всегда может быть выбран так, что часть 1 не будет разрезаться. Поэтому будет получено самое большее пять частей. Если разрезается только часть 3, то будут получены только четыре части.

Голдберг и Стюарт [28] показали, что $k(n) \leqslant 4$ и что это число не может быть улучшено. Для первых 99 квадратов они построили табл. 2.2. Число квадратов получается путем взятия первой цифры

Таблица 2.2

	0	1	2	3	4	5	6	7	8	9
0		1	2	3	1	2	3	4	2	1
1	2	4	3	2	3	3	1	3	2	4
2	2	3	4	4	3	1	3	3	3	3
3	3	4	2	3	3	3	1	3	4	4
4	2	2	3	4	3	2	4	4	3	1
5	2	4	2	3	3	3	3	3	3	4
6	3	2	4	3	1	3	3	4	3	3
7	3	4	2	3	2	3	3	3	3	4
8	2	1	3	4	3	2	4	4	3	3
9	2	3	3	3	4	3	3	3	2	3

из левого столбца, а второй цифры из верхней строки. Например, в случае 25 квадратов не требуется делать разрезы и, следовательно, каждый маленький квадрат остается как одна часть. Они указывают,

что аналогично можно исследовать эту задачу для равностороннего треугольника, правильного шестиугольника и L-образной области, образованной путем удаления угловой четверти квадрата.

2.5. Примеры изопериметрических задач и задач поиска кратчайшего пути

В этом разделе рассматриваются задачи оптимизации, но в большинстве из них требуется находить решения с определенными свойствами. За исключением некоторых специальных случаев, в этом разделе будет мало вычислений. Однако методы, использованные в доказательствах, полезны и представляют интерес. Некоторые утверждения только формулируются. Приводится несколько доказательств.

В своей интересной монографии Казаринофф [47] приводит большое число элементарных задач о замкнутых периметрах и площадях, ограниченных ими на плоскости. В некоторых случаях задается периметр и требуется максимизировать площадь; в других — задается площадь и надо минимизировать периметр.

Задачи поиска среди всех простых замкнутых фигур определенного вида с заданным периметром фигуры, которая имеет наибольшую площадь на плоскости, называются *изопериметрическими* (с одним и тем же периметром) задачами. Например, среди всех плоских фигур с заданным периметром наибольшую площадь ограничивает окружность (доказательство приводится ниже в теореме 2.6).

Если задана площадь, то наименьший периметр имеет окружность. Аналогично сфера имеет наибольший объем среди всех фигур с заданной площадью поверхности. Среди всех треугольников с одинаковым основанием и фиксированным периметром наибольшую площадь имеет равнобедренный треугольник. Равносторонний треугольник имеет наибольшую площадь среди всех треугольников с заданным периметром.

Упражнение 2.16. Докажите, что если углы n-угольника, вписанного в круг радиуса r, равны $\alpha_1, \ldots, \alpha_n$, то площадь его равна

$$\frac{r^2}{2}\sum_{i=1}^{n}\sin 2\alpha_i.$$

Указание. Докажите, что $\sin x$ — вогнутая функция на интервале $[0, \pi]$ и, следовательно, эта площадь ограничена сверху выражением

$$\frac{r^2}{2}\sin\frac{2\alpha_1+\ldots+2\alpha_n}{n}=\frac{nr^2}{2}\sin\frac{2\pi}{n},$$

которое представляет собой площадь правильного n-угольника [7].

Сейчас будет дана схема хорошо известного доказательства Штайнера общей изопериметрической задачи на плоскости, кото-

рую, конечно, можно решить с помощью вариационного исчисления (см. ниже).

Теорема 2.6. *Среди всех простых замкнутых кривых C заданной длины наибольшую площадь ограничивает окружность* [11] [1]).

Доказательство (*схема*). Хотя последующие рассуждения выглядят эстетически привлекательными, они содержат ошибку, которая заключается в предположении существования решения. Однако существование решения можно доказать другими методами.

Штайнер сначала показывает, что решение существует, доказывая, что при заданном периметре среди всех равносторонних многоугольников наибольшую площадь имеет правильный многоугольник. Затем он показывает, что все вершины этого правильного многоугольника лежат на окружности. При стремлении к бесконечности числа вершин многоугольника многоугольник стремится в пределе к окружности. Поэтому он заключает, что задача имеет решение.

Чтобы получить окружность, отправляясь от произвольной простой замкнутой кривой C, он показывает сначала, что область, ограниченная кривой C, выпуклая. Если эта область где-то вогнутая, то кривую на этом участке можно симметрично отобразить так, что фигура станет выпуклой и площадь ее увеличится. Затем кривая C делится на два участка равной длины и через точки деления проводится прямая линия. Выбирается та часть области, которая имеет бóльшую площадь. Затем показано, что наибольшую площадь ограничивает полукруг (другая половина может быть получена зеркальным отражением). Эта часть теоремы доказывается следующим образом. Выбирается точка на рассматриваемом полупериметре, которая соединяется с оконечными точками, связанными прямой линией. Таким образом получается треугольник. Затем угол, прилегающий к указанной точке, увеличивается или уменьшается; при этом точка используется как дверная петля, около которой строится прямой угол. Это дает другой (прямоугольный) треугольник, площадь которого, как легко видеть, больше площади первоначального треугольника. Таким образом, площадь под кривой увеличивается. Поскольку точка была выбрана произвольно и прямоугольный треугольник может быть образован любой точкой окружности, соединенной с концами диаметра, в любой точке периметра можно построить прямоугольный треугольник. Это дает полуокружность. Путем симметричного отражения получается целая окружность. Этим завершается доказательство.

Заметим, что если длина окружности равна c, то площадь A любой фигуры, длина периметра которой тоже равна c, удовлетворяет неравенству $A \leqslant \pi \left(\frac{c}{2\pi}\right)^2 = \frac{c^2}{4\pi}$, которое называется изопериметрическим неравенством на плоскости

---

[1]) См. также избранные труды Штайнера, Берлин, 1881—1882 гг.

Упражнение 2.17. Докажите методом от противного, что простой замкнутой кривой с наименьшим периметром, которая ограничивает заданную площадь, является окружность. Это обратная изопериметрическая задача для окружности.

Замечание. Вариационный подход к вышеизложенной задаче (в части необходимости) приводит [42] к максимизации $\int_0^1 y\,dx$ при ограничениях $y(0) = y(1) = 0$, $\int_0^1 (1 + y'^2)^{1/2}\,dx = L$, где L задано. Используя уравнение Эйлера относительно лагранжиана $y + \lambda \times (1 + y'^2)^{1/2}$, получаем соотношение

$$\lambda \frac{d}{dx}\left[\frac{y'}{(1+y'^2)^{1/2}}\right] - 1 = 0,$$

которое после интегрирования, решения относительно y' и вторичного интегрирования дает $y = \pm[\lambda^2 - (x - c_1)^2]^{1/2} + c_2$, что после упрощений определяет окружность. Условия $y(0) = y(1) = 0$ и заданное L определяют c_1, c_2 и λ. Если $L > \pi/2$, то решение не однозначно по x.

Упражнение 2.18. Докажите, используя соображения симметрии, что среди всех прямоугольников с заданной площадью наименьший периметр имеет квадрат.

Упражнение 2.19. Распространив доказательство упражнения 2.18 на параллелепипед и n-мерный куб, используя вместо площади понятие объема, покажите, что наименьший периметр имеет n-мерный куб. (Ответ к этому упражнению будет также получен в гл. 4 алгебраическим методом.)

Теорема 2.7. *Правильный n-угольник имеет наименьшую площадь среди всех n-угольников, описанных около окружности* [83].

Доказательство. Пусть P_n — произвольный n-угольник, $\overline{P}_n$ — правильный n-угольник, описанный около окружности c. Рассмотрим окружность C, описанную около $\overline{P}_n$. Пусть P_nC — часть, общая для P_n и C; $s_1, s_2, \ldots, s_n$ — сегменты круга C, отсекаемые последовательными сторонами P_n, а $s_{k,\,k+1}$ — общая часть сегментов s_k и s_{k+1}, причем $s_{n,\,n+1} = s_{n1}$; s — сегмент C, отсекаемый касательной к c. Тогда

$$P_nC = C - ns + (s_{12} + s_{23} + \ldots + s_{n1}).$$

Так как

$$P_nC \geqslant C - ns,$$

следовательно,

$$P_n \geqslant P_nC \geqslant C - ns = \overline{P}_n.$$

Равенство в $P_n \geqslant P_nC$ (в $P_nC \geqslant \overline{P}_n$) имеет место только в том случае, когда ни одна вершина P_n не лежит вне (внутри) C.

Замечание. Используя метод множителей Лагранжа, Демир [16] доказал, что максимальная площадь плоской области, ограниченной простым замкнутым многоугольником с заданными сторонами, имеет место, если многоугольник может быть вписан в окружность. В доказательстве многоугольник делится на треугольники путем проведения ребер, соединяющих выбранную вершину со всеми остальными. Стороны многоугольника, сходящиеся в этой вершине, обозначаются через r_1 и r_n, а новые стороны, лежащие между ними,— через r_2, $r_3, \ldots, r_{n-1}$. Измеряются все углы $\theta_1, \theta_2, \ldots, \theta_{n-1}$ при этой вершине по отношению к r_1. Площадь многоугольника, которая должна быть максимизирована относительно $\theta_1, \ldots, \theta_{n-1}, r_2, \ldots, r_{n-1}$, задается выражением

$$S = \frac{1}{2} \sum_{i=1}^{n-1} r_i r_{i+1} \sin \Delta\theta_i,$$

где

$$\Delta\theta_i = \theta_i - \theta_{i-1}, \quad i = 1, \ldots, n-1, \quad \theta_0 = 0,$$

при ограничениях на r_i

$$g_i \equiv r_i^2 + r_{i+1}^2 - a_i^2 - 2r_i r_{i+1} \cos \Delta\theta_i = 0, \quad i = 1, \ldots, n-1,$$

где a_i, $i = 0, \ldots, n$,— i-я сторона многоугольника, причем отсчет начинается от $a_0 = r_1$ и заканчивается при $a_n = r_n$.

Упражнение 2.20. Найдите четыре треугольника, для которых сумма периметров минимальна при следующих ограничениях [60]:

1. Длины всех сторон треугольников представляют собой целочисленные величины.
2. В каждом треугольнике имеется сторона длиной k.

Изопериметрические задачи в трехмерном пространстве

Задачу об окружности можно обобщить на случай трех измерений. В этом случае требуется доказать, что сфера имеет наибольший объем при заданной площади поверхности. Другая изопериметрическая задача в трехмерном пространстве формулируется следующим образом: среди всех кривых фиксированной длины, соединяющих две точки, требуется найти кривую, при вращении которой вокруг заданной оси образуется фигура с наименьшей поверхностью. В этом случае искомой кривой является (в стандартном положении) цепная линия [54а]. Задача Платона является обобщением этой идеи и заключается в поиске поверхности минимальной площади, ограниченной заданной пространственной кривой.

Упражнение 2.21. Пусть n — число граней многогранника в трехмерном пространстве. Требуется показать, что для каждого $n > 4$ существуют пирамида, основанием которой является многоугольник с $(n-1)$ сторонами, а боковыми сторонами — треугольники, и призма, нижним и верхним основаниями которой являются многоугольники с $(n-2)$ сторонами, а боковыми сторонами — четырехугольники. При каждом n имеется бесконечно много таких пирамид и призм. При $n = 4$ существуют только пирамиды.

Упражнение 2.22. Пусть n — число граней многогранника и $4 \leqslant n \leqslant 8$. Требуется построить все возможные многогранники в соответствии с числом ребер каждой грани, т. е. с использованием комбинаций треугольников, четырехугольников, пятиугольников и т. д.

Изопериметрическая задача для многогранника

Что представляет собой выпуклый многогранник, который имеет наибольший объем V при заданном числе граней и заданной площади поверхности S? Естественно, сразу же возникает мысль о правильном многограннике с соответствующим числом граней. Однако ни октаэдр, ни икосаэдр не принадлежат к классу решений. Ответ известен только для многогранников с 4, 5, 6, 7 и 12 гранями.

Пусть P — многогранник с n гранями, такой, что при заданной площади поверхности S его объем V наибольший. Можно назвать такой многогранник оптимальным. Кроме того, он обладает таким свойством, что при данном его объеме V площадь S минимальна (доказательство предоставляется читателю). Это свойство влечет за собой предыдущее. Таким образом, оптимальный многогранник можно определить двумя способами, которые можно объединить следующим образом: P — оптимальный многогранник тогда и только тогда, когда изопериметрическое отношение S^3/V^2 минимально (очевидно).

В теореме Линделефа утверждается, что грани оптимального многогранника при любом их числе являются в своих центрах тяжести касательными плоскостями к вписанной сфере [82]. Таким образом, задача сводится к следующей: среди всех многогранников с n гранями, описанных около единичной сферы, требуется найти оптимальный. Поскольку для этих многогранников $S^3/V^2 = 27V$ (это можно доказать путем рассечения многогранника на пирамиды с общей вершиной в центре сферы), оптимальными многогранниками будут те, которые имеют наименьший объем. Голдберг [27] первым получил следующий результат для произвольного многогранника с F гранями, описанного около сферы:

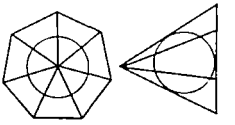
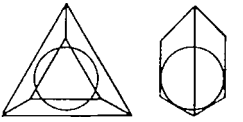
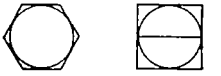
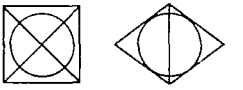
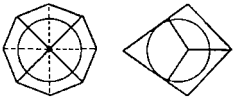
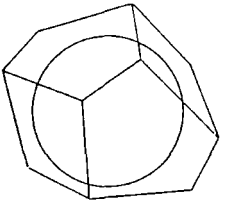
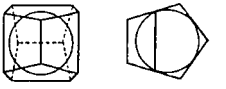
$$\frac{S^3}{V^2} \geqslant 54\,(F-2)\,\operatorname{tg}\omega_F\,(4\sin^2\omega_F - 1),$$

$$\omega_F = \frac{F}{F-2}\,\frac{\pi}{6}\,.$$

Равенство имеет место для четырехгранника, куба и десятигранника. Этот результат независимо получил Тот [82] и доказал Флориан.

Для случая восьми граней, для которого ответ еще не известен, Голдберг разработал экспериментальную схему, приведенную ниже, которая позволяет предположить, что сдвоенная скошенная пирамида дает искомый ответ. Задачу можно сформулировать следующим образом.

Какова форма и чему равен объем восьмигранника наименьшего объема, который можно описать около сферы единичного радиуса? Примеры восьмигранников приведены на фиг. 2.7.

Форма	Объем	Проекция на горизонтальную плоскость	Проекция на боковую поверхность
Простая пирамида	8,9894		
Усеченная двойная пирамида	8,0467		
Шестиугольная призма	6,9282		
Правильный октаэдр	6,9282		
Двойная скошенная пирамида	6,7241		
	6,7009		

Фиг. 2.7.

Задача Малфатти (нерешенная)

Другим примером геометрической задачи оптимизации при наличии ограничений является нерешенная до сих пор задача Малфатти, поставленная в 1803 г. Здесь будет дана только формулировка этой задачи.

Как надо расположить внутри прямой треугольной призмы, например из мрамора, три круговых цилиндра с высотой, равной высоте призмы, чтобы пришлось удалить наименьшее количество материала?

Эта задача сводится к плоской задаче вырезания трех кругов из заданного треугольника так, чтобы их суммарная площадь была максимальной [29].

Две задачи о путях

О кратчайшем пути через n точек на плоскости. Хорошо известна и сейчас является уже классической задача комбинаторной математики о коммивояжере. Коммивояжер должен проехать n городов и вернуться в исходную точку, посетив каждый город только один раз и проделав при этом самый короткий путь (см. в гл. 5 алгебраическую формулировку). Для случая, когда n городов расположены в единичном квадрате евклидовой плоскости так, что расстояния между ними соответствуют расстояниям на плоскости, теорема 2.8 дает интересную границу для суммарного расстояния, которое коммивояжер должен преодолеть при условии, что он не должен возвращаться в исходную точку. Эта оценка неприменима в случае, когда расстояния между парами точек приписываются произвольно (такие задачи будут рассматриваться в разд. 2.6).

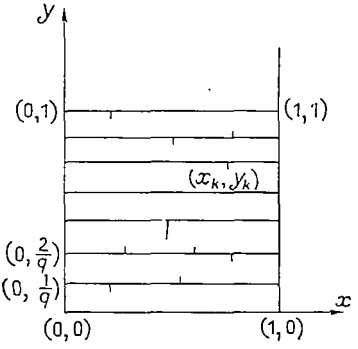
Замечание. В этой связи интересно отметить, что кратчайший замкнутый путь, соединяющий n точек на плоскости, не все из которых лежат на одной прямой, образует простой многоугольник. В частности, если выпуклая оболочка множества не содержит внутри ни одной из n точек, то его граница представляет собой кратчайший замкнутый путь. (Поэтому в соответствующей задаче о коммивояжере не должно быть пересечений путей.) Чтобы доказать это, соединим n точек любым замкнутым путем и заметим, что более короткий путь получится, если соединить точки в том же порядке отрезками прямой линии с данными точками как единственными вершинами. Если сегменты v_iv_{i+1} и v_jv_{j+1} пересекаются в некоторой точке v, предположим, что путь имеет вид $v_ivv_{i+1} \ldots v_jvv_{j+1} \ldots v_i$. Если v не принадлежит множеству точек, то $v_iv_j \ldots v_{i+1}v_{j+1} \ldots v_i$ является более коротким путем, который не содержит пересечения v. Если, напротив, v принадлежит данному множеству точек, то $v_ivv_j \ldots \ldots v_{i+1}v_{j+1} \ldots v_i$ представляет собой многоугольный путь, в котором v не является точкой пересечения. (Этот материал поясняет рассуждения, приведенные в работе [86].)

Следующую теорему предложил Фью [23].

Теорема 2.8. *При условии, что в единичном квадрате расположены произвольно n ($n \geqslant 2$) точек, найдется путь через эти n точек, длина которого не превышает* $\sqrt{2n} + 1{,}75$.

Доказательство. Пусть дан единичный квадрат $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, и координаты n точек имеют вид $(x_1, y_1), \ldots, (x_n, y_n)$. Рассмотрим q горизонтальных линий $y = 0, 1/q, 2/q, \ldots, 1$, где q — произвольное число, и проведем от каждой из n точек перпендикуляр к ближайшей из $(q+1)$ линий (фиг. 2.8). Повторим построение, используя q линий:

$$y = 0, 1/2q, 3/2q, \ldots, (2q-1)/2q.$$



Фиг. 2.8.

Каждое построение дает путь, состоящий из соответствующих отрезков линий $x = 0$, $0 \leqslant y \leqslant 1$, и $x = 1$, $0 \leqslant y \leqslant 1$, и отрезков перпендикуляров, учитываемых дважды — один раз в направлении от горизонтальной линии к точке, а второй раз в обратном направлении. Если обозначить эти два пути через L_1 и L_2 соответственно, то получим

$$L_1 = q + 1 + 2\sum_{i=1}^{n} q^{-1} \| qy_i \| + 1,$$

$$L_2 = q + 2\sum_{i=1}^{n} q^{-1} \left\| qy_i - \frac{1}{2} \right\| + 1,$$

где $\| \alpha \|$ — абсолютное значение разности между α и ближайшим целым числом. Заметим, что

$$\| \alpha \| + \left\| \alpha - \frac{1}{2} \right\| = \frac{1}{2}.$$

Имеем

$$L_1 + L_2 = 2q + 3 + 2q^{-1}\frac{n}{2} = 2q + 3 + nq^{-1}.$$

Выберем целое число q так, чтобы минимизировать эту величину. Это целое число, ближайшее к $\sqrt{n/2}$. Таким образом, $n = 2(q+\theta)^2$, где $|\theta| \leqslant 1/2$. Подставляя это выражение для n, получаем

$$L_1 + L_2 = 2q + 3 + nq^{-1} \leqslant 2(2n)^{1/2} + \frac{7}{2}.$$

Следовательно, длина одного из двух путей не превышает $\sqrt{2n} + 1{,}75$.

Упражнение 2.23. Используя построение, аналогичное описанному выше, покажите, что если вертикальная линия, ведущая к каждой

точке, учитывается только один раз, то длина одного из двух построенных «путей» не превышает $\sqrt{n} + 1{,}75$.

Следующая теорема является обобщением построения, предложенного в последнем примере.

Теорема 2.8а. *При условии, что в единичном k-мерном кубе даны n точек, при фиксированном k и $n \to \infty$ найдется путь (без повторений отрезков линии) через n точек, длина которого не превосходит*

$$k\,[8\,(k-1)]^{(1-k)/2k}\,n^{1-1/k} + O\,(n^{1-2k}).$$

Доказательство. Как и при получении оценки в доказательстве упражнения, которое близко к доказательству предыдущей теоремы, найдем в k-мерном случае путь, длина которого не превосходит

$$(q+1)^{k-1} + q^{-1}\,[(q+1)^{k-1} - 1] + \frac{1}{2}\,nq^{-1}\left(\frac{k-1}{2}\right)^{1/2}.$$

Выбор $q = \{[n^2/8\,(k-1)]^{1/2k}\} + 1$ дает верхнюю оценку для длины пути, задаваемой выражением в формулировке теоремы.

Асимптотическая оценка длины кратчайшего пути между n точками в единичном k-мерном кубе имеет вид

$$\left[\Gamma\left(1+\frac{k}{2}\right)\right]^{1/k}\pi^{-1/2}n^{1-1/k}.$$

Чтобы убедиться в этом, заметим, что объем сферы радиуса r в k-мерном пространстве равен

$$V_k = \frac{\pi^{k/2}r^k}{\Gamma\left(\frac{k}{2}+1\right)}.$$

Рассмотрим n точек в k-мерном единичном кубе, которые расположены таким образом, что расстояние между любыми двумя из них не менее $2r$. Таким образом, если каждая точка является центром сферы радиуса r, то сферы образуют упаковку (см. ниже в этой главе). Плотность этой упаковки при фиксированном k и $n \to \infty$ равна $\delta = nV_k$. Это отношение суммы объемов сфер к объему куба. Длина пути равна $2r\,(n-1)$. Выражая r через δ, получим искомый результат, учитывая из обсуждения, которое будет приведено ниже в этой главе, что $\delta \geqslant 1/2^k$.

Задача о максимине. Для случая, когда C является конфигурацией из пяти точек P_i, $i = 1, \ldots, 5$, в замкнутой области D, тре-

буется показать, что [15]

$$\max_{C}\min_{i\neq j}\overline{P_iP_j}=\begin{cases}2^{-1/2}, & \text{если } D\text{ — единичный квадрат,}\\ \frac{1}{2}, & \text{если } D\text{ — единичный равносторонний треугольник,}\\ 2\sin\frac{\pi}{5}, & \text{если } D\text{ — единичный круг,}\\ \frac{\pi}{2}, & \text{если } D\text{ — поверхность единичной сферы,}\end{cases}$$

где $\overline{P_iP_j}$ — расстояние от P_i до P_j.

Решение.

1. Разделим единичный квадрат на четыре квадрата со стороной $^1/_2$. По крайней мере один из них содержит две из пяти точек. В нем максимальное расстояние $2^{-1/2}$ наблюдается между двумя точками, расположенными на противоположных концах диагоналей. Такая же величина получается для единичного квадрата в случае, когда пять точек представляют собой вершины и центр квадрата.

2. Разделим треугольник на четыре равносторонних треугольника со стороной $^1/_2$ и расположим точки в пяти из шести вершин треугольников со стороной $^1/_2$.

3. Разделим круг на пять конгруэнтных секторов и расположим точки по окружности на равном расстоянии одна от другой. Если две точки попадают в сектор с дугой ϕ ($\phi\leqslant 2\pi/5$), то наибольшее расстояние между ними будет

$$\max\left(1,\ 2\sin\frac{\phi}{2}\right)\leqslant 2\sin\frac{\pi}{5}.$$

4. Помещаем одну точку P_1 на северный полюс, а другие на экватор со сдвигом 90° одна от другой. Если четыре точки удалены более чем на $\pi/2$ от P_1, то они будут лежать в южном полушарии; если поверхность единичной сферы разделить на четыре конгруэнтные области, то максимин для этих четырех точек был бы меньше, чем $\pi/2$.

2.6. Графы и сети [1)]

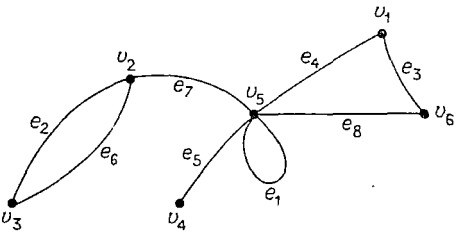
Обобщение задач о кратчайших расстояниях приводит к задачам о кратчайших путях. Здесь удобно ввести графы и приступить к рассмотрению различных задач оптимизации, связанных с ними. Затем будут даны некоторые приложения этих задач к многогранникам.

Граф представляет собой множество точек, называемых *вершинами*, и множество простых кривых, называемых *ребрами*, такое, что каждое замкнутое ребро содержит в точности одну вершину, каждое открытое ребро содержит в точности две вершины, которые являются

[1)] Рекомендуем читателю дополнительно [95*].— *Прим. ред.*

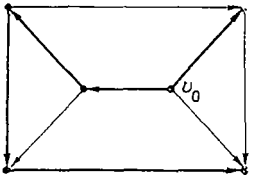
его конечными точками, и никакие ребры не имеют общих точек, помимо вершин (фиг. 2.9). С каждым ребром можно ассоциировать направление, указываемое стрелкой. В этом случае полученный граф называется *направленным графом*, а его ребра — *дугами* (фиг. 2.10).

Граф называется *двудольным*, если его вершины можно разбить на два несвязных множества, таких, что единственными ребрами в графе являются те, которые связывают вершины одного множества с вершинами другого.



Фиг. 2.9. Ненаправленный граф.

Простая цепь (путь) в ненаправленном (направленном) графе представляет собой последовательность ребер (дуг) и вершин, в которой не повторяется ни одна вершина; контур (цикл) представляет собой цепь (путь), начальная и конечная вершины которого совпадают. Говорят, что граф является связным без учета направленности, если существует простая цепь между любой парой вершин. Граф с $(n+1)$ вершинами является n-кратно связным, если удаление $(n-1)$ или меньшего числа вершин не делает его несвязным. Говорят, что две цепи не пересекаются, если они не имеют общих вершин, за исключением конечных точек. Дерево представляет собой связный подграф, в котором отсутствуют контуры. Стягивающее дерево (остов) представляет собой (максимальное) дерево, которое содержит все вершины графа. Ребро графа, которое входит в дерево, называется *ветвью*. Ребро графа, которое не входит в дерево, называется *хордой*. На фиг. 2.10 показано стягивающее дерево направленного графа. Дерево начинается в v_0, откуда исходят все пути дерева.



Фиг. 2.10. Направленный граф, в котором выделено дерево.

Сеть представляет собой связный граф, в котором каждому ребру и каждой вершине поставлено в соответствие неотрицательное число, называемое *пропускной способностью* [29]. Часто в сети выделяются две вершины v_α и v_ω; v_α называется *источником*, а v_ω — *стоком*. Пара, состоящая из пути от v_α до v_ω и неотрицательного числа (которое не должно превышать пропускную способность ни одного ребра или вершины вдоль пути), изображающая поток вдоль этого пути от v_α к v_ω, называется *путевым потоком от* v_α *к* v_ω.

Поток в сети представляет собой такой набор путевых потоков, что сумма чисел всех путевых потоков через любую вершину или ребро сети не превышает пропускную способность вершины или ребра. Величина потока представляет собой сумму чисел путевых

потоков, которые составляют его. Разъединяющим множеством в сети называется такой набор ребер и вершин, удаление которых разъединяет v_α и v_ω. Сумма пропускных способностей этих ребер и вершин является величиной разъединяющего множества.

Приведем здесь без доказательства теорему Форда и Фалкерсона [25] о максимальном потоке и минимальном разрезе. Экстремальные задачи о потоках в сетях не будем обсуждать. (Дополнительную информацию можно получить в [25a], где дается глубокое изложение вопросов, связанных с максимальными потоками, и смежных задач.)

Теорема 2.9. *Максимальная величина среди всех потоков (называемая максимальным потоком) от источника v_α к стоку v_ω сети равна минимальной величине среди всех величин разъединяющих множеств (называемой минимальным разрезом).*

В гл. 5 дается алгебраическая формулировка задачи о потоках в сетях.

Вершины и ребра многогранника можно рассматривать как вершины и ребра соответствующего графа в n-мерном пространстве. В связи с этим может возникнуть интерес к исследованию смежных задач в теории графов.

Возвращаясь к обсуждению линейного программирования (раздел 2.3), заметим, что в симплексном процессе не требуется, чтобы линейная функция была максимальной во всех вершинах ограничивающего многогранника; достаточно, чтобы она максимизировалась в некоторых из них. В этом случае могут существовать различные пути, ведущие из начальной вершины в конечную. Иногда для практического рассмотрения полезно бывает знать границы для числа таких путей. Следствие к теореме 2.11, которая будет сформулирована ниже, обеспечивает нас такой информацией. Дадим без доказательства следующую теорему:

Теорема 2.10 (Балинский). *Вершины, рассматриваемые как точки, и ребра, рассматриваемые как линии выпуклого многогранного множества S [встречающегося в задаче линейного программирования, в которой выпуклая оболочка вершин представляет собой многогранник S, не лежащий внутри никакой $(n-1)$-мерной гиперплоскости], образуют n-связный граф.*

Теорема 2.11 (Уитни). *Граф G является n-связным тогда и только тогда, когда существует n непересекающихся путей между любой парой точек* [88].

Доказательство. В нижеследующем доказательстве, данном Балинским, используются идеи теории потоков в сетях.

Введем сеть, обозначив произвольную пару точек через v_α и v_ω. Припишем пропускную способность 1 каждой вершине G, за исключением v_α и v_ω, и пропускную способность $(n+1)$ каждому ребру G, за исключением ребер, соединенных с v_α и v_ω. Если такие ребра существуют, припишем им пропускную способность 1. В полу-

ченной сети примем, что максимальный поток меньше n. Тогда минимальный разрез должен быть меньше n. Так как это противоречит тому, что G — n-связный граф, максимальный поток должен быть не менее n. Ни одна вершина не может иметь двух единичных путевых потоков, проходящих через нее. Поэтому существует n непересекающихся путей от v_α к v_ω. Так доказывается необходимость. Достаточность очевидна.

Следствие. *Существует не менее n непересекающихся путей между любой парой вершин многогранного выпуклого множества S.*

Замечание. Степень вершины равна числу ребер, инцидентных с ней. Дирак [18] доказал, что если каждая вершина графа с не более чем $2n$ вершинами имеет наименьшую степень $n > 1$, то граф имеет цикл, который проходит через каждую вершину в точности один раз.

Примеры задач о многогранниках и других экстремальных задач, в которых используются идеи теории графов

Определение. Диаметром многогранника является наибольшее число ребер на любом пути между любыми двумя вершинами.

Упражнение 2.24. Найдите диаметры пяти правильных многогранников с единичными ребрами в трехмерном пространстве.

Теорема 2.12. *Максимальный диаметр d трехмерных выпуклых многогранников с n вершинами задается формулой*

$$d = \left[\frac{n+1}{3}\right].$$

Доказательство [45]. По теореме Балинского, соответствующий граф является трехсвязным; по теореме Уитни, любая пара вершин связана тремя непересекающимися путями. Если расстояние между парой вершин равно диаметру d, то длина каждого из трех путей $\geqslant d$; следовательно, каждый путь имеет не менее чем $d - 1$ вершин, помимо данной пары. Поскольку все вершины различны, получаем

$$3(d-1) + 2 \leqslant n$$

или, так как d — целое число,

$$d \leqslant \left[\frac{n+1}{3}\right].$$

Равенство имеет место для треугольной призмы с четырехгранными шапками на треугольниках в верхнем и нижнем основаниях в случае $n \equiv 2 \bmod 3$. В других случаях одна или обе шапки могут быть удалены.

Определение. Наименьшее целое r, такое, что длина пути от некоторой вершины до любой другой вершины не превышает r, называется *радиусом многогранника.*

Гипотеза (Юкович и Мун). *Максимальный радиус трехмерных выпуклых многогранников с* $n \geqslant 6$ *вершинами больше или равен* $[(n+4)/4]$.

Мун и Мозер [58] изучили следующую задачу в k-мерном пространстве. Назовем максимальное число вершин на любом простом пути в многограннике, у которого не все из n ($n \geqslant k+1$) вершин лежат в $(k-1)$-мерном пространстве, *длиной пути*. Если $p(n, k)$ обозначает минимальную длину пути всех таких многогранников, то $p(n, 3) \leqslant (2n+13)/3$ [см. 8а].

В общем случае получаем [58].

Теорема 2.13.

$$p(n, k) < (2k+3)\{[1-2/(k+1)]n-(k-2)\}^{\log 2/\log k} - 1 < 3kn^{\log 2/\log k}.$$

Теорема 2.14 (Эрдёш). *Пусть* r *и* r' *— максимальное и минимальное расстояния, определенные* n *точками на плоскости. Тогда* r *может встречаться не более чем* n *раз, а* r' *— не более чем* $3n-6$ *раз* [21].

Доказательство. Если $P_1P_2 = r$ и $P_3P_4 = r$, то линии P_1P_2 и P_3P_4 должны пересекаться; в противном случае диаметр многоугольника $P_1P_2P_3P_4$ превышал бы r, что противоречит предположению о максимальности r. Соединим только те вершины, расстояние между которыми равно r. Надо рассмотреть два случая:

1. Если каждая точка связана не более чем с двумя другими точками, то существует не более чем $2n$ связанных упорядоченных пар; следовательно, число сегментов линии длины r не превосходит n.

2. Если точка P_1 связана с тремя точками P_2, P_3, P_4, причем отрезок P_1P_3 лежит между P_1P_2 и P_1P_4 (заметим, что угол $P_2P_1P_4 \leqslant \leqslant \pi/3$; в противном случае $P_2P_4 > r$, что невозможно, поскольку r — максимальное расстояние), то точка P_3 не может быть связана ни с какой другой точкой P_i, так как отрезок P_3P_i должен пересекать и P_1P_2 и P_1P_4. Если опустить P_3, то число точек и число максимальных расстояний уменьшается на единицу, и для завершения доказательства можно применить индукцию.

Рассмотрим теперь все точки, расстояния между которыми равны r'. Соединим только эти точки. Никакие из соединяющих сегментов не пересекаются в точке, которая не является одной из данных n точек; в противном случае некоторая пара из четырех точек, определяющих эти сегменты, была бы сближена больше, чем на r'. Получающийся граф является планарным, и, так как для такого графа $E \leqslant 3V-6$ (очевидно, что формула Эйлера также имеет место для плоской карты и поэтому соотношения, приведенные для многогранников, также выполняются), получаем искомый результат $3n-6$.

Теорема 2.15. *В трехмерном пространстве максимальное расстояние между* n *точками не может встречаться более чем* $2n-2$ *раз.*

Доказательство приводится в [33].

В k-мерном пространстве частота появления максимального расстояния между n точками имеет вид [22]

$$\frac{n^2}{2}\left(1-\frac{1}{[k/2]}\right)+O(n^{2-\varepsilon})$$

при некотором $\varepsilon > 0$.

Все вышеизложенное является частным случаем задачи Борсука (см. [32], стр. 418).

Представляется ли каждое ограниченное множество A в n-мерном евклидовом пространстве в виде $A \bigcup_{i=0}^{n} A_i$, где диаметр A_i меньше, чем диаметр A, $i = 0, 1, \ldots, n$? Диаметром множества является наибольшее расстояние (или супремум) между любыми двумя точками множества.

Замечание. Пусть $f(n)$ — наименьшее число различных расстояний между n произвольными точками на плоскости. Например, если $n = 3$, то для равностороннего треугольника $f(3) = 1$. Квадрат и правильный пятиугольник соответственно дают $f(4) = f(5) = 2$. В общем случае

$$f(n) > \frac{n^{2/3}}{2\sqrt[3]{9}} - 1.$$

Мозер [59] показал, что если точки являются вершинами выпуклого многоугольника, то

$$f(n) \geqslant \left[\frac{n+2}{3}\right].$$

Кратчайшие пути и некоторые задачи в теории графов

Обычная задача, которая встречается в теории графов, заключается в том, что требуется найти путь, состоящий из наименьшего числа дуг между любыми двумя точками. Она называется задачей *о кратчайшем пути*. Иногда может потребоваться отыскание кратчайшего простого цикла между двумя точками, когда ни одна из дуг или вершин на обратном пути не повторяет дуг или вершин на прямом пути.

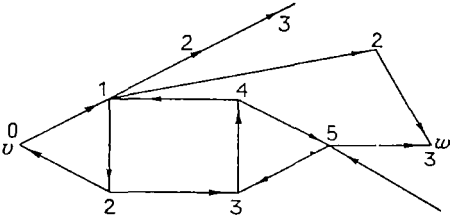
Более трудной задачей является задача отыскания кратчайшего пути между двумя точками, который проходит через каждую из заданного количества точек. Иногда требуется найти кратчайший цикл, проходящий через заданное число точек. Еще более трудной задачей является задача о коммивояжере, которая возникает, когда в каждой из предыдущих задач ребрам приписываются расстояния.

Упражнение 2.25. Сформулируйте задачу о коммивояжере с n городами в матричной форме и покажите, что существует $n!$ разных путей, каждый из которых проходит через n городов. Глобальный оптимум (т. е. кратчайший путь) является одним из них.

Дадим теперь два алгоритма отыскания кратчайших путей.

Алгоритм кратчайшего пути

Пусть v и w — две вершины графа G; требуется найти кратчайший путь между ними. Используем следующий алгоритм. *Шаг за шагом будем приписывать каждой вершине x графа G число m, равное кратчайшему расстоянию от v до x, следующим образом. На нулевом шаге припишем v нулевое расстояние. Если все вершины, которым приписано расстояние m, образуют известное множество $E(m)$ на $(m+1)$-м шаге, то припишем расстояние $(m+1)$ вершинам множества* $E(m+1) = \{x \mid x \notin E(k),\ k \leqslant m,$ *а x является конечной вершиной дуги, начальной вершиной которой является точка $y \in E(m)$}.*



Фиг. 2.11.

Процесс оканчивается после достижения w. Если $w \in E(m)$, можно проследить кратчайший путь от v к w, возвращаясь назад от w следующим образом. В качестве предпоследней вершины выбираем любую, предшествующую w (т. е. любую вершину, которая находится непосредственно перед w на любом пути, ведущем в w), которой приписано число $(m-1)$. В качестве предпредпоследней вершины выбираем любую предшествующую из тех, которым приписан вес $(m-2)$, и т. д.

Мы применили этот алгоритм к примеру, который иллюстрируется фиг. 2.11. Заметим, что вершине, которой приписано число 1, уже не приписывается другое расстояние 5, а вершине w, которой ранее приписано расстояние 3, не приписывается другое расстояние, в данном случае 6. Длина кратчайшего пути равна 3. Как проходит этот путь, очевидно.

Упражнение 2.26. В приведенном выше решении воспользуйтесь предыдущим алгоритмом, чтобы найти кратчайший путь от v к w после исключения последней дуги, ведущей к w.

Изложенный выше алгоритм непригоден для решения задачи, в которой задаются действительные длины (иногда они называются *пропускными способностями*) дуг. В этом случае отыскивается путь с наименьшей суммарной длиной независимо от числа дуг, которые его составляют. Однако решение предыдущей задачи будет получено, если каждой дуге приписана единичная длина.

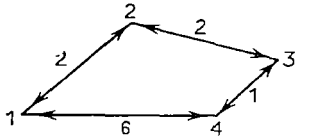
Следующий алгоритм предложил Флойд [24а]. *Требуется найти кратчайшие расстояния между всеми упорядоченными парами вершин одновременно.* Эта задача возникает при рассмотрении направленного графа. Если не существует дуги, соединяющей какую-то пару точек,

то фиктивной дуге, соединяющей их, приписывается бесконечная длина. При применении алгоритма к графу с n вершинами используется примерно n^3 сложений и сравнений. Этот алгоритм основан на идее, состоящей в том, что при рассмотрении расстояния вдоль пути, проходящего через некоторую вершину, ее можно исключить путем использования длины каждого пути, состоящего из дуги, ведущей от данной вершины к последующей. При вычислениях каждая упорядоченная пара вершин считается соединенной дугой, длина которой равна текущему значению элемента матрицы, так что для исключения вершины требуется $(n-1)^2$ операций. Заметим, что необходимо определять расстояния и до исключенной вершины.

Алгоритм Флойда для случая, когда дугам приписывается длина

Пусть m_{ij} — текущая длина кратчайшего пути из i в j, $(m_{ik} + m_{kj})$ — текущая длина пути из i в j, проходящего через k; если последняя величина меньше m_{ij}, то она используется в дальнейшем. Формально алгоритм Флойда состоит из следующих шагов:

1. Полагаем $i = 1$.
2. $(\forall j \neq i \ni m_{ji} + m_{ik} < m_{jk})$ $(\forall k \neq i)$, заменяем m_{jk} на $m_{ji} + m_{ik}$.
3. Увеличиваем i на 1.
4. Если $i \leqslant n$, переходим к шагу 2; в противном случае прекращаем процесс решения.



Фиг. 2.12.

Финальная (предельная) матрица дает длину кратчайшего пути между любыми двумя точками. Ведя учет изменений, вносимых для получения конечной формы, можно сохранить достаточно информации для того, чтобы проследить кратчайший путь между любыми двумя вершинами.

Алгоритм будет давать правильный ответ, если некоторые дуги имеют отрицательную длину, при условии, что в графе нет контуров отрицательной длины. При рассмотрении диагональных элементов финальной матрицы становится ясной ситуация, так как они дают длину кратчайшего контура для соответствующей вершины. Если отрицательные контуры не встречаются, то результаты являются приемлемыми.

Рассмотрим пример на фиг. 2.12, в котором вершины пронумерованы, а рядом с дугами проставлены длины каждой из симметричных пар дуг. Матрица примера $[m_{ij}]$ имеет вид

	1	2	3	4
1	∞	2	∞	6
2	2	∞	2	∞
3	∞	2	∞	1
4	6	∞	1	∞

Например, $m_{13} = \infty$. Эта величина заменяется на 11-м шаге алгоритма величиной $m_{12} + m_{23} = 4$. Аналогично $m_{14} = 6$. На 27-м шаге эта величина заменяется на $(m_{13} + m_{34}) = 5$ и т. д. Таким образом, кратчайший путь от 1 к 3 проходит от 1 к 2 и затем от 2 к 3. Кратчайший путь от 1 к 4 сначала проходит как кратчайший путь от 1 к 3 и затем от 3 к 4. Однако кратчайший путь от 1 к 3 уже известен. Следовательно, чтобы пройти от 1 к 4, нужно сначала пройти от 1 к 2, затем к 3 и, наконец, к 4 при помощи $m_{21} + m_{14} = 8$ на 3-м шаге. Однако эта величина снова заменяется на $m_{23} + m_{34} = 3$ на 25-м шаге.

Теорема 2.16. *Алгоритм Флойда дает матрицу кратчайшего пути, если граф не имеет контуров отрицательной длины.*

Доказательство (*Мерчлэнд*) [60а]. В бесконтурном пути между двумя вершинами ни одна вершина не повторяется. Если путь не является бесконтурным, то он содержит один или несколько контуров (или петель), но всегда существует вложенный путь без контуров. (Таких путей может быть несколько.) Так как ни один контур в графе не имеет отрицательную длину, среди всех путей, соединяющих две вершины, найдется по крайней мере один, который является одновременно и бесконтурным и кратчайшим.

Пусть основная операция, которая состоит в замене m_{ij} на $(m_{ik} + m_{kj})$, $m_{ik} + m_{kj} < m_{ij}$, называется *утроением* (i, k, j). Назовем k *промежуточной вершиной*.

Алгоритм, как указывалось ранее, содержит каждую возможную тройку в точности один раз; тройки, в которых промежуточная вершина является начальной или конечной, опускаются, так как они бесполезны.

Важным для доказательства является то обстоятельство, что эти $n\,(n-1)^2$ троек сгруппированы в n групп, где k-ю группу составляет *каждая полезная тройка с промежуточной вершиной* k.

Вычисления начинаются с того, что в матрице проставляются длины дуг. Легко видеть, что каждый элемент финальной матрицы соответствует реальному пути. В доказательстве следует показать, что ни одно расстояние не является слишком большим.

Каждое кратчайшее расстояние, соответствующее единственной дуге, сохраняется до конца вычислений. Первая группа троек с промежуточной вершиной 1 гарантирует, что кратчайшие расстояния, образованные путями из двух дуг с промежуточной вершиной 1, получены правильно.

Примем в качестве предположения математической индукции, что алгоритм позволяет найти кратчайшее расстояние между каждой парой вершин, между которыми имеется кратчайший путь, промежуточной вершиной которого с наибольшим номером является $(k-1)$. Ввиду выбранного расположения троек эти конечные значения расстояний должны быть получены до того, как будет достигнута группа k.

Пусть k — наивысшая вершина на кратчайшем бесконтурном пути от вершины v к вершине w или, если существует несколько таких путей, пусть k — наименьшая из наивысших промежуточных вершин на любом из этих путей. Тройка (v, k, w) встречается в k-й группе троек. Подпути от v к k и от k к w пути, который содержит k, имеют более низкие промежуточные вершины, чем k. По предположению индукции эти расстояния получаются правильно и, как отмечалось выше, до того, как достигается группа k. Следовательно, выполнение операции утроения (v, k, w) позволяет найти кратчайшее расстояние от v до w, так как кратчайшие расстояния от v до k и от k до w уже будут определены.

Следовательно, если предположение индукции справедливо для $(k - 1)$, то оно справедливо и для k. Поскольку известно, что оно справедливо для 1, можно заключить, что алгоритм позволяет правильно находить кратчайшие расстояния при любой наибольшей промежуточной вершине на кратчайшем пути.

Построение дерева минимальной полной длины

Легко представить себе задачи, в которых требуется построить пути между несколькими центрами, если существует один и только один путь, соединяющий любые два центра. Из всех таких возможных систем путей между центрами нужно найти тот, который имеет минимальную полную длину. Заметим, что необходимым условием того, чтобы дерево имело минимальную полную длину, является то, что длина каждой хорды должна быть не менее максимальной из длин ветвей фундаментального контура, который она определяет. В противном случае, используя эту хорду, можно сделать единственную замену. Оказывается, что необходимое условие является и достаточным, но доказать это сложно.

Чтобы выбрать дерево минимальной полной длины, сначала нумеруем ребра в порядке возрастания длин так, что длина e_i не превосходит длины e_j, если $i < j$. Затем начинаем с отбора e_1 и прибавляем e_2, если e_2 не образует контур с e_1. Продолжая рассматривать ребра последовательно в порядке возрастания номеров, отбираем ребро, если оно не образует контур с множеством ранее отобранных ребер, и отбрасываем его в противном случае. Этот процесс всегда приводит к построению дерева минимальной полной длины (см. доказательство в [50a]).

Другие примеры задач оптимизации в теории графов:

1. Требуется найти кратчайший простой контур (так называемый *гамильтонов контур*), проходящий через каждую вершину данного связного графа с n вершинами. Эта задача связана с задачей о коммивояжере, однако в ней ничего не говорится о длине ребер, а учитывается только их число. При рассмотрении этой задачи Гамильтон взял додекаэдр, воткнул иголки в его вершины и обвертывал нитку вокруг иголок, двигаясь от одной иголки к другой, чтобы получить гамильтонов контур.

2. *Задача Штайнера.* Пусть дано n точек на плоскости. Требуется найти кратчайший путь, соединяющий эти точки так, что из любой данной точки можно достичь любой другой точки, проходя через некоторые из точек, если это необходимо. Это может быть минимальное стягивающее дерево, если граф реально задан.

3. *Задача о китайском почтальоне (впервые изучена Куаном).* Требуется найти последовательность замкнутых ребер, которая включает каждое ребро связного графа по меньшей мере один раз и имеет минимальный полный вес. (Таким образом, задача состоит в том, чтобы удвоить некоторое количество ребер так, чтобы получился уникурсальный граф, добавляя минимальный вес в ходе этого процесса.) Граф является уникурсальным, если можно пройти вдоль всех его ребер, не повторив ни одно ребро [9].

Полные графы без треугольников

Определение. Полный граф на плоскости представляет собой множество из n вершин на плоскости, попарно связанных ребрами так, что два ребра пересекаются не более чем в одной точке и два ребра с общей вершиной не пересекаются.

Теорема 2.17. *Минимальное число ребер, которое можно удалить из полного графа на плоскости для того, чтобы оставшийся граф не имел контуров, состоящих из трех ребер, равно* $C_n^2 - A$, где

$$A = \textit{наибольшее целое число, не превосходящее} \begin{cases} \dfrac{n^2}{4}, & n \text{ четное} \\ \dfrac{n+1}{2}\,\dfrac{n-1}{2} = \dfrac{n^2-1}{4}, & n \text{ нечетное.} \end{cases}$$

Доказательство. Разбивая все вершины на два множества с одинаковым числом элементов (в случае нечетного n в одном из множеств на одну вершину больше) и связывая каждую вершину одного множества с каждой вершиной другого множества, получаем *двудольный* граф. Число ребер этого графа представляет собой наибольшую целую часть вышеприведенного выражения. Поэтому в обоих случаях число оставшихся ребер e_n не меньше этого числа. Чтобы доказать, что эти значения являются наилучшими среди всех возможных, воспользуемся методом математической индукции. Результат справедлив при $n = 2$. Для любого простого графа (т. е. для графа с не более чем одним ребром между парой вершин) с числом ребер $e_n \geqslant 1$, n вершинами (n четное) и без треугольников $e_n \leqslant n^2/4$. Чтобы убедиться в этом, разобьем все вершины на два множества, как указано выше, и рассмотрим пару вершин, по одной из каждого множества. Соединим эту пару ребром. Соединим также одну из них со всеми вершинами соответствующего ей множества, а другую со всеми вершинами другого множества. Ни одна вершина одного множества не связана с вершинами другого множества, так как тогда получа-

лись бы треугольники. Поэтому выбранные две вершины связаны с остающимися $(n-2)$ вершинами $(n-2)+1$ ребрами. Таким образом,

$$(n-2)+1+e_{n-2} \geqslant e_n.$$

Но по индукции левая часть равна

$$(n-1)+\frac{(n-2)^2}{4}=\frac{n^2}{4}.$$

Аналогичные рассуждения можно провести и в случае нечетного n. Этим завершается доказательство, так как C_n^2 является числом ребер в полном графе.

Максимальное число трехдуговых циклов в направленном графе [3, 9]

Снова рассмотрим полный граф с n вершинами на плоскости. Каждому ребру теперь можно приписать направление, которое указывается стрелкой. Направленное ребро называется *дугой*. При каждом задании направлений всем ребрам можно пересчитать простые циклы, которые образуют эти дуги. Такой цикл можно пройти, начиная с какой-то вершины и следуя в направлении дуги, идущей к другой вершине, затем аналогично переходя по дуге к третьей вершине и т. д. до тех пор, пока, наконец, дуга не приведет в исходную вершину. Если направления не заданы, циклы называются *контурами*. Очевидно, что по крайней мере один из методов задания направлений позволяет получить наибольшее число таких циклов. Чему равно это число? Рассмотрим здесь случай циклов, каждый из которых состоит из трех дуг.

Существует способ представления такой конфигурации (называемой *направленным графом*) при помощи матрицы *вершин*. Чтобы построить такую матрицу, обозначим вершины через $v_1, \ldots, v_n$ и запишем их по вертикали, каждую слева от строки матрицы; затем запишем их в том же порядке над столбцами. Элемент матрицы, стоящий на пересечении i-й строки и j-го столбца, равен 1, если существует дуга, направленная из вершины i в сторону вершины j. В противном случае элемент равен нулю. По главной диагонали все элементы равны нулю.

Теорема 2.18. *Максимальное число циклов, каждый из которых состоит из трех дуг, в полном графе с n вершинами при нечетном n равно*

$$C_n^3+\frac{1}{2}C_n^2-\frac{n}{2}\left(\frac{n-1}{2}\right)^2.$$

Доказательство. Рассмотрим матрицу вершин, соответствующую этому графу. В матрице i-я строка определяет отношение инцидентности для дуг, положительно инцидентных с i-й вершиной, т. е. исходящих из этой вершины, тогда как i-й столбец определяет отношение инцидентности для дуг, отрицательно инцидентных с вершиной,

т. е. входящих в эту вершину. Элемент такой матрицы a_{ij} равен единице, если существует дуга, направленная из вершины i к вершине j. В противном случае элемент равен нулю. Если обозначить через r_i сумму элементов в i-й строке, а через c_i — соответствующую сумму в i-м столбце, то $r_i + c_i = n - 1$, так как i-я вершина связана $n - 1$ ребрами с остальными $n - 1$ вершинами.

Полное число контуров с тремя ребрами равно C_n^3. Однако эта величина не совпадает с числом циклов. В цикле все ребра должны быть направлены в одну сторону. Поэтому, если две дуги положительно инцидентны с вершиной, они не могут входить в один цикл, потому что их направления взаимно противоположны. Обратно, каждый контур из трех дуг, который не является циклом, имеет в точности два ребра, положительно (отрицательно) инцидентных с одной и той же вершиной.

Так как сумма r_i элементов i-й строки дает число дуг, направленных из i-й вершины, следует исключить из полного числа контуров величину $\sum_{i=1}^{n} C_{r_i}^2$, т. е. сумму по всем строкам числа сочетаний по два из r_i. Это дает следующее выражение для числа циклов:

$$C_n^3 - \sum_{i=1}^{n} C_{r_i}^2 = C_n^3 - \frac{1}{2} \sum_{i=1}^{n} (r_i^2 - r_i).$$

Поскольку граф является полным, число его ребер равно C_n^2, и должно иметь место соотношение $\sum_{i=1}^{n} r_i = C_n^2$, потому что полная сумма по всем строкам должна учитывать все ребра графа.

Теперь получаем для числа циклов выражение

$$C_n^3 + \frac{1}{2} C_n^2 - \frac{1}{2} \sum_{i=1}^{n} r_i^2,$$

и задача состоит в том, чтобы определить r_i так, чтобы эта величина была максимальной. Выбор r_i соответствует полному графу со специальной ориентацией, которая максимизирует число циклов. Достаточно определить r_i так, чтобы $\sum_{i=1}^{n} r_i^2$ была минимальна, потому что эта величина вычитается из постоянной величины в вышеприведенном выражении, которое должно быть максимизировано. Предыдущее рассуждение было бы пригодно и в том случае, если бы мы использовали сумму элементов c_i столбца и тот факт, что две дуги с одинаковой ориентацией относительно вершины, с которой они инцидентны, не могут входить в один цикл. Тогда надо было бы найти c_i,

которые максимизируют

$$C_n^3 + \frac{1}{2} C_n^2 - \frac{1}{2} \sum_{i=1}^{n} c_i^2.$$

Таким образом, мы имели бы максимальное число циклов, если найдены c_i, которые минимизируют

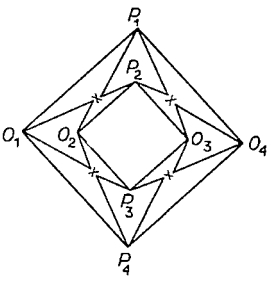
$$\sum_{i=1}^{n} c_i^2.$$

Итак, максимальное число циклов симметрично относительно r_i и c_i, т. е. они должны быть равны. Так как $r_i + c_i = n - 1$, при нечетном n получаем

$$r_i = (n - 1)/2.$$

Минимальное число пересечений на пути от пункта отправления к пункту назначения [9, 93]

На кирпичной фабрике имеется m печей, где обжигают кирпичи. Затем кирпичи грузят в маленький специальный вагон у каждой печи, направляющийся к любой из n платформ, к которым подходят грузовики. Так как каждая печь должна быть связана железной дорогой с каждой из погрузочных платформ, железнодорожные линии имеют очень много пересечений. Проходя через пересечения, вагоны часто сходят с рельсов, что приводит к падению кирпичей и транспортным заторам на фабрике. Задача состоит в том, чтобы проложить пути от печей к пунктам назначения с минимальным числом пересечений и тем самым минимизировать вероятность того, что вагон сойдет с рельсов.



Фиг. 2.13.

Эту задачу можно решить в рамках теории графов, где железнодорожным линиям соответствуют ребра графа, связывающие вершины, которые соответствуют печам, с другими вершинами, соответствующими погрузочным платформам. Накладывается одно условие, что никакие три ребра не могут пересекаться в одной и той же точке, если она не является вершиной. Однако два ребра могут пересекаться в промежуточной точке. Например, в случае, когда имеются четыре печи $O_1, \ldots, O_4$ и четыре платформы $P_1, \ldots, P_4$, получается четыре пересечения, которые помечены знаком $\times$ на фиг. 2.13.

Гипотеза (Царанкевич). *Минимальное число внутренних пересечений ребер, соединяющих каждую из p вершин с каждой из q вершин двудольного графа $C_{p,q}$ на плоскости (предполагается, что два ребра*

пересекаются не более чем в одной точке), больше либо равно

$$I(C_{p,q}) = \begin{cases} (r^2 - r)(s^2 - s), & \text{если } p = 2r, \quad q = 2s, \\ (r^2 - r)s^2, & \text{если } p = 2r, \quad q = 2s + 1, \\ r^2(s^2 - s), & \text{если } p = 2r + 1, \quad q = 2s, \\ r^2 s^2 & \text{если } p = 2r + 1, \quad q = 2s + 1, \end{cases}$$

или просто

$$I(C_{p,q}) = \left[\frac{p}{2}\right]\left[\frac{p-1}{2}\right]\left[\frac{q}{2}\right]\left[\frac{q-1}{2}\right],$$

где $[x]$ *означает наибольшее целое число, не превосходящее* x.

Замечание. Двудольный граф $C_{p,q}$ называется полным, если все p вершин связаны со всеми q вершинами.

В 1954 г. Царанкевич [93] дал доказательство вышеизложенного утверждения (оно приведено также в [9]), в котором Пауль Кайнен в 1964 г. нашел ошибку. Царанкевич также предложил следующую общую схему реализации для получения предполагаемого числа пересечений.

Построение схемы с минимальным числом пересечений производится следующим образом: рассмотрим систему декартовых координат на плоскости. Если $m = 2r$, выберем на оси x точки с абсциссами

$$-r, -(r-1), \ldots, -2, -1, 1, 2, \ldots, r,$$

а если $m = 2r + 1$, выберем на этой оси точки с абсциссами

$$-r, -(r-1), \ldots, -2, -1, 1, 2, \ldots, r, (r+1).$$

Если $n = 2s$, выберем на оси y точки с ординатами

$$-s, -(s-1), \ldots, -2, -1, 1, 2, \ldots, s,$$

а при $n = 2s + 1$ выберем точки с ординатами

$$-s, -(s-1), -2, -1, 1, 2, \ldots, s, (s+1)$$

и затем соединим прямолинейными отрезками все точки на оси x со всеми точками на оси y. В этом случае легко можно подсчитать все пересечения.

Минимальное число пересечений [100*]

С предыдущим материалом связана следующая давно сформулированная гипотеза:

Гипотеза. *Минимальное число пересечений* I_n *ребер полного графа* C_n *с* n *вершинами, изображенного на плоскости, задается формулой*

$$I_n = \begin{cases} \dfrac{n(n-2)^2(n-4)}{64}, & \text{если } n \text{ четное}, \\ \dfrac{(n-1)^2(n-3)^2}{64}, & \text{если } n \text{ нечетное}, \end{cases}$$

или просто

$$I_n = \frac{1}{4}\left[\frac{n}{2}\right]\left[\frac{n-1}{2}\right]\left[\frac{n-2}{2}\right]\left[\frac{n-3}{2}\right].$$

Докажем сначала эту гипотезу для случаев $n = 1, \ldots, 10$. Затем сформулируем общий принцип симметрии, обобщающий представление случаев малых n, из которого (при условии, что он доказан) будет следовать гипотеза [70].

Рассмотрим каждую вершину графа C_n и $n-1$ ребер, которые соединяют ее с остальными вершинами (т. е. звезду этой вершины). Удаление этой вершины и ее звезды приводит к исключению всех пересечений, приходящихся на эти ребра. В результате получается полный граф с $n-1$ вершинами. Звезда любой другой вершины полного графа с n вершинами имеет такое же или отличное число пересечений. В общем случае, если x_i, $i = 1, \ldots, n$, представляет собой число пересечений, приходящихся на ребра i-й вершины, то полное число пересечений в графе C_n равно $1/4 \sum_{i=1}^{n} x_i$, так как каждое пересечение учитывается четыре раза, по одному разу при рассмотрении каждой из четырех вершин для двух ребер, определяющих пересечение. Таким образом, получаем следующую теорему:

Теорема 2.19. *Если x_i, $i = 1, \ldots, n$,— число пересечений, приходящихся на звезду i-й вершины в полном графе C_n, то полное число пересечений в графе задается формулой* $^1/_4 \sum_{i=1}^{n} x_i$.

Теорема 2.20. *Минимальное число пересечений I_n в полном графе C_n необходимо удовлетворяет соотношениям*

$$I_n \geqslant \frac{n}{n-4} I_{n-1}, \quad n \geqslant 5,$$

$$I_n = 0, \quad n < 5.$$

Доказательство. Среднее число пересечений, приходящихся на одну вершину, в I_n равно $4I_n/n$. После удаления вершины с числом пересечений, не меньшим среднего значения, остается C_{n-1} со средним числом пересечений не менее чем I_{n-1}. Таким образом,

$$I_n - \frac{4I_n}{n} \geqslant I_{n-1},$$

откуда следует искомое соотношение.

Повторяя эту процедуру, получаем

$$I_n \geqslant \frac{n(n-1)(n-2)\ldots(n-k+1)}{(n-4)(n-5)(n-6)\ldots(n-k-3)} I_{n-k},$$

откуда можно найти нижнюю границу для I_n, если известно минимальное значение I_{n-k}. Приведенное выше выражение можно переписать в виде [34]

$$I_n \geqslant I_t \frac{C_n^t}{C_{n-4}^{t-4}}, \quad 4 < t \leqslant n.$$

Рассуждая по индукции, можно показать, что гипотетическая формула для I_n справедлива в случае четного n при условии, что она выполняется для непосредственно предшествующего нечетного значения n. Это следует из того, что

$$I_k \geqslant \frac{k}{k-4} I_{k-1} = \frac{k}{k-4} \frac{(k-2)^2 (k-4)^2}{64} = \frac{k (k-2)^2 (k-4)}{64}.$$

Однако при нечетном k получаем

$$I_k \geqslant \frac{k}{k-4} \frac{(k-1) (k-3)^2 (k-5)}{64} = \frac{(k-1) (k-3)^2}{64} \frac{k (k-5)}{k-4} =$$
$$= \frac{(k-1) (k-3)^2}{64} \left[(k-1) - \frac{4}{k-4} \right] = \frac{(k-1)^2 (k-3)^2}{64} - 4 \frac{k-1}{k-4} \frac{(k-3)^2}{64},$$

и индукция неприменима для перехода от четного случая к нечетному. На фиг. 2.14 показано, что $I_4 = 0$.

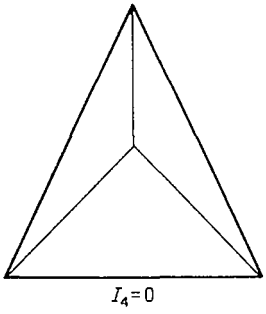
Теорема 2.21. $I_5 = 1$.

Доказательство. См. [51] или [9] и фиг. 2.15.

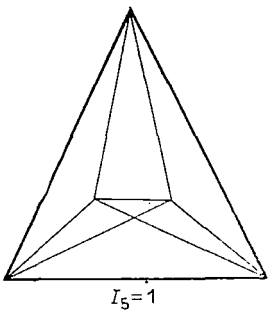
Теорема 2.22. $I_6 = 3$.

Доказательство. Теоремы 2.20 и 2.21 дают $I_6 \geqslant 3$. Однако на фиг. 2.16 показано, что $I_6 = 3$ реализуемо.

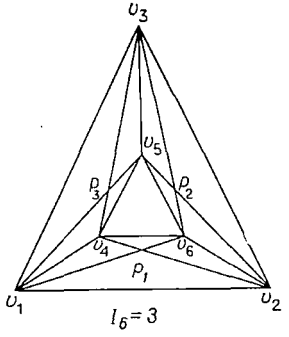
Изоморфизмом между двумя графами называется взаимно однозначное соответствие между их вершинами и ребрами, при котором



Фиг. 2.14.



Фиг. 2.15.



Фиг. 2.16.

сохраняется инцидентность соответствующих ребер с соответствующими вершинами.

Теорема 2.23. *Каждый полный граф с шестью вершинами и минимальным числом пересечений изоморфен графу, изображенному на фиг.* 2.16, *если его пересечения рассматривать как вершины.*

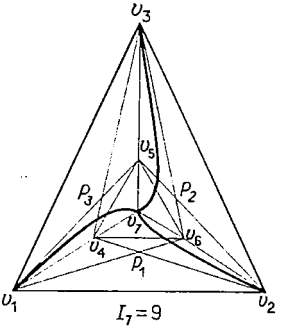
Доказательство. См. [69].

При *стереографическом проектировании* граф сначала отображается на сферу, южный полюс которой находится на внутренней области, представляющей собой часть плоскости, а северный полюс является начальной точкой лучей, которые проходят через сферу и каждой точке сферы ставят в соответствие точку на плоскости. Сфера, на которую отображен граф, теперь переворачивается так, что северный полюс касается плоскости и граф отображается со сферы на плоскость, причем в качестве источника лучей используется южный полюс. Таким образом, определенную внутреннюю область можно преобразовать во внешнюю область.

Теорема 2.24. $I_7 = 9$.

Доказательство.

Применяя теорему 2.19 последовательно к $I_7 = 3, 4, 5, 6$, каждый раз получаем вершину с таким числом пересечений, определенным звездой вершины, что после ее удаления вместе со звездой остается граф C_6 с менее чем тремя пересечениями, что противоречит теореме 2.22.



Фиг. 2.17.

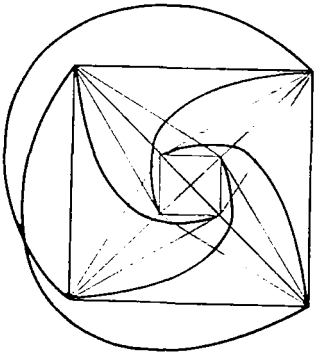
Докажем теперь, что $I_7 \neq 8$ (случай $I_7 = 7$ доказывается аналогично). По теореме 2.19 имеем $(8 \times 4)/7 > 4$. Поэтому существует по меньшей мере одна вершина с пятью пересечениями. Заметим, что наличие вершины с шестью или бо́льшим числом пересечений противоречило бы теореме 2.22. Единственная возможность здесь состоит в том, чтобы взять, например, $x_i = 5$, $i = 1, \ldots, 4$, и $x_i = 4$, $i = 5, \ldots, 7$. Во всяком случае, после удаления вершины с пятью пересечениями остается подграф C_6, который является минимальным. Используя фиг. 2.16, находим, что седьмая вершина может попасть в любую из трех областей, связанных с остальной частью графа. Этими областями являются: 1) внутренняя область внутреннего треугольника $v_4v_5v_6$ (внешняя часть внешнего треугольника $v_1v_2v_3$ симметрична этой области при стереографическом проектировании); 2) внутренняя область треугольника, подобного $v_2v_6p_2$; 3) внутренняя область треугольника, подобного $v_5v_6p_2$. Во всех трех случаях ребра, соединяющие седьмую вершину с любой из остальных вершин, должны иметь по меньшей мере шесть дополнительных пересечений, что противоречит тому факту, что при введении вершины в граф по предположению должно прибавляться в точности пять пересечений. Таким образом, $I_7 \geqslant 9$. Однако на фиг. 2.17 показано, что граф с $I_7 = 9$ реализуем. Такое представление не является единственным.

Теорема 2.25. $I_8 = 18$.

Доказательство. Из теорем 2.20 и 2.21 имеем $I_8 \geqslant 18$. Однако на фиг. 2.18 показано, что граф с $I_8 = 18$ реализуем. Такое представление не является единственным.

Теорема 2.26. $I_9 = 36$.

Доказательство. По теореме 2.20 $I_9 \geqslant 33$. Предположим, что $I_9 = 35$. При помощи метода, использованного в теореме 2.24, можно показать, что граф C_9 с 33, 34 и 35 пересечениями должен содержать минимальные числа пересечений графов C_7, C_5, C_3, что противоречит следующей лемме, доказанной в [71].



$I_8 = 18$

Фиг. 2.18.

Лемма. Если минимальный граф C_n содержит минимальный подграф C_{n-k} с $k = 2, 4, 6$, то I_n относится к графам, о которых говорилось в гипотезе, приведенной на стр. 110. Доказательство для случаев $I_9 = 33, 34$ проводится аналогично. Так как на фиг. 2.19 показано, что $x = 36$ реализуемо, $I_9 = 36$. Такое представление не является единственным.

Теорема 2.27. $I_{10} = 60$.

Доказательство. По теоремам 2.20 и 2.26 $I_{10} \geqslant 60$. Однако на фиг. 2.20 показано, что граф с $I_{10} = 60$ реализуем. Такое представление не является единственным.

Задачу о пересечениях в случае графа C_n алгебраически можно сформулировать следующим образом. Требуется найти неотрицательные целые числа x_i, $i = 1, \ldots, n$, такие, что существует геометрическая реализация (т. е. возможность начертить на плоскости) графа C_n с x_i пересечениями, которые соответствуют i-й вершине, и что эти x_i доставляют минимальное (целочисленное) значение

$$\frac{1}{4}\sum_{i=1}^{n} x_i.$$

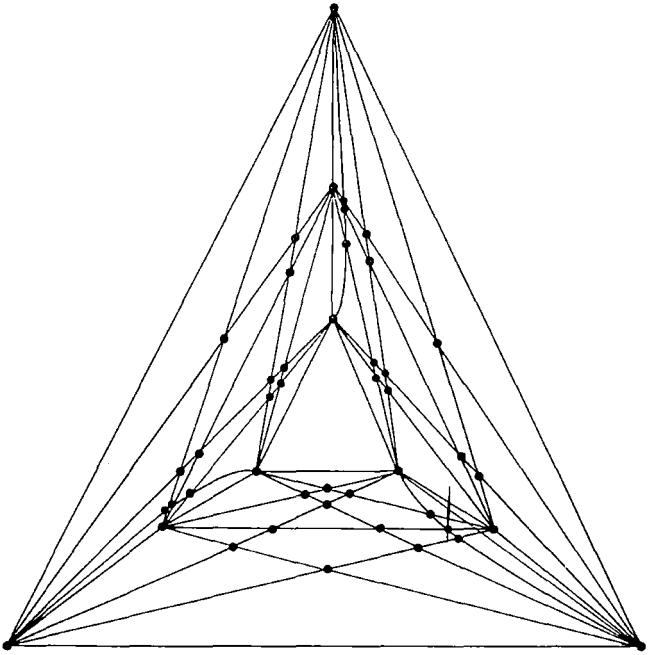
Заметим, что этот реализуемый минимум I_n удовлетворяет условию

$$I_n \geqslant \frac{n}{n-4} I_{n-1},$$

которое дано в теореме 2.20.

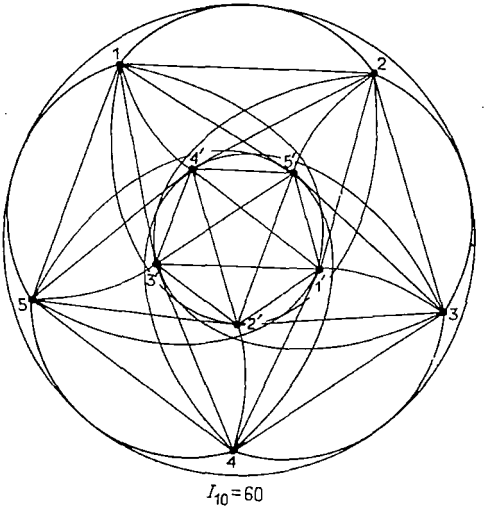
Теорема 2.28. *Максимальное число пересечений в графе C_n получается, если расположить его вершины на многоугольнике и соединить их ребрами, проходящими во внутренней части многоугольника.*

Доказательство. Поскольку каждое пересечение определяется четырьмя вершинами, максимально возможное число пересечений равно C_n^4 и ребра, внутренние по отношению к многоугольнику,



$I_9 = 36$

Фиг. 2.19.



$I_{10} = 60$

Фиг. 2.20.

дают в точности это число пересечений, так как диагонали каждого четырехугольника пересекаются во внутренней части многоугольника.

Можно доказать следующие теоремы, первая из которых является следствием теоремы 2.28:

Теорема 2.29. *В представлении графа C_n с минимальным числом пересечений $x_i > 0$, $i = 1, \ldots, n$.*

Теорема 2.30. *Полный граф с максимальным числом пересечений имеет такую реализацию, что $x_i = x_j$ при всех i и j.*

Доказательство. Из представления в виде многоугольника получаем

$$x_i = \frac{1}{4} C_n^4, \quad i = 1, \ldots, n.$$

Комбинаторная симметрия

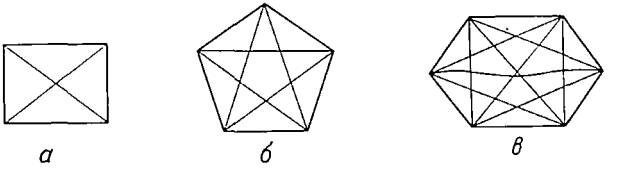
В геометрии симметрия определяется при помощи изометрий, которые называются *операциями симметрии* (и их групп автоморфизмов), оставляющих фигуры неизменными, или инвариантными, хотя их отдельные части перемещаются. Растяжения, переносы и повороты фигуры и ее частей играют важную роль при описании ее симметрии.

В этом разделе будут рассмотрены некоторые методы решения задач, в качестве основополагающего общего принципа которых предполагается существование более широкого понятия симметрии. Эти идеи еще далеки от точной формулировки и излагаются здесь, чтобы стимулировать теоретическое развитие в этой области.

Общая симметрия определяется правилами построения, применяемыми для соответствующего разбиения элементов множества на непересекающиеся подмножества, и полугруппой отображений между членами разбиения. Данное комбинаторное свойство сохраняется при отображениях, образующих полугруппу. Таким образом, операции симметрии не изменяют комбинаторное свойство, тогда как члены разбиения множества заменяются поэлементно. Напомним, что полугруппой называется множество элементов, замкнутое относительно некоторой операции, ассоциативное и содержащее единичный элемент.

В качестве примера можно рассмотреть представление графа C_n с максимальным числом пересечений. Здесь множество разбивается на отдельные элементы (тривиальное разбиение), каждый член представляет собой вершину. Представление инвариантно относительно преобразования, которое переводит вершину и ее звезду в другую вершину и ее звезду, сохраняя порядок ребер, в котором они располагаются в звезде. Вообще можно рассматривать любое разбиение.

На фиг. 2.21 представлены графы C_k, $k = 4, 5, 6$, с максимальным числом пересечений; после удаления любой вершины и ее звезды



Фиг. 2.21.

из C_n остается C_{n-1} с максимальным числом пересечений. В данном случае каждый граф C_n содержит все графы более низкого порядка C_k, $k = 1, 2, \ldots, n$. Число графов порядка k равно C_n^k.

Принцип регенерации в комбинаторной симметрии

Отметим, что максимальность сохраняется, если в фигуру с меньшим числом вершин ввести дополнительные вершины. Поэтому если в представлении фигуры, которая является расширением комбинаторно симметричной фигуры с меньшим числом элементов, обладающей определенным комбинаторным свойством (например, максимальностью числа пересечений), симметрия сохраняется, то это свойство сохраняется и для большего множества. Процесс расширения можно рассматривать также как суперпозицию фигуры с бо́льшим числом элементов на соответствующую фигуру с меньшим числом элементов.

Таким образом, бо́льшая фигура является расширением меньшей фигуры с сохранением симметрии. Меньшая фигура затем может быть получена из большей путем удаления соответствующих элементов, из которых в свою очередь можно получить еще меньшие фигуры и т. д. Таким образом, получается убывающая последовательность фигур, каждая из которых сохраняет соответствующее комбинаторное свойство. Процесс расширения с сохранением симметрии и соответствующих свойств называется *регенерацией*.

Возможность симметричной регенерации подсказывает следующая гипотеза:

Гипотеза. *Если известно, что данное комбинаторное свойство выполняется в симметричном представлении графа наименьшего порядка, использованном для разбиения множества, а также в соответствующем подмножестве семейства его непосредственных потомков (чтобы быть уверенным в том, что правила симметричного построения позволяют последовательное расширение исходного подмножества с сохранением комбинаторного свойства), то оно выполняется для произвольного потомка.*

Возможно, наиболее существенный критерий разработки правил симметризации для построения минимального или максимального графа C_n основан на следующем определении:

Определение. Симметричной реализацией полного графа является симметричный рисунок на плоскости, для которого $\sum_{i,\,j=1}^{n} (x_i - x_j)^2 = \min$.

Возвращаясь к задаче о минимальном C_n, предположим, что n четно. Симметризация при переходе от C_{n-2} к C_n путем введения двух вершин должна удовлетворять критерию симметрии, который уже использовался для построения симметричного графа C_{n-2}. Можно сказать, что граф C_n, так же как и граф C_{n-2}, симметричен, если его вершины можно разбить на несвязанные пары с сохранением взаимных соотношений между парами.

Пары являются взаимно заменяемыми в соответствии с числом пересечений, приходящихся на их ребра, и с правилом или схемой изображения этих ребер. Таким образом, пара вершин с ребрами, соединяющими их с остальными вершинами C_{n-2} (двойная звезда), образует конфигурацию, идентичную конфигурации, которую можно образовать из любых других $n/2 - 1$ пар. Любое ребро в двойной звезде пары имеет такое же число пересечений в соответствии с правилами симметричного построения, как и соответствующее ребро в любой другой паре.

Удаление одной вершины любой пары вместе с ее звездой из графа C_n (n четное) должно давать искомый граф C_{n-1} с предполагаемым числом пересечений. Чтобы получить C_{n-2}, следует удалить другую вершину пары.

Построение симметричного графа C_n с предполагаемой величиной I_n, который удовлетворяет требованиям к парам вершин, подробно описано в доказательстве теоремы 1, приведенном в работе [71]. Например, для I_{10} пары определяются так, как на фиг. 2.20. Одна вершина принадлежит внешнему многоугольнику, а парная с ней вершина представляет собой самую удаленную вершину (или одну из самых удаленных, если имеет место совпадение во внутреннем многоугольнике.

После удаления любой пары из графа C_n, естественно, остается C_{n-2}. Вслед за дополнением пар к графу C_n более низкого порядка следует использовать правила симметрии. Принцип регенерации влечет за собой следующую гипотезу:

Гипотеза. *Если граф C_n минимальный, то можно выбрать $r = \lfloor n/2 \rfloor$ пар вершин $P_1, \ldots, P_r$ из C_n, таких, что если C_{2s} — полный подграф C_n, определенный вершинами $P_1, \ldots, P_s, 1 \leqslant s \leqslant r - 1$,* то C_{2s} — *минимальный граф. Кроме того,* если n нечетно, то в графе с вершинами $V, P_1, \ldots, P_r$ любой граф C_7, определенный вершинами V, P_i, P_j, P_k, *минимальный.*

Задача о минимальном числе пересечений в графе C_n является частным случаем следующей общей задачи. Рассмотрим связный граф G с n вершинами, соединенными ребрами на плоскости следующим образом: каждая вершина соединена с k другими вершинами ($2 \leqslant k \leqslant n-1$) так, что максимальное число вершин имеет степень k. Определим минимальное число пересечений $I_n(k)$ ребер в G для случая, когда любые два ребра могут пересекаться самое большее один раз в точке, отличной от вершины.

Решение этой задачи должно следовать из наших предшествующих рассуждений о симметричной реализации. Исходную реализацию можно использовать для того, чтобы удалить необходимое число ребер. Очевидно, например, что при удалении ребер, соединяющих заданные пары, в случае четного n уменьшается наибольшее число пересечений. В случае нечетного n должна быть одна вершина, ни одно ребро которой не удаляется; в противном случае у другой вершины пришлось бы удалять $(n-2)$ ребер и т. д. Это дает

$$I_n(n-2) = \begin{cases} \dfrac{n(n-2)(n-4)(n-6)}{64} & \text{при } n \text{ четном,} \\ \dfrac{(n-1)(n-3)^2(n-5)}{64} & \text{при } n \text{ нечетном,} \end{cases}$$

$$I_n(n-3) = \begin{cases} \dfrac{n(n-4)(n-6)^2}{64} & \text{при } n \text{ четном,} \\ \dfrac{(n-3)^2(n-5)^2}{64} & \text{при } n \text{ нечетном,} \end{cases}$$

. . . . . .

$$I_n(3) = 0,$$
$$I_n(2) = 0,$$
$$I_n(1) = 0,$$
$$I_n(0) = 0.$$

Индукция по k может оказаться полезной при переходе от случая меньших n к бо́льшим.

Замечание. Если имеет место принцип регенерации, то доказательство задачи Царанкевича может быть получено при помощи его симметричного построения. Кайнен [46] доказал следующую теорему:

Теорема 2.30а. *Если гипотеза Царанкевича справедлива, то* $I_n \sim n^4/64$ *при* $n \to \infty$.

Доказательство. Запишем $S_n = n^{-4} I_n$ и покажем, что $\lim\limits_{n\to\infty} S_n = {}^1/_{64}$. Заметим, что при фиксированном p полный граф C_n можно разложить на полный двудольный граф $C_{p,\,n-p}$, в котором остальные ребра исключены. Обозначим множества вершин $C_{p,\,n-p}$ через V_p и V_{n-p}. Пусть x — пересечение на рисунке D графа C_n. Если оно принадлежит разложению $D_{p,\,n-p}$ рисунка D, соответствующего

разложению $C_{p,\,n-p}$ графа C_n, то в точности две из четырех вершин, связанных с x (т. е. четыре конечные точки двух ребер, пересекающихся в x), принадлежат V_p; это можно реализовать в точности четырьмя разными способами. (Напомним, что если обе конечные точки ребра в D принадлежат V_p или V_{n-p}, то ребро не принадлежит $D_{p,\,n-p}$.) Можно выбрать максимальную систему из m множеств вершин в D, состоящих из p элементов, обладающую тем свойством, что никакие два множества в m не имеют общего подмножества из двух элементов. Обозначим через $\overline{m} = \overline{m}(n, p)$ число элементов в m. Таким образом, можно найти систему $D_1, \ldots, D_{\overline{m}}$ подмножеств D, такую, что каждое D_j представляет собой рисунок $C_{p,\,n-p}$ и каждое пересечение x в D принадлежит самое большее четырем D_j. Поэтому число пересечений в D не может превосходить $d_1 + \ldots + d_{\overline{m}}$, где d_j — число пересечений в D_j.

По гипотезе Царанкевича получаем

$$I(C_{p,\,n-p}) = \left[\frac{p}{2}\right]\left[\frac{p-1}{2}\right]\left[\frac{n-p}{2}\right]\left[\frac{n-p-1}{2}\right].$$

Поскольку каждое D_j содержит по меньшей мере $I(C_{p,\,n-p})$ пересечений, D содержит не менее чем

$$\frac{1}{4}\overline{m}I(C_{p,\,n-p})$$

пересечений ребер. Но D и p выбирались произвольно, так что получаем

$$I_n \geqslant \frac{1}{4}\overline{m}I(C_{p,\,n-p})$$

при любом $1 < p < n$.

Эрдёш и Ханани [22a] показали, что

$$\lim_{n\to\infty} \overline{m}(n, p)(C_n^2)^{-1} C_p^2 = 1$$

при любом $2 \leqslant p \leqslant n$. Используя эти соотношения, для любого фиксированного целого $p > 1$ получаем

$$\lim_{n\to\infty} S_n \geqslant \lim_{n\to\infty} \frac{1}{4}\frac{n(n-1)}{2}\frac{2}{p(p-1)}\left[\frac{p}{2}\right]\left[\frac{p-1}{2}\right]\left[\frac{n-p}{2}\right]\left[\frac{n-p-1}{2}\right] n^{-4} =$$

$$= \lim_{n\to\infty} \frac{1}{16}\left(\frac{2}{p}\right)\left[\frac{p}{2}\right]\left(\frac{2}{p-1}\right)\left[\frac{p-1}{2}\right]\left(\frac{1}{n}\right)\left[\frac{n-p}{2}\right]\left(\frac{1}{n}\right)\times$$

$$\times\left[\frac{n-p-1}{2}\right] = \frac{1}{64}q(p),$$

где

$$q(p) = \left(\frac{2}{p}\right)\left[\frac{p}{2}\right]\left(\frac{2}{p-1}\right)\left[\frac{p-1}{2}\right].$$

Очевидно, что $q(p) \leqslant 1$, и легко видеть, что при данном $\varepsilon > 0$ можно выбрать p достаточно большим, так что $q(p) \geqslant 1 - \varepsilon$. Поэтому

$$\lim_{n\to\infty} S_n \geqslant \frac{1}{64},$$

и это вместе с уже установленным фактом, что предполагаемое значение I_n реализуемо и может быть использовано в качестве верхней границы, дающей $\lim_{n\to\infty} S_n \leqslant {}^1/_{64}$, приводит к доказательству теоремы.

2.7. Покрытие шахматной доски [62, 72, 91]

Задача. Какое наибольшее число ладей можно разместить на шахматной доске 8×8 так, чтобы никакие две ладьи не били одна другую? Решите эту задачу для ферзей, коней и слонов. Обобщите ее на случай доски $n \times n$.

Решение для доски 8×8. Заметим, что, если на доске находятся более чем восемь ладей или ферзей, по крайней мере две из этих фигур должны находиться на одной горизонтали или вертикали и, следовательно, они будут бить друг друга. Максимальное число фигур в каждом случае равно восьми. Например, ладьи можно разместить на восьми квадратах любой из двух центральных диагоналей. Всего возможно 8! размещений ладей, которые удовлетворяют указанному выше требованию. Задачу о размещении восьми ферзей на доске мы оставляем в качестве элементарного упражнения.

Что касается коней, заметим, что так как конь ходит с белого квадрата на черный, то, если все 32 коня размещены на белых квадратах, ни один из них не будет бить другого. С другой стороны, каждое черное поле может быть бито по меньшей мере двумя конями. Чтобы показать, что 32 — максимальное число, рассмотрим шахматную доску 2×4. На ней конь, стоящий на любом поле, бьет в точности одно другое поле и, следовательно, можно расположить не более четырех коней так, чтобы они не били друг друга. Шахматную доску можно покрыть восемью такими досками 2×4, что дает максимум 32 коня. Поскольку на маленькой доске могут мирно сосуществовать не более четырех коней, нельзя найти лучшего расположения коней на большой доске 8×8, так как ограничение на число коней на доске 2×4 всегда должно выполняться.

Принцип максимума

Если известно, что стандартный элемент обладает свойством максимальности, т. е. используя его точные копии, можно покрыть больший блок без перекрытий, то это можно использовать для получения верхней границы размеров большего элемента, обладающего таким же свойством максимальности.

Займемся теперь расположением на шахматной доске слонов. Заметим, что на доске имеются 15 черных (белых) параллельных диагоналей, но на них можно разместить только 14 слонов, потому что слоны, стоящие на противоположных угловых полях, угрожают друг другу. Эти 14 слонов можно разместить на первой горизонтали и на неугловых полях верхней горизонтали.

Упражнение 2.27. Покажите, что максимальное число слонов равно 14.

Упражнение 2.28. При помощи аналогичных рассуждений покажите, что на доске $n \times n$ максимальное число коней равно

$$\frac{n^2}{2}, \qquad n \text{ четное},$$

$$\frac{n^2+1}{2}, \qquad n \text{ нечетное}.$$

Покажите также, что максимальное число ладей равно n, а максимальное число слонов равно $2n - 2$. Что касается ферзей, то при помощи сложных рассуждений можно показать, что максимальное число их равно n. Одно из доказательств этого приводится в [42a], в нем используется понятие максимальных внутренне устойчивых множеств из теории графов.

Упражнение 2.29. Покажите, что максимальное число королей равняется $[(n + 1)/2]^2$ при n нечетном и $(n/2)^2$ при n четном.

Упражнение 2.30. Покажите, что минимальное число королей, которое можно разместить на шахматной доске $n \times n$ так, что каждое поле будет находиться под контролем хотя бы одного короля, равно

$$\frac{n^2}{9}, \qquad \text{если } n \equiv 0 \bmod 3,$$

$$\frac{(n+2)^2}{9}, \qquad \text{если } n \equiv 1 \bmod 3,$$

$$\frac{(n+1)^2}{9}, \qquad \text{если } n \equiv 2 \bmod 3.$$

Покажите, что ответ в такой же задаче, поставленной для ладей, равен n (существует $2n^n - n!$ таких расположений) и что для слонов это число тоже равно n.

Алгебраическая постановка. Предыдущие задачи о максимумах представляют собой частные случаи следующей постановки [72]. Требуется найти

$$\max \sum_{i=1}^{n} x_i, \quad x_i = 0 \text{ или } 1, \; i = 1, \ldots, n,$$

при ограничениях

$$x_i \sum_{j=1}^{n} a_{ij} x_j = 0, \quad i = 1, \ldots, n,$$

где $a_{ij} = 0$ или 1, $i, j = 1, \ldots, n$,— элементы симметрической матрицы, которая называется *матрицей угроз*, в каждой из строк которой цифрами «1» указываются позиции, которым фигура угрожает с данного квадрата.

Замечание. Следующую задачу поставил Хейлброн. Можно ли разместить $2n$ ферзей на шахматной доске $n \times n$ так, чтобы никакие три из них не стояли на одной линии? Несмотря на значительные усилия, эта задача не была решена. Согласно теореме Эрдёша, если p — простое число, то на доске $p \times p$ можно выбрать p полей так, что никакие три из них не будут лежать на одной линии.

Задача о покрытии шахматной доски пластинками домино. Алгебраическая формулировка [47a]. Дадим теперь алгебраическое решение задачи, которая уже была поставлена и решена ранее (см. предисловие): чему равно максимальное число пластинок домино, необходимое для того, чтобы покрыть без наложения как можно больше квадратов шахматной доски, в которой удалены правый верхний и левый нижний квадраты? Каждая пластинка домино покрывает два смежных квадрата.

Покажем сначала, что этого нельзя сделать с помощью 31 пластинки. Сопоставим каждому квадрату шахматной доски одночлен так, как указано в табл. 2.3. Таким образом, квадрату i, j сопоставим $x^i y^j$.

Таблица 2.3

7	y^7	xy^7	.	.	.	.	.	x^7y^7
6	.							.
5	.							.
4	.							.
3	.			.				.
2	y^2	xy^2	.	.	.	.	.	x^7y^2
1	y	xy	x^2y	x^3y	x^4y	x^5y	x^6y	x^7y
0	1	x	x^2	x^3	x^4	x^5	x^6	x^7
	0	1	2	3	4	5	6	7

Если перемножить $(1 + x + x^2 + \ldots + x^7)(1 + y + y^2 + \ldots + y^7)$, то получим все позиции на шахматной доске. Алгебраически исключение левого нижнего квадрата и правого верхнего дает

$$(1 + x + \ldots + x^7)(1 + y + \ldots + y^7) - 1 - x^7y^7.$$

Пластинки домино можно класть горизонтально или вертикально. Например, если положить пластинку домино в левый нижний

угол, то получим покрытие $(1 + x)$, если же положить ее вертикально в этот же угол, то получим покрытие $(1 + y)$. В общем случае пластинка занимает два смежных по горизонтали квадрата $x^a y^b$ и $x^{a+1} y^b$. Таким образом,

$$x^a y^b + x^{a+1} y = x^a y^b (1 + x).$$

Другая пластинка, расположенная горизонтально в каком-то другом месте, дает $x^c y^d (1 + x)$. Для двух пластинок получается покрытие $x^a y^b (1 + x) + x^c y^d (1 + x)$. Горизонтально расположенные пластинки домино дают выражение вида $(1 + x) f(x, y)$. Для вертикально расположенных пластинок соответствующее выражение имеет вид $(1 + y) g(x, y)$. Чтобы учесть оба выражения, составим сумму $(1 + x) f(x, y) + (1 + y) g(x, y)$. Если пластинки домино перекрываются, то может появиться коэффициент, больший единицы, а если какой-то квадрат остается непокрытым, то соответствующий коэффициент равен нулю. Таким образом, требуется найти $f(x, y)$ и $g(x, y)$ с коэффициентами 0 и 1, такие, что

$$(1 + x + \ldots + x^7)(1 + y + \ldots + y^7) - 1 - x^7 y^7 = \\ = (1 + x) f(x, y) + (1 + y) g(x, y).$$

Это тождество должно быть справедливо при любых x и y. В частности, оно должно выполняться при $x = -1$ и $y = -1$. Эти значения обращают правую часть в нуль. Выражение $(1 + x + \ldots + x^7)$ с левой стороны равно нулю. $(1 + y + \ldots + y^7)$ тоже равно нулю, а $(-1) - (-1)^7 (-1)^7 = -2$. Следовательно, $-2 = 0$, т. е. получено противоречие. Легко проверить, что на доску можно положить максимум 30 пластинок.

Замечание. Если сделать коэффициенты f и g произвольными, а не равными 0 и 1, то это будет означать использование пластинок различной толщины, причем толщина определяется коэффициентом, как, например, $^1/_2$ в $^1/_2 x^a y^b (1 + x)$. Можно было бы поставить задачу о равномерном по высоте покрытии доски пластинками домино.

Следующая теорема является интересным и полезным обобщением задачи о покрытии шахматной доски пластинками домино [10].

Теорема 2.31 (**де Брейн**) [15а]. *Пусть в n-мерном пространстве дан прямоугольный ящик со сторонами p_i, $i = 1, \ldots, n$; если его можно заполнить блоками со сторонами a_j, $j = 1, \ldots, n$, где a_j делит a_{j+1}, $j = 1, \ldots, n$ (такой блок назовем гармоническим), то его можно заполнить блоками, которые все ориентированы в одном направлении, т. е. a_1 делит p_1, a_2 делит p_2, $\ldots$, a_n делит p_n, возможно, после переупорядочивания p.*

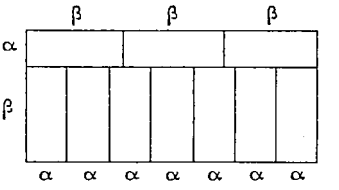
Замечание. Условие, что a_j делит a_{j+1}, необходимо. В противном случае, как показано на фиг. 2.22, ящик не может быть заполнен блоками, ориентированными параллельно какой-то одной из сторон. Если блок имеет размеры α, β, где α не делит β, ни α, ни β не будут делителями ширины ящика $\alpha + \beta$ (его длина равна $\alpha\beta$).

Доказательство теоремы 2.31. Докажем сначала, что теорема справедлива для блоков со сторонами 1, 1, . . ., 1, c. Такой блок назовем *плиткой*.

Предположим, что данный прямоугольный ящик заполнен единичными кубами, и припишем прямоугольные координаты $k_1, k_2, \ldots, k_n$ каждому кубу. Заметим, что выражение

$$\sum_A \Big(\sum_B e^{2\pi i (k_1+k_2+\ldots+k_n)/c} \Big),$$

где A — множество всех плиток, а B — множество кубов в каждой плитке, определяет покрытие ящика. Каждая плита укладывается параллельно некоторой стороне p_j ящика. Поскольку первая сумма берется по всем кубам в плитке, в выражении $e^{2\pi i k_j/c}$ величина k_j пробегает все последовательные значения, относящиеся к этим кубам. Так как всего имеется c таких кубов, k_j/c порождают все остатки по mod c, дающие все c корней единицы. Поскольку весь ящик покрывается плитками, эта двойная сумма должна обращаться в нуль, потому что каждый куб учитывается в какой-то плитке. Однако двойная сумма представляет собой краткую запись выражения $\sum_{k_n=1}^{p_n} \sum_{k_{n-1}=1}^{p_{n-1}} \ldots \sum_{k_1=1}^{p_1} e^{2\pi i (k_1+\ldots+k_n)/c}$, которое описывает покрытие ящика; это выражение равно $\sum_{k_n} e^{2\pi i k_n/c} \sum_{k_{n-1}} e^{2\pi i k_{n-1}/c} \ldots \sum_{k_1} e^{2\pi i k_1/c}$. Таким образом, один из членов этого произведения должен быть равен нулю. Такой член, поскольку он представляет собой сумму, может обратиться в нуль только в том случае, если его индекс суммирования пробегает по всем целочисленным значениям, кратным c. При этом сумма c корней единицы равняется нулю, т. е. некоторое p_i, $i = 1, \ldots, n$, кратно c. Таким образом, c должно делить длину одной из сторон ящика, а весь ящик можно заполнить плитками параллельно этой стороне.



Фиг. 2.22.

Если предположить, что ящик может быть заполнен блоками размерами $a_1 \times a_2 \times \ldots \times a_n$, то очевидно, что его можно заполнить плитками размерами $1 \times 1 \times \ldots \times 1 \times a_n$, где a_n делит p_i при некотором i. Без потери общности примем, что a_n делит p_n. Любая грань ящика со сторонами $p_1 \ldots p_{n-1}$ заполняется $(n-1)$-мерными плитками размерами

$$a_1 \times a_2 \times \ldots \times a_{n-2} \times a_{n-1},$$
$$a_1 \times a_2 \times \ldots \times a_{n-1} \times a_n,$$
$$a_1 \times a_2 \times \ldots \times a_{n-3} \times a_{n-1} \times a_n,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_2 \times \ldots \times a_{n-2} \times a_{n-1} \times a_n,$$

где каждый раз опускается один сомножитель. Так как каждый размер a_i делит a_{i+1}, то грань можно заполнить плитками размерами $a_1 \times a_2 \times \ldots \times a_{n-1}$. Стороны a_n плиток должны быть все ориентированы в направлении p_n. Таким образом, блоки полностью заполняют часть ящика между одной из его граней и параллельным сечением, которые имеют размеры $p_1 \times p_2 \times \ldots \times p_{n-1}$. Но теперь по индукции те же самые рассуждения можно применять, начиная с этого сечения, параллельного грани ящика, до тех пор, пока весь ящик не будет заполнен блоками, ориентированными в одну сторону.

В [15а] доказано, что если блок не является гармоническим, то существует коробка, которую можно заполнить, хотя она не кратна блоку. Коробка называется кратной блоку, если длины ее сторон кратны длинам сторон блока, возможно после переупорядочивания.

2.8. Дискретная геометрия: упаковка, покрытие, заполнение [11, 13, 41, 56, 68, 80]

Задачи упаковки максимально возможного числа предметов (без наложения) в контейнере определенной формы и определенного объема встречаются во многих приложениях. Область этих приложений простирается от хранения консервных банок в маленьком буфете до упаковки кабелей (цилиндров) в большой цилиндрической трубе. Известно, что однажды одна компания по перевозке цитрусовых наняла в качестве консультанта математика, чтобы он отыскал наилучшие способы упаковки апельсинов в корзины, что позволило бы максимизировать использование корзин и минимизировать количество раздавленных апельсинов. Он помог уменьшить количество используемых корзин и, следовательно, стоимость перевозки и, конечно, количество испорченных апельсинов.

Плотность, или эффективность, упаковки определяется как отношение полной меры (например, площади или объема) упаковываемых предметов к мере пространства, в которое упаковываются эти объекты. Очевидно, эффективность не превосходит единицы. С задачей упаковки связана задача покрытия некоторого объекта множеством меньших предметов, которые могут или не могут перекрываться, так что каждая точка объекта принадлежит по крайней мере одному из покрывающих предметов. Эффективность в этом случае определяется так же, как и ранее, за исключением того, что здесь она может превышать единицу.

Чтобы оценить плотность упаковки на бесконечной плоскости, рассмотрим произвольный круг радиуса R. Множество точек этого круга, общих с упаковываемыми предметами, имеет определенную площадь. Возьмем отношение этой площади к полной площади круга и устремим R к бесконечности. Этот предел не зависит от местоположения центра круга.

Упражнение 2.31. Рассмотрим другой круг радиуса R, центр которого находится на расстоянии d от центра первого круга. Покажите, что плотность упаковки при использовании этого круга ограничена снизу плотностью упаковки, соответствующей кругу радиуса $R - d$, и ограничена сверху плотностью упаковки, соответствующей кругу радиуса $R + d$.

Принятие пищи и паркование автомобилей тоже могут служить примерами упаковки. Застилание постели является примером покрытия, которым люди занимаются с раннего детства. Одевание — это еще один пример покрытия; объект, который должен быть покрыт, составляет часть бо́льшего объекта, т. е. человеческого тела. Если покрывающие объекты просто заполняют пространство, не перекрываясь, то это выглядит так, как если бы они были упакованы в пространство. Будем называть этот пограничный случай *заполнением*.

Часто задачи об упаковке и покрытии изучались в легче поддающихся решению постановках, в которых использовались правильные конфигурации. Под *целочисленной решеткой* будем понимать все линейные комбинации с целочисленными коэффициентами n линейно независимых векторов в n-мерном пространстве. Векторы целочисленной решетки можно использовать для сдвигов любого измеримого по Лебегу множества K. Полученные множества образуют решеточную упаковку множеств K. Сдвиги K образуют покрытия, если каждая точка пространства принадлежит хотя бы одному из этих множеств. Если a — вектор, то *сдвиг* K на a равен $(K + a)$.

Упаковка на плоскости

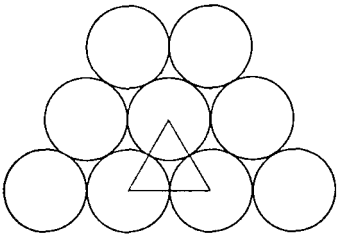
Конгруэнтные правильные шестиугольники можно использовать для заполнения плоскости таким образом, что в каждой вершине будут сходиться три шестиугольника. Существуют еще два способа покрытия плоскости правильными многоугольниками. В одном из них используются правильные конгруэнтные треугольники, причем в каждой вершине сходятся по шесть треугольников, а в другом используются конгруэнтные квадраты, причем в каждой вершине сходятся по четыре квадрата. Из этих трех способов покрытия плоскости покрытие с помощью шестиугольников является наилучшим для упаковки вписанных единичных кругов внутри таких многоугольников. Мы вычислим эффективность такой упаковки, которую обозначим через E_h, а также эффективность упаковки единичных кругов в квадраты, которую обозначим через E_s, а эффективность упаковки третьего типа оставим в виде упражнения. Рассмотрим задачу упаковки равных кругов радиуса $r = {}^1/_2$ на плоскости таким образом, что каждый круг касается шести других кругов, как показано на фиг. 2.23. Каждый такой круг можно рассматривать как круг, вписанный в правильный шестиугольник с тем же центром; эти шестиугольники заполняют плоскость.

Эффективность E_h получается путем сравнения площади правильного треугольника, вершинами которого служат центры трех смежных кругов, как показано на фиг. 2.23 (заметим, что благодаря симметрии этот треугольник повторяется на всей плоскости), с площадями секторов трех кругов, содержащихся внутри треугольника.

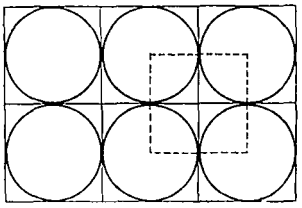
Площадь треугольника равна $\sqrt{3}/4$. Площадь каждого сектора равна $\pi r^2/6$. Поэтому площадь трех секторов равна $3\pi(1/2)^2/6 = \pi/8$. Эффективность такой упаковки равна

$$E_h = \frac{\pi/8}{\sqrt{3}/4} = \frac{\pi}{\sqrt{12}} \approx 0{,}9069.$$

Замечание. Из работы Тху [79], выполненной в 1910 г., следует, что круги не могут быть упакованы каким-то произвольным способом



Фиг. 2.23.



Фиг. 2.24.

на плоскости так, чтобы получилась эффективность, большая чем E_h. (Ниже мы дадим разные доказательства этого факта.)

Снова рассмотрим круги радиуса $r = 1/2$ на решетке точек с целочисленными координатами и предположим, что каждый круг вписан в единичный квадрат с тем же центром из множества квадратов, покрывающих плоскость (фиг. 2.24). Чтобы определить эффективность такой упаковки, воспользуемся симметрией структуры на всей плоскости. Сравним площадь единичного квадрата, углами которого служат центры четырех смежных кругов (как показано на фиг. 2.24), с площадью четвертей кругов, покрывающих части квадрата. Получаем

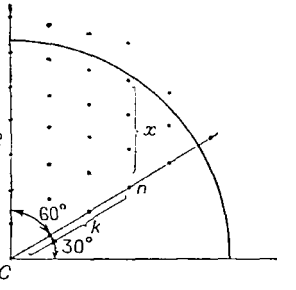
$$E_s = 4\,\frac{\pi\left(\frac{1}{2}\right)^2/4}{1} \approx \frac{3{,}142}{4} \approx 0{,}786.$$

Таким образом, упаковка кругов в шестиугольники, обладающая большей эффективностью, является более плотной упаковкой, чем упаковка кругов в квадраты.

Упражнение 2.32. Определите эффективность покрытия плоскости кругами, вписанными в конгруэнтные правильные треугольники, покрывающие плоскость.

Упражнение 2.33. Заметим, что радиус круга, вписанного в правильный многоугольник с n сторонами, каждая из которых имеет длину L, равен $r = L \operatorname{ctg}(\pi/n)$. Покажите, что плотность упаковки круга в такой многоугольник равна $(\pi/n)\operatorname{tg}(\pi/n)$. Воспользуйтесь этим результатом для того, чтобы определить эффективность упаковки кругов в трех указанных выше случаях.

Задача. Определите число кругов радиуса r, упакованных так, что каждый из них касается шести других и содержится частично или полностью в круге радиуса $R > r$, который имеет общий центр с одним из кругов [17].



Фиг. 2.25.

Решение. Для простоты примем сначала, что расстояние между центрами двух смежных кругов равно единице. Поскольку центры располагаются в треугольнике, вычислим число центров в одном из шести подобных секторов большого круга. Пусть C — центр этого круга. Тогда C не принадлежит ни одному сектору (фиг. 2.25). Надо определить число центров на произвольной вертикальной линии в секторе и значения этой величины вдоль вертикалей от 1 до n, где n — число центров, расположенных вдоль радиуса R.

Пусть k — число центров, расположенных на радиусе вплоть до произвольно выбранной вертикальной линии, число центров на которой x теперь надо вычислить. Получаем

$$\left(x + k \sin \frac{\pi}{6}\right)^2 + \left(k \cos \frac{\pi}{6}\right)^2 = n^2.$$

Таким образом,

$$x = \left[\sqrt{n^2 - \frac{3}{4}k^2} - \frac{k}{2}\right],$$

и полное число центров N получается путем суммирования по k, умножения на шесть (по числу секторов) и прибавления единицы для учета центра C, т. е.

$$N = \left[1 + 6 \sum_{k=0}^{[n]} \sqrt{n^2 - \frac{3}{4}k^2} - \frac{k}{2}\right].$$

Квадратные скобки означают взятие целой части. Если центры отстоят более чем на $2r$, то $R = 2rn$ будет содержать N центров и $R = 2rn + r$ будет содержать N кругов. Таким образом, $n = \{{}^1/_2 [(R/r) - 1]\}$, где квадратные скобки означают взятие целой части.

Упражнение 2.34 [50]. Рассмотрим два круга единичного радиуса. Пусть (ρ_1, θ_1) и (ρ_2, θ_2) — полярные координаты их центров. Тогда расстояние d между их центрами равно

$$d = [\rho_1^2 + \rho_2^2 - 2\rho_1\rho_2 \cos(\theta_1 - \theta_2)]^{1/2}.$$

Докажите, что два круга можно упаковать в круг радиуса 2, если $d \leqslant 4$. Заметим, что равенство имеет место в случае касания двух кругов.

Упражнение 2.35. N кругов, каждый из которых имеет радиус r, упакованы внутри круга радиуса R так, что каждый круг касается большого круга и не существует пустого пространства между кругами. Покажите, что

$$\frac{R}{r} = \frac{2}{1 - \operatorname{tg}^2[(\pi/4) - (\pi/2N)]}.$$

Упаковка кругов на плоскости

Здесь будет доказана следующая теорема:

Теорема 2.32. *Плотность наиболее плотной упаковки равных кругов на плоскости равна* $\pi/\sqrt{12}$.

Доказательство 1. Этапы этого доказательства будут даны ниже. Предположим, что имеется система кругов радиуса 1 с центрами в точках $P_1, P_2, \ldots$, которая образует упаковку на плоскости. Сопоставим каждой точке P_i многоугольник Вороного $\Pi(P_i)$, который состоит из всех точек плоскости, более близких к P_i, чем к любому другому центру. Эти многоугольники являются выпуклыми и в совокупности покрывают всю плоскость без перекрытий. Докажем, что

$$\text{Площадь } [\Pi(P_i)] \geqslant \sqrt{12}. \tag{2.1}$$

Лемма 2.3. Пусть P_i, P_j, P_k — *произвольные три центра, и пусть* X — *произвольная точка. Тогда или* $|X - P_i| \geqslant \sqrt{4/3}$, *или* $|X - P_j| \geqslant \sqrt{4/3}$, *или* $|X - P_k| \geqslant \sqrt{4/3}$.

Доказательство. Без потери общности можно принять точку X в качестве начала отсчета. Предположим, что $|P_i| < \sqrt{4/3}$,

$$|P_j| < \sqrt{4/3}, \quad |P_k| < \sqrt{4/3};$$

получится противоречие

Так как P_i, P_j, P_k — центры кругов радиуса 1, которые являются частью упаковки, получаем $|P_i - P_j| \geqslant 2$, $|P_i - P_k| \geqslant 2$ и $|P_j - P_k| \geqslant 2$. Поэтому

$$4 \leqslant |P_i - P_j|^2 = |P_i|^2 + |P_j|^2 - 2|P_i||P_j| \cos\theta_{ij},$$

где θ_{ij} — угол, образованный отрезками OP_i и OP_j. Следовательно,

$$4 < \frac{4}{3} + \frac{4}{3} - 2\,|P_i|\,|P_j| \cos\theta_{ij},$$

$$\cos\theta_{ij} < -\frac{2}{3}\,(|P_i|\,|P_j|)^{-1} < -\frac{1}{2}.$$

Отсюда вытекает, что $2\pi/3 < \theta_{ij} \leqslant \pi$ и аналогичные соотношения имеют место для θ_{ik} и θ_{jk}. Но это невозможно, и лемма доказана.

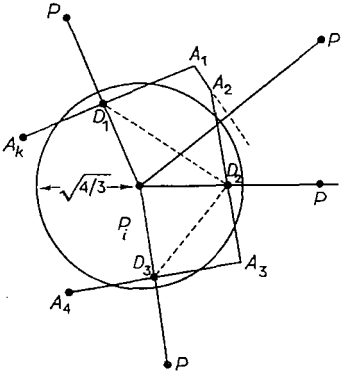
Пусть $\Pi\,(P_i) = A_1, A_2, \ldots, A_k$, где $A_1, A_2, \ldots, A_k$ — вершины многоугольника $\Pi\,(P_i)$. Тогда каждая вершина равноудалена по крайней мере от трех точек P, одной из которых является P_i; обозначим другие через P_j и P_k.

Из леммы следует, что если A — какая-то вершина $\Pi\,(P_i)$, то хотя бы одна из величин $|A - P_i|$, $|A - P_j|$, $|A - P_k|$ превосходит $\sqrt{4/3}$. Так как

$$|A - P_i| = |A - P_j| = |A - P_k|,$$

получаем

$$|A - P_i| \geqslant \sqrt{\frac{4}{3}}. \qquad (2.2)$$



Фиг. 2.26.

Предположим теперь, что круг с центром P_i радиуса 1 заменен на круг радиуса $\sqrt{4/3}$ и с центром в P_i (фиг. 2.26). Тогда в силу неравенства (2.2) вершины $\Pi\,(P_i)$ не являются внутренними точками этого круга. Пусть S означает область, общую для этого увеличенного круга и $\Pi\,(P_i)$. Очевидно, что площадь $\Pi\,(P_i)$ не меньше площади S; поэтому для того, чтобы доказать неравенство (2.1), достаточно доказать, что площадь

$$S \geqslant \sqrt{12}. \qquad (2.3)$$

Пусть m — число прямолинейных отрезков, ограничивающих S.

Случай 1 (Тот). $m \leqslant 6$.

Пусть T_1 — площадь области, лежащей внутри увеличенного круга, но по другую, если смотреть от P_i, сторону от одного из прямолинейных отрезков, ограничивающих S (фиг. 2.27). Тогда

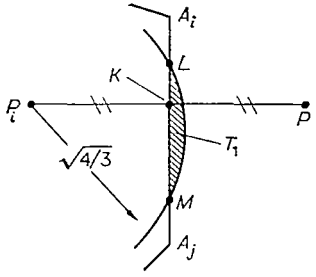
$$\text{Площадь } S = \frac{4}{3}\pi - (T_1 + T_2 + \ldots + T_m).$$

Если A_iA_j — та сторона $\Pi\,(P_i)$, которая составляет часть границы T_1, то отрезок A_iA_j перпендикулярен отрезку P_iP, соединяющему P_i с каким-то другим центром P, и, таким образом, $P_iP \geqslant 2$, так что $P_iK \geqslant 1$ (при отсутствии каких-либо предположений относи-

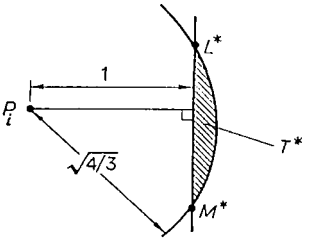
тельно m). Таким образом,

$$T_1 \leqslant T^*,$$

где T^* — площадь сегмента круга радиуса $\sqrt{4/3}$, отсекаемого хордой, проходящей на расстоянии 1 от P_i. Это показано на фиг. 2.28.



Фиг. 2.27.



Фиг. 2.28.

Простой расчет дает $\angle L^*P_iM^* = \pi/6$, так что

$$T_1 \leqslant T^* = \frac{1}{2} \cdot \frac{4}{3} \left(\frac{\pi}{3} - \sin \frac{\pi}{3} \right) = \frac{2}{3} \left(\frac{\pi}{3} - \frac{\sqrt{3}}{2} \right).$$

Аналогичные соотношения имеют место для $T_2, \ldots, T_m$. Следовательно,

$$\text{Площадь } S \geqslant \frac{4}{3}\pi - m\frac{2}{3}\left(\frac{\pi}{3} - \frac{\sqrt{3}}{2}\right) \geqslant \frac{4}{3}\pi - 4\left(\frac{\pi}{3} - \frac{\sqrt{3}}{2}\right) = \sqrt{12},$$

поскольку $m \leqslant 6$.

(Равенство достигается только в том случае, если S является правильным шестиугольником, описанным около круга радиуса 1 с центром в P_i и, следовательно, вписанного в круг радиуса $\sqrt{4/3}$.)

Случай 2 (Фью). $m \geqslant 7$.

Здесь S состоит из m треугольников, а также из секторов круга радиуса $\sqrt{4/3}$. Площадь сектора равна

$$\frac{1}{2}\sqrt{\frac{4}{3}}\sqrt{\frac{4}{3}}\theta = \frac{1}{2}\sqrt{\frac{4}{3}} \times \text{Длина дуги сектора} > \frac{1}{2} \times$$

$$\times \text{Длина дуги сектора}.$$

Площадь каждого треугольника равна

$$\frac{1}{2} \times \text{Основание} \times \text{Высота} \geqslant \frac{1}{2} \times \text{Основание},$$

поскольку, как отмечалось выше, высота каждого треугольника не меньше единицы. Таким образом,

$$\text{Площадь } S \geqslant \frac{1}{2} \times \text{Периметр } S. \tag{2.4}$$

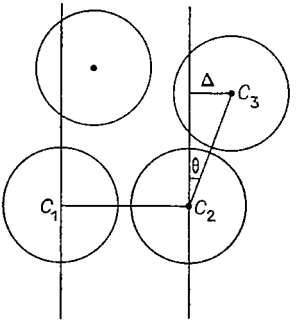
Пусть $D_1D_2 \ldots D_k$ — выпуклый многоугольник, где каждая точка D является серединой одного из прямолинейных отрезков, ограничивающих S. Тогда $D_iD_{i+1} \geqslant 1$ (так как расстояние между любыми двумя точками P больше либо равно 2 и каждая точка D является средней точкой P_iP). Поэтому периметр $S \geqslant T_1T_2 + \ldots + T_mT_1 \geqslant 7$, и в силу неравенства (2.4) площадь $S \geqslant 3{,}5 > \sqrt{12}$. Следовательно, неравенства (2.3) и (2.1) доказаны.

Теперь легко показать, что, поскольку площадь каждого круга равна π, плотность любой упаковки кругов не превосходит $\pi/\sqrt{12}$.

Равенство, согласно этому методу, достигается тогда и только тогда, когда используется упаковка в шестиугольники. Этим завершается доказательство.

Доказательство 2 [55].

1. Предположим, что имеется упаковка неперекрывающихся кругов на плоскости и что эта упаковка является насыщенной, т. е. нельзя добавить ни одного круга так, чтобы он не имел перекрытий.



Фиг. 2.29.

2. Центры смежных кругов могут быть соединены прямыми линиями, так что образуется сеть треугольников, покрывающих плоскость, причем ни в одном треугольнике нет угла, превышающего 120°. Чтобы убедиться в этом, возьмем круг с центром в C_1 и другой, ближайший к нему круг с центром в C_2. Рассмотрим полосу между двумя параллельными линиями, перпендикулярными к отрезку C_1C_2 (фиг. 2.29). Пусть C_3 — центр круга, ближайшего к C_1 или C_2 и к одной из параллельных линий (на расстоянии Δ от нее). Ввиду того что система насыщенная, такой круг должен существовать.

Если C_3 лежит внутри полосы, то угол $C_1C_2C_3$ должен быть меньше 120°. Если C_3 лежит вне полосы, то при помощи диаграммы и вычислений можно показать, что каждый угол треугольника $C_1C_2C_3$ также должен быть меньше 120°.

Заметим, что

$$\sin\theta = \frac{\Delta}{C_2C_3} < \frac{1}{2}.$$

Таким же образом строится вся сеть.

3. Если A — площадь треугольника, то

$$\sqrt{3} \leqslant \frac{1}{2} \cdot 2 \cdot 2 \cdot \sin\alpha \leqslant A < 8.$$

Правое неравенство выполняется, поскольку система кругов насыщенная, откуда следует, что основание и высота каждого треугольника не превосходят 4. Левое неравенство справедливо, так как

длина каждой стороны не менее чем 2, а угол α такой, что $60° \leqslant \alpha < 120°$. Напомним, что площадь треугольника равна половине произведения его сторон на синус угла между ними.

4. Возьмем круг C большого радиуса R с центром в точке O. Рассмотрим сеть треугольников, составленных из отрезков, соединяющих центры кругов, которые частично или полностью лежат в C. Некоторые ребра будут пересекать C в направлении центров, лежащих вне C (внутри круга радиуса $R+1$). Формула Эйлера дает $V - E + F = 1$ для плоской сети. (Докажите этот факт путем сведения задачи к треугольнику, для которого, как было показано ранее в этой главе, $V - E + F = 1$, где V, E, F — вершины, ребра и области сети.) Если E' — количество ребер, входящих в E и лежащих на границе сети в C, то каждое из этих ребер ограничивает только одну область в сети. Получаем $3F = 2E - E'$, потому что каждая грань ограничена тремя ребрами и каждое ребро засчитывается дважды; ребра E' ограничивают только одну область. Поэтому $F = 2V - E' - 2$.

5. Если обозначить площадь сети через $n(C)$, которая равна AF, то из пп. 3 и 4 следует

$$\sqrt{3}\,F \leqslant n(C) < 8F, \quad \text{или} \quad 1 \leqslant \frac{n(C)}{\sqrt{3}\,F} = \frac{n(C)}{\sqrt{3}\,(2V - E' - 2)}.$$

Очевидно, что если рассмотреть два концентрических круга, один из которых внутренний, а другой внешний по отношению к C, то получим

$$\pi\,(R-3)^2 < n(C) < \pi\,(R+1)^2$$

и, следовательно,

$$\lim_{R\to\infty} \frac{n(C)}{\pi R^2} = 1;$$

этот факт будет использован в п. 6.

6. Пусть δ_R обозначает плотность упаковки. Тогда по определению

$$\delta_R = \lim_{R\to\infty} \frac{\pi V}{\pi R^2}.$$

Однако из п. 5 следует, что

$$\frac{\pi V}{\pi R^2} \leqslant \frac{\pi V}{\sqrt{3}\,(2V - E' - 2)} \frac{n(C)}{\pi R^2}. \qquad (*)$$

Наша цель состоит в том, чтобы показать, что

$$\delta_R = \lim_{R\to\infty} \frac{\pi V}{\pi R^2} \leqslant \lim_{R\to\infty} \frac{\pi V}{\sqrt{3}\,(2V - E' - 2)} \lim_{R\to\infty} \frac{n(C)}{\pi R^2} = \frac{\pi}{\sqrt{12}}.$$

7. Деля числитель и знаменатель в правой части (*) на V и замечая, что из $R \to \infty$ следует, что $V \to \infty$, получим, что $\pi/\sqrt{12}$ являет-

ся пределом правой части при $R \to \infty$, если

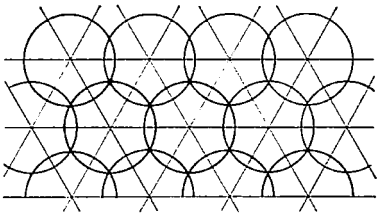
$$\lim_{R\to\infty} \frac{E'}{V} = 0.$$

Здесь E' представляет собой также число вершин (а также число центров кругов) граничного многоугольника сети. Площадь этих кругов равна $\pi E'$ и содержится в кольце между кругами радиусов $R + 2$ и $R - 4$. Таким образом, $\pi E' < 12\pi\,(R - 1)$. Из п. 4 получаем $F = 2V - E' - 2$, или $2V > F$, а из п. 5 получаем $n\,(C) < 8F$. Следовательно,

$$\frac{\pi E'}{V} < \frac{24\pi\,(R-1)}{F} < \frac{192\pi\,(R-1)}{n\,(C)} < \frac{192\,(R-1)}{(R-3)^2} \text{ при } R \to \infty.$$

Этим заканчивается доказательство. Как мы уже видели ранее, граница $\pi/\sqrt{12}$ достигается в случае шестиугольной упаковки кругов.

Теорема 2.33. *Если каждая точка плоскости покрыта (принадлежит) конечным числом кругов из совокупности единичных кругов, покрывающих плоскость, то плотность этого покрытия больше либо равна* 1,209.



Фиг. 2.30.

Доказательство. Построим сеть треугольников, как в задаче об упаковке, но начнем с двух кругов, которые, пересекаясь, образуют лунку. Одна из двух точек пересечения должна быть покрыта третьим кругом (если их несколько, то выберем один), центр которого используем в качестве третьей вершины треугольника. Каждый треугольник имеет описанную окружность, радиус которой не превосходит единицы. Но равносторонний треугольник имел бы бо́льшую площадь, равную $^3/_4\sqrt{3}$. Таким образом, $n\,(C) \leqslant {}^3/_4\sqrt{3}\,F$. Имеют место соотношения

$$\frac{\pi V}{\pi R^2} = \frac{\pi V}{n\,(C)} \frac{n\,(C)}{\pi R^2} \geqslant \frac{4\pi V}{3\sqrt{3}\,2V} \frac{n\,(C)}{\pi R^2};$$

при $R \to \infty$ получаем доказательство теоремы. (Покрытие, которое реализует эту плотность, показано на фиг. 2.30 [55].)

Применения

Задача о парковании [40]. Имеется много реальных задач, в которых встречаются различные геометрические понятия, например регулярные и нерегулярные геометрические фигуры. Примером

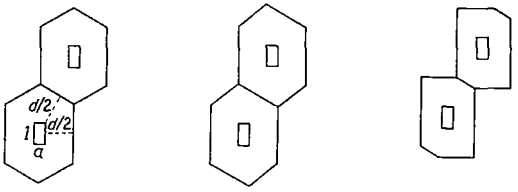
может служить задача о парковании автомобилей. Даже несмотря на то, что автомобили располагаются под разными углами и это может привести к необходимости модифицировать приводимые ниже рассуждения, ответ заслуживает внимания.

Постановка. Дана большая прямоугольная область. Требуется определить наибольшее число конгруэнтных (сравнительно маленьких) кругов, которые можно упаковать в ней так, что любой круг можно передвинуть, не трогая других кругов. Что это за расположение?

Ответ состоит в том, что следует использовать двойные ряды кругов, где круг в одном ряду касается двух кругов в соседнем ряду и ширина промежутков между двойными рядами достаточно велика для того, чтобы можно было выкатить круг. Вблизи границ следует располагать круги в один ряд, но, поскольку предполагается, что площадь прямоугольника очень велика, этот вопрос не представляет большого интереса. Плотность этой упаковки равна

$$\frac{1}{2}(\sqrt{5}-1)\pi/\sqrt{12}.$$

Задача о земельном участке [81]. Дан земельный участок большой площади и n конгруэнтных прямоугольных домов, расположенных на нем (стороны каждого дома равны a и b). Как разместить дома



Фиг. 2.31.

так, чтобы расстояния между любыми двумя домами были максимальными? Обратно, при данном оптимальном желательном расстоянии d между домами, какое максимальное число домов можно разместить на этой площади? Оказывается, что решения распадаются на три класса в зависимости от отношения сторон прямоугольника. Примем, что бо́льшая сторона каждого дома имеет длину $b = 1$.

Пусть S — параллельная область на единичном расстоянии от прямоугольника, меньшая сторона которого имеет длину a. Для удобства примем, что $d = 2$. Тогда центрально симметричный шестиугольник наименьшей площади, содержащий S, будет относиться к типам 1, 2 или 3 в зависимости от того, какое из следующих трех

соотношений выполняется (фиг. 2.31):

$$2 - \sqrt{2} \leqslant a,$$
$$a \leqslant 4 - \sqrt{12} = 0{,}536,$$
$$4 - \sqrt{12} < a < 2 - \sqrt{2} = 0{,}586$$

соответственно.

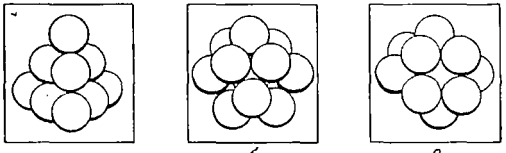
Расстояние от каждого дома, обозначенного прямоугольником, до ограничивающего шестиугольника равно половине разделяющего соседние дома расстояния. Грубо говоря, следовало бы ожидать интуитивно такого решения, так как шестиугольники заполняют всю плоскость.

Упаковка в трехмерном и n-мерном пространстве

Рассмотрим теперь упаковку сфер в трехмерном пространстве. Один из путей состоит в том, чтобы разделить (или разбить) все пространство на единичные кубы и вписать сферу в каждый куб. Так как каждый куб имеет по шесть соседей, имеющих с ним по одной общей грани, каждая вписанная сфера будет касаться шести других сфер.

Упражнение 2.36. Покажите, что эффективность такой упаковки равна приблизительно 0,524.

Рассмотрим снова вышеуказанное разбиение трехмерного пространства и представим себе, что все грани кубов удалены, но ребра



Фиг. 2.32.

и вершины остались. Поместим сферу внутрь куба так, чтобы она касалась всех ребер и проникала в соседние кубы через просветы, образованные после удаления граней.

Естественно, что некоторые кубы, соседние с данным, содержат сферические шапки и, следовательно, в них нельзя поместить сферу (фиг. 2.32, *а*). Однако сфера может быть помещена в другие соседние кубы, так что сферические шапки будут проникать в кубы, которые уже содержат сферические шапки. Продолжим упаковку таким же образом. Очевидно, что теперь каждая сфера касается двенадцати других сфер по ребрам куба.

Представим себе, что шесть сферических шапок, проникающих в соседние кубы, отсечены и расположены таким образом, что они проникают внутрь шести граней одного из кубов; таким образом, каждой сфере можно сопоставить два куба. В предположении, что стороны квадрата имеют длину a, объем двух кубов равен $2a^3$ и объем сферы равен $V = (4/3)\pi\,(a\sqrt{2}/2)^3$. Эффективность, или плотность, такой упаковки при $a = 1$ равна $V/2a^3 = \pi/3\sqrt{2} \sim 0{,}74048$; это наиболее плотная среди всех известных упаковок сфер. Стейнхауз [75], получивший это значение плотности упаковки, отметил, что если бензин находится во взвешенном состоянии в мыльной эмульсии, причем доля эмульсии превышает 75%, то, вероятно, бензин не сгорит полностью, так как его сферы (или шарики) не все могут соприкасаться.

Известны два других способа упаковки сфер в пространстве (фиг. 2.32, *б*, *в*), которые позволяют получить такую же плотность. Один из них состоит в том, что сферы упаковываются в плоском слое так, что каждая касается шести других, как это делается в случае наиболее плотной упаковки кругов. Над этим слоем размещается следующий слой сфер так, что каждая сфера расположена над полостью между тремя сферами. Третий слой размещается так, что его сферы располагаются над полостями первого слоя, над которыми нет сфер второго слоя. Каждая сфера второго слоя касается 12 сфер. Точки касания образуют фигуру, которая называется кубооктаэдром. Она получается путем отсечения вершин правильного октаэдра кубом, грани которого проходят через середины ребер октаэдра. Эта фигура имеет шесть квадратов и восемь равносторонних треугольников. При третьем способе упаковки производятся те же действия, что и выше, за исключением того, что здесь сферы в третьем слое лежат над сферами в первом слое. В этом случае точки касания образуют многогранник с 14 гранями, который получается из кубооктаэдра путем рассечения последнего пополам плоскостью, проходящей через ребра, образующие шестиугольник, поворота одной половины по отношению к другой на 60° и соединения этих половин [50а].

Было показано [68], что эффективность наиболее плотной упаковки равных сфер в трехмерном пространстве независимо от расположения их центров, как, например, в случае с решеточными упаковками, не может превышать 0,7797. Нижняя граница числа сфер при наиболее плотных решеточных упаковках известна для $2 \leqslant n \leqslant 12$, но неизвестна для бо́льших значений n [68].

Исследуем теперь границы для плотностей упаковки и покрытия n-мерного пространства равными сферами. Рассмотрим $n + 1$ единичных сфер радиуса 1, упакованных в n-мерном пространстве, центры которых совпадают с вершинами правильного $(n + 1)$-симплекса с длиной ребер, равной 2. Пусть σ_n обозначает отношение объема части симплекса, покрытой сферами, к объему всего симплекса.

Роджерс [68] показал, что эффективность наболее плотной упаковки единичных сфер в n-мерном пространстве не превосходит σ_n. Даннелс (см. также [68], стр. 90) предложил следующую асимптотическую формулу для σ_n:

$$\sigma_n \sim \frac{n}{2^{n/2}\, e}.$$

Если δ_n^L — плотность наиболее плотной решетчатой упаковки, а δ_n — плотность наиболее плотной упаковки, то

$$\frac{nC}{2^n} \leqslant \delta_n^L \leqslant \delta_n \leqslant \sigma_n,$$

где C — некоторая константа. Величину σ_n теоретически можно вычислить при всех n, а практически σ_n вычислена для $n = 2, 3, 4$. Верхняя граница $\delta_n \leqslant \sigma_n$ является наилучшей из всех известных границ *при произвольном* n. Границы, лучшие чем $\delta_n^L \geqslant nC/2^n$, вычислены для $9 \leqslant n \leqslant 12$ (см. в [68] ссылку на Чаунди, Барнза, Коксетера и Тодда), а также для некоторых бо́льших величин n [2a].

Нижняя граница для плотности *покрытия* n-мерного пространства $n + 1$ равными сферами дана в работе [12], где рассматриваются единичные сферы, центры которых расположены в вершинах правильного $(n + 1)$-симплекса с ребром $[2\,(n + 1)/n]^{1/2}$. Сферы покрывают весь симплекс (с перекрытиями).

Если τ_n — отношение суммы объемов частей («секторов») сфер, лежащих в симплексе, к объему всего симплекса, то нижняя граница плотности покрытия пространства единичными сферами равна

$$\tau_n = \left(\frac{2n}{n+1}\right)^{n/2} \sigma_n.$$

Задача (нерешенная). Дана сфера в трехмерном пространстве. Требуется определить минимальное число сфер того же размера, которые можно расположить вокруг данной сферы так, чтобы полностью покрыть ее, т. е. чтобы все продолжения радиусов упирались в одну из сфер. Естественно, некоторые сферы можно расположить над сферами, непосредственно прилегающими к данной, или еще дальше. В работе [40a] показано, что это число больше или равно 24. В другой работе, упоминаемой в [40a], доказано, что это число меньше или равно 42.

Покажем теперь, как вычислять границы в n-мерном пространстве.

Упаковка равных сфер в n-мерном пространстве: метод Блихфельдта

Лемма 2.4 (очень полезная в разнообразных модификациях метода Блихфельдта [6] для других задач упаковки). *Пусть* $X_1, X_2, \ldots, X_m$ *представляют собой* m *точек в* n*-мерном евклидовом пространстве, причем расстояние между ними* $|X_i - X_j| \geqslant 2$, *т. е. если* $X_i = (x_{1i}, \ldots, x_{ni})$, $X_j = (x_{1j}, \ldots, x_{nj})$, *то*

$[\sum_{k=1}^{m}(x_{ki}-x_{kj})^2]^{1/2}\geqslant 2$ (*при* $i\neq j$). *Пусть* X — *произвольная точка в пространстве. Тогда*

$$\sum_{i=1}^{m}|X-X_i|^2\geqslant\sum_{i=1}^{m}|\overline{X}-X_i|^2=\frac{1}{m}\sum_{1\leqslant i<j\leqslant m}|X_i-X_j|^2\geqslant 2(m-1),$$

где $\overline{X}$ — *центр тяжести* m *точек, т. е.* $\overline{X}=(1/m)\sum_{i=1}^{m}X_i$.

Доказательство.

$$\sum_{i=1}^{m}|X-X_i|^2=\sum_{i=1}^{m}|(X-\overline{X})+(\overline{X}-X_i)|^2=$$

$$=\sum_{i=1}^{m}|X-\overline{X}|^2+2(X-\overline{X})\sum_{i=1}^{m}(\overline{X}-X_i)+\sum_{i=1}^{m}|\overline{X}-X_i|^2=$$

$$=m|X-\overline{X}|^2+\sum_{i=1}^{m}|\overline{X}-X_i|^2\geqslant$$

[так как $\sum_{i=1}^{m}(\overline{X}-X_i)=0$ по определению $\overline{X}$]

$$\geqslant\sum_{i=1}^{m}|\overline{X}-X_i|^2=m|\overline{X}|^2-2\overline{X}\sum_{i=1}^{m}X_i+\sum_{i=1}^{m}|X_i|^2=$$

$$=-m|\overline{X}|^2+\sum_{i=1}^{m}|X_i|^2=-\frac{1}{m}\Big(\sum_{i=1}^{m}X_i\Big)^2+\sum_{i=1}^{m}|X_i|^2=$$

$$=\Big(1-\frac{1}{m}\Big)\sum_{i=1}^{m}|X_i|^2-\frac{2}{m}\sum_{1\leqslant i<j\leqslant m}X_iX_j=$$

$$=\frac{1}{m}\sum_{1\leqslant i<j\leqslant m}|X_i-X_j|^2\geqslant\frac{4}{m}\sum_{1\leqslant i<j\leqslant m}1=\frac{4}{m}\,\frac{m(m-1)}{2}=2(m-1).$$

Этим заканчивается доказательство леммы.

Пусть единичные сферы, образующие упаковку, заменены на сферы с центрами в тех же точках, но радиусами $\sqrt{2}$ и плотностью $2-\rho^2$ на расстоянии ρ от центра. Обозначим через X произвольную точку в пространстве. Докажем, что

Плотность в точке $X\leqslant 2$. (2.5)

Свой вклад в плотность в точке X вносят только те сферы, центры которых находятся на расстоянии не более $\sqrt{2}$ от X. Обозначим эти центры через $X_1, X_2, \ldots, X_m$. Тогда в силу закона плотности

по лемме плотность в точке $X = \sum_{i=1}^{m} (2 - |X - X_i|^2) =$

$$= 2m - \sum_{i=1}^{m} |X - X_i|^2 \leqslant 2m - 2(m-1) = 2.$$

Пусть теперь M — масса увеличенной сферы. Тогда

$$M = \int_0^{\sqrt{2}} (2 - \rho^2)\, d(J_n \rho^n),$$

где J_n — объем единичной n-мерной сферы. Интегрирование по частям дает

$$M = (2 - \rho^2) J_n \rho^n \Big|_0^{\sqrt{2}} + \int_0^{\sqrt{2}} 2 J_n \rho^{n+1}\, d\rho = \frac{2(\sqrt{2})^{n+2}}{n+2} J_n. \qquad (2.6)$$

Пусть N — число центров в большом кубе со стороной $2t$, центр которого находится в точке O, а стороны параллельны координатным осям. Плотность упаковки определяется, согласно [6], как

$$\delta = \sup_{t \to \infty} \frac{N J_n}{(2t)^n} \equiv \frac{\text{Объем сфер}}{\text{Объем куба}}. \qquad (2.7)$$

Теперь увеличенные сферы содержатся в кубе со стороной $2(t + \sqrt{2})$. Масса увеличенных сфер равна NM и по (2.5)

$$NM \leqslant 2\{2(t + \sqrt{2})\}^n. \qquad (2.8)$$

Следовательно, в силу равенств (2.6), (2.7) и неравенства (2.8)

$$\delta \leqslant \sup_{t \to \infty} 2 \left(\frac{t + \sqrt{2}}{t}\right)^n \frac{n+2}{2(\sqrt{2})^{n+2}} = \frac{n+2}{2(\sqrt{2})^n}.$$

Другой метод, дающий несколько более слабый результат [80]

Если $U = (u_1, \ldots, u_n)$ и $V = (v_1, \ldots, v_n)$, то

$$UV = u_1 v_1 + \ldots + u_n v_n.$$

Лемма 2.5. *Существует не более чем $n + 1$ точек $X_1, X_2, \ldots$ в n-мерном пространстве, таких, что*

$$X_i X_j < 0 \quad \text{при} \quad i \neq j.$$

Доказательство. Предположим, что существует $n + 2$ таких точек. Поскольку

$$2X_i X_j = X_i^2 + X_j^2 - (X_i - X_j)^2,$$

получаем, что $(X_i - X_j)$ зависит от начальной точки O, но не зависит от выбора (взаимно ортогональных) осей. Выберем оси так, что

$$X_1 = (a, 0, 0, \ldots, 0), \quad a \geqslant 0.$$

Пусть $X_2 = (x_1, \ldots, x_n)$. Если $X_1X_2 < 0$, то $ax_1 < 0$, так что $x_1 < 0$. Если $X_3 = (y_1, \ldots, y_n)$, то $y_1 < 0$. Если $X_2X_3 < 0$, то $x_1y_1 + \ldots + x_ny_n < 0$. Но $x_1 < 0$, $y_1 < 0$, так что

$$x_2y_2 + \ldots + x_ny_n < 0.$$

Таким образом, имеем по меньшей мере $n + 1$ точек $X_2^*, \ldots$ в $(n-1)$-мерном пространстве [$X_2^* = (x_2, \ldots, x_n)$, $X_3^* = (y_2, \ldots, y_n)$ и т. д.], где $X_i^*X_j^* < 0$ [в $(n-1)$-мерном пространстве]. Продолжая этот процесс, получим по меньшей мере четыре точки в двумерном пространстве, таких, что $X_iX_j < 0$, т. е. угол при O больше $\pi/2$. Это противоречие, и лемма 2.5 доказана.

Теперь возьмем сферы из упаковки единичных сфер и заменим каждую сферу концентрической сферой радиуса $\sqrt{2}$. Пусть X — произвольная точка пространства. Докажем, что X находится внутри не более чем $n + 1$ увеличенных сфер. Обозначим центры этих сфер через $X_1, X_2, \ldots$. Поскольку $X_iX_j \geqslant 2$ (упаковка единичных сфер),

$$4 \leqslant |X_i - X_j|^2 = |(X_i - X) + (X - X_j)|^2 = |X_i - X|^2 + |X - X_j|^2 + 2(X_i - X)(X_j - X) < 4 + 2(X_i - X)(X_j - X),$$

так как $|X_j - X| < \sqrt{2}$, если X находится внутри увеличенной сферы с центром в X_i. Таким образом,

$$(X_i - X)(X_j - X) < 0.$$

Положим $Y_i = X_i - X$. Тогда

$$Y_iY_j < 0.$$

Следовательно, по лемме существует не более чем $n + 1$ таких точек Y_i и, следовательно, не более чем $n + 1$ точек X_i.

Если δ — плотность исходной упаковки, то плотность «упаковки» увеличенных сфер равна $\delta(\sqrt{2})^n$. Поскольку ни одна точка не покрывается больше чем $n + 1$ раз, плотность упаковки

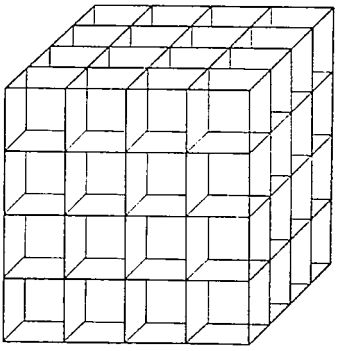
$$\delta(\sqrt{2})^n \leqslant n+1, \quad \text{или} \quad \delta \leqslant \frac{n+1}{(\sqrt{2})^n}.$$

Заполнение

Задача. Требуется разделить n-мерное пространство на конгруэнтные выпуклые многогранники. Обратно, найдите выпуклые многогранники, которыми можно заполнить пространство, используя переносы и вращения. До настоящего времени эта задача полностью решена только для случая плоскости (см. ниже).

Многие заполнения (но не все) известны для трехмерного пространства. Например, если взять любое заполнение плоскости

и построить над ним призмы, то получится заполнение трехмерного пространства конгруэнтными призмами. В некоторых случаях призмы можно разбить на конгруэнтные части и любые такие части рассматривать как стандартные блоки. Например, каждую призму построенную над квадратом из числа квадратов, заполняющих плоскость, можно разбить на кубы. На фиг. 2.33 показано заполнение трехмерного пространства кубами. Аналогично конгруэнтные треугольные призмы (бесконечно разнообразные) можно получить из каждой призмы, построенной над треугольником из числа треугольников, заполняющих плоскость. Известны четыре вида тетраэдров, которые можно использовать для заполнения трехмерного пространства.



Фиг. 2.33. Заполнение пространства кубами.

Следующая лемма будет полезна при обсуждении вопроса о заполнении плоскости конгруэнтными многоугольными областями.

Лемма 2.6. *Среднее число граничных ребер области на конечной плоской карте меньше шести.*

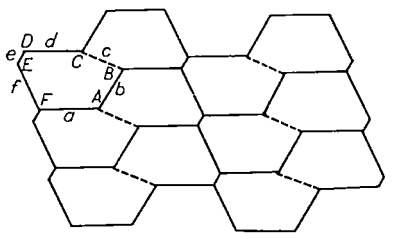
Доказательство. Если воспользоваться формулой Эйлера без учета внешней области, то получим $V - E + F = 1$. Поскольку $3V \leqslant 2E$, имеем $E \leqslant 3F - 3$. Если области пронумерованы как $i = 1, \ldots, F$ и число ребер в i-й области равно E_i, то можно записать $\sum_{i=1}^{F} E_i \leqslant 2E \leqslant 6F - 6$, так как некоторые, но не все ребра принадлежат двум граням. Деля на F, получаем

$$\frac{1}{F}\sum_{i=1}^{F} E_i \leqslant 6 - \frac{6}{F} < 6,$$

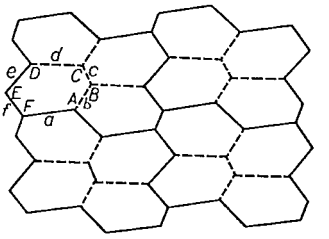
что доказывает лемму 2.6.

Следствие. *На конечной плоской карте существует по крайней мере одна область, ограниченная не более чем пятью ребрами. При $F \to \infty$ среднее число ребер достигает, но не превышает шести.*

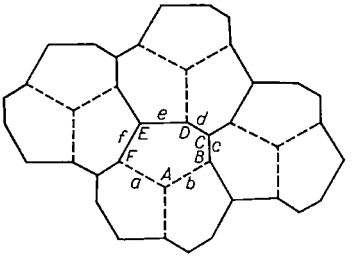
Отсюда следует, что если требуется получить заполнение плоскости конгруэнтными выпуклыми многоугольниками без промежутков и без перекрытий, то не надо рассматривать многоугольники, в которых число сторон превышает шесть. К таким задачам относятся задачи о замещении, о покрытии крыши или о разбиении на соты. Очевидно, что плоскость можно покрыть равносторонними треугольниками, квадратами или правильными шестиугольниками.



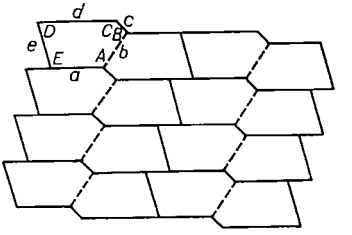
Шестиугольник типа 1; $A+B+C=2\pi$, $a=d$



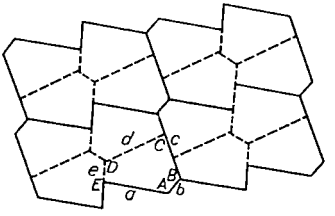
Шестиугольник типа 2; $A+B+D=2\pi$, $a=d$, $c=e$



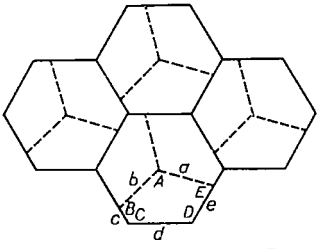
Шестиугольник типа 3; $A=C=E=\frac{2}{3}\pi$, $a=b$, $c=d$, $e=f$



Пятиугольник типа 1; $A+B+C=2\pi$



Пятиугольник типа 2; $A+B+D=2\pi$, $a=d$



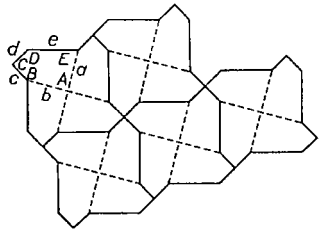
Пятиугольник типа 3; $A=C=D=\frac{2}{3}\pi$, $a=b$, $d=c+e$

Фиг. 2.34.

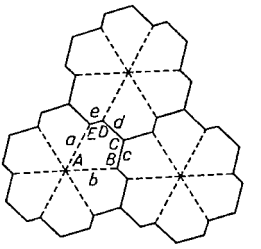
Упражнение 2.37. Покажите, что плоскость можно замостить треугольниками, которые являются конгруэнтными повторениями любого треугольника, и четырехугольниками, которые представляют собой конгруэнтные повторения любого четырехугольника.

Указание. Нужно исказить покрытие плоскости правильными треугольниками или квадратами.

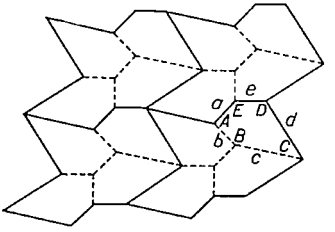
Кершнер [48] показал, что существуют три типа замощения шестиугольниками и восемь типов замощения пятиугольниками; первые три из последних можно рассматривать как частные случаи первых, так как они могут быть превращены в шестиугольники путем соответствующего введения вершины на одной из сторон.



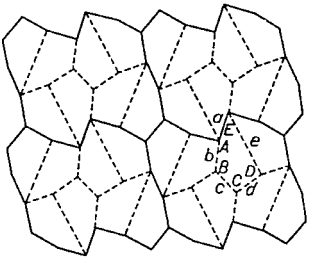
Пятиугольник типа 4;
$A=C=\frac{1}{2}\pi$, $a=b$, $c=d$



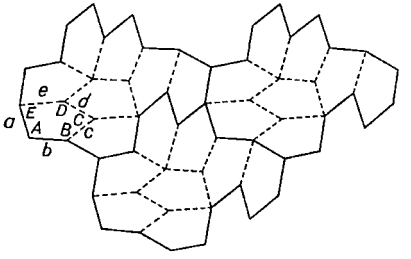
Пятиугольник типа 5;
$A=\frac{1}{3}\pi$, $C=\frac{2}{3}\pi$, $a=b$, $c=d$



Пятиугольник типа 6;
$A+B+D=2\pi$, $A=2C$, $a=b=e$, $c=d$



Пятиугольник типа 7;
$2B+C=2\pi$, $2D+A=2\pi$, $a=b=c=d$



Пятиугольник типа 8;
$2A+B=2\pi$, $2D+C=2\pi$, $a=b=c=d$

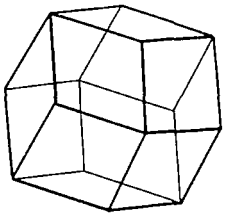
Продолжение фиг. 2.34.

Это дает полное замощение плоскости выпуклыми многоугольниками. На фиг. 2.34 показаны заполнения с указанием условий на углы (прописные буквы) и ребра (строчные буквы).

Упражнение 2.38. Рассмотрите наиболее плотную упаковку кругов на плоскости, где каждый круг касается шести других. Присоединяя к каждому кругу по две треугольные дольки, образованные промежутками между кругами, покажите, что получается замощение плоскости идентичными фигурами. Докажите, что возможны три таких замощения.

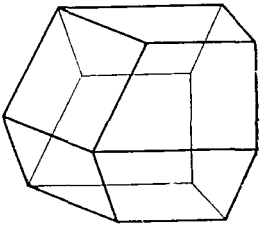
Применения

Кристаллы [49]. Кристаллы состоят из очень большого числа повторяющихся крошечных единиц, имеющих форму многогранника. Предполагается, что единицы состоят из атомов и ионов, плотно упакованных силами притяжения. Экспериментально проверено, что такое представление справедливо для металлов, сплавов и неорганических солей — большинства твердых тел. Например, в гранате эта единица представляет собой ромбический додекаэдр (фиг. 2.35). При кристаллизации меди и ртути наблюдается упаковка, в которой единичной ячейкой является куб. В кристалле повторения получаются путем последовательных параллельных перемещений единичного стандартного блока. Аргентит (Ag_2S) представляет собой кристалл, единицей которого является кубооктаэдр.



Фиг. 2.35. Ромбический додекаэдр.

Медовые соты. Заполнение медовыми сотами представляет собой заполнение части пространства конгруэнтными ячейками, каждая из которых является частью шестиугольной призмы. Один конец ячейки представляет собой шестиугольник, который является входом в нее, а другой конец состоит из трех равных ромбов, составленных так, как показано на фиг. 2.36. Каждая ромбовидная грань является общей с другой ячейкой, шестиугольный вход которой находится на противоположной стороне.



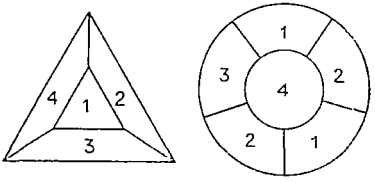
Фиг. 2.36.

Таким образом, два слоя ячеек отделяются не плоскостью, а изогнутой полосой ромбов. Долгое время люди удивлялись, откуда взялась у пчел эта способность строить правильные структуры для медовых сот [78, 84]. Некоторые предлагали физические аргументы, что каждая пчела, работающая в первоначально трубчатой ячейке, пытается оттолкнуть восковую поверхность как можно дальше. Поскольку соседние пчелы делают то же самое в своих трубчатых ячейках, в результате получается шестиугольная сота. Однако ячейки, в которых работали пчелы без соседей, как выяснилось, тоже имели правильные формы, и, насколько нам известно, дебаты еще не закончились.

Тот показал, что пчелиная ячейка не является решением следующей задачи: среди всех открытых ячеек объемом V, образующих медовые соты (любой ширины), требуется найти ячейку наименьшей площади [84]. Заметим, что две ячейки можно составить их открытыми (шестиугольными) концами. Это дает продолженный ромбический додекаэдр. Федоров показал, что ромбические додекаэдры можно использовать для заполнения пространства. Однако если взять октаэдр и соответствующим образом отсечь его вершины плоскими сечениями, сохранив правильность и конгруэнтность его граней, и затем разрезать получившуюся фигуру пополам плоскостью, ортогональной одной из его шестиугольных граней, то в результате получится ячейка, образующая медовые соты. Среди открытых ячеек данного объема, образующих медовые соты, такая ячейка имеет наименьшую площадь. Пчелы по некоторым причинам строят ячейки с несколько большей, чем оптимальная, длиной стен ромбических додекаэдров.

Анализ, проведенный Тотом, показывает, что для пчел выгоднее было бы заменить три ромба на замкнутом конце ячейки на два шестиугольника и два ромба. Это уменьшило бы площадь открытого конца на 35%. Возможно, что пчелы знают правильное решение, но оно им не нравится. В противном случае, почему такие трудолюбивые создания, как пчелы, отказываются от экономного решения!

Раскрашивание [9]. Картой называется граф на плоскости и области, окруженные его простыми контурами, к которым добавляется внешняя область, так что каждая вершина имеет степень, большую двух, и каждое ребро ограничивает в точности две области.



Фиг. 2.37.

Рассмотрим произвольную карту на плоскости. Карту можно рассматривать как замощение плоскости неправильными фигурами. Раскрашивание карты так, чтобы никакие две области с общей граничной кривой (так называемые *смежные области*) не были одного цвета, эквивалентно отысканию некоторого вторичного замощения, конгруэнтного с первичным замощением, в котором цветные плитки располагаются таким образом, что никакие две смежные плитки не имеют одинакового цвета.

Карты можно рисовать на плоскостях, сферах и многих других поверхностях. Формулировка задачи о раскрашивании во всех случаях одна и та же: требуется отыскать наиболее экономичное число цветов для раскрашивания произвольной карты на поверхности. Ответ известен для многих видов поверхностей. Однако, как уже отмечалось в гл. 1, он не получен для плоскости.

Следующие примеры показывают, что на плоскости необходимо не менее четырех цветов (фиг. 2.37).

Задача о раскрашивании областей на плоской карте в четыре цвета эквивалентна задаче о раскрашивании в четыре цвета вершин на двойственной карте, так что никакие две вершины, инцидентные с одним и тем же ребром (так называемые *смежные вершины*) не могут быть одного цвета. Двойственная карта получается путем сопоставления вершины каждой области и связывания пары вершин ребром, если соответствующие области имеют общую границу. Если соответствующие области имеют кратные граничные линии, то используются кратные ребра. Углы исходной карты, в которых сходятся только два граничных ребра, должны быть преобразованы в единственное граничное ребро.

Теорема 2.34. *Пяти цветов достаточно для раскрашивания областей плоской карты.*

Доказательство. Докажем теорему по индукции на двойственной карте, т. е. будем раскрашивать вершины. Предположим, что это справедливо для $n - 1$ вершин. Граф имеет по крайней мере одну вершину v, степень которой меньше или равна 5. Если бы это предположение не было верно, то, используя соотношения $3r \leqslant 2m$ и $6n \leqslant 2m$ (т. е. принимая противоположное утверждение, что каждая вершина имеет степень не менее 6) и подставляя их в формулу Эйлера, получим

$$0 = \frac{2m}{6} - m + \frac{2m}{3} \geqslant 2,$$

что является противоречием. (Аналогичное заключение было получено в разд. 2.3 путем рассмотрения соотношений между ребрами и гранями многогранника.) Если удалить вершину v из графа, то полученный граф может быть раскрашен пятью цветами в силу предположения индукции. Рассмотрим теперь худший случай, т. е. когда пять вершин $v_1, \ldots, v_5$ смежны с v, и пронумеруем их по часовой стрелке. Предположим, что при раскрашивании графа без v (если v удалена, то ребра, инцидентные с ней, тоже удалены) каждой из этих вершин приписывается определенный цвет (в противном случае один из оставшихся цветов можно было бы приписать v и теорема была бы доказана). Пусть соответствующие цвета будут $c_1, \ldots, c_5$. Покажем теперь, что можно перераспределить цвета так, что вершина v получит цвет, отличный от цвета вершин, с которыми она инцидентна, т. е. по крайней мере двум из этих вершин можно приписать один цвет. Рассмотрим подграф, составленный из вершин, окрашенных в цвета c_1 и c_3 (в те же цвета, что и v_1 и v_3). Если v_1 и v_3 не связаны (т. е. между ними нет цепочки) в этом подграфе, то окрашенные в цвет c_1 вершины компоненты связности, которая содержит v_1, могут быть окрашены в c_3, а вершины, окрашенные в c_3, в этой компоненте связности теперь тоже окрашиваются в c_3. Таким образом, и v_1 и v_3 окрашиваются в c_3, а вершина v может быть окрашена в c_1. Если, с другой стороны, v_1 и v_3 связаны в подграфе вершин,

окрашенных в цвета c_1 и c_3, то, например, v_2 и v_4 не могут быть связаны в подграфе вершин, окрашенных в c_2 и c_4. В противном случае цепь, связывающая их, должна встречаться с цепью, связывающей v_1 и v_3, и вершине, в которой сходятся обе цепи, должен быть приписан один цвет одного подграфа и второй цвет второго подграфа. Подграф, который включает v_2 и v_4, может быть перекрашен таким же способом, как и в рассмотренном выше случае v_1 и v_3, не связанных между собой, и, следовательно, вершинам v_2 и v_4 можно приписать один цвет. Другой цвет затем приписывается v. Этим заканчивается доказательство.

Задача о раскрашивании исходной карты в четыре цвета эквивалентна задаче о раскрашивании областей трехвалентной карты. В такой карте каждая вершина (точка пересечения граничных линий) имеет степень 3. Это означает, что она является точкой пересечения *в точности трех граничных линий*.

Задача раскрашивания областей произвольной карты может быть сведена к задаче раскрашивания областей трехвалентной карты. Для этого вершины исходной карты, степень которых отличается от трех, преобразуются в вершины степени 3. Это достигается путем замены какой-то вершины степени, отличной от трех, на замкнутую многоугольную область с числом вершин, равным числу ребер, инцидентных с исходной вершиной. Каждая из новых вершин имеет степень 3, и к ней инцидентно одно из этих ребер. В результате имеем трехвалентную карту. Раскраска исходной карты получается из раскраски трехвалентной карты путем стягивания каждой из новых областей снова в исходную вершину. Таким образом, если четырех цветов достаточно для раскрашивания трехвалентной карты, то их достаточно и для раскрашивания исходной карты.

Теперь можно рассмотреть карту, двойственную к трехвалентной, и сконцентрировать внимание на раскрашивании ее вершин. Заметим, что знание наиболее экономного числа цветов не помогает при раскрашивании конкретной карты. Эта задача все еще является источником разочарования.

Покрытие точек координатной решетки

Рассмотрим теперь некоторые элементарные задачи упаковки и покрытия для точек решетки евклидовой плоскости (точки, все координаты которых — целые числа).

Теорема 2.35 (Ньюмен). *Квадрат S размером $n \times n$, произвольно расположенный на плоскости, покрывает не более чем $(n+1)^2$ точек решетки* [61].

Доказательство. Если C — наименьшее выпуклое множество, содержащее точки решетки в S, то легко получить, что площадь $C \leqslant n^2$, а периметр $C \leqslant 4n$. Пик доказал (см. [13]), что площадь простого многоугольника (никакие два ребра не пересекают друг

друга), вершинами которого являются точки решетки и на границе которого расположено b точек решетки, а во внутренней области — c точек решетки, равна $b/2 + c - 1$. Поэтому площадь $C = b/2 + c - 1 \leqslant n^2$. Поскольку расстояния между любыми двумя точками решетки больше или равны 1, кривая длиной b содержит b точек решетки и, следовательно, $b \leqslant$ периметра $C \leqslant 4n$, откуда $b/2 \leqslant 2n$. Поэтому $b + c \leqslant n^2 + 2n + 1 = (n + 1)^2$.

Упражнение 2.39. Покажите, что число квадратов, образованных в евклидовой плоскости множествами из четырех точек (вершинами квадратов), на совокупности точек решетки размером $m \times n$ $(m \geqslant n)$ равно [52]

$$\frac{(n-1)\,n\,(n+1)\,(2m-n)}{12}.$$

(Для наклоненных квадратов используйте их стороны как гипотенузы треугольников со сторонами a и b.)

Теорема 2.36 (Бликфельдт). *Пусть S — множество на плоскости, площадь которого >1. Тогда S содержит две различные точки (x_1, x_2), (y_1, y_2), такие, что $(y_1 - x_1)$ и $(y_2 - x_2)$ — целые числа.*

Доказательство. Сопоставим каждой целочисленной точке решетки на плоскости единичный квадрат решетки, в котором точка решетки служит правым нижним углом. Рассмотрим пересечение S с этими квадратами. Перенесем все единичные квадраты решетки, которые пересекаются с S, так, чтобы они совпадали с квадратом Δ, правая нижняя вершина которого находится в начале координат. Правые нижние вершины всех квадратов должны попадать в начало координат. Рассмотрим пересечение S с этими единичными квадратами. Все эти пересечения теперь лежат в Δ. Так как площадь S превышает единицу, два из этих пересечений должны перекрываться в Δ. Пусть (α, β) — координаты точки в области перекрытия. Очевидно, что $0 \leqslant \alpha, \beta \leqslant 1$. Рассмотрим точки решетки, которые являются правыми нижними углами двух квадратов, пересечения которых с S перекрываются в Δ, и пусть их координатами будут (a, b) и (c, d). Тогда $(x_1, x_2) = (\alpha + a, \beta + b)$ представляют собой координаты точки, принадлежащей пересечению S с одним из квадратов исходной конфигурации, и

$$(y_1, y_2) = (\alpha + c, \beta + d)$$

— ее координаты, так как эта точка принадлежит другому пересечению. Таким образом, обе точки содержатся в S. Очевидно, что $(y_1 - x_1)$ и $(y_2 - x_2)$ — целые числа, так как $(c - a)$ и $(d - a)$ — целые числа [91].

Теорема 2.37 (Минковский). *Выпуклое множество S на плоскости, центрально симметричное относительно начала координат O с площадью > 4, содержит точки решетки, отличные от O.*

Доказательство. Если $X = (x_1, x_2) \in S$ и $X' = (x_1/2, x_2/2)$, то X' лежит на OX и $\overline{OX'} = {}^1/_2\overline{OX}$. Если сопоставить точку X' каждой точке X, то множество S' всех X' подобно S в отношении $1:2$ и, поскольку отношение площадей пропорционально квадрату отношения линейных размеров, получаем

$$|S'| = \frac{1}{4}|S| > \frac{1}{4} \cdot 4 = 1,$$

где вертикальные черточки означают площадь.

По теореме Бликфельдта, множество S' содержит две точки $Y' = (y_1', y_2')$ и $Z' = (z_1', z_2')$, так что $(z_1' - y_1')$ и $(z_2' - y_2')$ — целые числа. Этим точкам в множестве S соответствуют точки $Y = (2y_1', 2y_2')$ и $Z = (2z_1', 2z_2')$. Ввиду симметрии множества S относительно O оно должно содержать $-Y = (-2y_1', -2y_2')$, зеркальное отображение $(2y_1', 2y_2')$ относительно O. Из выпуклости S следует, что середина отрезка $\overline{(-Y)\,Z}$ тоже принадлежит S. Координаты этой точки равны

$$\left(\frac{-2y_1' + 2z_1'}{2}, \frac{-2y_2' + 2z_2'}{2}\right) = (z_1' - y_1', z_2' - y_2').$$

Координаты точки справа — целые числа (не все равные нулю), и, следовательно, это точка решетки, лежащая в S. Поскольку S содержит также зеркальное отображение предыдущей точки, оно содержит по крайней мере две точки решетки. Этим заканчивается доказательство.

Замечание. Можно показать, что эллипс

$$\frac{(x - a/3)^2}{a^2} + \frac{y^2}{b^2} = \frac{5^{2k}}{9}, \quad k = 0, 1, \ldots,$$

проходит в точности через $2k + 1$ точек решетки, а эллипс

$$\frac{(x - a/2)^2}{a^2} + \frac{y^2}{b^2} = \frac{5^{k-1}}{4}, \quad k = 1, 2, \ldots,$$

проходит в точности через $2k$ точек решетки. Здесь a и b — разные положительные целые числа. В первом случае b должно быть простым числом вида $4\alpha + 3$ и не должно делиться на 3. Во втором случае a не должно делиться на 2, а b должно быть простым числом вида $4(2k) + 3$.

Аналогичные результаты известны и для эллипсоида [90].

2.9. Максимумы и минимумы в теории множеств

Иногда возникает необходимость оптимизации в задачах теории множеств. Приведем здесь три примера и формулировки некоторых хорошо известных теорем. Хорошо известной задачей такого вида является задача определения минимального числа носков, которое надо вытащить в темной комнате из шкафа, заполненного белыми

и черными носками, чтобы из них наверняка можно было составить пару. Ниже даются более тонкие примеры. Докажем сначала следующую теорему (см. гл. 1, в [15a]):

Теорема 2.38. *Максимальное число троек, которое можно составить из n объектов так, чтобы любые две тройки имели самое большее один общий объект, асимптотически равно $n^2/6$.*

Доказательство. Число троек не меньше чем $n(n-3)/6$. Чтобы убедиться в этом, пронумеруем объекты как $1, 2, \ldots, n$ и составим все тройки, сумма номеров которых равна $0 \pmod n$. Такие тройки имеют не более чем один общий объект, так как если число общих объектов больше двух, то третьи номера должны быть равны по модулю n и тогда тройки совпадают. Число таких троек можно оценить по полному числу решений уравнения $x + y + z \equiv 0 \pmod n$, в которых $x \not\equiv y \not\equiv z$. Величину x можно выбрать n способами, а y надо выбирать так, чтобы выполнялись следующие соотношения по модулю n: $y \not\equiv x$, $y \not\equiv z \equiv -x - y$, т. е. y выбирается так, что $x \not\equiv z$, поскольку из $x \equiv z$ следует, что $y \equiv -2x$, чего не должно быть. Поэтому y можно выбирать по меньшей мере $n - 3$ способами. Величина z определяется автоматически. Таким образом, полное число способов выбора x и y не менее чем $n(n-3)$, что дает не менее чем $n(n-3)/6$ троек, так как x, y, z можно переставлять внутри тройки шестью способами.

Число троек не превышает $n(n-1)/6$. Чтобы убедиться в этом, заметим, что в каждой тройке существуют три пары и никакие две пары во множестве всех пар не совпадают, в противном случае их тройки имели бы два общих объекта. Так как полное число пар равно C_n^2 и в каждой тройке насчитывается три пары, число троек не превышает $^1/_3 C_n^2 = n(n-1)/6$. Отсюда следует асимптотический результат.

Предположим, что каждый раз производится выбор k точек из n, что дает C_n^k подмножеств. Эти подмножества затем разбиваются на r классов $C_1, \ldots, C_r$. Если m — целое число, $m > k$, то можно ли найти m точек, из которых можно составить подмножества по k элементов в каждом, относящиеся к одному классу C_i? Получаем следующую теорему [67]:

Теорема 2.39. (Рамсей). *Существует наименьшее целое число $n_0 = n_0(k, m, r)$, такое, что если $n \geqslant n_0$, то при любом разбиении подмножеств из k точек на r классов существует множество из m точек, для которого все подмножества из k точек, относятся к одному классу.*

Функция $n_0(k, m, r)$ называется функцией Рамсея и известна только для простых классов $n_0(2, 3, 2) = 6$, $n_0(2, 4, 2) = 18$, $n_0(2, 3, 3) = 17$. (Эти идеи обобщили Эрдёш и Радо [20].)

Теорема 2.40. *Если X — множество из n точек, $n \geqslant 1$, и A_i, $i = 1, \ldots, k$, — набор подмножеств X, таких, что каждое подмножество X может быть выражено через A_i с помощью теоретико-*

множественных операций объединения, пересечения и дополнения, то минимальная величина k равна $-[-\log_2 n]$, где квадратные скобки означают наибольшее целое число, не превосходящее данного [19].

Доказательство. Эта задача эквивалентна задаче о выражении каждого подмножества определенным образом через одиночный элемент. A_i называется *носителем* $x \in X$ тогда и только тогда, когда $x \in A_i$. Так как все одиночные элементы различны, то необходимо и достаточно, чтобы каждый $x \in X$ имел отличное от других множество носителей. Теперь k множествам соответствуют 2^k множеств носителей (т. е. каждый $x \in X$ должен удовлетворять условию: если $y \in X$, то x имеет носитель, который не является носителем y), и должно выполняться неравенство $2^k \geqslant n$, откуда и следует результат.

Замечание. Имеем $\{x\} = \bigcap\limits_{x \in A_i} A_i \cap (\sim \bigcup\limits_{x \in A_j} A_j)$.

Упражнение 2.40. Докажите, что если $A_i \subset S$, $i = 1, \ldots, m$,— подмножества множества S из N элементов и каждое A_i содержит не менее чем n элементов, то некоторый элемент S должен принадлежать не менее чем mn/N подмножествам A_i.

Харари сделал следующее чрезвычайно полезное наблюдение, касающееся одного из вариантов теоремы Менгера [9]. Один из этих вариантов (Уитни) уже был ранее рассмотрен в этой главе. Пути здесь иногда означают также цепи.

В 1927 г. Менгер [9] опубликовал следующую замечательную теорему: (I) Для любых двух несмежных вершин графа максимальное число не имеющих общих вершин путей, соединяющих данные несмежные вершины, равно минимальному числу вершин, разделяющих их. Только в 1956 г. (II) был получен соответствующий результат (Фордом и Фалкерсоном) для не имеющих общих ребер путей, соединяющих пары вершин, и ребер, которые разделяют их. Другая теорема, эквивалентная теореме Менгера, была сформулирована в 1932 г.: (III) Граф является n-связным тогда и только тогда, когда каждая пара вершин соединяется не менее чем n путями, отличающимися вершинами. Теоремы (I), (II) и (III) являются следствиями теоремы Форда и Фалкерсона о максимальном потоке и минимальном разрезе. Харари указывает, что несколько других теорем, которые на первый взгляд отличаются от теоремы Менгера, на самом деле эквивалентны ей. Речь идет о следующих теоремах:

Теорема (IV) (Халл, 1935). Пусть $I = \{1, 2, \ldots, n\}$ — конечное множество индексов, S_i, $i = 1, \ldots, n$,— подмножества множества S. Существуют различные x_i, $i = 1, \ldots, n$, $x_i \in S_i$, $x_i \neq x_j$ при $i \neq j$ тогда и только тогда, когда для каждого $k = 1, \ldots, n$ и при любом выборе k различных индексов $i_1, \ldots, i_k$ среди подмножеств $S_{i_1}, \ldots, S_{i_k}$ содержатся по меньшей мере k различных элементов.

Теорема (V) (Кениг, 1931). В матрице, состоящей из нулей и единиц, минимальное число строк, которые содержат единичный элемент, равно максимальному числу единиц, которые можно выбрать так, чтобы никакие две из них не принадлежали одной строке.

Теорема (VI) (Дилуорт, 1948). В частично упорядоченном множестве E минимальное число различных цепей, которые в совокупности содержат все элементы E, равно максимальному числу элементов в подмножестве E, элементы которого попарно несравнимы, если это число конечно.

Теорема (VII) (Минти, 1960). Если ребра несепарабельного графа (в котором нет вершины, удаление которой вместе с ее ребрами делает граф несвязным) произвольно ориентированы и затем окрашены в красный, белый и голубой цвета и если одно из белых ребер названо «Джордж», чтобы его можно было отличить от остальных, то существует или цикл, в который входит ребро «Джордж» и в котором нет красных ребер, или сопряженный цикл, в который входит «Джордж» и не входят голубые ребра, такой, что направление всех белых ребер либо в цикле, либо в сопряженном цикле определяется направлением ребра «Джордж».

Теорема (VIII) (Байнеке и Харари, 1967). Если (m, n) — пара связности (число ребер, инцидентных с каждой из них) двух вершин графа, то существует $m + n$ не имеющих общих ребер путей, соединяющих данные вершины, из которых m являются путями, не имеющими общих вершин.

Харари также указывает, что другие варианты теоремы Менгера встречаются в литературе по линейному программированию и теоремам двойственности. Халл написал обзор статей, в котором имеется много вариантов теоремы (IV).

ЛИТЕРАТУРА

1. Adler C. F., An Isoperimetric Problem with an Inequality, *Am. Math. Monthly*, 52, 59 (1945).
2. Balinski M. L., On the Graph Structure of Convex Polyhedra in n-Space, *Pacific J. Math.*, **11**, № 2 (1961).

2a. Barnes E. S., Wall G. E., Some Extreme Forms Defined in Terms of Abelian Groups, *J. Aust. Math. Soc.*, 1, 47 (1959).

3. Berge C., The Theory of Graphs, Wiley, N. Y., 1962; русский перевод: Берж К., Теория графов и ее применения, ИЛ, 1962.
4. Blachman N. M., The Closest Packing of Equal Spheres in a Larger Sphere, *Am. Math. Monthly*, 70, № 5, 526 (1963).
5. Blažek J., Koman M., A Minimal Problem Concerning Complete Plane Graphs, in: «Theory of Graphs and Its Applications», Academic Press, N. Y., 1964; *Proc. Symp. Smolenice* (June 1963).
6. Blichfeldt H. F., The Minimum Value of Quadratic Forms, and the Closest Packing of Spheres, *Math. Ann.* 101, 605 (1929).
7. Болтянский В. Г., Равновеликие и равносоставленные фигуры, Гостехиздат, 1956.
8. Bricard R., Sur un Question de Géométrie Relative aux Polyèdres, *Nouvelles Ann. Math.*, 55, 331 (1896).

8a. Brown T. A., Simple Paths on Convex Polyhedra, *Pac. J. Math.*, 11, 1211 (1961).
9. Busacker R. G., Saaty T. L., Finite Graphs and Networks: An Introduction with Applications, McGraw-Hill, N. Y., 1965.
10. Chang T., Much Ado About Nothing, *Eng. News*, 27 (Oct. 1966).
11. Courant R., Robbins H., What is Mathematics?, Oxford Univ. Press, London, 1961; русский перевод: Курант Р. и Роббинс Г., Что такое математика?, Элементарный очерк идей и методов, 2-е изд., изд-во «Просвещение», 1967.
12. Coxeter H. S. M., Few L., Rogers C. A., Covering Space with Equal Spheres, *Mathematika*, 6, 147 (1959).
13. Coxeter H. S. M., Introduction to Geometry, Wiley, N. Y., 1962; русский перевод: Коксетер Г. С., Введение в геометрию, изд-во «Наука», 1966.
14. Coxeter H. S. M., Greening M. G., Graham R., Sets of Points with Given Minimum Separation, *Am. Math. Monthly*, 75, № 2 (Fev. 1968).
15. Croft H. T., Marsh D. C. B., A Max-min Problem, *Am. Math. Monthly*, 86 (Jan. 1967).
15a. de Bruijn N. G., Filling Boxes with Bricks, *Am. Math. Monthly*, 76, № 1 (1969).
16. Demir H., Maximum Area of a Region Bounded by a Closed Polygon with Given Sides, *Math. Mag.* 39, № 4, 228 (Sept. 1966).
17. Demir H., Sutcliffe A., Circle Packing, *Math. Mag.*, 427 (Nov. —Dec. 1961).
18. Dirac G. A., Some Theorems of Abstract Graphs, *Proc. London Math. Soc.*, ser. 3, 2, 69 (1952).
19. Dwinger P., Marcus D. A., A Minimum Number of Subsets, *Am. Math. Monthly*, 75, № 4, 410 (April 1968).
20. Erdös P., Rado R., A Partition Calculus in Set Theory, *Bull. Am. Math. Soc.*, 62, 427 (1956).
21. Erdös P., On Sets of Distances of n Points, *Am. Math. Monthly*, 248 (1946).
22. Erdös P., On Sets of Distances of n Points in Euclidean Space, *Magy. Tud. Akad. Mat. Kut. Int. Közl*, 5, 165 (1960).
22a. Erdös P., Hanani H., On a Limit Theorem in Combinatorial Analysis, *Publ. Math. Debrecen*, 10, 10 (1963).
23. Few L., The Shortest Path and the Shortest Road Through n Points, *Mathematika*, 2, 141 (1955).
24. Few L., Covering Space by Spheres, *Mathematika*, 3, 136 (1956).
24a. Floyd R. W., Algorithm 97, Shortest Path, *Commun. ACM*, 5, № 6, 345 (1962).
25. Ford L. R., Fulkerson D. R., Maximal Flow Through a Network, *Can. Math. J.*, 8, 399 (1956).
25a. Frank H., Frisch I. T., Communication, Transmission and Transportation Networks, Addison Wesley Publ. Co., Inc., Reading, Mass. (в печати).
26. Goldberg M., Covering by Dissected Squares, *Am. Math. Monthly*, 59, 699 (Dec. 1952).
27. Goldberg M., The Isoperimetric Problem for Polyhedra, *Tôhoku Math. J.*, 40, pt. I, 226 (Dec. 1934).
28. Goldberg M., Stewart B. M., A Dissection Problem for Sets of Polygons, *Am. Math. Monthly*, 71, № 10, 1077 (1964).
29. Goldberg M., On the Original Malfatti Problem, *Math. Mag.*, 40, № 5, 241 (Nov. 1967).
30. Golomb S. W., Polyominoes, Charles Scribner's Sons, N.Y., 1965.
31. Goodman A. W., On Sets of Acquaintances and Strangers at any Party, *Am. Math. Monthly*, 778 (Nov. 1959).
32. Grünbaum B., Convex Polytopes, Interscience Publishers, N.Y., 1967.
33. Grünbaum B., A Proof of Vazsonyi's Conjecture, *Bull. Res. Council Israel*, 6A, № 1, 77 (Oct. 1956).
34. Guy R. K., A Combinatorial Problem, *Bull. Malay. Math. Soc.*, 7, № 2, 68 (July 1960).

35. Guy R. K., Jenkyns T., Schaer J., The Toroidal Crossing Number of the Complete Graph, Univ. of Calgary (Canada), Dept. of Mathematics, Res. Paper № 18, May 1967.
36. Hall M. Jr., Combinatorial Theory, Blaisdell Publ. Co., 1967; русский перевод: Холл М., Комбинаторика, изд-во «Мир», 1970.
37. Hancock H., Development of the Minkowski Geometry of Numbers, Vols. 1 and 2, Dover Publications, Inc., N.Y., 1939.
38. Harary F., A Seminar on Graph Theory, Holt, Rinehart and Winston, Inc., N.Y., 1967.
39. Heesch H., Kienzle O., Flächenschluss, Springer-Verlag OHG, Berlin-Göttingen, Heidelberg, 1963.
40. Heppes A., On the Densest Packing of Circles not Blocking Each Other, *Stud. Sci. Math. Hung.*, 2, № 1—2, 257 (1967).
40a. Heffer A., On the Number of Spheres Which Can Hide a Given Sphere, *Can J. Math.*, 19, 413 (1967).
41. Hilbert D., Cohn-Vossen S., Geometry and the Imagination, Chelsea Publ. Co., N.Y., 1952; русский перевод: Гильберт Д., Кон-Фоссен С., Наглядная геометрия, 2-е изд., Гостехиздат, М.—Л., 1951.
42. Hildebrand F. B., Methods of Applied Mathematics, Prentice-Hall, Englewood Cliffs, N. J., 1961.
42a. Hoffman E. J., Loessi J. C., Moore R. C., Constructions for the Solution of the *m* Queens Problem, *Math. Mag.*, 66 (March 1969).
43. Jacobson R. A., Yocom K. L., Paths of Minimal Length within a Cube, *Am. Math. Monthly*, 634 (June—July 1966).
44. Jacobson R. A., Yocom K. L., Shortest Paths within Polygons, *Math. Mag.*, 290 (Nov.—Dec. 1966).
45. Jucovič E., Moon J. W., The Maximum Diameter of a Convex Polyhedron, *Math. Mag.*, 38, № 1 (1965).
46. Kainen P., On a Problem of Erdös, *J. Comb. Theory*, 5, 374 (1968).
47. Kazarinoff N. D., Geometric Inequalities, Random House, Inc., N. Y., 1961.
47a. Kelley J. B., Polynomials and Polyominoes, *Am. Math. Monthly*, 73, 464 (May 1966).
48. Kershner R. B., On Paving the Plane, *Am. Math. Monthly*, 75, № 8, 839 (Oct. 1968).
49. Китайгородский А. И., Порядок и беспорядок в мире атомов, 4-е изд., изд-во «Наука», 1966.
50. Kravitz S., Packing Cylinders into Cylindrical Containers, *Math. Mag.*, 40, № 2, 65 (1967).
50a. Kruskal J. B., Jr., On the Shortest Spanning Subtree of a Graph and the Travelling Salesman Problem, *Proc. Am. Math. Soc.*, 48 (1956).
51. Kuratowski K., Sur le problème des courbes gauches en topologie, *Fund. Math.*, 15, 271 (1930).
52. Langman H., Play Mathematics, Hafner Publ. Co., Inc., N.Y., 1962.
53. Lietzmann W., Visual Topology, American Elsevier Publ. Co., Inc., N.Y., 1965.
54. Lindsey II, J. H., Assignment of Numbers to Vertices, *Am. Math. Monthly*, 508 (May 1964).
54a. Люстерник Л. А., Кратчайшие линии, вариационные задачи, Гостехиздат, 1955.
55. Meschkowski H., Unsolved and Unsolvable Problems in Geometry, Oliver and Boyd Ltd., London, 1966.
56. Minkowski H., Geometrie der Zahlen, Chelsea Publ. Co., N.Y., 1953.
57. Moon J. W., On the Distribution of Crossings in Random Complete Graphs, *J. Soc. Ind. Appl. Math.*, 13, 506 (1965).
58. Moon J. W., Moser L., Simple Paths on Polyhedra, *Pacific J. Math.*, 13, № 2, 629 (1963).
59. Moser L., On the Different Distances Determined by *n* Points, *Am. Math. Monthly*, 59, 85 (1952).

60. Mott-Smith J., Mathematical Puzzles for Beginners and Enthusiasts, McGraw-Hill, N.Y., 1946.
60a. Murchland J. D., Shortest Distances by a Fixed Matrix Method, Transport Network Theory Unit. Publication, London, 1968.
61. Newman D. J., A Maximum Covering of Lattice Points by a Square, *Am. Math. Monthly*, 75, № 5 (May 1968).
62. Newman I., Patenaude R., Nonattacking Knights on a Chessboard, *Am. Math. Monthly*, 210 (Fev. 1964).
63. Oler N., The Slackness of Finite Packings in E_2, *Am. Math. Monthly*, 69, № 6, 511 (June—July 1962).
64. Ore O., The Four Color Problem, Academic Press, Inc., N.Y., 1967.
65. Page Y., Selfridge J. L., Another Squre-covering Problem, *Am. Math. Monthly*, 185 (Feb. 1960).
66. Pinzka C. F., Leetch J. F., Moran D. A., The Soap Contest, *Am. Math. Monthly*, 68, № 3, 295 (March 1961).
67. Ramsey F. P., On a Problem of Formal Logic, *Proc. London Math. Soc.* (2), 30, 264 (1930).
68. Rogers C. A., Packing and Covering, Cambridge Univ. Press, N.Y., 1964; русский перевод: Роджерс К., Укладки и покрытия, изд-во «Мир», 1968.
69. Saaty T. L., Remarks on the Four Color Problem, The Kempe Catastrophe, *Math. Mag.* (Jan. 1967).
70. Saaty T. L., Symmetry and the Crossing Number for Complete Graphs, *J. Nat. Bur. Standards* (April—May 1969).
71. Saaty T. L., Two Theorems on the Minimum Number of Intersections for Complete Graphs, *J. Comb. Theory*, 2, № 4, 571 (June 1967).
72. Scheid F., Some Packing Problems, *Am. Math. Monthly*, 231 (March 1960).
73. Segre B., Mahler K., On the Densest Packing of Circles, *Am. Math. Monthly*, 261 (May 1944).
74. Smalley I., Simple Regular Sphere Packings in Three Dimensions, *Math. Mag.*, 36, № 5, 295 (Nov. 1963).
75. Steinhaus H., Mathematical Snapshots, Oxford Univ. Press, Fairlawn, N.J., 1950.
76. Stover D. W., Mosaics, Houghton Mifflin Co., Boston, 1966.
77. Соминский И. С., Метод математической индукции, 6-е изд. Физматгиз, 1961.
78. Thompson D'Arcy W., On Growth and Form, Cambridge Univ. Press, N.Y., 1966.
79. Thue A., Über die dichteste Zusammenstellung von kongruenten Kreisen in einer Ebene, *Christiania Vid. Sel'sk. Skr.*, 1, 3 (1910).
80. Tóth L. F., Lagerungen in Der Ebene Auf Der Kugel Und Im Raum, Springer-Verlag OHG, Berlin, 1953; русский перевод: Тот Л. Ф., Расположения на плоскости, на сфере и в пространстве, Физматгиз, 1958.
81. Tóth L. F., On the Arrangement of Houses in a Housing Estate, *Stud. Sci. Math. Hung.*, 2, № 1—2, 37 (1967).
82. Tóth L. F., Regular Figures, Pergamon Press, N.Y., 1964.
83. Tóth L. F., New Proof of a Minimum Property of the Regular n-Gon, *Am. Math. Monthly*, 589 (1947).
84. Tóth L. F., What the Bees Know and What They Do Not Know, *Bull. Am. Math. Soc.*, 70, № 4, 468 (July 1964).
85. Trigg C. W., Mathematical Quickies, McGraw-Hill, N.Y., 1967.
86. Ungar P., Minimal Path Connecting n Points, *Am. Math. Monthly*, 57, 261 (1950).
87. Weyl H., Symmetry, Princeton Univ. Press, Princeton, N.J., 1952; русский перевод: Вейль Г., Симметрия, изд-во «Наука», 1968.
88. Whitney H., Congruent Graphs and The Connectivity of Graphs, *Am. J. Math.*, 54, 150 (1932).
89. Witgen G., Découpage du Cube, *Rev. Franc. Rech. Opération*, 7, No. 26, 92 (1963).

90. Wulczyn G., Ellipses Passing through a Prescribed Number of Lattice Points, *Am. Math. Monthly*, 75, 671 (June—July 1968).
91. Яглом А. М., Яглом И. М., Неэлементарные задачи в элементарном изложении, Гостехиздат, 1954.
92. Яглом И. М., Болтянский В. Г., Выпуклые фигуры, Гостехиздат, М.—Л., 1951.
93. Zarankiewicz K., On a Problem of P. Turán Concerning Graphs, *Fund. Math.*, 41, 137 (1954).
94. Zassenhaus H., Modern Developments in the Geometry of Numbers, *Bull. Am. Math. Soc.*, 67, № 5 (1961).
95*. Грюнбаум Б., Этюды по комбинаторной геометрии и теории выпуклых тел, изд-во «Наука», 1971.
96*. Зыков А. А., Теория конечных графов, изд-во «Наука», Новосибирск, 1969.
97*. Касселс Дж. В. С., Введение в геометрию чисел, изд-во «Мир», 1965.
98*. Оре О., Графы и их применения, изд-во «Мир», 1965.
99*. Оре О., Теория графов, изд-во «Наука», 1968.
100*. Саати Т. Л., О числе пересечений в полных графах, *Изв. АН СССР*, сер. Техническая кибернетика, № 6 (1971).
101*. Форд Л., Фалкерсон Д., Потоки в сетях, изд-во «Мир», 1966.
102*. Болтянский В. Г. и Гохберг И. П., Разбиение фигур на меньшие части, изд-во «Наука», 1971.

Глава 3

Некоторые элементарные приложения

3.1. Введение

Некоторые задачи оптимизации, требующие целочисленного решения, которые имеют практическое значение, можно сформулировать непосредственно, не обращаясь к основополагающим понятиям теории множеств и соответствующим геометрическим представлениям. В этой главе будут приведены некоторые примеры и их решения.

3.2. Теория информации

С точки зрения теории информации существование задачи означает наличие неопределенности. Полное или частичное решение задачи — это полное исключение или уменьшение неопределенности. Таким образом, сначала задачу надо рассмотреть с точки зрения количества содержащейся в ней неопределенности. Решение можно получить с помощью логических рассуждений и экспериментов, собирая информацию, которая позволяет уменьшить неопределенность (или полностью исключить ее). Количество неопределенности или энтропия (эта величина с противоположным знаком определяется как количество информации) множества из n взаимно исключающих событий задается соотношением

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log p_i, \quad \sum_{i=1}^{n} p_i = 1.$$

Таким образом, в задаче рассматриваются n различных факторов, и вероятность появления i-го фактора равна p_i. В непрерывном случае энтропия задается формулами [11]

$$H[f(x)] = -E[\log f(x)] = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx,$$

$$\int_{-\infty}^{\infty} f(x)\, dx \equiv 1, \quad \int_{-\infty}^{\infty} x f(x)\, dx = 1, \quad \int_{-\infty}^{\infty} x^2 f(x)\, dx = \sigma^2.$$

Аналогичные соотношения можно привести и для случая нескольких переменных.

Логарифм в определении энтропии появляется из требования аддитивности в том смысле, что сумма энтропий, полученных для

двух событий, равна энтропии этих событий, рассматриваемых совместно. В алгебраической форме это требование выражается в виде функционального уравнения

$$f(x) + f(y) = f(xy),$$

решением которого является $f(x) = \log x$.

Теорема 3.1. *В дискретном случае неопределенность максимальна, если $p_i = p_j$ для всех i и j.*

Доказательство. Для любой непрерывной выпуклой функции $F(x)$ справедливо соотношение

$$F\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right) \leqslant \frac{1}{n}\sum_{i=1}^{n} F(a_i),$$

где a_i, $i = 1, \ldots, n$,— положительные числа. Положив $a_i = p_i$ $F(x) = x \log x$, получим

$$F\left(\frac{1}{n}\right) = \frac{1}{n}\log\frac{1}{n} \leqslant \frac{1}{n}\sum_{i=1}^{n} p_i \log p_i = -\frac{1}{n} H(p_1, \ldots, p_n),$$

откуда следует

$$H(p_1, \ldots, p_n) \leqslant \log n \leqslant H\left(\frac{1}{n}, \ldots, \frac{1}{n}\right).$$

На этом доказательство заканчивается [8].

Замечание. Предположив непрерывность H по всем p_i, для доказательства теоремы можно использовать метод множителей Лагранжа. Однако тогда сначала надо доказать, что максимум не достигается на границе симплекса $\sum_{i=1}^{n} p_i = 1$. Затем следует положить одну или несколько вероятностей p_i равными нулю и использовать соотношение $\lim_{p_i \to 0} p_i \log p_i = 0$ и тот факт, что $-\log 1/k < -\log 1/n$ при $k < n$. Отсюда можно сделать вывод, что оптимум находится во внутренней области $0 < p_i < 1$, $i = 1, \ldots, n$, и, следовательно, его можно найти с помощью дифференцирования.

Упражнение 3.1. Проведите вышеизложенное доказательство. Докажите следующую теорему:

Теорема 3.2. *В непрерывном случае наибольшей энтропией обладает нормальное распределение.*

Задание функции должно увеличивать информацию о ее регулярности и уменьшать работу, необходимую для получения достаточного для исключения неопределенности количества информации. Поэтому очевидно, что максимум функции труднее всего найти

если она совершенно нерегулярна. Если допускается проведение ограниченного числа экспериментов, то можно рассчитывать только на уменьшение неопределенности и получение оценки для максимума среди рассмотренных значений.

3.3. Фальшивые монеты и фальшивомонетчики [1, 4, 10, 11] [1)]

Используем определение энтропии для отыскания априорной верхней границы N, где N — число монет, среди которых за определенное число взвешиваний n на рычажных весах нужно выделить фальшивую монету; установим также, легче или тяжелее фальшивая монета, чем настоящая.

Можно предположить, что с равной вероятностью любая из монет фальшивая и, следовательно, легче или тяжелее, чем все остальные монеты. Так как всего N монет, то имеется $2N$ разных возможностей. Поскольку все эти возможности равновероятны, получаем

$$p_i = \frac{1}{2N} \quad \text{при} \quad i = 1, \ldots, 2N.$$

Теперь легко вычислить, что полная энтропия равна $\log 2N$. Отметим, что энтропия была бы меньше, если предположение о фальшивости монеты не с равными вероятностями относилось к каждой монете. Чтобы выделить фальшивую монету за n взвешиваний, необходимо, чтобы энтропия взвешиваний была бы не меньше чем $\log 2N$. Если же энтропия взвешиваний меньше чем $\log 2N$, вообще говоря, фальшивая монета не может быть обнаружена.

Если на каждой чашке весов помещать равное количество монет, то в предположении, что фальшивая монета может появиться при любом таком взвешивании, получим границу величины энтропии. Затем будем улучшать эту границу.

После одного взвешивания возможны три исхода: левая чашка весов легче или тяжелее правой или они уравновешены. Каждый из трех исходов имеет определенную вероятность w_i, и количество информации, получаемое при взвешивании, максимально, если все w_i равны друг другу. Поэтому предположим, что взвешивания производятся таким образом, чтобы (приблизительно) было выполнено это условие, так что энтропия при одном взвешивании задается соотношением

$$-\sum_{i=1}^{3} w_i \log w_i = \log 3, \quad \text{где} \quad w_i = \frac{1}{3}.$$

Заметим, что число значений i равно числу возможных исходов. Энтропия при n взвешиваниях равна сумме энтропий при каждом

[1)] Достаточно полное изложение этой задачи можно найти в [15*]. — *Прим. ред.*

взвешивании:

$$\log 3 + \ldots + \log 3 = \log 3^n.$$

Итак, фальшивая монета будет обнаружена, когда энтропия взвешиваний превысит $\log 2N$ или, другими словами, $\log 2N$ должен быть меньше $\log 3^n$. Предположим, что эти две величины равны. Тогда $\log 3^n = \log 2N$, или $N = {}^1/_2 3^n$. Так как это не целое число, возьмем от него целую часть, что дает

$$N = \frac{3^n - 1}{2}.$$

Покажем теперь, используя понятие энтропии, что n взвешиваний недостаточно для обнаружения фальшивой монеты среди $(3^n - 1)/2$ монет. Заметим сначала, что N имеет вид $3M + 1$, где M — целое число. Нетрудно заметить, что наибольшее приращение энтропии достигается, если прежде всего разделить монеты на три равные группы. На каждую чашку весов помещается по M монет, так что остается $M + 1$ монет. Если чашки уравновешены, то энтропия, получаемая при оставшихся $(n - 1)$ взвешиваниях, должна быть не меньше, чем

$$\log 2 (M + 1) = \log (3^{n-1} + 1),$$

поскольку $M = (N - 1)/3$ и $N = (3^n - 1)/2$. Так как $\log (3^{n-1} + 1)$ больше, чем максимальная энтропия, получаемая при $(n - 1)$ взвешиваниях, т. е. $\log 3^{n-1}$, очевидно, что задача не может быть решена для

$$N = \frac{3^n - 1}{2}.$$

Таким образом, количество монет не может превышать

$$N = \frac{3^n - 3}{2}.$$

Можно показать, что задача действительно решается для такого количества монет. На этом простом примере можно эффективно продемонстрировать, как с помощью абстрактных рассуждений получается решение задачи в виде количественных соотношений.

Используя теорию информации, определим стратегию решения задачи в случае $N = 9$. Стратегия состоит в том, чтобы получать как можно большее количество информации при каждом взвешивании.

Пусть $p_л$, $p_п$ и $p_р$ — вероятности того, что перевесит соответственно левая или правая чашка или весы останутся в равновесии. Тогда информация, полученная при одном взвешивании, равна величине

$$H = -p_л \log p_л - p_р \log p_р - p_п \log p_п,$$

которая по предыдущей теореме максимальна, если все вероятности равны друг другу. Таким образом, наилучшая стратегия состоит

в том, чтобы проводить взвешивания так, чтобы все три альтернативы были равновероятны.

Если на левую чашку поместить k монет, на правую — тоже k монет и отложить $9 - 2k$ монет, то

$$p_{\text{р}} = \frac{9-2k}{9}, \quad p_{\text{л}} = p_{\text{п}} = \frac{k}{9}.$$

Эти три величины равны при $k = 3$. Поэтому надо на каждую чашку класть по три монеты и три откладывать. Если весы уравновешены, то фальшивая монета находится среди отложенных. Теперь таким же образом можно исследовать эти три монеты, не обращая внимания на первые шесть монет.

Если при первом взвешивании равновесия не будет, то для того, чтобы с вероятностью $^1/_3$ чашки весов уравновесились при втором взвешивании, надо с каждой чашки снять по монете. С другой стороны, чтобы уравнять вероятности того, что перевесят монеты на левой или правой чашке, следует взять по одной монете с каждой чашки и поменять их местами. Если теперь весы уравновесятся, одна из двух снятых монет фальшивая и т. д. В любом случае трех взвешиваний достаточно для того, чтобы обнаружить фальшивую монету и установить, легче или тяжелее она, чем настоящая.

Упражнение 3.2. Примените аналогичную стратегию в случае $N = 27$, используя три взвешивания для обнаружения более легкой фальшивой монеты.

Фальшивомонетчики [6]

Предположим, что N мастеров занимаются изготовлением монет одного достоинства; некоторые из них фальшивомонетчики. Все фальшивые монеты имеют один и тот же вес, но он несколько отличается от веса настоящей монеты. Каждый мастер изготавливает или только хорошие монеты, или только фальшивые. Располагая одной заведомо хорошей монетой, набором всевозможных гирь и весами, определим за три взвешивания, есть ли среди мастеров фальшивомонетчики и кто они.

Решение. При первом взвешивании определим W_g — вес настоящей монеты. Возьмем у каждого мастера по одной монете и при втором взвешивании определим полный вес T. Разность D равна $T - NW_g$. Если $D = 0$, то все мастера работают честно. Если $D \neq 0$, то берем 2^{i-1} монет от i-го мастера, $i = 1, \ldots, N$, и определяем полный вес T'. Разность теперь равна $D' = T' - (2^N - 1)\, W_g$. Находим целое число S, такое, что $D'/D = S/\beta\,(S)$, где $\beta\,(S)$ — число единиц в двоичном разложении S. Пусть $S = \sum_{i=1}^{N} B_i 2^{i-1}$, где B_i равно 0 или 1. Тогда $\beta\,(S)$ есть число фальшивомонетчиков, а i-й мастер честный или нечестный в зависимости от того, равно B_i нулю или единице.

3.4. Задача справедливого дележа (как справедливо разрезать пирог) [2, 5]

Эта задача широко известна в следующей формулировке: два человека хотят разделить справедливо пирог на две части. Решение задачи состоит в том, что один делит пирог, а другой выбирает себе один из кусков.

Для случая $n > 2$ задача была решена Банахом и Кнастером. Нож перемещают параллельно в выбранном направлении. Как только намечается доля, которая удовлетворяет хотя бы одного из игроков, этот кусок отрезают и передают ему, после чего процесс продолжается. Если несколько игроков желают получить один и тот же кусок, его разыгрывают между ними. Можно показать, что это справедливый метод. Ниже излагается еще один метод решения.

С этой задачей тесно связана и «задача о Ниле». Каждый год Нил разливается, опустошая земли древнего Египта. Ценность различных частей страны зависит от уровня воды (предполагается, что всего возможно n уровней). Как разделить страну между k наместниками, чтобы каждый получил $1/k$-ю часть земли по стоимости при любом уровне воды?

Третья задача — это задача о «бутерброде с ветчиной». Бутерброд с ветчиной, в котором на хлеб намазано масло, надо разделить ножом на две части так, чтобы все ингредиенты (т. е. хлеб, масло и ветчина) распределились поровну между обеими порциями. При помощи теории меры доказана теорема существования для решения этой задачи.

В задаче о дележе пирога требуется разделить пирог между n игроками так, что каждый, по его мнению, получает долю, не меньшую $1/n$ от всего пирога. Пронумеруем игроков как $1, 2, \ldots, n$. Пусть $m_j(p)$ — мера, которую j-й игрок приписывает куску p. Все меры взаимно независимы. Если игрок j получает долю p_j, то $m_j(p_j) \geqslant 1/n$. Предполагается, что меры аддитивны и всему пирогу приписывается мера, равная единице. (Строго говоря, надо требовать, чтобы мера была такой, чтобы среди частей куска нельзя было найти такие, мера которых равна мере всего этого куска, или чтобы им нельзя было приписать никакого значения меры.)

Рассмотрим задачу [5], в которой требуется разделить пирог на k частей между двумя игроками 1 и 2 так, что игрок 1 получает одну часть, а игрок 2 получает $k - 1$ частей. Меры соответствующих кусков p_1 и p_2 должны удовлетворять критерию справедливости, т. е.

$$m_1(p_1) \geqslant \frac{1}{k},$$

$$m_2(p_2) \geqslant \frac{k-1}{k}.$$

Лемма. *Справедливым является следующий порядок: игрок 2 делит пирог на k частей, а игрок 1 выбирает одну из них.*

Доказательство. Чтобы доказать, что этот порядок является справедливым, заметим, что игрок 2 может разделить пирог на k кусков $p_1, \ldots, p_k$ с $m_2(p_j) \geqslant 1/k$, $i = 1, \ldots, k$. Так как $\sum_{i=1}^{k} m_1(p_i) = 1$, должен найтись кусок p_j, такой, что $m_1(p_j) \geqslant 1/k$.

Обратимся теперь снова к общей задаче, в которой требуется справедливо разделить пирог между n игроками, так что если игрок j получает кусок p_j, то $m_j(p_j) \geqslant 1/n$, $j = 1, \ldots, n$.

Теорема 3.3. *Справедливым является следующий порядок дележа: на первом шаге пирог отдают игроку 1. На втором шаге игрок 2 вместе с игроком 1 делит пирог на два куска в соответствии с леммой при $k = 2$. Если куски p_1 и p_2 кажутся равными для игроков 1 и 2, то*

$$m_1(p_1) \geqslant \frac{1}{2},$$

$$m_2(p_2) \geqslant \frac{1}{2}.$$

На третьем шаге игрок 3 вместе с игроком 1 делит кусок p_1 на три части, из которых выбирает q_1, так что $m_3(q_1) \geqslant {}^1/_3 m_1(p_1)$. Совместно с игроком 2 он делит на три части и кусок p_2, из которого забирает долю q_2, такую, что $m_3(q_2) \geqslant {}^1/_3 m_2(p_2)$. Если r_1 и r_2 — остатки от кусков p_1 и p_2 соответственно, то

$$m_1(r_1) \geqslant \frac{2}{3} m_1(p_1),$$

$$m_2(r_2) \geqslant \frac{2}{3} m_2(p_2).$$

На k-м шаге игрок k отдельно с каждым из игроков 1, 2, . . ., $k - 1$ делит принадлежащие им куски. Каждый из игроков 1, 2, . . ., $k - 1$ делит свой кусок на k частей, а k-й игрок выбирает себе из каждого куска одну часть.

На n-м шаге процесс заканчивается.

Доказательство. Такой порядок дележа справедлив для игрока j. Заметим, что, так как он не принимает участия в разделах до j-го шага, достаточно показать, что на k-м шаге ($j \leqslant k \leqslant n$) раздел происходит справедливо для j, т. е. на k-м шаге игрок j получает кусок q, такой, что $m_j(q) \geqslant 1/k$. Доказательство проводится по индукции. При $j = 1$ на j-м шаге раздел является справедливым для игрока j. Если $p_1, \ldots, p_{j-1}$ — куски, полученные игроками $1, \ldots, j - 1$ к j-му шагу, то игрок j получает от каждого из них при порядке раздела, изложенном в лемме, кусок q_i, $i = 1, \ldots, j - 1$, такой, что

$$m_j(q_i) \geqslant \frac{1}{j} m_j(p_i).$$

Но

$$\sum_{i=1}^{j-1} m_j(p_i) = 1,$$

поэтому

$$\sum_{i=1}^{j-1} m_j(q_i) \geqslant \frac{1}{j}$$

и, следовательно, раздел является справедливым на $(j-1)$-м шаге. Покажем, что если на $(k-1)$-м шаге раздел справедлив для игрока j, $j \leqslant k \leqslant n$, то и на k-м шаге раздел производится справедливо. Если игрок j после $(k-1)$-го шага получает кусок p, то по предположению индукции $m_j(p) \geqslant 1/(k-1)$. Так как игрок j делит куски с каждым из ранее игравших участников, то в соответствии с леммой ему гарантировано получение куска q, такого, что $m_j(q) \geqslant \geqslant [(k-1)/k]\, m_j(p) \geqslant 1/k$. Таким образом, порядок дележа является справедливым на любом шаге k, и, в частности, при $k = n$ выполняется условие $m_j(p) \geqslant 1/n$.

3.5. Количество тестов, метод исчерпания

Задача [7]. Рассмотрим тесты, включающие n вопросов, $n \geqslant 10$, в каждом из которых надо сделать выбор между k ответами, $k \geqslant 1$. Сколько можно построить подобных тестов так, чтобы количество набранных очков (оно задается формулой $[(r - w)/(k-1)]\, 100/n$, где r — число правильных ответов, а w — число неправильных ответов) было целым числом?

Решение.

1. $r + w < n$, т. е. не на все вопросы получен ответ. Здесь $|r - w| \leqslant n$ и $|r - w|$ принимают целые значения, не меньшие 0, но и не большие n. Величины $|100(r - w)|$ также принимают значения 0, 100, 200, ... и т. д., не превышающие $100n$, с ограничением $100n \geqslant 1000$. Число $(k-1)n$ должно быть делителем всех этих величин. В частности, оно должно быть делителем 100, т. е., так как $n \geqslant 10$, оно должно принимать значения 10, 20, 25, 50 или 100. Таким образом, из условий

$$(k-1)n = 10, \quad (k-1)n = 20, \quad (k-1)n = 25,$$
$$(k-1)n = 50, \quad (k-1)n = 100$$

получаем следующие возможные тесты:

(n, k): (10, 2), (10, 3), (10, 6), (10, 11), (20, 2), (20, 6), (25, 2), (25, 3), (25, 5), (50, 2), (50, 3), (100, 2).

2. Если на все вопросы получен ответ, то $r + w = n$ и в выражении для подсчета очков можно заменить w на $(n - r)$. Если известно,

что n — четное число, то $(2r - n)$ — тоже четное и, следовательно, $(k - 1)\,n$ должно быть делителем 200. В этом случае дополнительно возможны тесты:

(n, k): (10, 5), (10, 21), (20, 3), (20, 11), (40, 2), (40, 6), (50, 5), (100, 3), (200, 2).

3.6. Задача о джипе [1] [3, 11]

Здесь речь пойдет о задаче минимизации функции, на которую наложены ограничения, задаваемые в виде разностных уравнений.

Задача. Требуется проехать в автомобиле 1000 километров по пустыне с минимальным расходом топлива. Бак автомобиля может вместить самое большее 500 единиц топлива. Расход топлива составляет одну единицу на километр. Автомобилист должен сам постепенно устраивать промежуточные склады, используя топливо из собственного бака. Определим расположение складов на пути, при котором минимизируется расход топлива на все путешествие, число рейсов между каждой парой складов и минимальное количество топлива, которое ему надо взять со старта.

Формулировка и решение. Пусть s_i — количество топлива, заготовленного на i-м складе ($i = 0, 1, \ldots, n$), d_i — расстояние между $(i - 1)$-м и i-м складами и k_i — число рейсов между этими двумя точками. Тогда

$$s_{i-1} = s_i + 2k_i d_i + d_i, \quad i = 1, \ldots, n. \tag{3.1}$$

Таким образом, количество топлива, заготовленного на $(i - 1)$-м складе, равно сумме количества топлива на i-м складе, количества топлива, затраченного во время рейсов k_i между этими складами, и топлива еще на один рейс.

Нетрудно видеть, что используется минимальное количество топлива, если перед каждым рейсом автомобиль полностью загружает свой бак. В этом случае сокращается общее число рейсов и, следовательно, уменьшается расход топлива на все путешествие. Кроме того, чтобы в конце пути топливо не оставалось, необходимо, чтобы автомобилист забрал 500 единиц топлива на 500-километровой отметке. Из этих двух условий следует, что i-й склад надо располагать таким образом, чтобы автомобиль делал k_{i+1} рейсов между i-м и $(i + 1)$-м складами в обе стороны и один рейс вперед всегда полностью загруженным и не оставлял после себя топливо на i-м складе. Решая задачу с конца, придем к выводу, что последнее утверждение справедливо для всех складов, включая первый, но не распространяется на исходную точку, так как ее положение задано заранее. Следовательно, автомобиль сделает последний рейс между исходной

[1] Джип — одна из распространенных марок американских автомобилей.— *Прим. ред.*

точкой и первым складом с загрузкой $c \leqslant 500$. Таким образом, получаем

$$s_i = k_i\,(500 - 2d_i) + 500 - d_i, \quad i = 2, \ldots, n,$$
$$s_1 = k_1\,(500 - 2d_1) + c - d_1. \tag{3.2}$$

Требуется минимизировать s_0, заданное соотношением

$$s_0 = s_1 + 2k_1d_1 + d_1.$$

Теперь, используя для s_1 данное ранее значение, получим

$$s_0 = 500k_1 + c. \tag{3.3}$$

Так как автомобиль может проехать последние 500 км, не нуждаясь в складах топлива, то, чтобы минимизировать s_0, достаточно положить $s_n = 500$ и разместить склады вдоль 500-километрового пути. Таким образом,

$$\sum_{i=1}^{n} d_i = 500.$$

Теперь из (3.2) получим

$$d_i = \frac{500k_i + 500 - s_i}{2k_i + 1}, \quad i = 2, \ldots, n.$$

Заменив в формуле (3.1) i на $i + 1$ и подставив полученный результат вместо s_i в выражение для d_i, получим

$$d_i = \frac{500k_i + 500 - s_{i+1} - 2k_{i+1}d_{i+1} - d_{i+1}}{2k_i + 1}, \quad i = 2, \cdots, n, \; k_{n+1} \equiv 0.$$

Наконец, заменив s_{i+1} на равное выражение из (3.2) и упрощая, получаем

$$d_i = \frac{500\,(k_i - k_{i+1})}{2k_i + 1}, \quad i = 2, \ldots, n, \; k_{n+1} \equiv 0.$$

Аналогично

$$d_1 = \frac{500\,(k_1 - k_2) + c - 500}{2k_1 + 1}.$$

Так как $d_i > 0$, то $k_i > k_{i+1}$. Таким образом, нужно минимизировать

$$s_0 = 500k_1 + c$$

при ограничении

$$\sum_{i=1}^{n} d_i = 500 \sum_{i=1}^{n} \frac{k_i - k_{i+1}}{2k_i + 1} + \frac{c - 500}{2k_1 + 1} = 500.$$

Его можно переписать в виде

$$\sum_{i=1}^{n} \frac{k_i - k_{i+1}}{2k_i + 1} - \frac{1 - c/500}{2k_1 + 1} = 1, \quad k_{n+1} \equiv 0.$$

Поскольку $k_1 > k_2 > \ldots > k_i > \ldots > k_n$, второй член в левой части меньше наименьшего значения, которое может принимать любой член в сумме, т. е.

$$\frac{k_i - k_{i+1}}{2k_i + 1} > \frac{1}{2k_1 + 1}, \quad i = 2, \ldots, n.$$

Короче говоря, это соотношение определяет выбор k_i, таких, что

$$\sum_{i=1}^{n} \frac{k_i - k_{i+1}}{2k_i + 1}$$

превосходит единицу и k_1 принимает минимальное значение.

Покажем, что минимальное значение k_1 получается, если положить $k_i - k_{i+1} = 1$ $(i = 1, \ldots, n)$ и $k_n = 1$. Предположим, что для $i = i_0$ величина $k_{i_0} - k_{i_0+1} = m$ — целое число, большее 1. Тогда для i_0-го члена получаем

$$\frac{k_{i_0} - k_{i_0+1}}{2k_{i_0} + 1} = \frac{m}{2(k_{i_0+1} + m) + 1},$$

где знаменатель превращается из $2k_{i_0+1} + 1$ в $2(k_{i_0+1} + m) + 1$, когда i уменьшается от $(i_0 + 1)$ до i_0.

Теперь, приняв $k_{i_0} - k_{i_0+1} = 1$, этот скачок можно заменить на постепенно возрастающую сумму членов

$$\frac{1}{2(k_{i_0+1} + 1) + 1} + \frac{1}{2(k_{i_0+1} + 2) + 1} + \ldots + \frac{1}{2(k_{i_0+1} + m) + 1}.$$

Эта сумма больше, чем предыдущее выражение, которое в m раз превышает наименьший член в сумме. Следовательно, минимум k_1 достигается при $k_i - k_{i+1} = 1$, т. е. при использовании единичных разностей и $k_n = 1$. Ввиду свойства монотонности k_i число n выбирается таким, что $k_n = 1$, $k_{n-1} = 2$, ... и в конце концов получаем k_2 из условия

$$\sum_{i=2}^{n} \frac{k_i - k_{i+1}}{2k_i + 1} < 1 < \sum_{i=1}^{n} \frac{k_i - k_{i+1}}{2k_i + 1}.$$

В этом случае $k_2 = 6$. Затем вычисляются соответствующие d_i, начиная с 500-километровой отметки. Оставшееся расстояние принимается за d_1. Значение k_1 полагается равным семи. Такой выбор k_1 минимизирует s_0, и соответствующий выбор d_i удовлетворяет ограничению на расстояние.

Заметим, что если исходная полная разность выбрана соответствующим образом, то d_{n-i} образуют гармонические ряды с нечетными знаменателями.

То обстоятельство, что эти ряды расходятся весьма медленно, проливает некоторый свет на экономические трудности организации торговли со слаборазвитыми странами.

Упражнение 3.3. Покажите, что $k_i = 8 - i$, $d_i = 500/(17 - 2i)$, $i = 1, \ldots, 7$ и $s_0 = 38\,336{,}45$.

3.7. Задача о кокосовых орехах (гл. 1) [9]

Обозначим первоначальное количество кокосовых орехов через X, и пусть i-й матрос берет x_i штук (после m-го раздела каждый получает x_m штук). Тогда

$$x_1 = \frac{X - d_1}{n}, \tag{3.4}$$

$$x_i = \frac{(n-1)\,x_{i-1} - d_i}{n}, \quad i = 2, \ldots, m. \tag{3.5}$$

Заметим, что d_i при $i > m$ не влияют на решение. Удобно положить их k-ю разность равной нулю, т. е. $\Delta^k d_i = 0$, $k > m$. Заметим также, что $d_i \geqslant 0$, $i = 1, \ldots, m$, и, следовательно, $x_i \leqslant x_{i-1}$. Таким образом, если x_m положительное, то x_i при $i < m$ и X тоже положительные.

Если обозначить

$$\Delta x_i = x_{i+1} - x_i,$$

то уравнение (3.5) можно переписать в виде

$$n\Delta x_i + x_i = -d_{i+1}, \quad i = 1, 2, \ldots, m. \tag{3.6}$$

Частное решение этого уравнения имеет вид

$$P_i = -\sum_{k=0}^{m-1} (-n)^k \Delta^k d_{i+1}.$$

Используя формулу Ньютона, получаем

$$d_{i+1} = \sum_{j=0}^{m-1} \Delta^j d_1 C_i^j,$$

где

C_i^j — число сочетаний из i по j.

Используя тот факт, что $\Delta^k C_i^j = C_i^{j-k}$ и $C_i^j = 0$, $j < 0$, получаем после подстановки и изменения порядка суммирования

$$P_i = -\sum_{j=0}^{m-1} \sum_{k=0}^{j} (-n)^k \Delta^j d_1 C_i^{j-k}. \tag{3.7}$$

Решение уравнения (3.6) представляет собой сумму общего решения однородного уравнения и частного решения P_i. Получаем

$$x_i = \frac{X - d_1 - nP_1}{n-1} \left(\frac{n-1}{n}\right)^i + P_i, \quad i = 1, \ldots, m.$$

Теперь, чтобы определить X, заметим, что x_m должно быть положительным целым числом r. Полагая $i = m$, приравнивая к r и решая относительно X, получаем

$$X = \frac{(r - P_m)\, n^m}{(n-1)^{m-1}} + d_1 + nP_1. \tag{3.8}$$

Последние два члена — целые числа. Так как n и $(n-1)$ взаимно просты, X будет целым числом тогда и только тогда, когда для некоторого целого s имеет место соотношение

$$r - P_m = s\,(n-1)^{m-1}. \tag{3.9}$$

Чтобы это X было наименьшим положительным целым числом, r должно быть наименьшим и, следовательно, s должно быть выбрано так, чтобы r было наименьшим. Таким образом, получаем

$$s = \left[\frac{-P_m}{(n-1)^{m-1}}\right] + 1. \tag{3.10}$$

Это дает

$$x_m = s\,(n-1)^{m-1} + P_m,$$

и при помощи исходного уравнения получаем

$$x_{m-k} = s\,(n-1)^{m-1-k}\, n^k + P_{m-k}.$$

Теперь, подставляя (3.10) и (3.9) в (3.8), вычисляя P_i при $i = 1$ и $i = m$ и подставляя полученные выражения также в (3.8), находим

$$X = \left\{\left[\frac{\sum_{j=0}^{m-1}\sum_{k=0}^{j}(-n)^k\,\Delta^j d_1 C_m^{j-k}}{(n-1)^{m-1}}\right] + 1\right\} n^m + (1-n)\sum_{j=0}^{m-1}(-n)^j\,\Delta^j d_1.$$

Чтобы упростить вычисления при больших n и m, можно вычислить наименьшее целое C, для которого $X > 0$, $x_i > 0$, $i = 1, \ldots, m$, и написать

$$X = Cn^m + (1-n)\sum_{j=0}^{m-1}(-n)^j\,\Delta^j d_1.$$

Это соотношение справедливо, даже если $d_i < 0$ при некотором i.

Упражнение 3.4.

1. Найдите X, если $d_i = d$, $i = 1, \ldots, m$.
2. Пусть $n = 5$, $m = 4$, $d_1 = 4$, $d_2 = 3$, $d_3 = 1$, $d_4 = 1$. Покажите, что $X = 314$.

Упражнение 3.5. Аквариум вмещает n единиц воды. В течение недели одна единица воды испаряется и надо подливать свежую воду. Поскольку свежая вода содержит некоторое количество солей, во избежание увеличения в аквариуме концентрации соли до уровня, опасного для рыбы, из него выливают одну единицу воды [остается

(n — 2) единиц] и затем добавляют две единицы свежей воды. Докажите, что в стационарном состоянии максимальная концентрация соли на единицу воды в аквариуме вдвое превышает концентрацию соли в свежей воде. (Указание: воспользуйтесь моделью разностных уравнений.)

3.8. Задача о неограниченном сверху максимуме [12]

Пусть даны k пластинок домино, имеющих форму параллелепипеда. Предположим, что эти пластинки положены в точности одна на другую. Если теперь по очереди параллельно сдвигать пластинки, то получится конфигурация, похожая на лестницу. Требуется определить, насколько можно сдвигать пластинки, не уронив их, чтобы горизонтальная проекция этой конфигурации была как можно длиннее. Примем длину одной пластинки в рассматриваемом направлении за единицу и определим максимальную длину проекции как функцию числа пластинок.

Решение. Рассмотрим k-ю сверху пластинку. Центр тяжести множества из (k — 1) пластинок, расположенных выше ее, по отношению к (k — 1)-й пластинке вычисляется следующим образом. В центре тяжести системы из (k — 2) пластинок, расположенном над одним из ребер (k — 1)-й пластинки, приложен вес, который уравновешен в центре тяжести системы из (k — 1) пластинок весом (k — 1)-й пластинки, приложенном в ее центре тяжести. Таким образом, если расстояние от центра тяжести (k — 2) верхних пластинок, который располагается над одним из ребер (k — 1)-й пластинки, до центра тяжести системы из (k — 1) пластинок обозначить через x и измерить расстояние от центра тяжести системы из (k — 1) пластинок до центра тяжести (k — 1)-й пластинки, то получится уравнение $(k-2)\,x = 1\cdot(0{,}5 - x)$. Это дает условие $x = 1/[2\,(k-1)]$, $k > 1$. Первая сверху пластинка выдвигается над второй на половину ее длины; вторая — на четверть ее длины над третьей, третья — на $^1/_6$ ее длины над четвертой и т. д. В результате получается сумма

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\ldots+\frac{1}{2(k-1)}=\frac{1}{2}\left(1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{k-1}\right).$$

Величина в правой части определяет максимальное расстояние от края k-й пластинки. Так как сумма справа при $k \to \infty$ представляет собой расходящийся гармонический ряд, длина проекции не ограничена сверху и, следовательно, не имеет максимума.

3.9. Существование выигрывающей стратегии [14]

Рассмотрим кризисную ситуацию между двумя нациями A и B и предположим, что кризис требуется разрешить путем переговоров. Предположим также, что в этой конкретной ситуации невозможна

ничья, что предложения или действия стороны совершают по очереди и что число раундов переговоров n ограничено (т. е. что переговоры имеют конец). Под *стратегией нации* мы подразумеваем последовательность действий от начала до конца, каждое из которых может зависеть от действий другой стороны. Выигрывающая стратегия нации — это стратегия, которая в конце концов приводит к победе независимо от действий другой стороны. Докажем следующую теорему:

Теорема 3.4. *Существует или стратегия, обеспечивающая победу стороне A, или стратегия, обеспечивающая победу стороне B.*

Доказательство. Из того, что существует конечная последовательность действий, легко видеть, что одна из сторон имеет выигрывающую стратегию, определяемую первым же действием. Первое действие совершается в выигрывающей позиции, для которой характерно то, что при правильной игре сторона побеждает независимо от действий противника.

Прежде всего отметим, что условие *отсутствия ничьей* означает, что одна сторона должна победить (независимо от того, кто начинает). Если одна из сторон должна победить где-то по ходу игры, то она находится в выигрышном положении. Рассмотрим случай, когда одна из сторон, например A, получила выигрышное положение; тогда действие стороны B на предыдущем шаге должно быть ошибочным, иначе сторона A должна была получить выигрышное положение раньше. Таким образом, если A выигрывает, сторона B должна была совершить ошибку. С другой стороны, если B не ошибается, то A не может победить. Точно так же, если A не ошибается, то B не может победить. Таким образом, ни одна из сторон не может победить, если другая не совершит ошибку. Но это противоречит тому, что игра должна закончиться через определенное время и ничья невозможна. Таким образом, одна из сторон должна с самого начала иметь выигрывающую стратегию. Этим завершается доказательство.

3.10. Игральный столик [13]

Задача. На столике имеются N квадратов, на которых сделаны надписи: «выигрыш 2 : 1», «выигрыш 3 : 1»,... и «выигрыш $(N+1)$: 1». На каждом квадрате помещается некоторая сумма денег, и затем наудачу выбирается один из квадратов. Игроку выплачивается сумма, равная ставке на выигрышном квадрате и выигрышу, который пропорционален числу, проставленному на этом квадрате. Игрок проигрывает суммы, помещенные на других квадратах. Чему равно максимальное N, при котором игрок еще может располагать деньги так, что в итоге он ни при каком исходе не останется в убытке?

Решение. Если S_k — сумма, помещенная на k-й квадрат и T — полная сумма, поставленная на игру, то игрок не окажется в про-

игрыше тогда и только тогда, когда $(k+2)\,S_k \geqslant T$ или $1/(k+2) \leqslant S_k/T$ $(k = 1, 2, \ldots, N)$. Суммирование по k справа и слева дает $\sum_{k=1}^{N} 1/(k+2) \leqslant 1$. Максимальное N, при котором это возможно, равно 4. Игрок может делать ставки так, что чистый выигрыш не будет зависеть от выбранного квадрата.

ЛИТЕРАТУРА

1. Eves D., Schell E. D., Rosenbaum J., The Extended Coin Problem, *Am. Math. Monthly*, 46 (Jan. 1947).
2. Dubins L. E., Spanier E. H., How to Cut a Cake Fairly, *Am. Math. Monthly*, 1 (Jan. 1961).
3. Fine N. J., The Jeep Problem, *Am. Math. Monthly*, 24 (Jan. 1947).
4. Fine N. J., The Generalized Coin Problem, *Am. Math. Monthly*, 489 (Oct. 1947); *Math. Gazette*, 227 (1945); 231 (1946).
5. Fink A. M., A Note on the Fair Division Problem, *Math. Mag.*, 341 (Nov.—Dec. 1964).
6. Ford L. R., Jr., Baun J., The Counterfeiters of Lower Slobovia, *Am. Math. Monthly*, 61, 472 (Sept. 1954).
7. Howell J. M., Sevier F. A. C., Integral Test Scores, *Math. Mag.*, 224 (March—April, 1957).
8. Хинчин А. Я., Понятие энтропии в теории вероятностей, *УМН*, VIII, вып. 3 (1953).
   Хинчин А. Я., Об основных теоремах теории информации, *УМН*, XI, вып. 1 (1956).
9. Kirchner R. B., The Generalized Coconut Problem, *Am. Math. Monthly*, 516 (June—Jule 1960).
10. Raisbeck G., Information Theory, The M. I. T. Press, Cambridge, Mass., 1963.
11. Saaty T. L., Mathematical Methods of Operations Research, McGraw-Hill, N.Y., 1959; русский перевод: Саати Т. Л., Математические методы исследования операций, Воениздат, 1963.
12. Sharp R. T., Piled Dominos, *Pi Mu Epsilon J.*, 322 (April 1953); 411 (April 1954).
13. Van Voohris W. R., Pinzka C. F., A Gambling Table, *Am. Math. Monthly*, 61, 474 (1964).
14. Westwick R., McWorter W. A., Quiring D., Games with a Winning Strategy, *Am. Math. Monthly*, 604 (May 1967).

15*. Гарднер М., Математические головоломки и развлечения, изд-во «Мир», 1971.

16*. Яглом А. М., Яглом И. М., Вероятность и информация, Физматгиз, 1960.

Глава 4

Оптимизация при диофантовых ограничениях

4.1. Введение

Диофантовым уравнением называется полиномиальное уравнение с несколькими переменными с рациональными коэффициентами (например, $6x + 4y + z = 40$), которое требуется решить в целых (часто положительных) числах. Заметим, что уравнение с рациональными коэффициентами эквивалентно уравнению с целыми коэффициентами. Например, уравнение

$$\frac{6}{5}x^2 + \frac{4}{3}y + z = \frac{1}{2}$$

можно переписать в виде

$$36x^2 + 40y + 30z = 15.$$

Уравнение, для которого требуется отыскать рациональное решение, тоже иногда называется диофантовым уравнением.

Иногда это понятие распространяется и на уравнения более общего вида. Система диофантовых уравнений представляет собой систему уравнений с несколькими переменными с рациональными коэффициентами, для которых требуется одновременно найти целочисленные (или рациональные) решения. Целочисленное решение, в котором наибольший общий делитель значений переменных равен единице, называется *примитивным решением.*

Решение линейного диофантова уравнения тесно связано с задачей отыскания числа возможных способов разбиения положительного целого числа N на слагаемые, принадлежащие множеству S, элементами которого являются положительные целые числа $a_1, a_2, \ldots, a_n$, т. е. это задача отыскания числа целочисленных решений уравнения

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = N, \quad x_j \geqslant 0, \quad j = 1, \ldots, n.$$

Часто диофантово уравнение или система таких уравнений

$$g_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, m,$$

могут определять множество ограничений в задачах оптимизации. Требование, чтобы все переменные были неотрицательными, т. е. $x_j \geqslant 0$, $j = 1, \ldots, n$, часто является существенным ограничением. Максимизация или минимизация ограниченной функции

$f(x_1, \ldots, x_n)$, определенной над множеством дискретных решений такой системы ограничений, часто является трудной задачей. Естественно, что такая задача неразрешима, если ограничения не допускают решения. Некоторое представление о решении можно получить, если рассмотреть вместо поставленной задачи соответствующую непрерывную задачу и использовать теорему об оптимизации полунепрерывных функций над компактными множествами. Как было показано в гл. 2, в некоторых задачах оптимизации требуется, чтобы целочисленное решение уравнения, которое является ограничением, давало максимальное (или минимальное) значение $f(x_1, \ldots, x_n)$, которое тоже должно быть целым числом. Требование, что f должно принимать целочисленные значения, является дополнительным осложнением. Очевидно это не так, если f — полиномиальная форма с целыми коэффициентами.

Вообще говоря, системы ограничений с трудом поддаются решению. Если ограничения линейные, значительную информацию можно получить, рассматривая вершины соответствующего выпуклого многогранника.

Существует немного общих теорем, которые относились бы сразу к большим классам диофантовых уравнений. Вовсе не обязательно, чтобы данное алгебраическое уравнение, которое требуется решить в целых числах, имело такое решение. Например, проблема существования решения в положительных целых числах хорошо известного уравнения Ферма $x^n + y^n = z^n$ при $n > 2$ окончательно еще не решена.

Среди двадцати нерешенных проблем Гильберта, предложенных Международному математическому конгрессу в Париже в 1901 г., была следующая нерешенная до сих пор проблема [1]).

Десятая проблема Гильберта. Найти алгоритм, при помощи которого можно после конечного числа операций определить, имеет ли данное диофантово уравнение целочисленное решение.

Некоторые диофантовы уравнения были классифицированы, и исследованы их решения. Если такие уравнения появляются в виде ограничений в оптимизационной задаче, эту информацию можно использовать на первом шаге в процессе оптимизации. Рассмотрим теперь кратко некоторые методы решения элементарных диофантовых уравнений и приведем примеры оптимизации функции, принимающей дискретные значения, на которую наложены простые диофантовы ограничения. В гл. 5 излагаются общие методы линейного программирования, используемые при решении задач линейной оптимизации. В этой главе будут даны некоторые примеры нелинейных задач, хотя для них до сих пор общей теории не существует.

---

[1]) Десятая проблема Гильберта решена советским математиком Ю. В. Матиясевичем. Он доказал, что требуемого в проблеме алгоритма не существует [38].

4.2. О разрешимости диофантовых уравнений

Существует ли алгоритм, о котором говорится в проблеме Гильберта, не известно, но, как мы увидим позже, такой алгоритм может существовать для специальных классов диофантовых уравнений.

Диофантовы уравнения могут не иметь решения. Рассмотрим, например, уравнение $x^2 - 3y^2 = 17$. Заметим, что всякое целое число x представляется в виде $3n$, $3n \pm 1$. Подстановка их в уравнение дает соответственно

$$3(3n^2 - y^2) = 17,$$
$$3(3n^2 \pm 2n - y^2) = 16.$$

Легко видеть, что эти уравнения невозможно разрешить в целых числах. Например, в первом уравнении при любом выборе целых чисел n и y левая часть делится на 3, а правая нет. Аналогично уравнение

$$x^4 + y^4 + z^4 - x^4y^2 - y^4z^2 - z^4x^2 \pm x^2y^2z^2 = 0$$

нельзя разрешить в целых числах. Так, если принять, что x, y, z не имеют общих делителей, они не могут все иметь вид $3n$. Но если все или некоторые x, y, z имеют вид $3n \pm 1$, то левая часть уравнения сравнима с $\pm 1 \bmod 3$.

Кроме того, диофантово уравнение может иметь только конечное число решений. Например, уравнение пятой степени

$$x^5 + x - 1 = y^2$$

имеет только два решения: $x = 1$, $y = \pm 1$.

Упражнение 4.1. Найдите все целочисленные решения уравнения $\sum_{i=1}^{n} x_i^3 = 100$, $1 \leqslant n \leqslant 100$.

Диофантовы уравнения могут иметь бесконечно много решений. Хорошо известным уравнением с бесконечным числом решений в положительных целых числах является уравнение Пелля

$$x^2 - Ay^2 = 1,$$

где $A > 0$ и A не является полным квадратом. Если (x_1, y_1) и (x_2, y_2) — два решения этого уравнения, то из тождеств

$$(x_1^2 - Ay_1^2)(x_2^2 - Ay_2^2) = (x_1x_2 + Ay_1y_2)^2 - A(x_1y_2 + x_2y_1)^2 = (x_1x_2 - Ay_1y_2)^2 - A(x_1y_2 - x_2y_1)^2$$

заключаем, что $(x_1x_2 + Ay_1y_2,\ x_1y_2 + x_2y_1)$ также является решением, а $(x_1x_2 - Ay_1y_2,\ x_1y_2 - x_2y_1)$ — еще одно решение. Можно продолжить получение новых решений таким способом.

Если записать уравнение Пелля в виде

$$(x + y\sqrt{A})(x - y\sqrt{A}) = 1,$$

то получим

$$(x+y\sqrt{A})^n(x-y\sqrt{A})^n=1$$

для любого n, в частности для целочисленных n. Можно записать

$$(x+y\sqrt{A})^n=X_n+Y_n\sqrt{A}.$$

Здесь в разложении левой части сгруппированы члены X_n, в которые не входит $\sqrt{A}$, и члены, в которые входит $\sqrt{A}$; коэффициент при $\sqrt{A}$ обозначен через Y_n. Запишем также

$$(x-y\sqrt{A})^n=X_n-Y_n\sqrt{A},$$

следовательно,

$$(X_n+Y_n\sqrt{A})(X_n-Y_n\sqrt{A})=1.$$

Таким образом, если x и y — решения, то X_n и Y_n — тоже решения.

Упражнение 4.2. Выразите X_n через x и y. Проделайте то же для Y_n.

Можно показать, что, отправляясь от фундаментального решения (x_1, y_1), где x_1 и y_1 — ненулевые целые числа, для которых $x+y\sqrt{A}$ принимает наименьшее значение, можно получить бесконечное число решений при помощи рекуррентных соотношений

$$x_{n+1}=x_1x_n+Ay_1y_n,$$
$$y_{n+1}=y_1x_n+x_1y_n,$$

которые можно переписать в матричной форме

$$\begin{bmatrix}x_{n+1}\\y_{n+1}\end{bmatrix}=\begin{bmatrix}x_1 & Ay_1\\y_1 & x_1\end{bmatrix}\begin{bmatrix}x_n\\y_n\end{bmatrix}=\begin{bmatrix}x_1 & Ay_1\\y_1 & x_1\end{bmatrix}^{n+1}\begin{bmatrix}1\\0\end{bmatrix}.$$

Доказано, что фундаментальное решение всегда существует, но приводить здесь доказательство мы не будем. Это доказательство не позволяет конкретно строить фундаментальные решения.

Упражнение 4.3. Проверьте следующие фундаментальные решения для указанных значений A:

A	(x_1, y_1)
2	(3, 2)
3	(2, 1)
5	(9, 4)
6	(5, 2)

Интересно отметить, что если $A=46$, то $(x, y)=(24\,335, 3588)$. Такое решение трудно получить угадыванием.

Не все уравнения вида $x^2-Ay^2=-1$ можно разрешить. Например, $x^2-3y^2=-1$ не имеет решения.

Если уравнение с одной неизвестной и с целыми коэффициентами

$$a_nx^n+a_{n-1}x^{n-1}+\ldots+a_1x+a_0=0$$

имеет целое решение, то это решение должно быть делителем a_0. Таким образом, один из путей отыскания целочисленного решения состоит в том, чтобы найти все делители a_0 и подстановкой проверить, какие из них удовлетворяют уравнению. Если ни один делитель не подходит, то уравнение не имеет целочисленного решения. Если найдется такой делитель, то он, очевидно, и будет решением.

Очевидно, что любое диофантово неравенство можно представить как диофантово уравнение путем добавления неотрицательной вспомогательной переменной. С другой стороны, каждое диофантово уравнение можно представить как два диофантовых неравенства, т. е. записать уравнение $f = g$ в виде $f \geqslant g$ и $g \geqslant f$. Таким образом, проблемы существования и решения систем диофантовых неравенств или уравнений тесно связаны между собой.

Ниже приводятся примеры того, как диофантовы уравнения возникают при формулировке и решении задач.

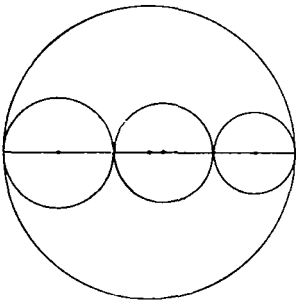
Задача. Оптовый торговец получает от изготовителя радиоприемники в коробках двух типов, в которых содержится разное количество радиоприемников. Клерк, ведающий отправкой, обнаружил, что, правильно подбирая количество коробок какого-то одного или обоих типов, он может выполнить почти любой заказ, не вскрывая коробки. Возможны были в точности шесть вариантов заказов, для выполнения которых приходилось открывать коробки [21а].

С некоторого времени изготовитель перестал использовать более маленькие коробки и перешел на упаковку нового размера. Клерк подсчитал, что в этом случае возможны десять видов заказов, для выполнения которых надо распаковывать коробки. Сколько радиоприемников содержится в коробке нового размера?

Решение. Если $(m, n) = 1$, то количество положительных чисел, которые не могут быть представлены в виде $Am + Bn$, где A и B — неотрицательные целые числа, равно $(m - 1)(n - 1)/2$. Кроме того, значение каждого такого целого числа меньше, чем $(m - 1)(n - 1)$.

Пусть a — количество радиоприемников в маленькой коробке, b — количество радиоприемников в большой коробке, c — количество радиоприемников в коробке нового типа, которой заменили меньшую коробку. Получаем $(a - 1)(b - 1)/2 = 6$ и $(c - 1) \times (b - 1)/2 = 10$. Следовательно, $5(a - 1) = 3(c - 1)$, или $5a - 3c = 2$. Только одно решение дает соответствующее целое значение b, которое является целым положительным числом, большим a. Это единственное решение имеет вид $a = 4$, $c = 6$, $b = 5$. Таким образом, в коробку нового типа входит шесть радиоприемников. При использовании коробок исходного типа с четырьмя и пятью приемниками клерк не мог отправлять, не распаковывая коробки, 1, 2, 3, 6, 7 или 11 радиоприемников. При использовании коробок с пятью и шестью радиоприемниками он не мог отправлять, не распаковывая коробки, 1, 2, 3, 4, 7, 8, 9, 13, 14 и 19 радиоприемников.

Задача. Рассмотрим фиг. 4.1, на которой изображены три круга с разными целочисленными диаметрами [18]. В сумме эти три диаметра равны диаметру большого круга, на котором расположены все три центра. Найдите диаметры трех меньших кругов при условии, что их площади в сумме составляют половину площади большого круга, диаметр которого равен $2r$.



Фиг. 4.1.

Решение. Пусть x, y, z — искомые диаметры, тогда

$$x + y + z = 2r,$$

$$x^2 + y^2 + z^2 = \frac{(x+y+z)^2}{2}.$$

Совместное решение этой системы дает

$$x = \left(\frac{\sqrt{z} \pm \sqrt{z+4(r-z)}}{2}\right)^2,$$

$$y = \left(\frac{-\sqrt{z} \pm \sqrt{z+4(r-z)}}{2}\right)^2,$$

откуда при любом данном r после выбора целого неотрицательного z можно получить неотрицательные целые значения для x и y. Так, при $r = 13$ и $z = 1$ получаем $x = 9$ и $y = 16$.

Упражнение 4.4. Найдите положительные целые числа x, y, z, которые максимизируют $x^2 + y^2 + z^2$ при ограничении $x + y + z = a$, где a — данное положительное целое число.

Задача. Авиакомпания покупает самолеты типа Боинг 707 по цене 6 млн. долл., Боинг 727 по 4 млн. долл. и Каравелла по 1 млн. долл. Сколько самолетов каждого типа купила компания, если она уплатила всего 40 млн. долл. за 20 самолетов, причем были куплены самолеты всех типов?

Если x, y, z означают количества самолетов каждого типа, то требуется найти целые неотрицательные x, y, z, которые удовлетворяют системе уравнений

$$x + y + z = 20,$$

$$6x + 4y + z = 40.$$

Упражнение 4.5. Сформулируйте на языке алгебры следующую задачу [35].

Тринадцать пиратов добыли некоторое количество золотых монет. Они попытались разделить их поровну, но оказалось, что остается 8 штук. Когда они снова стали поровну делить монеты, после того как умерли от оспы два пирата, оказалось, что остается 3 монеты.

Потом в перестрелке погибли еще 3 пирата, но когда 8 пиратов стали делить монеты, оказалось, что остается 5 монет. Сколько всего было монет?

Указание. Получите соотношение вида $13x + 8 = 11y + 3 = 8z + 5$ и сведите его к паре уравнений $13x - 11y = -5$, $11y - 8z = 2$.

Упражнение 4.6. Найдите длины сторон всех прямоугольных треугольников с целочисленными сторонами, площади которых численно равны периметру.

Проверьте, что эта задача сводится к следующей системе уравнений:

$$\frac{xy}{2} = x + y + z,$$

$$z^2 = x^2 + y^2.$$

4.3. Линейные диофантовы уравнения [22]

Приведенные ниже уравнения определяют условия, при которых уравнения или системы линейных уравнений разрешимы в целых, не обязательно положительных числах.

Теорема 4.1. *Необходимое и достаточное условие разрешимости в целых числах линейного уравнения*

$$\sum_{j=1}^{n} a_j x_j = b,$$

где a_j $(j = 1, \ldots, n)$ *и* b *— целые числа, состоит в том, что наибольший общий делитель* $(a_1, \ldots, a_n)$ *чисел* $a_1, \ldots, a_n$ *делит* b.

Доказательство. Пусть d — наибольший общий делитель $a_1, \ldots, a_n$. Будем считать, что $b \geqslant 0$, так как, если $b < 0$, можно умножить обе части уравнения на -1.

Докажем сначала, что теорема справедлива при $b = d$. В ходе доказательства придем к выводу, что существование целочисленных решений невозможно при $b < d$. Наконец, дадим доказательство для $b > d$. Итак, предположим сначала, что $b = d$.

Рассмотрим множество S всех целых чисел вида

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n, \tag{4.1}$$

где $x_1, \ldots, x_n$ принимают всевозможные целочисленные значения: положительные, отрицательные и нулевые. Суммы и разности таких чисел также входят в S. Числа $a_1, \ldots, a_n$ тоже принадлежат S. Наибольший общий делитель всех принадлежащих S чисел равен d. Это следует из того, что d делит все эти числа, а ни одно число, большее d, не делит их, так как тогда оно делило бы также $a_1, \ldots, a_n$,

поэтому наибольший общий делитель не может быть больше d. Если $\bar{s}$ — наименьшее натуральное число в S и если s — любое другое число в S, то можно записать $s = \bar{s}q + r$, где q и r — целые числа и $0 \leqslant r \leqslant \bar{s}$. Тогда $r = -\bar{s}q$ и, следовательно, r также принадлежат S. Так как $r < \bar{s}$, следует положить $r = 0$ и тогда каждое число в S кратно $\bar{s}$. Таким образом, $\bar{s}$ делит $a_1, \ldots, a_n$. По предположению d принадлежит S и, следовательно, $\bar{s}$ делит d, или $\bar{s} \leqslant d$. Но неравенство $d > \bar{s}$ невозможно, так как d делит $\bar{s}$. Поэтому $d = \bar{s}$. Так как

$$\bar{s} = a_1x_1 + \ldots + a_nx_n,$$

уравнение разрешимо в целых числах при $b = d$; в этом случае доказательство закончено.

Предположим теперь, что $b > d$. Так как d делит b, получаем $b = kd$, где k — целое число. Поскольку уравнение $a_1x_1 + \ldots + a_nx_n = d$ разрешимо в целых числах, то уравнение, полученное из него умножением обеих частей на k, также разрешимо в целых числах. Обратное утверждение следует из того, что, как показано выше, любое целое значение $a_1x_1 + \ldots + a_nx_n$ должно делиться на d.

Упражнение 4.7. Покажите, что уравнение

$$2x + 6y = 9$$

неразрешимо в целых числах.

Теорема 4.2. *Система независимых линейных уравнений*

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

с целочисленными коэффициентами и целочисленными b_i ($i = 1, \ldots, m$), имеет целочисленное решение $x = (x_1, \ldots, x_n)$ тогда и только тогда, когда наибольший общий делитель всех определителей m-го порядка матрицы коэффициентов (a_{ij}) равен наибольшему общему делителю всех определителей m-го порядка расширенной матрицы.

Упражнение 4.8. Покажите, что система уравнений

$$\begin{aligned} 19x + 10y + 7z + 3 &= 0, \\ 5x + 2y + 2z + 1 &= 0 \end{aligned}$$

не имеет целочисленного решения, хотя каждое уравнение в отдельности разрешимо в целых числах. Обратите внимание, что одно из уравнений системы можно заменить на уравнение $3x + 6y - 1 = 0$, которое не имеет целочисленного решения.

В качестве простой геометрической иллюстрации идеи разрешимости в неотрицательных целых числах рассмотрим следующую задачу:

Задача. Требуется найти неотрицательные целые числа x_j, $j = 1, \ldots, n$, такие, что

$$\sum_{j=1}^{n} a_j x_j = \max$$

при ограничении

$$\sum_{j=1}^{n} b_j x_j = C.$$

Заметим, что если знак равенства заменить знаком неравенства, то задача называется задачей об упаковке ранца (она обсуждается в гл. 5). Чтобы существовало целочисленное решение этой задачи (не обязательно положительное), необходимо и достаточно, чтобы C делилось на наибольший общий делитель $\{b_j\}$, $j = 1, \ldots, n$. Для существования положительного целочисленного решения необходимо, например, чтобы дополнительно выполнялось условие $C \geqslant \sum_{j=1}^{n} b_j$. Чтобы существовало неотрицательное решение, необходимо выполнение условия $C \geqslant \min_j b_j$. Однако эти условия не являются достаточными.

Геометрически эта задача сводится к отысканию вектора с неотрицательными целыми компонентами, такого, что $ax = \max$ и $bx = C$. Приведем ограничение (которое определяет гиперплоскость в n-мерном пространстве) к канонической форме, разделив его на $|b| = \left(\sum_{j=1}^{n} b_j^2\right)^{1/2}$, модуль b. В этом случае $b/|b|$ определяет направляющие косинусы, и само b ортогонально к гиперплоскости. Расстояние от начала координат до этой гиперплоскости равно $C/|b|$. Все векторы x в гиперплоскости имеют одну и ту же проекцию на вектор b, длина которой равна $C/|b|$. Задача состоит в том, чтобы найти вектор x, направленный из начала координат в некоторую точку с целочисленными координатами, лежащую на гиперплоскости $bx = C$, который принадлежит также гиперплоскости $ax/|a| = M/|a|$ (где M — некоторая константа), такой, что расстояние до последней гиперплоскости от начала координат, равное $M/|a|$, максимально. Заметим, что, вообще говоря, решение находится в той части гиперплоскости $bx = C$, которая вместе с координатными плоскостями определяет $(n+1)$-мерный симплекс. Таким образом, решение находится в ограниченной области.

Метод непрерывных дробей [16, 26]

Одним из методов решения линейных и некоторых нелинейных диофантовых уравнений является метод непрерывных дробей.

Понятие о непрерывных дробях дает следующий пример:

$$\frac{33}{7}=4+\frac{5}{7}=4+\frac{1}{7/5}=4+\frac{1}{1+2/5}=4+\frac{1}{1+\dfrac{1}{5/2}}=$$

$$=4+\frac{1}{1+\dfrac{1}{2+1/2}}=4+\frac{1}{1+\dfrac{1}{2+\dfrac{1}{1+1}}}.$$

Такое представление называется *непрерывной дробью* [1]). В компактной форме последнее выражение можно записать в виде $4+\frac{1}{1+}\frac{1}{2+}\frac{1}{1+1}$. Очевидно, что непрерывные дроби могут быть и конечными и бесконечными. В последнем случае они называются *бесконечными непрерывными дробями.*

В общем случае если записать непрерывную дробь в виде

$$a_1+\frac{1}{a_2+}\frac{1}{a_3+}\frac{1}{a_4+}\cdots,$$

то дроби, которые получаются при взятии первого члена, первых двух членов, первых трех членов и т. д., называются *подходящими дробями непрерывной дроби.* Они задаются выражениями

$$a_1,\ a_1+\frac{1}{a_2},\quad a_1+\frac{1}{a_2+1/a_3},\ \ldots,$$

или

$$\frac{a_1}{1},\ \frac{a_1a_2+1}{a_2},\ \frac{a_3(a_1a_2+1)+a_1}{a_3a_2+1},\ \ldots,$$

которые можно просто обозначить c_1, c_2, c_3, . . ., c_n. Если ввести обозначения $c_{n-1}=p_{n-1}/q_{n-1}$ и $c_n=p_n/q_n$, то по индукции можно показать, что

$$p_n=a_np_{n-1}+p_{n-2},\quad q_n=a_nq_{n-1}+q_{n-2}.$$

Кроме того,

$$p_nq_{n-1}-q_np_{n-1}=(-1)^n.$$

Многие диофантовы уравнения можно решить методом непрерывных дробей. Однако здесь невозможно достаточно подробно описать все случаи использования этого метода. Только на примере будет показано, как он применяется. Рассмотрим уравнение

$$\alpha x-\beta y=1$$

[1]) Теория непрерывных, или цепных, дробей в доступной и строгой форме изложена в книге [39*].— *Прим. ред.*

и предположим, что α/β разлагается в непрерывную дробь $a_1, \ldots, a_n$ с подходящими дробями $c_1, \ldots, c_n$. Теперь $p_n = \alpha$ и $q_n = \beta$, и, следовательно, $\alpha q_{n-1} - \beta p_{n-1} = (-1)^n$. Если n — четное, т. е. если существует четное количество величин a, то $(-1)^n = 1$, и частное решение рассматриваемого уравнения имеет вид $\bar{x} = q_{n-1}$, $\bar{y} = p_{n-1}$. Если n — нечетное, то $(-1)^n = -1$, и неполное частное разложение может быть модифицировано путем замены

$$\frac{1}{a_n} \text{ на } \frac{1}{(a_n - 1) + 1/1}, \quad \text{если } a_n > 1,$$

или замены

$$\frac{1}{a_{n-1} + 1/a_n} \text{ на } \frac{1}{a_{n-1} + 1}, \quad \text{если } a_n = 1.$$

В обоих случаях получается четное число частных отношений. Пересчитывая p_{n-1} и q_{n-1}, находим

$$\alpha q_{n-1} - \beta p_{n-1} = 1.$$

Чтобы найти общее решение, вычтем $\alpha\bar{x} - \beta\bar{y} = 1$ из $\alpha x - \beta y = 1$. В результате получим

$$\alpha (x - \bar{x}) = \beta (y - \bar{y}).$$

Так как α и β взаимно просты, β должно делить $x - \bar{x}$, т. е. $x - \bar{x} = \beta t$, или $x = \bar{x} + \beta t$. Аналогично $y = \bar{y} + \alpha t$, $t = 0, \pm 1, \pm 2, \ldots$, что дает общее решение.

Упражнение 4.9. Покажите, что общее решение уравнения $\alpha x - \beta y = \gamma$ может быть записано в виде $x = c\bar{x} + \beta t$, $y = c\bar{y} + \alpha t$, где $\bar{x}$, $\bar{y}$ — частные решения уравнения $\alpha x - \beta y = 1$.

Метод наименьшего коэффициента для решения линейных уравнений

Метод наименьшего коэффициента состоит в последовательном исключении одной из неизвестных. В прошлом он использовался для получения неотрицательных решений систем. С разработкой этого метода связаны имена Эйлера и Сильвестра.

Чтобы проиллюстрировать применение этого метода, рассмотрим уравнение

$$11x + 3y = 23.$$

Сначала, решая его относительно переменной y, так как у нее наименьший коэффициент, получаем

$$y = \frac{23 - 11x}{3} = 7 - 3x + \frac{2 - 2x}{3} = 7 - 3x + t,$$

где $t = (2 - 2x)/3$ или $3t + 2x = 2$. Поскольку x и y должны быть целыми числами, t тоже должно быть целым числом.

Последнее уравнение разрешим относительно переменной x, так как у нее наименьший коэффициент. В результате получим

$$x = \frac{2-3t}{2} = 1 - t - \frac{t}{2} = 1 - t - u,$$

где $u = t/2$, или $t = 2u$. Так как x и t должны быть целыми числами, u тоже должно быть целым числом.

Пусть теперь u — целое число, тогда $t = 2u$ и, следовательно, t — тоже целое число. Отсюда x и y имеют вид

$$x = 1 - 3u,$$
$$y = 4 + 11u.$$

Таким образом, придавая u целые значения 0, ± 1, ± 2, . . ., получим всевозможные целочисленные значения x и y, которые удовлетворяют уравнению. Требование положительности x и y влечет за собой условие $1 - 3u > 0$; $4 + 11u > 0$, т. е. $-4/11 < u < 1/3$. Это дает единственное целочисленное решение $u = 0$ и, следовательно, $x = 1$, $y = 4$.

Предыдущие рассуждения можно слегка модифицировать. Чтобы решить уравнение [31]

$$31x_1 + 14x_2 = 7,$$

рассмотрим член с бо́льшим коэффициентом и затем запишем, используя коэффициент при x_2,

$$31x_1 \equiv 7 \bmod 14.$$

Это дает

$$3x_1 \equiv 7 \bmod 14, \quad \text{или} \quad x_1 \equiv 7 \bmod 14.$$

Таким образом,

$$x_1 = 7 + 14t,$$

и подстановка в уравнение дает

$$x_2 = -15 - 31t.$$

Как и ранее, придавая целочисленные значения t, получим требуемое решение.

Чтобы решить уравнение $a_1x_1 + a_2x_2 + a_3x_3 = b$, сначала решим $a_3x_3 \equiv b \bmod d_1$, где $d_1 = (a_1, a_2)$ — наибольший общий делитель a_1 и a_2. При этом вводится параметр t_1 и соотношение $x_3 = \bar{x}_3 + dt_1$, где $d = (d_1, a_3)$. Затем этот результат используется в уравнении и вводится второй целочисленный параметр t_2, соответствующий t в предыдущем примере. Рассмотрим уравнение $20x_1 + 34x_2 + 40x_3 + 77x_4 = 127$. Так как $(20, 34, 40, 77) = 1$, существует целочисленное решение. Далее, поскольку $(20, 34, 40) = 2$, рассмотрим уравнение $77x_4 \equiv 127 \bmod 2$, или $x_4 \equiv 1 \bmod 2$. Таким образом, $x_4 = 1 + 2t_1$. Подставляя в уравнение, получаем $10x_1 + 17x_2 + 20x_3 = 25 - 77t_1$. Поскольку $(10, 17, 20) = 1$, то уравнение раз-

решимо при всех t_1. В частности, так как (10, 17) = 1, можно положить $x_3 = t_2$ и получить

$$10x_1 + 17x_2 = 25 - 77t_1 - 20t_2.$$

Рассматривая это уравнение по mod 10, получим $7x_2 \equiv 5 - 7t_1$. Чтобы решить это уравнение, умножим его на 3; при этом получим $x_2 \equiv 5 + 9t_1$, так как $-21 \equiv -1 \equiv 9 \bmod 10$. Итак, $x_2 = 9t_1 + 10t_3 + 5$. Подстановка в предыдущее уравнение дает

$$10x_1 = 25 - 77t_1 - 20t_2 - 17\,(9t_1 + 10t_3 + 5).$$

Отсюда получаем

$$\begin{aligned} x_1 &= -6 - 23t_1 - 2t_2 - 17t_3, \\ x_2 &= 9t_1 + 10t_3 + 5, \\ x_3 &= t_2, \\ x_4 &= 1 + 2t_1. \end{aligned}$$

Чтобы найти все неотрицательные целочисленные решения x_i, $i = 1, \ldots, 4$, заметим, что из условия $40x_3 \leqslant 127$ следует, что $0 \leqslant x_3 \leqslant 3$ и, следовательно, $0 \leqslant t_2 \leqslant 3$. Аналогично из $1 + 2t_1 \geqslant 0$ следует, что $t_1 \geqslant 0$. Заметим, что из $-6 - 23t_1 - 17t_3 \geqslant 2t_2$ вытекает, что $t_3 \leqslant -1$, иначе при $t_1 = 0$ левая часть неравенства становится отрицательной. В целом неравенства $t_1 \geqslant 0$, $0 \leqslant t_2 \leqslant 3$, $-6 - 23t_1 - 17t_3 \geqslant 2t_2$ определяют область допустимых значений, которая, как мы сейчас покажем, является пустым множеством. Из уравнения видно, что $x_4 = 0$ или 1. Если $x_4 = 0$, то уравнение не имеет решения. Если $x_4 = 1$, то полученное уравнение, очевидно, не имеет положительного решения. Таким образом, уравнение неразрешимо в целых положительных числах.

Упражнение 4.10. Рассмотрите задачу о покупке самолетов из разд. 4.2. Исключите y, потом x. Получите $x = z - 20 + z/2$, откуда $z = 2u$ для некоторого целочисленного u. Таким образом,

$$\begin{aligned} x &= 3u - 20 > 0, \\ y &= 40 - 5u > 0, \\ z &= 2u > 0. \end{aligned}$$

Покажите, что искомое решение имеет вид $u = 7$, $x = 1$, $y = 5$, $z = 14$.

Упражнение 4.11. Докажите, что минимальным решением в задаче о золотых монетах из разд. 4.2 являются 333 монеты.

Чтобы решить в целых числах систему линейных диофантовых уравнений, надо найти все решения первого уравнения и подставить эти решения (с их параметрами) во второе, затем решить второе уравнение и подставить решение в третье и т. д. Если требуется получить положительное целочисленное решение, то на последнем

шаге накладывается это условие и проверка решений осуществляется в обратном порядке. Получается весьма громоздкая процедура. Система уравнений [32]

$$6x_1 + 3x_2 + 2x_3 + x_4 = 20,$$
$$x_1 + x_2 + x_3 + x_4 = 10$$

имеет следующие решения: (0, 0, 10, 0), (0, 1, 8, 1), (0, 2, 6, 2), (0, 3, 4, 3), (0, 4, 2, 4), (0, 5, 0, 5), (1, 0, 5, 4), (1, 1, 3, 5), (1, 2, 1, 6), (2, 0, 0, 8).

Упражнение 4.12. Покажите, что набор (1, 1, 3, 5) минимизирует $\sum_{i=1}^{3} |x_i - x_{i+1}|$ при условии, что последние два уравнения служат ограничениями.

Общее решение. Если линейное диофантово уравнение [2]

$$a_1x_1 + \ldots + a_nx_n = b$$

с целочисленными коэффициентами имеет целочисленное решение, то b должно делиться на наибольший общий делитель d_n чисел $a_1, \ldots, a_n$. Таким образом, $b = d_nt_n$, где t_n — произвольное целое число.

Теорема 4.3. *Индукцией по n можно показать (см. ниже вывод), что общее решение уравнения*

$$a_1x_1 + \ldots + a_nx_n = d_nt_n, \quad n \geqslant 2,$$

имеет вид

$$\begin{aligned}
x_1 &= t_nx_1^{(n)} + t_{n-1}\frac{a_n}{d_n}x_1^{(n-1)} + t_{n-2}\frac{a_{n-1}}{d_{n-1}}x_1^{(n-2)} + \ldots + t_1\frac{a_2}{d_2},\\
x_2 &= t_nx_2^{(n)} + t_{n-1}\frac{a_n}{d_n}x_2^{(n-1)} + t_{n-2}\frac{a_{n-1}}{d_{n-1}}x_2^{(n-2)} + \ldots - t_1\frac{d_1}{d_2},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
x_{n-1} &= t_nx_{n-1}^{(n)} + t_{n-1}\frac{a_n}{d_n}x_{n-1}^{(n-1)} - t_{n-2}\frac{d_{n-2}}{d_{n-1}},\\
x_n &= t_nx_n^{(n)} - t_{n-1}\frac{d_{n-1}}{d_n},
\end{aligned} \tag{4.2}$$

где $t_1, \ldots, t_n$ — произвольные целые числа; $d_1 = a_1$; d_2 — наибольший общий делитель чисел a_1 и a_2; $\ldots$; d_n — наибольший общий делитель чисел $a_1, \ldots, a_n$; $x_i^{(j)}$, $i, j = 1, \ldots, n$, представляют собой решения уравнений

$$\begin{aligned}
d_2 &= a_1x_1^{(2)} + a_2x_2^{(2)},\\
d_3 &= a_1x_1^{(3)} + a_2x_2^{(3)} + a_3x_3^{(3)},\\
&\ldots\ldots\ldots\ldots\ldots\ldots\\
d_n &= a_1x_1^{(n)} + a_2x_2^{(n)} + \ldots + a_nx_n^{(n)}.
\end{aligned}$$

Метод решения любого уравнения из последнего множества уравнений основывается на следующей теореме [37]:

Теорема 4.4. *Для любых данных положительных целых чисел* $a_1, \ldots, a_n$ *наименьшим положительным числом вида* $a_1x_1 + \ldots + a_nx_n$ ($x_1, \ldots, x_n$ — *целые числа*) *является* d — *наибольший общий делитель* $a_1, \ldots, a_n$.

Доказательство. Предположим, что $0 < a_1x_1 + \ldots + a_nx_n = s \neq d$. Так как d должно делить s, хотя бы один из коэффициентов a_k имеет вид $a_k = qs + r$, $0 < r < s$. Это дает

$$a_k = q(a_1x_1 + \ldots + a_nx_n) + r,$$

или

$$0 < a_1(-qx_1) + a_2(-qx_2) + \ldots + a_k(1 - qx_k) + \ldots$$
$$\ldots + a_n(-qx_n) = r < s.$$

Положим теперь $x_j^1 = -qx_j$, $j = 1, \ldots, n$, $j \neq k$, и $x_k^1 = 1 - qx_k$ и будем продолжать этот процесс, пока не достигнем d.

Частичное доказательство теоремы 4.3. Бонд [2] показал, что существенным моментом при выводе по индукции системы уравнений (4.2) является следующее. Пусть результат справедлив при $n = k$, $k \geqslant 2$. Если $x_1^{(k+1)}, \ldots, x_{k+1}^{(k+1)}$ — решение уравнения

$$a_1x_1 + \ldots + a_{k+1}x_{k+1} = d_{k+1},$$

а $x_1, \ldots, x_{k+1}$ — решение уравнения

$$a_1x_1 + \ldots + a_{k+1}x_{k+1} = t_{k+1}d_{k+1},$$

где $a_1, \ldots, a_{k+1}$, t_{k+1} — ненулевые целые числа, то

$$a_1(x_1 - t_{k+1}x_1^{(k+1)}) + \ldots + a_{k+1}(x_{k+1} - t_{k+1}x_{k+1}^{(k+1)}) = 0.$$

Разделив это уравнение на d_{k+1}, перенеся последний член левой части в правую часть и заметив, что $a_1/d_{k+1}, \ldots, a_{k+1}/d_{k+1}$ — взаимно просты, приходим к выводу, что последнее уравнение имеет место тогда и только тогда, когда

$$x_{k+1} - t_{k+1}x_{k+1}^{(k+1)} = (-t_k)\frac{d_k}{d_{k+1}}.$$

При каждом выборе целочисленного значения t_k по индукции получаем

$$x_1 - t_{k+1}x_1^{(k+1)} = t_k\frac{a_{k+1}}{d_{k+1}}x_1^{(k)} + t_{k-1}\frac{a_k/d_{k+1}}{d_k/d_{k+1}}x_1^{(k-1)} + \ldots + t_1\frac{a_2/d_{k+1}}{d_2/d_{k+1}},$$

$$x_2 - t_{k+1}x_2^{(k+1)} = t_k\frac{a_{k+1}}{d_{k+1}}x_2^{(k)} + t_{k-1}\frac{a_k/d_{k+1}}{d_k/d_{k+1}}x_2^{(k-1)} + \ldots - t_1\frac{d_1/d_{k+1}}{a_2/d_{k+1}},$$

. . . . . .

$$x_k - t_{k+1}x_k^{(k+1)} = t_k\frac{a_{k+1}}{d_{k+1}}x_k^{(k)} - t_{k-1}\frac{d_{k-1}/d_{k+1}}{d_k/d_{k+1}},$$

где

$$\frac{a_1}{d_{k+1}} x_1^{(i)} + \ldots + \frac{a_i}{d_{k+1}} x_i^{(i)} = \frac{d_i}{d_{k+1}} \text{ при } i = 2, \ldots, k.$$

После перенесения членов, содержащих t_{k+1}, вправо и упрощений получим $x_1, \ldots, x_{k+1}$ в искомом виде. Легко проверить, что каждый выбор целых чисел $t_1, \ldots, t_k$ дает целочисленное решение уравнения $a_1x_1 + \ldots + a_{k+1}x_{k+1} = t_{k+1}d_{k+1}$.

Пример. Можно начать с произвольного выбора $x_1^{(0)}, \ldots, x_n^{(0)}$ и последовательно выбирать $x_1^{(i)}, \ldots, x_n^{(i)}$ в соответствии с вышеизложенной процедурой до тех пор, пока не будет получено значение d и, таким образом, найдено решение диофантова уравнения с коэффициентами $a_1, \ldots, a_n$, в правой части которого стоит d. Эта процедура полезна и при отыскании наименьшего общего делителя $a_1, \ldots, a_n$.

При использовании этого метода для отыскания общего решения уравнения

$$3x_1 + 2x_2 + x_3 = 13$$

сначала следует использовать его для нахождения наибольшего общего делителя 3, 2, 1. Получаем $d_3 = 1$, откуда $t_3 = 13$. Нужно также найти наибольший общий делитель 3, 2, который равен 1. Затем решаем уравнения

$$3x_1^{(2)} + 2x_2^{(2)} = 1,$$
$$3x_1^{(3)} + 2x_2^{(3)} + x_3^{(3)} = 1.$$

В ходе этого процесса находим частные решения

$$(x_1^{(2)}, x_2^{(2)}) = (1, -1),$$
$$(x_1^{(3)}, x_2^{(3)}, x_3^{(3)}) = (1, -1, 0).$$

Теперь для определения общего решения уравнения имеем

$$x_1 = 13 + t_2 + 2t_1,$$
$$x_2 = -13 - t_2 - 3t_1,$$
$$x_3 = -t_2.$$

Чтобы получить неотрицательные целочисленные решения, полагаем

$$13 + t_2 + 2t_1 \geqslant 0,$$
$$13 + t_2 + 3t_1 \leqslant 0, \quad t_2 \leqslant 0.$$

Из условий $13 + t_2 \geqslant -2t_1$ и $13 + t_2 \leqslant -3t_1$ получаем неравенство $-3t_1 \geqslant -2t_1$, которое справедливо тогда и только тогда, когда $t_1 \leqslant 0$. Таким образом, все решения лежат на плоскости t_1, t_2 в треугольнике с вершинами (13/3, 0), (—13/2, 0) и (0, —13).

Максимизация $8x_1 + 5x_2 + x_3$ при ограничении $3x_1 + 2x_2 + x_3 = 13$, где x_i, $i = 1, 2, 3$,— неотрицательные целые числа, сводится к следующему: находятся неотрицательные целые числа u_1 и u_2, которые максимизируют

$$8(13 - u_2 - 2u_1) + 5(-13 + u_2 + 3u_1) + u_2 = -u_1 - 2u_2 + 39$$

при ограничениях $13 - u_2 - 2u_1 \geqslant 0$, $-13 + u_2 + 3u_1 \geqslant 0$, $u_2 \geqslant 0$, что представляет собой задачу целочисленного линейного программирования. Вообще говоря, система из $m \leqslant n$ диофантовых уравнений с n переменными приводит к системе из mn неравенств с $(n - 1)$ переменными. Решение исходной системы в неотрицательных целых числах может быть получено, если соответствующая система неравенств решается в целых числах.

4.4. Некоторые нелинейные уравнения

Покажем, что уравнение [19]

$$x^2 + y^2 = z^2$$

имеет бесконечно много решений в целых числах. Можно принять, что x, y, z — попарно взаимно просты, так как, если это не так и d — их наибольший общий делитель, можно поделить все члены уравнения на d^2 и добиться желаемого условия. Таким образом, два члена соотношения должны быть нечетными. Покажем сначала, что этой парой не могут быть x и y. Если переменная x — нечетная, то $x \equiv 1$ или 3 по модулю 4 и, следовательно, $x^2 \equiv 1$ по модулю 4. Аналогично если переменная y нечетная, то $y^2 \equiv 1 \bmod 4$ и, следовательно, $x^2 + y^2 \equiv 2 \bmod 4$. Но $x^2 + y^2 = z^2$, тогда как для каждого целого z имеем $z^2 \equiv 0$ или $1 \bmod 4$. Таким образом, либо x, либо y является нечетной, например y — нечетная, а x — четная. Итак,

$$x^2 = 4t^2 = z^2 - y^2 = (z - y)(z + y).$$

Теперь члены $(z - y)$ и $(z + y)$ имеют наибольший общий делитель, равный 2. Таким образом, следует записать

$$z - y = 2m^2,$$
$$z + y = 2n^2,$$

чтобы x^2 было квадратом произведения целых чисел. Тогда x будет целым числом.

Итак, $z = m^2 + n^2$, $y = n^2 - m^2$ и $x = 2mn$. Целые числа m и n называются *задающими числами треугольника*. Общее решение задается соотношениями

$$x = 2kmn,$$
$$y = k(n^2 - m^2), \quad m - n \equiv 1 \bmod 2,$$
$$z = k(m^2 + n^2),$$

где k — произвольное целое число.

Рассмотрим теперь положительные целочисленные решения уравнения $a^b = b^a$ [34]. Заметим сначала, что

$$\sqrt[n]{n+1} = \begin{cases} 2, & n=1, \\ \sqrt{3}, & n=2, \end{cases}$$
$$< \sqrt{3}, \; n>2.$$

Последнее неравенство следует из рассмотрения первых трех членов разложения $(1+2)^n$, т. е.

$$3^n = (1+2)^n > 1 + n2 + \frac{n(n-1)}{2!}2^2 = (1+2n+n^2) + n(n-2) > (n+1)^2.$$

Следовательно, $\sqrt[n]{n+1} < \sqrt{3}$. Таким образом, для $n \geqslant 1$ единственное целочисленное значение $\sqrt[n]{n+1}$ равно 2. Предположим, что $a > b$; тогда, поскольку $a^b = b^a$,

$$a = b^{a/b} = bb^{(a/b)-1}.$$

Так как a и b — целые числа, $b^{(a/b)-1}$ также должно быть целым числом, которое обозначим через n и запишем $a = nb$. Таким образом, $(nb)^b = (b^n)^b$, откуда следует $nb = b^n$ или $b = \sqrt[n-1]{n}$. Итак, чтобы b было целым числом, ввиду изложенного выше должно выполняться равенство $n - 1 = 1$, или $n = 2$, откуда $b = 2$ и $a = 4$. Поэтому, если $a \neq b$, положительные целочисленные решения имеют вид $a = 4$, $b = 2$ или $a = 2$, $b = 4$.

Исследуем теперь вопрос о существовании положительного целочисленного решения квадратичной формы с двумя переменными [16]:

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_1x_1 + 2a_2x_2 + a_3 = 0,$$

шесть коэффициентов которой являются целыми числами.

Разрешив это уравнение относительно x_1, получим

$$a_{11}x_1 + a_{12}x_2 + a_1 = \pm\left[(a_{12}^2 - a_{11}a_{22})x_2^2 + 2(a_{12}a_1 - a_{11}a_2)x_2 + (a_1^2 - a_{11}a_3)\right]^{1/2}.$$

Чтобы существовали положительные целочисленные значения для x_1 и x_2, дискриминант должен быть полным квадратом, т. е. он должен иметь вид

$$px_2^2 + 2qx_2 + r = y^2,$$

где y — переменная. Решая это уравнение относительно x_2, получаем

$$px_2 + q = \pm\sqrt{q^2 - pr + py^2}.$$

Условие, что дискриминант должен быть полным квадратом, приводит к уравнению

$$x^2 - py^2 = q^2 - pr,$$

где x — переменная, которое надо разрешить в целых числах. Это уравнение можно переписать в виде

$$x^2 \pm Ny^2 = \pm a,$$

где N и a — положительные целые числа. Уравнение

$$x^2 - Ny^2 = \pm 1$$

является, конечно, уравнением Пелля.

Замечание. Если a_{11}, a_{22} и a_{12} — все положительные, то для больших x_1 и x_2 квадратный член является доминирующим и, следовательно, левая часть уравнения не может обратиться в нуль. Отсюда вытекает, что число положительных целочисленных решений конечно

Упражнение 4.13. Покажите, что если $a_{12}^2 - a_{11}a_{22} < 0$, то число целочисленных решений конечно.

Упражнение 4.14. Покажите, что уравнение $x^2 + Ny^2 = -a$ не имеет действительных решений, а уравнение $x^2 + Ny^2 = a$ имеет конечное число решений.

Нас интересует уравнение

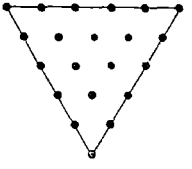
$$x^2 - Ny^2 = \pm a.$$

Если $N = M^2$ — полный квадрат, то уравнение $x^2 - M^2y^2 = a$ можно представить в виде $(x + My)(x - My) = a$. Если $a = f_1 f_2$, $f_1 > f_2$, то положим $x + My = f_1$, $x - My = f_2$. Таким образом, целочисленные решения, если они существуют, получаются путем различных факторизаций a.

Упражнение 4.15. Покажите, что уравнение $x^2 - y^2 = 60$ имеет решения (x, y), равные (8, 2) и (16, 14).

Случай, когда N не является полным квадратом, более сложен. Читателям, которых интересует этот вопрос, следует обратиться к специальной литературе.

Пример. Объекты некоторого множества расположены так, что они образуют правильный треугольник: один объект расположен в вершине, за ним на двух сторонах треугольника расположены еще два объекта, за ними еще три объекта на одном уровне и т. д. до N объектов (фиг. 4.2). Объекты этого множества можно перераспределить так, что они будут образовывать два идентичных правильных треугольника. Найдите точное число объектов, если их количество в первом треугольнике не менее 1000 и не более 10 000.



Фиг. 4.2.

Решение. В первом треугольнике имеется $S = N(N + 1)/2$ объектов, а во втором и третьем — по $M(M + 1)/2$. Получаем

уравнение [13]

$$\frac{N(N+1)}{2} = \frac{2M(M+1)}{2}. \tag{4.3}$$

Изложим один из методов решения этого уравнения. После дополнения до полного квадрата обеих частей равенства и упрощений получаем

$$(2N+1)^2 - 2(2M+1)^2 = -2.$$

Если подставить $x = 2N + 1$, $y = 2M + 1$, то уравнение принимает простой вид:

$$x^2 - 2y^2 = -2.$$

(Этим уравнением мы не будем пользоваться в дальнейшем.) Разрешив относительно N уравнение второй степени (4.3), получим

$$2N = -1 \pm \sqrt{1 + 8M(M+1)}.$$

Пусть $M = N - k$, тогда уравнение принимает вид

$$2N = 4k - 1 \pm \sqrt{1 + 8k^2}. \tag{4.4}$$

Следовательно, $1 + 8k^2$ должно быть полным квадратом. Заметим, что $k = 1$ дает $N = 0$ или $N = 3$, и полное число объектов равно шести; их можно разделить на два треугольника по три объекта в каждом. Для произвольного k обозначим через P меньшее из двух значений N, а через Q — большее. Если подставить в уравнение (4.4) Q вместо N и упростить, то получится уравнение второй степени относительно k:

$$8k^2 - 8k(1 + 2Q) + 4Q(Q + 1) = 0.$$

Таким образом, Q дает два значения k, k_1 и k_2, причем $k_1 < k_2$. Теперь для каждого из этих двух значений k получаем соответствующие P и Q, но две из четырех этих величин совпадают, так как для получения k_1 и k_2 использовалось данное значение Q. Одно из этих двух равных значений соответствует Q в случае k_1 и P в случае k_2. Теперь значение Q, соответствующее k_2, снова используем для получения двух значений k, одним из которых будет k_2, а другим — новое $k_3 > k_2$ и т. д.

Получаем следующие рекуррентные соотношения:

$$2Q_i = 4k - 1 - \sqrt{1 + 8k^2},$$

$$2P_{i+1} = 4k - 1 + \sqrt{1 + 8k^2},$$

$$Q_i + P_i = 4k - 1.$$

Последнее соотношение представляет собой сумму двух корней [см. уравнение (4.4)] и дает возможность переходить от P к Q, соответствующему такому же значению k. Заметим, что если один из корней k_1 — целое число, то k_2 тоже будет целым числом, так как $k_1 + k_2 = 1 + 2Q$ — целое число.

Начнем с $k = 1$. Используя формулу (4.4), получаем $P_1 = 0$, $Q_1 = 3$ и $S_1 = 6$, где S_i — полное число объектов, вычисленное для k_i, $i = 1, 2, \ldots$. Теперь $Q_1 = P_2 = 3$ и $S_2 = 6$, так как $k_1 + k_2 = 1 + 2Q$. При $k_2 = 6$ получаем $Q_2 = 20$, так как $P_2 + Q_2 = 4k_2 - 1$, откуда $S_2 = 210$. Наконец, $k_3 = 35$, $Q_3 = 119$ и $S_3 = 7140$, что и дает искомый ответ.

Из соотношения

$$S_i = \frac{Q_i(Q_i+1)}{2}$$

получается полезная формула

$$S_i = k_i k_{i+1}.$$

Кроме того, из равенств

$$k_i + k_{i+1} = 1 + 2Q_i,$$
$$Q_{i+1} + Q_i = 4k_{i+1} - 1$$

имеем

$$k_{i+2} - 6k_{i+1} + k_i = 0, \tag{4.5}$$

где $k_0 = 0$, $k_1 = 1$. Это позволяет избежать непосредственного вычисления P_i и Q_i. Кроме того, $\alpha_{i+1}/\alpha_i \to 3 + 2\sqrt{2}$ при $i \to \infty$ и, следовательно, $S_{i+1}/S_i \to 17 + 12\sqrt{2}$.

Вышеприведенное решение является единственным, так как при любом N в уравнении (4.3) требуется отыскивать значения k; при помощи уравнения (4.5) получаем наименьшее положительное k_0, которое необходимо равно нулю.

4.5. Оптимизация при диофантовых ограничениях

До сих пор известно немного общих методов решения диофантовых уравнений и естественно поэтому использовать имеющуюся в распоряжении информацию при решении соответствующих задач оптимизации с ограничениями. Например, известно, что общее целочисленное решение (не обязательно неотрицательное) уравнения

$$\alpha x + \beta y = \gamma$$

имеет вид

$$x = \bar{x} + \frac{\beta}{(\alpha, \beta)}\, t, \quad y = \bar{y} - \frac{\alpha}{(\alpha, \beta)}\, t,$$

где $(\bar{x}, \bar{y})$ — частное решение и (α, β) — наибольший общий делитель α и β.

Может возникнуть необходимость отыскания значения t, которое минимизирует

$$f(x, y) = c_1 x^2 + c_2 y^2 \qquad c_1, c_2 > 0,$$

при условии, что приведенное выше уравнение служит ограничением. Подстановка x и y, выраженных через t, приводит к уравнению второй степени относительно t. Можно использовать методы математического анализа и затем взять ближайшее целое значение t. Ввиду выпуклости $f(x, y)$ этот метод пригоден для отыскания минимума.

Ниже приводится несколько задач, представляющих теоретический и практический интерес, которые легко формулируются, но общие решения которых до сих пор не известны:

1. Пусть дано положительное целое число C. Найти разложение на положительные целочисленные слагаемые

$$\sum_{j=1}^{n} x_j = C,$$

такое, что наименьшее общее кратное чисел $x_1, \ldots, x_n$ максимально.

Если $p_1, \ldots, p_s$ — все простые числа, которые встречаются в разложениях $x_1, \ldots, x_n$, и

$$x_j = p_1^{\alpha_{1j}} \ldots p_s^{\alpha_{sj}}, \quad j = 1, \ldots, n,$$

то наименьшее общее кратное $x_1, \ldots, x_n$ имеет вид

$$L = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_s^{\alpha_s},$$

где

$$\alpha_i = \max(\alpha_{i1}, \ldots, \alpha_{in}), \quad i = 1, \ldots, s.$$

2. Следующая задача возникла при изучении структуры телефонной связи [32]: минимизировать в положительных целых числах нелинейное выражение

$$\sum_{j=2}^{n} |x_j - x_{j-1}|$$

при ограничениях

$$\sum_{j=1}^{n} x_j = C \quad \text{и} \quad \sum_{j=1}^{n} a_j x_j = C_1,$$

где целые числа C и C_1 удовлетворяют условию

$$C \leqslant C_1 \leqslant \alpha C$$

при данном четном целом числе α и где положительные целые коэффициенты α_j, $j = 1, \ldots, n$, делят α. Точное решение этой задачи не известно, но британское правительственное издательство опубликовало обширные таблицы. Однако Р. Сиски указал, что в таблицах не приводятся неединственные решения.

3. Если $x_j = m_{0j}/m_{1j}$ — отношение исходной массы к загрузке j-й ступени n-ступенчатой ракеты и $a_j = m_{0j}/m_{2j}$, где m_{2j} — масса j-й ступени без топлива, то при конструировании ракет возникает задача определения неотрицательных целых чисел x_j, $j = 1, \ldots, n$,

которые минимизируют $\prod_{j=1}^{n} x_j$ при ограничении

$$\log \prod_{j=1}^{n} \frac{a_j x_j}{a_j + x_j} = C,$$

где a_j, $j = 1, \ldots, n$ и C заданы.

Упражнение 4.16. Найдите целые числа x, y, z (положительные или отрицательные), которые удовлетворяют уравнению

$$\frac{z}{2} - \frac{x}{3} + 2y = 11$$

и максимизируют

$$| x | 10^2 + | y | 10 + | z |.$$

Ответ: $x = 9$, $y = 9$, $z = -8$ и максимальное значение равно 998.

Упражнение 4.17. Если x и y — целые числа, может ли быть целым числом $(x/y + y/x)$?

Указание. Если $x/y + y/x = k$ — целое число, то $x^2 + y^2 = kxy$, $x^2 = y(kx - y)$; таким образом, y делит x^2, что возможно, если только $y = \pm x$. Таким образом, сумма может быть равна 2 или -2.

Известно несколько критериев получения неотрицательного целочисленного решения в нелинейных задачах оптимизации при наличии ограничений. Полезный критерий, который будет изложен ниже, принадлежит О. Гроссу [15].

Теорема 4.5. *Необходимое и достаточное условие того, что вектор с неотрицательными целочисленными компонентами* $x = (x_1, \ldots, x_n)$ *минимизирует выражение*

$$\sum_{j=1}^{n} \varphi_j(x_j), \tag{4.6}$$

где φ_j, $j = 1, \ldots, n$, — *выпуклые функции, при ограничении*

$$\sum_{j=1}^{n} x_j = m, \tag{4.7}$$

где m — *заданное положительное целое число, состоит в том, что*

$$\min_{j \in I} [\varphi_j(x_j + 1) - \varphi_j(x_j)] \geqslant \max_{j \in S^+(x)} [\varphi_j(x_j) - \varphi_j(x_j - 1)], \tag{4.8}$$

где $I = \{1, \ldots, n\}$ *и* $S^+(x) = \{j \in I \mid x_j > 0\}$.

Доказательство. Достаточность. Пусть x — допустимое решение (такое, что

$$\sum_{i=1}^{n} x_i = m,$$

$x_i \geqslant 0$ — целые числа), которое удовлетворяет неравенству (4.8) Пусть x' — другое допустимое решение. Покажем, что

$$\sum_{j=1}^{n} \varphi_j(x_j') \geqslant \sum_{j=1}^{n} \varphi_j(x_j).$$

Пусть

$$\lambda \equiv \min_{j \in I} [\varphi_j(x_j + 1) - \varphi_j(x_j)]. \tag{4.9}$$

Тогда

$$\lambda \leqslant \varphi_j(x_j + 1) - \varphi_j(x_j) \text{ для всех } j \in I. \tag{4.10}$$

Из неравенства (4.8) получаем

$$\lambda \geqslant \varphi_j(x_j) - \varphi_j(x_j - 1) \text{ для всех } j \in S^+(x). \tag{4.11}$$

Из выпуклости φ_j следует

$$\frac{1}{2}[\varphi_j(k+1) + \varphi_j(k-1)] \geqslant \varphi_j(k). \tag{4.12}$$

Таким образом,

$$\varphi_j(k+1) - \varphi_j(k) \geqslant \varphi_j(k) - \varphi_j(k-1) \tag{4.13}$$

и, следовательно, $[\varphi_j(k+1) - \varphi_j(k)]$ — неубывающая функция на целых числах. Таким образом, если $k \geqslant x_j$, неравенство (4.10) дает

$$\lambda \leqslant \varphi_j(k+1) - \varphi_j(k), \tag{4.14}$$

и если $0 < k \leqslant x_j$, то $j \in S^+(x)$ и неравенство (4.11) дает

$$\lambda \geqslant \varphi_j(k) - \varphi_j(k-1). \tag{4.15}$$

Просуммируем неравенство (4.14) по всем k, таким, что $x_j \leqslant k \leqslant x_j' - 1$ при $x_j' > x_j$, и просуммируем неравенство (4.15) по всем значениям k, таким, что $x_j' + 1 \leqslant k \leqslant x_j$ при $x_j' < x_j$. Первое суммирование дает

$$\varphi_j(x_j') - \varphi_j(x_j) \geqslant \lambda(x_j' - x_j), \quad x_j' > x_j,$$

а второе

$$\varphi_j(x_j) - \varphi_j(x_j') \leqslant \lambda(x_j - x_j'), \quad x_j' < x_j.$$

В обоих случаях и, очевидно, даже при $x_j = x_j'$ $\varphi_j(x_j') \geqslant \varphi_j(x_j) + \lambda x_j' - \lambda x_j$ для всех $j \in I$.

Суммируя по $1 \leqslant j \leqslant n$ и используя $\sum_{j=1}^{n} x_j' = m = \sum_{j=1}^{n} x_j$, мы доказываем достаточное условие, т. е. x минимизирует выражение (4.6).

Доказательство. Необходимость. Пусть x минимизирует выражение (4.6) и удовлетворяет условию (4.7) в неотрицательных целых числах; предположим, что знак неравенства (4.8) изменен на обратный. Так как $m > 0$, $S^+(x) \neq \varnothing$, правая часть неравенства (4.8) достигает максимума при некотором целочисленном α, а левая часть достигает минимума при некотором целочисленном значении β. Так как мы предполагаем, что знак неравенства изменен на обратный, получаем

$$\varphi_\beta(x_\beta + 1) - \varphi_\beta(x_\beta) < \varphi_\alpha(x_\alpha) - \varphi_\alpha(x_\alpha - 1). \qquad (4.16)$$

Если $\alpha = \beta$, это соотношение не может иметь места, так как функция φ — монотонная. Таким образом, $\alpha \neq \beta$. Рассмотрим

$$x'_\alpha = x_\alpha - 1,$$
$$x'_\beta = x_\beta + 1$$

и $x'_j = x_j$ для остальных значений j. Очевидно, что $x'_j \geqslant 0$ — все целые числа и $\sum_{j=1}^{n} x'_j = m$. Кроме того,

$$\sum_{j=1}^{n} \varphi_j(x'_j) - \sum_{j=1}^{n} \varphi_j(x_j) = \varphi_\alpha(x_\alpha - 1) - \varphi_\alpha(x_\alpha) + \varphi_\beta(x_\beta + 1) - \varphi_\beta(x_\beta).$$

Это выражение меньше нуля ввиду неравенства (4.16), что противоречит тому факту, что x_j, $j = 1, \ldots, n$, минимизируют функцию. Таким образом, неравенство (4.8) должно иметь место, что и доказывает необходимость.

Решение получается при помощи итераций, начиная с

$$x_j^{(0)} = 0, \quad j = 1, \ldots, n.$$

Пусть $j^*(k)$ для $k \geqslant 0$ будет номером, таким, что

$$\min_{j \in I} [\varphi_j(x_j^{(k)} + 1) - \varphi_j(x_j^{(k)})],$$

и положим

$$x_{j^*}^{(k+1)} = x_{j^*}^{(k)} + 1,$$
$$x_j^{(k+1)} = x_j^{(k)}, \quad j \neq j^*.$$

Можно показать, что решение получается путем распределения m единиц среди φ_i, причем на каждом шаге одна единица придается аргументу той функции φ_i, которая имеет наименьшее приращение. Таким образом, вектор $x^{(m)}$ удовлетворяет ограничениям и минимизирует целевую функцию.

Упражнение 4.18. В задаче, сформулированной в теореме 4.5, положим $\varphi_i = \varphi_j$, $i, j = 1, \ldots, n$. Покажите, что

$$x_j = \left[\frac{m}{n}\right], \quad j \in S,$$

$$x_j = \left[\frac{m}{n}\right] + 1 \text{ в противном случае,}$$

где $[m/n]$ — целая часть m/n, а S — подмножество целых чисел, состоящее в точности из $n - m + n\,[m/n]$ элементов.

Упражнение 4.19. Решите задачу из упражнения 4.18 при помощи метода множителей Лагранжа.

Упражнение 4.20. Покажите, что предыдущие рассуждения применимы к задаче минимизации выражения

$$\sum_{j=1}^{n} w_j p_j^{x_j}$$

при ограничениях $\sum_{j=1}^{n} x_j = m$, где $x_j \geqslant 0$, $j = 1, \ldots, n$,— целые числа, m — положительное целое число, $w_j > 0$ — действительные и $0 < p_j < 1$.

Здесь m — число единиц данного оружия, которое необходимо для уничтожения n вражеских целей, $x_j \geqslant 0$, $j = 1, \ldots, n$,— количество оружия, предназначенное для поражений j-й цели, которой приписана стоимость w_j, а вероятность поражения цели при использовании одной единицы оружия равна q_j. Вероятность выживания цели в ходе атаки равна $p_j^{x_j} = (1 - q_j)^{x_j}$, и задача заключается в максимизации ожидаемой стоимости пораженных целей (или в минимизации стоимости сохранившихся целей). ∎

Следующая теорема интересна тем, что она полезна в некоторых приложениях, и доказательство ее можно провести несколькими различными способами.

Теорема 4.6. *Если* $x_1, \ldots, x_n$ — *положительные целые числа, такие, что*

$$\prod_{i=1}^{n} x_i \text{ } \textit{максимально при } n, \textit{ которое надо определить}, \tag{4.17}$$

и

$$\sum_{i=1}^{n} x_i = C, \tag{4.18}$$

где C *— данное положительное целое число, и если*

$$C = 3k, \quad \text{то} \quad x_i = 3, \qquad i = 1, \ldots, n,$$
$$C = 3k + 2, \quad \text{то} \quad x_1 = 2,\ x_i = 3, \qquad i = 2, \ldots, n,$$
$$C = 3k + 1, \quad \text{то} \quad x_1 = 2,\ x_2 = 2,\ x_i = 3, \qquad i = 3, \ldots, n.$$

Напомним, что n *должно быть найдено.*

Доказательство 1. Заметим сначала, что в наборе x_i, доставляющих максимум произведению, $x_i \neq 1$. Если $x_i = 4$ при некотором i, то x_i можно заменить на $x_j = 2$, $x_k = 2$; если $x_i > 4$, то x_i можно заменить на $x_j = 3$ и $x_k = x_i - 3$, произведение которых $3\,(x_i - 3)$ превосходит x_i. Следовательно, все множители равны 2 или 3. Кроме того, сомножители числа 2^3 в сумме дают столько же, сколько сомножители 3^2, и, следовательно, 2^3 можно заменить на 3^2. Оставшаяся часть доказательства вытекает непосредственно.

Заметим, например, что если $C = 100$, то, поскольку x_i не равно 1 ни при каком i, нельзя брать 3^{33}, и, следовательно, $x_1 = 2$, $x_2 = 2$, $x_i = 3$, $i = 3, \ldots, 34$. Произведение равно $2^2 \cdot 3^{32}$. Другое решение имеет вид $x_1 = 4$, $x_i = 3$, $i = 2, \ldots, 33$.

Доказательство 2. Другое рассуждение доказывает, что x_i должны быть равны (или приблизительно равны ввиду дискретности). Приведем здесь только доказательство равенства.

Рассмотрим экспоненциальное среднее k-го порядка

$$X_k = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^k\right)^{1/k};$$

X_2 — среднеквадратическое значение, X_1 — арифметическое среднее, X_0 — среднее геометрическое значение и X_{-1} — гармоническое среднее. Заметим, используя первые два члена разложения в ряд экспоненты, что

$$X_0 = \lim_{k\to 0} X_k = \lim_{k\to 0}\left(\frac{1}{n}\sum_{i=1}^{n} e^{k \ln x_i}\right)^{1/k} = \lim_{k\to 0}\left(1 + \frac{k}{n}\sum_{i=1}^{n} \ln x_i\right)^{1/k} =$$

$$= \lim_{k\to 0}\left(1 + \frac{(1/n)\sum_{i=1}^{n} \ln x_i}{1/k}\right)^{1/k} = \exp\left(\frac{1}{n}\sum_{i=1}^{n} \ln x_i\right) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}.$$

Покажем теперь, что $X_k \geqslant X_j$ при $k > j$. Заметим, что равенство имеет место только при $x_1 = x_2 = \ldots = x_n$. Пусть

$$0 < r < x,\ r = s\alpha,\ 0 < \alpha < 1,$$
$$p_i a_i^s = u_i,\ p_i = v_i,\ v_i > 0$$

и

$$p_i a_i^{s\alpha} = (p_i a_i^s)^\alpha\, p_i^{1-\alpha} = u_i^\alpha v_i^{1-\alpha};$$

тогда

$$\sum u_i^{\alpha} v_i^{1-\alpha} < \left(\sum u_i\right)^{\alpha} \left(\sum v_i\right)^{1-\alpha},$$

кроме того случая, когда u_i/v_i не зависят от i или просто a_i не зависят от i. Это следует из того, что [17]

$$\gamma_j > 0,\ j = 1, \ldots, n,\ \sum_{j=1}^{n} \gamma_j = 1$$

влечет за собой

$$\sum a_i^{\gamma_1} b_j^{\gamma_2} \ldots l_m^{\gamma_n} < \left(\sum a_i\right)^{\gamma_1} \left(\sum b_j\right)^{\gamma_2} \ldots \left(\sum l_m\right)^{\gamma_n},$$

за исключением случая, когда все a_i, b_j, . . .; l_m пропорциональны между собой или одна из переменных равна нулю при всех значениях индексов. Итак,

$$\left(\frac{\sum p_i a_i^{s\alpha}}{\sum p_i}\right)^{1/s\alpha} < \left(\frac{\sum p_i a_i^{s}}{\sum p_i}\right)^{1/2}.$$

Мы получим искомый результат, если заменим $s\alpha$ на r и положим $p_i = 1/n$.

Теперь требуется, чтобы геометрическое среднее достигало максимума при ограничении на арифметическое среднее. Предыдущие рассуждения показывают, что $X_1 \geqslant X_0$ и, следовательно, X_1 достигается при тех значениях ограничивающей константы, которые делятся на n, тогда и только тогда, когда все x_i равны между собой. Чтобы выполнялись граничные условия, надо действовать так же, как и ранее

Доказательство 3. Индуктивный подход. Разобьем множество решений $S = \{x_1, \ldots, x_n\}$ на два подмножества $S_1 = \{y_1, \ldots, y_r\}$ и $S_2 = \{z_1, \ldots, z_s\}$, где $r + s = n$. Предположим, что $\sum_{i=1}^{r} y_i = Q_1$, $\sum_{i=1}^{s} z_i = Q_2$ и $Q_1 + Q_2 = C$. Тогда $\prod_{i=1}^{r} y_i$ максимально для всех разбиений Q_1 и $\prod_{i=1}^{s} z_i$ максимально для всех разбиений Q_2. В противном случае S не было бы решением, так как Q_1 и Q_2 — произвольные числа с единственным ограничением $Q_1 + Q_2 = C$. Таким образом, достаточно изучить одно произведение Πy_i. Мы приходим к задаче максимизации этого произведения при меньшем значении ограничивающей константы Q_1.

Теорема 4.7. *Пусть Q — целое число, не меньшее двух.*

а) Если $Q_1 \equiv 0 \bmod 3$, то $\max \prod_{i=1}^{r} y_i = 3^k$ *и* $Q_1 = 3k = 3 + \ldots + 3$ (k *троек*).

б) Если $Q_1 \equiv 1 \bmod 3$, *то* $\max \prod_{i=1}^{r} y_i = 4 \cdot 3^{k-1}$ и $Q_1 = 3k + 1 = 4 + (3 + \ldots + 3)$ ($k - 1$ *троек*), или $Q_1 = 3k + 1 = 2 + 2 + (3 + \ldots + 3)$ ($k - 1$ *троек*).

в) Если $Q_1 \equiv 2 \bmod 3$, *то* $\max \prod_{i=1}^{r} y_i = 2 \cdot 3^k$ *и* $Q_1 = 3k + 2 = 2 + (3 + \ldots + 3)$ (k *троек*).

Доказательство 4. Предположим, что x_i, $i = 1, \ldots, n$,— решение. Выберем два сомножителя x_j и x_k и составим $z_j = x_j - 1$, $z_k = x_k + 1$, приняв, что $x_j \leqslant x_k$. Тогда

$$\sum_{i \neq j,\, k}^{n} x_i + z_j + z_k = C,$$

$$\prod_{i \neq j,\, k}^{n} x_i z_j z_k = (x_j - 1)(x_k + 1) \prod_{i \neq j,\, k}^{n} x_i \leqslant \prod_{i=1}^{n} x_i.$$

После сокращения получаем

$$x_j x_k \geqslant (x_j - 1)(x_k + 1), \quad \text{или} \quad x_k + 1 \geqslant x_j.$$

С другой стороны, если записать $z_j = x_j + 1$, $z_k = x_k - 1$, то получим

$$x_j x_k \geqslant (x_j + 1)(x_k - 1), \quad \text{или} \quad x_j \geqslant x_k - 1.$$

Отсюда

$$x_k + 1 \geqslant x_j \geqslant x_k - 1,$$

что имеет место для любых двух сомножителей x_j и x_k. Таким образом, некоторые (возможно, все) сомножители должны быть равны целому числу x_0, а остальные сомножители должны быть равны $x_0 - 1$. Задача сводится к определению x_0, n_1 и n_2, таких, что они максимизируют

$$x_0^{n_1} (x_0 - 1)^{n_2}$$

при ограничении

$$n_1 x_0 + n_2 (x_0 - 1) = C.$$

Выяснив, что сомножители должны быть равны 2 или 3, доказательство можно завершить следующим образом (в данном случае $C = 100$). Прежде всего, чтобы определить степени, запишем произведение $2^r 3^s$, где $2r + 3s = 100$, или $r = {}^1/_2 (100 - 3s)$. Затем найдем s, которое максимизирует $2^{50 - 3s/2}/3^s$. Эта величина максимальна, когда ее логарифм достигает максимума. Выражение

$$\left[50 - \frac{3}{2} s\right] \log 2 + s \log 2 = s \left(\log 3 - \frac{3}{2} \log 2\right) + 50 \log 2$$

достигает максимума при наибольшем возможном s, таком, что 3^s не превышает 100. Так как x_i не может быть равно 1, $s \neq 33$. Тогда $s = 32$ и решение имеет вид $3^{32} \cdot 2^2$.

Эвристический подход. Откажемся от требования, что x_i должно быть целым числом. Кроме того, ограничения в виде неравенств $x_i > 0$ сведем к ограничениям в виде равенств $x_i - y_i^2 = 0$, где y_i — действительное число. Задача состоит в том, чтобы найти y_i, которые максимизируют

$$\prod_{i=1}^{n} y_i^2 \tag{4.19}$$

при ограничении

$$\sum_{i=1}^{n} y_i^2 = 100. \tag{4.20}$$

Заметим, что в исходной задаче требуется определить точку на поверхности симплекса, а во второй — точку на поверхности сферы.

Чтобы максимизировать $\sum_{i=1}^{n} \log y_i^2$ при ограничении $\sum_{i=1}^{n} y_i^2 = 100$, составим функцию Лагранжа

$$F(y_1, \ldots, y_n, \lambda) = \sum_{i=1}^{n} \log y_i^2 + \lambda \Big(\sum_{i=1}^{n} y_i^2 - 100 \Big).$$

Тогда

$$\frac{\partial F}{\partial y_i} = \frac{2}{y_i} + 2\lambda y_i = 0, \quad i = 1, \ldots, n.$$

Таким образом, $y_i^2 = -1/\lambda$. Подставляя это значение в уравнение (4.20), получим $\lambda = -(n/100)$, $y_i^2 = 100/n$. Определим теперь n, которое максимизирует

$$\prod_{i=1}^{n} \frac{100}{n} = \Big(\frac{100}{n}\Big)^n.$$

Снова, рассматривая непрерывную задачу, находим, что $(100/x)^x$ максимально при $x = 100/e = 36{,}8$. Используя ближайшее целое число, чтобы получить оценку для n (т. е. $n = 36$ или $n = 37$) и затем для y_i^2, находим $y_i^2 = 3$. Такой выбор y_i^2 приводит к уменьшению значения n, чтобы выполнялось условие $\sum_{i=1}^{n} y_i^2 = 100$. Однако $3^{32} \cdot 2^2 > 3^{33} \cdot 1$. Последняя часть доказательства представляет собой тщательное рассмотрение эффектов на концах. Заметим, например, что $1^{100} < 2^{50} < e^{100/e} > 3^{100/3} > 4^{25} > 5^{20} > 6^{100/6} > \ldots$.

Замечание. Полезно отметить, что в непрерывном случае (который дает представление о том, каким должен быть ответ) известно следующее. Если $x_1, \ldots, x_n$ — положительные числа и $a_1, \ldots, a_n$,

C — положительные константы, то максимум $\prod_{i=1}^{n} x_i$ при ограничении $\sum_{i=1}^{n} a_i x_i = C$ достигается при $a_1 x_1 = \ldots = a_n x_n$. При тех же условиях минимум $\sum_{i=1}^{n} a_i x_i$ при ограничении $\prod_{i=1}^{n} x_i = C$ достигается при $a_1 x_1 = \ldots = a_n x_n$. Эти два утверждения можно доказать при помощи метода множителей Лагранжа. Если $a_i = 1$, $i = 1, \ldots, n$, то соотношение между арифметическим и геометрическим средними (см. разд. 4.6) дает возможность решить эти задачи.

Упражнение 4.21. Минимизируйте в неотрицательных целых числах $\sum_{i=1}^{n} x_i^{-1}$ при ограничении $\sum_{i=1}^{n} x_i = C$, где C — данное положительное целое число.

Упражнение 4.22. Максимизируйте в неотрицательных целых числах $\sum_{i=1}^{n} x_i$ при ограничении $\sum_{i=1}^{n} x_i^{-1} = C$, где C — данное положительное целое число.

Теорема 4.8. *Положительные целые числа* $x_1, \ldots, x_n$ *максимизируют* $\prod_{i=1}^{n} x_i^i$ *при ограничении* $\sum_{i=1}^{n} x_i = C$ *тогда и только тогда, когда* $x_i = 1$, 2 или 3, $i = 1, \ldots, n$. (*Здесь* C — *положительное целое число, а* n *не фиксировано.*)

Теорема 4.9. *Если* α, β *и* γ — *количества единиц, двоек и троек соответственно в решении, данном теоремой* 4.8, *то максимум достигается в точках целочисленной решетки, определенных следующим образом* [28]:

$$(2C + 1)\log 2 - (C + 1)\log 3 \leqslant (3\beta + 6\gamma)\log 2 - (\beta + 3\gamma)\log 3 \leqslant (2C + 1)\log 2 - C \log 3,$$

$$(C - 1)(\log 3 - \log 2) \leqslant (\beta + 4\gamma)\log 3 - 3\gamma \log 2 \leqslant (C + 2)(\log 3 - \log 2),$$

$$(C - 2)\log 3 \leqslant 3\beta \log 2 + (\beta + 5\gamma)\log 3 \leqslant (C + 3)\log 3,$$

$$(C - 4)\log 2 \leqslant 3(\beta + \gamma)\log 2 + \gamma \log 3 \leqslant (C + 2)\log 2.$$

Прежде чем доказывать теоремы 4.8 и 4.9, сформулируем лемму.

Лемма 4.1. *Если решение представляет собой вектор* $x^0 = (x_1^0, \ldots, x_n^0)$, *то* $x_i^0 \geqslant x_j^0$ *для* $i > j$.

Доказательство. Очевидно, выгоднее самое большое число поместить в самом конце, перед ним второе по величине и т. д.

Доказательство теоремы 4.8. Если $x_i = 4$, то 4^k можно заменить на k-й позиции на 2^k, а другой множитель 2^{k+1} поставить на $(k+1)$-й позиции; тогда получим $4^k < 2^k \cdot 2^{k+1}$. Если $x_i > 4$, то $x_i = 3 + (x_i - 3)$ и $x_i < 3(x_i - 3)$. Этот процесс можно продолжить, переходя к $(x_i - 3)$. Тогда запишем $x_i - 3 = 3 + [(x_i - 3) - 3]$, $x_i - 6 = 3 + (x_i - 6) - 3$ и т. д. Заметим также, что $x_i^i < 3^i (x_i - 3)^i$ и т. д.

Доказательство теоремы 4.9. Пусть α, β и γ показывают, сколько раз встречаются множители 1, 2 и 3 соответственно; рассмотрим произведение

$$\prod_{i=1}^{n} x_i^i = 1 \cdot 1^2 \ldots 1^{\alpha} \cdot 2^{\alpha+} \ldots 2^{\alpha+\beta} \cdot 3^{\alpha+\beta+1} \ldots 3^{\alpha+\beta+\gamma} =$$

$$= 2^{\beta(2\alpha+\beta+1)/2} \cdot 3^{\gamma(2\alpha+2\beta+\gamma+1)/2},$$

где $\alpha + 2\beta + 3\gamma = C$. Задача состоит в том, чтобы найти α, β и γ, которые максимизируют произведение. Предположим, что при некоторых значениях α, β и γ достигается максимум. Введем отклонения δ_1 в α, δ_2 в β и δ_3 в γ. Тогда

$$\delta_1 + 2\delta_2 + 3\delta_3 = 0.$$

Мы ищем фундаментальное множество решений этого уравнения (существует конечное множество таких решений), таких, что любое другое решение представляет собой линейную комбинацию этих решений, взятых с целочисленными коэффициентами. Чтобы получить эти решения для вышеприведенного уравнения, воспользуемся производящей функцией [20]

$$\frac{1}{(1-\delta_1^m\delta_2)(1-\delta_1^b\delta_3^a)}$$

уравнения

$$a\delta_1 = ma\delta_2 + b\delta_3,$$

где b взаимно просто с a.

Фундаментальные решения состоят из степеней δ_i в каждом сомножителе знаменателя. Для данной задачи $a = 1$, $m = -2$, $b = -3$, и мы получаем два фундаментальных решения $(-2, 1, 0)$ и $(-3, 0, 1)$. Поскольку, как мы увидим позже, нашей целью является построение области, в которой находится решение задачи оптимизации, возьмем дополнительные решения. (Идеи станут понятнее по мере продолжения разбора примера.) Будем использовать следующие решения уравнения в качестве малых отклонений от опти-

мума:

$$\begin{array}{ll} (1,\ \ 1,\ -1) & (-1,\ -1,\ \ 1) \\ (1,\ -2,\ \ 1) & (-1,\ \ 2,\ -1) \\ (3,\ \ 0,\ -1) & (-3,\ \ 0,\ \ 1) \\ (2,\ -1,\ \ 0) & (-2,\ \ 1,\ \ 0) \end{array}$$

Подставляя в произведение сначала α, β и γ, потом $\alpha + \delta_1$, $\beta + \delta_2$, $\gamma + \delta_3$ и деля второй результат на первый, получим выражение, не превосходящее единицы. Логарифмируя обе части, получаем

$$\frac{\beta+\delta_2}{2}[2(\alpha+\delta_1)+\beta+\delta_2+1]\log 2+$$

$$+\frac{\gamma+\delta_3}{2}[2(\alpha+\delta_1)+2(\beta+\delta_2)+\gamma+\delta_3+1]\log 3-\frac{\beta}{2}(2\alpha+\beta+1)\log 2-$$

$$-\frac{\gamma}{2}2\alpha+2\beta+\gamma+1)\log 3\leqslant 0$$

или

$$\left[\alpha\delta_2+\beta(\delta_1+\delta_2)+\delta_1\delta_2+\frac{1}{2}\delta_2^2+\frac{1}{2}\delta_2\right]\log 2+$$

$$+\left[\alpha\delta_3+\beta\delta_3+\gamma(\delta_1+\delta_2+\delta_3)+\delta_1\delta_3+\delta_2\delta_3+\frac{1}{2}\delta_3^2+\frac{1}{2}\delta_3\right]\log 3\leqslant 0.$$

Если положить $\alpha = C - 2\beta - 3\gamma$ и $\delta_1 = -2\delta_2 - 3\delta_3$, то после упрощений будем иметь

$$[3\beta(\delta_2+\delta_3)+3\gamma\delta_2]\log 2+[\beta\delta_3+\gamma(\delta_2+5\delta_3)]\log 3\geqslant$$

$$\geqslant\left(C\delta_2-\frac{3}{2}\delta_2^2-3\delta_2\delta_3+\frac{1}{2}\delta_2\right)\log 2+$$

$$+\left(C\delta_3-\delta_2\delta_3-\frac{5}{2}\delta_3^2+\frac{1}{2}\delta_3\right)\log 2.$$

Воспользуемся теперь значениями $(\delta_1, \delta_2, \delta_3)$ и запишем после каждого из них результирующее неравенство: $(1, -2, 1)$:

$$(3\beta + 6\gamma)\log 2 - (\beta + 3\gamma)\log 3 \leqslant (2C + 1)\log 2 - C\log 3;$$

$(-1, 2, -1)$:

$$(3\beta + 6\gamma)\log 2 - (\beta + 3\gamma)\log 3 \geqslant (2C + 1)\log 2 - (C + 1)\log 3.$$

Комбинируя эти два результата, получаем

$$(2C + 1)\log 2 - (C + 1)\log 3 \leqslant (3\beta + 6\gamma)\log 2 - (\beta + 3\gamma)\log 3 \leqslant$$
$$\leqslant (2C + 1)\log 2 - C\log 3$$

$(1, 1, -1)$:

$$3\gamma\log 2 - (\beta + 4\gamma)\log 3 \geqslant (C + 2)\log 2 - (C + 2)\log 3;$$

$(-1, -1, 1)$:

$$-3\gamma\log 2 + (\beta + 4\gamma)\log 3 \geqslant -(C - 1)\log 2 + (C - 1)\log 3.$$

Комбинируя эти два результата, получаем

$$(C-1)(\log 3-\log 2)\leqslant(\beta+4\gamma)\log 3-3\gamma\log 2\leqslant$$
$$\leqslant(C+2)(\log 3-\log 2).$$

(3, 0, —1):

$$-3\beta\log 2-(\beta+5\gamma)\log 3\geqslant-(C+3)\log 3;$$

(—3, 0, 1):

$$3\beta\log 2+(\beta+5\gamma)\log 3\geqslant(C-2)\log 3.$$

Комбинируя эти два результата, получаем

$$(C-2)\log 3\leqslant 3\beta\log 2+(\beta+5\gamma)\log 3\leqslant(C+3)\log 3.$$

(2, —1, 0):

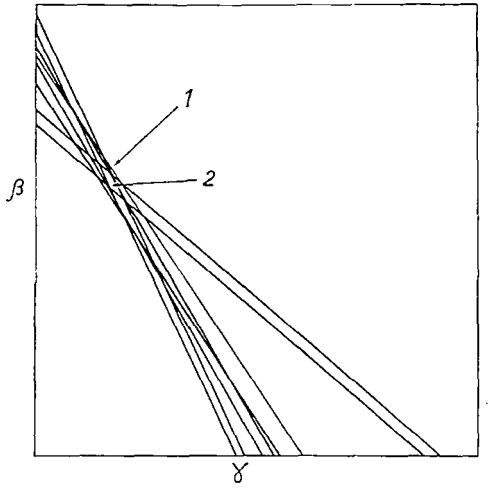
$$-3(\beta+\gamma)\log 2-\gamma\log 3\geqslant-(C+2)\log 2;$$

(—2, 1, 0):

$$3(\beta+\gamma)\log 2+\gamma\log 3\geqslant(C-4\log 2.$$

Комбинируя эти два результата, получаем

$$(C-4)\log 2\leqslant 3(\beta+\gamma)\log 2+\gamma\log 3\leqslant(C+2)\log 2.$$



Фиг. 4.3.

1 — решение (6, 24); *2* — область допустимых решений — две точки решетки (6, 24) и (7, 23).

Для примера возьмем $C=100$. При этом получим

$$27{,}6\leqslant 0{,}97\beta+0{,}84\gamma\leqslant 28{,}7,$$
$$40{,}19\leqslant 1{,}1\beta+2{,}33\gamma\leqslant 41{,}82,$$
$$107{,}8\leqslant 3{,}17\beta+5{,}5\gamma\leqslant 113{,}3,$$
$$66{,}24\leqslant 2{,}08\beta+3{,}18\gamma\leqslant 70{,}7.$$

Заметим, что средние части выражений имеют один и тот же вид при любом значении C. Кроме того, может оказаться желательным построение большего числа неравенств при наличии дополнительных значений (δ_1, δ_2, δ_3). Один из подходов к решению состоит в том, что эти неравенства представляются графически. Это облегчает поиск решения, которое должно быть точкой с целочисленными координатами в общей области, определяемой пересечениями прямых. На фиг. 4.3 графически представлена область решения, определяемая восемью неравенствами. В ней имеются две точки с целочисленными координатами. В качестве решения получаем $\beta=6$, $\gamma=24$ и, следовательно, $\alpha=34$. Область сохраняет свой вид при любом C, но положение ее на плоскости меняется.

Неправильное использование метода множителей Лагранжа. Использование функции Лагранжа и, следовательно, непрерывный подход к этой задаче при $C = 100$ дает результат, далекий от истинного. Рассмотрим кратко этот подход. Составим функцию

$$F(x_1, \ldots, x_n, \lambda) = x_1 x_2^2 x_3^3 \ldots : x_n^n - \lambda\left(\sum x_i - 100\right),$$

$$\frac{\partial F}{\partial x_1} = x_2^2 x_3^3 \ldots x_n^n - \lambda = 0,$$

$$\frac{\partial F}{\partial x_2} = x_1 2 x_2 x_3^3 \ldots x_n^n - \lambda = 0,$$

$$\frac{\partial F}{\partial x_i} = x_1 x_2^2 \ldots i x_i^{i-1} x_{i+1}^{i+1} \ldots x_n^n - \lambda = 0.$$

Умножая i-е уравнение на x_i и складывая, получаем

$$\frac{n(n+1)}{2} \prod_{i=1}^{n} x_i^i - 100\lambda = 0.$$

Таким образом,

$$\lambda = \frac{n(n+1)}{200} \prod_{i=1}^{n} x_i^i$$

и i-е уравнение после умножения на x_i дает

$$x_i = i \prod_{i=1}^{n} \frac{x_i^i}{\lambda},$$

или

$$x_i = \frac{200i}{n(n+1)}.$$

Задача теперь состоит в том, чтобы найти n, которое максимизирует $\prod_{i=1}^{n} [200i/n(n+1)]^i$, или просто его логарифм. Отсюда приходим к соотношению

$$\sum_{i=1}^{n} \left(i \log \frac{200}{n(n+1)} + i \log i\right) = \frac{n(n+1)}{2} \log \frac{200}{n(n+1)} + \sum_{i=1}^{n} i \log i.$$

Теперь из того, что $\int_z^{z+1} \log \Gamma(\zeta)\, d\zeta = z \log z - z + {}^1/_2 \log 2\pi$, $|\arg z| < \pi$, предыдущее выражение равно $[n(n+1)/2]\,[\log 200 - \log n(n+1) + 1] + \int_1^{n+1} \log \Gamma(\zeta)\, d\zeta - (n/2) \log 2\pi$. После диф-

ференцирования по n и приравнивания к нулю, получим

$$\frac{(2n+1)}{2}\left[\log\frac{200}{n(n+1)}\right]+\log\Gamma(n+1)-\log\frac{2\pi}{2}=0,$$

или

$$\frac{\Gamma(n+1)}{\sqrt{2\pi}}=\left(\frac{n(n+1)}{200}\right)^{n+1/2}.$$

Используя формулу Стирлинга, окончательно получаем равенство

$$e^{-1/2}=\left(\frac{200}{e(n+1)}\right)^{n+1/2},$$

которое выполняется, если n находится между 73 и 74. Поскольку $n=64$ является правильным ответом, можно сделать вывод, что этот подход не очень помогает при поиске решения. Кроме того, возникают трудности при подборе целочисленных значений для x_i.

Метод, предложенный в теореме 4.9, можно обобщить на задачу максимизации $\prod_{j=1}^{n} x_j$ при ограничении $\sum_{j=1}^{n} x_j^p = C$, $p>1$ — целое число. Эта задача сводится к отысканию α единиц, β двоек и γ троек, таких, что

$$\alpha+2^p\beta+3^p\gamma=C.$$

Отклонения приводят к уравнению

$$\delta_1+2^p\delta_2+3^p\delta_3=0,$$

которое имеет производящую функцию

$$\frac{1}{(1-\delta_1^{-2^p}\delta_2)(1-\delta_1^{-3^p}\delta_3)}$$

и функциональные решения $(-2^p, 1, 0)$ и $(-3^p, 0, 1)$.

Такой же метод применим и для случая максимизации $\prod_{j=1}^{n} x_j^j$ при ограничении $\sum_{j=1}^{n} x_j^p = C$, $p>1$ — целое число. Решение будет произведением единиц, двоек и троек. Если $C=100$, то решение второй задачи имеет вид

$$\begin{array}{ll} p=2 & (44, 14, 0)\\ p=3 & (52, 6, 0)\\ p=4 & (52, 3, 0)\\ p=5 & (68, 1, 0)\\ p=6 & (100, 0, 0)\end{array}$$

Замечание. Положительное целочисленное решение, максимизирующее $\sum_{i=1}^{n} a_i x_i$ при ограничении $\prod_{i=1}^{n} x_i = C$, где $a_k = \max(a_1, \ldots, a_n)$ без ограничения общности, имеет вид $x_k = C$, $x_i = 1$, $i \neq k$. В этом легко убедиться, если максимизировать частное от деления $\sum_{i=1}^{n} a_i x_i$ на $\prod_{i=1}^{n} x_i = C$. Этот результат легко обобщить на случай максимизации $\sum_{j=1}^{n} a_j x_j^p$, где p — положительное целое число при том же ограничении.

Соответствующую задачу минимизации можно решить следующим образом. Предположим, что $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$ и пусть $C = p_1^{\alpha_1} p_2^{\alpha_2} \ldots p_s^{\alpha_s}$, где $p_1 < p_2 < \ldots < p_s$. Если $\alpha_1 + \alpha_2 + \ldots + \alpha_s \leqslant n$, то минимум достигается путем приписывания значений простых чисел p_s, p_{s-1} и т. д. каждому x_i в порядке убывания и затем использования единиц в конце, если это окажется необходимым. В общем случае если мы пишем $x_j = p_1^{\alpha_{1j}} \ldots p_s^{\alpha_{sj}}$, $j = 1, \ldots, n$, то задача состоит в том, чтобы найти неотрицательные целые числа α_{ij}, которые минимизируют $\sum_{j=1}^{n} a_j x_j$ при линейных ограничениях

$$\sum_{j=1}^{n} \alpha_{ij} = \alpha_i, \quad i = 1, \ldots, s.$$

Рассмотрим теперь задачу максимизации в целых положительных числах $\sum_{j=1}^{n} \alpha_j x_j^p$, где α_j — данные натуральные числа, p — положительное целое число при ограничении $\prod_{j=1}^{n} x_j^j = C$, где C — данное положительное целое число.

Лемма 4.2. *Пусть k — такой номер, что $\alpha_k \geqslant \alpha_j$, $j = 1, \ldots, n$, тогда $x_j = 1$, $j = k + 1, \ldots, n$.*

Доказательство. При условии, что задан вектор $x_1, \ldots, x_n$, определим новый вектор соотношениями

$$\begin{aligned} \bar{x}_1 &= x_1 x_{k+1}, \\ \bar{x}_k &= x_k x_{k+1}, \\ \bar{x}_j = x_j, \quad & j = 2, \ldots, n, \quad j \neq k, \\ \bar{x}_{k+1} &= 1. \end{aligned}$$

Очевидно, что произведение не изменяется, так как

$$\prod_{j=1}^{n} \bar{x}_j^j = (x_1 x_{k+1})\, \bar{x}_2^2 \ldots (x_k x_{k+1})^k \cdot 1 \ldots \bar{x}_n^n = x_1 x_2^2 \ldots x_k^k x_{k+1}^{k+1} \ldots x_n^n = C.$$

Теперь покажем, что сумма не уменьшилась после преобразования, т. е.

$$\alpha_1 \bar{x}_1^p + \sum_{j=2}^{k-1} \alpha_j \bar{x}_j^p + \alpha_k \bar{x}_k^p + \alpha_{k+1} + \sum_{j=k+2}^{n} \alpha_j \bar{x}_j^p \geqslant \sum_{j=1}^{n} \alpha_j x_j^p.$$

Это следует из того, что

$$\alpha_k \bar{x}_k^p = \alpha_k x_k^p x_{k+1}^p \geqslant \alpha_k x_k^p + \alpha_k x_{k+1}^p \geqslant \alpha_k x_k^p + \alpha_{k+1} x_{k+1}^p,$$

где $x_k \neq 1$, $x_{k+1} \neq 1$, и из

$$\alpha_1 \bar{x}_1 \geqslant \alpha_1 x_1.$$

Результат получается непосредственно, если $x_k = 1, x_{k+1} = 1$. Те же рассуждения справедливы для всех $j \geqslant k$.

Лемма 4.3. *При тех же предположениях, что и в лемме* 4.2, *из*

$$\prod_{j=1}^{n} x_j^j = \prod_{\substack{j=1 \\ j \neq i}}^{n} x_j^j \,(x_i^{i/k})^k = C, \quad 1 \leqslant i \leqslant k,$$

если положить

$$\bar{x}_k = x_k x_i^{i/k}, \quad \bar{x}_i = 1, \quad \bar{x}_j = x_j, \quad j \neq k, \tag{4.21}$$

где x_i *удовлетворяет неравенству*

$$(x_i^{i/k})^p > \frac{x_i^p - 1}{x_k^p} + 1 \tag{4.22}$$

и $x_i^{i/k}$ — *целое число, следует, что*

$$\sum_{\substack{i=1 \\ j \neq i}}^{k-1} \alpha_j \bar{x}_j^p + \alpha_i + \alpha_k \bar{x}_k^p + \sum_{j=k+1}^{n} \alpha_j \geqslant \sum_{j=1}^{k} \alpha_j x_j^p + \sum_{j=k+1}^{n} \alpha_j.$$

Доказательство. Из $\alpha_k \geqslant \alpha_i$ и неравенства (4.22) получаем

$$\alpha_k x_k^p x_i^{(i/k)p} - \alpha_k x_k^p \geqslant \alpha_i x_i^p - \alpha_i.$$

Соотношение (4.21) дает

$$\alpha_k \bar{x}_k^p + \alpha_i > \alpha_i x_i^p + \alpha_k x_k^p,$$

откуда следует результат. Получаем следующую теорему:

Теорема 4.10. *Необходимым и достаточным условием того, что допустимое решение можно преобразовать так, что оно приблизится к оптимальному решению, является существование* x_i, $1 \leqslant i \leqslant k$, *такого, что* $x_i^{i/k}$ — *целое число, удовлетворяющее условию* (*4.22*). *В* (*4.21*) *используется наибольшее такое* x_i.

Упражнение 4.23. Найдите положительные целые x_i, которые максимизируют следующие выражения:

а) $x_1^m \prod\limits_{i=2}^{n} x_i$ при ограничении $\prod\limits_{i=1}^{n} x_i = 100$ для целого $m > 1$.

Ответ. Положите $x_i = 3$, $x_1 = [mx_i]$. Квадратные скобки означают наибольшее целое число, меньшее mx_i, такое, что $100 - [mx_i]$ делится на 3. Таким образом, если $m = 3$, то $mx_i = 9$, $100 - 9 \not\equiv 0 \bmod 3$ и $100 - 8 \not\equiv 0 \bmod 3$, но $100 - 7 \equiv 0 \bmod 3$; следовательно, $x_1 = 7$.

б) $\prod\limits_{i=1}^{n} x_i$ при ограничении $x_1^p + \sum\limits_{i=2}^{n} x_i = 100$ при целом $p > 1$.

Ответ. Положите $x_1 = 1$, $x_i = 3$.

в) $x_1^m \prod\limits_{i=2}^{n} x_i$ при ограничении $x_1^p + \sum\limits_{i=2}^{n} x_i = 100$, если $m > 1$ и $p > 1$ — целые числа.

г) $\prod\limits_{i=1}^{n} x_i^i$ при ограничении $\sum\limits_{i=1}^{n} x_i = 100$.

д) $\prod\limits_{i=1}^{n} x_i$ при ограничении $\sum\limits_{i=1}^{n} x_i + x_i^2 = 100$.

е) $\prod\limits_{i=1}^{n} x_i^p$ при ограничении $\prod\limits_{i=1}^{n} x_i^i = C$, C — данное положительное целое число.

ж) $\prod\limits_{i=1}^{n} x_i^i$ при ограничении $\prod\limits_{i=1}^{n} x_i = C$, C — данное положительное целое число.

Упражнение 4.24. Пусть p, N, r — положительные числа и $r > 1$. Найдите [14] разложение N на p неотрицательных целых слагаемых x_i, которые минимизируют сумму биномиальных коэффициентов $\sum\limits_{i=1}^{p} C(x_i, r)$, где

$$C(x_i, r) = \frac{x_i!}{r!\,(x_i - r)!}, \quad C(x_i, r) = 0, \text{ если } r > x_i.$$

Указание. Используя тот факт, что x_i должны быть по возможности равны друг другу, положите $N = pq + r$, где $p > r \geqslant 0$. Выберите

$$x_1 = x_2 = \ldots = x_r = q + 1 \quad \text{и} \quad x_{r+1} = \ldots = x_p = q.$$

4.6. Полезные неравенства

При анализе многих задач оптимизации для определения положения оптимума необходимо оценивать и уточнять границы. Для этой цели могут оказаться очень полезными некоторые классические неравенства. Для того чтобы читателям удобнее было пользоваться ими, приведем здесь некоторые из них. Рассмотрим выражение

$$\sum_{i=1}^{n} (a_i - \lambda b_i)^2,$$

которое представляет собой неотрицательное квадратичное выражение относительно λ, где a_i и b_i, $i = 1, \ldots, n$,— произвольные действительные числа. После разложения на слагаемые это выражение принимает вид

$$\sum_{i=1}^{n} a_i^2 - 2\lambda \sum_{i=1}^{n} a_i b_i + \lambda^2 \sum_{i=1}^{n} b_i^2.$$

Так как выражение неотрицательно при любом λ, дискриминант этого квадратичного уравнения относительно λ должен быть неположительным, т. е. удовлетворять соотношению (которое называется неравенством Коши — Шварца)

$$\Big(\sum_{i=1}^{n} a_i b_i\Big)^2 - \sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 \leqslant 0.$$

Обобщением неравенства Коши — Шварца является неравенство Гельдера

$$\Big|\sum_{i=1}^{n} a_i b_i\Big| \leqslant \Big(\sum_{i=1}^{n} |a_i|^p\Big)^{1/p} \Big(\sum_{i=1}^{n} |b_i|^q\Big)^{1/q},$$

где $1/p + 1/q = 1$ и $p > 0$, $q > 0$.

Отсюда можно получить

$$\frac{1}{n}\Big|\sum_{i=1}^{n} a_i b_i\Big| \leqslant \Big(\frac{1}{n}\sum_{i=1}^{n} |a_i|^p\Big)^{1/p} \Big(\frac{1}{n}\sum_{i=1}^{n} |b_i|^q\Big)^{1/q}.$$

Неравенство Минковского дает

$$\Big(\sum_{i=1}^{n} |a_i + b_i|^p\Big)^{1/p} \leqslant \Big(\sum_{i=1}^{n} |a_i|^p\Big)^{1/p} + \Big(\sum_{i=1}^{n} |b_i|^p\Big)^{1/p},$$

где $p \geqslant 1$. При $p \leqslant 1$ выполняется противоположное неравенство. Заметим, что равенство имеет место, если $p = 1$ или если $a_i = cb_i$ ($i = 1, \ldots, n$), где c — константа. И неравенство Гельдера (следовательно, и неравенство Коши — Шварца), и неравенство Минковского имеют место, если суммы заменить интегралами, a_i — функцией $f(x)$, а b_i — функцией $g(x)$. В случае неравенства Гельдера

требуется интегрируемость $|f(x)|^p$ и $|g(x)|^q$, а в случае неравенства Минковского требуется интегрируемость $|f(x)|^p$ и $|g(x)|^p$.

Неравенство Иенсена имеет вид

$$\left(\sum_{i=1}^{n} |a_i|^r\right)^{1/r} > \left(\sum_{i=1}^{n} |a_i|^s\right)^{1/s}, \text{ если } 0 < r < s.$$

Другие полезные неравенства:

$$(1+x)^p > 1 + px, \quad \text{если} \quad x > -1, \quad x \neq 0 \quad \text{и} \quad p > 1,$$

$x^p - 1 > p(x-1)$, если $x > 1$ и $p > 1$ — действительное число,

$$x^p - 1 < p(x-1), \quad \text{если} \quad x > 1 \quad \text{и} \quad 0 < p < 1,$$

$x^a y^b < ax + by$, если $x \neq y$ и $a + b = 1$, $a > 0$, $b > 0$,

$$x \leqslant \frac{1-p^x}{1-p} \leqslant x p^{x-1}, \text{ если } 0 \leqslant x \leqslant 1 \text{ и } 0 < p < 1.$$

Заметим, что в последнем неравенстве p может быть, например, таким выражением, как $1 - ae^{-by/x}$, где $a, b > 0$. Таким образом, это неравенство можно использовать для доказательства соотношения

$$\lim_{x \to 0} \frac{1-(1-ae^{-by/x})^x}{x} = 0,$$

которое представляет собой производную числителя в нуле.

Между гармоническим, геометрическим и арифметическим средними, если a_i — действительные числа, $i = 1, \ldots, n$, имеет место следующее соотношение (слева направо):

$$\frac{n}{1/a_1 + \ldots + 1/a_n} < \sqrt[n]{a_1 \ldots a_n} < \frac{a_1 + \ldots + a_n}{n}.$$

Каждое из этих средних ограничено сверху наибольшим из a_i, а снизу наименьшим из a_i. Если эти три средних определены для интегрируемых функций, то вышеприведенное соотношение принимает вид

$$\frac{b-a}{\int_a^b dx/f(x)} \leqslant e^{1/(b-a)} \int_a^b \log f(x)\, dx \leqslant \frac{1}{b-a} \int_a^b f(x)\, dx.$$

Чтобы геометрическое среднее было определено, функция $f(x)$ должна быть положительной.

Согласно теореме Иенсена, для выпуклой функции $f(x)$ и положительных a_i $(i = 1, \ldots, n)$ имеет место неравенство

$$f\left(\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i}\right) \leqslant \frac{\sum_{i=1}^{n} a_i f(x_i)}{\sum_{i=1}^{n} a_i},$$

где x_i $(i = 1, \ldots, n)$ — n произвольных значений x. Если положить $n = 2$, $a_1 = a_2 = 1$ и $x_1 < x_2$, то получается определение выпуклой функции.

Эта теорема имеет много интересных следствий и приложений как в случае сумм, так и в случае интегралов. Теорема применима к интегралам, определенным на отрезке (a, b), если суммы заменяются интегралами, x_i — ограниченной функцией $g(x)$ и a_i — неотрицательной функцией $a(x)$ с положительным значением интеграла на отрезке (a, b).

Общая теорема о гармонических, геометрических и арифметических средних дает при положительных a_i и положительных b_i, $i = 1, \ldots, n$, соотношение

$$\frac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} b_i/a_i} \leqslant \exp\left(\frac{\sum_{i=1}^{n} b_i \log a_i}{\sum_{i=1}^{n} b_i}\right) \leqslant \frac{\sum_{i=1}^{n} a_i b_i}{\sum_{i=1}^{n} b_i}.$$

Кроме того, получаем

$$\exp\left[\frac{\sum_{i=1}^{n} (b_i/a_i) \log a_i}{\sum_{i=1}^{n} b_i/a_i}\right] < \frac{\sum_{i=1}^{n} b_i}{\sum_{i=1}^{n} b_i/a_i}$$

и

$$\frac{\sum_{i=1}^{n} a_i b_i}{\sum_{i=1}^{n} b_i} < \exp\left(\frac{\sum_{i=1}^{n} a_i b_i \log a_i}{\sum_{i=1}^{n} a_i b_i}\right).$$

Эти три результата легко обобщаются на непрерывный случай. Более подробно эти вопросы изложены в книге [17].

4.7. Теория максимина

Существуют игровые задачи, которые не могут быть решены методами теории игр. Часто в таких задачах исследуется целевая функция $F(x, y)$, выражающая интересы двух соперников, при наличии ограничений на векторное решение каждого соперника, например вида $x \geqslant 0$, $y \geqslant 0$, $\sum_{i=1}^{n} x_i = X$, $\sum_{i=1}^{n} y_i = Y$. Функция $F(x, y)$ задается в аналитической форме, и решение, дающее минимум по y и максимум по x, должно быть целочисленным. Порядок

оптимизации не может быть изменен. Общей теории отыскания таких целочисленных решений не существует.

Пример. Сторона A должна защитить n городов, выделяя $x_i \geqslant 0$ единиц оружия i-му городу, причем $\sum_{i=1}^{n} x_i = X$. Каждый город может быть атакован стороной B, которая выделяет $y_i \geqslant 0$ единиц оружия для нападения на i-й город, $\sum_{i=1}^{n} y_i = Y$. Отношение x_i/y_i определяет относительную эффективность защиты i-го города. Если $k_i > 0$ — эффективность защиты и α_i, $0<\alpha_i<1$, — вероятность, что наступательное оружие поразит цель, то $\{1-\alpha_i \exp[-k_i(x_i/y_i)]\}^{y_i}$ — — вероятность того, что i-я цель переживет атаку с применением y_i единиц оружия. Если $v_i > 0$ — ценность i-й цели, то B хочет минимизировать ожидаемое значение потерь:

$$\sum_{i=1}^{n} v_i \left[1-\alpha_i \exp\left(-\frac{k_i x_i}{y_i}\right)\right]^{y_i},$$

тогда как A хочет максимизировать это минимальное значение. Каков наилучший выбор векторов с целочисленными координатами x и y?

Данскин [4а] разработал интересную тонкую теорию для решения задач такого типа (но не целочисленных). Изложим кратко материал, в котором приводится его главный результат.

Если $\varphi(x) = \min_{y} F(x, y)$, где $F(x, y)$ и ее частные производные $F_{x_i}(x, y)$ непрерывны по x и y и оба вектора x и y принадлежат некоторому подмножеству евклидова пространства, то $\varphi(x)$ может быть недифференцируемой функцией. Чтобы обойти эту трудность, необходимы новые методы. Направление $\gamma = (\gamma_1, \ldots, \gamma_n)$ (единичный вектор направляющих косинусов) в x-пространстве, исходящее из точки x^0, является *допустимым*, если существует дуга, один конец которой находится в точке x^0 и которая целиком лежит в множестве допустимых значений x (это может быть все пространство или подпространство, определяемое ограничениями), таких, что для любой последовательности x^m вдоль дуги, для которой $x^m \to x^0$, получаем $\gamma^m \to \gamma$, где γ определяется соотношением

$$\gamma_i^m \equiv \frac{x_i^m - x_i}{|x^m - x^0|}.$$

Теорема 4.11. *Необходимое условие того, что x^0 является максимумом, состоит в том, что производная по направлению*

$$D_\gamma \varphi(x^0) = \min_{y \in Y(x)} \sum_{i=1}^{n} \gamma_i F_{x_i}(x^0, y)$$

(*можно показать, что она существует в каждом допустимом направлении*) *неположительна.*

Здесь $Y(x)$ — множество векторов y, принадлежащих замкнутому и ограниченному подмножеству, которое дает минимум относительно $x = x^0$.

ЛИТЕРАТУРА

1. Andersson J., All Nonnegative Integer Solution $x + 2y + 3z + 5t = n$, *Mathesis*, 383 (1953).
2. Bond J., Calculating the General Solution of a Linear Diophantine Equation, *Am. Math. Monthly*, 74, 955 (Oct. 1967).
3. Carlitz L., Solution of a Problem Posed by L. Moser, *Pi Mu Epsilon J.*, 3, 232 (1961).
4. Carmichael R. D., The Theory of Numbers and Diophantine Analysis, Dover Publications, Inc., N.Y., 1915.

4a. Danskin J. M., The Theory of Max-Min, Springer-Verlag, New York, Inc., 1967.

5. Davenport H., Roth K. F., The Solubility of Certain Diophantine Inequalities, *Mathematika*, 2, 81 (Dec. 1955).
6. Delcourte M., *Mathesis*, 383 (1958); 272 (1959).
7. denBroeder G. G., Ellison R. E., Emerling L., On Optimum Target Assignments, *Operations Res.*, 7, 322 (May—June 1959).
8. Dickson L. E., History of the Theory of Numbers, Vols. 1—3, Chelsea Publ. Co., N.Y.
9. Erdös P., Some Results on Diophantine Approximation, *Act. Arithmet.*, 359 (1959).
10. Fox B., Discrete Optimization Via Marginal Analysis, *Management Sci.*, 13, 210 (Nov. 1966).
11. Garver R., The Solution of Problems in Maxima and Minima by Algebra, *Am. Math. Monthly*, 42, 435 (1935).
12. Гельфонд А. О., Решение уравнений в целых числах, Гостехтеориздат, 1956.
13. Gibrat M. R., Escadrilles d'avions, *Rev. Franc. Rech. Opération.*, № 27 (1963).
14. Goldman A. J., Langford E. S., Minimal Sum of Binomial Coefficients, *Am. Math. Monthly*, 785 (Sept. 1965).
15. Gross O., Class of Discrete Type Minimization Problems, pt. 30, RM-1644, Rand Corp., Santa Monica, Calif., Feb. 1956.
16. Hall H. S., Knight S. R., Higher Algebra, Macmillan, London, 1948.
17. Hardy G. H., Littlewood J. E., Polya G., Inequalities, Cambridge Univ. Press, N.Y., 1934; русский перевод: Харди Г. Г., Литтлвуд Д. Е., Полиа Г., Неравенства, ИЛ, 1948.
18. Horner W. W., Brousseau A., The Pancake Problem, *Math. Mag.*, 100 (March—April 1968).
19. Itard J., Arithmétique et Théorie Des Nombres, Presses Universitaires de France, Paris, 1963.
20. MacMahon P. A., Combinatory Analysis, Vols. 1 and 2, Chelsea Publ. Co., N.Y., 1960.
21. Mordell L. J., Some Diophantine Inequalities, *Mathematika*, 2, 145 (1955).

21a. Moser L., Sandwick C. M., Sr., Packaged Radios, *Am. Math. Monthly*, 59, 637 (Nov. 1952).

22. Nagell T., Number Theory, Chelsea Publ. Co., N.Y., 1964.
23. Nagell T., Sur quelques catégories d'equations diophantiennes résolubles par des identités, *Act. Arithmet.*, 9, 227 (1964).
24. Netto E., Lehrbuch der Combinatorik, Chelsea Publ. Co., N.Y., 1901.
25. Niven I., Diophantine Approximations, Interscience Publishers, N.Y., 1963.
26. Olds C. D., Continued Fractions, Random House, Inc., N.Y., 1963.
27. Ryser H. J., Combinatorial Mathematics, Wiley, N.Y., 1963; русский перевод: Райзер Г. Дж., Комбинаторная математика, изд-во «Мир», 1966.

28. Saaty T. L., On Nonlinear Optimization in Integers, *Naval Res. Log. Quart.*, 15, № 1, 1 (March 1968).
29. Silverman D. L., Goldberg M., A Diophantine Equation, *Am. Math. Monthly*, 74, 1013 (Oct. 1967).
30. Skolem Th., Diophantische Gleichungen, Chelsea Publ. Co., N.Y., 1950.
31. Stewart B. M., Theory of Numbers, 2nd ed., The Macmillan Co., N.Y., 1965.
32. Syski R., Algebraic Properties of Optimum Gradings, 3d Intern. Teletraffic Cong., Paris, 1961.
33. Takács L., Combinatorial Methods in the Theory of Stochastic Processes, Wiley, N.Y., 1967; русский перевод: Такач Л., Комбинаторные методы в теории случайных процессов, изд-во «Мир», 1971.
34. Thebault V., Positive Integer Solutions of $a^b = b^a$, *Mathesis*, 67 (1960).
35. Utz W. R., Diophantine Equations, *Pi Mu Epsilon J.* (Nov. 1954).
36. Veidinger L., On the Distribution of the Solutions of Diophantine Equations with Many Unknowns, *Act. Arithmet.* (1958).
37. Weinstock R., Greatest Common Divisor of Several Integers and An Associated Linear Diophantine Equations, *Am. Math. Monthly*, 67, 664 (1960).
38*. Матиясевич Ю. В., Диофантовы множества, *УМН*, **XXVII**, вып. 5 (1972).
39*. Хинчин А. Я., Цепные дроби, Физматгиз, 1961.
40*. Беккенбах Э., Беллман Р., Введение в неравенства, изд-во «Мир», 1965.
41*. Беккенбах Э., Беллман Р., Неравенства, изд-во «Мир», 1965.
42*. Бухштаб А. А., Теория чисел, 2-е изд, изд-во «Просвещение», 1966.
43*. Дэвенпорт Г., Высшая арифметика, введение в теорию чисел, изд-во «Наука», 1965.
44*. Лэнг С., Введение в теорию диофантовых приближений, изд-во «Мир», 1970.

Глава 5

Целочисленное программирование

5.1. Введение

В наиболее общей форме задача целочисленной оптимизации имеет следующий вид: найти вектор x с неотрицательными компонентами x_j, $j = 1, \ldots, n$, в E_n, который максимизирует целевую функцию $f(x_1, \ldots, x_n)$ при ограничениях $g_i(x_1, \ldots, x_n) \leqslant 0$, $i = 1, \ldots, m$. С геометрической точки зрения ищется точка с целочисленными координатами в области, которая удовлетворяет ограничениям (в так называемой *области допустимых значений*) и минимизирует f. В гл. 4 уже говорилось о том, что задачи оптимизации в целых положительных числах с ограничениями в виде равенств часто приводятся к задачам с ограничениями в виде неравенств и, следовательно, целочисленное программирование имеет важное значение при оптимизации диофантовых задач. В некоторых задачах требуется, чтобы только некоторые из компонент x были целыми. Другие компоненты должны быть рациональными. Этот случай носит название *частично целочисленного программирования*. Случай, когда все компоненты x должны быть целыми, иногда называется *полностью целочисленным программированием*.

В некоторых задачах целочисленного программирования требуется определять вектор x, компоненты которого принимают только двоичные значения 0 или 1; в этом случае говорят о *бивалентном программировании*.

Задачи, в которых переменные являются неотрицательными целыми числами, можно свести к случаю двоичных переменных, заменяя каждую переменную x_j выражением

$$x_j = x_{j1} + 2x_{j2} + 2^2 x_{j3} + \ldots + 2^{k-1} x_{jk},$$

где $x_{jp} = 0$ или 1 ($p = 1, \ldots, k$) и k достаточно велико, так что x_j может принимать наибольшее значение в допустимой области. Таким образом, 2^{k-1} становится верхней границей для x_j. При таком преобразовании резко увеличивается число переменных, так что решение задачи целочисленного программирования, сведенной к бивалентному программированию, становится очень громоздким или вообще невозможным. Однако многие задачи, естественно, формулируются как задачи бивалентного программирования; для решения задач такого типа разработаны алгоритмы.

Пример 5.1 [31]. Рассмотрим следующую задачу целочисленного программирования. Требуется минимизировать в неотрицательных

целых числах x_1, x_2 выражение

$$x_1 + 3x_2$$

при ограничениях

$$3x_1 - x_2 \geqslant 4,$$
$$x_1 + x_2 \leqslant 3.$$

Заключая из второго соотношения, что $x_1 \leqslant 3$, $x_2 \leqslant 3$, положим

$$x_1 = x_{11} + 2x_{12}, \quad x_2 = x_{21} + 2x_{22};$$

при этом задача сводится к отысканию переменных

$$x_{11}, \; x_{12}, \; x_{21}, \; x_{22},$$

которые принимают значения 0 или 1 и минимизируют выражение

$$x_{11} + 2x_{12} + 3x_{21} + 6x_{22}$$

при ограничениях

$$3x_{11} + 6x_{12} - x_{21} - 2x_{22} \geqslant 4,$$
$$x_{11} + 2x_{12} + x_{21} + 2x_{22} \leqslant 3.$$

О разрешимости в целых числах

Рассмотрим систему из m уравнений с n неизвестными

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

или в матричной форме $Ax = b$. Пусть $\bar{x} = (\bar{x}_1, \; \ldots, \; \bar{x}_n)$ служит решением. Каждому x_j соответствует вектор-столбец из коэффициентов при этом x_j. Рассмотрим все ненулевые $\bar{x}_j$ в $\bar{x}$. Если их соответствующие столбцы линейно независимы (т. е. из каждого соотношения вида $\sum_{i=1}^{h} \alpha_i v_i = 0$, где v_i — рассматриваемые вектор-столбцы, следует, что $\alpha_i = 0$, $i = 1, \; \ldots, \; k$), то $\bar{x}$ называется *крайним решением*. Линейная независимость гарантирует единственность решения для плоскостей, пересекающихся в точке (решения), компонентами которой являются эти ненулевые значения x_i. Если система задана в виде множества неравенств, то вершины соответствующего выпуклого многогранника могут служить крайними решениями системы, полученной заменой неравенств на равенства. Эти вершины называются *крайними точками* многогранника. Рассмотрим теперь, какие условия следует накладывать на матрицу A для того, чтобы все экстремальные точки имели целочисленные координаты.

Определение. Базисом матрицы A размерностью $m \times n$ с целочисленными элементами (назовем ее *целочисленной матрицей*),

m строк которой линейно независимы, является набор из m столбцов, ранг которого равен m.

Определение. Базис A называется унимодулярным, если его определитель равен $+1$ или -1.

Определение. Матрица A называется вполне унимодулярной, если каждая невырожденная подматрица A унимодулярна.

Запишем

$$X(A, b) \equiv \{x: Ax = b,\ x \geqslant 0\}$$

и

$$X^*(A, b) = \{x: Ax \leqslant b,\ x \geqslant 0\}.$$

Достаточным условием того, что крайние точки $X(A, b)$ будут целочисленными при любом целочисленном векторе b, является унимодулярность базиса. Пользуясь правилом Крамера, можно решить систему относительно всех x_i, и, так как определитель матрицы коэффициентов, стоящий в знаменателе, равен ± 1, а b имеет целочисленные компоненты, x_i должны быть целочисленными. Докажем теперь, что унимодулярность является и необходимым условием.

Теорема 5.1 (Вейнот и Данициг) [36]. *Если A — целочисленная матрица с линейно независимыми строками, то следующие утверждения эквивалентны:*

а) *Каждый базис является унимодулярным.*

б) *Крайние точки $X(A, b)$ являются целочисленными при любом целочисленном b.*

в) *Каждый базис имеет целочисленную обратную матрицу.*

Доказательство а $\Rightarrow$ б. Пусть B — базис, связанный с ненулевыми компонентами x_B крайней точки x из множества $X(A, b)$. Тогда по предположению $Bx_B = b$ и $\det B = \pm 1$. Следовательно, по правилу Крамера получаем, что x_B — целочисленный вектор.

б $\Rightarrow$ в. Пусть B — базис, 1_i обозначает вектор-столбец, в котором на i-й позиции стоит 1, а все остальные элементы — нули, и пусть y — любой целочисленный вектор, такой, что

$$z \equiv y + B^{-1}1_i \geqslant 0.$$

Теперь $Bz = By + 1_i \equiv b$ — целочисленный вектор и z содержит ненулевые компоненты крайней точки множества $X(A, b)$; следовательно, z — целочисленный вектор. Так как левая часть равенства $z - y = B^{-1}1_i$ имеет целочисленные компоненты, i-й столбец матрицы B^{-1} целочисленный. Те же рассуждения можно использовать при любом i, поэтому B^{-1} — целочисленная матрица.

в $\Rightarrow$ а. Пусть B — базис. Так как B и B^{-1} — целочисленные матрицы, их определители должны быть ненулевыми целыми числами, и из

$$(\det B)(\det B^{-1}) = 1$$

следует, что

$$\det B = \det B^{-1} = \pm 1.$$

Следствие (Хофман и Краскал). *Если A — целочисленная матрица, то следующие утверждения эквивалентны:*

а*) *A — вполне унимодулярная матрица.*

б*) *Крайние точки множества $X^*(A, b)$ являются целочисленными при любом целочисленном b.*

в*) *Каждая невырожденная подматрица A имеет целочисленную обратную матрицу.*

Доказательство. Дополним A единичной матрицей, так что матрица $A' = (A, I)$ имеет m строк. Эти строки линейно независимы. Если в теореме 5.1 вместо A использовать A', то утверждения этой теоремы в отношении A' эквивалентны утверждениям следствия относительно A. Таким образом, а $\Rightarrow$ а*, так как если C — невырожденная подматрица A, ранг которой равен $(m - k)$, то базис B в A' можно получить путем перестановки строк:

$$B = \begin{bmatrix} C & 0, \\ D & I_k \end{bmatrix},$$

где I_k — единичная матрица размерности $k \times k$. Таким образом, $\det B = \det C$ и, следовательно, $\det B = \pm 1$ тогда и только тогда, когда $\det C = \pm 1$.

Транспонированную матрицу A будем обозначать A^T.

Упражнение 5.1. Покажите, что если одна из матриц A, A^T, $-A$ вполне унимодулярна, то и остальные матрицы вполне унимодулярны

Следующая теорема дает некоторые достаточные условия унимодулярности матрицы A.

Теорема 5.2 (Хеллер — Томпкинс). *Каждый базис матрицы A является унимодулярным, если строки A можно разбить на два непересекающиеся подмножества R_1 и R_2, такие, что*

а) *Каждый столбец содержит не более двух ненулевых элементов.*

б) *Каждый элемент равен 0, $+1$ или -1.*

в) *Два ненулевых элемента в столбце, знаки которых совпадают, не входят в одно и то же множество R_i строк.*

г) *Два ненулевых элемента в столбце, знаки которых не совпадают, входят в одно и то же множество R_i строк.*

Доказательство. Теорема доказывается по индукции.

Упражнение 5.2. Заметим, что каждый столбец матрицы коэффициентов в транспортной задаче, приведенной в гл. 1, имеет два единичных элемента, а остальные равны нулю. Докажите, что эта матрица унимодулярна и, следовательно, задача имеет целочисленное решение.

Предыдущие три результата являются специальным случаем общей идеи, высказанной Хеллером [26]. Матрица, в которой каждый

базис, образованный столбцами, является унимодулярным, рассматриваемая как множество столбцов, называется *унимодулярным множеством*. Это приводит к общему определению унимодулярных множеств векторов в векторном пространстве (или элементов в свободной абелевой группе).

Определение. Множество A векторов в m-мерном пространстве является унимодулярным, если для любого базиса в A разложение каждого вектора из A по базисным векторам имеет коэффициенты -1, 0, 1. Очевидно, что если множество A унимодулярно, то унимодулярно и каждое подмножество A.

Теорема 5.2а (Хеллер).

а) *Множество ребер (т. е. одномерных граней, которые ориентированы в любом направлении и рассматриваются как векторы) симплекса является унимодулярным.*

б) *Унимодулярное множество размерности m содержит не более $m(m+1)$ элементов (не считая нулевого элемента), т. е. если оно содержит $m(m+1)$ элементов, то оно является максимальным.*

в) *Если унимодулярное множество A размерности m содержит $m(m+1)$ векторов (не считая нулевого вектора), то A является множеством ребер m-мерного симплекса.*

г) *При $m \geqslant 4$ существуют максимальные унимодулярные множества с числом элементов менее чем $m(m+1)$.*

Упор здесь сделан на максимальные множества, поскольку каждое подмножество унимодулярного множества тоже унимодулярно.

Выбор базиса среди ребер m-мерного симплекса и представление части или всех оставшихся ребер в виде линейной комбинации элементов этого базиса дают множество столбцов A, в котором каждый базис унимодулярен. Например, если в качестве базиса выбрать звезду (т. е. все ребра исходят из данной вершины), то получается матрица транспортной задачи; если выбрать гамильтонов цикл ребер, то получается матрица, в которой каждый столбец состоит из нескольких последовательных единиц, за которыми или перед которыми стоят последовательно несколько нулей.

5.2. Задача о ранце [16]

Одной из простейших задач целочисленной оптимизации, которую можно представить алгебраически, является линейная задача, в которой надо максимизировать или минимизировать в неотрицательных целых числах линейное выражение при ограничениях в виде линейных неравенств. Задача формулируется следующим образом. Найти

$$\max \sum_{j=1}^{n} c_j x_j$$

при ограничении

$$\sum_{j=1}^{n} a_j x_j \leqslant b,$$

где все $x_j \geqslant 0$ — целые числа. В векторной записи задача состоит в том, чтобы максимизировать cx при ограничении $Ax \leqslant b$, где $c = (c_1, \ldots, c_n)$, $x^T = (x_1, \ldots, x_n)^T$, $A = (a_1, \ldots, a_n)$ и компоненты вектора x — неотрицательные целые числа. Название задачи дано по одной частной формулировке, состоящей в том, что в ранец емкостью b надо упаковать n видов предметов с весами $c_1, \ldots, c_n$ и размерами $a_1, \ldots, a_n$ соответственно так, чтобы загрузка ранца была максимальной.

Естественно, что предметы с наибольшим весом на единицу объема следует упаковывать в первую очередь, оставшийся объем надо заполнять предметами со следующим по величине весом на единицу объема и т. д. Поэтому без потери общности можно принять, что предметы упорядочены следующим образом:

$$\frac{c_1}{a_1} \geqslant \frac{c_2}{a_2} \geqslant \ldots \geqslant \frac{c_n}{a_n}.$$

Выше уже упоминалось следующее определение.

Определение. Лексикографическим упорядочением во множестве векторов одинаковой размерности называется упорядочение векторов по величине первых компонент, а если они равны, то по величине вторых компонент и т. д.; вектор с большей компонентой считается бóльшим.

Так, (3, 7, 1) при лексикографическом упорядочении больше, чем (2, 10, 50), и меньше, чем (3, 8, 0).

Используем этот подход при решении задачи об упаковке ранца. Сначала построим лексикографически наибольший вектор $x = (x_1, \ldots, x_n)$, затем следующий по величине вектор и будем продолжать, пока не достигнем $x = 0$. Тот факт, что лексикографически наибольший вектор не обязательно обеспечивает максимум, очевиден из примера задачи максимизации выражения

$$8x_1 + 5x_2 + x_3$$

при ограничении

$$3x_1 + 2x_2 + x_3 \leqslant 13.$$

где $x_i \geqslant 0$, $i = 1, 2, 3$,— целые числа. Здесь $^8/_3 \geqslant {}^5/_2 \geqslant 1$ и лексикографически наибольший вектор получается путем вычисления целой части 13/3, т. е. наибольшего целого, не превышающего 13/3; это число равно 4. Остается $13 - 4 \cdot 3 = 1$. Задача теперь состоит в том, чтобы распределить 1 среди оставшихся переменных, т. е. решаем уравнение $2x_2 + x_3 = 1$, что можно сделать, положив $x_2 = 0$, $x_3 = 1$. Таким образом, (4, 0, 1) — искомый вектор. Но если взять $x_1 = 3$ и решать уравнение $2x_2 + x_3 = 4$, то наибольший вектор

будет иметь вид (3, 2, 0). Вычисляя целевую функцию на этих векторах, получаем, что вектор (4, 0, 1) дает для этой функции значение 33, а (3, 2, 0) — значение 34. Второе решение является максимальным, что легко проверить перебором.

Наибольшее в смысле лексикографического упорядочения решение $x^0 = (x_1^0, \ldots, x_n^0)$ получается следующим образом:

$$x_1^0 = \left[\frac{b}{a_1}\right], \quad x_2^0 = \left[\frac{b - a_1 x_1^0}{a_2}\right], \quad \ldots, \quad x_n^0 = \left[\frac{b - \sum_{j=1}^{n-1} a_j x_j^0}{a_n}\right].$$

Если найден лексикографически упорядоченный вектор x^i, который является решением, наибольший вектор x^{i+1}, меньший x^i при таком упорядочении, который также является решением, получается уменьшением последней положительной компоненты x^i на единицу и затем увеличением следующей компоненты на максимально возможную величину и т. д. Если при этом ограничение не выполняется, то последняя компонента x^i уменьшается на два и процесс продолжается. Чтобы найти максимум, надо перебрать все векторы от наибольшего до наименьшего, т. е. до $x = 0$. Однако нет необходимости выписывать все векторы в лексикографическом порядке. Здесь будет дан способ отбора как можно меньшего числа векторов. Заметим, что если $x_s^i > 0$ — последняя отличная от нуля компонента x_i, а

$$x_{s+1}^i = \ldots = x_n^i = 0,$$

и если определить вектор y^i так, что он совпадает с x^i, за исключением того, что

$$y_s^i = x_s^i - 1 \geqslant 0,$$

то максимум $\sum_{j=1}^{n} c_j x_j$ будет иметь вид

$$\delta^i = \sum_{j=1}^{n} c_j y_j^i + \frac{c_{s+1}}{a_{s+1}} \left(b - \sum_{j=1}^{n} a_j y_j^i\right).$$

При этом игнорируется условие целочисленности $x_{s+1}, \ldots, x_n$ и учитывается, что $c_{s+1}/a_{s+1} \geqslant \ldots \geqslant c_n/a_n$ и $b - \sum_{j=1}^{n} a_j y_j^i$ — остающаяся емкость. При наличии требования целочисленности максимум не может превышать δ^i. Теперь если x^* доставляет максимальное значение среди всех векторов, которые стоят по порядку раньше x^i, и если $\delta^i \leqslant cx^*$, то нет необходимости рассматривать все векторы x, первые s компонент которых совпадают с соответствующими компонентами y^i. Таким образом, если $\delta^i \leqslant cx^*$, принимается $x^{i+1} = y^i$ и процесс продолжается. Если $\delta^i > cx^*$, то в качестве x^{i+1} выбирается следующий в лексикографическом порядке вектор [2, 4а, 16а].

Отправляясь от лексикографически наибольшего векторного решения $x^1 = (4, 0, 1)$ в рассмотренном выше примере, получаем $x_3^1 = 1 > 0$, $y_3^1 = x_3^1 - 1 \geqslant 0$. Таким образом, $y_1^1 = 4$, $y_2^1 = 0$, $y_3^1 = 0$. Кроме того, $c_1 y_1^1 = 32$, $c_2/a_2\,(13 - 4 \times 3) = {}^5/_2$, $\delta^1 = 32 + {}^5/_2 = 34{,}5$, а $34{,}5 \geqslant 33$, где 33 означает максимальное значение, соответствующее $x^* = (4, 0, 1)$. Таким образом, выбираем следующий в лексикографическом порядке вектор, удовлетворяющий ограничению. Это вектор (3, 2, 0), который дает значение 34. Снова получаем $x_1^2 = 3$, $x_2^2 = 2$, $x_3^2 = 0$; $y_1^2 = 3$, $y_2^2 = 1$, $y_3^2 = 0$; $c_1 y_1^2 = 24$, $c_2 y_2^2 = 5$. Отсюда

$$\delta^2 = 29 + \frac{5}{2}[13 - (3 \times 3 + 1 \times 2)] = 34,$$

т. е. это то же значение, которое дает вектор x^2. Поэтому выбираем $x^3 = y^2 = (3, 1, 0)$, откуда получаем $y^3 = (3, 0, 0)$, $\delta^3 = 24 + 1 \times (13 - 3 \times 3) = 28 \leqslant 34$. Отсюда следует, что в качестве x^4 нужно брать вектор (2, 3, 1), который дает величину 31. Получаем $y^4 = (2, 3, 0)$ и $\delta^4 = 31 + {}^5/_2\,(13{-}12) = 33{,}5 < 34$, откуда вытекает, что вектор $x^5 = (2, 3, 0)$ является следующим в лексикографическом порядке вектором. Он дает величину 31. Теперь $y^5 = (2, 2, 0)$, $\delta^5 = 29 < 34$ и $x^6 = (2, 2, 0)$. Снова $x^7 = (2, 1, 5)$ дает значение 26, $y^7 = (2, 1, 4)$, $\delta^7 = 25 < 34$, $x^8 = (2, 1, 4)$, $x^9 = (2, 1, 3)$ и т. д.

Упражнение 5.3. Максимизируйте в неотрицательных целых числах выражение

$$2x_1 + 7x_2 + 3x_3 + x_4$$

при ограничении

$$6x_1 + 3x_2 + 2x_3 + x_4 \leqslant 20.$$

Ответ: (0, 6, 1, 0).

Если целью задачи является минимизация, то сначала распределяется переменная с наименьшим относительным весом и т. д. (Заметим, что задача минимизации может быть решена путем умножения целевой функции на -1 и последующей максимизацией.)

Упражнение 5.4. Покажите, что решением задачи минимизации выражения

$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6$$

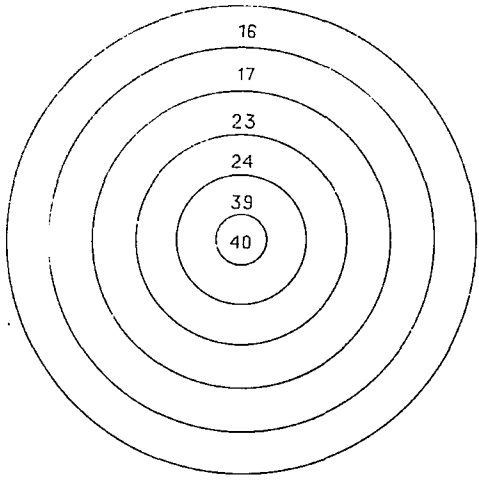
при ограничении в виде равенства

$$16x_1 + 17x_2 + 23x_3 + 24x_4 + 39x_5 + 40x_6 = 100,$$

где $x_i \geqslant 0$, $i = 1, \ldots, 6$,— целые числа, является вектор (2, 4, 0, 0, 0, 0).

Это алгебраическая формулировка задачи, приведенная в журнале «New Yorker» за апрель 1967 г. В задаче требуется определить

число стрел, которое надо использовать для того, чтобы набрать 100 очков на мишени, изображенной на фиг. 5.1. Заметим, что,



Фиг. 5.1. Сколько необходимо стрел, чтобы набрать ровно 100 очков?

поскольку существует единственное решение, здесь не требуется определять наименьшее число стрел.

Упражнение 5.5. Максимизируйте выражение $3x_1 - x_2$ при ограничениях

$$\begin{aligned} 3x_1 - 2x_2 &\leqslant 3, \\ -5x_1 - 4x_2 &\leqslant -10, \\ 2x_1 + x_2 &\leqslant 5, \end{aligned}$$

где x_i, $i = 1, 2,$ — целые числа.

Оптимальное решение (1, 2). *Указание.* Приведите ограничения к виду

$$\begin{aligned} 3x_1 - 2x_2 + x_3 &= 3, \\ -5x_1 - 4x_2 + x_4 - x_6 &= -10, \\ 2x_1 + x_2 + x_5 &= 5. \end{aligned}$$

Затем исключите все переменные, кроме одной. Оптимальное решение имеет вид (1, 2, 4, 3, 1, 0).

Упражнение 5.6. Максимизируйте выражение $3x_1 + x_2$ при ограничениях

$$\begin{aligned} 2x_1 + 3x_2 &\leqslant 6, \\ 2x_1 - 3x_2 &\leqslant 3, \end{aligned}$$

где x_i, $i = 1, 2,$ — целые числа.

5.3. Общее линейное программирование

Симплекс-метод — наиболее широко используемый метод решения задач линейного программирования при отсутствии требований целочисленности. Однако модификации этого метода используются и в целочисленном программировании. Важно понять ход этого процесса, чтобы быть подготовленным к использованию его в последующих разделах для решения целочисленных задач.

В задаче линейного программирования требуется максимизировать функцию f, называемую *целевой функцией*, при наличии линейных ограничений g_i, $i = 1, \ldots, m$, о которых говорилось в разд. 5.1, т. е.

$$f(x) = c_1x_1 + \ldots + c_nx_n$$

и

$$g_i(x) \leqslant 0 \quad \text{или} \quad a_{i1}x_1 + \ldots + a_{in}x_n - b_i \leqslant 0,$$
$$i = 1, \ldots, m, \quad x_j \geqslant 0, \quad j = 1, \ldots, n.$$

Здесь эта хорошо известная задача не рассматривается, будет дан только краткий обзор симплекс-метода, который наиболее часто используется для решения таких задач. Этот метод также иногда используется в качестве вспомогательного средства при решении задач нелинейного программирования.

В задачах линейного программирования *область допустимых значений*, которая определяется ограничивающими неравенствами, представляет собой многогранник (так как все g_i линейны и, следовательно, являются гиперплоскостями в n-мерном пространстве); целевая функция определяет гиперплоскость в $(n + 1)$-мерном пространстве. Для каждого постоянного значения f получаем линию уровня в n-мерном пространстве. Вейль показал, что решение находится на границе и обычно является вершиной многогранника. Это интуитивно ясно в случае трехмерного пространства. Если производится максимизация, то линия уровня, проходящая через вершину, в которой достигается максимум (линии уровня могут покрывать n-мерное пространство), находится на наибольшем расстоянии от начала координат. Очевидно, что при отыскании решения можно проверить все вершины, так как их число конечно. Однако на практике используются методы, которые быстрее дают решение.

На фиг. 5.2 дано геометрическое представление следующей задачи линейного программирования: требуется минимизировать выражение

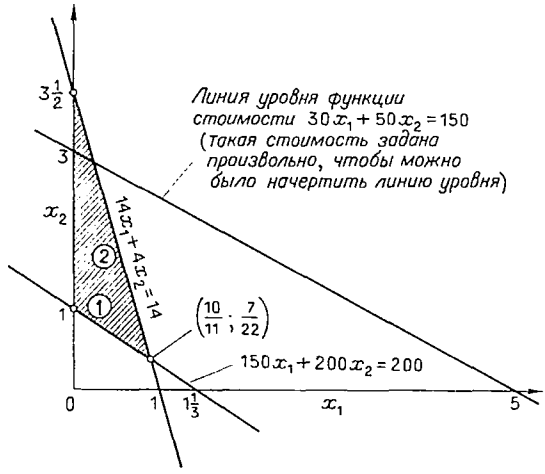
$$30x_1 + 50x_2$$

при наличии ограничений

$$150x_1 + 200x_2 \geqslant 200,$$
$$14x_1 + 4x_2 \leqslant 14,$$
$$x_1, x_2 \geqslant 0.$$

Упражнение 5.7. Найдите, в какой из вершин многоугольника, определенного ограничениями, достигается минимум.

В матричной форме задача линейного программирования ставится следующим образом: найти вектор-столбец $x \geqslant 0$, т. е. с компонентами $x_j \geqslant 0$, $j = 1, \ldots, n$, который удовлетворяет ограничениям $Ax \leqslant b$, а также максимизирует линейную функцию cx, где $A = (a_{ij})$, $i = 1, \ldots, m$, $j = 1, \ldots, n$, $c = (c_1, \ldots, c_n)$, а b — вектор-столбец с компонентами $b_1, \ldots, b_m$. С исходной задачей, которая



Фиг. 5.2.

называется *прямой*, связана обратная, или *двойственная*, задача $yA \geqslant c$, $y \geqslant 0$; yb достигает минимума, где $y = (y_1, \ldots, y_m)$.

В хорошо известной в линейном программировании теореме двойственности утверждается, что если прямая и обратная задачи имеют решение, что значения целевых функций в обеих задачах в оптимальной точке совпадают. Решение одной задачи может быть получено из решения другой. Иногда для облегчения вычислений удобнее решать двойственную задачу.

При использовании симплекс-метода Данцига для решения задач линейного программирования сначала выбираются базисные m-мерные векторы, где m — число неравенств. Базисы выбираются последовательно, пока, наконец, не получится базис, который является решением задачи. С каждой итерацией, т. е. переходом к новому базису, значение целевой функции приближается к оптимальному значению, возможному при итерациях в области допустимых значений, или, в худшем случае, не изменяется. Так как число возможных базисов конечно, очевидно, что, за исключением патологических случаев (например, зацикливание процесса, когда базисы начинают

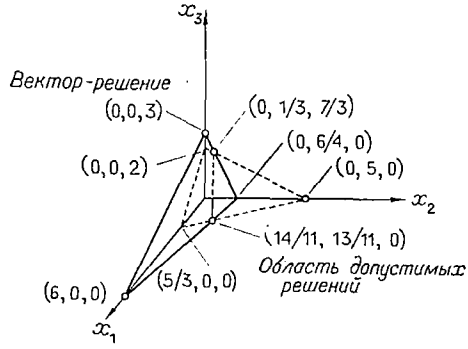
повторяться), оптимум достигается за конечное число шагов. Существуют методы, предназначенные для того, чтобы избегать таких затруднений. Здесь будет изложена часть теории и даны примеры использования симплекс-метода. Нижеследующая теорема существования хорошо известна.

Теорема 5.3. *Если существует одно допустимое решение, то существует допустимое решение (называемое базисным допустимым решением) с не более чем m точками P_i, имеющими положительные веса x_i, и не менее чем с n — m точками P_i с весами $x_i = 0$.*

Замечание. Базисное решение называется *вырожденным*, если хотя бы один вес x_i, соответствующий базисному вектору, равен нулю.

Теорема 5.4. *Если значения целевой функции на классе допустимых решений имеют конечную верхнюю границу, то существует максимальное допустимое решение, которое является базисным допустимым решением.*

Симплекс-метод будет использован при решении задачи, приведенной для иллюстрации. Чтобы разъяснить идеи, здесь приводятся



Фиг. 5.3.

подробные вычисления. Табличный метод будет проиллюстрирован в одном из последующих разделов на примере решения задачи квадратичного программирования. Множество ограничений в задаче задано неравенствами

$$x_1 + 4x_2 + 2x_3 \geqslant 6,$$
$$3x_1 + x_2 + 2x_3 \geqslant 5,$$
$$x_i \geqslant 0, \quad i = 1, 2, 3,$$

а линейная форма (функция стоимости), которую надо минимизировать, имеет вид

$$2x_1 + 9x_2 + x_3.$$

Геометрическая интерпретация этой задачи дана на фиг. 5.3.

1. Заменяем ограничивающие неравенства на равенства путем введения вспомогательных неотрицательных переменных

$$x_1 + 4x_2 + 2x_3 - x_4 = 6,$$
$$3x_1 + x_2 + 2x_3 - x_5 = 5,$$
$$-2x_1 - 9x_2 - x_3 = \max,$$

где

$x_4 \geqslant 0$ и $x_5 \geqslant 0$ — вспомогательные переменные.

2. Записываем матрицу

$$[P_1, P_2, P_3, P_4, P_5; P_0] = \left[\begin{array}{ccccc|c} 1 & 4 & 2 & -1 & 0 & 6 \\ 3 & 1 & 2 & 0 & -1 & 5 \end{array}\right]$$

и

$$c_1 = -2, \; c_2 = -9, \; c_3 = -1, \; c_4 = 0, \; c_5 = 0.$$

3. Начинаем выбор базиса, который представляет собой множество линейно независимых векторов (т. е. определитель матрицы, составленной из этих векторов, не равен нулю); любой другой вектор можно представить в виде линейной комбинации базисных векторов. Линейная комбинация векторов P_1 и P_2, например, записывается в виде $aP_1 + bP_2$, где a и b — действительные числа. Как видно из матрицы, в данном случае для образования базиса нужны два вектора.

Замечание. Использование в качестве базиса множества искусственных векторов позволит избежать на начальном этапе выбора недопустимых векторов, т. е. множества векторов, для которых соответствующие переменные принимают отрицательные значения. Простым примером базиса из искусственных векторов является (1, 0) и (0, 1).

Предположим, что в качестве начального базиса выбраны P_1 и P_2. Выразим P_0 в виде их линейной комбинации. (Ниже показано, как это делать.) В данном случае получаем $P_0 = {}^{14}/_{11}P_1 + {}^{13}/_{11}P_2$. В этой задаче функция стоимости будет отрицательной при положительных x_i. Следовательно, если для вектора P_0, представленного в виде комбинации выбранных базисных векторов, получится положительное значение функции стоимости, то по крайней мере один базисный вектор следует изменить. Итак, выражая оставшиеся векторы в виде линейной комбинации P_1 и P_2, получим

	P_1	P_2
P_1	1	0
P_2	0	1
P_3	$\frac{6}{11}$	$\frac{4}{11}$
P_4	$\frac{1}{11}$	$-\frac{3}{11}$
P_5	$-\frac{4}{11}$	$\frac{1}{11}$

Пусть β_{ij} — коэффициент в вышеуказанной таблице, первый индекс которого совпадает с индексом вектора, определяющего верхний вход в таблицу, а второй индекс совпадает с индексом вектора, определяющего вход в таблицу слева. Например, чтобы получить P_4 в виде линейной комбинации P_1 и P_2, записываем $P_4 = aP_1 + bP_2$. Заметим, что $\beta_{14} = a$, $\beta_{24} = b$. Чтобы показать, как получаются a и b, запишем соотношение между векторами

$$\begin{matrix} P_4 & P_1 & P_2 \end{matrix}$$
$$\begin{bmatrix} -1 \\ 0 \end{bmatrix} = a\begin{bmatrix} 1 \\ 3 \end{bmatrix} + b\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} a \\ 3a \end{bmatrix} + \begin{bmatrix} 4b \\ b \end{bmatrix} = \begin{bmatrix} a+4b \\ 3a+b \end{bmatrix}.$$

Приравнивая координаты в крайнем правом и крайнем левом векторах, получаем $-1 = a + 4b$, $0 = 3a + b$. Решая эти два уравнения совместно относительно a и b, находим $a = {}^1/_{11}$, $b = -{}^3/_{11}$. Аналогично все другие векторы представляются в виде линейных комбинаций базисных векторов P_1 и P_2.

4. Рассмотрим $z_j = \beta_{1j}c_1 + \beta_{2j}c_2$ $(j = 1, \ldots, 5)$.

Замечание. В общем случае, если базисные векторы имеют индексы p, q, r и т. п., записываем

$$z_j = \beta_{pj}c_p + \beta_{qj}c_q + \beta_{rj}c_r + \ldots,$$

$$z_1 = c_1 \qquad = -2 \quad \text{сравниваем с} \quad c_1 = -2,$$

$$z_2 = c_2 \qquad = -9 \quad \text{сравниваем с} \quad c_2 = -9,$$

$$z_3 = \frac{6}{11}c_1 + \frac{4}{11}c_2 \quad = -\frac{48}{11} \quad \text{сравниваем с} \quad c_3 = -1,$$

$$z_4 = \frac{1}{11}c_1 - \frac{3}{11}c_2 \quad = \frac{25}{11} \quad \text{сравниваем с} \quad c_4 = 0,$$

$$z_5 = -\frac{4}{11}c_1 + \frac{1}{11}c_2 = -\frac{1}{11} \quad \text{сравниваем с} \quad c_5 = 0.$$

Сравниваем z_j с c_j, как указано.

Если $z_j \geqslant c_j$ при всех j, то процесс оканчивается, т. е. $P_0 = {}^{14}/_{11}P_1 + {}^{13}/_{11}P_2$, а стоимость (при максимизации отрицательная) будет задаваться следующим образом: если $P_0 = aP_1 + bP_2$, то стоимость равна $ac_1 + bc_2$. В этом случае получаем

$$\frac{14}{11}(-2) + \frac{13}{11}(-9) = -\frac{145}{11} \text{ единиц}$$

т. е. 11 единиц в качестве минимума целевой функции. Если производится непосредственная минимизация целевой функции, то критерием должно служить соотношение $z_j \leqslant c_j$.

Замечание. Очевидно, что $z_j < c_j$ при некотором j. Рассмотрим максимум $(c_j - z_j)$. В данном случае $c_3 - z_3 = \max$. Поэтому при

выборе следующего базиса с P_3 нужно использовать P_1 или P_2. Вопрос об использовании P_1 или P_2 решается следующим методом (если $c_j \geqslant z_j$ при всех j и $\beta_{ij} \leqslant 0$ при всех i, то максимальное допустимое решение бесконечно).

5. Представляем P_0 в виде линейной комбинации P_1 и P_2:

$$P_0 = \frac{14}{11} P_1 + \frac{13}{11} P_2.$$

Представляем P_3 в виде линейной комбинации P_1 и P_2 и умножаем на θ:

$$\theta P_3 = \theta \frac{6}{11} P_1 + \theta \frac{4}{11} P_2.$$

Вычисляем разность первого и второго уравнений

$$P_0 = \theta P_3 + \left(\frac{14}{11} - \frac{6}{11}\theta\right) P_1 + \left(\frac{13}{11} - \frac{4}{11}\theta\right) P_2.$$

Выбираем

$$\theta = \min\left(\frac{14/11}{6/11}, \frac{13/11}{4/11}\right) = \frac{14}{6}.$$

Следовательно, получаем

$$P_0 = \frac{14}{6} P_3 + \frac{2}{6} P_2.$$

Новый базис будет состоять из P_2 и P_3. Функция стоимости в этом случае принимает значение

$$\frac{14}{6} c_3 + \frac{2}{6} c_2 = \frac{14}{6}(-1) + \frac{2}{6}(-9) = -\frac{32}{6},$$

что, очевидно, является улучшением по сравнению с предыдущим значением, так как решается задача максимизации.

Замечание. В общем случае, если $P_0 = \alpha_1 P_1 + \ldots + \alpha_q P_q$, $P_1, \ldots, P_q$ — базисные векторы и $c_j - z_j = \max$, в число базисных векторов надо включить P_j, равный

$$P_j = \beta_{1j} P_1 + \ldots + \beta_{qj} P_q,$$

а затем выбрать

$$\theta = \min_i \left(\frac{\alpha_i}{\beta_{ij}}\right), \quad \beta_{ij} > 0.$$

Таким образом, один из векторов в соотношении

$$P_0 = \theta P_j + (\alpha_1 - \theta\beta_{1j}) P_1 + \ldots + (\alpha_q - \theta\beta_{qj}) P_q$$

исключается.

6. Представляем остальные векторы в виде линейной комбинации P_2 и P_3:

	P_2	P_3
P_1	$-\frac{2}{3}$	$\frac{11}{6}$
P_2	1	0
P_3	0	1
P_4	$-\frac{1}{3}$	$\frac{1}{6}$
P_5	$\frac{1}{3}$	$-\frac{2}{3}$

Рассмотрим теперь $z_j = \beta_{2j}c_2 + \beta_{3j}c_3$.

$$z_1 = \left(-\frac{2}{3}\right)(-9) + \frac{11}{6}(-1) = \frac{25}{6} \quad \text{сравниваем с} \quad c_1 = -2,$$

$$z_2 = c_2 \qquad = -9 \quad \text{сравниваем с} \quad c_2 = -9,$$

$$z_3 = c_3 \qquad = -1 \quad \text{сравниваем с} \quad c_3 = -1,$$

$$z_4 = \left(-\frac{1}{3}\right)(-9) + \frac{1}{6}(-1) = \frac{17}{6} \quad \text{сравниваем с} \quad c_4 = 0,$$

$$z_5 = \left(\frac{1}{3}\right)(-9) - \frac{2}{3}(-1) \quad = -\frac{7}{3} \quad \text{сравниваем с} \quad c_5 = 0.$$

Здесь получаем $c_5 - z_5 = \max$. Следовательно, P_2 или P_3 надо заменить на P_5 при следующем выборе базиса. Чтобы произвести выбор, запишем

$$P_0 = \frac{14}{6}P_3 + \frac{2}{6}P_2,$$

$$\theta P_5 = -\theta\frac{2}{3}P_3 + \theta\frac{1}{3}P_2.$$

Вычитая второе равенство из первого, получаем

$$P_0 = \theta P_5 + \left(\frac{14}{6} + \frac{2}{3}\theta\right)P_3 + \left(\frac{2}{6} - \frac{1}{3}\theta\right)P_2.$$

Таким образом, $\theta = 1$, так как рассматриваются только значения β_{ij}, большие нуля. Это дает

$$P_0 = P_5 + \frac{18}{6}P_3,$$

и функция стоимости равна

$$1 \times 0 + \frac{18}{6} \times (-1) = -3,$$

что больше предыдущих значений.

7. Снова представляем оставшиеся векторы в виде линейной комбинации P_3 и P_5:

	P_3	P_5
P_1	$\frac{1}{2}$	-2
P_2	2	3
P_3	1	0
P_4	$-\frac{1}{2}$	-1
P_5	0	1

Рассмотрим еще раз $z_j = \beta_{3j}c_3 + \beta_{5j}c_5$.

$$z_1 = -\frac{1}{2} \quad \text{сравниваем с} \quad c_1 = -2,$$

$$z_2 = -2 \quad \text{сравниваем с} \quad c_2 = -9,$$

$$z_3 = -1 \quad \text{сравниваем с} \quad c_3 = -1,$$

$$z_4 = \frac{1}{2} \quad \text{сравниваем с} \quad c_4 = 0,$$

$$z_5 = 0 \quad \text{сравниваем с} \quad c_5 = 0.$$

Очевидно, что все $z_j \geqslant c_j$, и это решение является окончательным. Другими словами, поскольку $P_0 = {}^{18}/_{6}P_3 + P_5$, получаем $(x_1, x_2, x_3, x_4, x_5) = (0, 0, {}^{18}/_{6}, 0, 1)$. Полная стоимость равна 3 (если требуется получить минимум, знак изменяется на обратный). Вектор P_5, имеющий нулевую стоимость, ничего не вносит. Нет предмета, который соответствовал бы ему. На фиг. 5.3 вершина (0, 0, 3) представляет собой решение.

Следует отметить, что симплекс-метод не гарантирует целочисленности компонент вектора решения. Нет никаких оснований заранее считать, что любая вершина многогранника, представляющая собой допустимое решение, должна иметь целочисленные координаты. Аппроксимация значений координат целыми числами не обязательно дает допустимый вектор, так как вершина, в которой достигается решение, может быть ближайшей к точке с целочисленными координатами, лежащей вне области допустимых решений, а ближайшая в области допустимых значений точка с целочисленными координатами может лежать так далеко, что аппроксимация окажется неоправданной.

Решение двойственной задачи

Если базис, дающий максимальное решение прямой задачи, получен при помощи симплекс-метода, то решение двойственной задачи находится следующим образом. Пусть B — матрица, столбца-

ми которой являются базисные векторы, дающие решение, с обратной матрицей B^{-1}, а C^* — вектор стоимости, соответствующий базисному решению. Тогда решением двойственной задачи является вектор $y = C^*B^{-1}$, компоненты которого составляют решение $(y_1^0, \ldots, y_m^0)$.

Доказательство. Заметим, что $z_j = C^*(B^{-1}A)$. Так как $z_j - c_j \geqslant 0$, имеем $C^*(B^{-1}A) - c \geqslant 0$. Последнее соотношение после подстановки $y = C^*B^{-1}$ переходит в $yA - c \geqslant 0$ или $yA \geqslant c$, а это и есть двойственная задача.

Если поставлена задача линейного программирования с тремя ограничениями и n переменными, то можно получить симплексное решение путем перехода к двойственной задаче, которая решается геометрически. Чтобы показать это, предположим для простоты, что (y_1, y_2, y_3) — величины, составляющие решение. Тогда P_1, P_2 и P_3 — базисные векторы в симплексном решении двойственной задачи. Если теперь взять подматрицу, составленную из этих векторов, из исходной матрицы задачи найти обратную к ней матрицу, умножить на вектор, компонентами которого являются стоимостные коэффициенты, соответствующие P_1, P_2, P_3, то получится решение (x_1, x_2, x_3) прямой задачи, т. е.

$$(x_1, x_2, x_3) = \text{Вектор стоимости} \times B^{-1}.$$

Если значения x_1, x_2 и x_3 известны, соответствующие векторы матрицы образуют базис, приводящий к решению прямой задачи. Аналогичные рассуждения применимы и в более простом случае двух ограничений в прямой задаче.

Упражнение 5.8. Используя описанный выше симплекс-метод, покажите, что выражение

$$2x_1 + 3x_2 + x_3$$

достигает минимума в точке $(0, {}^{7}/_{20}, {}^{51}/_{40})$ при наличии ограничений

$$4x_1 + x_2 + 6x_3 \geqslant 8,$$
$$3x_1 + 7x_2 + 2x_3 \geqslant 5,$$
$$x_1, x_2, x_3 \geqslant 0.$$

Упражнение 5.9. Получите решение упражнения 5.8 геометрически, т. е. начертите рисунок, дающий решение задачи.

Упражнение 5.10. Сформулируйте задачу, двойственную к предыдущей; получите решение геометрическим и алгебраическим путем, используя решение прямой задачи.

Обоснование симплексного решения

1. Для решения задачи максимизации $f = cx$ при ограничении $Ax \leqslant b$, $x \geqslant 0$, введем вспомогательные переменные $x_{n+1}, \ldots, x_{n+m}$ с соответствующими стоимостными коэффициентами $c_{n+1} = \ldots$

$\ldots = c_{n+m} = 0$. Тогда задача принимает вид: максимизировать $f = cx$ при ограничении $[A; I]\, x = b \geqslant 0$, где теперь $c = \{c_1, \ldots, c_n, c_{n+1}, \ldots, c_{n+m}\}$, $x^T = \{x_1, \ldots, x_{n+m}\} \geqslant 0$ и I — единичная матрица размерностью $m \times m$.

2. Используем тот факт, что решение задачи линейного программирования достигается в одной из вершин области, определенной ограничениями, и что в любой вершине n из $(n + m)$ компонент равны нулю. Остальные m компонент вектора x, которые могут принимать положительные значения, образуют «базис», и симплексный метод по существу заключается в переходе от одной вершины к сопряженной. При этом множество ненулевых компонент меняется за счет того, что в «старом» базисе одна из них полагается равной нулю, а вместо нее выбирается ненулевой другая компонента, прежде равная нулю. Переход от данного базиса к новому осуществляется таким образом, чтобы выполнялись ограничения и увеличивалась целевая функция f. Возможность перехода от некоторой вершины к сопряженной так, чтобы при этом выполнялись ограничения, гарантируется ввиду выпуклости «области допустимых значений», определяемой ограничениями. Если базисное решение вырожденное, то можно изменить базис, не увеличивая f.

Ниже дается краткое изложение симплекс-метода. Обозначим пробное решение, выбранное в некоторой точке в ходе итеративной процедуры, через x, а x' определим так, что

$$x'^T = \{\bar{y}^T; \bar{\bar{y}}^T\},$$

т. е. x' представляет собой перегруппировку координат x, такую, что $\bar{y}$ — базис, выбранный для пробного решения, а $\bar{\bar{y}}$ содержит компоненты x, не входящие в базис ($\bar{y}_j = 0$, $j = 1, \ldots, n$).

Аналогично определив соответствующие перегруппировки элементов матриц c и $[A; I]$, можно записать

$$f = \bar{c}\bar{y} + \bar{\bar{c}}\bar{\bar{y}}, \tag{5.1}$$

$$b = \bar{A}\bar{y} + \bar{\bar{A}}\bar{\bar{y}}. \tag{5.2}$$

Теперь желательно изменить одну из компонент $\bar{\bar{y}}$ так, чтобы произошли следующие события:

1. f возрастает.

2. Ограничение (5.2) по-прежнему выполняется.

3. Одна из компонент $\bar{y}$ обращается в нуль, но ограничение $x \geqslant 0$ выполняется. Ограничение (5.2) влечет за собой требование, состоящее в том, что при любом изменении $\Delta\bar{\bar{y}}$ вектора $\bar{\bar{y}}$

$$\Delta b = \bar{A}\Delta\bar{y} + \bar{\bar{A}}\Delta\bar{\bar{y}} = 0, \tag{5.3}$$

или

$$\Delta\bar{y} = -(\bar{A})^{-1}\bar{\bar{A}}\Delta\bar{\bar{y}} = -D\Delta\bar{\bar{y}}, \tag{5.4}$$

где

$$D = (\bar{A})^{-1}\,\bar{\bar{A}} \tag{5.5}$$

и существование $(\bar{A})^{-1}$ гарантируется существованием решения уравнения (5.2), где $\bar{y}$ — вектор решения.

Поскольку только одна компонента $\bar{\bar{y}}$, скажем $\bar{\bar{y}}_\alpha$, должна быть сделана положительной, формулу (5.4) можно переписать в виде

$$\Delta\bar{y}_i = -D_{i\alpha}\Delta\bar{\bar{y}}_\alpha \tag{5.6}$$

Как видно из формул (5.1) и (5.6), соответствующее изменение целевой функции f должно быть равно

$$\Delta f_0 = (\bar{\bar{c}}_\alpha - z_\alpha)\,\Delta\bar{\bar{y}}_\alpha, \tag{5.7}$$

где

$$z_\alpha = \sum_{i=1}^{m} \bar{c}_i D_{i\alpha}, \tag{5.8}$$

или

$$z = \bar{c}D. \tag{5.8а}$$

Таким образом, если надо увеличить f, то необходимо делать положительной такую компоненту $\bar{\bar{y}}$, для которой $(\bar{\bar{c}}_\alpha - z_\alpha)$ больше нуля. При отсутствии вырожденности естественно выбирать компоненту с наибольшей положительной разностью. Если все разности $(\bar{\bar{c}}_\alpha - z_\alpha)$ отрицательны, то решение не может быть улучшено и $\bar{y}$ является решением. Так как одна из компонент $\bar{y}$ должна обратиться в нуль, $\Delta\bar{\bar{y}}_\alpha$ следует выбирать в виде [см. (5.6)]

$$\Delta\bar{\bar{y}}_\alpha = \left(\frac{\bar{y}_\beta}{D_{\beta\alpha}}\right) \tag{5.9}$$

для некоторого β. Чтобы обеспечить выполнение неравенства $x \geqslant 0$ для нового пробного решения, величину β следует выбирать так, чтобы

$$\left(\frac{\bar{y}_\beta}{D_{\beta\alpha}}\right) = \min_i \left(\frac{\bar{y}_i}{D_{i\alpha}}\right), \tag{5.10}$$

так как тогда новые компоненты старого базиса $\bar{y}$ будут иметь вид

$$\bar{y}_i^* = \bar{y}_i + \Delta\bar{y}_i = \bar{y}_i - D_{i\alpha}\left[\min_i \frac{\bar{y}_i}{D_{i\alpha}}\right] \geqslant 0. \tag{5.11}$$

Наконец, запишем новый базис

$$\bar{y}_i^* = \bar{y}_i - D_{i\alpha}\frac{\bar{y}_\beta}{D_{\beta\alpha}}, \quad i \neq \alpha,$$

$$\bar{y}_\alpha^* = \frac{\bar{y}_\beta}{D_{\beta\alpha}}. \tag{5.12}$$

Итеративная процедура продолжается до тех пор, пока не будет получено решение.

Дадим теперь изложение алгоритма, описанного выше.

Этап 1. Вводятся вспомогательные переменные, чтобы перейти от неравенств к уравнениям.

Этап 2. Записывается расширенная матрица $[A; I]$ и вектор ограничений b, который может быть использован для вычисления значения целевой функции на каждом этапе итерации.

Этап 3. Определяется матрица D, задаваемая соотношением (5.5), хотя процедура в действительности определяет матрицу $\beta = [I; D]$. Но единичная матрица, составляющая часть матрицы β, не используется на последующих этапах процедуры. Соответствие между β и D становится очевидным, если записать соотношение (5.5) в виде

$$\bar{\bar{A}} = \bar{A}D. \tag{5.5а}$$

Этап 4. Производится вычисление величин z_α (или вектора z) [(5.8) или (5.8а)] и разностей $(\bar{\bar{c}}_\alpha - z_\alpha)$. Кроме того, вычисляется значение целевой функции

$$f = \bar{\bar{c}}\,\bar{\bar{y}}$$

при помощи соотношения

$$f = \bar{c}\,(\bar{A})^{-1}\,b,$$

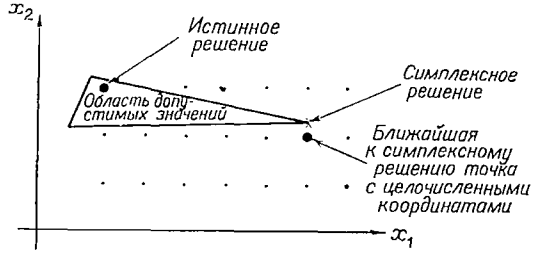
так как $\bar{y}$ должен в силу уравнения (5.2) удовлетворять соотношению $\bar{y} = (\bar{A})^{-1}\,b$. [Коэффициенты a, b линейной комбинации являются компонентами вектора $(\bar{A})^{-1}\,b$.]

Этап 5. Определяется β [для которого $\bar{y}_\beta/D_{\beta\alpha}$ является минимальным среди $\bar{y}_i/D_{i\alpha}$] и вычисляется $\Delta\bar{\bar{y}}_\alpha = \bar{\bar{y}}_\beta/D_{\beta\alpha}$. Остальное очевидно.

Почему аппроксимации, основанные на выборе ближайшего целого, обычно дают плохое решение

Одна из причин, по которой аппроксимации, основанные на выборе ближайшего целого, в линейном программировании не дают правильного решения иллюстрируется фиг. 5.4, где оба ограничения выполняются в вершине, а прямые, изображающие ограничения,

проходят между двумя рядами целочисленной решетки. Предположим, что симплексный процесс дает эту вершину (помеченную крестиком на фиг. 5.4) в качестве решения. Ближайшая целочисленная аппроксимация этой вершины лежит вне области допустимых значений (обведенной отрезками прямых линий) и, следовательно, не



Фиг. 5.4.

может быть использована. Как видно из рисунка, целочисленное решение задачи в действительности может далеко отстоять от симплексного решения. В данном случае в допустимой области есть только одна точка с целочисленными координатами.

5.4. Использование симплексного процесса при решении транспортной задачи [10, 13]

При использовании симплексного процесса для решения транспортной задачи с целыми коэффициентами получается целочисленное решение. Это иллюстрируется следующим примером. Пусть

$$a_1 = \sum_{j=1}^{5} x_{1j} = 24, \quad b_1 = \sum_{i=1}^{4} x_{i1} = 10,$$

$$a_2 = \sum_{j=1}^{5} x_{2j} = 18, \quad b_2 = \sum_{i=1}^{4} x_{i2} = 20,$$

$$a_3 = \sum_{j=1}^{5} x_{3j} = 20, \quad b_3 = \sum_{i=1}^{4} x_{i3} = 10,$$

$$a_4 = \sum_{j=1}^{5} x_{4j} = 16, \quad b_4 = \sum_{i=1}^{4} x_{i4} = 18,$$

$$b_5 = \sum_{i=1}^{4} x_{is} = 20$$

и матрица стоимости имеет вид

$$(c_{ij}) = \begin{bmatrix} 4 & 9 & 8 & 10 & 12 \\ 6 & 10 & 3 & 2 & 3 \\ 3 & 2 & 7 & 10 & 3 \\ 3 & 5 & 5 & 4 & 8 \end{bmatrix}.$$

Таким образом, стоимость перевозки единицы груза из второго пункта отправления в первый пункт назначения равна 6, из четвертого

пункта отправления в пятый пункт назначения равна 8 и т. д. Ставится задача выбрать для перевозки грузы величиной x_{ij} (т. е. выбрать величины в матрице перевозок) при заданных ограничениях так, чтобы функция стоимости, которая теперь может быть записана в виде

$$4x_{11} + 9x_{12} + 8x_{13} + 10x_{14} + 12x_{15} + 6x_{21} + 10x_{22} + 3x_{23} + 2x_{24} + 3x_{25} + 3x_{31} + 2x_{32} + 7x_{33} + 10x_{34} + 3x_{35} + 3x_{41} + 5x_{42} + 5x_{43} + 4x_{44} + 8x_{45},$$

была минимальной.

Ограничения составляют множество из $4 + 5 = 9$ уравнений с 20 неизвестными. Из этих уравнений $m + n - 1$ независимы (напомним, что $\sum_{i=1}^{4} a_i = \sum_{j=1}^{5} b_j$). Поэтому в минимальном решении должно быть не более $m + n - 1$ путей с положительными величинами перевозок.

Найти допустимое решение, в котором не более восьми элементов матрицы перевозок отличны от нуля, несложно. Это делается следующим образом: составляется таблица с соответствующими номерами пунктов отправления и назначения; проставляются данные значения a_i в дополнительный столбец с правой стороны и значения b_j в дополнительную строку в нижней части таблицы так, как показано в табл. 5.1.

Таблица 5.1

	j / i	Пункты назначения 1	2	3	4	5	Итого
Пункты отправления	1						$24 = a_1$
	2						$18 = a_2$
	3						$20 = a_3$
	4						$16 = a_4$
	Итого	$10 = b_1$	$20 = b_2$	$10 = b_3$	$18 = b_4$	$20 = b_5$	78

Клетки в таблице заполняются числами, которые в сумме дают значения, приведенные в указанных строке и столбце. При этом в клетку (1,1) записывается минимальное число из (a_1, b_1), т. е. минимум из двух значений 24 и 10, в данном случае 10. Сдвинемся на одну клетку в направлении максимума (a_1, b_1), в данном случае в клетку (1, 2), стоящую на пересечении первой строки и второго

столбца. Запишем в эту клетку минимальное значение из $(a_1 - 10, b_2)$, равное 14 при $a_1 = 24$, $b_2 = 20$. Затем сдвинемся на одну клетку в направлении максимума $(a_1 - 10, b_2)$, т. е. в данном случае в клетку (2, 2), и запишем здесь минимум $(b_2 - 14, a_2)$, который равен $\min(20 - 14, 18) = 6$. Продолжая этот процесс, можно получить допустимое решение. Очевидно, что потребуется не более $m + n - 1 = 8$ ненулевых значений из матрицы перевозок.

Таблица 5.2

Первая таблица перевозок

Пункты отправления i \ Пункты назначения j	1	2	3	4	5	Итого
1	10		10	4		24
2					18	18
3		20				20
4				14	2	16
Итого	10	20	10	18	20	78

Хотя изложенный выше метод всегда дает допустимое решение и, следовательно, исходную точку для симплексного процесса, вообще говоря, можно получить более удобное решение; объем последующей работы сильно сокращается. При этом не просто выбираются клетки с наименьшими стоимостями перевозок, а берутся клетки, стоимости перевозок в которых после проверки дают «самый выгодный вариант». Чтобы показать это, рассмотрим первый столбец в матрице стоимостей; в третьей и четвертой клетках этого столбца стоят наименьшие в этом столбце значения, но выбор первой клетки оказывается наилучшим. Это следует из того факта, что в третьей строке можно выбрать лучший элемент, чем 3, а именно 2, и в четвертой строке можно выбрать почти такой же хороший элемент, как 3, а именно 4. Однако в первой строке, если не удается использовать первый элемент, который равен 4, следует выбирать между стоимостями, превышающими это число на величины от 4 до 8 единиц.

С учетом всех этих соображений в табл. 5.1 проставляется допустимое решение, которое дано в табл. 5.2. Пустые клетки означают отсутствие перевозок между этими пунктами. Обратите внимание, что использовано менее чем $m + n - 1$ значений. Отметим также, что стоимость при таком выборе равна

$$40 + 80 + 40 + 54 + 40 + 56 + 16 = 326.$$

Чтобы улучшить это допустимое решение, составим новую таблицу, в которой столько же элементов, как и в матрице стоимостей, проставим значения стоимостей c_{ij} в каждую клетку, содержащую ненулевое значение объема перевозки груза, и заключим такие числа

Таблица 5.3

Первая таблица псевдостоимостей

$(\bar{c}_{ij}) =$

[4]	4	[8]	[10]	14
−7	−7	−3	−1	[3]
2	[2]	6	8	12
−2	−2	2	[4]	[8]

в квадратики (см. табл. 5.3). Элементы такой таблицы псевдостоимостей будут обозначаться через $\bar{c}_{ij}$. Следовательно, $c_{ij} = \bar{c}_{ij}$ для всех величин, которые имеют те же индексы, что и ненулевые члены матрицы перевозок. Что касается остальных значений, построим вспомогательную таблицу для значений U_i $(i = 1, \ldots, m)$ и V_j $(j = 1, \ldots, n)$. Для этого выберем любую заключенную в квадратик величину $\bar{c}_{ij}$ и припишем такие произвольные действительные значения U_i и V_j, что

$$U_i + V_j = \bar{c}_{ij} = c_{ij}.$$

Найдем заключенную в квадратик величину, один из индексов которой такой же, как у c_{ij}, скажем c_{ik}, и выберем V_k, такое, что выбранное ранее значение U_i в сумме с V_k составляет c_{ik}. Если таких значений нет, то выберем случайным образом среди заключенных в квадратики значений новое c_{kp} и припишем U_k и V_p такие значения, что $U_k + V_p = c_{kp}$. Продолжим этот процесс до тех пор, пока не будет заполнена вспомогательная табл. 5.4.

Таблица 5.4

Первая вспомогательная таблица

	U_i	V_j
1	4	0
2	−7	0
3	2	4
4	−2	6
5		10

Чтобы посмотреть, как строится эта таблица, выберем клетку (1, 1) и произвольно положим $U_1 = c_{11} = 4$. Так как $U_1 + V_1 = c_{11}$, должно быть $V_1 = 0$. В клетке (1, 3) также проставлено ненулевое значение объема перевозки. Поскольку U_1 и V_3 соответствуют этой клетке, должно иметь место соотношение $U_1 + V_3 = c_{13}$ или $4 + V_3 = 8$, откуда следует $V_3 = 4$. Аналогично $U_1 + V_4 = c_{14}$, что дает $V_4 = 6$. Однако в этом столбце также стоит ненулевое значение объема перевозки в клетке (4, 4). Следовательно, $U_4 + V_4 = c_{44}$. Так как $V_4 = 6$, то $U_4 = -2$ и т. д. После того как вспомогательная таблица заполнена таким путем, все $\bar{c}_{ij} = U_i + V_j$ теперь могут быть вычислены и затем их можно сравнить с исходными c_{ij}. Если все $\bar{c}_{ij} \leqslant c_{ij}$, распределение объемов перевозок дает минимальную стоимость. С другой стороны, рассмотрим только те клетки, в которых разность $\bar{c}_{ij} - c_{ij}$ положительна, и выберем клетку, в которой эта разность максимальна. Получаем табл. 5.5.

Таблица 5.5

$$(\bar{c}_{ij} - c_{ij}) = \begin{bmatrix} 0 & -5 & 0 & 0 & 2 \\ -13 & -17 & -6 & -3 & 0 \\ -1 & 0 & -1 & -2 & 9 \\ -5 & -7 & -3 & 0 & 0 \end{bmatrix}$$

В данном случае $\bar{c}_{ij} > c_{ij}$ только в клетках (1, 5) и (3, 5); в последней из них наблюдается наибольшая разность. Перепишем старую таблицу объемов перевозок, в которой в клетке (3, 5) проставлено значение θ_1. Чтобы получить суммарное ограничение a_3, θ_1 следует вычесть из ненулевого значения объема перевозки в клетке (3, 2). Но чтобы получить b_2, θ_1 надо поместить где-то во втором столбце. Поэтому поместим θ_1 в клетке (4, 2) и вычтем из величины,

Таблица 5.6

Таблица перевозок

Пункты отправления \ Пункты назначения	1	2	3	4	5	Итого
1	10		10	4		24
2					18	18
3		$20-\theta_1$			θ_1	20
4		θ_1		14	$2-\theta_1$	16
Итого	10	20	10	18	20	78

проставленной в клетке (4, 5). Теперь суммарные ограничения везде правильно выполняются. Если имеются две возможные ненулевые величины, к которым можно прибавить θ_1, как это имеет место в четвертой строке, то используется значение, дающее меньшую суммарную стоимость. Таким образом получается табл. 5.6.

Среди выражений, в которых появляется θ_1, отбираем те, которые имеют вид $\alpha - \theta_1$. Затем приписываем θ_1 значение, наименьшее среди этих значений α, и записываем эту величину в таблицу. В результате получаем новую таблицу перевозок. В данном случае $\theta_1 = 2$, и в таблице перевозок остается восемь ненулевых величин, как это требуется из условия, что $m + n - 1$ ограничений должны быть независимы. В результате получается табл. 5.7.

Таблица 5.7

Вторая таблица перевозок

	i \ j	Пункты назначения 1	2	3	4	5	Итого
Пункты отправления	1	10		10	4		24
	2					18	18
	3		18			2	20
	4		2		14		16
	Итого	10	20	10	18	20	78

Процесс повторяется до тех пор, пока не будет выполнено условие $\bar{c}_{ij} < c_{ij}$ для всех c_{ij}. После того как это достигнуто, решение дает минимальную стоимость перевозок. Решение приводится в табл. 5.8—5.17 без дальнейших объяснений.

Таблица 5.8

Вторая таблица псевдостоимостей

[4]	11	[8]	[10]	12
−5	2	−1	1	[3]
−5	[2]	−1	1	[3]
−2	[5]	2	[4]	6

Таблица 5.9

Вторая вспомогательная таблица

U_i	V_i
4	0
—5	7
—5	4
—2	6
	8

Таблица 5.10

Таблица перевозок

Пункты отправления	Пункты назначения					
	1	2	3	4	5	Итого
1	10	θ_2	10	$4-\theta_2$		24
2					18	18
3		18			2	20
4		$2-\theta_2$		$14+\theta_2$		16
Итого	10	20	10	18	20	78

Таблица 5.11

Третья таблица перевозок

Пункты отправления	Пункты назначения					
	1	2	3	4	5	Итого
1	10	2	10	2		24
2					18	18
3		18			2	20
4				16		16
Итого	10	20	10	18	20	78

Таблица 5.12

Третья таблица псевдостоимостей

[4]	[9]	[8]	[10]	10
—3	2	1	3	[3]
—3	[2]	1	3	[3]
—2	3	2	[4]	4

Таблица 5.13

Третья вспомогательная таблица

U_i	V_i
4	0
—3	5
—3	4
—2	6
	6

Таблица 5.14

Таблица перевозок

Пункты отправления i \ j	Пункты назначения 1	2	3	4	5	Итого
1	10	$2+\theta_3$	10	$2-\theta_3$		24
2				θ_3	$18-\theta_3$	18
3		$18-\theta_3$			$2+\theta_3$	20
4				16		16
Итого	10	20	10	18	20	78

Таблица 5.15

Окончательная таблица перевозок

Пункты отправления i \ Пункты назначения j	1	2	3	4	5	Итого
1	10	4	10			24
2				2	16	18
3		16			4	20
4				16		16
Итого	10	20	10	18	20	78

Таблица 5.16

Окончательная таблица псевдостоимостей

[4]	[9]	[8]	9	10
—3	2	1	[2]	[3]
—3	[2]	1	2	[3]
—1	4	3	[4]	5

Таблица 5.17

Окончательная вспомогательная таблица

U_i	V_i
4	0
—3	5
—3	4
—1	5
	6

Новое значение стоимости, соответствующее табл. 5.7, равно

$$40 + 80 + 40 + 54 + 36 + 6 + 10 + 56 = 322.$$

Единственное значение $\bar{c}_{ij}$, большее c_{ij},— это $\bar{c}_{12} = 11$, которому соответствует $c_{12} = 9$. Наименьшее неотрицательное θ_2 — это $\theta_2 = 2$; соответствующие значения приводятся в табл. 5.11.

Стоимость теперь равна

$$40 + 18 + 80 + 20 + 54 + 36 + 6 + 64 = 318.$$

Это несколько лучше, чем в предыдущем случае. Здесь $\bar{c}_{24} > c_{24}$.

Наименьшее неотрицательное значение θ_3 равно 2; соответствующие значения приводятся в табл. 5.15.

Стоимость теперь равна

$$40 + 36 + 80 + 4 + 48 + 32 + 12 + 64 = 316.$$

Это еще немного меньше, чем ранее.

Все $\bar{c}_{ij} \leqslant c_{ij}$, так что последняя таблица перевозок дает оптимальное решение со стоимостью 316.

Если поставки превышают спрос в постановке задачи, то

$$\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j.$$

Если спрос превышает поставки, то указанное неравенство направлено в другую сторону. В первом случае в матрицы стоимостей и объемов перевозок можно добавить лишний столбец со значениями стоимости, равными нулю (в этих столбцах указывается запас). В последнем случае к матрицам стоимостей и объемов перевозок можно добавить лишнюю строку со значениями стоимости, равными M в каждой ячейке. Величина M должна быть очень большой, так что строка используется только для того, чтобы указать пункты потребления, требования которых не выполняются.

Отличительной чертой транспортных задач является то, что они не обязательно имеют единственное решение, и может существовать целое семейство планов перевозок, каждый из которых дает стоимость, не большую, чем любой план, не входящий в это семейство.

5.5. Алгоритм целочисленного программирования

Существует много подходов к решению задач целочисленного программирования. Эти методы излагаются и сравниваются в работах [2, 3, 4, 15, 17, 20, 31, 45]. Наша цель состоит в том, чтобы изложить основной алгоритм Гомори так, чтобы студент, которому надо решить задачу целочисленного программирования, мог бы приступить к работе, владея некоторыми знаниями. На протяжении всей книги мы намеренно избегали глубокого рассмотрения алгоритмов ввиду недостатка места, предпочитая более широко и разнообразно

излагать вопросы математики целочисленных величин. Ниже в этой главе будет дан другой, более общий подход, который называется методом бивалентного программирования. Для задач большой размерности этот метод становится слишком громоздким ввиду огромного количества операций.

Рассмотрим теперь задачу максимизации линейной целевой функции, которую для удобства запишем в виде

$$z = c_0 - \sum_{j=1}^{n} c_j x_j \tag{5.13}$$

(часто c_0 может обращаться в нуль),
при ограничениях

$$y_i \equiv b_i - \sum_{j=1}^{n} a_{ij} x_j \geqslant 0, \quad i = 1, \ldots, m, \tag{5.14}$$

с условием $x_j \geqslant 0$, которое можно переписать как

$$x_j = (-1)\,(-x_j) \geqslant 0, \quad j = 1, \ldots, n, \tag{5.15}$$

где компоненты вектора $x = (x_1, \ldots, x_n)$ — целые числа. Эта задача называется *полностью целочисленной задачей целочисленного программирования*, если a_{ij}, b_i, c_j и x_j — целые числа, и задачей *полностью целочисленного программирования*, если наложено только условие, что x_j — целые числа. Задачи, в которых только некоторые из x_j должны быть целыми числами, называются *задачами частично целочисленного программирования*.

Заметим, что если опустить условие целочисленности для вектора решения x, то остается обычная задача линейного программирования (ЛП), которую можно решить при помощи симплекс-метода. При использовании основного метода программирования поставленная задача целочисленного программирования (ЦП) решается без учета целочисленности. Если решение, полученное таким образом, удовлетворяет условию целочисленности, то оно и является оптимальным решением данной задачи ЦП, поскольку каждое допустимое решение задачи ЦП является также решением соответствующей задачи ЛП, получаемой после отбрасывания условия целочисленности. Однако если оптимальное решение задачи ЛП не является допустимым решением задачи ЦП (условие целочисленности нарушено), то формулируется новая задача ЛП путем добавления новых ограничений. Новое ограничение выбирается так, что множество допустимых решений новой задачи ЛП не включает оптимальное решение первоначальной задачи ЛП, но включает все допустимые решения задачи ЦП. Затем решается новая задача ЛП. Если полученное оптимальное решение является допустимым решением задачи ЦП, то задача решена. В противном случае процесс продолжается с добавлением новых ограничений. Эти дополнительные ограничения называются отсечениями. Они отсекают часть множества допустимых

решений задачи ЛП. Алгоритм Гомори дает такой метод выбора этих отсечений, что процесс отсекания приводит к оптимальному решению задачи ЦП за конечное число шагов [19]. (В [4] дается обзор некоторых других алгоритмов, например метода ветвей и границ. Еще один алгоритм описан в работе [1].)

Чтобы показать, как вводятся отсечения, рассмотрим типичное ограничение, которое записывается следующим образом (здесь $a_{i0} = b_i$):

$$y_i = a_{i0} + \sum_{j=1}^{n} a_{ij}(-x_j), \tag{5.16}$$

или (опуская для удобства индекс i)

$$a_0 + \sum_{j=1}^{n} a_j(-x_j) + 1(-y) = 0.$$

Пусть λ — положительное число, которое будет определено позже. Выражая каждое число a_j/λ, $j = 0, 1, \ldots, n$, в виде суммы целой и дробной частей, можно записать

$$a_j = \left[\frac{a_j}{\lambda}\right]\lambda + r_j, \quad j = 0, 1, \ldots, n.$$

Очевидно, получаем

$$1 = \left[\frac{1}{\lambda}\right]\lambda + r,$$

где

$$0 \leqslant r_j < \lambda \quad \text{и} \quad 0 \leqslant r < \lambda.$$

Подстановка и перегруппировка членов равенства (5.16) дают

$$\sum_{j=1}^{n} r_j x_j + ry = r_0 + \lambda\left\{\left[\frac{a_0}{\lambda}\right] + \sum_{j=1}^{n}\left[\frac{a_j}{\lambda}\right](-x_j) + \left[\frac{1}{\lambda}\right](-y)\right\}. \tag{5.17}$$

Так как r_j, $j = 1, \ldots, n$, и r — неотрицательные числа, левая часть (5.17) неотрицательна. Следовательно, правая часть тоже неотрицательна. Кроме того, для целочисленных значений x_j и y выражение в фигурных скобках в правой части (5.17) является целочисленным, так как все коэффициенты — целые числа. Поскольку $r_0 \not< \lambda$, при отрицательном целочисленном значении выражения в фигурных скобках вся правая часть является отрицательной, что противоречит условию неотрицательности правой части. Таким образом, выражение в фигурных скобках является одновременно целочисленным и неотрицательным. Поэтому

$$y' = \left[\frac{a_0}{\lambda}\right] + \sum_{j=1}^{n}\left[\frac{a_j}{\lambda}\right](-x_j) + \left[\frac{1}{\lambda}\right](-y) \tag{5.18}$$

является новой неотрицательной целочисленной переменной; это выражение следует из (5.16). В первом алгоритме Гомори $\lambda = 1$, но мы не будем рассматривать этот случай. Во втором алгоритме Гомори $\lambda > 1$, откуда получаем $[1/\lambda] = 0$. Таким образом,

$$y' = \left[\frac{a_0}{\lambda}\right] + \sum_{j=1}^{n} \left[\frac{a_j}{\lambda}\right](-x_j) \equiv a_0' - \sum_{j=1}^{n} a_j' x_j. \qquad (5.19)$$

Это новое уравнение, которому должно удовлетворять любое целочисленное решение исходной задачи ЦП. Чтобы показать, как выбирается λ, и, следовательно, дать также геометрическую интерпретацию метода, следует привести все шаги алгоритма, приводящие к выбору λ.

Решим задачу в табличной форме. Типичное i-е ограничение записывается в виде строки матрицы

$$(b_i \mid a_{i1}, a_{i2}, \ldots, a_{in}), \quad i = 1, \ldots, m. \qquad (5.20)$$

Все векторы (5.20) теперь располагаются систематически так, что получается таблица, которая приводится ниже.

0	$c_1 \ldots c_n$	Строка коэффициентов целевой функции
b_1	$a_{11} \ldots a_{1n}$	m базисных ограничений
b_2	$a_{21} \ldots a_{2n}$	
...	.....	
b_m	$a_{m1} \ldots a_{mn}$	
0	$-1 \quad 0 \ldots \quad 0$	Контрольные строки
0	$0 \quad -1 \ldots \quad 0$	
...	.....	
0	$0 \quad 0 \ldots \quad -1$	
		Строка, зарезервированная для дополнительных ограничений

(5.21)

Целевую функцию можно представить в виде $y_0 \equiv z$; она записывается в верхней строке таблицы. Нижняя строка резервируется за векторами дополнительных ограничений. В крайнем слева столбце записаны свободные члены. Таблица (5.21) называется допустимой для прямой задачи, если в столбце свободных членов, возможно, за исключением верхнего элемента, стоят только неотрицательные

элементы. Очевидно, свойство допустимости для прямой задачи означает, что точка $x = 0$ является допустимой. Аналогично таблица называется допустимой для обратной (двойственной) задачи, если вектор целевой функции, возможно, за исключением первого элемента, содержит только неотрицательные элементы.

Лемма. *Если таблица является допустимой и для прямой и для двойственной задач, то $x = 0$ представляет собой оптимальное целочисленное решение задачи целочисленного программирования.*

	1	$-x_1$	$-x_2$	. . .	$-x_n$
$z =$	c_0	c_1	c_2	. . .	c_n
$y_1 =$	b_1	a_{11}	a_{12}	. . .	a_{1n}
$y_2 =$	b_2	a_{21}	a_{22}	. . .	a_{2n}
. . . . .	. . .	. . . . .	. . . . .		. . . . .
$y_m =$	b_m	a_{m1}	a_{m2}	. . .	a_{mn}

Доказательство. Пусть таблица является допустимой для прямой задачи, т. е. $b_i \geqslant 0$, $i = 1, \ldots, m$. Тогда $x_j = 0$, $j = 1, \ldots, n$, является допустимым решением, поскольку $y_i = b_i \geqslant 0$. Если одновременно таблица является допустимой для двойственной задачи, то $c_j \geqslant 0$, $j = 1, \ldots, m$. Максимальное значение выражение $z = c_0 - \sum_{j=1}^{n} c_j x_j$ принимает при $x = (0, \ldots, 0)$, потому что x_j могут принимать только неотрицательные значения. Отсюда следует, что $x = 0$ является оптимальным решением. Этим завершается доказательство.

Этап замены. Как ранее было показано при обсуждении симплекс-метода, задача программирования может быть преобразована путем введения этапа замены, который заключается в введении новой переменной

$$y = d_0 - d_1 x_1 - d_2 x_2 - \ldots - d_n x_n$$

за счет исключения одной из исходных переменных, скажем x_r. Вследствие этого линейное выражение

$$\alpha_0 - \alpha_1 x_1 - \ldots - \alpha_n x_n$$

преобразуется в выражение

$$\bar{\alpha}_0 - \bar{\alpha}_1 x_1 - \ldots - \bar{\alpha}_{r-1} x_{r-1} - \bar{\alpha}_r y - \bar{\alpha}_{r+1} x_{r+1} - \ldots - \bar{\alpha}_n, x_n,$$

где

$$\bar{\alpha}_i = \begin{cases} \alpha_i - \dfrac{d_i}{d_r} a_r & \text{при } i \neq r, \\ -\dfrac{a_r}{d_r} & \text{при } i = r. \end{cases}$$

Отметим, что этап замены — это просто преобразование Гаусса — Иордана, примененное к строке

$$(\alpha_0 \mid \alpha_1, \ldots, \alpha_n)$$

нерасширенной таблицы относительно направляющей строки

$$(d_0 \mid d_1, \ldots, d_n)$$

и направляющего значения d_r.

Опишем теперь подробнее этап замены, чтобы разъяснить введенные ранее понятия. Рассмотрим две линейные формы с двумя переменными x_1, x_2:

$$y_1 = \alpha_0 - \alpha_1 x_1 - \alpha_2 x_2, \tag{5.22}$$

$$y_2 = \beta_0 - \beta_1 x_1 - \beta_2 x_2. \tag{5.23}$$

Предположим, что $\alpha_1 \neq 0$. Из формулы (5.22) получаем

$$x_1 = \frac{\alpha_0}{\alpha_1} - \frac{1}{\alpha_1} y_1 - \frac{\alpha_2}{\alpha_1} x_2. \tag{5.24}$$

Подставляя это выражение для x_1 в (5.23), получаем

$$y_2 = \beta_0 - \beta_1 \left[\frac{\alpha_0}{\alpha_1} - \frac{1}{\alpha_1} y_1 - \frac{\alpha_2}{\alpha_1} x_2\right] - \beta_2 x_2,$$

или

$$y_2 = \left(\beta_0 - \frac{\beta_1 \alpha_0}{\alpha_1}\right) + \frac{\beta_1}{\alpha_1} y_1 - \left(\beta_2 - \frac{\beta_1 \alpha_2}{\alpha_1}\right) x_2. \tag{5.25}$$

Уравнения (5.24) и (5.25) представляют собой две новые линейные формы, которые характеризуются тем, что ранее независимая переменная x_1 теперь стала функцией (зависимой переменной), тогда как y_1 стала независимой переменной. Другими словами, x_1 и y_1 поменялись ролями. Назовем эту операцию «заменой». В принятой здесь табличной записи получаем:

Исходная таблица

	1	$-x_1$	$-x_2$
$y_1 =$	α_0	α_1	α_2
$y_2 =$	β_0	β_1	β_2

(5.26)

Таблица после замены

	1	$-y_1$	$-x_2$
$x_1=$	$\frac{\alpha_0}{\alpha_1}$	$\frac{1}{\alpha_1}$	$\frac{\alpha_2}{\alpha_1}$
$y_2=$	$\left(\beta_0-\frac{\beta_1\alpha_0}{\alpha_1}\right)$	$-\frac{\beta_1}{\alpha_1}$	$\left(\beta_2-\frac{\beta_1\alpha_2}{\alpha_1}\right)$

(5.27)

Определение. Столбец x и строка y замененных переменных пересекаются по элементу α_1; этот элемент называется *направляющим элементом* на данном шаге.

Правила выполнения замены

1. Направляющий элемент заменяется на обратную величину.
2. Остальные элементы направляющего столбца должны быть поделены на направляющий элемент.
3. Другие элементы направляющей строки делятся на направляющий элемент и умножаются на минус единицу.
4. Элементы в оставшейся части матрицы преобразуются следующим образом. Добавляется новая направляющая строка в нижней части таблицы (так называемая фундаментальная строка). К элементу прибавляется произведение элемента фундаментальной строки, стоящего непосредственно под рассматриваемым, и элемента направляющего столбца, который находится в той же строке, что и рассматриваемый.

Как уже отмечалось, этап замены представляет собой просто преобразование Гаусса — Иордана. При использовании алгоритма Гомори исходная таблица должна обладать некоторыми специальными свойствами.

1. Столбцы C_i, $i = 1, \ldots, n$, исходной таблицы, т. е. все столбцы, за исключением столбца свободных членов, должны быть лексикографически положительными. Столбец называется *лексикографически положительным*, если его первый сверху отличный от нуля элемент положителен. Один столбец лексикографически больше другого, если их разность лексикографически положительна.

2. На каждом шаге свободный член b_i в i-й контрольной строке означает величину, принимаемую переменной x_i в точке, в которой все текущие небазисные переменные обращаются в нуль. Очевидно, что исходная таблица является допустимой для двойственной задачи. Если эта таблица также является допустимой и для прямой задачи, то она оптимальна. С другой стороны, рассмотрим ограничение, имеющее отрицательный элемент в столбце свободных членов. В таком случае надо действовать по следующим правилам:

Правило 1. Рассмотрим ограничение с отрицательным элементом в столбце свободных членов. Обозначим его как

$$(p_0 \mid p_1, \ldots, p_n). \tag{5.28}$$

Правило 2. Выберем в качестве направляющего столбца C_r лексикографически наименьший столбец, который имеет отрицательный элемент p_r в строке (5.28).

Правило 3. Для каждого столбца C_k с $p_k < 0$ найдем наибольшее целое $u_k \leqslant C_k/C_r$, т. е. возьмем отношения соответствующих элементов и вычислим $\lambda_k = -p_k/u_k$. Наконец, выберем

$$\lambda = \max_k \lambda_k.$$

Заметим, что $u_r = 1$.

Правило 4. Добавим ограничение $d_0 - d_1x_1 - \ldots - d_nx_n$, где

$$d_j = \left[\frac{p_j}{\lambda}\right], \quad j = 0, 1, \ldots, n,$$

т. е. запишем коэффициенты d_j в нижней строке таблицы. По правилу 3 величина λ выбирается так, что $d_r = -1$. Используем это d_r в качестве направляющего значения при замене.

Замечания

1. Так как на каждом шаге выбирается направляющее значение, равное -1, целочисленность таблицы сохраняется при таком преобразовании.

2. Можно показать следующее:

а) столбцы новой таблицы лексикографически положительны;

б) элементы столбца свободных членов лексикографически убывают.

3. Процесс решения заканчивается, если:

а) таблица становится допустимой для прямой задачи;

б) все элементы строки

$$(p_0 \mid p_1, \ldots, p),$$

выбранные в соответствии с правилом 1, становятся отрицательными, за исключением p_0; при этом условии задача не имеет допустимого решения.

4. Процесс решения заканчивается через конечное число шагов при условии, что существует допустимое целочисленное решение.

Заметим, что в этом алгоритме направляющими могут быть только строки, в которых константы b_i отрицательны. Кроме того, в табличной форме только отрицательные элементы могут быть направляющими. Таким образом, строка может быть направляющей только в том случае, если она начинается с отрицательного свободного члена и содержит другие отрицательные элементы.

Но из $a_j < 0$ следует, что $[a_j/\lambda] < 0$, поэтому если строка, определяемая выражением (5.16), может быть выбрана в качестве направ-

ляющей, то, очевидно, строка, определяемая формулой (5.19), которая получается из нее, тоже является приемлемой.

Таким образом, если в таблице остается хотя бы одна строка, которая может быть выбрана в качестве направляющей, то из любой такой строки при помощи формулы (5.19) можно получить новую приемлемую строку. Если же нет ни одной строки, которую можно было бы выбрать в качестве направляющей, то задача или уже решена, или не имеет решения.

Замечание. Выбор λ (правило 3) дает возможность подправить элементы новой строки таким образом, чтобы направляющий элемент стал равен -1. Конечно, это возможно, поскольку для достаточно больших значений λ все отрицательные d_j становятся равными -1 или обращаются в нуль. Правила для определения λ были получены из следующих двух требований:

1. Направляющее значение должно быть равно -1.
2. Величина λ должна быть как можно меньшей.

Алгоритм

Предположим, что полностью целочисленная исходная таблица является допустимой для двойственной задачи. Если исходная таблица не удовлетворяет условию $c_1 > 0, \ldots, c_n > 0$, то необходимо применить какой-нибудь специальный прием. Один из способов состоит в добавлении еще одного ограничения:

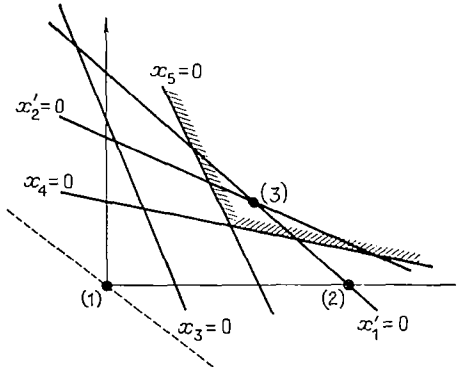
$$y_{m+1} = b_{m+1} - x_1 - x_2 - \ldots - x_n \geqslant 0,$$

где b_{m+1} очень велико; тогда направляющее значение находится в этой строке и в лексикографически наименьшем столбце. Решение этой расширенной линейной системы является решением исходной системы.

Выбираем строку с отрицательным значением свободного члена (если таких строк нет, то задача решена); для такой строки выбираем $\lambda_{\text{мин}}$ и направляющий столбец в соответствии с первыми тремя правилами, приведенными выше. Строим новую строку типа (5.19) и присоединяем эту строку снизу к таблице. Затем делаем замену в новой строке. Так как направляющий элемент равен -1, матрица остается целочисленной. Повторяем эти этапы до тех пор, пока не будет получено решение, допустимое для прямой задачи, или пока не появится ограничение, при котором не существует допустимого решения.

Пример 5.2. Рассмотрим задачу, исходная таблица которой приведена ниже. Первое пробное решение $x_1 = 0$, $x_2 = 0$ не является допустимым. Поэтому выбираем строку x_5, соответствующую максимальному отклонению от условия допустимости, и строим новое ограничение x_1'. Следующее пробное решение $x_1 = 3$, $x_2 = 0$ находится на гиперплоскости $x_1' = 0$. Это решение тоже не является допустимым. Теперь выберем строку x_4, соответствующую макси-

мальному отклонению от условия допустимости, и построим ограничение x_2'. Следующая пробная точка $x_1 = 2$, $x_2 = 1$ лежит на



Фиг. 5.5.

гиперплоскости $x_2 = 0$ и является допустимым и, следовательно, оптимальным решением (фиг. 5.5 и табл. 5.18—5.20).

Таблица 5.18

	1	$-x_1$	$-x_2$	
x_0	0	4	5	(а) Направляющий столбец 1
x_3	−2	−3	−1	(б) (1/1) $4 \geqslant 4$; $\lambda_1 = 3$
x_4	−5	−1	−4	(1/1) $5 \geqslant 4$; $\lambda_2 = 2$
x_5	−7	−3	−2	←
x_1	0	−1	0	(в) $\lambda = 3$
x_2	0	0	−1	
x_1'	−3	−1*	−1	

Таблица 5.19

	1	$-x_1'$	$-x_2$
x_0	−12	4	1
x_3	7	−3	2
x_4	−2	−1	−3 ←
x_5	2	−3	1
x_1	3	−1	1
x_2	0	0	−1
x_2'	−1	−1	−1*

Таблица 5.20

	1	$-x_1'$	x_2'
x_0	—13	3	1
x_3	5	—5	2
x_4	1	2	—3
x_5	1	—4	1
x_1	2	—2	1
x_2	1	1	—1

5.6. Алгоритм полностью целочисленного программирования с параболическими ограничениями [40]

Будет показано, что задача целочисленного программирования минимизации линейной целевой функции при линейных и параболических ограничениях может быть решена за конечное число шагов при помощи небольшой модификации вышеизложенного алгоритма Гомори.

Задача. Требуется минимизировать

$$z = \sum_{i=1}^{n} c_i x_i - c_0, \tag{5.29}$$

где $x_1, \ldots, x_n$ — неотрицательные целые числа, при параболических ограничениях с целочисленными коэффициентами

$$P_j(x) \geqslant 0, \quad j = 1, \ldots, m. \tag{5.30}$$

Коэффициенты c_i, $i = 0, 1, \ldots, n$,— целые числа.

Определение (параболические ограничения). Параболическим ограничением k-го ранга называется квадратичная форма

$$a_{00} - L_0(x) - \sum_{s=1}^{k} b_s (L_s(x))^2 \geqslant 0, \tag{5.31}$$

где b_s — положительные числа, а $L_s(x)$, $s = 0, 1, \ldots, k$,— линейно независимые, однородные, линейные формы от n переменных вида

$$L_s(x) = \sum_{i=1}^{n} a_{si} x_i. \tag{5.32}$$

Как и в алгоритме Гомори, будем решать задачу в табличной форме:

$a_{10}=0$	$a_{11}\ \ldots\ a_{1n}$	b_1
$\ldots$	$\ldots\ \ldots\ \ldots$	$\ldots$
$a_{k0}=0$	$a_{k1}\ \ldots\ a_{kn}$	b_k
a_{00}	$a_{01}\ \ldots\ a_{0n}$	

(5.33)

Верхние k строк составлены из коэффициентов линейных форм

$$L_s(x)=\sum_{i=1}^{n} a_{si}x_i, \quad s\neq 0,$$

которые возводятся в квадрат. Они однородны, так как все $a_{s0}=0$. В нижней строке записаны коэффициенты линейной формы L_0.

Очевидно, если ограничение линейное, то оно занимает только одну линию.

Преобразование. Параболическое выражение

$$a_{00}-L_0(x)-b_1(L_1(x))^2-\ldots-b_k(L_k(x))^2 \tag{5.34}$$

преобразуется путем замены (как в алгоритме Гомори).

В результате преобразования Гаусса — Иордана при направляющем значении d_r в таблице

0	$a_{1j}\ \ldots\ a_{1n}$	b_1
$\ldots$	$\ldots\ \ldots\ \ldots$	$\ldots$
0	$a_{k1}\ \ldots\ a_{kn}$	b_k
a_{00}	$a_{01}\ \ldots\ a_{0n}$	
d_0	$d_1\ \ldots\ d_n$	

(5.35)

и при неизменных значениях b_s получается следующее представление параболического выражения в новых переменных:

$\bar{a}_{10}$	$\bar{a}_{11}\ \ldots\ \bar{a}_{1n}$	b_1
$\ldots\ \ldots$	$\ldots\ \ldots\ \ldots$	$\ldots$
$\bar{a}_{k0}$	$\bar{a}_{k1}\ \ldots\ \bar{a}_{kn}$	b_k
$\bar{a}_{00}$	$\bar{a}_{01}\ \ldots\ \bar{a}_{0n}$	

(5.36)

Однако элементы столбца свободных членов не равны нулю, $\bar{a}_{s0} \neq 0$, $s \neq 0$. Поэтому необходим еще один шаг, чтобы вернуться от неоднородной формы к стандартному виду (восстановление).

Правило восстановления. Заменяем

$$\begin{gathered}\bar{a}_{0i} \text{ на } \bar{a}_{0i} - 2\sum_{s=1}^{k} b_s \bar{a}_{s0}\bar{a}_{si} \quad \text{при} \quad i \neq 0,\\ \bar{a}_{00} \text{ на } \bar{a}_{00} - \sum_{s=1}^{k} b_s \bar{a}_{s0}^2, \qquad (5.37)\\ \bar{a}_{s0} \text{ на } 0 \quad \text{при} \quad s \neq 0.\end{gathered}$$

Другие значения $\bar{a}_{si}$ и b_s остаются неизменными при $i \neq 0$ и $s \neq 0$.

Алгоритм

Как и в алгоритме Гомори, таблица должна удовлетворять следующим требованиям:

1. n верхних ограничений должны быть линейными и линейно независимыми. В качестве таких строк по возможности выбираются контрольные строки или линейные базисные ограничения, но иногда необходимо добавлять подходящие фиктивные ограничения. Причина, по которой может оказаться невозможным использование контрольных строк таблицы в качестве n верхних строк, связана с требованием 2.

2. Все столбцы, за исключением столбца свободных членов, должны быть лексикографически положительными (для сравнения см. алгоритм Гомори). Предположим, что исходная таблица удовлетворяет этим условиям. Тогда алгоритм будет представлять собой последовательность преобразований, применяемых к этой таблице. Рассмотрим самое верхнее ограничение, имеющее отрицательный элемент в столбце свободных членов. Если это ограничение

$$(a_0 \mid a_1, \ldots, a_n)$$

линейное, то поступаем так же, как при использовании алгоритма Гомори. Если это параболическое ограничение, то поступаем так же, как при использовании алгоритма Гомори. Можно показать, что алгоритм заканчивается после конечного числа шагов.

Пример 5.3. Требуется минимизировать $z = 3x + y$, где x и y — неотрицательные целые величины при ограничениях

$$\begin{gathered}6x - 2y - x^2 - 3 \geqslant 0,\\ 5(x + y) - (x - y)^2 - 16 \geqslant 0.\end{gathered}$$

Решение. На первом шаге составляем таблицу. Ограничение

$$5(x + y) - (x - y)^2 - 16 \geqslant 0$$

записываем в виде

	0	−1	1 | 1
Линейная часть →	−16	−5	−5

Не изменяя задачу, добавляем ограничения $x \geqslant 0$, $y \geqslant 0$. Эти ограничения одновременно будут служить и контрольными строками, и двумя верхними линейно независимыми ограничениями. Таким образом, исходная таблица имеет такой вид, как таблица 5.21.

Таблица 5.21

	Столбец свободных членов			
	0	3	1	
Контрольные строки	0	−1	0	
	0	0	−1	
Первое ограничение	0	−1	0	1
	−3	−6	2	
Второе ограничение	0	−1	1	1
	−16	−5	−5	
	−1	−1*	0	

Находим направляющий элемент, необходимый для замены; выбираем самое верхнее ограничение с отрицательным элементом в столбце констант. Заметим, что это вектор-строка

−3	−6	2

Здесь направляющим столбцом служит C_1, соответствующий отрицательному элементу −6. Из правила 3 следует, что

$$u_1 = \left[\frac{3}{3}\right] = 1, \quad \lambda = \left[\frac{+6}{1}\right] = 6.$$

В соответствии с правилом 4 добавляем ограничение $d_0 - d_{1x} - d_{2y}$; здесь

$$d_0 = \left[\frac{-3}{6}\right] = \left[-\frac{1}{2}\right] = -1, \quad d_1 = -1, \quad d_2 = 0.$$

Выбрав в качестве направляющего значения d_1, делаем замену, что дает табл. 5.22, а затем восстановление, что дает табл. 5.23. Предоставляем читателю проверить этап восстановления. Продолжаем процесс, выбирая в качестве нового опорного значения $d_2 = -1$. Новые таблицы приведены ниже (табл. 5.24—5.26).

Табл. 5.26 является оптимальной и даст решение $x = y = 2$.

В качестве практического примера ниже рассмотрена задача о спутнике.

Таблица 5.22

−3	3	1	
1 0	−1 0	0 −1	
1	−1	0	1
3	−6	2	
1	−1	1	1
−11	−5	−5	

Таблица 5.23

−3	3	1	
1 0	−1 0	0 −1	
0	−1	0	1
2	−4	2	
0	−1	1	1
−12	−3	−7	
−2	−1	−1*	

Таблица 5.24

−5	2	1	
1 2	−1 1	0 −1	
0	−1	0	1
−2	−6	2	
0	−2	1	1
−2	−4	−3	
−1	−1*	0	

Таблица 5.25

−7	2	1	
2 1	−1 1	0 −1	
0	−1	0	1
3	−4	2	
0	−2	1	1
−2	4	−7	
−1	0	−1*	

Таблица 5.26

	−6	2	1	
$x \rightarrow$ $y \rightarrow$	2 2	−1 1	0 −1	
	0	−1	0	1
	1	−4	2	
	0	−2	1	1
	4	2	−5	
	*	*	*	

5.7. Алгебраическая формулировка задач

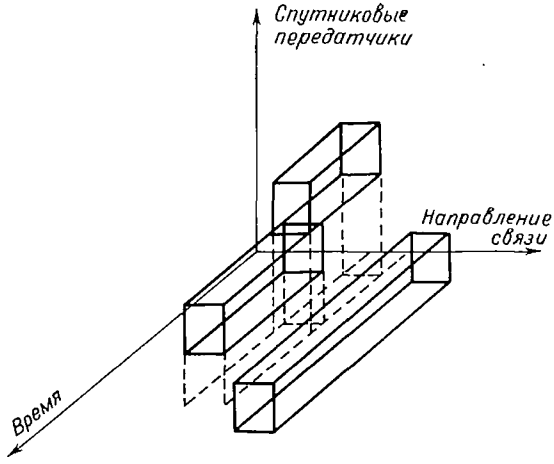
Задача с использованием спутников связи [33]

В гл. 1 в качестве примера приводилась задача о связи через спутники. Здесь она будет сформулирована в виде задачи двоичного программирования.

Прежде чем представить модель, рассмотрим трехмерную диаграмму (фиг. 5.6), иллюстрирующую возможность использования спутникового передатчика в данном направлении связи в течение некоторого времени t, где t разделено на дискретные интервалы равной продолжительности. Примем, что направления связи и спутники упорядочены и пронумерованы и что всего имеется n направлений связи и m спутников. Без потери общности можно принять, что на каждом спутнике установлено k передатчиков и, следовательно,

что все km передатчики пронумерованы в естественном порядке. Из фиг. 5.6 видно, что на спутниках может быть установлено разное количество передатчиков. Наша модель пригодна при любом количестве передатчиков на спутнике. Приемно-передающим устройством будем называть систему приемник — передатчик.

В качестве начального шага дадим упрощенную формулировку задачи для случая одного периода и простой целевой функции. Затем



Фиг. 5.6.

обобщим формулировку на случай бо́льших отрезков времени и более сложных целевых функций.

Обозначим спутниковые передатчики через i, $i = 1, \ldots, m$, направления связи — через j, $j = 1, \ldots, n$, а время в дискретных единицах — через t. Далее положим

$$a_{ijt} = \begin{cases} 1, \text{ если передатчик } i \text{ можно использовать} \\ \text{в направлении связи } j \text{ в момент } t \text{ и он} \\ \text{должен вести передачу,} \\ -p \text{ в противном случае } (p \text{ — произволь-} \\ \text{ное положительное число),} \end{cases}$$

$$X_{ijt} = \begin{cases} 1, \text{ если передатчик } i \text{ используется в на-} \\ \text{правлении связи } j \text{ в момент } t, \\ 0 \quad \text{в противном случае} \end{cases}$$

и пусть r_{ijt} означает количество спутниковых передатчиков, которые должны использоваться в направлении связи j в момент t. Обозначив стоимость, получаемую в результате использования передатчика i в направлении связи j в момент t, через c_{ijt}, можно сформулировать следующую задачу о назначениях в момент t_0.

Задача. Найти X_{ijt_0}, такое, что

$$\sum_j X_{ijt_0} = 1$$

(передатчик в данный момент можно использовать только в одном направлении связи),

$$\sum_i X_{ijt_0} = r_{jt_0}$$

(в направлении связи j следует использовать определенное число передатчиков) и

$$\sum_{i,j} c_{ijt_0} a_{ijt_0} X_{ijt_0} = \max.$$

Замечание. Отметим, что в вышеизложенной формулировке ограничения заданы в виде равенств. Очевидно, что это не обязательно, так как число передатчиков, которые можно использовать в данном направлении связи, может превышать заданную величину или наоборот. Наша задача является специальным случаем транспортной задачи. Ограничениями в такой задаче являются следующие:

1. Если в какой-то момент времени для всех направлений связи число требуемых передатчиков превышает доступное число передатчиков, то в алгоритм надо ввести один искусственный (или фиктивный) спутниковый передатчик, чтобы можно было записать ограничение на числа требуемых и доступных передатчиков в виду равенства, а не неравенства.

2. Если для передачи информации требуется меньше передатчиков, чем количество передатчиков, с которыми возможна связь, то надо ввести искусственное направление связи, чтобы превратить неравенство в равенство. Для искусственного направления связи q полагаем $a_{qjt} = 1$ при всех j и t. Стоимость, соответствующая фиктивному передатчику, полагается равной нулю.

Если количество приемно-передающих устройств ограничено возможностями антенн, можно учесть и это ограничение путем добавления к приведенным выше суммам ограничений на $\sum_i X_{ijt_0}$. В таком случае эта сумма не должна превышать количества приемно-передающих устройств, которые можно использовать в направлении связи j. Наложение таких ограничений не выводит нашу задачу за пределы транспортной задачи.

Модифицированная модель. Модифицируем модель в двух существенно важных направлениях. Во-первых, переформулируем ее таким образом, чтобы расписание составлялось для нескольких периодов. Кроме того, будем использовать целевую функцию, которая включает другие важные факторы, помимо рассмотренных ранее. Если обозначить матрицу $(a_{ij,\,t+u})$ в момент $t + u$ через A_{t+u}, то получится более сложная задача о назначениях для нескольких

периодов, в которой используется блочно-диагональная матрица вида

$$\text{Передатчики спутника} \quad \begin{matrix} \text{Направления связи} \\ \left[\begin{array}{ccccc} A_t & 0 & \dots & 0 & I \\ 0 & A_{t+1} & \dots & 0 & I \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_{t+s} & I \\ \hline R_t & R_{t+1} & \dots & R_{t+s} & \end{array}\right] \end{matrix},$$

где величины I в правой части представляют собой вектор-столбцы, состоящие из единичных элементов, указывающих на возможность использования передатчиков в определенное время; R_{t+u} — вектор-строка, элементами $r_{j,\,t+u}$ которой являются числа требуемых передатчиков в направлении связи j в момент $t+u$.

Задача формулируется теперь следующим образом:

Задача. Найти $X_{ij,\,t+u}$, такие, что

$$\sum_j X_{ij,\,t+u} = 1,$$

$$\sum_i X_{ij,\,t+u} = r_{j,\,t+u} \text{ для всех } j \text{ и } u$$

и

$$\sum_{i,\,j,\,u} c_{ij,\,t+u} a_{ij,\,t+u} X_{ij,\,t+u} = \max.$$

Целевая функция тоже модифицируется путем введения двух дополнительных факторов. Первый заключается в том, что требуется составить расписание таким образом, чтобы станция в общем случае не переключалась с одного спутника на другой, если один из них уже используется в каком-то направлении связи, так как переключение увеличивает расходы. Пусть $d_{ij,\,t+u}$ — стоимость переключения. Тогда вторая наша цель состоит в минимизации расходов на переключение, которые можно выразить с помощью функции

$$\sum_{i,\,j} \sum_{u=0}^{s} d_{ij,\,t+u} \left(a_{ij\ t+u+1} X_{ij\ t+u+1} - a_{ij\ t+u} X_{ij,\,t+u}\right)^2$$

Квадрат здесь используется потому, что при наличии переключения стоимость должна быть положительной. Если передача занимает один период времени, то стоимость переключения учитывается в моменты, когда передача заканчивается и, возможно, также когда передача начинается. По-видимому, это не является серьезным недостатком. (В противном случае фактор переключения можно учесть при помощи какого-нибудь другого соотношения.) Другая цель состоит в том, чтобы минимизировать число случаев, когда передатчики не используются или когда их не хватает. Если обозначить стоимость, связанную с этим фактором, через $e_{j,\,t+u}$, то надо минимизи-

ровать

$$\sum_{j,\,u} e_{j,\,t+u}\Big(\sum_{i} a_{ij,\,t+u}X_{ij,\,t+u} - r_{j,\,t+u}\Big)^2.$$

Эти три целевые функции, доход, убытки от переключений и невыполнений заявок можно объединить в одну, умножая первую целевую функцию на -1 и переходя от задачи максимизации к задаче минимизации взвешенного среднего.

Теперь задача формулируется следующим образом:

Задача. Требуется найти $X_{ij,\,t+u}$, такие, что

$$\sum_{j} X_{ij,\,t+u} = 1,$$

$$\sum_{i} X_{ij,\,t+u} = r_{j,\,t+u}$$

и

$$-\alpha \sum_{i,\,j,\,u} c_{ij,\,t+u}a_{ij,\,t+u}X_{ij,\,t+u} +$$

$$+\beta \sum_{i,\,j,\,u} d_{ij,\,t+u}\left(a_{ij,\,t+u+1}X_{ij,\,t+u+1} - a_{ij,\,t+u}X_{ij,\,t+u}\right)^2 +$$

$$+(1-\alpha-\beta)\sum_{j,\,u} c_{j,\,t+u}\Big(\sum_{i} a_{ij,\,t+u}X_{ij,\,t+u} - r_{ij,\,t+u}\Big)^2 = \min,$$

где $\alpha + \beta < 1$, $\alpha \geqslant 0$, $\beta \geqslant 0$.

Эта задача представляет собой задачу о назначениях с квадратичной целевой функцией. Отметим, что, так как $X_{ij,\,t+u}$ принимает двоичные значения, можно записать $X^2_{ij,\,t+u} = X_{ij,\,t+u}$; однако этого недостаточно для линеаризации целевой функции.

Вводя переменную

$$z = a + \sum_{i=1}^{n} a_iX_i + \sum_{i,\,j=1}^{n} a_{ij}X_iX_j,$$

получаем задачу минимизации линейной целевой функции z при дополнительном параболическом ограничении

$$z - a - \sum_{i=1}^{n} a_iX_i - \sum_{i,\,j=1}^{n} a_{ij}X_iX_j \geqslant 0,$$

к которой применим метод, обсуждавшийся в разд. 5.6.

Блочно-диагональная матрица может быть использована для составления расписания на один или более периодов времени, например для периода $(t,\ t+v)$, где $v \leqslant s$. Затем блочные матрицы других периодов используются в необходимом по времени порядке для образования следующей блочно-диагональной матрицы, дополнительными членами которой являются блоки, описывающие возможность использования передатчиков в определенных направлениях связи в последующие периоды времени. Таким образом, количество передатчиков, которые можно применять в некотором направлении связи, и число

требуемых передатчиков используются для того, чтобы повлиять на расписание в настоящее время, в частности в направлении сокращения числа переключений.

Линеаризация. Работа по решению (в целых числах) квадратичных задач о назначениях, таких, как описанная выше, продолжается. Однако в настоящее время наиболее полезным алгоритмом, дающим целочисленные значения, является алгоритм транспортной задачи, который может быть применен к задаче о назначениях с линейной целевой функцией.

Так как целевая функция является квадратичной, желательно было бы изменить эту функцию так, чтобы она стала линейной. Например, если $\alpha = 1$ (и, следовательно, $\beta = 0$) и требуется максимизировать доход, то задача является линейной. Однако при помощи подходящего выбора коэффициентов стоимости можно ввести дополнительный линейный член. Рассмотрим временной интервал $(t, t + s)$, в котором величина s выбирается достаточно малой для того, чтобы существовало допустимое расписание. Пусть b_{jt} — время в интервале $(t, t + s)$, остающееся до момента прекращения использования передатчика в направлении связи j; f_{ijt} — длительность отрезка времени, начиная с момента t, в течение которого i-й передатчик может быть использован в j-м направлении связи. Определим

$$g_{ijt} = \begin{cases} b_{jt} - f_{ijt}, & \text{если разность } \geqslant 0, \\ 0 & \text{в противном случае.} \end{cases}$$

При этом линейная функция стоимости, которую надо минимизировать (и к которой применим алгоритм транспортной задачи), имеет вид

$$-\alpha \sum_{i,\, j,\, u} c_{ij,\, t+u} a_{ij,\, t+u} X_{ij,\, t+u} + (1-\alpha) \sum_{i,\, j,\, u} g_{ij,\, t+u} X_{ij,\, t+u},$$

где $0 \leqslant \alpha \leqslant 1$.

Удобным методом отыскания решений задачи в различных изложенных выше формулировках может служить бивалентное программирование, которое будет описано в разд. 5.8.

Замечание. Если задача решена для одного периода, в течение которого возможна связь, т. е. величина X_{ijt} получена для одного значения t, то переключения могут быть учтены путем модификации матрицы $(a_{ij,\, t+1})$ на следующий период времени с использованием расписания на данный период времени. В этом случае введем

$$a^*_{ij,\, t+1} = a_{ij,\, t+1}(1 + CX_{ijt}),$$

где C — константа, значение которой выбирается так, чтобы были уравновешены стоимость переключения и стоимость невыполнения требований. В рассматриваемом случае $0 \leqslant C \leqslant 1$. При этом целевой функцией является доход и ставится задача максимизации при ограничениях в момент $t+1$.

Задача о потоках в сетях

Пусть вершины v_0 и v_n направленного графа означают источник и сток некоторой субстанции, проходящей по дугам. Кроме того, примем, что дуга, соединяющая вершины v_i и v_j, имеет определенную пропускную способность, или верхний предел потока c_{ij}. Наконец, пусть C_{ij} — стоимость прохождения единицы потока по дуге. Задача о потоке формулируется как задача линейного программирования, в которой требуется минимизировать $\sum\limits_{i,\,j} C_{ij}x_{ij}$ при полном потоке c из v_0 в v_n и при ограничениях

$$\sum_j (x_{0j} - x_{j0}) = c,$$

$$\sum_j (x_{ij} - x_{ji}) = 0, \quad i = 1, \ \ldots, \ n-1,$$

$$\sum_j (x_{nj} - x_{jn}) = -c,$$

$0 \leqslant x_{ij} \leqslant c_{ij}$ для каждой дуги.

Задачи линейного программирования такого типа, которые ставятся как транспортные задачи, иногда удобно рассматривать как задачи о потоках в сетях.

Задача о коммивояжере [11, 31a]

Пусть задана матрица d_{ij}, $i \neq j$, расстояний между любыми двумя вершинами графа. Требуется проложить маршрут, который начинается в «базовом городе» с индексом 0, проходит через все намеченные точки в определенном направлении и заканчивается в исходной точке. Если положить $x_{ij} = 1$ или 0 в зависимости от того, проходит или нет направленная дуга (на схеме, представляющей возможные пути) из вершины i в вершину j, то можно записать следующие условия на маршрут (частный случай элегантной формулировки, данной в [31a]):

$$\sum_{\substack{i=0 \\ i \neq j}}^{n} x_{ij} = 1, \qquad j = 1, \ \ldots, \ n,$$

$$\sum_{\substack{j=0 \\ j \neq i}}^{n} x_{ij} = 1, \qquad i = 1, \ \ldots, \ n,$$

$$\sum_{i=1}^{n} x_{i0} = 1,$$

$u_i - u_j + nx_{ij} \leqslant n-1$, u_i, $i = 1, \ \ldots, \ n$, — произвольные действительные числа.

В этой задаче требуется минимизировать длину маршрута

$$\sum_{0 \leqslant i \neq j \leqslant n} \sum d_{ij} x_{ij}.$$

Задача о четырех цветах [11]

Пусть области на плоской карте пронумерованы как $r = 1, 2, \ldots, n$. Переменная t_r принимает целые значения $0 \leqslant t_r \leqslant 3$. Таким образом, t_r приписывает области с номером r один из четырех цветов, обозначенных как 0, 1, 2, 3. Если две области r и s имеют общую границу, то $t_r - t_s \neq 0$. Такое соотношение записывается для каждой пары смежных областей. Соотношение для одной пары может быть записано как

$$\text{либо } t_r - t_s \geqslant 1, \quad \text{либо } t_s - t_r \geqslant 1.$$

Эту пару неравенств можно переписать в виде

$$t_r - t_s \geqslant 1 - 4\delta_{rs}$$

и

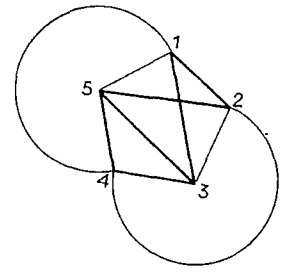
$$t_s - t_r \geqslant -3 + 4\delta_{rs},$$

где $\delta_{rs} = 0$ или 1. Изменяя r и s от 1 до n, получаем систему таких неравенств. Задача тогда заключается в том, чтобы определить, можно ли выбрать целые числа $0 \leqslant t_r \leqslant 3$, $r = 1, \ldots, n$, и двоичные переменные δ_{rs}, $r, s = 1, \ldots, n$, так, чтобы система неравенств имела решение. Если нет, то предположение о том, что t_r принимает только четыре значения, неправильно.

Алгебраическая формулировка задачи о пересечении для полного графа

Задача о пересечении для полного графа уже обсуждалась в гл. 2; дадим теперь изящную алгебраическую формулировку, предложенную Бейсином [4]. Легко показать, что в полном графе всегда существует гамильтонов цикл. Однако подход Бейсина предполагает, что граф с минимальным числом пересечений содержит гамильтонов цикл H с дополнительным условием, что на его ребрах нет пересечений. Причина этого условия поясняется схемой реализации (см. работу [5] в списке литературы к гл. 2). Эта схема является альтернативой к представлению в виде многоугольника, описанному в гл. 2. Из рассмотрения цикла H, который представляет собой простую замкнутую кривую, очевидно, что ребра, не входящие в H, могут быть или внутренними, или внешними по отношению к H. Все ребра, внутренние по отношению к H, вместе с любым подмножеством ребер H определяют подграф G_1. Ребра, внешние по отношению к H, вместе с остальными ребрами H определяют подграф G_2, который является дополнением к G_1 во всем графе, т. е. вме-

сте они составляют весь граф. Вершины, которые все принадлежат H, пронумерованы как 1, 2, . . ., n в направлении движения часовой стрелки. На фиг. 5.7 изображен полный граф с пятью вершинами, подграф выделен жирными линиями.



Ф и г. 5.7.

Заметим, что пересекаться могут только ребра, внутренние или внешние по отношению к H. В силу выбранного способа нумерации ребра пересекаются тогда и только тогда, когда номера их соответствующих вершин перекрываются, т. е. ребра e_{ij} и e_{hl} пересекаются тогда и только тогда, когда они находятся в одном подграфе (G_1 или G_2) и $1 \leqslant i < l < j < k \leqslant n$. Для всех i и j, которые удовлетворяют условию $1 \leqslant i < j \leqslant n$, положим

$$a_{ij} = \begin{cases} 1, & \text{если } e_{ij} \in G_1, \\ 0, & \text{если } e_{ij} \in G_2, \end{cases}$$

а для всех i и j, которые удовлетворяют условию $1 \leqslant j < i \leqslant n$, положим

$$a_{ij} = \begin{cases} 1, & \text{если } e_{ij} \in G_2, \\ 0, & \text{если } e_{ij} \in G_1. \end{cases}$$

Для примера рассмотрим матрицу (a_{ij}), соответствующую фиг. 5.7:

$$(a_{ij}) = \begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 0 & 1 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 1 \\ 3 & 0 & 1 & 0 & 1 & 1 \\ 4 & 1 & 1 & 0 & 0 & 1 \\ 5 & 1 & 0 & 0 & 0 & 0 \end{array}$$

На главной диагонали все элементы равны нулю, так как в графе отсутствуют петли. Две диагонали, проходящие снизу и сверху от главной, определяют, принадлежит ли некоторое ребро H подграфу G_1 или G_2. Отметим, что элементы треугольника над главной диагональю определют G_1, а элементы нижнего треугольника определяют G_2. Это следует из определения a_{ij} и связано со свойством перекрывания индексов.

Теперь $a_{ij}a_{hl} = 1$, если ребра e_{ij} и e_{hl} принадлежат одному подграфу. Кроме того, эти ребра пересекаются, если $1 \leqslant i < k < j < l \leqslant n$ (т. е. их индексы перекрываются). Таким образом, общее число пересечений можно записать следующим образом:

$$\underset{\text{Пересечения в } G_1}{\sum_{1 \leqslant i < k < j < l \leqslant n}} a_{ij}a_{hl} + \underset{\text{Пересечения в } G_2}{\sum_{1 \leqslant l < j < k < i \leqslant n}} a_{ij}a_{hl}.$$

Заметим, что если ребро, соединяющее i с j, принадлежит G_1, то ребро, соединяющее j с i, которое является тем же ребром, не принадлежит G_2; следовательно, $a_{ji} = 1 - a_{ij} \equiv \bar{a}_{ij}$. Чтобы облегчить вычисление по этой формуле, отметим, что, для того чтобы e_{ij} и e_{kl} пересекались, должны пересекаться их индексы. Следовательно, в e_{ij} индекс j должен превышать i не менее чем на 2, так что i не может превышать $n - 3$, потому что тогда j может принимать только значения $n - 1$ (и n), и, чтобы получилось пересечение, k должно быть равно $n - 2$, l должно быть равно n и, следовательно, j в действительности не может быть равно n. Таким образом, j может быть равно самое большее $n - 1$. Аналогично k меняется между $i + 1$ и $j - 1$, а l — между $j + 1$ и n. Таким образом, имеем

$$\sum_{i=1}^{n-3}\sum_{j=i+2}^{n-1} a_{ij}\Big\{\sum_{k=i+1}^{j-1}\sum_{l=j+1}^{n} a_{kl}\Big\} + \sum_{i=1}^{n-3}\sum_{j=i+2}^{n-1} \bar{a}_{ij}\Big\{\sum_{k=i+1}^{j-1}\sum_{l=j+1}^{n} \bar{a}_{kl}\Big\}.$$

Используя $\bar{a}_{ij} = 1 - a_{ij}$, получаем, что число пересечений составляет

$$C_n^4 - \sum_{i=1}^{n-2}\sum_{j=i+2}^{n} (j-i-1)(n-j+i-1)\,a_{ij} + 2\sum_{i=1}^{n-3}\sum_{j=i+2}^{n-1} a_{ij}\Big\{\sum_{k=i+1}^{j-1}\sum_{l=j+1}^{n} a_{kl}\Big\}.$$

Задача теперь состоит в том, чтобы придать a_{ij} такие значения (0 или 1), которые минимизировали бы написанное выше выражение. Заметим, что в полном графе максимальное число пересечений равно C_n^4. Этот метод подразумевает выборочное «изъятие» ребер из внутренней области многоугольника с максимальным пересечением, описанного в гл. 2.

Дихотомическая задача [11] (случай двух взаимно исключающих условий)

Предположим, что задача ставится таким образом, что должно быть выполнено условие типа либо — либо, т. е. выполняется либо условие $G(x_1, \ldots, x_n) \geqslant 0$, либо условие $H(x_1, \ldots, x_n) \geqslant 0$ для $x_1, \ldots, x_n$, принадлежащих некоторому множеству S. Пусть L_G — нижняя граница для G, а L_H — нижняя граница для H; δ — двоичная переменная, которая принимает значения 0 или 1. Тогда задача формулируется следующим образом: найти $x_1, \ldots, x_n \in S$ и δ, такие, что неравенства

$$G(x_1, \ldots, x_n) - \delta L_G \geqslant 0$$

и

$$H(x_1, \ldots, x_n) - (1 - \delta)\,L_H \geqslant 0$$

выполняются. Заметим, что если $\delta = 1$, то первое условие выполняется, а если $\delta = 0$, то второе условие выполняется. Условия $0 \leqslant x_1 \leqslant 2$, $0 \leqslant x_2 \leqslant 2$ и либо $x_1 \leqslant 1$, либо $x_2 \leqslant 1$ можно заменить

на условия

$$0 \leqslant x_1 \leqslant 1 + \delta, \quad 0 \leqslant x_2 \leqslant 2 - \delta, \quad 0 \leqslant \delta \leqslant 1, \quad \delta = 0, 1.$$

В общем случае если решение

$$G_1(x_1, \ldots, x_n) \geqslant 0, \ \ldots, \ G_m(x_1, \ldots, x_n) \geqslant 0$$

должно быть таким, чтобы выполнялись одновременно k условий, то систему можно заменить на

$$G_1 - \delta_1 L_1 \geqslant 0, \ \ldots, \ G_m - \delta_m L_m \geqslant 0,$$

где L_i — нижняя граница для G_i, $i = 1, \ldots, m$, а δ_i — целочисленные переменные, которые удовлетворяют условию

$$\delta_1 + \delta_2 + \ldots + \delta_m = m - k, \quad 0 \leqslant \delta_i \leqslant 1, \quad i = 1, \ldots, m.$$

Например, условие, что x_1 принимает одно из значений $a_1, a_2, \ldots, a_k$, можно переписать в виде

$$x_1 = a_1\delta_1 + \ldots + a_k\delta_k, \quad \delta_1 + \ldots + \delta_k = 1, \quad \delta_i = 0 \text{ или } 1.$$

Если на другие переменные наложены аналогичные условия, то они могут быть представлены в таком же виде.

5.8. Псевдобулевы методы в бивалентном программировании [1)]

Изложение метода бивалентного программирования разбито на три части. В первой части изучаются решения линейных систем псевдобулевых уравнений и неравенств. Во второй части обсуждается случай нелинейных систем псевдобулевых уравнений и неравенств. В последней, третьей части рассматривается задача минимизации, но не максимизации, псевдобулевой функции с ограничениями и без них. Главным источником информации здесь является работа Хэммера (Ивэнеску) и Рудеану [22].

Булева алгебра B_2 двух элементов состоит из множества $\{0, 1\}$, над которым определены две бинарные и одна унитарная операции, которые называются дизъюнкция ($\cup$), конъюнкция ($\cap$) и дополнение ($\bar{x}$) соответственно. Эти операции определяются следующими таблицами:

Дизъюнкция

$\cup$	0	1
0	0	1
1	1	1

Конъюнкция

$\cap$	0	1
0	0	0
1	0	1

Дополнение

x	0	1
$\bar{x}$	1	0

Определение. Псевдобулева функция представляет собой отображение f, такое, что

$$f\colon B_{2^n} \to \mathrm{Re},$$

[1)] Автор выражает благодарность Алану Р. Кертису за его большой вклад в подготовку этого раздела.

где $B_{2^n} = \{(x_1, \ldots, x_n) \mid x_i \in B_2, i = 1, \ldots, n\}$, а Re означает поле действительных чисел.

Установив соответствие

$$x \cup y = x + y - xy,$$
$$x \cap y = xy$$

и

$$\bar{x} = 1 - x,$$

любую булеву функцию g, g: $B_2{}^n \to B_2$ можно рассматривать как псевдобулеву функцию.

Так как аргументы псевдобулевой функции являются бивалентными величинами, можно показать, что каждая псевдобулева функция линейна по каждой переменной и, следовательно, представляется в виде полинома с коэффициентами из поля действительных чисел. Таким образом, имеем следующую теорему:

Теорема 5.5. *Каждая псевдобулева функция f линейна по каждой из своих переменных и, кроме того, может быть представлена в виде полинома над полем действительных чисел, который после приведения подобных членов однозначно определяется с точностью до порядка сумм и произведений.*

Доказательство. Чтобы доказать линейность функции $f(x_1, \ldots, x_n)$ по каждой из ее переменных, надо показать, что существуют псевдобулевы функции h_i, g_i, такие, что

$$f(x_1, \ldots, x_n) = x_i g_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) + h_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n), \quad i = 1, \ldots, n. \tag{5.38}$$

Выберем

$$g_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n)$$

и

$$h_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = f(x_1, \ldots, x_{i-1}, 0, x_{i+1}, \ldots, x_n).$$

Тогда получаем

$$f(x_1, \ldots, x_n) = x_i g_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) + h_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n).$$

Таким образом, функция $f(x_1, \ldots, x_n)$ линейна по каждой из своих переменных x_i. Следовательно, из определения

$$g_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

и

$$h_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

получаем, что левая часть уравнения (5.38) представляет собой псевдобулеву функцию переменных $x_1, \ldots, x_n$. Последовательно применяя соотношение (5.38), можно выразить функцию $f(x_1, \ldots, x_n)$ в виде полинома от переменных $x_1, \ldots, x_n$, коэффициенты которых представляют собой суммы и произведения $f(\alpha_1^i, \ldots, \alpha_n^i)$, где $\alpha_k^i \in B_2$, и получаются в процессе приведения. После приведения подобных членов единственность представления следует из того факта, что полином является псевдобулевой функцией и, следовательно, линеен по каждой из своих переменных. Таким образом, если $P_1(x_1, \ldots, x_n)$ и $P_2(x_1, \ldots, x_n)$ — два представления в виде полиномов функции $f(x_1, \ldots, x_n)$, то

$$P_1(x_1, \ldots, x_n) = x_i Q_1 + R_1 = x_i Q_2 + R_2 = P_2(x_1, \ldots, x_n), \tag{5.39}$$

где Q и R — полиномы от $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$. Тогда из уравнения (5.39) получаем $R_1 = R_2$ и $Q_1 = Q_2$. Таким образом, представление единственно, так как Q_1, R_1, Q_2, R_2 — минимальные полиномы и делителями их служат только 1 и они сами. Этим заканчивается доказательство.

Как и в случае булевых функций, существует каноническая форма псевдобулевой функции, аналогичная дизъюнктивной форме булевой функции.

Теорема 5.6 (М. Карвалло). *Всякую псевдобулеву функцию можно записать в виде*

$$f(x_1, \ldots, x_n) = \sum_{\alpha \in B_{2^n}} C(\alpha)\, x_1^{\alpha_1} \ldots x_n^{\alpha_n}, \tag{5.40}$$

где

$$\alpha = (\alpha_1, \ldots, \alpha_n),$$
$$x_i^1 = x_i, \quad x_i^0 = \bar{x}_i$$

и

$$C(\alpha) = f(\alpha_1, \ldots, \alpha_n).$$

Чтобы привести пример псевдобулевой функции, выберем $n = 2$ и определим f соотношением

$$f(x_1, x_2) = e^{x_1 + x_2}.$$

По теореме 5.5

$$f(x_1, x_2) = (e^2 - 2e + 1)\, x_1 x_2 + (e - 1)\, x_1 + (e - 1)\, x_2 + 1,$$

а по теореме 5.6 можно составить следующую таблицу:

$\alpha = (\alpha_1, \alpha_2)$	$C(\alpha) = e^{\alpha_1 + \alpha_2}$
(0, 0)	1
(0, 1)	e
(1, 0)	e
(1, 1)	e^2

Используя уравнение (5.40), получаем

$$f(x_1, x_2) = 1x_1^0x_2^0 + ex_1^0x_2^1 + ex_1^1x_2^0 + e^2x_1^1x_2^1 = \bar{x}_1\bar{x}_2 + e\bar{x}_1x_2 + ex_1\bar{x}_2 + e^2x_1x_2.$$

При помощи соотношения $\bar{x}_i = 1 - x_i$ можно показать, что эти две формы $f(x_1, x_2)$ эквивалентны.

I. Линейные псевдобулевы уравнения, неравенства и системы

Линейные псевдобулевы уравнения

Пусть $f(x_1, \ldots, x_n)$ — псевдобулева функция от n переменных $x_1, \ldots, x_n$. Псевдобулево уравнение тогда определяется как соотношение $f(x_1, \ldots, x_n) = 0$. Поскольку каждая псевдобулева функция может быть представлена в виде полинома от переменных $x_i, \bar{x}_i$, рассмотрим теперь общее линейное псевдобулево уравнение вида

$$\sum_{i=1}^{n} a_i x_i + b_i \bar{x}_i = k, \tag{5.41}$$

$$k,\ a_i,\ b_i \in \mathrm{Re},\ x_i,\ \bar{x}_i \in B_2, \quad \text{при} \quad i = 1, \ldots, n.$$

Без потери общности можно принять, что $a_i \neq b_i$, так как выражение

$$a_i x_i + b_i \bar{x}_i = (a_i - b_i)\, x_i + b_i$$

сводится к константе, если $a_i = b_i$, и, следовательно, его можно включить в k.

Используя преобразование T, определяемое соотношениями

$$y_i = \begin{cases} x_i, & \text{если } a_i > b_i, \\ \bar{x}_i, & \text{если } a_i < b_i, \end{cases}$$

$$a_i x_i + b_i \bar{x}_i = \begin{cases} (a_i - b_i)\, y_i + b_i, & \text{если } a_i > b_i, \\ (b_i - a_i)\, y_i + a_i, & \text{если } a_i < b_i, \end{cases} \tag{5.42}$$

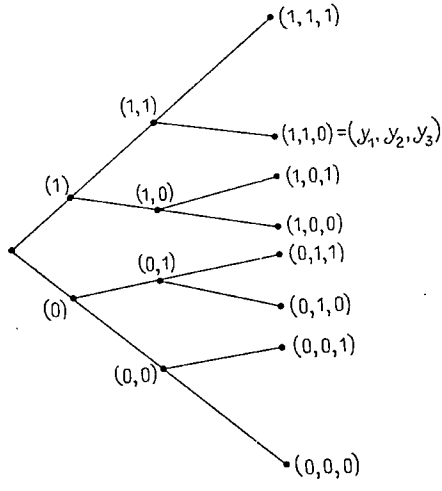
уравнение (5.41) можно переписать в виде

$$\sum_{i=1}^{n} c_i y_i = d, \tag{5.43}$$

$$c_i > 0,\ c_i,\ d_i \in \mathrm{Re},\ y_i \in B_2,\ i = 1, \ldots, n.$$

Изменяя нумерацию c_i, можно добиться того, чтобы $c_1 \geqslant \ldots \geqslant c_n > 0$, и тогда уравнение (5.43) будет называться канонической формой уравнения (5.41).

Вместо того чтобы проверять все 2^n возможных решений (5.43), используем более систематический подход к получению решений при помощи бинарного дерева, изображенного на фиг. 5.8.



Фиг. 5.8.

Существуют восемь взаимно исключающих случаев, соответствующих различным значениям c_i и d, из которых можно сделать выводы о решениях уравнения (5.43).

Теорема 5.7. *Все решения линейного псевдобулева уравнения* (5.43) *распадаются на восемь непересекающихся классов*:

1. *Если* $d < 0$, *то не существует решений.*

2. *Если* $d = 0$, *то существует единственное решение* $y_1 = \ldots = y_n = 0$.

3. *Если* $d > 0$ *и* $c_1 \geqslant \ldots \geqslant c_p > d \geqslant c_{p+1} \geqslant \ldots \geqslant c_n$, *то решениями, если они существуют, являются следующие:* $y_1 = \ldots = y_p = 0$ *и решения уравнения* $\sum_{i=p+1}^{n} c_i y_i = d$.

4. *Если* $d > 0$ *и* $c_1 = \ldots = c_p = d > c_{p+1} \geqslant \ldots \geqslant c_n$, *то существуют* $p+1$ *типов возможных решений*:
1) $y_1 = \ldots = y_{j-1} = y_{j+1} = \ldots = y_n = 0$ *и* $y_j = 1$ *для* $j = 1, \ldots, p$ и 2) $y_1 = \ldots = y_p = 0$ *и решения уравнения* $\sum_{i=p+1}^{n} c_i y_i = d$ (*если они существуют*).

5. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$ *и* $\sum_{i=1}^{n} c_i < d$, *то не существует решений.*

6. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$ *и* $\sum_{i=1}^{n} c_i = d$, *то единственное решение имеет вид* $y_1 = \ldots = y_n = 1$.

7. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$ *и* $\sum_{i=1}^{n} c_i > d$, $\sum_{i=2}^{n} c_i < d$, *то решениями* (*если они существуют*) *являются следующие*:
$y_1 = 1$ *и решения уравнения* $\sum_{j=2}^{n} c_i y_i = d - c_1$.

8. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$ *и* $\sum_{i=1}^{n} c_i > d$, $\sum_{i=2}^{n} c_i > d$, *то решениями являются*: 1) $y_1 = 1$ *и решения уравнения* $\sum_{i=2}^{n} c_i y_i = d - c_1$ (*если они существуют*) и 2) $y_1 = 0$ *и решения равенства* $\sum_{i=2}^{n} c_i y_i = d$ (*если они существуют*).

Теорема 5.7 дает систематический метод получения решений псевдобулева уравнения. Этот метод в виде алгоритма формулируется следующим образом.

Алгоритм решения линейных псевдобулевых уравнений

Шаг 1. Преобразуем линейное псевдобулево уравнение в каноническую форму [уравнение (5.43)] с $c_1 \geqslant \ldots \geqslant c_n > 0$. Переходим к шагу 2.

Шаг 2. Сравнивая каноническое уравнение (5.43) со случаями, рассмотренными в теореме 5.7, можно получить следующие выводы:

а. Если уравнение (5.43) соответствует случаям 1 или 5, то решений нет. Переходим к шагу 3.

б. Если уравнение (5.43) соответствует случаям 2 или 6, то существует единственное решение, задаваемое теоремой 5.7. Переходим к шагу 3.

в. Если уравнение (5.43) такое, как в случаях 3, 4 или 7, то оно заменяется аналогичным уравнением, которое дается в выводах к случаям 3, 4 или 7; исключенным переменным придаются соответствующие значения. Переходим к шагу 3.

г. Если уравнение (5.43) такое, как в случае 8, то оно заменяется одним из двух уравнений, которые содержат на одну переменную меньше и даются в выводах к случаю 8 в зависимости от значения, которое придается исключенной переменной. Переходим к шагу 3.

Шаг 3. Если после шага 2 не появились новые уравнения типа уравнения (5.43) и не осталось уравнений, которые надо решать, то процесс решения закончен и, применяя преобразование T^{-1} к полученному решению, получаем исходное решение. Если после шага 2 появляются новые уравнения или остаются еще не решенные уравнения, повторяем шаг 2 до тех пор, пока все уравнения не будут решены.

Вышеизложенный алгоритм позволяет получить все решения линейного псевдобулева уравнения, так как он исчерпывает все возможности.

Пример 5.4. Требуется найти решения линейного псевдобулева уравнения

$$6x_1 + 3\bar{x}_1 + 5x_2 - 4\bar{x}_3 + 2x_4 + \bar{x}_5 - x_6 = 7.$$

Решение. Применяя преобразование T и изменяя пумерацию переменных, находим уравнение

$$5y_1 + 4y_2 + 3y_3 + 2y_4 + y_5 + y_6 = 9 \quad \text{(случай 8)}, \tag{5.44}$$

где

$$y_1 = x_2,\ y_2 = x_3,\ y_3 = x_1,\ y_4 = x_4,\ y_5 = \bar{x}_5,\ y_6 = \bar{x}_6.$$

Так как уравнение (5.44) такое, как в случае 8 теоремы 5.7, получаются следующие два частичных решения и новые уравнения:

Частичное решение $(y_1, y_2, \ldots, y_i)$

$y_1 = 1$ и $4y_2 + 3y_3 + 2y_4 + y_5 + y_6 = 4$ (случай 4), (5.44а) (1)

$y_1 = 0$ и $4y_2 + 3y_3 + 2y_4 + y_5 + y_6 = 9$ (случай 7). (5.44б) (0)

Уравнение (5.44а) соответствует случаю 4 теоремы 5.7, и получаются следующие частичные решения и уравнения:

$y_2 = 1$ и $y_3 = y_4 = y_5 = y_6 = 0$, (1, 1, 0, 0, 0, 0)

$y_2 = 0$ и $3y_3 + 2y_4 + y_5 + y_6 = 4$. (5.44в) (1, 0)

Уравнение (5.44в) соответствует случаю 8 теоремы 5.7, и получаются следующие частичные решения и новые уравнения:

$y_3 = 1$ и $2y_4 + y_5 + y_6 = 1$ (случай 3), (5.44г) (1, 0, 1)

$y_3 = 0$ и $2y_4 + y_5 + y_6 = 4$ (случай 6). (5.44д) (1, 0, 0)

Уравнение (5.44г) соответствует случаю 3, и получаются следующие уравнение и частичное решение:

$y_4 = 0$ и $y_5 + y_6 = 1$ (случай 4). (5.44е) (1, 0, 1, 0)

Уравнение (5.44е) соответствует случаю 4, и получаются следующие решения:

$y_5 = 0,\ y_6 = 1$, (1, 0, 1, 0, 0, 1)

$y_5 = 1,\ y_6 = 0$. (1, 0, 1, 0, 1, 0)

Уравнение (5.44д) соответствует случаю 6, и получается решение

$$y_4 = y_5 = y_6 = 1. \qquad (1, 0, 0, 1, 1, 1)$$

Уравнение (5.44б) соответствует случаю 7, и получаются следующие уравнение и ча-

стичное решение:

$y_2 = 1$ и $3y_3 + 2y_4 + y_5 + y_6 = 5$ (случай 7). (5.44ж) (0, 1)

Уравнение (5.44ж) соответствует случаю 7, и получаются следующие уравнение и частичное решение:

$y_3 = 1$ и $2y_4 + y_5 + y_6 = 2$ (случай 4). (5.44з) (0, 1, 1)

Уравнение (5.44з) соответствует случаю 4, и получаются следующее решение, а также новое уравнение и частичное решение:

$y_4 = 1$ и $y_5 = y_6 = 0$, (0, 1, 1, 1, 0, 0)

$y_4 = 0$ и $y_5 + y_6 = 2$ (случай 6). (5.44и) (0, 1, 1, 0)

Уравнение (5.44и) соответствует случаю 6 и имеет решение

$y_5 = y_6 = 1$. (0, 1, 1, 0, 1, 1)

Решения уравнения (5.44)

y_1	y_2	y_3	y_4	y_5	y_6
1	1	0	0	0	0
1	0	1	0	0	1
1	0	1	0	1	0
1	0	0	1	1	1
0	1	1	1	0	0
0	1	1	0	1	1

Решения исходной задачи

x_1	x_2	x_3	x_4	x_5	x_6
0	1	1	0	1	1
1	1	0	0	1	0
1	1	0	0	0	1
0	1	0	1	0	0
1	0	1	1	1	1
1	0	1	0	0	0

Упражнение 5.11. Найдите псевдобулеву функцию от четырех переменных x_1, x_2, x_3, x_4, определенную следующими таблицами:

x_1	x_2	x_3	x_4	$f(x_1, x_2, x_3, x_4)$
0	0	0	0	—5
0	0	0	1	0
0	0	1	0	7
0	0	1	1	4
0	1	0	0	2
0	1	0	1	—11
0	1	1	0	1
0	1	1	1	0

x_1	x_2	x_3	x_4	$f(x_1, x_2, x_3, x_4)$
1	0	0	0	3
1	0	0	1	0
1	0	1	0	0
1	0	1	1	—3
1	1	0	0	8
1	1	0	1	5
1	1	1	0	0
1	1	1	1	—4

Указание. Воспользуйтесь теоремой 5.6.

Упражнение 5.12. Найдите решения псевдобулева уравнения

$$4x_1 + \bar{x}_1 - 3x_2 + \bar{x}_2 + 5x_3 - 2x_4 + 5x_5 + 2x_6 - x_7 = 7.$$

Линейные псевдобулевы неравенства

Рассмотрим вместо уравнения $f(x_1, \ldots, x_n) = 0$ общие линейные псевдобулевы неравенства

$$\sum_{i=1}^{n} a_i x_i + b_i \bar{x}_i \geqslant k \tag{5.45}$$

и

$$a_i,\ b_i,\ k \in \mathrm{Re},\ x_i,\ \bar{x}_i \in B_2, \qquad i = 1, \ldots, n,$$

$$\sum_{i=1}^{n} a_i x_i + b_i \bar{x}_i > k, \qquad a_i \neq b_i. \tag{5.46}$$

Если a_i, b_i и k — целые числа, неравенство (5.46) сводится к неравенству (5.45), где k заменяется на $k - 1$.

Неравенства (5.45) и (5.46) можно переписать в канонической форме при помощи преобразования T, определяемого соотношением (5.42). Это дает

$$\sum_{i=1}^{n} c_i y_i \geqslant d, \quad c_1 \geqslant \ldots \geqslant c_n > 0,\ c_i,\ d \in \mathrm{Re},\ y_i \in B_2, \tag{5.47}$$

$$\sum_{i=1}^{n} c_i y_i > d, \quad c_1 \geqslant \ldots \geqslant c_n > 0, \quad c_i,\ d \in \mathrm{Re},\ y_i \in B_2. \tag{5.48}$$

Рассмотрение всех решений (5.47) и (5.48) показывает, что эти решения могут быть сгруппированы в семейства, которые характеризуются тем, что каждое семейство образуется путем фиксирования некоторого количества переменных, причем остальные переменные принимают произвольные значения 0 или 1.

Определение. Пусть $\alpha^* \in B_{2^n}$ удовлетворяет уравнению (5.47) или (5.48), а K означает подмножество во множестве целых чисел $\{1, 2, \ldots, n\}$. Определим

$$\Omega(\alpha^*, K) = \{\beta = (\beta_1, \ldots, \beta_n) \mid \beta \in B_{2^n},\ \beta \text{ — решение (5.47) или (5.48)}\}$$

и

$$\beta_i = \begin{cases} \alpha_i^*, & \text{если } i \in K, \\ \text{Произвольное значение}, & \text{если } i \notin K. \end{cases} \tag{5.49}$$

Будем называть множество $\Omega(\alpha^*, K)$ семейством решений неравенства (5.47) или (5.48), порожденных α^*, а α^* — генератором $\Omega(\alpha^*, K)$ с множеством индексов K; число элементов в K будем называть индексом $\Omega(\alpha^*, K)$.

Определение. Решение $\alpha^* = (\alpha_1^*, \ldots, \alpha_n^*)$ неравенства (5.47) или (5.48) называется *базисным решением* неравенства (5.47) или (5.48) тогда и только тогда, когда для каждого i, такого, что $x = 1$, вектор

$$\alpha' = (\alpha_1^*, \ldots, \alpha_{i-1}^*, 0, \alpha_{i+1}^*, \ldots, \alpha_n^*)$$

не является решением неравенства (5.47) или (5.48).

Используя определение базисного решения, можно легко доказать следующую теорему о некоторых свойствах базисного решения:

Теорема 5.8 (свойства базисных решений).

а. *Если* $y = (y_1^*, \ldots, y_n^*)$ — *базисное решение неравенства* (5.47) [или (5.48)], *то* $(y_{p+1}^*, \ldots, y_n^*)$, $1 \leqslant p \leqslant n$, *является базисным решением неравенства*

$$\sum_{i=p+1}^{n} c_i y_i \geqslant d - \sum_{i=1}^{p} c_i y_i \quad \left(> d - \sum_{i=1}^{p} c_i y_i\right). \tag{5.50}$$

б. *Если* $(y_{p+1}^*, \ldots, y_n^*)$ — *базисное решение неравенства*

$$\sum_{i=p+1}^{n} c_i y_i \geqslant d \quad (> d), \tag{5.51}$$

то вектор $(0, \ldots, 0, y_{p+1}^*, \ldots, y_n^*)$ *является базисным решением неравенства* (5.47) [*или* (5.48)].

в. *Если* $(y_2^*, \ldots, y_n^*)$ — *базисное решение неравенства*

$$\sum_{i=2}^{n} c_i y_i \geqslant d - c_1 \quad (> d - c_1), \tag{5.52}$$

то вектор $(1, y_2^*, \ldots, y_n^*)$ *является базисным решением уравнения* (5.47) [или (5.48)].

Используя теорему (5.8) и рассматривая неравенства (5.47) и (5.48), получаем следующую теорему, аналогичную теореме 5.7:

Теорема 5.9 [**базисные решения неравенства** (5.47)]. *Любое из базисных решений псевдобулева неравенства* (5.47) *относится к одному из шести непересекающихся классов.*

1. *Если* $d \leqslant 0$, *существует единственное базисное решение* $y_1 = \ldots = y_n = 0$.

2. *Если* $d > 0$ *и* $c_1 \geqslant \ldots \geqslant c_p \geqslant d > c_{p+1} \geqslant \ldots \geqslant c_n$, *то решения, если они существуют, имеют вид:*

а. $y_k = 1$ *для некоторого целого* $k \leqslant p$ *и* $y_1 = \ldots = y_{k-1} = y_{k+1} = \ldots = y_n = 0$. (*Для каждого* $k \leqslant p$ *решения различны.*)

б. $y_1 = \ldots = y_p = 0$ *и* $(y_{p+1}, \ldots, y_n)$ — *базисное решение неравенства*

$$\sum_{i=p+1}^{n} c_i y_i \geqslant d.$$

3. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$ *и* $\sum_{i=1}^{n} c_i < d$, *то решения не существует.*

4. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$ *и* $\sum_{i=1}^{n} c_i = d$, *то единственное базисное решение имеет вид* $y_1 = \ldots = y_n = 1$.

5. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$, $\sum_{i=1}^{n} c_i > d$ *и* $\sum_{i=2}^{n} c_i < d$, *то базисными решениями, если хотя бы одно существует, являются* $y_1 = 1$ *и остальные* $n - 1$ *компонент, которые представляют собой базисные решения неравенства* $\sum_{i=2}^{n} c_i y_i \geqslant d - c_1$.

6. *Если* $d > 0$ *и* $c_i < d$ *при* $i = 1, \ldots, n$, $\sum_{i=1}^{n} c_i > d$ *и* $\sum_{i=2}^{n} c_i y_i \geqslant d$, *то базисными решениями, если они существуют, являются следующие*:

а. $y_1 = 1$ *и* $(y_2, \ldots, y_n)$ — *базисное решение* $\sum_{i=2}^{n} c_i y_i \geqslant d - c_1$.

б. $y_1 = 0$ *и* $(y_2, \ldots, y_n)$ — *базисное решение* $\sum_{i=2}^{n} c_i y_i \geqslant d$.

Для неравенств типа неравенства (5.48) можно получить теорему, подобную теореме 5.9, которая позволяет разделить все базисные решения неравенства (5.48) на пять непересекающихся классов. Если неравенство имеет целочисленные коэффициенты, то, как указывалось выше, можно преобразовать строгое неравенство к виду (5.47).

Используя теорему 5.9, можно сформулировать следующий алгоритм, аналогичный алгоритму решения линейных псевдобулевых неравенств.

Алгоритм отыскания базисных решений линейных псевдобулевых неравенств

Шаг 1. Применяем преобразование T, определенное соотношением (5.42), к неравенству, чтобы получить каноническую форму [неравенство (5.47)]. Переходим к шагу 2.

Шаг 2. Сравниваем каноническую форму неравенства со случаями, указанными в теореме 5.9.

а. Если неравенство (5.47) соответствует случаю 3 теоремы 5.9, то не существует решений. Переходим к шагу 3.

б. Если неравенство (5.47) соответствует случаю 1 или 4 теоремы 5.9, то существует единственное базисное решение, определяемое этой теоремой.

в. Если неравенство (5.47) соответствует случаю 2 или 5 теоремы 5.9, то оно заменяется на аналогичное неравенство с меньшим числом переменных, приводимое в соответствующем случае вместе

со значениями исключенных переменных. Если неравенство (5.47) соответствует случаю 2, то получается множество базисных решений, приводимых в соответствующем случае. Переходим к шагу 3.

г. Если неравенство (5.47) или (5.48) соответствует случаю 6 теоремы 5.9, то оно заменяется на два аналогичных неравенства, в которых число переменных на единицу меньше; эти неравенства приводятся в соответствующем случае вместе со значением исключенной переменной. Переходим к шагу 3.

Шаг 3. Если в результате шага 2 не появляются новые неравенства и все имевшиеся неравенства решены, то применяем преобразование T^{-1} к базисным решениям, полученным на шаге 2, если хотя бы одно существует, чтобы найти решения исходных неравенств. Если на шаге 2 возникают новые неравенства, то повторяем шаг 2 до тех пор, пока все неравенства не будут решены. Если какое-либо неравенство не имеет решения, то соответствующее частичное решение и это неравенство следует исключить.

Описанный алгоритм определяет все базисные решения неравенств (5.45), так как он исчерпывает все возможные решения.

После определения базисных решений неравенства (5.47) или (5.48) множество всех решений (5.47) [или (5.48)] строится следующим образом:

Теорема 5.10. *Множество всех базисных решений* α^j, $j = 1, \ldots, m$, *неравенства* (5.47) [*или* (5.48)] *порождает класс непересекающихся семейств* $\Omega(\alpha^j, K_j)$, $j = 1, \ldots, m$, *и каждое решение неравенства* (5.47) [*или* (5.48)] *принадлежит к одному из этих семейств.*

Доказательство. Определим

$$K_j = \left\{ p \mid 0 < p \leqslant q, \text{ где } q \text{ — номер последней ненулевой компоненты } \alpha^j. \right\}$$

Если $x = (x_1, \ldots, x_n) \in \Omega(\alpha^j, K_j)$ при некотором j, то x удовлетворяет неравенству (5.47) [или (5.48)], так как $\sum_{i=1}^{r} c_i\alpha_i \geqslant d$ $(>d)$, где r — номер семейства $\Omega(\alpha^j, K_j)$, к которому принадлежит x.

Поскольку $c_i \geqslant 0$, $i = 1, \ldots, n$, получаем

$$\sum_{i=1}^{n} c_i\alpha_i = \sum_{i=1}^{r} c_i\alpha_i + \sum_{i=r+1}^{n} c_i\alpha_i \geqslant d \quad (>d).$$

Вследствие выбранного способа построения базисных решений α^j, $j = 1, \ldots, m$, семейства $\Omega(\alpha^j, K_j)$ должны быть различными. Каждое решение должно принадлежать к одному из вышеуказанных семейств, поскольку либо оно является базисным решением и, следовательно, порождает некоторое семейство $\Omega(\alpha^j, K_j)$ (так как все базисные решения определяются алгоритмом), либо оно принадлежит некоторому семейству $\Omega(\alpha^j, K_j)$, что можно показать путем изме-

нения последних ненулевых компонент вектора $x = (x_1, \ldots, x_n)$ до тех пор, пока он не станет базисным решением.

Пример 5.5. Требуется найти решения псевдобулева неравенства

$$2\bar{x}_1 - 8x_2 + 7\bar{x}_2 + 5x_3 - 6\bar{x}_4 + 8x_5 \geqslant 8. \qquad (5.53)$$

Применяя преобразование T к неравенству (5.53) и изменяя нумерацию, получаем

$$15y_1 + 8y_2 + 6y_3 + 5y_4 + 2y_5 \geqslant 22 \text{ (случай 5 теоремы 5.9)}, \qquad (5.54)$$

где

$$y_1 = \bar{x}_2,\ y_2 = x_5,\ y_3 = x_4,\ y_4 = x_3,\ y_5 = \bar{x}_1.$$

Неравенство (5.54) соответствует случаю 5 теоремы 5.9, поэтому получаются следующие частичное решение и новое неравенство:

Частичное базисное решение

$$8y_2 + 6y_3 + 5y_4 + 2y_5 \geqslant 7 \text{ (случай 2)}$$
$$\text{и } y_1 = 1. \qquad (5.54а) \qquad (1)$$

Неравенство (5.54а) соответствует случаю 2 теоремы 5.9, поэтому получаем следующее базисное решение, а также новое неравенство и частичное решение:

$$y_2 = 1,\ y_3 = y_4 = y_5 = 0, \qquad \underline{(1, 1, 0, 0, 0)}$$
$$6y_3 + 5y_4 + 2y_5 \geqslant 7 \text{ (случай 6) и } y_2 = 0. \qquad \underline{(1, 0)} \qquad (5.54б)$$

Неравенство (5.54б) соответствует случаю 6 теоремы 5.9, поэтому получаются следующие два неравенства и частичные решения:

$$5y_4 + 2y_5 \geqslant 1 \text{ (случай 2) и } y_3 = 1, \qquad (5.54в) \qquad (1, 0, 1)$$
$$5y_4 + 2y_5 \geqslant 7 \text{ (случай 4) и } y_3 = 0. \qquad (5.54г) \qquad (1, 0, 0)$$

Неравенство (5.54в) соответствует случаю 2 теоремы 5.9, в результате получаются следующие базисные решения:

$$y_4 = 1,\quad y_5 = 0, \qquad \underline{(1, 0, 1, 1, 0)}$$
$$y_4 = 0,\quad y_5 = 1. \qquad \underline{(1, 0, 1, 0, 1)}$$

Неравенство (5.54г) соответствует случаю 4 теоремы 5.9, поэтому получается базисное решение $y_4 = y_5 = 1$. $\underline{(1, 0, 0, 1, 1)}$

Базисные решения имеют вид $\alpha^1 = (1, 1, 0, 0, 0)$, $\alpha^2 = (1, 0, 1, 1, 0)$, $\alpha^3 = (1, 0, 1, 0, 1)$ и $\alpha^4 = (1, 0, 0, 1, 1)$. Используя эти базисные решения, можно построить множества индексов K_i: $K_1 = \{1, 2\}$, $K_2 = \{1, 2, 3, 4\}$, $K_3 = \{1, 2, 3, 4, 5\}$ и $K_4 = \{1, 2, 3, 4, 5\}$. Используя

значения K_i и α_i, строим следующие четыре семейства решений неравенства (5.54):

$$\Omega(\alpha^1, K_1) = \{(1, 1, 0, 0, 0), (1, 1, 1, 0, 0), (1, 1, 0, 1, 0),$$
$$(1, 1, 1, 1, 0), (1, 1, 0, 0, 1), (1, 1, 1, 0, 1), (1, 1, 0, 1, 1), (1, 1, 1, 1, 1)\};$$
$$\Omega(\alpha^2, K_2) = \{(1, 0, 1, 1, 0), (1, 0, 1, 1, 1)\};$$
$$\Omega(\alpha^3, K_3) = \{(1, 0, 1, 0, 1)\};$$
$$\Omega(\alpha^4, K_4) = \{(1, 0, 0, 1, 1)\}.$$

После применения преобразования T^{-1} получаем решение неравенства (5.53), приведенное в следующих таблицах, где на местах прочерка может стоять 0 или 1:

y_1	y_2	y_3	y_4	y_5
1	1	—	—	—
1	0	1	1	—
1	0	1	0	1
1	0	0	1	1

x_1	x_2	x_3	x_4	x_5
—	0	—	—	1
—	0	1	1	0
0	0	0	1	0
0	0	1	0	0

Упражнение 5.13. Используя теорему 5.9, решите строгое псевдобулево неравенство

$$2\bar{x}_1 - 5x_2 + 3x_3 + 4\bar{x}_4 - 7x_5 + 16x_6 - x_7 < 4.$$

Упражнение 5.14. При условии, что величины c_i и d в (5.47) — целые числа, постройте семейства решений строгого неравенства (5.48), зная базисные решения неравенства (5.47).

Указание. Воспользуйтесь решениями $\sum_{i=1}^{n} c_i y_i = d$ для того, чтобы модифицировать семейства решений неравенства (5.48). Сделайте упражнение (5.13), используя указанные выше результаты.

Системы линейных псевдобулевых уравнений и (или) неравенств

Здесь будет дано краткое изложение методов, используемых для решения псевдобулевых уравнений и (или) неравенств; более подробное изложение можно найти в [22].

Методы решения псевдобулевых уравнений и неравенств уже разработаны. Очевидно, что решения системы таких уравнений и неравенств должны удовлетворять каждому из этих уравнений и неравенств. Если каждое семейство F решений каждого уравнения или неравенства исследуется отдельно, то решение J системы

получается в виде пересечения всех семейств F_i, т. е.

$$J = \bigcap_{i=1}^{m} F_i, \tag{5.55}$$

где m — число уравнений или неравенств в системе, а F_i — семейство решений i-го уравнения или неравенства.

Вместо того чтобы вычислять семейство решений каждого члена системы, можно использовать более систематический подход. Путем правильного выбора одного из членов системы можно решить систему относительно одной или нескольких переменных и свести исходную систему к одной или нескольким новым системам с меньшим числом переменных. Последовательно применяя этот метод, можно получить все решения исходной системы. Подробно этот метод описан в [22, стр. 37—52].

II. Нелинейные псевдобулевы уравнения и неравенства

Методы получения решений линейных псевдобулевых уравнений, неравенств и систем уравнений и (или) неравенств, изложенные выше, здесь будут использованы для решения нелинейных задач.

Характеристическая функция псевдобулевых уравнений, неравенств и систем

Каждому псевдобулеву уравнению, неравенству или системе $G(x_1, \ldots, x_n) \geqslant 0$ соответствует булево уравнение, которое имеет те же решения, что и $G \geqslant 0$, а вид его задается следующей теоремой:

Теорема 5.11. *Каждому псевдобулеву уравнению, неравенству или каждой системе* $G(x_1, \ldots, x_n) \geqslant 0$ *соответствует булева функция* ψ, *которая называется характеристической функцией* $G(x_1, \ldots, x_n)$; ψ: $B_{2^n} \to B_2$, *такая, что решения* $G(x_1, \ldots, x_n) \geqslant 0$ *являются решением булева уравнения*

$$\psi(x_1, \ldots, x_n) = 1. \tag{5.56}$$

Доказательство. Определим $\psi(x_1, \ldots, x_n)$ следующим образом:

$$\psi(x_1, \ldots, x_n) = \bigcup_{F \in \mathscr{F}} \chi_F \tag{5.56а}$$

(объединение берется по всем $F \in \mathscr{F}$). Здесь

$$\chi_F = \bigcup_{\alpha \in F} (x_1^{\alpha_1} \cap x_2^{\alpha_2} \cap \ldots \cap x_n^{\alpha_n}) = \bigcup_{\alpha \in F} x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n},$$

$\mathscr{F}$ — класс всех семейств F решений $G(x_1, \ldots, x_n) \geqslant 0$,

$$x_i^{\alpha_i} = \begin{cases} x_i, & \text{если } \alpha_i = 1, \\ \bar{x}_i, & \text{если } \alpha_i = 0, \end{cases}$$

где α_i — i-я компонента решения $\alpha \in F$ неравенства $G(x_1, \ldots, x_n) \geqslant 0$. Если α — решение $G(x_1, \ldots, x_n) \geqslant 0$, то по определе-

нию ψ оно является решением уравнения (5.56). Обратно, если α — решение уравнения (5.56), то по определению ψ существует по крайней мере одна функция χ_F такая, что $\chi_F = 1$. Так как χ_F является по определению объединением минимальных булевых полиномов (см. [5a]), α должно быть решением хотя бы одного из членов объединения и, следовательно, по определению χ_F является решением $G(x_1, \ldots, x) \geqslant 0$.

Характеристическую функцию $\psi(x_1, \ldots, x_n)$ можно переписать в другой форме, зависящей от вида $G(x_1, \ldots, x_n) \geqslant 0$. Ниже даются примеры представления $\psi(x_1, \ldots, x_n)$ в другой форме.

ψ для линейных псевдобулевых неравенств $G(x_1, \ldots, x_n) \geqslant 0$.

Для

$$\psi(x_1, \ldots, x_n) = \bigcup_{\alpha^p \in Q} x_{p_1}^{\alpha_{p_1}} \ldots x_{p_m}^{\alpha_{p_m}} \tag{5.57}$$

$p_i \in K_p$, которое является множеством индексов $\Omega(\alpha^p, K_p)$, а Q — множество всех генераторов решений $G(x_1, \ldots, x_n) \geqslant 0$ и $\alpha^p \in Q$ — генератор. Заметим, что только элементы, содержащие генераторы решений $G(x_1, \ldots, x_n) \geqslant 0$, определяют характеристическую функцию.

ψ для линейных псевдобулевых систем неравенств и уравнений. Пусть $G(x_1, \ldots, x_n) \geqslant 0$ определяется как следующая система линейных псевдобулевых уравнений и неравенств:

$$g_j(x_1, \ldots, x_n) = 0, \quad j = 1, \ldots, m, \tag{5.58а}$$
$$g_j(x_1, \ldots, x_n) \geqslant 0, \quad j = m+1, \ldots, p, \tag{5.58б}$$
$$g_j(x_1, \ldots, x_n) > 0, \quad j = p+1, \ldots, q. \tag{5.58в}$$

Если $\psi_j(x_1, \ldots, x_n)$ — характеристическая функция $g_j(x_1, \ldots, x_n)$, $j = 1, \ldots, q$, то получаем характеристическую функцию, связанную с системой $G(x_1, \ldots, x_n) \geqslant 0$:

$$\psi(x_1, \ldots, x_n) = \bigcap_{j=1}^{q} \psi_j(x_1, \ldots, x_n). \tag{5.59}$$

Упражнение 5.15. Проверьте уравнения (5.57), (5.58а), (5.58б) и (5.58в), используя определение характеристической функции.

Примеры характеристических функций. Характеристическая функция в примере 5.4 имеет вид

$$\psi(x_1, \ldots, x_6) = \bar{x}_1 x_2 x_3 \bar{x}_4 x_5 x_6 \cup x_1 x_2 \bar{x}_3 \bar{x}_4 x_5 \bar{x}_6 \cup x_1 x_2 \bar{x}_3 \bar{x}_4 \bar{x}_5 x_6 \cup$$
$$\cup \bar{x}_1 x_2 \bar{x}_3 x_4 \bar{x}_5 \bar{x}_6 \cup x_1 \bar{x}_2 x_3 x_4 x_5 x_6 \cup x_1 \bar{x}_2 x_3 \bar{x}_4 \bar{x}_5 \bar{x}_6.$$

Характеристическая функция в примере 5.5 имеет вид

$$\psi(x_1, \ldots, x_5) = \bar{x}_2 x_5 \cup \bar{x}_2 x_3 x_4 \bar{x}_5 \cup \bar{x}_1 \bar{x}_2 \bar{x}_3 x_4 \bar{x}_5 \cup \bar{x}_1 \bar{x}_2 x_3 \bar{x}_4 \bar{x}_5.$$

Решения нелинейных псевдобулевых уравнений, неравенств и систем уравнений и (или) неравенств

Используя результаты теоремы 5.6, можно записать общее псевдобулево уравнение с n переменными $x_1, \ldots, x_n$ в виде

$$a_1 P_1(x_1, \ldots, x_n) + \ldots + a_m P_m(x_1, \ldots, x_n) = b, \quad (5.60)$$
$$a_i b \in \mathrm{Re}(x_1, \ldots, x_n) \in B_2,$$

где

$$P_i(x_1, \ldots, x_n) = x_{i_1}^{\beta_{i_1}} \ldots x_{i_{k(i)}}^{\beta_{i_{k(i)}}},$$

$$x_i^{\beta_i} = \begin{cases} x_i, & \text{если } \beta_i = 1, \\ \bar{x}_i, & \text{если } \beta_i = 0. \end{cases}$$

В общем случае $P_i(x_1, \ldots, x_n)$ не зависит от всех переменных $x_1, \ldots, x_n$, и это видно из определения $P_i(x_1, \ldots, x_n)$. Полагая

$y_j = P_j(x_1, \ldots, x_n)$, $j = 1, \ldots, n$, перепишем (5.60) в виде

$$a_1 y_1 + \ldots + a_m y_m = b. \quad (5.61)$$

Так как $P_j(x_1, \ldots, x_n) \in B_2$, все y_j можно рассматривать как новые бивалентные переменные, а уравнение (5.61) можно рассматривать как линейное псевдобулево уравнение относительно $y_1, \ldots, y_m$. Используя методы части I, можно найти все решения уравнения (5.61).

Чтобы получить решения уравнения (5.60), зная решения уравнения (5.61), рассмотрим характеристическое уравнение, связанное с уравнением (5.61):

$$\psi(y_1, \ldots, y_m) = 1. \quad (5.62)$$

Заменяя в уравнении (5.62) y_j на $P_j(x_1, \ldots, x_n)$, $j = 1, \ldots, m$, получаем новое характеристическое уравнение, которое имеет такие же решения, что и уравнение (5.60):

$$\psi[P_1(x_1, \ldots, x_n), \ldots, P_m(x_1, \ldots, x_n)] = 1. \quad (5.63)$$

Решая булево уравнение (5.63), находим все решения уравнения (5.60); если некоторые из переменных $x_1, \ldots, x_n$ не входят в решение, им можно придавать произвольные значения.

В случае нелинейных неравенств и систем применяются те же методы линеаризации, как и в случае нелинейного уравнения, а для решения линейных неравенств или систем записывается соответствующее характеристическое уравнение, с помощью которого получаются решения нелинейной задачи.

Решение характеристического уравнения

На основании законов булевой алгебры любое характеристическое уравнение можно свести к виду

$$\psi(x_1, \ldots, x_n) = Q_1 \cup \ldots \cup Q_m, \tag{5.64}$$

где Q_i имеет вид

$$Q_i = x_{i_1}^{\beta_{i_1}} \ldots x_{i_{h(i)}}^{\beta_{i_{h(i)}}}, \quad i = 1, \ldots, m,$$

$$x_i^{\beta_i} = \begin{cases} x_i, & \text{если } \beta_i = 1, \\ \bar{x}_i, & \text{если } \beta_i = 0. \end{cases}$$

После приведения $\psi(x_1, \ldots, x_n)$ к виду (5.64) решения характеристического уравнения (5.56) можно получить, полагая каждое $Q_i = 1$ и отыскивая решения путем проверки. Эта процедура позволяет получить все решения нелинейных псевдобулевых уравнений, но для нелинейных псевдобулевых неравенств и систем она дает только базисные решения. Множество всех решений нелинейных псевдобулевых неравенств можно построить следующим образом.

Для каждого решения α^i уравнения (5.64) строим $\Omega(\alpha^j, K_i)$, где K_i — множество индексов, соответствующих компонентам α^i. Очевидно, что каждое семейство $\Omega(\alpha^i, K_i)$ является семейством решений, и если вектор

$$\hat{\alpha} = (\hat{\alpha}_1, \ldots, \hat{\alpha}_n)$$

представляет собой решение нелинейной задачи, то он должен принадлежать по крайней мере одному семейству $\Omega(\alpha, K)$; если это не так, то по определению $\psi(x_1, \ldots, x_n)$ должно иметь место

$$Q_1 = \ldots = Q_m = 0,$$

откуда следует, что $\psi(\hat{\alpha}_1, \ldots, \hat{\alpha}_n) = 0$, но это противоречит уравнению (5.56).

Семейства $\Omega(\alpha^i, K_i)$ могут оказаться одинаковыми; совпадающие семейства исключаются следующим образом.

Прежде чем решать уравнение (5.64), проверяем Q_i, $i = 1, \ldots, m$, следующим образом. Если $Q_i \cap Q_j = \phi$, $i \neq j$, оставляем Q_i. Если $Q_i \cap Q_j \neq \phi$, $i \neq j$, исключаем Q_i из уравнения (5.64).

Указанная процедура дает (при использовании обозначений $R_i = Q_i$ или $R_i = \phi$, как объяснено выше) равенство

$$\psi(x_1, \ldots, x_n) = R_1 \cup \ldots \cup R_m, \tag{5.65}$$

которое имеет различные решения; следовательно, семейства $\Omega(\alpha^j, K_j)$ различны.

Пример 5.6. Найдем решения нелинейного псевдобулева неравенства

$$-10x_1x_2x_3 + 6x_1x_2 - 5x_1x_3 + 7x_2x_3 - 2x_1 - 2x_2 - 2x_3 \leqslant 2. \tag{5.66}$$

Положим $y_1 = x_1x_2x_3$, $y_2 = \bar{x}_2\bar{x}_3$, $y_3 = \bar{x}_1\bar{x}_2$, $y_4 = x_2x_3$, $y_5 = x_1$, $y_6 = x_2$, $y_7 = x_3$. Тогда неравенство (5.66) примет вид

$$10y_1 + 7y_2 + 6y_3 + 5y_4 + 2y_5 + 2y_6 + 2y_7 \geqslant 11. \qquad (5.67)$$

Неравенство (5.67) имеет следующие решения:

y_1	y_2	y_3	y_4	y_5	y_6	y_7
1	1	—	—	—	—	—
1	0	1	—	—	—	—
1	0	0	1	—	—	—
1	0	0	0	1	—	—
1	0	0	0	0	1	—
1	0	0	0	0	0	1
0	1	1	—	—	—	—

y_1	y_2	y_3	y_4	y_5	y_6	y_7
0	1	0	1	—	—	—
0	1	0	0	1	1	—
0	1	0	0	1	0	1
0	1	0	0	0	1	1
0	0	0	1	1	1	1
0	0	1	1	—	—	—
0	0	1	0	1	1	1

Здесь на местах прочерка может стоять 0 или 1. Тогда характеристическая функция, связанная с неравенством (5.67), имеет вид

$$\psi(y_1, \ldots, y_7) = y_1y_2 \cup y_1\bar{y}_2y_3 \cup y_1\bar{y}_2\bar{y}_3y_4 \cup y_1\bar{y}_2\bar{y}_3\bar{y}_4y_5 \cup$$
$$\cup y_1\bar{y}_2\bar{y}_3\bar{y}_4\bar{y}_5y_6 \cup y_1\bar{y}_2\bar{y}_3\bar{y}_4\bar{y}_5\bar{y}_6y_7 \cup \bar{y}_1y_2y_3 \cup \bar{y}_1y_2\bar{y}_3y_4 \cup$$
$$\cup \bar{y}_1y_2\bar{y}_3\bar{y}_4y_5y_6 \cup \bar{y}_1y_2\bar{y}_3\bar{y}_4y_5\bar{y}_6y_7 \cup \bar{y}_1y_2\bar{y}_3\bar{y}_4\bar{y}_5y_6y_7 \cup \bar{y}_1\bar{y}_2\bar{y}_3y_4y_5y_6y_7 \cup$$
$$\cup \bar{y}_1\bar{y}_2y_3y_4 \cup \bar{y}_1y_2y_3\bar{y}_4\bar{y}_5y_6y_7.$$

Подставляя вместо $y_1, \ldots, y_7$ их выражения через x_1, x_2, x_3, получаем после приведения

$$\psi(x_1, x_2, x_3) = x_1 \cup \bar{x}_1\bar{x}_2 \cup \bar{x}_1x_2\bar{x}_3. \qquad (5.68)$$

Решая равенство $\psi(x_1, x_2, x_3) = 1$, находим решения

$$(x_1, x_2, x_3) = (1, —, —), (0, 0, —), (0, 1, 0).$$

Упражнение 5.16. Решите псевдобулево неравенство

$$7x_1x_2x_3 + 5x_2x_4x_6x_7x_8 - 3x_3x_8 - 2\bar{x}_1x_4x_8 - x_4\bar{x}_5x_6 \leqslant 3.$$

III. Минимизация псевдобулевых функций

Задачи минимизации без ограничений

По определению вектор $x^* = (x_1^*, \ldots, x_n^*)$ является точкой минимизации псевдобулевой функции $f(x_1, \ldots, x_n)$ тогда и только тогда, когда

$$f(x_1^*, \ldots, x_n^*) \leqslant f(x_1, \ldots, x_n) \quad \text{при всех } x \in B_{2^n}. \qquad (5.69)$$

Используя определение точки минимизации, получаем следующее необходимое условие того, что вектор x^* является точкой минимизации:

$$f(x_1^*, \ldots, x_n^*) \leqslant f(x_1^*, \ldots, x_{i-1}^*, \overline{x_i^*}, x_{i+1}^*, \ldots, x_n^*), \quad i = 1, \ldots, n. \tag{5.70}$$

При помощи теоремы 5.5 соотношение (5.70) может быть переписано в виде

$$(x_i^* - \overline{x_i^*})\, g_i(x_1^*, \ldots, x_{i-1}^*, x_{i+1}^*, \ldots, x_n^*) \leqslant 0, \quad i = 1, \ldots, n, \tag{5.71}$$

где

$$f(x_1^*, \ldots, x_n^*) =$$
$$= x_i^* g_i(x_1^*, \ldots, x_{i-1}^*, x_{i-1}^*, \ldots, x_n^*) + h_i(x_1^*, \ldots, x_{i-1}^*, x_{i+1}^*, \ldots, x_n^*).$$

Упражнение 5.17. Докажите, что соотношение (5.71) не является достаточным условием.

Указание. Неравенство (5.71) характеризует локальный минимум.

Рассматривая условие (5.71), можно заметить, что

$$x_i^* = \begin{cases} 1, & \text{если } g_i < 0, \\ 0, & \text{если } g_i > 0, \\ p_i, & \text{если } g_i = 0, \text{ где } p_i \text{ — произвольный бивалентный параметр.} \end{cases} \tag{5.72}$$

Используя (5.72), можно определить булеву функцию $\Omega_i(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n)$ так, что

$$\Omega_i(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n) =$$
$$= \begin{cases} 1, & \text{если } g_i < 0, \\ 0, & \text{если } g_i > 0, \\ p_i, & \text{если } g_i = 0, \text{ где } p_i \text{ — произвольный бивалентный параметр.} \end{cases} \tag{5.73}$$

Если заменить $x \cup y$ на $x + y - xy$ и $x \cap y$ на xy, то соотношение (5.73) можно рассматривать как псевдобулеву функцию и, следовательно,

$$x_i = \Omega(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n). \tag{5.74}$$

Соотношения (5.71) — (5.74) дают следующий результат.

Следующие три теоремы сформулированы Хэммером и Рудеану [22].

Теорема 5.12. *Пусть дан вектор* $(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n) \in B_{2^n}$, *тогда если*

$$x_i = \Omega(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n),$$

то вектор $(x_1, \ldots, x_i, \ldots, x_n) \in B_{2^n}$ *удовлетворяет уравнению* (5.71), *и обратно, если* $(x_1, \ldots, x_i, \ldots, x_n) \in B_{2^n}$ *является решением уравне-*

ния (5.71), *то существует* p_i (*бивалентный параметр, принадлежащий* B_2), *такой, что уравнение* (5.73) *выполняется.*

Построение булевой функции $\Omega_i(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n)$ легко выполнить следующим образом. Пусть Ω' — характеристическая функция $g_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) < 0$, а Ω_i'' — характеристическая функция

$$g_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = 0.$$

Тогда

$$\Omega_i(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n) = \Omega_i' \cup (p_i \cap \Omega_i''), \qquad (5.75)$$

поскольку

$$\Omega_i' = \begin{cases} 1, & \text{если } g_i < 0, \\ 0, & \text{если } g_i \geqslant 0, \end{cases}$$

$$\Omega_i'' = \begin{cases} 1, & \text{если } g_i = 0, \\ 0 & \text{в противном случае.} \end{cases}$$

Из теоремы (5.12) и соотношения (5.71) получается следующее необходимое и достаточное условие того, что x^* является точкой минимизации.

Теорема 5.13. *Вектор* $x^* = (x_1^*, \ldots, x_n^*)$ *является точкой минимизации функции* $f(x_1, \ldots, x_n)$ *тогда и только тогда, когда выполняются следующие два условия:*

а. *Существует параметр* $p_i^* \in B_2$, *такой, что*

$$x_i^* = \Omega_i(x_1^*, \ldots, x_{i-1}^*, p_i^*, x_{i+1}^*, \ldots, x_n^*). \qquad (5.76)$$

б. *Вектор* $(x_1^*, \ldots, x_{i-1}^*, x_{i+1}^*, \ldots, x_n^*)$ *является точкой минимизации функции* (Ω_i *рассматривается как псевдобулева функция*)

$$\Omega_i(x_1, \ldots, x_{i-1}, p_i, x_{i+1}, \ldots, x_n)\, g_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n), \\ + h_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \qquad (5.77)$$

Упражнение 5.18. Докажите теорему 5.13.

Указание. Воспользуйтесь теоремой 5.12 и определением точки минимизации.

Путем повторного применения теоремы 5.13 можно получить алгоритм отыскания минимума $f(x_1, \ldots, x_n)$, использующий следующую теорему:

Теорема 5.14. *Вектор* $x^* = (x_1^*, \ldots, x_n^*)$ *является точкой минимизации псевдобулевой функции* $f(x_1, \ldots, x_n)$ *тогда и только тогда, когда существуют значения* $p_1^*, \ldots, p_n^*$ *такие, что*

$$x_1^* = \Omega_1(p_1^*, x_2^*, \ldots, x_n^*),$$
$$x_2^* = \Omega_2(p_2^*, x_3^*, \ldots, x_n^*),$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$x_n^* = \Omega_n(p_n^*)$$

и $f(x_1^*, \ldots, x_n^*) = f(\Omega_1, \Omega_2, \ldots, \Omega_n)$ — *минимум функции* $f(x_1, \ldots, x_n)$.

Алгоритм отыскания точек минимизации и минимума псевдобулевой функции

Шаг 1. Если $f_i = \text{const}$, то $\min f = f_i$. Переходим к шагу 4. Если $f_i \neq \text{const}$, то переходим к шагу 2.

Шаг 2. Находим g_i, h_i, такие, что $f_i = x_i g_i + h_i$, и строим Ω_i' и Ω_i''. Полагаем $x_i = \Omega_i' \cup p_i \Omega_i''$, где $\Omega_i' \cup p_i \Omega_i''$ выражается в виде псевдобулевой функции. Переходим к шагу 3.

Шаг 3. Образуем $f_{i+1} = f_i[\Omega_i'(x_{i+1}, \ldots, x_n), x_{i+1}, \ldots, x_n]$. Полагаем $i = i + 1$ и переходим к шагу 1.

Шаг 4. Если $i = 1$, то все значения x_i можно брать произвольными, $i = 1, \ldots, n$. Если $i \neq 1$, скажем $i = m$, то

$$\begin{aligned} x_m &= \Omega_m(p_m), \\ x_{m-1} &= \Omega_{m-1}(p_{m-1}, x_m), \\ &\ldots\ldots\ldots\ldots\ldots \\ x_1 &= \Omega_1(p_1, x_2, \ldots, x_m). \end{aligned}$$

Если $m < n$, то $x_{m+1}, \ldots, x_n$ могут принимать произвольные значения. При $m = n$ величины $x_1, \ldots, x_n$ определены выше.

В изложенном здесь алгоритме при отыскании минимального значения f можно использовать Ω_i вместо Ω_i'. В этом алгоритме Ω_i' использовалось ввиду того, что желательно было получить минимум $f(x_1, \ldots, x_n)$; следовательно, параметр p_i полагается равным нулю.

Пример 5.7. Требуется найти минимум функции

$$-5x_1x_2x_3 + 10x_1x_2 + 3x_1x_3 + 3x_2x_3 - 3x_1 - 8x_2 - 3x_3 + 3. \qquad (5.78)$$

Записываем

$$f_1(x_1, x_2, x_3) = x_1(-5x_2x_3 + 10x_2 + 3x_3 - 3) + \\ + 3x_2x_3 - 8x_2 - 3x_3 + 3 \qquad (5.79)$$

и решаем

$$-5x_2x_3 + 10x_2 + 3x_3 - 3 \leqslant 0.$$

Полагаем

$$y_1 = x_2x_3, \quad y_2 = \bar{x}_2, \quad y_3 = \bar{x}_3.$$

Тогда надо решить

$$10y_2 + 5y_1 + 3y_3 \geqslant 10. \qquad (5.79а)$$

Решением неравенства (5.79а) является $y_2 = 1$, $y_1 = y_3 = 0$ (базисное решение), а $\Omega_1' = \bar{x}_2\bar{x}_3$, $\Omega_1'' = 0$.

$$f_2(x_2, x_3) = f_1(\Omega_1', x_2, x_3) = -5x_2.$$

Из того что $g_2 = 1$, следует, что $\Omega_2' = 1$ и $\Omega_2'' = 0$. Поэтому $f_3 = f_2(\Omega_2', x_3) = -5 = \min f(x_1, x_2, x_3)$. Точки минимизации имеют вид

$$x_2 = 1,$$
$$x_1 = \bar{x}_2\bar{x}_3 = 0\bar{x}_3 = 0,$$
$$x_3 = p \text{ (произвольный бивалентный параметр).}$$

Упражнение 5.19. Минимизируйте следующую функцию f и найдите все точки минимизации:

$$f = 2x_1 + 3x_2 - 7x_3 - 5x_1x_2x_3 + 3x_2x_4 + 9x_4x_5 - 2x_1x_5.$$

Решение.

$$x_1 = x_2 = x_3 = x_5 = 1,\ x_4 = 0,\ \min f = -9.$$

Минимизация произвольной целевой функции при наличии системы ограничений

Рассмотрим следующую задачу минимизации. Требуется минимизировать псевдобулеву функцию $f(x_1, \ldots, x_n)$ при наличии системы ограничений $G(x_1, \ldots, x_n) \geqslant 0$.

Пусть $F = (\Omega(\alpha^1, K_1), \ldots, \Omega(\alpha^p, K_p)$ — множество всех различных базисных решений $G(x_1, \ldots, x_n) \geqslant 0$. Тогда указанную задачу минимизации можно решать как последовательность не более чем p задач минимизации, используя следующий алгоритм.

Алгоритм

Шаг 1. Образуем псевдобулевы функции f_j, $j = 1, \ldots, p$, которые получаются путем подстановки z_i в $f(x_1, \ldots, x_n)$ вместо x_i, где

$$z_i = \begin{cases} \alpha_i, & \text{если } i \in K \text{ — множество индексов } \Omega(\alpha^j, K_j), \\ x_i, & \text{если } i \notin K \text{ — множество индексов } \Omega(\alpha^j, K_j). \end{cases}$$

Переходим к шагу 2.

Шаг 2. Минимизируем каждую функцию f_j, $j = 1, \ldots, p$, используя метод, изложенный в части III, и обозначим минимумы через $\delta^1, \ldots, \delta^p$. Переходим к шагу 3.

Шаг 3. Находим $\min \delta^j = \min f$. Точки минимизации $f(x_1, \ldots, x_n)$ являются такими точками x^*, что $f(x_1^*, \ldots, x_n^*) = \min f$.

Упражнение 5.20. Минимизируйте функцию

$$f = 3x_1\bar{x}_2 - 8\bar{x}_1x_3x_6 + 4x_2x_5\bar{x}_6 - 7x_5\bar{x}_6 + 3x_4 - 5x_4x_5x_6$$

при ограничениях

$$2x_1 - 3x_2 + 5x_3 - 4x_4 + 2x_5 - x_6 \leqslant 2,$$
$$4x_1 + 2x_2 + x_3 - 8x_4 - x_5 - 3x_6 \geqslant 4.$$

Решение. $\min f = -12$, $x = (0, 1, 1, 1, 0, 1)$ и $(0, 0, 1, 1, 0, 1)$.

С другими методами минимизации псевдобулевых функций при наличии ограничений можно познакомиться в [22].

ЛИТЕРАТУРА

1. Balas E., An Additive Algorithm for Solving Linear Programs with Zero-One Variables, *Operations Res.*, 13, 517 (1965).
1a. Balas E., Duality in Discrete Programming, Stanford Univ. Techn. Rep. № 67-5, Nov. 1967.
1b. Balas E., Discrete Programming by the Filter Method, *Operations Res.*, 15, 915 (1967).
2. Balinski M. L., Integer Programming: Methods, Uses, Computation, *Management Sci.*, 12, 253 (Nov. 1965).
3. Balinski M. L., On Finding Integer Solutions to Linear Programs, H. Koenig, ed., *Proc. IBM Sci. Symp. Combinatorial Problems*, 225 (1965).
4. Balinski M. L., Some General Methods in Integer Programming, in: «Nonlinear Programming» (NATO Summer School, Menton 1964), ch. 9, North-Holland Publ. Co., Amsterdam, 1967, p. 221.
4a. Basin S., Algebraic Formulation of the Minimum Crossing Problem for Complete Graphs (частное сообщение) 1967.
4b. Benders J. F., Catchpole C. A. R., Kuiken C., Discrete Variables, Optimization Problems, The Rand Corp., Symposium on Mathematical Programming, March 16—20, 1959.
5. Ben-Israel A., Charnes A., On Some Problems of Diophantine Programming, *Cahiers Centr. d'Etud. Rech. Operation* (Bruxelles), 4, 215 (1962).
5a. Birkhoff G., MacLane S., A Survey of Modern Algebra, The Macmillan Co., N.Y., 1953.
6. Bradley G. H., Equivalent Integer Programs I: Basic Theory, Operations Research House, Stanford Univ., Techn. Rept. № 68-4, March 28, 1968.
7. Bradley G. H., Equivalent Integer Programs II: The Special Problem, Operations Research House, Stanford Univ., Techn. Rept. № 68-6, April 18, 1968.
8. Breuer M. A., The Minimization of Boolean Functions Containing Unequal and Nonlinear Cost Functions, Electronics Res. Lab., Univ. of California, ser. 60, № 431, Jan. 22, 1962.
9. Cooper L., Drebes C., An Approximate Solution Method for the Fixed Charge Problem, Naval Res. Log. Quart., 14, № 1, 101 (March 1967).
10. Dantzig G. B., Linear Programming and Extensions, Princeton Univ. Press, Princeton, N.J., 1963; русский перевод: Данциг Дж. Б., Линейное программирование, его применения и обобщения, изд-во «Мир», 1966.
11. Dantzig G. B., On the Significance of Solving Linear Programming Problems with Some Integer Variables, *Economet.*, 28, № 1, 30 (Jan. 1960).
12. Fox B., Discrete Optimization Via Marginal Analysis, *Management Sci.*, 13, 210 (Nov. 1966).
13. Gass S. I., Linear Programming Methods and Applications, 3d ed., McGraw-Hill, N.Y., 1969; русский перевод: Гасс С. И., Линейное программирование (методы и приложения), Физматгиз, 1961.
14. Gass S. I., Recent Developments in Linear Programming, in: 1961. Advances in Computers», Academic Press, Inc., N.Y., vol. 2, 1961, pp. 295—377.
15. Geoffrion A. M., Integer Programming by Implicit Enumeration and Balas' Method, SIAM Rev., 9, № 2, 178 (April 1967).
16. Gilmore P., Gomory R. E., The Theory and Computation of Knapsack Functions, *Operations Res.*, 14, 1045 (1966).
16a. Gilmore P., Gomory R. E., A Linear Programming Approach to the Cutting Stock Problem—Part II, *Operations Res.*, 863 (1963).
17. Glover F., A Multiphase-dual Algorithm for the Zeor-One Integer Programming Problem, *Operations Res.*, 13, 879 (1965).

18. Glover F., Generalized Cuts in Diophantine Programming, *Management Sci.*, 13, 254 (Nov. 1966).
19. Gomory R. E., All Integer-Integer Programming Algorithm, IBM Research Center, RC-189, Jan. 1960.
20. Gomory R. E., Outline of an Algorithm for Integer Solutions to Linear Programs, *Bull. Am. Math. Soc.*, 64, 275 (1958).
21. Gordon R. B., Selecting Different Dropping Variables in the Simplex Algorithm, Operations Research House, Stanford Univ., Techn. Rept. № 68-3, March 1968.
22. Hammer (Ivănescu) P. L., Rudeanu S., Pseudo-Boolean Methods for Bivalent Programming, Springer-Verlag OHG, Berlin, 1966.
23. Hammer (Ivănescu) P. L., Rudeanu S., Boolean Methods in Operations Research and Related Areas, Springer-Verlag OHG, Berlin, 1968.
24. Heller I., Tompkins C. B., An Extension of a Theorem of Dantzig's in: Kuhn H. W., Tucker A. W (eds.), Linear Inequalities and Related Systems, Princeton Univ. Press, Princeton, N.J., 1956; русский перевод: Линейные неравенства и смежные вопросы, Сб. статей под ред. Куна Г. У. и Таккера А. У., ИЛ, 1959.
25. Heller I., Hoffman A. J., On Unimodular Matrices, *Pacific J., Math.*, 12, № 4, 1321 (1962).
26. Heller I., On Linear Systems with Integral Valued Solutions, *Bull. Am. Math. Soc.*, 1351 (Oct. 1956).
27. Heller I., On Unimodular Sets of Vectors, in: Graves R. L., Wolfe P. (eds), Recent Advances in Mathematical Programming, McGraw-Hill, N.Y., 1963.
28. Hoffman A. J., On Simple Linear Programming Problems, IBM Research Center, RC-541, Sept. 1961.
29. House R. W., Nelson L. D., Rado T., Computer Studies of a Certain Class of Linear Integer Problems; Chapter in: Lavi A., Vogl T. P. (eds.), Recent Advances in Optimization Techniques, Wiley, N.Y., 1966.
30. Kaplan S., Solution of the Lorie-Savage and Similar Integer Programming Problems by the Generalized Lagrange Multiplier Method, *Operations Res.*, 14, 6, 1130 (Nov.-Dec. 1966).
31. Lawler E. L., Bell M. D., A Method for Solving Discrete Optimization Problems, *Operations Res.*, 14, 6, 1098 (Nov.—Dec. 1966).
31a. Miller C. E., Tucker A. W., Zemlin R. A., Integer Programming Formulation of Traveling Salesman Problems, *Journal of the A. C. M.*, 7, № 4, (Oct. 1960).
32. Rockafellar R. T., A Combinatorial Algorithm for Linear Programs in the General Mixed Form, *J. Soc. Ind. Appl. Math.*, 12 № 1 (March 1964).
33. Saaty T. L., Suzuki G., A Nonlinear Programming Model in Optimum Communication Satellite Use, *SIAM Rev.*, 7, 403 (July 1965).
34. Saaty T. L., Bram J., Nonlinear Programming, in: Nonlinear Mathematics, ch. 3, McGraw-Hill, N.Y., 1964.
35. Shapiro J. F., Wagner H. M., A Finite Renewal Algorithm for the Knapsack and Turnpike Models, *Operations Res.*, 15, № 2, 319 (March—April 1967).
35a. Simmonard M., Linear Programming, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966.
36. Veinott A. F. Jr., Dantzig G. B., Integral Extreme Points, *SIAM Rev.*, 10, № 3 (July 1968).
37. Wagner H. M., An Integer Linear-Programming Model for Machine Scheduling, *Naval Res. Log. Quart.*, 6, № 2, 131 (June 1959).
38. Wegner P., Doig A., Symmetric Solutions of the Postage Stamp Problem, *Rev. Franc. Rech. Operation.*, 41, 353 (1966).
39. Weissman J., Boolean Algebra, Map Coloring, and Interconnections, *Am. Math. Monthly*, 69, № 7, 608 (Aug.—Sept. 1962).
40. Witzgall C., An All Integer Programming Algorithm with Parabolic Constraints, *J. Soc. Ind. Appl.* Math., 11, 855 (Dec. 1963).
41. Young R. D., A Primal (All-Integer) Integer Programming Algorithm, *J. Res. Nat. Bur. Standards* (U.S.), 69B, 213 (Sept. 1965).

42*.Гольштейн Ю. Г., Юдин Д. Б., Новые направления в линейном программировании, изд-во «Сов. радио», 1966.
43*.Емеличев В. А., Дискретная оптимизация, последовательные схемы решения, *Кибернетика*, № 6 (1971); № 2 (1972).
44*.Корбут А. А., Финкельштейн Ю. Ю., Дискретное программирование, изд-во «Наука», 1969.
45*.Balas E., Intersection Cuts—A New Type of Cutting Planes for Integer Programming, Management Sci., Res. Rep., № 187, Oct. 1969.
46*.Balas E., Bowman V. T., Lover F. G., Sommer D., An Intersection Cut from the Dual of the Unite Hipercube, *Operations Res.*, 19, № 1 (1971).
47*.Balas E., Integer Programming and Convex Analysis, Management Sci., Res. Rep., № 246, Carnegie-Mellon Univ., April 1971.

Оглавление

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., 2, издательство «Мир».

Т. Саати

ЦЕЛОЧИСЛЕННЫЕ МЕТОДЫ
ОПТИМИЗАЦИИ И СВЯЗАННЫЕ С НИМИ
ЭКСТРЕМАЛЬНЫЕ ПРОБЛЕМЫ

Редактор Л. П. Якименко
Художник А. В. Карпов
Художественный редактор Ю. С. Урманчеев
Технический редактор Л. П. Бирюкова
Корректор Т. С. Лаврова

Сдано в набор 24/IV 1973 г.
Подписано к печати 1/X 1973 г.
Бумага № 2 60 × 90¹/₁₆=9,5 бум. л.
19 усл. печ. л.
Уч.-изд. л. 17,80. Изд. № 20/6998.
Цена 1 р. 44 к. Зак. 01037

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