

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РЕСПУБЛИКИ КАЗАХСТАН**

**Казахский национальный исследовательский
технический университет имени К. И. Сатпаева**

Ж. Е. ЕГИНБАЕВ

ТЕОРИЯ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ

**Рекомендовано Научно-методическим советом университета в
качестве учебного пособия**

Алматы 2016

УДК 681.5.015.23 (075)

ББК 32.965 я7

Е 29

Рецензенты:

Бегалиев И. Т. – проф., президент ТОО «Полиграфкомбинат»

Тураев И. И. – проф., ген. директор ТОО AZ PRIN CENTRE

Сураишов Н. Т. – зав. каф., д-р техн. наук, проф. КазННТУ
имени К. И. Сатпаева, академик НАЕН РК

Егинбаев Ж. Е.

Е29 Теория решений задач оптимизации. Учебное пособие. / Ж. Е. Егинбаев.
– Алматы: КазННТУ имени К. И. Сатпаева, 2016. – 239 с.: Ил.–48.
Табл. – 14. Библиогр. – 9 наим.

ISBN 978–601–228–919–0

Печатается по плану издания Министерства образования и науки
Республики Казахстан на 2016 г.

В данном учебном пособии освещено систематическое изложение основ методов оптимизации, а также оно и имеет прикладную инженерно-техническую направленность. Основное внимание уделено прикладным и вычислительным аспектам оптимизации, связанным с разработкой численных методов решения задач и построением алгоритмов их реализации, математическому моделированию полиграфических процессов, их оптимизации и их практическим применениям. Изложенный материал позволяет научить магистра исследовать процессы методом оптимизации, включая составление математического описания, выбор метода решения, программную реализацию модели и проверку адекватности модели реальному объекту. Приведены конкретные примеры теории решения задач оптимизации.

Предлагаемое учебное пособие «Теория решений задач оптимизации» рассчитано на специалистов в области математического моделирования и смежных с ней научно-технических дисциплин, а также на преподавателей и студентов соответствующих специальностей.

Учебное пособие предназначено для магистров, обучающихся по специальности 6М072200 «Полиграфия», а также может быть полезно инженерно-техническим работникам, занимающимся расчетами и применением систем моделирования и оптимизации.

УДК 681.5.015.23 (075)

ББК 32.965 я7

ISBN 978–601–228–919–0

© Егинбаев Ж. Е., 2016

© КазННТУ им. К. И. Сатпаева, 2016

ВВЕДЕНИЕ

В современном мире ведущие позиции занимают те государства, в которых целенаправленно стимулируется развитие инновационного потенциала на базе самых современных научно-технических достижений. Однако невозможно создавать инновационные технологии без построения индустрии генерации новых идей.

Решение этих задач обеспечивает принципиально новый подход к инженерной деятельности, расширяет влияние новых достижений на экономику и создает условия для принятия стратегических концептуальных решений в области техники.

Владение методами поиска новых технических решений дает возможность наметить стратегические цели развития отрасли техники, так как эти методы обладают предсказательной силой и позволяют прогнозировать тенденции научно-технического прогресса.

В современном мире накопление технических знаний происходит с колоссальной скоростью. Области техники приобретают все большую специализацию. При этом каждая специальность создает свой профессиональный язык, свои способы отражения достижений. Выходом из этой ситуации может стать использование теории решения задач оптимизации и других методов поиска решений.

По мере развития техники, усложнения характера производственной деятельности связь математики и практики становилась все более опосредованной. Постепенно возникла логика собственного развития, порождающая задачи которые стимулировали появление новых идей и методов дисциплины. Тем не менее, практическая деятельность людей и другие науки всегда оказывали решающее влияние на эволюцию математики, на выбор новых направлений, на формирование шкалы ценностей.

Решение задачи – сложный и многогранный мыслительный процесс, в котором важными составляющими являются, во-первых, понимание задачи и, во-вторых, психологическая готовность к решению. Последнее достигается уверенностью в своих силах, которая основана на осознанном владении приемами и методами решения задач.

Очень важно, что в процессе развития общества возникла потребность в «культуре мышления», и эту нагрузку, в своей значительной части, приняла на себя математика. Она превратилась в определенную школу мышления и анализа. Математика играет еще одну важную роль – она основу того языка, который объединяет различные направления науки, облегчает миграцию идей; проникая в различные дисциплины, она постепенно становится их составной частью.

1. ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ ТЕОРИИ РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ

Могущество математики в ее единстве, в ее целостности. Вот почему, говоря о прикладной математике, следует иметь в виду не какую-либо специальную дисциплину, а вопросы использования математических методов анализа для решения прикладных задач. Термин «прикладная математика» («*mathematique appliquee*») возник в XIX в. именно в связи с использованием математики для решения задач небесной механики, статистики, инженерного дела и т. д. Крупнейшими представителями прикладной математики прошлого были Пуанкаре, Чебышев, Ляпунов и др.

Во второй половине XX в. возник целый ряд разделов математики, в название которых входит слово программирование: линейное программирование, нелинейное программирование, динамическое программирование, стохастическое программирование и т. д. Термины динамическое программирование, линейное программирование и так далее стали общепринятыми и поэтому будут сохранены.

Актуальность всего того направления, которое иногда называется теория оптимизации, возникла лишь в эпоху ЭВМ, ибо без электронных машин методы оптимизации лишены какой-либо перспективы своего практического использования – реализация алгоритмов отыскания экстремумов чересчур трудоемка. Но возникла эта дисциплина как глава анализа, как развитие тех классических подходов, которые связаны с именами Ньютона, Эйлера, Лагранжа. Конечно, в процессе развития появилась и новые методы и новые идеи, но основа все-таки лежит в анализе.

Моделирование – одна из основных категорий теории познания: на идее моделирования по существу базируется любой метод научного исследования – как теоретический (при котором используются различного рода знаковые, абстрактные модели), так и экспериментальный (использующий предметные модели); является одним из методов изучения окружающего мира; предполагает построение действующей материальной или

математической модели, обладающей свойствами, подобными свойствам рассматриваемой системы.

Моделирование используется:

- для экспериментирования или численной оценки, т. е. для предсказания последствий изменения параметров системы или условий ее работы, когда осуществление такого изменения в реальной обстановке связано с большими затратами или определенным риском;
- для изучения систем с целью их переделки или усовершенствования;
- для ознакомления с системами или условиями их работы, которые пока еще не существуют в реальной действительности;
- для проверки или демонстрации новой идеи, системы или метода.

Под *математическим моделированием* понимают изучение свойств объекта на математической модели. Математическое моделирование как альтернатива пробному физическому эксперименту по мере усложнения техники и решаемых инженерных задач все прочнее входит в практику производства, как достаточно эффективный и экономически выгодный путь в подготовке любого нововведения. В широком смысле математическое моделирование есть процесс оперирования созданной по исходным физическим параметрам моделью, обеспечивающий получение данных о поведении интересующего объекта. Его целью является определение оптимальных условий протекания процесса, управление им на основе математической модели и перенос результатов на объект.

Методы оптимизации – курс, ориентированный на решение задач, возникающих в практической деятельности: в экономике, физике, инженерном деле и т. д.; это мост, ведущий от математики к анализу конкретных задач; этот курс состоит из трех частей: конечномерные задачи оптимизации, вариационное исчисление, включающее в себя теорию оптимального управления и теория принятия решений. Приведем несколько прямых методов оптимизации одномерной минимизации функций: *метод перебора* – здесь рассматриваемая область покрывается сеткой с шагом h и определяются значения функции в ее узлах и, сравнивая полученные числа между

собой, находят среди них наименьшее и принимается оно приближенно за наименьшее значение функции для всей области; *метод парабол* – использующий полиномы второго порядка, на каждой итерации этого метода строится квадратный трехчлен, график которого (парабола) проходит через три выбранные точки графика функции $f(x)$; *метод дихотомии* (половинного деления; от греческого *dichotomia* – разделение надвое; способ классификации: классы, множества, понятия, термины и другие разбиваются на пары «соподчинённых» элементов) – поиск минимума функции начинается с двух экспериментов и выбирается один; *метод золотого сечения* – здесь рассматривается симметричное расположение точек x_1 и x_2 на отрезке $[a, b]$, при котором одна из них становится пробной точкой на новом отрезке, полученном после исключения части исходного отрезка и *метод поразрядного поиска* – здесь перебор точек отрезка происходит сначала с шагом $\Delta = x_{i+1} - x_i > \varepsilon$ до тех пор, пока не выполнится условие $f(x_i) \leq f(x_{i+1})$ или пока очередная из этих точек не совпадает с концом отрезка. Главным потребителем и «поставщиком» оптимизационных задач является исследование операций – дисциплина, занимающаяся проблемами принятия решений (деятельностью, завершающей любое исследование).

Основная литература: 1 [5–11]; 2 [5–12], 3–7.

Контрольные вопросы:

1. Чем стимулируется развитие инновационного потенциала?
2. Возникновение и сущность термина «прикладная математика».
3. Сущность понятия «моделирование».
4. Сущность понятия «математическое моделирование».
5. Методы оптимизации.

2. МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ОСНОВНОЙ МЕТОД РЕШЕНИЯ ЗАДАЧ ОПТИМИЗАЦИИ ПОЛИГРАФИЧЕСКИХ ПРОЦЕССОВ

2.1. Этапы математического моделирования

Математическое моделирование включает три взаимосвязанных этапа: 1) составление математического описания изучаемого объекта; 2) выбор метода решения системы уравнений математического описания и реализация его в форме моделирующей программы; 3) установление соответствия (адекватности) модели объекту. *Адекватность* (правильное соответствие) (от латинского *adaequatus* «адекватус» – приравненный, равный, тождественный) – способность отображать заданные свойства объектов с погрешностью не выше заданной.

На этапе составления математического описания предварительно выделяют основные явления и элементы в объекте и затем устанавливают связи между ними. Далее, для каждого выделенного элемента и явления записывают уравнение (или систему уравнений), отражающее его функционирование. Кроме того, в математическое описание включают уравнения связи между различными выделенными явлениями. В зависимости от процесса математическое описание может быть представлено в виде системы алгебраических, дифференциальных, интегральных и интегродифференциальных уравнений.

Этап выбора метода решения и разработки моделирующей программы подразумевает выбор наиболее эффективного метода решения из имеющихся (под эффективностью имеются в виду быстрота получения и точность решения) и реализацию его сначала в форме алгоритма решения, а затем – в форме программы, пригодной для расчета на ЭВМ. *Язык программирования*, формальный язык для описания данных (информации) и алгоритма (программы) их обработки на ЦВМ. *Алгоритм* (алгорифм) (от *algorithmi*, *algorismus*, первоначально – латинская транслитерация имени математика аль-Хорезми) конечный набор правил, позволяющих чисто механически

решать любую конкретную задачу из некоторого класса однотипных задач. *Алгоритмический язык*, набор символов и система правил составления и истолкования языковых конструкций из этих символов, предназначенные для записи алгоритмов. *Аппроксимация* (от латинского *approximo* – приближаюсь), замена одних математических объектов (например, чисел или функции) другими, более простыми и в том или ином смысле близкими к исходным (например, кривых линий близкими к ним ломаными).

Построенная на основе физических представлений модель должна верно качественно и количественно описывать свойства моделируемого процесса, т. е. она должна быть адекватна моделируемому процессу. Для проверки адекватности математической модели реальному процессу нужно сравнить результаты измерений на объекте в ходе процесса с результатами предсказания модели в идентичных условиях.

Этап установления адекватности модели является заключительным в последовательности этапов, выполняемых при ее разработке. На рис. 2.1 изображена общая схема разработки математической модели.

При построении математической модели реальное явление упрощается, схематизируется и полученная схема описывается в зависимости от сложности явлений с помощью того или иного математического аппарата.

От правильности учета в модели характерных черт рассматриваемого процесса зависят успех исследования и ценность полученных результатов моделирования.

В модели должны быть учтены все наиболее существенные факторы, влияющие на процесс, и вместе с тем она не должна быть загромождена множеством мелких, второстепенных факторов, учет которых только усложнит математический анализ и сделает исследование либо чрезмерно громоздким, либо вообще нереализуемым. *Анализ* (от греческого *analysis* – разложение, расчленение) – процедура разложения объекта (процесса, явления) на составные части. *Синтез* (от греческого *synthesis* – соединение, сочетание, составление) – метод научного исследования, состоящий в познании рассматриваемого объекта как единого целого, в единстве и взаимной связи его частей.

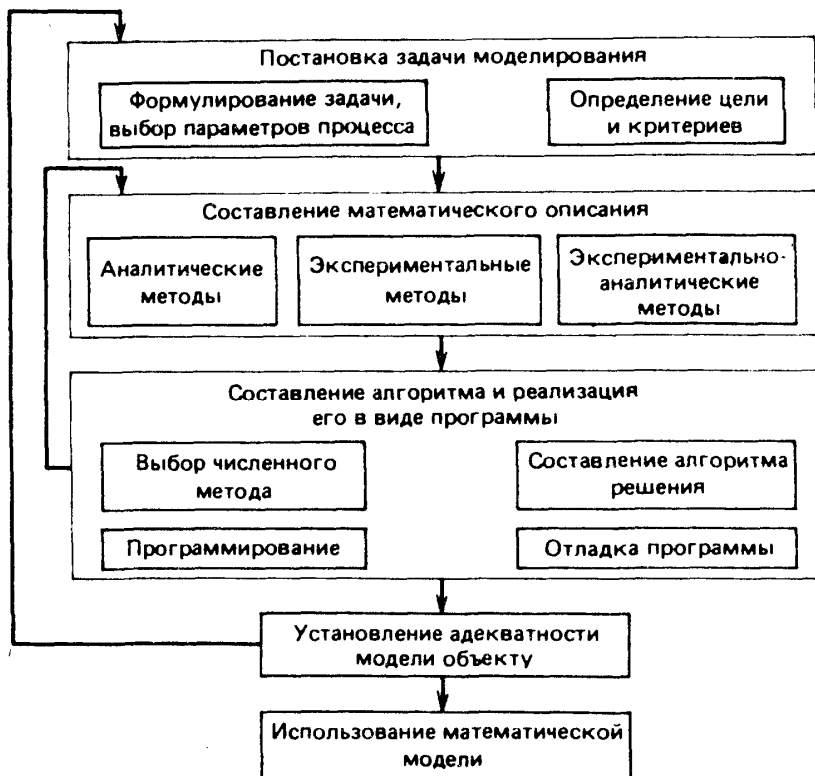


Рис. 2.1. Этапы разработки математической модели

Метод математического моделирования применяют при изучении свойств процессов, для которых имеется достаточно точное математическое описание. В зависимости от степени полноты математического описания можно выделить два предельных случая: а) известны полная система уравнений, описывающая все основные стороны моделируемого процесса, и все числовые значения параметров этих уравнений; б) полное математическое описание процесса отсутствует. Этот второй случай типичен для решения кибернетических задач, в которых приходится иметь дело с управлением процессами при наличии неполной информации об объекте и действующих на него возмущениях. *Информационный процесс* – процесс сбора,

хранения, передачи и обработки информации. При отсутствии достаточной информации об исследуемых явлениях их изучение начинается с построения простейших моделей, но без нарушения основной (качественной) специфики исследуемого процесса.

2.2. Физическое описание природы объекта

Построение любой математической модели начинают с физического описания объекта моделирования. При этом выделяют «элементарные» процессы, протекающие в объекте моделирования, которые подлежат отражению в модели, и формулируют основные допущения, принимаемые при их описании. В свою очередь, перечень учитываемых «элементарных» процессов определяет совокупность явлений, описывающих объект, которые включают в математическую модель. В данном случае под «элементарным» процессом понимается физико-химический процесс, относящийся к определенному классу явлений, например массообмен, теплопередача и т. д. Здесь следует отметить, что название «элементарные» процессы отнюдь не означает, что данные процессы являются простейшими и описываются несложными уравнениями. Так, массообмен является предметом целой теории, до настоящего времени еще далекой до полного завершения. Это название означает лишь, что такие процессы являются составляющими много более сложного всего химико-технологического процесса.

Обычно при математическом моделировании объектов химической технологии принимаются во внимание следующие «элементарные» процессы: 1) движение потоков фаз; 2) массообмен между фазами; 3) теплопередача; 4) изменение агрегатного состояния (испарение, конденсация, растворение и т. д.); 5) химические превращения.

Полнота математического описания «элементарных» процессов в модели зависит от их роли во всем химико-технологическом процессе, степени изученности, глубины взаимосвязи «элементарных» процессов в объекте и желаемой точности всего описания. Взаимосвязь «элементарных»

процессов может быть чрезвычайно сложной. Поэтому на практике часто делают различные допущения относительно характера связей, что позволяет избежать необходимости введения в модель недостаточно изученных зависимостей и, следовательно, излишнего усложнения описания.

Например, при физическом описании процесса ректификации смесей выделяют следующие «элементарные» процессы: 1) гидродинамика потоков жидкости и пара в колонне; 2) массообмен между жидкостью и паром; 3) теплопередача между жидкостью и паром; 4) испарение жидкости и конденсация пара. Все указанные «элементарные» процессы протекают либо на тарелке, либо в насадочной секции колонн и прямо связаны между собой. Полное описание этих процессов представляет собой чрезвычайно сложную систему уравнений. Только описание гидродинамики потока жидкости на тарелке (либо в насадке) с помощью уравнения Навье-Стокса представляет собой задачу чрезмерной вычислительной сложности. Не менее сложно и решение задачи полного описания массообмена между потоками жидкости и пара. Вместе с тем эти задачи должны решаться совместно как единая система уравнений. Отсюда следует, что без разумных упрощающих допущений здесь не обойтись. Поэтому обычно принимают идеализированное представление относительно движения потоков пара и жидкости (пар движется в режиме полного вытеснения, а жидкость полностью перемешивается на тарелке), а массопередачу выражают через эффективность ступеней разделений, определяемую в большинстве случаев полуэмпирическими методами, либо вообще не рассматривают ее, считая, что на каждой ступени разделения достигается равновесие.

Следует отметить, что иногда физическое описание объекта моделирования устанавливается в результате математического моделирования. Так, математическое моделирование используется для проверки некоторых гипотез о механизме процессов, протекающих в объекте. Для этого в состав модели вводят исследуемые соотношения, чтобы по результатам последующего моделирования судить о справедливости того или иного физического предположения. Например, механизмы

каталитических химических превращений в большинстве случаев неизвестны исследователям. Закладывая в математическую модель тот или иной механизм протекания химической реакции и сравнивая результаты моделирования с экспериментальными, можно отыскать наиболее близкий к истинному механизм.

Литература: 2[7–12], 7 [313–316].

Контрольные вопросы:

1. Назовите трех взаимосвязанных этапов математического моделирования.
2. В зависимости от процесса математическое описание может быть представлено в виде системы каких уравнений?
3. Что представляет собой описание гидродинамики потока жидкости на тарелке (либо в насадке) с помощью уравнения Навье-Стокса?
4. Этапы математического моделирования.
5. Анализ и синтез.
6. Информационный процесс.
7. Физическое описание природы объекта.

3. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ. МОДЕЛЬ И КОНСТРУКЦИЯ ПЕЧАТНЫХ ИЗДАНИЙ

3.1. Математическая модель

Любая модель представляет собой некоторое отображение объекта в форме, отличной от формы его реального существования, и служит средством, помогающим в объяснении или понимании происходящих процессов и при решении задач.

В широком смысле модель – это любой образ, умозрительный или материальный, замещающий рассматриваемый объект при его изучении. Это может быть чертеж, конспект, график, план, выкройка, таблица, макет, шаблон и т. д. Использование моделей позволяет упростить рассматриваемый объект, выделить существенные свойства, сделать его более обозримым и наглядным для изучения.

Математической моделью называется приближенное описание какого-либо явления или процесса внешнего мира, выраженное с помощью математической символики. В отличие от реально поставленного эксперимента математическая модель имеет ряд преимуществ, которые связаны с тремя основными аспектами. Во-первых, это экономия материальных ресурсов, требуемых для постановки и проведения физического эксперимента; во-вторых, возможность апробации системы в экстремальных условиях и даже за их пределами; в-третьих, оценка работоспособности систем с длительными технологическими (дни, недели, месяцы, годы) циклами в существенно более сжатые сроки (например, исследование ниже десятых долей градуса в час). Для рассматриваемой узкой проблемы математическая модель по форме представляет собой систему дифференциальных уравнений в частных производных, описывающих поведение объекта в окружающей среде с установленными краями и начальными условиями. Существенным являются свойство *полноты* математической модели и ее *продуктивность*. *Точность модели*, это количественная адекватность. *Постулаты модели*, это то же, что гипотезы.

Формулировка математической модели объекта должна быть продуктом совместной работы технолога (конструктора) и

математика при ведущей на этом этапе роли технолога (конструктора); последнее связано, прежде всего, с тем, что подавляющее большинство прикладных задач по моделированию технологических процессов являются оптимизационными, и постановщик задачи (технолог или конструктор) должен указать требуемое множество переменных технологических параметров, определяющих заданное качество объекта, а также сформулировать обязательные ограничения, связанные с реализацией в производстве поставленной задачи (например, размеры конечного продукта, предельные усилия или скорости резания).

Минор (итальянский *minore*, от латинского *minor* – меньший; также *moll* от латинского *mollis* – мягкий) (муз.), лад, в основе которого лежит малое (минорное) трезвучие; определитель, составленный из элементов, стоящих на пересечении произвольно выделенных k строк и k столбцов данной матрицы или определителя.

Гистограмма (от греческого *histos*, здесь – столб и ...*грамма*) (столбчатая диаграмма), один из видов графического изображения статистических распределений какой-либо величины по количественному признаку. Гистограмма представляет собой совокупность смежных прямоугольников, построенных на одной прямой; площадь каждого из них пропорциональна частоте нахождения данной величины в интервале, на котором построен данный прямоугольник.

Синектика, слово «синектика» (*synectics*, иногда *synactic*) составлено из двух древнегреческих слов: *syn* – вместе и *actics* – действовать, совершать, поступать. Это словосочетание означает «совместно действовать», а применительно к методу решения задач имеет смысл «совместно искать», «вместе находить», «сообща придумывать», «вместе изобретать».

Сепаратриса (от латинского *separabu*) – траектория динамической системы с двумерным фазовым пространством, стремящаяся к седловому состоянию равновесия при времени $t \rightarrow \infty$ (устойчивая сепаратриса) или при $t \rightarrow -\infty$ (неустойчивая сепаратриса). Если устойчивая сепаратриса стремится к седлу при $t \rightarrow \pm \infty$, то её (вместе с седлом) называют петлей сепаратрисы.

3.2. Модель и конструкция печатных изданий

Проектирование начинается с моделирования в соответствии с характером издания и его технической характеристикой. Для специалиста широкого профиля это может быть воображаемая модель, а вообще это физическая модель конкретного или аналогичного изделия.

Модель создается по конструкции изделия, зависящей от формата, объема, характера, назначения и других технических требований. Технологический и производственный процессы проектируются на выпуск изделий определенной конструкции.

При выпуске книжно-журнальной продукции элементами конструкции являются: толщина основных тетрадей (8-, 16-, 32-страничные); способ соединения с основной частью издания отдельно печатаемых цветных иллюстраций (отдельные тетради, наклейки, вкладки, вклейки); способ комплектовки и крепления составных частей (вкладка, подборка, вразъем, втачку, потетрадно, клеевое); оформление (обложка, переплет, суперобложка); форзац (приклепной, пришивной и т. п.).

В настоящее время роль модели повышается. Прогрессивное развитие техники и технологии, интенсификация производства посредством механизации и автоматизации ускоряют производственный процесс. Усложняется управление производством и, в свою очередь, возникает необходимость механизации и автоматизации процесса управления и подготовки производства. Повышается роль оперативного планирования. Из моделей изданий могут быть получены все необходимые данные для планирования и управления производством: физический объем работы на этапах комплексного процесса в натуральных единицах; трудоемкость изготовления продукции на этапах комплексного процесса; необходимое сырье и материалы; количество полуфабриката в заделах, необходимые устройства для его укладки, потребные площади для установки; межцеховые и внутрицеховые грузопотоки и т. п.

При заданном формате и объеме издания основным вопросом конструкции является способ соединения и крепления составных частей изделия, т. е. работа, выполняемая на

последнем этапе изготовления продукции. Конструкция издания должна быть по возможности технологична, т. е. иметь наименьшее число соединяемых деталей, представляющих элементы печатного процесса, без дополнительной брошюровочной обработки.

Конструкцию издания следует ориентировать на поточную систему производства, оснащенную машинами–агрегатами и автоматами; на максимально возможную эффективность производства; на обеспечение заданных параметров качества продукции.

Конструкция должна быть согласована с технологическими возможностями печатных машин, а именно: форматом печати, красочностью машин и т. п. Ее следует ориентировать на простую и надежную технологию, обеспечивающую стабильное качество продукции.

Основная литература: 3 [5–14].

Контрольные вопросы:

1. Математическая модель.
2. Модель печатных изданий.
3. Конструкция печатных изданий.

4. КЛАССИФИКАЦИЯ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

4.1. Различные подходы к классификации

Математические модели классифицируют по ряду признаков.

Доминантный признак. Иногда доминирование выражают величиной органической массы, создаваемой данными видами, или числом особей. *Доминанта* (от латинского *dominans*, родительный падеж *dominantis* – господствующий), господствующая идея, основной признак или важнейшая составная часть чего-либо.

Роль декомпозиции. Слово «декомпозиция» на гуманитарном уровне трактуется как разложение некоторого целого на части, предполагающее дальнейшее соединение этих составных частей в единое целое, т. е. в исходный объект. Точно эквивалентна, если речь идет о точной декомпозиции, и приближенно эквивалентна, если идет речь о приближенной декомпозиции.

Особое значение имеет декомпозиция сложных управляемых процессов. Все реальные системы управления сложными техническими, социально-экономическими, биологическими процессами основаны на некоторой иерархической декомпозиции задачи назначения управляющих воздействий на процесс: каждый орган в системе управления, имеющим возможность принимать решения, ведает своим, в большинстве случаев достаточно четко очерченным, кругом вопросов.

4.2. Функциональные и структурные модели

По характеру отображаемых свойств проектируемого объекта модели делят на функциональные и структурные.

Если же модель отражает только то, как объект функционирует – например, как он реагирует на внешние воздействия, – то она называется *функциональной* или, образно, *черным ящиком*. *Модель черного ящика, системы типа «вход – выход».* «Черный ящик», термин, употребляемый главным

образом в системотехнике для обозначения систем, структура и внутренние процессы которых неизвестны или очень сложны; метод изучения таких систем основан на исследовании их реакций на известные (заданные) входные воздействия (сигналы).

Функциональные модели отображают процессы функционирования объекта. Они чаще всего имеют форму систем уравнений. В зависимости от физической природы отображаемых явлений среди функциональных моделей различают модели тепловые, электрические, оптические, гидравлические, электромеханические, газодинамические, кинематические и др.

Обычно в математической модели отражается структура (устройство) моделируемого объекта, существенные для целей исследования свойства и взаимосвязи компонентов этого объекта; такая модель называется *структурной*. *Структурные модели* отображают только структурные (в частности, геометрические) свойства объекта. Эти модели могут иметь форму матриц, графов, списков векторов и выражать взаимное расположение элементов в пространстве. Они могут характеризовать наличие непосредственных связей между элементами и в виде каналов, проводников, трубопроводов и т. п. Структурные модели обычно используют в случаях, когда задачи структурного синтеза удастся ставить и решать, абстрагируясь от особенностей физических процессов в объекте, например при оформлении конструкторской документации.

Связь структурных и функциональных моделей. Поскольку структурные и функциональные свойства объектов тесно взаимосвязаны, в большинстве проектных процедур требуются модели с отображением особенностей как структуры объекта, так и характера физических или в функциональных моделях, которые по этой причине следует отнести к основному типу моделей в САПР (система автоматизированного проектирования).

Возможны и модели комбинированного типа. *Различные подходы к выбору подсистем.* Основную функциональную нагрузку несет технологическая часть подсистемы, но

остальные подсистемы совершенно необходимы, чтобы эта часть действовала. Элементы подсистем должны обладать такими параметрами, чтобы функционировала с заданными характеристиками, например с заданными скоростью, точностью, чувствительностью и т. д. Все элементы системы находятся в строгих взаимосвязях.

4.3. Дискретные и непрерывные модели

Как известно, величины могут быть двух типов – *дискретные (логические)*, т. е. принимающие «оторванные» друг от друга значения, допускающие естественную нумерацию (с помощью логических отношений), и *непрерывные*, принимающие все значения из некоторого интервала (при непрерывности переменных). Возможен также смешанный случай, например, когда величина на каком-то интервале своих значений ведет себя, как дискретная, а на другом – как непрерывная. (Эти определения не являются исчерпывающими). Дискретная модель принимает «оторванные» друг от друга значения, допускающие естественную нумерацию. Дискретность (от латинского *discretus* – разделённый, прерывистый), прерывность; противопоставляется непрерывности.

Подобным образом, модели – как содержательные, так и математические – могут быть либо дискретными, либо непрерывными, либо смешанными. Между этими типами нет принципиального барьера и при уточнении или видоизменении модели дискретная картина может стать непрерывной и обратно; то же может произойти в процессе решения математической задачи. Таким образом, во многих задачах при составлении математической модели, а также при выборе метода ее исследования надо учитывать возможность применения как «дискретного», так и «непрерывного» аппаратов (например, для дискретных моделей характерно применение сумм, а для непрерывных – производных и интегралов) независимо от характера исходной картины.

4.4. Динамические и статические модели

Различают *статические* и *динамические (эволюционные)* модели; для второго типа моделей предметом изучения является изменение рассматриваемого объекта во времени. Промежуточное место занимают квазистатические, стационарные и квазистационарные модели.

Квазистатическое приближение. В квазистатической модели принимается, что изменение объекта происходит столь медленно, что при рассмотрении ситуации в каждый момент можно в первом приближении объект считать статическим (грубо говоря, пренебречь инерционными силами), а время считать добавочным параметром.

Статические модели. Статистика – выборочная характеристика. Модели, неизменные во времени – статические модели. *Статические модели.* Статические модели отражают работу объекта в стационарных условиях, т. е. когда параметры процесса не меняются во времени. Соответственно математическое описание в статических моделях не включает время как переменную и состоит из алгебраических уравнений либо дифференциальных уравнений в случае объектов с распределенными параметрами. Примером объекта, описываемого статической моделью, служит аппарат полного смешения объемом V в установившемся режиме работы, в который непрерывно подаются реагенты A и B в количестве ν_A , ν_B ($\nu_A + \nu_B = \nu$) и отводится продукт реакции P .

Математическое описание аппарата включает следующие уравнения материального баланса (для простоты тепловой баланс не рассматривается):

$$\nu(C_{A_0} - C_A) = Vkc_AC_B, \quad \nu(C_{B_0} - C_B) = Vkc_AC_B, \quad (4.1)$$

здесь k – константа скорости реакции.

В стационарной модели считается, что процессы происходят, но изучаемый объект во времени не меняется; простейший пример – электрическая цепь с постоянным током. Установившимся процессом обычно называют стационарный или периодический процесс; переходным процессом называют

процесс перехода от одного статического состояния или установившегося процесса к другому.

Динамические модели. Динамическая модель отражает изменение объекта во времени. Математическое описание таких моделей обязательно включает производную по времени. Часто динамическую модель объекта строят в виде передаточных функций, связывающих входные и выходные переменные (представление динамических моделей в виде передаточных функций особенно удобно для целей управления объектом). Примером динамической модели может служить модель рассмотренного выше аппарата полного смешения, но работающего в неустановившемся режиме. В этом случае математическое описание аппарата включает следующие уравнения материального баланса:

$$\frac{dC_A}{dt} = \frac{v}{V}(C_{A_0} - C_A) - kC_A C_B, \quad (4.2)$$

$$\frac{dC_B}{dt} = \frac{v}{V}(C_{B_0} - C_B) - kC_A C_B, \quad (4.3)$$

а также начальные условия

$$C_A = C_{A_0}, C_B = C_{B_0}, \text{ при } t = 0. \quad (4.4)$$

Непрерывные динамические модели. Непрерывные модели принимают все значения из некоторого интервала. Для динамических моделей предметом изучения является изменение рассматриваемого объекта во времени.

4.5. Детерминированные и стохастические модели

Математическая модель может включать случайные компоненты – случайные скалярные или векторные величины, случайные последовательности или функции, случайные структуры и т. п., удовлетворяющие статистическим законам. Такие модели называются *вероятностными* или *стохастическими*, а *детерминированные* (от латинского слова «детермино» – определяю; детерминант от латинского determinans – определяющий, то же, что определитель) модели

таких компонентов не содержат. Между случайными величинами обычно существуют такая связь, при которой с изменением одной величины меняется распределение другой. Такая связь называется *стохастической*. Вероятные модели изучаются с помощью методов теории вероятностей. При построении вероятностных моделей надо уделять существенное внимание источнику вероятностных характеристик случайных компонентов.

4.6. Некоторые возможные математические модели

Так как математические модели являются отражением соответствующих объектов, то для них характерны и такие классы, а именно: *модели с сосредоточенными и существенными параметрами; модели с распределенными параметрами.*

Модели с сосредоточенными параметрами. Для данного класса моделей характерно постоянство переменных в пространстве. Математическое описание включает алгебраические уравнения либо дифференциальные уравнения первого порядка для нестационарных процессов. Примером объекта, описываемого данным классом моделей, может служить аппарат с идеальным (полным) перемешиванием потока. Скорость мешалки такова, что концентрация во всех точках аппарата одинакова (рис. 4.1).

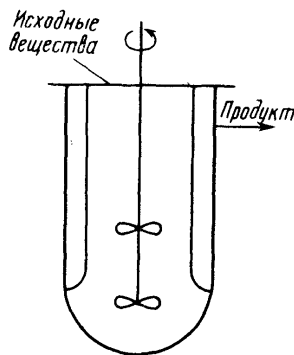


Рис. 4.1. Пример схемы аппарата, реализующего модель идеального смешения

Модели с распределенными параметрами. Если основные переменные процесса изменяются как во времени, так и в пространстве, или если указанные изменения происходят только в пространстве, то модели, описывающие такие процессы, называются моделями с распределенными параметрами. Их математическое описание включает обычно дифференциальные уравнения в частных производных, либо обыкновенные дифференциальные уравнения в случае стационарных процессов с одной пространственной переменной.

Совокупность, имеющая k параметров (или степеней свободы) называется k -мерным многообразием (говорят также – пространством), также многомерным многообразием. *Многообразие состояний* – фазовое состояние. Примером процесса, описываемого такими моделями, служит трубчатый аппарат с большим отношением длины к диаметру и значительной скоростью движения реагентов (рис. 4.2).

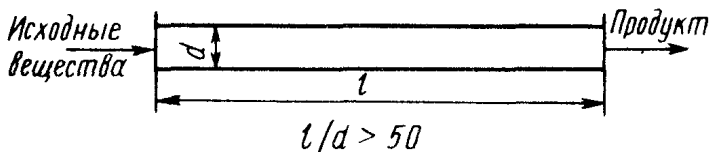


Рис. 4.2. Пример схемы аппарата, реализующего модель идеального вытеснения

По характеру переменных, фигурирующих в модели, различают модели фазовые и факторные. *Фазовые модели* – модели, в которых фигурируют фазовые переменные. Если предмет исследования – процессы функционирования объекта, т. е. переменные в моделях рассматриваются как функции времени, то фазовые модели называют *имитационными моделями*. *Факторная модель* – модель вида $F(Q, Y) = 0$ в явной форме $Y = F(Q)$, где Q – вектор внешних параметров; Y – вектор выходных параметров объекта.

Содержательная модель, строение модели на языке той или иной науки.

Качественные методы, с помощью качественных методов свойства решения изучаются без его построения, путем анализа свойств заданного уравнения.

Качественная модель по необходимости применяется, где количественные закономерности проявляются не вполне четко (например, в некоторых социальных или биологических науках).

Абстрагирование, этот метод научного исследования заключается в отвлечении от несущественных сторон и признаков рассматриваемого объекта.

Евклидово пространство, пространство, свойства которого изучаются в евклидовой геометрии. В более широком смысле евклидовом пространством называется n -мерное векторное пространство, в котором определено скалярное произведение.

Верификация (проверка) модели – контроль правильности модели на основе сравнения результата с другими известными фактами, в частности, с экспериментальными данными и т. д.

Идентификация (от позднелатинского *identifico* – отождествляю), в технике, математике установление соответствия распознаваемого предмета своему образцу (знаку), называемой идентификатором.

Инвариантность, неизменность какой-либо величины при изменении физической условий или по отношению к некоторым преобразованиям, обычно преобразованиям координат и времени при переходе от одной инерциальной системы отчёта к другой; т. е. не зависит от выбора системы координат.

Итерационные методы (от латинского «итерацио» – повторение), в каждом из таких методов некоторая единообразная вычислительная процедура повторяется вновь и вновь, причем каждый раз результат проведенного вычисления кладется в основу последующего вычисления.

Имитация (от латинского *imitatio* – подражание), подражание кому-либо или чему-либо, воспроизведение; подделка.

Существенные параметры, т. е. при любом малом изменении параметров рассматриваемый объект фактически меняется.

Атрибутивным (от латинского *attributio* – свойство, признак) называется высказывание о признаке предмета.

«Блуждающая трубка» как алгоритм имеет характер метода последовательных приближений.

«Веником» называется процедура отыскания минимума аддитивной функции, основанная на последовательном «отметании» неконкурентоспособных вариантов.

«Киевский веник» как алгоритм дает возможность отыскать глобальный экстремум, причем для функций $f_i(x_i, x_{i+1})$ произвольного вида (например, не делалось никаких предположений об их выпуклости и пр.).

Модели, в которых искомые переменные явно выражены через известные величины, называют *аналитическими моделями*. В отличие от них в *алгоритмических моделях* вычисление значений искомых величин связано с необходимостью решения систем уравнений.

Математическая модель является системой уравнений математического описания, отражающей сущность протекающих в объекте явлений, для которой определен алгоритм решения, реализованный в форме моделирующей программы. Согласно этому определению математическая модель должна рассматриваться в совокупности трех ее аспектов: смыслового, аналитического и вычислительного.

Смысловой аспект представляет собой физическое описание природы моделируемого объекта.

Аналитический аспект является математическим описанием процесса в виде некоторой системы уравнений, отражающей протекающие в объекте явления и функциональные связи между ними.

Наконец, вычислительный аспект есть метод и алгоритм решения системы уравнений математического описания, реализованные как моделирующая программа на одном из языков программирования.

4.7. Линейные и нелинейные модели

Линейная зависимость одной величины от другой – это пропорциональность их приращений, т. е. зависимость вида $y = ax + b$, откуда получаем $\Delta y = a\Delta x$ (Δ – обычное обозначение приращения); аналогично, линейная зависимость величины от двух других – это зависимость вида $z = ax + by + c$, откуда $\Delta z = a\Delta x + b\Delta y$, и т. д. Типичные

линейные зависимости между физическими величинами – закон Гука (удлинение пропорционально силе растяжения), закон Ома, закон теплового расширения и т. д.

Понятие линейной модели применяется для моделей объектов, рассматриваемых как преобразователи, для которых каждому входу соответствует некоторый выход. Будем считать, что начала отчета входа и выхода выбраны так, что нулевому входу отвечает нулевой выход. Тогда модель называется *линейной*, если в ней выполнен принцип суперпозиции (наложения) т. е. при сложении входов складываются и выходы, а при умножении входа на любое число выход умножается на то же число. Если этот принцип не выполнен, модель называется *нелинейной*. Линейные модели обычно описываются линейными неоднородными уравнениями – алгебраическими, дифференциальными и т. д., в которых неоднородный член отвечает входу, а решение – выходу.

Метод матричных функций Грина. Грина функция, связанная с аналитическим представлением применяется также при решении краевых задач для обыкновенных дифференциальных уравнений математической физики. Грина функция названа по имени Джорджа Грина, впервые рассмотревшего один её частный случай в своём исследовании по теории потенциала (1828 г.). Грина функция описывает распространение полей от источников и от порождающих (поэтому её называют также функцией распределения).

Выгоды линейности бывают столь велики, что приближенная замена нелинейных соотношений на линейные, нелинейных моделей на линейные, т. е. *линеаризация* соотношений, моделей и т. д. весьма распространена. Такая линеаризация обычно проводится в двух случаях: либо если эксперимент показывает (как, например, для закона Гука), что отклонение от линейности в рассматриваемых диапазонах изменения переменных невелико и несущественно, либо же, если эти диапазоны малы и заменяем приращения переменных на их дифференциалы, отбрасывая члены высшего порядка малости. Во втором случае применяется также линейное интерполирование.

Необходимо иметь в виду, что существуют принципиально нелинейные объекты (в том числе явления), для которых

применение линейных моделей приводит к грубым искажением картины. Это, прежде всего системы, для которых изменение масштаба воздействия приводит к качественному изменению результата. Типичным примером могут служить механические системы с сухим трением, для которых малая сила не порождает движения, а большая – порождает; вообще, наличие любых барьеров подобного рода – это типичный нелинейный эффект. Существенно нелинейный является также задача об изучении околосреднего состояния объекта, зависящего от параметров, когда при их изменении тип движения – другим и т. п. Во всех таких случаях надо применять методы нелинейного анализа, которые можно найти в специальной литературе.

Основная литература: 2[7–12], 4[313–316], 5, 6.

Контрольные вопросы:

1. Доминантный признак и декомпозиция.
2. Функциональные, структурные, дискретные и непрерывные модели.
3. Предельные переходы: континуализация и дискретизация моделей.
4. Динамические и статические модели.
5. Квазистатическое приближение.
6. Детерминированные и стохастические модели.
7. Модели с сосредоточенными и распределенными параметрами.
8. Имитационные, фазовые и факторные модели.
9. Аналитические и алгоритмические модели.
10. Линейные и нелинейные модели.
11. Аналитический, смысловой и вычислительный аспекты.
12. Грина функция.
13. Линеаризация.

5. ЗАДАЧИ ОПТИМИЗАЦИИ. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ

5.1. Задачи оптимизации

В любой сфере человеческой деятельности как на сугубо личном, так и на общегосударственном уровне, явно или неявно мы встречаемся с оптимизацией. Экономическое планирование, управление, проектирование сложных объектов всегда направлено на поиск наилучшего варианта с точки зрения намеченной цели.

При всем многообразии задач оптимизации дать общие методы их решения может только математика, резкое расширение приложений которой связано с появлением ЭВМ, что привело к математизации не только физики, но и химии, биологии, экономики, психологии, медицины – практически всех наук. Суть математизации состоит в построении математических моделей процессов и явлений и в разработке методов их исследования.

Использование математического аппарата при решении задач оптимизации предполагает формулировку интересующей проблемы на языке математики, придание количественных оценок возможным вариантам вместо слов «лучше», «хуже».

Многие задачи оптимизации сводятся к отысканию наименьшего или наибольшего значения некоторой функции, которую принято называть *целевой функцией*.

Такая постановка задачи будет обычной при дальнейшем изложении. В этом случае методы исследования существенно зависят от свойств целевой функции и той информации о ней, которая может считаться доступной до решения задачи и в процессе решения.

Наиболее просты с математической точки зрения случаи, когда целевая функция является дифференцируемой функцией. В этом случае для исследования ее свойств (участки возрастания и убывания, точки локального экстремума) может быть использована производная.

В последние десятилетия в условиях научно-технического прогресса круг задач оптимизации, поставленных практикой,

существенно расширился. Во многих из них значения целевой функции могут получаться в результате численных расчетов или браться из эксперимента. Такие задачи являются более сложными, при их решении нельзя исследовать целевую функцию с помощью производной. Это привело к разработке специальных методов, рассчитанных на широкое применение ЭВМ. Следует также иметь в виду, что сложность решения задачи существенно зависит от размерности целевой функции, т. е. от числа ее аргументов.

Пример 5.1. Указать наилучший вариант консервной банки фиксированного объема V , имеющей обычную форму прямого кругового цилиндра.

Рассмотрим два варианта этой задачи.

1. Наилучшая банка должна иметь наименьшую поверхность S (на изготовление пойдет наименьшее количество жести).

2. Наилучшая банка должна иметь наименьшую длину швов l (швы нужно сваривать, и эта работа должна быть минимальной).

Запишем формулы объема банки, площади поверхности и длины швов:

$$V = \pi r^2 h, \quad S = 2\pi r^2 + 2\pi r h, \quad l = 4\pi r + h.$$

Объем банки задан, это устанавливает связь между радиусом r и высотой h . Выразив высоту через радиус $h = V/\pi r^2$ и подставив полученное выражение в формулы для поверхности и длины швов, получим:

$$S(r) = 2\pi r^2 + \frac{2V}{r}, \quad 0 < r < \infty, \quad (5.1)$$

$$l(r) = 4\pi r + \frac{V}{\pi r^2}, \quad 0 < r < \infty. \quad (5.2)$$

Таким образом, с математической точки зрения задача о наилучшей консервной банке сводится к определению значения r , при котором достигает своего наименьшего значения в одном

случае функция $S(r)$, в другом – $l(r)$. Поскольку $S(r)$ и $l(r)$ дифференцируемы, то задача решается просто.

1. Первый вариант постановки задачи: из формулы (5.1) следует

$$S'(r) = 4\pi r - \frac{2V}{r^2} = \frac{2}{r^2}(2\pi r^3 - V).$$

При $0 < r < r_1 = \sqrt[3]{V/2\pi}$ $S'(r) < 0$, т. е. функция убывает, при $r_1 < r < \infty$ $S'(r) > 0$, т. е. функция возрастает. Следовательно, $S = S_{min}$ при $r = r_1$, где

$$r_1 = \sqrt[3]{\frac{V}{2\pi}}, \quad h_1 = 2r_1, \quad (5.3)$$

а $S(r_1) = 3\sqrt[3]{2\pi V^2} \leq S(r)$. График функции $S(r)$ показан на рис. 5.1.

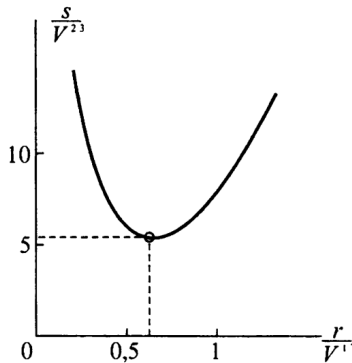


Рис. 5.1. График функции $S(r)$

2. Второй вариант постановки задачи: из формулы (5.2) следует

$$l'(r) = 4\pi - \frac{2V}{\pi r^3} = \frac{2}{\pi r^3}(2\pi^2 r^3 - V).$$

При $0 < r < r_2 = \sqrt[3]{V/2\pi^2}$ $l'(r) < 0$, т. е. функция убывает, при $r_2 < r < \infty$ $l'(r) > 0$ функция возрастает. Следовательно, $l = l_{\min}$ при $r = r_2$, где

$$r_2 = \sqrt[3]{\frac{V}{2\pi^2}}, \quad h_2 = 2r_2, \quad (5.4)$$

при этом $l(r_2) = 3\sqrt[3]{4\pi V} \leq l(r)$. График функции $l(r)$ показан на рис. 5.2.

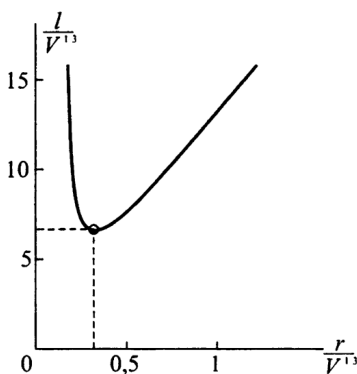


Рис. 5.2. График функции $l(r)$

Очевидно, что при разных критериях оптимизации получаются различные ответы. В первом варианте (5.3) высота наилучшей банки равна ее диаметру, а во втором – (5.4) она в π раз больше него.

Перейдем от конкретного примера к общей постановке задач и определениям.

5.2. Минимум функции одной переменной

Рассмотрим математическую модель оптимизации, в которой целевая функция зависит от одной переменной, определенной на множестве U вещественной оси,

$$\begin{aligned} f(x) &\rightarrow \min, \\ x &\in U \end{aligned} \quad (5.5)$$

Максимизация целевой функции ($f(x) \rightarrow \max$) эквивалентна минимизации противоположной величины ($-f(x) \rightarrow \min$), поэтому, не умаляя общности, будем рассматривать только задачи минимизации.

Определение. Число $x^* \in U$ называется *точкой глобального* (абсолютного) *минимума* или просто *точкой минимума* функции $f(x)$ на множестве U , если $f(x) \leq f(x^*)$ для всех $x \in U$. Множество всех точек минимума $f(x)$ на U будем в дальнейшем обозначать через U^* .

Определение. Число $x \in U$ называется *точкой локального минимума* функции $f(x)$, если $f(\tilde{x}) \leq f(x)$ для всех $x \in U$, достаточно близких к $\tilde{x}$, т. е. если существует $\varepsilon > 0$ такое, что это неравенство выполняется для любого

$$x \in \{x \in U, |x - \tilde{x}| < \varepsilon\}.$$

Определение. Пусть функция $f(x)$ определена и ограничена снизу на множестве U , т. е. $f(x) \geq A > -\infty$ для всех $x \in U$. Число f_* называется *точной нижней гранью* функции $f(x)$ на множестве U ($f_* = \inf_U f(x)$), если $f(x) \geq f_*$ при всех $x \in U$ и для любого $\varepsilon > 0$ найдется точка $x_\varepsilon \in U$ такая, что $f(x_\varepsilon) < f_* + \varepsilon$ (т.е. среди значений $f(x)$ на множестве U найдутся сколь угодно близкие к f_*).

Для ограниченных функций $f(x)$ позволяют $f_* = -\infty$.

Необходимо отметить следующее.

1. Глобальный минимум $f(x)$ является и локальным минимумом, а обратное, вообще говоря, неверно.

2. Множество точек минимума U^* функции $f(x)$ на множестве U может быть пустым, состоять из конечного или бесконечного числа точек.

Например:

а) если $f(x) = \ln x$, $U = [0, 1]$, то $U^* = \emptyset$;

б) если $f(x) = x^2$, $U = [-1, 1]$, то $U^* = \{0\}$ – конечное множество;

в) если $f(x) = \sin^2 nx$, $U = \mathbb{R}$, то $U^* = \mathbb{Z}$ – бесконечное множество;

3. Если $U^* = \emptyset$, то $\inf_U f(x) = \min_U f(x)$. Таким образом, точная нижняя грань обобщает понятие минимума функции на случай $U^* = \emptyset$.

Широкий класс функций, для которых $U^* \neq \emptyset$, определяет известная из математического анализа теорема Вейерштрасса, утверждающая, что непрерывная на замкнутом ограниченном множестве функция достигает на нем своих минимального и максимального значений. Таким образом, задача отыскания минимума (5.5) для непрерывных функций всегда имеет решение.

5.3. Унимодальные функции

Если функция $f(x)$, определенная на множестве U , имеет, кроме глобального, еще и локальные минимумы, отличные от него, то минимизация $f(x)$, как правило, сильно затрудняется. Большинство методов поиска минимума $f(x)$ приспособлено только для тех функций, у которых каждый локальный минимум является одновременно и глобальным. Этим свойством обладают унимодальные функции.

Определение. Функция $f(x)$ называется *унимодальной* на отрезке $[a, b]$, если она непрерывна на $[a, b]$ и существуют числа α и β , $a \leq \alpha \leq \beta \leq b$, такие, что:

1) если $a < \alpha$, то на отрезке $[a, \alpha]$ функция $f(x)$ монотонно убывает;

2) если $\beta < b$, то на отрезке $[\beta, b]$ функция $f(x)$ монотонно возрастает;

3) при $x \in [\alpha, \beta]$ выполняется $f(x) = f_* = \min_{[a, b]} f(x)$.

Отметим, что возможно вырождение в точку одного или двух отрезков из $[a, \alpha]$, $[\alpha, \beta]$ и $[\beta, b]$. Некоторые варианты расположения и вырождения в точку отрезков монотонности и постоянства унимодальной функции показаны на рис. 5.3.

Из определения вытекают следующие свойства унимодальных функций.

1. Любая из точек локального минимума унимодальной функции является и точкой ее глобального минимума на отрезке $[a, b]$.

2. Функция, унимодальная на отрезке $[a, b]$ унимодальна и на любом меньшем отрезке $[c, d] \subset [a, b]$.

3. Пусть $f(x)$ унимодальна на $[a, b]$ и $a \leq x_1 < x_2 \leq b$. Тогда:

если $f(x) \leq f(x)$, то $x^* \in [a, x_2]$,

если $f(x) > f(x)$, то $x^* \in [x_1, b]$,

где x^* — одна из точек минимума $f(x)$ на отрезке $[a, b]$.

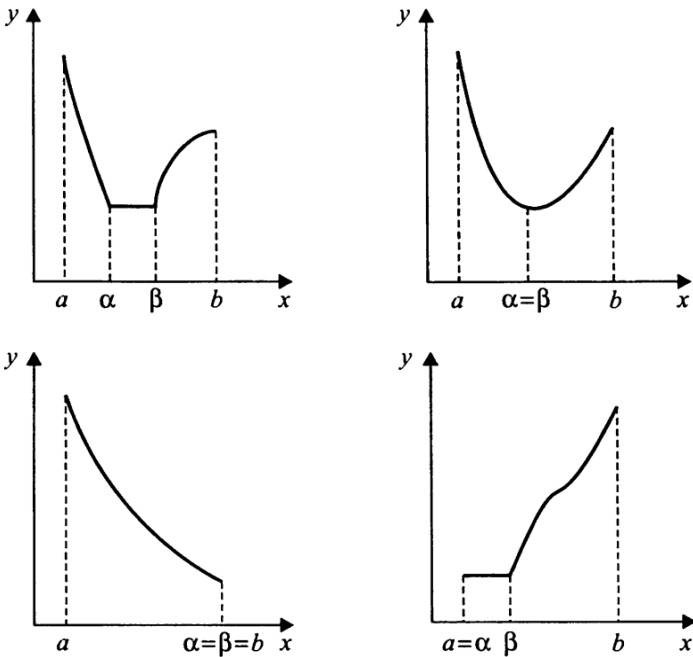


Рис. 5.3. Графики унимодальных функций

Пример 5.2. На какие три части следует разбить отрезок $[-1, 2]$, чтобы на каждой из них функция $f(x) = ||x(x-1)|-1|$ была унимодальной?

Строя поэтапно график функции $f(x)$, получим три отрезка: $[-1, 0]$, $[0, 1]$ и $[1, 2]$, на каждом из которых функция унимодальна.

Пример 5.3. Найти максимальное значение b , при котором функция $f(x) = -x^2 + 5x - 6$ унимодальна на отрезке $[-5, b]$.

Корни квадратного трехчлена $x_1 = 2$, $x_2 = 3$; вершина параболы $x_b = 2,5$. Поэтому $b = 2,5$.

5.4. Выпуклые функции

Функция $f(x)$, заданная на отрезке $[a, b]$ называется *выпуклой* на этом отрезке, если для всех $x', x'' \in [a, b]$ и произвольного числа $\alpha \in [0, 1]$ выполняется неравенство

$$f(\alpha x' + (1 - \alpha) x'') \leq \alpha f(x') + (1 - \alpha) f(x''). \quad (5.6)$$

Свойство 1. Если функция $f(x)$ выпукла на $[a, b]$, то на любом $[x', x''] \subset [a, b]$ ее график расположен не выше хорды, проведенной через точки графика с абсциссами x' и x'' (см. рис. 5.4).

Свойство 2. Можно показать, что всякая выпуклая непрерывная функция на $[a, b]$ является унимодальной на этом отрезке. Обратное, вообще говоря, неверно.

Из курса математического анализа известны следующие условия выпуклости функции:

а) для того чтобы дифференцируемая на отрезке $[a, b]$ функция $f(x)$ была выпуклой на этом отрезке, необходимо и достаточно, чтобы ее производная $f'(x)$ не убывала на $[a, b]$

(например: $y = x^2$, $x \in R$; $y = \cos x$, $x \in [\frac{\pi}{2}, \frac{3\pi}{2}]$);

б) для того чтобы дважды дифференцируемая на отрезке $[a, b]$ функция $f(x)$ была выпуклой на этом отрезке, необходимо и достаточно, чтобы при всех $x \in [a, b]$ выполнялось

неравенство $f''(x) \geq 0$ ($y = x^2$, $x \in R$; $y = \cos x$, $x \in [\frac{\pi}{2}, \frac{3\pi}{2}]$).

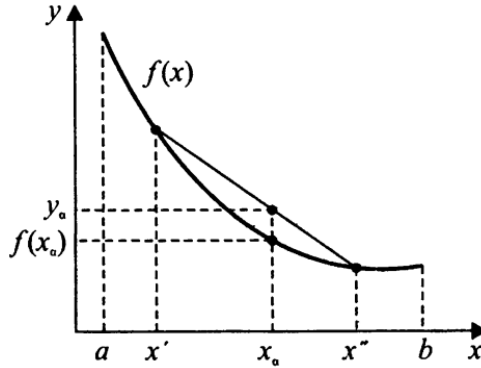


Рис. 5.4. Взаимное расположение графика выпуклой функции и хорды

При исследовании выпуклости функций на практике неравенство (5.6) удается использовать только в редких случаях. Поэтому для дифференцируемых достаточно количество раз функций обычно применяют вышеприведенные дифференциальные критерии выпуклости.

Непосредственная проверка унимодальности с помощью определения этого понятия также в большинстве случаев вызывает затруднения. Поэтому для обоснования унимодальности достаточно гладких функций используют те же критерии выпуклости. Если функция оказывается выпуклой, то можно утверждать, что она унимодальна. Однако при отрицательном результате проверки функции на выпуклость нельзя сделать вывод, что она не унимодальна.

Пример 5.4. Показать, что функция $f(x) = x^4 - 10x^3 + 36x^2 + 5x$ унимодальна на отрезке $[3, 5]$.

Вторая производная функции $f(x)$ равна $f''(x) = 12x^2 - 60x + 72$. Корни полученного квадратного трехчлена $x_1 = 2$ и $x^2 = 3$. Следовательно, $f''(x) \geq 0$, если $x \geq 3$, в частности при $x \in [3, 5]$. По дифференциальному критерию «б» получаем, что $f(x)$ выпукла, а значит, и унимодальна на этом отрезке.

Пример 5.5. В следующих задачах убедиться в унимодальности функции $f(x)$ на указанных отрезках $[a, b]$

- 1) $f(x) = x^2 - 3x + x \ln x$, $x \in [1, 2]$.

2) $f''(x) = 2 + \frac{1}{x} > 0$ при $x \in [1, 2]$. По дифференциальному критерию «б» $f(x)$ выпуклая, а значит, и унимодальная при $x \in [1, 2]$.

$$3) f(x) = \ln(1 + x^2) - \sin x, x \in [0, \frac{\pi}{4}].$$

$$f''(x) = \frac{2(1 - x^2)}{(1 + x^2)^2} + \sin x > 0 \quad \text{при} \quad x \in [0, \frac{\pi}{4}]. \quad \text{По}$$

дифференциальному критерию

«б» $f(x)$ выпуклая, а значит, и унимодальная при $x \in [0, \frac{\pi}{4}]$.

$$3) f(x) = \frac{1}{2}x^2 - \sin x, x \in [0, 1].$$

$f''(x) = 1 + \sin x \geq 0$ при $x \in [0, 1]$. По дифференциальному критерию «б» $f(x)$ выпуклая, а значит, и унимодальная при $x \in [0, 1]$.

Пример 5.6. Привести примеры функций $f(x)$, унимодальных на отрезке $[a, b]$ но невыпуклых на нем.

$$f(x) = -\frac{1}{1 + x^2}, x \in \mathbb{R} \text{ или } x \in [a, b],$$

где $[a, b]$ – такой отрезок, что $0 \in [a, b]$.

Очевидно, что функция унимодальна на $[a, b]$, но $f'(x) = \frac{2x}{(1 + x^2)^2}$, вторая производная $f''(x) = \frac{2(1 - 3x^2)}{(1 + x^2)^3} < 0$, например

при $|x| > \frac{1}{\sqrt{3}}$. Функция $f(x)$ не является выпуклой по

дифференциальному критерию выпуклости «б».

$f(x) = |\ln x|, x \in (0, +\infty)$. Очевидно, что функция унимодальна при $x \in (0, +\infty)$, но не является выпуклой при $x \in (1, +\infty)$, т. к.

на этом интервале $f''(x) = -\frac{1}{x^2} < 0$.

Пример 5.7. Будет ли функция $f(x) = ax^3 - 3x^2 - 10$ унимодальной на отрезке $[1, 2]$ при $a > 3$?

Первая производная функции $f'(x) = 3ax^2 - 6x$, вторая $-f''(x) = 6a(x - \frac{1}{a})$. Если $x \in [1, 2]$, а величина $a > 3$, т. е. $0 < \frac{1}{a} < \frac{1}{3}$, то $f''(x) > 0$ и функция $f(x)$ унимодальна по дифференциальному критерию «б» выпуклости функций.

Пример 5.8. Доказать, что из выпуклости функции $f(x)$ на отрезке $[a, b]$ следует ее унимодальность на $[a, b]$, если ограничиваться только дифференцируемыми на $[a, b]$ функциями.

Пусть $f(x)$ выпукла на $[a, b]$, тогда $f(x)$ не убывает на этом отрезке. Допустим противное, т. е. что $f(x)$ не унимодальна на $[a, b]$. Тогда существуют $x_1, x_2, x_3 \in [a, b]$ такие, что $f(x_1) < f(x_2)$ и $f(x_3) < f(x_2)$. А это противоречит дифференциальному критерию «а» выпуклости $f(x)$ на $[a, b]$.

5.5. Условие Липшица

Применение некоторых методов одномерной минимизации возможно только в том случае, если скорость изменения целевой функции $f(x)$ на любых участках отрезка $[a, b]$ ограничена некоторым числом, одним и тем же для всех участков. В этом случае говорят, что $f(x)$ удовлетворяет на отрезке $[a, b]$ *условию Липшица*. Целевые функции большинства практических задач оптимизации таким свойством обладают.

Определение. Функция $f(x)$ удовлетворяет на отрезке $[a, b]$ условию Липшица, если существует такое число $L > 0$ (константа Липшица), что

$$|f(x') - f(x'')| \leq L|x' - x''| \quad (5.7)$$

для всех x' и x'' , принадлежащих $[a, b]$.

Необходимо обратить внимание на следующее.

1. Если неравенство (5.7) выполняется с константой L , то оно справедливо и для всех $L' > L$. Поэтому для функции, удовлетворяющей условию Липшица, существует бесконечное множество констант L из (5.7). При использовании алгоритмов минимизации, включающих L как параметр, наилучшие

результаты достигаются, как правило, если в качестве L берется минимальная из констант Липшица.

2. Из условия (5.7) сразу следует непрерывность $f(x)$ на отрезке $[a, b]$.

Поэтому согласно теореме Вейерштрасса функция $f(x)$, удовлетворяющая на отрезке $[a, b]$ условию Липшица, имеет на нем хотя бы одну точку минимума, хотя не является, вообще говоря, унимодальной.

3. Условие (1.7) означает, что модуль углового коэффициента любой хорды графика $f(x)$ не превосходит L . Переходя в (5.7) к пределу при $|x' - x''| \rightarrow 0$, убеждаемся, что если в некоторой точке существует касательная к графику $f(x)$, то модуль ее углового коэффициента также не может превышать L . Так, функция $f(x) = \sqrt{x}$ на отрезке $[0, 1]$ условию Липшица не удовлетворяет, потому что при $x \rightarrow +0$ угловой коэффициент касательной к ее графику k неограниченно возрастает (рис. 5.5).

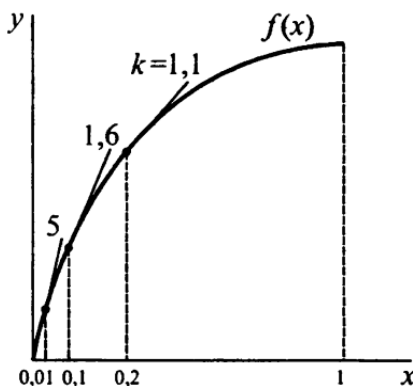


Рис. 5.5. График функции $f(x) = \sqrt{x}$, $x \in (0, 1)$, не удовлетворяющей условию Липшица

4. Если функция $f(x)$ имеет на отрезке $[a, b]$ непрерывную производную, то она удовлетворяет на этом отрезке условию Липшица с константой $L = \max_{[a,b]} |f'(x)|$.

Пример 5.9. Найти наименьшую из констант Липшица функции $f(x) = \frac{1}{3}x^3 + 2x^2 - 5x + 6$ на отрезках: а) $x \in [0, 1]$, б) $x \in [0, 10]$.

Производная функции $f'(x) = x^2 + 4x - 5 = (x + 5)(x - 1)$, поэтому в случае «а» $L = \max_{[0,1]} |f'(x)| = |f'(0)| = 5$, а в случае «б» $L = \max_{[0,10]} |f'(x)| = |f'(10)| = 135$.

Пример 5.10. Показать, что для дифференцируемой на отрезке $[a, b]$ функции $f(x)$ величина $L = \max_{[a,b]} |f'(x)|$ представляет собой $[a, b]$ минимальную из констант Липшица $f(x)$ на $[a, b]$.

Пусть существует $L_1 < L = \max_{[a,b]} |f'(x)|$ такая, что

$$|f(x') - f(x'')| \leq L_1 |x' - x''| \text{ для всех } x', x'' \in [a, b].$$

Тогда, фиксируя x' и переходя к пределу при $x'' \rightarrow x'$ получим $|f'(x')| \leq L_1$, а вследствие произвольности точки $x' \in [a, b]$ получим $\max_{[a,b]} |f'(x)| \leq L_1 < L$. Полученное противоречие доказывает утверждение.

Пример 5.11. Найти наименьшую из констант Липшица функции $f(x) = 4x^3 - 30x^2 + 72x + 12$ на отрезках: а) $x \in [0, 2]$, б) $x \in [2, 3]$.

Производная функции $f'(x) = 12x^2 - 60x + 72 = 12(x - 2)(x - 3)$, поэтому в случае «а» $L = \max_{[0,2]} |f'(x)| = |f'(0)| = 72$, а в случае «б»

$$L = \max_{[2,3]} |f'(x)| = |f'(2, 5)| = 3.$$

Пример 5.12. Найти наименьшую из констант Липшица функции $f(x) = x^3 + 6x^2 - 15x$ на отрезках: а) $x \in [0, 1]$, б) $x \in [0, 10]$.

Производная функции $f'(x) = 3x^2 + 12x - 15 = 3(x + 5)(x - 1)$, поэтому в случае «а» $L = \max_{[0,1]} |f'(x)| = |f'(0)| = 15$, а в случае «б»

$$L = \max_{[0,10]} |f'(x)| = |f'(0)| = 405.$$

5.6. Классическая минимизация функции одной переменной

Из математического анализа известны следующие условия локального экстремума функции $f(x)$, дифференцируемой достаточное количество раз.

1. Если функция $f(x)$ дифференцируема в точке x и достигает в ней локального экстремума, то $f'(\tilde{x}) = 0$ (необходимое условие экстремума).

2. Пусть функция $f(x)$ n раз дифференцируема в точке $\tilde{x}$ и в этой точке все производные $f(x)$ до $(n - 1)$ -го порядка включительно равны нулю, а $f^{(n)}(\tilde{x}) \neq 0$. Тогда, если n нечетно, то точка $\tilde{x}$ не является точкой локального экстремума функции $f(x)$. Если же n – четное число, то:

а) при $f^{(n)}(\tilde{x}) > 0$ $\tilde{x}$ – точка локального минимума $f(x)$;

б) при $f^{(n)}(\tilde{x}) < 0$ $\tilde{x}$ – точка локального максимума $f(x)$ (достаточные условия экстремума).

Перечисленные условия позволяют предложить следующий путь решения задачи минимизации (5.5).

Шаг 1. С помощью условия 1 находим все точки возможного экстремума функции $f(x)$ на интервале (a, b) , т. е. корни уравнения

$$f'(x) = 0, \quad (5.8)$$

(стационарные точки функции $f(x)$, принадлежащие интервалу (a, b)).

Шаг 2. Найденные стационарные точки исследуем в соответствии с условием 2, выделяя из них только точки локальных минимумов $f(x)$.

Шаг 3. Значения $f(x)$ в точках локальных минимумов и на концах отрезка $[a, b]$ сравниваем между собой. Наименьшему

из этих значений соответствует точка глобального минимума $f(x)$ на $[a, b]$.

Применение условия 2 требует вычисления высших производных функции $f(x)$, поэтому в большинстве случаев бывает проще сравнить значения $f(x)$ во всех стационарных точках, не интересуясь их характером. С учетом этого можно предложить следующий алгоритм минимизации $f(x)$ на отрезке $[a, b]$ *классический метод*, который разберем на примере).

Пример 5.13. Решить задачу $f(x) = x^3 - 3x + 1 \rightarrow \min, x \in [-2, 2]$.

Шаг 1. Находим корни уравнения $f'(x) = 3x^2 - 3 = 0$ из интервала $(-2, 2)$: $x_1 = -1, x_2 = 1$. Полагаем $x_0 = -2, x_3 = 2$.

Шаг 2. Вычисляем значения $f(x)$ в точках $x_i, i = 0, \dots, 3$: $f(x_0) = -17, f(x_1) = 3, f(x_2) = -1, f(x_3) = 1$.

Шаг 3. Находим $f^* = \min(-17, 3, -1, 1) = -17 = f(x_0)$. Поэтому

$$x^* = -2, f^* = -17.$$

Пример 1.14. Найти $f(x) = x^3 - 27x + 5 \rightarrow \min, x \in [-4, 4]$.

Найдем корни уравнения $f'(x) = 3x^2 - 27 = 0$ из интервала $x \in [-4, 4]$: $x_1 = -3, x_2 = 3$. Положим $x_0 = -4, x_3 = 4$.

Так как $f''(x_1) = f''(-3) = -18 < 0$, то x_1 — точка локального максимума. Так как $f''(x_2) = f''(3) = 18 > 0$, то x_2 — точка локального минимума. $f(3) = -49, f(-4) = 49, f(4) = -39$. Производя перебор, получим

$$x^* = 3, f^* = -49.$$

Пример 5.15. Решить задачу

$$f(x) = \sin 2\pi x - 2\pi \cdot \exp(x - 1) \rightarrow \min, x \in \left[\frac{1}{2}, \frac{3}{2}\right].$$

Найдем корни уравнения $f'(x) = 2\pi \cos 2\pi x - 2\pi \cdot \exp(x - 1) = 0$ из промежутка $x \in \left[\frac{1}{2}, \frac{3}{2}\right]$. Точка $x_1 = 1$ является корнем (показать, что других корней на промежутке нет). Далее, полагая $x_0 = \frac{1}{2}, x_2 = \frac{3}{2}$, вычислим

$$f(x_0) = -2\pi \cdot \exp(-\frac{1}{2}), f(x_1) = -2\pi, f(x_1) = -2\pi \cdot \exp(\frac{1}{2}).$$

Производя перебор, получим $x^* = x_2 = \frac{3}{2}$,

$$x^* = f(x_2) = -2\pi \cdot \exp(\frac{1}{2}).$$

Пример 5.16. Найти $f(x) = -\sin \frac{x}{2} - \cos \frac{x}{2} \rightarrow \min, x \in [\frac{\pi}{3}, \frac{2\pi}{3}]$.

Найдем корни уравнения

$$f'(x) = -\frac{1}{2} \cos \frac{x}{2} + \frac{1}{2} \sin \frac{x}{2} = 0, x \in [\frac{\pi}{3}, \frac{2\pi}{3}].$$

Точка $x_1 = \frac{\pi}{2}$ является корнем (показать, что других корней

на промежутке нет). Положим $x_0 = \frac{\pi}{3}, x_2 = \frac{2\pi}{3}$. Производя

перебор, $f(x_0) = -\frac{1}{2} - \frac{\sqrt{3}}{2}, f(x_1) = -\sqrt{2}, f(x_2) = -\frac{1}{2} - \frac{\sqrt{3}}{2}$.

Поскольку $1 + \sqrt{3} < 2\sqrt{2}$, получим $x^* = x_1 = \frac{\pi}{2}, f^* = f(x_1) = -\sqrt{2}$.

Основная литература: 7[6–19].

Контрольные вопросы:

1. Какая функция называется целевой?
2. Дать определение локального и глобального минимумов функции.
3. Что такое точная нижняя грань функции на множестве?
4. Как соотносятся точная нижняя грань и минимум функции на множестве?
5. Какая функция называется унимодальной на отрезке $[a, b]$?
6. Сформулировать свойства унимодальных функций.
7. Какая функция называется выпуклой на отрезке $[a, b]$?
8. Каков геометрический смысл выпуклости функции?

9. Сформулировать два необходимых и достаточных дифференциальных условий выпуклости функций.

10. Сформулировать условие Липшица для функции $f(x)$ на отрезке $[a, b]$.

11. Всякая ли унимодальная на отрезке $[a, b]$ функция $f(x)$ удовлетворяет на нем условию Липшица?

12. Всякая ли функция $f(x)$, удовлетворяющая условию Липшица на отрезке $[a, b]$, унимодальна на нем?

13. Сформулировать свойства функций, удовлетворяющих на отрезке $[a, b]$ условию Липшица.

14. В чем заключается классический метод минимизации функций?

15. Для каких целей разработан классический метод минимизации функций?

16. Какова практическая ограниченность применимости этого метода?

6. ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

В современном мире возник целый ряд разделов математики, в название которых входит слово программирование: *линейное программирование, нелинейное программирование, динамическое программирование, стохастическое программирование* и т. д. Совокупность этих дисциплин часто объединяется единым термином *математическое программирование*. Одновременно под программированием понимают кодирование математических алгоритмов на машинном языке. Термины динамическое программирование, линейное программирование и т. д. стали общепринятыми и поэтому будут сохранены.

Задачи линейного программирования были первыми подробно изученными задачами оптимизации при наличии ограничений типа неравенств. Термин «линейное программирование» возник в результате неточного перевода английского «*linear programming*». Одно из значений слова «*programming*» – составление планов, планирование. Следовательно, правильным переводом этого термина было бы не «линейное программирование», а «линейное планирование», «планирование на основе линейных соотношений», что более точно отражает содержание дисциплины.

6.1. О постановках задачи линейного программирования и ее приложениях

1. *Основные определения.* В этой теме мы будем рассматривать задачу максимизации линейной функции

$$f(x) = c^1 x^1 + c^2 x^2 + \dots + c^n x^n = (c, x).$$

Эта функция неограничена, и поэтому искать ее максимум, не налагая никаких ограничений на область изменения вектора x , бессмысленно. Интерес представляет задача максимизации $f(x)$ при условии, что x принадлежит некоторому множеству. Те из соответствующих задач, в которых область изменения x – многогранник, т. е. выбор значений вектора x стеснен условиями

типа линейных равенств и неравенств, составляют предмет линейного программирования.

В общем случае задача линейного программирования формулируется следующим образом: найти величины $x^1, \dots, x^n$, доставляющие максимум (минимум) линейной функции

$$f(x) = (c, x) = c^1 x^1 + c^2 x^2 + \dots + c^n x^n.$$

на множестве значений $x^1, \dots, x^n$, удовлетворяющих ограничениям, в числе которых могут присутствовать только равенства и неравенства вида

$$\begin{aligned} a_{i1}x^1 + \dots + a_{in}x^n &\leq b^i, \\ a_{k1}x^1 + \dots + a_{kn}x^n &= b^k, \\ a_{l1}x^1 + \dots + a_{ln}x^n &\geq b^l. \end{aligned}$$

Среди ограничений задач линейного программирования часто встречаются условия неотрицательности всех или части переменных:

$$x^j \geq 0, \quad j=1, \dots, p.$$

Хотя формально эти условия являются частным случаем представленных выше условий общего вида, на практике, при построении алгоритмов, их обычно выделяют в особую группу. Целевую функцию $f(x) = (c, x)$ принято называть *критерием* задачи линейного программирования (иногда ее называют также функционалом задачи линейного программирования). Эти термины отражают природу тех экономических задач, которые послужили источником излагаемой в этой теме математической теории. С некоторыми примерами подобных задач мы сейчас познакомимся.

2. *Примеры прикладных задач линейного программирования. Задача о выборе оптимального плана.* Допустим, что для создания некоторого продукта требуется m видов ресурсов и можно использовать n способов (технологий) производства. Обозначим через a_{ij} расход ресурса номера i при единичной интенсивности использования технологии номера j ; эту интенсивность можно измерять, например, долей суток, в течение которой эксплуатируется данная технология. Одним из ресурсов в таком случае будет время, а через c^j – количество

производимой при этом продукции, и будем считать, что зависимость расходов и выпусков от интенсивностей линейна. Предположим, далее, что производству выделено b^i единиц i -го ресурса, и обозначим через x^j интенсивность использования j -й технологии. Тогда под оптимизацией плана можно понимать поиск максимума объема выпуска продукции, равного

$$\sum_{j=1}^n c^j x^j,$$

при заданных расходах ресурсов:

$$\begin{array}{l} a_{11}x^1 + \dots + a_{1n}x^n = b^1, \\ \vdots \\ a_{m1}x^1 + \dots + a_{mn}x^n = b^m. \end{array}$$

Интенсивности использования технологий по смыслу неотрицательны, т. е. переменные x^j должны подчиняться еще и ограничениям

$$x^1 \geq 0, \dots, x^n \geq 0.$$

Таким образом, задача оптимального распределения ресурса по заданным способам производства может быть сформулирована как задача линейного программирования.

Задача о рационе. Предположим, что нужно выбрать суточный рацион питания, имея в распоряжении n продуктов, содержащих m питательных веществ.

Пусть a_{ij} – количество j -го питательного вещества в единице i -го продукта, b^j – суточная потребность организма в j -м питательном веществе, c^i – цена единицы i -го продукта. Если обозначить суточное потребление i -го продукта через x^i стоимость рациона будет равна $c^1x^1 + \dots + c^nx^n$. Естественно поставить задачу минимизации при условии полного удовлетворения потребности организма в питательных веществах:

$$\begin{array}{l} a_{11}x^1 + \dots + a_{1n}x^n \geq b^1, \\ \vdots \\ a_{m1}x^1 + \dots + a_{mn}x^n \geq b^m. \end{array}$$

Потребление i -го продукта должно быть, разумеется, неотрицательно:

$$x^i \geq 0, \quad i = 1, \dots, n.$$

Конечно, постановка такой задачи не учитывает многих важных факторов, и ее решение часто приводит к парадоксам. Например, рацион коровы, если не учитывать емкость ее желудка, должен состоять из одной соломы. Поэтому формулировка задачи о рационе для того, чтобы ее решение представляло практический интерес, должна быть уточнена.

Транспортная задача. Транспортная задача является одной из самых распространенных задач линейного программирования специального вида. Предположим, что имеется m кладов, где хранится некоторый продукт в количествах $a_1, \dots, a_m$, и n пунктов реализации этого продукта, потребности которых равны $b^1, \dots, b^n$. Требуется найти наиболее экономичный способ перевозки продукта со складов к потребителям, если затраты на перевозку единицы продукта с i -го склада в j -й пункт потребления равны c_{ij} . Обозначим через x_{ij} количество продукта, перевозимое с i -го склада в j -й пункт. Тогда задача минимизации транспортных расходов формулируется так: найти

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

при ограничениях

$$\sum_{j=1}^n x_{ij} \leq a_i, \quad i = 1, \dots, m,$$

$$\sum_{i=1}^m x_{ij} \geq b^j, \quad j = 1, \dots, n,$$

$$x_{ij} \geq 0, \quad i = 1, \dots, m, \quad j = 1, \dots, n.$$

Первое неравенство означает, что количество продукта, вывезенного со склада, не должно превышать запаса, второе – что потребности каждого пункта должны быть удовлетворены. Третье условие обеспечивает неотрицательность планируемых объемов перевозок.

Мы получили задачу линейного программирования, называемую транспортной и отличающуюся от рассмотренных ранее двойной индексацией переменных. Специфика ограничений этой задачи позволила разработать весьма эффективные алгоритмы ее решения. Этим оправданы значительные усилия, которые нередко тратят математики на преобразование с помощью всевозможных искусственных приемов той или иной задачи в транспортную.

3. Стандартные формы представления задач линейного программирования. Существует немало различных вариантов формулировок задач линейного программирования, но наиболее употребительны две формы их записи. И к той, и к другой любую задачу можно привести с помощью простых приемов. Некоторые из таких приемов перечислены ниже.

Переход от задачи минимизации к задаче максимизации. Для этого, нужно изменить знак критериальной функции, т. е. задача минимизации функции

$$f(x) = (c, x)$$

эквивалентна задаче на поиск максимума функции

$$f(x) = -(c, x).$$

Переход к эквивалентной системе неравенств. Меняя знаки свободного члена и коэффициентов в ограничении – неравенстве, можно поменять знак этого неравенства на обратный. Например, ограничение

$$d^1 x^1 + d^1 x^1 + \dots + d^n x^n \geq d$$

можно заменить следующим:

$$(-d^1)x^1 + \dots (-d^n)x^n \leq -d.$$

Переход от ограничения – неравенства к равенству. Любое ограничение в форме неравенства введением дополнительной неотрицательной переменной может превратиться в ограничение – равенство. Так, к примеру, условие

$$a^1x^1 + \dots + a^nx^n \leq a$$

эквивалентно двум ограничениям

$$a^1x^1 + \dots + a^nx^n + x^{n+1} = a, \quad x^{n+1} \geq 0.$$

Переменные типа x^{n+1} называют *фиктивными* или *дополнительными*.

Представление ограничения – равенства парой неравенств.
Ограничение

$$b^1x^1 + \dots + b^nx^n = b$$

можно представить парой ограничений

$$b^1x^1 + \dots + b^nx^n \leq b, \quad (-b^1)x^1 + \dots + (-b^n)x^n \leq -b.$$

Переход к неотрицательным переменным. Если на знак переменной x^j не наложено ограничений, можно заменить ее разностью двух неотрицательных переменных

$$x^j = v^j - w^j, \quad v^j \geq 0, \quad w^j \geq 0.$$

Переход от переменных, ограниченных снизу, к неотрицательным переменным. Пусть переменная ограничена снизу:

$$x^j \geq b^j.$$

Заменив ее по формуле

$$x^j = y^j + b^j,$$

перейдем к задаче, в которой фигурирует неотрицательная переменная

$$y^j \geq 0.$$

Используя перечисленные и подобные им несложные приемы, можно свести произвольную задачу линейного программирования к любой из двух общеупотребительных форм: канонической и с однотипными условиями.

Каноническая форма задачи линейного программирования. Будем говорить, что задача линейного программирования записана в канонической форме, если она формулируется следующим образом:

$$\max (c^1 x^1 + \dots + c^n x^n) \quad (6.1.1)$$

при ограничениях

$$\begin{aligned} & a_{1l}x^1 + \dots + a_{ln}x^n = b^l, \\ & \vdots \\ & a_{ml}x^1 + \dots + a_{mn}x^n = b^m. \end{aligned} \tag{6.1.2}$$

$$x^i \geq 0, \quad i = 1, 2, \dots, n. \quad (6.1.3)$$

В матричной форме эта задача записывается так:

$$\max (c, x), \quad (6.1.1')$$

$$Ax = b, \quad (6.1.2')$$

$$x \geq 0. \quad (6.1.3')$$

Матрица (математическая), прямоугольная таблица каких-либо элементов a_{ik} (чисел, математических выражений), состоящая из m строк и n столбцов. Здесь $A = \|a_{ij}\|$ есть $(m \times n)$ -матрица условий, ее столбцы $\{a_{1j}, \dots, a_{mj}\}^T, j = 1, 2, \dots, n$, называются векторами условий, вектор $b = \{b^1, \dots, b^m\}^T$ – вектором правых частей, а $\{c^1, \dots, c^n\}$ – вектором коэффициентов линейной формы. Далее всегда будем считать, что ранг матрицы A равен числу ее строк. Если это не так и задача, тем не менее, разрешима, некоторые из уравнений (6.1.2) должны быть линейными комбинациями остальных. Они «лишние» и их следовало бы отбросить. Точку x , удовлетворяющую ограничениям (6.1.2), (6.1.3), называют допустимым решением. Допустимое решение, на котором линейная форма (6.1.1) принимает максимальное значение, называется оптимальным решением или просто решением задачи (6.1.1)–(6.1.3).

Пересечение указанных полупространств является многогранником, который и будет областью допустимых решений задачи. Последний впредь будем обозначать через X .

Линии уровня минимизируемой функции

$$c^1x^1 + \dots + c^nx^n = const \quad (6.2.1)$$

образуют семейство параллельных плоскостей. Вектор нормали к этим плоскостям

$$c = \{c^1, \dots, c^n\}^T$$

определяет направление возрастания линейной формы (рис. 6.1).

Выберем из семейства (6.2.1) любую плоскость, пересекающую многогранник допустимых векторов X , и будем смещать ее в направлении c до такого предельного положения, когда многогранник X окажется в одном из полупространств, порожденных нашей плоскостью, и хотя бы одна точка из X все еще будет ей принадлежать. Полученная предельная плоскость называется

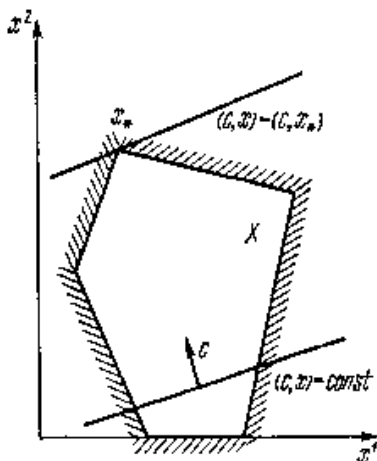


Рис. 6.1

опорной для многогранника X , а его точки, лежащие в этой плоскости, будут решениями задачи.

Заметим, что множество точек, удовлетворяющих неравенствам (6.1.5), может быть пустым, ограниченным и неограниченным. В первом случае (рис. 6.2) задача не имеет решения. Во втором случае она заведомо разрешима и имеет либо единственное решение (рис. 6.1), совпадающее с одной из вершин допустимого многогранника, либо бесконечное множество решений (рис. 6.3) – ребро или грань многогранника, параллельные плоскостям семейства (6.2.1).

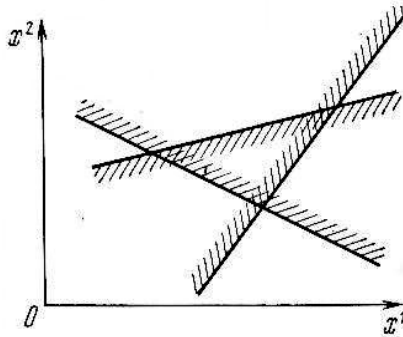


Рис. 6.2

Если допустимое множество задачи линейного программирования неограничено, ответ на вопрос о существовании ее решения зависит от того, ограничена сверху на этом множестве целевая функция или нет. Если ограничена – задача разрешима, причем возможны те же ситуации, что и во втором из рассмотренных выше случаев. Если нет – решение в обычном понимании отсутствует (рис. 6.4).

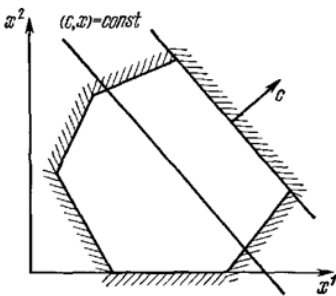


Рис. 6.3

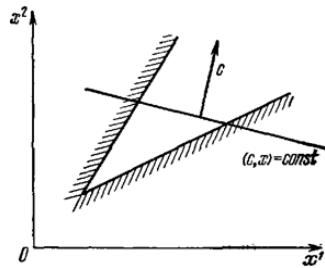


Рис. 6.4

На рис. 6.5 это множество показано для случая, когда $m = 2$, т. е. задача имеет два ограничения типа равенства.

Неотрицательные линейные комбинации расширенных векторов-условий образуют многогранный конус K , ребрами которого будут векторы $\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_5$. Проекция их на плоскость $u^0 = 0$ есть, соответственно, $a_1, \dots, a_5$. Длина перпендикуляра, соединяющего конец вектора $\tilde{a}_j$ с концом его проекции a_j , равна коэффициенту линейной формы c^j .

При произвольном значении x^0 вектор $\tilde{b}$ указывает в некоторую точку, лежащую на вертикальной прямой Q (см. рис. 2.5), проходящей через конец вектора

$$b = \{b^1, \dots, b^m\}^T$$

на плоскости $u^0 = 0$ (точку с координатами $\{0, b^1, \dots, b^m\}^T$). Если же x^0 — компонента какого-нибудь допустимого решения задачи (6.2.2)–(6.2.4), вектор $\tilde{b}$ будет принадлежать также конусу K . Ясно и то, что если вектор $\tilde{b}$ указывает

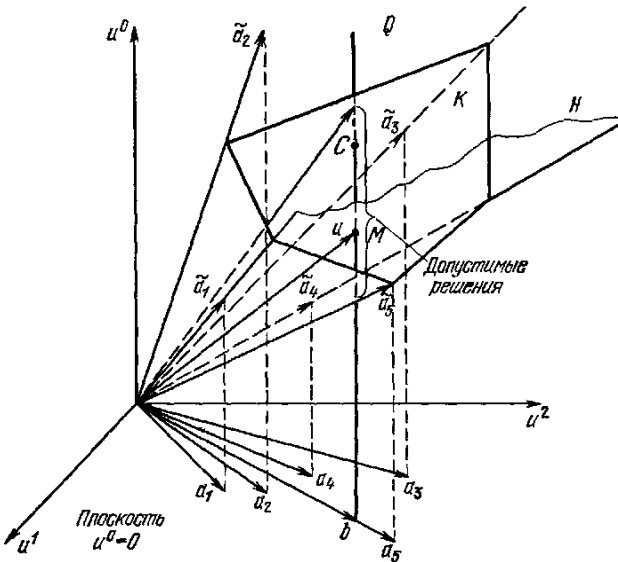


Рис. 6.5

в точку пересечения конуса K и прямой Q , соответствующее значение x^0 будет компонентой некоторого допустимого решения. Таким образом, допустимое множество задачи (6.2.2)–(6.2.4) отражается на отрезок прямой Q , принадлежащей конусу K . Если прямая Q проходит вне этого конуса, задача не имеет допустимых решений. Когда конус K содержит ось u^0 , значение критерия на допустимом множестве не ограничено сверху (рис. 6.6).

Значения целевой функции задачи (6.2.2)–(6.2.4) на любом из ее допустимых решений, соответствующих определенному вектору $\tilde{b}$, совпадают и равны его первой координате. Оптимальные решения (их может быть целое множество) являются прообразами верхней точки пересечения прямой Q с конусом K (точка C на рис. 6.5). Коэффициенты произвольного разложения вектора $\tilde{b}$, указывающего в эту точку, в неотрицательную линейную комбинацию расширенных векторов условий $\tilde{a}_j$ будут компонентами одного из оптимальных решений. Ясно (в частности, из рис. 6.5), что среди них ненулевыми могут быть лишь компоненты, которым отвечают векторы $\tilde{a}_j$, принадлежащие той же грани конуса K , что и точка C (на рис. 6.5 это ребра $\tilde{a}_1, \tilde{a}_2$). Если число таких векторов не превосходит m и соответствующие им a , линейно независимы, задача имеет единственное решение. В противном случае будет существовать континуум решений.

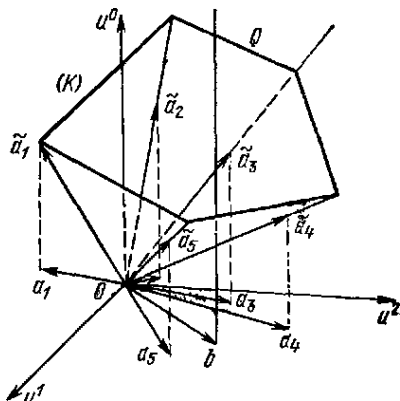


Рис. 6.6

В дальнейшем нам понадобится специфическое понятие невырожденности, используемое в линейном программировании. Мы будем говорить, что задача (6.2.2)–(6.2.4) не вырождена, если вектор b нельзя разложить в неотрицательную линейную комбинацию менее чем m векторов a_j .

На основе представленной выше интерпретации можно наметить путь решения задачи линейного программирования (6.2.2)–(6.2.4). Построим некоторый допустимый вектор такой задачи, все компоненты которого, за исключением $x^{s_1}, \dots, x^{s_m}$, равны нулю, причем выберем индексы $s_1, \dots, s_m$ так, чтобы векторы условий $a_{s_i}, i = 1, \dots, m$, были линейно независимы. Всевозможные линейные комбинации соответствующих им расширенных векторов условий $\tilde{a}_{s_1}, \dots, \tilde{a}_{s_m}$ образуют в $(m+1)$ -мерном пространстве гиперплоскость H . Первая координата ее точки пересечения с прямой Q равна значению целевой функции на рассматриваемом допустимом решении (на рис. 5.5 это решение имеет ненулевые компоненты x^1, x^4 , гиперплоскость H натянута на векторы $\tilde{a}_1, \tilde{a}_4$ и пересекается с прямой Q в точке M).

По предположению, векторы a_{s_i} линейно независимы, т. е. образуют в m -мерном пространстве базис. Разложим остальные векторы условий по этому базису:

$$a_k = a_{s_1} z_{1k} + \dots + a_{s_m} z_{mk} = \sum_{i=1}^m a_{s_i} z_{ik}. \quad (6.2.6)$$

Этому разложению в $(m+1)$ -мерном пространстве соответствуют точки

$$\tilde{a}_k^0 = \tilde{a}_{s_1} z_{1k} + \dots + \tilde{a}_{s_m} z_{mk} = \sum_{i=1}^m \tilde{a}_{s_i} z_{ik}, \quad (6.2.7)$$

принадлежащие плоскости H . Расстояние от точки $\tilde{a}_k^0$ до плоскости $u^0 = 0$ определяется ее первой координатой и, согласно формуле (6.2.7), равно

$$c_k^0 = c^{s_1} z_{1k} + \dots + c^{s_m} z_{mk}.$$

Если все разности $c_k^0 - c^k$ неотрицательны, конус K «лежит под» гиперплоскостью H и наше допустимое решение оптимально. Если же для какого-нибудь индекса k разность $c_k^0 - c^k$, носящая название *оценки замещения*, меньше нуля, точка $\tilde{a}_k$ окажется «выше» плоскости H . Отсюда следует, что если точка пересечения прямой Q и плоскости, натянутой на вектор $\tilde{a}_k$ и какие-нибудь $(m-1)$ векторов a_{s_1} , не выйдет за пределы конуса K , то она будет лежать не ниже прежней точки пересечения. Таким образом, построив допустимое решение с m , вообще говоря, отличными от нуля компонентами, среди которых будет присутствовать x^k и отсутствовать одна из старых компонент, мы тем самым, возможно, увеличим значение целевой функции. Так, для задачи, изображенной на рис. 5.5, имея допустимое решение с ненулевыми компонентами x^1, x^4 , можно перейти к другому, в котором отличны от нуля x^2 или x^3 . Точка пересечения прямой Q с плоскостью, натянутой, например, на векторы $\tilde{a}_2, \tilde{a}_4$, будет лежать между точками M и C , т. е. значение критерия при переходе к допустимому решению с ненулевыми x^2, x^4 увеличится.

В описанной схеме возрастание критерия на каждой итерации гарантируется только, когда задача не вырождена. В противном случае, при формальном переходе от одного индекса s к другому допустимая точка и, соответственно, критерий могут остаться неизменными. Это произойдет, например, в задаче, изображенной на рис. 6.5, если перейти от $s_1 = 2, s_2 = 5$ к $s_4 = 4, s_2 = 5$. Вопросы, связанные с вырожденностью, подробнее будут рассмотрены ниже, когда мы перейдем к обсуждению конкретной реализации указанной схемы — так называемого симплекс-метода решения задач линейного программирования. Однако прежде чем заняться им, необходимо рассмотреть некоторые общие свойства этих задач.

$$Ax = b, \quad x \geq 0, \quad Ay = b, \quad y \geq 0,$$

$$\alpha \geq 0, \beta \geq 0, \alpha + \beta = 1$$

следует, что

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay = (\alpha + \beta)b = b, \\ \alpha x + \beta y \geq 0.$$

Таким образом, точка $z = \alpha x + \beta y$ удовлетворяет тем же ограничениям, что и точки x, y , т. е. является допустимой. Теорема доказана.

Итак, мы показали, что допустимое множество задачи (6.3.1) – выпуклый многогранник. Среди всех его точек для нас особый интерес будут представлять так называемые крайние точки.

Определение 3.2. Точка z называется *крайней точкой* выпуклого множества X , если не существует векторов $x, y \in X, x \neq y$, таких, что

$$z = \alpha x + \beta y,$$

где α, β – некоторые числа, удовлетворяющие условиям

$$\alpha + \beta = 1, \quad \alpha > 0, \quad \beta > 0.$$

Иными словами, точка z называется *крайней точкой* выпуклого множества X , если она не принадлежит внутренности ни одного из отрезков, целиком лежащих в этом множестве. На рис. 6.8 крайними будут точки A, B и все точки дуги C, D . Крайние точки выпуклого многогранника принято называть вершинами. Мы тоже будем использовать этот термин.

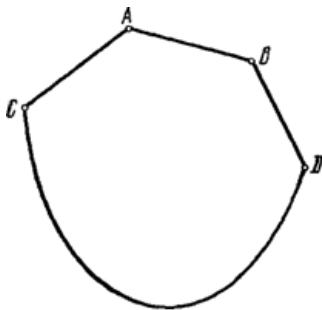


Рис. 6.8

Интерес к вершинам допустимого многогранника задачи линейного программирования объясняется тем, что, как будет установлено ниже, среди ее оптимальных решений обязательно найдется хотя бы одна вершина. Последнее, в силу ограниченности числа вершин, позволяет строить конечные алгоритмы линейной оптимизации.

2. *Существование базисных оптимальных решений.* В линейном программировании принят термин *базисное допустимое решение задачи* (6.3.1). Так называют точку x , удовлетворяющую ограничениям этой задачи и обладающую тем свойством, что векторы условий, соответствующие ее положительным координатам, линейно независимы. (Само собой, таких координат может быть не больше m .) Покажем, что базисные допустимые решения и вершины допустимого многогранника суть одно и то же.

Теорема 3.2. Для того чтобы точка x была вершиной множества допустимых решений задачи (6.3.1), необходимо и достаточно, чтобы векторы условий, отвечающие ее положительным координатам, были линейно независимы.

Доказательство. И необходимость и достаточность будем доказывать от противного. Начнем с необходимости. Пусть x – вершина многогранника, не являющаяся базисным допустимым решением. Не умаляя общности, будем считать, что ненулевыми являются ее первые s координат. Тогда

$$a_1x^1 + \dots + a_sx^s = b,$$

и векторы $a_1, \dots, a_s$ линейно зависимы. Последнее означает, что можно подобрать s чисел $z^1, \dots, z^s$, среди которых есть ненулевые, так, что

$$a_1z^1 + \dots + a_sz^s = 0.$$

Ясно, что при этом для любого ε выполнено равенство

$$a_1(x^1 \pm \varepsilon z^1) + \dots + a_s(x^s \pm \varepsilon z^s) = b.$$

Если же взять ε достаточно малым по модулю, кроме того, будет

$$x^1 \pm \varepsilon z^1 \geq 0, \dots, x^s \pm \varepsilon z^s \geq 0.$$

Таким образом, при малых ε точки $x + \varepsilon z$, $x - \varepsilon z$, где

$$z = \{ z^1, z^2, \dots, z^s, 0, 0, \dots, 0 \}^T,$$

допустимы, причем

$$x = \frac{1}{2} (x + \varepsilon z) + \frac{1}{2} (x - \varepsilon z).$$

Последнее противоречит определению точки x как вершины допустимого многогранника. Необходимость доказана.

Докажем достаточность. Пусть точка x – допустимое базисное решение, но при этом найдутся отличные от x допустимые точки y и z такие, что при некоторых $\alpha > 0$, $\beta > 0$, $\alpha + \beta = 1$ имеем

$$x = \alpha y + \beta z.$$

Из этого равенства следует, что у точек y и z ненулевыми будут те же самые координаты, что и у точки x . Считая, как и прежде, что это – первые s координат, получим

$$\begin{aligned} a_1 y^1 + \dots + a_s y^s &= b, \\ a_1 z^1 + \dots + a_s z^s &= b, \end{aligned}$$

откуда

$$a_1 (y^1 - z^1) + \dots + a_s (y^s - z^s) = 0.$$

Так как среди чисел $y^1 - z^1, \dots, y^s - z^s$ есть ненулевые, это равенство означает, что векторы $a_1, \dots, a_s$ линейно зависимы, что противоречит определению точки x . Полученное противоречие завершает доказательство теоремы.

Теперь мы можем доказать основное утверждение данного параграфа.

Теорема 3.3. Среди оптимальных решений любой задачи (3.1) всегда найдется хотя бы одна вершина допустимого многогранника.

Доказательство. Допустим, что теорема не верна, и мы построили такую задачу (6.3.1), среди решений которой нет

вершин. Возьмем какое-нибудь из этих решений $\hat{x}$. В силу предыдущей теоремы векторы условий, соответствующие его ненулевым компонентам, линейно зависимы. Стало быть (смотрите доказательство теоремы 3.2), можно построить ненулевой вектор z такой, что его компоненты, индексы которых совпадают с индексами нулевых компонент вектора $\hat{x}$, будут нулевыми, причем

$$Az = 0.$$

Отсюда следует, что при малых по модулю $\rho > 0$ векторы $\hat{x} + \rho z$, $\hat{x} - \rho z$ будут допустимы. При этом, так как $\hat{x}$ – решение задачи, имеем

$$(c, \hat{x} + \rho z) \leq (c, \hat{x}),$$

$$(c, \hat{x} - \rho z) \leq (c, \hat{x}),$$

что возможно только при соблюдении равенства

$$(c, z) = 0.$$

Последнее означает, что при тех ρ , для которых точки $\hat{x} + \rho z$ или $\hat{x} - \rho z$ допустимы, они будут и оптимальными

У вектора z есть хотя бы одна ненулевая компонента z^q . Без ограничения общности можно считать, что она отрицательна. Это значит, что, двигаясь из точки $\hat{x}$ в направлении z и не желая покидать допустимое множество, мы сможем сделать шаг, не превышающий $x^q/|z^q|$. (При таком шаге q -я координата точки $x^q + z^q$ станет нулевой, а при большем – отрицательной.) Следовательно, при движении вдоль z одна из координат точки $\hat{x} + \rho z$, соответствующая положительной координате точки $\hat{x}$, в какой-то момент станет нулевой. Точка, в которой это произойдет, в силу сказанного ранее будет решением задачи. При этом у нее будет на одну положительную координату меньше, чем у $\hat{x}$. Применяя к ней все прежние рассуждения, мы придем к решению, число положительных координат которого меньше, чем у $\hat{x}$, на две и т. д. Если предположение о

существовании задач (6.3.1), среди решений которых нет вершин, справедливо, процесс будет бесконечным. Однако это невозможно, так как у $\hat{x}$ есть только конечное число положительных координат. Следовательно, упомянутых задач не существует. Теорема доказана.

Итак, мы показали, что хотя бы одно из решений задачи (6.3.1) будет базисным, т. е. – вершиной допустимого многогранника, а еще точнее – точкой, число положительных координат которой не превосходит m и соответствующие этим координатам векторы условий линейно независимы. Коль скоро рассматриваемая задача не вырождена, у каждой вершины будет ровно m положительных координат. В противном случае их число может оказаться меньше m . Однако, если ранг матрицы условий равен m (а мы всегда предполагаем это), и здесь с каждой вершиной можно связать набор (причем, вообще говоря, неоднозначно) из m индексов таких, что столбцы матрицы A с этими индексами линейно независимы и координаты вершины с номерами, не вошедшими в этот набор, равны нулю. Для этого, возможно, придется пополнить список номеров положительных координат вершины одним или несколькими номерами ее нулевых координат.

Сказанное приводит к новому, эквивалентному предыдущему, определению допустимого базисного решения: допустимое решение задачи (6.3.1) называется базисным, если все его координаты удастся разбить на две группы так, чтобы в первую попало ровно m координат и соответствующие им векторы условий были линейно независимы, а вторая состояла бы только из нулей. Список координат первой группы принято называть допустимым базисом, а сами эти координаты и отвечающие им векторы условий – базисными. Координаты, относящиеся ко второй группе, и их векторы условий называют небазисными. Если все координаты, составляющие допустимый базис, положительны, говорят, что он не вырожден. Такими будут все допустимые базисы невырожденной задачи. При этом между ними и вершинами допустимого многогранника существует взаимно однозначное соответствие. Если же задача

1. *Алгоритм с использованием симплекс-таблиц.* Пусть поставлена невырожденная задача (6.4.1) и известно одно из ее допустимых базисных решений $\tilde{x}$. Среди координат $\tilde{x}$ будет ровно m положительных (мы договорились называть их базисными). Не ограничивая общности, можно считать, что это – первые m координат. Тогда столбцы $a_1, a_2, \dots, a_m$ линейно независимы (по определению базиса) и, соответственно, уравнения задачи (6.4.1) разрешимы относительно базисных переменных $x^1, \dots, x^m$. Выражения последних через остальные переменные должны быть линейны, т. е. выглядят так:

$$x^i = \mu^i - \sum_{k=m+1}^n z_{ik} x^k, \quad i = 1, \dots, m,$$

где μ^i, z_{ik} – некоторые параметры. Поскольку эти равенства должны выполняться и для $x^i = \tilde{x}^i, i = 1, \dots, m$, а небазисные координаты $x^i = \tilde{x}^i, i = m + 1, \dots, n$, равны нулю, окончательно получим

$$x^i = \tilde{x}^i - \sum_{k=m+1}^n z_{ik} x^k, \quad i = 1, \dots, m. \quad (6.4.2)$$

Параметры z_{ik} в этом выражении можно вычислить, например, методом исключения Гаусса. Они являются коэффициентами разложения небазисных столбцов по базисным.

В силу (6.4.2) значение критерия задачи (6.4.1) в любой точке x , удовлетворяющей ее ограничениям – равенствам, равно

$$(c, x) = \sum_{i=1}^m c^i \tilde{x}^i - \sum_{i=1}^m c^i \sum_{k=m+1}^n z_{ik} x^k + \sum_{k=m+1}^n c^k x^k = \sum_{i=1}^m c^i \tilde{x}^i - \sum_{k=m+1}^n \left(\sum_{i=1}^m c^i z_{ik} - c^k \right) x^k \quad (6.4.3)$$

Величины

$$\sigma^k = \sum_{i=1}^m c^i z_{ik} - c^k, \quad k = m + 1, \dots, n,$$

называют оценками замещения, а z_{ik} – коэффициентами замещения (смысл этих терминов станет ясен чуть ниже). И те, и другие, а также значения базисных координат точки $\tilde{x}$ и критерия в ней сводят в так называемую симплекс-таблицу:

		столбец s			столбец k		столбец j				
	$\tilde{x}^0$	0	0	...	0	...	σ^k	...	σ^j	...	σ^n
	$\tilde{x}^1$	1	0	...	0	...	z_{ik}	...	z_{1j}	...	z_{1n}
	$\tilde{x}^2$	0	1	...	0	...	z_{2k}	...	z_{2j}	...	z_{2n}
	.	.	.								.
	.	.	.								.
строка i	.	.	.								.
	$\tilde{x}^i$	0	0	...	0	...	z_{ik}	...	z_{ij}	...	z_{in}
	.	.	.								.
	.	.	.								.
	.	.	.								.
строка s	.	.	.								.
	$\tilde{x}^s$	0	0	...	1	...	z_{sk}	...	z_{sj}	...	z_{sn}
	.	.	.								.
	.	.	.								.
	.	.	.								.
	$\tilde{x}^m$	0	0	...	0	...	z_{mk}	...	z_{mj}	...	z_{mn}

6.4.4

Она образована расширением матрицы системы уравнений (6.4.2). Слева к ней приписан столбец с компонентами $\tilde{x}^i$, $i = 1, \dots, m$; сверху – оценки замещения и, наконец, в верхнем левом углу стоит значение критерия в точке $\tilde{x}$. Столбцы, соответствующие базисным координатам, в рассматриваемом случае составляют единичную подматрицу симплекс-таблицы. Вообще же, они бывают расположены в ней произвольным образом.

При анализе симплекс-таблицы могут встретиться три случая:

Случай 1. Все оценки замещения σ^k неотрицательны. Тогда $\tilde{x}$ – решение задачи (6.4.1). Действительно, при $\sigma^k \geq 0$, $k = m + 1, \dots, n$, из равенства (6.4.3) следует, что

$$(c, x) \leq (c, \tilde{x}) \quad (6.4.5)$$

для любой точки x , удовлетворяющей ограничениям – равенствам задачи (6.4.1) и имеющей неотрицательные координаты $x^{m+1}, \dots, x^n$. Множество таких точек включает допустимый многогранник задачи (6.4.1) и поэтому неравенство (6.4.5) заведомо будет выполняться для любого допустимого решения. Это и означает, что точка $\tilde{x}$ оптимальна.

Случай 2. Для некоторого k оценка замещения σ^k отрицательна и среди коэффициентов замещения z_{ik} , $i = 1, \dots, m$, нет ни одного положительного. Тогда у задачи (6.4.1) нет конечного решения. Чтобы убедиться в этом, рассмотрим точки $x(\rho)$ с координатами

$$\begin{aligned} x^i(\rho) &= \tilde{x}^i - z_{ik}\rho, \quad i = 1, \dots, m, \\ x^k(\rho) &= \rho, \end{aligned} \quad (6.4.6)$$

$$x^i(\rho) = 0, \quad i = m + 1, \dots, k - 1, k + 1, \dots, n.$$

При любом ρ точка $x(\rho)$ является решением уравнений (6.4.2) и, следовательно, удовлетворяет ограничениям – равенствам задачи (6.4.1). Если $\rho \geq 0$, в силу неположительности z_{ik} имеем $x(\rho) \geq \tilde{x} \geq 0$. Таким образом, при любом $\rho \geq 0$ точка $x(\rho)$ допустима, а из (6.4.3) следует, что

$$(c, x(\rho)) = \sum_{i=1}^m c^i \tilde{x}^i - \rho \sigma^k. \quad (6.4.7)$$

Поскольку оценка σ^k , по предположению, отрицательна, отсюда ясно, что, выбирая соответствующие значения $\rho > 0$, можно, не выходя из допустимого многогранника, получить сколь угодно большие значения критерия.

Случай 3. Среди оценок замещения есть отрицательные и для каждого $\sigma^k < 0$ существует $z_{ik} > 0$. Тогда нетрудно найти допустимое базисное решение со значением критерия большим, чем $(c, \tilde{x})$. В качестве такового можно взять точку $x(\rho)$, вычисленную по формулам (6.4.6) при любом k , для которого $\sigma^k > 0$, полагая

$$\bar{\rho} = \min_{\{i: z_{ik} > 0\}} \frac{\tilde{x}^i}{z_{ik}} > 0. \quad (6.4.8)$$

Здесь $\bar{\rho}$ – максимальное для выбранного k положительное значение величины ρ в (6.4.6), при котором координаты $x^i(\bar{\rho})$, $i = 1, \dots, m$, неотрицательны. Для построенной данным способом точки $x(\bar{\rho})$ справедливо неравенство

$$(c, x(\bar{\rho})) = \sum_{i=1}^m c^i \tilde{x}^i - \bar{\rho} \sigma^k > (c, \tilde{x}).$$

То, что $x(\bar{\rho})$ будет допустимым решением, мы уже установили, разбирая случай 2. То, что это допустимое решение будет базисным, т. е. столбцы матрицы условий, соответствующие его неположительным координатам, будут линейно независимы, следует из неравенства нулю величины z_{sk} , где

$$\tilde{x}^s > 0, \quad \bar{\rho} = \frac{\tilde{x}^s}{z_{sk}}, \quad \tilde{x}^s(\bar{\rho}) = 0.$$

Эта величина, как уже было сказано ранее, есть коэффициент при a_s в разложении по a_i , $i = 1, \dots, m$, столбца a_k и должна была бы быть нулем, окажись векторы условий

$$a_1, a_2, \dots, a_{s-1}, a_{s+1}, \dots, a_m, a_k$$

линейно зависимыми.

Итак, имея симплекс-таблицу для некоторого допустимого базисного решения задачи (6.4.1), легко установить, является ли оно оптимальным, и если нет, то вычислить координаты нового базисного допустимого решения с большим значением критерия. Осталось указать способ определения компонент

симплекс–таблицы в новой точке – и алгоритм решения задачи (6.4.1) построен. Основу данной таблицы составляет матрица системы уравнений, получающейся, если разрешить условия равенства в (6.4.1) относительно новых базисных переменных. Это можно сделать методом исключения Гаусса, исходя как из самих равенств (6.4.1), так и из имеющихся, эквивалентных им, равенств (6.4.2), а попросту говоря – из старой симплекс–таблицы. Второй путь значительно более экономен. Он сводится к последовательному умножению на определенные коэффициенты строки старой таблицы, связывающей выводимую из базиса переменную с небазисными, и вычитанию результатов этих умножений из остальных строк. Цель указанных операций – добиться, чтобы в столбце, соответствующем вводимой в базис переменной, отличалась от нуля и была равной единице только одна компонента – та, которая принадлежит упомянутой выше строке. Тем самым и будет получена новая симплекс–таблица (включая новые значения базисных координат, оценок замещения и критерия).

Суммируя сказанное выше, можно предложить алгоритм перебора допустимых базисных решений невырожденной задачи (6.4.1), на каждом шаге которого выполняются следующие операции:

1) *Определение ведущих столбца и строки.* Просматриваются оценки замещения σ^k . Если все они неотрицательны, текущая вершина оптимальна. В противном случае по какому-либо признаку (обычно по признаку минимальности σ^k) выбирается *ведущий столбец* k с оценкой замещения $\sigma^k < 0$. Далее просматриваются значения коэффициентов замещения z_{ik} , $i = 1, \dots, m$. Если среди них нет положительных, задача не имеет конечного решения. В противном случае перебором тех значений индекса i , для которых $z_{ik} > 0$, определяется, причем единственным образом, номер s такой, что

$$\frac{\tilde{x}_{i_s}}{z_{sk}} = \min_{\{i: z_{ik} > 0\}} \frac{\tilde{x}_{i_i}}{z_{ik}}, \quad (6.4.9)$$

где $\tilde{l}_s, \tilde{l}_i$ – номера базисных координат. Теперь нужно перейти в вершину, соответствующую новому допустимому базису, в котором есть k -я переменная и нет переменной с номером $\tilde{l}_s$. Это осуществляется при выполнении второй группы операций.

2) *Пересчет симплекс-таблицы.* Определяется новая s -я строка, равная частному от деления старой на *ведущий элемент* z_{sk} . Для каждого $i \neq s$ вычисляется новая i -я строка – разность старой и результата умножения z_{ik} на новую s -ю строку. Короче говоря, формулы расчета элементов $\bar{z}_{ik}$ новой симплекстаблицы выглядят так:

$$\bar{z}_{ij} = \begin{cases} z_{ij} - \frac{z_{sj} z_{ik}}{z_{sk}}, & i \neq s, \\ \frac{z_{ij}}{z_{ik}}, & i = s, \end{cases}$$

$$i = 0, \dots, m, j = 0, \dots, n.$$

В этой записи $\bar{z}_{0i} = \bar{x}^{\tilde{l}_i}$, $i = 1, \dots, m$, причем $\tilde{l}_s = k$, $\tilde{l}_i = \tilde{l}_i$ для $i \neq s$ и $\bar{z}_{00} = (c, \tilde{x})$, где $\tilde{x}$ – новая вершина.

Сходимость описанного алгоритма решения невырожденной задачи гарантируется тем, что при сменах допустимых базисов значение критерия возрастает. Поэтому повторяться они не могут, и так как их число не больше, чем C_n^m , за конечное число шагов либо будет найдено решение задачи, либо установлено, что ограниченного решения не существует. Алгоритм применим и в вырожденных случаях, но тогда формальная замена одного допустимого базиса другим не обязательно приведет к изменению точки. Соответственно, возможно так называемое *заикливание* – бесконечное повторение набора допустимых базисов, отвечающих одной вершине. Реализуется оно или нет – зависит от того, по каким правилам выбираются номера k и $\tilde{l}_s$ вводимой и выводимой из базиса координат (формула (6.4.9) в вырожденном случае, вообще говоря, не определяет s единственным образом).

Проиллюстрируем работу метода с использованием симплекс-таблиц на задаче

$$\begin{aligned} & \max (-2x^1 - x^2 - x^3), \\ & \begin{cases} x^1 & - x^4 & - 2x^6 = 5, \\ x^2 & + 2x^4 & - 3x^5 + x^6 = 3, \\ x^3 & + 2x^4 & - 5x^5 + 6x^6 = 5, \\ x^i \geq 0, & i = 1, \dots, 6. \end{cases} \end{aligned}$$

В качестве исходного базисного допустимого решения здесь можно взять $\tilde{x}^1 = 5$, $\tilde{x}^2 = 3$, $\tilde{x}^3 = 5$, $\tilde{x}^4 = 0$, $i = 4, 5, 6$. Элементы z_{ij} , $i = 1, 2, 3$, $j = 1, 2, 6$, связанной с ним симплекс-таблицы совпадают с соответствующими элементами матрицы условий. В целом таблица, отвечающая выбранному допустимому базису, выглядит так:

-13	0	0	0	-3	8	-5
5	1	0	0	-1	0	-2
3	0	1	0	2	-3	1
5	0	0	1	2	-5	6

Поскольку среди оценок замещения есть отрицательные, его можно улучшить. Для этого введем в базис переменную x^4 (т. е. возьмем в качестве ведущего столбец с оценкой замещения, равной -3). Тогда ведущей строкой будет вторая, т. е. из базиса будет выведена переменная x^2 . Чтобы вычислить элементы новой симплекс-таблицы, ведущую строку нужно разделить на ведущий элемент $z_{24} = 2$ и, умножив результат на 3, 1, -2, сложить получающиеся строки с нулевой, первой и третьей строкой, соответственно. Это приведет к таблице:

-8,5	0	1,5	0	0	3,5	-3,5
6,5	1	0,5	0	0	-1,5	-1,5
1,5	0	0,5	0	1	-1,5	0,5
2	0	-1	1	0	-2	5

Таким образом, базисные компоненты новой допустимой вершины $\tilde{x}^{l_1} = 6,5$, $\tilde{x}^{l_2} = 1,5$, $\tilde{x}^{l_3} = 2$, где $l_1 = 1$, $l_2 = 4$, $l_3 = 3$. Она

первые n координат полученного решения определяют оптимальную вершину задачи (6.4.10). Строго доказывать этот факт мы не будем, но поясним его на простом примере. Рассмотрим такую задачу:

$$\begin{aligned} \max c^1 x^1, \\ x^1 + x^2 = \bar{x}, \\ x^1 - x^2 = \underline{x}, \\ x^i \geq 0, \quad i = 1, 2, 3. \end{aligned} \quad (6.4.13)$$

где $\bar{x} > 0$ и $c^1 < 0$. Множество допустимых для нее значений переменной x^1 представляет собой отрезок $[\underline{x}, \bar{x}]$. Максимальное значение критерия реализуется при $x^1 = \underline{x}$, $x^2 = \bar{x} - \underline{x}$, $x^3 = 0$. Связанная с (6.4.13) M -задача выглядит так:

$$\begin{aligned} \max (c^1 x^1 - M(x^4 + x^5)) \\ x^1 + x^2 + x^4 = \bar{x}, \\ x^1 - x^3 + x^5 = \underline{x}, \\ x^i \geq 0, \quad i = 1, \dots, 5. \end{aligned} \quad (6.4.14)$$

Значение первой координаты ее оптимальной точки есть решение задачи

$$\begin{aligned} \max F(x^1), \\ x^1 \geq 0, \end{aligned}$$

где $F(x^1)$ – максимальное значение критерия в задаче

$$\begin{aligned} \max_{x^2, x^3, x^4, x^5} (c^1 x^1 - M x^4 - M x^5), \\ x^2 + x^4 = \bar{x} - x^1, \\ -x^3 + x^5 = \underline{x} - x^1, \\ x^i \geq 0, \quad i = 2, \dots, 5. \end{aligned}$$

(Если допустимое множество последней пусто, а это будет при $x^1 > \bar{x}$, полагаем $F(x^1) = -\infty$.) График функции $F(x^1)$ при $M > |c^1|$ показан на рисунке 6.9.

Точкой ее максимума является $x^1 = \underline{x}$. Соответственно, решением задачи (6.4.14) будет $x^1 = \underline{x}$, $\bar{x}^2 = \bar{x} - \underline{x}$, $x^3 = x^4 = x^5 = 0$. Его первые три координаты составляют решение задачи (6.4.13).

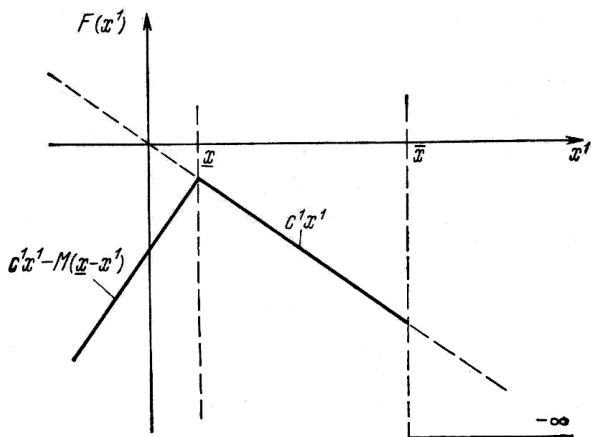


Рис. 6.9

3. *Вырожденность.* Описание функционирования простейшей версии симплекс-метода, приведенное в начале параграфа, дано применительно к невырожденной задаче линейного программирования. Невырожденность обеспечивает смену вершин на каждой итерации, на основе чего доказывается сходимость алгоритма. Последний, будучи дополнен правилом выбора ведущей строки среди строк с номерами s , удовлетворяющих (6.4.9), реализуем и в вырожденных случаях. Однако здесь нет гарантии того, что на каждой итерации вершина допустимого многогранника сменится, — базис, и симплекс-таблица, разумеется, будут меняться, но если список положительных базисных координат при этом сохраняется, точка останется неизменной. В данном случае возникает опасность вырождения процесса поиска решения задачи в бесконечный циклический перебор нескольких допустимых базисов, отвечающих одной, вообще говоря, не оптимальной вершине. Например, если решение задачи

$$\begin{aligned} \max & \left(-\frac{3}{4}x^1 + 150x^2 - \frac{1}{50}x^3 + 6x^4 \right), \\ \frac{1}{4}x^1 - 60x^2 - \frac{1}{25}x^3 + 9x^4 + x^5 &= 0, \\ \frac{1}{2}x^1 - 90x^2 - \frac{1}{50}x^3 + 3x^4 + x^6 &= 0, \\ x^3 + x^7 &= 0, \\ x^i \geq 0, \quad i &= 1, \dots, 7, \end{aligned}$$

начать с допустимого базиса, содержащего пятую, шестую и седьмую координаты, причем ведущий столбец k выбирать по принципу минимума оценки замещения, а среди строк с номерами s , для которых ведущей всегда

$$\frac{\tilde{x}_{\tilde{l}_s}}{z_{sk}} = \min_{\{i: z_{ik} > 0\}} \frac{\tilde{x}_{\tilde{l}_i}}{z_{ik}} = 0,$$

считать строку с наименьшим номером, симплекс-алгоритм заикнется в исходной точке (которая, как легко видеть, не оптимальна; смотрите табл. 6.1).

Таблица 6.1

№ итерации	Коэффициенты нулевой строки (оценка замещения)							Базисные переменные	Значение критерияльной функции
1	-3/4	150	-1/50	6	0	0	0	$x^5 x^6 x^7$	0
2	0	-30	-7/50	33	3	0	0	$x^1 x^6 x^7$	0
3	0	0	-2/25	18	1	1	0	$x^1 x^2 x^7$	0
4	1/4	0	0	-3	-2	3	0	$x^3 x^2 x^7$	0
5	-1/2	120	0	0	-1	1	0	$x^3 x^4 x^7$	0
6	-7/4	330	1/50	0	0	-2	0	$x^5 x^4 x^7$	0
7	-3/4	150	-	6	0	0	0	$x^5 x^6 x^7$	0
			1/50						

Нужно сказать, что заикливание на практике встречается крайне редко, хотя среди решаемых задач вырожденные составляют большинство (в частности, если вектор правых частей ограничений – равенств задачи имеет нулевые

компоненты, процедуру поиска ее решения по схеме, описанной в предыдущем пункте, предлагается начинать с вырожденного допустимого базиса). Однако раз это возможно, нужно иметь средства борьбы с ним, а точнее – правила вывода из базиса, исключающие вероятность его повторения. С одним из таких правил мы сейчас познакомимся.

Пусть дана вырожденная задача, матричная запись которой выглядит так:

$$\begin{aligned} \max (c, x), \\ Ax = b, \\ x \geq 0. \end{aligned} \quad (6.4.15)$$

Рассмотрим наряду с ней задачу вида

$$\begin{aligned} \max (c, x), \\ Ax = b + \sum_{i=1}^n \varepsilon^i a_i, \\ x \geq 0, \end{aligned} \quad (6.4.16)$$

где ε – положительное число, а ε^i – его i -я степень. При достаточно малых ε задача (6.4.16) не вырождена и каждый из ее допустимых базисов будет допустимым базисом исходной задачи, причем оптимальный допустимый базис для задачи (6.4.16) будет оптимальным и для задачи (6.4.15). Следовательно, найти решение последней можно, перебирая ее допустимые базисы в том порядке, в котором симплекс–алгоритм перебирал бы их, решая задачу (6.4.16). Зацикливание при этом исключено, поскольку задача (6.4.16) не вырождена.

Посмотрим теперь, как сменяются допустимые базисы при решении симплекс–алгоритмом задачи (6.4.16). Обозначим через $\tilde{x}$, $\tilde{x}(\varepsilon)$ допустимые базисные решения задач (6.4.15), (6.4.16), соответствующие очередному из них. Тогда, как легко понять,

$$\tilde{x} \tilde{l}_i(\varepsilon) = \tilde{x} \tilde{l}_i + \varepsilon z_{i1} + \dots + \varepsilon^n z_{in}, \quad i = 1, \dots, m,$$

где $\tilde{l}_i$ – номера базисных координат, а z_{ij} – элементы симплекс–таблицы для задачи (4.15). Сразу отметим, что эта таблица отличается от симплекс–таблицы для задачи (4.16) только крайним левым столбцом. Поэтому в качестве номера

координаты, которую следовало бы ввести в базис на текущей итерации симплекс-алгоритма решения задачи (4.16), мы можем взять любой номер k , для которого оценка замещения σ^k в симплекс-таблице задачи (4.15) отрицательна. При этом номер $\tilde{l}_s$ выводимой из базиса координаты определится из равенства

$$\frac{\tilde{x}^{\tilde{l}_s}}{z_{sk}} = \min_{\{i: z_{ik} > 0\}} \frac{\tilde{x}^{\tilde{l}_i}(\varepsilon)}{z_{ik}} = \min_{\{i: z_{ik} > 0\}} \frac{1}{z_{ik}} (\tilde{x}^{\tilde{l}_i} + \varepsilon z_{i1} + \dots + \varepsilon^n z_{in}).$$

Когда величина $\varepsilon > 0$ близка к нулю, процедура определения ведущей строки s состоит в следующем: сравниваются значения $\tilde{x}^{\tilde{l}_i} / z_{ik}$ для i таких, что $z_{ik} > 0$; если одно из них меньше, чем все остальные, соответствующая строка и будет ведущей; в противном случае для тех i , при которых $z_{ik} > 0$ и $\tilde{x}^{\tilde{l}_i} / z_{ik}$ минимально, сравниваются отношения z_{i1}/z_{ik} опять-таки, если минимальным будет только одно из них, отвечающая ему строка – ведущая; если же минимальных z_{i1}/z_{ik} несколько, для них сравниваются отношения z_{i2}/z_{ik} и так далее.

Мы пришли к так называемому *лексикографическому* правилу выбора ведущей строки, исключающему возможность закливания симплекс-алгоритма при решении вырожденной задачи: в качестве ведущей строки надо брать ту, для которой вектор

$$\left\{ \frac{\tilde{x}^{\tilde{l}_i}}{z_{ik}}, \quad \frac{z_{i1}}{z_{ik}}, \quad \dots, \quad \frac{z_{in}}{z_{ik}} \right\}^T$$

лексикографически минимален. Напомним, что вектор a *лексикографически* меньше вектора b , если при сравнении их координат слева направо первая несовпадающая координата вектора a меньше, чем соответствующая координата вектора b .

4. *Алгоритм с обратной матрицей.* В изложенной выше простейшей версии симплекс-метода мы оперировали симплекс-таблицами, основу которых составляют матрицы систем уравнений, получающихся из исходных ограничений – равенств задачи

$$\begin{aligned} \max (c, x), \\ Ax = b, \\ x \geq 0, \end{aligned} \quad (6.4.17)$$

если разрешать их относительно базисных переменных, т. е. умножать A слева на матрицу A'^{-1} , где A' составлена из базисных столбцов. Допустим, что на очередной итерации это – первые m столбцов, и обозначим через x' , x'' векторы, образованные из первых m и последних $n - m$ координат вектора x , через c' , c'' – соответствующие им части вектора c , а через A'' – матрицу, составленную из последних, небазисных, $n - m$ столбцов матрицы A . Тогда ограничения-равенства задачи (6.4.17) можно переписать так:

$$A'x' + A''x'' = b.$$

Эквивалентная им система уравнений, разрешенных относительно базисных переменных (т. е. вектора x'), имеет вид

$$x' + A'^{-1}A''x'' = A'^{-1}b = \tilde{x}', \quad (6.4.18)$$

где $\tilde{x}'$ – вектор, образованный координатами рассматриваемой вершины. При этом значение критерия в любой точке $x = \{x', x''\}^T$, удовлетворяющей (6.4.18), равно

$$(c', x') + (c'', x'') = (c', \tilde{x}') - (c', A'^{-1}A''x'') + (c'', x'') = (c', \tilde{x}') - ((A'^{-1}A'')^T c' - c'', x'').$$

Сопоставляя это равенство и (6.4.18) с (6.4.3), (6.4.2), видим, что симплекс-таблицу в новых обозначениях можно записать так:

$$\left[\begin{array}{c|cc} (c', \tilde{x}') & 0 & c'^T & A'^{-1}A'' - c''^T \\ \tilde{x}' & E & A'^{-1}A'' & \end{array} \right], \quad (6.4.19)$$

где F – единичная матрица.

Из полученного представления симплекс-таблицы ясно, что для перехода к новому допустимому базису по прежней схеме достаточно знать матрицу A'^{-1} и вектор-строку $\lambda = c'^T A'^{-1}$:

последовательно перемножая λ , на небазисные столбцы и вычитая каждый раз из результата соответствующий элемент вектора c'' , вычислим оценки замещения и выделим ведущий столбец; умножив на него матрицу A'^{-1} , найдем те коэффициенты замещения, которые используются в формуле для определения ведущей строки, и выделим эту строку; затем по старым формулам вычислим значения координат нового базиса. Таким образом, можно предложить версию симплекс-метода, при реализации которой на ЭВМ в памяти машины вместо симплекс-таблиц придется хранить только $(m \times m)$ -матрицу A'^{-1} , m -мерный вектор λ , текущие значения базисных координат и критерия. Как правило, эта версия требует значительно меньшего объема памяти, чем первоначальная. Чтобы она была, кроме того, эффективна в смысле быстродействия, надо заложить в нее экономный алгоритм пересчета (при сменах допустимых базисов) матрицы, обратной к базисной. Вывести такой алгоритм совсем несложно.

Вернемся к системе уравнений (6.4.18). Ее матрица,

$$[E | A'^{-1} A''] = A'^{-1} A$$

составляет основной блок симплекс-таблицы (6.4.19). Представленные в параграфе 6.1 формулы преобразования последней при переходе к новому базису есть не что иное, как формулы умножения ее на матрицу

$$M = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & -z_{0k}/z_{sk} & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & -z_{ik}/z_{sk} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & 1 & -z_{s-1k}/z_{sk} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 & 1/z_{sk} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 & -z_{s+1k}/z_{sk} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & -z_{mk}/z_{sk} & 0 & \dots & 1 \end{bmatrix} = \left[\begin{array}{c|ccc} 1 & 0 & \dots & \\ \hline - & - & - & - \\ 0 & & & \\ \hline 0 & & & M \end{array} \right].$$

$s+1$ -я строка

При этом $A'^{-1}A$ умножается слева на M' , и в результате получается матрица эквивалентной (6.4.18) системы уравнений, разрешенных относительно новых базисных переменных, т. е.

$$M'A'^{-1}A = \bar{A}'^{-1}A,$$

где $\bar{A}'$ – новая базисная матрица. Отсюда следует, что

$$M'A'^{-1} = \bar{A}'^{-1},$$

и, стало быть, матрица, обратная к базисной, пересчитывается по тем же формулам, что и симплекс–таблица, т. е.

$$\bar{\beta}_{ij} = \begin{cases} \frac{\beta_{ij} - \beta_{sj}z_{ik}}{z_{sk}}, & i \neq s, \\ \frac{\beta_{sj}}{z_{sk}}, & i = s, \end{cases} \quad i = 1, \dots, m, \quad j = 1, \dots, m,$$

где через β_{ij} , $\bar{\beta}_{ij}$, $i = 1, \dots, m, j = 1, \dots, m$, обозначены (i, j) -е элементы матриц A'^{-1} , $\bar{A}'^{-1}$. Полагая $j = 0$, получим формулы пересчета вектора λ , которые выглядят так:

$$\bar{\lambda}^i = \bar{\beta}_{i0} = \begin{cases} \frac{\lambda^i - \lambda^s z_{ik}}{z_{sk}}, & i \neq s, \\ \frac{\lambda^s}{z_{sk}}, & i = s, \end{cases} \quad i = 1, \dots, m.$$

Описанная версия симплекс-метода используется на практике значительно чаще, чем алгоритм с симплекс–таблицами. Наиболее эффективные из существующих реализаций симплекс-метода представляют собой различные ее воплощения, отличающиеся друг от друга методами хранения обратной матрицы в компактной форме.

6. 5. Двойственные задачи и методы.

1. *Теоремы двойственности в линейном программировании.*
Как уже было сказано в предыдущем параграфе, оценки замещения $\hat{\sigma}^k$, соответствующие некоторому базисному допустимому решению $\hat{x}$ задачи

$$\begin{aligned} \max (c, x), \\ Ax = b, \\ x \geq 0, \end{aligned} \quad (6.5.1)$$

можно вычислить, в матричной записи, по формуле

$$\begin{aligned} \hat{\sigma} &= A''^T \hat{\lambda} - c'', \\ \hat{\lambda} &= (A'^{-1})^T c', \end{aligned}$$

т. е.

$$\hat{\sigma}^k = \sum_{r=1}^m \hat{\lambda}^r a_{rk} - c^k,$$

где величины $\hat{\lambda}^r$, $r = 1, \dots, m$, определяются из решения системы уравнений

$$\sum_{r=1}^m \hat{\lambda}^r a_{ri} - c^i = 0, \quad i = 1, \dots, m. \quad (6.5.2)$$

Здесь l_i – номера базисных переменных, $\hat{\sigma}$ – вектор, составленный из оценок замещения для небазисных переменных, c' , c'' – векторы, координатами которых являются компоненты вектора c , отвечающие базисным и небазисным переменным соответственно, A' – матрица, образованная из базисных, а A'' – из небазисных столбцов матрицы A . При этом вектор $\hat{x}'$, состоящий из базисных координат вектора $\hat{x}$, удовлетворяет равенствам

$$\begin{aligned} A' \hat{x}' &= A \hat{x} = b, \\ (c', \hat{x}') &= (c, \hat{x}) \end{aligned}$$

и, следовательно,

$$(b, \hat{\lambda}) = (\hat{\lambda}, \hat{x}) = (c' \hat{\lambda}, \hat{x}') = (c', \hat{x}') = (c, \hat{x}). \quad (6.5.3)$$

Пусть теперь $\hat{x}$ – не просто допустимое, а оптимальное базисное решение задачи (6.1), найденное симплекс-методом. Тогда соответствующие ему оценки замещения неотрицательны, т. е. для вектора $\hat{\lambda}$, связанного с $\hat{x}$, помимо равенств (6.2) справедливы неравенства:

$$\sum_{r=1}^m \hat{\lambda}^r a_{rj} \geq c^j, \quad 1 \leq j \leq n; \quad j \neq i, \quad i = 1, \dots, m.$$

Таким образом, вектор $\hat{\lambda}$ является допустимым для системы ограничений

$$\sum_{r=1}^m \hat{\lambda}^r a_{rj} \geq c^j, \quad j = 1, \dots, n, \quad (6.5.4)$$

и мы покажем сейчас, что $\hat{\lambda}$ – решение задачи минимизации линейной формы (b, λ) на множестве точек, удовлетворяющих данным ограничениям. Эту задачу, матричная запись которой имеет вид

$$\begin{aligned} \min (b, \lambda), \\ A^T \lambda \geq c, \end{aligned} \quad (6.5.5)$$

называют двойственной по отношению к задаче (6.5.1).

Чтобы установить оптимальность рассматриваемого вектора $\hat{\lambda}$ в задаче (6.5.5), сравним значения критериев прямой (6.5.1) и двойственной (6.5.5) задач на их произвольных допустимых решениях x , $\hat{\lambda}$. Для этого умножим векторное неравенство

$$A^T \lambda \geq c,$$

которому, по определению, подчиняется вектор λ , скалярно на вектор x . Все компоненты последнего неотрицательны, и поэтому после умножения знак неравенства сохранится, т. е.

$$(A^T \lambda, x) = (\lambda, Ax) \geq (c, x),$$

а так как

$$Ax = b,$$

отсюда следует, что

$$(b, \lambda) \geq (c, x). \quad (6.5.6)$$

Это – фундаментальное неравенство теории двойственности в линейном программировании. В частности, полагая в нем $x = \hat{x}$ и учитывая (6.5.3), для любого допустимого для задачи (6.5.5) вектора a получим

$$(b, \lambda) \geq (c, \hat{x}) = (\hat{\lambda}, b),$$

что и требовалось доказать.

Итак, коль скоро прямая задача (6.5.1) разрешима, можно утверждать, что разрешима и двойственная по отношению к ней задача (6.5.5). Справедливо и обратное утверждение: из разрешимости задачи (6.5.5) следует разрешимость задачи (6.5.1). Для доказательства этого факта снова обратимся к симплекс-методу. Чтобы можно было говорить о его применении в том виде, как он был описан ранее, для поиска решения задачи (6.5), преобразуем ее в следующую, эквивалентную ей, задачу:

$$\begin{aligned} \max \left(- \sum_{i=1}^m b^i \lambda^i \right), \\ \sum_{i=1}^m a_{ik} \lambda^i - \lambda^{m+k} = c^k, \quad k = 1, \dots, n, \\ \lambda^{m+k} \geq 0, \quad k = 1, \dots, n. \end{aligned} \quad (6.5.7)$$

Здесь λ^{m+k} – вспомогательные переменные. Эту задачу от канонической, на которую были рассчитаны представленные выше версии симплекс-метода, отличает одно – отсутствие условий неотрицательности для первых m переменных. Однако это не должно смущать читателя. Легко понять, что достаточно изменить в алгоритмах, о которых идет речь, чтобы приспособить их к задаче (6.5.7), – нужно при выборе ведущей строки искать минимум отношения базисной координаты к коэффициенту замещения только по координатам с номером

большим, чем m , а прочие координаты всегда считать базисными. Вся остальная техника поиска решения сохраняется.

Если задача (6.5.5) разрешима, таковой будет и задача (6.5.7). Пусть $\hat{\lambda}^i, i = 1, \dots, m + n$, есть координаты ее базисного оптимального решения, найденного симплекс-методом. Тогда оценки замещения для небазисных переменных λ^{μ_i} ,

$i = 1, \dots, m, \dots, m < \mu_i \leq m + n$, неотрицательны. Они, по определению, равны умноженным на минус единицу коэффициентам при λ^{μ_i} в выражениях для критерия задачи (6.5.7) через небазисные координаты произвольного вектора λ , удовлетворяющего ее ограничениям-равенствам. Чтобы получить это выражение, добавим к целевой функции линейную комбинацию разностей

$$\sum_{i=1}^m a_{ik} \lambda^i - \lambda^{m+k} - c^k$$

с некоторыми весами $\hat{x}^k$. Это не изменит ее значения при рассматриваемых λ , т. е. будет

$$\begin{aligned} -\sum_{i=1}^m b^i \lambda^i &= -\sum_{i=1}^m b^i \lambda^i + \sum_{k=1}^m \hat{x}^k \left(\sum_{i=1}^m a_{ik} \lambda^i - \lambda^{m+k} - c^k \right) = \\ &= \sum_{i=1}^m \lambda^i \left(\sum_{k=1}^m a_{ik} \hat{x}^k - b^i \right) - \sum_{\substack{\mu_i \\ m < \mu_i \leq m+n}} \lambda^{\mu_i} \hat{x}^{\mu_i-m} - \sum_{\substack{m < j \leq n \\ j \neq \mu_i}} \lambda^j \hat{x}^{j-m} - \sum_{k=1}^m c^k \hat{x}^k. \end{aligned}$$

Подберем теперь $\hat{x}^k$ так, чтобы коэффициенты перед базисными переменными в правой части равенства обратились в ноль:

$$\begin{aligned} \sum_{k=1}^m a_{ik} \hat{x}^k &= b^i, \\ \hat{x}^{j-m} &= 0, \quad m < j \leq n, \quad j \neq \mu_i. \end{aligned} \tag{6.5.8}$$

(Эта система уравнений разрешима относительно $\hat{x}^k$, так как ее матрица есть транспонированная к базисной матрице решения $\hat{\lambda}^i, i = 1, \dots, m + n$, и, соответственно, не вырождена.) Тогда получим

$$-\sum_{i=1}^m b^i \lambda^i = -\sum_{\mu_i} \lambda^{\mu_i} \widehat{x}^{\mu_i-m} - \sum_{k=1}^n c^k \widehat{x}^k. \quad (6.5.9)$$

Следовательно, оценки замещения есть просто-напросто $\widehat{x}^{\mu_i-m}$. Так как они неотрицательны, вектор $\widehat{x}$, с учетом (6.5.8), будет допустимым для задачи (6.5.1), а поскольку $\widehat{\lambda}^{\mu_i} = 0$, из (6.5.9) имеем

$$\sum_{i=1}^m b^i \lambda^i = \sum_{k=1}^n c^k \widehat{x}^k.$$

Вектор, составленный из первых m компонент решения задачи (6.5.7), является решением задачи (6.5.5). Поэтому из последнего равенства и неравенства (6.5.6) следует, что $\widehat{x}$ – решение задачи (6.5.1).

Итак, мы установили, что из разрешимости одной из задач (6.5.1), (6.5.5) следует разрешимость другой, причем оптимальные значения критериев совпадают. Этот результат называют первой теоремой двойственности. В ее доказательстве центральную роль играло неравенство (6.5.6). Из него вытекает и вторая теорема двойственности: для того чтобы задачи (6.5.1), (6.5.5) были разрешимы, достаточно, чтобы их допустимые множества содержали хотя бы по одной точке.

Действительно, в данном случае в силу (6.5.6) значение критерия прямой задачи ограничено сверху на ее допустимом многограннике величиной (b, λ) , где λ – некоторое допустимое решение двойственной задачи (6.5.5). Но линейная форма, ограниченная сверху на многограннике, достигает на нем максимума и, следовательно, прямая задачи разрешима. Аналогично устанавливается и разрешимость двойственной задачи.

Как утверждает первая теорема двойственности, если $\widehat{x}$, $\widehat{\lambda}$ – решения задачи (6.5.1), (6.5.5), выполнено равенство

$$(b, \widehat{\lambda}) = (c, \widehat{x}),$$

т. е. совпадение значений критериев пары двойственных задач в их допустимых точках есть необходимое условие оптимальности этих точек. Это же условие в силу (6.5.6)

является и достаточным. Таким образом, мы получили простой критерий оптимальности двух допустимых решений задачи (6.5.1), (6.5.6): для того чтобы допустимые решения $\hat{x}$, $\hat{\lambda}$ прямой и двойственной задач были оптимальными, необходимо и достаточно, чтобы значения прямого и двойственного критериев на них совпадали.

Поскольку для любых допустимых x, λ имеем

$$(b, \lambda) - (c, x) = (Ax, \lambda) - (c, x) = (A^T \lambda, x) - (c, x) = (A^T \lambda - c, x),$$

полученный критерий оптимальности можно сформулировать и так: для того чтобы допустимые $\hat{x}$, $\hat{\lambda}$ были решениями прямой и двойственной задач, необходимо и достаточно, чтобы выполнялось равенство

$$(A^T \lambda - c, x) = 0, \quad (6.5.10)$$

или, что то же самое, равенства

$$\left(\sum_{r=1}^m \hat{\lambda}^r a_{rj} - c^j \right) \hat{x}_j = 0, \quad j = 1, \dots, n. \quad (6.5.11)$$

Последние принято называть условиями дополняющей нежесткости, а эквивалентность (6.5.10), (6.5.11) следует из неотрицательности сомножителей скалярного произведения в (6.5.10).

Наконец, отметим, что если строки матрицы условий A в задаче (6.5.1) линейно независимы, допустимые множества этой и двойственной к ней задач одновременно пустыми быть не могут. (Установить этот факт несложно, но для этого нужна теорема об отделимости, которая будет доказана лишь в следующей главе. Поэтому мы приводим его без доказательства.) Таким образом, в силу теорем двойственности при анализе пары задач (6.5.1), (6.5.5) с матрицей A полного ранга могут реализоваться только два случая:

а) обе задачи имеют допустимые решения, и тогда обе они разрешимы;

б) у одной из задач нет допустимых решений, и тогда значения критерия второй задачи на ее допустимом множестве не ограничены сверху, если это – задача максимизации, и снизу – в противоположном случае.

2. *Геометрическая интерпретация теорем двойственности.* Вспомним геометрическую интерпретацию задачи (6.5.1), изложенную в третьем параграфе данной главы. Мы рассматривали преобразование n -мерного пространства векторов x в $m+1$ -мерное пространство векторов $u = \{u^0, u^1, \dots, u^m\}^T$ по формулам

$$\begin{aligned} u^0 &= \sum_{j=1}^n c^j x^j, \\ u^i &= \sum_{j=1}^n a_{ij} x^j, \quad j = 1, \dots, n. \end{aligned} \quad (6.5.12)$$

Это преобразование переводит положительный ортант пространства E_n (т. е. множество векторов x таких, что $x^j \geq 0, j = 1, \dots, n$) в многогранный конус K в пространстве E_{m+1} . Направляющими ребер этого конуса будут векторы $\{c^j, a_{1j}, \dots, a_{mj}\}^T, j = 1, \dots, n$. Образом в E_{m+1} множества допустимых решений задачи (6.1) является заключенный в конусе K отрезок прямой Q , проходящей через точку $\{0, b^1, \dots, b^m\}^T$ параллельно оси $0u^0$. Координаты точек этой прямой определяются так:

$$-\infty < u^0 < +\infty, \quad u^1 = b^1, \dots, u^m = b^m.$$

Рассмотрим теперь в пространстве E_{m+1} семейство проходящих через начало координат гиперплоскостей с уравнениями

$$\sum_{i=1}^n \lambda^i u^i - u^0 \geq 0, \quad (6.5.13)$$

где λ^i – некоторые числа. Нетрудно убедиться, что если λ^i – координаты какого-нибудь допустимого решения задачи (6.5.5), двойственной к (6.5.1), конус K будет лежать с одной стороны от гиперплоскости (6.5.13). Действительно, возьмем произвольный вектор u из K . Его координаты, по определению конуса K , есть линейные комбинации координат некоторого

неотрицательного вектора x и связаны с последними преобразованиями (6.5.12). Таким образом, для $u \in K$ имеем

$$\sum_{i=1}^m \lambda^i u^i - u^0 = \sum_{i=1}^m \lambda^i \sum_{j=1}^n a_{ij} x^j - \sum_{j=1}^n c^j = \sum_{j=1}^n (\lambda^i a_{ij} - c^j) x^j, \quad (6.5.14)$$

Если же λ^i – координаты допустимого решения двойственной задачи, т. е.

$$\sum_{i=1}^n \lambda^i a_{ij} - c^j \geq 0, \quad i = 1, \dots, n,$$

из (6.5.14) следует, что

$$\sum_{i=1}^n \lambda^i u^i - u^0 \geq 0. \quad (6.5.15)$$

Данное неравенство, полученное для произвольного вектора из конуса K , и означает, что этот конус лежит с одной стороны от гиперплоскости (6.5.13), причем с той же стороны, что и вектор $\{-1, 0, 0\}^T$, т. е. «под гиперплоскостью».

Мы показали, что каждому допустимому решению A двойственной к (6.5.1) задачи (6.5.5) отвечает гиперплоскость (6.5.13), расположенная «над конусом K ». Есть и обратная связь – параметры A' любой гиперплоскости (6.5.13), лежащей «над конусом K », образуют допустимое решение задачи (6.5.5). Действительно, соблюдение при некоторых A' неравенства (6.5.15) для каждого и из конуса K или, что то же самое, соблюдение неравенства

$$\sum_{j=1}^n (\lambda^i a_{ij} - c^j) x^j \geq 0$$

для любых $x^j \geq 0$ возможно лишь в том случае, если

$$\lambda^i a_{ij} - c^j \geq 0, \quad i = 1, \dots, m.$$

Итак, допустимое множество двойственной задачи отождествляется с множеством направляющих векторов семейства тех

гиперплоскостей вида (6.5.15), которые лежат над конусом K . При этом координаты точек пересечения каждой из них с прямой Q определяются уравнениями

$$\sum_{i=1}^n \lambda^i u^i - u^0 \geq 0, \\ u^1 = b^1, \dots, u^m = b^m,$$

т. е. первая координата u^0 равна $\sum_{i=1}^m \lambda^i b^i$ и, соответственно,

двойственная задача интерпретируется как задача поиска гиперплоскости из указанного семейства с «наинизшей» точкой пересечения прямой Q . Полученный в предыдущем пункте критерий оптимальности позволяет утверждать, что эта точка будет принадлежать конусу K .

Чтобы пояснить сказанное рисунками, рассмотрим задачу

$$\begin{aligned} \max & (c^1 x^1 + c^2 x^2 + c^3 x^3 + c^4 x^4), \\ & a_{11} x^1 + a_{12} x^2 + a_{13} x^3 + a_{14} x^4 = b^1, \\ & a_{21} x^1 + a_{22} x^2 + a_{23} x^3 + a_{24} x^4 = b^2, \\ & x^1 \geq 0, x^2 \geq 0, x^3 \geq 0, x^4 \geq 0. \end{aligned} \quad (6.5.16)$$

Соответствующий ей конус K с направляющими векторами ребер $\tilde{a}_i = \{c^i, a_{1i}, a_{2i}\}^T$, изображен на рисунке 6.10. Рассечем его плоскостью с уравнением $u^1 = b^1$

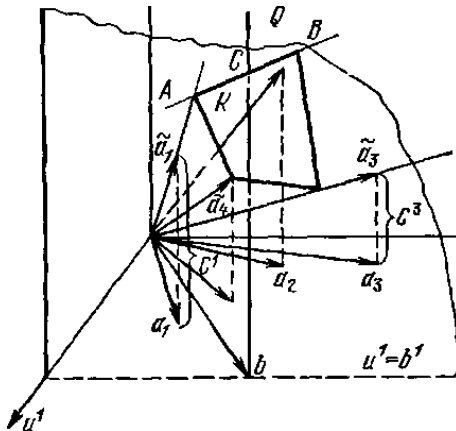


Рис. 6.10

(см. рис. 6.10), содержащей прямую Q . В сечении получим четырехугольник, вершинами которого являются точки пересечения ребер конуса K (т. е. прямых, состоящих из точек $u = \tilde{a}t, t \geq 0$) с рассматриваемой плоскостью. Этот четырехугольник представлен на рис. 6.11. Допустимым решениям задачи (6.5.16) соответствуют принадлежащие ему точки прямой Q . Оптимальному решению – точка C .

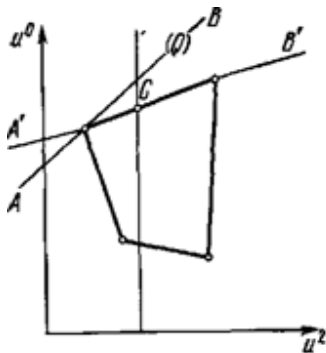


Рис. 6.11

На рисунке 6.11 показаны также прямые $AB, A'B'$, по которым с плоскостью $u^1 = b^1$ пересекаются две плоскости с уравнениями

$$u^0 = \lambda^1 u^1 + \lambda^2 u^2, \quad u^0 = \lambda'^1 u^1 + \lambda'^2 u^2,$$

расположенные над конусом K . Им отвечают допустимые решения

$$\lambda^1 = \{\lambda^1, \lambda^2\}^T, \quad \lambda' = \{\lambda'^1, \lambda'^2\}^T$$

двойственной к (5.5.16) задачи. Значения критерия последней на этих решениях есть ординаты точек пересечения прямых $AB, A'B'$ с прямой Q . Из рисунка 6.11 видно, что любая плоскость, лежащая над конусом K , пересечет эту прямую выше, чем та, которой соответствует прямая $A'B'$. Следовательно, λ' – оптимальное решение двойственной задачи. При этом оптимальное значение двойственного критерия есть ордината точки C , равная по построению оптимальному значению прямого критерия.

3. *Двойственный симплекс-метод.* В предыдущем параграфе были описаны две версии симплекс-метода, применяя которые к задаче.

$$\begin{aligned} \max (c, x), \\ Ax = b, \quad x \geq 0, \end{aligned} \quad (6.5.17)$$

мы за конечное число шагов получим ее решение. Последнее можно отыскать и по-другому, применив, к примеру, алгоритм с использованием симплекс-таблиц к задаче

$$\begin{aligned} \max (-b^1\lambda^1 - b^2\lambda^2 - \dots - b^m\lambda^m), \\ \sum_{r=1}^m a_{ri}\lambda^r - \lambda^{m+1} = c^i, \quad i = 1, \dots, n, \\ \lambda^{m+1} \geq 0, \quad i = 1, \dots, n. \end{aligned} \quad (6.5.18)$$

Тогда базисные координаты $\hat{x}^{s_i}$, $i = 1, \dots, m$, оптимальной вершины задачи (6.5.17) будут получены на последней итерации как оценки замещения для небазисных переменных λ^{s_i+1} , $i = 1, \dots, m$, $i = 1, \dots, m$, задачи (6.5.18) (справедливость данного утверждения была установлена в п. 1 настоящего параграфа). При этом нетрудно построить алгоритм решения задачи (6.5.18), оперирующий симплекс-таблицами задачи (6.5.17). К описанию этого алгоритма мы сейчас и перейдем.

Пусть дан некоторый допустимый базис задачи (6.5.18), т. е. известен список номеров n координат $(n+m)$ -мерного вектора λ такой, что

- а) в него входят первые m координат;
- б) столбцы матрицы условий задачи (6.5.18), отвечающие попавшим в него координатам, линейно независимы;
- в) значения координат из этого списка с номерами больше m , получающиеся из решения уравнений задачи (6.5.18) при нулевых значениях, не включенных в него координат, неотрицательны.

Обозначим номера небазисных координат через μ_i , $i = 1, \dots, m$, $m < \mu_i \leq n + m$. Тогда, как было показано в п. 1, соответствующие им оценки замещения $\hat{x}^{\mu_i-m}$ будут решением уравнений

$$\sum_{\substack{i=\mu_i-m \\ i=1,...,m}} a_{ij} \hat{x}^j = b^r, \quad i = 1, ..., m.$$

Если все они неотрицательны, рассматриваемый базис оптимален для задачи (6.5.18) и вектор $\hat{x}$ ($\hat{x}^j = 0, j \neq \mu_i - m$) – решение задачи (6.5.17). В противном случае можно, используя технику симплекс-метода, найти лучший базис. Включить в него нужно переменную λ^{μ_i} такую, что $\hat{x}^{\mu_i-m} < 0$, а какую из старых базисных переменных с номерами, большими m , исключить – определяется, во-первых, их величинами и, во-вторых, связанными с ними коэффициентами замещения. Последние, по определению, представляют собой параметры системы уравнений, получающихся, если разрешить относительно базисных переменных уравнения (6.18), которые можно переписать так:

$$\sum_{r=1}^b a_{r(\mu_i-m)} \lambda^r = c^{\mu_i-m} + \lambda^{\mu_i}, \quad i = 1, ..., m,$$

$$\lambda^{j+m} = \sum_{r=1}^m a_{ri} \lambda^r - c^i, \quad 1 \leq j \leq n, \quad i + m \neq \mu_i, \quad (6.5.19)$$

$$i = 1, ..., m.$$

Первая группа данных уравнений позволяет выразить $\lambda^r, i = 1, ..., m$, через небазисные переменные λ^{μ_i} . Подстановка получающихся при этом выражений во вторую группу уравнений дает зависимость базисных переменных $\lambda^j, j > m$, от λ^{μ_i} . Параметры этой зависимости и есть с точностью до знака интересующие нас коэффициенты замещения.

Обозначим через $A', c', \lambda', \hat{\lambda}$ матрицу и векторы, составленные из столбцов a_{μ_i-m} и координат $c^{\mu_i-m}, \lambda^{\mu_i}, \lambda^r$, соответственно. Тогда из первой строки в (5.5.19)

$$\hat{\lambda} = (A'^T)^{-1} (c' + \lambda'),$$

получим второй следует, что

$$\lambda'' = A''^T \hat{\lambda} - c'' = A''^T (A'^T)^{-1} (c' + \lambda') - c'' = (A'^{-1} A'')^T (c' + \lambda') - c'', \quad (6.5.20)$$

где A'' , c'' , λ'' – матрица и векторы, образованные столбцами a_i и коэффициентами c^j , λ^{j+m} с номерами $j \neq \mu_i - m$, $i = 1, \dots, m$. Таким образом, коэффициенты замещения есть элементы матрицы $-[A'^{-1}A'']$ или, что то же самое, множители в разложениях столбцов $(-a_{js})$, $j \neq \mu_i - m$, $i = 1, \dots, m$, по столбцам $a_{\mu_i - m}$. Но эти множители есть с точностью до знака элементы симплекс-таблицы задачи (6.5.17), отвечающей ее *недопустимому базису*, в который входят координаты с номерами $\mu_i - m$, $i = 1, \dots, m$, причем в верхней (нулевой) строке этой таблицы будут стоять значения координат задачи (6.5.18) с номерами, большими m , а в крайнем левом столбце – величины $\hat{x}^{\mu_i - m}$ – оценки замещения для небазисных переменных λ^{μ_i} . Следовательно, имея данную симплекс-таблицу, мы можем указать номера координат, которые следует ввести и вывести из базиса задачи (6.5.18) в соответствии с симплекс-методом. Первой из них отвечает строка таблицы с номером $s \geq 1$, элемент которой z_{s0} отрицателен. Номер второй, $m+k$, где $1 \leq k \leq n$, определяется из условия

$$-\frac{z_{0k}}{z_{sk}} = \min_{\{1: z_{sj} < 0\}} -\frac{z_{0j}}{z_{sj}} \quad (6.5.21)$$

или

$$\frac{z_{0k}}{z_{sk}} = \max_{\{1: z_{sj} < 0\}} \frac{z_{0j}}{z_{sj}}.$$

После того как новый допустимый базис задачи (6.5.18) (и отвечающий ему новый, вообще говоря, недопустимый базис задачи (6.5.17)) найден, симплекс-таблица пересчитывается по обычным формулам (см. тему 6.3), и процедура повторяется.

Если задача (6.5.18) не вырождена, ее оптимальный базис и решение задачи (6.5.17) будут получены описанным способом за конечное число шагов. В противном случае в принципе возможно закливание, но, как уже было сказано ранее, это явление крайне редкое.

Представленный алгоритм является простейшей версией *двойственного симплекс-метода*. (Название последнего

отражает тот факт, что поиск с его помощью оптимума в задаче (6.5.17) состоит в решении задачи (6.5.18), эквивалентной двойственной к (6.5.17) задаче.) На практике чаще используют другую его версию: *двойственный симплекс-метод с обратной матрицей*. При реализации ее на ЭВМ в памяти машины хранятся только матрица A^{-1} , обратная к составленной из столбцов $a_{\mu i} - m$ A' (где μ_i – номера небазисных переменных задачи (6.5.18)), а также текущие значения базисных переменных λ^r , $r = i = 1, \dots, m$, и оценок замещения $x^{\mu - m}$, $i = 1, \dots, m$. В данном случае используется формула (6.5.20), в соответствии с которой величина базисной переменной λ^{j+m} , $j \neq \mu_i - m$, $\mu_i = 1, \dots, m$, есть

$$\lambda^{j+m} = z_{0j} = \sum_{r=1}^m a_{rj} \lambda^r - c^j, \quad (6.5.22)$$

а связанные с ней коэффициенты замещения равны

$$z_{ij} = - \sum_{r=1}^m a_{rj} \beta_{ir}, j \neq \mu_i - m, i = 1, \dots, m$$

где β_{ir} – элемент матрицы A^{-1} . После того как номер s строки, отвечающей вводимой в базис переменной $\lambda^{\mu s}$, определен, при каждом $j \geq 1$, $j \neq \mu_i - m$, $i = 1, \dots, m$, по формулам (6.5.22), (6.5.23) вычисляются величины z_{0j} , z_{si} и из равенства (6.5.21) определяется номер переменной, которую следует ввести в базис. При переходе к следующей итерации матрица A^{-1} и величины λ^r , $r = 1, \dots, m$, $x^{\mu - m}$, $i = 1, \dots, m$, пересчитываются по формулам, полученным в 6.4 для прямого симплекс-метода с обратной матрицей.

Чтобы приступить к решению задачи (6.5.17) двойственным симплекс-методом, нужно знать какой-либо допустимый базис задачи (6.5.18). Если специфика последней не позволяет сразу указать такой базис, исходную задачу (6.5.17) модифицируют, вводя дополнительные переменную x^{n+1} и ограничения

$$\sum_{i=1}^{n-m} x^{s_i} + x^{n+1} = b^{n+1}, x^{n+1} \geq 0,$$

где b^{m+1} – большое положительное число и индексы s_i подобраны так, чтобы столбцы a_j , $j \neq s_i$, $i = 1, \dots, n - m$ были линейно независимыми. Решения такая модификация не изменит. При этом задача типа (6.5.18), соответствующая «расширению» исходной задачи, выглядит так:

$$\begin{aligned} \max \left(- \sum_{r=1}^n \lambda^r b^r - \lambda^{m+1} b^{m+1} \right), \\ \sum_{r=1}^m \lambda^r a_{rj} - \lambda^{m+1+j} = c^j, \quad j \neq s_i, \quad i = 1, \dots, n - m, \\ \sum_{r=1}^m \lambda^r a_{rj} + \lambda^{m+1} - \lambda^{m+1+j} = c^j, \quad j = s_i, \quad i = 1, \dots, n - m. \end{aligned}$$

Легко проверить, что вектор $\bar{\lambda}$, координаты которого определяются уравнениями

$$\begin{aligned} \sum_{r=1}^m \bar{\lambda}^r a_{rj} = c^j, \quad \bar{\lambda}^{m+1+j} = 0, \quad j \neq s_i, \quad i = 1, \dots, n - m, \\ \bar{\lambda}^{m+1} = \max_{\substack{j=s_i, \\ i=1, \dots, n-m}} (c^j - \sum_{r=1}^m \bar{\lambda}^r a_{rj}) = c^k - \sum_{r=1}^m \bar{\lambda}^r a_{rj}, \quad \bar{\lambda}^{m+1+k} = 0, \\ \bar{\lambda}^{m+1+j} = \bar{\lambda}^{m+1+j} - c^j + \sum_{r=1}^m \bar{\lambda}^r a_{rj}, \quad j = s_i, \quad i = 1, \dots, n - m \end{aligned}$$

является ее допустимым базисным решением, причем в базис войдут переменные с номерами от 1 до $m+1$ и переменные λ^{m+1+j} , $j = s_i$, $j \neq k$. Таким образом, расширение исходной задачи указанным способом позволяет сразу выявить допустимый начальный базис для двойственного симплекс-метода. Единственная трудность, которая может здесь возникнуть, связана с необходимостью выделения t линейно независимых столбцов матрицы A . Однако, как правило, это не требует больших усилий.

В заключение данного пункта следует сказать, что двойственный симплекс-метод обычно оказывается значительно эффективнее прямого в случаях, когда число переменных исходной задачи существенно больше числа ее уравнений.

4. *Прямодвойственный метод (метод последовательного сокращения невязок).* И в прямом, и в двойственном симплекс-методах с каждой итерацией связана пара векторов $\hat{x}$, $\hat{\lambda}$ (в методах с обратными матрицами и тот и другой простонапросто вычисляются) таких, что значения функционалов прямой и двойственной задач на них совпадают, или, что то же самое, – векторы $\hat{x}$, $\hat{\lambda}$ удовлетворяют условиям дополняющей нежесткости. В прямом симплекс-методе вектор $\hat{x}$ есть базисное допустимое решение прямой задачи, а ограничения двойственной задачи в точках $\hat{\lambda}$ на всех итерациях, кроме последней, не выполняются. В двойственном симплекс-методе все наоборот – вектор $\hat{\lambda}$, составленный из первых t координат очередного базисного решения задачи (6.18), указывает в допустимую вершину двойственной к (6.17) задачи, а x на всех итерациях, кроме последней, хотя и подчиняется уравнениям прямой задачи, имеет отрицательные компоненты, и поэтому недопустим.

В этом пункте мы рассмотрим алгоритм, в котором, как и в двойственном симплекс-методе, перебираются допустимые (правда, не базисные) решения $\hat{\lambda}$ двойственной задачи, а точки $\hat{x}$, связанные с $\hat{\lambda}$ условиями дополняющей нежесткости, ни на одной итерации, кроме последней, ограничениям прямой задачи не удовлетворяют. Однако теперь в $\hat{x}$ будут нарушаться не условия неотрицательности, а ограничения – равенства этой задачи.

Итак, пусть дана задача

$$\max \left(\sum_{j=1}^n c^j x^j \right),$$

$$\sum_{j=1}^n a_{ri} x^j = b^r, \quad r = 1, \dots, m, \quad (6.5.24)$$

$$x^j \geq 0, \quad j = 1, \dots, n,$$

причем все b^r неотрицательны (этого всегда можно добиться, поменяв, если нужно, знаки коэффициентов и правых частей отдельных уравнений) Пусть, кроме того, известно допустимое решение $\hat{\lambda}$ двойственной к (6.5.24) задачи

$$\max \left(\sum_{r=1}^n b^r \lambda^r \right),$$

$$\sum_{r=1}^m a_{ri} \lambda^r \geq c^j, \quad j = 1, \dots, n. \quad (6.5.25)$$

Обозначим через $\hat{J}$ набор тех индексов j , для которых выполнены равенства

$$\sum_{r=1}^m a_{ri} \hat{\lambda}^r \geq c^j,$$

и рассмотрим множество векторов x , связанных с $\hat{\lambda}$ условиями дополняющей нежесткости, т. е. таких x , что

$$x^j \geq 0, \quad k \notin \hat{J}.$$

Если среди них найдется неотрицательный и удовлетворяющий ограничениям задачи (6.5.24), он будет ее решением (тогда $\hat{\lambda}$ – решение задачи (6.5.25). Чтобы отыскать такой вектор или убедиться в том, что его не существует, достаточно решить (заведомо разрешимую) задачу

$$\max \left(- \sum_{r=1}^m x^{n+r} \right),$$

$$\sum_{j \notin \hat{J}} a_{ri} x^j + x^{n+r} = b^r, \quad r = 1, \dots, m, \quad (6.5.26)$$

$$x^j \geq 0, \quad j \in \hat{J}, \quad x^{n+r} \geq 0, \quad r = 1, \dots, m,$$

Коль скоро максимальное значение критерия этой задачи окажется нулем, вектор x , j -е координаты которого для $j \in \hat{J}$ совпадают с соответствующими координатами ее решения, а остальные координаты равны нулю, и есть то, что нам нужно. В противном случае в рассматриваемом множестве векторов x оптимального для задачи (6.24) нет, и если мы все же хотим искать его, решая задачу типа (6.26), необходимо «уточнить набор $\hat{J}$ ». Посмотрим, как это можно сделать.

Получив решение задачи (6.26) любым из описанных выше алгоритмов, мы одновременно найдем и решение $r = 1, \dots, m$, двойственной к ней задачи вида

$$\max \left(\sum_{r=1}^n b^r \mu^r \right),$$

$$\sum_{r=1}^m a_{ri} \mu^r \geq 0, \quad j \in \hat{J}, \quad \mu^r \geq -1, \quad r = 1, \dots, n.$$

Минимальное значение критерия последней равно максимуму критерия задачи (6.5.26) и в исследуемой ситуации отрицательно:

$$\sum_{r=1}^m b^r \hat{\mu}^r < 0. \quad (6.5.27)$$

Это значит, что при движении вдоль $\hat{\mu}$ критерий задачи (6.5.25) будет убывать. Максимальный шаг $\hat{t}$ который можно сделать из точки $\hat{\lambda}$ по направлению r не выходя из допустимого многогранника задачи (6.5.25), определяется ее ограничениями с номерами $j \notin \hat{J}$. Действительно,

$$\sum_{r=1}^m a_{ri} \mu^r \geq 0, \quad j \in \hat{J},$$

и поэтому неравенства

$$\sum_{r=1}^m a_{ri}(\hat{\lambda}^r + t\hat{\mu}^r) = c^j + t \sum_{r=1}^m a_{ri}\hat{\mu}^r \geq c^j, \quad j \in \hat{J},$$

выполнены при любом $t \geq 0$. Если окажется, что справедливы неравенства

$$\sum_{r=1}^m a_{ri}\mu^r \geq 0, \quad j \notin \hat{J},$$

шаг $\hat{t}$ будет равен бесконечности, т. е. точки $\hat{\lambda} + t\hat{\mu}$ допустимы для всех $t \geq 0$. Тогда, учитывая (5.27), можно утверждать, что критерий задачи (6.5.25) не ограничен на ее допустимом множестве и, следовательно, задача (6.5.24) не имеет решения. Если же есть индексы j такие, что

$$\sum_{r=1}^m a_{ri}\mu^r < 0,$$

то шаг $\hat{t}$ определяется по формуле

$$\hat{t} = \min_{\left\{i: \sum_{r=1}^m a_{ri}\hat{\mu}^r < 0\right\}} \left\{ \frac{\sum_{r=1}^m a_{rj}\hat{\lambda}^r - c^j}{-\sum_{r=1}^m a_{rj}\hat{\mu}^r} \right\} > 0.$$

При этом точка $\lambda = \hat{\lambda} + \hat{t} \hat{\mu}$ допустима для задачи (6.25)

и

$$\sum_{r=1}^m b^r \lambda^r = \sum_{r=1}^m b^r (\lambda^r + \hat{t} \hat{\mu}^r) = \sum_{r=1}^m b^r \lambda^r + \hat{t} \sum_{r=1}^m b^r \hat{\mu}^r < \sum_{r=1}^m b^r \lambda^r,$$

т. е. $\bar{\lambda}$ «ближе» к ее решению, чем $\hat{\lambda}$. Это позволяет надеяться на то, что набор индексов $\bar{J}$, для которых

$$\sum_{r=1}^m a_{ri} \bar{\lambda}^r = c^j,$$

«лучше» набора $\widehat{J}$ с точки зрения возможности приблизиться к оптимуму задачи (6.5.24), решая аналог задачи (6.5.26) с $\bar{J}$ вместо $\widehat{J}$. В определенном смысле так оно и есть, и мы сейчас покажем это.

Обозначим через $\widehat{x}^{n+r}$, $\bar{x}^{n+r}$, $r = 1, \dots, m$, соответствующие координаты решений задачи (6.5.26) и задачи

$$\begin{aligned} \max \left(- \sum_{r=1}^m x^{n+r} \right), \\ \sum_{j \in \bar{J}} a_{ri} x^j + x^{n+r} = b^r, \quad r = 1, \dots, m, \\ x^j \geq 0, \quad j \in \bar{J}, \quad x^{n+r} \geq 0, \quad r = 1, \dots, m, \end{aligned} \quad (6.5.28)$$

а через μ^r , $r = 1, \dots, m$, – координаты решения двойственной к (6.5.28) задачи вида

$$\begin{aligned} \min \left(\sum_{r=1}^m b^r \mu^r \right), \\ \sum_{r=1}^m a_{ri} \mu^r \geq 0, \quad j \in \bar{J}, \quad \mu^r \geq -1, \quad r = 1, \dots, m. \end{aligned} \quad (6.5.29)$$

Тогда выполнены равенства

$$\begin{aligned} \sum_{r=1}^m b^r \bar{\mu}^r &= - \sum_{r=1}^m \bar{x}^{n+r}, \\ \sum_{r=1}^m b^r \widehat{\mu}^r &= - \sum_{r=1}^m \widehat{x}^{n+r}. \end{aligned}$$

Ясно, что вектор $\widehat{\mu}$ является решением задачи

$$\begin{aligned} \min \left(\sum_{r=1}^m b^r \mu^r \right), \\ \sum_{r=1}^m a_{ri} \mu^r \geq 0, \quad j \in \widehat{J}, \quad \mu^r \geq -1, \quad r = 1, \dots, m. \end{aligned} \quad (6.5.30)$$

где $\hat{J}'$ – множество индексов существенных ограничений из $\hat{J}$ таких, что

$$\sum_{r=1}^m a_{ri} \hat{\mu}^r = 0, \quad j \in \hat{J}',$$

а из определения вектора $\bar{\lambda}$ следует вложение

$$\hat{J}' \subset \bar{J}.$$

(т. е. в задаче (6.5.29) больше ограничений, чем в (6.5.30)). Поэтому

$$-\sum_{r=1}^m \bar{x}^{n+r} = \sum_{r=1}^m b^r \bar{\mu}^r \geq \sum_{r=1}^m b^r \hat{\mu}^r = -\sum_{r=1}^m \hat{x}^{n+r}$$

или, что то же самое,

$$\sum_{r=1}^m \bar{x}^{n+r} \leq \sum_{r=1}^m \hat{x}^{n+r}. \quad (6.5.31)$$

Таким образом, решив задачу (6.5.28), мы получим значение ее критерия, а попросту говоря, сумму невязок ограничений исходной задачи (6.5.24) не большую, чем для задачи (6.5.26). При этом оптимальные базисы задач (6.5.26), (6.5.28) будут допустимыми базисами задачи

$$\begin{aligned} & \max \left(-\sum_{r=1}^m x^{n+r} \right), \\ & \sum_{r=1}^m a_{ri} x^j + x^{n+r} = b^r, \quad r = 1, \dots, m, \\ & x^j \geq 0, \quad j = 1, \dots, n+m, \end{aligned} \quad (6.5.32)$$

и если она не вырождена, неравенство в (6.5.31) будет строгим. Последнее, в свою очередь, гарантирует сходимость процесса поиска решения задачи (6.5.24), на каждом шаге которого решается задача вида (6.5.26) и множество $\hat{J}$ переопределяется описанным выше способом, за конечное число шагов. Этот

процесс и называется *методом сокращения невязок* (в названии отражено неравенство (6.5.31)) Если задача (6.5.32) вырождена, метод может заиклиться, что, однако, случается крайне редко.

В заключение хотелось бы только добавить, что машинная реализация представленных методов решения линейных задач проста и надежна лишь в тех случаях, когда число ограничений невелико. Чем больше это число, тем большую роль будут играть ошибки округления, неизбежные при машинном счете, и тем менее надежными будут программы, построенные на основе описанных схем. Здесь нужны специальные модификации.

Основная литература: 2[95–150]

Контрольные вопросы:

1. О постановках задачи линейного программирования и ее приложениях.
2. Геометрическая интерпретация задач линейного программирования.
3. Некоторые свойства задач линейного программирования.
4. Симплекс–метод.
5. Двойственные задачи и методы.

7. ОДНОМЕРНАЯ МИНИМИЗАЦИЯ. МЕТОДЫ, ИСПОЛЬЗУЮЩИЕ ИНФОРМАЦИЮ О ПРОИЗВОДНЫХ ЦЕЛЕВОЙ ФУНКЦИИ

Пусть целевая функция $f(x)$ является дифференцируемой или дважды дифференцируемой выпуклой функцией и возможно, вычисление производных $f'(x)$ в произвольно выбранных точках. В этом случае эффективность поиска точки минимума можно существенно повысить.

Рассмотрим три метода минимизации, в которых используются значения производных целевой функции:

- метод средней точки;
- метод хорд;
- метод Ньютона.

Напомним, что для выпуклой дифференцируемой функции равенство $f'(x) = 0$ является не только необходимым, но и достаточным условием глобального минимума. Поэтому если известно, что x^* является внутренней точкой отрезка $[a, b]$, то приближенное равенство $f'(x) = 0$ или $|f'(x)| \leq \varepsilon$ может служить условием остановки вычислений в рассматриваемых трех методах.

7.1. Метод средней точки

Если определение производной $f'(x)$ не представляет затруднений, то в процедуре исключения отрезков методом дихотомии вычисление двух значений $f(x)$ вблизи середины очередного отрезка можно заменить вычислением одного

значения $f(x)$ в его средней точке $\bar{x} = \frac{a+b}{2}$.

В самом деле, если $f'(x) > 0$, то точка $\bar{x}$ лежит на участке монотонного возрастания $f(x)$, поэтому $x^* < \bar{x}$ и точку минимума следует искать на отрезке $[a, \bar{x}]$. При $f'(\bar{x}) < 0$ имеем противоположную ситуацию и переходим к отрезку $[\bar{x}, b]$. Равенство $f'(\bar{x}) = 0$ означает, что точка минимума найдена точно: $x^* = \bar{x}$ (рис. 7.1).

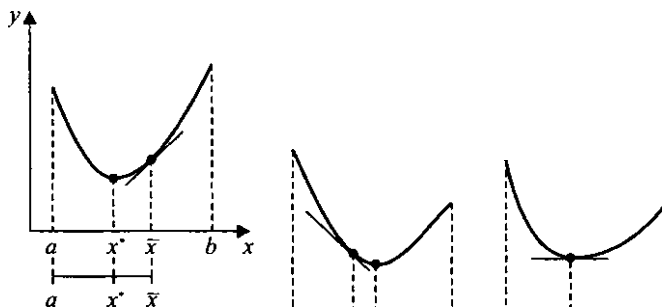


Рис. 7.1. Иллюстрация исключения отрезков методом средней точки

Такое исключение отрезков требует на каждой итерации только одного вычисления $f'(x)$ и уменьшает отрезок поиска точки x^* ровно вдвое.

Алгоритм метода средней точки следующий.

Шаг 1. Положить $\bar{x} = \frac{a+b}{2}$. Вычислить $f'(x)$. Перейти к шагу 2.

Шаг 2. Проверка на окончание поиска: если $|f'(\bar{x})| < \varepsilon$, то положить $x^* = \bar{x}$, $f^* = f(x)$ и завершить поиск, иначе перейти к шагу 3.

Шаг 3. Сравнить $f'(\bar{x})$ с нулем: если $f'(\bar{x}) > 0$, то продолжить поиск на отрезке $[a, \bar{x}]$, положив $b = \bar{x}$, иначе – на отрезке $[\bar{x}, b]$, положив $a = \bar{x}$. Перейти к шагу 1.

Пример 7.1. Методом средней точки решить задачу $f(x) = x^4 + e^{-1} \rightarrow \min, x \in [0, 1]$ с точностью до $|f'(\bar{x})| \leq 0,02$.

Рассмотрим детально действия при выполнении итерации 1.

Шаг 1. Находим $x = 0,5$, $f'(x) = -0,107$, переходим к шагу 2.

Шаг 2. $|f'(\bar{x})| > 0,02$, поэтому переходим к шагу 3.

Шаг 3. $f'(x) < 0$, значит, полагаем $a = \bar{x} = 0,5$ и переходим к следующей итерации начиная с шага 1 и т.д., пока не достигнем необходимой точности.

Метод средней точки напоминает метод дихотомии, но сходится к искомому значению x^* быстрее, поскольку, в отличие от него, после вычисления n значений производной для длины

интервала неопределенности минимизируе мой на $[a, b]$ функции $f(x)$ получаем $\Delta_n = \frac{b-a}{2^n}$.

Исходные данные и результаты вычислений при выполнении итераций приведены в табл. 7.1.

Таблица 7.1
Результаты вычислений примера методом средней точки

Номер итерации	A	B	X	$f'(\bar{x})$	Знак $f'(\bar{x})$
1	0	1	0,5	-0,107	-
2	0,5	1	0,75	1,215	+
3	0,5	0,750	0,625	0,441	+
4	0,5	0,625	0,563	0,142	+
5	0,5	0,563	0,531	0,012	Точность достигнута

Таким образом, $x^* \approx 0,531, f^* \approx 0,668$.

Таким образом, для одинакового уменьшения Δ_n при использовании метода средней точки нужно вычислить вдвое меньше значений функции по сравнению с числом значений самой функции при методе дихотомии.

7.2. Метод хорд

Равенство $f'(x) = 0$ является необходимым и достаточным условием глобального минимума выпуклой дифференцируемой функции $f(x)$. Если на концах отрезка $[a, b]$ производная $f'(x)$ имеет разные знаки, т.е. $f'(a) \cdot f'(b) < 0$, и она непрерывна, то на интервале (a, b) найдется точка, в которой $f'(x)$ обращается в нуль. В этом случае поиск точки минимума $f(x)$ на отрезке $[a, b]$ эквивалентен решению уравнения

$$f'(x) = 0, x \in (a, b). \quad (7.2.1)$$

Таким образом, при $f'(a) \cdot f'(b) < 0$ любой приближенный метод решения уравнения (7.2.1) можно рассматривать как метод минимизации выпуклой, непрерывно дифференцируемой функции $f(x)$ на отрезке $[a, b]$.

Сущность приближенного решения уравнения $F(x) = 0$ на отрезке $[a, b]$ при $F(a) \cdot F(b) < 0$ методом хорд состоит в исключении отрезков путем определения x — точки пересечения с осью Ox хорды графика функции $F(x)$ на $[a, b]$, представленного на рис. 7.2.

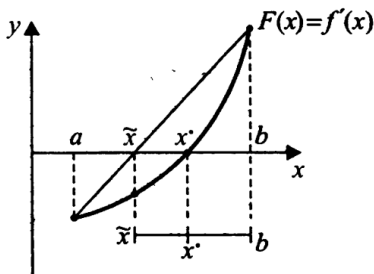


Рис. 7.2. Иллюстрация исключения отрезков методом хорд

Полагая $F(x) = f'(x)$, запишем координату точки $\bar{x}$:

$$\bar{x} = a - \frac{f'(a)}{f'(a) - f'(b)}(a - b). \quad (7.2.2)$$

Отрезок дальнейшего поиска точки x^* ($[a, \bar{x}]$ или $[\bar{x}, b]$) выбирается в зависимости от знака $f'(x)$ так же, как в методе средней точки. На каждой итерации, кроме первой, необходимо вычислять только одно новое значение $f'(x)$.

Алгоритм метода хорд следующий.

Шаг 1. Найти x по формуле (6.2.2). Вычислить $f'(x)$ и перейти к шагу 2.

Шаг 2. Проверка на окончание поиска: если $|f'(\bar{x})| \leq \varepsilon$, то положить $x^* = \bar{x}$, $f^* = f'(\bar{x})$ и завершить поиск, иначе перейти к шагу 3.

Шаг 3. Переход к новому отрезку. Если $f'(x) > 0$, то положить $b = \bar{x}$, $f'(b) = f'(\bar{x})$, иначе положить $a = \bar{x}$, $f'(a) = f'(\bar{x})$. Перейти к шагу 1.

Пример 7.2. Методом хорд решить задачу

$$f(x) = x^4 + e^{-1} \rightarrow \min, x \in [0, 1] \text{ с точностью до } |f'(x)| \leq 0,05.$$

Условие $f'(0)f'(1) < 0$ выполняется.

Рассмотрим детально действия при выполнении итерации 1.
 Шаг 1. Находим: $\bar{x} := 0,216, f(x) = -0,766$. Переходим к шагу 2.
 Шаг 2. $|f'(\bar{x})| > 0,05$, поэтому переходим к шагу 3.
 Шаг 3. $f'(\bar{x}) < 0,05$ поэтому полагаем $a = 0,216, f(a) = -0,766$.
 Переходим к итерации 2 начиная с шага 1 и т. д., пока не достигнем требуемой точности.

До сих пор предполагалось, что $f'(a) \cdot f'(b) < 0$, т.е. производная $f'(x)$ на концах отрезка имеет разные знаки. При нарушении ахого условия точку x^* можно указать сразу.

Так, если $f'(a) \cdot f'(b) > 0$, то $f(x)$ возрастает на $[a, b]$, следовательно, $x^* = a$, а при $f'(a) \cdot f'(b) < 0$ она убывает и $x^* = b$.

В случае $f'(a) \cdot f'(b) = 0$ $x^* = a$ или $x^* = b$ в зависимости от того, на каком из концов отрезка $[a, b]$ производная $f'(x) = 0$.

Результаты вычислений при выполнении итераций записаны в табл. 7.2.

Таблица 7.2
 Результаты вычислений примера методом хорд

Номер итерации	A	b	$\bar{x}$	$f'(\bar{x})$	Знак $f'(\bar{x})$
1	0	1	0,216	-0,766	–
2	0,216	1	0,352	-0,528	–
3	0,352	1	0,435	-0,319	–
4	0,435	1	0,480	-0,175	–
5	0,480	1	0,504	-0,091	–
6	0,504	1	0,516	-0,046	точность достигнута

Отсюда $x^* \approx 0,516, f^* \approx 0,668$.

7.3. Метод Ньютона

Если выпуклая на отрезке $[a, b]$ функция $f(x)$ дважды непрерывно дифференцируема на этом отрезке, то точку $x^* \in [a, b]$ минимума этой функции можно найти, решая уравнение $f'(x) = 0$ методом Ньютона (другое название – метод касательных). Пусть $x_0 \in [a, b]$ – нулевое (начальное) приближение к искомой точке x^* . Линеаризуем функцию $F(x) = f'(x)$ в окрестности

начальной точки, приближенно заменив дугу графика этой функции касательной в точке $(x_0, f'(x_0))$

$$F(x) \approx F(x_0) + F'(x_0)(x - x_0). \quad (7.3.1)$$

Выберем в качестве следующего приближения к x^* точку x_1 пересечения касательной с осью абсцисс. Приравняв к нулю правую часть в (7.3.1), получим первый элемент $x_1 = x_0 - \frac{F(x_0)}{F'(x_0)}$ и итерационной последовательности $\{x_k\}$, $k = 1, 2, \dots$

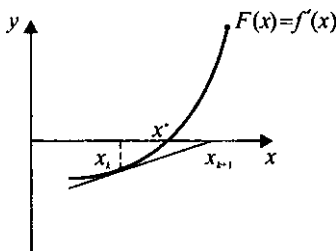


Рис. 7.3. Иллюстрация метода касательных

В очередной точке x_k построим линейную аппроксимирующую функцию для $F(x)$ и определим точку, в которой эта функция обращается в нуль, используя в качестве следующего приближения x_{k+1} (рис. 7.3).

Уравнение касательной к графику $F(x)$ в точке $x = x_k$ имеет вид $y = F(x_k) + F'(x_k)(x - x_k)$, поэтому точка $x = x_{k+1}$, найденная из условия $y = 0$, определяется формулой $x_{k+1} = x_k - \frac{F(x_k)}{F'(x_k)}$.

Поскольку $F(x) = f'(x)$, получим, что для решения уравнения $f'(x) = 0$ необходимо построить последовательность

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}, \quad k = 1, 2, \dots \quad (7.3.2)$$

где x_0 – точка, выбранная в качестве начального приближения. Вычисления по формуле (7.3.2) производятся до тех пор, пока не

выполнится неравенство $|f'(x_k)| \leq \varepsilon$, после чего полагают $x^* \approx x_k$, $f'(x_k) \approx f(x_k)$.

Формулу (7.3.2) можно получить также из иных соображений. Запишем квадратичную аппроксимацию $q_k(x)$ функции $f(x)$ в точке x_k с помощью формулы Тейлора:

$$q_k(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{f''(x_k)}{2}(x - x_k)^2.$$

Функция $q_k(x)$ – квадратный трехчлен, аппроксимирующий $f(x)$ в окрестности точки x_k . Находя точку минимума $q(x)$ из условия $q'(x) = 0$ и считая ее следующим приближением к x^* , получаем искомую формулу (рис. 7.4).

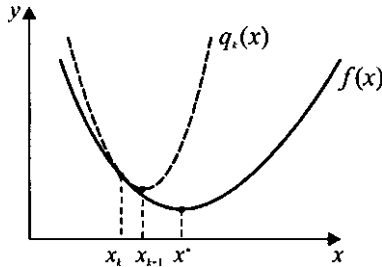


Рис. 7.4. Иллюстрация вывода формулы (7.3.2)

Для квадратичной функции $f(x)$ функция $f'(x)$ линейна, поэтому в (7.3.2) равенство будет точным, а метод Ньютона будет сходиться за один шаг при любом выборе точки x_0 из области определения этой функции.

Для выяснения достаточных условий сходимости рассмотрим метод Ньютона в случае, если $f(x)$ – трижды непрерывно дифференцируемая выпуклая функция. Напишем формулу Тейлора для $f'(x^*)$, где x^* – искомый корень, в окрестности k -го приближения

$$f'(x^*) = 0 = f'(x_k) + f''(x_k) \cdot (x^* - x_k) + \frac{f'''(\xi)}{2}(x^* - x_k)^2,$$

а точка $\xi \in [x^\bullet, x_k]$.

Разделив последнее соотношение на $f''(x_k)$ и перенеся первые два слагаемых из правой части в левую, получим

$$\left[x_k - \frac{f'(x_k)}{f''(x_k)} \right] - x^\bullet = \frac{1}{2} \frac{f'''(\xi)}{f''(\xi)} (x_k - x^\bullet)^2,$$

что, учитывая (7.3.2), переписываем в виде

$$x_{k+1} - x^\bullet = \frac{1}{2} \frac{f'''(\xi)}{f''(x_k - x^\bullet)^2},$$

откуда

$$|x_{k+1} - x^\bullet| = \frac{1}{2} \frac{|f'''(\xi)|}{|f''(x_k)|} |x_k - x^\bullet|^2. \quad (7.3.3)$$

Из (7.3.3) следует оценка

$$|x_{k+1} - x^\bullet| \leq \frac{1}{2} \frac{M_3}{m_2} |x_k - x^\bullet|^2, \quad (7.3.4)$$

где $M_3 = \max_{[a,b]} |f'''(x)|$ $m_2 = \min_{[a,b]} |f''(x)|$.

Ошибка убывает на каждом шаге в том случае, если

$$\frac{1}{2} \frac{M_3}{m_2} |x_0 - x^\bullet| < 1. \quad (7.3.5)$$

Полученное условие означает, что сходимость зависит от начального приближения. Таким образом, метод Ньютона имеет *локальную сходимость*. Если начальное приближение взято далеко, на сходимость рассчитывать не приходится. С другой стороны, всегда можно добиться выполнения условия (7.3.5) за счет более точного выбора начального приближения x_0 , например, с помощью нескольких итераций методами золотого сечения или поразрядного поиска.

Оценка (7.3.4) характеризует скорость убывания погрешности для метода Ньютона: на каждом шаге погрешность пропорциональна квадрату предыдущей (квадратичная скорость сходимости). Это очень высокий темп, например, если в некотором приближении получена одна точная цифра после запятой, то в следующем можно ожидать две точные цифры, затем – четыре и т. д.

Сформулируем теперь *достаточное условие монотонной сходимости* метода Ньютона. Пусть $x^* \in [a, b]$ и $f(x)$ трижды непрерывно дифференцируемая и выпуклая на отрезке $[a, b]$ функция. Ясно, что итерационная последовательность $\{x_k\}$ будет сходиться к пределу x^* монотонно, если

$$0 < \frac{x^* - x_{k+1}}{x^* - x_k} < 1.$$

В соответствии с формулой Тейлора с остаточным членом в форме Лагранжа

$$f'(x^*) = 0 = f'(x_k) + f''(x_k)(x^* - x_k) + \frac{f'''(x)}{2} (x^* - x_k)^2,$$

где точка $x \in [x^*, x_k]$. Поэтому с учетом основной формулы (6.3.2) имеем

$$\frac{x^* - x_{k+1}}{x^* - x_k} = \frac{x^* - x_k + \frac{f'(x_k)}{f''(x_k)}}{x^* - x_k} = 1 - \frac{2}{2 + \frac{f'''(x)(x^* - x_k)^2}{f'(x_k)}}.$$

Отсюда следует, что последовательность $\{x_k\}$ монотонная, если $\frac{f'''(x)}{f'(x_k)} > 0$, т. е. достаточным условием монотонной

сходимости метода Ньютона является постоянство в диапазоне $x \in [x^*, x_k]$ знака производной $f'''(x)$ и совпадение его со знаком $f'(x_0)$. При этом квадратичная скорость сходимости не

гарантируется. Если кроме того выполняется условие (7.3.4), то скорость сходимости метода Ньютона становится квадратичной.

Высокую скорость сходимости метода Ньютона можно объяснить и так: квадратный трехчлен $q_k(x)$, построенный с учетом информации как о первой, так и о второй производной $f(x)$ в точке x_k с высокой точностью аппроксимирует выпуклую дважды дифференцируемую функцию $f(x)$ в достаточно малой окрестности этой точки. Поэтому если очередное приближение x_k оказывается достаточно близким к x^* , то точки минимума x_{k+1} и x^* функций $q_k(x)$ и $f(x)$ практически совпадают.

Пример 7.3. Методом Ньютона найти точку минимума функции $f(x) = x \arctg x, \frac{1}{2} \ln(1+x^2)$ с точностью $|f'(x)| \leq 10^{-7}$.

Производные функции равны $f'(x) = \arctg x, f''(x) = \frac{1}{1+x^2} > 0, f'''(x) = -\frac{2x}{(1+x^2)^2}$. Видно, что при всех значениях

переменной $f'(x) \cdot f'''(x) < 0$, т. е. достаточные условия монотонной сходимости не выполняются. Выберем начальное приближение $x_0 = 1$ и построим приближения x_k по формуле (7.3.2), результаты записаны в табл. 7.3.

Таблица 7.3
Результаты минимизации функции методом Ньютона

К	x_k	$f'(x_k)$
0	1	0,785
1	-0,570	-0,519
2	0,117	0,116
3	$-1,061 \cdot 10^{-3}$	$-1,061 \cdot 10^{-3}$
4	$9 \cdot 10^{-8}$	$9 \cdot 10^{-8}$

Откуда $x^* \approx 9 \cdot 10^{-8} \approx 0$.

Решение примера демонстрирует высокую скорость сходимости последовательности $\{x_k\}$. Можно провести неслож-

ные расчеты и убедиться, что оценка (7.3.4) и условие (7.3.5) выполняются, ошибка на каждом шаге убывает и имеет место квадратичная скорость сходимости.

Достаточные условия монотонной сходимости не выполняются, и последовательные приближения лежат в данном случае по разные стороны от корня.

Если начальное приближение выбрать недостаточно близким к точке x^* то последовательность (7.3.2) метода Ньютона будет расходиться. Так, положив при решении примера $x_0 = 3$, получим последовательность $x_1 = -9,5$, $x_2 = 124$, $x_3 = -23\,905$, $x_4 = 8,97 \cdot 10^8$, $x_5 = -1,27 \cdot 10^{18}$,... которая очевидно расходится.

В подобных случаях необходимо найти лучшее начальное приближение x_0 , например, с помощью нескольких итераций методами золотого сечения или поразрядного поиска.

Пример 7.4. Найти $f(x) = x^2 + \frac{16}{x} \rightarrow \min, x \in [1, 4]$.

Вычислим $f'(x) = 2x - 16x^{-2}$, $f''(x) = 2 + 32/x^3$, $f'''(x) = -96/x^4$. Всюду при $x \in [1, 4]$ имеем: $f''(x) > 0$, $f'''(x) < 0$. Поэтому если начальную точку x_0 выбрать исходя из условия $f'(x) < 0$, то построенная итерационная последовательность будет монотонной.

Выберем $x_0 = 1$ и вычислим $f'(x_0) = -14$, $f''(x_0) = 34$, а затем по формуле (6.3.2) получим

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)} = 1 - \frac{-14}{34} \approx 1,4118.$$

Далее при помощи формулы (6.3.2) находим

$$x_2 = 1,8010, x_3 = 1,9790, x_4 \approx 1,9998.$$

Величина $|x_4 - x^*| < |x_3 - x_4| \approx 0,0208$. Действительно, точное значение минимума $x^* = 2$.

7.4. Возможные модификации метода Ньютона

Метод Ньютона-Рафсона. При переходе к очередной итерации новая точка x_{k+1} рассчитывается по формуле

$$x_{k+1} = x_k - \tau_k \frac{f'(x_k)}{f''(x_k)}, \quad 0 < \tau_k \leq 1.$$

В простейшем варианте метода $\tau_k = \tau = \text{const}$ ($\tau = 1$ соответствует исходному методу Ньютона). Оптимальный набор параметров τ_k может быть найден из решения задачи минимизации $\varphi(\tau) = f\left(x_k - \tau \frac{f'(x_k)}{f''(x_k)}\right) \rightarrow \min$.

На практике для параметров τ_k обычно используется приближенное решение последней задачи $\tau_k = \frac{(f'(x_k))^2}{(f'(x_k))^2 + (f'(\tilde{x}))^2}$,

где $\tilde{x} = x_k - \frac{f'(x_k)}{f''(x_k)}$.

Метод Марквардта. При переходе к очередной итерации новая точка x_{k+1} рассчитывается по формуле $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k) + \mu_k}$, $\mu_k > 0$.

Значение параметра μ_0 выбирается как минимум на порядок больше значения $f''(x_0)$. При переходе к очередной итерации новое значение μ_{k+1} полагают равным $\mu_{k+1} = \mu_k / 2$, если $f(x_{k+1}) < f(x_k)$, либо $\mu_{k+1} = 2\mu_k$ в противном случае.

7.5. Методы минимизации многомодальных функций

При использовании описанных до сих пор методов численной минимизации существенную роль играла унимодальность целевой функции $f(x)$. Если $f(x)$ этим свойством не обладает, т.е. является многомодальной, то поиск минимума может оказаться безрезультатным или привести к ошибке (будет найдена точка локального экстремума, отличного от

глобального минимума $f(x)$). Поэтому представляют интерес методы минимизации многомодальных функций.

Метод перебора. Применение этого метода строго обосновано лишь для унимодальной на (a, b) функции $f(x)$. Однако если вместо унимодальности потребовать, чтобы $f(x)$ удовлетворяла на $[a, b]$ условию Липшица

$$|f(x_1) - f(x_2)| \leq L|x_1 - x_2| \quad (7.5.1)$$

для всех $x_1, x_2 \in [a, b]$, то можно гарантировать определение минимального значения f^* методом перебора с любой заданной точностью. Сформулируем утверждение более строго.

Пусть функция $f(x)$ удовлетворяет на отрезке $[a, b]$ условию Липшица (7.5.1) с константой L и приближенные значения $x^* \approx x_m, f^* \approx f(x_m)$ найдены методом перебора с разбиением отрезка $[a, b]$ на n частей. Тогда для погрешности δ_n определения минимального значения f^* справедлива оценка

$$\delta_n = f(x) - f^* \leq L \frac{b-a}{2^n}. \quad (7.5.2)$$

Пусть x^* – точка минимума $f(x)$ на $[a, b]$ (существование точки минимума вытекает из теоремы Вейерштрасса для непрерывной функции). Среди пробных точек $x_i = a + \frac{b-a}{n}i, i = 0, \dots, n$ метода перебора найдется такая n точка x_i ,

что $|x_i - x^*| \leq \frac{b-a}{2^n}$. Поэтому с учетом условия Липшица (7.3.1)

можно записать

$$0 \leq f(x_m) - f^* \leq f(x_i) - f^* \leq L|x_i - x^*| \leq L \frac{b-a}{2^n}.$$

Если функция $f(x)$ многомодальна, то погрешность определения ее точки минимума может быть значительной, несмотря на то что сам минимум f^* найден достаточно точно (рис. 7.5).

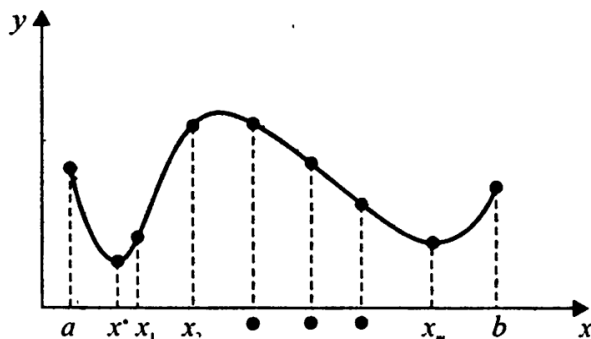


Рис. 7.5. Определение минимума многомодальной функции

Однако во многих случаях практический интерес представляют те значения аргумента x (возможно, далекие от x^*), при которых целевая функция принимает значения, достаточно близкие к минимальному. Это позволяет использовать метод перебора для многомодальных функций.

Основная литература: 7 [37–56]

Контрольные вопросы:

1. Метод средней точки.
2. Метод хорд.
3. Метод Ньютона.
4. Возможные модификации метода Ньютона. Метод Ньютона-Рафсона. Метод Марквардта.
5. Методы минимизации многомодальных функций. Метод перебора.

8. ПРЯМЫЕ МЕТОДЫ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ МНОГОМЕРНЫХ ЗАДАЧ

8.1. Проблема минимизации многомерных задач

Подавляющее число реальных задач оптимизации, представляющих практический интерес, являются многомерными, причем число аргументов целевой функции иногда может быть весьма большим.

Например, в задаче о химическом производстве, где целевая функция зависит от температуры, при определенном ее выборе производительность (выход интересующего нас продукта) оказывается максимальной. Однако наряду с температурой производительность зависит также от давления, соотношения между концентрациями вводимого сырья, катализаторов и ряда других факторов. Таким образом, задача выбора наилучших условий химического производства – это типичная многомерная задача оптимизации.

Как и в одномерном случае, характер задачи и возможные методы решения существенно зависят от информации о целевой функции, которая доступна в процессе ее исследования. В одних случаях, рассмотренных выше в градиентных методах минимизации, функция задается аналитически, являясь при этом дифференцируемой функцией. Тогда можно вычислить ее частные производные, получить явное выражение для градиента, определяющего в каждой точке направления возрастания и убывания функции, и использовать эту информацию для решения задачи. В других случаях никакого аналитического выражения для целевой функции нет, а имеется лишь возможность определить ее значение в любой точке рассматриваемой области (с помощью расчетов, в результате эксперимента и т. д.). В таких задачах в процессе решения можно найти значение целевой функции лишь в конечном числе точек, и по этой информации необходимо приближенно установить ее наименьшее значение для всей области.

В многомерном случае задачи становятся более сложными и трудоемкими, причем трудности при их решении обычно возрастают при увеличении размерности. Для того чтобы лучше почувствовать это, возьмем самый простой по своей идее

приближенный метод поиска наименьшего значения функции, который уже обсуждался для одномерных задач. Покроем рассматриваемую область сеткой с шагом h и определим значения функции в ее узлах. Сравнивая полученные числа между собой, найдем среди них наименьшее и примем его приближенно за наименьшее значение функции для всей области.

Для одномерных задач этот метод носит название метода перебора. Иногда он применяется также для решения двумерных, реже трехмерных задач. Однако для задач большей размерности он практически непригоден ввиду резкого увеличения временных затрат, необходимых для проведения расчетов и вычисления значений целевой функции в большом числе узлов.

Действительно, предположим, что целевая функция зависит от семи переменных, а область определения является семимерным кубом, каждая сторона которого при построении сетки делится на 40 частей. Тогда общее число узлов сетки будет равно $41^7 - 10^{11}$. Пусть вычисление значений функции в одной точке требует 1000 арифметических операций (это немного для функции семи переменных). В этом случае общее число операций составит 10^{14} . Если в распоряжении имеется компьютер с быстродействием 3,2 ГГц, то для решения задачи с помощью данного метода потребуется $3 \cdot 10^4$ с, что составляет сутки непрерывной работы. Добавление еще одной независимой переменной увеличит это время в 40 раз.

Проведенная оценка показывает, что для больших задач оптимизации метод перебора непригоден. Иногда его заменяют случайным поиском, при котором точки сетки просматриваются не подряд, а в случайном порядке. В результате поиск наименьшего значения целевой функции существенно ускоряется, но теряет свою надежность.

Рассмотрим прямые методы минимизации, позволяющие вести поиск наименьшего значения функции многих переменных целенаправленно. Необходимо отметить, что алгоритмы прямых методов решения задачи безусловной минимизации опираются только на вычисление значений функции $f(x)$.

Сначала рассмотрим один из наиболее простых алгоритмов минимизации функции $f(x)$, определенной в E_n . Выберем точку

$x^0 \in E_n$, называемую обычно базовой. Вычислим в ней значение $f(x^0)$, а затем построим n -мерный куб с центром в этой точке и ребрами длиной l . Множество вершин этого куба вместе с точкой x^0 составляют так называемый *шаблон*. Вычислив значения функции в вершинах куба, выберем в качестве новой базовой точки ту из вершин, в которой значение функции меньше $f(x^0)$, и повторим процедуру выбора следующей базовой точки и построения шаблона. Если такой вершины не оказалось, то оставим прежнюю базовую точку x^0 и построим шаблон с уменьшенной (например, вдвое) длиной ребер куба. Процесс поиска закончим тогда, когда длина ребра куба станет меньше заданного числа $\varepsilon > 0$. Геометрическая иллюстрация такого алгоритма, называемая обычно *поиском при помощи шаблона*, для случая двух переменных представлена на рис. 8.1.

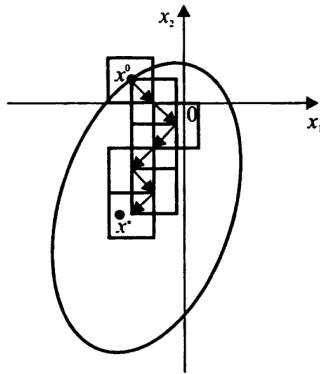


Рис. 8.1. Иллюстрация поиска точки минимума при помощи шаблона

Основной недостаток поиска при помощи шаблона состоит в резком росте количества вычислений значений целевой функции при увеличении размерности пространства n (порядка 2^n), а главное преимущество – простота алгоритма. Эффективность этого алгоритма можно повысить. Геометрически это соответствует процедуре построения вокруг базовой точки в качестве шаблона не n -мерного куба, а конфигурации с меньшим количеством вершин.

Реализация такой идеи привела к разработке методов *симплексного поиска*, получивших к началу 1970-х годов широкое распространение. Симплекс (от латинского *simplex* –

простой) – это простейший выпуклый многогранник в E_n с $n + 1$ вершинами (в E_2 симплекс – совокупность вершин треугольника, в E_3 – тетраэдра). Каждую вершину многогранника называют *вершиной симплекса*, отрезок, соединяющий две вершины, – *ребром симплекса*.

В первых разработанных вариантах симплексного поиска был использован *правильный симплекс*, у которого все вершины равноудалены друг от друга (в E_2 правильный симплекс – совокупность вершин равностороннего треугольника, в E_3 – правильного тетраэдра). Позднее был рассмотрен произвольный (деформируемый) симплекс и предложен способ расширения или сжатия симплекса в зависимости от того, удачно ли выбран шаг поиска в направлении убывания функции. Изменение формы симплекса позволяет улучшить процесс адаптации используемой конфигурации к рельефу графика минимизируемой функции.

8.2. Минимизация функций по правильному (регулярному) симплексу

Рассмотрим простейший алгоритм симплексного поиска минимума унимодальной целевой функции $f(x)$, $x^0 \in E_n$ с использованием регулярного симплекса постоянных размеров. На первом шаге поиска строим регулярный симплекс с вершинами $x^{1,i}$, $i = 1, \dots, n + 1$, который обозначим S_1 . Построение симплекса можно провести двумя простыми способами.

Первый способ. Если $x_j^{1,i} \in S_1$ – заданная базовая точка, то координаты $x_j^{1,i}$, $i = 2, \dots, n + 1, j = 1, \dots, n$ остальных n вершин $x_j^{1,i} \in S_j$ регулярного симплекса S_1 , имеющего ребра длиной l , можно вычислить по формулам

$$x_j^{1,i} = \begin{cases} x_j^{i,1} + \frac{\sqrt{n+1}-1}{n\sqrt{2}}l, & i = j+1, \\ x_j^{i,1} + \frac{\sqrt{n+1}+n-l}{n\sqrt{2}}l, & i \neq j+1 \end{cases} \quad (8.2.1)$$

где $x_j^{1,i} = 1, \dots, n$ – координаты вершины $x^{1,i} \in S_1$. Например, если в E_2 выбрана базовая точка $x^{1,1} = (0, 0)^T$, то при длине ребра $l = 2$ остальные две вершины $x^{1,2}$ и $x^{1,3}$ имеют координаты $x_1^{1,2} = 0,518$, $x_2^{1,2} = 1,932$, $x_1^{1,3} = 1,932$, $x_2^{1,3} = 0,518$. На рис. 8.2 приведен построенный симплекс.

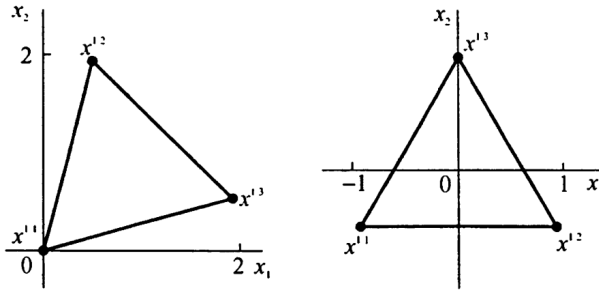


Рис. 8.2. Построение регулярного симплекса двумя способами

Второй способ. Если $x^0 \in E_n$ – заданная базовая точка, определяемая как центр регулярного симплекса S_1 имеющего ребра длиной l , то координаты $x_j^{1,i}$,

$i = 1, \dots, n$, $j = 1, \dots, n$ всех вершин $x^{1,i} \in S_1$ находятся по формулам

$$x_j^{1,i} = \begin{cases} x_j^0, & j < i-1; \\ x_j^0 + \sqrt{\frac{j}{2(j+1)}} l, & j = i-1; \\ x_j^0 + \frac{1}{\sqrt{2j(j+1)}} l, & j > i-1, \end{cases} \quad (8.2.2)$$

где x_j^0 , $j = 1, \dots, n$ – координаты точки x^0 . Например, если в E_2 задана точка $x^0 = (0, 0)^T$ и $l = 2$, то вершинами правильного симплекса будут точки $x^{1,1} = (-1, 0)^T$, $x^{1,2} = (1, 0)^T$, $x^{1,3} = (0, 2)^T$. На рис. 8.2 приведен построенный симплекс.

После построения симплекса S_x на первом шаге поиска в вершинах $x^{1,i}$, $i = 1, \dots, n+1$ вычисляют значения минимизируемой функции и выбирают вершину, в которой значение функции является наибольшим. Пусть для определенности это будет вершина $x^{1,n+1}$ если таких вершин несколько, то может быть взята любая из них. Далее по формуле

$$x^{2,n+1} = 2x_c^1 - x^{1,n+1} = \frac{2}{n} \sum_{i=1}^n x^{1,i} - x^{1,n+1} \quad (8.2.3)$$

находят точку $x^{2,n+1}$ симметричную вершине $x^{1,n+1}$ относительно гиперплоскости, в которой лежат остальные вершины симплекса S_1 . Точка x_c^1 равноудалена от вершин, т. е. является центром регулярного симплекса с n вершинами, или, иначе говоря, центром масс системы материальных точек, расположенных в вершинах симплекса и имеющих одинаковую массу.

Нахождение точки $x^{2,n+1}$ называется *отражением вершины*. В результате получают новый регулярный симплекс S_2 , образованный новой вершиной $x^{2,n+1}$ и n вершинами $x^{1,i}$ $i = 1, \dots, n$, принадлежавшими симплексу S_1 . На рис. 8.3 представлена процедура построения нового регулярного симплекса в случае E_2

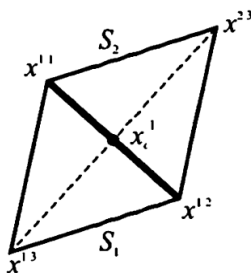


Рис. 8.3. Построение нового симплекса отражением вершины в пространстве E_2

На втором шаге поиска в новой вершине $x^{2,n+1} \in S_2$ симплекса S_2 вычисляют значение $f(x^{k,n+1})$, и если $f(x^{2,n+1}) < f(x^{1,n+1})$, то процедуру поиска повторяют далее. При реализации

алгоритма удобно на каждом k -м шаге поиска отражаемой вершине симплекса S_k присваивать номер $n + 1$, т. е. обозначать ее $x^{k, n+1} \in S_k$. Нумерацию вершин симплекса S_k назовем правильной, если выполнены неравенства

$$f(x^{k,1}) \leq \dots \leq f(x^{k,i}) \leq \dots \leq f(x^{k,n+1}). \quad (8.2.4)$$

Если на k -м шаге после отражения вершины $x^{k, n+1} \in S_k$ в точку $x^{k+1, n+1}$ и вычисления значения $f(x^{k+1, n+1})$ окажется, что $f(x^{k+1, n+1}) > f(x^{k, n+1})$, то следует вернуться к симплексу S_k , считая отражение вершины $x^{k, n+1}$ неудачным. После этого, предполагая нумерацию вершин симплекса S_k правильной, проводят отражение вершины $x^{k, n} \in S_k$, получают точку $x^{k+1, n}$ и сравнивают значения $f(x^{k+1, n})$ и $f(x^{k, n})$. Если $f(x^{k+1, n}) \geq f(x^{k, n})$, то отражение и этой вершины считают неудачным и проводят отражение вершины $x^{k, n-1} \in S_k$ и т. д.

При фиксированной длине ребра симплекса поиск прекращают на шаге с номером K , если отражения всех вершин симплекса S_K оказались неудачными. Тогда в качестве оценки искомого наименьшего значения f^* минимизирующей функции $f(x)$ можно взять наименьшее из вычисленных значений $f(x^{K, j})$, $x^{K, i} \in S_K$, $i = 1, \dots, n + 1$. В процессе поиска построена последовательность симплексов S_k , $k = 1, \dots, K$. Ей соответствует невозрастающая последовательность минимизируемой функции $\{f(x^{k,1})\}_K$.

Исследовать свойства последовательности $\{f(x^{k,1})\}_K$ сложно. Поэтому на практике вместо нее изучают свойства последовательности значений минимизируемой функции $f_k = f(x^k)$, каждое из которых вычислено в центре x^k симплекса S_k , $k = 1, \dots, K$. Процедуру поиска точки $x^* \in E_n$, в котором функция $f(x)$ достигает наименьшего значения, в этом случае проводят в соответствии с рекуррентным соотношением

$$x^{k+1} = x^k + \alpha p^k, \quad k = 1, \dots, K,$$

где p^k — единичный n -мерный вектор, определяющий направление смещения центра симплекса на k -м шаге; $\alpha_k > 0$ —

смещение центра симплекса в направлении вектора p_k . Геометрическая иллюстрация процедуры такого поиска в случае E_2 представлена на рис. 8.4.

Рассмотрим путь построения эффективных алгоритмов, связанный с идеей изменения размера регулярного симплекса в процессе поиска (алгоритмы *управляемого прямого поиска*). Это важно, когда наименьшее значение функции $f(x)$ необходимо найти с высокой точностью. Действительно, чем

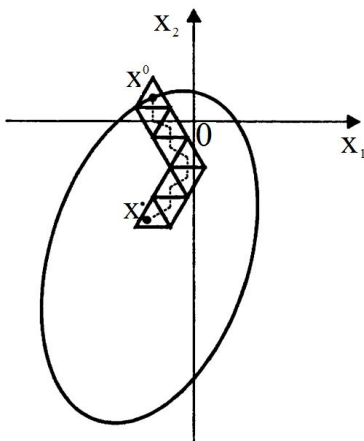


Рис. 8.4. Поиск минимума в E_2 при помощи правильного симплекса

меньше размер симплекса, тем точнее можно локализовать точку x^* в которой эта функция достигает наименьшего значения. При этом малый размер симплекса приводит к его уменьшенному смещению в направлении точки x^* т. е. к росту числа шагов поиска, связанному с увеличением вычислительных затрат. Большой размер симплекса, наоборот, позволяет за каждый шаг поиска осуществить большее смещение центра симплекса, но обеспечивает лишь грубую локализацию точки x^* . Поэтому изменение размера симплекса наделяет алгоритм поиска адаптивными свойствами. Рассмотрим подход, при котором размер симплекса изменяется при выполнении определенного условия, проверяемого на каждом шаге поиска.

Пусть $S_k \in E_n$ – регулярный симплекс на k -м шаге поиска, имеющий правильную нумерацию вершин. Процедура отыскания вершины $x^{k+1; n+1} \in S_{k+1}$ нового симплекса S_{k+1} включает в себя два этапа – отражение вершины $x^{k; n+1} \in S_{k+1}$, если необходимо, уменьшение размера S_{k+1} , называемое *редукцией симплекса*.

Первый этап. Отражение вершины $x^{k; n+1} \in S_k$ осуществляют в соответствии с формулой, аналогичной (8.2.3):

$$x^{k+1; n+1} = 2x_c^k - x^{k; n+1} = -x^{k; n+1}, \quad (8.2.5)$$

где x_c^k равноудалена от вершин $x^{k; j}$, $j = 1, \dots, n$. Затем вычисляют $f(x^{k+1; n+1})$.

Второй этап. Редукцию симплекса S_k проводят только при выполнении неравенства $f(x^{k+1; n+1}) > f(x^{k; n+1})$. При этом длину всех ребер симплекса уменьшают в $1/\delta$ раз, где $\delta \in (0, 1)$ – заданный коэффициент редукции, и находят вершины нового симплекса S_{k+1} по формуле

$$x^{k+1, i} = x^{k, i} + \delta (x^{k, j} - x^{k, 1}), \quad i = 2, \dots, n+1,$$

сжимая симплекс в $1/\delta$ раз к вершине $x^{k, 1}$ в которой значение целевой функции меньше, чем в других вершинах симплекса. Затем осуществляют переход к первому этапу, полагая $k = k + 1$.

Условие прекращения поиска зависит от конкретной задачи минимизации. Например, поиск можно вести, до тех пор пока длина ребер симплекса не станет меньше заранее выбранного значения. Или пока не выполнится неравенство

$$\left(\frac{1}{n+1} \sum_{i=1}^{n+1} (f(x^{k, j}) - f(x^k))^2 \right)^{\frac{1}{2}} < \varepsilon,$$

где $\varepsilon > 0$ – заданное достаточно малое число.

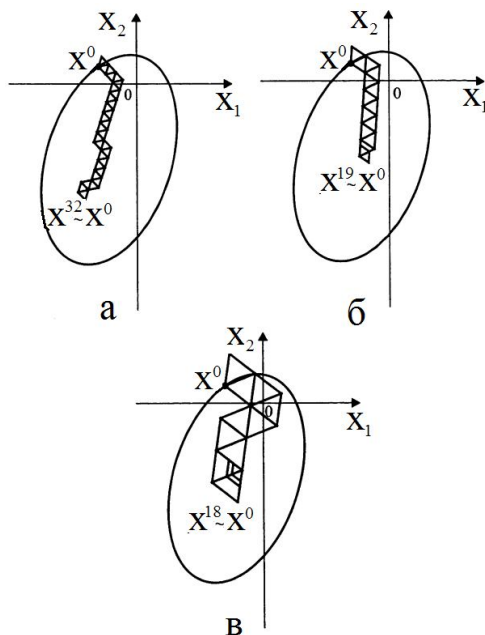


Рис. 8.5. Иллюстрация к примеру 8.1

Пример 8.1. Используя метод регулярного симплекса, найти решение задачи минимизации функции

$$f(x_1, x_2) = 6x_1^2 - 4x_1x_2 + 3x_2^2 + 4\sqrt{5}(x_1 + 2x_2) + 22.$$

Функция квадратичная, имеет минимум $f'' = -28$ при $x = (-\sqrt{5}, -2\sqrt{5})^T$. Выберем базовую точку $x^0 = (-2, 1)^T$ и положим $\varepsilon = 0,01$. На рис. 8.5 дана графическая иллюстрация процесса поиска при коэффициенте редукции $\delta = 1/2$ и трех значениях начальной длины ребра симплекса, указанных в табл. 8.1.

Таблица 8.1

l	x^*	$f(x^*)$	N
0,5	$(-2,233; -4,470)$	-28,0	32
1,0	$(-2,233; -4,470)$	-28,0	19
2,0	$(-2,233; -4,470)$	-28,0	18

8.3. Минимизация функций при помощи нерегулярного симплекса

Симплексные процедуры безусловной минимизации, в которых используются регулярные симплексы, эффективны лишь в тех случаях, когда топография поверхностей уровня целевой функции достаточно проста. В противном случае эффективность применения таких процедур значительно снижается. В частности, это характерно для функций, скорость убывания которых по одному или нескольким направлениям значительно больше, чем по остальным. В таком случае говорят, что график функции имеет овражную структуру.

Примером функции, график которой обладает овражной структурой, является функция Розенброка

$$f(x_1, x_2) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

с минимумом в точке $x^* = (1, 1)^T$. Препятствием для успешного поиска минимума $f(x)$ является в этом случае форма симплекса: регулярный симплекс нельзя изменить так, чтобы он «вытянулся» вдоль «оврага», а это не позволяет продолжить поиск минимума. В этих случаях естественной выглядит идея деформирования симплекса в процессе поиска, т. е. изменения его формы и размера.

Рассмотрим простейшую схему алгоритма прямого поиска по деформируемому симплексу – алгоритм Нелдера-Мида. Этот алгоритм допускает возможность отражения только одной вершины симплекса на каждом шаге поиска. К процедуре отражения при построении нового симплекса добавляются процедуры *сжатия* и *растяжения*. А именно положение новой вершины $x^{k+1, n+1}$ вместо вершины $x^{k, n+1}$ соответствующей наибольшему значению функции, находится сравнением и выбором наименьшего среди значений целевой функции в точках

$$z^1 = x_c^k - \alpha(x_c^k - x^{k, n+1}) = \frac{1-\alpha}{n} \sum_{i=1}^n x^{k, i} + \alpha x^{k, n+1}, \quad 0 < \alpha < 1;$$

$$z^2 = x_c^k + \alpha(x_c^k - x^{k,n+1}) = \frac{1+\alpha}{n} \sum_{i=1}^n x^{k;j} - \alpha x^{k,n+1}, \quad 0 < \alpha < 1;$$

$$z^3 = x_c^k + \beta(x_c^k - x^{k,n+1}) = \frac{1+\beta}{n} \sum_{i=1}^n x^{k,j} - \beta x^{k,n+1}, \quad \beta \approx 1; \quad (8.3.1)$$

$$z^4 = x_c^k + \gamma(x_c^k - x^{k,n+1}) = \frac{1+\gamma}{n} \sum_{i=1}^n x^{k,i} - \gamma x^{k,n+1}, \quad \gamma > 1.$$

Геометрическая иллюстрация этих процедур для пространства E_2 приведена на рис. 8.6 и 8.7. Так как величина $\alpha \in (0,1)$, то выбор точек z^1 и z^2 соответствует сжатию симплекса; $\beta \approx 1$, поэтому выбор точки z^3 соответствует отражению, а $\gamma > 1$ и выбор точки z^4 приводит к растяжению симплекса. Численные эксперименты показывают, что этот алгоритм хорошо работает в пространстве E_n для $n \leq 6$.

Необходимо отметить, что при деформациях утрачивается свойство правильности исходного симплекса. Поэтому, не стремясь к правильности начального симплекса, его строят из произвольной базовой точки $x^0 \in E_n$ по формулам

$$x^i = x^0 + a e^i, \quad (8.3.2)$$

где $e^i, i = 1, \dots, n$ – базисные векторы, a – параметр симплекса.

На практике хорошо зарекомендовал себя следующий набор параметров α, β и γ для выбора пробных точек z^i в формулах (8.3.1): $\alpha = 1/2, \beta = 1, \gamma = 2$.

Для того чтобы избежать сильной деформации симплекса, алгоритм дополняют процедурой обновления. Например, после N шагов алгоритма из точки x^0 снова строят симплекс по формулам (8.2.1), (8.2.2) или (8.3.2), полагая $a = \|x^0 - x''\|$. С теоретической точки зрения симплексные методы минимизации слабо исследованы, но практика показывает их работоспособность.

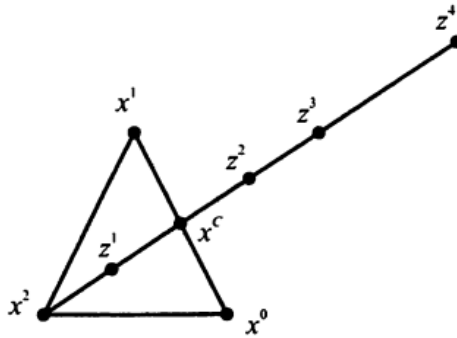


Рис. 8.6. Расположение точек z^1, z^2, z^3, z^4 для перехода к новому симплексу

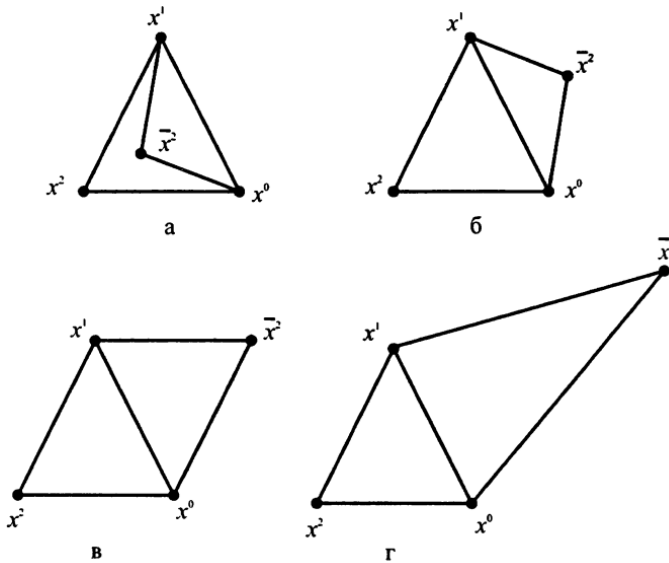


Рис. 8.7. Новые симплексы, полученные в результате процедур сжатия (а, б); отражения (в); растяжения (г)

Вернемся теперь к описанию некоторых вычислительных процедур вида $x^{k+1} = x^k + \alpha_k p^k$. В зависимости от способа выбора направления спуска p^k и шага α_k получаются различные алгоритмы минимизации.

8.4. Метод циклического покоординатного спуска

Этот метод заключается в последовательной минимизации целевой функции $f(x)$ с исчерпывающим спуском сначала по направлению первого базисного вектора e^1 затем второго – e^2 и т. д. После окончания минимизации по направлению последнего базисного вектора e^n цикл повторяется.

Алгоритм метода следующий.

Шаг 1. Выбрать $x \in E_n$, критерий достижения точности $\rho(x, x'') < \varepsilon_1$ или $\rho|f(x) - f(x'')| < \varepsilon_2$, величины $\varepsilon_1, \varepsilon_2$. Найти $f(x)$, положить $j = 1$. Перейти к шагу 2.

Шаг 2. Решить задачу одномерной минимизации $\Phi(a) = f(x + \alpha e^j) \rightarrow \min, \alpha \in E$, т. е. найти α^* . Положить $\bar{x} = x + a^* e^j$ вычислить $f(\bar{x})$. Перейти к шагу 3.

Шаг 3. Если $j < n$, то положить $x = \bar{x}, j = j + 1$ и перейти к шагу 2, иначе перейти к шагу 4.

Шаг 4. Проверить условие достижения точности $\rho(x, \bar{x}) < \varepsilon_1$ или $|f(\hat{x}) - f(\bar{x})| < \varepsilon_2$. Если оно выполняется, то положить $x^* = \bar{x}, f(x^*) = f(\bar{x})$ и закончить поиск. Иначе положить $x = \bar{x}, f(x) = f(\bar{x}), j = 1$ и перейти к шагу 2.

Для приближенного решения вспомогательной задачи одномерной минимизации на шаге 2 на практике удобно использовать метод поразрядного поиска. Алгоритм этого одномерного метода не требует обязательного задания границ интервала минимизации.

Алгоритм метода циклического покоординатного спуска прост, но его эффективность заметно зависит от свойств целевой функции. Метод эффективен в случае, если минимизируемая функция является *сепарабельной*, т. е. представляет собой сумму функций, каждая из которых зависит лишь от одной координаты

$$f(x) = \sum_{j=1}^n h_j(x_j), x \in E_n.$$

В этом случае решение задачи минимизации можно получить за один шаг поиска. Действительно, так как

$$f(x^{\bullet}) = \min_{x \in E_n} f(x) = \sum_{j=1}^n \min_{x_j \in E} h_j(x_j) = \sum_{j=1}^n h_j(x_j^{\bullet}),$$

то достаточно на первом шаге поиска последовательно решить n задач одномерной минимизации функций $h_j(x_j)$, $j = 1, \dots, n$, что позволит найти все n координат искомой точки $x^{\bullet}$. Для функций общего вида эффективность поиска минимума методом циклического покоординатного спуска может быть значительно ниже.

Пример 8.2. Решить задачу $f(x) = x_1^2 + x_2^2 \rightarrow \min, x \in E_2$ методом циклического покоординатного спуска.

Эта функция сепарабельна, ее линии уровня – окружности с центром в начале координат (рис. 8.8). Выберем произвольную начальную точку x , например $x = (3, 3)^T$. Очевидно, два шага исчерпывающего спуска сначала по направлению e^1 , затем – e^2 приведут в точку минимума $x^{\bullet} = (0, 0)^T$.

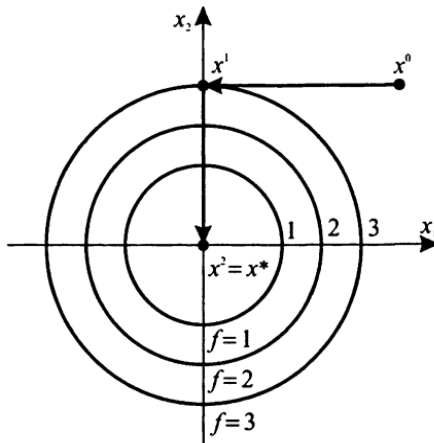


Рис. 8.8. Иллюстрация к примеру 8.2

В примере 8.2 точку минимума функции удалось найти точно за конечное число шагов. Это скорее исключение, чем правило.

Пример 7.3. Найти $f(x) = 5x_1^2 + 5x_2^2 + 8x_1x_2 \rightarrow \min, x \in E_2$ методом циклического покоординатного спуска.

Заметим, что поворот системы координат на угол -45° (замена переменных $x_1 = (y_1 + y_2)/\sqrt{2}$ и $x_2 = (-y_1 + y_2)/\sqrt{2}$) приводит функцию к виду $f_1(y) = y_1^2 + 9y_2^2$. Линии уровня этой целевой функции – эллипсы $y_1^2/9 + y_2^2 = c^2$, главные оси которых направлены вдоль осей (y_1, y_2) (рис. 8.9). Результаты решения представлены в табл. 8.2 (точность решения одномерной задачи минимизации выбрана 10^{-2}).

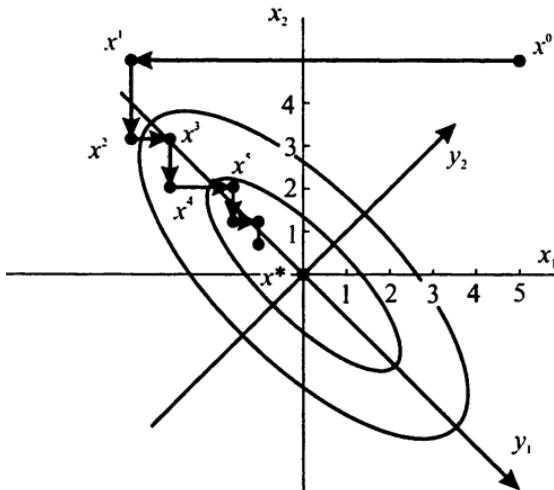


Таблица 8.2

Результаты расчетов в примере 8.3

Номер итерации	x_1	x_2	$f(x)$
0	5	5	450
1	-4	5	45
2	-4	3,2	28,8
3	-2,56	3,2	18,43
4	-2,56	2,05	11,8
5	-1,64	2,05	7,55
6	-1,64	1,31	4,83
7	-1,05	1,31	3,09
8	-1,05	0,84	1,98
9	-0,67	0,84	1,27
10	-0,67	0,54	0,81

Обратим внимание, что в примере 8.3 эффективность решения задачи методом циклического покоординатного спуска можно повысить. Для этого следует дополнить его алгоритм периодически повторяющимся поиском точки минимума в направлении $p^i = x^i - x^{i-2}$ из точек x^i . Так, например, если из точки x^4 провести исчерпывающий спуск в направлении $p^4 = x^4 - x^2$ (координаты точек приведены в таблице 8.2), то получим точку $(-2,2 \cdot 10^{-5}, 5,6 \cdot 10^{-3})^T$ расположенную значительно ближе к точке минимума $x^* = (0, 0)^T$, чем точки $x^5, x^6, \dots, x^{10}$.

Такой подход, состоящий в последовательном нахождении направлений убывания целевой функции и минимизации ее по этим направлениям, лежит в основе ряда алгоритмов. Рассмотрим один из них.

8.5. Метод Хука-Дживса

Эффективность прямого поиска точки минимума ограниченной снизу целевой функции можно повысить, если на каждом k -м шаге поиска последовательно выбирать направление спуска. Для этого на каждом k -м шаге выделяют предварительный этап *исследующего поиска*. Целью этого этапа

является выбор направления спуска путем исследования поведения целевой функции $f(x)$ в окрестности точки x^{k-1} найденной на предыдущем шаге. В результате исследующего поиска находится точка x^k , для которой $f(x^k) < f(x^{k-1})$. Направление спуска, завершающего k -й шаг поиска, определяется вектором $\bar{x}^k - x^{k-1}$. Такая стратегия поиска, предложенная в 1961 г., получила название *метода Хука-Дживса*. Это один из наиболее эффективных прямых методов.

Алгоритм метода Хука-Дживса на каждом шаге содержит две основные процедуры:

а) исследующий поиск в окрестности данной точки x для определения направления убывания функции $f(x)$. В результате получим точку $\bar{x}$;

б) перемещение в направлении убывания $(\bar{x} - x)$.

Поиск завершается, если в результате выполнения шага «а» получено, что $\bar{x} = x$.

Опишем один из возможных алгоритмов исследующего поиска – покоординатный поиск. Пусть задана точка x с приращениями по каждой координате Δ_j , $j = 1, \dots, n$.

Шаг 1. Положить $\bar{x} = x$, $j = 1$. Перейти к шагу 2.

Шаг 2. Сделать пробный шаг $y = \bar{x} - \Delta_j e^j$, где e^j – j -й базисный вектор. Если $f(\bar{x}) \leq f(y)$, то перейти к шагу 3, иначе – к шагу 4.

Шаг 3. Сделать пробный шаг $y = \bar{x} + \Delta_j e^j$. Если $f(\bar{x}) \leq f(y)$, то перейти к шагу 5, иначе – к шагу 4.

Шаг 4. Положить $\bar{x} = y$. Перейти к шагу 5.

Шаг 5. Положить $j = j + 1$. Если $j \leq n$, то перейти к шагу 2, иначе исследующий поиск окончен, т. е. получена точка $\bar{x}$, для которой $f(\bar{x}) < f(x)$, если $\bar{x} = x$.

В результате исследующего поиска может оказаться, что $\bar{x} = x$. Тогда исследующий поиск считается неудачным. Если при этом норма приращения $\Delta = (\Delta_1 \dots \Delta_n)$ мала, т. е. $\|\Delta\| < \varepsilon$, где ε – заданная точность, то полагают $x^\bullet = x$, $f(x^\bullet) = f(x)$. Если заданная точность не достигнута, то полагают $\Delta = \Delta / \gamma$ (где коэффициент дробления шага $\gamma > 1$) и повторяют исследующий поиск.

Полный алгоритм метода Хука–Дживса следующий.

Шаг 1. Выбрать начальную точку x^0 , вектор приращений $\Delta = (\Delta_1 \dots \Delta_n)$, коэффициент дробления шага $\gamma > 1$, параметр окончания поиска $\varepsilon > 0$. Перейти к шагу 2.

Шаг 2. Провести исследующий по координатный поиск из точки x^0 , т. е. найти точку $\bar{x}^0$. Если $\bar{x}^0 \neq x^0$, то перейти к шагу 4, иначе – к шагу 3.

Шаг 3. Проверка окончания поиска. Если $\|\Delta\| < \varepsilon$, то прекратить поиск и положить $x^* = x^0$. Иначе – положить $\Delta = \Delta / \gamma$ и перейти к шагу 2.

Шаг 4. Осуществить перемещение из точки $\bar{x}^0$ в направлении убывания $(\bar{x}^0 - x^0)$ в точку x^1

$$x^1 = x^0 + a_k(\bar{x}^0 - x^0)$$

и подобрать так называемый ускоряющий множитель $a_k > 0$, чтобы $f(x^1) < f(\bar{x}^0)$.

Часто $a_k \in [0, 1]$. С увеличением a_k увеличивается длина $a_k|\bar{x}^0 - x^0|$ шага спуска в направлении вектора $\bar{x}^0 - x^0$. Значение a_k можно подбирать из условия минимума функции $f(x)$ при смещении точки $\bar{x}^0$ в направлении этого вектора. В простейшем варианте метода Хука–Дживса значение a_k выбирают постоянным, обычно $a_k = 2$. В этом случае формула, по которой осуществляется спуск, имеет вид $x^1 = x^0 + 2(\bar{x}^0 - x^0) = 2\bar{x}^0 - x^0$.

8.6. Методы случайного поиска

Основой для этих методов служит итерационный процесс

$$x^{k+1} = x^k + \alpha_k \frac{\xi}{\|\xi\|}, \quad k = 0, 1, \dots \quad (8.6.1)$$

где $\alpha_k > 0$ – величина шага, $\xi = (\xi_1, \dots, \xi_n)$ – некоторая реализация n -мерного случайного вектора ξ .

Будем считать, что координаты вектора ξ – это независимые случайные величины, равномерно распределенные на отрезке $[-1, 1]$. На текущей итерации при помощи

генерирования случайных векторов ξ получаются точки, лежащие на гиперсфере радиуса α_k с центром в точке x^k . Если значение функции в полученной точке не меньше, чем в центре сферы, шаг считается неудачным. Если число неудачных шагов из данной точки достигает заданного числа M , поиск повторяется из той же точки с уменьшенным шагом, до тех пор пока шаг не станет меньше заданной точности. Если же значение функции в полученной точке меньше, чем в центре, шаг считается удачным, и полученное значение выбирается за новый центр поиска.

Приведем два возможных алгоритма метода случайного поиска. Они могут использоваться как самостоятельные минимизирующие процедуры или входить в состав других алгоритмов, например, использоваться для исследующего поиска в методе Хука-Дживса.

Алгоритм поиска с возвратом при неудачном шаге следующий.

Шаг 1. Выбрать параметр точности $\varepsilon > 0$, начальный шаг $\alpha > 0$, коэффициент уменьшения шага $\gamma > 1$, предельное число попыток M , точку x . Вычислить $f(x)$. Перейти к шагу 2.

Шаг 2. Положить счетчик числа неудачных попыток $j = 1$. Перейти к шагу 3.

Шаг 3. Получить реализацию случайного вектор ξ . Перейти к шагу 4.

Шаг 4. Найти пробную точку $y = x + \alpha \frac{\xi}{\|\xi\|}$, вычислить $f(y)$.

Перейти к шагу 5.

Шаг 5. Если $f(y) < f(x)$, то положить $x = y$, $f(x) = f(y)$ и перейти к шагу 4. Иначе – перейти к шагу 6.

Шаг 6. Положить $j = j + 1$. Если $j < M$, то перейти к шагу 3, иначе – к шагу 7.

Шаг 7. Если $\alpha < \varepsilon$, то поиск завершить, полагая $x^* = x$, $f(x^*) = f(x)$. Иначе – положить $\alpha = \alpha / \gamma$ и перейти к шагу 2.

Иллюстрация построения последовательности (6.8) с помощью описанного алгоритма для функции двух переменных приведена на рисунке 8.10, где пунктиром показаны неудачные попытки определения точки x^{k+1} из (7.6.1), не приводящие к уменьшению $f(x)$.

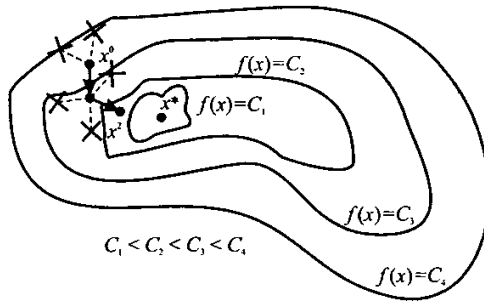


Рис. 8.10. Иллюстрация работы алгоритма с возвратом при неудачном шаге в E_2

На практике предельное число неудачных попыток M обычно полагают равным $3n$, где n – число переменных целевой функции.

Алгоритм метода случайного поиска наилучшей пробы следующий.

3 и 4.

Шаг 3. Получить реализаций случайного вектора $\xi: \xi^1, \dots, \xi^m$.

Шаг 4. Найти пробные точки $y^i = x + \alpha \frac{\xi^i}{\|\xi\|}$, $i = 1, \dots, m$,

вычислить $f(y^i)$.

Найти y^k из условия $f(y^k) = \min_i f(y^i)$ и положить $y = y^k$.

Этот алгоритм отличается от предыдущего только шагами.

Основная литература: 7 [101–121]

Контрольные вопросы:

1. Проблема минимизации многомерных задач.
2. Минимизация функций по правильному (регулярному) симплексу.
3. Минимизация функций при помощи нерегулярного симплекса.
4. Метод циклического покоординатного спуска.
5. Метод Хука-Дживса.
6. Методы случайного поиска.

9. УСЛОВНЫЙ ЭКСТРЕМУМ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ. НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ УСЛОВНОГО ЭКСТРЕМУМА

Пусть даны дважды непрерывно дифференцируемая целевая функция $f(x) = f(x_1, \dots, x_n)$ и функции ограничений $g_j(x) = g_j(x_1, \dots, x_n) = 0, j = 1, \dots, m$, определяющие множество допустимых решений U .

Требуется исследовать функцию $f(x)$ на экстремум, т. е. определить точки $x^* \in U$ ее локальных минимумов и максимумов

$$f(x^*) = \min_{x \in U} f(x), \quad f(x^*) = \max_{x \in U} f(x), \quad (9.1)$$

где $U = \{x: g_j(x) = 0, j = 1, \dots, m; m < n\}$. В первой части этой темы рассматриваются условные экстремумы с ограничениями только типа равенств.

9.1. Условный экстремум при ограничениях типа равенств

Определение. Функция

$$L(x, \lambda_0, \lambda) = \lambda_0 f(x) + \sum_{j=1}^m \lambda_j g_j(x) \quad (9.1.1)$$

называется *обобщенной функцией Лагранжа*, числа $\lambda_0, \lambda_1, \dots, \lambda_m$ – *множителями Лагранжа*. *Классической функцией Лагранжа* называется функция

$$L(x, \lambda) = f(x) + \sum_{j=1}^m \lambda_j g_j(x). \quad (9.1.2)$$

Определение. *Градиентами* обобщенной и классической функций Лагранжа по x называются вектор-столбцы, составленные из соответствующих частных производных первого порядка по $x_i, i = 1, \dots, n$:

$$\nabla, L(x, \lambda_0, \lambda) = \begin{pmatrix} \frac{\partial L(x, \lambda_0, \lambda)}{\partial x_1} \\ \cdots \\ \frac{\partial L(x, \lambda_0, \lambda)}{\partial x_n} \end{pmatrix}, \quad (9.1.3a)$$

$$\nabla, L(x, \lambda) = \begin{pmatrix} \frac{\partial L(x, \lambda)}{\partial x_1} \\ \cdots \\ \frac{\partial L(x, \lambda)}{\partial x_n} \end{pmatrix}. \quad (9.1.3б)$$

Определение. Вторым дифференциалом обобщенной (классической) функции Лагранжа $L(x, \lambda_0, \lambda)$ ($L(x, \lambda)$) называется функция

$$d^2 L(x, \lambda_0, \lambda) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 L(x, \lambda_0, \lambda)}{\partial x_i \partial x_j} dx_i dx_j. \quad (9.1.4a)$$

$$d^2 L(x, \lambda) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 L(x, \lambda)}{\partial x_i \partial x_j} dx_i dx_j. \quad (9.1.4б)$$

Определение. Первым дифференциалом ограничения $g_j(x)$ называется функция

$$dg_i(x) = \sum_{i=1}^n \frac{\partial g_i(x)}{\partial x_i} dx_i, \quad j = 1, \dots, m. \quad (9.1.5)$$

Пример 9.1. Выписать функции (9.1)–(9.5) для задачи поиска условного экстремума функции $f(x) = x_1^2 + x_2^2$ на множестве $U = \{x: x_2^2 - x_1 + 3 = 0\}$, заданном ограничением $g_j(x) = x_2^2 - x_1 + 3 = 0$.

Обобщенная функция Лагранжа

$$L(x, \lambda_0, \lambda_1) = \lambda_0(x_1^2 + x_2^2) + \lambda_1(x_2^2 - x_1 + 3).$$

Классическая функция Лагранжа

$$L(x, \lambda_1) = x_1^2 + x_2^2 + \lambda_1(x_2^2 - x_1 + 3).$$

Градиент функций Лагранжа

$$\nabla, L(x, \lambda_0, \lambda_1) = (2\lambda_0 x_1 - \lambda_1; 2\lambda_0 x_2 + 2\lambda_1 x_2)^T;$$

$$\nabla, L(x, \lambda_1) = (2x_1 - \lambda_1; 2x_2 + 2\lambda_1 x_2)^T.$$

Второй дифференциал функций Лагранжа

$$d^2 L(x, \lambda_0, \lambda_1) = 2\lambda_0 dx_1^2 + (2\lambda_0 + 2\lambda_1) dx_2^2;$$

$$d^2 L(x, \lambda_1) = 2dx_1^2 + (2 + 2\lambda_1) dx_2^2.$$

Первый дифференциал ограничения $dg_1(x) = -dx_1 + 2x_2 dx_2$.

Стратегия поиска локальных экстремумов

Находятся точки x^* локальных экстремумов с помощью необходимых и достаточных условий минимума и максимума первого и второго порядка при ограничениях типа равенств (порядок условий определяется порядком используемых производных). Вычисляются значения $f(x^*)$ в найденных точках локального экстремума.

Необходимые условия экстремума первого порядка. Пусть x^* – точка локального экстремума в задаче (9.1). Тогда найдутся числа $\lambda_0^*, \lambda_1^*, \dots, \lambda_m^*$, не равные одновременно нулю и такие, что выполняются следующие условия:

– стационарности обобщенной функции Лагранжа по x

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, \dots, n; \quad (9.1.6a)$$

– допустимости решения

$$g_j(x^*) = 0, \quad j = 1, \dots, m. \quad (9.1.6b)$$

Если при этом градиенты $\nabla g_i(x^*), \dots, \nabla g_m(x^*)$ в точке линейно независимы (выполняется условие регулярности), то $\lambda_0^* \neq 0$.

Важно отметить следующие моменты.

1. Условие (9.1.6a) может быть записано в векторной форме

$$\nabla L(x^*, \lambda_0^*, \lambda^*) = 0.$$

2. Система (9.6) содержит $(n + m)$ уравнений с $(n + m + 1)$ неизвестными $\lambda_0^*, \lambda_1^*, \dots, \lambda_m^*, x^*, \dots, x^*$. Точки x^* , удовлетворяющие системе при некоторых λ_0^*, λ_1^* называются *условно-стационарными*.

3. При решении задач проверка регулярности затруднена, так как точка x^* заранее неизвестна. Поэтому, как правило, рассматриваются два случая: $\lambda_0^* = 0$ и $\lambda_0^* \neq 0$. Если $\lambda_0^* \neq 0$, в системе (9.1.6a) полагают $\lambda_0^* = 1$. Это эквивалентно делению

системы уравнений (9.1.6a) на λ_0^* и замене $\frac{\lambda_j^*}{\lambda_0^*}$ на λ_j . При этом

обобщенная функция Лагранжа становится классической, а сама система (9.1.6) принимает вид

$$\frac{\partial L(x^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, \dots, n, \quad (9.1.7a)$$

$$g_j(x^*) = 0, \quad j = 1, \dots, m, \quad (9.1.7б)$$

где число уравнений равно числу неизвестных.

Определение. Точка экстремума, удовлетворяющая системе (9.1.6) при $\lambda_0^* \neq 0$ называется *регулярной*, а при $\lambda_0^* = 0$ – *нерегулярной*.

Случай $\lambda_0^* = 0$ отражает вырожденность ограничений. При этом в обобщенной функции Лагранжа исчезает член, содержащий целевую функцию, а в необходимых условиях экстремума не используется информация, представляемая градиентом целевой функции.

Необходимые условия экстремума второго порядка. Пусть x^* – регулярная точка минимума (максимума) в задаче (9.1) и имеется решение (x^*, λ^*) системы (9.1.7). Тогда второй дифференциал классической функции Лагранжа, вычисленный в точке (x^*, λ^*) , неотрицателен (неположителен):

$$d^2L(x^*, \lambda^*) \geq 0, (d^2L(x^*, \lambda^*) \leq 0) \quad (9.1.8)$$

для всех dx таких, что

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, \quad j = 1, \dots, m. \quad (9.1.9)$$

Достаточные условия экстремума. Пусть имеется точка (x^*, λ^*) , удовлетворяющая системе (9.1.7). Если в этой точке $d^2L(x^*, \lambda^*) > 0$ (соответственно $d^2L(x^*, \lambda^*) < 0$) для всех ненулевых dx таких, что

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, \quad j = 1, \dots, m,$$

то точка x^* является точкой локального минимума (максимума) задачи экстремума.

Достаточные и необходимые условия экстремума второго порядка проверяются в условно-стационарных точках, которые удовлетворяют системе (9.7) при $\lambda_0^* \neq 0$ или системе (9.8), так как для практики представляет интерес случай, когда в функции Лагранжа присутствует целевая функция, экстремум которой ищется.

Алгоритм решения задачи следующий.

Шаг 1. Составить обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda) = \lambda_0 f(x) + \sum_{j=1}^m \lambda_j g_j(x).$$

Шаг 2. Записать необходимые условия экстремума первого порядка

$$a) \frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, \dots, n;$$

$$б) g_j(x^*) = 0, \quad j = 1, \dots, m.$$

Шаг 3. Решить систему для двух случаев:

$$1) \lambda_0^* = 0;$$

$$2) \lambda_0^* \neq 0 \text{ (при этом поделить условие а) на } \lambda_0^* \text{ и заменить } \frac{\lambda_j^*}{\lambda_0^*} \text{ на } \lambda_j^*).$$

В результате найти условно стационарные точки x^* , выделив из них полученные при $\lambda_0^* \neq 0$ (они могут быть регулярными точками экстремума).

Шаг 4. Для выделенных на шаге 3 точек проверить достаточные условия экстремума:

а) записать выражение для второго дифференциала классической функции Лагранжа в точке (x^*, λ^*)

$$d^2 L(x^*, \lambda^*) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_i \partial x_j} dx_i dx_j;$$

б) записать систему дополнительных условий в точке x^*

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, \quad j = 1, \dots, m;$$

в) из последних условий выразить любые из m дифференциалов dx_i через остальные $(n-m)$ и подставить в $d^2 L(x^*, \lambda^*)$;

г) если $d^2 L(x^*, \lambda^*) > 0$ при ненулевых dx , то в точке x^* условный локальный минимум. Если $d^2 L(x^*, \lambda^*) < 0$ при ненулевых dx , то в точке x^* условный локальный максимум. Если достаточные условия экстремума не выполняются, следует проверить выполнение необходимых условий второго порядка,

следуя аналогичной процедуре. Если они выполняются, то требуется дополнительное исследование, в противном случае в точке x^* нет условного экстремума.

Шаг 5. Вычислить значения функции в точках условного экстремума.

Приведем в виде таблицы результаты исследования на наличие в точке x^* локальных экстремумов при соблюдении необходимых условий первого порядка

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*)}{\partial x_i} = 0, \quad i = 1, \dots, n,$$

$$g_j(x^*) = 0, \quad j = 1, \dots, m,$$

причем $\lambda_0^* \neq 0$ (табл. 9.1).

Таблица 9.1

Результаты исследования локальных экстремумов в точке x^* при соблюдении необходимых условий первого порядка

№	$d^2 L(x^*, \lambda^*)$	$g_j(x^*)=0, \\ j=1, \dots, m,$	Тип условноста- ционарной точки x^*
1	> 0	$0, dx \neq 0$	Условный локальный минимум
2	< 0	$0, dx \neq 0$	Условный локальный максимум
3	≥ 0	0	Может быть условный локальный минимум, требуется дополнительное исследование
4	≤ 0	0	Может быть условный локальный максимум, требуется дополнительное исследование
5	$= 0$	0	Требуется дополнительное исследование
6	$> 0, < 0$	0	Нет экстремума

Пример 9.2. Найти $f(x) = x_1^2 + x_2^2 \rightarrow \text{extr}$ при условии

$$g(x^*) = x_1 + x_2 = 0.$$

Проверим условие регулярности. Так как $\nabla g_1(x) = (1, 1)^T \neq 0$, то условие выполняется. Поэтому будем пользоваться классической функцией Лагранжа.

1. Составим функцию Лагранжа

$$L(x, \lambda_1) = x_1^2 + x_2^2 + \lambda_1(x_1 + x_2 - 2).$$

2. Выпишем необходимые условия экстремума первого порядка

$$а) \frac{\partial L(x, \lambda_1)}{\partial x_1} = x_1 + x_2 = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = 2x_2 + \lambda_1 = 0,$$

$$\text{откуда } x_1 = -\frac{\lambda_1}{2}, \quad x_2 = -\frac{\lambda_1}{2};$$

$$б) g_1(x) = x_1 + x_2 - 2 = 0.$$

3. Решение системы: $x_1^2 = x_2^2 = 1$, $\lambda_1^* = -2$ – условно стационарная точка.

4. Проверим достаточные условия экстремума:

$$а) d^2L(x^*, \lambda_1^*) = 2dx_1^2 + 2dx_2^2, \text{ т. к. } \frac{\partial g_1(x)}{\partial x_1} = \frac{\partial g_1(x)}{\partial x_2} = 1;$$

$$б) dg_1(x^*) = dx_1 + dx_2 = 0, \text{ т. к. } \frac{\partial g_1(x)}{\partial x_1} = \frac{\partial g_1(x)}{\partial x_2} = 1;$$

в) выразим дифференциал dx_1 через dx_2 : $dx_1 = -dx_2$ и подставим в d^2L ;

г) так как $d^2L(x^*, \lambda_1^*) = 4dx_2^2 > 0$ при $dx_2 \neq 0$, то в точке $x^* = (1, 1)^T$ регулярный локальный условный минимум.

5. Вычислим значение функции в точке условного экстремума: $f(x^*) = 2$.

Графическое решение задачи приведено на рис. 9.1.

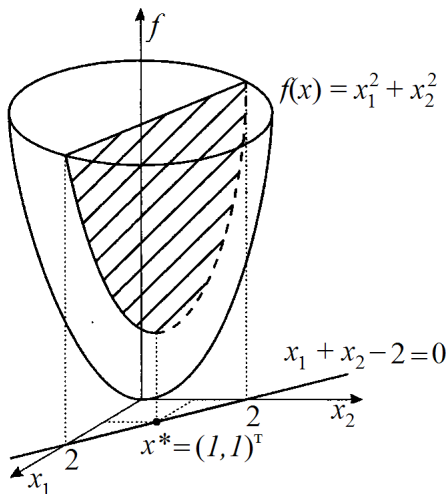


Рис. 9.1. Иллюстрация к решению задачи из примера 8.2

Пример 9.3. Найти условный экстремум

$$f(x) = x_1 + x_2 \rightarrow \text{extr},$$

$$g_1(x) = x_1^2 + x_2^2 - 2 = 0.$$

Проверим условие регулярности. Так как $\nabla g_1(x) = (2x_1, 2x_2)^T \neq 0$ для всех $x \in U$, то условие регулярности выполняется. Поэтому будем пользоваться классической функцией Лагранжа.

1. Составим функцию Лагранжа

$$L(x, \lambda_1) = x_1 + x_2 + \lambda_1(x_1^2 + x_2^2 - 2).$$

2. Выпишем необходимые условия экстремума первого порядка:

$$\text{а) } \frac{\partial L(x, \lambda_1)}{\partial x_1} = 1 + 2\lambda_1 x_1 = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = 1 + 2\lambda_1 x_2 = 0,$$

$$\text{откуда } x_1 = -\frac{1}{2\lambda_1}, \quad x_2 = -\frac{1}{2\lambda_1};$$

$$\text{б) } g_1(x) = x_1^2 + x_2^2 - 2 = 0.$$

3. Решением системы являются две условно стационарные точки

$$A: x_1^* = x_2^* = 1, \lambda_1^* = -\frac{1}{2} \quad B: x_1^* = x_2^* = -1, \lambda_1^* = \frac{1}{2}.$$

4. Проверим достаточные условия экстремума:

$$a) d^2L(x^*, \lambda_1^*) = 2\lambda_1^* dx_1^2 + 2\lambda_1^* dx_2^2, \text{ т. к.}$$

$$\frac{\partial^2 L(x, \lambda_1)}{\partial x_1^2} = \frac{\partial^2 L(x, \lambda_1)}{\partial x_2^2} = 2\lambda_1^*,$$

$$\frac{\partial^2 L(x, \lambda_1)}{\partial x_1^2} = \frac{\partial^2 L(x, \lambda_1)}{\partial x_2^2} = 2\lambda_1^*;$$

$$б) dg(x^*) = 2x_1^* dx_1 + 2x_2^* dx_2 = 0, \text{ т. к. } \frac{\partial g_1(x)}{\partial x_1} = 2x_1, \frac{\partial g_1(x)}{\partial x_2} = 2x_2;$$

$$в) \text{ Иследуем точку } A. \text{ Получаем } dg_1(A) = 2dx_1 + 2dx_2 = 0,$$

откуда $dx_1 = -dx_2$.

С учетом полученного соотношения $d^2L(A) = -dx_1^2 - dx_2^2 = -2dx_2^2 < 0$ при $dx_2 \neq 0$. Поэтому в точке $x^* = (1, 1)^T$ имеем регулярный условный локальный максимум.

Иследуем точку B . Получаем $dg_1(B) = -2dx_1 - 2dx_2 = 0$, откуда $dx_1 = -dx_2$.

С учетом полученного соотношения $d^2L(B) = dx_1^2 + dx_2^2 = 2dx_2^2 > 0$ при $dx_2 \neq 0$.

Поэтому в точке $x^* = (-1, -1)^T$ имеем регулярный условный локальный минимум.

5. Подсчитаем значения функции в точках экстремума: $f(A) = 2, f(B) = -2$.

Графическое решение задачи представлено на рис. 9.2.

Пример 9.4. Найти условный экстремум в задаче

$$f(x) = x_1 \rightarrow \text{extr},$$

$$g_1(x) = x_2^2 - x_1^3 = 0.$$

Проверим условие регулярности. Так как $\nabla g_1(x) = (-3x_1^2, 2x_2)^T = 0$ в точке $x^\bullet = (0, 0)^T$, то условие регулярности не выполняется. Будем пользоваться алгоритмом с использованием обобщенной функции Лагранжа.

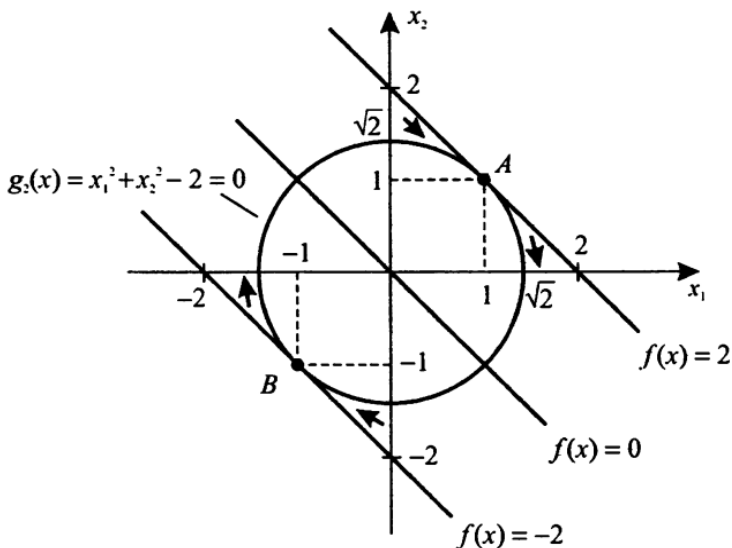


Рис. 9.2. Иллюстрация к решению задачи из примера 9.3

1. Составим обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda_1) = \lambda_0 x_1 + \lambda_1 (x_2^3 - x_1^3).$$

2. Выпишем необходимые условия экстремума первого порядка:

$$\text{а) } \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_1} = \lambda_0 - 3\lambda_1 x_1^2 = 0, \quad \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_2} = 2\lambda_1 x_2 = 0,$$

$$\text{б) } g_1(x) = x_2^2 - x_1^3 = 0.$$

3. Решим систему для двух случаев.

В первом случае $\lambda_0 = 0$. Тогда $\lambda_1 \neq 0$, так как все множители Лагранжа не могут быть одновременно равными нулю. Отсюда $x_1^* = x_2^* = 0$, $\lambda_0^* = 0$.

Во втором случае $\lambda_0 \neq 0$. Поделим уравнения системы, полученной на втором шаге, на λ_0 с заменой $\frac{\lambda_1}{\lambda_0}$ на λ_1

$$1 - 3\lambda_1 x_1^2 = 0, \quad 2\lambda_1 x_2 = 0;$$

$$x_2^2 - x_1^3 = 0.$$

Рассмотрим второе уравнение. Если $\lambda_0 = 0$, то система несовместна. Если $x_2 = 0$, то $x_1 = 0$ и система тоже несовместна. Таким образом, применение классической функции Лагранжа не дает результата.

4. Так как $\lambda_0^* = 0$, достаточные условия экстремума не проверяются. Точка x^* со значением целевой функции $f(x^*) = 0$ является точкой нерегулярного локального и глобального минимума (рис. 8.3).

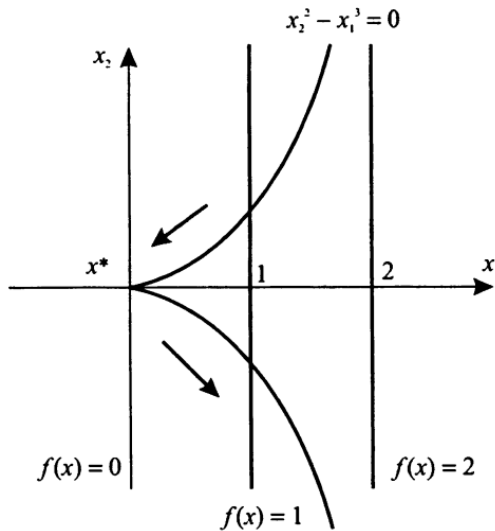


Рис. 9.3. Иллюстрация к решению задачи из примера 9.4

Пример 9.5. Найти условный экстремум

$$f(x) = x_1^2 + x_2^2 \rightarrow \text{extr},$$

$$g_1(x) = (x_1 - 1)^2 + x_2^2 - 4 = 0.$$

Будем следовать алгоритму, не проверяя условия регулярности.

1. Составим обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda_1) = \lambda_0(x_1^2 + x_2^2) + \lambda_1[(x_1 - 1)^2 + x_2^2 - 4].$$

2. Выпишем необходимые условия экстремума первого порядка:

$$\text{а) } \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_1} = 2\lambda_0 x_1 + 2\lambda_1(x_1 - 1) = 0, \quad \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_2} = 2\lambda_0 x_2 + 2\lambda_1 x_2 = 0,$$

$$2\lambda_1 x_2 = 0,$$

$$\text{б) } g_1(x) = (x_1 - 1)^2 + x_2^2 - 4 = 0.$$

3. Решим систему для двух случаев.

В первом случае $\lambda_0 = 0$. Тогда из п. 2 следует

$$2\lambda_1(x_1 - 1) = 0, \quad 2\lambda_1 x_2 = 0,$$

$$(x_1 - 1)^2 + x_2^2 - 4 = 0.$$

Так как $\lambda_0 = 0$, то $\lambda_1 \neq 0$; поэтому из первых двух уравнений $x_1 = 1$, $x_2 = 0$. Однако при этом ограничение не выполняется: $g_1(x) = -4 \neq 0$, следовательно, система несовместна.

Во втором случае $\lambda_0 \neq 0$. Поделим уравнения на λ_0 и заменим $\frac{\lambda_1}{\lambda_0}$ на λ_1 :

$$\text{а) } \frac{\partial L(x, \lambda_1)}{\partial x_1} = 2x_1 + 2\lambda_1(x_1 - 1) = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = 2x_2 + 2\lambda_1 x_2 = 0,$$

$$\text{б) } g_1(x) = (x_1 - 1)^2 + x_2^2 - 4 = 0.$$

Рассмотрим второе уравнение. Если $x_2 = 0$, то из третьего уравнения следует $x_1 = 3$, $x_1 = -1$, а из первого $\lambda_1 = -\frac{3}{2}$, $\lambda_1 = -\frac{1}{2}$ соответственно.

Если же $x_1 = -1$, то первое уравнение имеет вид $2 = 0$, т. е. система несовместна. Таким образом, найдены две условно стационарные точки

$$A: x_1^* = 3, x_2^* = 0, \lambda_1^* = -\frac{3}{2}; \quad B: x_1^* = -1, x_2^* = 0, \lambda_1^* = -\frac{1}{2}.$$

4. Проверим достаточные условия экстремума:

$$a) d^2L(x^*, \lambda_1^*) = 2(1 + \lambda_1^*)dx_1^2 + 2(1 + \lambda_1^*)dx_2^2;$$

$$б) dg_1(x^*) = 2(x_1^* - 1)dx_1 + 2x_2^*dx_2 = 0;$$

в, г) исследуем точку A . $dg_1(A) = 4dx_1 + 0 = 0$, откуда $dx_1 = 0$ и

$d^2L(B) = -dx_1^2 - dx_2^2 = -dx_2^2 < 0$, если $dx_2 \neq 0$. Поэтому в точке A регулярный условный максимум.

Исследуем точку B . $Dg_1(B) = 4dx_1 + 0 = 0$, откуда $dx_1 = 0$ и $d^2L(B) = dx_1^2 + dx_2^2 > 0$ при $dx_2 \neq 0$. Поэтому в точке B регулярный условный минимум.

5. Подсчитаем значения функции в точках экстремума:

$$f(A) = 9, f(B) = 9.$$

Графическое решение задач представлено на рис. 9.4.

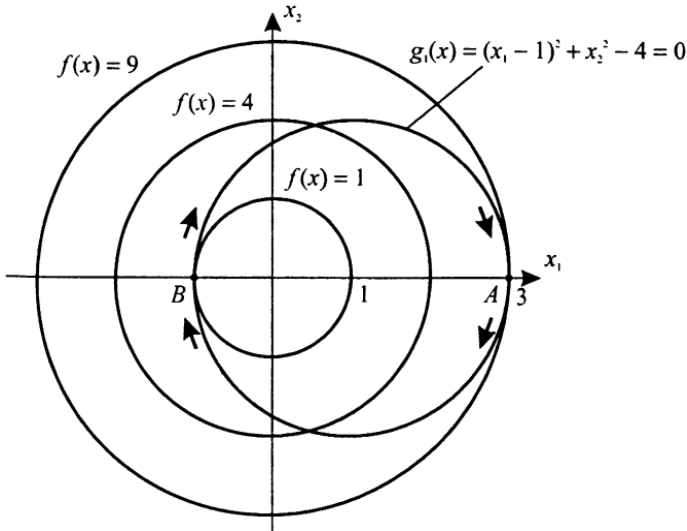


Рис. 9.4. Графическое решение задачи из примера 9.5

Пример 9.6. Найти условный экстремум в задаче

$$f(x) = x_1^2 - x_2^2 = \rightarrow extr,$$

$$g_I(x) = x_1^2 + x_2^2 - I = 0.$$

1. Составим обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda_1) = \lambda_0(x_1^2 - x_2^2) + \lambda_1(x_1^2 + x_2^2 - 1).$$

2. Запишем необходимые условия экстремума первого порядка:

$$a) \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_1} = 2\lambda_0 x_1 + 2\lambda_1 x_1 = 0, \quad \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_2} = -2\lambda_0 x_2 + 2\lambda_1 x_2 = 0,$$

$$+ 2\lambda_1 x_2 = 0,$$

$$б) g_I(x) = x_1^2 + x_2^2 - 1 = 0.$$

3. Решим систему для двух случаев.

В первом случае $\lambda_0 = 0$. Тогда $\lambda_1 \neq 0$, поэтому из первых двух уравнений следует $x_1 = x_2 = 0$. Однако условие «б» при этом не выполняется.

Следовательно, система несовместна.

Во втором случае $\lambda_0 \neq 0$. Поделим уравнения на λ_0 и заменим $\frac{\lambda_1}{\lambda_0}$ на λ_1 :

$$a) \frac{\partial L(x, \lambda_1)}{\partial x_1} = 2x_1(1 + \lambda_1) = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = 2x_2(-1 + \lambda_1) = 0,$$

$$б) g_I(x) = x_1^2 + x_2^2 - 1 = 0.$$

Пусть $\lambda_1 = -1$, тогда $x_2 = 0$, $x_1 = \pm 1$. Пусть $\lambda_1 = 1$, тогда $x_1 = 0$, $x_2 = \pm 1$. Других решений система не имеет. Таким образом, имеем четыре условно стационарные точки

$$A: x_1^* = 1, x_2^* = 0, \lambda_1^* = -1; \quad B: x_1^* = -1, x_2^* = 0, \lambda_1^* = -1;$$

$$C: x_1^* = 0, x_2^* = 1, \lambda_1^* = 1; \quad D: x_1^* = 0, x_2^* = -1, \lambda_1^* = 1.$$

4. Проверим достаточные условия экстремума

а) $d^2L(x^\bullet, \lambda_1^\bullet) = 2(1 + \lambda_1^\bullet)dx_1^2 + 2(\lambda_1^\bullet - 1)dx_2^2$;

б) $dg_1(x^\bullet) = 2x_1^\bullet dx_1 + 2x_2^\bullet dx_2 = 0$;

в, г) исследуем точку A . $dg_1(A) = 2dx_1 = 0$, откуда получаем $dx_1 = 0$ и $d^2L(A) = -4Adx_2^2 < 0$, если $dx_2 \neq 0$. Поэтому в точке A регулярный локальный условный максимум.

Исследуем точку B . $Dg_1(B) = -2dx_1 = 0$, откуда $dx_1 = 0$ и $d^2L(B) = -4dx_2^2 < 0$ при $dx_2 \neq 0$. Поэтому в точке B регулярный локальный условный максимум.

Исследуем точку C . $dg_1(C) = 2dx_2 = 0$, откуда $dx_2 = 0$ и $d^2L(C) = 4dx_1^2 > 0$ при $dx_1 \neq 0$. Поэтому в точке C регулярный локальный условный минимум.

Исследуем точку D . $Dg_1(D) = -2dx_2 = 0$, откуда $dx_2 = 0$ и $d^2L(D) = 4dx_1^2 > 0$ при $dx_1 \neq 0$. Поэтому в точке D регулярный локальный условный минимум.

4. Вычислим значения функции в точках условного экстремума

$$f(A) = f(B) = 1; f(C) = f(D) = -1.$$

Графическое решение задач представлено на рис. 9.5.

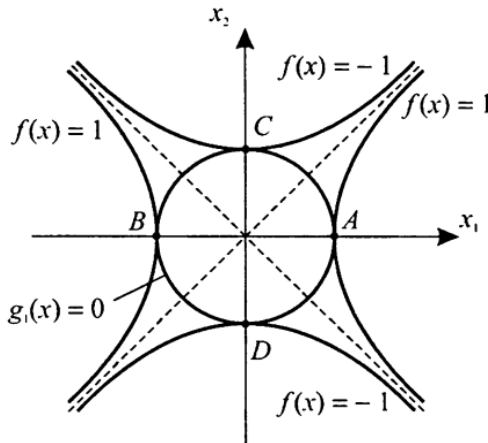


Рис. 9.5. Графическое решение задачи из примера 9.6

Пример 9.7. Найти условный экстремум в задаче

$$f(x) = \frac{1}{2} [(x_1-1)^2 + x_2^2] \rightarrow \text{extr},$$

$$g_1(x) = -x_1 + \alpha x_2^2 = 0.$$

при различных $\alpha > 0$.

Проверим условие регулярности. Так как $\nabla g_1(x) = (-1, 2\alpha x_2)^T \neq 0$ ни в одной точке множества U , то условие выполняется. Поэтому воспользуемся классической функцией Лагранжа.

1. Составим функцию Лагранжа

$$L(x, \lambda_1) = \frac{1}{2} [(x_1-1)^2 + x_2^2] + \lambda_1(-x_1 + \alpha x_2^2).$$

2. Запишем необходимые условия экстремума первого порядка:

$$а) \frac{\partial L(x, \lambda_1)}{\partial x_1} = x_1 - 1 - \lambda_1 = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = x_2 + 2\alpha\lambda_1 x_2 = 0;$$

$$б) g_1(x) = -x_1 + \alpha x_2^2 = 0.$$

3. Из второго уравнения следует $x_2(1 + 2\alpha\lambda_1) = 0$. Если $x_2 = 0$, то $x_1 = 0$, $\alpha\lambda_1 = -1$. Если, $\lambda_1 = -\frac{1}{2\alpha}$, то $x_1 = 1 + \lambda_1 = 1 - \frac{1}{2\alpha} =$

$$\frac{2\alpha - 1}{2\alpha} \text{ и при этом } x_2 = \pm \sqrt{\frac{2\alpha - 1}{2\alpha^2}}.$$

При $0 < \alpha < \frac{1}{2}$ решения не существует. Таким образом, имеются три условно стационарные точки (рис. 9.6)

$$A: x_1^* = 1, x_2^* = 0, \lambda_1^* = -1; B: x_1^* = \frac{2\alpha - 1}{2\alpha}, x_2^* = \sqrt{\frac{2\alpha - 1}{2\alpha^2}},$$

$$\lambda_1^* = -\frac{1}{2\alpha};$$

$$C: x_1^* = \frac{2\alpha - 1}{2\alpha}, x_2^* = -\sqrt{\frac{2\alpha - 1}{2\alpha^2}}, \lambda_1^* = -\frac{1}{2\alpha}.$$

3. Проверим достаточные условия экстремума:

$$a) d^2L(x^*, \lambda_1^*) = dx_1^2 + (l + 2\alpha\lambda_1^*)dx_2^2;$$

$$б) dg_l(x^*) = -dx_1 + 2\alpha x_2^2 dx_2 = 0;$$

$$в) \text{ выразим из последнего пункта } dx_1: dx_1 = 2\alpha x_2^2 dx_2;$$

$$г) d^2L(x^*, \lambda_1^*) = ((4\alpha x_2^2)^2 + (l + 2\alpha\lambda_1^*))dx_2^2.$$

$$\text{Исследуем точку } A \quad d^2L(x^*, \lambda_1^*) = (l - 2\alpha)dx_2^2.$$

При $0 < \alpha < \frac{1}{2}$ получаем, что $d^2L(x^*, \lambda_1^*) > 0$ при $dx_2 \neq 0$ и в точке A условный локальный минимум.

При $\alpha > \frac{1}{2}$ находим, что $d^2L(x^*, \lambda_1^*) < 0$ при $dx_2 \neq 0$ и в точке A условный локальный максимум.

При $\alpha = \frac{1}{2}$ получаем, что $d^2L(x^*, \lambda_1^*) = 0$ при любых dx_2 , и требуется дополнительное исследование.

Для этого воспользуемся *методом исключения переменных*. Выразим из уравнения ограничения одну из координат $x_1 = \frac{1}{2}x_2^2$. Подставим полученное выражение в целевую функцию

$$f(x) = \frac{1}{2}[(x_1 - 1)^2 + x_2^2] = \frac{1}{8}x_2^4 + \frac{1}{2}x_2^2 + \frac{1}{2}.$$

Исследуем полученную функцию на безусловный экстремум

$$\frac{df}{dx_2} = \frac{1}{2}x_2^3 + x_2 = 0,$$

отсюда $x_2^* = 0$ – стационарная точка. В этой точке первая отличная от нуля производная $f^{(4)}(x^*) = 3 > 0$ является четной и положительной. Следовательно, в

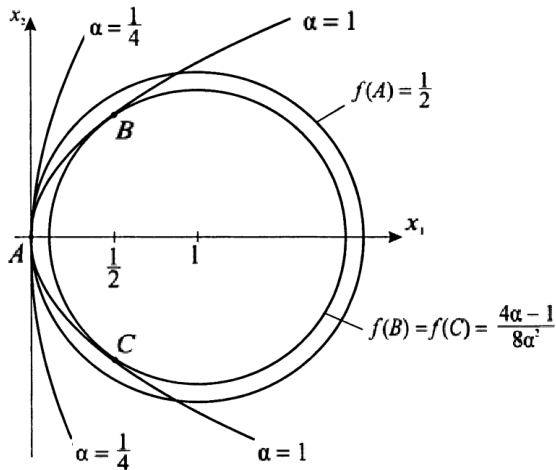


Рис. 9.6. Иллюстрация к решению задачи из примера 9.7

точке $x_2^\bullet$ локальный минимум. С учетом связи $x_1^\bullet = \frac{1}{2}(x_2^\bullet)^2 = 0$ получаем, что в точке $x^\bullet = (0, 0)^\top$ условный локальный минимум.

Рассмотрим точки B и C . $d^2L(x^\bullet, \lambda_1^\bullet) = 16\alpha x_2 \frac{2\alpha - 1}{2\alpha^2} dx_2^2 = 8(2\alpha - 1)dx_2^2$.

При $\alpha = \frac{1}{2}$ имеем $x_1^\bullet = x_2^\bullet = 0$ и $d^2L(x^\bullet, \lambda_1^\bullet) = 0$ при всех dx_2 . Значит, требуется дополнительное исследование, которое уже было выполнено.

При $\alpha > \frac{1}{2}$ имеем $d^2L(x^\bullet, \lambda_1^\bullet) > 0$ при $dx_2 \neq 0$ и в точках B и C условный локальный минимум.

4. Значение функции в точках экстремума

$$f(A) = \frac{1}{2}, f(B) = f(C) = \frac{4\alpha - 1}{8\alpha^2}.$$

Иллюстрация к решению задачи представлена на рис. 9.6.

9.2. Условный экстремум при ограничениях типа неравенств

Пусть требуется, исследовать функцию $f(x)$ на экстремум, т. е. определить точки $x^* \in U$ ее локальных минимумов и максимумов

$$f(x^*) = \min_{x \in U} f(x), \quad f(x^*) = \max_{x \in U} f(x), \quad (9.2.1)$$

где $U = \{x: g_j(x) \leq 0, j = 1, \dots, m\}$. Рассмотрим условный экстремум с ограничениями типа неравенств.

Стратегия решения задачи

Находятся точки x^* локального экстремума с помощью необходимых и достаточных условий минимума и максимума первого и второго порядка при ограничениях типа неравенств. Вычисляются значения $f(x^*)$ в найденных точках локального экстремума.

Необходимые условия минимума (максимума) первого порядка. Пусть x^* – точка локального минимума (максимума) в задаче (9.2.1). Тогда найдется такое число $\lambda_0^* \geq 0$ и вектор $\lambda^* = (\lambda_1^*, \dots, \lambda_m^*)^T$, не равные одновременно нулю и такие, что выполняются следующие условия:

– стационарности обобщенной функции Лагранжа по x

$$\frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_i} = 0, \quad i = 1, \dots, n; \quad (9.2.2a)$$

– допустимости решения

$$g_j(x^*) \leq 0, \quad j = 1, \dots, m; \quad (9.2.2б)$$

– неотрицательности для условного минимума

$$\lambda_j^* \geq 0, \quad j = 1, \dots, m \quad (9.2.2в)$$

(условие неположительности для условного максимума $\lambda_j^* \leq 0$, $j = 1, \dots, m$);

– дополняющей нежесткости

$$\lambda_j^\bullet g_j(x^\bullet) = 0, j = 1, \dots, m. \quad (9.2.2r)$$

Если при этом градиенты активных (т. е. $g_j(x^\bullet) = 0$) в точке $x^\bullet$ ограничений линейно независимы (выполняется условие регулярности), то $\lambda_0^\bullet \neq 0$. Точки $x^\bullet$ удовлетворяющие системе (9.2.2), называются *условно стационарными*.

Необходимо обратить внимание на следующее.

1. В отличие от ограничений типа равенств необходимые условия экстремума первого порядка формулируются отдельно для максимума и минимума. Эти условия впервые были доказаны Куном и Таккером.

2. Если в задаче ограничения записаны в форме $g_j(x^\bullet) \geq 0$, то их необходимо переписать как $-g_j(x^\bullet) \leq 0$ (8.2.2).

Обозначим через J_a множество индексов ограничений, активных в точке $x^\bullet$.

3. При $\lambda_0^\bullet \neq 0$ необходимые условия минимума (максимума) первого порядка для выпуклых функций $f(x)$, $g_j(x)$, $j = 1, \dots, m$, $(-f(x)$, $g_j(x)$, $j = 1, \dots, m$) будут одновременно и достаточными условиями глобального минимума (глобального максимума).

4. Из условия дополняющей нежесткости следует, что если ограничение в точке $x^\bullet$ пассивное, т. е. $g_j(x^\bullet) < 0$, то $\lambda_j^\bullet = 0$, а если активное, т. е. $g_j(x^\bullet) = 0$, то $\lambda_j^\bullet \geq 0$ (для минимума) и $\lambda_j^\bullet \leq 0$ (для максимума).

Достаточные условия минимума (максимума) первого порядка. Пусть имеется точка $(x^\bullet, \lambda^\bullet)$, удовлетворяющая системе (9.2.2) при $\lambda_0^\bullet \neq 0$, число активных ограничений в точке $x^\bullet$ совпадает с числом n переменных (при этом условие регулярности выполняется). Если $\lambda_j^\bullet > 0$ для всех $j \in J_a$, то $x^\bullet$ – точка условного локального минимума. Если $\lambda_j^\bullet < 0$ для всех $j \in J_a$, то $x^\bullet$ – точка условного локального максимума в задаче (9.2.1).

Необходимые условия минимума (максимума) второго порядка. Пусть x^* – точка локального минимума (максимума) в задаче (9.2.1) и имеется решение (x^*, λ^*) системы (9.2.2). Тогда второй дифференциал классической функции Лагранжа, вычисленный в точке (x^*, λ^*) , неотрицателен (неположителен)

$$d^2L(x^*, \lambda^*) \geq 0 \quad (d^2L(x^*, \lambda^*) \leq 0) \quad (9.2.3)$$

для всех dx таких, что

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, j \in J_a, \lambda_j^* > 0, (\lambda_j^* < 0);$$

$$dg_j(x^*) < 0, j \in J_a, \lambda_j^* = 0.$$

Достаточные условия минимума (максимума) второго порядка. Пусть имеется точка (x^*, λ^*) , удовлетворяющая системе (9.2.2) при $\lambda_0^* \neq 0$. Если в этой точке $d^2L(x^*, \lambda_1^*) > 0$ ($d^2L(x^*, \lambda_1^*) < 0$) для всех ненулевых dx таких, что

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, j \in J_a, \lambda_j^* > 0, (\lambda_j^* < 0);$$

$$dg_j(x^*) < 0, j \in J_a, \lambda_j^* = 0,$$

то x^* – точка локального минимума (максимума) в задаче (9.2.1).

Алгоритм решения задачи.

Шаг 1. Составить обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda) = \lambda_0 f(x) + \sum_{j=1}^m \lambda_j g_j(x).$$

Шаг 2. Записать необходимые условия минимума (максимума) первого порядка:

$$a) \frac{\partial L(x^*, \lambda_0, \lambda)}{\partial x_i} = 0, i = 1, \dots, n;$$

$$б) g_j(x^*) \leq 0, j = 1, \dots, m;$$

в) $\lambda_j^\bullet \geq 0, j = 1, \dots, m$ (для минимума), $\lambda_j^\bullet \leq 0, j = 1, \dots, m$ (для максимума);

г) $\lambda_j^\bullet g_j(x^\bullet) = 0, j = 1, \dots, m$.

Шаг 3. Решить систему для двух случаев

1) $\lambda_0^\bullet = 0$;

2) $\lambda_0^\bullet \neq 0$ (при этом поделить условия, записанные на шаге 2, на $\lambda_0^\bullet$ и заменить $\frac{\lambda_j^\bullet}{\lambda_0^\bullet}$).

В результате найти условно стационарные точки $x^\bullet$, выделив из них полученные при $\lambda_0^\bullet \neq 0$ (они могут быть регулярными точками экстремума).

В каждом из двух случаев следует начинать с рассмотрения $2m$ вариантов удовлетворения условия «г» – дополняющей нежесткости.

Шаг 4. Для выделенных на шаге 3 точек проверить достаточные условия экстремума первого или второго порядка.

Для проверки достаточных условий первого порядка следует:

а) определить число l активных в точке $x^\bullet$ ограничений;

б) если $l = n$ и $\lambda_j^\bullet > 0$ для всех $j \in J_a$, то в точке $x^\bullet$ локальный минимум; если $l = n$ и $\lambda_j^\bullet < 0$ для всех $j \in J_a$, то в точке $x^\bullet$ локальный максимум. Если $l < n$ или соответствующие множители Лагранжа не удовлетворяют достаточным условиям первого порядка, то проверить достаточные условия второго порядка.

Для проверки достаточных условий второго порядка следует:

а) записать выражение для второго дифференциала классической функции Лагранжа в точке $(x^\bullet, \lambda^\bullet)$

$$d^2 L(x^\bullet, \lambda^\bullet) = \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 L(x^\bullet, \lambda^\bullet)}{\partial x_i \partial x_j} dx_i dx_j ;$$

б) записать условия, накладываемые на первые дифференциалы активных ограничений

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i = 0, j \in J_a, \lambda_j^* > 0, (\lambda_j^* < 0); \quad (9.2.4)$$

$$dg_j(x^*) = \sum_{i=1}^n \frac{\partial g_j(x^*)}{\partial x_i} dx_i \leq 0, j \in J_a; \lambda_j^* = 0;$$

в) исследовать знак второго дифференциала функции Лагранжа для ненулевых dx , удовлетворяющих системе (9.2.4). Если $d^2L(x^*, \lambda^*) > 0$, то в точке x^* условный локальный минимум.

Если достаточные условия первого и второго порядка не выполняются, следует проверить выполнение необходимых условий второго порядка, следуя аналогичной процедуре. Если они выполняются, то требуется дополнительное исследование, а если нет, то в точке x^* нет условного экстремума.

Шаг 5. Вычислить значения функции в точках условного экстремума.

Условия экстремума в исходной задаче (9.2.1) приведены в табл. 9.2 и 9.3.

Таблица 9.2

Необходимые и достаточные условия первого порядка в задаче поиска условного экстремума при ограничениях

Необходимые условия первого порядка				Достаточные условия первого порядка $\lambda_0^* \neq 0$		
№	$\nabla, L(x^*, \lambda_0^*, \lambda^*);$ $\lambda_j^* g_j(x^*),$ $j = 1, \dots, m$	$g_j(x^*),$ $j = 1, \dots, m$	$\lambda_0^* \geq 0,$ $\lambda_j^*,$ $j = 1, \dots, m$	Число l	$\lambda_j^2,$ $j \in J_a$	Тип условно-стационарной точки x^*
1	0	≤ 0	≥ 0	n	> 0	условный локальный минимум
2	0	≤ 0	≤ 0	n	< 0	условный локальный максимум

Таблица 9.3

Необходимые и достаточные условия второго порядка в задаче поиска условного экстремума при ограничениях типа неравенств

№	$d^2(x^*, \lambda^*)$	$dg_j(x^*),$ $j \in J_a,$ $\lambda_j^* > 0$	$dg_j(x^*),$ $j \in J_a,$ $\lambda_j^* < 0$	$dg_j(x^*),$ $j \in J_a,$ $\lambda_j^* = 0$	Тип условно- стационарной точки x^*
1	> 0	$0, dx \neq 0$	–	≤ 0	условный локальный минимум
2	< 0	–	$0, dx = 0$	≤ 0	условный локальный максимум
3	≥ 0	0	–	≤ 0	может быть локальный минимум, требуется дополнительное исследование
4	≤ 0	–	0	≤ 0	может быть локальный максимум, требуется дополнительное исследование
5	$= 0$	0	–	≤ 0	требуется дополнительное исследование
6	$= 0$	–	0	≤ 0	требуется дополнительное исследование
7	$> 0, < 0$	0	–	≤ 0	нет экстремума
8	$> 0, < 0$	–	0	≤ 0	нет экстремума

Пример 9.8. Найти условный экстремум

$$f(x) = x_1^2 + x_2^2 \rightarrow \text{extr},$$

$$g_1(x) = x_1 + x_2 - 2 \leq 0.$$

1. Составим обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda_1) = \lambda_0(x_1^2 + x_2^2) + \lambda_1(x_1 + x_2 - 2).$$

2. Выпишем необходимые условия экстремума первого порядка:

$$а) \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_1} = 2\lambda_0 x_1 + \lambda_1 = 0, \quad \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_2} = 2\lambda_1 x_2 + \lambda_1 = 0,$$

$$б) x_1 + x_2 - 2 \leq 0;$$

$$в) \lambda_1 \geq 0 \text{ (для минимума)}, \lambda_1 \leq 0 \text{ (для максимума)};$$

$$г) \lambda_1(x_1 + x_2 - 2) = 0.$$

3. Решим систему для двух случаев.

В первом случае $\lambda_0 = 0$. Тогда из условия «а» следует, что $\lambda_1 = 0$. Это противоречит требованию существования ненулевого вектора $(\lambda_0, \lambda_1)^T$.

Во втором случае $\lambda_0 \neq 0$. Поделим систему на λ_0 и заменим $\frac{\lambda_1}{\lambda_0}$ на λ_1 .

Обобщенная функция Лагранжа при этом заменяется классической

$$а) \frac{\partial L(x, \lambda_1)}{\partial x_1} = 2x_1 + \lambda_1 = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = 2x_2 + \lambda_1 = 0;$$

$$б) x_1 + x_2 - 2 \leq 0;$$

$$в) \lambda_1 \geq 0 \text{ (для минимума)}, \lambda_1 \leq 0 \text{ (для максимума)};$$

$$г) \lambda_1(x_1 + x_2 - 2) = 0.$$

Из условия «г» дополняющей нежесткости следует:

1) $\lambda_1 = 0$ (фактически решается задача поиска безусловного экстремума). Тогда $x_1^* = x_2^* = 0$, $\lambda_1^* = 0$ и условие «б» выполняется. Выполняются необходимые условия и для минимума, и для максимума;

2) $\lambda_1 \neq 0$. Тогда из системы $x_1 + x_2 - 2 = 0$, $2x_1 + \lambda_1 = 0$, $2x_2 + \lambda_1 = 0$ получаем $x_1^* = x_2^* = 1$, $\lambda_1 = -2$. Так как $\lambda_1^* < 0$, то необходимое условие минимума не выполняется (в точке $(1, 1)^T$ нет минимума), но выполняется необходимое условие максимума. Таким образом, имеем две условно стационарные точки.

4. Проверим выполнение достаточных условий экстремума.

В точке $x^* = (0, 0)^T$ ограничение не является активным, так как $g_1(x) = -2 < 0$, поэтому достаточные условия первого порядка не удовлетворяются. Проверим условия второго порядка. Так как $d^2L(x^*, \lambda^*) = 2dx_1^2 + 2dx_2^2 > 0$ при $dx \neq 0$, то в точке $x^* = (0, 0)^T$ регулярный локальный условный минимум, совпадающий в данной задаче с безусловным.

В точке $x^* = (1, 1)^T$ ограничение является активным, но $l = 1 < n = 2$, поэтому достаточное условие первого порядка не выполняется. Проверим достаточное условие второго порядка.

$$d^2L(x^*, \lambda^*) = 2dx_1^2 + 2dx_2^2, dg_1(x) = dx_1 + dx_2 = 0,$$

откуда $dx_1 = -dx_2$.

Следовательно, $d^2L(x^*, \lambda^*) = 4dx_2^2 > 0$ при $dx_2 \neq 0$. Так как в этой точке $\lambda_1^* = -2 < 0$, то достаточное условие максимума не выполняется. Проверим достаточное условие максимума второго порядка. Так как $d^2L(x^*, \lambda^*) = 4dx_2^2 \geq 0$ при любых dx_2 , то необходимое условие максимума не выполняется, поэтому в точке $x^* = (1, 1)^T$ максимума нет.

5. Вычислим значение функции в точке условного минимума $f(x^*) = 0$ (рис. 9.7).

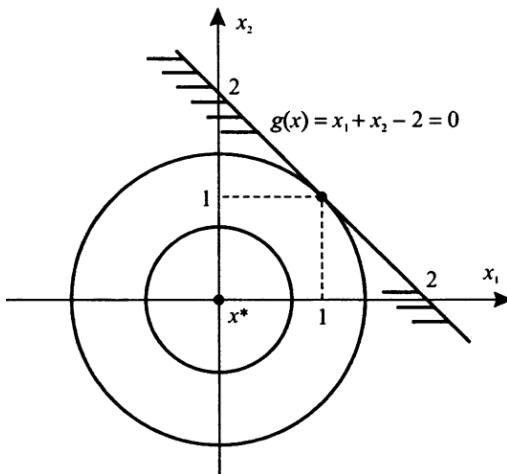


Рис. 9.7. Иллюстрация к решению задачи из примера 8.8

Пример 9.9. Найти условный экстремум в задаче

$$f(x) = x_1 + x_2 \rightarrow \text{extr},$$

$$g(x) = x_1^2 + x_2^2 - 1 \leq 0.$$

1. Составим обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda_1) = \lambda_0(x_1 + x_2) + \lambda_1(x_1^2 + x_2^2 - 1).$$

2. Выпишем необходимые условия экстремума первого порядка:

$$\text{а) } \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_1} = \lambda_0 + 2\lambda_1 x_1 = 0, \quad \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_2} = \lambda_0 + 2\lambda_1 x_2 = 0,$$

$$\text{б) } x_1 + x_2 - 1 \leq 0;$$

$$\text{в) } \lambda_1 \geq 0 \text{ (для минимума), } \lambda_1 \leq 0 \text{ (для максимума);}$$

$$\text{г) } \lambda_1(x_1^2 + x_2^2 - 1) = 0.$$

3. Решим задачу для двух случаев.

В первом случае $\lambda_0 = 0$. Тогда согласно необходимым условиям экстремума первого порядка требуется, чтобы $\lambda_1 \neq 0$. При этом $x_1 = x_2 = 0$ и не удовлетворяется условие «г» дополняющей нежесткости.

Во втором случае $\lambda_0 \neq 0$. Поделим уравнения на λ_0 и заменим $\frac{\lambda_1}{\lambda_0}$ на λ_1 :

$$\text{а) } \frac{\partial L(x, \lambda_1)}{\partial x_1} = 1 + 2\lambda_1 x_1 = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = 1 + 2\lambda_1 x_2 = 0;$$

$$\text{б) } x_1^2 + x_2^2 - 1 \leq 0;$$

$$\text{в) } \lambda_1 \geq 0 \text{ (для минимума), } \lambda_1 \leq 0 \text{ (для максимума);}$$

$$\text{г) } \lambda_1(x_1^2 + x_2^2 - 1) = 0.$$

Из условия «г» дополняющей нежесткости следует два варианта:

1) $\lambda_1 = 0$. Тогда условие «а» не выполняется.

2) $\lambda_1 \neq 0$. Тогда $x_1^2 + x_2^2 - 1 = 0$ и система имеет два решения (рис. 7.8)

точка A : $x_1^\bullet = x_2^\bullet = -\frac{\sqrt{2}}{2}$, $\lambda_1^\bullet = \frac{\sqrt{2}}{2}$ (в точке A может быть минимум);

точка B : $x_1^\bullet = x_2^\bullet = \frac{\sqrt{2}}{2}$, $\lambda_1^\bullet = -\frac{\sqrt{2}}{2}$ (в точке B может быть максимум);

4. Проверим достаточные условия экстремума. В точках A и B ограничения являются активными, но $l = 1 < n = 2$. Поэтому условия первого порядка не выполняются. Проверим условия второго порядка:

$$d^2L(x^\bullet, \lambda_1^\bullet) = 2\lambda_1^\bullet dx_1^2 + 2\lambda_1^\bullet dx_2^2.$$

$$dg_l(x) = 2x_1^2 dx_1 + 2x_2^2 dx_2 = 0.$$

В точках A и B выполняется $dx_1 = -dx_2$. Так как $d^2L(A) = 4\lambda_1^\bullet dx_2^2 = 2\sqrt{2} \cdot dx_2^2 > 0$ при $dx_2 \neq 0$, то в точке A регулярный локальный условный минимум.

Так как $d^2L(B) = 4\lambda_1^\bullet dx_2^2 = -2\sqrt{2} \cdot dx_2^2 < 0$ при $dx_2 \neq 0$, то в точке B регулярный локальный условный максимум. С другой стороны, функции $f(x)$ и $-f(x) = -x_1 - x_2$ — выпуклые и ограничение выпуклое, поэтому в точках A и B достигается глобальный экстремум. Достаточные условия первого и второго порядка проверялись для демонстрации методики.

5. Вычислим значение целевой функции в точках условного экстремума: $f(A) = -\sqrt{2}$, $f(B) = \sqrt{2}$ (рис. 8.8).

Пример 9.10. Найти условный экстремум

$$f(x) = (x_1 - 2)^2 + (x_2 - 3)^2 \rightarrow \text{extr},$$

$$g_l(x) = x_1^2 + x_2^2 - 52 \leq 0.$$

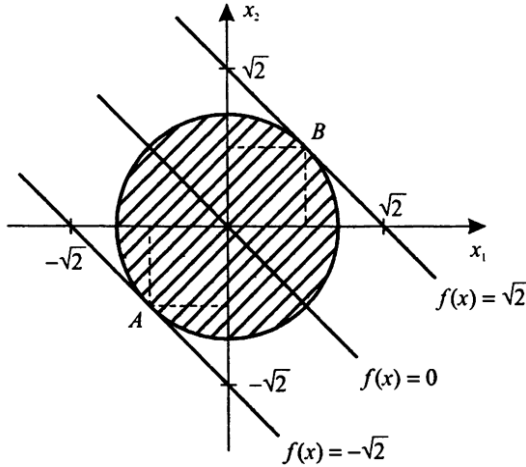


Рис. 9.8. Иллюстрация к решению задачи из примера 9.9

1. Составим обобщенную функцию Лагранжа

$$L(x, \lambda_0, \lambda_1) = \lambda_0((x_1 - 2)^2 + (x_2 - 3)^2) + \lambda_1(x_1^2 + x_2^2 - 52).$$

2. Выпишем необходимые условия экстремума первого порядка:

$$\text{а) } \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_1} = 2\lambda_0(x_1 - 2) + 2\lambda_1 x_1 = 0,$$

$$\frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_2} = 2\lambda_0(x_2 - 3) + 2\lambda_1 x_2 = 0;$$

$$\text{б) } x_1^2 + x_2^2 - 52 \leq 0;$$

$$\text{в) } \lambda_1 \geq 0 \text{ (для минимума), } \lambda_1 \leq 0 \text{ (для максимума);}$$

$$\text{г) } \lambda_1(x_1^2 + x_2^2 - 52) = 0.$$

3. Решим задачу для двух случаев.

В первом случае $\lambda_0 = 0$. Тогда согласно необходимым условиям экстремума первого порядка требуется, чтобы $\lambda_1 \neq 0$. При этом $x_1 = x_2 = 0$ и не выполняется условие «г» дополняющей нежесткости.

Во втором случае $\lambda_0 \neq 0$. Поделим уравнения на λ_0 и заменим $\frac{\lambda_1}{\lambda_0}$ на λ_1

$$\text{а) } \frac{\partial L(x, \lambda_1)}{\partial x_1} = 2(x_1 - 2) + 2\lambda_1 x_1 = 0,$$

$$\frac{\partial L(x, \lambda_1)}{\partial x_2} = 2(x_2 - 3) + 2\lambda_1 x_2 = 0;$$

$$\text{б) } x_1^2 + x_2^2 - 52 \leq 0;$$

$$\text{в) } \lambda_1 \geq 0 \text{ (для минимума), } \lambda_1 \leq 0 \text{ (для максимума);}$$

$$\text{г) } \lambda_1(x_1^2 + x_2^2 - 52) = 0.$$

Из условия «г» следуют два варианта:

1) $\lambda_1 = 0$. Тогда $x_1 = 2$, $x_2 = 3$ и выполняются необходимые условия и для минимума, и для максимума. Имеем условно стационарную точку $A: x_1^\bullet = 2, x_2^\bullet = 3, \lambda_1^\bullet = 0$;

2) $\lambda_1 \neq 0$. Тогда $x_1^2 + x_2^2 - 52 = 0$ и система имеет решение:

$$\text{точка } B: x_1^\bullet = 4, x_2^\bullet = 6, \lambda_1^\bullet = -\frac{1}{2}; \text{ точка } C: x_1^\bullet = -4, x_2^\bullet = -6, \lambda_1^\bullet = -\frac{3}{2}.$$

Так как $\lambda_1^\bullet < 0$ в обеих точках, то в них минимума нет, но может быть максимум.

4. Проверим достаточные условия экстремума. В обеих условно стационарных точках ограничение превращается в равенство, т.е. активно. Так как число активных ограничений $l = 1 < 2 = n$, то условия первого порядка не выполняются. Так как функция $-f(x) = -(x_1 - 2)^2 - (x_2 - 3)^2$ не является выпуклой, то необходимые условия не являются достаточными. Поэтому проверим условия второго порядка

$$d^2 L(x^\bullet, \lambda_1^\bullet) = (2 + 2\lambda_1^\bullet)dx_1^2 + (2 + 2\lambda_1^\bullet)dx_2^2.$$

В точке A ограничение не является активным. Так как $\lambda_1^\bullet = 0$, то $d^2 L(x^\bullet, \lambda_1^\bullet) = 2dx_1^2 + 2dx_2^2 > 0$ при $dx \neq 0$. Поэтому в точке A условный локальный минимум. Так как целевая функция

выпуклая и множество допустимых значений выпукло, то можно заключить, что в данном случае необходимое условие минимума является достаточным. В точке A глобальный минимум.

В точках B и C ограничение активно. Поэтому $dg_1(x) = 2x_1^2 dx_1 + 2x_2^2 dx_2 = 0$. В точках B и C выполняется $dx_1 = -\frac{3}{2} dx_2$.

Так как $d^2L(B) = [(-\frac{3}{2})dx_2]^2 + dx_2^2 > 0$ при $dx_2 \neq 0$, а $\lambda_1^\bullet = -\frac{1}{2} < 0$,

то достаточные условия максимума не выполняются. Так как $d^2L(B) \notin 0$ при всех dx_2 , то и необходимое условие максимума второго порядка в точке B не выполняется. Поэтому в ней нет экстремума. Так как $d^2L(C) = -[(-\frac{3}{2})dx_2]^2 - dx_2^2 < 0$ при $dx \neq 0$, то достаточные условия максимума выполняются. В точке C условный локальный максимум.

5. Вычислим значения функций в точках экстремума $f(A) = 0$, $f(C) = 117$.

Графическое решение задачи приведено на рис. 9.9.

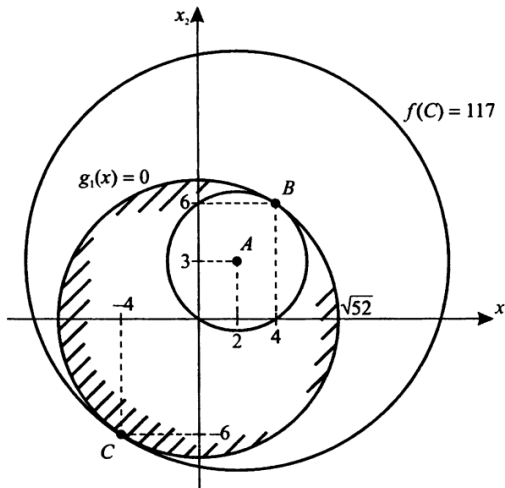


Рис. 9.9. Графическое решение задачи из примера 9.10

Основная литература: 7 [101–123].

Контрольные вопросы:

1. Условный экстремум при ограничениях типа равенств.
2. Условный экстремум при ограничениях типа неравенств.

10. ЗАДАЧА МИНИМИЗАЦИИ ФУНКЦИИ МНОГИХ ПЕРЕМЕННЫХ. НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ БЕЗУСЛОВНОГО ЭКСТРЕМУМА

10.1. Постановка задачи и определения

Постановка задачи поиска минимума функций.

Найти такой вектор $x^* = (x_1, \dots, x_n)^T$, принадлежащий области допустимых значений $U \subset E_n$ (E_n – евклидово пространство), целевая функция от которого

$$f(x^*) = \min_{x \in U} f(x). \quad (10.1)$$

Необходимо обратить внимание на следующее.

1. Поиск максимума функции $f(x)$ сводится к задаче определения минимума путем замены знака перед функцией на противоположный

$$f(x^*) = \max_{x \in U} f(x) = \min_{x \in U} (-f(x)).$$

2. Задача определения минимума и максимума целевой функции $f(x)$ называется задачей поиска *экстремума* $f(x^*) = \text{extr} f(x)$.

3. Если множество допустимых решений U задается ограничениями (условиями), накладываемыми на вектор x , то решается задача поиска *условного экстремума*. Если $U = E_n$, то ограничения (условия) на вектор x отсутствуют и решается задача поиска *безусловного* экстремума.

4. Решением задачи поиска условного экстремума является пара $(x^*, f(x^*))$ которая включает точку x^* и значение целевой функции в ней.

Напомним, что $x^* \in U$ называется точкой глобального минимума, или просто минимума, функции $f(x)$ на множестве U , если функция достигает в этой точке своего наименьшего значения, т.е. $f(x^*) \leq f(x)$ для всех $x \in U$;

точка $x^* \in U$ называется точкой *локального минимума* функции $f(x)$ на множестве U , если существует такое $\varepsilon > 0$, что если $x^* \in$

U и $\|x - x^*\| < \varepsilon$, то $f(x^*) \leq f(x)$. Здесь $\|x\| = \sqrt{x_1^2 + \dots + x_n^2}$ – евклидова пространства норма векторов.

Определение. Поверхностью уровня функции $f(x)$ называется множество точек, в которых функция принимает постоянное значение, т. е. $f(x) = \text{const}$. Если $n = 2$, то поверхность уровня изображается линией уровня на плоскости E_2 .

Пример 10.1. Построить линии уровня функций:

$$\text{а) } f(x) = x_1^2 + x_2^2;$$

$$\text{б) } f(x) = \frac{x_1^2}{4} + x_2^2; \text{ в) } f(x) = x_2^2 - x_1^2; \text{ г) } f(x) = x_2^2.$$

* а) $f(x) = x_1^2 + x_2^2 = \text{const} = r^2$ – уравнение окружности с центром в точке $(0, 0)^T$ и радиусом, равным r ;

б) $\frac{x_1^2}{4} + x_2^2 = \text{const}$ – уравнение эллипса. Если $\text{const} = 1$, то $a = 2, b = 1$ – большая и малая полуоси эллипса;

в) $f(x) = x_2^2 - x_1^2 = \text{const}$ – уравнение гиперболы;

г) $f(x) = x_2^2 = \text{const}$ – уравнение двух параллельных оси Ox_1 прямых.

Определение. Градиентом $\nabla f(x)$ непрерывно дифференцируемой функции $f(x)$ в точке x называется вектор–столбец, элементами которого являются частные производные первого порядка, вычисленные в данной точке

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \dots \\ \frac{\partial f(x)}{\partial x_n} \end{pmatrix}.$$

Градиент функции направлен по нормали к поверхности уровня, т. е. перпендикулярно к касательной плоскости, проведенной в точке x , в сторону наибольшего возрастания функции в данной точке.

Определение. Матрицей Гессе $H(x)$ дважды непрерывно дифференцируемой в точке x функции $f(x)$ называется матрица частных производных второго порядка, вычисленных в данной точке

$$H(x) = \begin{pmatrix} \frac{\partial^2 f(x)}{\partial x^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x^2} & \dots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \dots & \dots & \dots & \dots \\ h_{1n} & h_{n2} & \dots & h_{nn} \end{pmatrix},$$

где $h_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$, $i, j = 1, \dots, n$.

Необходимо обратить внимание на следующее.

1. Матрица Гессе является симметрической размера $n \times n$.

Вместе с градиентом можно определить вектор антиградиента, равный по модулю вектору градиента, но противоположный по направлению. Он указывает в сторону наибольшего убывания функции в данной точке.

2. С помощью градиента и матрицы Гессе, используя разложение в ряд Тейлора, приращение функции может быть записано в форме

$$\Delta f(x) = f(x + \Delta x) - f(x) = \nabla f(x)^T \Delta x + \frac{1}{2} \Delta x^T H(x) \Delta x + o(\|\Delta x\|^2), \quad (10.2)$$

где $o(\|\Delta x\|^2)$ – сумма всех членов разложения, имеющих порядок выше второго, $\Delta x^T H(x) \Delta x$ – квадратичная форма.

Пример 10.2. Для функции $f(x) = x_1^2 + x_2^2$ вычислить градиент в точках $x^{(0)} = (0, 1)^T$, $x^{(1)} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$, $x^{(2)} = (1, 0)^T$, $x^{(3)} = (0, -1)^T$ и найти матрицу Гессе.

По определению градиента и матрицы Гессе имеем

$$\nabla f(x) = (2x_1, 2x_2)^T, \nabla f(x^{(0)}) = (0, 2)^T, \nabla f(x^{(1)}) = (\sqrt{2}, \sqrt{2})^T, \nabla f(x^{(2)}) = (2, 0)^T, \\ \nabla f(x^{(3)}) = (0, -2)^T, H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

Заметим, что матрица Гессе квадратичной функции не зависит от x .

Пример 10.3. Для функции $f(x) = x_1^2 + x_2^4$ вычислить градиент и найти матрицу Гессе в точках $x^{(0)} = (0, 0)^T$, $x^{(1)} = (1, 1)^T$.

Согласно определениям имеем

$$\nabla f(x) = (2x_1, 4x_2^3)^T, H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix};$$

$$\nabla f(x^{(0)}) = (0, 0)^T, H(x^{(0)}) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix};$$

$$\nabla f(x^{(1)}) = (2, 4)^T, H(x^{(1)}) = \begin{pmatrix} 2 & 0 \\ 0 & 12 \end{pmatrix}.$$

Определение. Квадратичная форма (и соответствующая ей матрица Гессе $H(x)$) называется:

– *положительно определенной* ($H(x) > 0$), если для любого ненулевого Δx выполняется неравенство $\Delta x^T H(x) \Delta x > 0$;

– *отрицательно определенной* ($H(x) < 0$), если для любого ненулевого Δx выполняется неравенство $\Delta x^T H(x) \Delta x < 0$;

– *положительно полуопределенной* ($H(x) \geq 0$), если для любого ненулевого Δx выполняется неравенство $\Delta x^T H(x) \Delta x \geq 0$ и имеется отличный от нуля вектор Δx , для которого $\Delta x^T H(x) \Delta x = 0$;

– отрицательно полуопределенной ($H(x) \leq 0$), если для любого Δx выполняется неравенство $\Delta x^T H(x) \Delta x \leq 0$ и имеется отличный от нуля вектор Δx , для которого $\Delta x^T H(x) \Delta x = 0$;

– неопределенной, если существуют такие векторы Δx , $\Delta \tilde{x}$, что выполняется неравенство $\Delta x^T H(x) \Delta x > 0$, $\Delta \tilde{x}^T H(x) \Delta \tilde{x} < 0$;

– тождественно равной нулю ($H(x) \equiv 0$), если для любого Δx выполняется $\Delta x^T H(x) \Delta x = 0$.

Пример 10.4. Классифицировать квадратичную форму и матрицу Гессе $H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

Выпишем квадратичную форму

$$\Delta x^T H(x) \Delta x = (\Delta x_1, \Delta x_2) \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = 2\Delta x_1^2 + 2\Delta x_2^2.$$

Очевидно, $\Delta x^T H(x) \Delta x > 0$ для любого ненулевого Δx . Согласно определению, квадратичная форма (матрица Гессе) положительно определена.

Пример 10.5. Классифицировать квадратичную форму и матрицу Гессе $H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$.

Выпишем квадратичную форму

$$\Delta x^T H(x) \Delta x = (\Delta x_1, \Delta x_2) \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = 2\Delta x_1^2.$$

Очевидно, $\Delta x^T H(x) \Delta x \geq 0$ для любого вектора Δx и $\Delta x^T H(x) \Delta x = 0$ для $\Delta x_1 = 0$ и $\Delta x_2 \neq 0$. Согласно определению, квадратичная форма (матрица Гессе) положительно полуопределена.

Пример 10.6. Найти матрицу Гессе функции $f(x) = x_1^2 - x_2^2$ и классифицировать ее.

Следуя определению матрицы Гессе, получаем $H(x) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$.

Выпишем соответствующую ей квадратичную форму

$$\Delta x^T H(x) \Delta x = (\Delta x_1, \Delta x_2) \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \Delta x_1 \\ \Delta x_2 \end{pmatrix} = 2\Delta x_1^2 - 2\Delta x_2^2.$$

При $\Delta x_1 = 0$ и любых $\Delta x_2 \neq 0$ квадратичная форма отрицательна, а при $\Delta x_1 \neq 0$ и $\Delta x_2 = 0$ положительна. Квадратичная форма (матрица Гессе) неопределенная.

Решение задач минимизации в E_n , как правило, сопряжено со значительными трудностями, особенно для многоэкстремальных функций. Неотерые из этих трудностей устраняются, если ограничиться рассмотрением только *выпуклых* целевых функций.

10.2. Свойства выпуклых множеств и выпуклых функций

Определение. Пусть $x, y \in E_n$. Множество точек $\{z\} \subset E_n$, $z = \alpha x + (1 - \alpha)y$, $\alpha \in [0, 1]$ называется *отрезком*, соединяющим точки x и y .

В пространстве E_n , $n \leq 3$ это соотношение определяет обычный отрезок, соединяющий точки x и y .

Определение. Множество $U \subset E_n$ называется *выпуклым*, если вместе с любыми точками x и $y \in U$ оно содержит и весь отрезок $z = \alpha x + (1 - \alpha)y$, $\alpha \in [0, 1]$.

Необходимо отметить, что проверка условия выпуклости в большинстве случаев требует громоздких выкладок, поэтому на практике при исследовании выпуклости множества в пространствах E_2 и E_3 часто прибегают к решениям в графическом виде.

Образно говоря, выпуклыми являются множества, которые не содержат «вмятин», «дырок» и состоят из одного «куска».

Примерами выпуклых множеств являются само пространство E_n , отрезок, прямая, шар. Множество $x^2 + y^2 = 1$ в E_2 не является выпуклым.

Определение. Функция $f(x)$, заданная на выпуклом множестве $U \subset E_n$, называется:

– *выпуклой*, если для любых точек $x, y \in U$ и любого $\alpha \in [0, 1]$ выполняется неравенство

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y). \quad (10.3)$$

– *строго выпуклой*, если для всех $\alpha \in (0, 1)$ неравенство (10.3) выполняется как строгое

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y). \quad (10.4)$$

– *сильно выпуклой* с константой $l > 0$, если для любых точек $x, y \in U$ и любого $\alpha \in [0, 1]$ выполняется неравенство

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{l}{2} \alpha(1 - \alpha) \|x - y\|^2. \quad (10.5)$$

Необходимо обратить внимание на следующее.

1. Функцию $f(x)$ называют выпуклой, если ее график целиком лежит не выше отрезка, соединяющего две ее произвольные точки. Функцию называют строго выпуклой, если ее график целиком лежит ниже отрезка, соединяющего две ее произвольные, но не совпадающие точки.

2. Если функция сильно выпуклая, то она одновременно строго выпуклая и выпуклая. Если функция строго выпуклая, то она одновременно выпуклая.

3. Выпуклость функции можно определить по матрице Гессе:

- если $H(x) \geq 0$ для любого $x \in E_n$, то функция выпуклая;
- если $H(x) > 0$ для любого $x \in E_n$, то функция строго выпуклая;
- если $H(x) \geq lE$ для любого $x \in E_n$, где E – единичная матрица, то функция сильно выпуклая.

Пример 10.7. Исследовать на выпуклость функцию $f(x) = x^2$, определенную на множестве $U = \{x \in [1, -1]\}$.

Функция является строго выпуклой, так как ее график целиком лежит ниже отрезка, соединяющего две ее произвольные, но не совпадающие точки. Более того, функция является сильно выпуклой, так как выписывается условие $H(x) = f''(x) = 2 \geq l$ при $0 \leq l \leq 2$.

Пример 10.8. Исследовать на выпуклость функцию $f(x) = x$, определенную на множестве $U = \{x \in [0, 1]\}$.

Функция является выпуклой, так как ее график целиком лежит не выше отрезка, соединяющего две ее произвольные точки, но не является строго выпуклой и тем более сильно выпуклой.

Пример 10.9. Исследовать на выпуклость функцию $f(x) = x_1^2 + x_2^2$, определенную на множестве E_2 .

Матрица Гессе функция удовлетворяет условию $H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \geq l \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ при $0 \leq l \leq 2$.

Следовательно, эта функция сильно выпукла. Одновременно она является строго выпуклой и выпуклой.

В дальнейшем часто используются следующие свойства выпуклых функций.

1. Если $f(x)$ выпуклая функция на выпуклом множестве U , то всякая точка локального минимума является точкой ее глобального минимума на U .

2. Если выпуклая функция достигает своего минимума в двух различных точках, то она достигает минимума во всех точках отрезка, соединяющего эти точки.

3. Если $f(x)$ выпуклая функция на выпуклом множестве U , то она может достигать своего глобального минимума на U не более чем в одной точке.

10.3. Необходимые и достаточные условия безусловного экстремума

Необходимые условия экстремума первого порядка. Пусть $x^* \in E_n$ есть точка локального минимума (максимума) функции $f(x)$ на множестве E_n и $f(x)$ дифференцируема в точке x^* . Тогда градиент функции $f(x)$ в точке x^* равен нулю

$$\nabla f(x) = 0, \quad (10.6)$$

и

$$\frac{\partial f(x^*)}{\partial x_i} = 0, \quad i = 1, \dots, n. \quad (10.7)$$

Точки x^* , удовлетворяющие условию (10.6) и (10.7), называются *стационарными*.

Необходимые условия экстремума второго порядка. Пусть $x^* \in E_n$ есть точка локального минимума (максимума) функции $f(x)$ на множестве E_n , и функция $f(x)$ дважды дифференцируема в этой точке. Тогда матрица Гессе $H(x)$ функции $f(x)$, вычисленная в точке x^* , является положительно определенной (отрицательно полуопределенной), т. е. $H(x) \geq 0$, $H(x) \leq 0$.

Достаточные условия экстремума. Пусть функция $f(x)$ в точке $x^* \in E_n$ дважды дифференцируема, ее градиент равен нулю, а матрица Гессе является положительно полуопределенной (отрицательно полуопределенной), т. е.

$$\nabla f(x) = 0 \text{ и } H(x) > 0, H(x) < 0. \quad (10.8)$$

Точки x^* — точка локального минимума (максимума) функции $f(x)$ на E_n .

Проверка выполнения условий экстремума. Рассмотрим матрица Гессе, составленную для стационарной точки x^*

$$H(x) = \begin{pmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \dots & \dots & \dots & \dots \\ h_{n1} & h_{n2} & \dots & h_{nn} \end{pmatrix}.$$

Угловыми минорами k -го порядка матрицы размера $n \times n$, где $k \leq n$, называются определители, составленные из элементов исходной матрицы, стоящих в ней на пересечении k верхних строк и k левых столбцов.

Главными минорами m -го порядка матрицы размера $n \times n$, где $m \leq n$, называются определители, составленные из элементов исходной матрицы, оставшихся после вычеркивания в ней любых $(n - m)$ строк и $(n - m)$ столбцов с одинаковыми номерами.

Существуют два основных способа проверки выполнения условий экстремума. *Первый способ* связан с исследованием положительной или отрицательной определенности угловых и главных миноров матрицы Гессе. *Второй способ* основан на анализе собственных значений этой матрицы.

В первом случае при анализе знаков угловых и главных миноров матрицы Гессе доказаны и обычно используются на практике два критерия – проверки достаточных и необходимых условий второго порядка. Приведем их без доказательства.

Критерий Сильвестра проверки достаточных условий экстремума

1. Для того чтобы матрица Гессе $H(x^*)$ была положительно определенной ($H(x) \geq 0$) и стационарная точка x^* являлась точкой локального минимума, необходимо и достаточно, чтобы знаки угловых миноров были строго положительны

$$(\Delta_1 > 0, \Delta_2 > 0, \dots, \Delta_n > 0).$$

2. Для того чтобы матрица Гессе $H(x^*)$ была положительно полуопределенной ($H(x) \geq 0$) и стационарная точка x^* являлась точкой локального максимума, необходимо и достаточно, чтобы все главные миноры определителя матрицы Гессе были знаки угловых миноров чередовались, начиная с отрицательны

$$(\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \dots, (-1)^n \Delta_n > 0).$$

Критерий проверки необходимых условий экстремума второго порядка

1. Для того чтобы матрица Гессе $H(x^*)$ была положительно полуопределенной ($H(x) \geq 0$) и стационарная точка x^* являлась бы точкой локального минимума, необходимо и достаточно, чтобы все главные миноры определителя были отрицательны

$$(\Delta_1 \geq 0, \Delta_2 \geq 0, \dots, \Delta_n \geq 0).$$

2. Для того чтобы матрица Гессе $H(x^*)$ была положительно полуопределенной ($H(x) \geq 0$) и стационарная точка x^* являлась бы точкой локального максимума, необходимо и достаточно, чтобы все главные миноры четного порядка были неотрицательны, а все главные миноры нечетного порядка – неположительны

$$(\Delta_1 \leq 0, \Delta_2 \geq 0, \Delta_3 \leq 0, \dots, (-1)^n \Delta_n \geq 0).$$

Второй способ проверки достаточных условий экстремума связан с анализом собственных значений матрицы Гессе и применим только в случае, если эти значения удастся вычислить.

Определение. Собственные значения $\lambda_i, i = 1, \dots, n$ матрицы $H(x^*)$ размера $n \times n$ находятся как корни характеристического уравнения (алгебраического уравнения n -й степени) $|H(x^*) - \lambda E| = 0$.

Необходимо отметить, что собственные значения вещественной симметрической матрицы $H(x^*)$ вещественны.

Если собственные значения матрицы Гессе вычислены, то дальнейшая проверка достаточных условий экстремума не составляет затруднений. Приведем в виде таблицы порядок применения двух способов, позволяющих установить достаточные условия экстремума и необходимые условия второго порядка при соблюдении необходимых условий первого порядка $\nabla f(x^*) = 0$ (табл. 10.1).

Алгоритм решения задачи следующий.

Таблица 10.1

Критерии проверки достаточных и необходимых условий
второго порядка при поиске безусловного экстремума

№	$H(x^*)$	Условия	Первый способ	Второй способ	Тип стационарной точки x^*
1	> 0	Достаточные условия экстремума	$\Delta_1 > 0, \Delta_2 > 0, \dots \Delta_n > 0$	$\lambda_1 > 0, \dots \lambda_n > 0$	Локальный минимум
2	< 0	Достаточные условия экстремума	$\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \dots (-1)^n \Delta_n > 0$	$\lambda_1 < 0, \dots \lambda_n < 0$	Локальный минимум
3	≥ 0	Необходимые условия второго порядка	Все главные миноры определителя матрицы $H(x^*)$ неотрицательны	$\lambda_1 \geq 0, \dots \lambda_n \geq 0$	Может быть локальный минимум, требуется дополнительное исследование
4	≤ 0	Необходимые условия второго порядка	Все главные миноры четного порядка ≥ 0 , нечетного порядка ≤ 0	$\lambda_1 \leq 0, \dots \lambda_n \leq 0$	Может быть локальный максимум, требуется дополнительное исследование
5	$= 0$	Необходимые условия второго порядка	Матрица Гессе состоит из нулевых элементов	$\lambda_1 = 0, \dots \lambda_n = 0$	Требуется дополнительное исследование
6	$> 0, < 0$	Необходимые условия второго порядка	Не выполняются условия 1–5	λ_1 имеют разные знаки	Нет экстремума

Шаг 1. Записать необходимые условия экстремума первого порядка и найти стационарные точки x^* в результате решения системы n в общем случае нелинейных алгебраических

уравнений с n неизвестными. Для численного решения системы могут использоваться методы простой итерации, Зейделя, Ньютона.

Шаг 2. В найденных стационарных точках x^* проверить выполнение достаточных условий, а если они не выполняются, то необходимых условий второго порядка с помощью одного из двух способов (табл. 10.1).

Шаг 3. Вычислить значения $f(x^*)$ в точках экстремума.

Необходимо обратить внимание на следующее.

1. Если требуется определить глобальные экстремумы, то они находятся в результате сравнения значений функции в точках локальных экстремумов с учетом ограниченности функции на E_n .

2. Для функции одной переменной $f(x)$ можно пользоваться следующим правилом, заменяющим операции на шаге 2.

Если $f(x)$ и ее производные непрерывны, то точка x^* является точкой экстремума тогда и только тогда, когда m – четное, где m -порядок первой не обращающейся в нуль в точке x^* производной. Если $f^{(m)}(x^*) > 0$, то в точке x^* – локальный минимум, а если $f^{(m)}(x^*) < 0$, то в точке x^* – локальный максимум. Если число m нечетное, то в точке x^* нет экстремума.

Часто на практике при применении численных методов поиска экстремума требуется проверить, выполняются ли необходимые и достаточные условия экстремума в некоторой точке. Такая проверка необходима, т. к. многие численные методы позволяют найти лишь стационарную точку, тип которой требует уточнения.

Пример 10.10. Найти экстремум функции $f(x) = x_1^2 + x_2^2$, определенной на множестве E_2 .

1. Запишем необходимые условия экстремума первого порядка

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 = 0; \quad \frac{\partial f(x)}{\partial x_2} = 2x_2 = 0.$$

В результате решения системы получаем стационарную точку $x^* = (0, 0)^T$.

2. Проверим выполнение достаточных условий экстремума.

Первый способ. Матрица Гессе имеет вид $H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. Так

как $\Delta_1 = h_{11} = 2 > 0$, $\Delta_2 = \begin{vmatrix} 2 & 0 \\ 0 & 2 \end{vmatrix} = 4 > 0$, то в точке x^* локальный минимум (строка 1 табл. 9.1).

Второй способ. Найдем собственные значения матрицы Гессе

$$\begin{vmatrix} 2 - \lambda & 0 \\ 0 & 2 - \lambda \end{vmatrix} = 0. \text{ Отсюда } (2 - \lambda)^2 = 0, \lambda_1 = \lambda_2 = 2 > 0.$$

Так как все собственные значения положительны, то в точке x^* локальный минимум (строка 1 табл. 10.1). Так как $H(x) > 0$, то функция строго выпукла на множестве E_2 . Поэтому точка локального минимума является и точкой глобального минимума.

3. Вычислим значение функции в точке глобального минимума $f(x^*) = 0$.

Пример 9.11. Найти экстремум функции $f(x) = x_1^2 - x_2^2$ на множестве E_2 .

1. Запишем необходимые условия экстремума первого порядка

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 = 0; \quad \frac{\partial f(x)}{\partial x_2} = -2x_2 = 0.$$

В результате решения системы получаем стационарную точку $x^* = (0, 0)^T$.

2. Проверим выполнение достаточных условий экстремума и необходимых условий второго порядка.

Первый способ. Матрица Гессе имеет вид $H(x) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$.

Так как $\Delta_1 = h_{11} = 2 > 0$, $\Delta_2 = \begin{vmatrix} 2 & 0 \\ 0 & -2 \end{vmatrix} = -4 < 0$, то достаточные условия экстремума не выполняются (строки 1 и 2 табл. 10.1).

Проверим выполнение необходимых условий второго порядка. Главные миноры первого порядка ($m = 1$) получаются из Δ_2 в результате вычеркивания $n - m = 2 - 1 = 1$ строк и столбцов с одинаковыми номерами, а их значения равны соответственно -2 и 2 . Главный минор второго порядка получается из Δ_2 в результате вычеркивания $n - m = 0$ строк и столбцов и совпадает с $\Delta_2 = -4$. Отсюда следует, что необходимые условия экстремума второго порядка не выполняются (строки 3 и 4 в табл. 10.1). Так как матрица Гессе не является нулевой, то делаем вывод, что в точке x^* экстремума нет (строка 6 в табл. 10.1).

Второй способ. Найдем собственные значения матрицы Гессе:

$$\begin{vmatrix} 2-\lambda & 0 \\ 0 & -2-\lambda \end{vmatrix} = 0. \text{ Отсюда } (2-\lambda)(-2-\lambda) = 0, \lambda_1 = -2 < 0, \lambda_2$$

$= 2 > 0$, т. е. собственные значения имеют разные знаки. Поэтому точка x^* не является точкой минимума и максимума (строка 6 таблица 10.1).

3. Так как экстремума нет ни в одной точке, значение $f(x^*)$ не вычисляется.

Пример 10.12. Найти экстремум функции $f(x) = x_1^2 + x_2^4$, определенной на множестве E_2 .

1. Запишем необходимые условия экстремума первого порядка

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 = 0; \quad \frac{\partial f(x)}{\partial x_2} = 4x_2^3 = 0.$$

В результате решения системы получаем стационарную точку $x^* = (0, 0)^T$.

2. Проверим выполнение достаточных условий второго порядка. Матрица Гессе имеет вид $H(x) = \begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$.

Так как $\Delta_1 = h_{11} = 2 > 0$, $\Delta_2 = \begin{vmatrix} 2 & 0 \\ 0 & 0 \end{vmatrix} = 0$, то достаточные условия

экстремума не выполняются. Проверим выполнение необходимых условий экстремума второго порядка. Аналогично решению предыдущего примера получаем значения главных миноров первого порядка соответственно 2 и 0 и главный минор второго порядка 0. Так как все главные миноры неположительные, то в точке x^* может быть минимум и требуется дополнительное исследование.

3. Вычислим значение целевой функции в стационарной точке $f(x^*) = 0$ и рассмотрим ε -окрестность точки x^* , а также поведение функции $f(x)$ на множестве E_2 . При любых $x \in E_2$ имеем $f(x) \geq f(x^*) = 0$. Поэтому точка x^* является точкой глобального минимума.

Пример 10.13. Найти экстремум функции

$$f(x) = -x_1^2 - x_2^3 - x_3^2 - x_1 + x_1x_2 + 2x_3 \text{ на множестве } E_3.$$

1. Запишем необходимые условия экстремума первого порядка

$$\frac{\partial f(x)}{\partial x_1} = -2x_1 - 1 + x_2 = 0; \quad \frac{\partial f(x)}{\partial x_2} = -2x_2 + x_1 = 0, \quad \frac{\partial f(x)}{\partial x_3} = -2x_3 + 2 = 0.$$

В результате решения системы получаем стационарную точку

$$x^* = \left(-\frac{2}{3}, -\frac{1}{3}, 1\right)^T.$$

2. Проверим выполнение достаточных условий экстремума и необходимых условий второго порядка.

Первый способ. Матрица Гессе имеет вид

$$H(x^*) = \begin{pmatrix} 2 & 1 & 0 \\ 1 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \text{ Так как } \Delta_1 = -2 < 0, \Delta_2 = \begin{vmatrix} -2 & 1 \\ 1 & -2 \end{vmatrix} = 4 - 1 > 0,$$

$\Delta_3 = (-2) \cdot 3 = -6 < 0$, т. е. знаки угловых миноров чередуются, начиная с отрицательного, то x^* – точка локального максимума.

Второй способ. Найдем собственные значения матрицы Гессе

$$\det(H - \lambda E) = \begin{vmatrix} -2 - \lambda & 1 & 0 \\ 1 & -2 - \lambda & 0 \\ 0 & 0 & -2 - \lambda \end{vmatrix} = 0.$$

Отсюда $(-2 - \lambda)[(-2 - \lambda)^2 - 1] = 0$ и $\lambda_1 = -2 < 0$, $\lambda_2 = -1 < 0$, $\lambda_3 = -3 < 0$. Так как все собственные значения матрицы Гессе отрицательны, то в точке x^* локальный максимум.

3. Вычислим значение функции в точке локального максимума $f(x^*) = \frac{4}{3}$.

Пример 10.14. Найти экстремум функции

$$f(x) = x_1^3 + x_2^3 - 3x_1x_2 + 2x_3 \text{ на множестве } E_2.$$

1. Запишем необходимые условия экстремума первого порядка

$$\frac{\partial f(x)}{\partial x_1} = 3x_1^2 - 3x_2 = 0; \quad \frac{\partial f(x)}{\partial x_2} = 3x_2^2 - 3x_1 = 0.$$

Решая систему, найдем стационарные точки

$$x_1^* = (0, 0)^T; \quad x_2^* = (1, 1)^T.$$

2. Проверим выполнение достаточных условий второго порядка.

Первый способ. Матрица Гессе имеет вид

$$H(x_1^*) = \begin{pmatrix} 6x_1 & -3 \\ -3 & 6x_2 \end{pmatrix} = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix} \text{ при } x = x_1^*. \text{ Так как } \Delta_1 = h_{11} = 0,$$

$$\Delta_2 = \begin{vmatrix} 0 & -3 \\ -3 & 0 \end{vmatrix} = -9, \text{ то достаточные условия экстремума не}$$

выполняются. Проверим выполнение необходимых условий экстремума второго порядка. Главные миноры первого порядка равны 0, главный минор второго равен -9 . Необходимое условие второго порядка не выполняется, следовательно, $x_1^* = (0, 0)^T$

экстремума быть не может. В точке $x_2^\bullet = (1, 1)^T$ матрица Гессе имеет вид

$$H(x_2^\bullet) = \begin{pmatrix} 6x_1 & -3 \\ -3 & 6x_2 \end{pmatrix} = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

Так как $\Delta_1 = h_{11} = 6 > 0$, $\Delta_2 = \begin{vmatrix} 6 & -3 \\ -3 & 6 \end{vmatrix} = 27 > 0$, то

выполняются достаточные условия, следовательно, $x_2^\bullet = (1, 1)^T$ – точка локального минимума.

Второй способ. Найдем собственные значения матрицы Гессе в точке $x_1^\bullet = (0, 0)^T$:

$$\det(H - \lambda E) = \begin{vmatrix} -\lambda & -3 \\ -3 & -\lambda \end{vmatrix} = 0, \text{ откуда } \lambda_1 = -3, \lambda_2 = 3.$$

Так как все собственные значения положительны, следовательно, $x_2^\bullet = (1, 1)^T$ – точка локального минимума.

3. Вычислим значение функции в точке локального максимума $f(x_2^\bullet) = -1$.

Пример 10.15. Найти экстремум функции

$$f(x) = 5x^6 - 36x^5 + \frac{165}{2}x^4 - 60x^3 + 36 \text{ на } E_1.$$

1. Выпишем необходимое условие экстремума первого порядка

$$\frac{df(x)}{dx} = 30x^5 - 180x^4 + 330x^3 - 180x^2 = 30x^2(x-1)(x-2)(x-3) = 0.$$

Отсюда получаем стационарные точки $x_1^\bullet = 0$, $x_2^\bullet = 1$, $x_3^\bullet = 2$, $x_4^\bullet = 3$.

2. Проверим выполнение достаточных условий экстремума

$$\frac{d^2 f(x)}{dx^2} = 150x^4 - 720x^3 + 990x^2 - 360x,$$

$$(x_1^{\bullet}) = 0, f''(x_2^{\bullet}) = 60 > 0, f''(x_3^{\bullet}) = -120 < 0, f''(x_4^{\bullet}) = 540 > 0,$$

поэтому в точках $x_2^{\bullet}$, $x_4^{\bullet}$ – локальный минимум, а в точке $x_3^{\bullet}$ локальный максимум. В точке $x_1^{\bullet}$ достаточные условия не выполняются, поэтому вычислим третью производную

$$f'''(x) = 600x^3 - 2160x^2 + 1980x - 360, \quad f'''(x_1^{\bullet}) = -360.$$

Так как эта производная отлична от нуля и имеет нечетный порядок, то в точке $x_1^{\bullet}$ экстремума нет.

2. Выпишем значения экстремумов функции

$$f(1) = \frac{55}{2}; f(2) = -108; f(3) = \frac{551}{2}.$$

Основная литература: 7 [57–73].

Контрольные вопросы:

1. Постановка задачи и определения.
2. Свойства выпуклых множеств и выпуклых функций.
3. Необходимые и достаточные условия безусловного экстремума.
4. Критерий Сильвестра проверки достаточных условий экстремума.
5. Критерий проверки необходимых условий экстремума второго порядка.

11. ОПТИМИЗАЦИЯ КОМПЛЕКСНОГО ПРОИЗВОДСТВЕННОГО ПРОЦЕССА ИЗГОТОВЛЕНИЯ ПЕЧАТНОЙ ПРОДУКЦИИ

Здесь рассматриваются математические методы, используя которые при проектировании производственных процессов можно получить наилучшие результаты с точки зрения какого-либо технико-экономического показателя. Эти методы можно применять при условии, что результаты процесса линейно зависят от исходных величин и его можно описать матричным выражением $Y = Ax$, причем элементы матричного уравнения линейны.

11.1. Техничко-экономические модели

В технико-экономических моделях в математической форме представлены количественные отношения и зависимости между технологическими и экономическими параметрами производственного процесса. В расчетах по модели используется аппарат матричной алгебры. Основой для построения модели служит схема технологического процесса, в которой показаны связи между отдельными его элементами, и для каждой операции установлены нормативы затрат живого и овеществленного труда, приходящиеся на конечный продукт производства. Для полиграфического производства таким конечным продуктом является тираж издания, его часть (завод) или (если речь идет о малотиражных, близких по своим характеристикам изданиях) группа тиражей, объединяемых в одну совокупность. При построении модели в схеме технологического процесса указывается количество продукции, произведенной на каждой операции (стадии) технологического процесса, количество продукции, которое использовано в качестве полуфабриката на последующей операции, и количество продукции, изготовленной в результате осуществления всего технологического процесса. На рис. 11.1 показана схема условного

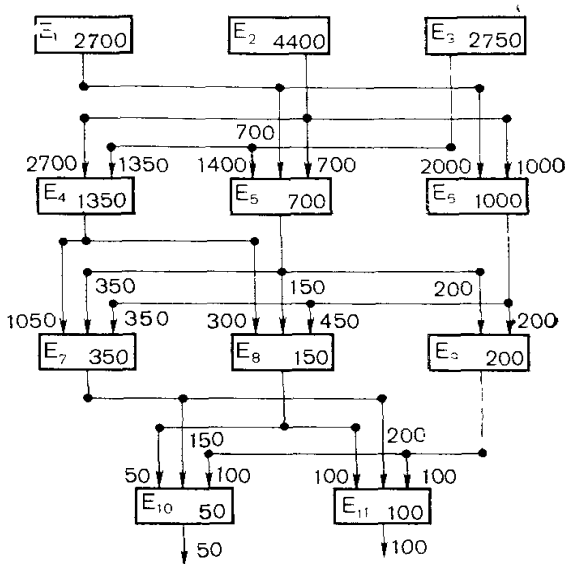


Рис. 11.1. Схема технологического процесса

технологического процесса в виде прямоугольников, обозначающих операцию, и стрелок, характеризующих связи между операциями. Число, стоящее в прямоугольнике, показывает количество продукции, произведенной на данной операции в условных единицах. Эта продукция поступает в виде полуфабрикатов на другие операции, ее количество выражено числами, стоящими около стрелок, либо является конечной для всего технологического процесса. Маршрут продукции можно проследить по направлениям стрелок.

Представленную на рис. 11.1 схему можно интерпретировать следующим образом:

E₁, E₂, E₃ – операции изготовления набора (1-я стадия процесса);
 E₄, E₅, E₆ – операции изготовления печатных форм (2-я стадия);
 E₇, E₈, E₉ – печать продукции (3-я стадия);
 E₁₀, E₁₁ – брошюровочно-переплетные и отделочные операции (4-я стадия).

На рис. 11.1 показано, что элементы, принадлежащие одной стадии, не связаны друг с другом, т. е. изготовленная на данной стадии продукция не может быть использована здесь же в

качестве полуфабриката. Она целиком потребляется на следующей стадии технологического процесса. Схема технологического процесса служит основой для составления таблицы производственных связей (табл. 11.1).

Таблица 11.1

Передача от $E_i \setminus nE_j$	E_1	E_2	E_3	E_4	E_5	E_6	E_7	E_8	E_9	E_{10}	E_{11}	$\underline{m}$	$\underline{\omega}$	$\underline{x}$
E_1					700	2000						2700	–	2700
E_2				2700	700	1000						4400	–	4400
E_3				1350	1400							2750	–	2750
E_4							1050	300				1350	–	1350
E_5							350	150	200			700	–	700
E_6							350	450	200			1000	–	1000
E_7										150	200	350	–	350
E_8										60	100	150	–	150
E_9										100	100	200	–	200
E_{10}													50	50
E_{11}													100	100
Общий объем выпуска по E_i ($\underline{x}^T$)	2700	4400	2750	1350	700	1000	350	150	200	50	100			

Введем ряд обозначений. Пусть $\underline{m}$ – количество продукции, произведенной на определенной операции технологического процесса и использованной как полуфабрикат на следующей стадии; $\underline{\omega}$ – количество конечной продукции технологического процесса, $\underline{x}$ – объем производства, включающий в себя как полуфабрикаты, так и конечную продукцию. Тогда в векторной записи объем производства

$$\underline{x} = \underline{m} + \underline{\omega}.$$

На основе таблицы производственных связей (см. табл. 11.1) составляется матрица связей

$$M = \|m_{ij}\| \quad i, j = 1, 2, \dots, 11;$$

где m_{ij} обозначает количество продукции, произведенное на операции E_i и переданное на операцию E_j , причем оно отнесено ко всему количеству продукции, произведенному на операции E_j . Таким образом, m_{ij} является показателем расхода продукции на единицу продукции E_j . Матрица связей показана на рис. 11.2.

$$\underline{M} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Рис. 11.2. Матрица связей

Из матричных соотношений

$$\underline{x} = \underline{m} + \underline{\omega} \text{ и } \underline{m} = \underline{M}\underline{x}$$

вытекает, что производство конечной продукции связано с выпуском полуфабрикатов матричным соотношением $\underline{\omega} = (\underline{E} - \underline{M})\underline{x}$ (где $\underline{E}$ – единичная матрица), которое представляет собой производственное уравнение, а $(\underline{E} - \underline{M})$ есть производственная матрица, показанная на рисунке 11.3.

$$\underline{E} - \underline{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -3 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 & -3 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -3 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Рис. 11.3. Производственная матрица

Она соответствует технологической схеме с коэффициентами прямых затрат (рис. 11.4). Знак минус показывает, что на данной операции используется полуфабрикат, произведенный на предыдущей стадии технологического процесса.

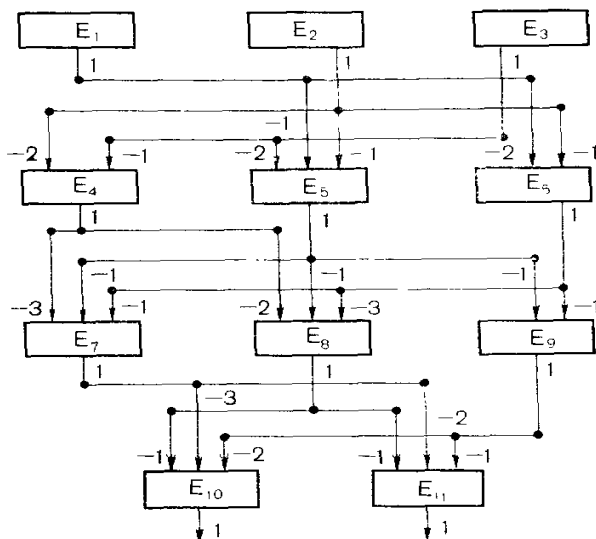


Рис. 11.4. Технологическая схема с коэффициентами прямых затрат

Нередко производственную матрицу называют технологической или матрицей коэффициентов прямых затрат (матрицей прямых затрат). Производственное уравнение показывает, сколько конечной продукции получится при заданном объеме производства. Однако более важна задача определения объема производства при заданном количестве конечной продукции. Для ее решения производственное уравнение следует перевести в форму модели планирования производства. С этой целью выполняются следующие матричные преобразования:

$$\underline{x} = (\underline{E} - \underline{M})^{-1} \underline{\omega},$$

где $(\underline{E} - \underline{M})^{-1}$ – матрица плана или матрица полных затрат (матрица коэффициентов полных затрат). Она показана на рис. 11.5. Из записи

$$(\underline{E} - \underline{M})^{-1} = \underline{B} = \|b_{ij}\|$$

следует, что b_{ij} – показывает, сколько единиц продукции следует произвести па каждой операции (стадии), чтобы в итоге получить единицу готовой продукции. Иначе говоря, это есть расход полуфабриката на единицу готового изделия.

$$(\underline{E} - \underline{M})^{-1} = \left\| \begin{array}{cccccccccccc} 1 & 0 & 0 & 0 & 1 & 2 & 3 & 7 & 3 & 22 & 16 \\ 0 & 1 & 0 & 2 & 1 & 1 & 8 & 8 & 2 & 36 & 26 \\ 0 & 0 & 1 & 1 & 2 & 0 & 5 & 4 & 2 & 23 & 16 \\ 0 & 0 & 0 & 1 & 0 & 0 & 3 & 2 & 0 & 11 & 8 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 6 & 4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 3 & 1 & 8 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right\|$$

Рис. 11.5. Матрица коэффициентов полных затрат

Примеры расчета величин m_{ij} :

$$m_{15} = \frac{700}{700};$$

$$m_{25} = \frac{700}{700};$$

$$m_{35} = \frac{1400}{700};$$

$$m_{18} = \frac{2000}{1000};$$

$$m_{28} = \frac{1000}{1000}.$$

Вообще для определения затрат всех видов ресурсов на производство должны использоваться нормативные данные или, если их нет, статистические данные. При этом следует помнить, что расход некоторых ресурсов пропорционален количеству выпускаемой продукции, а расход других ресурсов пропорционален времени, затраченному на ее производство. Наглядное представление об этом дает рис. 11.6. Рассмотрен только один элемент массива $E_j = E_{11}$ (затраты приведены без расшифровки по статьям).

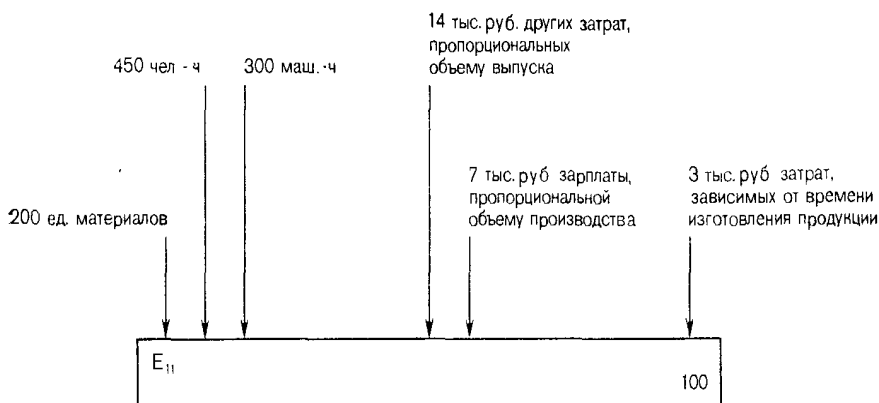


Рис. 11.6. Введение ресурсов в технико-экономическую модель

Коэффициенты прямых затрат, т. е. расход ресурсов на единицу продукции, рассчитываются делением величины затрат на количество произведенной продукции. В табл. 11.2 величина затрат показана в столбце E_{11} , а количество продукции – в строке «Общий объем производства».

Введем обозначения:

Q_1 – матрица коэффициентов прямых затрат по статьям, зависящим от объема производства; Q_2 – матрица коэффициентов прямых затрат по статьям, зависящим от затрат времени, $\underline{x}$ – вектор объема производства; $\underline{t}$ – вектор времени выпуска продукции $\overline{E_j}$; $\underline{a}_1$ – вектор полных затрат, зависящих от объема производства; $\underline{a}_2$ – вектор полных затрат, зависящих от времени выпуска.

Таблица 11.2

Затраты	Зависимые от производства	Не зависимые от производства
E_i	$E_1 E_2 \dots E_{11}$	$E_1 E_2 \dots E_{11}$
Вид затрат		
Материалы	200	
Трудоемкость чел.ч	450	
Трудоемкость маш. ч	300	
Зарплата, тыс руб	7	
Другие затраты, тыс руб	14	3
Общий объем производства по E	100	
Время изготовления		200

Соответствующие уравнения имеют вид: $\underline{Q}_1 \cdot \underline{x} = \underline{a}_1$; $\underline{Q}_2 \cdot \underline{t} = \underline{a}_2$.

Объединяя векторы $\underline{a}_1$ и $\underline{a}_2$ в вектор совокупных полных

затрат, получаем выражение: $\| \underline{Q}_1 \underline{Q}_2 \| \cdot \left\| \frac{\underline{x}}{\underline{t}} \right\| = \| \underline{a} \|$.

В нашем примере (рис. 11.6 и табл. 11.2) вектору $\underline{Q}_1$ соответствует вектор $q_{11} = (2 \ 4,5 \ 3, \ 0,07 \ 0,14)^T$, а вектору $\underline{Q}_2$ – вектор $p_{11} = (0 \ 0 \ 0 \ 0 \ 0 \ 15)^T$.

Объединяя производственное уравнение с уравнением затрат, получаем технико-экономическую модель производства

$$\left\| \begin{array}{c} \underline{E} - \underline{M} \\ \underline{Q}_1 \end{array} \right\| \cdot \left\| \frac{\underline{x}}{\underline{t}} \right\| = \left\| \frac{\underline{\omega}}{\underline{a}} \right\|.$$

где $\underline{O}$ – нулевая матрица. Технико-экономическая модель планирования производства получается путем обращения предыдущего матричного уравнения:

$$\left\| \begin{array}{c} (\underline{E} - \underline{M})^{-1} \\ \underline{Q}_1 (\underline{E} - \underline{M})^{-1} \end{array} \right\| \cdot \left\| \frac{\underline{\omega}}{\underline{t}} \right\| = \left\| \frac{\underline{x}}{\underline{a}} \right\|.$$

Элементы вектора $\underline{x}$ дают представление о количестве продукции, произведенной на каждом отдельном производственном участке E_i , учитываемом моделью; элементы вектора $\underline{a}$ – о необходимых затратах ресурсов, зависящих как от объема производства, так и от затраченного времени. Чтобы рассчитать, сколько полуфабрикатов, изготовленных на отдельном производственном участке, направляется на другие участки, и определить по участкам потребность в ресурсах для выпуска продукции, элементы векторов ω и t записываются в виде диагональных матриц $\underline{D}_\omega$ и $\underline{D}_t$, которые, в свою очередь, образуют диагональную матрицу. Тогда модель планирования производства примет следующий вид:

$$\left\| \begin{pmatrix} \underline{E} - \underline{M} \end{pmatrix}^{-1} \quad \underline{Q} \right\| \cdot \left\| \begin{pmatrix} \underline{D}_\omega & \underline{O} \\ \underline{O} & \underline{D}_t \end{pmatrix} \right\| = \left\| \begin{pmatrix} \underline{X} & \underline{O} \\ \underline{A}_1 & \underline{A}_2 \end{pmatrix} \right\|.$$

Расчеты по модели дают следующие результаты:

1. В i -й строке матрицы $\underline{X}$ дано распределение продукции (полуфабриката), изготовленной на i -м производственном участке, по другим производственным участкам, потребляющим эту продукцию, причем итог по i -й строке матрицы $\underline{X}$ равен количеству изготовленной на i -м участке продукции, т. е. элементу x_i вектора $\underline{x}$.

2. В j -м столбце матрицы $\underline{X}$ представлено количество полуфабрикатов, получаемых j -м производственным участком от других участков. При $i=j$ элемент x_{ii} матрицы $\underline{X}$ показывает количество изготовленной на j -м участке продукции.

3. В i -й строке матрицы $\underline{A}_1$ дано распределение i -го ресурса, расход которого пропорционален объему производства, по производственным участкам E_i , причем итог по i -й строке матрицы $\underline{A}_1$ равен полным затратам i -го ресурса на выпуск продукции, т. е. элементу a_i вектора $\underline{a}$.

4. В j -м столбце матрицы $\underline{A}_1$ представлено количество различных ресурсов, пропорциональных объему производства и необходимых для выпуска продукции на j -м участке.

5. i -я строка матрицы $\underline{A}_2$ содержит данные, аналогичные приведенным в п. 3, но в этом случае речь идет о ресурсах, расход которых пропорционален затраченному времени.

6. В j -м столбце матрицы $\underline{A}_2$ находятся показатели, аналогичные представленным в п. 4, но относящиеся к ресурсам, расходуемым пропорционально затратам времени.

7. Сложение соответствующих элементов a_i векторов $\underline{a}_1$ и $\underline{a}_2$ дает элемент a_i вектора $\underline{a}$.

Технико-экономические модели позволяют определить возможности производства по выпуску данного вида продукции или потребности в разного рода ресурсах для реализации заданной производственной программы. В общем случае вопрос об обеспеченности программы ресурсами решается, исходя из соотношений:

$$\underline{x} = (\underline{E} - \underline{M})^{-1} \underline{\omega},$$

$$\underline{a}_1 = \underline{Q}_1 (\underline{E} - \underline{M})^{-1} \underline{\omega}.$$

Рассчитывая варианты производства при различных заданных значениях $\underline{\omega}^1$ ($\underline{\omega}^1$ – количество конечной продукции по i -му варианту производства), можно определить общий объем производства $\underline{x}^1$ и потребности в ресурсах $\underline{a}_1^i$. Сопоставляя их соответственно с наличными мощностями $\underline{k}$ производственных участков и имеющимися ресурсами $\underline{r}$, определяют возможность реализации рассматриваемых вариантов. Необходимыми условиями реализации являются соотношения:

$$\underline{x}^i \leq \underline{k},$$

$$\underline{a}_1^i \leq \underline{r} - \underline{a}_2 = \underline{\tilde{r}}.$$

Соблюдение данных условий создает предпосылки для отыскания оптимального варианта, но это не означает, что вариант производства, отвечающий этим условиям, оптимален. Ответ на этот вопрос дают только оптимизационные методы.

11.2. Оптимизационные модели

Поиск оптимального решения – важнейшая проблема экономики и технологии. Для ее решения используется разнообразный математический аппарат, в частности, линейное программирование.

Линейное программирование основывается на таких математических методах, которые обеспечивают нахождение максимума или минимума линейной функции при наличии ограничений на переменные.

Области применения линейного программирования: оперативно-календарное планирование; выбор технологии или альтернативных вариантов; раскрой материалов; определение маршрутов внутризаводского транспорта; размещение машин, установок, строений.

При разработке математической модели линейного программирования в первую очередь необходимо решить вопрос о выборе критерия оптимальности, т. е. определить задачу оптимизации. В этом случае говорят о задании целевой функции.

В качестве критерия оптимальности могут быть приняты: максимум производства продукции в оптовых ценах (реализация); максимум производства продукции в натуральных единицах; максимум прибыли; минимум приведенных затрат; максимум использования производственных мощностей; минимум себестоимости; минимум материальных затрат; минимум протяженности транспортных маршрутов.

Так как оптимизация линейных функций происходит при наличии ограничений, то в состав экономико-математической модели входят и эти ограничения в виде неравенств или уравнений.

Ограничения накладываются на ресурсы, в качестве которых могут выступать рабочая сила (фонд времени работы персонала), материалы, оборудование (фонд времени работы оборудования).

В общем случае требующиеся ресурсы должны по величине быть меньше или равны имеющимся в наличии. Однако наряду с ограничениями вида «меньше – равно» могут

существовать ограничения вида «больше – равно» (минимальная потребность) и в форме равенств.

Экономико-математическая модель содержит требование неотрицательных переменных. Это понятно: варианты решения имеют смысл тогда, когда они неотрицательны.

В модель, как правило, вносятся упрощения (например, средние отпускные цены, средние затраты на выпуск продукции, средние нормы времени и выработки и т. п.). Это позволяет резко сократить трудоемкость расчетов, уменьшить затраты времени на подготовку и ввод исходных данных в ЭВМ, свести к минимуму затраты машинного времени (если расчеты ведутся на ЭВМ). Важно, чтобы вносимые допущения не сказались решающим образом на принципиальных решениях при проектировании производственных процессов.

Таким образом, модель линейного программирования включает в себя: целевую функцию, ограничения, условие неотрицательности.

Введем следующие обозначения:

z – значение целевой функции,

$\underline{\omega} = \|\omega_j\| \quad j = 1, 2, \dots, n$ вектор конечной продукции производства в решении модели;

$\underline{p} = \|p_j\| \quad j = 1, 2, \dots, n$ вектор критериальной оценки;

$\underline{x} = \|x_j\| \quad j = 1, 2, \dots, n$ вектор объема производства;

$\underline{b} = \|b_j\| \quad j = 1, 2, \dots, n$ вектор общественной потребности в продукции, причем индекс *min* соответствует минимальной потребности, индекс *max* – максимальной;

$\underline{k} = \|k_j\| \quad j = 1, 2, \dots, n$ вектор мощности производственных участков;

$\underline{r} = \|r_i\| \quad i = 1, 2, \dots, n$ вектор ресурсов, расход которых пропорционален объему выпуска продукции (материалы, рабочее время и т. п.)

$\underline{Q} = \|q_{ij}\| \quad i = 1, 2, \dots, n$ матрица коэффициентов прямых
 $j = 1, 2, \dots, n$ затрат по статьям, зависшим от объема производства;

$\underline{E} - \underline{M} = \underline{\tilde{M}}$ производственная матрица (см. раздел 10.1), где $\underline{M} = \|m_{ij}\|$ $i, j = 1, 2, \dots$.

Математическая модель оптимизации планирования производства имеет вид:
целевая функция

$$z = \underline{p}^T \underline{\omega} \rightarrow opt;$$

ограничения

$$\underline{Q}_1 \underline{x} \leq \underline{\tilde{r}}$$

$$\underline{x} \leq \underline{k}$$

$$\underline{\omega} = (\underline{E} - \underline{M}) \underline{x}$$

$$\underline{b}_{\min} \leq \underline{\omega} \leq \underline{b}_{\max};$$

условие неотрицательности

$$\underline{\omega} \geq \underline{O}.$$

Благодаря такой записи мы получаем вместо неравенств и n уравнений только $m + 3$ неравенств. Модель может быть сокращена еще на n неравенств, если принять, что $\underline{\omega} \geq \underline{b}_{\min}$.

Тогда

$\underline{\tilde{\omega}} = \underline{\omega} - \underline{b}_{\min}$ – вектор количества конечной продукции, произведенной сверх минимального уровня;

$\underline{\tilde{b}} = \underline{b}_{\max} - \underline{b}_{\min}$ – вектор общественных потребностей, превышающих минимальный уровень;

$\underline{\tilde{r}} = \underline{\tilde{r}} - \underline{Q}_1 (\underline{E} - \underline{M}) \underline{b}_{\min}$ – вектор ресурсов, необходимых для изготовления продукции сверх минимального количества;

$\underline{\tilde{k}} = \underline{k} - (\underline{E} - \underline{M})^{-1} \underline{b}_{\min}$ – вектор мощностей, не задействованных при изготовлении минимального количества продукции;

$c = \underline{p}^T \cdot \underline{b}_{\min}$ – значение критерия оптимальности при минимальном количестве продукции.

Так как $\underline{x} = (\underline{E} - \underline{M})^{-1} \underline{\omega}$, то модель описывается следующими соотношениями:

целевая функция

$$z = \underline{p}^T \underline{\omega} \rightarrow opt;$$

ограничения

$$\underline{Q}_1(\underline{E} - \underline{M})^{-1} \underline{\omega} \leq \underline{\tilde{r}}$$

$$(\underline{E} - \underline{M})^{-1} \underline{\omega} \leq \underline{k}$$

$$\underline{b}_{\min} \leq \underline{\omega} \leq \underline{b}_{\max};$$

условие неотрицательности

$$\underline{\omega} \geq \underline{O}.$$

С этими обозначениями сокращенная модель записывается следующим образом:

целевая функция

$$z = \underline{p}^T \underline{\tilde{\omega}} + c \rightarrow opt;$$

ограничения

$$\underline{Q}_1(\underline{E} - \underline{M})^{-1} \underline{\tilde{\omega}} \leq \underline{\tilde{r}};$$

$$(\underline{E} - \underline{M})^{-1} \underline{\tilde{\omega}} \leq \underline{\tilde{k}};$$

$$\underline{\tilde{\omega}} = \underline{\tilde{b}};$$

условие неотрицательности

$$\underline{\tilde{\omega}} \geq \underline{O}.$$

Обозначив через

$$\underline{F} = \left\| \begin{array}{c} \underline{Q_1}(\underline{E} - \underline{M})^{-1} \\ (\underline{E} - \underline{M})^{-1} \\ \underline{E} \end{array} \right\| \quad \text{и} \quad \underline{f} = \left\| \begin{array}{c} \underline{\tilde{r}} \\ \underline{\tilde{k}} \\ \underline{\tilde{b}} \end{array} \right\|.$$

получаем

$$z = \underline{p}^T \underline{\tilde{\omega}} + c \rightarrow opt;$$

$$\underline{F} \underline{\tilde{\omega}} < \underline{f};$$

$$\underline{\tilde{\omega}} \geq \underline{O}.$$

В следующем разделе эта модель рассматривается в несколько упрощенном виде:

$$z = \underline{p}^T \underline{\omega} \rightarrow opt;$$

$$\underline{A} \underline{\omega} \leq \underline{a};$$

$$\underline{\omega} \geq \underline{O}.$$

Решение по этой модели отыскивается симплекс-методом. [Симплекс (от латинского *simplex* – простой), простейший выпуклый многогранник данного числа измерений n . При $n = 3$ трёхмерный симплекс представляет собой произвольный тетраэдр. Под двухмерным симплексом понимают произвольный треугольник, а под одномерным – отрезок. Нульмерный симплекс есть просто точка. Для этого в модель наряду с основными переменными ω_j вводятся дополнительные переменные v_k . Число дополнительных переменных равно числу накладываемых ограничений. Таким образом, в сокращенную модель вводятся $m + 2n$ дополнительных переменных v_k ($k = 1, 2, \dots, p$ при $p = m + 2n$).

В экономико-математической модели могут быть учтены различные способы изготовления продукции. Существуют две возможности для этого: 1) расширение индикации переменных в модели путем введения индекса S , тогда j -я продукция,

изготовленная по технологии S , будет обозначаться ω_{jS} ($S = 1, 2, \dots, t$); 2) одна и та же продукция, но изготовленная различными технологическими способами, рассматривается как разные виды продукции, т. е. продукция ω_{jS} расчленяется на $S\omega_j$ видов продукции.

11.3 Примеры постановки и решения оптимизационных задач

Пример 1.

Предприятие выпускает продукцию трех видов – E_1 , E_2 и E_3 .

Расход некоторого материала на каждый из видов продукции равен соответственно 2 т, 4 т и 1 т. Максимальная потребность предприятия в этом материале определена в размере 240 т.

Поставлены следующие условия: продукция видов E_1 и E_2 должна быть изготовлена в равных количествах; продукции вида E_3 должно быть изготовлено вдвое больше, чем продукции E_2 .

Требуется найти оптимизированную по прибыли производственную программу, причем прибыль по каждому виду продукции равна соответственно 5, 1 и 2 условным денежным единицам.

Математическая модель:
целевая функция

$$z = 5\omega_1 + 1\omega_2 + 2\omega_3 \longrightarrow \max; \quad (11.3.1)$$

ограничения:

$$2\omega_1 + 4\omega_2 + 1\omega_3 \leq 240 \quad (11.3.2)$$

$$\omega_1 = \omega_2; \quad (11.3.3)$$

$$2\omega_2 = \omega_3; \quad (11.3.4)$$

условие неотрицательности

$$\omega_1, \omega_2, \omega_3 \geq 0.$$

В данном примере можно, воспользовавшись равенствами (11.3.3) и (11.3.4), заменить переменные ω_1 и ω_3 и записать выражения (11.3.1) и (11.3.2) в следующем виде:

$$z = 10\omega_2 + 0\nu_1 \rightarrow \max; \quad (11.3.5)$$

$$8\omega_2 + 1\nu_1 = 240. \quad (11.3.6)$$

Задача решается симплексным методом. Не приводя подробного решения, представим полученные результаты: $z = 300$, $\omega_2 = 30$, $\nu_1 = 0$.

Таблица 11.3

Технико-экономический показатель (на единицу продукции)	Номер машины	Вид продукции		
		Е ₁	Е ₂	Е ₃
Затраты времени	1	1	2	—
	2	2	1	1
	3	1	—	1
	4	—	3	2
Себестоимость		15	10	21
Прибыль		5	1	9
Отпускная цена		20	11	30

Отсюда окончательное решение: $\nu_1 = 0$, $\omega_1 = 30$, $\omega_2 = 30$, $\omega_3 = 60$, $z = 300$, т. е. обеспечено получение прибыли в размере 300 условных денежных единиц при выпуске 30 единиц продукции видов E_1 и E_2 и 60 единиц продукции вида E_3 (z , ω_1 , ω_2 , ω_3); материальные ресурсы использованы полностью ($\nu_1 = 0$). Если $\nu_1 > 0$, то это бы означало, что ресурсы исчерпаны не полностью.

Пример 2.

Предприятие выпускает три вида продукции. Для этой цели оно располагает четырьмя машинами. Фонд основного рабочего времени машины № 1 составляет 30 условных единиц времени, машины № 2 – 45 единиц, у машины № 3 – 20 единиц и машины № 4 – 50 единиц. Технико-экономические показатели процесса представлены в табл. 11.3.

Производственным планом предприятия предусматривается выпуск продукции в следующих количественных диапазонах: $E_1 - (2-15)$, $E_2 - (6-20)$, $E_3 - (1-8)$.

Предполагается составить математическую модель, оптимизируя: а) производственную программу по прибыли; б) выпуск продукции в стоимостном выражении; в) выпуск продукции в натуральном выражении; г) производственную программу по себестоимости.

Математическая модель

Целевые функции:

$$\text{а) } z = 5\omega_1 + 1\omega_2 + 2\omega_3 \rightarrow \max;$$

$$\text{б) } z = 20\omega_1 + 11\omega_2 + 30\omega_3 \rightarrow \max;$$

$$\text{в) } z = \omega_1 + \omega_2 + \omega_3 \rightarrow \max;$$

$$\text{г) } z = 15\omega_1 + 10\omega_2 + 21\omega_3 \rightarrow \max;$$

Ограничения (действительны для всех вариантов):

$$\omega_1 + 2\omega_2 \leq 30; \quad (11.3.1')$$

$$2\omega_1 + 4\omega_2 + 1\omega_3 \leq 45; \quad (11.3.2')$$

$$1\omega_1 + 1\omega_3 \leq 20; \quad (11.3.3')$$

$$3\omega_1 + 2\omega_3 = 50 \quad (11.3.4')$$

$$\omega_1 \geq 2; \quad (11.3.5')$$

$$\omega_2 \geq 6; \quad (11.3.6')$$

$$\omega_3 \geq 1; \quad (11.3.7')$$

$$\omega_1 \leq 15; \quad (11.3.8')$$

$$\omega_2 \leq 20; \quad (11.3.9')$$

$$\omega \geq 8. \quad (11.3.10')$$

При наличии выражений (11.3.5'), (11.3.6') и (11.3.7') запись условий неотрицательности не нужна.

В этом примере также можно сократить число переменных и уравнений. Для этого рассмотрим неравенства (11.3.5'), (11.3.6') и (11.3.7'), входящие в математическую модель:

$$\omega_1 \geq 2;$$

$$\omega_2 \geq 6;$$

$$\omega \geq 1.$$

Запишем их в форме:

$$\omega_1^* = \omega_1 - 2 \geq 0;$$

$$\omega_2^* = \omega_2 - 6 \geq 0; \quad \omega_1^*, \omega_2^*, \omega_3^* \geq 0,$$

$$\omega_3^* = \omega_3 - 1 \geq 0;$$

т. е.

$$\omega_1^* + 2 = \omega_1;$$

$$\omega_2^* + 6 = \omega_2;$$

$$\omega_3^* + 1 = \omega_3.$$

Если теперь ввести в целевую функцию и ограничения (11.3.1'), (11.3.2'), (11.3.3'), (11.3.4'), (11.3.8'), (11.3.9') и (11.3.10') вместо ω_j значения ω_j^* , то получим:

$$z = 5\omega_1^* + \omega_2^* + 9\omega_3^* + 25 \rightarrow \max;$$

$$\omega_1^* + 2\omega_2^* \leq 30; \quad (11.3.1'')$$

$$2\omega_1^* + \omega_2^* + \omega_3^* \leq 34; \quad (11.3.2'')$$

$$\omega_1^* + \omega_3^* \leq 17; \quad (11.3.3'')$$

$$3\omega_2^* + 2\omega_3^* \leq 30; \quad (11.3.4'')$$

$$\omega_1^* \leq 13; \quad (11.3.8') \rightarrow (11.3.5'')$$

$$\omega_2^* \leq 14; \quad (11.3.9') \rightarrow (11.3.6'')$$

$$\omega_3^* \leq 7. \quad (11.3.10') \rightarrow (11.3.7'')$$

Это соответствует сокращению модели по нижней границе диапазона выпуска продукции. Тем самым сокращается и число неравенств. С введением в модель дополнительных переменных имеем (демонстрируется на примере варианта оптимизации «а»):

$$\begin{aligned} z &= 5\omega_1^* + \omega_2^* + 9\omega_3^* + 25 \rightarrow \max; \\ \omega_1^* + 2\omega_2^* + 1v_1 &= 16; \\ 2\omega_1^* + \omega_2^* + \omega_3^* + 1v_2 &= 34; \\ \omega_1^* + \omega_3^* + 1v_3 &= 17; \\ 3\omega_2^* + 2\omega_3^* + 1v_4 &= 30; \\ \omega_1^* + 1v_5 &= 13; \\ \omega_2^* + 1v_6 &= 14; \\ \omega_3^* + 1v_7 &= 7. \end{aligned}$$

Решение дает следующие результаты:

$$\begin{aligned} \omega_1^* &= 10; \omega_2^* = 3; \omega_3^* = 7; \\ v_1 &= 0; v_2 = 4; v_3 = 0; v_4 = 7; v_5 = 3; v_6 = 11; v_7 = 0; \\ z &= 141. \end{aligned}$$

Возвращаясь к первоначальным переменным, имеем:

$$\begin{aligned} \omega_1 &= 12; \\ \omega_2 &= 9; \\ \omega_3 &= 8; \\ z &= 141. \end{aligned}$$

При этом переменные $v_1, \dots, v_7$ сохраняют свои значения.

Поставленная задача решена и можно сделать следующие выводы:

1. Предприятие получит максимальную прибыль в размере 141 условной денежной единицы, если будет изготавливать 12, 9 и 8 условных единиц продукции соответственно вида E_1 , E_2 и E_3 .

2. Машины № 1 и № 3 загружены полностью ($v_1 = 0$, $v_3 = 0$), в то время как машины № 2 и № 4 располагают резервом времени в размере соответственно 4 и 7 единиц ($v_2 = 4$, $v_4 = 7$). Эти машины можно догрузить другими заказами.

3. Если увеличить фонд времени основной работы оборудования, то можно изготовить еще 3 единицы продукции E_1 ($\nu_5 = 3$) и 11 единиц продукции E_2 ($\nu_6 = 11$). Продукция E_3 изготавливается в максимальном количестве ($\nu_7 = 0$).

4. Минимальный уровень производства превзойден по всем видам продукции.

Пример 3. Оптимизация производственной программы при различных вариантах технологии изготовления.

Предприятие может изготавливать три вида продукции двумя различными технологическими способами. Требуемые показатели приведены в табл. 11.4.

Таблица 11.4

Продукция		1		2		3			
Потребность в продукции, ед.		200		100		250			
Затраты времени, ч/ед., по технологии		1	2	1	2	1	2	Фонд времени, ч	
								1	2
Технологические участки	1	0,4	0,8	0,4	–	0,4	0,2	150	150
	2	0,4	–	0,6	0,4	0,2	1,2	84	316
	3	0,4	0,6	0,8	2,0	1,4	1,8	160	440
Себестоимость (ден. ед./ед.)		20	24	25	20	30	23		

Первоначальная модель

$$z = 20\omega_{11} + 24\omega_{12} + 25\omega_{21} + 20\omega_{22} + 30\omega_{31} + 23\omega_{32} \rightarrow \min;$$

$$0,4\omega_{11} + 0,4\omega_{21} + 0,4\omega_{31} \leq 150;$$

$$0,8\omega_{12} + 0,2\omega_{32} \leq 150;$$

$$0,4\omega_{11} + 0,6\omega_{21} + 0,2\omega_{31} \leq 84;$$

$$+ 0,4\omega_{22} + 1,2\omega_{32} \leq 316;$$

$$0,4\omega_{11} + 0,8\omega_{21} + 1,4\omega_{31} \leq 160;$$

$$0,6\omega_{12} + 2,0\omega_{21} + 0,8\omega_{32} \leq 440;$$

$$\omega_{11} + \omega_{12} = 200;$$

$$\omega_{21} + \omega_{22} = 100;$$

$$\omega_{31} + \omega_{32} = 250.$$

Уравнения приведения:

$$\begin{aligned}\omega_{11} &= 200 - \omega_{12}, \text{ причем } \omega_{12} \leq 200, \\ \omega_{21} &= 100 - \omega_{22}, \text{ причем } \omega_{22} \leq 100, \\ \omega_{31} &= 250 - \omega_{32}, \text{ причем } \omega_{32} \leq 250.\end{aligned}$$

Сокращенная модель:

$$\begin{aligned}z &= 4\omega_{12} - 5\omega_{22} - 7\omega_{32} + 14000 \rightarrow \min; \\ 0,8\omega_{12} + 0,2\omega_{32} &\leq 150; \\ 0,4\omega_{22} + 1,2\omega_{32} &\leq 316; \\ 0,6\omega_{12} + 2,0\omega_{22} + 0,8\omega_{32} &\leq 440; \\ 0,4\omega_{12} + 0,4\omega_{22} + 0,4\omega_{32} &\leq 70; \\ 0,4\omega_{12} + 0,6\omega_{22} + 0,2\omega_{32} &\leq 106; \\ 0,4\omega_{12} + 0,8\omega_{22} + 1,4\omega_{32} &\leq 350; \\ \omega_{12} &\leq 200; \\ \omega_{22} &\leq 100; \\ \omega_{32} &\leq 250.\end{aligned}$$

Результаты решения:

$$\begin{aligned}z &= 11890; \\ \omega_{12} &= 0; \\ \omega_{22} &= 100; \\ \omega_{32} &= 230.\end{aligned}$$

Комментарий к решению:

количество продукции 1, изготавливаемой по 1-й технологии – 200 ед.

»	1,	»	2-й	»	0 ед.
»	2,	»	1-й	»	0 ед.
»	2,	»	2-й	»	100 ед.
»	3,	»	1-й	»	20 ед.
»	3,	»	2-й	»	230 ед.

Загрузка (свободные мощности):

на 1-м технологическом участке при работе по 1-й технологии – 88 (62) ед.

на 1-м	»	по 2-й	»	46(104)
на 2-м	»	по 1-й	»	84(0)
на 2-м	»	по 2-й	»	316(0)
на 3-м	»	по 1-й	»	108(52)
на 3-м	»	по 2-й	»	384(56)

Любая модель линейного программирования имеет число решений, равное числу сочетаний из n по m , где n – количество основных и дополнительных переменных, а m – число равенств и неравенств в модели. Если первоначальная модель, у которой $m = 9$ и $n = 15$, могла дать 5005 решений, то сокращенная модель при $m = 9$ и $n = 12$ имеет только 220 решений.

Пример 4. В примере рассматривается загрузка печатного и брошюровочно-переплетного участков. Фонд времени основной работы оборудования на каждом участке равен соответственно 1100 и 470 ч. В отчетном периоде на участке печати из 2400 единиц печатных материалов было получено 1100 тысяч листов печатной продукции. Из этого количества 800 тысяч листов идут в дальнейшую обработку на брошюровочно-переплетный участок партиями по 80 тысяч экземпляров, а остаток – партиями по 50 тысяч экземпляров (издания E_2 и E_3). Обработка 1 тысяч экземпляров издания E_2 занимает здесь 4 ч, а издания E_3 – 3 ч. На каждую тысячу экземпляров издания E_2 расходуется 1,05 единиц переплетного материала, на 1 тысяч экземпляров издания E_3 – 0,52 единицы. Запасы переплетного материала в типографии составляют 110 единиц.

В плановом периоде при тех же материальных ресурсах, но при лучшем их использовании (2 единицы печатных материалов на 1 тыс. листов, 1 единица переплетного материала на 1 тысяч экземпляров издания E_2 и 0,5 единицы переплетного материала на 1 тысяч экземпляров издания E_3 , сокращение простоев на 100 ч на участке печати и на 30 ч на брошюровочно-переплетном участке) предполагается увеличить объем производства.

Требуется ответить на следующие вопросы: а) возможно ли выполнение планового задания, включающего в себя 82 тысяч экземпляров издания E_2 и 60 тысяч экземпляров издания E_3 ? б) каково максимально возможное увеличение объема производства? в) как изменится максимально возможный объем производства, если: 1) потребуется максимально 80 тысяч экземпляров издания E_3 ? 2) нужно выпустить по меньшей мере 50 тысяч экземпляров издания E_2 и 40 тысяч экземпляров издания E_3 ? 3) должны сохраниться пропорции отчетного периода?

Решение:

(по п. а)

исходя из данных отчетного периода, которые принимаются на плановый период, получаем

$$m_{12} = \frac{800}{80} = 10 \quad \text{и} \quad m_{13} = \frac{300}{50} = 6.$$

Из этого следует:

$$\underline{E} - \underline{M} = \begin{vmatrix} 1 & -10 & -6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \quad \text{или} \quad ((\underline{E} - \underline{M})^{-1} = \begin{vmatrix} 1 & 10 & 6 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}.$$

Для планового периода материала материальных затрат выглядит следующим образом:

$$Q_1 = \begin{vmatrix} 2 & 0 & 0 \\ 0 & 1 & 0,5 \\ 1 & 0 & 0 \\ 0 & 4 & 3 \end{vmatrix} \begin{array}{l} \text{печатные материалы;} \\ \text{переплетные материалы;} \\ \text{время на печать;} \\ \text{время на брошуровочно-переплетные работы} \end{array}$$

наличие ресурсов:

$$\tilde{r} = \begin{vmatrix} 2 & 4 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 & 0 \\ 5 & 0 & 0 \end{vmatrix} \begin{array}{l} \text{печатные материалы;} \\ \text{переплетные материалы;} \\ \text{время на печать;} \\ \text{время на брошуровочно-переплетные работы} \end{array}$$

плановое задание:

$$\underline{\omega} = \begin{vmatrix} 0 \\ 82 \\ 60 \end{vmatrix}.$$

Необходимые для выполнения планового задания ресурсы рассчитываются из соотношения:

$$\underline{a}_1 = \underline{Q}_1(\underline{E} - \underline{M})^{-1} \underline{\omega} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0,5 \\ 1 & 0 & 0 \\ 0 & 4 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 10 & 6 & 0 \\ 0 & 1 & 0 & 82 \\ 0 & 0 & 1 & 60 \end{pmatrix} = \begin{pmatrix} 2360 \\ 112 \\ 1180 \\ 508 \end{pmatrix}.$$

Так как

$$\underline{a}_1 = \begin{pmatrix} 2360 \\ 112 \\ 1180 \\ 508 \end{pmatrix} \neq \begin{pmatrix} 2400 \\ 110 \\ 1200 \\ 500 \end{pmatrix} = \underline{\tilde{r}}$$

II

лановое задание полностью выполнено быть не может (для этого не хватает 2 ед. брошюровочно-переплетных материалов и 8 ч фонда времени основной работы брошюровочно-переплетного оборудования).

(по п. б)

максимально возможный выпуск продукции может быть рассчитан при использовании методов линейного программирования. Математическая модель при этом имеет вид: целевая функция

$$z = \omega_2 + \omega_3 \rightarrow \max;$$

ограничения:

$$2x_1 \leq 2400 \text{ (11.3.1''')} \text{ (материал для печатания);}$$

$$\omega_2 + 0,5\omega_3 \leq 110 \text{ (11.3.2''')} \text{ (переплетный материал);}$$

$$x_1 \leq 1200 \text{ (11.3.3''')} \text{ (время на печать);}$$

$$4\omega_2 + 3\omega_3 \leq 500 \text{ (11.3.4''')} \text{ (время на брошюровочно-переплетные работы);}$$

$$10\omega_2 + 6\omega_3 = x_1 \text{ (9.3.5''')} \text{ (межцеховые отношения);}$$

условие неотрицательности

$$\omega_2, \omega_3 > 0.$$

Так как неравенства (11.3.1''') и (11.3.3''') пропорциональны, одно из них лишнее. Кроме того, можно равенство (11.3.5''') ввести в (11.3.3''') и записать модель следующим образом:

$$\begin{aligned}z &= \omega_2 + \omega_3 \rightarrow \max; \\ \omega_2 + 0,5\omega_3 &\leq 110; \\ 10\omega_2 + 6\omega_3 &\leq 1200; \\ 4\omega_2 + 3\omega_3 &\leq 500; \\ \omega_2, \omega_3 &\geq 0.\end{aligned}$$

Методом графических построений или симплекс-методом получаем решение:

$$\omega_3 \geq 0, \quad \omega_3 = \frac{500}{3} \approx 166,7 \quad \text{и} \quad z = \frac{500}{3} \approx 166,7.$$

Таким образом, максимальный выпуск продукции достигается тогда, когда производственная программа включает только один вид продукции E_3 . Объем производства при этом составляет 166,7 тысяч экземпляров изданий вида E_3 .

(по п. в)

1. В модель, составленную на предыдущем этапе (п. б), достаточно ввести неравенство, в котором записано ограничение на выпуск изданий типа E_3 , не превышающий 80 тысяч экземпляров:

$$\begin{aligned}z &= \omega_2 + \omega_3 \rightarrow \max; \\ \omega_2 + 0,5\omega_3 &\leq 110; \\ 10\omega_2 + 6\omega_3 &\leq 1200; \\ 4\omega_2 + 3\omega_3 &\leq 500; \\ \omega_3 &\leq 80; \\ \omega_2, \omega_3 &\geq 0.\end{aligned}$$

Результат решения $\omega_2 = 65$, $\omega_3 = 80$ и $z = 145$, т. е. максимально возможный выпуск в этом случае составит 145 тысяч экземпляров, в т. ч. 65 тысяч экземпляров издания E_2 и 80 тысяч экземпляров издания E_3 .

2. В модель (п. б) вводятся два неравенства, отражающие поставленные условия; при этом необходимость в условиях неотрицательности переменных отпадает:

$$\begin{aligned}z &= \omega_2 + \omega_3 \rightarrow \max; \\ \omega_2 + 0,5\omega_3 &\leq 110;\end{aligned}$$

$$10\omega_2 + 6\omega_3 \leq 1200;$$

$$4\omega_2 + 3\omega_3 \leq 500;$$

$$\omega_2 \geq 50;$$

$$\omega_3 \geq 40.$$

Результат решения $\omega_2 = 50$, $\omega_3 = 100$, $z = 150$, следовательно, при поставленных ограничениях возможен выпуск 150 тысяч экземпляров, в том числе 50 тысяч экземпляров изданий типа E_2 , 100 тысяч экземпляров изданий типа E_3 .

3. В отчетном периоде пропорция выпуска изданий была:

$$\omega_2 : \omega_3 = 80:50 \quad \text{или} \quad 5\omega_2 = 8\omega_3.$$

В модель (п. б) вводится указанное равенство:

$$z = \omega_2 + \omega_3 \rightarrow \max;$$

$$\omega_2 + 0,5\omega_3 \leq 110;$$

$$10\omega_2 + 6\omega_3 \leq 1200;$$

$$4\omega_2 + 3\omega_3 \leq 500;$$

$$5\omega_2 \leq 8\omega_3;$$

$$\omega_2, \omega_3 \geq 0.$$

Результат решения $\omega_2 \approx 83,8$; $\omega_3 \approx 52,4$; $z = 136,2$.

В этом случае максимально возможный выпуск составит 136,2 тысяч экземпляров, из них 83,8 тысяч экземпляров изданий типа E_2 и 52,4 тысяч экземпляров изданий типа E_3 .

Основная литература: 7 [101–123].

Контрольные вопросы:

1. Техничко-экономическис модели.
2. Оптимизационные модели.
3. Примеры постановки и решения оптимизационных задач.

СЛОВАРЬ ТЕРМИНОВ

Абстрагирование, этот метод научного исследования заключается в отвлечении от несущественных сторон и признаков рассматриваемого объекта.

Адекватность (правильное соответствие) (от латинского *adaequatus* «адекватус» – приравненный, равный, тождественный) – способность отображать заданные свойства объектов с погрешностью не выше заданной.

Алгебраическое уравнение, уравнение получающееся при приравнении двух алгебраических выражений. Например, алгебраическим уравнением являются

$$\frac{x}{x+1} = 2, \quad x^2 + xy + y^2 - x + 2y + 1 = 0.$$

Алгебраическое уравнение с одним неизвестным может быть преобразовано к виду $a_0x^n + a_1x^{n-1} + \dots + a_n = 0$.

Алгоритм (алгорифм) (от *algorithmi*, *algorismus*, первоначально – латинская транслитерация имени математика аль-Хорезми) конечный набор правил, позволяющих чисто механически решать любую конкретную задачу из некоторого класса однотипных задач.

Алгоритмические модели, в них вычисление значений искомых величин связано с необходимостью решения систем уравнений.

Алгоритмический язык, набор символов и система правил составления и истолкования языковых конструкций из этих символов, предназначенные для записи алгоритмов.

Анализ (от греческого *analysis* – разложение, расчленение) – процедура разложения объекта (процесса, явления) на составные части.

Аналитические модели, в которых искомые переменные явно выражены через известные величины.

Аналитический аспект является математическим описанием процесса в виде некоторой системы уравнений, отражающей протекающие в объекте явления и функциональные связи между ними.

Аппроксимация (от латинского *approximo* – приближаюсь), замена одних математических объектов (например, чисел или функции) другими, более простыми и в том или ином смысле близкими к исходным (например, кривых линий близкими к ним ломаными).

Аспект (от латинского *aspectus* – вид), точка зрения с которой рассматривается какое-либо явление, понятие, перспектива.

Атрибут (от латинского *attribuo* – придаю, наделяю), 1) необходимое, существенное, неотъемлемое свойство объекта (например, атрибут материи в философии – движение). 2) В языкознании – то же, что определение. *Атрибутивным* (от латинского *attributio* – свойство, признак) называется высказывание о признаке предмета.

«**Блуждающая трубка**» как алгоритм имеет характер метода последовательных приближений.

Вариационное исчисление – раздел математики, посвященный нахождению наибольших и наименьших значений переменных величин, зависящих от выбора одной или нескольких функции (такие величины называются *функционалами*). К числу задач вариационных исчислений относятся, например, изопериметрические задачи.

«**Веником**» называется процедура отыскания минимума аддитивной функции, основанная на последовательном «отметании» неконкурентоспособных вариантов.

Верификация (проверка) модели – контроль правильности модели на основе сравнения результата с другими известными фактами, в частности, с экспериментальными данными, и т. д.

Вероятностная модель может включать случайные компоненты – случайные скалярные или векторные величины, случайные последовательности или функции, случайные структуры и т. п., удовлетворяющие статическим законом.

Вычислительный аспект есть метод и алгоритм решения системы уравнений математического описания, реализованные как моделирующая программа на одном из языков программирования.

Гистограмма (от греческого *histos*, здесь – столб и ...*грамма*) (столбчатая диаграмма), один из видов графического изображения статистических распределений какой-либо величины по количественному признаку. Гистограмма

представляет собой совокупность смежных прямоугольников, построенных на одной прямой; площадь каждого из них пропорциональна частоте нахождения данной величины в интервале, на котором построен данный прямоугольник.

Грина функция – функция, связанная с аналитическим представлением применяется также при решении краевых задач для обыкновенных дифференциальных уравнений математической физики. Грина функция названа по имени Джорджа Грина, впервые рассмотревшего один её частный случай в своём исследовании по теории потенциала (1828 г.). Грина функция описывает распространение полей от источников от порождающих (поэтому её называют также функцией распределения).

Декомпозиция, это деление объекта на составные части.

Детерминированная модель случайных компонентов не содержат.

Динамическая модель, предметом изучения является изменение рассматриваемого объекта во времени.

Динамическое программирование, это специфический класс задач с дискретным временем (в частности задачи с непрерывным временем).

Дифференциальное уравнение, уравнение, связывающее искомую функцию, её производные (или дифференциалы) и независимые переменные.

Дихотомия – (от греческого *dichotomia* – разделение надвое), способ классификации: классы, множества, понятия, термины и другие разбиваются на пары «соподчинённых» элементов (подклассов, подмножеств и другие).

Доминанта (от латинского *dominans*, родительный падеж *dominantis* – господствующий), господствующая идея, основной признак или важнейшая составная часть чего-либо.

Доминантный признак, иногда доминирование выражают величиной органической массы, создаваемой данными видами, или числом особей.

Достаточная простота модели, модель является достаточно простой, если имеющиеся в нашем распоряжении средства исследования дают возможность провести в приемлемые сроки и экономно по затратам труда и средств, но с разумной

точностью качественный или количественный – в зависимости от постановки задачи – анализ исследуемых свойств и осмыслить результат.

Евклидово пространство, пространство, свойства которого изучаются в евклидовой геометрии. В более широком смысле евклидовом пространством называется n -мерное векторное пространство, в котором определено скалярное произведение.

Идентификация (от позднелатинского *identifico* – отождествляю), в технике, математике установление соответствия распознаваемого предмета своему образцу (знаку), называемой идентификатором.

Изопериметрические задачи (от *изо...* и *периметр*), класс задач вариационных исчисления. Простейшие изопериметрические задачи: нахождение треугольников и многоугольников заданного периметра, имеющих наибольшую площадь, замкнутой кривой данной длины, ограничивающей наибольшую площадь.

Имитационное моделирование применяется для анализа поведения сложных экономических и т. п. задач, для которых математическую модель в виде системы уравнений даже выписать затруднительно.

Имитация (от латинского *imitatio* – подражание), подражание кому-либо или чему-либо, воспроизведение; подделка.

Инвариантность, неизменность какой-либо величины при изменении физической условий или по отношению к некоторым преобразованиям, обычно преобразованиям координат и времени при переходе от одной инерциальной системы отчёта к другой; т.е. не зависит от выбора системы координат.

Интерфейс (английский *interface*), система унифицированных связей (по виду передаваемой информации, параметрам сигналов, аппаратуре), предназначен для обмена информацией между устройствами вычислительной системы (например, между устройством ввода данных и запоминающим устройством). Применяется в ЭВМ 3-го поколения.

Информационный процесс – процесс сбора, хранения, передачи и обработки информации.

Итерационные методы (от латинского «итерацио» – повторение), в каждом из таких методов некоторая единообразная вычислительная процедура повторяется вновь и

вновь, причем каждый раз результат проведенного вычисления кладется в основу последующего вычисления.

Итерация (от латинского *iteratio* – повторение), повторное применение какой-либо математической операции.

Качественные методы, с помощью качественных методов свойства решения изучаются без его построения, путем анализа свойств заданного уравнения.

Качественная модель по необходимости применяется, где количественные закономерности проявляются не вполне четко (например, в некоторых социальных или биологических науках).

Квази...(от латинского *quasi* – якобы, как будто), часть сложных слов, соответствующая по значению словам «мнимый» (например, квазиучёный), «почти», «близко» (например, квазиоптика).

Квазиньютоновскими методами называются алгоритмы, которые объединяют достоинства метода наискорейшего спуска и метода Ньютона.

Квазистатическая модель, здесь принимается, что изменение объекта происходит столь медленно, что при рассмотрении ситуации в каждый момент можно объект считать статическим (грубо говоря, пренебречь инерционными силами), а время считать добавочным параметром.

«**Киевский веник**» как алгоритм дает возможность отыскать глобальный экстремум, причем для функций $f_i(x_i, x_{i+1})$ произвольного вида (например, не делалось никаких предположений об их выпуклости и прочие).

Классификация (от латинского *classis* – разряд, класс и ...фикация) представляет собой распределения объектов по группам (классам), причем каждый класс и типовые представители в каждом классе имеют свое, строго определенное место, которое определяется выбранным ОЛД (основание логического деления). Классификация (в логике), система соподчинённых понятий (классов объектов) какой-либо области знания или деятельности человека, используемая как средство для установления связей между этими понятиями или классами объектов.

Коммивояжер (французский *commis voyageur*), устаревший разъездной представитель торговой фирмы, предлагающий покупателям товары по имеющимся у него образцам.

Конгрев, сокращённое название рельефного (конгревного) тиснения на переплётных крышках, реже на плотной бумаге. Изображение получается с помощью штампа с углублённым рисунком и выпуклого контрштампа – матрицы, расположенной с обратной стороны крышки. Назван от имени английского изобретателя (конструктор, полковник) Уильяма Конгрева (*Kongreve*) (1772–1828).

Контекст – обстановка, сложившаяся вокруг конкретного явления.

Линеаризация (от латинского *linearis* – линейный), один из наиболее распространённых методов анализа нелинейных систем (или зависимостей), при котором они рассматриваются (с определёнными допущениями) как линейные.

Линейная модель, в ней выполнен принцип суперпозиции (наложения), т. е. при сложении входов складываются и выходы, а при умножении входа на любое число выход умножается на то же число.

Линейное программирование (*«linear programming»*), один из разделов математического программирования, в котором изучаемая функция линейна (1-й степени) и задана на множестве, заданном линейными равенствами и неравенствами.

Математическое моделирование, как термин часто применяется по отношению к области прикладной математики, включающей в себя как построение и исследование математических моделей, так и создание вычислительных алгоритмов и программ, реализующих эти алгоритмы на ЭВМ.

Математическая модель – приближенное описание какого-либо явления или процесса внешнего мира, выраженное с помощью математической символики.

Математическое программирование, математическая дисциплина, посвящённая теории и методам решения задач о нахождении экстремумов функций на множествах, определяемых некоторыми ограничениями (равенствами и неравенствами).

Матрица (немецкий *Matrize*, от латинского *matrix* – матка) (полиграфическая), 1) металлическая пластинка с углублённым прямым изображением буквы или знака, изготовленная штамповкой пуансоном или гравированием. Служит формой для отливки литер в шрифтолитейном и наборном производствах. 2) Лист пластиночного материала с углублённым изображением текста и рисунков, полученный прессованием с оригинальной формы высокой печати. Применяют при изготовлении стереотипов; хранится вместо набора (для целей переиздания, децентрализации печатания многотиражных изданий). 3) Рельефная копия штампа (контрштамп), используемая при конгревном тиснении. *Матрица* (математическая), прямоугольная таблица каких-либо элементов a_{ik} (чисел, математических выражений), состоящая из m строк и n столбцов:

$$\left\| \begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \dots & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{array} \right\| = \|a_{ik}\|.$$

Если $m = n$, то матрица называется квадратной. Над матрицей можно производить действия по правилам матричной алгебры. Матрицы используются во многих разделах математики и физики, в частности при исследовании систем m линейных уравнений с n неизвестными.

Метод дихотомии (половинного деления) – поиск минимума функции начинается с двух экспериментов и выбирается один.

Метод золотого сечения, здесь рассматривается симметричное расположение точек x_1 и x_2 на отрезке $[a, b]$, при котором одна из них становится пробной точкой на новом отрезке, полученном после исключения части исходного отрезка. Использование таких точек позволяет на каждой итерации этого метода, кроме первой, ограничиться определением только одного значения $f(x)$, так как другое значение уже найдено на одной из предыдущих итераций.

Метод наименьших квадратов, один из методов ошибок теории для оценки неизвестных величин по результатам измерений, содержащим случайные ошибки. Применяется при обработке наблюдений.

Метод Марквардта, в этом методе при переходе к очередной итерации новая точка рассчитывается.

Метод Ньютона (другое название – *метод касательных*) имеет локальную сходимость; итерационный процесс минимизации функции многих переменных, начатый из произвольной точки $x^0 \in E_n$ и является обобщением метода Ньютона в одномерном случае; интерпретируется как последовательный поиск точек минимума квадратичных аппроксимирующих функций, имеющих конечную точку минимума.

Метод парабол, использующий полиномы второго порядка, на каждой итерации этого метода строится квадратный трехчлен, график которого (парабола) проходит через три выбранные точки графика функции $f(x)$.

Метод перебора, здесь рассматриваемая область покрывается сеткой с шагом h , и определяются значения функции в ее узлах и, сравнивая полученные числа между собой, находят среди них наименьшее и принимается оно приближенно за наименьшее значение функции для всей области.

Метод порядного поиска, здесь перебор точек отрезка происходит сначала с шагом $\Delta = x_{i+1} - x_i > \varepsilon$ до тех пор, пока не выполнится условие $f(x_i) \leq f(x_{i+1})$ или пока очередная из этих точек не совпадает с концом отрезка. После этого шаг уменьшается (обычно четыре раза), и перебор точек с новым шагом производится в противоположном направлении до тех пор, пока значения $f(x)$ не перестанут уменьшаться или очередная точка не совпадает с концом отрезка и т. д. Процесс завершается, когда перебор в данном направлении закончен.

Метод средней точки – вычисление из двух значений одного в его средней точке $\bar{x} = \frac{a+b}{2}$.

Метод хорд, сущность: в исключении отрезков путем определения x – точки пересечения с осью Ox хорды графика

функции $F(x)$ на $[a, b]$, на каждой итерации, кроме первой, необходимо вычислять только одно новое значение $f'(x)$.

Метод Хука-Дживса – стратегия поиска, предложенная в 1961 г., согласно которой направление спуска завершающего k -й шаг поиска, определяется вектором $\bar{x}^k - x^{k-1}$.

Метод циклического покоординатного спуска, этот метод заключается в последовательной минимизации целевой функции $f(x)$ с исчерпывающим спуском сначала по направлению первого базисного вектора e^1 , затем второго – e^2 и т. д.

Методы оптимизации – курс, ориентированный на решение задач, возникающих в практической деятельности: в экономике, физике, инженерном деле и т. д.; это мост, ведущий от математики к анализу конкретных задач.

Методы случайного поиска, для которых основой служит итерационный процесс.

Минор (итальянский *minore*, от латинского *minor* – меньший; также *moll* от латинского *mollis* – мягкий) (муз.), лад, в основе которого лежит малое (минорное) трезвучие; определитель, составленный из элементов, стоящих на пересечении произвольно выделенных k строк и k столбцов данной матрицы или определителя.

Моделирование, исследование каких либо явлений, процессов или систем объектов путём построения и изучения их моделей; использование моделей для определения или уточнения характеристик и рационализации способов построения вновь конструируемых объектов. Моделирование – одна из основных категорий теории познания: на идее моделирования по существу базируется любой метод научного исследования – как теоретический (при котором используются различного рода знаковые, абстрактные модели), так и экспериментальный (использующий предметные модели); является одним из методов изучения окружающего мира.

Модель, представляет собой некоторое отображение объекта в форме, отличной от формы его реального существования, и служит средством, помогающим в объяснении или понимании происходящих процессов и при решении задач; в широком смысле модель – это любой образ, мысленный или предметный,

замещающий рассматриваемый объект при его изучении. Это может быть чертеж, конспект, график, план, выкройка, таблица, макет, шаблон и т. д.

Морфология (от греческого – форма + *logos* – слово, понятие, учение) – наука, изучающая форму и строение различных систем.

Навье-Стокса уравнения, дифференциальные уравнения движения вязкой жидкости или газа, выведены в 1822 году. Названы по имени А. Навье (французский инженер и учёный) и Дж. Стокса (английский физик и математик).

Нелинейная модель, в ней принцип суперпозиции не выполнен.

Оптимизация (математика) – нахождение экстремума (минимума или максимума) целевой функции в некоторой области конечномерного векторного пространства.

Политропа (от *поли...* и греческого *tropos* – поворот, направление), кривая на диаграммах состояния, изображающая политропный процесс. *Политропный процесс* (политропический процесс), термодинамический процесс, характеризующийся постоянной теплоёмкостью; *политропный газ*, для идеального газа описывается уравнением $pV^n = const$, где n – постоянная, называемая показателем политропы.

Постулаты модели, это то же, что гипотезы.

Продуктивность модели связана с тем, что изучаемый объект может включать различные параметры – такие, как массы, длины и т. п. его компонент, включать функциональные зависимости, которые считаются заданными и описывают связи между рассматриваемыми величинами (например, связь между усилием и перемещением в случае нелинейного закона упругости).

Прямые методы вариационного исчисления, с помощью их приближенное решение задачи на экстремум функционала сводится без обращения к уравнению Эйлера к аналогичной задаче с конечным числом степеней свободы.

Прямыми методами минимизации называются методы, использующие только значения функции и не требующие вычисления ее производных.

Ресурс – средство, к которому обращаются в случае необходимости.

Решение задачи – сложный и многогранный мыслительный процесс, в котором важными составляющими являются, во-первых, понимание задачи и, во-вторых, психологическая готовность к решению.

Сепаратриса (от латинского *separabu*) – траектория динамической системы с двумерным фазовым пространством, стремящаяся к седловому состоянию равновесия при времени $t \rightarrow \infty$ (устойчивая сепаратриса) или при $t \rightarrow -\infty$ (неустойчивая сепаратриса). Если устойчивая сепаратриса стремится к седлу при $t \rightarrow \pm \infty$, то её (вместе с седлом) называют петлей сепаратрисы.

Сильвестр (*Sylvester*) Джеймс Джозеф (1814–1897), английский математик; труды по алгебре, теории чисел, математической физике.

Симплекс (от латинского *simplex* – простой), простейший выпуклый многогранник данного числа измерений n . При $n = 3$ трёхмерный симплекс представляет собой произвольный тетраэдр. Под двухмерным симплексом понимают произвольный треугольник, а под одномерным – отрезок. Нульмерный симплекс есть просто точка.

Симплекс-метод, в основе которого лежит идея, упорядочного перебора вершин допустимого многогранника.

Синектика, слово «синектика» (*synectics*, иногда *synactic*) составлено из двух древнегреческих слов: *syn* – вместе и *actics* – действовать, совершать, поступать. Это словосочетание означает «совместно действовать», а применительно к методу решения задач имеет смысл «совместно искать», «вместе находить», «сообща придумывать», «вместе изобретать».

Синтез (от греческого *synthesis* – соединение, сочетание, составление) – метод научного исследования, состоящий в познании рассматриваемого объекта как единого целого, в единстве и взаимной связи его частей.

Система – совокупность взаимосвязанных компонентов (элементов), характеризующаяся постоянством определенных

связей и их устойчивостью. Термин *система* произошел от греческого *systema* – составленное из частей, соединение.

Смысловой аспект представляет собой физическое описание природы моделируемого объекта.

Содержательная модель, строение модели на языке той или иной науки.

Сплайн – это кусочно-полимиальная функция, в точках сопряжения кусков которой удовлетворяются условия непрерывности для нее самой и определенного числа производных.

Статическая модель, модель неизменная во времени

Стационарная модель, в этой модели считается, что процессы происходят, но изучаемый объект во времени не меняется.

Стохастическая модель, то же, вероятностная модель.

Стохастический (от греческого *stochastikos* – умеющий угадывать), случайный, вероятностный.

Стохастический процесс, то же, что случайный процесс.

Структурная модель, в которой отражается структура (устройство) моделируемого объекта.

Существенные параметры, т. е. при любом малом изменении параметров рассматриваемый объект фактически меняется.

Структура (от латинского *structura* – взаиморасположение, строение) – состав и взаимосвязь составных частей рассматриваемого объекта.

Существенные параметры, т. е. при любом малом изменении параметров рассматриваемый объект фактически меняется.

Точность модели, это количественная адекватность.

Унимодальность, поиск минимума функции одной переменной.

Унимодальные функции, это функции, имеющие на заданном интервале $[a, b]$ единственный минимум.

Фазовые модели – модели, в которых фигурируют фазовые переменные.

Факторная модель – модель вида $F(Q, Y) = 0$ в явной форме $Y = F(Q)$, где Q – вектор внешних параметров, Y – вектор выходных параметров объекта.

Фибоначчи (*Fibonacci*) Леонардо Пизанский (*Leonardo Pisano*) (1180–1240), итальянский математик, первым систематически

изложил достижения арабской математики, чем способствовал знакомству с ними в Западной Европе.

Функциональная модель, которая отражает только то, как объект функционирует – например, как он реагирует на внешние воздействия.

Функция (от латинского *functio* – исполнение, осуществление), деятельность, обязанность, работа; в математике зависимая переменная величина.

Характеристическое уравнение, алгебраическое уравнение степени n имеющее вид $\det(A - \lambda I) = 0$, где A – квадратная матрица, I – единичная матрица порядка n .

Целевая функция, в которой отыскивается наименьшее или наибольшее значение.

Целочисленное программирование, т. е. задач для которого аргументы целевой функции по своему смыслу могут принимать только целочисленные значения.

«**Черный ящик**», термин, употребляемый главным образом в системотехнике для обозначения систем, структура и внутренние процессы которых неизвестны или очень сложны; метод изучения таких систем основан на исследовании их реакций на известные (заданные) входные воздействия (сигналы).

Эволюционная модель, то же, что динамическая модель.

Эпистемология (от греческого *epistole* – письмо и *...графия*), то же, что *теория познания* (гносеология), раздел философии, в котором изучаются закономерности и возможности познания, отношения знания (ощущений, представлений, понятий) к объективной реальности, исследуются ступени и формы процесса познания, условия и критерии его достоверности и истинности.

Язык программирования, формальный язык для описания данных (информации) и алгоритма (программы) их обработки на ЦВМ.

Список сокращений

БМ – безусловная минимизация;
ДИСО – диалоговая система оптимизации;
НЛП – нелинейное программирование;
ОЛД – основание логического деления;
САПР – система автоматизированного проектирования;
ЦВМ – цифровая вычислительная машина.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Ревенков А. В., Резчикова Е. В.* Теория и практика решения технических задач : учеб. пособие– 2-е изд., испр. и доп. – М. : ФОРУМ, 2009. – 384 с.
2. *Моисеев Н. Н., Иванюков Ю. П., Столярова Е. М.* Методы оптимизации. – М., Главная редакция физико-математической литературы: Наука, 1978. – 352 с.
3. *Левин Ю. С.* Производственные процессы в полиграфии: проектирование и расчет. [Текст]: Учебник для вузов по специальности «Технология полиграфического производства» / *Ю. С. Левин, П. А. Матвеев, К. Д. Маудрих.* –М. –Лейпциг: Книга, 1985. – 320 с.
4. *Одинокова Е. В.* Проектирование полиграфических машин: Учебник для вузов / *Е. В. Одинокова, Г. Б. Куликов, И. Ш. Герценштейн;* Моск. гос. ун-т печати. – М.: МГУП, 2003. – 411 с. ISBN 5-8122-0307-5.
5. *Самарский А. А., Михайлов А. П.* Математическое моделирование: Идеи. Методы. Примеры. – 2-е изд., испр. – М.: Физматлит, 2001. – 320 с. – ISBN 5-9221-0120-X.
6. *Мышкис А. Д.* Элементы теории математических моделей. Изд. 3-е, исправленное. –М.: КомКнига, 2007. – 192 с.
7. *Гончаров В. А.* Методы оптимизации: учеб. пособие. – М.: Высшее образование, 2009. – 191 с. – (Основы наук).
8. *Васильев Ф. П.* Методы оптимизации. – М.: Факториал пресс, 2002. –824 с.
9. *Аттетков А. В., Галкин В.С., Зарубин В. С.* Методы оптимизации. – М. Изд-во МГТУ им. Н. Э. Баумана, 2003. – 440 с.

СОДЕРЖАНИЕ

Введение	3
1. Основные понятия и определения	
Теории решений задач оптимизации	5
2. Математическое моделирование основной метод	
решения задач оптимизации полиграфических	
процессов	8
2.1. Этапы математического моделирования.....	8
2.2. Физическое описание природы объекта	11
3. Математическая модель. Модель и конструкция	
печатных изданий	14
3.1. Математическая модель	14
3.2. Модель и конструкция печатных изданий	16
4. Классификация математических моделей	18
4.1. Различные подходы к классификации	18
4.2. Функциональные и структурные модели	18
4.3. Дискретные и непрерывные модели	20
4.4. Динамические и статические модели	21
4.5. Детерминированные и стохастические модели ...	22
4.6. Некоторые возможные математические модели...	23
4.7. Линейные и нелинейные модели	26
5. Задачи оптимизации. Основные определения	29
5.1. Задачи оптимизации	29
5.2. Минимум функции одной переменной	32
5.3. Унимодальные функции	34
5.4. Выпуклые функции	36
5.5. Условие Липшица	39
5.6. Классическая минимизация функции одной	
переменной	42
6. Линейное программирование	46
6.1. О постановках задачи линейного	
программирования и ее приложениях	46
6.2. Геометрическая интерпретация задач линейного	
программирования	53
6.3. Некоторые свойства задач линейного	
программирования	61
6.4. Симплекс-метод	67

6.5.	Двойственные задачи и методы	85
7.	Одномерная минимизация. Методы, использующие информацию о производных целевой функции ...	107
7.1.	Метод средней точки	107
7.2.	Метод хорд	109
7.3.	Метод Ньютона	111
7.4.	Возможные модификации метода Ньютона. Метод Ньютона-Рафсона. Метод Марквардта	118
7.5.	Методы минимизации многомодальных функций. Метод перебора	118
8.	Прямые методы безусловной минимизации многомерных задач	121
8.1.	Проблема минимизации многомерных задач	121
8.2.	Минимизация функций по правильному (регулярному) симплексу.....	124
8.3.	Минимизация функций при помощи нерегулярного симплекса	131
8.4.	Метод циклического покоординатного спуска ...	134
8.5.	Метод Хука-Дживса	137
8.6.	Методы случайного поиска	139
9.	Условный экстремум функции многих переменных. Необходимые и достаточные условия условного экстремума	142
9.1.	Условный экстремум при ограничениях типа равенств	142
9.2.	Условный экстремум при ограничениях типа неравенств	161
10.	Задача минимизации функции многих переменных. Необходимые и достаточные условия безусловного экстремума	175
10.1.	Постановка задачи и определения	175
10.2.	Свойства выпуклых множеств и выпуклых функций	180
10.3.	Необходимые и достаточные условия безусловного экстремума	183
11.	Оптимизация комплексного производственного процесса изготовления печатной продукции	194
11.1.	Технико-экономические модели	194
11.2.	Оптимизационные модели	204

11.3. Примеры постановки и решения оптимизационных задач	209
Словарь терминов	221
Список сокращений	234
Библиографический список	235
Содержание	236

Учебное издание

Жетписбай Егинбаевич Егинбаев

ТЕОРИЯ РЕШЕНИЙ ЗАДАЧ ОПТИМИЗАЦИИ

Учебное пособие

Нач. РО УИЦ

Редактор

Компьютерная верстка

З. А. Губайдулина

Г. М. Дюсенбаева

А. С. Сайбулатова

Подписано в печать 28.03.2016 г.

Тираж 300 экз. Формат 60х84 1/16. Бумага типографская № 1.

Объем 14.9 п. л. 13.9 Заказ № 688. Цена договорная

Издание Казахского национального исследовательского
технического университета имени К. И. Сатпаева

Учебно-издательский центр КазНТУ

г. Алматы, ул. Сатпаева, 22

ISBN 978-601-228-919-0



9

7 8 6 0 1 2 2 8 9 1 9 0